Article

Decision Model for Predicting Social Vulnerability Using Artificial Intelligence

**Francisco Javier Abarca-Alvarez [1,2,]* [iD], Rafael Reinoso-Bellido [1,2] and
Francisco Sergio Campos-Sánchez [1,2] [iD]**

[1] Department of Urban and Spatial Planning, University of Granada, 18071 Granada, Spain;
rafaelreinoso@ugr.es (R.R.-B.); scampos@ugr.es (F.S.C.-S.)
[2] Higher Technical School of Architecture, University of Granada, 18071 Granada, Spain
* Correspondence: fcoabarca@ugr.es

check for
updates

Abstract: Social vulnerability, from a socio-environmental point of view, focuses on the identification of disadvantaged or vulnerable groups and the conditions and dynamics of the environments in which they live. To understand this issue, it is important to identify the factors that explain the difficulty of facing situations with a social disadvantage. Due to its complexity and multidimensionality, it is not always easy to point out the social groups and urban areas affected. This research aimed to assess the connection between certain dimensions of social vulnerability and its urban and dwelling context as a fundamental framework in which it occurs using a decision model useful for the planning of social and urban actions. For this purpose, a holistic approximation was carried out on the census and demographic data commonly used in this type of study, proposing the construction of (i) a knowledge model based on Artificial Neural Networks (Self-Organizing Map), with which a demographic profile is identified and characterized whose indicators point to a presence of social vulnerability, and (ii) a predictive model of such a profile based on rules from dwelling variables constructed by conditional inference trees. These models, in combination with Geographic Information Systems, make a decision model feasible for the prediction of social vulnerability based on housing information.

Keywords: social vulnerability; predictive models; urban model; dwelling; decision model; artificial neural network; self-organizing maps; decision trees

1. Introduction

Vulnerability is often defined as the potential for physical or economic loss or damage [1] located in a specific territory. When vulnerability is approached from a social point of view, it focuses sharply on its human aspect of population application [2]. In recent years, a multitude of lines of work have emerged around the concept of social vulnerability, some more linked to natural risks and disasters [3,4], some to environmental factors [5], and others, closer to the concept of poverty [4]. From an approach that initially considered natural events as the main focus, there has been a gradual shift to one that considered that the effects on the population were conditioned by its own mitigation capacity [6]. Mitigation, in this sense, is considered to be the ability of an individual or community to anticipate, resist, and overcome the impact of unforeseen events [7]. Thus, the approach to the concept of social vulnerability has been opened up, placing people at the center. This approach concerns the people who have or do not have the capacity to overcome [8] or adapt to vicissitudes, which are not exclusively linked to environmental risks, and even incorporates a spatial aspect [9].

Social vulnerability presents various challenges, such as multidimensionality [4,5,10–12], or the fact that many of the variables or dimensions to be evaluated are not generally directly observable [10]. Among the studies that have tried to identify indicators of social vulnerability, one by Cutter et al. in

2003 stands out in which they incorporated, as variables of the so-called Social Vulnerability Index (SoVI), a whole series of indicators including socio-economic factors, age, commercial or industrial development indicators, unemployment, rurality indicators, residential property, level of infrastructure, level of income, occupation, access to medical services, gender factors, race and ethnicity, family structure, educational level, vegetative growth, dependence on social services, and the presence of a population with special needs [5]. There are studies that integrated similar techniques in data interpretation methodologies [13,14], and many other studies [14–19] that integrated or compiled indicators with the same objective.

With the challenge of developing a decision model connected to the prediction of social vulnerability, together with the concept of a decision model, the decision support system (DSS) is adopted because of its capacity, beyond the use of information technologies (ITs) [20], to amplify the capacities of decision-makers [21]. The DSS concept was introduced by George Anthony Gorry and Michael S. Scott Morton [22]. Linked in our case to social vulnerability, it is proposed as a tool in which a Geographic Information System (GIS) must be connected in a fully integrated manner.

The integration of massive data, in what some call the new quantitative geography based on GIS tools, is elevating the granularity of geographic data to the extreme, in an authentic "n-dimensionality" of the data [23]. In most cases, according to Pragya Agarwal and André Skupin, GISs have focused on traditional statistical analysis to solve spatial autocorrelation problems, leaving many other areas totally unexplored. Some of these spaces are being addressed by emerging approaches such as artificial intelligence (AI) or artificial neural networks (ANN), machine learning (ML), or specifically geo-computing. ANNs are usually included as a category of ML methods frequently used for prediction, classification, and pattern recognition [24], with multiple practical applications, e.g., monitoring and control of industrial or medical instrumentation, in telecommunication networks, etc. [25] These new techniques and approaches are propitiating a change of paradigm in DSSs, considering that at present they can be useful for the understanding of reality, detection of its problems, and in short, for the formulation of new hypotheses and not only as an instrument to verify those previously established.

Within this conceptual framework, the main aim of the research is the construction of a predictive model that allows the identification of territories with high social vulnerability from a limited set of variables that are easy to access for the decision-maker. Specifically, the creation of a model-based exclusively on residential information as the basis and fundamental support of social reality is proposed. The main contribution to the field of social vulnerability consists of evidencing the viability of such a model of social vulnerability, constructed from residential indicators that are simple to obtain. The residential model can be obtained by means of an ocular inspection in situ, in contrast to the information necessary to evaluate social vulnerability, which is notably more complex and costly to obtain. To this end, techniques for interpreting reality using artificial intelligence and machine learning supported by a geographic information system will be used. Specifically, as a case study of the proposed methodology, the social vulnerability of the population of Andalusia (Spain) is characterized on the basis of residential information in which the population resides, validating the model by evaluating its predictive performance compared with research on social vulnerability in the region.

2. Literature Review

2.1. Social Vulnerability

Social vulnerability is a complex concept, which requires an approach that includes multiple dimensions and factors. With the intention of synthesizing some of the main contributions with respect to it, a review of the literature is carried out, organizing its approaches and indicators of measurement or evaluation, highlighting among all of them the indicators of social vulnerability (SoVI) [5]. In order to carry out a systematic approach, the main references are organized around the classification recently proposed by Lee [15], all of which are summarized in Table 1.

Table 1. Integrated factors of social vulnerability and resources. Classification based on [5,15]. Source: Compiled by the authors based on cited references.

Types of Capital [1]	Description of Factors [1]	Concept [2]	Resources of References
Human capital	Demographic characteristics	Age	[15,16,18,26–33]
		Gender	[1,15,16,27,30,32,34–41]
		Race and ethnicity	[16,18,39,41,42]
		Occupation	[15,27,31,43,44]
		Population growth and mortality	[18,28,31,32,43,45,46]
	Social and economics characteristics	Socioeconomic status (income, political power, prestige)	[15–18,27,28,33,34,39,41,45–48]
		Employment	[15,17,31,33,40,41,44,49,50]
		Education	[16,17,31–33,41,43,44,48,50]
		Social dependence	[16,17,30,32,40,41,43–45,51,52]
		Special needs populations	[16,41,44,45,53]
Social capital	Community development	Commercial and industrial development	[16,43,44]
		Rural/urban	[16,28,54,55]
		Residential property	[15,18,28,31,42–44,50]
		Renters	[15,18,31,33,41,43,45,50]
		Family and social structure	[16,18,34,43,45,46]
Public resource provision and public security	Public infrastructure and resources that belong to inhabitants and its safety	Infrastructure and lifelines	[16,18,32,41,43,50,56,57]
		Medical services	[17,18,27,31,32,43,45,57]

[1] Categories based on [15]. [2] Categories based on [5,15].

It is important to highlight the existence of a reference work on social vulnerability in Andalusia, which is the area where the methodological proposal is assessed. This research identifies deprived urban areas [58] throughout the Andalusian region. Deprived urban areas are understood as those areas that present a series of weaknesses in their socio-demographic structure and/or in the environmental qualities of their physical space. The authors explain that these are neighborhoods with a structural social and economic weakness in which any threat, external risk or even social intervention without prior analysis can turn them into a vulnerable area.

2.2. Decision Model and Decision Support System

The DSS assists and guides decision making [59], allowing the amplification of the decision maker's abilities to interpret information and knowledge [21], reducing improvisation and indeterminacy [60]. DSSs have been used in multiple fields such as business intelligence [61], health [62], and fleet management [63], and with the help of a GIS, for the management of means of transport [64]. On the other hand, GISs have been used assiduously by governments, researchers, and companies as a decision-making tool in which the spatial dimension reaches a certain repercussion and influence [65].

GIS emerged at the end of the 1960s, with a particularly important development occurring in the 1980s [66], and reached the generalization of its use from the 1990s, coinciding with the arrival of GPS technologies to the civilian population in 1993 [67]. Today, GIS has been fully integrated into social media [68], in a space-time integration [69]. They have been established in society in what is called "spatial thinking" [70], making it easier for citizens such as urban planners to become important actors in planning.

It must be borne in mind, that GISs have not always been prepared to act as a DSS, as they require the integration of complex realities and problems for decision support [71], consolidating as flexible and resilient systems. Likewise, DSSs have recently moved from a focus on technology and systems to one focused on decision-makers [20], with the intention of helping them to process knowledge [21] and facilitate decision-making based on technology [59] in an accessible and affordable way.

Moreover, since its origin, the informational reality on which GISs are based has also changed rapidly, increasing the presence of spatial data and information of free access, configuring itself practically as a discipline in itself, which some call GIScience [72–75], a term introduced by Goodchild in 1992 [76] based on the idea of a new quantitative geography, fundamentally spatial, tending towards planning and management. With time, GISs have evolved from an emphasis on the "S" for the computational problems (1960s–1970s), to the "I" for the interest in the information (1980s–1990s), to, from 2000, focusing on the "G", due to a need for geographical interpretation, materializing in the "society of the geographic information" and opening a new stage for the history of geography [77]. It is at this point that the GIS approaches the DSS concept.

In order to achieve the aim of the research, a decision framework or decision model is created. This decision framework is conceptually fed by the idea of DSS, flowing between a more focused approach to DSS towards the IT techniques that support the decision [78], and a more focused vision on broadening the capacity of decision-makers [20,21]. In the first approximation, a DSS could be described as "a set of interactive and expandable IT techniques and tools for data processing and analysis, that supports managers in decision making" [78]. In the second approximation, it can be described in the words of Power, Sharde, and Burstein [20]: "More broadly, DSS is not exclusively based on the use of the computing technologies, instead it is focused on the "ability to relax cognitive, temporal, and economic limits of decision-makers—amplifying decision-makers' capacities for processing knowledge which is the lifeblood of decision making "[21]".

This research is framed in two of the different types of DSS described by Power et al. [20]: (1) knowledge-focused and (2) model-oriented [20]. This research is framed within the DSS paradigms of knowledge-focused and model-oriented [20]. The first type focuses on the construction of a knowledge discovery system based on institutional databases on the demographic and social qualities of Andalusia. It allows the identification of socially vulnerable areas. The second DSS type focuses on the creation

and management of a quantitative model of social reality. It is aimed at providing decision support and drawing it up from the dwelling properties of the territories under study.

Initially, DSSs were not thought of as an autonomous discipline, but rather as a method for bringing intelligence to decisions at the productive and environmental level [79], it has undergone an important development and evolution in recent years in parallel with the emergence of data science. DSSs use data and parameters provided by decision-makers to help them analyze a situation, although they do not have to be based on massive data [20]. In our case, the model was built with massive data (both demographic and residential) but can be used for decision making with very limited information—even scarce information.

The research, as it has progressed, consists of the construction of two models linked to tools from information technologies [80], as it is described in the following sections:

2.3. Models of Knowledge Discovery and Clustering through Non-Supervised Learning—Self-Organizing Maps (SOM)

The study used self-organizing maps (SOM) methodologies. They were initially proposed by Teuvo Kohonen [81,82]. The SOM methodology is a knowledge discovery or data mining technique consisting of an artificial neural network.

It is based on non-supervised learning, obtaining from the input data (input layer) the organization of them in a representation of the space of M neurons, which are arranged in a lattice of size $a \cdot b$, where $M = a \cdot b$. This lattice, which has the capacity of evidencing the topological relations and similarity between the subjects under study, locates those instances that present properties or attributes with greater similarity closer to each other. By means of an iterative process, the topological distance between the neurons is evaluated. Each neuron presents a prototype representing a cluster of input samples. At each time step, a new sample is presented to the network, and a winner neuron is declared, and the prototypes are adjusted. The process is stopped with a predicted number of iterations or when a decaying learning rate is reached.

SOM come from the field of knowledge of artificial intelligence, having shown itself to be very effective and robust in numerous disciplines. SOM show diverse capacities, among which we can highlight two, initially: (i) it is capable of showing and visualizing the starting information in a clear and ordered way, (ii) it allows the clustering and, therefore, labelling of study subjects in classes that do not require their definition, characterization or previous nominative labeling (non-supervised learning). Compared to other pattern discovery methodologies, such as cluster analysis, the SOM methodology has the advantage of (i) allowing a large set of statistical data to be visualized [83], (ii) showing the topological relationships of similarity or difference among the items under study, (iii) being graphically interpretable, and (iv) constituting by itself a knowledge system of a DSS for the analysis and visualization of statistical indicators [83].

By means of these techniques, on the one hand, the labeling is obtained as classes or profiles of the different fragments of the Andalusian territory studied, paying attention to the multi-variable analysis of the demographic and social attributes of the study. Based on the SOM methodology, analysis and interpretation of the profiles obtained are carried out, which is materialized in thematic cartographies of the different attributes included in the neural network and in different tables and statistical data that allow the differentiating characteristics of each profile to be known. To facilitate its use as part of a DSS, such classes are represented, and in particular, the social vulnerability profile through GIS.

The SOM methodology has been widely used in numerous fields. In the field of image interpretation, we can highlight its use for the analysis and classification of multiple satellite images of 200 different bands [84] and the classification of soils and minerals using spectral radio images and GIS [85]. It has also been used in transport for the graphical analysis of spatial interactions and obtaining patterns in US air transport structures [86], or for the classification of the sustainability of transport in cities according to TOD (Transit-Oriented Development) criteria [87]. The use of SOM to classify and recognize patterns of epidemiology with the help of GIS, or in distribution of ecological

risk by contamination is described in [88]. It has been frequently used to classify and locate, for example, patterns of pesticide contamination in the Asour-Garonne river basin in France [88], to create models of plant location for the treatment of wood residues [89] or regarding the environmental quality of soils [90]. The SOM and GIS have also been used to classify community health based on environmental conditions variables [91] and to show quality of life trends in the neighborhoods of Charlotte (USA) [92].

In areas of knowledge with a more social and demographic aspect, certain studies with the use of SOM focused on the representation of data stand out, such as, for example, the joint visualization with GIS of the demographic changes of the counties of Texas (USA) over time by means of SOM [93], SOM visualization of spatial–temporal patterns of geographical variables in the USA [94], or the use of the SOM for the realization of an alternative and complementary holistic representation to the spatial representation of the GIS in which information of 69 census attributes in the USA is simultaneous, with information on climate, topography, soil, geology, land use, and population [95]. SOM have been used as a classifier to determine homogeneous demographic regions from data from the Athens census [96], to classify European adaptation strategies [97], for characterizing neighborhoods by tagging New York census sections from 79 geo-demographic attributes [98], and for non-supervised classification of geospatial data from German communities in terms of population, migration, taxes, residence, employment, and transportation [99]. SOM have also been used to make a semantic representation and characterization of exemplary neighborhoods from the recent history of urbanism [100], to identify and characterize the urban sprawl of Milan (Italy) [101], and for the analysis of the residential market from variables of prices, qualities, and characteristics of housing, density, inhabitants, etc., of Finland, Hungary, and The Netherlands [24].

As indicated, the state of the art shows that SOM is very often used as a methodology for reduction and classification [102] and also for entity labeling [103]. Compared to other dimensional reduction methods such as PCA (principal component analysis) or MDS (multidimensional scaling), the ability of SOM to preserve the topology of the data results in more efficient use of the available space in the map representation, with the consequence of greater distortion in relative distances [104]. On the other hand, the SOM has notable advantages over other techniques or methods. SOM is relatively insensitive to missing values while tolerating data with a non-normal distribution, which allows you to dispense with checks that are difficult to comply with, making it valid for any data distribution. On the other hand, as a clustering method, SOM is more robust than, for example, the K-means, although it requires more computing time [90,105].

2.4. Construction of Predictive Models through Supervised Learning—Decision Trees

By means of a machine learning process, a series of rules was obtained that allows for the prediction of one of the profiles that were determined with the SOM model, using only attributes on the dwelling reality of the territories under study. These dwelling variables were not taken into consideration in the evaluation of the SOM neural network, nor could they consequently affect or correlate with the definition of the profiles obtained by the SOM neural network. An approach to the problem of learning is proposed through the "divide and conquer" paradigm, which, when carried out on a set of independent instances, naturally leads to a style of representation called the decision tree [106]. In each node of the tree, a particular attribute intervenes, typically comparing each instance of the attribute with the value of a constant, and, usually, generating two branches attending to the instances that fulfill or do not fulfill such a rule. The decision trees suppose a simple and user-friendly representation to interpret and use in the prediction of the demographic and social reality of a territory and are consequently useful for the decision making about it. The model uses a limited portion of the available features and generally, it is easy and economical to obtain the residential reality of the place under study. Likewise, when evaluating the "value" of the profiles reached in Phase 1, in their spatial characterization by means of GIS, the usefulness of the proposed methodology is verified.

Decision trees are machine learning techniques that generate models that are very easy to understand and use. Decision trees are (i) models insofar as they construct a hypothesis or representation of the regularity of the data, (ii) understandable by symbolically expressing a set of conditions, and (iii) propositional by establishing "attribute-value" rules in their construction in which the conditions are expressed over the value of a single attribute [107].

There are many variants of decision tree algorithms. Here, we highlight just a few of them. As historical antecedents of the most used decision trees, we can find, for instance, the algorithms CHAID, CART, ID3, and C4.5. CHAID stands for automatic interaction detection by Chi-squared automatic interaction detection. It is an original by Kass [108] based on Bonferroni's significance test and characterized by its ease of graphic interpretation and for not being a parametric analysis. CHAID is a multivariate technique in which there is a single variable to explain and several explanations, in which different categories are identified to serve as a division in each branch, selecting for each of them the variable that discriminates most and the classes that, when combined, provide the greatest discrimination in the dependent variable under analysis, i.e., detects the interactions that most discriminate. CART stands for classification and regression trees, proposed by Breiman, Friedman, Olshen and Stone [109]. ID3 later gave rise to algorithm C4.5, both of which are very popular for their non-parametric approach and their interpretability [110].

Among the algorithms that generate decision trees, we highlight the recursive partitioning methods that have become very popular and widely used in recent years for nonparametric regression and for classification in many scientific fields [111].

The method specifically chosen is that of conditional inference trees, which is based on the permutation test [112], using nonparametric tests as criteria for branch division. It should be noted that the selection of this method is mainly due to its high comprehensibility and ease of interpretation of the rules obtained. There are other methods that are more precise, such as random forests. This method shows great precision in its predictions, often greater than other recent statistical learning techniques such as support vector machines (SVMs) or boosting [111], but it offers high illegibility and complexity of the rules that make it very difficult for the analyst to reproduce the model by hand.

2.5. Hybrid Model—SOM and Decision Trees

This section provides a few examples of research that present a methodological framework similar to the one used in this paper, i.e., the implementation of a hybrid model in which a decision tree is applied to the SOM, i.e., the decision tree uses the non-supervised clustering provided by the SOM as information to be predicted.

With this approach, we can find some methodological work [113,114] and others related to biology and medicine, such as a mining study on biological data [115]. Concerning engineering, there is a selection of variables to group road samples [116] and post-processing of accident scenarios [117]. Related to economics and business, there is a discovery of preferences in stock trading [118], focused on this approach of "SOM + decision trees". To conclude this section, we find some similarities in the approach with our research for the selection of properties in the analysis of census data by SOM and decision trees [80,119].

3. Materials and Methods

For an optimal understanding of the methodology (Figure 1) and to obtain the best results from the DSS, the following phases [120] were followed: (i) information and processing functions, (ii) data sets, (iii) models, and (iv) visual representations.
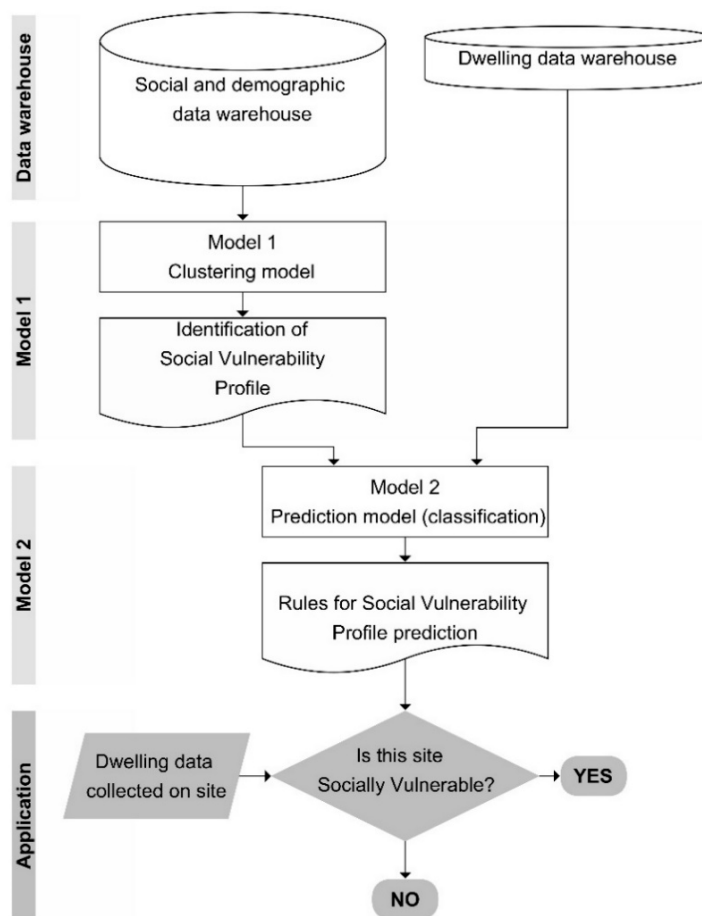
Figure 1. Simplified scheme of the methodology. Source: Compiled by the authors.

3.1. Materials. Processing Information, and Functions

The information used in this research came from the 2001 Population Census of Andalusia provided by the regional government of Andalusia through the "Instituto de Estadística y Cartografía de Andalucía" (IECA). Data from 2011 was not used as much as this update was based on interpolations rather than survey data. Intense data preparation was carried out on this information with data integration and cleaning, the transformation of attributes through the creation of aggregated indicators that synthesize the main demographic qualities of the original data in an objective and compact way. Due to the robustness of the SOM, it is not necessary to carry out their typification or normalization [121] prior to aggregation and incorporation into the model.

1. Instances: The unit of territory on which the data were obtained is the Census Section, reaching the totality of the 5381 census sections of Andalusia, representing the totality of the surface and population censused in the Andalusian region, not initially carrying out any kind of sampling.
2. Attributes: Table 2 lists the indicators elaborated from the Andalusian Population Census used as Modeling Phase 1, measuring instruments to identify the factors and concepts of the state-of-the-art of social vulnerability with which they are related. The attributes used in Modeling Phase 2 (Table 3) were composed of variables of the residential dimension not being used in Modeling Phase 1.

Table 2. Indicators used in research (in Model 1), their relationship (positive or negative) with social vulnerability, and their connection with state-of-the-art concepts. Abbreviations: ● = Connected; ○ = Non-connected. Source: Compiled by the authors.

Measurements (Indicators)	Relation to Social Vulnerability	Age	Gender	Race and Ethnicity	Occupation	Population Growth	Socioeconomic Status	Employment	Education	Social Dependence	Spetial Needs Populations	Commercial & ind. Dev.	Rural/ Urban	Residential Property	Renters	Family and Social Structure	Infrastructure & Lifelines	Medical Services
A01_Health facilities/1 k inhabitants	−	○	○	○	○	○	○	○	○	●	●	○	○	○	○	○	○	●
A02_Education facilities/1 k inhabitants	−	○	○	○	○	○	●	○	●	●	○	○	○	○	○	○	○	○
A03_Well-being facilities/1 k inhabitants	−	○	○	○	○	○	●	○	○	●	○	●	○	○	○	○	○	○
A04_Cultural or sport facilities/1 k inhabitants	−	○	○	○	○	○	●	○	○	○	○	○	○	○	○	○	○	●
A05_Facilities/1 k inhabitants	−	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	○
B01_Percentage of dwellings no running water	+	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○	○	○
B02_Percentage of dwellings with gas	−	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○	○	○
B03_Percentage of dwellings with telephone	−	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○	○	○
C01_Percentage of street cleaning complaints	+	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○	○	○
C02_Percentage of crime complaints	+	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○	○	○
D01_Population	+/−	○	○	○	○	●	○	○	○	○	○	○	○	○	○	○	○	○
D02_Average age population	+	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
D03_Percentage of births	−	●	○	○	○	●	○	○	○	○	○	○	○	○	○	○	○	○
D04_Percentage of men	−	○	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
D05_Percentage of women	+	○	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
E01_People per building	+/−	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○
E02_Percentage of households 1 adult	+	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○
E03_Porcent. of households 1 adult and minor	+	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○
E04_Percentage of households with 2 adults	−	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○
E05_Percentage of households with 3 adults	+	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○
E06_Percentage of households with 4 adults	+	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○
E07_Homes	−	○	○	○	○	●	○	○	○	○	○	○	○	○	○	○	○	○
E08_Inhabitants per household	+/−	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○
E09_Ratio of residential buildings/household	+	○	○	○	○	●	○	○	○	○	○	○	○	○	○	○	○	○
F01_Percentage of rooted population	−	○	○	●	○	●	○	○	○	○	○	○	○	○	○	○	○	○
F02_Percent. of provincial immigrant populat.	+	○	○	●	○	●	○	○	○	○	○	○	○	○	○	○	○	○
F03_Percent. of regional immigrant population	+	○	○	●	○	●	○	○	○	○	○	○	○	○	○	○	○	○
F04_Percent. of national immigrant population	+	○	○	●	○	●	○	○	○	○	○	○	○	○	○	○	○	○
F05_Percent. of foreign immigrant population	+	○	○	●	○	●	○	○	○	○	○	○	○	○	○	○	○	○
F06_Percentage from Spain	−	○	○	●	○	●	○	○	○	○	○	○	○	○	○	○	○	○
F07_Percentage from EU	−	○	○	●	○	●	○	○	○	○	○	○	○	○	○	○	○	○
F08_Percentage from non-EU Europe	−	○	○	●	○	●	○	○	○	○	○	○	○	○	○	○	○	○
F09_Percentage from North America	−	○	○	●	○	●	○	○	○	○	○	○	○	○	○	○	○	○
F10_Percentage from Central America	+	○	○	●	○	●	○	○	○	○	○	○	○	○	○	○	○	○
F11_Percentage from South America	+	○	○	●	○	●	○	○	○	○	○	○	○	○	○	○	○	○
F12_Percentage from Asia	+	○	○	●	○	●	○	○	○	○	○	○	○	○	○	○	○	○
F13_Percentage from Africa	+	○	○	●	○	●	○	○	○	○	○	○	○	○	○	○	○	○
F15_Percentage from Oceania	+/−	○	○	●	○	●	○	○	○	○	○	○	○	○	○	○	○	○
F16_Percentage Statelessness	+	○	○	●	○	●	○	○	○	○	○	○	○	○	○	○	○	○
G01_Percentage working in the province	−	○	○	○	●	●	○	●	○	○	○	○	○	○	○	○	○	○
G02_Percentage working in region	+	○	○	○	●	●	○	●	○	○	○	○	○	○	○	○	○	○
G03_Percentage working in Spain	+	○	○	○	●	●	○	●	○	○	○	○	○	○	○	○	○	○
G04_Percentage working in another country	+	○	○	○	●	●	○	●	○	○	○	○	○	○	○	○	○	○
H01_Percentage of employed population	−	○	○	○	●	●	○	●	○	○	○	○	○	○	○	○	○	○
H02_Percentage of population unemployed	+	○	○	○	●	●	○	●	○	○	○	○	○	○	○	○	○	○
H03_Percentage of inactive population	+	○	○	○	●	●	○	●	○	○	○	○	○	○	○	○	○	○
I01_Commercial establishment/1 k inhabitants	+/−	○	○	○	●	○	○	○	○	○	○	●	○	○	○	○	○	○
I02_Office and services/1 k inhabitants	−	○	○	○	●	○	○	○	○	○	○	●	○	○	○	○	○	○
I03_Industrial/1 k inhabitants	+	○	○	○	●	○	○	○	○	○	○	●	○	○	○	○	○	○
I04_Premises dedicated to farming/1 k inhab.	+/−	○	○	○	●	○	○	○	○	○	○	●	○	○	○	○	○	○
I05_Inactive premises/1 k inhabitants	+	○	○	○	●	○	○	○	○	○	○	●	○	○	○	○	○	○

Table 2. *Cont.*

Measurements (Indicators)	Relation to Social Vulnerability	Age	Gender	Race and Ethnicity	Occupation	Population Growth	Socioeconomic Status	Employment	Education	Social Dependence	Spetial Needs Populations	Commercial & ind. Dev.	Rural/ Urban	Residential Property	Renters	Family and Social Structure	Infrastructure & Lifelines	Medical Services
J01_Combined HDI -	+	○	○	●	○	●	●	○	○	○	○	○	○	○	○	○	○	○
J02_Combined HDI +	−	○	○	●	○	●	●	○	○	○	○	○	○	○	○	○	○	○
J03_Combined HDI	+/−	○	○	●	○	●	●	○	○	○	○	○	○	○	○	○	○	○
K01_Percent. active pop. employed in farming	+/−	○	○	○	●	○	○	○	○	○	○	●	●	○	○	○	○	○
K02_Percent. active pop. employed in fisheries	+/−	○	○	○	●	○	○	○	○	○	○	●	○	○	○	○	○	○
K03_Percent. active pop. employed in industry	+/−	○	○	○	●	○	○	○	○	○	○	●	○	○	○	○	○	○
K04_P. active p. employed in construction	+/−	○	○	○	●	○	○	○	○	○	○	●	○	○	○	○	○	○
K05_P. active p. employed in the service sector	−	○	○	○	●	○	○	○	○	○	○	●	○	○	○	○	○	○
K06_Percent. active population unemployed	+	○	○	○	●	○	○	○	○	○	○	●	○	○	○	○	○	○
L01_Percentage of dwelling owned	−	○	○	○	○	○	○	○	○	○	○	○	○	●	●	○	○	○
L02_Percentage of dwellings for rent	+	○	○	○	○	○	○	○	○	○	○	○	○	●	●	○	○	○
L03_P. of housing not rented and not owned	+	○	○	○	○	○	○	○	○	○	○	○	○	●	●	○	○	○
M01_Construction status (M)	+	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○	○	○
M02_Construction status (SD)	+	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○	○	○
N01_Percentage of illiteracy	+	○	○	○	○	○	○	○	●	○	○	○	○	○	○	○	○	○

Table 3. Indicators used in Model 2 as residential dimension variables. Source: Compiled by the authors.

Measurements
O01_Percentage of dwellings with garage
O02_Percentage of non-accessible dwellings
P01_Average age of constructions (year) (M)
P02_Average age of Constructions (year) (SD)
Q01_Percentage of complaints outside noise
Q02_Percentage of pollution complaints
R01_Average height of constructions (M)
R02 Average building height (SD)
S01_Housing buildings
T01_Percentage of single-family dwellings
T02_Percent. grouped single-family dwelling
T03_P. single-family dwellings with commercials
T04_Percentage of multi-family dwellings
T05_Percentage commercial b. with dwellings
T06_Percentage of commercial buildings
T07_Percentage of housing buildings
U01_ P. complaints poor communications
U02_P. of complaints about scarce green areas
V01_Percentage of dwellings with no toilets

3.2. Data Warehouse

Initially, it operates with two disconnected databases: one for Modeling Phase 1, with a mainly demographic and social dimension, and another for Modeling Phase 2, based on the dwelling dimension. The functioning is fundamentally independent, connecting only after Modeling Phase 1 to assess how the social vulnerability profiles fit and for the construction of the decision tree.

3.3. Methods-Models

As progress was made, we distinguished between two phases of modeling:

1. Modeling Phase 1: Clustering and knowledge model. Its objectives include the clustering and labeling of demographic and social dimension data, as identified above. In this phase, an artificial neural network was used, specifically, the SOM methodology. This methodology, since it is non-supervised, allows clustering without attributing, a priori, a label with previously attributed definitions and meanings which is useful to reduce the enormous complexity of the data [98].

The clustering of the entities was carried out by means of additional Ward-cluster analysis on the map. In this way, profiles or prototypes are generated by modeling patterns and trends in information [122]. To choose the number of clusters or profiles to be reached, there are multiple different methods and criteria, sometimes using a combination of them [123]. There are statistical approaches that use validation metrics, such as measures of "sums of squares" or dispersion. These include the Ball and Hall indices [124], Calinski and Harabasz [125], the Davies–Bouldin (DB) [126], the Silhouette Coefficient [127], the Cubic Clustering Criterion (CCC) [123], and the method based on the observation of dendrograms [123]. We can highlight approaches that are not strictly based on statistical criteria, such as the a priori method described by Joseph F. Hair Jr. et al. [128], in which an adjusted range was initially defined with which it was expected that it would be possible to interpret the groupings based on manageability criteria and efficiency in the communication and interpretation of the results. After that, by means of practical judgment based on common sense and theoretical foundations, the researcher increases or reduces the final number based on conceptual aspects of the problem. According to the author, this methodology provides a better probability solution than

those based exclusively on statistical criteria [128]. Considering the above and due to the descriptive nature of the research, it is considered pertinent to restrict the number of profiles to a conceptual criterion of the problem, choosing the number of profiles from which a relevant interpretation of them can be obtained. In view of the above, and considering that this research presents a clear intention descriptive of reality, it was considered appropriate to restrict the solution of the number of profiles to an exclusively conceptual criterion of the problem, proposing to reach a number of profiles on which it is possible to make a relevant and useful interpretation of the data. In this way, an iterative process was carried out in which the number of profiles grew, evaluating meaning and relevance. The process ends when it is no longer possible to explain, with the necessary clarity, the meaning of a new profile, or, on the contrary, its fragmentation presents little value at a practical and conceptual level in research.

In order to facilitate the understanding of the profiles obtained, each cluster was characterized with its basic statistics, such as the Mean, Standard Deviation, Maximum and Minimum [88], with the main aim of obtaining two additional results, (i) the factor or variable that is most important for the effect and (ii) the value of such a factor [129]. In addition to the statistical information that defines the profiles, monovariable SOM Maps are valuable for the analysis of the profiles, since they allow, according to the distribution of values in the same position, the evaluation of relationships and correlations between variables. Following the recommendations of the American Statistical Association [130] for each variable and profile, in addition to the statistical significance, its effect size (ES) was calculated [131]. Statistical significance was calculated using the bilateral T-Student Test (p-value ≤ 0.05). The ES is a measure of how the values reached in the variable are influenced by whether or not they are within the profile in question. It is calculated as the quotient of the difference of the mean between the experimental group (profile) and the mean of the control group (population mean) divided by the standard deviation of the population [131]. In the corresponding tables, the effect sizes are indicated for each attribute/variable that intervenes in the construction of the profile: +++ large positive effect, ++ medium positive effect, + low positive effect, $-$ low negative effect, $- -$ medium negative effect, $- - -$ large negative effect [132], obtaining very relevant information of the effect that the variables have on the definition and singularity of each profile. The Viscovery SOMine 5.0.2.t. software was used in this work for the construction of the SOM model, due to its good results at the visual representation level [133].

2.　Modeling Phase 2: Prediction model. For the construction of the model that allows for the prediction of social vulnerability, a decision tree based on rules was evaluated, identifying through the representation of successive conditions, the degree of probability of the existence of the vulnerability pattern obtained in the Modeling Phase 1. For this purpose, the data was partitioned into 70/30 (training/test) and conditional inference trees were used based on the permutation test [112] using non-parametric tests as branch division criteria, not requiring pruning. For this purpose, the "rpart" package of the statistical software R-Project [134] was used, using minimum division = 20, maximum depth = 2, and minimum cube = 7 as the parameters.

3.4. Visual Representations

One of the main qualities of the SOM is its ability to represent the resulting information in a very powerful and synthetic way and, at the same time, in a way that is relatively simple to understand and interpret, by showing a two-dimensional representation of the starting instances with the characteristic that each one of them has as a "neighbor", the instance with the most similar qualities. The same cartography usually represents the groupings of the instances in the different conformed profiles. This representation is usually completed with a map for each of the attributes or variables that helped build the ANN of the SOM.

As each evaluated instance has its identity and form in space, in our case, the spaces that made up the profiles in Modeling Phase 1 are represented through a GIS. This return to the GIS of the instances, once classified into classes, has been frequent, for example, in medical research in non-linear analysis of multiple variables in certain diseases [91], in the representation of SOM clustering results

on the ecological risk of contamination [88] or experimentally applying them to data from official socio-demographic information of the Lisbon Metropolitan Area [135].

Finally, once the graphs of the decision trees from Modeling Phase 2 were obtained, it was possible to move on to the Application Phase of the learned models, enabling the prediction from certain residential variables, whether there is a greater or lesser probability of social vulnerability.

4. Results

According to the methodology, two independent databases were obtained. A descriptive synthesis of these baseline variables can be seen in the first two columns of data (population) in Table 4 for the main demographic data and Table 5 for the dwelling dimension. Continuing with the following section of the methodology, Modelling Phase 1 was carried out using 66 variables of the demographic, social, labor, facilities, and services, etc., dimensions. The profiles that characterize the demographic reality and the main dimensions of social vulnerability were obtained.

Once the profiles were obtained, they were spatially represented (Figure 2).
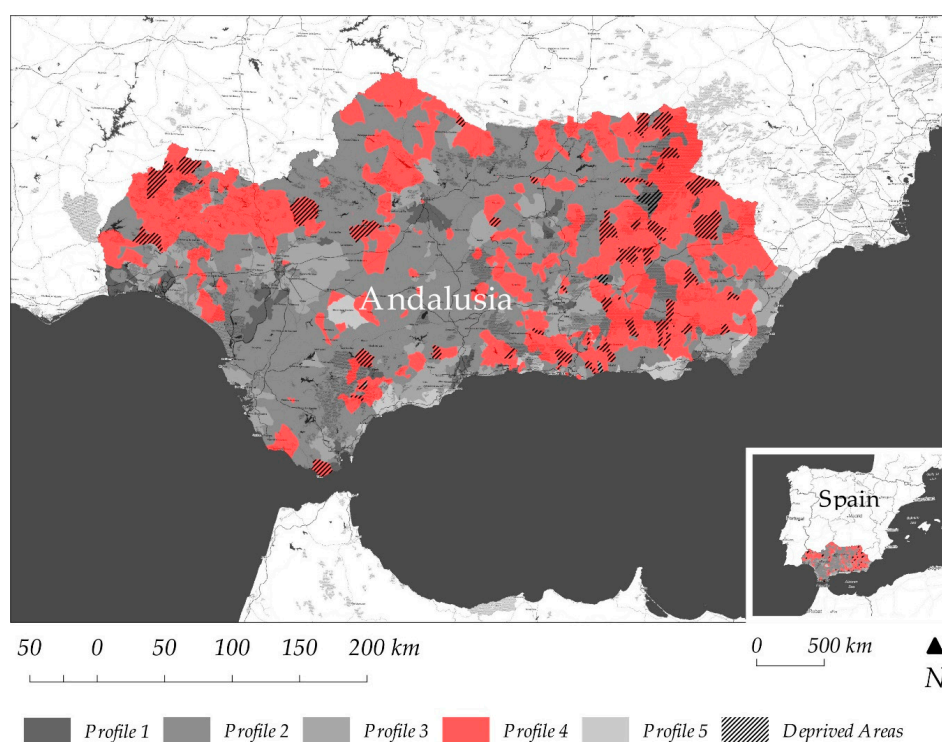
Figure 2. GIS representation of the clustering generated by SOM. From the statistical analysis of all the profiles, Profile 4 coincides with the social vulnerability traits of the state-of-the-art. The deprived areas are represented by a stripe hatch according to [58]. Source: Compiled by the authors.

A simple observation of the figure shows that there are several areas on the map that are known and recognized as areas of some degree of vulnerability [58]. Among them, the areas of Northeast Granada, North of Huelva, and Interior of Almería can be highlighted.

Beyond this rough verification, validation of Model 1 is carried out, comparing it with the results of an investigation in which deprived urban areas of Andalusia are identified. The authors of that study describe that any threat in certain circumstances could make such deprived areas as vulnerable [58]. In order to obtain it, the authors used aggregated indicators in a simple way (sum) based on their typification and without any type of weighting or complex statistical analysis. It can, therefore, be approximated that the areas proposed by this work are areas exposed to a degree of vulnerability, although a certain weakness in the methodological approach could be criticized. In any case, they are considered adequate to validate the results of Model 1.

Table 4. Variables of the demographic, social, labor, facilities, and services dimensions. Data from the full sample and Profile 4 with social vulnerability traits are shown. Abbreviations: cs: census tracks; M: Mean; SD: Standard Deviation; C: Confidence (*p*-value: ns: *p* > 0.05; * *p* ≤ 0.05; ** *p* ≤ 0.01; *** *p* ≤ 0.001; ns: Non-significant); ES: Size of Effect (+++: Large positive; ++: Medium positive; +: Small positive; − − −: Large negative; − −: Medium negative; −: Small negative). Source: Compiled by the authors.

Measurements	Population		Profile 4					
	5381cs (100%)		550cs (10.22%)					
	M	**SD**	**M**	**SD**	**C**	**T**	**ES**	**ES**
A01_Health facilities/1 k inhabitants	1.39	6.16	1.56	2.45	ns	1.62	0.027	
A02_Education facilities/1 k inhabitants	1.14	2.36	2.16	3.2	***	7.51	0.433	+
A03_Well-being facilities/1 k inhabitants	0.831	1.682	1.845	3.774	***	6.30	0.602	++
A04_Cultural or sport facilities/1 k inhabitants	0.715	1.468	1.941	3.067	***	9.37	0.835	+++
A05_Facilities/1 k inhabitants	4.08	8.31	7.51	7.25	***	11.1	0.412	+
B01_Percentage of dwellings no running water	0.906	2.836	3.156	6.471	***	8.15	0.793	++
B02_Percentage of dwellings with gas	22.77	33.76	12.31	28.29	***	8.67	−0.30	−
B03_Percentage of dwellings with telephone	86.96	17.48	63.45	23.69	***	23.2	−1.34	− − −
C01_Percentage of street cleaning complaints	35.09	19.82	21.22	22	***	14.7	−0.69	− −
C02_Percentage of crime complaints	25.32	23.66	4.942	10.61	***	45.0	−0.86	− − −
D01_Population	1367	518	900.2	410.5	***	26.6	−0.90	− − −
D02_Average age population	38.00	4.407	42.73	4.015	***	27.6	1.073	+++
D03_Percentage of births	11.31	3.209	9.541	2.78	***	14.9	−0.55	− −
D04_Percentage of men	0.491	0.023	0.519	0.019	***	9.85	0.432	+
D05_Percentage of women	0.509	0.023	0.481	0.019	***	9.85	−0.43	−
E01_People per building	14.46	22.91	2.43	3.79	***	74.5	−0.52	− −
E02_Percentage of households 1 adult	18.80	7.548	27.86	7.369	***	28.8	1.200	+++
E03_Porcent. of households 1 adult and minor	1.832	1.115	1.278	0.923	***	14.0	−0.49	−
E04_Percentage of households with 2 adults	41.24	6.749	41.00	5.11	ns	1.07	−0.03	
E05_Percentage of households with 3 adults	18.40	3.363	16.53	3.129	***	14.0	−0.55	− −
E06_Percentage of households with 4 adults	19.71	6.639	13.31	4.701	***	31.9	−0.96	− − −
E07_Homes	449.2	167.5	340.2	145.7	***	17.5	−0.65	− −
E08_Inhabitants per household	3.035	0.359	2.618	0.305	***	32.0	−1.15	− − −
E09_Ratio of residential buildings/household	0.742	0.564	1.417	0.476	***	33.1	1.195	+++
F01_Percentage of rooted population	80.10	9.491	82.19	4.996	***	9.78	0.219	+
F02_Percent. of provincial immigrant populat.	3.871	5.813	3.833	3.524	ns	0.25	−0.01	
F03_Percent. of regional immigrant population	1.349	1.278	0.966	1.088	***	8.24	−0.29	−
F04_Percent. of national immigrant population	1.582	1.326	1.886	1.586	***	4.50	0.229	+
F05_Percent. of foreign immigrant population	1.317	2.794	1.131	1.946	*	2.23	−0.06	
F06_Percentage from Spain	97.92	4.58	98.58	2.68	***	5.79	0.144	
F07_Percentage from EU	0.77	2.967	0.819	2.205	ns	0.52	0.016	
F08_Percentage from non-EU Europe	0.181	0.56	0.119	0.376	***	3.85	−0.11	
F09_Percentage from North America	0.049	0.181	0.016	0.076	***	10.2	−0.18	
F10_Percentage from Central America	0.036	0.088	0.014	0.048	***	10.8	−0.25	−
F11_Percentage from South America	0.412	0.827	0.202	0.473	***	10.3	−0.25	−
F12_Percentage from Asia	0.083	0.337	0.019	0.118	***	12.6	−0.18	
F13_Percentage from Africa	0.543	2.011	0.226	1.052	***	7.08	−0.15	
F14_Percentage from Oceania	0.003	0.022	0.001	0.011	***	3.53	−0.08	
F16_Percentage Statelessness	0.000	0.006	0	0	***	−	−0.04	
G01_Percentage working in the province	6.428	5.484	6.808	4.43	*	2.00	0.069	
G02_Percentage working in region	0.867	0.999	0.975	0.948	**	2.65	0.107	
G03_Percentage working in Spain	0.472	0.568	1.001	1.256	***	9.87	0.929	+++
G04_Percentage working in another country	0.12	0.289	0.162	0.534	ns	1.85	0.146	
H01_Percentage of employed population	33.19	6.505	27.37	7.108	***	19.1	−0.89	− − −
H02_Percentage of population unemployed	10.47	5.106	12.94	7.532	***	7.68	0.483	+
H03_Percentage of inactive population	38.02	7.242	43.94	7.253	***	19.1	0.817	+++
I01_Commercial establishment/1 k inhabitants	26.9	98.9	19.4	20.9	***	8.40	−0.07	
I02_Office and services/1 k inhabitants	10.7	44.6	8.1	12.2	***	5.08	−0.05	
I03_Industrial/1 k inhabitants	3.07	12.11	3.63	7.94	ns	1.66	0.046	
I04_Premises dedicated to farming/1 k inhab.	0.65	6.71	3.14	19.84	**	2.95	0.371	+
I05_Inactive premises/1 k inhabitants	13.37	20.22	19.63	29.47	***	4.98	0.309	+
J01_Combined HDI -	−0.002	0.004	0.001	0.002	***	8.98	0.222	+
J02_Combined HDI +	0.000	0.000	0.000	0.000	ns	1.29	−0.03	
J03_Combined HDI	−0.002	0.004	0.000	0.003	***	8.79	0.220	+
K01_Percent. active pop. employed in farming	4.161	6.062	7.275	5.736	***	12.7	0.513	++
K02_Percent. active pop. employed in fisheries	0.139	0.593	0.026	0.123	***	21.6	−0.19	
K03_Percent. active pop. employed in industry	3.809	2.46	2.814	2.178	***	10.7	−0.40	−
K04_P. active p. employed in construction	4.471	2.442	4.352	2.081	ns	1.34	−0.04	
K05_P. active p. employed in the service sector	20.60	8.115	12.90	4.829	***	37.3	−0.95	− − −
K06_Percent. active population unemployed	24.09	11.26	31.61	16.40	***	10.7	0.668	++
L01_Percentage of dwelling owned	82.28	13.3	82.73	10.1	ns	1.02	0.033	
L02_Percentage of dwellings for rent	9.393	11.28	5.046	5.163	***	19.7	−0.38	−
L03_P. of housing not rented and not owned	8.324	8.686	12.22	9.616	***	9.52	0.449	+
M01_Construction status (M)	1.165	0.206	1.230	0.197	***	7.77	0.316	+
M02_Construction status (SD)	0.395	0.233	0.501	0.208	***	11.9	0.451	+
N01_Percentage of illiteracy	4.631	3.766	6.811	3.825	***	13.3	0.579	++

Table 5. Residential dimension variables. Full sample and Profile 4 data are displayed. Please note that this information did intervene with the constitution of the profile. Abbreviations: cs: census tracks; M: Mean; SD: Standard Deviation; C: Confidence (*p*-value: ns: $p > 0.05$; * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; ns: Non-significant); ES: Size of Effect (+++: Large positive; ++: Medium positive; +: Small positive; − − −: Large negative; − −: Medium negative; −: Small negative). Source: compiled by the authors.

Measurements	Population		Profile 4					
	5381cs (100%)		550cs (10.22%)					
	M	**SD**	**M**	**SD**	**C**	**T**	**ES**	**ES**
O01_Percentage of dwellings with garage	19.03	18.79	16.7	14.07	***	3.88	−0.12	
O02_Percentage of non-accessible dwellings	72.81	30.4	87.94	22.67	***	15.6	0.497	+
P01_Average age of constructions (year) (M)	1963	20	1942	28.1	***	17.1	−1.01	− − −
P02_Average age of Constructions (year) (SD)	23.38	14.79	37.00	14.78	***	21.6	0.920	+++
Q01_Percentage of complaints outside noise	32.52	18.35	13.15	15.82	***	28.7	−1.06	− − −
Q02_Percentage of pollution complaints	19.46	14.82	8.114	11.65	***	22.8	−0.77	− −
R01_Average height of constructions (M)	2.678	1.603	1.701	0.507	***	45.2	−0.61	− −
R02 Average building height (SD)	5.463	5.272	1.943	1.338	***	61.7	−0.67	− −
S01_Housing buildings	327.5	292.7	460.1	225.3	***	13.8	0.453	+
T01_Percentage of single-family dwellings	61.1	33.88	87.48	13.05	***	47.4	0.779	++
T02_Percent. grouped single-family dwelling	21.66	21.86	8.52	8.7	***	35.4	−0.60	− −
T03_P. single-family dwell. with commercials	17.04	24.09	3.94	6.38	***	48.1	−0.54	− −
T04_Percentage of multi-family dwellings	0.196	0.795	0.057	0.2	***	16.3	−0.17	
T05_Percentage commercial b. with dwellings	0.8	2.585	0.239	0.742	***	17.7	−0.22	−
T06_Percentage of commercial buildings	8.93	19.87	6.95	7.79	***	5.95	−0.10	
T07_Percentage of housing buildings	0.08	1.047	0.22	1.79	ns	1.84	0.134	
U01_ P. complaints poor communications	13.94	15.95	15.06	20.53	ns	1.27	0.069	
U02_P. of complaints about scarce green areas	48.84	25.97	37.59	31.06	***	8.49	−0.43	−
V01_Percentage of dwellings with no toilets	1.319	2.947	2.197	5.114	***	4.03	0.298	+

In order to validate Model 1, the data presented in the Tables in [58] are used, which are presented in aggregate form by municipality. This has meant that only data from municipalities with only one census section and those that can be extrapolated to each census section without the possibility of error can be used for the validation. Table 6 shows the confusion matrix of Model 1.

Table 6. Confusion matrix of Model 1. This does not include all the data incorporated in the construction of Model 1. It excludes the data of the document used [59], from which the true condition attribution cannot be guaranteed for each specific census section. This is because the information in such a document is presented aggregated at the municipal level and not at the census section level as in our models. Normally excluded are municipalities with more than one census section, i.e., bigger municipalities. Abbreviations: TP = True Positive; FP = False Positive; FN = False Negative; TN = True Negative.

		True Condition (Population in Deprived Areas) [1]		Total Cumulative
		Condition Positive	**Condition Negative**	
Predicted condition (Predicted as Profile 4)	**Predicted positive**	**60** (TP)	**412** (FP)	472
	Predicted negative	**4** (FN)	**1499** (TN)	1503
	Total cumulative	64	1911	1975

[1] Source: Prepared by the authors on the basis of [58].

From the previous table, the following indicators of the classification performance of Model 1 can be obtained:

$$\text{Recall, Sensitivity, or True Positive Rate (TPR)} = TP/(TP + FN) = 0.9375, \tag{1}$$

$$\text{Precision, or Positive Predictive Value (PPV)} = TP/(TP + FP) = 0.1271 \tag{2}$$

$$\text{Specificity, or True Negative Rate (TNR)} = TN/(FP + TN) = 0.7844 \tag{3}$$

$$\text{Accuracy (ACC)} = (TP + TN)/(TP + TN + FP + FN) = 0.7894 \tag{4}$$

$$\text{Balanced Accuracy (bAAC)} = (TPR + TNR)/2 = 0.8610 \tag{5}$$

Because data is imbalanced, the indicator considered most suitable for evaluating performance is (4) balanced accuracy (bAAC). A bACC = 0.8610 is obtained, which is considered a good accuracy. You also get one (1) recall = 0.9375, quite good performance, which shows that true positive predictions are high. On the other hand, the indicator (2) Precision = 0.1271, which denotes that the model is predicting many more cases with social vulnerability than the reference considers as such. This weakness of the predictions is considered tolerable since it is somehow predicting practically all "real" cases and others in which with certain probability situations tending towards social vulnerability are taking place.

If we analyze the five profiles obtained in Clustering Model 1, comparing both the statistical information that characterizes each of them (Table 3) and the spatialization of the profiles in the Andalusian region (Figure 2), we obtain the following results:

- Profile 1: Statistically, it is verified that the census sections contained in this profile present, compared with the other profiles, a greater presence of delinquency, a greater number of persons per building, a greater dedication in service employment, and a lower number of dwellings per occupied household. Through spatial representation through GIS, coincidences are observed with the main urban areas and their closest conurbations throughout the region. This profile shows the urban connotations of a well-consolidated city.

- Profile 2: A clear diversification of employment is observed, with little presence of the service sector, an eminently Spanish population, with few immigrants and a high number of illiterates, with little presence of households with only one adult and minors. This profile is spatially identified with a population located in rural environments, differing with respect to the other rural profile (Profile 4) in that its population is younger than in the former, with a larger active population with more activities typical of that reality, such as, for example, a greater dedication to construction or industry, and with households with a greater number of inhabitants.

- Profile 3: It stands out for a greater number of births, a greater number of immigrants of provincial origin, and to a lesser extent, regional or national. They usually work in the province, with a high percentage of employed—a low unemployment rate. It is below average age, with few single-person households, and a low level of rootedness. Spatially, they are located in the main cities' outskirts.

- Profile 4: The statistical analysis reveals that this population profile presents a high average age, a large number of households with a single occupant, an abundance of empty dwellings, and with issues such as a greater proportion of lack of running water than the rest. Statistical data reveal that they live in settlements with good ratios of cultural equipment and well-being per population, probably derived from the low number of inhabitants of such populations and acceptable distribution of such functions. Spatially, it is observed that they correspond to the most isolated rural sites and at a greater distance from the main cities. Comparing this profile with Profile 2, it is observed that it coincides with an older rural population, which often lives alone in urban environments with a small population, with little occupation of the dwellings and with high rates of illiteracy, unemployment, and inactivity. We can locate this profile, among other areas, prominently in Hoya de Baza (Granada), in Campos de Tabernas (Almería), in Altos de

Sierra de Gádor (Almería) or in Sierra de Aracena (Huelva). As we observed in the state-of-the-art, this profile is identified with most of the factors that trigger social vulnerability.

- Profile 5: It stands out for a high number of dwellings occupied by one person, on many occasions with some minor in charge, a high presence of immigrants from the rest of Andalusia, the rest of Spain and especially, foreigners with the consequent low rootedness of its population. They have a high employment rate, low unemployment, and low inactivity, working primarily in the service sector or in agriculture. They are spatially recognized and identified as well-known urban areas with a strong and unique presence of foreign residents. It is shown in tourist enclaves, such as the coast of Málaga and Granada, and in a very intensive agricultural production zone, such as the greenhouse area of the coast of Almería (Campo de Dalías).

Then, in Modeling Phase 2, a tree was obtained that allowed "predicting" how to identify Profile 4, from the variables that were introduced as predictors (dwelling variables). In other words, it is a question of identifying the belonging or probability of belonging to the socially vulnerable profile from certain dwelling qualities that can be observed with certain ease in the scope of the corresponding census section (Figure 3).
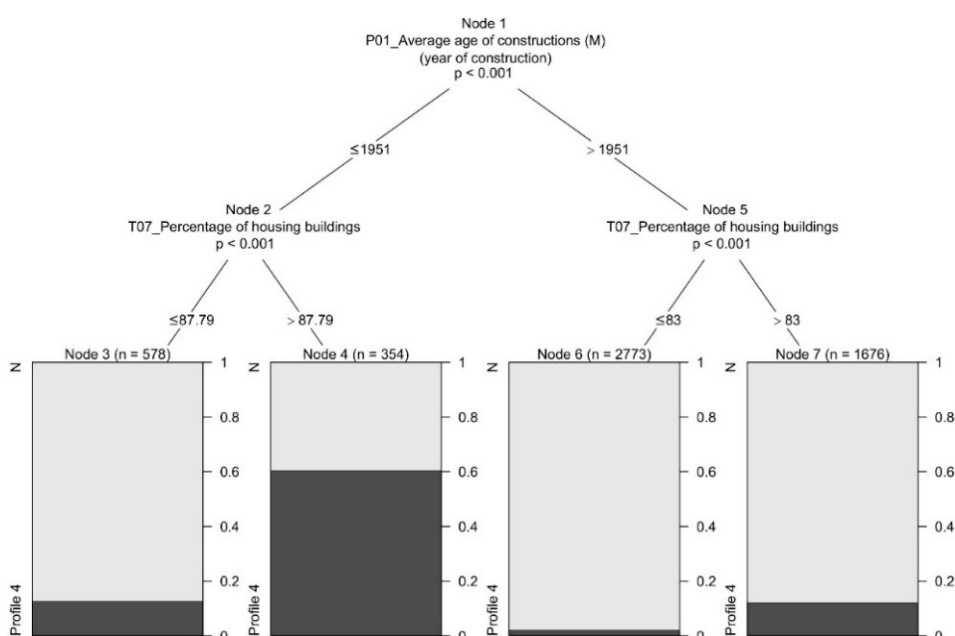
Figure 3. Conditional decision tree. The predicted variable is the membership in Profile 4 that has been connected to social vulnerability traits. N: Belonging to a profile other than 4. Source: Compiled by the authors.

The conditional or decision tree obtained in Figure 3 represents, at the bottom, the probability (ratio: 1 = 100%) of presenting the profile with social vulnerability (marked in black) as opposed to the probability of belonging to other profiles, which, as we previously verified, show other well-differentiated characteristics. In the tree obtained, it was observed that with only two variables observable in situ—the average age of construction and the percentage of housing building—it is possible to predict whether or not they belong to the socially vulnerable profile, which, as we verified, requires the use of numerous variables and indicators, often difficult and costly to access.

To evaluate the predictive capabilities of the model obtained by means of a conditional classification tree, the ROC (receiver operating characteristic) curve was calculated (Figure 4), obtaining AUC = 0.78 (area under the curve).

Finally, Table 7 provides a preview of the predictive capabilities of Models 1 and 2 when predicting the presence of municipalities with more than 50% of the population in deprived areas. Table 6 shows

the four municipalities with a false negative (Table 5) extracted from the municipalities with more than 50% of the population in deprived areas. Two municipalities with an imprecise Model 1 prediction are added to the previous ones. It can also be observed that most of the probability predictions of Model 2 are close to those of Model 1 and the reference data [58]. It should be noted that in Model 2, the highest predicted probabilities are 60%.

Table 7. Municipalities with populations greater than 50% in deprived areas [58] and predictive performance of Model 1 and Model 2 are shown. Adequate predictions are shown in bold font.

Province	Municipality	Population 2006 [1]	Population in Deprived Areas (%) [1]	Model 1: Population in Profile 4 (%)	Model 2: Predicted Probability of Profile 4 (%)
Almería	Almócita	156	100	**100**	**60 (max.)**
	Alsodux	131	100	**100**	**60 (max.)**
	Beires	128	100	100	13.50
	Benitagla	66	100	**100**	**60 (max.)**
	Canjáyar	1561	58.62	100	**60 (max.)**
	Cóbdar	192	100	**100**	**60 (max.)**
	Ohanes	765	100	**100**	**60 (max.)**
	Sta. Cruz de Marchena	245	100	**100**	**60 (max.)**
	Turrillas	249	100	**100**	**60 (max.)**
	Tres Villas (Las)	581	100	**100**	**60 (max.)**
Cádiz	San José del Valle	4244	67.95	0	3 to 13.5 [2]
Córdoba	Conquista	486	100	**100**	**60 (max.)**
Granada	Agrón	283	100	**100**	**60 (max.)**
	Albondón	914	100	**100**	12.5
	Albuñán	448	100	**100**	13.5
	Almegíjar	421	100	**100**	**60 (max.)**
	Cástaras	259	100	**100**	**60 (max.)**
	Cortes de Baza	2206	55.03	100	**13.5 to 60** [2]
	Darro	1438	100	0	13.5
	Freila	1074	100	**100**	12.5
	Gorafe	526	100	**100**	**60 (max.)**
	Itrabo	1117	100	**100**	12.5
	Lobras	121	100	**100**	**60 (max.)**
	Lugros	367	100	**100**	**60 (max.)**
	Lújar	491	100	**100**	13.5
	Orce	1387	100	**100**	**60 (max.)**
	Polopos	1557	100	0	12.5
	Soportújar	265	100	**100**	13.5
	Villanueva Torres	778	100	**100**	3
	Nevada	1179	100	**100**	**60 (max.)**
	Guajares (Los)	1337	100	**100**	13.5
Huelva	Cumbres Enmedio	44	100	**100**	**60 (max.)**
	Cumbres S. Bartolomé	490	100	**100**	**60 (max.)**
	Valdelarco	237	100	**100**	**60 (max.)**
Jaén	Chiclana de Segura	1191	60.20	100	**60 (max.)**
	Espelúy	750	100	**100**	**60 (max.)**
	Génave	565	100	**100**	**60 (max.)**
	Hinojares	446	100	**100**	**60 (max.)**
	Hornos	663	100	**100**	**60 (max.)**
	Santiago Calatrava	883	100	**100**	13.5
Málaga	Atajate	142	100	0	13.5
	Benadalid	258	100	**100**	13.5
	Benarrabá	538	100	**100**	13.5
	Sedella	646	100	**100**	**60 (max.)**
Sevilla	Villanueva Río-M	5217	83.21	0	3 to 12.5 [2]

[1] Source: [58]. [2] In this municipality, there are several census sections with different predictions.
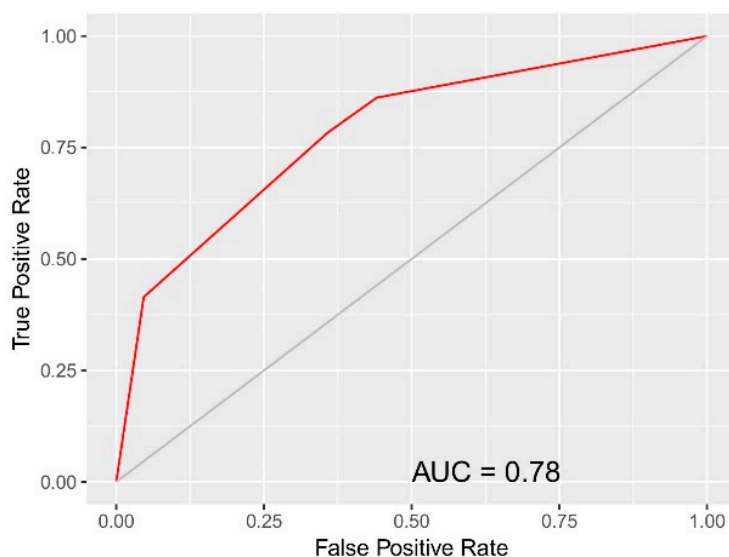
Figure 4. Receiver operating characteristic (ROC) curve of the decision tree model. Source: Compiled by the authors.

5. Discussion

The main contribution of this research is that it has been possible to predict, with a certain level of precision both in Model 1 (Balanced Accuracy = 0.8610) and for Model 2 (AUC = 0.78), the probability of social vulnerability based on such simple residential indicators as the age of the buildings (year of construction) and the percentage of residential housing. The indicators that can be used to predict social vulnerability are (1) P01 Average age of constructions (year of construction), and (2) T07 Percentage of housing buildings. There is no doubt that such immediate approaches to such complex problems can have weaknesses, but they also allow us to have an almost immediate first approximation that can be extremely useful when carrying out approaches with greater depth of analysis and knowledge for the development and implementation of social and urban policies.

From the analysis of the state-of-the-art on the application of the clustering and knowledge model by means of the SOM methodology and corroborated by our own experience, it can be concluded that the SOM methodology is useful to carry out an exploratory analysis [98] to make the descriptive classifications more powerful, robust, and more complete [102], and to help understand the patterns of spatial distribution [88], facilitating explorations and visual evaluations [88,100], effectively analyzing complex geographic and demographic data sets. It also allows inferring spatial considerations from the taxonometric groups found [88], coding classifications in a GIS to approximate them to a wider audience not familiar with AI [24], overcoming the traditional challenges associated with studies of the complexity of environmental communities and showing their value by integrating SOM and GIS [91]. This study verified the ability to label geographic reality without the need to name such categories, suppressing the inherent problems of factor analysis [98], making it possible to evaluate the effects of the concurrence of certain variables under study [88], constituting a powerful alternative solution in a time characterized by information technologies and data proliferation [96], and that can be used as a decision support system to analyze and visualize sets of statistical indicators for various applications [83].

Moreover, the methodology based on decision trees from SOM clustering proved useful to attribute, in a very simple way, behavior patterns that can be very complex in order to effectively predict behaviors of variables that present a certain cost or difficulty of evaluation, such as demographic or social variables, from other variables with less complexity and cost of evaluation, such as residential variables. Its usefulness was verified to generate and verify hypotheses on complex realities and behaviors, without the user's participation is necessary for its formulation, making decision support systems accessible to

a non-expert public, and allowing the identification of variables that are significantly related and their weight or size of the effect on the studied reality.

However, it is necessary to bear in mind certain precautions and limitations in the use of these methodologies and in their concrete implementation. These include the fact that the data used may already be obsolete, and that not all the dimensions of social vulnerability [5] were represented, such as rural/urban differentiation, although, as we have seen, it was implicit in some way with the rest of the indicators. In addition, an analysis of the population of a census section is not an analysis of the population itself, and extreme caution should be exercised and inference should be limited to the scale of observation, not directly reaching individuals [98], i.e., the conclusions obtained from the study of groups of individuals should not be extrapolated to individuals. Moreover, the complete integration between SOM and GIS is complex [136], being limited to a more or less manual connection. Except for a few connection attempts, a "friendly" direct connection between none of the main GIS and SOM software has been implemented to date, requiring the combination of both expert knowledge and creativity [24]. Likewise, the methodologies based on knowledge-based systems are not developed for direct integration into urban and territorial development and planning processes [99,137], which suggests, in conjunction with the previous one, that there is an important technological gap that can become a space for technical and technological development and for research and/or business opportunities. Another limitation that should be highlighted is that the results of "Prediction model" are specific to the territory under study, i.e., Andalusia. They will probably not fit to the specific features of other regions. However, the methodology for obtaining such a model can be used and applied in other geographical contexts.

6. Conclusions

Through research applied to the case study of the region of Andalusia, we obtained a decision tree oriented to the prediction of a model of social vulnerability. This model was constructed using a clustering methodology non-supervised by Self-Organizing Maps. Both techniques proved to be simple to use, as well as useful and able to predict, with relatively low error (Model 1: Balanced Accuracy = 0.8610; Model 2: AUC = 0.78), complex and relevant demographic phenomena, such as social vulnerability. For such a prediction, once the models were trained, only residential reality information was used.

In the methodological process, a series of socio-demographic profiles were obtained in Andalusia. In these models, it is worth highlighting that the presence of an eminently urban profile was distinguished (Profile 1); two suburban profiles, among which we can differentiate a Profile (3) in which there abounds a young and active population with families, short-distance immigrants (provincial), with housing and work in the province, as opposed to another Profile (5) characterized fundamentally by the abundance of long-distance immigrants (regional, national or foreign), who are very active in jobs linked to agriculture or services and who predominantly live in rented housing. Finally, two eminently rural profiles stood out, one in which a certain vitality was observed, youth and economic activity (Profile 2) and another in clear depression, ageing of its population and recession (Profile 4), in which a whole series of indications were evidenced that according to the state-of-the-art, predict a high social vulnerability.

Together with this statistical approximation, by representing the spatial profiles in the region, the areas that could be affected by social vulnerability were detected, evidencing what could be called "another Andalusia", an eminently rural Andalusia, with signs of isolation from the opportunities for employability, etc., offered by cities. Urban areas framed in the social vulnerability profile are certainly scarce. This could be a weakness of the model, and it would be advisable to adjust it to modify the vulnerability threshold and thus encompass areas that the state-of-the-art identifies as such.

Nevertheless, the decision tree obtained was interesting and relevant in that it allowed, in a simple way and with a certain level of precision, prediction of the probability that the inhabitants of an area are socially vulnerable, using a small number of variables that could be observed practically in situ

without costly analysis or surveys. Specifically in the region evaluated, it was observed that only with the age of the buildings and the amount of single-family housing in the place under study was it possible to predict belonging to an urban profile related to situations of social vulnerability, with a probability that can be evaluated with the indicator AUC = 0.78.

Therefore, it can be concluded that there is a connection and relationship between demographic and social vulnerability phenomena and the residential configuration of Andalusia, being cautious and avoiding a priori a cause–effect establishment between such phenomena, which would require other differentiated tests that are far from being the objective of this research. It can be summarized that the main contribution that this work contributes to the field of social vulnerability consists of the prediction with a certain level of precision of the complex phenomenon from easily obtained dwelling information, almost by means of a simple ocular inspection.

References

1. Cutter, S.L. Vulnerability to environmental hazards. *Prog. Hum. Geogr.* **1996**, *20*, 529–539. [CrossRef]
2. Wisner, B.; Blaikie, P.; Cannon, T.; Davis, I. *At Risk: Natural Hazards, People's Vulnerability, and Disasters*; Routledge: New York, NY, USA, 2004; ISBN 0415084768.
3. Ebert, A.; Kerle, N.; Stein, A. Urban social vulnerability assessment with physical proxies and spatial metrics derived from air- and spaceborne imagery and GIS data. *Nat. Hazards* **2009**, *48*, 275–294. [CrossRef]
4. Prowse, M. *Towards a Clearer Understanding of 'Vulnerability' in Relation to Chronic Poverty*; Chronic Poverty Research Centre: Manchester, UK, 2003; ISBN 1904049230.
5. Cutter, S.L.; Boruff, B.J.; Shirley, W.L. Social vulnerability to environmental hazards. *Soc. Sci. Q.* **2003**, *84*, 242–261. [CrossRef]
6. Sánchez-González, D.; Egea-Jiménez, C. Enfoque de vulnerabilidad social para investigar las desventajas socioambientales. Su aplicación en el estudio de los adultos mayores. *Papeles de población* **2011**, *17*, 151–185.
7. Blaikie, P.; Cannon, T.; Davis, I.; Wisner, B. *Vulnerabilidad: El Entorno Social, Político y Económico de los Desastres*; Prim. Edición Julio; LA RED: Ciudad de Panamá, República de Panamá, 1996; p. 292.
8. CEPAL-ECLAC. *Vulnerabilidad Sociodemográfica: Viejos y Nuevos Riesgos Para Comunidades, Hogares y Personas*; CEPAL-ECLAC: Brasilia, Brazil, 2002.
9. Gu, H.; Du, S.; Liao, B.; Wen, J.; Wang, C.; Chen, R.; Chen, B. A hierarchical pattern of urban social vulnerability in Shanghai, China and its implications for risk management. *Sustain. Cities Soc.* **2018**, *41*, 170–179. [CrossRef]
10. Tate, E. Social vulnerability indices: A comparative assessment using uncertainty and sensitivity analysis. *Nat. Hazards* **2012**, *63*, 325–347. [CrossRef]
11. Maharani, Y.N.; Lee, S.; Ki, S.J. Social vulnerability at a local level around the Merapi volcano. *Int. J. Disaster Risk Reduct.* **2016**, *20*, 63–77. [CrossRef]
12. Maharani, Y.N.; Lee, S. Assessment of social vulnerability to natural hazards in South Korea: Case study for typhoon hazard. *Spat. Inf. Res.* **2017**, *25*, 99–116. [CrossRef]
13. Kleinosky, L.R.; Yarnal, B.; Fisher, A. Vulnerability of hampton roads, Virginia to storm-surge flooding and sea-level rise. *Nat. Hazards* **2007**, *40*, 43–70. [CrossRef]

14. Nelson, K.S.; Abkowitz, M.D.; Camp, J.V. A method for creating high resolution maps of social vulnerability in the context of environmental hazards. *Appl. Geogr.* **2015**, *63*, 89–100. [CrossRef]

15. Lee, Y.-J. Social vulnerability indicators as a sustainable planning tool. *Environ. Impact Assess. Rev.* **2014**, *44*, 31–42. [CrossRef]

16. Fatemi, F.; Ardalan, A.; Aguirre, B.; Mansouri, N.; Mohammadfam, I. Social vulnerability indicators in disasters: Findings from a systematic review. *Int. J. Disaster Risk Reduct.* **2017**, *22*, 219–227. [CrossRef]

17. Khazai, B.; Merz, M.; Schulz, C.; Borst, D. An integrated indicator framework for spatial assessment of industrial and social vulnerability to indirect disaster losses. *Nat. Hazards* **2013**, *67*, 145–167. [CrossRef]

18. Rufat, S.; Tate, E.; Burton, C.G.; Maroof, A.S. Social vulnerability to floods: Review of case studies and implications for measurement. *Int. J. Disaster Risk Reduct.* **2015**, *14*, 470–486. [CrossRef]

19. Schmidtlein, M.C.; Deutsch, R.C.; Piegorsch, W.W.; Cutter, S.L. A sensitivity analysis of the social vulnerability index. *Risk Anal.* **2008**, *28*, 1099–1114. [CrossRef]

20. Power, D.J.; Sharda, R.; Burstein, F. Decision Support Systems. In *Wiley Encyclopedia of Management*; Cooper, C.L., Ed.; John Wiley & Sons, Ltd.: Chichester, UK, 2015; pp. 1–4. ISBN 9781118785317.

21. Burstein, F.; Holsapple, C. *Handbook on Decision Support Systems 1: Basic Themes*; Springer: Berlin, Germany, 2008; ISBN 9783540487135.

22. Gorry, G.A.; Scott Morton, M.S. *A Framework for Management Information System*; Massachusetts Institute of Technology: Cambridge, MA, USA, 1971; pp. 458–470.

23. Agarwal, P.; Skupin, A. *Self-Organising Maps: Applications in Geographic Information Science*; John Wiley & Sons, Ltd.: Chichester, UK, 2008; ISBN 978-0-470-02167-5.

24. Kauko, T. Using the self-organising map to identify regularities across country-specific housing-market contexts. *Environ. Plan. B Plan. Des.* **2005**, *32*, 89–110. [CrossRef]

25. Kohonen, T. *Self-Organizing Maps*; Springer: Berlin, Germany, 1995; ISBN 978-3-540-62017-4.

26. O'Brien, P.W.; Mileti, D.S. Citizen participation in emergency response following the Loma Prieta earthquake. *Int. J. Mass Emerg. Disasters* **1992**, *10*, 71–89.

27. Hewitt, K. *Regions of Risk: A Geographical Introduction to Disasters*; Routledge: New York, NY, USA, 1997; ISBN 9781315844206.

28. Cutter, S.L.; Mitchell, J.T.; Scott, M.S. Revealing the vulnerability of people and places: A case study of georgetown county, South Carolina. *Ann. Assoc. Am. Geogr.* **2000**, *90*, 713–737. [CrossRef]

29. Ngo, E.B. When Disasters and Age Collide: Reviewing Vulnerability of the Elderly. *Nat. Hazards Rev.* **2001**, *2*, 80–89. [CrossRef]

30. Khan, S. Vulnerability assessments and their planning implications: A case study of the Hutt Valley, New Zealand. *Nat. Hazards* **2012**, *64*, 1587–1607. [CrossRef]

31. Fekete, A. Social vulnerability change assessment: Monitoring longitudinal demographic indicators of disaster risk in Germany from 2005 to 2015. *Nat. Hazards* **2019**, *95*, 585–614. [CrossRef]

32. Wu, C.C.; Jhan, H.T.; Ting, K.H.; Tsai, H.C.; Lee, M.T.; Hsu, T.W.; Liu, W.H. Application of social vulnerability indicators to climate change for the southwest coastal areas of Taiwan. *Sustainability* **2016**, *8*, 1270. [CrossRef]

33. Fekete, A. Social Vulnerability (Re-) Assessment in Context to Natural Hazards: Review of the Usefulness of the Spatial Indicator Approach and Investigations of Validation Demands. *Int. J. Disaster Risk Sci.* **2019**, *10*, 220–232. [CrossRef]

34. Blaikie, P.; Cannon, T.; Davis, I.; Wisner, B. *At Risk: Natural Hazards, People's Vulnerability and Disasters*; Routledge: London, UK, 2014.

35. Enarson, E.P.; Morrow, B.H. *The Gendered Terrain of Disaster: Through Women's Eyes*; Praeger: London, UK, 1998; ISBN 0275961109.

36. Enarson, E.P.; Scanlon, J. Gender Patterns in Flood Evacuation: A Case Study in Canada's Red River Valley. *Appl. Behav. Sci. Rev.* **1999**, *7*, 103–124. [CrossRef]

37. Fothergill, A. Gender, Risk, and Disaster. *Int. J. Mass Emerg. Disasters* **1996**, *14*, 33–56.

38. Morrow, B.H.; Phillips, B. What's Gender 'Got to Do With It'? *Int. J. Mass Emerg. Disasters* **1999**, *17*, 5–11.

39. Peacock, W.G.; Morrow, B.H.; Gladwin, H. *Hurricane Andrew: Ethnicity, Gender, and the Sociology of Disasters*; Routledge: Hoboken, NJ, USA, 1997; ISBN 0415168112.

40. Dwyer, A.; Zoppou, C.; Nielsen, O.; Day, S.; Roberts, S. *Quantifying Social Vulnerability: A Methodology for Identifying Those at Risk to Natural Hazards*; Record 200; Goescience Australia: Camberra, Australia, 2014; ISBN 1-920871-09-8.

41. Fischer, A.P.; Frazier, T.G. Social Vulnerability to Climate Change in Temperate Forest Areas: New Measures of Exposure, Sensitivity, and Adaptive Capacity. *Ann. Am. Assoc. Geogr.* **2018**, *108*, 658–678. [CrossRef]

42. Bolin, R.C.; Stanford, L. *The Northridge Earthquake: Vulnerability and Disaster*; Routledge: London, UK, 1998; ISBN 9780203028070.

43. Heinz Center for Science Economics and the Environment. *The Hidden Costs of Coastal Hazards: Implications for Risk Assessment and Mitigation*; Island Press: Washington, DC, USA, 2000; ISBN 1559637560.

44. Colburn, L.L.; Jepson, M.; Weng, C.; Seara, T.; Weiss, J.; Hare, J.A. Indicators of climate change and social vulnerability in fishing dependent communities along the Eastern and Gulf Coasts of the United States. *Mar. Policy* **2016**, *74*, 323–333. [CrossRef]

45. Morrow, B.H. Identifying and mapping community vulnerability. *Disasters* **1999**, *23*, 1–18. [CrossRef]

46. Rubayet, K.R.; Lourenco, J.M.; Viegas, J.M. Perceptions of Pedestrians and Shopkeepers in European Medium-Sized Cities: Study of Guimaraes, Portugal. *J. Urban Plan. Dev.* **2012**, *138*, 26–34. [CrossRef]

47. Burton, I.; Kates, R.W.; Robert, W.; White, G.F. *The Environment as Hazard*; Guilford Press: New York, NY, USA, 1993; ISBN 9780898621594.

48. Combes, P.; Gaillard, M.C.; Pellet, J.; Demongeot, J. A score for measurement of the role of social vulnerability in decisions on abortion. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **2004**, *117*, 93–101. [CrossRef] [PubMed]

49. Mileti, D. *Disasters by Design: A Reassessment of Natural Hazards in the United States*; Joseph Henry Press: Washington, DC, USA, 1999; ISBN 978-0-309-26173-9.

50. De Oliveira Mendes, J.M. Social vulnerability indexes as planning tools: Beyond the preparedness paradigm. *J. Risk Res.* **2009**, *12*, 43–58. [CrossRef]

51. Drabek, T.E. *Disaster Evacuation Behavior: Tourists and Other Transients*; Institute of Behavioral Science, University of Colorado: Boulder, CO, USA, 1996; ISBN 9781877943133.

52. Hewitt, K. Safe place or 'catastrophic society'? Perspectives on hazards and disasters in Canada. *Can. Geogr./Le Géographe Can.* **2000**, *44*, 325–341. [CrossRef]

53. Tobin, G.A.; Ollenburger, J.C. *Natural Hazards and the Eldery*; University of Colorado, Natural Hazards Research and Applications Information Center: Boulder, CO, USA, 1993.

54. Cova, T.J.; Church, R.L. Modelling community evacuation vulnerability using GIS. *Int. J. Geogr. Inf. Sci.* **1997**, *11*, 763–784. [CrossRef]

55. Mitchell, J.K. *Crucibles of Hazard: Mega-Cities and Disasters in Transition*; United Nations University Press: Tokyo, Japan, 1999; ISBN 9280809873.

56. Platt, R.H. Lifelines: An Emergency Management Priority for the United States in the 1990s. *Disasters* **1995**, *15*, 172–176. [CrossRef]

57. Holand, I.S. Lifeline Issue in Social Vulnerability Indexing: A Review of Indicators and Discussion of Indicator Application. *Nat. Hazards Rev.* **2015**, *16*, 1–12. [CrossRef]

58. Jiménez, C.E.; Calmaestra, J.A.N.; Clemente, J.D.; Rego, R.A.G. *Vulnerabiliad del Tejido Social de Los Barrios Desfavorecidos de Andalucía. Análisis y potenciales*; Centro de Estudios Andaluces, Consejería de la Presidencia, Junta de Andalucía: Sevilla, Spain, 2008; ISBN 9788469144060.

59. Power, D.J. *Decision Support Systems: Concepts and Resources for Managers*; Quorum Books: London, UK, 2002; ISBN 156720497X.

60. Ayeni, B. The design of spatial decision support systems in urban and regional planning. In *Decision Support System in Urban Planning*; Timmermans, H., Ed.; Taylor and Francis: Abingdon, UK, 1997; pp. 3–15.

61. Negash, S.; Gray, P. Business Intelligence. In *Handbook on Decision Support Systems 2: Variatio*; Burstein, F., Holsapple, C., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 175–193.

62. Kohli, R.; Piontek, F. DSS in Healthcare: Advances and Opportunities. In *Handbook for Decision Support Systems 2*; Burstein, F., Holsapple, C., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 483–497.

63. Kek, A.G.H.; Cheu, R.L.; Meng, Q.; Fung, C.H. A decision support system for vehicle relocation operations in carsharing systems. *Transp. Res. Part E Logist. Transp. Rev.* **2009**, *45*, 149–158. [CrossRef]

64. Arampatzis, G.; Kiranoudis, C.T.; Scaloubacas, P.; Assimacopoulos, D. A GIS-based decision support system for planning urban transportation policies. *Eur. J. Oper. Res.* **2004**, *152*, 465–475. [CrossRef]

65. Jarupathirun, S.; Zahedi, F. GIS as Spatial Decision Support Systems. In *Geographic Information Systems in Business*; Pick, J.B., Ed.; Idea Group Pub: Hershey, PA, USA, 2005; ISBN 9781591404019.

66. Yeh, A.G.-O. Urban planning and GIS. In *Geographical information systems: principles, techniques, applications and management*; Long Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W., Eds.; Wiley: New York, NY, USA, 2005.

67. Goodchild, M.F. Two decades on: Critical GIScience since 1993. *Can. Geogr.* **2015**, *59*, 3–11. [CrossRef]

68. Sui, D.; Goodchild, M. The convergence of GIS and social media: Challenges for GIScience. *Int. J. Geogr. Inf. Sci.* **2011**, *25*, 1737–1748. [CrossRef]

69. Kwan, M.P.; Neutens, T. Space-time research in GIScience. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 851–854. [CrossRef]

70. National Research Council. *Learning to Think Spatially*; National Academies Press: Washington, DC, USA, 2006; ISBN 978-0-309-09208-1.

71. Keen, P.G.W. Decision support systems: The next decade. *Decis. Support Syst.* **1987**, *3*, 253–265. [CrossRef]

72. Agarwal, P. Ontological considerations in GIScience. *Int. J. Geogr. Inf. Sci.* **2005**, *19*, 501–536. [CrossRef]

73. Elwood, S. Geographic Information Science: New geovisualization technologies - Emerging questions and linkages with GIScience research. *Prog. Hum. Geogr.* **2009**, *33*, 256–263. [CrossRef]

74. Leszczynski, A. Rematerializing GIScience. *Environ. Plan. D Soc. Space* **2009**, *27*, 609–615. [CrossRef]

75. Yang, C.; Raskin, R.; Goodchild, M.; Gahegan, M. Geospatial Cyberinfrastructure: Past, present and future. *Comput. Environ. Urban Syst.* **2010**, *34*, 264–277. [CrossRef]

76. Goodchild, M.F. Geographical Information Science. *Int. J. Geogr. Inf. Sci.* **1992**, *6*, 31–45. [CrossRef]

77. Buzai, G.D.; Cacace, G.; Humacata, L.; Lanzelotti, S.L. *Teoría y Métodos de la Geografía Cuantitativa: Libro 1: Por una Geografía de lo Real*; MCA Libros: Buenos Aires, Argentina, 2015; ISBN 9789874598622.

78. Golfarelli, M.; Rizzi, S. *Datawarehouse design. Modern Principles and Methodologies*; Tata McGraw Hill Education Private Limited: Bologna, Italy, 2009; ISBN 978-0-07-067752-4.

79. Cao, L. Introduction to domain driven data mining. In *Data Mining for Business Applications*; Cao, L., Philip, S.Y., Zhang, C., Zhang, H., Eds.; Springer: Dordrecht, The Netherlands, 2009; pp. 3–10. ISBN 9780387794198.

80. Abarca-Alvarez, F.J.; Campos-Sánchez, F.S.; Reinoso-Bellido, R. Methodology of Decision Support through GIS and Artificial Intelligence: Implementation for Demographic Characterization of Andalusia based on Dwelling. *Estoa* **2017**, *6*, 33–51. [CrossRef]

81. Koskela, T.; Varsta, M.; Heikkonen, J.; Kaski, K. Temporal Sequence Processing using Recurrent SOM. *Proc. Knowl.-Based Intell. Electron. Syst.* **1998**, *1*, 1689–1699. [CrossRef]

82. Ritter, H.; Kohonen, T. Self-organizing semantic maps. *Biol. Cybern.* **1989**, *61*, 241–254. [CrossRef]

83. Kaski, S.; Kohonen, T. Exploratory Data Analysis By The Self-Organizing Map: Structures Of Welfare And Poverty In The World. In Proceedings of the Third International Conference on Neural Networks in the Capital Markets, London, UK, 11–13 October 1995; pp. 498–507.

84. Villmann, T.; Merényi, E.; Hammer, B. Neural maps in remote sensing image analysis. *Neural Netw.* **2003**, *16*, 389–403. [CrossRef]

85. Tayebi, M.H.; Hashemi Tangestani, M.; Vincent, R.K. Alteration mineral mapping with ASTER data by integration of coded spectral ratio imaging and SOM neural network model. *Turk. J. Earth Sci.* **2014**, *23*, 627–644. [CrossRef]

86. Yan, J.; Thill, J.-C. Visual data mining in spatial interaction analysis with self-organizing maps. *Environ. Plan. B Plan. Des.* **2009**, *36*, 466–486. [CrossRef]

87. Campos-Sánchez, F.S.; Abarca-Álvarez, F.J.; Serra-Coch, G.; Chastel, C. Evaluación comparativa del nivel de Desarrollo Orientado al Transporte (DOT) en torno a nodos de transporte de grandes ciudades: Métodos complementarios de ayuda a la decisión. *EURE. Rev. Latinoam. Estud. Urbanos Reg.* **2019**, *45*, 5–30. [CrossRef]

88. Faggiano, L.; de Zwart, D.; García-Berthou, E.; Lek, S.; Gevrey, M. Patterning ecological risk of pesticide contamination at the river basin scale. *Sci. Total Environ.* **2010**, *408*, 2319–2326. [CrossRef]

89. Gomes, H.; Ribeiro, A.B.; Lobo, V. Location model for CCA-treated wood waste remediation units using GIS and clustering methods. *Environ. Model. Softw.* **2007**, *22*, 1788–1795. [CrossRef]

90. Yang, C.; Guo, R.; Wu, Z.; Zhou, K.; Yue, Q. Spatial extraction model for soil environmental quality of anomalous areas in a geographic scale. *Environ. Sci. Pollut. Res.* **2014**, *21*, 2697–2705. [CrossRef]

91. Basara, H.G.; Yuan, M. Community health assessment using self-organizing maps and geographic information systems. *Int. J. Health Geogr.* **2008**, *7*, 67. [CrossRef]

92. Delmelle, E.C.; Thill, J.C.; Furuseth, O.; Ludden, T. Trajectories of Multidimensional Neighbourhood Quality of Life Change. *Urban Stud.* **2012**, *50*, 923–941. [CrossRef]

93. Skupin, A.; Hagelman, R. Visualizing Demographic Trajectories with Self Organizing Maps. *Geoinformatica* **2005**, *9*, 159–179. [CrossRef]

94. Guo, D.; Chen, J.; MacEachren, A.M.; Liao, K. A Visualization System for Space-Time and Multivariate Patterns (VIS-STAMP). *IEEE Trans. Vis. Comput. Graph.* **2006**, *12*, 1461–1474. [CrossRef] [PubMed]

95. Skupin, A.; Esperbé, A. An alternative map of the United States based on an n-dimensional model of geographic space. *J. Vis. Lang. Comput.* **2011**, *22*, 290–304. [CrossRef]

96. Hatzichristos, T. Delineation of demographic regions with GIS and computational intelligence. *Environ. Plan. B Plan. Des.* **2004**, *31*, 39–49. [CrossRef]

97. Abarca-Alvarez, F.J.; Navarro-Ligero, M.L.; Valenzuela-Montes, L.M.; Campos-Sánchez, F.S. European Strategies for Adaptation to Climate Change With the Mayors Adapt Initiative by Self-Organizing Maps. *Appl. Sci.* **2019**, *9*, 3859. [CrossRef]

98. Spielmans, S.E.; Thill, J.-C. Social area analysisss, data mining, and GIS. *Comput. Environ. Urban Syst.* **2008**, *32*, 110–122. [CrossRef]

99. Behnisch, M.; Ultsch, A. Urban data-mining: Spatiotemporal exploration of multidimensional data. *Build. Res. Inf.* **2009**, *37*, 520–532. [CrossRef]

100. Abarca-Alvarez, F.J.; Osuna-Pérez, F. Cartografías semánticas mediante redes neuronales: Los mapas auto-organizados (SOM) como representación de patrones y campos. *EGA. Rev. Expresión Gráfica Arquit.* **2013**, *18*, 154–163. [CrossRef]

101. Diappi, L.; Bolchim, P.; Buscema, M. Improved Understanding of Urban Sprawl Using Neural Networks. In *Recent Advances in Design and Decision Support Systems in Architecture and Urban Planning*; Van-Leeuwen, J.P., Timmermans, H.J.P., Eds.; Politecn Milan, Dept Architecture and Planning: Milan, Italy, 2004; pp. 33–49. ISBN 1-4020-2408-8.

102. Hamaina, R.; Leduc, T.; Moreau, G. Towards Urban Fabrics Characterization based on Buildings Footprints. In *Bridging the Geographic Information Sciences*; Gensel, J., Josselin, D., Vandenbroucke, D., Eds.; Springer: Berlin, Germany, 2012; pp. 231–248. ISBN 978-3-642-29063-3.

103. Salah, M.; Trinder, J.; Shaker, A. Evaluation of the self-organizing map classifier for building detection from lidar data and multispectral aerial images. *J. Spat. Sci.* **2009**, *54*, 15–34. [CrossRef]

104. Skupin, A.; Agarwal, P. Introduction: What is a Self-Organizing Map? In *Self-Organising Maps: Applications in Geographic Information Science*; Agarwal, P., Skupin, A., Eds.; Wiley: Chichester, UK, 2008; pp. 1–20. ISBN 0470021675.

105. Bação, F.; Lobo, V.; Painho, M. Self-organizing maps as substitutes for k-means clustering. *Comput. Sci.* **2005**, *3516*, 476–483. [CrossRef]

106. Witten, I.H.; Frank, E.; Hall, M. *A Data Mining: Practical Machine Learning Tools and Techniques*; Elsevier Science & Technology Books: San Diego, CA, USA, 2011; ISBN 0080890369.

107. Hernández Orallo, J.; Ramírez Quintana, M.J.; Ferri Ramírez, C. *Introducción a la Minería de Datos*; Pearson Prentice Hall: Madrid, Spain, 2004; ISBN 8420540919.

108. Kass, G.V. An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Appl. Stat.* **1980**, *29*, 119–127. [CrossRef]

109. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Chapman & Hall: London, UK, 1984; ISBN 0-412-04841-8.

110. Quinlan, J.R. Induction of Decision Trees. *Mach. Learn.* **1986**, *1*, 81–106. [CrossRef]

111. Strobl, C.; Malley, J.; Tutz, G. An Introduction to Recursive Partitioning: Rationale, Application and Characteristics of Classification and Regression Trees, Bagging and Random Forests. *Psychol. Methods* **2010**, *14*, 323–348. [CrossRef] [PubMed]

112. Strasser, H.; Weber, C. On the Asymptotic Theory of Permutation Statistics. *Math. Methods Stat.* **1999**, *8*, 220–250. [CrossRef]

113. Astudillo, C.A.; John Oommen, B. Imposing tree-based topologies onto self organizing maps. *Inf. Sci. (Ny).* **2011**, *181*, 3798–3815. [CrossRef]

114. Astudillo, C.A.; Oommen, B.J. On achieving semi-supervised pattern recognition by utilizing tree-based SOMs. *Pattern Recognit.* **2013**, *46*, 293–304. [CrossRef]

115. Yang, Z.R.; Chou, K.-C. Mining biological data using self-organizing map. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1748–1753. [CrossRef]

116. Gómez-Carracedo, M.P.; Andrade, J.M.; Carrera, G.V.S.M.; Aires-de-Sousa, J.; Carlosena, A.; Prada, D. Combining Kohonen neural networks and variable selection by classification trees to cluster road soil samples. *Chemom. Intell. Lab. Syst.* **2010**, *102*, 20–34. [CrossRef]

117. Di Maio, F.; Rossetti, R.; Zio, E. Postprocessing of Accidental Scenarios by Semi-Supervised Self-Organizing Maps. *Sci. Technol. Nucl. Install.* **2017**. [CrossRef]

118. Tsai, C.-F.; Lin, Y.-C.; Wang, Y.-T. Discovering Stock Trading Preferences By Self-Organizing Maps and Decision Trees. *Int. J. Artif. Intell. Tools* **2009**, *18*, 603–611. [CrossRef]

119. Shanmuganathan, S.; Li, Y. An AI based approach to multiple census data analysis for feature selection. *J. Intell. Fuzzy Syst.* **2016**, *31*, 859–872. [CrossRef]

120. Silver, M.S. On the Design Features of Decision Support Systems: The Role of System Restrictiveness and Decisional Guidance. In *Handbook on Decision Support Systems 2: Variations*; Burstein, F., Holsapple, C., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 261–291.

121. Demartines, P.; Blayo, F. Kohonen Self-Organizing Maps: Is the Normalization Necessary? *Complex Syst.* **1992**, *6*, 105–123.

122. Weiss, S.M.; Indurkhya, N. *Predictive Data Mining: A Practical Guide*; Morgan Kaufmann: San Francisco, CA, USA, 1998; ISBN 1558604030.

123. Ketchen, D.J.; Shook, C.L. The Application Of Cluster Analysis In Strategic Management Reseach: An Anlysis and Critique. *Strateg. Manag. J.* **1996**, *17*, 441–458. [CrossRef]

124. Ball, G.H.; Hall, D.J. *A Novel Method of Data Analysis Andpattern Classification*; SRI International: Menlo Park, CA, USA, 1965.

125. Calinski, T.; Harabasz, J. A Dendrite Method for Cluster Analysis. *Commun. Stat.* **1974**, *3*, 1–27.

126. Davies, D.L.; Bouldin, D.W. A Cluster Separation Measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1979**, *PAMI-1*, 224–227. [CrossRef]

127. Lletí, R.; Ortiz, M.C.; Sarabia, L.A.; Sánchez, M.S. Selecting variables for k-means cluster analysis by using a genetic algorithm that optimises the silhouettes. *Anal. Chim. Acta* **2004**, *515*, 87–100. [CrossRef]

128. Hair, J.F., Jr.; Black, W.C.; Babin, B.J.; Anderson, R.E. *Multivariate Data Analysis*, 7th ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2009; ISBN 9780138132637.

129. Wu, P.K.; Hsiao, T.C. Factor Knowledge Mining Using the Techniques of AI Neural Networks and Self-Organizing Map. *Int. J. Distrib. Sens. Netw.* **2015**, *11*, 412418. [CrossRef]

130. Wasserstein, R.L.; Lazar, N.A. The ASA's statement on p-values: Context, process, and purpose. *Am. Stat.* **2016**, *70*, 129–133. [CrossRef]

131. Coe, R.; Merino, C. Magnitud del efecto: Una guía para investigadores y usuarios. *Rev. Psicol.* **2003**, *21*, 147–177.

132. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 1998; ISBN 0-8058-0283-5.

133. Sarlin, P. Exploiting the self-organizing financial stability map. *Front. Artif. Intell. Appl.* **2012**, *243*, 248–257. [CrossRef]

134. Hothorn, T.; Hornik, K.; Zeileis, A. Unbiased recursive partitioning: A conditional inference framework. *J. Comput. Graph. Stat.* **2006**, *15*, 651–674. [CrossRef]

135. Bação, F.; Lobo, V.; Painho, M. The Self-Organizing Map and it's variants as tools for geodemographical data analysis: The case of Lisbon's Metropolitan Area. *Comput. Geosci.* **1995**, *31*, 155–163. [CrossRef]

136. Skupin, A.; Hagelman, R. Attribute space visualization of demographic change. In Proceedings of the 11th ACM International Symposium on Advances in Geographic Information Systems, New Orleans, LA, USA, 7–8 November 2003; pp. 56–62. [CrossRef]

137. Streich, B. *Stadtplanung in der Wissensgesellschaft Ein Handbuch*; Verlag für Sozialwissenschaften: Wiesbaden, Germany, 2005, ISBN 9783663114802.