

UNIVERSIDAD DE GRANADA
E.T.S. Ingenieros de Caminos, Canales y Puertos
Departamento de Ingeniería Civil



PROGRAMA DE DOCTORADO:
INGENIERÍA CIVIL Y ARQUITECTURA

TESIS DOCTORAL

ANÁLISIS DE EFICIENCIA DE REDES NEURONALES EN EL ESTUDIO DE LA CALIDAD DEL SERVICIO EN EL TRANSPORTE PÚBLICO. APLICACIÓN AL ÁREA METROPOLITANA DE GRANADA.

Para la obtención del
GRADO DE DOCTOR POR LA UNIVERSIDAD DE GRANADA

AUTORA:
M^a CONCEPCIÓN GARRIDO RODRÍGUEZ

Granada, 2015

Editor: Universidad de Granada. Tesis Doctorales
Autor: María Concepción Garrido Rodríguez
ISBN: 978-84-1306-208-2
URI: <http://hdl.handle.net/10481/55743>

UNIVERSIDAD DE GRANADA
E.T.S. Ingenieros de Caminos, Canales y Puertos
Departamento de Ingeniería Civil



TESIS DOCTORAL

**ANÁLISIS DE EFICIENCIA DE REDES NEURONALES EN EL ESTUDIO DE LA
CALIDAD DEL SERVICIO EN EL TRANSPORTE PÚBLICO. APLICACIÓN AL ÁREA
METROPOLITANA DE GRANADA.**

AUTORA:

M^a CONCEPCIÓN GARRIDO RODRÍGUEZ

DIRECTOR:

JUAN DE OÑA LÓPEZ

Doctor Ingeniero de Caminos, Canales y Puertos
Universidad de Granada

AGRADECIMIENTOS

Mediante estas líneas me gustaría agradecer a todas aquellas personas que con su apoyo y colaboración me han animado y ayudado a elaborar esta tesis doctoral. Son muchos los amigos, compañeros de trabajo, compañeros de profesión y también jefes los que me han apoyado en este largo y tedioso camino, que no ha estado exento de dificultades, desánimos y contratiempos. Esta tesis doctoral es un trabajo no sólo de investigación, sino una labor de compañerismo, altruismo y colaboración. Gracias a todos.

Quiero agradecer a mi director de tesis, D. Juan de Oña López el tiempo invertido en esta investigación. Muchas gracias por guiarme y asesorarme.

A mis compañeros del Departamento de Ingeniería Civil, especialmente a Rocío y a José Luis. Gracias por vuestra ayuda.

Por supuesto, no habría sido posible culminar este trabajo sin la ayuda siempre incondicional de mis padres, que han estado a mi lado en todo momento y me han alentado en los momentos difíciles. Sois unos luchadores incansables, y el coraje, saber hacer y constancia con el que habéis desempeñado vuestra labor como padres me ha dado la fuerza necesaria para conseguir éste y muchos otros objetivos en la vida. Simplemente gracias por estar siempre ahí.

También imprescindible ha sido la ayuda de mi marido Gardier. Gracias por entender que esta tesis doctoral es importante para mí, y ayudarme y apoyarme siempre, y disculpa por los ratos que te he robado para dedicarlo a este trabajo en lugar de a ti. Gracias por ser como eres. Mientras escribo estas líneas, nuestro hijo crece en mis entrañas. Este trabajo y a partir de ahora todo lo que consiga en la vida irá dedicado a él.

No quiero olvidarme de darle las gracias a mi hermana, M^a Dolores, y a mi abuela Concha, os quiero y sois un pilar fundamental en mi vida. Abuela, a pesar de ser nonagenaria, eres una mujer moderna, de mente abierta, y siempre has tenido la convicción de que tanto mi hermana como yo llegaríamos hasta donde quisiéramos con nuestro trabajo y esfuerzo. Aunque a ti la vida no te ha permitido tener las mismas oportunidades de realizarte profesionalmente, siempre las has visto cumplidas a través de nosotras. Esta tesis también es logro tuyo.

Por último, quería dar las gracias al Consorcio de Transportes de Granada por facilitarnos los datos necesarios para desarrollar esta investigación.

RESUMEN

El éxito de un servicio de transporte público depende en gran medida del número de usuarios que es capaz de captar, y de la influencia que ejerce en la movilidad de una determinada ciudad o área.

Los planificadores, administraciones y empresas gestoras de un modo de transporte público deben encontrar la fórmula de atraer a nuevos pasajeros y crear fidelidad en los que ya son usuarios, para rentabilizar el servicio y además para mitigar los problemas que el uso masivo del vehículo privado generan actualmente en las ciudades: contaminación, atascos, salud, accidentes, etc. En los últimos años parece que existe un consenso generalizado en los expertos en la materia sobre la necesidad de satisfacer las demandas y las expectativas de la población si se quiere que el transporte público tenga un papel relevante en la sociedad actual, y es por ello que el concepto de calidad del servicio, íntimamente relacionado con la satisfacción y el grado de cumplimiento de dichas expectativas está siendo estudiado en profundidad y desde distintos ámbitos y puntos de vista.

La valoración que los usuarios actuales y/o potenciales tienen de un determinado servicio de transporte público es captada normalmente por medio de encuestas, en las que se obtiene información relacionada tanto con la valoración global de la calidad del servicio como de los distintos atributos que caracterizan al mismo. Partiendo de ellas, se han desarrollado numerosos estudios en los que se han aplicado diversos métodos para analizar tanto la valoración global de un determinado servicio de transporte público como la importancia que los usuarios asignan a cada uno de los atributos que lo definen. Todos ellos han aportado luz para mejorar y profundizar el conocimiento en este ámbito del transporte, pero cuentan con las limitaciones propias e inherentes de los propios métodos, de ahí que sea necesario seguir investigando en la materia y aplicando nuevas técnicas que puedan salvar las dificultades presentes en los métodos empleados hasta la actualidad.

En esta tesis doctoral se realiza un estudio de la calidad del servicio en el transporte público por autobús en el área metropolitana de Granada utilizando una técnica novedosa y no convencional, las Redes Neuronales Artificiales. Su elevada capacidad de predicción, clasificación y tolerancia a fallos la ha convertido en una herramienta válida y de uso generalizado en numerosos y muy variados campos de investigación. El principal objetivo de esta tesis doctoral es validar si las Redes Neuronales Artificiales son una técnica adecuada para el estudio de la calidad del servicio en el transporte público, y pueden convertirse en una herramienta válida para que los planificadores y gestores de un servicio público puedan extraer información interesante sobre los atributos que cualifican la calidad del servicio.

Los datos utilizados para el desarrollo de este estudio proceden de una encuesta de satisfacción realizada por el Consorcio de Transportes de Granada en el año 2007.

Además, en este trabajo de investigación se ha propuesto y desarrollado una nueva metodología para determinar la importancia relativa de los atributos de calidad del servicio considerados en este estudio.

Los resultados de esta investigación han mostrado que las Redes Neuronales Artificiales son una técnica válida para el estudio de la calidad del servicio en el transporte público por autobús, y que la metodología para determinar la importancia relativa de los atributos que la definen es asimismo válida y robusta, suponiendo esta nueva metodología una aportación novedosa y un valor añadido al campo de las Redes Neuronales Artificiales, que puede ser utilizada en cualquier estudio en el que se emplee esta técnica.

ABSTRACT

Nowadays the success of a public transport system depends largely on the amount of users that can attract and on the influence that it exerts on the mobility of a city or area.

Planners, administrations and operators of a public transport system need to find the formula to attract new customers and to build consumers loyalty, in order to make profitable the service, and to mitigate the problems that currently a massive use of the private vehicle generates in the cities: contaminations, traffic jams, health, accidents, etc. In the last few years, experts agree that population's requests and expectations need to be satisfied to achieve that public transportation plays an important role in the society. That is why the service quality concept, closely linked to the satisfaction and expectation's concepts, is been deeply studied from different points of view.

Normally, the information referred to a public transportation service's assessment is collected by means of the surveys carried out on current and/or potential customers, and information related both to the global valuation of the service quality and to the several specific attributes is gathered from them. Several studies, which involve different methods for analyzing both service quality and attributes' assessments, have been developed from surveys. All them have made substantial contributions to improve and to get more fully acquainted with this transportation field, but there are some inevitable limitations in each one of them, so that a further research on this issue needs to be carried out, and new techniques that overcome the current restrictions needs to be developed.

This PhD thesis studies the service quality of a metropolitan public bus transportation service in Granada by using a new and non-conventional technique, Artificial Neural Networks. Its high capability for prediction, classification a fault tolerance has converted it into a valid tool, and it has been used in an extremely wide range of areas. The main goal of this PdD thesis is to validate that Artificial Neural Networks is an adequate methodology for analyzing service quality in a public bus transportation system, and whether this methodology can become a potential tool for transport planners, and for extracting interesting information from the attributes that qualify the quality of the service.

The data used for the development of this study come from a customer satisfaction survey conducted by the Granada Consorcio de Transportes in 2007.

Furthermore, in this research work a new methodology for determining the relative importance of the service quality's attributes has been proposed and developed.

The results of this research work showed that Artificial Neural Network are a valid technique for studying the service quality in the public bus transportation system, and that the methodology for determining the relative importance of the quality service's attributes is likewise novel. It is an added value to the Artificial Neural Network field, that can be used in any research field where this technique is applied.

ÍNDICE GENERAL

1. INTRODUCCIÓN.....	14
1.1. Visión general del problema.....	15
1.2. Objetivos	16
1.3. Estructura de la tesis doctoral.....	17
1.4. Principales contribuciones.....	18
2. ESTADO DEL ARTE.....	20
2.1. Calidad del servicio en el transporte público	21
2.1.1. Concepto de calidad del servicio	21
2.1.2. Atributos.....	22
2.1.3. Encuestas.....	23
2.1.4. Métodos de estudio de la calidad del servicio en el transporte público... ..	24
2.2. Redes neuronales artificiales.....	29
2.2.1. Fundamento biológico de las redes neuronales artificiales	29
2.2.2. ¿Qué son las redes neuronales artificiales?	30
2.2.3. Clasificación de las redes neuronales artificiales.....	33
2.2.4. Métodos de cálculo de la importancia relativa de las variables consideradas en los modelos de redes neuronales artificiales.....	34
2.2.5. Aplicación de las redes neuronales artificiales a la calidad del servicio .	40
2.2.6. Ventajas y desventajas de las redes neuronales artificiales.....	42
2.3. Consideraciones finales.....	43
3. OBJETIVOS.....	46
3.1. Objetivo general.....	47
3.2. Objetivos específicos	47
4. CASO DE ESTUDIO.....	48
4.1. Descripción del sistema de transporte en estudio	49
4.1.1. Marco territorial y urbanístico	49
4.1.2. El consorcio de transportes metropolitano del Área de Granada (CMTAG)	52
4.1.3. Los servicios de autobús metropolitano.....	53
4.1.4. Evolución del sistema tarifario: la tarjeta de transporte.....	56
4.2. Datos.....	58

5. MATERIALES Y MÉTODOS.....	61
5.1. Métodos.....	62
5.2. Metodología seguida.....	72
6. RESULTADOS.....	75
6.1. Capacidad de generalización de las ANN entrenadas.....	76
6.2. Importancia relativa de las variables.....	78
6.3. Análisis estadístico.....	86
6.4. Análisis de la robustez de la metodología de cálculo de importancia relativa propuesta en esta investigación.....	89
6.5. Comparación entre las técnicas de ANN y DT aplicadas a la calidad del servicio.....	98
6.6. Discusión.....	101
7. CONCLUSIONES.....	104
7.1. Conclusiones.....	105
7.2. Futuras líneas de investigación.....	108
8. REFERENCIAS BIBLIOGRÁFICAS.....	109
APÉNDICE 1. FORMULARIO DE ENCUESTA.....	124
APÉNDICE 2. ROBUSTEZ Y ESTABILIDAD DEL MÉTODO DE CÁLCULO DE IMPORTANCIA RELATIVA DE LAS VARIABLES PROPUESTO.....	130
APÉNDICE 3. ARTÍCULOS PUBLICADOS.....	143

LISTA DE TABLAS

Tabla 1. Funciones de activación habituales	31
Tabla 2. Evolución de la población en Granada capital y en su Área Metropolitana	50
Tabla 3. Población de los municipios integrantes del Área Metropolitana en el año 2006.....	51
Tabla 4. Operadores y líneas de autobuses del transporte interurbano en el Área Metropolitana.....	55
Tabla 5. Evolución del uso de la tarjeta de transportes.....	57
Tabla 6. Características de la muestra encuestada.....	59
Tabla 7. Ratios de importancia y de percepción media de los atributos de calidad del servicio.....	60
Tabla 8. Rango de valores de MAPE para los conjuntos de ANN entrenados.....	78
Tabla 9. Importancia relativa y ranking de los atributos de calidad del servicio según los métodos PROFILE, PERTURB, CW y PaD.....	85
Tabla 10. Grupos homogéneos con valores medios (n=50) significativamente diferentes entre otros grupos.....	87
Tabla 11. Ranking de importancia relativa de las variables FRECUENCIA, VELOCIDAD e INFORMACIÓN por métodos.....	91
Tabla 12. Ranking de importancia relativa de las variables SEGURIDAD, PROXIMIDAD y PUNTUALIDAD por métodos.....	92
Tabla 13. Ranking de importancia relativa de las variables TRATO, TEMPERATURA y ESPACIO por métodos.....	93
Tabla 14. Ranking de importancia relativa de las variables PRECIO, LIMPIEZA y ACCESIBILIDAD por métodos.....	94
Tabla 15. Porcentajes máximo y mínimo de importancia relativa de los atributos por métodos.....	95
Tabla 16. Porcentajes máximo y mínimo de desviación típica de los valores de importancia relativa de los atributos por métodos.....	95

Tabla 17. Ratio de importancia relativa (%) y ranking de los atributos de calidad del servicio según la técnica DT.....98

Tabla 18. Comparación de las ventajas y desventajas de las técnicas de ANN y DT.....100

LISTA DE FIGURAS

Figura 1. Representación de neurona biológica.....	29
Figura 2. Paralelismo entre neurona biológica y neurona artificial.....	32
Figura 3. Estructura jerárquica de un sistema de ANN.....	33
Figura 4. Localización del Área Metropolitana de Granada.....	50
Figura 5. Evolución de viajeros en las líneas interurbanas del Área de Granada.....	53
Figura 6. Esquema de red de servicios del transporte público metropolitano.....	54
Figura 7. Líneas del sistema de transporte por autobús del Área de Granada.....	56
Figura 8. Arquitectura de una red neuronal artificial tipo MLP.....	63
Figura 9. Algoritmo de aprendizaje de descenso por gradiente.....	64
Figura 10. Evolución de los errores de entrenamiento y validación durante el entrenamiento de un MLP.....	66
Figura 11. Esquema de entrenamiento de DT y de cálculo de la importancia relativa de las variables de estudio.....	71
Figura 12. Arquitecturas de ANNs.....	77
Figura 13. Diagrama de cajas de importancia relativa de las variables predictoras según el método PERTURB.....	79
Figura 14. Diagrama de cajas de importancia relativa de las variables predictoras según el método PROFILE.....	79
Figura 15. Diagrama de cajas de importancia relativa de las variables predictoras según el método CW.....	80
Figura 16. Diagrama de cajas de importancia relativa de las variables predictoras según el método PaD.....	80
Figura 17. Importancia relativa de los atributos de calidad del servicio según el método PROFILE.....	81
Figura 18. Perfiles de variación de los atributos de calidad del servicio según el método PROFILE.....	82
Figura 19. Perfiles de variación del MSE de los atributos de calidad del servicio según el método PERTURB.....	83
Figura 20. Importancia relativa de los atributos de calidad del servicio según el método PERTURB.....	84

Figura 21. Importancia relativa de los atributos de calidad del servicio según el método CW.....	84
Figura 22. Importancia relativa de los atributos de calidad del servicio según el método PaD.....	85
Figura 23. Ranking de importancia relativa de los atributos de calidad del servicio según los métodos PROFILE, PERTURB, CW y PaD.....	86
Figura 24. Importancia relativa de la variable FRECUENCIA en función de H según el método aplicado.....	131
Figura 25. Desviación típica de la importancia relativa de la variable FRECUENCIA en función de H según el método aplicado.....	131
Figura 26. Importancia relativa de la variable VELOCIDAD en función de H según el método aplicado.....	132
Figura 27. Desviación típica de la importancia relativa de la variable VELOCIDAD en función de H según el método aplicado.....	132
Figura 28. Importancia relativa de la variable INFORMACIÓN en función de H según el método aplicado.....	133
Figura 29. Desviación típica de la importancia relativa de la variable INFORMACIÓN en función de H según el método aplicado.....	133
Figura 30. Importancia relativa de la variable SEGURIDAD en función de H según el método aplicado.....	134
Figura 31. Desviación típica de la importancia relativa de la variable SEGURIDAD en función de H según el método aplicado.....	134
Figura 32. Importancia relativa de la variable PROXIMIDAD en función de H según el método aplicado.....	135
Figura 33. Desviación típica de la importancia relativa de la variable PROXIMIDAD en función de H según el método aplicado.....	135
Figura 34. Importancia relativa de la variable PUNTUALIDAD en función de H según el método aplicado.....	136
Figura 35. Desviación típica de la importancia relativa de la variable PUNTUALIDAD en función de H según el método aplicado.....	136
Figura 36. Importancia relativa de la variable TRATO en función de H según el método aplicado.....	137

Figura 37. Desviación típica de la importancia relativa de la variable TRATO en función de H según el método aplicado.....	137
Figura 38. Importancia relativa de la variable TEMPERATURA en función de H según el método aplicado.....	138
Figura 39. Desviación típica de la importancia relativa de la variable TEMPERATURA en función de H según el método aplicado.....	138
Figura 40. Importancia relativa de la variable ESPACIO en función de H según el método aplicado.....	139
Figura 41. Desviación típica de la importancia relativa de la variable ESPACIO en función de H según el método aplicado.....	139
Figura 42. Importancia relativa de la variable PRECIO en función de H según el método aplicado.....	140
Figura 43. Desviación típica de la importancia relativa de la variable PRECIO en función de H según el método aplicado.....	140
Figura 44. Importancia relativa de la variable LIMPIEZA en función de H según el método aplicado.....	141
Figura 45. Desviación típica de la importancia relativa de la variable LIMPIEZA en función de H según el método aplicado.....	141
Figura 46. Importancia relativa de la variable ACCESIBILIDAD en función de H según el método aplicado.....	142
Figura 47. Desviación típica de la importancia relativa de la variable ACCESIBILIDAD en función de H según el método aplicado.....	142

CAPÍTULO 1. INTRODUCCIÓN

1.1. Visión general del problema

En la actualidad, uno de los objetivos fundamentales de las administraciones públicas es conseguir que los diferentes modos de transporte público sean de uso extendido por los ciudadanos. La realidad, sin embargo, muestra que los ciudadanos consideran el transporte público como una alternativa al vehículo privado, en lugar de desempeñar el rol predominante y generalizado en el movimiento de pasajeros (Simoes, 2013). Los desplazamientos en vehículo privado son percibidos en general como más cómodos, flexibles y rápidos, adecuándose mejor al estilo de vida imperante en nuestros días (Jakobsson Bergstad et al., 2011). Sin embargo, el uso excesivo del vehículo privado genera problemas medioambientales y sociales en las ciudades (contaminación, congestiones de tráfico, ruido, etc.) que en ocasiones convierte la movilidad de las ciudades en un sistema insostenible.

Los gestores de los servicios de transporte público tienen que demostrar a la población que pueden competir con los demás modos de transporte, garantizando efectividad y calidad en el servicio prestado. Las administraciones públicas, por su parte, están tendiendo a incentivar a los operadores de los servicios públicos en función de la calidad del servicio que éstos ofrecen (Mouwen y Rietveld, 2013). Puesto que los servicios de transporte público se prestan directamente a los pasajeros, la calidad del servicio debe ser resultado de la percepción que el usuario tiene de ella (Das y Pandit, 2013), porque como afirmó Bordagaray et al. (2013), “sin el consumidor, el mercado no tiene razón para existir”. Por tanto, el nivel de calidad de un servicio será elevado cuando cubra las necesidades y expectativas que los pasajeros han fijado.

En las últimas décadas, investigadores, planificadores y gestores han centrado su atención en este aspecto, intentando conocer con mayor detalle cómo evalúan los pasajeros un servicio público, teniendo en cuenta la influencia de los atributos que lo caracterizan. Varios autores han puesto de manifiesto que la calidad del servicio es un concepto complejo, difuso y abstracto (Carman, 1990; Parasuraman et al., 1985), debido fundamentalmente a las tres propiedades que lo caracterizan: intangibilidad, heterogeneidad e inseparabilidad, pero también por la subjetividad inherente en la medición de la opinión de los pasajeros sobre la calidad del servicio. En la literatura, existen diferentes métodos para determinar esta influencia, pero no existe consenso sobre cuál es el mejor, por lo que la medición de la calidad del servicio sigue siendo un reto para investigadores y planificadores del transporte.

La influencia de las características que definen la calidad del servicio sobre la evaluación global del mismo han sido estudiadas en los últimos años a través de distintas técnicas: modelos de ecuaciones estructurales (de Oña et al., 2013; Eboli y Mazzulla, 2007; 2012; Irfan et al., 2011; Ngatia et al., 2010) y modelos de elección discreta (Bordagaray et al. 2013; Eboli y Mazzulla, 2008;2010; Hensher, 2014; Hensher y Prioni, 2002; Hensher et al.

2003; Marcucci y Gatta, 2007) fundamentalmente. Sin embargo, la mayoría de estos modelos parten de premisas y de limitaciones en cuanto a las relaciones subyacentes entre las variables dependientes e independientes, como la normalidad de los datos, las relaciones de dependencia entre variables, baja multi-colinealidad, etc. Según Garver (2003), estos requisitos casi nunca se cumplen en el campo de satisfacción del cliente.

Kikuchi (2012) afirma que “los paradigmas tradicionales de predicción, diagnóstico, regulación u optimización, no son suficientes para tratar los sistemas humanos y sociales extremadamente complejos, de los que la ingeniería del transporte forma parte”. Siguiendo esta idea, de Oña et al. (2012) aplicaron exitosamente árboles de decisión al estudio de la calidad del servicio en el transporte público, ya que esta técnica no sufre algunas de las limitaciones presentes en las usadas tradicionalmente. Así, en esta tesis doctoral, se plantea el estudio de la calidad del servicio en el transporte público con una técnica de minería de datos, las Redes Neuronales Artificiales, que al igual que los árboles de decisión, supera varias de las debilidades que presentan los métodos que hasta ahora se han utilizado para su análisis.

Las Redes Neuronales Artificiales son sistemas de procesamiento de la información que han sido aplicados en otros campos de la calidad del servicio ajenos al de los transportes con resultados muy satisfactorios (Lee et al., 2010; Xie et al., 2003), debido a su elevada capacidad de predicción y tolerancia a fallos, sin necesidad de disponer de un modelo de partida con limitaciones inherentes al mismo o requisitos previos, y sin verse afectado por las posibles relaciones de dependencia entre las variables consideradas para el desarrollo de los modelos. Sin embargo, presentan una importante limitación debido a su incapacidad para determinar de forma clara y explícita la influencia que cada una de las variables dependientes del modelo ejerce sobre las dependientes, por lo que las Redes Neuronales Artificiales se han incluido durante mucho tiempo dentro del conjunto de técnicas denominadas “cajas negras” (Cao y Qiao, 2008; Cortez y Embrechts, 2013). Existen diversos métodos de cálculo de la influencia de las variables predictoras respecto a las dependientes, pero éstos presentan problemas de inestabilidad en los resultados y no se ha llegado a alcanzar un consenso sobre cuál de ellos presenta un mejor comportamiento global.

Por tanto, con este estudio se pretende determinar si la técnica de las Redes Neuronales Artificiales es válida para analizar con eficacia la calidad del servicio de transporte público por autobús e identificar los factores claves que la afectan.

1.2. Objetivos

El principal objetivo de esta tesis doctoral es validar si la técnica de las Redes Neuronales Artificiales es una metodología apropiada para analizar la calidad del servicio en el transporte público.

Las Redes Neuronales Artificiales son una técnica de minería de datos novedosa para analizar la calidad del servicio en el campo del transporte público, que puede resultar muy útil para los planificadores y gestores del servicio, debido a su elevada capacidad de predicción y de captar las relaciones altamente no lineales subyacentes en los atributos considerados en los modelos.

Para completar este objetivo principal, en este trabajo de investigación también se proponen una serie de objetivos específicos, tales como el desarrollo de una nueva metodología para determinar la importancia relativa de las variables predictoras consideradas en los modelos de Redes Neuronales Artificiales entrenados, y que mitigue la inestabilidad que actualmente presentan los modelos de cálculo existentes. Esta nueva metodología podría aplicarse, no sólo a la calidad del servicio en el transporte, sino a cualquier otro ámbito de estudio en el que se utilicen Redes Neuronales Artificiales. En este estudio se analiza la robustez de esta metodología, para comprobar si puede ser aplicada de forma genérica, de modo que se consiga agilizar el tiempo empleado aplicando esta técnica hasta obtener los resultados adecuados, sin que éstos se vean afectados por esta mejora. Además, se realiza una comparación entre las Redes Neuronales Artificiales y los árboles de decisión en este caso de estudio, para poder determinar las fortalezas y debilidades que ofrecen cada una de estas técnicas.

Con la verificación de estos objetivos, se proporciona a los planificadores, gestores y operadores de transporte público nueva información sobre la calidad del servicio, que servirá a su vez para diseñar líneas de acción enfocadas a la mejora del servicio prestado, y por ende a un uso más extendido del transporte público.

1.3. Estructura de la tesis doctoral

En este apartado se realiza una breve descripción de la estructura de esta tesis doctoral, que está compuesta por ocho capítulos y un apéndice:

Capítulo 1, incluye una introducción de la tesis, una breve descripción de los objetivos propuestos, de la estructura del documento y de las principales contribuciones de esta investigación.

Capítulo 2, presenta una visión general sobre las características que definen la calidad del servicio en el transporte público, y una recopilación de las principales técnicas aplicadas hasta el momento para su análisis. También se realiza una descripción de la técnica de las Redes Neuronales Artificiales, así como una recopilación de los diferentes métodos de cálculo de importancia relativa de las variables que se han desarrollado y aplicado en diversos estudios de la literatura científica que han usado esta técnica. Además, se han descrito brevemente las principales aplicaciones de las Redes Neuronales Artificiales al campo de la calidad del servicio, sus principales ventajas e inconvenientes.

Capítulo 3, presenta los objetivos a verificar en este trabajo de investigación.

Capítulo 4, describe las principales características del sistema de transporte que se ha tomado como base en esta investigación, así como el contexto experimental que da lugar a la base de datos de este estudio.

Capítulo 5, incluye los materiales y métodos usados en esta tesis doctoral. Se describe la tipología de red neuronal artificial usada en este trabajo de investigación, y la técnica utilizada para validar y evaluar los modelos y los métodos de cálculo de la importancia relativa de las variables que definen la calidad del servicio que han sido aplicados. También se describe la técnica de los árboles de decisión, que será comparada con la de redes neuronales artificiales. En este capítulo además se presenta la metodología desarrollada en este trabajo de investigación.

Capítulo 6, muestra los resultados obtenidos tras aplicar la técnica de Redes Neuronales Artificiales a la calidad del servicio, así como aquéllos que indican el grado de validez y de robustez de la nueva metodología de cálculo de la importancia relativa de las variables que se plantea en esta investigación. También se indican los resultados extraídos tras aplicar una técnica estadística para determinar la existencia de grupos homogéneos de variables con diferencias estadísticamente significativas, y aquéllos obtenidos tras aplicar la técnica de árboles de decisión a la base de datos de este estudio.

Capítulo 7, presenta las principales conclusiones de este trabajo y las líneas futuras de investigación.

Capítulo 8, incluye todas las referencias usadas en esta tesis.

Finalmente, este documento incluye varios apéndices:

Apéndice 1, reproduce la encuesta que se elaboró para los trabajos de campo en los que se recabó la información que posteriormente se ha utilizado en esta investigación.

Apéndice 2, incluye los datos obtenidos tras analizar la robustez y la estabilidad de la metodología novedosa desarrollada en esta investigación por medio de gráficos y tablas.

Apéndice 3, en el que aparecen los artículos relacionados con este trabajo de investigación y publicados en revistas científicas.

1.4. Principales contribuciones

Las principales contribuciones de esta tesis son tres artículos científicos, que han sido publicados o están en revisión, y un poster presentado en un congreso internacional de transportes. Ellos incluyen las principales aportaciones obtenidas de este trabajo de investigación. Las contribuciones mencionadas son las siguientes:

Artículo 1. Juan de Oña y Concepción Garrido. Extracting the contribution of independent variables in neural network models: a new approach to handle instability. Publicado en Neural Computing and Applications, 24(5). DOI: 10.1007/s00521-014-1573-5.

Artículo 2. Concepción Garrido, Rocío de Oña y Juan de Oña. Neural Networks for analyzing service quality in public transportation. Publicado en Expert Systems with Applications, 41, 6830-6838.

Artículo 3. Juan de Oña, Rocío de Oña y Concepción Garrido. How to extract the attributes' importance from satisfaction surveys with data mining techniques: A comparison between Neural Networks and Decision Trees. En revisión en Transportation Letters.

Poster. Juan de Oña, Rocío de Oña y Concepción Garrido. Extracting importance of attributes from customer satisfaction surveys with data mining: decision trees and neural networks. Transportation Research Board 94th Annual Meeting. January 11-15 2015. Washington D.C.

CAPÍTULO 2. ESTADO DEL ARTE

CAPÍTULO 2. ESTADO DEL ARTE

Este capítulo incluye una revisión del estado actual del conocimiento que concierne al análisis de la calidad del servicio en el transporte público y de las principales metodologías que hasta el momento se han desarrollado para su estudio. Asimismo, se realiza una breve descripción de la técnica de redes neuronales artificiales, su evolución y el desarrollo de sus diferentes tipologías a lo largo de las últimas décadas, así como una revisión de los métodos de cálculo de importancia relativa de las variables analizadas utilizando las redes neuronales artificiales. Finalmente, concluye con un breve recorrido por la literatura científica en el que se hace referencia a los estudios relacionados con la calidad del servicio en los que se ha aplicado la técnica de las redes neuronales artificiales, para los cuales ha demostrado ser una herramienta potente y versátil, proporcionando buenos resultados, por lo que podría ser aplicada al análisis de la calidad del servicio en el transporte público en el futuro.

2.1. Calidad del servicio en el transporte público

Este apartado se basa en el estado del arte desarrollado en la tesis doctoral “Análisis de la calidad del servicio del transporte público mediante árboles de decisión” (De Oña, 2013), que posteriormente ha dado lugar a la publicación “Quality of service in public transport based on customer satisfaction surveys: a review and assessment of methodological approaches” (De Oña y De Oña, 2015).

2.1.1. Concepto de calidad del servicio

El concepto de calidad del servicio en el transporte público es complejo y difuso, y además el estudio del mismo ha sufrido una transformación en las últimas décadas, en las que se ha pasado de centrarse en él desde el punto de vista de la eficacia de su gestión por parte de las administraciones y las compañías encargadas de la prestación del servicio (Carter y Lomax, 1992; Ozment y Morash, 1998; Pullen, 1993; Wipper, 1993), a preocuparse más por el análisis de la percepción que tienen de él los usuarios, y a profundizar en el conocimiento de aquellos aspectos que lo caracterizan y que más influyen en la calidad del servicio percibida por los pasajeros.

La calidad del servicio es un concepto intangible, ya que no puede medirse a través de atributos físicos, sino a través de atributos que expresan las experiencias de los usuarios; heterogéneo, ya que cada individuo lo percibe de forma diferente; e inseparable en cuanto a producción y consumo, puesto que el usuario experimenta el servicio al mismo tiempo en que éste es producido.

Por otra parte, no existe consenso sobre la relación entre el concepto de calidad del servicio y los conceptos de percepción y satisfacción (De Oña, 2013; De Oña y De Oña, 2015). Mientras unos autores consideran que la calidad del servicio es el resultado de una comparación entre las expectativas del usuario y su percepción del servicio prestado (Grönroos, 1988, Lehtinen y Lehtinen, 1982; Parasuraman et al., 1985; Sasser et al., 1978), otros consideran que este concepto debe centrarse solamente en la percepción de los pasajeros (Cronin y Taylor, 1992) o incluso en la percepción de los gestores del servicio de transportes (Eboli y Mazzulla, 2011; 2012; Nathanail, 2008; Tyrinopoulos y Aifadopoulou, 2008). Además, también existen dos corrientes principales que tienen puntos de vista diferentes sobre las relaciones causa-efecto entre la calidad del servicio y la satisfacción del usuario: aquella que considera que la satisfacción del usuario define la calidad percibida, frente a la que mantiene que la calidad del servicio es el vehículo que permite definir el grado de satisfacción del usuario (Chen, 2008; Chou y Kim, 2009; Chou et al., 2011b; Croning y Taylor, 1992).

La calidad del servicio es también un concepto multidimensional (Brady y Cronin Jr, 2011). Según Grönroos (1982, 1984; 1988) y Lehtinen y Lehtinen (1991) es tridimensional, porque cuenta con una componente técnica (la calidad que un usuario objetivamente recibe), una componente funcional (la calidad que el usuario percibe respecto a la componente técnica recibida) y una componente de imagen (la calidad atribuida por los gestores del servicio y “vendida” a los usuarios actuales y potenciales). De estas tres componentes, la funcional es la que está presente en mayor medida en los atributos que definen la calidad del servicio (Parasuraman et al., 1985, 1988).

2.1.2. Atributos

La calidad del servicio queda caracterizada por medio de atributos. Los diversos autores que han estudiado este aspecto del transporte, han definido los servicios a través de un número muy elevado de variables o atributos, que a su vez han sido agrupados en grupos más pequeños en función de aspectos o elementos en común, denominados dimensiones (de Oña y de Oña, 2015). Aunque hay atributos prácticamente comunes en todos los estudios de calidad del servicio en el transporte, existen otros específicos y particulares para cada caso de estudio. Por eso, la idea inicial defendida por Parasuraman et al. (1988) de la creación de una lista genérica de atributos y dimensiones para analizar la calidad del servicio ha sido cuestionada por numerosos autores (Babakus y Boller, 1992; Carman, 1990), que consideran necesario incluir en cada estudio atributos específicos para cada caso concreto, tales como aspectos socioeconómicos, demográficos, geográficos o motivo del viaje (Andreassen, 1995), u otras particularidades cualesquiera que presente la prestación de un determinado servicio. De hecho, Carrillat et al. (2007) demostraron que los estudios en los que se particularizan los atributos que definen la calidad de un servicio concreto arrojan mejores resultados que aquéllos que trabajan únicamente con los genéricos. Debido al elevado número de atributos que pueden caracterizar un determinado servicio de transporte público. Murray et al. (2010) definieron

166 y la UNE-EN 13816 (2003) considera 117), se han adoptado diversos criterios para seleccionar un número más reducido (de Oña y de Oña, 2014), permitiendo así acotar los estudios y hacerlos factibles. En todos ellos, se realiza un análisis detallado de los atributos más influyentes en cada caso, para lo que los investigadores se basan en diversos criterios: revisión bibliográfica de los atributos utilizados en otros estudios similares, encuestas de los operadores, encuestas a los usuarios, o test estadísticos. También puede utilizarse una combinación de varias de estas opciones (Chau y Kao, 2009; Dell’Olio et al., 2010). Según de Oña y de Oña (2014), las encuestas ad hoc son el método más apropiado para caracterizar la calidad del servicio.

Existen diversas clasificaciones de categorías de atributos, en función del mayor o menor impacto que tienen en la calidad del servicio y en la satisfacción del usuario. Philip y Hazlett (1997) propusieron un modelo con una estructura jerárquica basada en tres clases de atributos: pivotaes, centrales y periféricos. Los atributos pivotaes ejercen la mayor influencia en el nivel de satisfacción de los usuarios, los centrales representan la conjunción de los aspectos de organización, funcionamiento e interacción con el pasajero que permiten interactuar o recibir los atributos pivotaes, y los periféricos son atributos adicionales que permiten que la calidad del servicio percibido pueda llegar a su máximo nivel. La UNE-EN 13186 (2003) clasifica las características de un servicio de transporte público en básicas, proporcionales y atractivas, dependiendo de cómo afecta el grado de cumplimiento de estos atributos a la satisfacción del cliente: los atributos básicos se consideran implícitos en la prestación del servicio de transporte público (puntualidad, frecuencia, seguridad, etc.), y el incumplimiento o baja calidad de alguno de ellos afecta enormemente a la satisfacción del usuario; los proporcionales normalmente están presentes en el servicio y aportan calidad al mismo (limpieza, confort, etc.) y los atractivos son atributos extra que aportan un plus diferenciador de calidad a ese servicio.

El Transportation Research Board (2004) agrupa los atributos en factores de disponibilidad y en factores de confort o de conveniencia. Las deficiencias de calidad de alguno de los pertenecientes al primer grupo afecta en gran medida a la satisfacción del usuario, de modo que la calidad del servicio percibida por éste ya no cambia, independientemente de las características que presenten los demás atributos; en el caso del segundo grupo, las características de los atributos incluidos en él afectan a la percepción de la calidad del servicio pero no son vinculantes para su valoración global. Eboli y Mazzulla (2008) demostraron empíricamente la existencia de dos categorías de atributos (básicos y no básicos) en las preferencias mostradas por los usuarios: los básicos son fundamentales para la valoración de la calidad del servicio, mientras que los no básicos quedan en una posición de importancia secundaria.

2.1.3. Encuestas

Las encuestas son la herramienta más utilizada para recabar la información necesaria sobre la calidad del servicio, que posteriormente será la base sobre la que se aplicarán los diferentes métodos de estudio. En estos cuestionarios, normalmente se pide a los

usuarios que puntúen la calidad global del servicio y/o cada uno de los atributos que lo caracterizan. En estas encuestas también se puede pedir al usuario que puntúe el servicio global y sus atributos característicos desde dos puntos de vista diferentes: la percepción y las expectativas que éste servicio había generado en ellos. De Oña y de Oña (2015) elaboraron un detallado estudio en el que realizaron una revisión profunda sobre las tipologías de encuestas utilizadas para evaluar posteriormente la calidad del servicio, así como los criterios de valoración indicados a los usuarios en ellas, pudiendo ser éstos numéricos (asignación de una puntuación dentro de un rango de valores) o lingüísticos.

También es frecuente incluir en la encuesta preguntas sobre aspectos socioeconómicos y demográficos (por ej. edad, sexo, disponibilidad de vehículo, frecuencia de uso del servicio de transporte público, etc.) (Andreassen, 1995; Dell’Ollio et al., 2010), para tener en cuenta en las fases posteriores de estudio de la información recopilada la gran heterogeneidad que existe entre los encuestados; es decir, existe una diversidad importante en aspectos como los gustos del cliente, su nivel socioeconómico, sus gustos personales y preferencias respecto al uso del transporte público, etc. Resulta adecuado incluir en las encuestas información relacionada sobre estos aspectos, para poder realizar un estudio desagregado de los distintos grupos de usuarios, que proporcione una información más detallada.

2.1.4. Métodos de estudio de la calidad del servicio en el transporte público

Según de Oña y de Oña (2015), existen dos corrientes de estudio de la calidad del servicio en el transporte público: los métodos que tienen en cuenta la percepción y las expectativas de los usuarios, y aquéllos que sólo tienen en cuenta la percepción de los mismos respecto a la calidad del servicio recibido. Del mismo modo, también establecen otra clasificación, basada en el grado de detalle con el que se analiza la calidad del servicio, distinguiendo así entre métodos agregados y desagregados: los primeros tienen como objetivo obtener una valoración o índice global de la calidad del servicio, mientras que los segundos analizan individualmente cada uno de los atributos que la caracterizan. Cada uno de ellos presenta ventajas e inconvenientes, esto es, mientras que los modelos desagregados permiten obtener un conocimiento más detallado de aspectos concretos de la calidad del servicio, aportando información valiosa a los planificadores y gestores del transporte sobre los elementos en los que debe incidirse para mejorar la calidad del servicio público prestado (Weinstein, 2000), los agregados proporcionan valores globales de la valoración de esta calidad del servicio, que resultan muy útiles para realizar un seguimiento de su evolución a lo largo del tiempo y para comparar diferentes servicios.

Por otra parte, también se han utilizado técnicas basadas en métodos matemáticos para analizar la calidad del servicio en el transporte público. Muchos de los estos estudios están destinados a verificar la información proporcionada por las encuestas (de Oña, 2013), de modo que a lo largo de la literatura científica se han venido aplicando métodos estadísticos clásicos, como test-t, ANOVA, MANOVA, etc., para confirmar las hipótesis (Drea y Hanna, 2000; Ganesan-Lim et al., 2008; Oyewole, 2001; Pedersen et al., 2011),

modelos de ecuaciones estructurales (SEM) (Andreassen, 1995; Chen, 2008; Cheng et al., 2008; Chou y Kim, 2009; Chou et al., 2011b; Friman y Gärling, 2001; Friman et al., 2001; Jen et al., 2011; Joewono y Kubota, 2007a; Joewono y Kubota, 2007b; Joewono y Kubota, 2007c, Kim y Lee, 2011; Lai y Chen, 2011; Minser y Webb, 2010; Park et al., 2006; Saha y Theingi, 2009; Stuart et al., 2010; Jen y Hu, 2003; Tripp y Drea, 2002; Wen et al., 2005, Yang et al., 2012), o análisis path (Forgas et al., 2010; Jen y Hu, 2003; Park et al., 2004; Ringle et al., 2011). Recientemente se han aplicado árboles de decisión (de Oña et al., 2012).

También son numerosos los estudios destinados a analizar las diferencias de percepción de calidad del servicio de los usuarios en función de la heterogeneidad de los grupos (características socioeconómicas, de población, o antes y después de llevar a cabo una determinada acción, como por ejemplo cuestionarios). Aunque para estos estudios se han aplicado de forma generalizada métodos estadísticos y modelos de ecuaciones estructurales, también se han aplicado modelos de elección discreta (por ejemplo, Dell'Ollio et al., 2010; Hensher et al., 2010; Huse y Evangelho, 2007).

2.1.4.1. Métodos agregados basados en la percepción y las expectativas

El método más representativo y empleado en el estudio de la calidad del servicio de transporte público que considera tanto la percepción como las expectativas de los usuarios fue el desarrollado por Parasuraman et al. (1985). En él se define un índice global de valoración de la calidad del servicio (SQ) que es función de la diferencia entre la percepción de calidad P_{ij} que tiene un usuario sobre un atributo j , y las expectativas E_{ij} de ese mismo usuario sobre ese mismo atributo. Este índice se expresa mediante la ecuación:

$$SQ = \sum_{j=1}^k (P_{ij} - E_{ij}) \quad (1)$$

Para medir la calidad del servicio global, Parasuraman et al. (1988) desarrollaron una escala, denominada SERVQUAL, que fue usada por varios autores en la evaluación de la calidad del servicio durante años (Abdlla et al., 2007; Kiatcharoenpol y Laosirihongthong, 2006; Jun y He, 2007). Los principales problemas achacados a este método residen en el hecho de que éste asigna igual importancia a todos los atributos que caracterizan la calidad del servicio considerados en el modelo, lo cual no se ajusta a la realidad. Para solventar este inconveniente, diversos autores introdujeron como modificación al método unos pesos que marcaran diferencias en cuanto a la importancia relativa de los atributos (Pakdil y Aydın, 2007), y también la técnica de la lógica difusa (Chou et al., 2011a). Hill et al. (2003) presentaron una nueva ecuación para medir la calidad del servicio considerando importancias diferentes para los distintos atributos, en la que se calcula un índice CSI:

$$CSI = \sum_{k=1}^N [\bar{S}_k \cdot W_k] \quad (2)$$

siendo \bar{S}_k la media de los ratios de satisfacción expresados por los usuarios para un atributo k y W_k el peso del atributo k calculado como el ratio entre la importancia media expresada por los usuarios para el atributo k y la suma de los ratios de importancia media asignados a los demás atributos considerados en el estudio. Aunque este método considera pesos diferentes para cada atributo, presenta el inconveniente de que no puede medir la heterogeneidad que pueden presentar los ratios de ciertos atributos (de Oña y de Oña, 2015). Eboli y Mazzulla (2009) introdujeron un nuevo índice HCSI para tener en cuenta la heterogeneidad de los atributos:

$$HCSI = \sum_{k=1}^N S_k^c \cdot W_k^c \quad (3)$$

$$S_k^c = \bar{S}_k \cdot \frac{\bar{S}_k / \text{var}(S_k)}{\sum_{k=1}^N (\bar{S}_k / \text{var}(S_k))} \cdot N \quad (4)$$

$$W_k^c = \frac{\bar{I}_k / \text{var}(I_k)}{\sum_{k=1}^N (\bar{I}_k / \text{var}(I_k))} \quad (5)$$

donde S_k^c es la media de los ratios de satisfacción expresados por los usuarios para un atributo k corregidos teniendo en cuenta la desviación respecto al valor medio de los valores registrados y W_k^c tiene el mismo significado que W_k en la fórmula (2), pero está corregido teniendo en cuenta la dispersión I de los ratios del atributo k.

Por otra parte, varios autores han incluido en estos métodos análisis multicriterio, incluso combinado con lógica difusa (Kuo et al., 2007; Chang y Yeh, 2002; Liou y Tzeng, 2007; Tsaur et al., 2002; Nejati et al., 2009).

2.1.4.2. Métodos agregados basados sólo en la percepción

Entre los métodos agregados que se basan sólo en la percepción de la calidad del servicio, el más conocido de ellos fue el desarrollado por Cronin y Taylor (1992), que determinaron la calidad del servicio a través de un índice SQ que sigue la expresión:

$$SQ = \sum_{j=1}^k P_{ij} \quad (6)$$

Donde P_{ij} es la percepción del estímulo i frente al atributo j. Al igual que ocurría con el método de Parasuraman et al. (1985), éste no tiene en cuenta pesos o importancias diferentes para cada atributo, por lo que, como en el mencionado método, diversos autores (Yeh et al., 2010; Awasthi et al., 2011; Kuo, 2011; Nathaniel, 2008) han aplicado variaciones al mismo incluyendo análisis multicriterio y lógica difusa para considerar este aspecto.

2.1.4.3. Métodos desagregados basados sólo en la percepción

Los métodos desagregados basados sólo en la determinación de la percepción de los usuarios han utilizado mayoritariamente cuadrantes para analizar la calidad del servicio. El más conocido es el análisis IPA (importance-performance analysis), que representa en coordenadas la importancia asignada por el usuario a cada atributo (eje y) versus el valor medio del ratio con el que el usuario caracteriza a ese atributo (eje x). Debido a su simplicidad, ha sido utilizado en numerosos estudios para evaluar la calidad del servicio de diversos modos de transporte (Weinstein, 2000; Chou et al., 2011b; Chen y Chang, 2005, Christopher et al., 1999; Figler et al., 2011; Foote and Stuart, 1998). Este método es muy fácil de interpretar visualmente y proporciona una valoración rápida y sencilla de los atributos que caracterizan un determinado servicio, haciendo posible incidir en aquellos aspectos que hayan presentado una valoración baja (Slack, 1994), pero los resultados que ofrecen son una guía que indica una jerarquía de importancia de los atributos según el usuario, no siendo posible cuantificar con exactitud los valores de esa importancia (Abalo et al., 2007). Recientemente, Eboli y Mazzulla (2011) han desarrollado un método que tiene en cuenta tanto la percepción de los pasajeros como la medida del rendimiento proporcionada por la entidad gestora del servicio. Según este método, a cada atributo le son asignados dos parámetros, uno subjetivo (S) resultado del valor medio de satisfacción de los usuarios, y uno objetivo (O), calculado como la media de las puntuaciones que los operadores del servicio le han asignado. Con estos dos parámetros, se obtiene un indicador combinado (X), y en función del valor de la varianza de este último indicador se puede conocer cuánta componente subjetiva u objetiva debe considerarse en la valoración de cada atributo, esto es, si la varianza del indicador objetivo es muy baja, con valores cercanos a cero, el valor X casi coincidirá con el parámetro objetivo O, pudiendo considerarse que los efectos del parámetro S son muy bajos, y viceversa.

2.1.4.4. Métodos desagregados basados en la percepción y en las expectativas

Básicamente, los métodos incluidos en este grupo definen una serie de índices relacionados con diversos conceptos de percepción y expectativas, y realizan comparaciones analizando los valores obtenidos tras calcular sus diferencias.

Parasuraman et al. (1991) definieron dos parámetros para caracterizar las expectativas de los usuarios: servicio deseado (DS), que es el servicio que el usuario espera recibir y servicio adecuado (AS), que es el nivel de calidad mínimo que el usuario considera aceptable, y uno para el servicio percibido (PS). La zona de tolerancia de expectativas (ZOT) se calcula como la diferencia entre DS y AS, la superioridad del servicio (SS) se calcula como la diferencia entre DS y PS y la idoneidad del servicio (SA) se calcula como la diferencia entre PS y AS.

Otros autores han utilizado otros índices para medir la calidad del servicio, también calculados como diferencia entre parámetros. Así, Hu y Jen (2006) definieron un índice para cuantificar la calidad del servicio, calculando la diferencia entre los parámetros calidad tolerable (similar al parámetro AS) y la calidad del servicio percibida (similar al

parámetro PS). Chang et al. (2012) introdujeron el concepto de ZOT difusa, Tsai et al. (2011) aplicaron de forma combinada un proceso analítico para jerarquizar la importancia de los atributos teniendo en cuenta el índice IPA y Hu (2010) propuso el concepto de expectativas para evaluar la calidad del servicio (ZSQ), basándose en los índices ZOT e IPA.

2.1.4.5. Métodos estadísticos para determinar la importancia relativa de los atributos

Para determinar la importancia de cada atributo, normalmente se recurre a la elaboración de encuestas al usuario a partir de las cuales se recaba información sobre la calidad del servicio prestado mediante la puntuación de cada atributo en una escala de valores comprendida en un determinado intervalo, es lo que se conoce como métodos de importancias manifestadas. Los principales inconvenientes achacados a estos métodos residen en que la puntuación que los usuarios asignan a cada atributo no tiene por qué coincidir con la importancia asignada al atributo, y además es necesario realizar un mayor número de encuestas para poder realizar un análisis aceptable (Weinstein, 2000). En otros casos, se pregunta a los encuestados sobre una valoración global del servicio, y posteriormente se emplean métodos estadísticos para determinar la importancia de las relaciones entre cada atributo (análisis factorial, análisis de regresión múltiple, modelos de ecuaciones estructurales, etc.) y la calidad del servicio global, conocidos como métodos de importancias derivadas (De Oña, 2013).

El análisis factorial es una técnica estadística que permite determinar si existen relaciones entre un conjunto de variables de estudio y un reducido número de factores no observables, considerando para ello que existe un modelo causal subyacente. Este método suele emplearse como un paso previo al empleo de otros modelos más complejos, puesto que permite detectar las similitudes existentes entre varios de los atributos de estudio, y por tanto reducir el número de variables a considerar en las siguientes fases del estudio (Kim y Lee, 2011).

El método de correlación bivalente es capaz de establecer un ranking de importancia relativa de los atributos que se consideren en un estudio asignando unos ratios o valores a cada uno de ellos. Sin embargo, no capta adecuadamente las posibles relaciones de dependencia entre estas variables, por lo que siempre resulta necesario matizar los resultados de los coeficientes obtenidos.

Los métodos de regresión han sido extensamente utilizados en el campo del transporte público en general y del estudio de la calidad del servicio en particular (Kim y Lee, 2011; Aksoy et al., 2003; Tyrinopoulos y Aifadopoulo, 2008; Tyrinopoulos y Antoniou, 2008; Dell'Ollio et al., 2010, Huse y Evangelho, 2007; Henser et al., 2010; Castillo y Benítez, 2012) para estudiar el efecto que cada variable independiente tiene sobre la adopción de cada uno de los posibles valores de la variable dependiente. Los modelos de ajuste responden a una combinación lineal de las variables independientes, afectadas por un coeficiente, más un término independiente.

Los modelos de ecuaciones estructurales han cobrado fuerza en los últimos años en el campo de la calidad del servicio en el transporte público (Andreasen, 1995; Eboli y Mazzulla, 2007; Friman y Gärling, 2001; Friman et al., 2001; Joewono y Kubota, 2007a; 2007c; Karlaftis et al., 2001; Lai y Chen, 2011; Minser y Webb, 2010; Nurul-Habib et al., 2011; Stuart et al., 2000), al ser una herramienta que combina técnicas de análisis de regresión, análisis factorial y análisis de la varianza para determinar las diversas relaciones de dependencia que comparten simultáneamente varias variables, analizando estas relaciones no sólo entre las variables que son directamente observables en el estudio, sino entre las variables latentes o subyacentes.

2.1.4.6. Árboles de decisión

Recientemente, de Oña y de Oña (2012) han aplicado una técnica no convencional de minería de datos para analizar la importancia relativa de los atributos que caracterizan la calidad de un servicio de transporte público: los árboles de decisión. Los resultados obtenidos tras aplicarse a una base de datos de transporte público metropolitano por autobús y a una base de datos de transporte público en tren (de Oña et al., 2014) han demostrado que es una herramienta válida para este tipo de estudios.

2.2. Redes neuronales artificiales

2.2.1. Fundamento biológico de las redes neuronales artificiales

Una neurona biológica está compuesta esquemáticamente por tres partes: dendritas (canal de entrada de la información a través de sinapsis), soma (procesador) y axón (canal de salida de la información) (Figura 1).

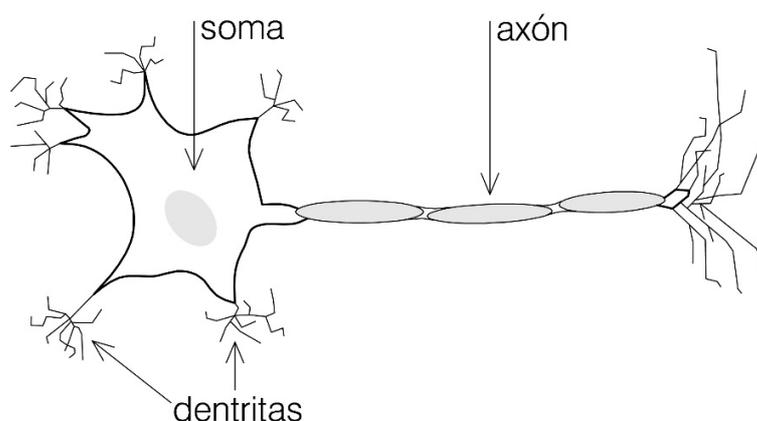


Figura 1. Representación de neurona biológica

En estado de reposo, el interior de la neurona está cargada negativamente, con un potencial de unos -60 mV. Los elementos químicos más importantes que influyen en la transmisión de la información son el fósforo, el sodio y el potasio (P²⁻, Na⁺ y K⁺). La membrana de la neurona tiene unos canales, que en el caso del potasio son simples aperturas, y en el caso del sodio son canales que permiten que éste entre a la neurona. El fósforo siempre está dentro debido a su gran tamaño.

Otra neurona, que transmite información a ésta, denominada presináptica, libera una serie de sustancias químicas con pequeños potenciales: excitadores (si son positivos) o inhibidores (si son negativos). Si se acumulan una cantidad suficiente de excitadores en el interior de la neurona postsináptica, y el potencial pasa de -60 mV a -45 mV, se activa el umbral de disparo y el Na⁺ penetra en la neurona postsináptica a través de su canal, elevando el potencial a +50 mV. En ese momento se genera una señal nerviosa que transmite información.

2.2.2. ¿Qué son las redes neuronales artificiales?

Las redes neuronales artificiales (ANN, Artificial Neural Networks en inglés) son sistemas de procesamiento de la información basados en el comportamiento biológico del cerebro humano (Martín del Brío y Sanz Molina, 2006).

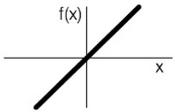
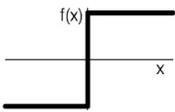
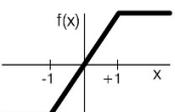
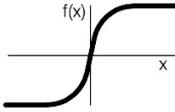
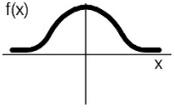
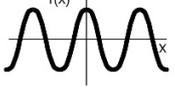
Una neurona artificial está compuesta por:

- **Conjunto de entradas $x_j(t)$** , que pueden ser de tipo binario [0,1] o analógicas (un intervalo continuo de valores).
- **Pesos sinápticos de la neurona i , w_{ij}** , que representan la intensidad de interacción entre cada neurona presináptica j y la neurona postsináptica i . Si el peso es positivo, tenderá a excitar a la neurona postsináptica, y si es negativo tenderá a inhibirla.
- **Regla de propagación, $\sigma(w_{ij}, x_j(t))$** , que proporciona el valor del potencial postsináptico $h_i(t) = \sigma(w_{ij}, x_j(t))$ en función de los valores de entrada y de los pesos sinápticos.
- **Función de activación $a_i(t) = f_i(a_i(t-1), h_i(t))$** , que proporciona el estado de activación actual en función del estado de activación anterior y del potencial postsináptico.
- **Umbral o bias θ_i** , que se resta del potencial sináptico, y que ejerce una función asimilable al del umbral de disparo de la neurona biológica.
- **Función de salida $F_i(a_i(t))$** , que proporciona la salida actual de la neurona en función de su estado de activación actual.

En muchos modelos de ANN, se hace la simplificación de considerar que el estado actual de la neurona no depende de su estado anterior, es decir, que $a_i(t) = f_i(h_i(t))$, y que además la función de salida $F_i(a_i(t))$, y la propia salida $y_i(t)$, coincide con la función de activación:

$$y_i(t) = F_i(a_i(t)) = a_i(t) \quad (7)$$

Las funciones de activación más habituales son las que se representan en la Tabla 1:

	FUNCIÓN	RANGO	GRÁFICA
Identidad	$y = x$	$[-\infty, \infty]$	
Escalón	$y = \text{signo}(x)$ $y = H(x)$	$\{-1, 1\}$ $\{0, 1\}$	
Lineal a tramos	$y = -1$ si $x < -1$ $y = x$ si $-1 \leq x \leq 1$ $y = 1$ si $x > 1$	$\{-1, 1\}$	
Sigmoidea	$y = \frac{1}{1 + e^{-x}}$ $y = \text{tgh}(x)$	$[0, 1]$ $[-1, 1]$	
Gaussiana	$y = A \cdot e^{-B \cdot x^2}$	$[0, 1]$	
Sinusoidal	$y = A \cdot \text{sen}(w \cdot x + \varphi)$	$[-1, 1]$	

Fuente: Martín del Brío y Sanz, 2006

Tabla 1. Funciones de activación habituales

Con frecuencia, se añade al conjunto de pesos de las neuronas un parámetro denominado umbral o bias, θ_i , que se resta del potencial sináptico, de manera que el argumento de la función de activación es:

$$h_i(t) = \sum w_{ij} \cdot x_j(t) - \theta_i \quad (8)$$

Si se establece un paralelismo entre una neurona biológica y una neurona artificial (Figura 2), podría decirse que el conjunto de entradas $x_j(t)$ del modelo artificial ejerce una función asimilable a las dendritas de la neurona biológica, ya que actúan como canales de entrada de la información. El procesamiento de la información y la transformación de ésta mediante la determinación del potencial postsináptico $h_i(t)$ y el cálculo del estado de activación a través de la correspondiente función de activación tiene lugar en el cuerpo de la neurona artificial, equivalente al núcleo celular. Finalmente, la función de salida $y_i(t)$

ejerce la función de transmisora de la información transformada, al igual que el axón en la neurona biológica.

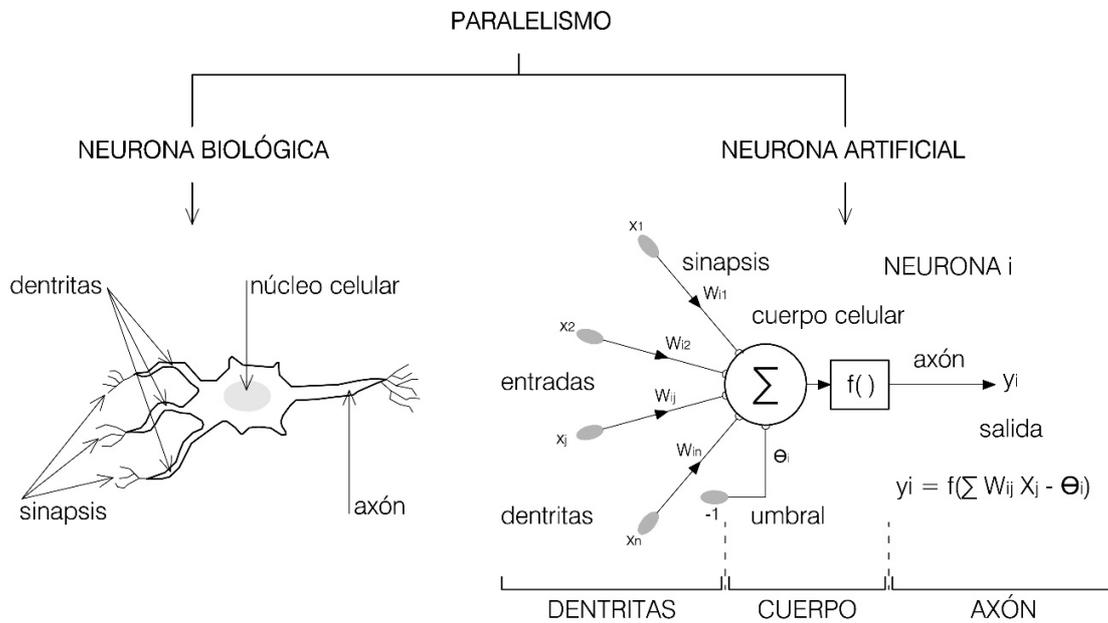


Figura 2. Paralelismo entre neurona biológica y neurona artificial

Ni las neuronas biológicas ni las artificiales funcionan por sí solas de forma independiente, sino que lo hacen en estructura de red. Una ANN está compuesta por unidades elementales de información, las neuronas artificiales, organizadas en capas e interconectadas entre sí a través de los pesos sinápticos. Cada peso sináptico representa la intensidad de la interacción entre cada par de neuronas. Así, los elementos que componen un sistema neuronal artificial (Figura 3) son (Martín del Brío y Sanz Molina, 2006; Rumelhart y McClelland, 1986):

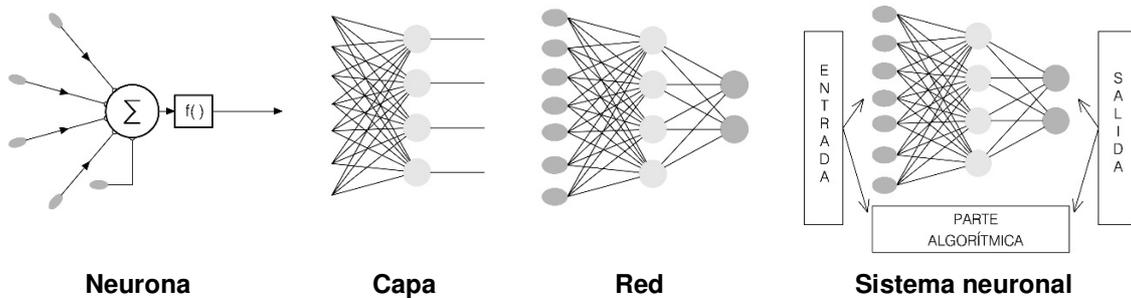
- Un conjunto de procesadores elementales o neuronas artificiales.
- Un patrón de conectividad o arquitectura.
- Una dinámica de activaciones.
- Una regla o dinámica de aprendizaje.
- El entorno donde opera.

Cualquier sistema neuronal, debe diseñarse para que funcione de la siguiente manera (Martín del Brío y Sanz Molina, 2006):

- Que procese la **información en paralelo**, y no en serie ejecutando de manera secuencial un paso tras otro.
- Que tenga **memoria distribuida**, es decir, en los sistemas neuronales biológicos la información se distribuye entre las sinapsis de la red, de modo que si una

neurona muere, el sistema se redistribuye para que todo siga funcionando bien. Los sistemas neuronales admiten errores.

- **Adaptabilidad** al entorno, pudiendo modificar las sinapsis en función de las condiciones de contorno y de lo aprendido.



Fuente: Martín del Brío y Sanz, 2006

Figura 3. Estructura jerárquica de un sistema de ANN

2.2.3. Clasificación de las redes neuronales artificiales

Atendiendo a cómo se mueve el flujo de datos en la red neuronal, los sistemas neuronales artificiales se clasifican en (Martín del Brío y Sanz, 2006):

- Redes unidireccionales (feedforward), la información circula en un único sentido, desde las neuronas de entrada hacia las de salida.
- Redes recurrentes o retroalimentadas (feedback), la información puede circular entre las capas en cualquier sentido, incluido el de salida-entrada.

Atendiendo a la dinámica de actualización del estado de las neuronas de los sistemas neuronales artificiales, ésta se clasifica en:

- Dinámica síncrona, cuando la información se actualiza siguiendo un cierto reloj común, es decir, en una misma capa toda la información se actualiza al mismo tiempo.
- Dinámica asíncrona, cuando no siguen un cierto reloj común, sino que cada neurona actualiza su estado sin atender a cuándo lo hacen las demás. Con esta dinámica es con la que funcionan los sistemas neuronales biológicos.
- Dinámica no determinista (estocástica), cuando se fuerza a que la salida de una neurona tenga carácter probabilístico.

Atendiendo a los modos de recuerdo y aprendizaje, existen dos modos de trabajo en las redes neuronales: aprendizaje o entrenamiento y recuerdo o ejecución.

El aprendizaje consiste en ajustar los grados de libertad que tiene la red neuronal, a partir de un proceso de estimulación por el entorno que rodea la red. En la mayoría de las ocasiones consiste en ajustar los valores de los pesos sinápticos, que a priori comienzan el entrenamiento con valores nulos o arbitrarios.

Existen 2 niveles de aprendizaje:

- El de modelado de la sinapsis, que consiste en modificar los pesos sinápticos siguiendo una cierta regla de aprendizaje, construida a partir de la optimización de una función de error (que mide la eficacia actual de la red). Este proceso normalmente es iterativo.
- El de creación o destrucción de neuronas, en el cual se modifica la propia arquitectura de la red.

Por otra parte, existen 4 tipos de aprendizaje:

- Supervisado, se define una función $E[W]$, que es función de los pesos sinápticos, y que representa el error esperado en la red. Se pretende estimar una función multivariable (desconocida en principio), que representará el funcionamiento de la red neuronal, a partir de valores x e y (entradas y salidas) tomados aleatoriamente, por medio de minimización iterativa de $E[W]$.
Es decir, se presentan a la red un conjunto de entradas con sus respuestas esperadas u objetivos, y se ajustan los pesos hasta que las salidas tiendan a ser las deseadas.
- No supervisado o auto-organizado, donde se introducen un conjunto de valores x de entrada, sin adjuntar la respuesta que se espera. Mediante una función de probabilidad $p(x)$, que describe la distribución de patrones que capta de los datos de partida, se pueden reconocer características comunes en el conjunto de entradas, y extraer rasgos o agrupar patrones según su similitud.
- Híbrido, es una mezcla de los dos anteriores. En unas capas se aplica el aprendizaje supervisado y en otras el auto-organizado.
- Aprendizaje reforzado, está a medio camino entre el supervisado y el auto-organizado, pero no es exactamente ninguno de ellos. En este caso, se introducen valores de entrada pero no los valores de salida esperados. Sin embargo, sí que se define, al igual que en los modelos supervisados, una función error que indica de manera global el funcionamiento de la red.

2.2.4. Métodos de cálculo de la importancia relativa de las variables consideradas en los modelos de redes neuronales artificiales

Las ANN no se basan en una ecuación predefinida o fórmula, sino en su capacidad para capturar la información inherente en los datos durante el proceso de entrenamiento. Su arquitectura contiene unos parámetros que son capaces de proporcionar, una vez que la red ha sido entrenada correctamente, respuestas correctas a los nuevos casos que se le presentan. En estos parámetros reside su conocimiento (Palmer y Montaña, 2002a). Esta

forma singular de aprendizaje permite capturar las relaciones altamente no lineales (Watts y Worner, 2008) y complejas (Mohammadipour y Alavi, 2009) existentes entre las variables predictoras, pero dificulta que sea fácil extraer una explicación explícita sobre la relación que existe entre las variables explicatorias (de entrada) y las dependientes (de salida). Sin embargo, las técnicas estadísticas clásicas sí que permiten conseguir este objetivo (Azadeh et al., 2011). Es por ello que las ANN se incluyen en el grupo de técnicas de minería de datos denominadas “cajas negras” (Cortez y Embrechts, 2013) ya que, para un fenómeno dado (variable de salida) teóricamente es muy difícil conocer la importancia relativa de cada variable predictora (variable de entrada).

Debido a este problema, en la literatura científica han proliferado en las dos últimas décadas diversos métodos para determinar la importancia relativa de cada variable predictora en los modelos de ANN.

Sung (1998) aplicó tres métodos: análisis de sensibilidad (Zurada et al., 1994; Engelbrecht et al., 1995), curvas difusas (Lin y Cunningham, 1995) y cambio del error cuadrático medio (MSE) (He et al., 1997) a una base de datos relacionada con la ingeniería del petróleo y comparó los resultados obtenidos, concluyendo que el método de las curvas difusas se comportaba mejor que los otros dos. Olden y Jackson (2002) describieron varios métodos que habían sido desarrollados por otros autores: el método del diagrama de interpretación neuronal (Özesmi y Özesmi, 1999), el algoritmo de Garson (Garson, 1991; Goh, 1995) y un método de análisis de sensibilidad (Lek et al., 1995; 1996a, 1996b), y propusieron uno nuevo, denominado “randomization approach”. Tras aplicar estos métodos a una base de datos de ecología, demostraron que los resultados diferían en función del método aplicado. Palmer y Montaña (2002b) también pusieron de manifiesto los problemas de aplicabilidad y de inestabilidad de resultados que presentaban tanto los métodos de cálculo basados en los valores de los pesos sinápticos como aquéllos basados en análisis de sensibilidad. Argumentaron que los primeros no eran efectivos (Garson, 1991; Rzepoluk, 1998; Hunter et al., 2000) y que los segundos presentaban algunos problemas en función de la naturaleza cualitativa o cuantitativa de las variables. Así, desarrollaron una nueva metodología de cálculo de la importancia relativa de las variables dependientes en modelos de ANN, denominada Análisis de Sensibilidad Numérico (NSA) (Palmer y Montaña, 2003), que determina la relación entre cada variable de entrada de la ANN y de salida, mediante el concepto de “pendiente”, sin tener en cuenta la naturaleza cuantitativa o cualitativa de las variables. Gevrey et al. (2003) analizaron y compararon los resultados obtenidos tras aplicar a una base de datos de ecología siete métodos ya existentes: derivadas parciales (Dimopoulos et al., 1995), método de Garson (Garson, 1991; Goh, 1995), método Perturb (Yao et al., 1998; Scardy y Harding, 1999), método Profile (Lek et al. 1995; 1996a; 1996b), y “stepwise” con algunas de sus variantes (Sung, 1998). El estudio concluyó que el método de derivadas parciales aportaba resultados más coherentes, mientras que el método “stepwise” clásico presentaba peor comportamiento. Olden et al. (2004) continuaron el trabajo de Gevrey et al. (2003) y aplicaron los mismos siete métodos más uno nuevo denominado Connection Weights (Olden y Jackson, 2002) a una base de datos simulada. En este caso, el nuevo método Connection Weights proporcionó los mejores resultados.

Gevrey et al. (2006) introdujeron una variación al método de las derivadas parciales (PaD), denominado PaD2, con el objetivo de analizar la contribución conjunta de cada combinación posible de parejas de variables, argumentando que puesto que en la naturaleza las variables normalmente interactúan entre sí, de modo que cuando una se modifica, las restantes también cambian. Kemp et al. (2007) implementaron el método Holdback Input Randomization (HIPR), basado en la alteración aleatoria de los valores de los parámetros de entrada de una ANN. Aplicaron este método a un sistema ecológico complejo, y los resultados obtenidos presentaron la misma eficacia que aquéllos obtenidos aplicando el método Connection Weights. Yeh y Cheng (2010) aportaron un punto de vista nuevo al análisis de contribución de las variables predictoras en los modelos de ANN, mediante la introducción de un método que considera no sólo los efectos lineales (derivadas de primer orden) entre las variables de estudio, sino los efectos de curvatura (derivadas de segundo orden). Cortez y Embrechts (2013) introdujeron tres nuevos métodos de análisis de sensibilidad (SA, Sensitivity Analysis): data-based SA (DSA), Monte-Carlo SA (MSA) y SA basado en análisis cluster (CSA), y realizaron una comparativa entre estos nuevos métodos y otros de SA ya existentes: SA unidimensional (1D-SA) (Kewley et al., 2000) y SA global (GSA) (Cortez y Embrechts, 2011). Además, desarrollaron nuevos índices y formas de representación de la importancia relativa de las variables.

Generalmente, el procedimiento seguido cuando se trabaja con ANN es seleccionar en primer lugar una arquitectura óptima, pero es muy difícil elegir el modelo óptimo de ANN debido a varios factores como la inicialización aleatoria de los valores de los pesos sinápticos, múltiples arquitecturas posibles de red y algoritmos de aprendizaje, los cuales convergen hacia mínimos locales en una superficie de error compleja. Cao y Quiao (2008) pusieron de manifiesto este hecho, y sugirieron que los métodos análisis de sensibilidad para determinar la importancia relativa de las variables deberían aplicarse no a una única ANN óptima o subóptima, sino a un conjunto de ANN que presentaran buen comportamiento. Siguiendo este criterio, propusieron un nuevo método denominado "Neural Network Committee (NNC)". Recientemente, otros autores como Paliwal y Kumar (2011) han resaltado también el problema que supone la elevada variabilidad de los pesos antes de empezar el entrenamiento. Puesto que se parte de un punto de la superficie de error (con varias crestas y valles) a partir del cual se desciende durante el entrenamiento hasta llegar a un mínimo, es difícil saber si la red entrenada se corresponde con un mínimo de la superficie de error local o global, y por tanto, saber si es la óptima o no. Estos autores propusieron un nuevo método denominado "intercuartil range" (IR), en el que cada arquitectura de red neuronal es entrenada un número de veces, y se utilizan los valores del primer y tercer cuartil de la distribución de pesos sinápticos obtenida para determinar la importancia relativa de cada variable.

Sin embargo, a pesar de la variedad de los métodos estudiados, no hay consenso general sobre cuál es el mejor modelo para determinar la contribución de las variables. Cuando se aplican varios métodos a arquitecturas de red óptimas o sub-óptimas, el ranking de importancia de las variables varía de un método a otro, e incluso aplicando varias veces el mismo método, indicando que los métodos actuales son inestables.

Muchos de los métodos existentes en la literatura científica para determinar la importancia relativa de las variables predictoras en un modelo de ANN han sido desarrollados para el MLP. Uno de los motivos puede achacarse al hecho de que esta tipología de red neuronal es la utilizada en aproximadamente un 70% de todos los estudios desarrollados con esta técnica (Gedeon et al., 1995). Hay dos grandes grupos de métodos para interpretar el efecto de las variables de entrada en un perceptrón MLP (Palmer y Montaña, 2002a):

- los que analizan la magnitud de los pesos.
- los que realizan un análisis de sensibilidad.

Los métodos basados en la magnitud de los pesos se centran en los valores de los pesos estáticos de las conexiones entre neuronas. Las técnicas de análisis de la magnitud de pesos han evolucionado a diferentes ecuaciones, y de ellas, una de las más conocidas y usadas es el método de Garson (Garson, 1991). En una ANN con una capa de entrada con $i \in [1, N]$ neuronas, una capa oculta con $j \in [1, L]$ neuronas y una de salida con $k \in [1, M]$ neuronas, la fórmula de Garson utiliza la expresión:

$$Q_{ij} = \frac{\sum_{j=1}^L \frac{w_{ij} \cdot w_{jk}}{\sum_{r=1}^N w_{rj}}}{\sum_{i=1}^N (\sum_{j=1}^L \frac{w_{ij} \cdot w_{jk}}{\sum_{r=1}^N w_{rj}})} \quad (9)$$

siendo:

w_{ij} = el valor del peso sináptico entre la neurona i de la capa de entrada y la neurona j de la capa oculta.

w_{jk} = el valor del peso sináptico entre la neurona j de la capa de oculta y la neurona k de la capa de salida.

$\sum_{r=1}^N w_{rj}$ = la suma de los pesos sinápticos entre las i neuronas de la capa de entrada y la neurona j de la capa oculta.

Q_{ij} = el porcentaje de influencia de la variable de entrada i sobre la neurona K de la capa de salida.

Una de las principales limitaciones del método de Garson (Olden y Jackson, 2002) es que considera el valor absoluto de los pesos de las conexiones cuando calcula la contribución de las variables, por lo que no puede saberse la dirección o signo que relaciona la variable predictora con la variable salida. Además, este método es poco fiable (Masters, 1993) debido a que la presencia de pesos altos en las conexiones entre la capa de entrada y la capa oculta indicarían que la importancia de esa variable es alta, cosa que no tiene porqué ocurrir. En los artículos en los que se compara este método de interpretación de variables con otros métodos (Palmer y Montaña, 2002b; Gevrey et al., 2003; Olden y Jackson, 2004), siempre es el que tiene mayores problemas para establecer una jerarquía en la importancia de las variables del modelo.

Para intentar solventar algunas de estas limitaciones, Olden y Jackson (2002) propusieron un nuevo método, denominado Connection Weights, que calcula la suma del producto de los pesos de conexión entre cada nodo de entrada y los nodos de la capa oculta, con los

pesos de conexión entre cada nodo de la capa oculta y el nodo de la capa de salida. Cuanto mayor sea el valor de la suma, mayor será la importancia relativa de la variable, que viene dada por la expresión:

$$R_{ik} = \sum_{j=1}^L w_{ij} \cdot w_{jk} \quad (10)$$

Siendo R_{ik} la importancia relativa de la variable de entrada x_i respecto a la variable de salida y_k .

Por otra parte, Tchaban et al. (1998) presentaron una variante de este análisis estático de pesos, denominado método Weight Product (WP), que calcula el ratio entre el valor de la variable de entrada x_i y el valor de la salida y_k en función del producto de sus correspondientes pesos sinápticos, mediante la expresión:

$$WP_{ik} = \frac{x_i}{y_k} \cdot \sum_{j=1}^L w_{ij} \cdot w_{jk} \quad (11)$$

En esta misma línea de trabajo, Özesmi y Özesmi (1999) propusieron el método del diagrama de interpretación neuronal (NID), basado en la idea de interpretar la importancia de las variables predictoras respecto a las variables de salida en función de la intensidad y el signo de los pesos de las conexiones de la red. Este método es más visual que los anteriores, ya que tras representar esquemáticamente el grafo de la arquitectura de red, se marcan con línea gruesa las conexiones con mayor valor absoluto, y con líneas más débiles aquéllas que tengan un valor pequeño. La importancia relativa asignada a cada variable es función de los valores excitadores (peso con signo positivo) o inhibidores (peso con signo negativo) de los pesos sinápticos que conectan cada par de neuronas, de modo que se considera que los efectos de una variable de entrada son importantes cuando el signo de las conexiones neurona entrada-neurona oculta y neurona oculta-neurona salida son ambos positivos o ambos negativos. Por el contrario, se considera que los efectos de una variable de entrada son negativos cuando las conexiones neurona entrada-neurona oculta y neurona oculta-neurona salida tienen signos opuestos. Este método presenta muchas limitaciones (Olden y Jackson, 2002), ya que no puede determinarse de manera cuantitativa el valor de las importancias de las variables, y además la complejidad aumenta mucho si el número de variables o de neuronas ocultas es elevado.

El método del Interquartil Range (IQR) (Paliwal y Kumar, 2011) aboga por trabajar con conjuntos de ANN y considerar la variabilidad de valores y signos que presentan los pesos sinápticos cuando se entrena una ANN de forma aleatoria un número de veces t . Así, la importancia relativa (R_i) de una variable de entrada x_i respecto a una variable salida k se mide como:

$$R_i = \frac{1}{J} \cdot \sum_{j=1}^J \text{Rango intercuartil } (W_{ij}) \quad (12)$$

Siendo Rango Intercuartil (W_{ij}) la diferencia entre el primer y el tercer cuartil de la distribución de pesos W_{ij} que conecta el nodo de entrada i con el nodo oculto j , y J el número total de nodos de la capa oculta. A mayor valor de R_i , mayor será la importancia relativa de la variable i respecto a la variable k .

El segundo grupo de métodos de cálculo de la importancia relativa de las variables, podría dividirse a su vez en dos: los basados en la variación del comportamiento de la red (error de generalización) en función de las Perturbaciones inducidas en los valores de las variables predictoras, y aquéllos en los que se analiza la magnitud de los cambios producidos en cada una de las variables predictoras en función de su pendiente de variación, es decir, de la pendiente de su primera derivada (Palmer y Montaña, 2002a).

Los métodos del primer subgrupo han sido bastante utilizados en numerosos estudios. Así, el método Perturb (Yao et al., 1998; Scardy and Harding, 1999) introduce ruido en los valores de una de las variables de entrada mientras las variables restantes mantienen sus valores originales y posteriormente se establece una comparativa entre el error de generalización que presenta la red antes de la perturbación y después de ésta, mediante la medida del error cuadrático medio (MSE). Es decir, el método aplica progresivamente un ruido δ en varios escalones de perturbación, de modo que la variable x_i cambia sus valores a $x_i = x_i + \delta$, y el rango de variación del error de generalización indica la importancia relativa de la variable; cuanto mayor sea el valor del MSE, mayor será la importancia relativa de la variable predictora i respecto a la variable salida k .

El método Profile (Lek et al., 1995; 1996a; 1996b) analiza la evolución de cada variable predictora a lo largo de una escala o rango de valores, mientras el resto de variables mantiene sus valores fijos. Tras repetir este procedimiento para todas las variables presentes en la ANN, se suele representar una curva denominada perfil de variación en la que se representan los valores de la mediana obtenidos tras aplicar las perturbaciones a cada una de las variables x_i . Las variables cuyo perfil de variación sufra mayor modificación, tendrán una importancia relativa más alta en el modelo, mientras que en aquéllas en los que las perturbaciones introducidas apenas presenten cambios en el perfil de variación indicarán tener una importancia relativa más baja en el modelo.

El método Stepwise, también perteneciente a este primer subgrupo, añade o elimina paso por paso cada una de las variables de entrada del modelo, y analiza el efecto que produce en el rendimiento de la red. Un criterio adoptado para determinar la importancia relativa es considerar que las variables que provocan mayores cambios en el error cuadrático medio MSE son más importantes (Gevrey et al. 2003). Otro criterio también usado con asiduidad es aquel que considera más importantes aquéllas variables que provocan una mayor disminución del error MSE (Sung 1998). Hay dos formas de aplicar este método:

A. Forward Stepwise: se generan tantos modelos I de ANN como variables predictoras se consideren en el estudio, de modo que en cada modelo se incluya sólo una de las variables en su capa de entrada. Posteriormente, se generan otros $I-1$ modelos de ANN, combinando la variable que proporcionó un menor error MSE cuando se creó el modelo de red en el que ésta era la única variable de la capa de entrada, con el resto de variables. Sucesivamente, se van generando modelos de ANN de dos neuronas en la capa de entrada, de tres, de cuatro..., hasta incluirlas todas en el modelo I . El orden de inclusión será función del error que proporcionaron en solitario, y este orden de inclusión también indicará el grado de contribución de las variables predictoras al modelo.

B. **Backward Stepwise:** Se crean I modelos de redes neuronales en los que se elimina una de las variables de entrada y se mantienen las restantes. En cada modelo se elimina una variable de entrada distinta y se mide su error MSE, considerando que la red que proporciona mayor error ha eliminado la variable más importante. Tras considerar esa variable como más importante, se crean $I-1$ modelos de redes, y en cada uno de ellos se va eliminando una de las variables de entrada. El orden de eliminación es función del error MSE que provocaron los I modelos originales en que esas variables no fueron incluidas (de mayor a menor error), y el orden de eliminación de esas variables indica el orden de importancia de su contribución.

El propio autor, consciente de una metodología tan tediosa, propuso dos nuevas variantes con objeto de disminuir el tiempo de procesamiento y de aplicación del mismo, en las que, por una parte, una vez entrenada una ANN, se va eliminando cada una de las variables y los pesos de sus conexiones, y posteriormente se mide el error de la red. De esta forma, a mayor error provocado por la eliminación de una variable concreta, mayor importancia relativa de la misma en el modelo. La otra variante se basa en, de nuevo una vez entrenada una ANN con todas las variables, fijar sucesivamente todos los valores de una variable de entrada en un valor, que suele ser la media, observando cómo afecta esta variación al error.

Los métodos de análisis de sensibilidad propuestos por Cortez y Embrechts (2013) también se plantean con un enfoque similar, presentando pequeñas variaciones respecto al Profile, en las que se mide el efecto de la perturbación de cada una de las variables (caso del método 1D-SA) o de varias variables que son perturbadas simultáneamente (método GSA). En el método MSA los datos para realizar el análisis de sensibilidad no se toman del conjunto de entrenamiento, sino que son construidos a partir de una distribución uniforme calculada a partir del rango de valores que adopta cada variable. El análisis CSA también sigue siendo similar al método Profile, con la diferencia de que utiliza análisis cluster para realizar la partición del rango de valores existente entre el máximo y el mínimo de cada variable, en un número de niveles L .

Dentro del segundo subgrupo de métodos para análisis de sensibilidad de las variables, se ha utilizado con bastante asiduidad el método PaD de derivadas parciales (Dimopoulos et al., 1995), que analiza los efectos de primer orden de las variables predictoras del modelo respecto a cada una de las variables de salida, utilizando la base de datos de entrenamiento.

2.2.5. Aplicación de las redes neuronales artificiales a la calidad del servicio

En la revisión bibliográfica realizada en este trabajo de investigación no se han encontrado estudios en los que se haya aplicado la técnica de las ANN para analizar la calidad del servicio en el transporte público, aunque sí ha sido utilizada con éxito en otros campos de la ingeniería del transporte. La aplicación de ANN al transporte proliferó a partir de los años 90 (Dougherty, 1995), y desde entonces han surgido numerosos

estudios en los que se ha aplicado esta técnica a las distintas ramas del mismo. En esta década se elaboraron estudios del comportamiento del conductor (Yang et al., 1992; Dougherty y Joint, 1992; Lyons y Hunt, 1993; Hunt y Lyons, 1993; Pant y Balakrishman, 1994; Kornhauser, 1991; Pomerleau et al., 1991) en los que se analizaban aspectos como la toma de decisiones ante los condicionantes que presentaban las distintas alternativas de viaje, el cambio de conducta cuando se proporcionaba a los conductores determinada información o comportamientos ante las condiciones de conducción programadas en simuladores. Así, Lee et al. (2010) aplicaron ANN y árboles de decisión para analizar los factores que afectan a las decisiones de los conductores al seleccionar la ruta de viaje entre varias alternativas, y Xie et al. (2003) modelaron diferentes modos de viajes al trabajo usando tres técnicas diferentes: ANN, árboles de decisión y modelos logísticos multinomiales.

Numerosos han sido también los estudios que han utilizado esta técnica para analizar los flujos de datos de tráfico, a través de matrices origen-destino, e incluso combinándola con otras técnicas como los algoritmos genéticos (Xiong y Schneider, 1992), los que han modelado el estado del tráfico (Bullock et al., 1992, 1993).

Costa y Marquellos (1997) analizaron la eficiencia en términos de productividad del metro de Londres entre los años 1970 y 1994, y los estudios llevados a cabo por Morato et al. (2001) y por Papagiannaki et al. (2003) utilizaron ANN para predecir el comportamiento del conductor ante las variaciones del tráfico. Del Hoyo et al. (2009) usaron las tipologías de ANN MLP, redes difusas y redes neuronales auto-organizadas para medir la calidad del servicio ofrecida por diversas aplicaciones de video y voz que pueden descargarse a través de internet y que indican en tiempo real el estado del tráfico de diferentes rutas alternativas de viaje.

Las ANN también han sido usadas con éxito para estudiar la calidad del servicio en campos no relacionados con la ingeniería, obteniéndose unos ajustes y exactitud de los modelos entrenados muy elevados. Principalmente se han utilizado en el sector terciario para medir la calidad percibida por los usuarios por el servicio percibido, debido fundamentalmente a la gran importancia que tiene para la viabilidad financiera y la competitividad en el mercado de la compañías del sector, el ofrecer un valor añadido a sus productos y conseguir que sus clientes queden satisfechos (Armstrong y Kotler, 2000; Weitz y Jap, 1995). En la literatura aparecen estudios en los que se analizan los factores más influyentes en la calidad del servicio de comunicación plantilla – cliente para empresas del sector servicios (Lin, 2007; Bejou et al., 1996), el grado de satisfacción de los clientes de hoteles en función del trato recibido por el personal (Deng et al., 2008; Deng y Pei, 2009), de los estudiantes usuarios de servicios de concesiones de restaurantes en colegios, institutos y universidades (Larasati et al., 2012). Moutinho et al. (1996) evaluaron el grado de satisfacción de los compradores de una marca de vehículos y su fidelidad hacia la misma; Mahatrapa y Khan (2006) aplicaron ANN para evaluar la calidad del servicio en el sector de la educación; y Azadeh et al. (2011) también aplicaron ANN para medir el grado de satisfacción de los trabajadores de diversas compañías de la industria de la refinera de gas.

2.2.6. Ventajas y desventajas de las redes neuronales artificiales

Tras la revisión literaria realizada sobre las ANN en los apartados anteriores, pueden señalarse las principales ventajas que presenta el uso de esta técnica:

- Las ANN no necesitan establecer un modelo o formulación previa (Palmer y Montaña, 2002a), sino que capturan la información inherente en los datos suministrados durante el proceso de entrenamiento y van ajustando los parámetros que la definen hasta que finaliza el entrenamiento. Esta forma singular de aprendizaje, permite captar relaciones intrínsecas altamente no lineales entre las variables de estudio (Watts y Worner, 2008) y complejas (Mohammadipour y Alavi, 2009).
- Los modelos de ANN entrenados presentan una elevada capacidad predictiva, en la inmensa mayoría de las ocasiones superior a los modelos lineales, debido a que han demostrado ser muy buenos aproximadores de funciones complejas (Hastie et al., 2001), y algunas de las tipologías de ANN son aproximadores universales de funciones (Funahashi 1989; Hornik et al., 1989).
- Su elevada capacidad asociativa la convierte en una técnica robusta cuando trabaja con casos de entrenamiento en los que faltan datos o en los que la información es imprecisa (Bejou et al., 1996) y presenta ruido (Kasabov, 1996).
- Las relaciones de colinealidad entre las variables predictoras consideradas en el estudio no afectan al comportamiento ni a la exactitud de las ANN, siendo ésta una de las mayores desventajas de las técnicas estadísticas convencionales (Bejou et al., 1996).

Sin embargo, esta técnica también presenta algunas desventajas, citadas a continuación:

- La capacidad explicatoria de los modelos de ANN es muy baja (Azadeh et al., 2011), esto es, es muy difícil conocer la influencia que cada una de las variables tiene en el modelo, y por tanto, su importancia relativa. Es por ello que en la literatura científica estas ANN se incluyen dentro del grupo de técnicas denominadas “cajas negras” (Cortez y Embrechts, 2013).
- No existe consenso sobre cuál de los métodos de cálculo de importancia relativa de las variables es más adecuado y proporciona mejores resultados, ya que cuando se aplican varios métodos a una arquitectura de ANN óptima ó sub-óptima entrenada, el ranking de importancia de las variables obtenido difiere de un método a otro, indicando la inestabilidad que presentan estos métodos (de Oña y Garrido, 2014).
- Es muy difícil seleccionar la ANN óptima debido a varios factores como la aleatoriedad de los valores iniciales de los pesos sinápticos, la existencia de múltiples arquitecturas posibles y algoritmos de aprendizaje que convergen hacia mínimos locales en una superficie de error compleja (Cao y Qiao, 2008).

- Lentitud en el proceso de entrenamiento (Martín del Brío y Sanz, 2006; Hagan et al., 1996).
- Es necesario realizar un pretratamiento de los datos, puesto que haciendo coincidir el rango de valores de las variables predictoras con el de las funciones de activación, se acelera la convergencia en la fase de entrenamiento (Martín del Brío y Sanz, 2006).

2.3. Consideraciones finales

De la revisión realizada sobre el campo de estudio de la calidad del servicio en el transporte público y sus metodologías de análisis y sobre la técnica de las Redes Neuronales Artificiales, se pueden establecer las siguientes consideraciones:

- El concepto de calidad del servicio en el transporte público es complejo. Las características que mejor lo definen son su intangibilidad, heterogeneidad e inseparabilidad. La calidad del servicio está sujeta a una gran subjetividad, debido a que está relacionada con la percepción y las expectativas de los usuarios de un determinado modo de transporte, y a una gran heterogeneidad en cuanto a las características socioeconómicas y de población de esos usuarios y a las particularidades de cada servicio.
- Se ha producido un cambio de tendencia en cuanto a qué aspecto de la calidad del servicio en el transporte público prestar más atención, pasándose de una perspectiva centrada fundamentalmente en el análisis de las administraciones y gestores sobre la eficacia y rendimiento del servicio, a otra más reciente en la que la mayoría de los estudios analizan la perspectiva de calidad percibida por los usuarios del mismo.
- El análisis de la calidad del servicio en el transporte público ha sido realizado desde dos puntos de vista principales: aquél que la caracteriza a través de un único índice global, que origina el conjunto de métodos denominados agregados, y el que lo hace analizando los diversos atributos que lo definen, que da lugar a los métodos denominados desagregados. Ambas tipologías de métodos aportan información valiosa sobre la calidad del servicio, ya que los modelos agregados son muy útiles para realizar un seguimiento temporal sobre las variaciones en la valoración de la calidad del servicio, mientras que los desagregados permiten obtener un conocimiento más profundo sobre los atributos que más influyen en la calidad del servicio percibida por el usuario.
- Existen varias interpretaciones en cuanto a las relaciones entre el concepto de calidad del servicio en el transporte público y los conceptos de percepción y expectativas de los usuarios, de modo que hay autores que consideran que ambos influyen en la calidad del servicio, mientras otros creen que solamente uno de estos conceptos la afecta. De ahí que los modelos agregados y desagregados mencionados en el punto anterior se desglosen a su vez en varios, que consideran en sus formulaciones unas u otras interpretaciones.

- La herramienta de uso más extendido para obtener información sobre aspectos de la calidad del servicio relacionados tanto con su valoración global como con la valoración asignada a cada uno de los atributos que la definen son las encuestas. Existen multitud de criterios para realizarlas y para que los usuarios encuestados evalúen la calidad global y sus atributos, utilizando diversos valores incluidos en un cierto rango o intervalo o mediante respuestas lingüísticas. A pesar de la diversidad de elementos que pueden analizarse a través de las encuestas, parece que hay cierto consenso en que es necesario incluir en los cuestionarios no sólo aspectos que son comunes en las encuestas de satisfacción de un servicio de transporte público, sino que deben tenerse en cuenta otros particulares de dicho servicio, que consideren aspectos socioeconómicos, de población o especiales de él.
- Es frecuente utilizar técnicas estadísticas para seleccionar las variables y la información más relevante recopilada en las encuestas, además de otras como análisis cluster o lógica difusa, así como el criterio de los expertos en la materia o el aplicado en estudios similares.
- Las técnicas estadísticas (p.e., análisis factorial, métodos de correlación bivalente, modelos de regresión y modelos de ecuaciones estructurales) han sido aplicadas extensamente en numerosos estudios de calidad del servicio en el transporte público, pero las limitaciones propias de cada una de ellas, que no ha hecho posible llegar a un consenso global sobre qué método es más adecuado para el estudio de la calidad del servicio en el transporte público, está llevando a la búsqueda de otros métodos que solventen al menos parte de éstas.
- Recientemente, se ha aplicado una técnica no paramétrica de minería de datos al análisis de la calidad del servicio en el transporte público, los árboles de decisión, que salva algunas de las limitaciones presentes en los métodos anteriores hasta la fecha, pero que sin embargo, presenta otras como problemas de estabilidad en los resultados y exactitud.
- Las Redes Neuronales Artificiales son una técnica matemática novedosa y no paramétrica que se basa en la simulación del funcionamiento biológico del cerebro humano. Ha sido aplicada en numerosos campos de investigación muy diferentes como la sanidad, la economía, el tráfico, la agricultura, etc., y en todos ellos ha proporcionado buenos resultados de exactitud, debido fundamentalmente a su gran capacidad de tratamiento de información inexacta, imprecisa y subjetiva, y de captar relaciones altamente no lineales inherentes en las variables de estudio.
- Hasta el momento, se han empleado Redes Neuronales Artificiales en el estudio de la calidad del servicio en otros campos diferentes al del transporte público, habiéndose obtenido en estos estudios muy buenos resultados. Sin embargo, no se tiene constancia de la aplicación de esta técnica al análisis de la calidad del servicio en el transporte público por autobús.
- Los expertos en esta técnica coinciden en apuntar que su principal debilidad reside en su incapacidad para conocer explícitamente el grado de importancia de cada variable independiente respecto a la/las variables dependientes consideradas en el estudio, por lo que se incluye dentro de las técnicas denominadas “cajas negras”.

Esta limitación, ha impedido su uso en diversos estudios en los que resulta muy importante conocer la importancia que un determinado atributo tiene en un resultado en concreto, sobre todo en ámbitos relacionados con la sanidad, la economía o la geotecnia. Es por ello que en las dos últimas décadas han proliferado numerosos métodos matemáticos con el objetivo de mejorar esta capacidad explicativa, pero no ha sido posible encontrar ninguno que pueda considerarse como el que presenta mejor comportamiento; de hecho, uno de los problemas achacados a estos métodos es la inestabilidad de sus resultados, es decir, cuando son aplicados a distintos modelos de Redes Neuronales Artificiales, los resultados de importancia relativa de cada variable independiente respecto a la/las variables dependientes difiere mucho en cada caso, siendo por tanto necesaria la opinión de un experto que evalúe la bondad de estos valores de importancia.

CAPÍTULO 3. OBJETIVOS

CAPÍTULO 3. OBJETIVOS

Esta tesis doctoral aplica Redes Neuronales Artificiales al campo de la calidad del servicio en el transporte público, dado que esta técnica ha demostrado ser eficiente para tratar problemas complejos en numerosos ámbitos de estudio. Estudios previos han utilizado técnicas estadísticas, y árboles de decisión para analizar la calidad del servicio en el transporte público. Sin embargo, la aplicación de Redes Neuronales Artificiales en este campo es nueva y proporciona una visión diferente que salva algunos de los inconvenientes que presentan las técnicas estadísticas y los árboles de decisión.

3.1. Objetivo general

El principal objetivo de esta tesis doctoral es validar la técnica de redes neuronales artificiales en el estudio de la calidad del servicio en el transporte público.

Conocer los atributos que son más valorados por los usuarios de un modo de transporte público es una necesidad fundamental de las administraciones y empresas encargadas de operar estos servicios, ya que es la herramienta que les permite mejorar las condiciones y expectativas de los usuarios, pudiendo captar así más viajeros y haciendo del transporte público un modo competitivo con el vehículo privado.

3.2. Objetivos específicos

Para evaluar la calidad del servicio en el transporte público mediante la técnica de redes neuronales artificiales, en esta tesis doctoral se han definido varios objetivos específicos.

- Determinar la importancia relativa de los atributos de calidad del servicio considerados en el estudio.
- Comprobar la coherencia de los resultados con los obtenidos **en otros estudios** presentes en la literatura científica.
- Desarrollar una nueva metodología para mitigar la inestabilidad inherente en las ANN al determinar la importancia relativa de las variables predictoras consideradas en el estudio.
- Comprobar la robustez y estabilidad de esta nueva metodología.
- Comparar la técnica de las redes neuronales artificiales con la técnica de los árboles de decisión, aplicadas ambas a la calidad del servicio en el transporte público por autobús.

CAPÍTULO 4. CASO DE ESTUDIO

CAPÍTULO 4. CASO DE ESTUDIO

Para estudiar un sistema de transporte es fundamental caracterizar tanto los aspectos relacionados con la prestación del servicio, esto es, líneas en funcionamiento, horarios, tiempos de recorrido y espera, etc., como aquéllos datos socioeconómicos que definen el contexto en el que el servicio opera, así como el grado de cumplimiento de las expectativas que los usuarios esperan satisfacer, ya que la capacidad para captar viajeros va a depender enormemente del conocimiento que se tenga de estos factores. Por ello, las entidades encargadas de gestionar los servicios realizan encuestas a los usuarios.

Este capítulo se desglosa en dos apartados. El apartado 4.1 realiza una descripción del sistema de transporte del que se han extraído los datos en los que se basa esta investigación, describiendo aspectos relacionados con el marco territorial, urbanístico y socioeconómico que justifican el modo en el que el sistema se ha ido desarrollando, y dando algunas pinceladas sobre la entidad encargada de su gestión, las líneas que operan el servicio y las características de funcionamiento del mismo. El apartado 4.2 se centra en la descripción del procedimiento seguido para la toma de datos que se recopiló en una encuesta realizada por el organismo gestor de este sistema de transporte, y que han servido como base para el desarrollo de este trabajo, y de las características de la muestra encuestada.

4.1. Descripción del sistema de transporte en estudio

4.1.1. Marco territorial y urbanístico

La ciudad de Granada se asienta en el contacto de Sierra Nevada con la Vega del Genil. Precisamente la relación de la ciudad con su vega y la forma de resolver ese contacto han centrado gran parte de la reflexión y el debate sobre el crecimiento de la ciudad en la sociedad y en la planificación urbanística y territorial desde los años 80 del siglo pasado.

El Plan de Ordenación del Territorio de Granada se elaboró en 2000 y fue revisado en 2005, estableciendo para el Área Metropolitana sus límites actuales.

La vega de Granada es un espacio protegido de excepcional valor que forma parte esencial del paisaje de la ciudad. La protección de este espacio es la justificación del trazado de la carretera de circunvalación, que se ciñe a la ciudad de forma espectacular conformando su borde este, un borde rotundo e impermeable entre la ciudad y su vega.

Paradójicamente, mientras el crecimiento de Granada hacia la vega se impedía por la planificación urbanística de la ciudad, las protecciones ambientales y territoriales han tenido una eficacia escasa en los pequeños municipios de la vega en los que la población se ha multiplicado. La ciudad, de hecho, ha crecido en la vega, de forma extensiva y con estructuras urbanas primarias que generan desplazamientos atraídos por la ciudad central.

Por tanto, el área metropolitana se caracteriza por la presencia de una ciudad, Granada, que actúa como centro neurálgico de un conjunto de 50 municipios circundantes (Figura 4), con fuertes relaciones de dependencia económica, laboral y social con la capital. Mientras la población de Granada está concentrada, la del área metropolitana ha continuado el proceso de dispersión poblacional iniciado en la década de los 80, disminuyendo el peso específico del volumen de población de Granada frente al conjunto que forma el resto de municipios del área.



Figura 4. Localización del Área Metropolitana de Granada

En total en el año 2006 se concentraba un total de 484.168 habitantes. En los 25 años anteriores al mencionado, ya se venía produciendo un proceso de expansión poblacional en los núcleos situados en la corona metropolitana, situados alrededor de la ciudad central.

Año	Granada (hab)	Resto Área Granada (hab)	Total Área Granada (hab)
1981	249.821	123.158	372.979
1991	257.580	152.065	409.645
2001	243.341	197.437	440.778
2006	237.929	246.239	484.168

Tabla 2. Evolución de la población en Granada capital y en su Área Metropolitana

Estos datos corresponden al conjunto de los municipios que constituyen el Área de Granada, sin embargo existen otros núcleos que establecen también una relación directa con la ciudad central y que pueden considerarse integrados igualmente en la estructura funcional de la aglomeración urbana (Tabla 2).

El conjunto de todos estos municipios, los integrantes es el Área de Granada y los que justificarían su integración representan una población superior al medio millón de habitantes (536.898) en el año 2006 (Tabla 3).

Área de Granada (2006):	484.168
Municipios externos (2006):	<u>52.730</u>
TOTAL (2006):	536.898

Municipio	Población 2006
Agrón	283
Alhama de Granada	6.137
Beas de Granada	1.092
Calicasas	565
Chimeneas	1.456
Cogollos de la Vega	2.095
Colomera	1.570
Deífontes	2.478
Dúdar	292
Dúrcal	6.856
Escúzar	738
Güejar Sierra	2.915
Huétor de Santillán	1.784
Íllora	10.304
La Malahá	1.703
Nigüelas	1.149
Nívar	668
Padul	7.871
Quéntar	1.036
Ventas de Huelma	734
Villamena	1.004
TOTAL	52.730

Tabla 3. Población de los municipios integrantes del Área Metropolitana en el año 2006

La creciente demanda de movilidad del vehículo privado y los problemas de congestión de tráfico han motivado, en las administraciones involucradas, mejoras en la capacidad de la red viaria mediante la ejecución de nuevos tramos y acondicionamiento de otros ya existentes. Este tipo de actuaciones, planificadas y ejecutadas respondiendo a las necesidades del transporte privado no han hecho más que fortalecer y promocionar el uso del mismo, generando en el tiempo nuevas y acuciantes situaciones de congestión. Por ello, sólo el fortalecimiento y promoción de los modos de transporte colectivos, con eficiencias espaciales, energéticas y ambientales mucho más optimizadas que el vehículo privado, permitirá un trasvase modal en el sistema de transportes hacia los modos colectivos, una optimización del espacio limitado del viario y la emisión de contaminantes, por tanto, una mejora sustancial de la actual calidad de vida de los ciudadanos del Área.

4.1.2. El consorcio de transportes metropolitanos del Área de Granada (CMTAG)

El Consorcio de Transporte Metropolitano del Área de Granada fue creado en diciembre de 2002 a través de la firma de un Convenio entre la Junta de Andalucía, la Diputación Provincial de Granada y 32 Ayuntamientos: Albolote, Alfacar, Alhendín, Armilla, Atarfe, Cájar, Cenes de la Vega, Chauchina, Churriana de la Vega, Cúllar Vega, Dílar, Fuente Vaqueros, Las Gabias, Gójar, Granada, Güevéjar, Huetor Vega, Jun, Maracena, Monachil, Ogíjares, Otura, Peligros, Pinos Genil, Pinos Puente, Pulianas, Santa Fe, Vegas del Genil, Víznar, La Zubia, Cijuela y Láchar. El Consorcio es una entidad de Derecho Público de carácter asociativo, dotada de personalidad jurídica independiente de la de sus miembros, patrimonio propio, administración autónoma y tan amplia capacidad jurídica de derecho público y de derecho privado como requiera la realización de sus fines, con especial atención a la puesta en marcha y gestión de infraestructuras y servicios del transporte de viajeros que permitan responder a las demandas y necesidades de movilidad, accesibilidad y calidad en los desplazamientos en transporte público regular de viajeros por carretera en el Área Metropolitana de Granada.

Con la creación del CMTAG, se realizaron diversas mejoras en los servicios prestados mediante la creación de nuevas líneas o el incremento de expediciones y/o ampliación de la franja horaria de funcionamiento en las ya existentes.

El resultado obtenido fue un incremento sostenido del 5 % anual durante los años 2004 – 2007 (Figura 5), alcanzándose en ese momento una cifra total en los servicios metropolitanos de 10.800.000 viajeros, 1.500.000 más que en el año anterior al periodo de actuación del CMTAG.

Este crecimiento contrasta con el reducido crecimiento registrado en los años anteriores a la creación del CMTAG (1,2%), debido fundamentalmente al incremento de la población metropolitana, más que a las mejoras en la calidad del sistema de transporte público regular de viajeros.

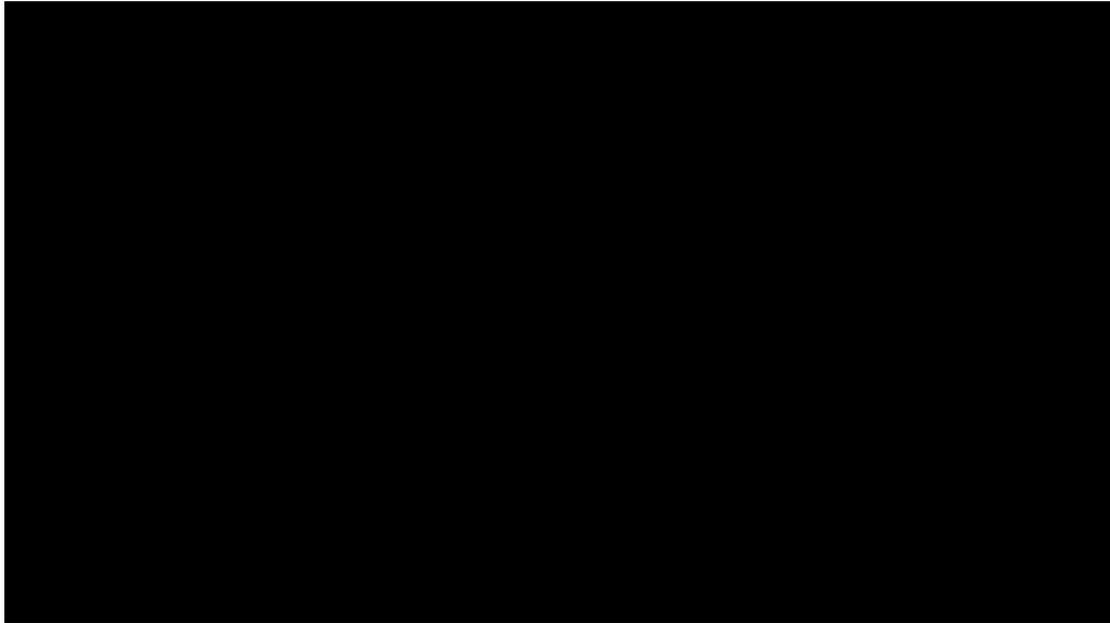


Figura 5. Evolución de viajeros en las líneas interurbanas del Área de Granada

4.1.3. Los servicios de autobús metropolitano

La red interurbana del Área de Granada se establece mediante 56 líneas de transporte prestada por 13 operadores (Tabla 4), las cuales movilizan más de 10 millones de viajeros. La flota de autobuses destinada es de 96 unidades. En el año 2006 se detectó una disminución de la tasa de crecimiento interanual del número de viajeros. El crecimiento durante el año 2005 fue del +5,1%, mientras que la registrada en el año 2006 resultó del +2,5 %.

La reducción de esta tasa es imputable a la ejecución masiva de obras en viarios urbanos del Área, especialmente en el caso de la ciudad central. No obstante, el número de viajeros a fecha de 2007 seguía siendo creciente, y la disminución no resultó muy acusada teniendo en cuenta la situación del sistema urbano de Granada, el cual está obteniendo actualmente tasas negativas por la misma causa.

El servicio de transporte regular en autobús en el Área Metropolitana de Granada se presta de manera indirecta mediante contratos de servicio público, cuya titularidad corresponde a la Junta de Andalucía aunque, a través de concesiones que discurren íntegramente en la aglomeración urbana de Granada. La mayoría de las competencias han sido delegadas por la Consejería de Fomento y Vivienda al Consorcio de Transporte Metropolitano del Área de Granada.

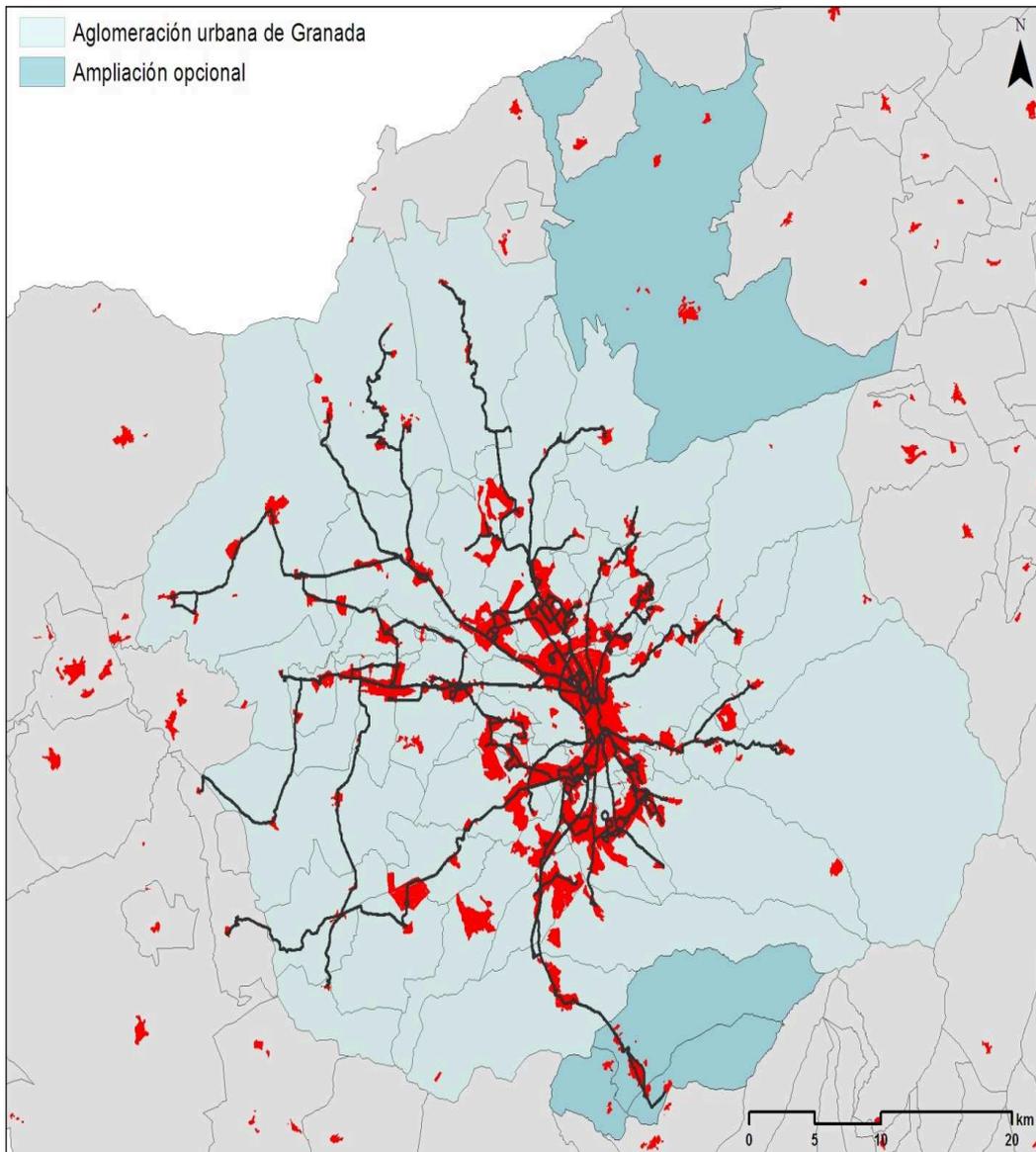


Figura 6. Esquema de red de servicios del transporte público metropolitano

Operador	Línea	Denominación
AUTOCARES FRAMAR, S.L.	175	Granada-Zubia (Laurel de la Reina)
	176	Granada-Zubia (Centro de Salud)
	177	Granada-Zubia (Campo de Fútbol)
	275	Granada-Zubia-Cumbres Verdes
AUTOCARES JOSÉ GONZÁLEZ, S.L.	245	Granada-Aeropuerto
	345	Granada-Chimeneas-Acula-Ventas de Huelma-Cacín
AUTOCARES MARCOS MUÑOZ, S.L.	117	Granada-Pantano Cubillas
	317	Granada-Pantano Cubillas -Iznalloz-D. Pérez-Montejicar-Guahortuna-Huelma
AUTOCARES MARTÍN CORRAL, S.L.	113	Granada - Monte Luz
	313	Granada-Monte Luz - Cjo. del Aire - Villas Blancas - El Chaparral - Calicasas
AUTOGRANADINA EMP. TORRES, S.L.	105	Granada-Pulianas
	305	Granada-Pulianas-Güejar-Nivar-Cogollos Vega
CARLOS FERNÁNDEZ DE LA TORRE, S.L.	100	Granada-Víznar
	101	Granada - Jun
	102	Granada - Jun - Alfacar
	103	Granada - Alfacar
EMPRESA LIÑAN, S.A.	180	Granada - Huétor Vega - Cájar - Bellavista - La Zubia
	181	Granada - Huétor Vega - Bº Monachil - Bellavista
	183	Granada - Huétor Vega - Cájar - Bº Monachil - Monachil
	184	Granada - Huétor Vega - Mundo Nuevo
	185	Granada - Huétor Vega - Las Chinas
	186	Granada - Cájar
	190	Granada - Pinos Genil
HEREDEROS DE GÓMEZ	390	Granada - Pinos Genil - Güejar Sierra
	395	Quéntar-Dúdar-Los Pinillos-Pinos Genil
TRANSPORTES ARANA, S.L.	170	Granada-Ogíjares-Lomalinda
	171	Granada-Ogíjares-Gojar-Dílar
	174	Circular Granada - Ogíjares - La Zubia - Granada
TRANVÍAS ELÉCTRICOS DE GRANADA, S.A.	110	Granada-Polígono Asegra-Peligros
	111	Granada-P. Asegra-La Joya-Peligros
	120	Granada-Maracena
	121	Granada-Maracena (por Ctra. de Jaén)
	122	Granada-Maracena-Albolote-Atarfe
	123	Granada-Polígono Juncaril-Albolote (por Ctra. de Jaén)
	125	Granada - Atarfe (por Ctra. Córdoba)
	155	Granada-Armilla
	156	Granada-Armilla-Churriana-Las Gabias
	156B	Granada-Armilla-Churriana-Las Gabias B
TRINIDAD NIEVAS, S.L.	158	Granada-Campus Ciencias de la Salud-Armilla
	300	Granada-El Fargue-Huétor Santillán-El Colmenar-Beas de Granada
TTES. ALSINA GRAELLS SUR, S.A.	140	Granada-Santa Fe
	150	Granada-Cúllar Vega-Vegas del Genil (Por Ambroz) - Híjar
	151	Granada-Cúllar Vega-Vegas del Genil (Por Belicena)
	152	Granada-Santa Fe-Vegas del Genil-Cúllar Vega
	160	Granada-Alhendín-Otura
	225	Granada-Pinos Puente
	226	Granada-Pinos Puente-Zujaira
	240	Granada-Santa Fe-Chauchina-Fte. Vaqueros
	241	Granada-Santa Fé-Chaucina-Cijuela
	242	Granada-Santa Fe-Chauchina-Cijuela-Lachar
	318	Granada - Colomera
	325	Granada-Pinos Puente-Zujaira-Ilora-Alomartes-Tocón-Brácana
	340	Granada-Santa Fe-Cijuela-Lachar-Trasmulas-Peñuelas-C.Tajarja-El Turro
	360	Granada-Alhendín-Padul-Marchena-Durcal-Nigüelas-Talará
UREÑA E HIJOS S.L.	335	Granada-Santa Fe-Fte. Vaqueros-Valderrubio-Escóznar-Obeilar
	336	Granada-Santa Fe-Chauchina-Fte.Vaqueros-Valderrubio-Obeilar
	535	Santa Fe-Pedro Ruíz-Fte.Vaqueros-Casa Real-Valderrubio

Tabla 4. Operadores y líneas de autobuses del transporte interurbano en el Área Metropolitana

En el sistema interurbano de transporte de la Aglomeración Urbana de Granada los servicios se establecen mediante dos haces radiales, uno norte y otro sur. Estos haces se concentran en dos zonas próximas al centro de la ciudad: el entorno de la Estación de Ferrocarril (Comedores Universitarios) y junto al extremo norte de la Gran Vía (Calles Capitán Moreno y Profesor Emilio Orozco, en los alrededores del edificio de la Normal) para el haz norte, y el del Palacio de Congresos, en el haz sur.

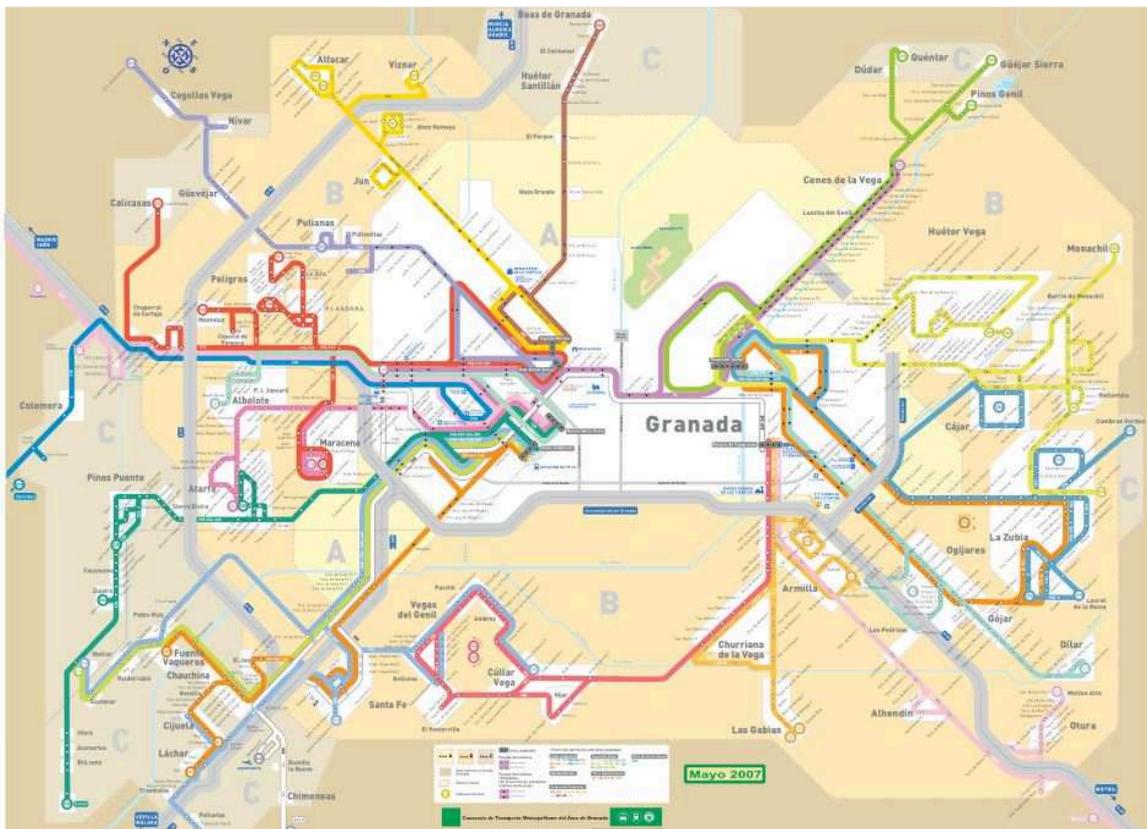


Figura 7. Líneas del sistema de transporte por autobús del Área de Granada

4.1.4. Evolución del sistema tarifario: la tarjeta de transporte

En diciembre de 2004 se implantó un sistema tarifario unificado para toda el Área de Granada mediante la distribución de una tarjeta de proximidad o sin contacto. Durante los tres años transcurridos desde la puesta en marcha de este sistema de tarifas se distribuyeron más de 71.500 tarjetas, con lo que uno de cada 7 habitantes del ámbito tarifario disponía en 2007 de una tarjeta de transporte del Consorcio, siendo válida para acceder a la totalidad de los servicios de transporte regular de viajeros, incluido los servicios urbanos de Granada, que discurren por el ámbito territorial del Área de Granada.

Esta tarjeta permite el acceso a todos los servicios interurbanos de cercanías de Granada, así como posibilidad de realizar transbordos a los autobuses urbanos de Granada, bonificando al viajero en estos movimientos mediante la aplicación de tarifas reducidas frente a la utilización de los tradicionales billetes univiaje. Desde el 1 de Junio de 2007 el uso recurrente de la tarjeta se bonifica con un 10% de descuento.

Igualmente, la mayoría de los municipios no pertenecientes al conjunto de los 32 que componen el Área y que están afectados por las líneas de cercanías han solicitado su adhesión al sistema a través de convenios con el Consorcio Metropolitano de Transportes, de modo que 12 municipios: Chimeneas, Quéntar, Dúdar, Güéjar Sierra, Calicasas, Nívar, Huétor Santillán, Íllora, Beas de Granada y Colomera, establecieron en 2007 este convenio y disfrutaban del mismo sistema tarifario que el resto del Área.

Estos resultados motivaron, en parte, la adopción o migración a este tipo de soporte mediante tarjeta sin contacto en otros Consorcios de Transportes de Andalucía como el de Málaga, Bahía de Cádiz, Sevilla y Campo de Gibraltar, así como en los sistemas de billeteaje de las futuras líneas de metro de Sevilla, Málaga y Granada (en aquel momento en fases de proyecto y/o construcción), constituyéndose a corto plazo como la tipología de tarjeta más aceptada en los servicios de transporte de Andalucía frente a otros formatos como el bono magnético o la tarjeta chip con contacto.

El sistema tarifario implantado ha sido claramente aceptado por los usuarios y agentes sociales. En una jornada laborable la tarjeta de transportes se utiliza para acceder a un autobús del Área de Granada en más de 27.000 ocasiones, 6.000 de ellas se efectúan en transbordos en los autobuses urbanos de la ciudad de Granada. La evolución del uso de la tarjeta en el servicio urbano de autobuses de Granada ha sido muy significativa desde el proceso de implantación del sistema (Tabla 5):

Año	Viajes con tarjeta en urbano de Granada
2005	920.000
2006	1.050.000
2007	1.300.000

Tabla 5. Evolución del uso de la tarjeta de transportes

En septiembre de 2007 se inició la puesta en funcionamiento de la interoperatividad entre las Áreas de Granada y Málaga, permitiendo a los poseedores de la tarjeta de transporte del Área de Granada el uso de cualquiera de los servicios adscritos al Consorcio de Transportes del Área de Málaga y viceversa. Posteriormente, los Consorcios de Bahía de Cádiz, Campo de Gibraltar y Sevilla han adaptado su sistema tarifario al de las Áreas de Granada y Málaga permitiendo la interoperatividad entre estas cinco Áreas de Andalucía.

4.2. Datos

La encuesta de la que provienen los datos utilizados en esta investigación fue realizada a usuarios del servicio de transporte público en autobús que opera en el área metropolitana de Granada (España). Este servicio consta de 13 compañías operadoras que conectan diferentes aglomeraciones urbanas del área metropolitana de Granada. El Consorcio de Transportes de Granada llevó a cabo entrevistas cara a cara en marzo de 2007, durante los cinco días de una semana, en las principales paradas de autobús, de modo que finalmente se recopilaron 858 encuestas.

Posteriormente se han realizado encuestas anuales, pero el año 2007 fue el único en el que se preguntó a los usuarios tanto por aspectos relacionados con la importancia asignada a las variables que definen el servicio recibido como por la percepción de la calidad del mismo. Este es el motivo por el que se han seleccionado los datos de la encuesta de ese año en lugar de los procedentes de encuestas de años más recientes.

El cuestionario se estructuró en dos secciones. El objetivo de la primera fue recopilar información relacionada con: información general del viaje (p.e., parada de autobús, línea, operador, origen, destino), características socioeconómicas de los pasajeros (p.e. sexo, edad, disponibilidad de vehículo privado) y hábitos de viaje (motivo del viaje, frecuencia, tipo de billete, modos complementarios usados por el usuario desde el origen hasta la parada de autobús y desde la parada de autobús hasta el lugar de destino).

La mayoría de los entrevistados fueron mujeres (Tabla 6). Más de la mitad tenían edades comprendidas entre los 18 y los 30 años, y sólo un 9,5% eran mayores de 60 años. La mayoría de los encuestados (61,1%) disponía de vehículo privado para hacer el viaje. Un 29,4% viajaba por motivos de trabajo y otro porcentaje similar por estudios, mientras que el resto de los encuestados lo hacía por otras razones, como visita al médico, compras, vacaciones, etc. La mayoría de ellos usaba el autobús con una frecuencia casi diaria, o con bastante frecuencia, mientras que sólo un 10% de la muestra encuestada lo usaba de forma esporádica. El modo de transporte complementario al autobús y que completaba sus trayectos era a pie. También indicaron otros modos que completaran el trayecto como el coche, el autobús urbano, la motocicleta, etc., pero su porcentaje de uso era mucho menor. En cuanto a la tipología de billete usado, el abono de transporte y el billete simple eran los más usados, representando entre ambos un porcentaje superior al 80%.

La segunda sección de la encuesta estaba más orientada a captar las opiniones de los usuarios sobre el servicio prestado, es decir, a obtener información sobre la percepción de calidad del servicio prestado. Para ello, entre otras variables, se definieron 12 atributos de calidad del servicio que lo caracterizaban, y una variable de valoración global del servicio. Cada atributo podía ser valorado por el usuario en una escala de valores del 0 al 10. Los atributos utilizados para caracterizar la calidad del servicio fueron: información, puntualidad, seguridad a bordo, amabilidad del conductor, limpieza del interior del autobús, espacio en el autobús, temperatura en el autobús, accesibilidad a/desde el autobús, tarifa, velocidad, frecuencia del servicio y proximidad a/desde el origen/destino.

VARIABLE	DISTRIBUCIÓN DE RESPUESTAS
Sexo	Hombre (33,0%), Mujer (67,0%)
Edad	18-30 (56,5%), 31-60 (34,1%); >60 años (9,5%)
Disponibilidad de vehículo privado	Sí (38,9%), No (61,1%)
Motivo del viaje	Trabajo (29,4%), Estudios (22,9%), Médico (14,2%), Compras (4,4%), Actividades personales (18,7%), Vacaciones (0,2%), Ocio (8,6%), Otros (1,5%)
Frecuencia	Casi diaria (67,9%), Frecuentemente (20,6%), Ocasionalmente (9,0%), Esporádicamente (2,5%)
Modos de transporte complementarios desde el origen hasta la parada de autobús	A pie (77,6%), Coche (1,9%), Autobús urbano (16,9%), Motocicleta (0,5%), Otros (3,1%)
Modos de transporte complementarios desde la parada de autobús hasta el destino	A pie (94,5%), Coche (2,1%), Autobús urbano (2,3%), Motocicleta (0,2%), Otros (0,9%)
Tipo de billete	Abono de transporte (49,6%), Billete sencillo (41,2%), tarjeta de la tercera edad (4,8%), Otros (4,4%)

Tabla 6. Características de la muestra encuestada

La Tabla 7 muestra los ratios de importancia media calculados a partir de los datos obtenidos. En esta tabla se reflejan las dos perspectivas de la encuesta: por una parte, el ratio de importancia media refleja cómo de importantes consideran los usuarios cada uno de los atributos o variables definidos, mientras que por otra, mediante el ratio de percepción media se obtiene un valor de la calidad que ese atributo o variable ofrece según el punto de vista del usuario. Puede observarse que hay una variación muy pequeña en cuanto a los ratios de importancia asignados por los pasajeros en la encuesta, lo que implica que todos los atributos son considerados como importantes. El valor medio de importancia está comprendido en el rango de valores de 8,5 a 9,5. Por lo tanto, la importancia es uniforme y prácticamente igual en todos los atributos, y ésta es una de las principales limitaciones que se plantean cuando se estudia la importancia de las variables de calidad del servicio basadas en la opinión de los pasajeros (Weinstein, 2000; de Oña et al., 2012).

Por el contrario, los ratios de percepción media muestran mayores diferencias entre atributos. En todos los casos se los valores son inferiores a los obtenidos en el cálculo de ratios de importancia media, concentrándose en valores comprendidos entre 6 y 8. Estos valores pueden considerarse bastante buenos porque todos los atributos son percibidos con al menos una calidad adecuada (>6), y algunos de ellos con una calidad bastante buena (>7). Los atributos que recibieron mejor valoración fueron Amabilidad del conductor, Seguridad a Bordo y Temperatura en el autobús, y el peor valorado fue la Tarifa. En general, los usuarios mostraron estar bastante satisfechos con el servicio, puesto que la evaluación de la valoración global media fue de 7,10.

ATRIBUTO	RATIO DE IMPORTANCIA MEDIA	RATIO DE PERCEPCIÓN MEDIA
(INF) Información	8,62	6,86
(PUN) Puntualidad	9,14	7,41
(SEG) Seguridad a bordo	8,98	7,73
(TRA) Amabilidad del conductor	8,77	7,96
(LIM) Limpieza del interior del autobús	8,86	7,46
(ESP) Espacio en el autobús	8,66	7,21
(TEM) Temperatura en el autobús	8,72	7,43
(ACC) Accesibilidad a/desde el autobús	8,91	6,90
(PRE) Tarifa	8,77	6,44
(VEL) Velocidad	8,73	7,30
(FRE) Frecuencia	9,05	6,99
(PRO) Proximidad desde/a origen/destino	8,71	7,43
(SQ) Calidad global del servicio		7,10

Tabla 7. Ratios de importancia y de percepción media de los atributos de calidad del servicio

En el **Apéndice 1** se incluye el formulario de encuesta elaborado para los trabajos de toma de datos.

CAPÍTULO 5. MATERIALES Y MÉTODOS

CAPÍTULO 5. MATERIALES Y MÉTODOS

En este capítulo se exponen las principales técnicas, métodos y herramientas de cálculo aplicados para conseguir los objetivos descritos en el capítulo 3 (sección 5.1), así como la metodología desarrollada en este estudio (5.2).

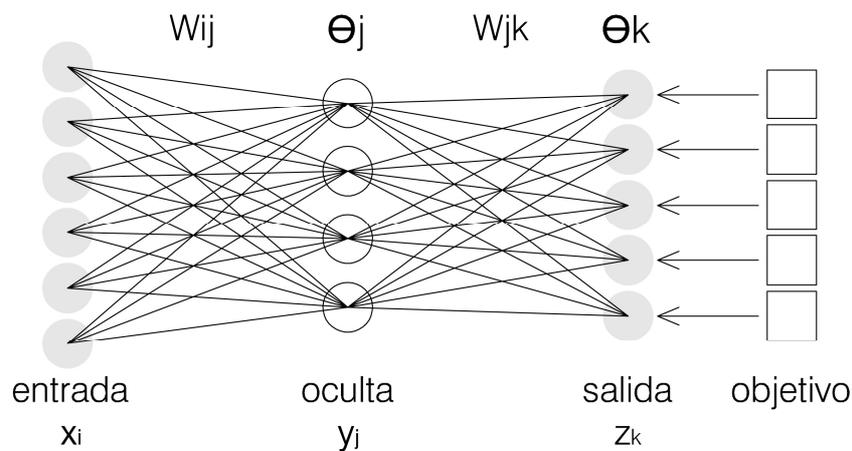
5.1. Métodos

En este apartado se describe la técnica de redes neuronales artificiales (ANN), que ha sido aplicada en esta investigación a la calidad del servicio en el transporte público (apartado 5.1.1), indicándose en él la división y tratamiento que se le da a los datos (5.1.1.1), los criterios utilizados para la parada de la fase de entrenamiento (5.1.1.2) y el método utilizado para medir la bondad o rendimiento de cada ANN entrenada (5.1.1.3). El apartado 5.1.2 describe los cuatro métodos que se han utilizado en esta investigación para calcular la importancia relativa de las variables que definen los atributos de calidad del servicio. También se ha descrito la técnica de los árboles de decisión (5.1.3), ya que ésta ha sido aplicada en la segunda fase de la investigación, descrita en el apartado 5.2.

5.1.1. El Perceptrón Multicapa (MLP)

En este estudio se ha utilizado la tipología de ANN denominada Perceptrón Multicapa MLP (Multilayer Perceptron) por dos motivos fundamentales: en primer lugar, diversos grupos de investigación (Funahashi 1989; Hornik et al., 1989) demostraron matemáticamente por separado que una ANN tipo MLP con una capa oculta es un aproximador universal de funciones, y en segundo lugar, porque ha sido usada aproximadamente en el 70% de los estudios realizados aplicando ANN, en numerosos y diversos campos de investigación, habiéndose obtenido buenos resultados.

El MLP es una tipología de ANN cuya arquitectura tiene una o varias capas ocultas (Figura 8) interconectadas entre sí por medio de pesos sinápticos, a través de los que fluye la información. Su arquitectura genérica puede grafarse según el siguiente esquema:



Fuente: Martín del Brío y Sanz, 2006

Figura 8. Arquitectura de una red neuronal artificial tipo MLP

Siendo x_i las entradas a la red, y_j las salidas de la capa oculta y z_k las salidas de la capa de salida. w_{ij} son los pesos sinápticos que conectan la capa de entrada con la capa oculta, y w_{jk} los pesos sinápticos que conectan la capa oculta con la capa de salida. El valor objetivo o salida que se desea obtener tras introducir cada patrón es denominada t . El parámetro θ_j es el bias de la capa oculta, y θ_k el bias de la capa de salida.

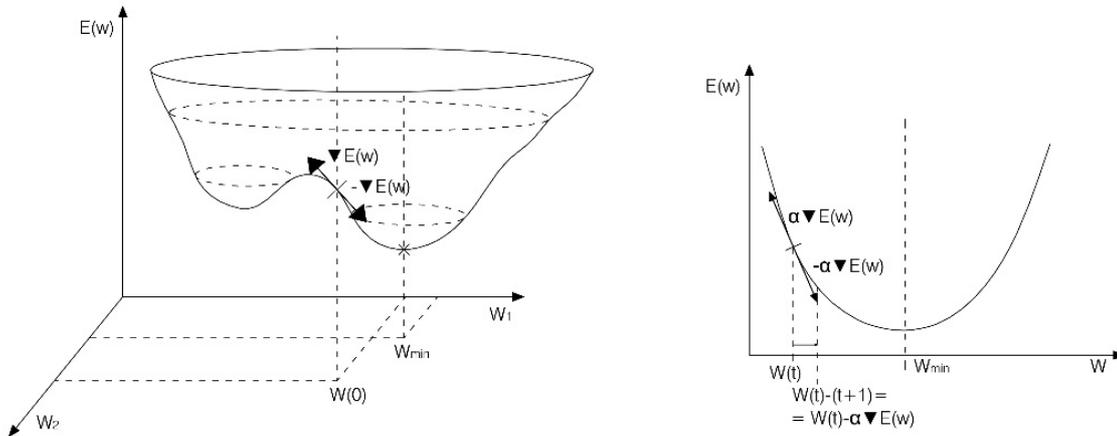
La función f representa a las funciones de transferencia de la capa oculta y de la capa de salida, de tipo sigmoideo, cuya función es transformar los valores de llegada de la capa de entrada y de la capa oculta en el intervalo $[0,1]$, según la siguiente formulación:

$$f(x) = \frac{1}{1+e^{-x}} \quad (13)$$

La regla de aprendizaje por excelencia del MLP es el algoritmo back-propagation (BP), primero introducida por Werbos (1974) y posteriormente desarrollado por Rumelhart y McClelland (1986a; 1986b). Es la regla de aprendizaje más popular (Azadeh et al., 2011), y además es la más utilizada y de uso más extendido en numerosas investigaciones previas (Gedeon, 1995).

La regla de aprendizaje BP se basa en la existencia de una función error $E(w)$ representada mediante una superficie de crestas y valles, dependiente de los pesos sinápticos w_{ij} y w_{jk} , que proporciona un valor del error actual de la red, como diferencia entre los valores obtenidos por la red en esa iteración y los valores objetivo o esperados. De forma iterativa, la función de error desciende a lo largo de la línea de máxima pendiente de esta superficie de error (Figura 9), modificando el valor de sus pesos sinápticos hasta que alcanza un mínimo, que puede ser local o global. La configuración de pesos finalmente adoptada es aquella que se corresponde con el mínimo alcanzado en la superficie de error. Así, en cada iteración:

$$W(t+1) = W(t) - \alpha \cdot \Delta E(W) \quad (14)$$



Fuente: Martín del Brío y Sanz, 2006

Figura 9. Algoritmo de aprendizaje de descenso por gradiente

Se denomina back-propagation porque funciona calculando los errores hacia atrás, de manera que en primer lugar se calcula el valor de la denominada señal de error, que es una medida del error proporcional al error de salida actual de la red, con el que se calcula la actualización de los pesos de w_{jk} . Seguidamente, se propagan hacia atrás los errores calculados en la capa de salida a través de las sinapsis, proporcionando las señales de error correspondientes a las sinapsis de la capa oculta; por último con éstas se calcula la actualización de los pesos sinápticos w_{ij} .

Uno de los inconvenientes achacado al MLP con aprendizaje tipo BP es la lentitud del algoritmo en la convergencia hacia un mínimo en la superficie de error (Martín del Brío y Sanz, 2006). Esto es debido a que la curvatura de la superficie de error varía mucho en algunas regiones del espacio, de modo que cuando se avanza en el algoritmo de aprendizaje modificando a un ritmo constante los pesos sinápticos en cada iteración, los valores de actualización de los pesos, y por tanto la trayectoria de descenso hacia un mínimo local, pueden verse ralentizados muchísimo en zonas casi planas en las que el gradiente es cercano a cero (Hagan et al., 1996). Así, es típico en el algoritmo de descenso por gradiente encontrar periodos prolongados de tiempo en los que la trayectoria hacia el mínimo local apenas avanza, mientras que cuando llega a una zona con un gradiente elevado, lo hace a buen ritmo. Existen dos modificaciones heurísticas del aprendizaje BP que aceleran notablemente el periodo de descenso hacia el mínimo local y la superficie de error: el método del momentum y el del ratio de aprendizaje.

El parámetro momentum es un factor constante que multiplica los valores de los pesos sinápticos durante el entrenamiento, y que tiene la función de mitigar las oscilaciones seguidas por la trayectoria de descenso a lo largo de la superficie de error, mejorando la estabilidad del método BP y acelerando la convergencia hacia el mínimo local, ya que ayuda a mantener la trayectoria en una misma dirección. Esto lo consigue modificando el

ritmo de aprendizaje fijado en función del signo de las actualizaciones de los pesos. Por su parte, el parámetro ratio de aprendizaje es un factor de valor también constante que controla el tamaño del cambio de los pesos. De este modo, si los incrementos en un determinado peso tienen siempre el mismo signo (lo que indica que la trayectoria de descenso va avanzando rápido a lo largo de la superficie de error porque el gradiente es elevado), las actualizaciones en cada iteración serán mayores; sin embargo, si los incrementos en cierto peso oscilan (a veces son positivos y a veces negativos), el incremento efectivo o acumulado se reduce al cancelarse. De este modo, en zonas estrechas y profundas de la superficie de error con forma de valle angosto, los pesos correspondientes a las dimensiones estrechas sufren incrementos pequeños, mientras que los de las direcciones que descienden directamente al fondo se ven potenciados (Bishop, 1994).

5.1.1.1. Conjuntos de entrenamiento, validación y test

Cuando se diseña una red neuronal tipo MLP, los datos de entrada suelen dividirse en tres grupos: aprendizaje, validación y test (Bishop, 1995; Haykin, 1999).

Durante el aprendizaje los pesos de la red se modifican de forma iterativa con objeto de minimizar el error cometido entre la salida obtenida por la red y la salida deseada u objetivo. Para cada patrón que se presenta a la red, los pesos se van ajustando en función de la diferencia entre la salida obtenida y la esperada u objetivo.

El problema que puede presentar este entrenamiento, sobre todo cuando el número de parámetros (pesos y bías) de la red es elevado, es que el modelo aprenda particularidades irrelevantes de los patrones de entrada que le son presentados, de modo que, si bien disminuye el error durante la fase de entrenamiento, posteriormente presenta errores de generalización importantes. Es decir, cuando a la red entrenada le son introducidos patrones nuevos diferentes a los que se le presentaron durante el entrenamiento, ésta tiende a cometer errores de clasificación o predicción importantes en las salidas que devuelve. A este efecto se le llama problema de sobreajuste. Para evitarlo, se utiliza el conjunto de validación, compuesto por un conjunto de datos diferente a los del entrenamiento, que controlan el proceso de aprendizaje. Con la introducción de este conjunto, durante el aprendizaje la red va modificando los pesos en función de los datos de entrenamiento, y de forma alternada se va obteniendo el error que comete la red ante los datos de validación.

El error de entrenamiento y el error de validación van disminuyendo a medida que aumenta el número de iteraciones, hasta alcanzar un mínimo en la superficie del error, momento en el que se para el aprendizaje de la red (Figura 10).

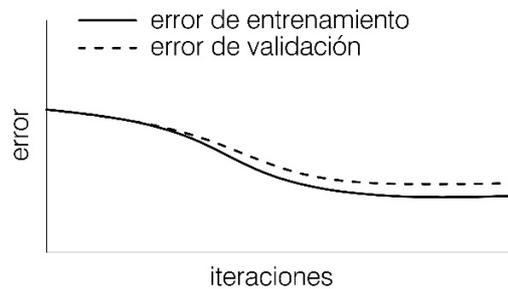


Figura 10. Evolución de los errores de entrenamiento y validación durante el entrenamiento de un MLP

Por último, para medir de forma completamente objetiva la eficacia de la ANN entrenada, no es objetivo basarse en el error cometido por la red que ha sido calculado con los datos de validación, ya que estos datos han participado en la fase de entrenamiento. Se debe contar con un tercer grupo de test, que proporcionará una estimación del error de generalización.

5.1.1.2. Criterios de parada

Normalmente se establecen simultáneamente varios criterios de parada, de modo que la ANN da por finalizado su entrenamiento cuando alcanza el primero de ellos. Estos criterios son:

- N° de iteraciones o epochs.
- Valores de error y porcentaje de acierto en los grupos de entrenamiento y validación.

Si durante el entrenamiento el error o el porcentaje de acierto en los grupos de entrenamiento y validación alcanza los valores marcados, el entrenamiento se detiene; si no es así, continúa hasta que agota el número de iteraciones.

5.1.1.3. Medida del error de generalización

El error de generalización, que define la exactitud que puede alcanzar la ANN entrenada, es decir, la capacidad de clasificar adecuadamente nuevos casos que se le presenten, se mide a través del Porcentaje de Error Medio Absoluto (MAPE) (Delen et al., 2006), por medio de la fórmula:

$$MAPE = \frac{1}{T} \cdot \sum_{i=1}^T \text{abs}\left(\frac{\text{Valor actual } i - \text{Valor objetivo } i}{\text{Valor objetivo } i}\right) \quad (15)$$

Es decir, para un conjunto de datos T del grupo test, este índice mide la media del error, en valor absoluto, entre los valores obtenidos tras introducir en la red entrenada un caso i , y su valor esperado. El valor del MAPE siempre estará comprendido en el intervalo $[0,1]$, y cuanto menor sea su valor, mejor será la capacidad de generalización de la red y, por tanto, su rendimiento.

5.1.2. La importancia relativa de las variables

Los métodos seleccionados en este estudio para determinar la importancia relativa de las variables en un modelo de ANN, han sido propuestos y aplicados por numerosos autores en diversos campos de investigación (Olden y Jackson, 2002; Gevrey et al., 2003). Éstos han sido: método Perturb, Profile, Connection Weights (CW) y Derivadas Parciales (PaD).

5.1.2.1. Método Perturb

Este método se basa en el principio de perturbar o introducir ruido a una de las entradas mientras que las restantes variables mantienen sus valores originales. Posteriormente, se compara el error cuadrático medio (MSE) entre las salidas obtenidas antes y después de la perturbación (Yao et al., 1998; Scardy y Harding, 1999).

Un ruido δ es progresivamente aplicado a cada variable en cinco pasos: 10%, 20%, 30%, 40% y 50% de su valor original. Así, la variable x_i cambia sus valores a $x_i = x_i + \delta$. Cuanto mayor sea el error medido al introducir las perturbaciones, mayor será la importancia relativa de la variable en ese modelo de red.

5.1.2.2. Método Profile

Este método analiza la evolución de cada variable de entrada a lo largo de un rango de valores, mientras las variables restantes se quedan bloqueadas en unos valores fijados (Lek et al., 1995; 1996a; 1996b).

Cada variable predictora x_i adopta 11 valores diferentes que son resultado de la división del rango comprendido entre sus valores máximo y mínimo en diez intervalos iguales. Seguidamente, todas las variables excepto una fijan inicialmente sus valores en su mínimo, y sucesivamente en su primer cuartil, mediana, tercer cuartil y valor máximo. En consecuencia, se obtienen 5 valores de la variable respuesta para cada uno de los 11 valores adoptados por x_i , y se calcula la mediana de esos 5 valores. Finalmente, para cada variable se representa una curva con el perfil de variación. Cuanto mayor sea la variación relativa de los valores representados en el perfil para cada variable, mayor será la importancia relativa de esa variable en el modelo estudiado.

5.1.2.3. Método Connection Weights (CW)

Este método determina la importancia relativa de las variables predictoras del modelo en función de los pesos sinápticos de la ANN, según la expresión matemática (Olden y Jackson, 2002):

$$R_{ik} = \sum_{j=1}^L w_{ij} \cdot w_{jk} \quad (16)$$

Siendo R_{ij} el valor de la importancia relativa de la variable predictora x_i respecto a la variable de la capa de salida k , w_{ij} el peso sináptico que conecta la neurona i de la capa de entrada con la neurona j de la capa oculta, y w_{jk} el peso sináptico que conecta la neurona j anterior con la neurona k de la capa de salida.

5.1.2.4. Método de las Derivadas Parciales (PaD)

Este método analiza los efectos de primer orden de las variables predictoras del modelo respecto a la variable salida, utilizando para ello el conjunto de datos de entrenamiento (Dimopoulos et al., 1995). La salida proporcionada por una neurona de la capa oculta de una ANN tipo MLP con funciones de activación tipo sigmoideo sigue la ecuación:

$$h_j = \frac{1}{(1+e^{-net_j})} \quad (17)$$

$$net_j = \sum_i w_{ij} \cdot x_i - \theta_j \quad (18)$$

donde h_j es la salida de la neurona j de la capa oculta, x_i es el valor de la variable predictora considerada en la capa de entrada, w_{ij} es el peso sináptico que conecta la variables predictora x_i y la neurona j de la capa oculta, y θ_k es el bias de la neurona j de la capa oculta.

La salida de una neurona k de la capa de salida vendrá dada por las siguientes expresiones

$$y_k = \frac{1}{(1+e^{-net_k})} \quad (19)$$

$$net_k = \sum_k w_{jk} \cdot h_j - \theta_k \quad (20)$$

donde w_{jk} es el valor del peso sináptico entre la neurona j de la capa oculta y la neurona k de la capa de salida, θ_k es el bias de la neurona k de la capa de salida e y_k es la salida que arroja la neurona k de la capa de salida.

La expresión que relaciona la variación de los valores de salida y_k respecto a la variación de la variable predictora x_i se obtiene con la aplicación de la regla de la cadena:

$$\frac{dy_k}{dx_i} = \sum_j \frac{dy_k}{dnet_k} \cdot \frac{dnet_k}{dh_j} \cdot \frac{dh_j}{dnet_j} \cdot \frac{dnet_j}{dx_i} = \sum_j f'_k \cdot w_{jk} \cdot f'_j \cdot w_{ij} \quad (21)$$

$$f'_k = y_k \cdot (1 - y_k) \quad (22)$$

$$f'_j = h_j \cdot (1 - h_j) \quad (23)$$

El valor de la sensibilidad de cada variable x_i viene dado por la expresión

$$L_i = \frac{\sum_P \frac{dy_k}{dx_i}}{P} \quad (24)$$

siendo P el número total de casos de entrenamiento.

Cuanto mayor sea el valor de L_i , mayor será la importancia relativa de la variable x_i en el modelo de ANN entrenado.

5.1.3. Los árboles de decisión (DT)

Los árboles de decisión (DT) son una técnica de minería de datos utilizada para la clasificación y predicción de una variable objetivo. Dependiendo de la naturaleza de la variable, pueden definirse dos modelos diferentes de DT: si la variable objetivo es discreta, se construye un árbol de decisión cuya variable salida a predecir es discreta, mientras que si la variable es continua, se genera un árbol de decisión que predice una cantidad numérica.

Existen muchos algoritmos diferentes para generar estos modelos, y la principal diferencia entre ellos reside en el criterio de partición utilizado para el crecimiento del árbol. El desarrollo de un árbol de decisión se caracteriza por la definición de los siguientes pasos (Montella et al., 2012): (a) un criterio de partición para definir la función óptima que realiza la mejor partición de los objetos en grupos homogéneos; (b) la regla de parada para detener el crecimiento del árbol; y (c) la regla asignada para identificar cada clase o valor con un nodo terminal.

El algoritmo CART (Breiman et al., 1984) ha sido utilizado para propósitos similares y ha proporcionado buenos resultados en otros trabajos desarrollados previamente (de Oña et al., 2012; 2014; 2015), por eso será el aplicado en este estudio. Es decir, se aplica un árbol de decisión a una base de datos que tiene como objetivo predecir la evaluación de la calidad del servicio percibida por un usuario a través de una variable continua, en una escala de valores numérica. Para esquematizar la secuencia seguida en el entrenamiento, se ha incluido el diagrama de flujo de la Figura 11 en el que se indica en 15 pasos (de DT01 a DT15) el procedimiento aplicado.

En primer lugar, la base de datos es dividida aleatoriamente en M subgrupos, conteniendo cada uno de ellos una porción $(M-1)/M$ del total de la muestra (paso DT01). Se construye un primer DT para el primer subgrupo m (paso DT02), usando como conjunto de entrenamiento los $(M-1)/M$ datos que contiene, y como conjunto de test los $1/M$ datos restantes. Este procedimiento es extensamente conocido como técnica de validación cruzada de m -capas (Witten y Frank, 2005). Este DT se desarrolla considerando i variables predictoras (que son las variables que definen los atributos que caracterizan la calidad del servicio), y teniendo en cuenta los siguientes aspectos (paso DT03):

El criterio de partición usado para evaluar el conjunto de reglas de división del árbol "candidatas" se basa en el criterio de error de mínimos cuadrados (LS). La función de

error LS mide la impureza de un nodo, de modo que la riqueza de una división se evalúa en función de la disminución del valor de impureza medido en un nodo padre según el criterio de la función LS. El algoritmo CART desarrolla todas las posibles particiones para cada una de las variables independientes consideradas, y selecciona aquella regla que más reduce la impureza del nodo padre. Esta impureza puede medirse según la siguiente fórmula (Yohannes y Webb, 1999):

$$E(t) = \frac{1}{N} \cdot \sum_{i=1}^{N_t} (y_{i(t)} - \bar{y}_{(t)})^2 \quad (25)$$

donde $E(t)$ es la función de impureza de un nodo t , $y_{i(t)}$ son los valores de la variable independiente del nodo t , $\bar{y}_{(t)}$ es el valor medio de la variable objetivo en el nodo t y N_t es el número de casos en el nodo t .

En el proceso de crecimiento se aplican dos reglas de parada: a) el mejor criterio de partición de entre todos los considerados proporciona un error no mayor de 0,0001; y b) el número de casos en uno o más nodos hijos es menor que el 1% del tamaño total de la muestra.

El valor asignado como referencia a cada nodo terminal es la media de los valores de la variable objetivo en dicho nodo.

La varianza de los datos explicada por el modelo se calcula a partir del conjunto de datos test, y se obtiene a partir del error cuadrático medio de los nodos terminales del modelo de DT construido (paso DT04). Es decir, la mejora que una variable i produce cuando es usada como principal divisor o divisor sustituto es añadida a cada partición del DT, midiéndose el número de casos afectados por esta mejora (paso DT06). La importancia de la variable i es almacenada (paso DT07), y este procedimiento se repite (de DT06 a DT07) hasta que i alcanza I (el valor obtenido alcanza al objetivo) (paso DT11). La importancia de las I variables es almacenada para el subconjunto m (paso DT10). Todo el procedimiento se repite para cada uno de los subconjuntos m (de DT03 a DT10), y finalmente la capacidad predictiva del modelo de DT (paso DT13) se mide como el valor medio de la varianza explicada para cada uno de los M modelos creados y cuyos resultados han sido almacenados previamente (en el paso DT04). Del mismo modo, la importancia media y la desviación estándar de cada variable se calculan (paso DT14) a partir de los M valores de importancia que se han ido almacenando en cada uno de los M modelos de DT generados con cada uno de los subconjuntos de datos m (almacenados en DT10). El ranking de importancia relativa de cada variable se calcula siguiendo el criterio de que cuanto mayor sea el valor medio para cada variable, mayor será la importancia relativa global en el modelo de DT generado (paso DT15).

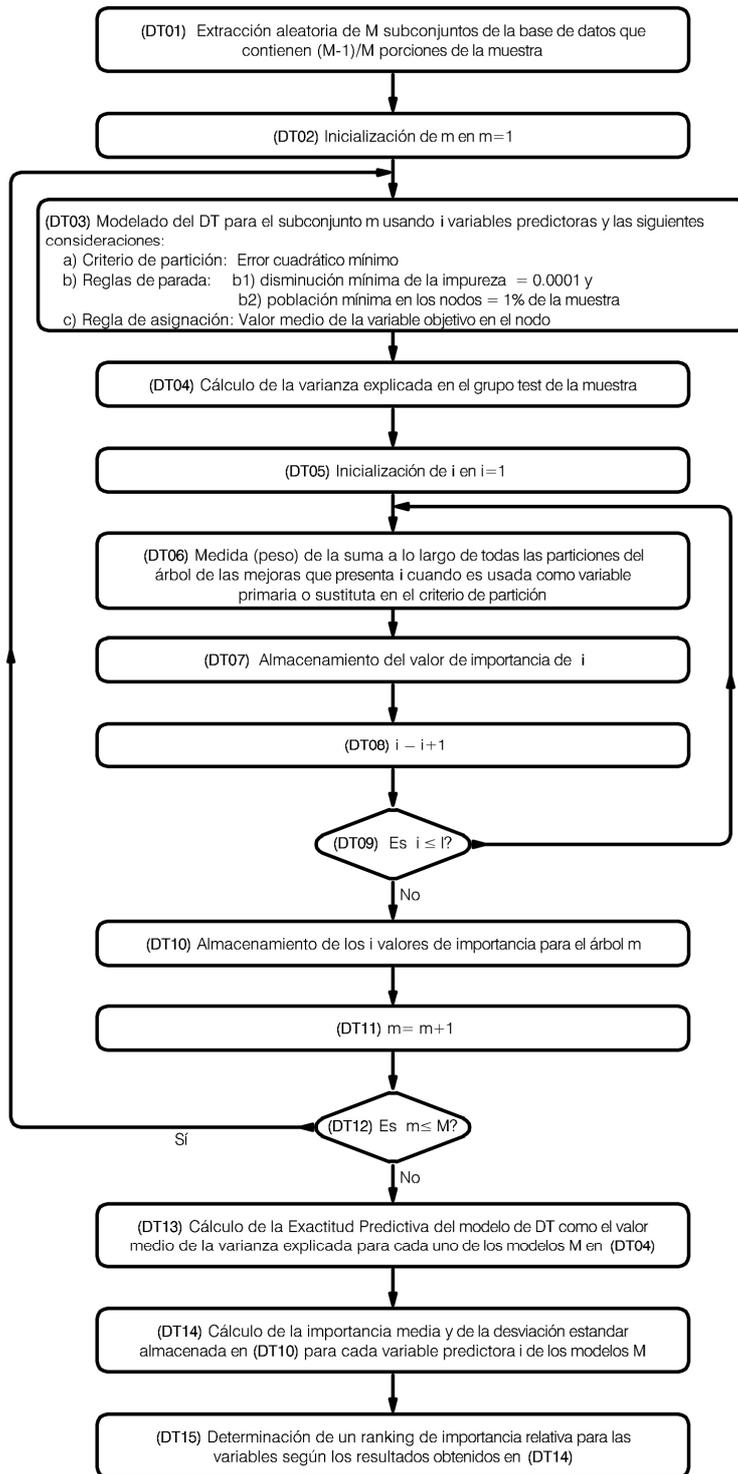


Figura 11. Esquema de entrenamiento de DT y de cálculo de la importancia relativa de las variables de estudio

5.2. Metodología seguida

Este trabajo de investigación se divide en dos fases. En una primera fase (5.2.1) se analiza la calidad del servicio del transporte público por autobús utilizando redes neuronales artificiales (ANN), desarrollando una nueva metodología para determinar la importancia relativa de los atributos considerados en el estudio. Al final de esta fase también se ha analizado la robustez y estabilidad de esta nueva metodología. En la segunda fase (5.2.2), se realiza una comparación entre las técnicas de redes neuronales artificiales (ANN) y los árboles de decisión (DT), evaluando las ventajas e inconvenientes de aplicación de cada una de ellas. Así, las dos fases de análisis siguen los siguientes pasos:

5.2.1. Fase 1. Análisis de la calidad del servicio en transporte público por autobús mediante ANN

En esta primera fase de la investigación se ha trabajado con conjuntos de ANN en lugar de con una única. Cada conjunto de ANN está compuesto por una serie de ANN con la misma arquitectura, que son entrenadas utilizando idénticos algoritmos de aprendizaje, con una partición aleatoria de los datos para cada entrenamiento en conjuntos de entrenamiento, validación y test en una proporción 70/15/15, funciones de activación (tipo logaritmo sigmoideo), valores de momentum de 0,9 y ratio de aprendizaje de 0,1 (Hagan et al., 1996), y número de iteraciones (20.000). La única diferencia entre cada una de las ANN incluidas en el mismo conjunto reside en los valores iniciales de los pesos con los que comienza su entrenamiento, ya que en cada una de ellas estos valores se obtienen de forma aleatoria. Una vez que cada una de las ANN ha sido entrenada, se aplican los cuatro métodos (CW, Perturb, Profile y PaD) descritos en el Capítulo 4, obteniéndose un ranking de importancia relativa de las variables consideradas por el modelo para cada uno de los métodos. Debido a la inestabilidad de los resultados de importancia relativa obtenidos cuando se aplica un único método a una única red (Cao y Qiao, 2008), esta investigación ha desarrollado un procedimiento basado en el cálculo del ranking de importancia relativa para cada método en función de los valores de importancia media obtenidos a partir de cada conjunto de ANN.

El software utilizado ha sido MATLAB (Beale et al., 2007).

La secuencia de pasos seguidos para desarrollar esta metodología ha sido:

1. Entrenamiento de cada una de las H arquitecturas de ANN consideradas, con 12 neuronas en la capa de entrada (una por cada atributo de calidad del servicio considerado), $H \in [1, N]$ neuronas en la capa oculta, y 1 neurona en la capa de salida (la calidad global del servicio) un número M de veces, y usando diferentes valores aleatorios para los pesos iniciales. Así, se obtienen $N \times M$ ANN entrenadas. En esta investigación se consideraron unos valores $N=30$ y $M=50$, resultando un total de 1.500 ANN entrenadas.

2. Cálculo de la exactitud o capacidad de generalización de las NxM ANN entrenadas, a través de la fórmula del MAPE, descrita en el apartado 5.1.1.3 del presente capítulo.
3. Cálculo del valor mínimo del MAPE para cada una de las arquitecturas (o conjuntos) de ANN, y selección del conjunto que obtiene el mínimo valor global de MAPE.
4. Aplicación de los cuatro métodos de cálculo de importancia relativa a cada una de las M ANN de la arquitectura (conjunto) seleccionado en el paso anterior (paso 3).
5. Cálculo del valor medio de importancia relativa para cada variable de cada una de las M ANN pertenecientes a la arquitectura considerada.
6. Determinación del ranking de importancia de las 12 variables consideradas en el estudio, en función de los valores obtenidos en el paso anterior para cada uno de los métodos aplicados.
7. Comparación de los resultados obtenidos según cada método.
8. Análisis estadístico de los resultados de importancia relativa obtenidos por cada método. Se han utilizado test no paramétricos, debido a que no se cumple el requisito de normalidad de los datos. Así, se ha aplicado el test Kruskal-Wallis para contrastar la hipótesis de la existencia de diferencias significativas en los ratios de importancia obtenidos para los atributos utilizados para definir la calidad del servicio, y el método de comparación múltiple de Dunn para identificar qué atributos son los responsables de estas diferencias significativas. Se han aplicado tres veces estos test no paramétricos, uno por cada uno de los cuatro métodos de contribución relativa. Así, han sido identificados diferentes grupos homogéneos con diferencias estadísticas entre ellos.
9. La metodología propuesta hasta este punto seleccionaba el conjunto de ANN cuyo valor global de MAPE es más bajo, pero para llegar a esta selección, es necesario realizar un proceso tedioso de entrenamiento de grupos de ANN con un número variable de neuronas en su capa oculta (en este caso $H \in [1,30]$). Es por ello que en esta investigación se ha querido comprobar en el presente punto de esta metodología si realmente la elección del conjunto H con un valor mínimo de MAPE global afecta a la determinación del ranking de importancia relativa de las variables, o si por el contrario, esta nueva metodología presenta la robustez y estabilidad suficientes como para obviar la selección del grupo de ANN con mejor comportamiento global, y es posible obtener resultados de importancia relativa similares sin necesidad de desarrollar el paso anterior. Así, para comprobar la robustez y la estabilidad de la metodología propuesta, se ha repetido la metodología descrita en los pasos 3 y 7 a todos y cada uno de los H conjuntos de ANN entrenados, determinando los valores de importancia relativa que se obtienen tras aplicar los cuatro métodos Perturb, Profile, CW y PaD a cada uno de los H conjuntos de ANN, con el fin de identificar si existen diferencias relevantes en cuanto al ranking de importancia relativa de las variables obtenido.

5.2.2. Fase 2. Comparación entre ANN y DT

En una segunda fase se compara la técnica de ANN aplicada en la fase 1 con los resultados obtenidos tras aplicar la técnica de los árboles de decisión (DT). Esto permite, no sólo observar diferencias en cuanto a exactitud de cada una de las técnicas, sino también analizar otros aspectos como fortalezas y debilidades de cada una de ellas, tiempo de cálculo requerido, complejidad computacional, estabilidad de resultados, etc.

CAPÍTULO 6. RESULTADOS Y DISCUSIÓN

CAPÍTULO 6. RESULTADOS Y DISCUSIÓN

En este capítulo se exponen los principales resultados obtenidos tras la aplicación de la técnica de ANN a la base de datos de calidad del servicio del transporte público de autobús descrita en el apartado 4.2. Se definen diferentes apartados, y en cada uno de ellos se pretende conseguir uno o varios objetivos específicos.

El apartado 6.1 analiza la capacidad de predicción y la exactitud de los modelos de ANN entrenados, con objeto de conocer el grado de validez de esta técnica (medido a través de su error de generalización).

El apartado 6.2 aborda el problema de la inestabilidad inherente en los modelos de cálculo de la importancia relativa de las variables predictoras, que se aplican a las ANN entrenadas, y expone los resultados extraídos tras aplicar la metodología propuesta en esta investigación, que pretende mitigar esta inestabilidad.

El apartado 6.3 analiza, mediante técnicas estadísticas, los resultados obtenidos en el apartado anterior con objeto de determinar si existen grupos homogéneos de variables que presenten diferencias estadísticamente significativas, y la concordancia de estos resultados con los obtenidos en otros estudios previos en los que se han aplicado otras técnicas diferentes.

El apartado 6.4 profundiza en la robustez de la metodología de cálculo de importancia relativa de las variables predictoras para los modelos de ANN desarrollado en el apartado 6.2.

Finalmente, el apartado 6.5 compara la técnica de ANN con la de DT, siendo ambas técnicas de minería de datos. No sólo se comparan los resultados obtenidos tras aplicar ambas, sino también se analizan las fortalezas y debilidades que presentan cada una de ellas.

6.1. Capacidad de generalización de las ANN entrenadas

Cada una de las $N=30$ arquitecturas de ANN fue entrenada independientemente un número $M=50$ veces, resultando un total de 1.500 ANN entrenadas. Su arquitectura queda reflejada en la Figura 12. Todas las variables predictoras presentaban unos valores de respuesta comprendidos en el intervalo $[0,10]$. Para hacer coincidir este intervalo de valores con el de las funciones de activación de tipo logarítmico sigmoideo, acelerando así la convergencia del entrenamiento (Masters, 1993; Delen et al., 2006), se consideraron directamente en un rango de valores comprendido en el intervalo $[0,1]$.

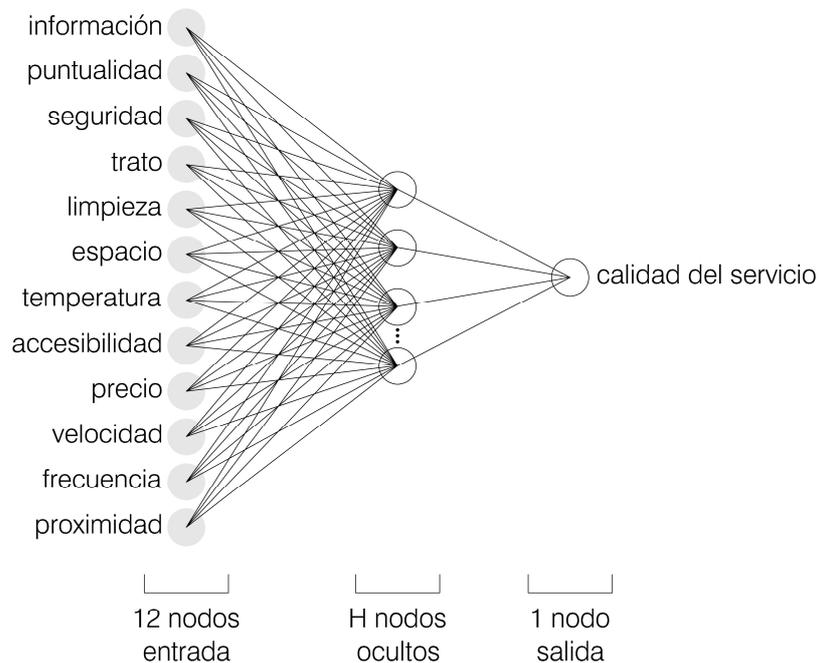


Figura 12. Arquitecturas de ANNs

Para medir su exactitud y capacidad de generalización, se calculó el valor del MAPE para cada una de ellas, obteniéndose el rango de valores para cada uno de los conjuntos que se recogen en la Tabla 8.

Todas las ANN entrenadas han alcanzado una exactitud muy elevada, superior al 90% en todos los casos. Estos resultados concuerdan con los obtenidos por otros autores que han aplicado la técnica de las ANN para evaluar la contribución relativa de las variables que definen la calidad del servicio en otros campos no relacionados con el transporte, como el sector de la educación (Mahapatra y Khan, 2006), el sector servicios (Lin, 2007; Deng et al., 2008; Deng y Pei, 2009; Larasati et al., 2012), o el grado de satisfacción en el trabajo percibido por los empleados de varias compañías de refinería de gas (Azadeh et al., 2011). En todos estos estudios los modelos de ANN alcanzaron una exactitud muy elevada. Estos resultados, junto con el obtenido en esta investigación, refuerzan la idea defendida por Garver (2002) de que las ANN son una técnica adecuada para evaluar la importancia relativa de las variables que definen la satisfacción de los clientes en cualquier sector, incluido por tanto el del transporte. Además, la exactitud alcanzada por todas las ANN entrenadas en este estudio es significativamente superior a la obtenida del estudio elaborado por de Oña et al. (2012), en el que se utilizaron árboles de decisión para analizar la contribución relativa de los atributos de calidad del servicio en el ámbito del transporte público, que estuvo comprendida entre el 59,72% y el 62,16%.

H	Mínimo	Media	Desviación estándar	Máximo
1	0,0400	0,0531	0,0082	0,0874
2	0,0401	0,0532	0,0090	0,0943
3	0,0438	0,0525	0,0045	0,0679
4	0,0372	0,0519	0,0067	0,0668
5	0,0424	0,0528	0,0062	0,0765
6	0,0325	0,0475	0,0064	0,0580
7	0,0417	0,0514	0,0052	0,0643
8	0,0437	0,0520	0,0044	0,0858
9	0,0436	0,0513	0,0036	0,0620
10	0,0368	0,0530	0,0074	0,0744
11	0,0399	0,0505	0,0051	0,0674
12	0,0384	0,0520	0,0086	0,0938
13	0,0396	0,0526	0,0106	0,0997
14	0,0427	0,0517	0,0043	0,0616
15	0,0354	0,0526	0,0065	0,0706
16	0,0365	0,0514	0,0087	0,0853
17	0,0379	0,0513	0,0056	0,0610
18	0,0416	0,0513	0,0070	0,0761
19	0,0406	0,0513	0,0046	0,0600
20	0,0419	0,0516	0,0049	0,0635
21	0,0395	0,0508	0,0056	0,0611
22	0,0346	0,0524	0,0078	0,0863
23	0,0424	0,0518	0,0700	0,0742
24	0,0393	0,0497	0,0047	0,0591
25	0,0385	0,0508	0,0067	0,0715
26	0,0405	0,0524	0,0063	0,0708
27	0,0417	0,0520	0,0078	0,0923
28	0,0375	0,0507	0,0056	0,0633
29	0,0427	0,0534	0,0075	0,0789
30	0,0413	0,0517	0,0091	0,0890

Tabla 8. Rango de valores de MAPE para los conjuntos de ANN entrenados

Aunque no existen grandes diferencias de exactitud en las arquitecturas entrenadas, aquella con H=6 neuronas en la capa oculta presenta menores valores mínimo y medio de MAPE (Tabla 8), así que fue la seleccionada para desarrollar las siguientes fases de esta investigación.

6.2. Importancia relativa de las variables

Los valores de importancia relativa obtenidos por cada uno de los métodos fueron escalados en el rango [0,100]. Es decir, tras conocer el rango de valores de importancia relativa obtenido para todas las variables de un método, se le asignó un valor de 100 al mayor de ellos, mientras que la importancia relativa de las restantes variables fueron escaladas en función del rango de valores de dicho método. Esta transformación es necesaria para comparar y analizar adecuadamente los resultados entre métodos, ya que

el rango de valores de importancia relativa obtenido por cada método difiere de los demás incluso en varios órdenes de magnitud.

Los diagramas de cajas representados en las Figuras 13 a 16 muestran las importancias relativas de las 12 variables independientes obtenidas tras aplicar cada uno de los cuatro métodos seleccionados para este estudio (Perturb, Profile, CW y PaD) a cada una de las 50 ANN pertenecientes al conjunto con H=6 neuronas en su capa oculta.

Las figuras muestran que el rango de valores de importancia relativa de cada variable predictora es muy amplio, no sólo tras aplicar diferentes métodos, sino al aplicar el mismo método a dos ANN distintas entrenadas partiendo de pesos iniciales aleatorios diferentes. Algunos autores (Zhou et al., 2002; Cao and Qiao, 2008) ya pusieron de manifiesto este problema, y sugirieron trabajar con conjuntos de ANN para mitigar los problemas de estabilidad derivados de los análisis de sensibilidad. Estas propuestas han reforzado la idea desarrollada en esta investigación de trabajar, no sólo con una única ANN, sino con un conjunto de ellas.

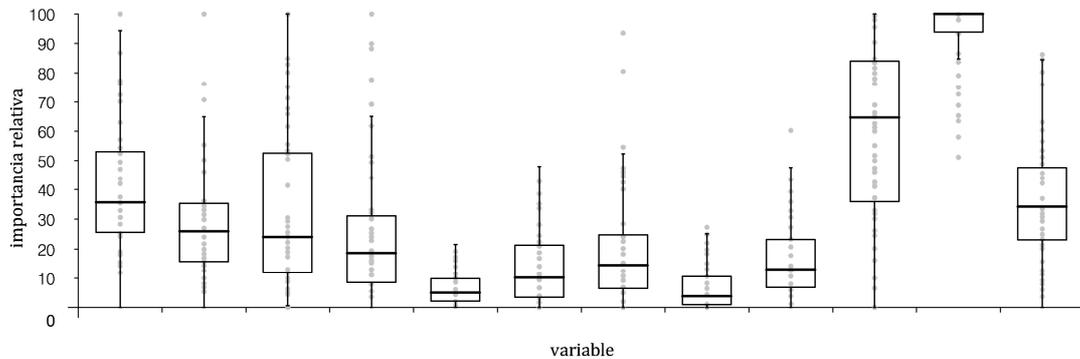


Figura 13. Diagrama de cajas de importancia relativa de las variables predictoras según el método PERTURB

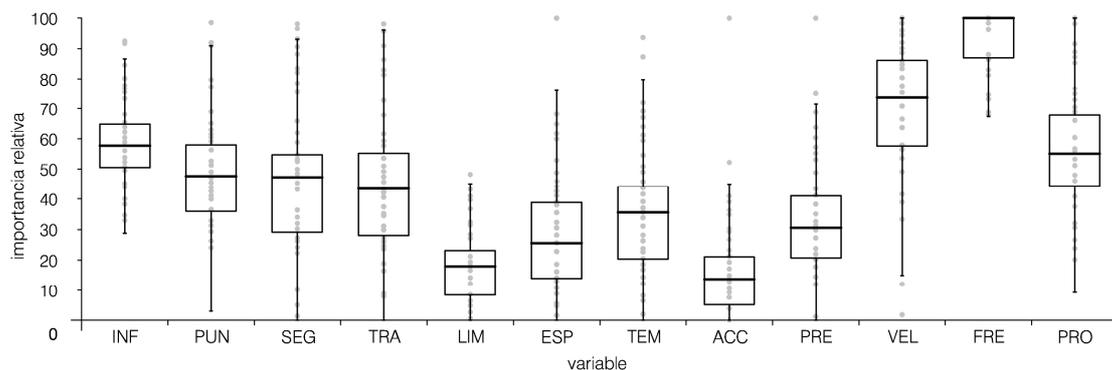


Figura 14. Diagrama de cajas de importancia relativa de las variables predictoras según el método PROFILE

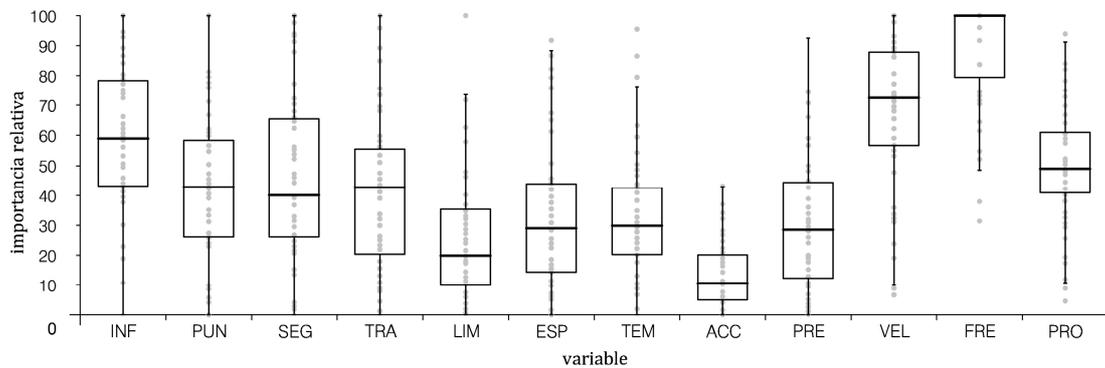


Figura 15. Diagrama de cajas de importancia relativa de las variables predictoras según el método CW

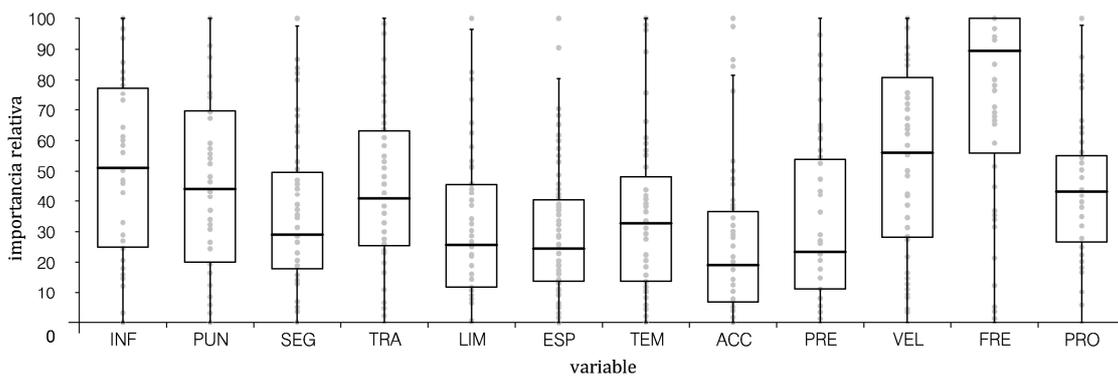


Figura 16. Diagrama de cajas de importancia relativa de las variables predictoras según el método PaD

Los métodos Profile y Perturb proporcionan rangos de valores de importancia relativa más restringidos para ciertas variables, como la FRECUENCIA, que siempre tiene una importancia relativa superior al 70% según el método Profile, y superior al 55% según el método Perturb; o la variable LIMPIEZA, cuya importancia relativa siempre está por debajo del 50% según el método Profile, y por debajo del 20% según el método Perturb.

Los resultados de los otros dos métodos, especialmente del PaD, muestran que cualquier variable puede alcanzar cualquier valor de importancia relativa comprendido entre el 0% y el 100%. En cualquier caso, el ranking de importancia obtenido tras aplicar los cuatro métodos para todas las variables es demasiado impredecible, lo que supone una limitación muy importante para determinar de forma adecuada un valor de importancia relativa y un orden de magnitud del mismo. Sin embargo, si se calculan los valores medios de importancia relativa obtenidos para cada variable por cada uno de los métodos, los resultados obtenidos son mucho más homogéneos.

En el método Profile, con los resultados de importancia relativa obtenidos en cada una de las N=50 ANN pertenecientes al conjunto con H=6 neuronas en la capa oculta, para cada una de las variables predictoras de la calidad del servicio, se generó un rango de 50 perfiles de variación en el intervalo [0,1] (Figura 18). Así, la importancia relativa de cada una de las variables viene determinada por la diferencia entre los valores máximo y mínimo (diferencia en el eje de ordenadas) de la línea que representa la media de valores del perfil de variación. La Figura 17 recoge los resultados obtenidos, que indican que la variable FRECUENCIA alcanza la importancia relativa más elevada (100,00%), seguida de la VELOCIDAD (77,72%), la INFORMACIÓN (64,15%) y la PROXIMIDAD (60,24%), por lo que estas variables tienen una importancia global en el modelo elevada. Un segundo nivel de importancia, considerado como medio-alto, puede ser asignado a las variables PUNTUALIDAD (54,45%), SEGURIDAD (53,28%) y TRATO (48,59%), seguido por un nivel de importancia media que contiene a las variables TEMPERATURA (38,44%), TARIFA (36,40%) y ESPACIO (27,22%). En el último escalón de importancia relativa quedarían las variables ACCESIBILIDAD (17,34%) y LIMPIEZA (3,36%).

Tras aplicar el método Perturb, también se obtienen 50 perfiles de variación para cada variable, que representan la variación del error MSE, en función de cómo afecta el porcentaje de ruido o perturbación introducida a la capacidad de respuesta de la ANN. (Figura 19). De nuevo, la diferencia entre los valores máximo y mínimo de la línea que representa los valores medios de cada perfil, indica la importancia relativa de las variables, de modo que cuanto mayor sea esta diferencia, mayor será la importancia relativa de la variable.

La Figura 20 muestra los resultados obtenidos tras aplicar este método. La FRECUENCIA (100,00%) es la variable globalmente más importante, seguida de la VELOCIDAD (63,60%). Las variables INFORMACIÓN (42,88%), PUNTUALIDAD (32,92%), SEGURIDAD (32,92%), TRATO (30,33%), PROXIMIDAD (23,28%) y TEMPERATURA (22,23%) se sitúan en un segundo nivel de importancia media. En las últimas posiciones del ranking de importancia quedan las variables PRECIO (17,90%), ESPACIO (14,51%), ACCESIBILIDAD (7,58%) y LIMPIEZA (7,12%).

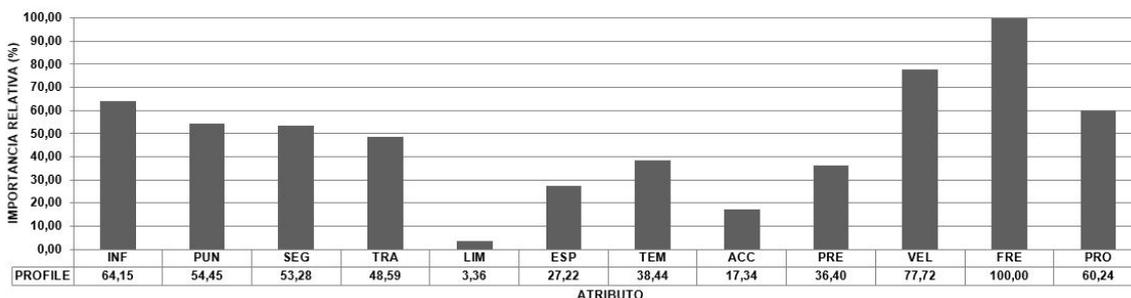


Figura 17. Importancia relativa de los atributos de calidad del servicio según el método PROFILE

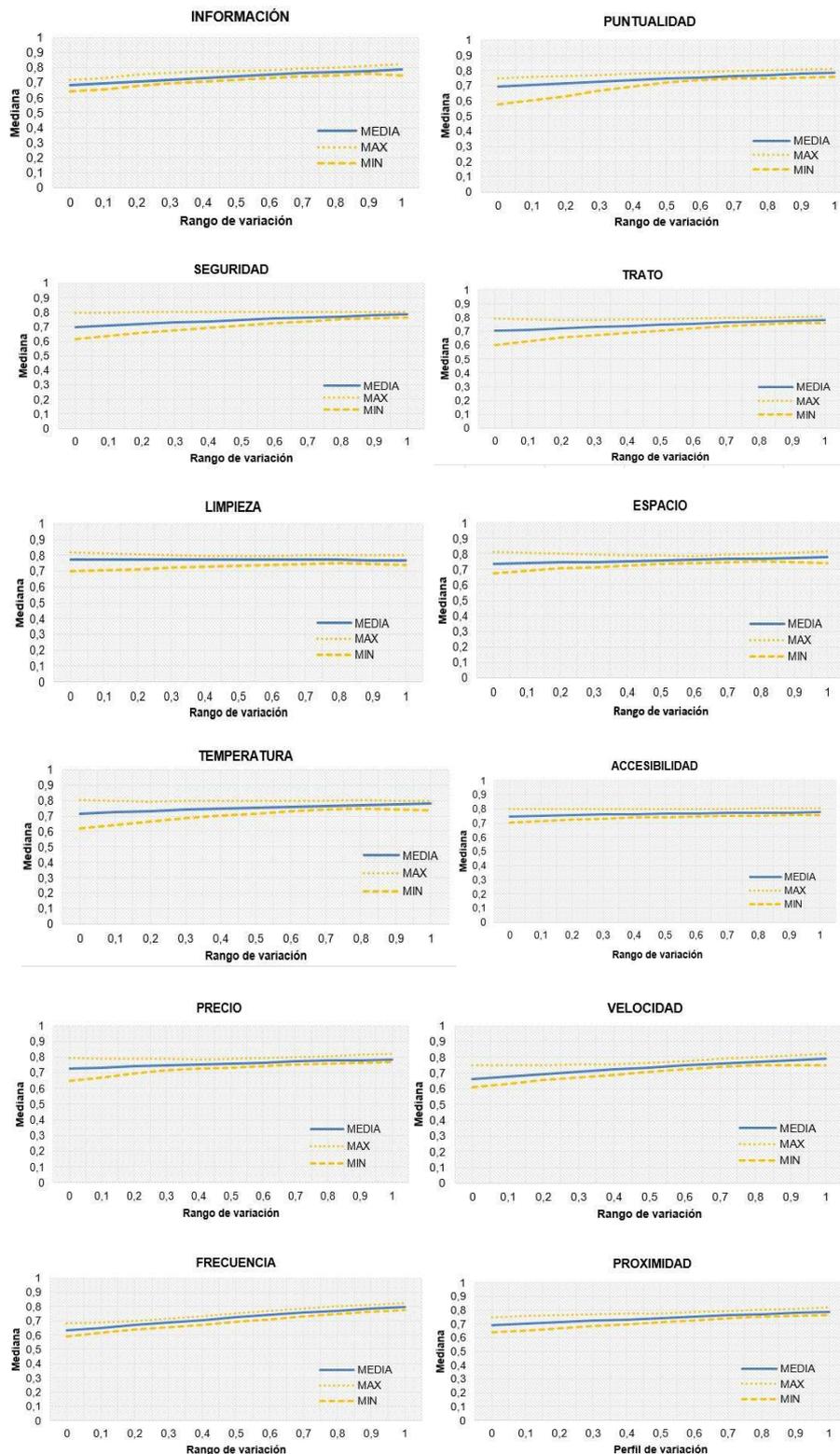


Figura 18. Perfiles de variación de los atributos de calidad del servicio según el método PROFILE

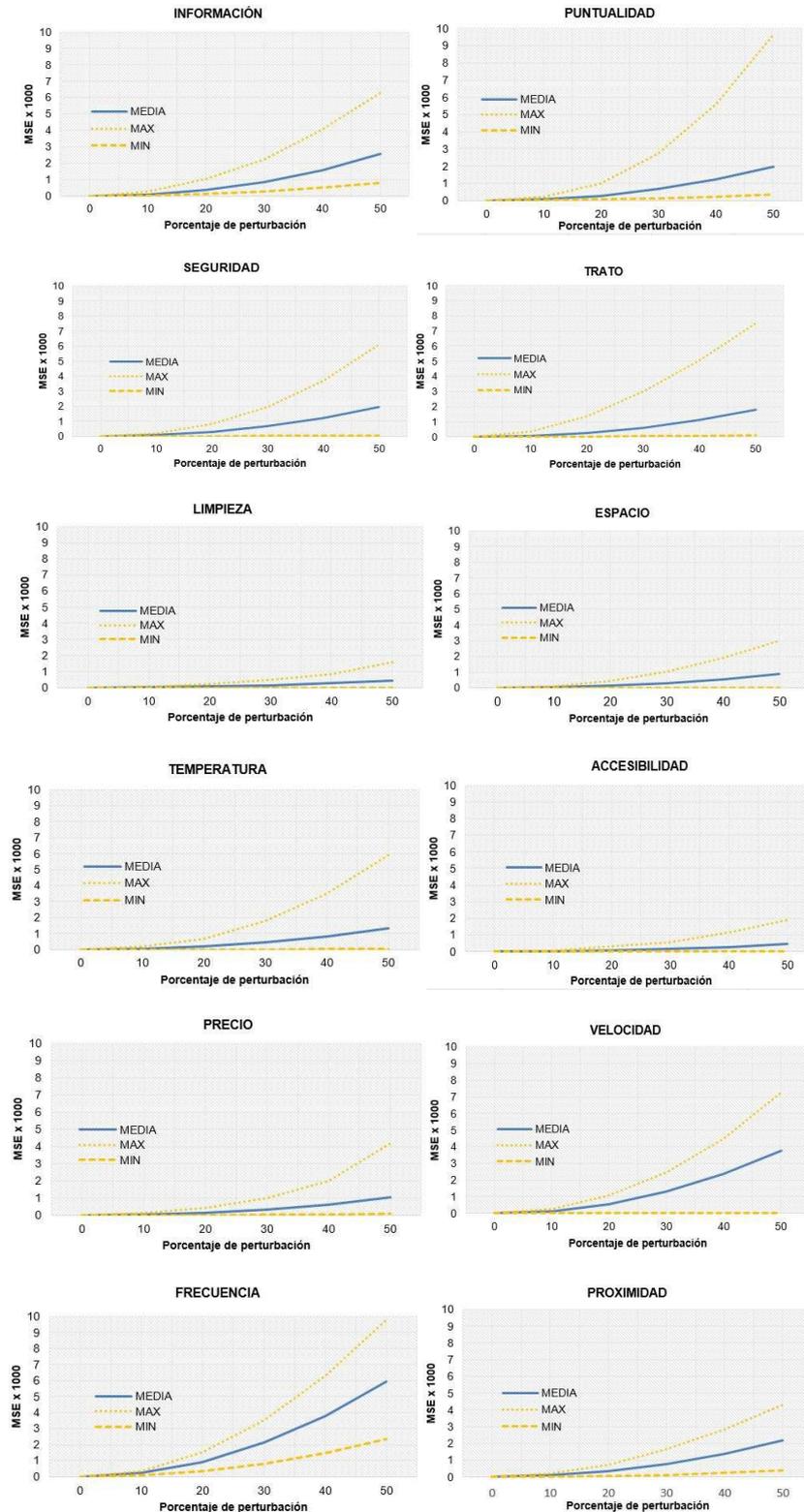


Figura 19. Perfiles de variación del MSE de los atributos de calidad del servicio según el método PERTURB

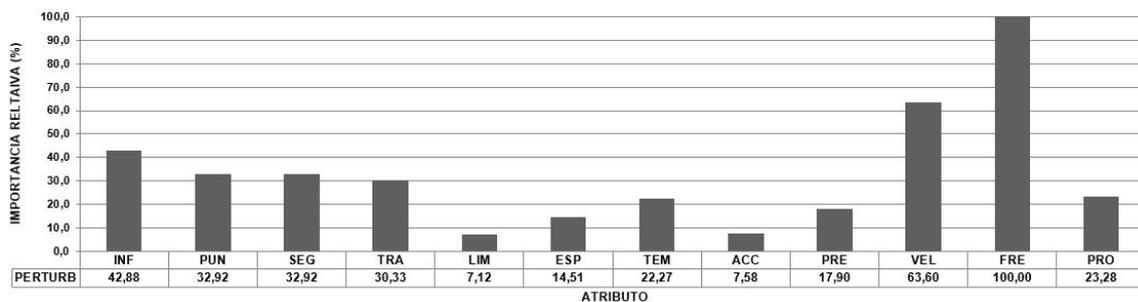


Figura 20. Importancia relativa de los atributos de calidad del servicio según el método PERTURB

En cuanto a los métodos CW y PaD, se ha aplicado la formulación correspondiente a cada uno de ellos indicada en los apartados 5.1.2.3 y 5.1.2.4 del capítulo 5 para determinar la importancia relativa de cada variable. Los valores de importancia relativa obtenidos se recogen en las Figuras 21 y 22.

Según el método CW, la variable con la mayor importancia relativa es la FRECUENCIA (100,00%), seguida de la VELOCIDAD (75,98%) y de la INFORMACIÓN (66,68%). En un segundo nivel de importancia, quedarían incluidas las variables PROXIMIDAD (55,49%), SEGURIDAD (51,38%), PUNTUALIDAD (51,35%) y TRATO (47,81%), seguido por el grupo de importancia media que contiene a las variables TEMPERATURA (36,65%), ESPACIO (36,45%), PRECIO (31,98%) y LIMPIEZA (27,39%). La variable con menor importancia relativa es la ACCESIBILIDAD (14,56%).

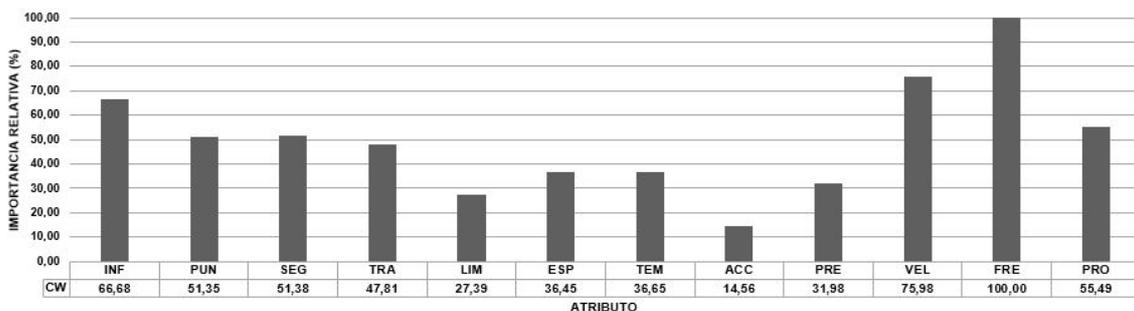


Figura 21. Importancia relativa de los atributos de calidad del servicio según el método CW

El método PaD considera como variables más importantes la FRECUENCIA (100,00%), la VELOCIDAD (72,90%), la INFORMACIÓN (71,01%), la PUNTUALIDAD (64,13%), el TRATO (60,67%) y la PROXIMIDAD (58,30%). La importancia relativa del resto de variables puede considerarse como media: TEMPERATURA (49,05%), SEGURIDAD (48,08%), PRECIO (45,43%), LIMPIEZA (44,01%), ESPACIO (43,35%) y ACCESIBILIDAD (37,41%).

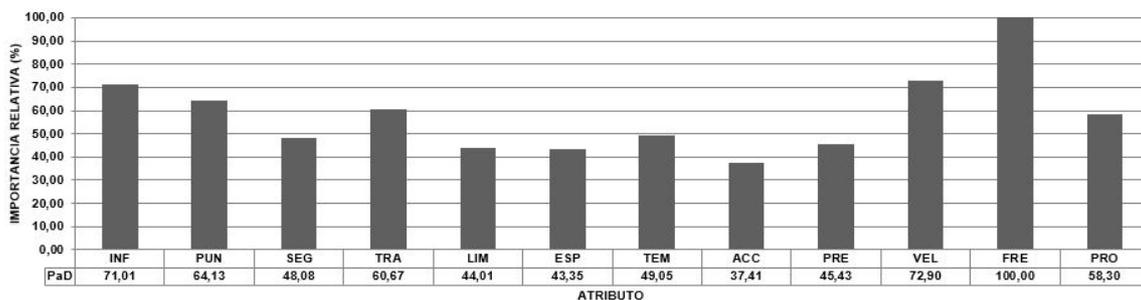


Figura 22. Importancia relativa de los atributos de calidad del servicio según el método PaD

La Tabla 9 resume los valores de importancia relativa y el ranking que ocupan los atributos de calidad del servicio estudiados en función de los cuatro métodos:

ATRIBUTO	PROFILE		PERTURB		CW		PaD	
	MEDIA (%)	RÁNKING	MEDIA (%)	RANKING	MEDIA (%)	RANKING	MEDIA (%)	RANKING
INF	64,15	3	42,88	3	66,68	3	71,01	3
PUN	54,45	5	32,92	4-5	51,35	6	64,13	4
SEG	53,28	6	32,92	4-5	51,38	5	48,08	8
TRA	48,59	7	30,33	6	47,81	7	60,67	5
LIM	3,36	12	7,12	12	27,39	11	44,01	10
ESP	27,22	10	14,51	10	36,45	9	43,35	11
TEM	38,44	8	22,27	8	36,65	8	49,05	7
ACC	17,34	11	7,58	11	14,56	12	37,41	12
PRE	36,40	9	17,90	9	31,98	10	45,43	9
VEL	77,72	2	63,60	2	75,98	2	72,90	2
FRE	100,00	1	100,00	1	100,00	1	100,00	1
PRO	60,24	4	23,28	7	55,49	4	58,30	6

Tabla 9. Importancia relativa y ranking de los atributos de calidad del servicio según los métodos PROFILE, PERTURB, CW y PaD

La Figura 23 muestra una comparación del ranking de importancia relativa de las 12 variables consideradas determinada por cada método. Esta figura refleja que las posiciones de importancia relativa son muy similares, y que por tanto aplicando la metodología propuesta en esta investigación se mitiga considerablemente la elevada inestabilidad inherente en estos métodos cuando se aplican a una única ANN. Los cuatro métodos coinciden en que las variables más influyentes son la FRECUENCIA, VELOCIDAD e INFORMACIÓN, y en que las variables menos influyentes son la LIMPIEZA, ACCESIBILIDAD y ESPACIO, con la única excepción del método CW, que considera a las variables ACCESIBILIDAD, LIMPIEZA y PRECIO como las menos importantes, seguidas de la variable ESPACIO. Las seis variables restantes ocupan posiciones intermedias en el ranking. Los métodos Profile y Perturb proporcionan una jerarquía de importancia similar, mientras que el método PaD es el que presenta mayores discrepancias.

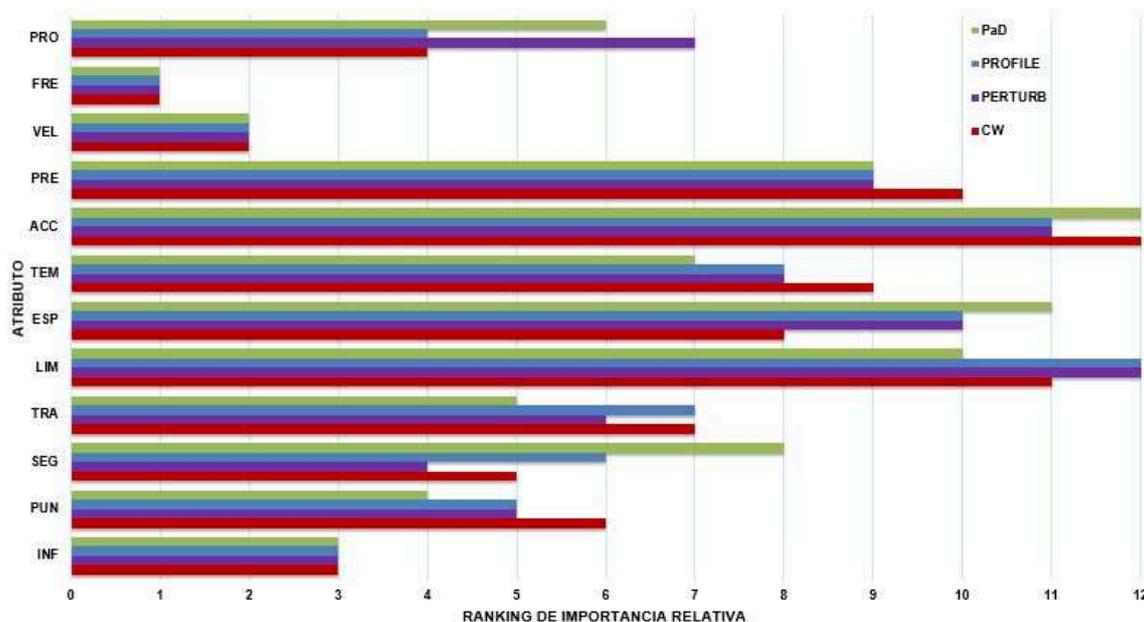


Figura 23. Ranking de importancia relativa de los atributos de calidad del servicio según los métodos PROFILE, PERTURB, CW Y PaD

Respecto al grado de importancia relativa asignado a cada variable y expresado en tanto por ciento, los métodos Profile y Perturb permiten establecer cuatro escalones de importancia: muy alta, alta, media y baja. El método CW establece tres niveles claros de importancia: alta, media y baja, y el método PaD establece unos porcentajes de influencia que pueden diferenciarse en dos niveles de importancia: alto y medio, ya que el menor valor de importancia relativa queda por encima del 37%.

6.3. Análisis estadístico

Se ha llevado a cabo un análisis estadístico para confirmar si existen diferencias significativas entre los valores de importancia relativa de los atributos de calidad del servicio considerados en esta investigación. Para ello ha sido necesario utilizar técnicas no paramétricas (tests de Kruskal-Wallis y Dunn) debido a la no-normalidad de los datos. Este análisis pone de manifiesto que existen diferencias significativas con un nivel de confianza del 95% entre algunos de estos atributos en función de los ratios de importancia relativa determinados por los métodos (Perturb, Profile, CW), tal y como puede apreciarse en la Tabla 10.

En el caso del método CW, se identificaron 7 grupos homogéneos. Las variables incluidas en cada grupo no presentan diferencias estadísticamente significativas intra-grupo, pero sí existen estas diferencias entre grupos diferentes. En el caso de los métodos Perturb y Profile, se han detectado 8 grupos estadísticamente significantes. La mayoría de los atributos que componen estos grupos homogéneos coinciden en los tres métodos

(Perturb, Profile y CW). Los grupos se han identificado con letras, de modo que sus ratios de importancia media decrecen a medida que se suceden las letras del abecedario.

ATRIBUTO	CW	PERTURB	PROFILE
INF	2,864 ^{b,c}	2,540 ^{b,c}	0,787 ^{b,c,d}
PUN	2,262 ^d	1,950 ^{d,e}	0,785 ^{c,d}
SEG	2,098 ^d	1,950 ^{d,e}	0,785 ^{c,d}
TRA	2,063 ^{d,e}	1,797 ^{e,f}	0,783 ^{d,e}
LIM	1,255 ^f	0,421 ^h	0,768 ^g
ESP	1,627 ^{e,f}	0,859 ^g	0,777 ^f
TEM	1,590 ^{e,f}	1,319 ^{f,g}	0,779 ^{e,f}
ACC	0,660 ^g	0,449 ^h	0,776 ^f
PRE	1,359 ^f	1,060 ^{f,g}	0,785 ^{c,d}
VEL	3,284 ^b	3,768 ^b	0,790 ^b
FRE	4,315 ^a	5,925 ^a	0,797 ^a
PRO	2,357 ^{c,d}	2,163 ^{c,d}	0,787 ^{b,c}

a,b,c,d,e,f,g,h: denota diferencias estadísticamente significantes ($p < 0,05$). Dos o más variables con la misma letra en la misma columna indican subgrupos homogéneos.

Tabla 10. Grupos homogéneos con valores medios (n=50) significativamente diferentes entre otros grupos

Tras observar las salidas que proporciona este análisis (Tabla 10), el atributo FRECUENCIA es identificado por los tres métodos como un grupo independiente (grupo a) que presenta diferencias estadísticamente significativas respecto al resto de grupos. Esta variable presenta el mayor porcentaje de influencia en la valoración global de la calidad del servicio en todos los casos. Del mismo modo, las variables VELOCIDAD, INFORMACIÓN y PROXIMIDAD también obtienen una importancia relativa elevada respecto a la valoración global de la calidad del servicio según los métodos CW y Perturb (grupos b y c), con diferencias estadísticamente significativas respecto al resto de los grupos. Respecto al método Profile, del análisis estadístico se extrae un grupo b, compuesto por las variables VELOCIDAD, INFORMACIÓN y PROXIMIDAD. Esto indica que, después del atributo FRECUENCIA, este grupo de tres atributos es el que provoca mayor impacto en la valoración global que los pasajeros hacen de la calidad del servicio.

Por el contrario, la ACCESIBILIDAD es la variable que presenta menores ratios de importancia. En el caso del método CW, el grupo g está compuesto sólo por esta variable, y presenta diferencias considerables respecto al resto de grupos de variables. Asimismo, esta variable se ubica dentro del grupo h del método Perturb y del grupo f del método Profile, existiendo diferencias estadísticamente significativas entre estos grupos y el resto de grupos en los que se han agrupado estadísticamente las variables para cada uno de los dos métodos. En ambos presentan ratios de importancia bajos.

En el método CW, el grupo f está compuesto por las variables LIMPIEZA interior del vehículo, ESPACIO a bordo, TEMPERATURA en el autobús y PRECIO. Se extraen diferencias estadísticamente significativas respecto al resto de grupos homogéneos,

excepto para los atributos que pertenecen al grupo e. Respecto al grupo g del método Perturb, se identifican las mismas variables que en el grupo f del método CW salvo la LIMPIEZA. En el método Profile, el grupo f está formado por las variables ESPACIO, TEMPERATURA y ACCESIBILIDAD. No hay coincidencia exacta entre los atributos identificados en este grupo y los que pertenecen a los grupos de los demás métodos. Sin embargo, si se consideran los atributos que pertenecen a los dos últimos grupos de manera conjunto en todos los métodos (atributos cuyos ratios de importancia media relativa son más bajos que los del resto de grupos), sí se identifican los mismos atributos (LIMPIEZA, ESPACIO, TEMPERATURA, ACCESIBILIDAD), excepto el PRECIO, que sólo pertenece a estos dos grupos homogéneos en el caso de los métodos CW y Perturb.

El resto de atributos (PUNTUALIDAD, SEGURIDAD, TRATO y PRECIO) son de importancia media, puesto que no producen el mayor impacto en la valoración global del servicio (como ocurre con la FRECUENCIA, VELOCIDAD, INFORMACIÓN Y PROXIMIDAD), ni tampoco un impacto muy bajo (como ocurre con la ACCESIBILIDAD, LIMPIEZA, ESPACIO y TEMPERATURA).

Diversos grupos de investigación han puesto de manifiesto la existencia de varios grupos de atributos en función de su mayor o menor impacto en la calidad del servicio y en la satisfacción del cliente (De Oña y De Oña, 2015). Philip y Hazlett (1997) propusieron un modelo con una estructura jerárquica que establecía tres clases de atributos: pivotaes, principales y periféricos. Este modelo fue contrastado tras aplicarse a la industria del transporte por ferrocarril por Tripp y Drea (2002), quienes extrajeron como conclusión que efectivamente existían una serie de atributos principales (información sobre servicios, comodidad de los asientos, sala de descanso, café bar, etc.) que ejercían mayor influencia sobre la satisfacción de los usuarios. Aunque en este caso no es posible establecer un paralelismo exacto entre el estudio de Tripp y Drea (2002) y el de la presente investigación, puesto que se trata de servicios de transporte diferentes que han sido caracterizados a través de variables diferentes, sí que ambos coinciden en que los aspectos relacionados con la INFORMACIÓN, son considerados como de importancia alta en el grado de satisfacción percibido por el usuario respecto al servicio recibido. Por el contrario, mientras en el estudio del transporte por ferrocarril las variables relacionadas con las comodidades del pasajero eran consideradas de importancia alta, las variables relacionadas con el bienestar en el habitáculo en esta investigación (LIMPIEZA, ESPACIO y TEMPERATURA) son consideradas como de importancia media-baja.

La UNE-EN 13186 estándar (Comité europeo de estandarización, 2002) clasifica las características de un servicio en básicas (PUNTUALIDAD, SEGURIDAD, etc.), proporcionales (COMODIDAD, LIMPIEZA, etc.) y atractivas (tarjetas sin contacto, navegadores, etc.), en función de cómo afectan a la confianza del usuario en cuanto al nivel de calidad del servicio recibido. Por su parte, el Manual de Capacidad de Tránsito y de Calidad del Servicio (Transportation Research Board, 2004) agrupa los atributos en factores de disponibilidad, considerados como más importantes (horarios, cobertura del servicio, INFORMACIÓN, etc.), y factores de conveniencia y comodidad para los pasajeros (aparición del servicio, grado de saturación, tarifa, etc), que son considerados como menos importantes. Los atributos considerados como importantes según estas dos

clasificaciones: INFORMACIÓN, PUNTUALIDAD y SEGURIDAD, coinciden con los que el análisis estadístico de este estudio incluye dentro de los grupos de importancia alta, mientras que otros como la TARIFA y los relacionados con el confort (que en el caso de esta investigación pueden englobar a las variables LIMPIEZA, TEMPERATURA y ESPACIO, son considerados como de importancia más baja.

Eboli y Mazzulla (2008) también distinguieron entre diferentes categorías de atributos en función de su impacto en la calidad del servicio utilizando una base de datos de transporte público por autobús en Cosenza (Italia). Estos autores definieron atributos básicos y no básicos basándose en las preferencias de los usuarios. La FRECUENCIA y la PROXIMIDAD del servicio, fueron identificadas como las variables con mayor impacto, y consideradas como atributos básicos. Estos resultados concuerdan con los obtenidos en la presente investigación. La LIMPIEZA y el TRATO fueron consideradas como atributos no básicos (Eboli y Mazzulla, 2008), y en esta investigación presentan un impacto medio o bajo en la calidad del servicio.

A diferencia de tres métodos anteriores, al realizar el análisis estadístico a los resultados de importancia relativa obtenidos tras aplicar el método PaD, no se aprecian diferencias significativas entre una variable y las restantes, por lo que no se pueden determinar varios grupos en función de su importancia relativa.

Así, este análisis estadístico permite identificar grupos homogéneos de atributos en el caso de los métodos Perturb, Profile y CW, en función del mayor o menor impacto que ocasionan en la percepción de los pasajeros de la calidad del servicio prestado, mientras que en el caso PaD no es posible distinguir entre grupos homogéneos.

6.4. Análisis de la robustez de la metodología de cálculo de importancia relativa propuesta en esta investigación

La metodología descrita en la sección 6.2 se desarrolla tras seleccionar el conjunto de ANN con $H=6$ neuronas en la capa oculta, basándose en el argumento de que este grupo es el que presenta un error de generalización, medido a través del MAPE, ligeramente menor. Sin embargo, para seleccionar este grupo se han considerado previamente $H=30$ conjuntos de ANN, que difieren unos de otros en el número de neuronas de su capa oculta, conteniendo cada grupo H un número de ANN $M=50$ que ha sido necesario entrenar. Este procedimiento es bastante tedioso y además, como es bien conocido en la literatura científica, es difícil que las ANN entrenadas en cada grupo sean las de comportamiento óptimo o sub-óptimo (Cao y Qiao, 2008), debido a las múltiples posibilidades de toparse con mínimos locales en la superficie de error durante la fase de entrenamiento, de modo que la solución alcanzada no sea la que corresponde al mínimo global. Es por ello que en este apartado se pretende comprobar la robustez de la metodología propuesta en la sección 6.2, es decir, cómo afecta la elección de uno u otro conjunto H a los valores de importancia relativa obtenidos para las variables consideradas en el estudio. Si existen diferencias importantes en cuanto a estos valores de importancia relativa según el conjunto H , no será posible obviar el paso de selección del conjunto H

con un mejor comportamiento global, pero si no existen diferencias importantes, esta metodología permitiría trabajar directamente con un conjunto de ANN cualquiera, considerando directamente un único grupo con H de neuronas en su capa oculta, sin necesidad de entrenar los grupos restantes, con lo que se agilizaría notablemente todo el procedimiento.

Para comprobar si existe estabilidad y robustez en los resultados, se ha repetido el mismo procedimiento que se aplicó en la sección 6.2 al conjunto con H=6 neuronas en la capa oculta a todos los conjuntos H de ANN entrenados, con $H \in [1,30]$, aplicando cada uno de los cuatro métodos de importancia relativa (Perturb, Profile, CW y PaD) a cada conjunto. Las Tablas 11-14 representan la posición o ranking de importancia que ocupa cada atributo de calidad del servicio en función del conjunto H al que pertenece y del método de cálculo de importancia relativa aplicado.

Como puede apreciarse, la FRECUENCIA ocupa la primera posición en el ranking de importancia para todos los conjuntos H de ANN y para los cuatro métodos, existiendo por tanto consenso total al considerarla como la variable más importante que define la valoración global de la calidad del servicio. La VELOCIDAD ocupa el segundo lugar en el ranking, existiendo consenso en los métodos Perturb, Profile y CW, y alguna disparidad en el método PaD, en el que para algunos conjuntos H esta variable ocupa la tercera posición. En el caso de la variable INFORMACIÓN, el método CW siempre considera que su posición en el ranking es la tercera, siendo ésta también la posición mayoritaria para el resto de métodos aplicados, aunque en ellos hay algunos conjuntos H en los que ocupa la cuarta posición. A continuación, se identifica un grupo de variables: SEGURIDAD, PROXIMIDAD, PUNTUALIDAD y TRATO, comprendidas entre las posiciones cuarta y séptima. En el caso de la variable SEGURIDAD, el método Perturb la sitúa mayoritariamente en el cuarto puesto, mientras que los métodos Perturb y CW la ubican en mayor número de conjuntos H en el quinto; el método PaD de nuevo presenta más discrepancias que los restantes, obteniéndose posiciones comprendidas entre el cuarto y el octavo puesto, según el subgrupo H considerado. La variable PROXIMIDAD queda enmarcada claramente en la cuarta posición según los métodos Profile y CW, mientras que el Perturb la ubica más en las posiciones quinta y sexta, y el PaD, con más diversidad de posiciones, la sitúa entre la tercera y la sexta. La PROXIMIDAD ocupa mayoritariamente la sexta posición, y el TRATO la séptima, especialmente según el método CW. En un tercer grupo del ranking quedarían las variables TEMPERATURA, que ocupa mayoritariamente el octavo puesto, con mayor diversidad de posiciones según los métodos Profile y PaD, y el ESPACIO y el PRECIO, que ocupan las posiciones novena ó décima, en los cuatro métodos, sin que los resultados permitan decantarse claramente por una u otra posición sea cual sea el método. La variable LIMPIEZA queda ubicada en la posición onceava por todos los métodos salvo por el Profile, que la coloca en la doceava, y la ACCESIBILIDAD ocupa la última posición del ranking según todos los métodos excepto el Profile, que la coloca en penúltima posición.

H	FRE				VEL				INF			
	Profile	Perturb	CW	PaD	Profile	Perturb	CW	PaD	Profile	Perturb	CW	PaD
1	1	1	1	1	2	2	2	2	3	3	3	3
2	1	1	1	1	2	2	2	2	3	3	3	3
3	1	1	1	1	2	2	2	3	3	3	3	2
4	1	1	1	1	2	2	2	2	3	3	3	3
5	1	1	1	1	2	2	2	2	3	3	3	3
6	1	1	1	1	2	2	2	2	3	3	3	3
7	1	1	1	1	2	2	2	2	3	3	3	3
8	1	1	1	1	2	2	2	3	3	3	3	2
9	1	1	1	1	2	2	2	2	3	3	3	3
10	1	1	1	1	2	2	2	2	3	3	3	3
11	1	1	1	1	2	2	2	2	3	3	3	3
12	1	1	1	1	2	2	2	3	3	3	3	2
13	1	1	1	1	2	2	2	2	3	3	3	3
14	1	1	1	1	2	2	2	3	3	3	3	2
15	1	1	1	1	2	2	2	2	3	3	3	4
16	1	1	1	1	2	2	2	2	3	3	3	4
17	1	1	1	1	2	2	2	2	3	3	3	3
18	1	1	1	1	2	2	2	2	3	3	3	3
19	1	1	1	1	2	2	2	2	3	3	3	3
20	1	1	1	1	2	2	2	2	4	3	3	3
21	1	1	1	1	2	2	2	2	5	4	3	3
22	1	1	1	1	2	2	2	2	3	3	3	3
23	1	1	1	1	2	2	2	2	3	3	3	3
24	1	1	1	1	2	2	2	2	3	3	3	3
25	1	1	1	1	2	2	2	2	3	3	3	3
26	1	1	1	1	2	2	2	2	3	3	3	3
27	1	1	1	1	2	2	2	3	3	3	3	2
28	1	1	1	1	2	2	2	2	3	3	3	3
29	1	1	1	1	2	2	2	2	3	3	3	3
30	1	1	1	1	2	2	2	3	3	3	3	2

Tabla 11. Ranking de importancia relativa de las variables FRECUENCIA, VELOCIDAD e INFORMACIÓN por métodos

H	SEG				PRO				PUN			
	Profile	Perturb	CW	PaD	Profile	Perturb	CW	PaD	Profile	Perturb	CW	PaD
1	5	4	5	5	4	6	4	4	6	5	6	6
2	5	4	5	5	4	6	4	4	6	5	6	6
3	5	4	5	6	4	6	4	4	6	5	6	5
4	5	4	6	6	4	6	4	4	6	5	5	5
5	4	4	4	6	5	6	4	5	6	5	6	4
6	6	4	5	8	4	7	4	6	5	5	6	4
7	5	4	5	5	4	6	4	6	6	7	6	4
8	4	4	4	4	5	5	4	6	6	6	6	5
9	5	4	5	5	4	7	4	4	6	5	6	6
10	5	4	5	4	4	6	4	5	7	7	6	7
11	4	4	4	5	5	7	4	6	6	5	6	4
12	5	4	6	5	4	6	4	6	6	5	5	4
13	4	4	5	5	5	5	4	4	6	6	6	6
14	5	4	6	6	4	6	4	5	6	5	5	4
15	5	4	6	7	4	7	4	5	7	6	4	3
16	5	4	6	6	4	6	4	3	6	5	5	5
17	5	4	5	4	4	5	4	5	6	6	6	6
18	5	4	5	5	4	5	4	4	6	6	6	6
19	5	4	5	5	4	5	4	4	6	6	6	6
20	3	4	4	4	5	7	4	5	6	6	6	6
21	4	3	5	4	3	5	4	5	6	6	6	6
22	5	4	5	5	4	6	4	4	7	7	6	6
23	4	4	4	5	5	5	4	4	6	6	6	6
24	5	4	5	6	4	5	4	4	6	6	6	5
25	5	4	5	6	4	6	4	4	7	7	6	5
26	5	4	5	5	4	6	4	4	7	7	7	6
27	4	4	5	5	5	6	4	4	7	7	6	6
28	5	4	6	6	4	6	4	4	6	5	5	5
29	5	4	5	5	4	6	4	4	7	7	7	7
30	5	4	6	6	4	5	4	5	6	7	5	4

Tabla 12. Ranking de importancia relativa de las variables SEGURIDAD, PROXIMIDAD y PUNTUALIDAD por métodos

H	TRA				TEM				ESP			
	Profile	Perturb	CW	PaD	Profile	Perturb	CW	PaD	Profile	Perturb	CW	PaD
1	7	7	7	7	10	8	10	10	8	9	8	8
2	7	7	7	7	10	9	10	9	8	8	8	8
3	7	7	7	8	8	8	8	7	9	9	9	10
4	7	7	7	7	9	9	9	8	8	8	8	10
5	7	7	7	7	10	10	10	11	8	8	8	8
6	7	6	7	5	8	8	9	7	10	10	8	11
7	7	5	7	7	10	10	10	8	8	8	8	9
8	7	7	7	8	10	8	10	7	9	9	9	9
9	7	6	7	7	10	9	10	8	8	8	8	10
10	6	5	7	6	9	8	10	8	8	9	8	9
11	7	6	7	7	8	9	8	8	10	10	10	10
12	7	7	7	7	8	8	8	11	10	10	9	8
13	7	7	7	7	9	8	10	10	10	9	8	8
14	7	7	7	7	8	8	9	8	10	9	8	9
15	6	5	7	6	8	8	8	8	10	10	9	10
16	7	7	7	7	8	9	9	11	9	8	8	8
17	7	7	7	9	10	9	10	10	8	8	8	7
18	7	7	7	7	8	8	9	9	10	10	10	10
19	7	7	7	7	8	8	9	9	9	9	8	8
20	7	5	7	7	8	8	8	8	9	9	10	11
21	7	7	7	7	8	8	8	10	9	9	9	8
22	6	5	7	8	9	9	10	11	10	10	9	10
23	7	7	7	7	8	8	8	8	10	10	10	10
24	7	7	7	7	8	8	8	8	9	9	9	11
25	6	5	7	7	9	10	8	8	8	8	10	9
26	6	5	6	7	8	8	9	10	9	10	8	8
27	6	5	7	7	8	8	8	9	9	9	9	8
28	7	7	7	7	8	8	9	9	10	10	10	10
29	6	5	6	6	9	8	8	8	10	9	10	9
30	7	6	7	7	8	8	8	9	10	9	10	10

Tabla 13. Ranking de importancia relativa de las variables TRATO, TEMPERATURA y ESPACIO por métodos

H	PRE				LIM				ACC			
	Profile	Perturb	CW	PaD	Profile	Perturb	CW	PaD	Profile	Perturb	CW	PaD
1	9	10	9	9	11	11	11	11	12	12	12	12
2	9	10	9	10	12	11	11	11	11	12	12	12
3	10	10	10	11	12	11	11	12	11	12	12	9
4	10	10	10	9	12	11	11	11	11	12	12	12
5	9	9	9	10	12	11	11	9	11	12	12	12
6	9	9	10	9	12	12	11	10	11	11	12	12
7	9	9	9	10	12	11	11	11	11	12	12	12
8	8	10	8	10	12	11	11	11	11	12	12	12
9	9	10	9	9	12	11	11	11	11	12	12	12
10	10	10	9	10	12	11	11	11	11	12	12	12
11	9	8	9	9	12	12	11	11	11	11	12	12
12	9	9	10	9	12	12	11	10	11	11	12	12
13	8	10	9	9	12	11	11	11	11	12	12	12
14	9	10	10	11	12	11	11	10	11	12	12	12
15	9	9	10	12	12	11	11	9	11	12	12	11
16	10	10	10	9	12	11	11	10	11	12	12	12
17	9	10	9	8	12	11	11	12	11	12	12	11
18	9	9	8	8	12	11	11	11	11	12	12	12
19	10	10	10	10	12	11	11	11	11	12	12	12
20	10	10	9	10	12	11	11	9	11	12	12	12
21	10	10	10	9	12	11	11	11	11	12	12	12
22	8	8	8	7	12	11	11	9	11	12	12	12
23	9	9	9	9	12	11	11	11	11	12	12	12
24	10	10	10	9	12	11	11	10	11	12	12	12
25	10	9	9	10	12	11	11	12	11	12	12	11
26	10	9	10	9	12	11	11	11	11	12	12	12
27	10	10	10	11	12	11	11	10	11	12	12	12
28	9	9	8	8	12	11	11	11	11	12	12	12
29	8	10	9	10	12	11	11	11	11	12	12	12
30	9	10	9	8	12	11	11	11	11	12	12	12

Tabla 14. Ranking de importancia relativa de las variables PRECIO, LIMPIEZA y ACCESIBILIDAD por métodos

Se han recogido en las Tablas 15 y 16 los resultados del porcentaje de importancia relativa y de desviación típica media obtenidos para cada variable en función del método aplicado para el conjunto total de ANN comprendidas en todos los subconjuntos H, con $H \in [1,30]$.

En el **Apéndice 2** se han representado gráficamente (Figuras 24 a 47) los valores tanto de importancia como de desviación típica medios obtenidos para cada variable y en función de cada uno de los métodos (Perturb, Profile, CW y PaD) aplicados.

ATRIBUTO	PERTURB		PROFILE		CW		PaD	
	Max	Min	Max	Min	Max	Min	Max	Min
INF	55,46	37,57	75,16	60,47	73,71	59,01	77,60	58,15
PUN	32,92	20,52	55,69	44,56	56,28	45,05	68,35	42,08
SEG	45,53	31,13	67,79	49,75	64,38	45,48	60,84	41,04
TRA	34,05	17,00	51,35	37,15	47,81	36,33	60,67	35,84
LIM	15,68	3,42	11,59	1,35	27,39	16,22	45,64	19,51
ESP	18,22	9,23	35,55	25,68	36,45	26,28	44,95	25,65
TEM	25,41	8,22	46,41	24,44	44,49	23,41	52,65	25,63
ACC	10,30	2,37	19,55	8,76	22,98	11,93	46,25	13,56
PRE	17,90	7,84	36,40	24,26	35,50	24,55	45,43	26,37
VEL	73,81	50,32	84,98	70,02	84,44	69,13	88,47	58,53
FRE	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
PRO	30,02	21,25	67,75	56,10	65,13	53,66	66,73	48,91

Tabla 15. Porcentajes máximo y mínimo de importancia relativa de los atributos por métodos

ATRIBUTO	PERTURB		PROFILE		CW		PaD	
	Max	Min	Max	Min	Max	Min	Max	Min
INF	30,67	14,32	17,89	8,88	24,45	14,30	30,85	18,88
PUN	28,12	10,55	19,75	10,75	26,49	15,48	30,52	17,10
SEG	32,05	18,28	24,62	15,91	28,04	18,82	30,89	20,30
TRA	32,54	13,09	27,59	15,23	27,49	15,81	30,34	19,57
LIM	24,73	3,85	31,13	13,61	21,15	11,66	26,64	15,89
ESP	39,99	8,58	24,03	11,38	23,69	15,29	29,33	15,39
TEM	22,64	8,57	23,98	13,27	24,57	14,82	28,03	17,55
ACC	16,02	3,08	17,88	8,60	18,64	8,42	27,82	9,68
PRE	24,11	6,80	19,26	10,05	20,34	13,72	27,81	16,50
VEL	40,09	22,56	29,50	11,68	28,83	17,84	32,10	20,81
FRE	38,66	24,55	21,05	8,47	21,54	7,75	33,28	9,70
PRO	15,43	8,17	20,31	9,05	24,51	14,76	29,04	17,01

Tabla 16. Porcentajes máximo y mínimo de desviación típica de los valores de importancia relativa de los atributos por métodos

La FRECUENCIA obtiene un 100,00% de importancia relativa independientemente del método de cálculo utilizado y del conjunto H considerado. En cuanto a la desviación típica, el método Perturb obtiene unos valores (en porcentaje) mayores que los demás métodos, estando su rango comprendido en el intervalo [24,55; 38,66]. Los métodos Profile y CW obtienen rangos de desviación típica más bajos, de [8,47; 21,05] y de [7,75; 21,54] respectivamente, mientras que los valores de desviación típica del método PaD oscilan en un rango de valores intermedio del [9,70; 33,28].

La segunda variable en el ranking de importancia relativa, la VELOCIDAD, tiene un rango de valores de importancia relativa comprendido en el intervalo [50,32; 88,47], aunque el método Perturb se diferencia de los tres métodos restantes en el hecho de que proporciona unos valores de importancia relativa más bajos ([50,32; 73,81]). Los métodos Profile y CW presentan unos valores de importancia relativa para esta variable muy similares, comprendidos en los intervalos [70,02; 84,98] y [69,13; 84,44], respectivamente; mientras que el método PaD proporciona valores intermedios del [58,53; 88,47] y en un rango más amplio que los demás métodos. En cuanto a la desviación típica respecto a la media, el método Perturb registra el rango de valores más elevado ([22,56; 40,09]), el método Profile el rango más bajo ([11,68; 29,50]), y los métodos CW y PaD valores intermedios entre los obtenidos por los otros dos métodos, de [17,84; 28,83] y [20,81; 32,10] respectivamente.

En cuanto a la variable INFORMACIÓN, los métodos Profile, CW y PaD presentan porcentajes de importancia relativa muy similares, comprendidos en el rango [58,15; 77,60], mientras el método Perturb presenta unos valores de importancia relativa considerablemente más bajos ([37,57; 55,46]). Los valores de desviación típica más bajos son los que ofrecen los resultados del método Profile ([8,88; 17,89]), los más altos los que presenta el método PaD ([18,88; 30,85]), mientras que los métodos Perturb ([14,32; 30,67]) y CW ([14,30; 24,45]) presentan valores intermedios.

El método que proporciona unos valores de importancia relativa mayores para la variable SEGURIDAD es el Profile ([49,75; 67,79]), seguido muy de cerca por los métodos CW ([45,48; 64,38]) y PaD ([41,04; 60,84]). De nuevo, el método Perturb se diferencia de los anteriores y proporciona unos valores de importancia más bajos ([31,13; 45,53]). El rango de valores de importancia relativa que ofrece una desviación típica más baja es el obtenido tras aplicar el método Profile ([15,91; 24,62]), mientras los restantes métodos presentan unos rangos de desviaciones típicas muy parecidos, del [18,28; 32,05] para el método Perturb, del [18,82; 28,04] para el CW y del [20,30; 30,89] para el PaD.

El mismo patrón en cuanto a asignación de importancia relativa se repite para la variable PROXIMIDAD, ya que el método Profile asigna unos valores ligeramente más elevados ([56,10; 67,75]) que los métodos CW ([53,66; 65,13]) y PaD ([48,91; 66,73]), aunque el rango de importancia extraído por estos tres métodos es notablemente superior al obtenido por el método Perturb ([21,25; 30,02]). Respecto a las otras variables analizadas, en este caso sí existe una jerarquía clara en cuanto al rango de desviación típica obtenido tras aplicar cada método. Así, el método PaD ofrece un espectro más elevado ([17,01; 27,04]), seguido del método CW ([14,76; 24,51]) y de los métodos Profile ([9,05; 20,31]) y Perturb ([8,17; 15,43]).

Para la variable PUNTUALIDAD, los rangos asignados para la importancia relativa siguen el siguiente orden: [42,08; 68,35] para el PaD, [45,05; 56,28] para el CW, [44,56; 55,69] para el Profile y [20,52; 32,92] para el Perturb. En cuanto a la desviación típica, los rangos obtenidos son: [17,10; 30,52] para el PaD, [15,48; 26,49] para el CW; [10,55; 28,12] para el Perturb y [10,75; 19,75] para el Profile.

Las variables TRATO, TEMPERATURA, ESPACIO y PRECIO presentan rangos de valores de importancia relativa muy parecidos entre sí, según los métodos PaD, CW y Perturb, y a su vez visiblemente superiores que los del método Perturb, mientras que los rangos de desviaciones típicas presentan diferencias más notables, siendo menor esta diferencia para la variable trato. Así, los rangos de importancia relativa por métodos para el TRATO son: [35,84; 60,67] para el PaD; [37,15; 51,35] para el Profile, [36,33; 47,81] para el CW y [17,00; 34,05] para el Perturb; para la TEMPERATURA: [25,63; 52,65] para el PaD; [46,41; 24,44] para el Profile; [44,49, 23,41] para el CW y [8,22; 25,41] para el Perturb; para el ESPACIO: [44,95; 25,65] para el PaD, [36,45; 26,28] para el CW; [25,68; 35,55] para el Profile y [9,23; 18,22] para el Perturb; para el PRECIO: [26,37; 45,43] para el PaD; [24,55; 35,50] para el CW; [24,26; 36,40] para el Profile y [7,84; 17,90] para el Perturb. Los rangos de desviación típica por métodos para el TRATO son: [19,57; 30,34] para el PaD; [15,81; 27,49] para el CW, [15,23; 27,59] para el Profile y [13,09; 32,54] para el Perturb; para la TEMPERATURA: [17,55; 28,03] para el PaD; [14,82; 24,57] para el CW; [13,27; 23,98] para el Profile y [8,57; 22,64] para el Perturb; para el ESPACIO: [8,58; 39,99] para el Perturb, [15,39; 29,33] para el PaD, [15,29, 23,69] y para el CW; [11,38; 24,03] para el Profile; para el PRECIO: [16,50; 27,81] para el PaD; [13,72; 20,34] para el CW; [10,05; 19,26] para el Profile y [6,80; 24,11] para el Perturb.

En las dos últimas variables en el ranking de importancia relativa, LIMPIEZA y ACCESIBILIDAD, no se repite la pauta de que el método Perturb asigna un rango de valores de importancia relativa notablemente a los tres restantes, con los cuales se obtienen rangos bastante similares, sino que existe una diferenciación más clara. En el caso de la variable LIMPIEZA, estos rangos de valores aparecen ordenados por métodos, de mayor a menor, en el siguiente orden: [19,51; 45,64] para el PaD; [16,22; 27,39] para el CW; [3,42; 15,68] para el Perturb y [1,35; 11,59] para el Profile. El orden para las importancias relativas también aparece más diferenciado: [15,89; 26,64] para el PaD; [13,61; 31,13] para el Profile, [11,66; 21,15] para el CW y [3,85; 24,73] para el Perturb. Del mismo modo, si se ordenan los rangos de valores de la variable ACCESIBILIDAD siguiendo el mismo criterio, para la importancia relativa se obtiene: [13,56; 46,25] para el PaD; [11,93; 22,98] para el CW; [8,76; 19,55] para el Profile y [2,37; 10,30] para el Perturb. Para la desviación típica los rangos por métodos siguen la siguiente jerarquía: [9,68; 27,82] para el PaD; [8,42; 18,64] para el CW; [8,60; 17,88] para el Profile y [3,08; 16,02] para el Perturb.

Tras analizar estos resultados, se puede afirmar que la elección en el número H de neuronas en la capa oculta no influye considerablemente en los resultados de importancia relativa obtenidos para cada variable, independientemente del método de cálculo de importancia relativa utilizado (Perturb, Profile, CW y PaD), con lo que es posible trabajar directamente con un único conjunto H de ANN concreto, sin necesidad de considerar un abanico de posibles grupos (H e [1,30]) para seleccionar el que presenta un mejor comportamiento global, ya que no se van a producir variaciones importantes en los resultados. De los cuatro métodos de cálculo de importancia relativa, los métodos Profile y CW son los más estables, puesto que proporcionan valores de porcentaje de importancia relativa más parecidos para cada una de las variables consideradas, independientemente

del conjunto H en el que han sido aplicados. El método que proporciona más disparidad en los resultados de importancia relativa, y por tanto es el menos estable, es el PaD. Por último, los resultados también ponen de manifiesto que el método Perturb proporciona unos valores de importancia relativa que son en general notablemente más bajos que los obtenidos por los tres métodos restantes para todas las variables, excepto para la FRECUENCIA, a la cual asigna una importancia relativa máxima, al igual que los restantes métodos, y para la variable LIMPIEZA, en el que asigna un porcentaje de importancia relativa muy similar al del método Profile.

6.5. Comparación entre las técnicas de ANN y DT aplicadas a la calidad del servicio

Se ha desarrollado un modelo de DT, obteniéndose una varianza explicada del 49,7% de la varianza total, lo que indica que la exactitud del modelo es baja (de Oña et al., 2012). La Tabla 17 muestra el ranking de importancia de las variables según el ratio normalizado obtenido tras aplicar las pautas metodológicas expuestas en el apartado 4.1.3, así como el ratio de importancia medio para cada atributo. El modelo ha sido calibrado usando el software SPSS.

ATRIBUTO	RATIO NORMALIZADO	RANKING
INF	61,6	5
PUN	86,3	3
SEG	60,6	6
TRA	59,4	7
LIM	38,1	11
ESP	45,8	8
TEM	65,2	4
ACC	21,6	12
PRE	42,6	9
VEL	86,8	2
FRE	100	1
PRO	41,6	10

Tabla 17. Ratio de importancia relativa (%) y ranking de los atributos de calidad del servicio según la técnica DT

Según este método, los factores clave que influyen en la satisfacción global del usuario son aquellos relacionados con la operación del servicio. Al igual que se hizo con los cuatro métodos de cálculo de importancia relativa de las ANN, la importancia relativa extraída del modelo de DT se ha normalizado asignando un valor del 100% al atributo con el mayor ratio de importancia relativa, para poder comparar estos resultados con los de la técnica de ANN. Así, la FRECUENCIA, VELOCIDAD y PUNTUALIDAD presentan valores de importancia normalizada por encima del 85%. La TEMPERATURA, INFORMACIÓN, SEGURIDAD y TRATO también son importantes, con ratios del orden del 60% (65,23%,

61,61%, 60,59% y 59,42%, respectivamente). El resto de variables tienen un impacto menor en la satisfacción, y están relacionados con el PRECIO, el confort (ESPACIO y LIMPIEZA) y la accesibilidad al servicio (ACCESIBILIDAD y PROXIMIDAD). De hecho, la ACCESIBILIDAD es la variable menos importante según este modelo de DT (21,59%).

Esta investigación encuentra algunas diferencias de comportamientos entre ambas técnicas de minería de datos (ANN y DT) en cuanto a la exactitud de los modelos, del 95% de las ANN frente al 49,7% del modelo DT. Estos resultados son consistentes con los obtenidos por Xie et al. (2003) y Lee et al. (2010), que compararon el comportamiento de ambas técnicas en otros campos de investigación.

La Tabla 18 muestra las principales ventajas y desventajas de ambos métodos.

Como se ha mencionado anteriormente, y como era de esperar, ambas técnicas proporcionan resultados similares respecto a los factores clave que definen un servicio de transporte público por autobús eficiente. Estos factores son la FRECUENCIA y la VELOCIDAD de operación. Por tanto, ambas metodologías deberían considerarse válidas y útiles en los procesos de planificación y en las fases de operación y dirección de estos servicios. Otros estudios, en los que se han aplicado otras técnicas, también ponen de manifiesto la importancia de la FRECUENCIA del servicio (de Oña et al., 2012; Dell’Olio, 2010; 2011; Del Castillo and Benítez, 2013; Tyrinopoulos and Antoniou, 2008). Por otro lado, la ACCESIBILIDAD y la LIMPIEZA son las características que menos influyen en la valoración global de la calidad del servicio para ambas técnicas. Eboli y Mazzulla (2008) también identificaron la LIMPIEZA como la variable con menor influencia sobre los usuarios del transporte urbano de autobús en Cosenza (Italia). Las diferencias más notables entre las distintas técnicas empleadas aparecen en las variables con un nivel de importancia media. En este caso, el ranking de importancia oscila entre varias posiciones en función de la metodología aplicada.

Si se comparan los ratios de importancia relativa obtenidos con las técnicas ANN y DT y aquéllos extraídos por los usuarios de las encuestas (Tabla 3), se observan algunas diferencias. Aunque existe poca variación en la importancia asignada por los pasajeros a los atributos de calidad del servicio, ya que consideran que todos ellos son importantes, en el ranking que establecen, la VELOCIDAD ocupa la séptima posición, mientras que ambas técnicas de minería de datos la consideran como una variable clave para la calidad del servicio de transporte metropolitano. Por otra parte, las variables ACCESIBILIDAD y LIMPIEZA, que fueron consideradas como las menos influyentes por ambas técnicas (ANN y DT), ocupan las posiciones cuarta y quinta, respectivamente, en el ranking de importancia. Esta falta de consenso también fue puesta de manifiesto por Weinstein (2000) en un estudio de la importancia de las variables basado también en las opiniones de los pasajeros.

	DT	ANN
VENTAJAS	<ul style="list-style-type: none"> • Baja complejidad en el procedimiento de cálculo de los ratios de importancia. • Poco tiempo computacional requerido para el cálculo de la importancia de las variables (segundos). • Simplicidad del modelo. • Resultados fácilmente interpretables debido a la posibilidad de representación gráfica. • Puede extraerse información de las reglas “si – entonces”. • El método no se ve afectado por las relaciones de dependencia entre las variables predictoras. 	<ul style="list-style-type: none"> • Elevada capacidad de exactitud de los modelos. • Mayor estabilidad en la determinación del cálculo de importancia relativa de las variables. • El método no se ve afectado por las relaciones de dependencia entre las variables predictoras.
DESVENTAJAS	<ul style="list-style-type: none"> • Bajos ratios de exactitud del modelo. • Inestabilidad de los modelos generados. • Las reglas de decisión no pueden revisarse o mejorarse (no existe una técnica de revisión de las mismas). • Los modelos no arrojan una significancia estadística de las variables. 	<ul style="list-style-type: none"> • Requiere un pretratamiento de los datos. • Elevada complejidad de los métodos de cálculo de los ratios de importancia relativa. • Más tiempo requerido para determinar la importancia relativa de las variables (del orden de horas). • El procedimiento para determinar la importancia relativa de las variables es tedioso. • Los modelos no arrojan una significancia estadística de las variables.

Tabla 18. Comparación de las ventajas y desventajas de las técnicas de ANN y DT

6.6. Discusión

En esta tesis doctoral se ha aplicado la técnica de ANN para analizar la calidad del servicio de transporte público por autobús. Esta metodología, aplicada por primera vez a este campo concreto del transporte, permite predecir la calidad global del servicio, así como determinar la influencia de los atributos o variables características que se han utilizado para su definición, lo que proporciona a los operadores de transporte público una nueva herramienta para analizar las fortalezas y debilidades que ofrecen al usuario a través del servicio prestado, lo que a su vez permite plantear mejoras concretas en aquellos aspectos que son detectados como mejorables o mal valorados. Con ello, podrá ofrecerse un mejor servicio a los actuales pasajeros y mejorar la competitividad de este modo de transporte frente a otros.

Esta investigación, al igual que otros estudios previos (Mussone et al., 1999; Delen et al., 2006; Mohammadipour y Alavi, 2009; Moghaddam et al., 2010; Akin y Akbaç, 2010), confirma la elevada capacidad de predicción y generalización de la técnica de las ANN, puesto que se obtiene un error de generalización, medido a través del MAPE, del orden de 0,05, lo que indica que los modelos de ANN entrenados son capaces de proporcionar al menos un 95% de salidas correctas.

Por otra parte, se aborda el problema de la inestabilidad que presentan los métodos de cálculo de la importancia relativa de las variables predictoras consideradas en los modelos ANN, ya expuesto previamente por varios autores (Gedeon, 1997; Sung 1998; Palmer y Montaña, 2002b; Olden y Jackson, 2002; Gevrey et al., 2003; Paliwal y Kumar, 2011). Hasta el momento no ha sido posible establecer un consenso sobre qué método es más estable, exacto o robusto, ni cuál proporciona resultados más ajustados a la realidad, lo que ha sembrado la duda de si estos métodos, que determinan la relación causa-efecto entre cada variable predictora – variable dependiente, son realmente válidos para clarificar la función que cada variable representa en un modelo de ANN (también denominado modelo de caja negra), y aún más importante, si es realmente útil aplicar esta técnica en aquellos campos de investigación en los que los resultados obtenidos pueden tener consecuencias serias y comprometedoras. Los valores de importancia relativa obtenidos tras aplicar los métodos Perturb, Profile, CW y PaD a una única ANN, muestran una variabilidad muy elevada, no sólo cuando se aplican diferentes métodos, sino cuando se aplica el mismo método a una misma arquitectura de ANN entrenada partiendo de diferentes valores de pesos iniciales aleatorios. Por tanto, trabajando de esta forma no existe garantía de que los resultados de importancia relativa obtenidos sean válidos.

Sin embargo, en esta investigación se ha trabajado con conjuntos de ANN con la misma arquitectura sobre los que se han aplicado diferentes métodos de cálculo de la importancia relativa, extrayéndose valores medios. Este procedimiento consigue obtener resultados de importancia relativa similares, independientemente del método utilizado, especialmente similares para los métodos Perturb y Profile.

Siguiendo esta metodología, cualquiera de los métodos de cálculo de la importancia relativa es válido, ya que todos ellos coinciden en identificar las variables más importantes y las menos importantes, aunque el método PaD es el menos recomendado de los cuatro analizados, puesto que muestra una mayor variabilidad en los valores de importancia relativa extraídos. Además, otro argumento que refuerza la robustez de esta nueva metodología es el hecho de que las variables FRECUENCIA, VELOCIDAD, PUNTUALIDAD y PROXIMIDAD son clasificadas como las más importantes tras aplicar los cuatro métodos, lo que concuerda con los resultados obtenidos por otros autores (Eboli y Mazulla, 2008; Eboli y Mazulla 2010; Dell’Olio et al.; 2010; Dell’Olio et al.; 2011; De Oña et al., 2012) que han aplicado otras técnicas, como modelos de regresión multinomial logística, modelos de elección discreta, modelos probit o ecuaciones estructurales.

Otro escollo que salva esta la metodología de cálculo de importancia relativa de las variables es la dificultad de seleccionar una ANN o conjunto de ANN óptimo o sub-óptimo, ya que, como es bien conocido en la literatura científica a través de los trabajos desarrollados por diversos autores (Cao y Qiao, 2008; Paliwal y Kumar, 2011), existen numerosas posibilidades de entrenamiento y de obtención de resultados diferentes de los modelos de ANN entrenados, sea cual sea su tipología, con lo que es muy complicado afirmar que en una determinada investigación se ha encontrado y entrenado el mejor modelo. Con este nuevo procedimiento, se ha demostrado que al trabajar con conjuntos de ANN, las oscilaciones en cuanto al porcentaje de importancia relativa de las variables predictoras del modelo se mitiga notablemente, y se pueden conseguir resultados parecidos. La principal debilidad de esta metodología es que, aunque los cuatro métodos proporcionan un ranking de importancia similar, los porcentajes de importancia relativa difieren fundamentalmente en las variables de importancia media, siendo necesaria por tanto la opinión de un experto para decidir qué método ofrece unos resultados más similares a los esperados. En general, el método Perturb asigna una importancia relativa más baja a las variables, mientras que los tres métodos restantes presentan valores más elevados y similares entre sí.

El análisis estadístico elaborado en esta investigación también revela la existencia de grupos de variables homogéneos en cuanto a diferencias significativas para los métodos Perturb, Profile y PaD analizados. En ellos queda claramente diferenciada la variable FRECUENCIA, como elemento exclusivo de un grupo homogéneo estadísticamente diferenciado del resto para todos los métodos estudiados, y las variables LIMPIEZA, ESPACIO, TEMPERATURA y ACCESIBILIDAD como pertenecientes a uno o dos grupos homogéneos (según el método de cálculo de la importancia relativa), también diferenciados estadísticamente de los demás, que ocupan las últimas posiciones en la jerarquía de importancia. Este análisis, que además concuerda con otros análisis estadísticos realizados a grupos de variables similares en otros estudios que aplicaban otras técnicas (Eboli y Mazzulla, 2008), refuerza la validez, tanto de la técnica de las ANN para el estudio de la calidad del servicio en el transporte público por autobús, como de la metodología de cálculo de importancia relativa de las variables propuesta y desarrollada en esta investigación.

CAPÍTULO 7. CONCLUSIONES Y LÍNEAS FUTURAS DE INVESTIGACIÓN

7.1. Conclusiones

En este capítulo se exponen las principales conclusiones de esta tesis doctoral.

En esta tesis doctoral se propone el empleo de Redes Neuronales Artificiales, una técnica no convencional de minería de datos, para estudiar la calidad del servicio en el transporte público metropolitano por autobús, debido a las considerables ventajas que presenta frente a otras técnicas convencionales paramétricas. Estas ventajas residen fundamentalmente en su elevada capacidad de predicción, generalización y para captar relaciones complejas y no lineales existentes entre las variables consideradas en un estudio. Además, no necesitan partir de un modelo definido a priori, ni la precisión de sus resultados se ve afectada por las relaciones de dependencia entre las variables.

Las Redes Neuronales Artificiales han demostrado, a través de numerosos estudios en los que han sido aplicadas, proporcionar buenos resultados cuando trabajan con variables difusas, con contenido subjetivo e incluso en las que falta parte de la información y su elevada tolerancia a los fallos presentes en los datos de partida, por lo que las cualidades de esta técnica se adecuan bien a las características de las variables y datos con los que se analiza la calidad del servicio en el transporte público por autobús.

Los datos utilizados para el estudio de la calidad del servicio provienen de una encuesta de satisfacción realizada a los usuarios del transporte público por autobús del área metropolitana de Granada en el año 2007. Partiendo de ellos, se han desarrollado diferentes conjuntos de Redes Neuronales Artificiales, siendo las principales conclusiones obtenidas tras su aplicación las siguientes:

- Las Redes Neuronales Artificiales son una técnica que proporciona unos resultados de exactitud superiores a los obtenidos con las técnicas utilizadas hasta el momento en el campo del análisis de la calidad del servicio del transporte público. Estos valores de exactitud coinciden con los alcanzados en otros estudios de calidad del servicio no relacionados con el transporte público en los que se ha aplicado esta técnica (Mahapatra y Khan, 2006; Lin, 2007; Deng et al., 2008; Deng y Pei, 2009; Larasati et al., 2012; Azadeh et al., 2011).
- La metodología propuesta en este trabajo de investigación para calcular la importancia relativa de las variables predictoras consideradas, trabajando con conjuntos de redes en lugar de con una sola de comportamiento sub-óptimo y con los valores medios de importancia relativa obtenidos tras aplicar varios métodos de cálculo de importancia tradicionales, mitiga considerablemente los problemas de inestabilidad presentes en estos métodos hasta el momento. Esta nueva metodología puede ser aplicada no sólo al campo de estudio de la calidad del servicio en el transporte público, sino a cualquier otro estudio desarrollado utilizando la técnica de las Redes Neuronales Artificiales. Además, en este estudio ha quedado demostrada la robustez de esta metodología, que permite minimizar el problema también consensuado en la literatura científica de tener que seleccionar la arquitectura de red óptima que presente mejor comportamiento en un determinado estudio, ya que la gran cantidad de posibilidades de entrenar redes es tan amplia, que en muchas ocasiones resulta complicado toparse con la que

presenta mayor exactitud y capacidad de predicción. Esta nueva metodología, al trabajar con conjuntos de redes, ofrece la posibilidad de obtener resultados similares independientemente de la arquitectura de red seleccionada.

- Se ha comprobado la robustez y estabilidad de esta nueva metodología, e modo que el procedimiento de cálculo de importancia relativa de las variables de estudio puede agilizarse considerablemente siendo posible trabajar directamente con un conjunto de ANN con un número determinado de neuronas en su capa oculta, sin necesidad de entrenar otros conjuntos con distinto número de neuronas en sus capas ocultas y seleccionar el que presenta un mejor comportamiento global. Esta mejora hace que esta técnica sea más competitiva, al disminuirse notablemente el tiempo necesario para su aplicación.
- Las Redes Neuronales Artificiales es una herramienta válida para el estudio de la calidad del servicio en el transporte público por autobús, y los resultados de importancia relativa obtenidos tras aplicar esta técnica concuerdan y son coherentes con los obtenidos en otros estudios de calidad del servicio de transporte público en los que se han aplicado otras técnicas diferentes a ésta (Dell'Ollio et al. 2010, 2011; de Oña et al., 2012; Eboli y Mazzulla, 2010).
- Se ha aplicado un análisis estadístico Kruskal-Wallis en el que se han identificado grupos homogéneos de atributos con diferencias estadísticamente significativas entre ellos desde el punto de vista de su influencia en la valoración global de la calidad del servicio, que coinciden con los grupos homogéneos identificados en otros estudios de servicio de transporte público por autobús (Eboli y Mazzulla, 2008), o con los genéricamente establecidos en varios manuales conocidos a nivel internacional (Comité europeo de estandarización, 2002; Transportation Research Board, 2004). De todos estos resultados se extrae como conclusión que las variables frecuencia y proximidad pueden considerarse atributos básicos, ya que tienen el mayor impacto en la valoración global de la calidad del servicio, mientras que la limpieza, la accesibilidad y el trato son las que menos impacto tienen en dicha valoración. En cambio, tras realizar este mismo análisis estadístico al método PaD, no es posible establecer grupos homogéneos que presenten diferencias significativas respecto a los demás.
- Los resultados de este estudio no deberían generalizarse a otro tipo de servicios de transporte público, puesto que las características de estos servicios y las necesidades requeridas por sus usuarios pueden diferir considerablemente de las que presenta este estudio. Sin embargo, el uso de esta técnica y de la metodología para el cálculo de la importancia relativa de los atributos que definen el servicio, pueden ser aplicados para estudiar la calidad del servicio en cualquier tipo y contexto de transporte público.
- Debe resaltarse que la encuesta de la que se ha extraído la información utilizada en este estudio, fue realizada por la entidad gestora del servicio para un fin ajeno a este estudio, esto es, para realizar un análisis de frecuencia estadística. Sin embargo, la aplicación de esta técnica no convencional y novedosa demuestra que este tipo de datos pueden revelar aspectos muy interesantes para los operadores del transporte público, y que estos estudios podrían incrementar la colaboración

entre empresas gestoras y centros de investigación, puesto que esta es una relación de simbiosis para ambas entidades.

- La comparación realizada entre las técnicas de minería de datos ANN y DT, confirma que ambas son válidas para estudiar la calidad del servicio en el transporte público por autobús, e incluso podrían considerarse complementarias, ya que algunas de las fortalezas de una pueden suplir algunas de las debilidades de la otra, es decir, aspectos como la elevada capacidad de predicción, la tolerancia a fallos o la estabilidad de los resultados son claramente más ventajosos si se aplica la técnica de las ANN. Sin embargo, aspectos como la complejidad computacional y el tiempo de desarrollo necesario para extraer resultados aconsejan el uso de DT. En función de la finalidad perseguida por los gestores/planificadores del servicio, puede optarse por utilizar una u otra técnica.
- Esta técnica proporciona a los planificadores y gestores de un servicio de transporte público de esta tipología la posibilidad de enfocar sus políticas y programas de actuación hacia aquellos aspectos que refuercen las posibilidades de obtener sus objetivos específicos.
- Las principales limitaciones halladas con la aplicación de Redes Neuronales Artificiales a la calidad del servicio en el transporte público por autobús residen en el hecho de que se han encontrado discrepancias en los valores de importancia relativa de algunos de los atributos considerados en el estudio, dependiendo del método de cálculo de importancia relativa utilizado, siendo más evidentes estas diferencias en cuanto al ranking de importancia relativa asignado a cada atributo en los métodos de derivadas parciales (PaD) y CW, mientras que los resultados de obtenidos en los métodos Perturb y Profile son mucho más parecidos. En cuanto a porcentajes de importancia relativa asignados a los atributos, el método Perturb, en general, proporciona unos valores más bajos que los obtenidos por los métodos restantes para todos los atributos de estudio. Además, el procedimiento de cálculo de estos valores de importancia relativa es tedioso, siendo necesario emplear mucho tiempo para la obtención de estos resultados, del orden de horas, mientras que en la mayoría de las técnicas usadas en este campo de estudio hasta la fecha sólo es necesario emplear minutos.
- Esta investigación puede ser útil para los planificadores y gestores de los servicios de transporte urbano por autobús, en una serie de aspectos:
 - Primero, para predecir el nivel de calidad del servicio prestado a los usuarios de este modo de transporte.
 - Segundo, para determinar la importancia de los atributos que caracterizan la calidad del servicio, proporcionando así una información valiosa y práctica.
 - Tercero, para servir como herramienta o técnica adicional a la ya aplicada en otros estudios de calidad del servicio a través de DT, complementando sus limitaciones y permitiendo comparar los resultados obtenidos mediante ésta, y por tanto su fiabilidad.

7.2. Futuras líneas de investigación

De la elaboración de esta tesis doctoral y tras analizar las principales conclusiones extraídas de este trabajo de investigación, es posible plantear el desarrollo de diversos estudios en un futuro cercano. Por tanto, puede ser interesante desarrollar las siguientes líneas de trabajo:

- En cuanto a la nueva metodología desarrollada en este trabajo de investigación para calcular la importancia relativa de las variables dependientes, podría estudiarse lo que ocurre cuando se aplica ésta a otros métodos de cálculo de importancia relativa existentes, tanto los que se han utilizado repetidamente en numerosos estudios a lo largo de las dos últimas décadas, como otros más recientes, como el métodos DSA (Cortez y Embrechts, 2013).
- También podría comprobarse la robustez de esta metodología cuando se utilizan otras funciones de activación u otros algoritmos de aprendizaje para el perceptrón multicapa, y extender este estudio a otras tipologías de Redes Neuronales Artificiales.
- Al igual que se ha realizado una comparación entre las técnicas de Redes Neuronales Artificiales y Árboles de Decisión, podría realizarse una comparación con los Modelos de Ecuaciones Estructurales, que son de uso extendido y validez reconocida por numerosos investigadores de la calidad del servicio en el transporte público.
- Reducir la heterogeneidad de los usuarios encuestados mediante la aplicación previa de otras técnicas como análisis cluster o métodos estadísticos, que permitan establecer grupos homogéneos a los que posteriormente se aplicará la técnica de las Redes Neuronales Artificiales. Este paso previo puede mejorar la exactitud de los modelos entrenados, y además proporcionar resultados interesantes y diferentes a los obtenidos en este estudio para algunos atributos que caracterizan la valoración de la calidad del servicio.
- Las Redes Neuronales Artificiales han sido aplicadas en este estudio al transporte público metropolitano por autobús, sin embargo podrían aplicarse al estudio de la calidad del servicio en otros modos de transporte público, estableciéndose los acuerdos de colaboración necesarios entre administraciones y gestores del servicio y universidad, y comparar los resultados obtenidos entre modos.

CAPÍTULO 8. REFERENCIAS BIBLIOGRÁFICAS

Abalo, J., Varela, J. y Manzano, V. (2007). Importance values for importance-performance analysis: a formula for spreading out values derived from preference rankings. *Journal of Business Research*, 60 (2), 115-121.

Abdlla, G., Mohamed A.R. y Mekawy, M.A. (2007) Managing tourists' needs and expectations: An empirical analysis of the Egyptian airline sector. *Tourism* 55(3):277–296.

Akin, D. y Akbaç, B. (2010). A neural network (NN) model to predict intersection crashes based upon driver, vehicle and roadway surface characteristics. *Scientific Research and Essays*, 5(19), 2837-2847.

Aksoy, S., Atilgan, E. y Akinci, S. (2003). Airline services marketing by domestic and foreign firms: differences from the customers' viewpoint. *Journal of Air Transport Management*, 9, 343-351.

Andreassen, T.W. (1995). Satisfaction with public services: The case of public transportation. *J. Services Marketing* 9 (5), 30–41.

Armstrong, G., y Kotler, P. (2000). *Marketing: An introduction*. New York: Prentice-Hall.

Awasthi, A., Chauhan, S.S., Omrani, H. y Panahi, A. (2011). A hybrid approach based on SERVQUAL and fuzzy TOPSIS for evaluating transportation service quality. *Computing. Industry Engineering*, 61 (3), 637–646.

Azadeh, A., Rouzbahman, M., Saberi, M. y Fam, I. M. (2011). An adaptative neural network algorithm for assessment and improvement of job satisfaction with respect to HSE and ergonomics program: the case of a gas refinery. *Journal of Loss Prevention in the Process Industries*, 24, 361-370.

Babakus, E. y Boller, G.W. (1992). An empirical assessment of the SERVQUAL scale. *J. Bus. Res.* 24(3):253–268.

Beale, M.H., Hagan, M.T. y Demuth, H.B. (2007). *Neural Network Toolbox 7. User's Guide*. MathWorks, Inc. 3 Apple Hill Drive Natic, MA 01760-2098.

Bejou, D., Wray, B. y Ingram, T.N. (1996). Determinants of relationships quality: an artificial neural network analysis. *Journal of Business Research*, 36, 137-143.

Bishop, C.M. (1995). *Neural networks for pattern recognition*. Oxford University Press.

Bordagaray, M, Dell'olio, L, Ibeas, A y Cecín, P. (2013). Modelling user perception of bus transit quality considering user and service heterogeneity. *Transportmetrica*, DOI: 10.1080/23249935.2013.823579.

Brady, M.K. y Cronin Jr., J.J. (2001). Some new thoughts on conceptualizing perceived service quality: a hierarchical approach. *Journal of Marketing*, 65 (3), 34-49.

Breiman, L., Friedman, J.H., Olshen, R.A. y Stone, C.J. (1984). *Classification and Regression Trees*. Wadsworth, Belmont (CA).

Bullock, D., Garrett, J., Hendrickson, C. y Pearce, A. (1992). A neural network for image base vehicle detection. Proceedings International Conference on Artificial Intelligence Applications in Transportation Engineering, San Buenaventura, CA.

Bullock D., Garrett, J. y Hendrickson, C. (1993) A neural network of image-based vehicle detection. Transportation. Research Part C, 1, 2355247.

Cao, M. y Qiao, P. (2008). Neural network committee-based sensitivity analysis strategy for geotechnical engineering problems. Neural Computing and Applications, 17, 509-519.

Carman, J. (1990). Consumer perceptions of service quality: an assessment of the SERVQUAL dimensions. Journal of Retailing, 66, 33-55.

Carrillat, F.A., Jaramillo, F. y Mulki, J.P. (2007).The validity of the SERVQUAL and SERVPERF scales a meta-analytic view of 17 years of research across five continents. International Journal of Service Industry Management 18 (5), 472–490.

Carter, D.N. y Lomar, T.J. (1992). Development and application of performance measures for rural public transportation operators. Transportation Research Record, 1338, 28-36.

Castillo, J.M.y Benítez, F.G. (2012). Determining a public transport satisfaction index from user surveys. Transportmetrica, 1-29, iFirst.

Chang, H., Liang, G., Chu, C. y Chou, C. (2012). Prioritizing service attributes for improvement using fuzzy zone of tolerance. International Journal of Innovative Computing Information and Control, 8 (1A), 75-89.

Chang, Y.H. y Yeh, C.H. (2002). A survey analysis of service quality for domestic airlines. European Journal Operational Research, 139 (1), 166–177.

Chau, V.S. y Kao, Y.Y. (2009). Bridge over troubled water or long and winding road? Gap-5 in airline service quality performance measures. Managing Service Quality, 19 (1),106–134.

Chen, C. (2008). Investigating structural relationships between service quality, perceived value, satisfaction, and behavioral intentions for air passengers: Evidence from Taiwan Transportation Research Part A, 42 (4), 709-717.

Chen, F.Y. y Chang, Y.H. (2005). Examining airline service quality from a process perspective. Journal of Air Transport Management, 11, 79-87.

Cheng, J-H., Chen, F-Y. y Chang, Y-H. (2008). Airline relationship quality: an examination of Taiwanese passengers. Tourism Research, 29, 487-499.

Chou, J.S. y Kim, C. (2009). A structural equation analysis of the QSL, relationship with passenger riding experience on high speed rail: an empirical study of Taiwan and Korea. Expert Systems with Applications, 36, 6945-6955.

Chou, J., Kim, C., Kuo, Y. y Ou, N. (2011a). Deploying effective service strategy in the operations stage of high-speed rail. Transportation Research Part E, 47 (4), 507-519.

Chou, CC, Liu L.J., Huang, S.F., Yih, J.M. y Han, T.C. (2011b) An evaluation of airline service quality using the fuzzy weighted SERVQUAL method. *Application Software Computing*. 11(2):2117–2128.

Christopher, M.K., Stuart, D. y Foote, P.J. (1999). Structuring and assessing transit management response to customer satisfaction surveys. *Transportation Research Record*, 1669, 99-108.

Cortez, P. y Embrechts, M.J. (2011). Opening black box data mining models using sensitivity analysis. *IEEE Synopsium Series in Computational Intelligence*, Paris, France, 4, 2011.

Cortez, P. y Embrechts, M.J. (2013). Using sensitivity analysis and visualization techniques to open black box data mining models. *Information Sciences*, 225, 1-17.

Costa, A. y Markellos, R.N. (1997). Evaluating public transport efficiency with neural network models. *Transportation Research, part C: Emerging Technologies*, 301-312.

Cronin, J. y Taylor, S. (1992). Measuring service quality: a reexamination and extension. *Journal of Marketing*, 56, 193-218.

Das, S. y Pandit, D. (2013). Importance of user perception in evaluating level of service for bus transit for a developing country like India: a review, *Transport Reviews: A Transnational Transdisciplinary Journal*, DOI:10.1080/01441647.2013.789571.

Del Castillo, J.M. y Benitez, F.G. (2013). Determining a public transport satisfaction index from user surveys. *Transportmetrica A: Transport Science*, 9(8), 713-741.

Delen, D., Sharda, R. y Bessonov, M. (2006). Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. *Accident Analysis and Prevention*, 38, 434-444.

del Hoyo, R., Martín del Brío, B., Medrano, N. y Fernández-Navajas, J. (2009). Computational intelligence tools for next generation quality of service management. *Neurocomputing*, 72, 3631-3639.

Dell'Olío, L., Ibeas, A. y Cecín, P. (2010). Modelling user perception of bus transit quality. *Transport Policy*, 17(6), 388-397.

Dell'Olío, L., Ibeas, A. y Cecín, P. (2011). The quality of service desired by public transport users. *Transport Policy*, 18(1), 217-227.

Deng, W-J., Chen, W-C. y Pei, W. (2008). Backpropagation neural network based importance-performance analysis for determining critical service attributes. *Expert Systems with Applications*, 34, 1115-1125.

Deng, W-J. y Pei, W. (2009). Fuzzy neural based importance-performance analysis for determining critical services attributes. *Experts Systems with Applications*, 36, 3774-3784.

de Oña, J., de Oña, R. y Calvo, F. J. (2012). A classification tree approach to identify key factors of transit service quality. *Expert Systems with Applications*, 39, 11164–11171.

de Oña, J., de Oña, R., Eboli, L. y Mazzulla, G. (2013). Perceived service quality in bus transit service: A structural equation approach. *Transport Policy* 29: 219-226.

de Oña, J. y de Oña, R. (2015). Quality of service in public transport based on customer satisfaction surveys: a review and assessment of methodological approaches. *Transportation Science*, 49(3), 605-622.

de Oña, R., Eboli, L. y Mazzulla, G. (2014). Key factors affecting rail service quality in the northern Italy: a decision tree approach. *Transport*, 29(1), 75-83. DOI: 10.3846/16484142.2014.898216.

de Oña, J. y Garrido, C. (2014). Extracting the contribution of independent variables in neural network models: a new approach to handle instability. *Neural Computing and Applications*, 24(5). DOI: 10.1007/s00521-014-1573-5.

De Oña, R. (2013). Analysis of service quality in public transportation using decision trees. Tesis doctoral.

Dimopoulos, Y., Bourret, P. y Lek, S. (1995). Use of some sensitivity criteria for choosing networks with good generalization ability. *Neural Processing Letters*, 2, 1-4.

Dougherty, M. (1995). A review of neural networks applied to transport. *Transportation Research, Part C*, 3(4), pp 247-260.

Dougherty, M. S. y Joint, M. (1992). A behavioural model of driver route choice using neural networks. *Proceedings International Conference on Artificial Intelligence Applications in Transportation Engineering*, San Buenaventura, CA.

Drea, J.T. y Hanna, J.B. (2000). Niche marketing in intrastate passenger rail transportation. *Transportation Journal*, 39 (3), 33-43

Eboli, L. y Mazzulla, G. (2007). Service quality attributes affecting customer satisfaction for bus transit. *Journal of Public Transportation* 10(3):21-34.

Eboli, L. y Mazzulla, G. (2008). A Stated Preference Experiment for Measuring Service Quality in Public Transport. *Transportation Planning and Technology*, 31(5), 509-523.

Eboli, L. y Mazzulla, G. (2009). A new customer satisfaction index for evaluating transit service quality. *Journal of Public Transportation*, 12(3), 21-37.

Eboli, L. y Mazzulla, G. (2010). How to capture the passengers' point of view on a transit service through rating and choice opinions. *Transport Review*, 30, 435-450.

Eboli, L. y Mazzulla, G. (2011). A methodology for evaluating transit service quality based on subjective and objective measures from the passenger's point of view. *Transport Policy*, 18, 172-181.

Eboli, L. y Mazzulla, G. (2012). Structural Equation Modelling for Analysing Passengers' Perceptions about Railway Services. *Procedia-Social and Behavioural Science*, 54, 96-106.

Engelbrecht, A.P., Cloete, I. y Zurada, J.M. (1995). Determining the significance of input parameters using sensitivity analysis, from natural to artificial neural computation. *Proceedings of International Workshop on Artificial Neural Networks*, pp. 382-388. Málaga-Torremolinos, Spain, Springer.

Fielding, G.J. (1992). European Committee for Standardization (2002) Transport – Logistics and Services – Public Passenger Transport – Service Quality Definition Target and Measurement. European Standard EN 13816 (European Committee for Standardization (CEN), Brussels). *Transit performance evaluation in the USA. Transportation Research*, 26A (6), 483-491.

Figler, S.A., Sriraj, P.S., Welch, E.W. y Yavuz, N. (2011). Customer loyalty and Chicago, Illinois, transit authority buses. *Transportation Research Record*, 2216, 148-156.

Foote, P.J. y Stuart, D.G. (1998). Customer satisfaction contrasts express versus local bus service in Chicago's North Corridor. *Transportation Research Record*, 1618, 143-152.

Friman, M., Edvardsson, B. y Gärling, T. (2001). Frequency of negative critical incidents and satisfaction with public transport services, I. *Journal of Retailing and Consumer Service*, 8 (2), 95-104.

Friman, M. y Gärling, T. (2001). Frequency of negative critical incidents and satisfaction with public transport services, II. *Journal of Retailing and Consumer Services*, 8 (2), 105-114.

Funahashi, K.I. (1989). On the approximate realization of continuous mappings by neural networks. *Neural Networks*, 2, 183-192.

Ganesan-Lim, C., Russell-Bennett, R. y Dagger, T. (2008). The impact of service contact type and demographic characteristics on service quality perceptions. *Journal of Services Marketing*, 22 (7), 550-561.

Garson, G.D. (1991). Interpreting neural-network connection weights. *Artificial Intelligence Expert*, 6, 47-51.

Garver, M.S. (2002). Using data mining for customer satisfaction research. *Marketing research*, 14(1), 8-12.

Gedeon, T.D., Wong, P.M. y Harris, D. (1995). Balancing the bias and variance: network topology and pattern set reduction techniques. *Proceedings of International Workshop on Artificial Neural Networks, IWANN95*, 550-558, Torremolinos, España.

Gedeon, T.D. (1997). Data mining of inputs: analyzing magnitude of functional measures. *International Journal of Neural Systems*, 8(2), 209-218.

Gevrey, M., Dimopoulos, I. y Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling*, 160, 249-264.

Gevrey, M., Dimopoulos, I. y Lek, S. (2006). Two-way interaction of input variables in the sensitivity analysis of neural network models. *Ecological Modelling*, 195, 43-50.

Goh, A.T.C. (1995). Back-propagation neural networks for modeling complex systems. *Artificial Intelligence in Engineering*, 9, 143-151.

Grönroos, C. (1982). Strategic management and marketing in service sector, Marketing Science Institute, Cambridge, M.A.

Grönroos, C. (1984). A service quality model and its marketing implications. *European Journal of Marketing*, 18 (4), 36-44.

Grönroos, C. (1988). Service quality: the six criteria of good service quality, review of business. New York: St. John's University Press.

Hagan, M.T., Demuth, H.B. y Beale, M.H. (1996). Neural network design. Campus Publishing Service. Colorado University Bookstore. ISBN 0-9717321-0-8.

Hastie, T., Tibshirani, R. y Friedman, J. (2001). The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer, New York.

Haykin, S. (1999). Neural networks. A comprehensive Foundation. 2nd Edition. Pentence-Hall.

He, F., Sung, A.H. y Guo, B. (1997). A neural network for prediction of oil well cement bonding quality. Proceedings of IASTED International Conference on Control, 417-420. Cancun-Mexico: IASTED-ACTA Press.

Hensher, D.A. (2014). The Relationship Between Bus Contract Costs, User Perceived Service Quality and Performance Assessment. *International Journal of Sustainable Transportation* 8(1): 5-27.

Hensher, D.A. y Daniels, R. (1995). Productivity measurement in the urban bus sector. *Transport Policy*, 2 (3), 179-194.

Hensher, D.A., Mulley, C. y Yahya, N. (2010). Passenger experience with quality-enhanced bus service: the Tyne and Wear "superoute" services. *Transportation*, 37 (2), 239-256.

Hensher, D. A., Stopper, P. y Bullock, P. (2003). Service quality-developing a service quality index in the provision of commercial bus contracts. *Transportation Research Part A*, 37, 499-517.

Hill, N., Brierley, G. y MacDougall, R. (2003). How to Measure Customer Satisfaction (Gower Publishing, Hampshire, UK).

Hornik, K., Stichcombe, M. y White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2, 359-366.

Hu, K-C. (2010). Evaluating city bus service based on zone of tolerance of expectation and normalized importance. *Transport Reviews*, 30 (2), 195-217.

Hu, K-C. y Jen, W. (2006). Passengers' perceived service quality of city buses in Taipei: scale development and measurement. *Transport Reviews*, 26 (5), 645-662.

Hunt J. G. y Lyons G. D. (1993). Modelling dual carriageway lane changing using neural networks. *Informing Technologies for Construction Civil Engineering and Transport*, Brunel.

Hunter, A., Kennedy, L., Henry, J. y Ferguson, I. (2000). Application of neural networks and sensitivity analysis to improved prediction of trauma survival. *Computer Methods and Programs in Biomedicine*, 62, 11-19.

Huse, C. y Evangelho, F. (2007). Investigating business traveler heterogeneity: low-cost vs full-service airline users?. *Transportation Research Part E*, 43, 259-268.

Irfan, S.M., Mui, D. y Shahbaz, S. (2011). Service Quality in Rail Transport of Pakistan: A Passenger Perspective. In *Proceedings of 3rd SAICON: International Conference on Management, Business Ethics and Economics (ICMBEE)*, December 28-29, 2011, Lahore, Pakistan.

Jakobsson Bergstad, C., Gamble, A., Hagman, O., Polk, M. y Garling, T. (2011) Affective-symbolic and instrumental-independence psychological motives mediating effects of socio-demographic variables on daily car use. *Journal of Transport Geography* 19 (1), 33-38.

Jen, W., Tu, R. y Lu, T. (2011). Managing passenger behavioral intention: an integrated framework for service quality, satisfaction, perceived value, and switching barriers. *Transportation*, 38, 321-342.

Joewono, T.B. y Kubota, H. (2007a). Exploring negative experiences and user loyalty in paratransit. *Transportation Research Record*, 2034, 134-142.

Joewono, T.B. y Kubota, H. (2007b). User perceptions of private paratransit operation in Indonesia. *Journal of Public Transportation*, 10 (4), 99-118.

Joewono, T.B. y Kubota, H. (2007c). User satisfaction with paratransit in competition with motorization in Indonesia: anticipation of future implications. *Transportation*, 34 (3), 337-354.

Jun, L. y He, G. (2007) Study on railway transport service quality evaluation. *International Conference on Wireless Community, Networking and Mobile Computing, 2007 IEEE WiCOM*, Shanghai, China, 3308-3311.

Karlaftis, M.G, Golias, J. y Papadimitriou, E. (2001). Transit quality as an integrated traffic management strategy: measuring perceived service. *Journal of Public Transportation*, 4 (1).

Kasabov, N.K. (1996). *Foundations of neural networks, fuzzy systems and knowledge engineering*. MIT Press.

Kemp, S.L., Zaradic, P. y Hansen, F. (2007). An approach for determining relative input parameter importance and significance in artificial neural networks. *Ecological Modelling*, 204, 326-334.

Kewley, R., Embrechts, M. y Breneman, C. (2000). Data strip mining for the virtual design of pharmaceuticals with neural networks. *IEEE Transactions on Neural Networks*, 11(3), 668-679.

Kiatcharoenpol, T. y Laosirihongthong, T. (2006) Innovations in service strategy: An evaluation of quality in airline service operations by using SERVQUAL model. *IEEE International Conference Management of Innovation Tech.*, 2006 IEEE ICMIT, Singapore, 748–752.

Kikuchi, S. (2012). Thoughts on the future of artificial intelligence and transportation. In *Transportation Research Circular E-C168: Artificial Intelligence Applications to Critical Transportation Issues*. Ed. Transportation Research Board, Washington DC, pp. 139-140.

Kim, Y.K. y Lee, H.R. (2011). Customer satisfaction using low cost carriers. *Tourism Management*, 32 (2), 235-243.

Kornhauser, A. (1991). Neural network approaches for lateral control of autonomous highway vehicles. *Proc. Vehicle Navigation and Information Systems Con*, pp. 1143-I 151, Dearborn, MI.

Kuo, M., Wu, J. y Pei, L. (2007). A soft computing method for selecting evaluation criteria of service quality. *Applications Mathematics Computation*, 189 (1), 241–254.

Kuo, M. (2011). A novel interval-valued fuzzy MCDM method for improving airlines' service quality in Chinese cross-strait airlines. *Transportation Research Part E*, 47 (6), 1177–1193.

Larasati, A., De Yong, C. y Slevitch, L. (2012). The application of neural network and logistics regression models on predicting customer satisfaction in a student-operated restaurant. *Procedia Social and Behavioral Sciences*, 65, 94-99.

Lai, W. y Chen, C. (2011). Behavioral intentions of public transit passengers, the roles of service quality, perceived value, satisfaction and involvement. *Transport Policy*, 18 (2), 318-325.

Lee, Ch., Ran, B., Yang, F. y Loh, W.Y. (2010). A hybrid tree approach to modeling alternate route choice behavior with online information. *Journal of Intelligent Transportation Systems*, 14(4), 209–219.

Lek, S., Beland, A., Dimopoulos, I., Lauga, J. y Moreau, J. (1995). Improved estimation, using neural networks, of the food consumption of fish populations. *Marine and Freshwater Research*, 46(8), 1229-1236.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J. y Aulagnier, S. (1996a). Application of neural networks to modeling nonlinear relationships in ecology. *Ecological Modelling*, 90, 39-52.

Lek, S., Beland, A., Baran, P., Dimopoulos, I. y Delacoste, M. (1996b). Role of some environmental variables in trout abundance models using neural networks. *Aquatic Living Resources*, 9, 23-29.

Lehtinen, U. y Lehtinen, J. (1982). *Service quality – a study of quality dimensions*, Service Management Institute, Helsingfors.

Lehtinen, U. y Lehtinen, J. (1991). Two approaches to service quality dimensions. *The Service Industries Journal*, 11 (3), 287-303.

Lin, Y. y Cunningham, G.A. (1995). A new approach to fuzzy-neural system modeling. *IEEE Transactions on Fuzzy Systems*, 3(2), 190-198.

Lin, W-B. (2007). An empirical of service quality model from the viewpoint of management. *Expert Systems with Applications*, 32, 364-375.

Lyons G. y Hunt J. (1993) Traffic modelling-a role for neural networks? Proc. Third Int. Conf on the Application of Artificial Intelligence to Civil and Structural Engineering, Edinburgh, U.K.

Liou, J.H. y Tzeng, G.H. (2007). A non-additive model for evaluating airline service quality. *Air Transport Management*, 13 (3), 131–138.

Mahapatra, S.S. y Khan, M.S. (2006). A methodology for evaluation of service quality using neural networks. Presented at the International Conference on Global Manufacturing and Innovation, Rourkela.

Marcucci, E. y Gatta, V. (2007). Quality and public transport service contracts. *European Transport*, 36, 92-106.

Martín del Brío, B. y Sanz, A. (2006). *Redes neuronales y sistemas borrosos*. Editorial Ra-Ma.

Masters, T. (1993). *Practical neural networks recipes in C++*. Academic Press.

Minser, J. y Webb, V. (2010). Quantifying the benefits: application of customer loyalty modeling in public transportation context. *Transportation Research Record*, 2144, 111-120.

Moghaddam, F.R., Afandizadeh, S. y Ziyadi, M. (2010). Prediction of accident severity using artificial neural networks. *Internativonal Journal of Civil Ingeniering*, vol 9, nº 1.

Mohammadipour, A.H. y Alavi, S.H. (2009). The optimization of the geometric cross-section dimensions of raised pedestrian crosswalks: a case of study in Qazvin. *Accident Analysis and Prevention*, 41, 314-326.

Montella, A., Aria, M., D'Ambrosio A. y Mauriello, F. (2012). Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery. *Accident Analysis and Prevention*, 49, 58-72.

Morato, D., Aracil, J. y Díez, L.A. (2001). On linear prediction of Internet traffic for packet and burst switching network. *Proceedings of 10th International Conference on Computer Communications Networks*, 138-143.

Moutinho, L., Davies, F. y Curry, B. (1996). The impact of gender on car buyer satisfaction and loyalty. *Journal of Retailing and Customer Services*, 3 (3), 135-144.

Mouwen, A. y Rietveld, P. (2013). Does competitive tendering improve customer satisfaction with public transport? A case study for the Netherlands. *Transportation Research Part A* 51, 29–45.

Murray, S.J., Walton, D. y Thomas, J.A. (2010). Attitudes towards public transport in New Zealand. *Transportation*, 37 (6), 915–929.

Mussone, L., Ferrari, A. y Oneta, M. (1999). An analysis of urban collisions using an artificial intelligence model. *Accident Analysis and Prevention*, 31, 705-718.

Nathanail, E. (2008). Measuring the quality of service for passengers on the Hellenic railways. *Transportation Research Part A*, 42, 48-66.

Nejati, M., Nejati, M. y Shafaei, A. (2009). Ranking airlines' service quality factors using a fuzzy approach: Study of the Iranian society. *International Journal of Quality Reliability Management*, 26 (3), 247–260.

Ngatia, G. J., Okamura, T. y Nakamura, F. (2010). The Structure of Users' Satisfaction on Urban Public Transport Service in Developing Country: the Case of Nairobi. *Journal of the Eastern Asia Society for Transportation Studies*, 8.

Nurul-Habib, K.M., Kattan, L. e Islaam, T. (2011). Model of personal attitudes towards transit service quality. *Journal of Advanced Transportation*, 45, 271-285.

Olden, J.D. y Jackson, D.A. (2002). Illuminating the "black-box": a randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*, 154, 135-150.

Olden, J.D., Joy, M.K. y Death, R.G. (2004). An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling*, 178, 389-397.

Oyewole, P. (2001). Consumer's socio-demographic characteristics and satisfaction with services in the airline industry. *Services Marketing Quarterly*, 23 (2), 61-80.

Özesmi, S.L. y Özesmi, U. (1999). An artificial neural network approach to spatial habitat modeling with interspecific interaction. *Ecological Modelling*, 116, 15-31.

Ozment, J. y Morash, E.A. (1998). Assessment of the relationship between productivity and performance quality in the U.S. domestic airline industry. *Transportation Research Record*, 1622, 22-30.

Pakdil F y Aydin Ö (2007). Expectations and perceptions in airline services: An analysis using weighted SERVQUAL scores. *J. Air Transport Management* 13(4):229–237.

Paliwal, M. y Kumar, U.A. (2011). Assessing the contribution of variables in feed forward neural network. *Applied Software Computing*, 3690-3696.

Palmer, A. y Montaña, J.J., (2002a). Redes neuronales artificiales aplicadas al análisis de datos. Tesis doctoral. Universidad de Palma de Mallorca.

Palmer, A. y Montaña, J.J., (2002b). Redes neuronales artificiales: abriendo la caja negra. *Metodología de las ciencias del comportamiento*, 4(1), 77-93.

Palmer, A. y Montaña, J.J., (2003). Numeric Sensitivity Analysis applied to feedforward neural networks. *Neural Computing and Applications*, 12, 119-125.

Pant, P. D. y Balakrishnan, P. (1994). Neural network for gap acceptance at stop-controlled intersections. *J. Transportation Engineering* 120(3), 432-446.

Papagiannaki, D., Taft, N., Zhang, Z.L. y Diot, C. (2003). Long-term forecasting of internet backbone traffic observations and initial model. *Proceeding of IEEE INFOCOM*, San Francisco.

Parasuraman, A., Zeithaml, V.A. y Berry, L.L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, 49, 41–50.

Parasuraman, A., Zeithaml, V.A. y Berry, L.L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *J. Retailing* 64 (1), 12–40.

Park, J., Robertson, R. y Wu, C.L. (2004). The effect of airline service quality on passengers' behavioral intentions: a Korea case study. *Journal of Air Transport Management*, 10, 435-439.

Pedersen, T., Friman, M. y Kristensson, P. (2011). Affective forecasting: predicting and experiencing satisfaction with public transportation. *Journal of Applications and Social Psychology*, 41 (8), 1926-1946.

Philip, G. y Hazlett, S.A. (1997). The measurement of service quality: A new P-C-P attributes model. *International Journal of Quality Reliability Management*, 14 (3), 260–286.

Pomerleau, D., Gowdy J. y Thorpe, C. (1991). Combining artificial neural networks and symbolic processing for autonomous robot guidance. *Engineering Applications and Artificial Intelligence*, 4 (4), 279-285.

Pullen, W.T. (1993). Definition and measurement of quality of service for local public transport management. *Transport Reviews*, 13 (3), 247-264.

Ringle, C.M., Sarstedt, M. y Zimmermann, L. (2011). Customer satisfaction with commercial airlines: the role of perceived safety and purpose of travel. *Journal of Marketing Theory and Practise*, 19 (4), 459-472.

Rumelhart, D.E. y McClelland, J.L. (1986a). *Parallel Distributed Processing. Vol 1: Foundations*. MIT Press.

Rumelhart, D.E., Hinton, G.E. y Williams, R.J. (1986b). Learning representations by backpropagation errors. *Nature*, 323, 533-536.

Rzempoluk, E.J. (1998). *Neural network data analysis using Simulnet*. New York: Springer-Verlag.

Sasser, W.E., Olsen, R.P. y Wyckoff, D.D. (1978). *Understanding service operations, Management of Service Operations*, Ally & Bacon, Boston, MA.

Scardi, M. y Harding, L.W. (1999). Developing an empirical model of phytoplankton primary production: a neural networks case study. *Ecological Modelling*, 120(2-3), 213-223.

Simoes, A. (2013). Mobility-impaired travelers and public transport: a framework to assess quality of service. *Theoretical Issues in Ergonomics Science*, 14:3, 247-257.

Slack, N. (1994). The importance-performance matrix as a determinant of improvement priority. *International Journal of Operations & Production Management*, 14 (5), 59-75.

Stuart, K.R., Mednick, M. y Bockman, J. (2000). Structural equation model of customer satisfaction for the New York city subway system. *Transportation Research Record*, 1735, 133-137.

Sung, A.H. (1998). Ranking importance of input parameters of neural networks. *Expert Systems with Applications*, 15, 405-411.

Tchaban, T., Taylor, M.J. y Griffin, A. (1998). Establishing impacts of the inputs in a feedforward network. *Neural Computing and Applications*, 7, pp. 309-317.

Transportation Research Board (2004). *Transit capacity and quality of service manual*, Second edition. Transit Cooperative Research Program (TCRP). Report 100, Transportation Research Board, Washington, DC.

Tripp, C. y Drea, J.T. (2002). Selecting and promoting service encounter elements in passenger rail transportation. *The Journal of Services Marketing*, 16 (5), 432-442.

Tsai, W., Hsu, W. y Chou, W. (2011). A gap analysis model for improving airport service quality. *Total Quality Management and Business Excellence*, 22 (10), 1025-1040.

Tsaur, S.H., Chang, T.Y. y Yen, C.H. (2002). The evaluation of airline service quality by fuzzy MCDM. *Tourism Management*, 23 (2), 107-115.

Tyrinopoulos, Y. y Antoniou, C. (2008). Public transit user satisfaction: Variability and policy implications. *Transport Policy*, 15(4), 260-272.

Tyrinopoulos, Y. y Aifadopoulou, G. (2008). A complete methodology for the quality control of passenger services in the public transport business. *European Transport*, 38, 1-16.

UNE-EN 13186 (2002). *Transportation. Logistics and Services. Public Passenger Transport. Service Quality Definition, Targeting and Measurement*. AENOR.

Watts, M.J. y Worner, S.P. (2008). Using artificial neural networks to determine the relative contribution of abiotic factors influencing the establishment of insect pest species. *Ecological Informatics*, 3, 64-74.

Weitz, B. A. y Jap, S. D. (1995). Relationship marketing and distribution channels. *Journal of the Academy of Marketing Science*, 23(4), 305–320.

Weinstein, A. (2000). Customer satisfaction among transit riders. How customer rank the relative importance of various service attributes. *Transportation Research Record*, 1735, 123–132.

Wipper, L.R. (1993). Performance measurement: producing results at the Oregon department of transportation. *Transportation Research Record*, 1395, 168-175.

Witten, I.H. y Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufman, Amsterdam.

Xie, C., Lu, J. y Parkany, E. (2003). Work travel mode choice modelling with data mining. In *Transportation Research Record: Journal of the Transportation Research Board*, 1854, 50-61.

Xiong, Y. y Schneider, J. B. (1992). Transportation network design using a cumulative genetic algorithm and neural network. *Transportation Research Record* 1364, pp. 31-44.

Yang H., Akiyama, T. y Sasaki, T. (1992). A neural network approach to the identification of real time origin-destination flows from traffic counts. *Proc. Int. Conf. on Artificial intelligence Applications in Transportation Engineering*, San Buenaventura, CA.

Yang, K. Hsieh, T., Li, H. y Yang, C. (2012). Assessing how service quality, airline image and customer value affect the intentions of passengers regarding low cost carriers. *Journal of Air Transport Management*, 20, 52-53.

Yao, J., Teng, N., Poh, H.L. y Tan, C.L., (1998). Forecasting and analysis of marketing data using neural networks. *Journal of Information Science and Engineering*, 14, 843-862.

Yeh, I. y Cheng, W. (2010). First and second order sensitivity analysis of MLP. *Neurocomputing*, 73, 2225-2233.

Yohannes, Y. y Webb, P. (1999). *Classification and Regression Trees. CART: A User Manual for Identifying Indicators of Vulnerability to Famine and Chronic Food Insecurity*. International Food Policy Research Institute. Washington D.C.

Zhou, Z.H., Wu, J. y Tang, W. (2002). Ensembling neural networks: many could be better than all. *Artificial Intelligence* 137 (1-2), 239-263.

Zurada, J.M., Malinowski, A. y Cloete, I. (1994). Sensitivity analysis for minimization of input data dimension for feed forward neural network. *Proceedings of IEEE International Symposium on Circuits and Systems*. London. IEEE Press.

APÉNDICE 1. FORMULARIO DE ENCUESTA



Datos del encuestador

Nombre.....Apellidos.....Nº de encuesta

Lugar (donde se realiza la encuesta).....

Empresa.....Línea..... Hora /
h m

1 ¿Podría realizar este viaje en vehículo privado?

- Si (Pasar a P 2) No (Pasar a P3)

2 ¿Por qué utiliza transporte público?

- Precio Aparcamiento Comodidad Otras razones

Indique cuales.....

3 ¿Usted utiliza esta línea...?

- Casi a diario [4 ó más días a la semana] Frecuentemente [1 a 3 días a la semana] Ocasionalmente [1 a 3 días al mes] Esporádicamente

4 ¿Podría decirnos cual es el motivo de su viaje?

- Trabajo Médico Asunto personal
 Estudios Ocio Acompañante
 Compras Otra situación (cual)

5 ¿Qué tipo de billete o título de transporte utiliza en este viaje?

- Con la tarjeta única En metálico (billete sencillo)

6. ¿Considera los horarios adecuados a sus necesidades?

- Si No
↓
→ Indique su propuesta _____

7 ¿Considera que el servicio público de autobús metropolitano es?:

- Muy buena Bueno Normal Mala Muy mala Ns/Nc

8 En relación con el servicio de autobuses que presta el CTAGR, podría decirnos sí:

- satisfecho Muy satisfecho Bastante satisfecho Poco satisfecho Nada satisfecho Ns/Nc

9 En una escala del 0 al 10 nos gustaría que valorase la importancia que usted le da, a los siguientes servicios:

Valoración sobre el servicio en general

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Información

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Puntualidad

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Seguridad a bordo de los autobuses (robos etc.)

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Limpieza en el interior de los autobuses

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Comodidad en los autobuses

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Descuentos

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Rapidez

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Frecuencia

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Proximidad

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

10. En una escala del 0 al 10 nos gustaría nos diese su Calificación sobre los siguientes aspectos:

Valoración sobre el servicio en general

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Información

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Puntualidad

.

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Seguridad a bordo de los autobuses (robos etc.)

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Limpieza en el interior de los autobuses

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Comodidad en los autobuses

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Descuentos

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Rapidez

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Frecuencia

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Proximidad

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

11 ¿Ha notado mejoras en el servicio?

- Sí en el último año
 No ha notado cambios
 Ha empeorado

PANEL

Necesitamos un grupo de personas dispuestas a hacer valer su opinión. Inscribirse no implica ninguna obligación, sólo la aceptación de ser invitado periódicamente a participar en alguno de los estudios que realice el CTAGR

NOMBRE..... APELLIDOS..... TELÉFONO.....Móvil..... Municipio/Residencia.....Actividad..... Edad <input type="checkbox"/> [18-30] <input type="checkbox"/> [30-60] <input type="checkbox"/> [Más de 60] Sexo <input type="checkbox"/> Hombre <input type="checkbox"/> Mujer	
---	--

Información complementaria:

Calle/Avd /Paseo.....Nº ... Población.....CP.....
--

Sus datos personales, facilitados a través de este formulario quedarán incorporados a los ficheros automatizados existentes en el *CONSORCIO DE TRANSPORTES ÁREA DE GRANADA* y serán tratados únicamente para la finalidad especificada en el mismo. Podrá, en todo momento, solicitar el acceso, cancelación o rectificación de los mismos.

APÉNDICE 2. ROBUSTEZ Y ESTABILIDAD DEL MÉTODO DE CÁLCULO DE IMPORTANCIA RELATIVA DE LAS VARIABLES PROPUESTO

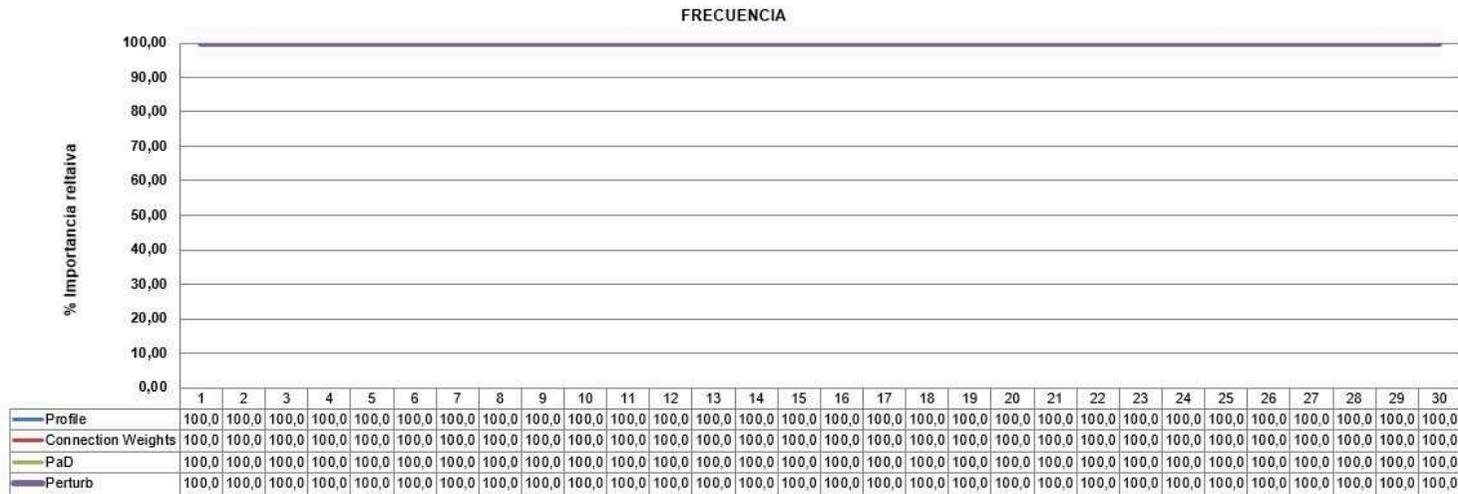


Figura 24. Importancia relativa de la variable FRECUENCIA en función de H según el método aplicado

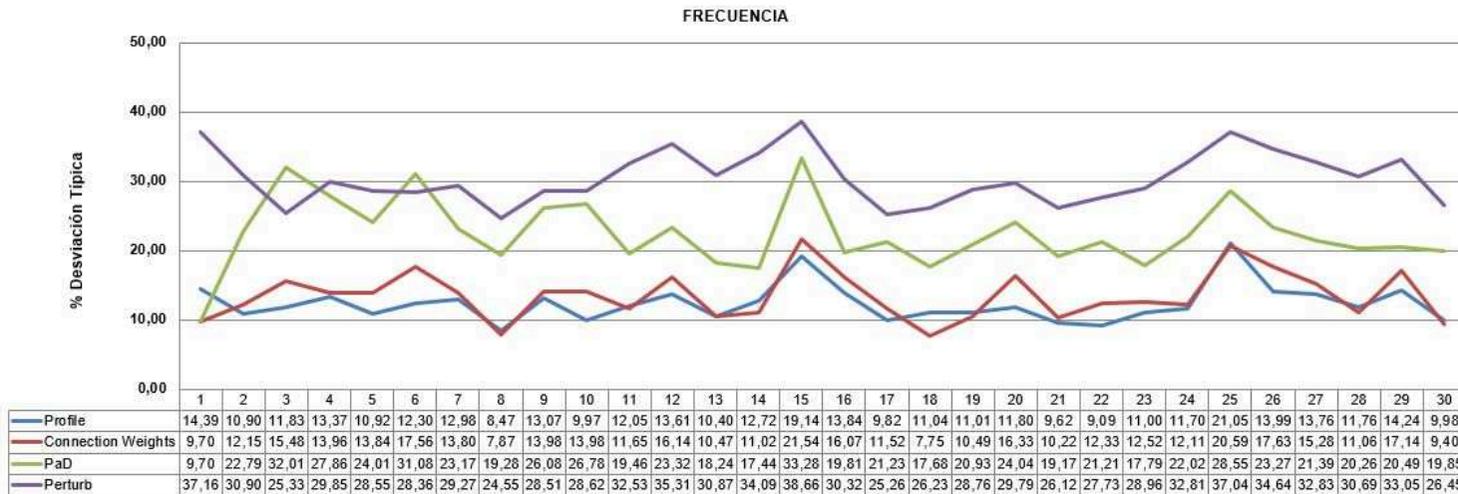


Figura 25. Desviación típica de la importancia relativa de la variable FRECUENCIA en función de H según el método aplicado

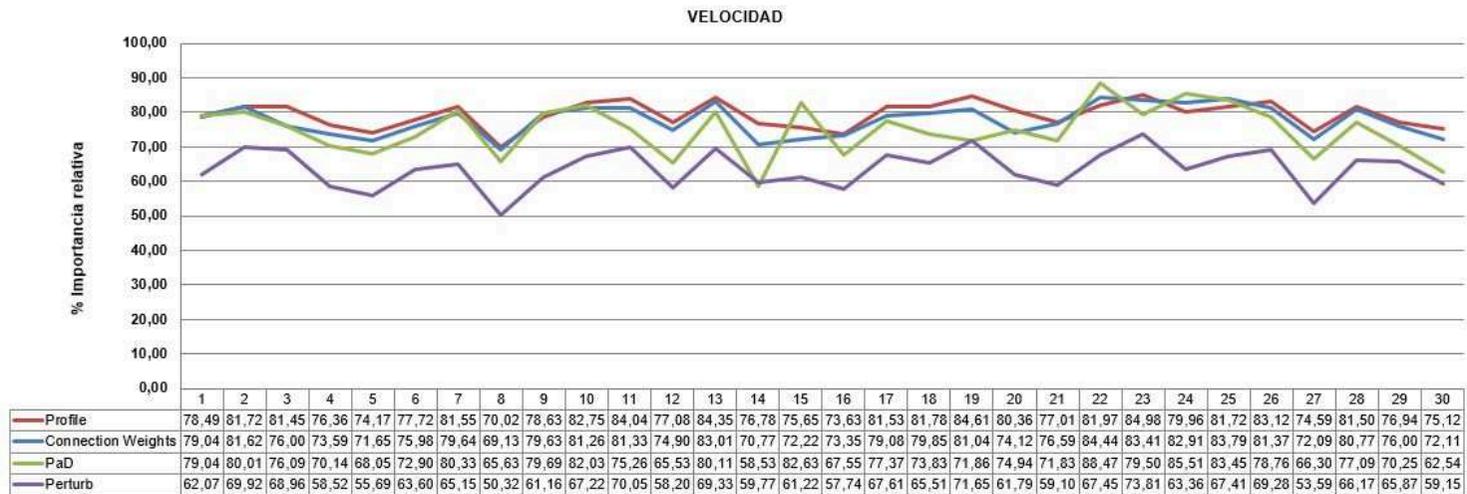


Figura 26. Importancia relativa de la variable VELOCIDAD en función de H según el método aplicado

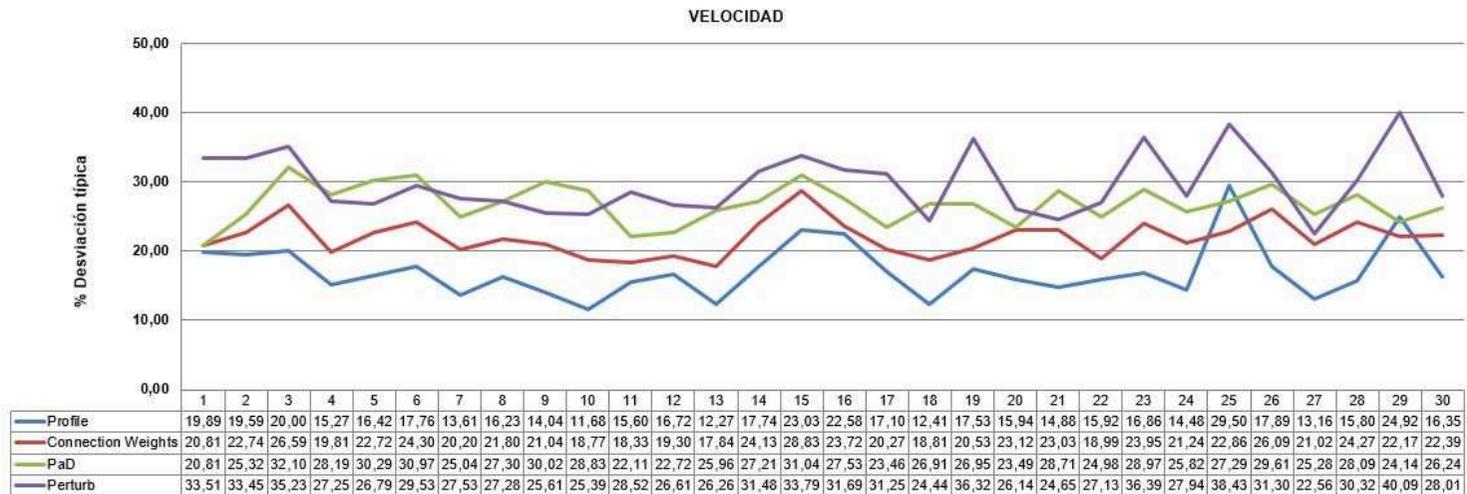


Figura 27. Desviación típica de la importancia relativa de la variable VELOCIDAD en función de H según el método aplicado

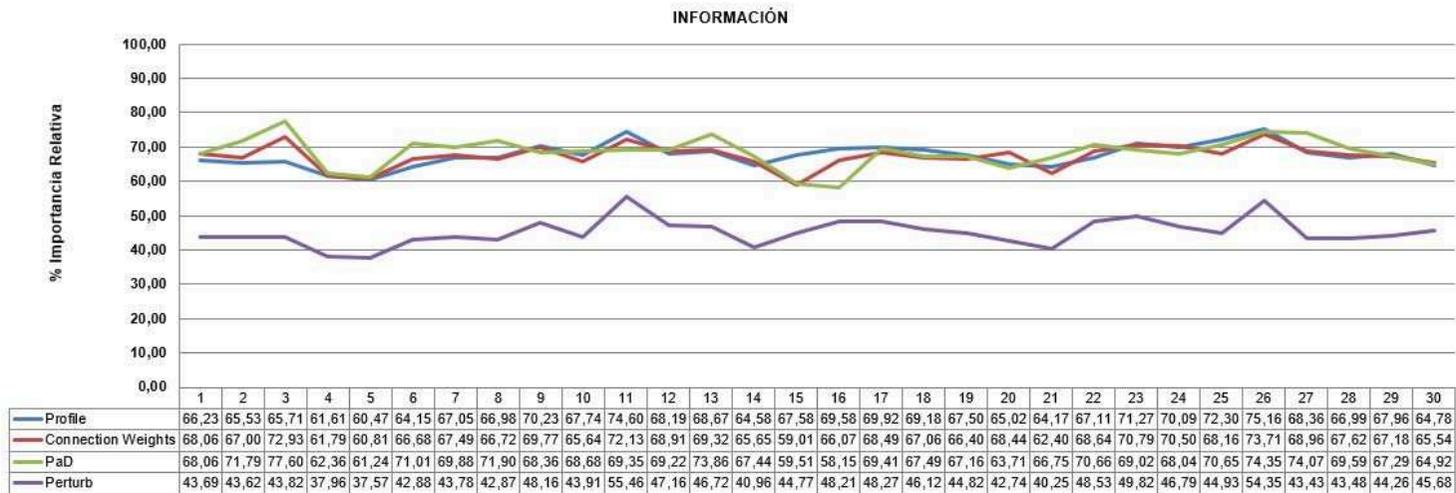


Figura 28. Importancia relativa de la variable INFORMACIÓN en función de H según el método aplicado

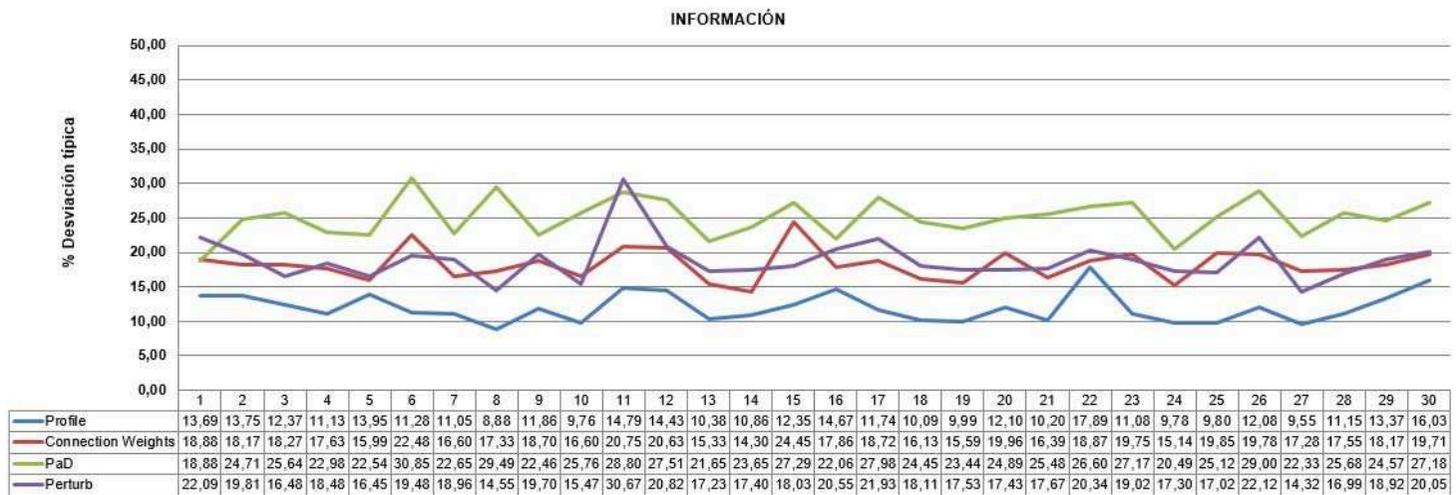


Figura 29. Desviación típica de la importancia relativa de la variable INFORMACIÓN en función de H según el método aplicado

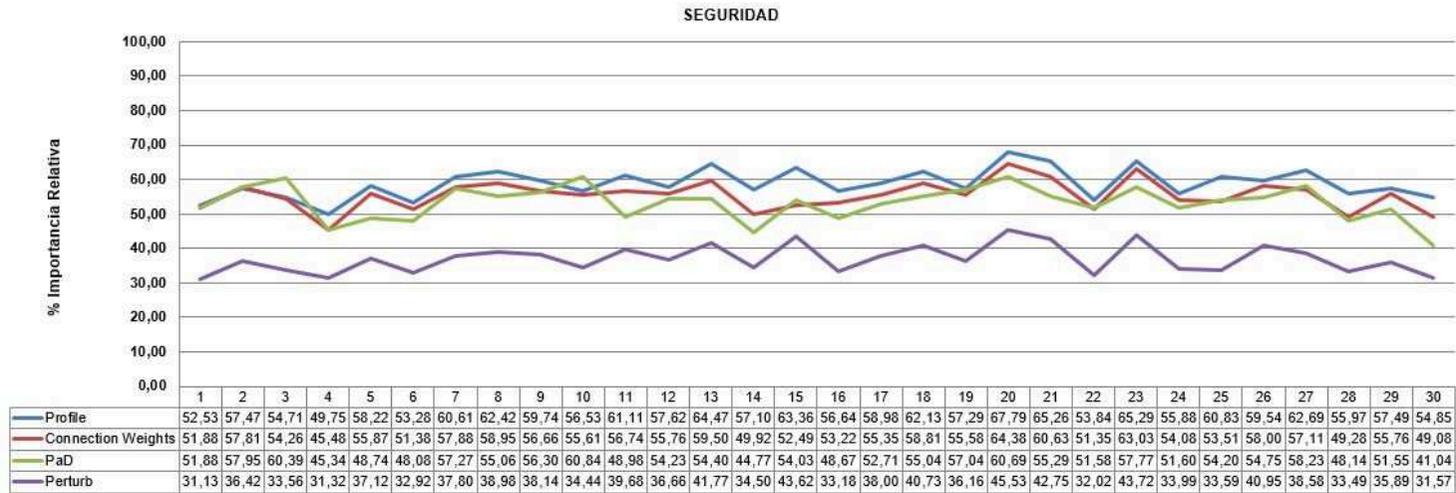


Figura 30. Importancia relativa de la variable SEGURIDAD en función de H según el método aplicado

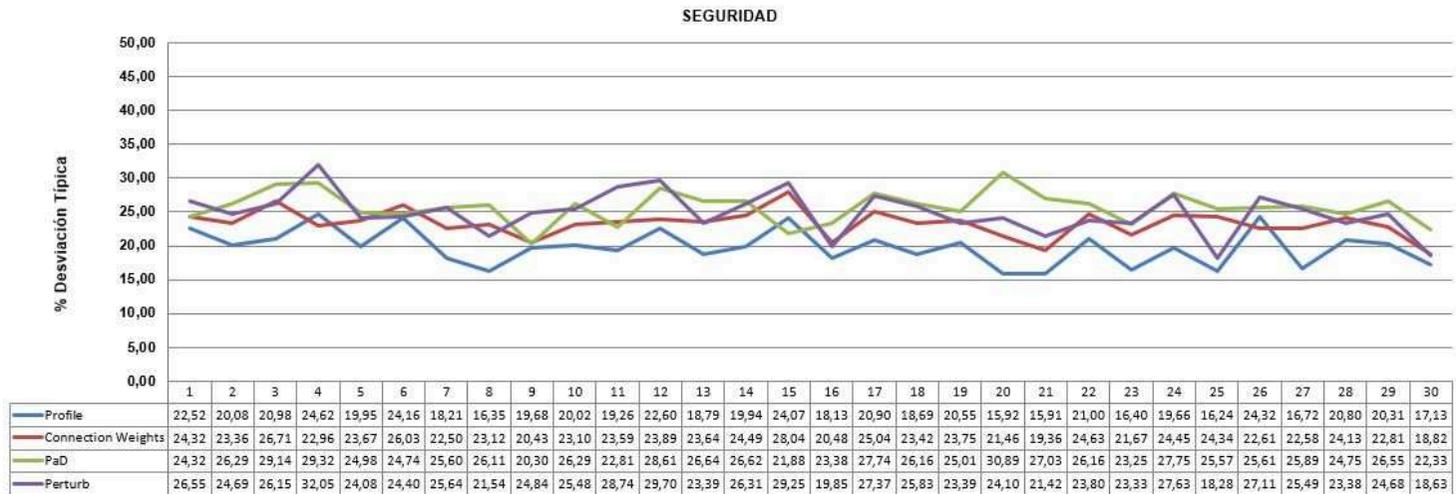


Figura 31. Desviación típica de la importancia relativa de la variable SEGURIDAD en función de H según el método aplicado

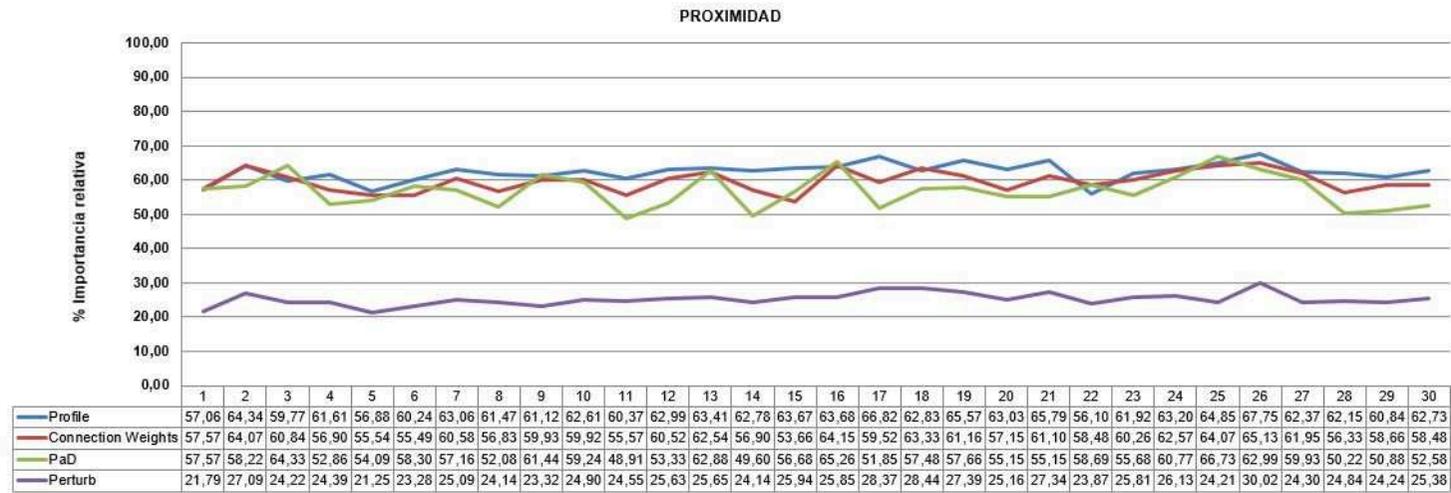


Figura 32. Importancia relativa de la variable PROXIMIDAD en función de H según el método aplicado

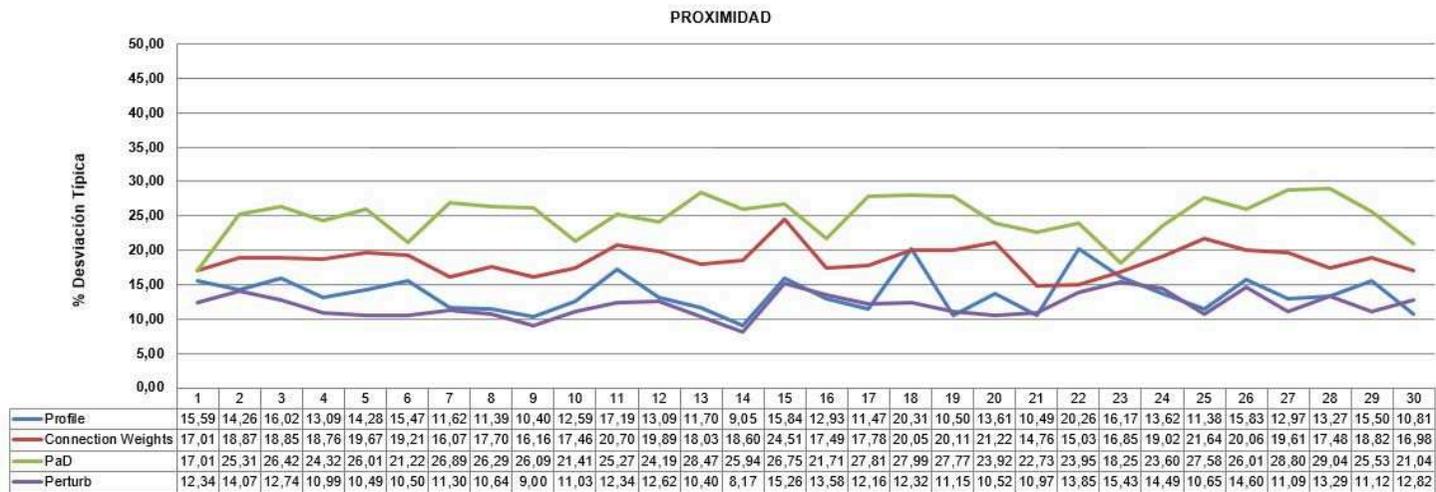


Figura 33. Desviación típica de la importancia relativa de la variable PROXIMIDAD en función de H según el método aplicado

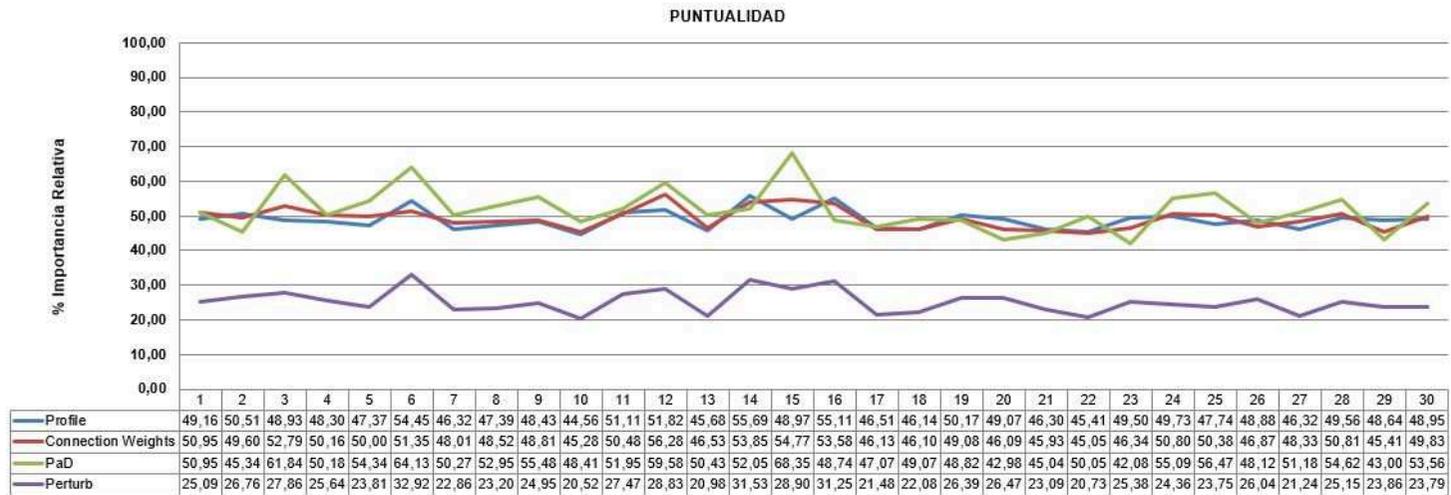


Figura 34. Importancia relativa de la variable PUNTUALIDAD en función de H según el método aplicado

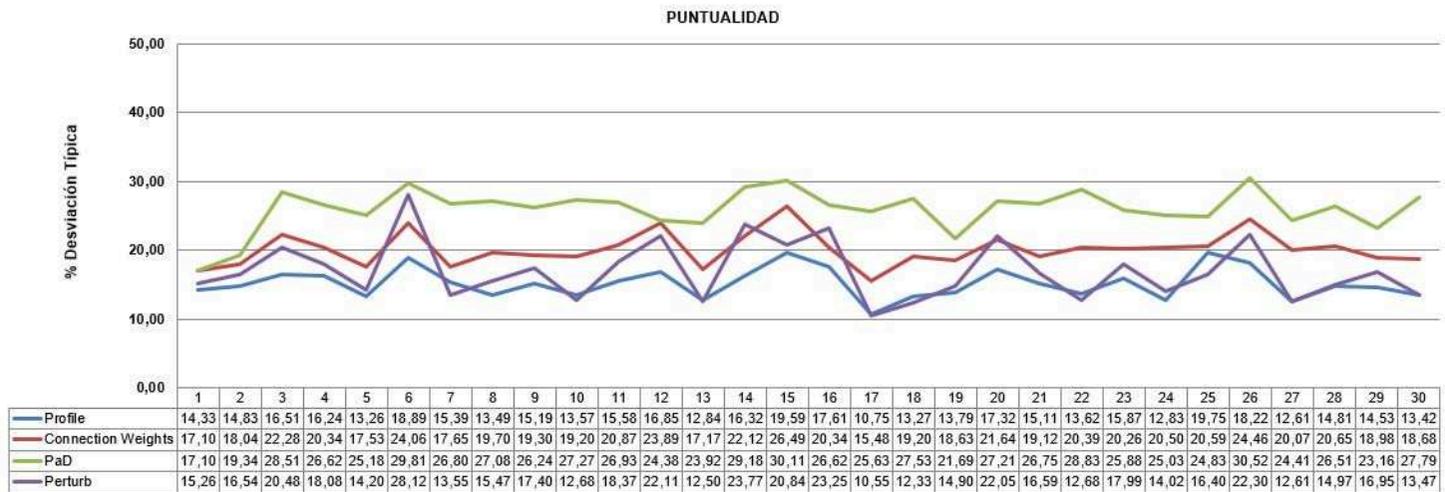


Figura 35. Desviación típica de la importancia relativa de la variable PUNTUALIDAD en función de H según el método aplicado

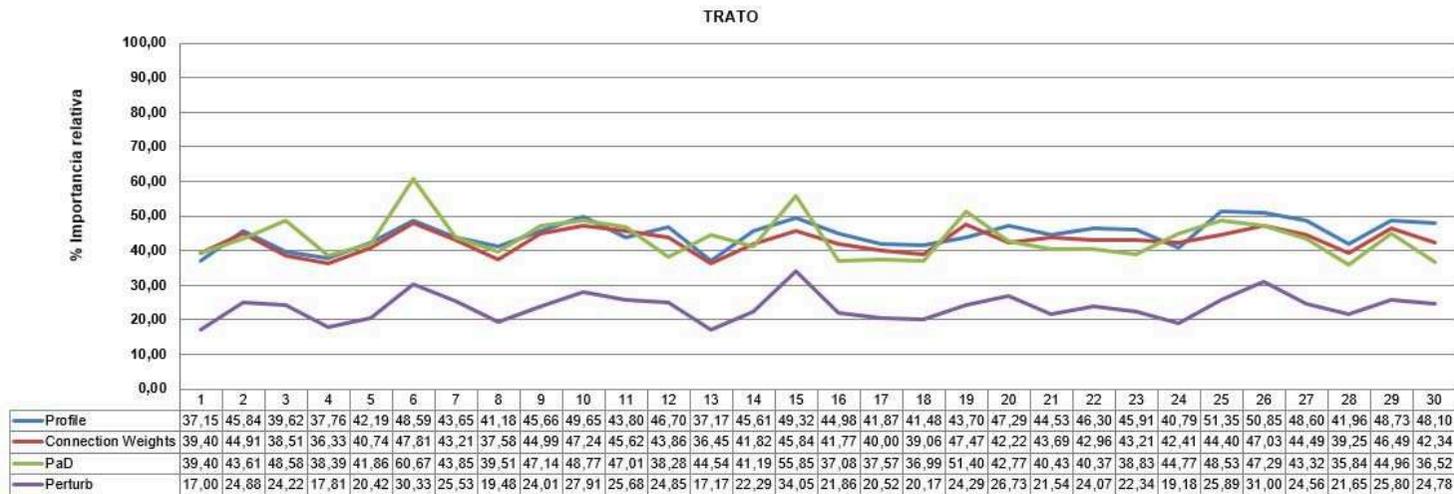


Figura 36. Importancia relativa de la variable TRATO en función de H según el método aplicado

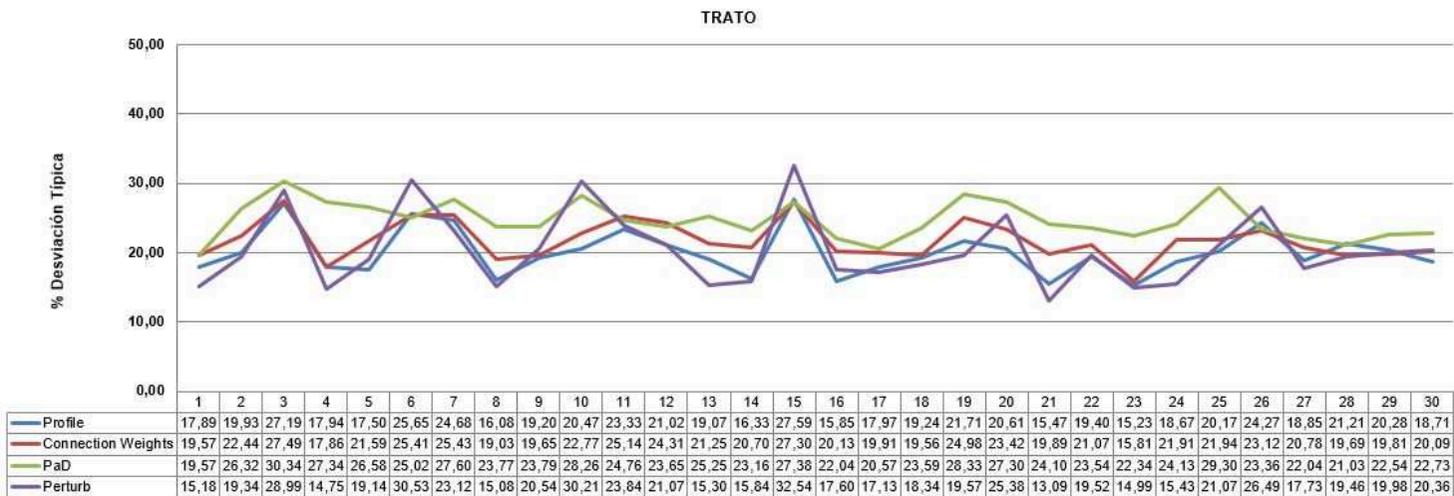


Figura 37. Desviación típica de la importancia relativa de la variable TRATO en función de H según el método aplicado

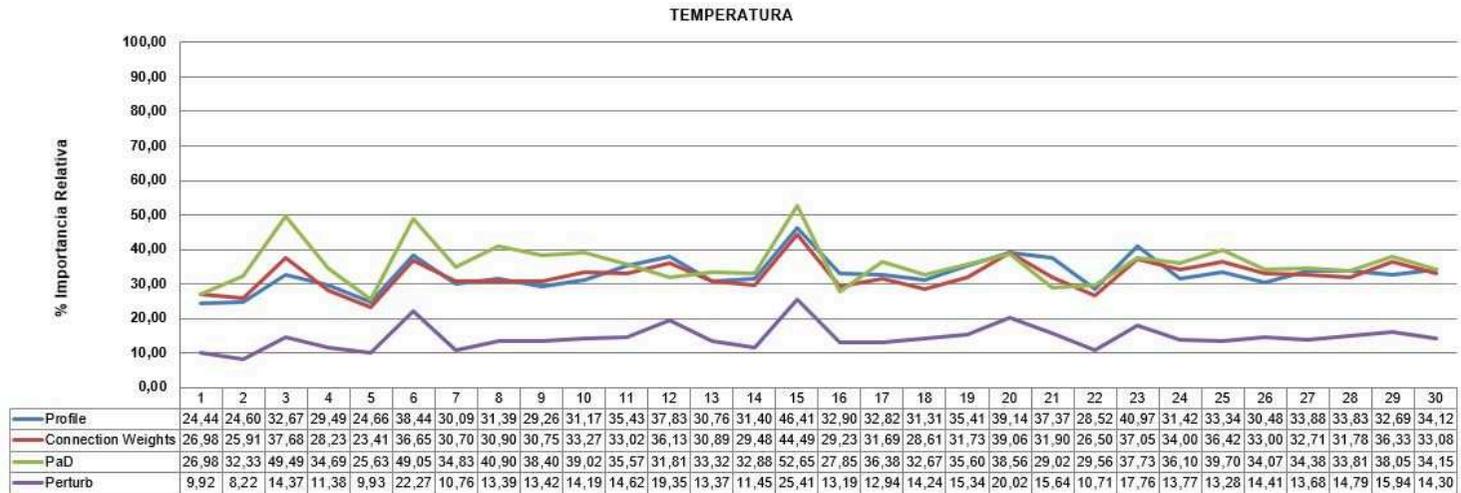


Figura 38. Importancia relativa de la variable TEMPERATURA en función de H según el método aplicado

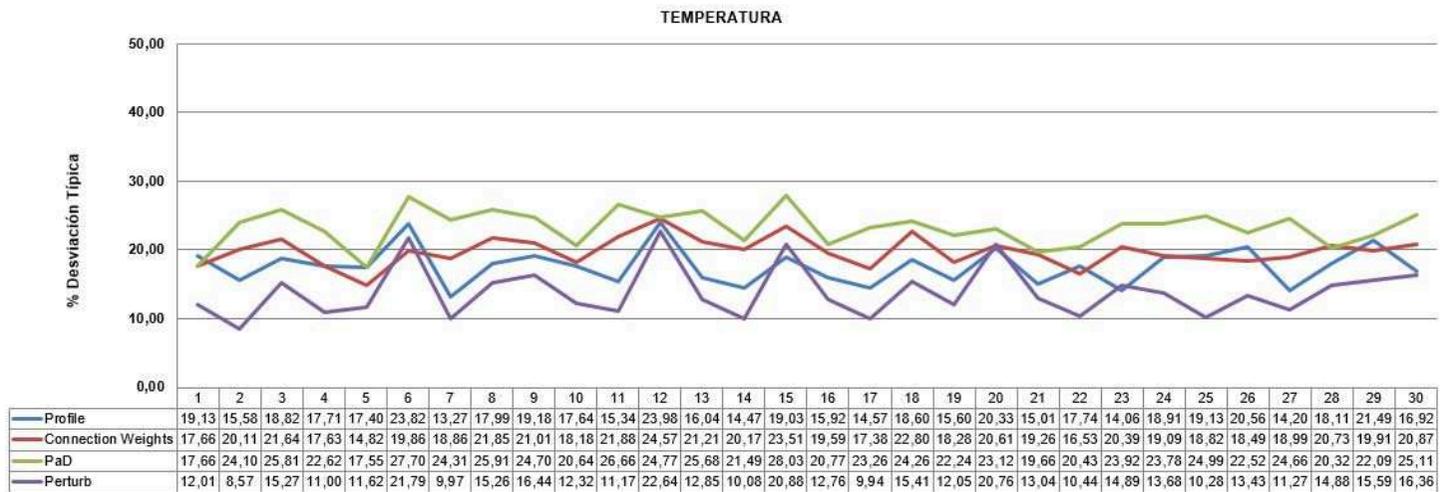


Figura 39. Desviación típica de la importancia relativa de la variable TEMPERATURA en función de H según el método aplicado

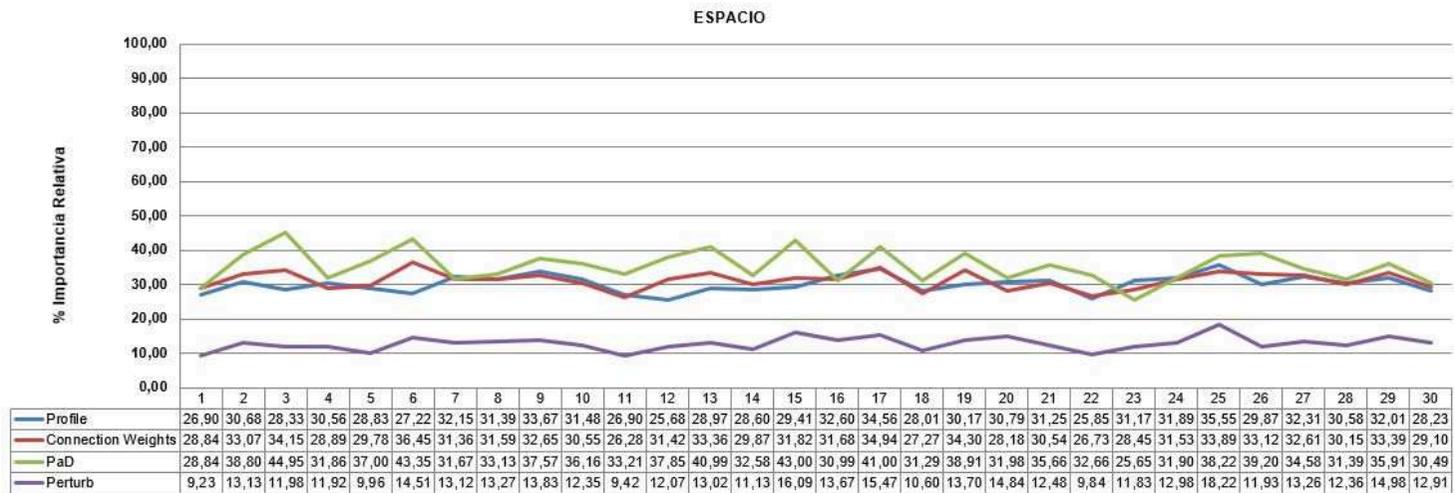


Figura 40. Importancia relativa de la variable ESPACIO en función de H según el método aplicado

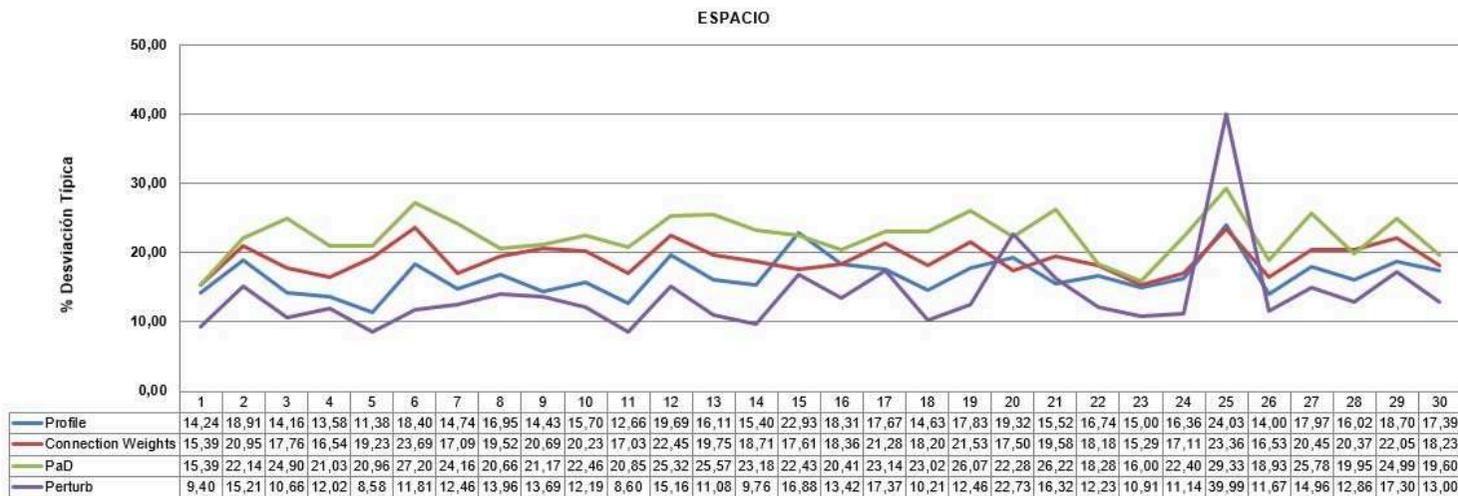


Figura 41. Desviación típica de la importancia relativa de la variable ESPACIO en función de H según el método aplicado

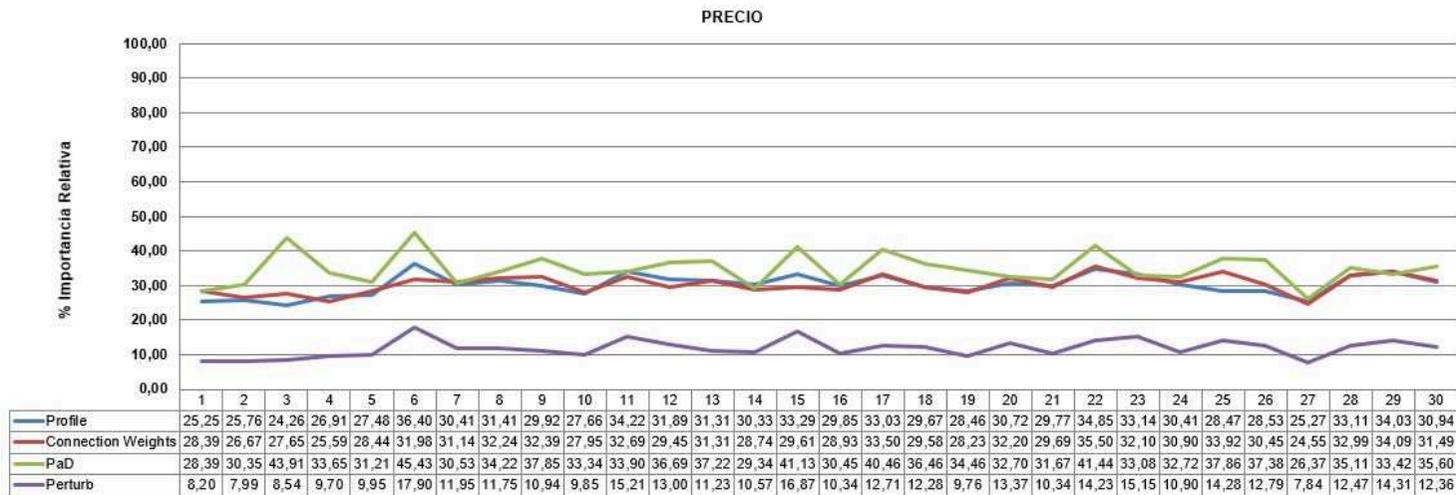


Figura 42. Importancia relativa de la variable PRECIO en función de H según el método aplicado



Figura 43. Desviación típica de la importancia relativa de la variable PRECIO en función de H según el método aplicado

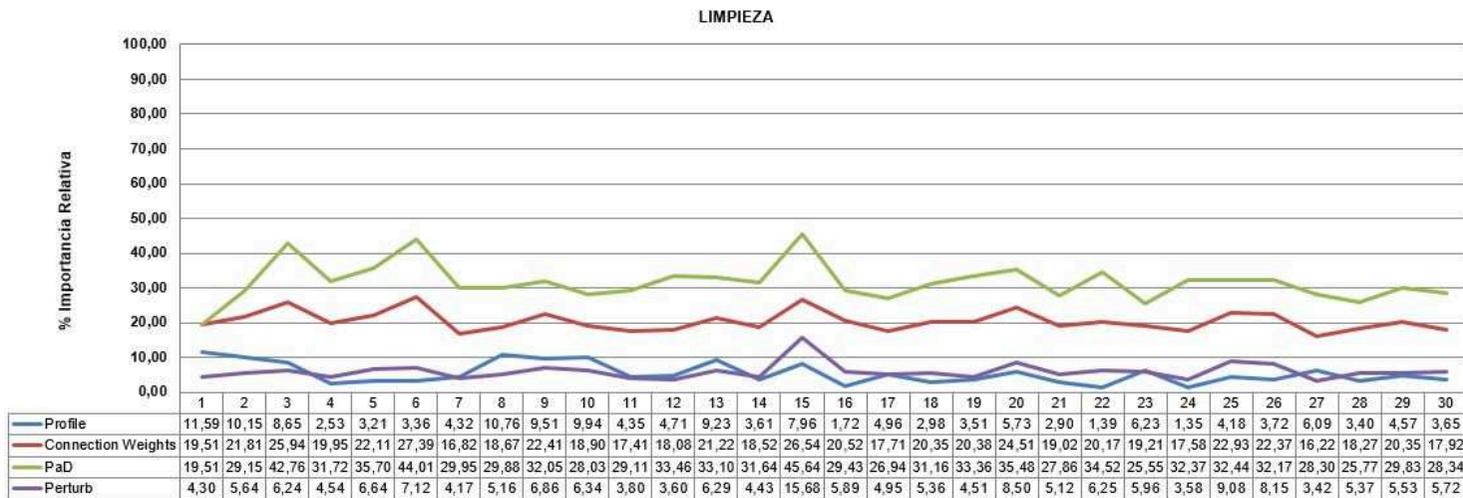


Figura 44. Importancia relativa de la variable LIMPIEZA en función de H según el método aplicado

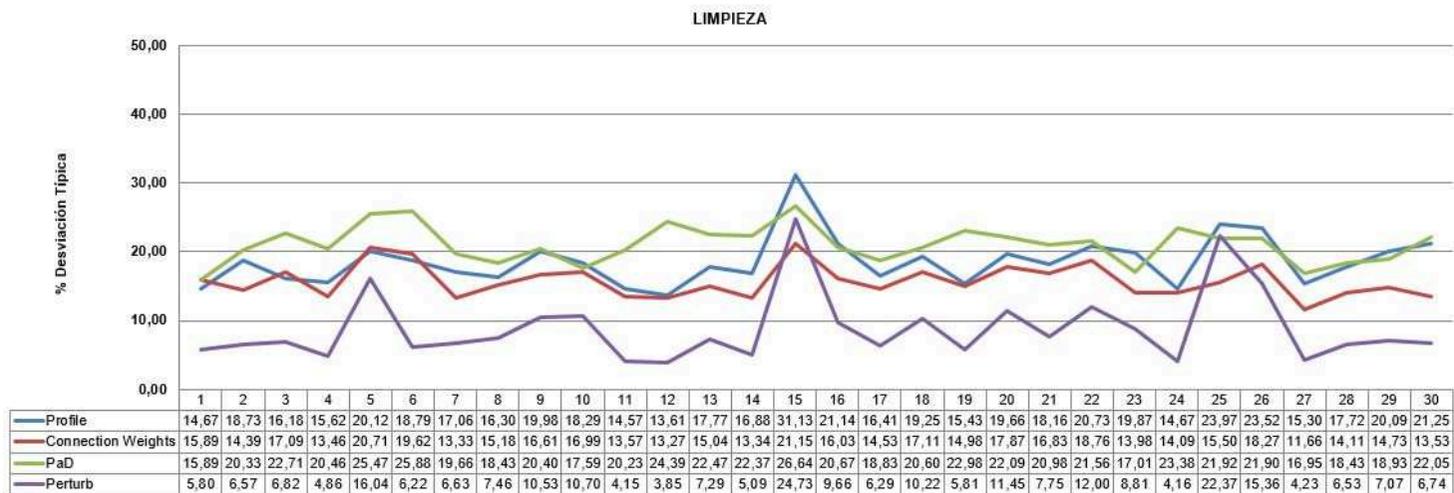


Figura 45. Desviación típica de la importancia relativa de la variable LIMPIEZA en función de H según el método aplicado

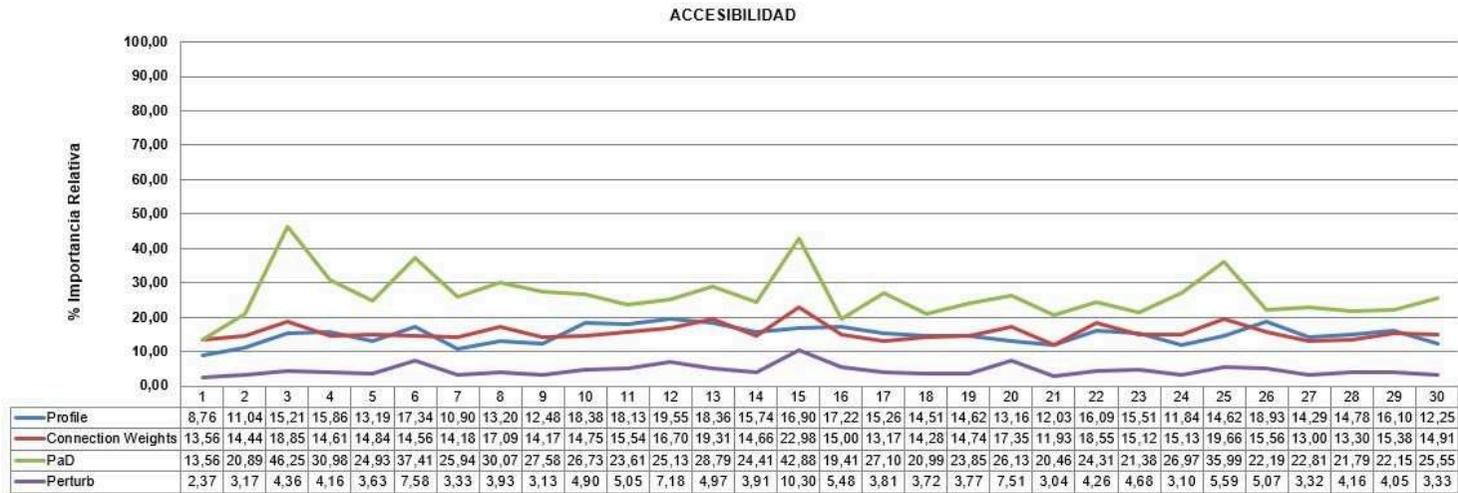


Figura 46. Importancia relativa de la variable ACCESIBILIDAD en función de H según el método aplicado

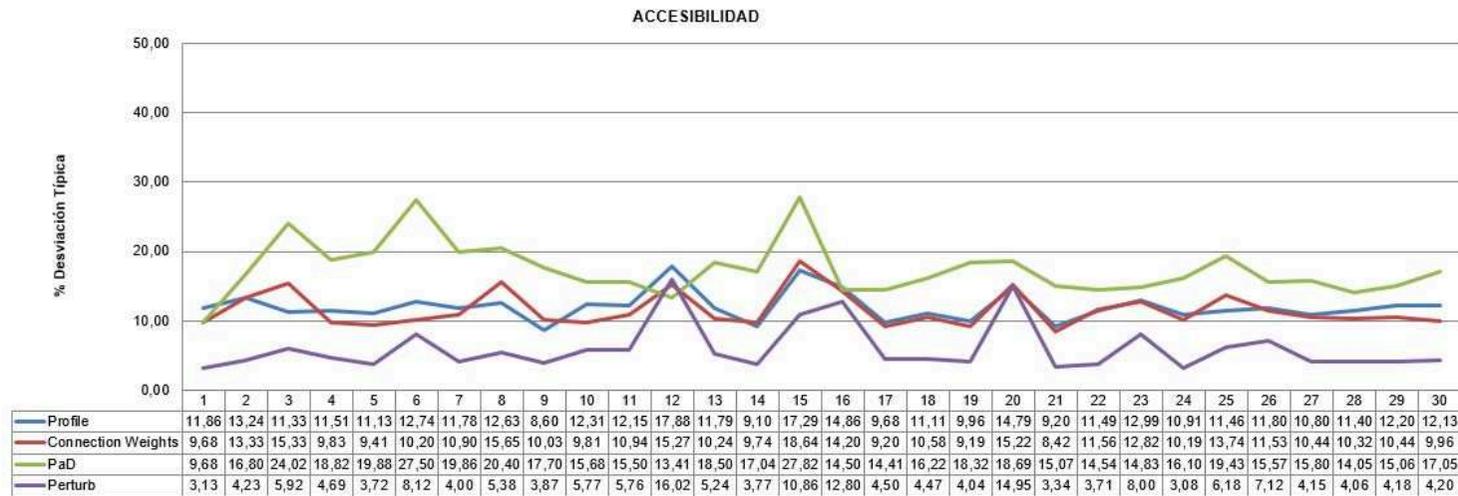


Figura 47. Desviación típica de la importancia relativa de la variable ACCESIBILIDAD en función de H según el método aplicado

APÉNDICE 3. ARTÍCULOS PUBLICADOS

ANÁLISIS DE EFICIENCIA DE REDES NEURONALES EN EL ESTUDIO DE LA CALIDAD DEL SERVICIO EN EL TRANSPORTE PÚBLICO. APLICACIÓN AL ÁREA METROPOLITANA DE GRANADA

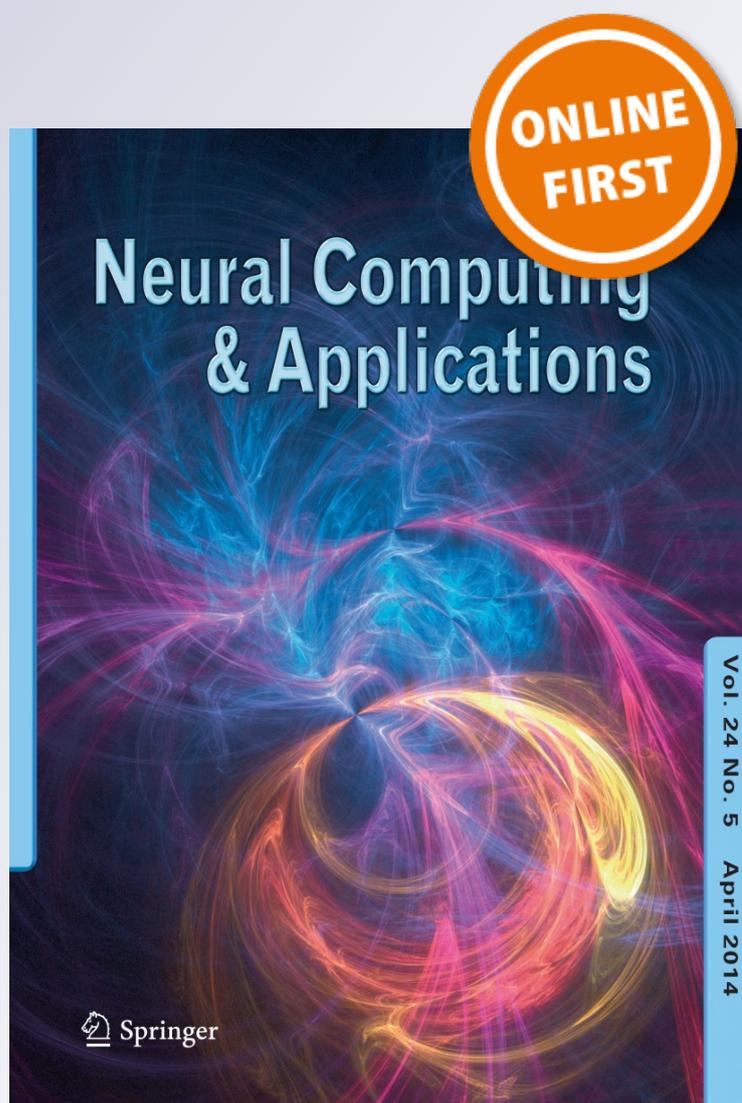
Extracting the contribution of independent variables in neural network models: a new approach to handle instability

Juan de Oña & Concepción Garrido

Neural Computing and Applications

ISSN 0941-0643

Neural Comput & Applic
DOI 10.1007/s00521-014-1573-5



Your article is protected by copyright and all rights are held exclusively by Springer-Verlag London. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Extracting the contribution of independent variables in neural network models: a new approach to handle instability

Juan de Oña · Concepción Garrido

Received: 12 November 2013 / Accepted: 12 March 2014
© Springer-Verlag London 2014

Abstract One of the main limitations of artificial neural networks (ANN) is their high inability to know in an explicit way the relations established between explanatory variables (input) and dependent variables (output). This is a major reason why they are usually called “black boxes.” In the last few years, several methods have been proposed to assess the relative importance of each explanatory variable. Nevertheless, it has not been possible to reach a consensus on which is the best-performing method. This is largely due to the different relative importance obtained for each variable depending on the method used. This importance also varies with the designed network architecture and/or with the initial random weights used to train the ANN. This paper proposes a procedure that seeks to minimize these problems and provides consistency in the results obtained from different methods. Essentially, the idea is to work with a set of neural networks instead of a single one. The proposed procedure is validated using a database collected from a customer satisfaction survey, which was conducted on the public transport system of Granada (Spain) in 2007. The results show that, when each method is applied independently, the variable’s importance rankings are similar and, in addition, coincide with the hierarchy established by researchers who have applied other techniques.

Keywords Instability · Neural networks · Black box · Variables contribution’s methods · Importance ranking

1 Introduction

Artificial neural networks (ANN) are information processing systems based on the biological behavior of the human brain and used in a growing number of multiple research fields. The strength of ANN compared to other techniques is their high capacity for classification, prediction and failure tolerance [33].

ANN are based not only on a predefined equation or formula, but also on their capacity to capture the information inherent to the data submitted during the training process. They create an architecture whose parameters are able to provide correct answers when some new cases are presented. These parameters are the key to their knowledge [42]. This singular way of learning allows them to capture highly nonlinear [50] and complex [36] relations, but prevents an explicit explanation of how explanatory variables (input) and dependent variables (output) are related. However, it is possible to achieve this goal with classical statistical techniques [2]. Therefore, ANN are included in the group of data mining techniques called “black boxes” [6], because for a given phenomenon (output), it is very difficult to know the relative importance of each variable (input).

Considering this problem, several methods have been proposed to determine the contribution of each independent variable in ANN models; many methods from the family of sensitivity analyses (SA), which basically changes input values and checks what happens in the output, and others specific neural networks (NN) methods have been developed by the researchers. SA methods perform a pure black-box approach over a data-driven model, which can be NN or other method, such as SVM. In contrast, specific NN methods can only be applied to one typology of NN [the multilayer perceptron (MLP)], so that they are not universal input relevance methods. Sung [49] applied three

J. de Oña (✉) · C. Garrido
TRYSE Research Group, Department of Civil Engineering,
University of Granada, ETSI Caminos, Canales y Puertos,
c/Severo Ochoa, s/n, 18071 Granada, Spain
e-mail: jdona@ugr.es

methods sensitivity analysis [15, 55], fuzzy curves [32] and change in mean square error (MSE) [24] to a database related to petroleum engineering and compared the results obtained. It was concluded that the fuzzy curves method performs better than the other two. Olden and Jackson [38] described the neural interpretation diagram [40], Garson's algorithm [17, 22] and sensitivity analysis [29–31] methods, and proposed a new one called randomization approach. After applying these methods to a database related to ecology, it was observed that the results were different depending on the method used. Palmer and Montaña [43] highlighted the problems of the methods studied so far, both those based on the weights of connections and those based on sensitivity analysis. They argued that several previous studies had demonstrated that the former group of methods is not effective [17, 25, 45], and that the latter presents some problems depending on the qualitative or quantitative nature of the variables. Therefore, they suggested a new approach called numeric sensitivity analysis (NSA) [44], which determines the relationship between each input and output variable through the slope and without taking into account the qualitative or quantitative nature of the variables. Gevrey et al. [20] analyzed and compared seven methods: partial derivatives [12], Garson [17, 22], perturb [48, 52], profile [29–31] and stepwise with some of its variants [49]. They used an empirical ecological database and concluded that the partial derivatives method (PaD) was the best, while the classical stepwise performed the worst. Olden et al. [39] continued the work of Gevrey et al. [20] and applied the same seven methods plus a new approach called connection weights [38] to a simulated database. The connection weights method provided the best results.

Gevrey et al. [21] introduced a new variation in the PaD, called PaD2, with the aim of analyzing the joint contribution of every possible pairwise combination of variables. They argued that in nature, variables normally interact with each other, so when a variable is modified the remaining variables also change. Kemp et al. [27] implemented the holdback input randomization (HIPR) method, based on the random alteration of NN input parameters. They applied this method to ecological complex systems, and the results obtained were as good as those extracted from the connection weights method. Yeh and Cheng [53] provided a new point of view to the contribution of variables in NN by introducing a method that considers not only linear effects (first-order derivatives) between the studied variables but also curvature effects (second-order derivatives). Cortez and Embrechts [6] introduced three new sensitivity analysis (SA) methods: data-based SA (DSA), Monte Carlo SA (MSA) and cluster-based SA (CSA), and they compared them with two other existing methods: one-dimensional SA (1D-SA) [28] and global SA (GSA) [5]. In

addition, they developed some new approaches to determine the relative importance of variables and also pairs of input variables with sensitivity analysis methods.

Commonly, the first step is to select an optimal NN architecture, and the second step consists in applying several variable contribution methods. But it is very difficult to choose the optimal NN model due to various factors such as random initialization of connection weights, multiple possible network architectures and learning algorithms that converge toward local minimums in a complex error surface. Cao and Qiao [4] underlined this issue and suggested that sensitivity analysis methods should be applied not only to a single NN, but also to a set of good-performing NN. The authors proposed a new application called neural network committee (NNC). Paliwal and Kumar [41] also referred to the high variability of weights before starting the training and proposed an approach named interquartile range, in which the network is trained a number of times, and the first and third quartiles of the weight distribution are used to determine the relative importance of each variable.

However, despite the variety of methods studied, there is no general consensus on which model is the best for determining the contribution of variables. When applying several methods to optimal or suboptimal NN architecture, the importance ranking of variables differs from method to method, indicating their inherent instability.

This paper intends to handle the instability problems derived from these methods and suggests a new systematic application of the existing methods in order to obtain similar results of the importance ranking of variables, independently of the method applied.

The database used in this paper is based on a customer satisfaction survey developed by the Transport Consortium of the Granada Metropolitan Area (Spain) in 2007. De Oña et al. [7, 8] have analyzed this database with other methodological approaches (e.g., decision trees and structural equation approaches) in order to identify the most important variables that contribute to the perception of service quality in a public transportation service.

The paper is structured in five sections. Section 2 describes ANN, the methods used to determine the contribution of variables, the methodology followed and the database used in this study. Sections 3 and 4 continue with the results and discussion. The paper concludes with a summary and directions for future research.

2 Neural networks and methods

2.1 Neural networks (NNs)

The MLP is a widely used NN typology, introduced by Werbos [51] and further developed and popularized by

Rumelhart and McClelland [46]. The multilayer feed-forward NN has been used in approximately 70 % of all ANN studies [18]. They are so successful because several research groups [16, 26] have mathematically demonstrated that a MLP neural network with a single hidden layer is a universal function approximator.

A gradient-descent supervised learning algorithm with a learning rate of 0.1 and a momentum of 0.9 was used to train the NN. This algorithm trains the NN by iteratively updating the synaptic weight values until the error function reaches a local minimum. The learning rate and momentum values help to accelerate the convergence [23, 47]. The weights were initialized before each training with small random values, and a number of 20,000 epochs were considered.

The database was randomly divided into training, validation and test sets, in a 70:15:15 ratio.

A wide range of NNs was trained, all of them characterized by a three-layer architecture: an input layer with I neurons (one neuron per input variable), a hidden layer with H neurons ($H \in [1, N]$) and an output layer with J neurons (one neuron per output variable). The neurons were activated using logarithmic sigmoidal transfer functions in all layers.

2.2 Methods for determining the contribution of variables

The methods selected in this study to determine the relative importance of variables in a NN model have been proposed and applied by numerous authors in several research fields [20, 38]: perturb, profile, connection weights and partial derivatives. The first two (perturb and profile) are pure SA methods, while the other two (connection weights and partial derivatives) are not.

2.2.1 Perturb method

This method is based on the principle of disturbing or introducing noise to one of the inputs, while the remaining variables keep their original values. Afterward, the MSE between the outputs obtained before and after the perturbation are compared [48, 52].

A noise δ was progressively applied to each variable in five steps: 10, 20, 30, 40 and 50 % of its original value. Thus, the variable x_i changes its values to $x_i = x_i + \delta$ due to the perturbation.

2.2.2 Profile method

This method analyses the evolution of each input along a scale or range of values, while the remaining variables keep their values fixed [29–31].

Each predictor variable x_i takes 11 different values resulting from the division of the range, between its minimum and maximum value, into 10 equal intervals. Furthermore, all variables except one are initially fixed at their minimum value and then successively at their first quartile, median, third quartile and maximum value. Thus, 5 values of the response variable are obtained for each of the 11 values adopted by x_i , and the median of those 5 values is calculated. Finally, a curve with the profile of variation is obtained for every variable.

2.2.3 Connection weights method

This method determines the relative importance of the predictor variables of the model as a function of the NN synaptic weights, according to the mathematical expression [38]:

$$R_{ij} = \sum_{H=1}^k W_{ik} \cdot W_{kj} \quad (1)$$

where R_{ij} is the relative importance of the variable x_i with respect to the output neuron j , H is the number of neurons in the hidden layer, W_{ik} is the synaptic connection weight between the input neuron i and the hidden neuron k , and W_{kj} is the synaptic weight between the hidden neuron k and the output neuron j .

2.2.4 Partial derivatives method

This method analyses the first-order effects of the predictor variables of the model with respect to the output variable, using all available training data [12].

The output provided by a neuron of the hidden layer of a MLP neural network with sigmoidal activation functions is given by the following equations:

$$h_k = \frac{1}{(1 + e^{-net_k})} \quad (2)$$

$$net_k = \sum_i W_{ik} \cdot x_i - \theta_k \quad (3)$$

where h_k is the output of the neuron k of the hidden layer, x_i is the value of the predictor variable of the considered input layer, W_{ik} is the connection weight between the predictor variable x_i and the neuron k of the hidden layer, and θ_k is the bias of the neuron k of the hidden layer.

The output of a neuron of the output layer is given by the following expressions:

$$y_j = \frac{1}{(1 + e^{-net_j})} \quad (4)$$

$$net_j = \sum_j W_{kj} \cdot h_k - \theta_j \quad (5)$$

where W_{kj} is the connection weight between the neuron k of the hidden layer and the neuron j of the output layer, θ_j is the bias of the output neuron j and y_j is the output of the neuron j of the output layer.

The expression that relates the variation in the output values y_j with respect to the variation in the predictor variable x_i is obtained through application of the chain rule:

$$\begin{aligned} \frac{dy_j}{dx_i} &= \sum_k \frac{dy_j}{dnet_j} \cdot \frac{dnet_j}{dh_k} \cdot \frac{dh_k}{dnet_k} \cdot \frac{dnet_k}{dx_i} \\ &= \sum_k f'_j \cdot W_{kj} \cdot f'_k \cdot W_{ik} \end{aligned} \tag{6}$$

$$f'_j = y_j \cdot (1 - y_j) \tag{7}$$

$$f'_k = h_k \cdot (1 - h_k) \tag{8}$$

The sensitivity value of every variable x_i is given by the expression:

$$L_i = \frac{\sum_P \frac{dy}{dx_i}}{P} \tag{9}$$

where P is the total number of training examples.

2.3 Methodology

The methodology proposed in this paper focuses on working with sets of NNs instead of a single NN. Every set is composed of a series of NNs with the same architecture, which are trained using an identical learning algorithm, activation functions, momentum value and learning ratio. NNs of the same set only differ in the initial random weight values considered in each training process. Once the NNs have been trained, the above four methods are applied to every one of them, and therefore, a ranking of relative importance is obtained for each NN and for each method. Due to the instability of the results when a single method is used, this methodology proposes an approach based on calculating the ranking of relative importance for each method as a function of the average importance values obtained from every NN in the set.

MATLAB software was used to develop the NNs [3].

The sequence of steps followed to develop this methodology was as follows:

Step 1 Train every one of the H NN architectures, with H $\in [1, N]$ neurons in the hidden layer, M number of times and using different random initial weights for each training. Thus, $N \times M$ trained ANNs are obtained.

Step 2 Determine the performance or capacity of generalization of the $N \times M$ trained NNs through the E error metric.

Step 3 Calculate the mean E and its standard deviation values for each one of the NN architectures, and select the

NN architecture that reaches the global minimum mean E value.

Step 4 Apply the methods of variable contribution to every one of the M NNs of the selected architecture.

Step 5 For every method applied, calculate the average value of the M NNs for every one of I predictor variables.

Step 6 Determine the ranking of importance of I variables considered in the study based on the values obtained in Step 5 for every applied method.

Step 7 Compare the results obtained for each method.

2.4 Data

The database used in this study was obtained through a customer service quality survey performed on public bus users by the Granada Area Transport Consortium in 2007. This Consortium was created to coordinate and organize the transit bus service of the Metropolitan Area of Granada (Spain).

A total of 858 surveys were conducted at the bus stops of different lines, with the aim of measuring the user satisfaction level regarding the service quality provided through 12 variables. The answers to these variables were scored from 0 to 10, as shown in Table 1.

In this case of study, the unitary value of the score obtained for each variable has been considered. That is, a range of values in the interval [0, 1] has been used as input values for every variable, instead of using the original interval [0, 10]. This translation allows to adapt them for subsequent treatment in the NN [33, 34], since the limits of

Table 1 Customer satisfaction survey's items

Variable	Symbol	Values
Input layer		
Information	INF	[0;10]
Punctuality	PUN	[0;10]
Safety	SAF	[0;10]
Courtesy	COU	[0;10]
Cleanliness	CLE	[0;10]
Space	SPA	[0;10]
Temperature	TEM	[0;10]
Accessibility	ACC	[0;10]
Fare	FAR	[0;10]
Speed	SPE	[0;10]
Frequency	FRE	[0;10]
Proximity	PRO	[0;10]
Output layer		
Quality of service	QS	[0;10]

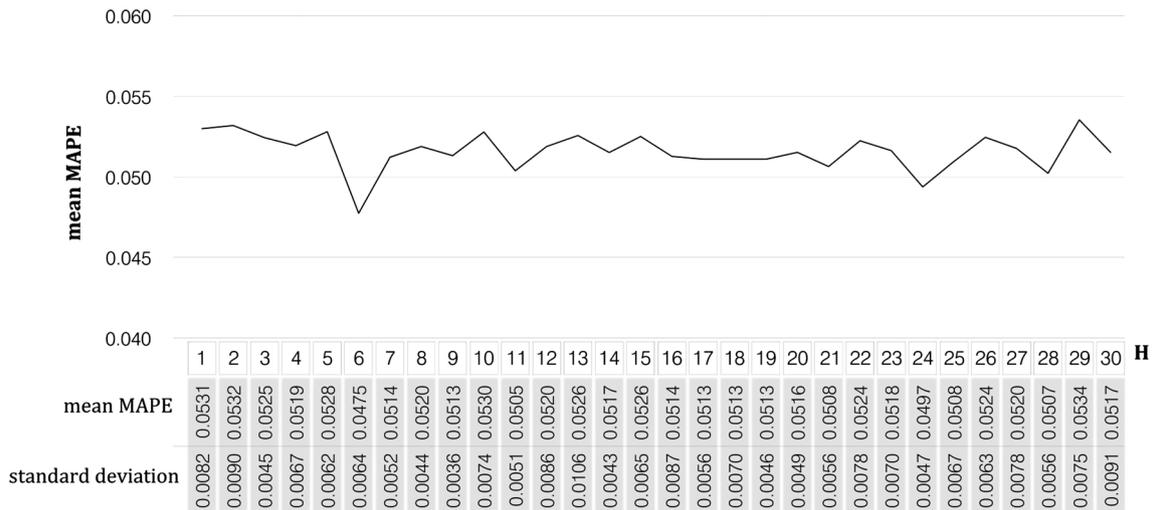


Fig. 1 Mean MAPE and standard deviation's values depending on H NN subsets

the value range of every variable directly coincide with the upper and lower limits of the sigmoidal activation functions used in NN models.

3 Results

In this study, it has been considered that $N = 30$ and $M = 50$, which came to a total of 1,500 networks, and the E error has been determined through the MAPE value, calculated according to the expression [9]:

$$MAPE = \frac{1}{T} \cdot \sum_{i=1}^T \text{abs} \left(\frac{\text{Actual value } i - \text{Set point value } i}{\text{Set point value } i} \right) \quad (10)$$

where T is the total number of considered cases in the test stage.

Figure 1 shows the mean and the standard deviation values of MAPE obtained for every NN architecture. The MLP neural network with 6 neurons in the hidden layer reaches the minimum mean MAPE value among all trained networks. This network architecture, along with its 50 MLP trained networks, is selected to analyze the outcomes of the four methods considered in this paper.

Figures 2, 3, 4 and 5 show the relative importance of the 12 independent variables of the 50 networks used, obtained after applying the four methods. A wide variability of the importance values has been observed; hence, every variable ranking varies greatly, depending not only on the method applied, but also on the different random initial values used for a same method. Some other researchers [4, 54] have already set out this problem and suggested the possibility of working with NN sets to control the instability problems derived from sensitivity analysis. This

supports the idea of working not only with a single NN but also with a set of them, as we are proposing in this paper.

The profile and perturb methods are able to enclose the relative importance of certain variables in a narrower range, as in the case of the variable Frequency, which always has a relative importance above 70 % according to the profile method, and above 55 % according to the perturb method, or the variable Cleanliness, whose importance is always under 50 % according to the profile method, and under 20 % according to the perturb method. The results of the other two methods, specially the PaD, show that any variable can reach any relative importance value between 0 % and 100 %. In either case, the ranking of importance of these variables is too unpredictable for the four methods. This is a major limitation for adequately determining the relative importance of the variables. However, if the values obtained individually for every variable are averaged for every method, results are more homogeneous.

In the profile method, a range of 50 profiles of variation within the interval [0, 1] was generated (Fig. 6). The relative importance of a variable is given by the difference between the maximum and minimum values (difference in the axis of ordinates) of the line representing the average of the profile of variation. Table 2 lists the results obtained, which indicate that the variable Frequency reaches the highest relative importance (100 %), followed by Speed (77.72 %), Information (64.15 %) and Proximity (60.24 %); so these variables have a very high global importance. A second level of importance, considered as high, includes the variables Punctuality (54.45 %), Safety (53.28 %) and Courtesy (48.59 %), followed by a third medium importance level containing the variables Temperature (38.44 %), Fare (36.40 %) and Space (27.22 %). In the last place of relative importance are the variables Accessibility (17.34 %) and Cleanliness (3.36 %).

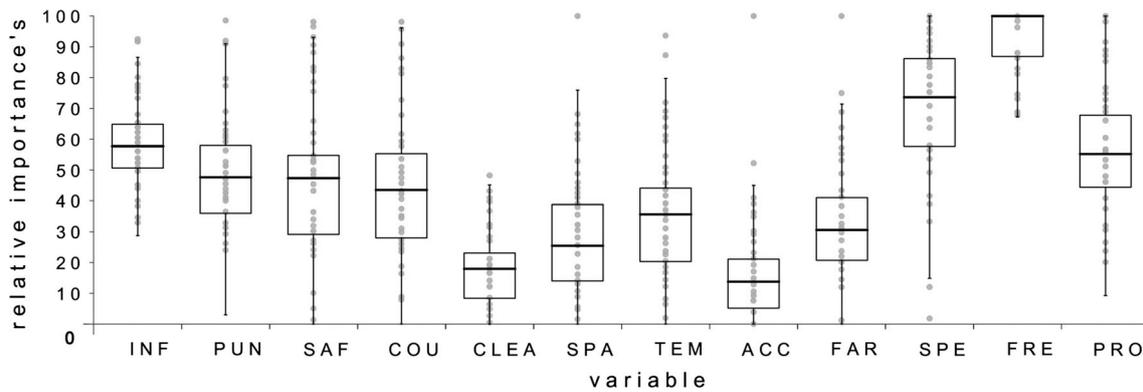


Fig. 2 Boxplot of variables' relative importance by profile method

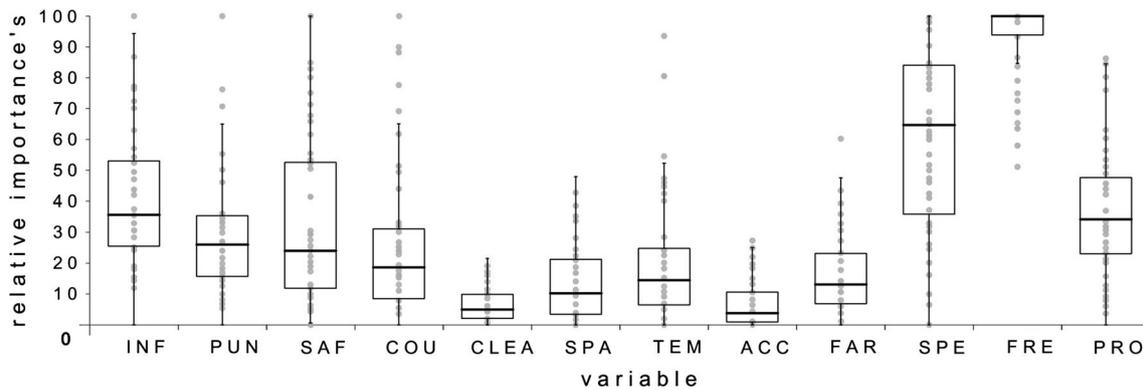


Fig. 3 Boxplot of variables' relative importance by perturb method

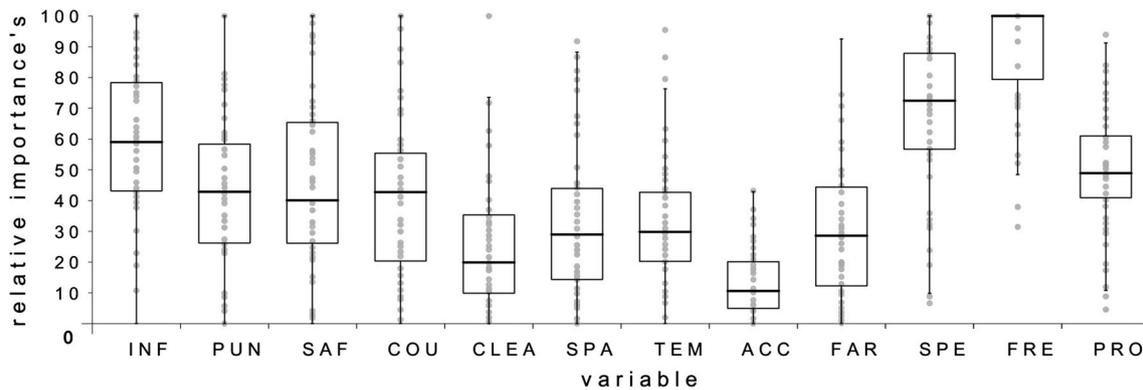


Fig. 4 Boxplot of variables' relative importance by connection weights method

A range of 50 profiles for every variable was also generated in the perturb method, representing the profiles of variation in the MSE error, as a function of the noise percentage or perturbation introduced (Fig. 7). Once again, the difference between the maximum and minimum values of the line representing the average of the profile values indicates the relative importance of the variables. The higher the variation in the MSE, the higher is the relative importance of a variable. Table 2 shows the results

obtained after applying this method. Frequency (100 %) is globally the most important variable, followed by Speed (63.60 %). The variables Information (42.88 %), Punctuality (32.92 %), Safety (32.92 %), Courtesy (30.33 %), Proximity (23.28 %) and Temperature (22.27 %) belong to a second medium importance level. In the last places of the ranking appear the variables Fare (17.90 %), Space (14.51 %), Accessibility (7.58 %) and Cleanliness (7.12 %).

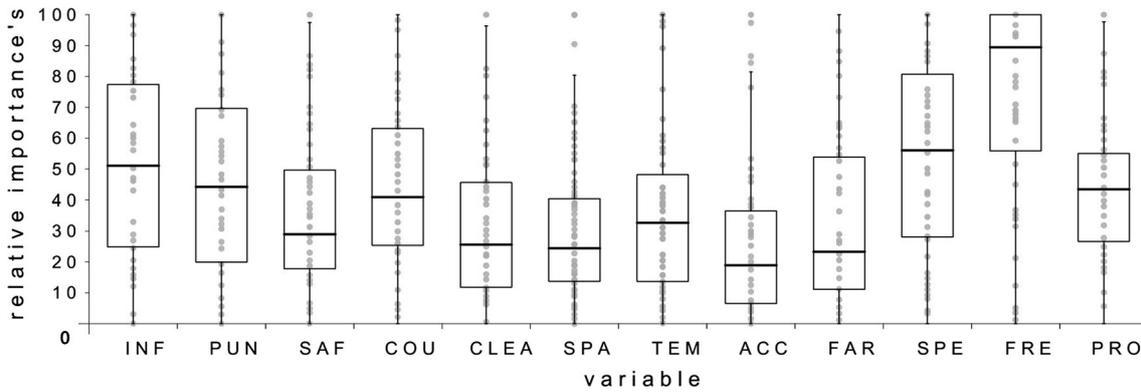


Fig. 5 Boxplot of variables' relative importance by partial derivatives method

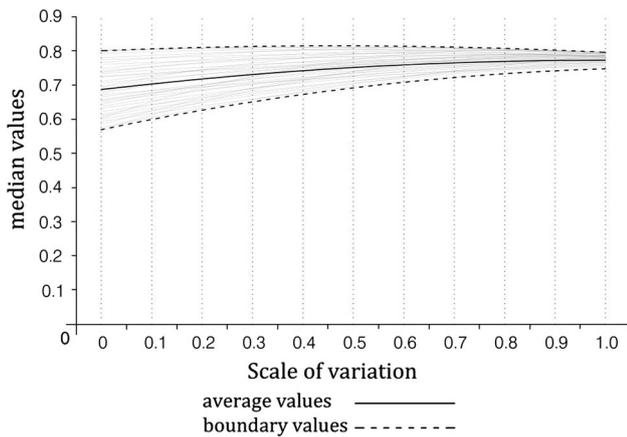


Fig. 6 Average and boundary values of the variable information profile method

Regarding the connection weights and PaDs, the corresponding formulas are applied to determine the relative importance of every variable (Eqs. 1, 9, respectively), and the average values and ranking have been obtained and compiled in Table 2.

The connection weights method shows that the variable with the highest importance is Frequency (100 %), followed by Speed (75.98 %) and Information (66.68 %). In a second level of high importance are included the variables Proximity (55.49 %), Safety (51.38 %), Punctuality (51.35 %) and Courtesy (47.81 %), followed by the group of medium importance variables containing Temperature (36.65 %), Space (36.45 %), Fare (31.98 %) and Cleanliness (27.39 %). The variable with less relative importance is Accessibility (14.56 %).

In the PaD, the most important variables are Frequency (100 %), Speed (72.90 %), Information (71.01 %), Punctuality (64.13 %), Courtesy (60.67 %) and Proximity (58.30 %). The relative importance of the remaining variables can be considered as medium: Temperature (49.05 %), Safety (48.08 %), Fare (45.43 %), Cleanliness (44.01 %), Space (43.35 %) and Accessibility (37.41 %).

Figure 8 shows a comparison of relative importance ranking of the 12 independent variables determined by every method. This figure shows that the relative importance is very similar, and therefore, the high instability inherent in these methods when applied to a single NN is considerably eliminated. The four methods agree on the fact that the most influencing variables are Frequency, Speed and Information, and that the least influencing variables are Cleanliness, Accessibility and Space, with the sole exception of the connection weights method, which considers the variables Accessibility, Cleanliness and Fare to be the least important, followed by the variable Space. The six remaining variables present intermediate positions in the ranking. The profile and perturb methods provide a more similar hierarchy of importance, while the PaD presents more discrepancies.

With regard to the degree of relative importance assigned to every variable, expressed as a relative percentage, it is noticed that in the profile and perturb methods, the differences allow the variables to be classified into four levels of importance: very high, high, medium and low. The perturb method allows a clear distinction to be made between three levels of importance: high, medium and low. The PaD allocates percentages of influence that can be differentiated into two levels of importance: high and medium, since the lowest value of importance is above 37 %.

4 Discussion

This paper, as demonstrated by many other previous studies [1, 9, 35–37], confirms the high prediction capacity in the generalization stage of NN, attaining MAPE's values below 0.05, which means more than 95 % of right answers in most of the trained NN.

Several authors [19, 20, 41, 38, 43, 49]) have analyzed the advantages and disadvantages of the existing methods

Table 2 Variables' relative importance by methods

Variable	Derived importance							
	Profile		Perturb		Connection weights		Partial derivatives	
	Average	Ranking	Average	Ranking	Average	Ranking	Average	Ranking
INF	64.15	3	42.88	3	66.68	3	71.01	3
PUN	54.45	5	32.92	4–5	51.35	6	64.13	4
SAF	53.28	6	32.92	4–5	51.38	5	48.08	8
COU	48.59	7	30.33	6	47.81	7	60.67	5
CLE	3.36	12	7.12	12	27.39	11	44.01	10
SPA	27.22	10	14.51	10	36.45	9	43.35	11
TEM	38.44	8	22.27	8	36.65	8	49.05	7
ACC	17.34	11	7.58	11	14.56	12	37.41	12
FAR	36.40	9	17.90	9	31.98	10	45.43	9
SPE	77.72	2	63.60	2	75.98	2	72.90	2
FRE	100.00	1	100.00	1	100.00	1	100.00	1
PRO	60.24	4	23.28	7	55.49	4	58.30	6

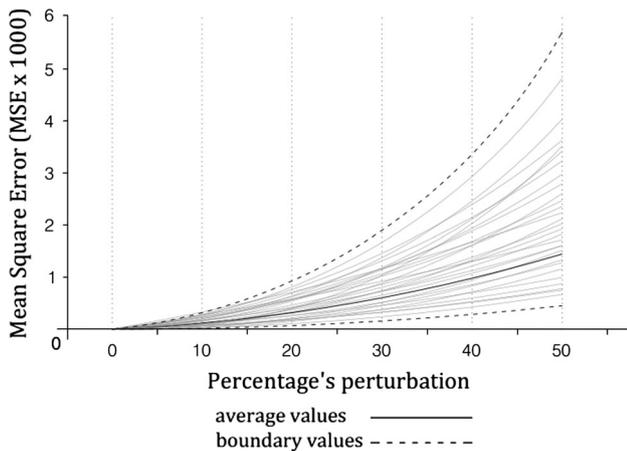


Fig. 7 Average and boundary values of MSE of the variable information perturb method

to determine the relative contribution of the variables in the NN models by comparing them, but it has not been possible to establish a global consensus on which method is the most stable, accurate or robust. This leaves it open to doubt whether these methods, which determine the cause–effect relationship between the predictor and dependent variables, are indeed clarifying approaches that explain the role of every variable in NN models, the so-called black boxes, and most importantly, whether they are reliable enough to be used in other research fields where their results can have serious and compromising consequences.

The values of relative importance obtained by applying the profile, perturb, connection weights and PaDs to a single NN show high variability, not only when different methods are applied to the same NN, but also when one of

them is applied several times to a certain NN architecture that has been trained with different initial random weights. Therefore, working this way does not guarantee the validity of the results of relative importance. However, in this study, we worked with NN sets of the same architecture on which the above methods were applied. This approach achieved a similar ranking of relative importance regardless of the method used, and the results are particularly similar to the profile and perturb methods. Thus, the main goal was attained.

Any method can be considered valid, since all of them clearly agree to identify the most and least important variables, although the PaD is the least recommended because it shows a higher variability of the relative importance values.

In addition, another argument that supports the robustness of this new approach is the fact that the variables Frequency, Speed, Punctuality and Proximity are classified as the most important by all methods, and this classification agrees with the results obtained by authors [7, 10, 11, 13, 14] who have used other techniques, such as multinomial logit models, multinomial discrete choice models, ordered probit models and decision trees.

The main weakness of this approach is that, even though the four methods provide similar results in terms of ranking variable importance, the percentages of relative importance are significantly different depending on the method used, and therefore, the opinion of an expert is necessary to decide which method offers the most similar results to those expected. Thus, while the profile and perturb methods deliver a wide range of values within the interval [3.36;100.0] (the former) and [7.1;100.0] (the latter), the connection weights method limits this range to the interval

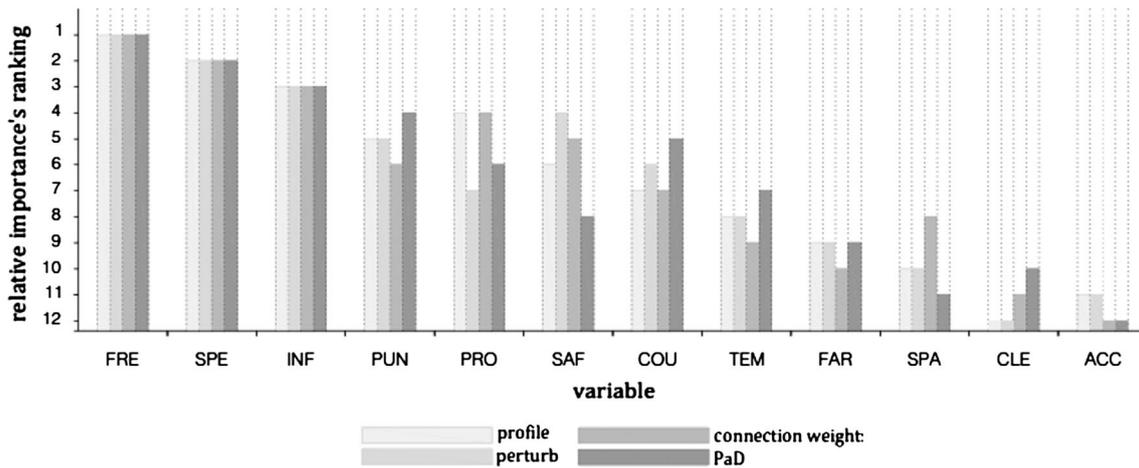


Fig. 8 Comparative of ranking's relative importance by methods

[14.5;100.0] and the PaD to the interval [37.4;100.0]. The importance assigned to intermediate-positioned variables also differs from method to method.

5 Conclusions

This paper presents an approach that mitigates the instability problems inherent in the methods used to determine the contribution of predictor variables in a NN model. The principle of this approach is based not only on the modification of existing methods or the introduction of new ones but also on the application of these methods to a set of NNs instead of a single NN.

A set of ANNs with the same architecture was selected on which several existing and previously used methods were applied to determine the contribution of variables. Afterward, a new treatment of the results was carried out, based on the calculation of the average values of the relative importance of variables.

The database used comes from a survey conducted on users of the bus transit service to know the service quality as perceived by them in the Metropolitan Area of Granada. This survey was carried out for different purposes to those pursued in this paper, and in addition, it has been analyzed with different techniques in other studies [7, 8]. We would like to point out that, to the authors' knowledge, this is the first time that ANNs are used to analyze service quality.

This approach seems to be stable, since all the methods used assign a very similar hierarchy of importance to the variables, especially to those that have a greater impact. Additionally, these results concur with those from other studies related to the subject, which support the validity of the results.

There are, however, some differences in the importance percentage assigned to the variables depending on the

method applied, since there is no single unique subset of the most important variables because of the inter-correlation of variables, and this importance is limited by the individual metric. Therefore, the opinion of an expert is necessary to evaluate which method shows importance values more concordant with the expected results.

The advantages of this approach overcome the drawbacks, since it is achieved the goal of significantly eliminating the high instability existing in current methods, as they are applied so far in the NN field.

This new perspective on the application of classical methods to NN sets offers great possibilities for future research, which could study what happens with other existing methods for determining the contribution of variables, such as the recent DSA method [6], with other pre-processing approaches for standardizing the input values or with other typologies and architectures of NNs, with the aim of checking the robustness showed by this approach.

In terms of future work, it would be interesting to measure, after identifying most relevant inputs by methods, how a particular input tends to affect the NN output response. This is valuable in real-world applications, and for instance, it has been applied in Cortez and Embrechts [6] using VEC curves, which are graphical representations that visualize these changes in the NN output response.

Acknowledgments Support from Consejería de Innovación, Ciencia y Economía of the Junta de Andalucía (Spain) (Research Project P08-TEP-03819, co-funded by FEDER) is gratefully acknowledged. The authors also acknowledge the Granada Consorcio de Transportes for making the data set available for this study.

References

1. Akin D, Akbaç B (2010) A neural network (NN) model to predict intersection crashes based upon driver, vehicle and roadway surface characteristics. *Sci Res Essays* 5(19):2837–2847

2. Azadeh A, Rouzbahman M, Saberi M, Fam IM (2011) An adaptative neural network algorithm for assessment and improvement of job satisfaction with respect to HSE and ergonomics program: the case of a gas refinery. *J Loss Prev Process Ind* 24:361–370
3. Beale MH, Hagan MT, Demuth HB (2007) *Neural Network Toolbox 7. User's Guide*. MathWorks, Inc. 3 Apple Hill Drive Natic, MA 01760-2098
4. Cao M, Qiao P (2008) Neural network committee-based sensitivity analysis strategy for geotechnical engineering problems. *Neural Comput Appl* 17:509–519
5. Cortez P, Embrechts MJ (2011) Opening black box data mining models using sensitivity analysis. *IEEE symposium series in computational intelligence*, Paris, France, 4, 2011
6. Cortez P, Embrechts MJ (2013) Using sensitivity analysis and visualization techniques to open black box data mining models. *Inf Sci* 225:1–17
7. De Oña J, De Oña R, Calvo FJ (2012) A classification tree approach to identify key factors of transit service quality. *Expert Syst Appl* 39:11164–11171
8. De Oña J, De Oña R, Eboli L, Mazzulla G (2013) Perceived service quality in bus transit service: a structural equation approach. *Transp Policy* 29:219–226
9. Delen D, Sharda R, Bessonov M (2006) Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. *Accid Anal Prev* 38:434–444
10. Dell'Olio L, Ibeas A, Cecín P (2010) Modelling user perception of bus transit quality. *Transp Policy* 17(6):388–397
11. Dell'Olio L, Ibeas A, Cecín P (2011) The quality of service desired by public transport users. *Transp Policy* 18(1):217–227
12. Dimopoulos Y, Bourret P, Lek S (1995) Use of some sensitivity criteria for choosing neural networks with good generalization ability. *Neural Process Lett* 2:1–4
13. Eboli L, Mazulla G (2008) Willingness-to-pay of public transport users for improvement in service quality. *Eur Transp* 38:107–118
14. Eboli L, Mazulla G (2010) How to capture the passengers' point of view on a transit service through rating and choice opinions. *Transp Rev* 30:435–450
15. Engelbrecht AP, Cloete I, Zurada JM (1995) Determining the significance of input parameters using sensitivity analysis, from natural to artificial neural computation. In: *proceedings of International Workshop on Artificial Neural Networks*. Málaga-Torremolinos, Springer, Spain, pp 382–388
16. Funahashi KI (1989) On the approximate realization of continuous mappings by neural networks. *Neural Netw* 2:183–192
17. Garson GD (1991) Interpreting neural-network connection weights. *Artif Intell Expert* 6:47–51
18. Gedeon, T.D., Wong, P.M. & Harris, D., (1995). Balancing the bias and variance: network topology and pattern set reduction techniques. In: *proceedings of International Workshop on Artificial Neural Networks*, IWANN95, Torremolinos, España, pp 550–558
19. Gedeon TD (1997) Data mining of inputs: analyzing magnitude of functional measures. *Int J Neural Syst* 8(2):209–218
20. Gevrey M, Dimopoulos I, Lek S (2003) Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecol Model* 160:249–264
21. Gevrey M, Dimopoulos I, Lek S (2006) Two-way interaction of input variables in the sensitivity analysis of neural network models. *Ecol Model* 195:43–50
22. Goh ATC (1995) Back-propagation neural networks for modeling complex systems. *Artif Intell Eng* 9:143–151
23. Hagan MT, Demuth HB, Beale MH (1996) *Neural network design*. Campus Publishing Service, Colorado University Bookstore, Colorado. ISBN 0-9717321-0-8
24. He F, Sung AH, Guo B (1997) A neural network for prediction of oil well cement bonding quality. In: *proceedings of IASTED international conference on control*, IASTED-ACTA Press, Cancun-Mexico, pp 417–420
25. Hunter A, Kennedy L, Henry J, Ferguson I (2000) Application of neural networks and sensitivity analysis to improved prediction of trauma survival. *Comput Methods Progr Biomed* 62:11–19
26. Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. *Neural Netw* 2:359–366
27. Kemp SL, Zaradic P, Hansen F (2007) An approach for determining relative input parameter importance and significance in artificial neural networks. *Ecol Model* 204:326–334
28. Kewley R, Embrechts M, Breneman C (2000) Data strip mining for the virtual design of pharmaceuticals with neural networks. *IEEE Trans Neural Netw* 11(3):668–679
29. Lek S, Beland A, Dimopoulos I, Lauga J, Moreau J (1995) Improved estimation, using neural networks, of the food consumption of fish populations. *Mar Freshw Res* 46(8):1229–1236
30. Lek S, Delacoste M, Baran P, Dimopoulos I, Lauga J, Aulagnier S (1996) Application of neural networks to modeling nonlinear relationships in ecology. *Ecol Model* 90:39–52
31. Lek S, Beland A, Baran P, Dimopoulos I, Delacoste M (1996) Role of some environmental variables in trout abundance models using neural networks. *Aquat Living Resour* 9:23–29
32. Lin Y, Cunningham GA (1995) A new approach to fuzzy-neural system modeling. *IEEE Trans Fuzzy Syst* 3(2):190–198
33. Martín del Bío B, Sanz Molina A (2006) *Neural networks and fuzzy systems*. Editorial RA-MA
34. Masters T (1993) *Practical neural networks recipes in C ++*. Academic Press, Waltham
35. Moghaddam FR, Afandizadeh S, Ziyadi M (2010) Prediction of accident severity using artificial neural networks. *Int J Civ Eng* 9:1
36. Mohammadipour AH, Alavi SH (2009) The optimization of the geometric cross-section dimensions of raised pedestrian crosswalks: a case of study in Qazvin. *Accid Anal Prev* 41:314–326
37. Mussone L, Ferrari A, Oneta M (1999) An analysis of urban collisions using an artificial intelligence model. *Accid Anal Prev* 31:705–718
38. Olden JD, Jackson DA (2002) Illuminating the “black-box”: a randomization approach for understanding variable contributions in artificial neural networks. *Ecol Model* 154:135–150
39. Olden JD, Joy MK, Death RG (2004) An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecol Model* 178:389–397
40. Özesmi SL, Özesmi U (1999) An artificial neural network approach to spatial habitat modeling with interspecific interaction. *Ecol Model* 116:15–31
41. Paliwal M, Kumar UA (2011) Assessing the contribution of variables in feed forward neural network. *Appl Soft Comput* 11:3690–3696
42. Palmer A, Montaña JJ (2002a) *Redes neuronales artificiales aplicadas al análisis de datos*. Doctoral Dissertation. University of Palma de Mallorca
43. Palmer A, Montaña JJ (2002) *Redes neuronales artificiales: abriendo la caja negra*. Metodología de las ciencias del comportamiento 4(1):77–93
44. Palmer A, Montaña JJ (2003) Numeric sensitivity analysis applied to feed forward neural networks. *Neural Comput Appl* 12:119–125
45. Rzepoluk EJ (1998) *Neural network data analysis using simulnet*. Springer, New York
46. Rumelhart DE, McClelland JL (1986) *Parallel distributed processing, vol 1: foundations*. MIT Press, Cambridge

47. Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by backpropagation errors. *Nature* 323:533–536
48. Scardi M, Harding LW (1999) Developing an empirical model of phytoplankton primary production: a neural networks case study. *Ecol Model* 120(2–3):213–223
49. Sung AH (1998) Ranking importance of input parameters of neural networks. *Expert Syst Appl* 15:405–411
50. Watts MJ, Worner SP (2008) Using artificial neural networks to determine the relative contribution of abiotic factors influencing the establishment of insect pest species. *Ecol Inform* 3:64–74
51. Werbos PJ (1974) Beyond regression: new tools for prediction and analysis in behavioral sciences. Doctoral Dissertation. Applied Mathematics, Harvard University
52. Yao J, Teng N, Poh HL, Tan CL (1998) Forecasting and analysis of marketing data using neural networks. *J Inf Sci Eng* 14:843–862
53. Yeh I, Cheng W (2010) First and second order sensitivity analysis of MLP. *Neurocomputing* 73:2225–2233
54. Zhou ZH, Wu J, Tang W (2002) Ensembling neural networks: many could be better than all. *Artif Intell* 137(1–2):239–263
55. Zurada JM, Malinowski A, Cloete I (1994) Sensitivity analysis for minimization of input data dimension for feed forward neural network. In: proceedings of IEEE international symposium on circuits and systems, IEEE Press, London



Neural networks for analyzing service quality in public transportation



Concepción Garrido, Rocío de Oña, Juan de Oña*

TRYSE Research Group, Department of Civil Engineering, University of Granada, ETSI Caminos, Canales y Puertos, c/Severo Ochoa, s/n, 18071 Granada, Spain

ARTICLE INFO

Keywords:

Service quality
Bus transit
Neural networks
ANN
MLP
Profile
Perturb
Connection Weights

ABSTRACT

It is essential to take into account the service quality assessment made by the passengers of a public transportation system, as well as the weight or relative importance assigned to each one of the attributes considered, in order to know its strengths and weaknesses. This paper proposes using Artificial Neural Networks (ANN) to analyze the service quality perceived by the passengers of a public transportation system. This technique is characterized by its high capability for prediction and for capturing highly non-linear intrinsic relations between the study variables without requiring a pre-defined model. First, an ANN model was developed using the data gathered in a Customer Satisfaction Survey conducted on the Granada bus metropolitan transit system in 2007. Next, three different methods were used to determine the relative contribution of the attributes. Finally, a statistical analysis was applied to the outcomes of each method to identify groups of attributes with significant differences in their relative importance. The results show that statistical significant differences exist among several categories of attributes that have a greater or lesser impact on service quality and satisfaction. All the methods agree that Frequency is the most influential attribute in the service quality, and that other attributes such as Speed, Information and Proximity are also important.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Currently, an extended use of public transport modes among citizens is one of the key aims of public administrations. Nowadays, the role of public transportation is viewed as an alternative to private cars instead of being just the support for the movement of passengers (Simoes, 2013). For individuals, car travel is generally perceived as more comfortable, flexible and faster for supporting busy lifestyles (Jakobsson Bergstad, Gamble, Hagman, Polk, & Garling, 2011). However, the excessive use of these private vehicles generates environmental and social problems in cities (e.g. pollution, traffic congestion, noise, etc.), greatly exacerbating of the unsustainability of citizens' mobility.

Public transport services have to prove that they can compete with other modes, by guaranteeing effective and high quality services. The authorities are attempting to impose strong incentives on operators (Mouwen & Rietveld, 2013) by using a good definition of service quality and a good measuring method. Given that public transport services are offered directly to customers, the resultant quality of a service should be seen as an outcome of user perception (Das and Pandit, 2013) because as Bordagaray, Dell'Olio, Ibeas, and Cecín (in press) stated, "Without the consumer, the

market has no reason to exist". Therefore, the level of quality in a service will be high when the performance of the service fits passengers' needs and expectations.

In recent decades, practitioners, managers and researchers have focused their attention on this point of view (De Oña & De Oña, in press), striving to learn more details about how passengers evaluate a service, by considering the impact of various attributes that characterize it. Several authors have stated that service quality is a complex, fuzzy and abstract concept (Carman, 1990; Parasuraman, Zeithaml, & Berry, 1985), mainly because of the three properties of service: intangibility, heterogeneity and inseparability, but also because of the subjective nature of considering passengers' opinions for measuring this quality. In the literature, there are very different methods for determining this influence, although there is no consensus as to which is the best one. That is why measuring service quality is still a challenge for researchers and transport planners.

Various authors pointed to the existence of several categories of attributes that have a greater or lesser impact on service quality and satisfaction. Philip and Hazlett (1997) proposed a model with a hierarchical structure, based on three classes of attributes: pivotal, core and peripheral attributes. This model was subsequently contrasted for the rail transportation industry (Tripp & Drea, 2002). The pivotal attributes exert the greatest influence on the satisfaction levels. The UNE-EN 13186 (2003) standard classifies

* Corresponding author. Tel.: +34 958 24 99 79; fax: +34 958 24 61 38.
E-mail address: jdona@ugr.es (J. de Oña).

the service's characteristics into basic, proportional and attractive, depending on how compliance and non-compliance affects customer satisfaction. The Transit Capacity and Quality of Service Manual (Transportation Research Board, 2004) groups attributes into availability factors (more important to passengers), and comfort and convenience factors (less important). Eboli and Mazzulla (2008) empirically demonstrated the existence of two categories of attributes (basic and not basic) in the preferences showed by users.

The influence of these characteristics on passengers' overall evaluation can be determined in different ways. For example, in recent years, structural equation models have gained popularity among researchers, such as De Oña, de Oña, Eboli, and Mazzulla (2013), Eboli and Mazzulla (2007, 2012), Irfan Syed Muhammad and Shahbaz Saman (2011) or Ngatia, Okamura, and Nakamura (2010). For others (e.g. Bordagaray, Dell'Olio, Ibeas, & Cecín, (in press); Eboli & Mazzulla, 2008, 2010; Hensher, 2014; Hensher & Prioni, 2002; Hensher, Stopper, & Bullock, 2003; Marcucci & Gatta, 2007) discrete choice models are a great method for deriving the importance of service quality attributes. However, most of these models have their own model assumptions and pre-defined underlying relationships between dependent and independent variables, such as normal data, linear relationships between dependent and independent variables, low multi-collinearity, and so on. According to Garver (2003), these assumptions are almost always violated in customer satisfaction research.

However, as Kikuchi (2012) says "the traditional paradigms of prediction, diagnosis, and regulation or optimization are not sufficient to deal with the extremely complex social and human systems, of which transportation engineering and planning are part". Following this idea, a novel insight has been proposed by De Oña, de Oña, and Calvo (2012); De Oña, de Oña, Eboli, and Mazzulla (in press) applying a data mining technique for overcoming the aforementioned weaknesses and analyzing quality of service for transit operation. The methodology used was a decision tree model, which needs neither model assumptions nor pre-defined underlying relationships between the independent and the dependent variables. Following this direction, and because of the powerful results obtained with the decision tree model, the authors' interest for other data mining techniques increased.

To the authors' knowledge, the neural network approach, which is also a non-parametric model with similar advantages to the tree models has not been used before for analyzing service quality in public transportation, although it has been successfully used in other transportation engineering fields such as choice behavior (Lee, Ran, Yang, & loh, 2010; Xie, Lu, & Parkany, 2003). As an example, Lee et al. (2010) applied the ANN and decision trees methodology to analyze the factors affecting car drivers' alternative route choice; while Xie et al. (2003) modeled work travel mode based on three different methodologies: decision trees, ANN and multinomial logit models. Both studies concluded that ANN achieved the best fitting of the problem, with higher accuracies than the decision tree models. Thus, the aim of the present study is to use an ANN approach to investigate the influence of service characteristics on passengers' overall evaluation of a service to know the relative importance assigned to the service quality attributes. Three different methods of relative contribution (Connection Weights, Perturb and Profile) will be used. Another objective of this paper is to verify the hypotheses of Eboli and Mazzulla (2008) on the existence of different categories of attributes, and to determine if significant differences exist among groups of attributes that have higher and lower importance in overall service quality.

This paper is structured in five sections. First, the experimental context is described, specifying the survey conducted for collecting the data, and the main characteristics of the sample and perceptions rates. The following section is about the methodology and

framework adopted in this context. In the fourth section the results of the analysis are shown and discussed. Finally, a brief concluding section is reported.

2. Experimental context

The survey supporting the research targeted sample of users of the metropolitan public transport service operating in Granada (Spain). The service was provided by a bus system in which 15 bus companies operated, connecting different urban agglomerations of the metropolitan area of Granada. The Transport Consortium of Granada carried face-to-face interviews on March 2007, during 5 days of a week. Passengers were interviewed at the main bus stops of the service, collecting a final random sample of 858 people.

The questionnaire was structured into two main sections. The aim of the first section was to collect data concerning: general information on the trip (e.g. time of the interview, bus stop, line, operator, origin, destination), socioeconomic characteristics of passengers (e.g. gender, age, private vehicle availability) and travel habits (travel reason, frequency, ticket, complementary modes from origin to bus stop and from bus stop to destination).

The majority of the respondents were female (Table 1). More than a half of the respondents were 18–30 years old, and only 9.5% were older than 60. Most of the people sampled (61.1%) had a private vehicle available for doing the trip. About a 29.4% of people traveled for business reasons and a similar percentage for studies purposes. The rest of the respondents traveled for other reasons, such as doctor, shopping, holidays and so on. Most of the passengers traveled with an almost daily or frequently frequency, while occasionally and sporadic passengers represented only about 10% of the sample. The most usual complementary mode used for reaching the bus stop or for reaching the destination from the bus stop was on foot. Other complementary modes had also been used, such as car, urban bus, motorbike, etc., although its representativeness was low. Finally, the consortium card and the standard ticket were the most widespread types of tickets among passengers, representing more than an 80% when combined.

The second section was more oriented to collect passenger opinions about the service. Specifically, users rated importance and perception on 12 service quality attributes that characterize the service, and also rated the overall quality of the service. A cardinal scale from 0 to 10 was used for state the ratings. The attributes used to characterize the service included information, punctuality, safety on board, driver courtesy, bus interior cleanliness, bus space, bus temperature, accessibility to/from the bus, fare, speed, frequency of service and stops proximity to/from origin/destination.

Table 2 shows the average rates calculated from the collected data. It can be seen that there is very little variation in the importance rates stated by the passengers in the survey, considering that all the attributes are highly important. The average value of importance is concentrated in the 8.5–9.5 range. Therefore, this importance is uniform and practically equal in all the attributes. This is one of the serious drawbacks encountered when studying the importance of variables based on the stated opinions of passengers (De Oña et al., 2012; Weinstein, 2000).

On the contrary, the average perception rates show higher differences among attributes. In all the cases they are lower than the mean values of the importance rates. They are concentrated in a range from 6 to 8. Nonetheless, these values are quite good because all the attributes are perceived with at least adequate quality (>6), and some of them with quite good quality (>7). The attributes characterized by the highest levels of quality were

Table 1
Sample characteristics.

Characteristics	Statistics
1. Gender	Male (33.0%), female (67.0%)
2. Age	18–30 (56.5%), 31–60 (34.1%), >60 year-olds (9.5%)
3. Private vehicle availability	Yes (38.9%), no (61.1%)
4. Travel reason	Occupation (29.4%), studies (22.9%), doctor (14.2%), shopping (4.4%), personal activities (18.7%), holidays (0.2%), leisure time (8.6%), others (1.5%)
5. Frequency	Almost daily (67.98%), frequently (20.58%), occasionally (8.94%), sporadic (2.49%)
6. Complementary modes from origin to bus stop	On foot (77.6%), car (1.9%), urban bus (16.9%), Motorbike (0.5%), others (3.1%)
7. Complementary modes from bus stop to destination	On foot (94.5%), car (2.1%), urban bus (2.3%), motorbike (0.2%), others (0.9%)
8. Type of ticket	Consortium card (49.6%), standard ticket (41.2%), senior citizen pass (4.8%), others (4.4%)

Table 2
Importance and perceptions average rates.

Attributes	Average importance rates	Average perception rates
Information	8.62	6.86
Punctuality	9.14	7.41
Safety on board	8.98	7.73
Driver courtesy	8.77	7.96
Bus interior cleanliness	8.86	7.46
Bus space	8.66	7.21
Bus temperature	8.72	7.43
Accessability to/from the bus	8.91	6.90
Fare	8.77	6.44
Speed	8.73	7.30
Frequency of service	9.05	6.99
Proximity to/from origin/destination	8.71	7.43
Overall service quality		7.10

Driver courtesy, Safety on board and Bus temperature, and the one judged with the lowest level of quality was the Fare. Finally, passengers are quite satisfied with the service, with an average overall evaluation rate of 7.10.

3. Methodology

3.1. Artificial Neural Networks (ANN)

ANN are information processing systems based on the behavior of the human brain (Martín del Bío & Sanz Molina, 2006). ANN capture the inherent information from the considered variables and learn from the existing data, even when noise is present (Kasabov, 1996), therefore no formulation or a priori model is required (Watts & Worner, 2008).

ANNs structure is composed of elemental information processing units, called neurons. They are organized into several layers and interconnected with each other through synaptic weights. Synaptic weights represent the intensity of the interaction between every pair of neurons, and the activation functions calculate the potential of every neuron (Martín del Bío & Sanz Molina, 2006).

The multilayer perceptron (MLP) with the back propagation learning algorithm is used in this study, since it is the most widely used type of ANN in numerous previous researches (Gedeon, Wong, & Harris, 1995), and it is also a universal function approximator (Funahashi, 1989). The information always flows from the input neurons to the output neurons, and no feedback exists. The back propagation algorithm, first introduced by Werbos (1974) and further developed by Rumelhart and McClelland (1986), is the most popular learning rule (Azadeh, Rouzbahman, Saberi, & Mohammad Fam, 2011).

A set of data with their target outputs is fed into the network during the supervised learning process, and an error function, represented by a hills and valleys surface, is defined. The synaptic weights values are iteratively updated until the provided output tends to be the desired, and the error function descends along the surface towards a local minimum. By defining the momentum and the learning ratio parameters, the learning convergence accelerates (Hagan, Demuth, & Beale, 1996), because the former modifies the fixed learning rhythm depending on the sign of the updated weights and the latter controls the size of the weights changes. A cross validation process has to be carried out during the learning phase to avoid overfitting. The training algorithm continues until a pre-defined number of 20,000 epochs. Thus, the database is randomly divided into three sets: training, validation and test (Bishop, 1995; Haykin, 1999), in a 70/15/15 ratio. The test set, which contains T data, determines the global performance of the trained ANN, according to the MAPE (Delen, Sharda, & Bessonov, 2006) equation (see Eq. (1)).

$$MAPE = \frac{1}{T} \cdot \sum_{i=1}^T abs \left(\frac{\text{Actual value } i - \text{Set point value } i}{\text{Set point value } i} \right) \quad (1)$$

A three layer MLP was implemented using MATLAB software (Beale, Hagan, & Demuth, 2007). The input layer was made up of twelve neurons, corresponding to the predictor variables (see Table 2 and Fig. 1), while a single neuron was in the output layer, which represents the overall quality of service. Several architectures with H C [1; 30] neurons in the hidden layer were trained (see Fig. 1). Logarithmic sigmoidal functions were used in every layer, and 0.1 and 0.9 values were selected for the learning rate and the momentum factor parameters, respectively.

The data have been pretreated prior to entry into the ANN in order to accelerate the training time and improve the convergence (Martín del Bío & Sanz Molina, 2006; Masters, 1993). Therefore, the min–max formula (Delen et al., 2006) has been used for scaling the initial variable values into the [0;1] interval, in a way that this range coincides with the limits of the activation functions (see Eq. (2)).

$$v' = \frac{v - \min_A}{\max_A - \min_A} (\text{new_max}_A - \text{new_min}_A) + \text{new_min}_A \quad (2)$$

The algorithm can stagnate in a local minimum because several valleys can exist in the error surface and there is no guarantee that the learning algorithm will descend towards the global minimum when the MLP is trained only once. To cover a larger number of possibilities of finding the best-behavior ANN, and in line with Paliwal and Kumar (2011), every architecture (with the same number of neurons in the hidden layer) was independently trained M times with different small initial random weights each time, since the learning algorithm can tend to reach different local minimums. Finally, the ANN architecture with the best global MAPE was selected.

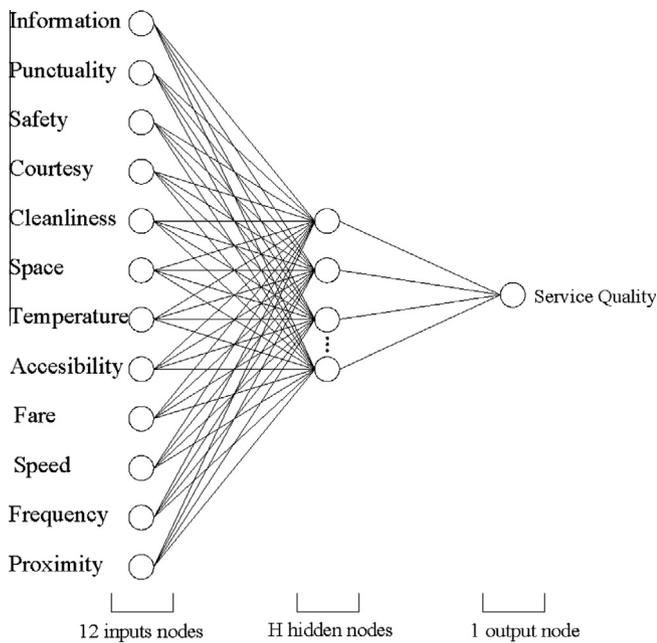


Fig. 1. Artificial Neural Network architecture.

3.2. Relative importance of the attributes

The following three methods will be employed in the analysis of the quality service attributes. These methods have been applied by several authors in different research fields (Gevrey, Dimopoulos, & Lek, 2003; Olden & Jackson, 2002).

3.2.1. Connection Weights

This method uses the synaptic weights of the trained ANN for quantifying input attributes contribution. For every input neuron, it calculates the sum of the product of the connection input neuron – hidden neuron with the connection hidden neuron – output neuron. The higher the sums for a given input neuron, the higher the relative importance of the corresponding attribute. For more information about the method see Olden and Jackson (2002).

3.2.2. Profile

This method divides the range of values of every variable $[0, 1]$ into a number of J intervals (e.g. 10 in this study, in a way that 11 values are obtained). All variables except one are fixed successively at their first quartile, median, third quartile and maximum, while the remaining variable successively adopts each of the $J + 1$ values of its interval for the five fixed values, so that the ANN gives five different values, of which their median is selected. Finally, the graphic representation of the median corresponding to every value of the interval is a curve of variation that indicates how the output variable (overall service quality) is affected by the changes in the input variables values. Thus, the larger variation in the ordinate axis values, the larger relative importance of the attribute. For more information about the method see Lek, Beland, Dimopoulos, Lauga, and Moreau (1995).

3.2.3. Perturb

Small amounts of noise are applied to each input neuron, until the 50% of the original value is Perturbed, while the remaining input neurons keep unaltered. The change in the Mean Square Error (MSE) assesses the relative importance of each attribute, therefore the larger the MSE for each input Perturbation, the more

the relative importance of the corresponding attribute. For more information about the method see Gevrey et al. (2003).

3.3. Statistical analysis

The statistical analysis of the results was performed by means of non-parametric tests, due to the data assumption of normality is not complied. Kruskal–Wallis test is used to explore the hypothesis that significant differences exist in the importance rates of the service quality attributes, and Dunn multiple comparison method is applied to identify which are the attributes responsible of these differences. In our analyses we performed three times the non-parametric tests, one for each relative contribution method applied (Connection Weights, Profile and Perturb), on a quantitative dependent variable (importance) and qualitative independent variable (service quality attributes). Thereby, homogeneous groups of attributes can be identified with statistical differences among each other.

4. Results

4.1. Neural Networks (NN)

Every NN architectures was independently trained $M = 50$ times, and a range of MAPÉs values were obtained for each of them (see Table 3). All the trained ANN reached a very high accuracy, above 90% in all the cases. This agrees with the results obtained by other authors who have applied ANN to evaluate the relative contribution of the service quality variables to other non-related with transportation fields, such as the education sector (Mahapatra & Khan, 2006), the service sector (Deng, Chen, & Pei, 2008; Deng & Pei, 2009; Larasati, De Yong, & Slevitch, 2012; Lin, 2007), or the level of job satisfaction perceived by the employees from several gas refinery enterprises (Azadeh et al., 2011). All of them also reached a very high accuracy in their outcomes. Moreover, this reinforces the idea maintained by Garver (2002) that ANN are an adequate technique for evaluating the relative importance of the customer satisfaction variables. In addition, their accuracy is significantly higher than that obtained by De Oña et al. (2012) when using decision trees for analyzing the relative contribution of the service quality attributes, which was between 59.72% and 62.16%.

Although there are not large differences in the accuracy of the trained architectures, those with 6 neurons in the hidden layer present lower values in the MAPE range, so this architecture was the selected to carry out the following phases of this research.

The attributes' importance obtained from each of the methods was scaled in the range $[0; 100]$. A value of 100 was assigned to the highest importance obtained from each of the methods, while the remaining variables' importance in the same method was scaled according to this relation. This transformation is necessary for adequately comparing and analyzing the results between methods, since the values of relative importance differ among the methods by several orders of magnitude. Fig. 2 shows the relative importance of the service quality attributes, expressed as a percentage, according to Connection Weights, Profile and Perturb methods.

The Connection Weights method commits the highest relative importance to Frequency (100.0%), followed by Speed (76.0%), Information (66.7%) and Proximity (55.5%). Safety (51.4%), Punctuality (51.4%) and Courtesy (47.8%) are also considered as of high importance, while Temperature (36.7%), Space (36.4%) and Fare (32.0%) are of middle importance, and Cleanliness (27.4%) and Accessibility (14.6%) are the least important attributes.

Table 3
MAPE's range for considered ANN architectures.

H	Min	Average	Standard deviation	Max
1	0.0400	0.0531	0.0082	0.0874
2	0.0401	0.0532	0.0090	0.0943
3	0.0438	0.0525	0.0045	0.0679
4	0.0372	0.0519	0.0067	0.0668
5	0.0424	0.0528	0.0062	0.0765
6	0.0325	0.0475	0.0064	0.0580
7	0.0417	0.0514	0.0052	0.0643
8	0.0437	0.0520	0.0044	0.0858
9	0.0436	0.0513	0.0036	0.0620
10	0.0368	0.0530	0.0074	0.0744
11	0.0399	0.0505	0.0051	0.0674
12	0.0384	0.0520	0.0086	0.0938
13	0.0396	0.0526	0.0106	0.0997
14	0.0427	0.0517	0.0043	0.0616
15	0.0354	0.0526	0.0065	0.0706
16	0.0365	0.0514	0.0087	0.0853
17	0.0379	0.0513	0.0056	0.0610
18	0.0416	0.0513	0.0070	0.0761
19	0.0406	0.0513	0.0046	0.0600
20	0.0419	0.0516	0.0049	0.0635
21	0.0395	0.0508	0.0056	0.0611
22	0.0346	0.0524	0.0078	0.0863
23	0.0424	0.0518	0.0700	0.0742
24	0.0393	0.0497	0.0047	0.0591
25	0.0385	0.0508	0.0067	0.0715
26	0.0405	0.0524	0.0063	0.0708
27	0.0417	0.0520	0.0078	0.0923
28	0.0375	0.0507	0.0056	0.0633
29	0.0427	0.0534	0.0075	0.0789
30	0.0413	0.0517	0.0091	0.0890

H: number of hidden neurons.

As with the Connection Weights method, the Profile method also assigns primary relative importance to Frequency (100.0%), and a quite high relative importance to Speed (77.7%), Information (64.2%) and Proximity (60.2%). The relative importance is also high to the Punctuality (54.4%); Safety (53.3%) and Courtesy (48.6%) attributes, and medium to Temperature (38.4%), Fare (36.4%) and Space (27.2%). Accessibility (17.3%) and Cleanliness (3.4%) are in the last positions in the ranking of importance.

In the Perturb method, Frequency (100.0%) reaches the highest relative importance, although Speed (63.6%) and Information (42.9%) are also important attributes. Unlike the results of the previous methods, Punctuality (33.0%), Safety (33.0%), Courtesy (30.3%), Proximity (23.3%) and Temperature (22.3%) are considered as of medium influence; Fare (17.9%) and Space (14.5%) as of low importance; and finally Accessibility (7.6%) and Cleanliness (7.1%) are the least important attributes.

By observing the outcomes of the application of the Connection Weights, Perturb and Profile methods, it can be said that all of them present similarities in percentage distribution terms and in the assigned position in the relative importance ranking (see Fig. 3), especially between the Connection Weights and the Profile methods. The Perturb method shows more discrepancies, mainly in the Proximity attribute, placing it in the seventh position in the relative importance ranking, while the other two methods raise it to the fourth position. All the methods agree that Frequency is the globally most important attribute for evaluating the service quality in the bus public transportation; that Speed, Information and Punctuality are of great importance; and that Proximity, Safety and Punctuality are also relevant attributes. On the other hand, Accessibility and Cleanliness are the less influential attributes.

These results are consistent with those extracted from other previous studies (de Oña et al., 2012; DelÓlio, Ibeas, & Cecin, 2010; Eboli & Mazzulla, 2010, 2011), in which different techniques (such as regression models, decision trees or relative importance

indexes) have been applied for analyzing the service quality in public transportation.

4.2. Statistical analysis

A statistical analysis was conducted to confirm the statistical differences between the importance rates of some of the service quality attributes describing the Granada public transport service. This process was carried out using non-parametric techniques (Kruskal–Wallis and Dunn tests) because of the non-normality of the data. The analysis pointed out that there are significant statistical differences with a 95% confidence level between some of these attributes and across the different importance rates deduced with each method (Connection Weights, Perturb and Profile) (see Table 4).

For the Connection Weights method, 7 homogeneous groups were identified. The variables included in each group do not present statistically significant intra-group differences, but they present statistically significant inter-groups differences. For the Perturb algorithm, 8 groups were also extracted, with the attributes that compose these homogeneous groups coinciding in almost all of them. Finally, the Profile data only allows 7 groups to be identified. These groups are marked with letters, and their average rates of importance decrease with the succession of the alphabet.

By observing the outcomes of the analysis (see Table 4), Frequency is identified in the three analysis as one independent group (group a), that has statistical significant differences with the rest of the attributes. Frequency presents the highest influence on overall service quality, with the highest average rate in all cases. In the same direction, at the Connection Weights and Perturb data, Speed, Information and Proximity also achieve high importance on service quality, making up group b and c, with significant differences compared to the rest of the groups. Regarding the results for the Profile's data, group b is also composed of Speed, Information and Proximity. This indicates that, after the attribute Frequency of the service, this group of three attributes is the one that produce the highest impact on the passengers' overall satisfaction.

On the contrary, Accessibility is the variable that presents the least average rate of importance. For the Connections Weights data, group g is composed of this variable only and presents considerable differences with the other groups of variables. At the same time, this pattern is identified with the data of the Perturb (group h) and the Profile (group f) algorithms. Accessibility is characterized by having significant differences with the rest of groups and being one of the least important attributes (the second lowest average rate for the Perturb data and the Profile data).

For Connections Weights, group f is composed of Bus interior cleanliness, Space on board, Bus temperature and Fare. Significant statistical differences are deduced with respect to the rest of homogeneous groups, except those attributes that belong to group e. With regards to group g for Perturb data, the same variables are identified, except for Cleanliness of the service. For Profile, the group f is formed by Space, Temperature and Accessibility. There is no exact coincidence in the attributes identified in this group among the different methods. However, if we consider the attributes belonging to the last two groups at the same time, in all of them (attributes whose importance average rates are lower to the rest of the groups), the same attributes are identified (Cleanliness, Space, Temperature, Accessibility), with exception of Fare, which is only considered to belong to these two homogeneous groups for the Connection Weights and Perturb data.

The rest of the attributes (Punctuality, Safety, Courtesy and Fare) are of average importance. They do not produce the highest impact on the overall evaluation (such as the Frequency, Speed, Information and Proximity) nor the lowest impact (such as

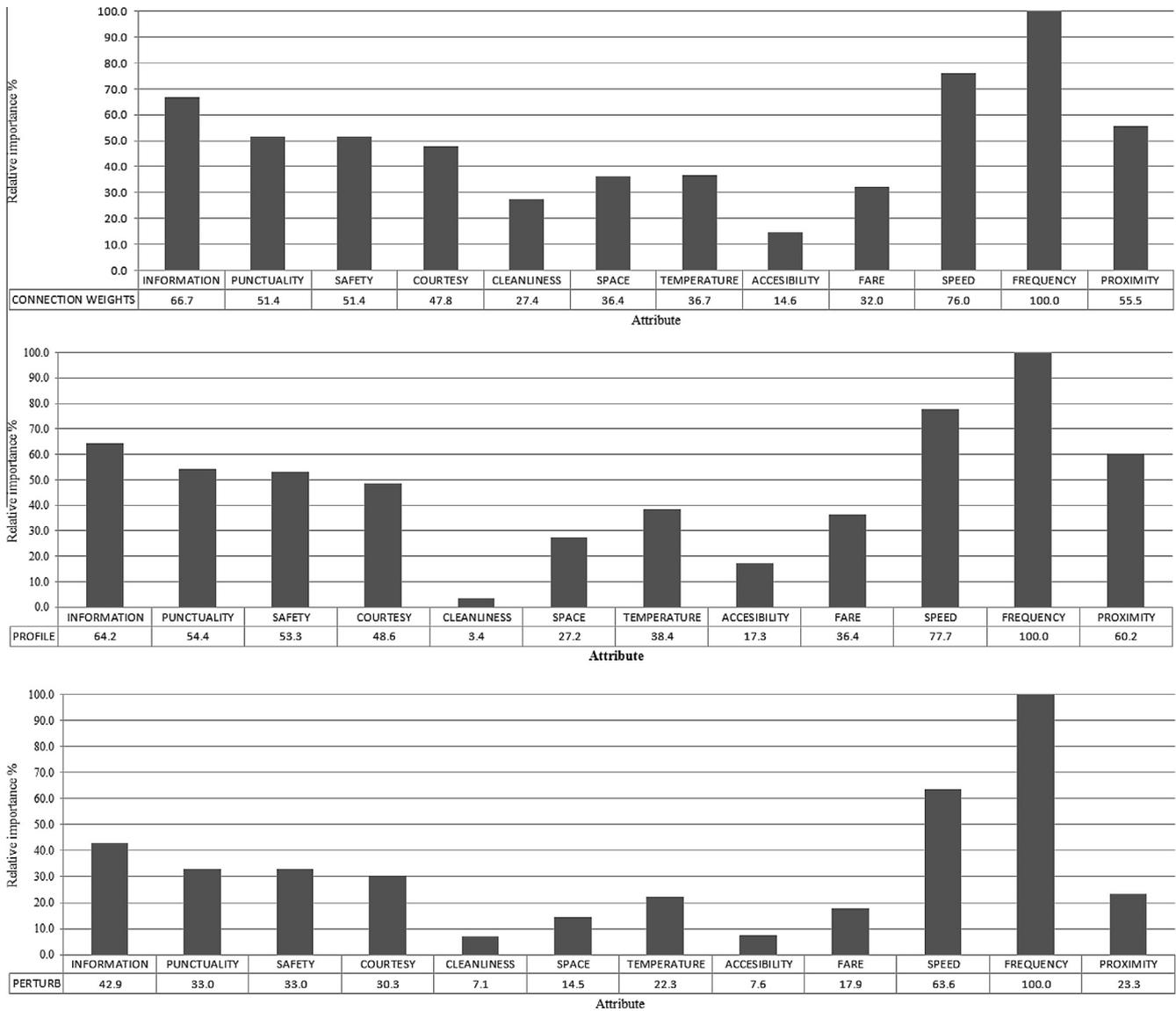


Fig. 2. Relative importance of each attribute for predicting service quality for each one of the methods (Connection Weights, Profile and Perturb).

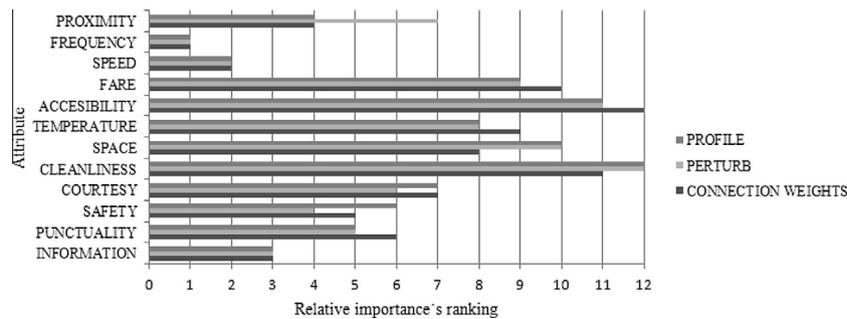


Fig. 3. Ranking of relative importance of each service quality attribute by methods (Connection Weights, Profile, Perturb).

Accessibility, Cleanliness, Space and Temperature). Eboli and Mazzulla (2008) also distinguished among different categories of attributes according to their impact on the overall service quality of a bus transit service in Cosenza (Italy). They defined basic and non-basic attributes based on the preferences showed by the users. The Frequency and Proximity of the service (named “walking distance” in their research) were identified as having the highest

impact and were defined as basic attributes. Our results agree with this point. Cleanliness and Courtesy were considered to be non-basic attributes (Eboli & Mazzulla, 2008), and these attributes were discovered to have a low and medium impact on service quality in the present research.

Thus, this statistical analysis can be used to identify homogeneous groups of attributes, according to their higher or lower

Table 4
Homogeneous groups with average values ($n = 50$) significantly different among other groups.

	Connection Weights	Perturb	Profile
Information	2.864 ^{b,c}	2.540 ^{b,c}	0.787 ^{b,c,d}
Punctuality	2.262 ^d	1.950 ^{d,e}	0.785 ^{c,d}
Safety on board	2.098 ^d	1.950 ^{d,e}	0.785 ^{c,d}
Driver courtesy	2.063 ^{d,e}	1.797 ^{e,f}	0.783 ^{d,e}
Bus interior cleanliness	1.255 ^f	0.421 ^h	0.768 ^g
Bus space	1.627 ^{e,f}	0.859 ^g	0.777 ^f
Bus temperature	1.590 ^{e,f}	1.319 ^{f,g}	0.779 ^{e,f}
Accessibility to/from the bus	0.660 ^g	0.449 ^h	0.776 ^f
Fare	1.359 ^f	1.060 ^{f,g}	0.785 ^{c,d}
Speed	3.284 ^b	3.768 ^b	0.790 ^b
Frequency of service	4.315 ^a	5.925 ^a	0.797 ^a
Proximity to/from origin/destination	2.357 ^{c,d}	2.163 ^{c,d}	0.787 ^{b,c}

^{a–h} Denotes differences statistically significant ($p < 0.05$). Two or more variables with the same letter in the same column denote homogeneous subgroup.

impact on passengers' perception of the service quality (importance rates).

5. Conclusions

The main objective of this paper was to validate the use of Artificial Neural Networks (ANN) for modeling the service quality of public transportation systems. More specifically, we used ANN to investigate the impact that several characteristics describing a transit service have on passengers' overall evaluation its quality. The metropolitan public bus service of Granada (Spain) was considered as a study case, using the data collected on a survey conducted by the Transport Consortium of Granada in 2007.

The ANN are proposed in this research because of its numerous advantages over more traditional parametric models (such as regression models, structural equation models or logit/probit models), but also over other non-parametric models, such as decision trees. ANNs provide higher fits of the phenomenon under study.

Three different methods were applied to determine the relative contribution of the service quality attributes, and both percentages and ranking positions show that Frequency, Speed, Information and Proximity are the users' most important attributes for evaluating the service quality perceived; although Punctuality, Safety and Courtesy must also be taken into account as relevant attributes. To the contrary, Cleanliness and Accessibility are considered to be the least important attributes. The validity of this methodology is corroborated by the fact that the outcomes agree with those obtained by other authors who have used different techniques such as regression models, decision trees or importance indexes (de Oña et al., 2012; DelÍOlío et al., 2010; Eboli & Mazzulla, 2010).

This study rises a well-understanding about which groups of attributes produce a higher contribution to passengers' perceptions of service quality, and which ones play a less important role in their evaluation. Various previous works (e.g. Eboli & Mazzulla, 2008; Philip & Hazlett, 1997; TRB, 2004; UNE, 2003) have pointed to the existence of several categories of attributes that have a greater or lesser impact on SQ and satisfaction. Philip and Hazlett (1997) propose a model with a hierarchical structure, based on three classes of attributes: pivotal, core and peripheral attributes. The pivotal attributes exert the greatest influence on the satisfaction levels. Core attributes are the amalgamation of the people, processes and the service organizational structure through which consumers must interact and/or negotiate so that they can achieve/receive the pivotal attribute. And the peripheral attributes can be defined as the "incidental extras" designed to add "roundness" to the service encounter and make the whole

experience for the consumer a complete delight. The UNE (2003) standard classifies the service's characteristics into basic, proportional and attractive, depending on how compliance and non-compliance affects customer satisfaction. The Transit Capacity and Quality of Service Manual (Transportation Research Board, 2004) groups attributes into availability factors (more important to passengers), and comfort and convenience factors (less important). Eboli and Mazzulla (2008) demonstrated the existence of two categories of attributes (basic and not basic) empirically from the preferences showed by users. Basic attributes compromise SQ when their level is low, and non-basic attributes are considered secondary service characteristics that affect SQ if they are present, but do not compromise it if they are absent.

In this study several homogeneous groups of attributes were identified, with significant statistical differences with respect to other groups regarding their influence on overall service quality. These homogeneous groups were similar across the three statistical analyses: one for the data derived from the Connection Weights method, and the other two from the Perturb and Profile methods. Seven homogeneous groups were identified with the Connection Weights and Profile data, and eight groups from the Perturb data. Cleanliness, Space, Temperature and Accessibility correspond to the attributes with the lowest impact on passengers' overall evaluation, while Frequency, Speed, Information and Proximity correspond to the categories with the highest influence. This is similar to research carried out in other bus public services (Eboli & Mazzulla, 2008) where Frequency and Proximity were also determined as basic attributes, having the highest impact on service quality, while Cleanliness and Courtesy were considered to be non-basic attributes.

These results provide powerful information for future policy-making due to the fact that once the transport planners know the variables that users value the most; they can define more efficient strategies for their investments. In the future, it would be also interesting to extend the interviews to non-users of the public transport service, in order to discover, not only the opinions of the current passengers about the level of quality provided, but also how non-users perceive service quality. This insight could be used to attract new users to the public transport system.

The main advantages of this approach can be summarized in the following points:

- This methodology allows to mitigate the inherent instability of ANN models, which is an important improvement in the field of "black-boxes" techniques, to which ANN belong, since until now there is no consensus about what method of relative importance must be used for determining the variables' relative importance (de Oña & Garrido, in press). This approach has demonstrated to solve this problem when using the Profile, Perturb and Connection Weight methods as methods for identifying the variables' contribution.
- The high capability of prediction and accuracy that characterize ANNs are met in the used database. And this accuracy is much higher than those obtained by some other methods as decision trees, regression models or importance indexes (de Oña et al., 2012; DelÍOlío et al., 2010; Eboli & Mazzulla, 2008, 2010) in the field of service quality in public transportation systems.
- The ranking of importance obtained for each variable is similar to those achieved by other techniques of widespread use in the field of service quality and transportation.

The main disadvantage of this approach is the duration of the calculus, since a large number of ANN must be trained and tested, and each method for determining the variables' relative importance (Profile, Perturb and Connection Weight) must be calculated as many times as ANN are included in the selected architecture.

Since it is not easy to select the optimal ANN set, due to the intrinsic difficulty of finding the optimal or sub-optimal ANN, as future research directions, it would be interesting to know how the selection of ANN architecture affects the ranking of relative importance of the variables. In this sense, it would be also interesting to analyze how the outcomes of the approach vary when a different database is used, or what happens when the accuracy obtained from the ANN is much lower than the obtained with the current database. Therefore, one could wonder whether the approach keeps the same effectiveness.

Acknowledgments

Support from Consejería de Innovación, Ciencia y Economía of the Junta de Andalucía (Spain) (Research Project P08-TEP-03819, co-funded by FEDER) is gratefully acknowledged. The authors also acknowledge the Granada Consorcio de Transportes for making the data set available for this study.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.eswa.2014.04.045>.

References

- Azadeh, A., Roubahman, M., Saberi, M., & Mohammad Fam, I. (2011). An adaptive neural network algorithm for assessment and improvement of job satisfaction with respect to HSE and ergonomics program: The case of a gas refinery. *Journal of Loss Prevention in the Process Industries*, 24, 361–370.
- Beale, M. H., Hagan, M. T., & Demuth, H. B. (2007). Neural network toolbox 7. User's Guide. MathWorks, Inc. 3 Apple Hill Drive Natick, MA 01760-2098.
- Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford University Press.
- Bordagaray, M., Dell'Olivo, L., Ibeas, A., & Cecin, P. (in press). Modelling user perception of bus transit quality considering user and service heterogeneity. *Transportmetrica A: Transport Science*, <http://dx.doi.org/10.1080/23249935.2013.823579>.
- Carman, J. (1990). Consumer perceptions of service quality: An assessment of the SERVQUAL dimensions. *Journal of Retailing*, 66, 33–55.
- Das, S., & Pandit, D. (2013). Importance of user perception in evaluating level of service for bus transit for a developing country like India: A review. *Transport Reviews: A Transnational Transdisciplinary Journal*, 33(4), 402–420.
- De Oña, J., & De Oña, R. (in press). Quality of service in public transport based on customer satisfaction surveys: A review and assessment of methodological approaches. *Transportation Science*.
- De Oña, J., & Garrido, C. (in press). Extracting the contribution of independent variables in neural network models: A new approach to handle instability. *Neural Computing and Applications*. <http://dx.doi.org/10.1007/s00521-014-1573-5>.
- De Oña, J., de Oña, R., & Calvo, F. J. (2012). A classification tree approach to identify key factors of transit service quality. *Expert Systems with Applications*, 39, 11164–11171.
- De Oña, J., de Oña, R., Eboli, L., & Mazzulla, G. (2013). Perceived service quality in bus transit service: A structural equation approach. *Transport Policy*, 29, 219–226.
- De Oña, J., de Oña, R., Eboli, L., & Mazzulla, G. (in press). Heterogeneity in perceptions of service quality among groups of railway passengers. *International Journal of Sustainable Transportation*. <http://dx.doi.org/10.1080/15568318.2013.849318>.
- Delen, D., Sharda, R., & Bessonov, M. (2006). Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. *Accident Analysis and Prevention*, 38, 434–444.
- Dell'Olivo, L., Ibeas, A., & Cecin, P. (2010). Modelling user perception of bus transit quality. *Transport Policy*, 17(6), 388–397.
- Deng, W.-J., Chen, W.-C., & Pei, W. (2008). Back-propagation neural network based importance–performance analysis for determining critical service attributes. *Expert Systems with Applications*, 34, 1115–1125.
- Deng, W.-J., & Pei, W. (2009). Fuzzy neural based importance–performance analysis for determining critical service attributes. *Expert Systems with Applications*, 36, 3774–3784.
- Eboli, L., & Mazzulla, G. (2007). Service quality attributes affecting customer satisfaction for bus transit. *Journal of Public Transportation*, 10(3), 21–34.
- Eboli, L., & Mazzulla, G. (2008). A stated preference experiment for measuring service quality in public transport. *Transportation Planning and Technology*, 31(5), 509–523.
- Eboli, L., & Mazzulla, G. (2010). How to capture the passengers' point of view on a transit service through rating and choice options. *Transport Reviews*, 30, 435–450.
- Eboli, L., & Mazzulla, G. (2011). A methodology for evaluating transit service quality based on subjective and objective measures from the passenger's point of view. *Transport Policy*, 18, 172–181.
- Eboli, L., & Mazzulla, G. (2012). Structural equation modelling for analysing passengers' perceptions about railway services. *Procedia-Social and Behavioural Science*, 54, 96–106.
- Funahashi, K. I. (1989). On the approximate realization of continuous mappings by neural networks. *Neural Networks*, 2, 183–192.
- Garver, M. S. (2002). Using data mining for customer satisfaction research. *Marketing Research*, 14(1), 8–12.
- Garver, M. S. (2003). Best practices in identifying customer-driven improvement opportunities. *Industrial Marketing Management*, 32(6), 455–466.
- Gedeon, T. D., Wong, P. M., & Harris, D. (1995). Balancing the bias and variance, network topology and pattern set reduction techniques. In *Proceedings of international workshop on artificial neural, IWANN95* (pp. 550–558). Torremolinos, España.
- Gevrey, M., Dimopoulos, I., & Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling*, 160, 249–264.
- Hagan, M. T., Demuth, H. B., & Beale, M. H. (1996). *Neural network design*. 0-9717321-0-8. Campus Publishing Service, Colorado University Bookstore.
- Haykin, S. (1999). *Neural networks. A comprehensive Foundation* (2nd ed.). Pentence-Hall.
- Hensher, D. A. (2014). The relationship between bus contract costs, user perceived service quality and performance assessment. *International Journal of Sustainable Transportation*, 8(1), 5–27.
- Hensher, D. A., & Prioni, P. (2002). A service quality index for an area-wide contract performance assessment regime. *Journal of Transport Economics and Policy*, 36(1), 93–113.
- Hensher, D. A., Stopper, P., & Bullock, P. (2003). Service quality—developing a service quality index in the provision of commercial bus contracts. *Transportation Research Part A*, 37, 499–517.
- Irfan Syed Muhammad, Mui Hung Kee Daisy, & Shahbaz Saman (2011). Service quality in rail transport of Pakistan: A passenger perspective. In *Proceedings of 3rd SAICON: International conference on management, business ethics and economics (ICMBEE), December 28–29, 2011*. Lahore, Pakistan.
- Jakobsson Bergstad, C., Gamble, A., Hagman, O., Polk, M., & Garling, T. (2011). Affective–symbolic and instrumental–independence psychological motives mediating effects of socio-demographic variables on daily car use. *Journal of Transport Geography*, 19(1), 33–38.
- Kasabov, N. K. (1996). *Foundations of neural networks, fuzzy systems and knowledge engineering*. MIT Press.
- Kikuchi, S. (2012). Thoughts on the future of artificial intelligence and transportation. In *Transportation Research Board (Ed.), Transportation research circular E-C168: Artificial intelligence applications to critical transportation issues* (pp. 139–140). Washington DC.
- Larasati, A., De Yong, C., & Slevitch, L. (2012). The application of neural network and logistics regression models on predicting customer satisfaction in a student-operated restaurant. *Procedia Social and Behavioral Sciences*, 65, 94–99.
- Lee, Ch., Ran, B., Yang, F., & Loh, W. Y. (2010). A Hybrid Tree Approach to Modeling Alternate Route Choice Behavior With Online Information. *Journal of Intelligent Transportation Systems*, 14(4), 209–219.
- Lek, S., Beland, A., Dimopoulos, I., Lauga, J., & Moreau, J. (1995). Improved estimation, using neural networks, of the food consumption of fish populations. *Marine and Freshwater Research*, 46, 1229–1236.
- Lin, W.-B. (2007). An empirical of service quality model from the viewpoint of management. *Expert Systems with Applications*, 32, 364–375.
- Mahapatra, S. S., & Khan, M. S. (2006). A methodology for evaluation of service quality using neural networks. In *Proceedings of the international conference on global manufacturing and innovation, July 27–29*.
- Marcucci, E., & Gatta, V. (2007). Quality and public transport service contracts. *European Transport*, 36, 92–106.
- Martin del Bío, B., & Sanz Molina, A. (2006). *Neural networks and fuzzy systems*. Editorial RA-MA.
- Masters, T. (1993). *Practical neural networks recipes in C++*. Academic Press.
- Mouwen, A., & Rietveld, P. (2013). Does competitive tendering improve customer satisfaction with public transport? A case study for the Netherlands. *Transportation Research Part A*, 51, 29–45.
- Ngatia, G. J., Okamura, T., & Nakamura, F. (2010). The structure of users' satisfaction on urban public transport service in developing country: The case of Nairobi. *Journal of the Eastern Asia Society for Transportation Studies*, 8, 1288–1300.
- Olden, J. D., & Jackson, D. A. (2002). Illuminating the “black-box”: A randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*, 154, 135–150.
- Paliwal, M., & Kumar, U. A. (2011). Assessing the contribution of variables in feed forward neural network. *Applied Soft Computing*, 3690–3696.
- Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, 49, 41–50.
- Philipp, G., & Hazlett, S. A. (1997). The measurement of service quality: A new P-C-P attributes model. *International Journal of Quality & Reliability Management*, 14(3), 260–286.

- Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. In *Foundations*. Cambridge, MA: MIT Press.
- Simoes, A. (2013). Mobility-impaired travelers and public transport: A framework to assess quality of service. *Theoretical Issues in Ergonomics Science*, 14(3), 247–257.
- Transportation Research Board (2004). *Transit Capacity and Quality of Service Manual*, 2nd ed.
- Tripp, C., & Drea, J. T. (2002). Selecting and promoting service encounter elements in passenger rail transport. *The Journal of Services Marketing*, 16(5), 432–442.
- UNE-EN 13186 (2003). *Transportation. Logistics and services. public passenger transport. Service quality definition, targeting and measurement*. AENOR.
- Watts, M. J., & Worner, S. P. (2008). Using artificial neural networks to determine the relative contribution of abiotic factors influencing the establishment of insect pest species. *Ecological Informatics*, 3, 64–74.
- Weinstein, A. (2000). Customer satisfaction among transit riders. How customer rank the relative importance of various service attributes. *Transportation Research Record*, 1735, 123–132.
- Werbos, P. I. (1974). *Beyond regression: New tools for prediction and analysis in the behavior sciences* (Ph.D. thesis). Cambridge, MA: Harvard University.
- Xie, C., Lu, J. Y., & Parkany, E. (2003). Work travel mode choice modeling with data mining: decision trees and neural networks. *Transportation Research Record*, 1854, 50–61.