# An analysis of occupational accidents involving national and international construction workers in Spain using association rule technique

María Martínez-Rojas[a] * (0000-0002-7050-3629), José Manuel Soto-Hidalgo[b] (0000-0003-4412-5449), María Dolores Martínez-Aires[c] (0000-0002-9292-5048), Juan Carlos Rubio-Romero[a] (0000-0002-5122-7526)

*a University of Málaga, Department of Economics and Business Management, E.I Industriales, C/Dr. Ortiz Ramos, s/n (Teatinos), 29071 Málaga, Spain*

*b University of Granada, Department of Computer Architecture and Technology, University of Granada, Granada, Spain*

*c University of Granada, Department of Architectural Construction, E.U. Arquitectura Técnica, C/Severo Ochoa s/n, 18071 Granada, Spain*

email: mmrojas@uma.es, jmsoto@ugr.es, aires@ugr.es, juro@uma.es

*Corresponding author

# An analysis of occupational accidents involving national and international construction workers in Spain using association rule technique

Worker safety awareness on construction sites is a major concern due to the hazardous work conditions. Additionally, globalization is increasing the cultural diversity of the workforce and this influences workers' attitudes, beliefs and behaviour. The growing number of migrant workers in this sector has become a distinctive feature of the industry's labour market. The objective of this paper is to analyse occupational accidents that occurred on Spanish construction sites while taking into consideration the nationality of the workers. Due to the large number of accidents and attributes associated with them, the use of association rules is proposed. Overall, results evince similar behaviour, although interesting differences can be observed regarding the occupation of workers. In addition, the results are in accordance with previous studies carried out in other countries. The analysis of these accidents will serve to establish initiatives that provide safer work environments.

Keywords: health and safety; construction; association rules; accidents

## 1. Introduction

Construction is one of the most hazardous industries due to its complex socio-technical system [1- 5]. On the one hand, as a result of its dynamic and transitory nature, and on the other, because it is a labour-intensive industry [6]. According to the European Statistics on Accidents at Work, "more than one fifth of all fatal accidents at work in the EU-28 took place within the construction sector" [7].

Therefore, occupational management is an important issue [8, 9], in particular at the lower end of the labour market where the people in the workforce often come from different cultural backgrounds [10]. The International Labor Organization (ILO) estimates that 164 million people worldwide are migrant workers [11]. In the European Union, there are 36.9 million people that were born outside the EU-28 as of January 1, 2017, while 20.4 million were born in EU Member States other than where they reside [12]. In Spain, migrant workers represented 11.43% of the total workforce in 2018 [13], with approximately 7.5% of these belonging to the construction sector. It should be noted that this percentage tripled from

2005 to 2009, when the construction boom took place, reaching a percentage of 21.7% for all migrants [14].

In the literature, different ways of defining foreign worker have been found: ethnic minority, foreign and migrant. In this paper, the definition of the ILO, which defines migrant workers as people who leave home to find work outside their hometown or home country [11] is accepted. This paper focuses specifically on persons who move to another country for work reasons. These people are known as international migrant workers.

Research focusing on the relationship between the nationality of workers and safety has aroused great interest over time [15]. Among these proposals, some authors examine differences in occupational injury rates among national or non-national workers in different nationalities [16-19]. Other studies attempt to identify factors contributing to management-related safety problems [20-27], as for example, differences in language, culture, training, education and living habits.

Nevertheless, to the best of our knowledge, there are no studies in the literature that explore associations in occupational accidents for national and international workers occurring in Spain. Only limited research has been carried out in construction accidents in Spain. For example, Camino López et al. [28] study eighteen variables such as age, type of contract, time of accident, length of service in the company, company size, day of the week, etc. in order to analyse the influence of each of them with respect to the severity or indeed fatality of the accident. The authors conclude that different training was needed, depending on the severity of accidents, for different age, length of service in the company, organisation of work, and time when workers work. Similarly, López Arquillos et al. [29] analyse construction sector accidents in Spain between 2003 and 2008. To do this, the authors select ten variables and evaluate the influence of each variable with respect to the severity of the accident. They draw relevant conclusions regarding the following variables: size of company, the experience of workers and the place of the accidents. However, none of these papers focus on national and international workers in spite of the interest that this issue currently provokes in other countries. Recently, García-Arroyo and Osca Segovia [27] presented a research paper on construction accidents from a cultural perspective, by exploring differences in languages and the cultural gap between countries. The authors highlight that the studies are inconclusive, citing that this is probably due to data heterogeneity. In addition, they only consider data from 2015, remarking that it would be interesting to include other information that reflects the reality lacking in the official figures.

Most models of the incidence of occupational accidents in the construction industry include multiple factors [30-33]. Although statistical techniques can be used to infer cause-and-effect relationships among these factors, the large number of factors involved, and the complexity of the relationships make it difficult for managers to identify potential hazards [34]. Nowadays, the ability to manage large amounts of data is becoming a key issue in a knowledge-based society [35]. In the same way, the ability to extract knowledge from large datasets is becoming increasingly important for organisations. For this purpose, techniques that have been widely applied in other domains are attracting attention among researchers for the analysis of occupational accidents [36]. Specifically, data mining techniques are highly effective in exploring associations in large datasets, especially when they contain many variables.

Therefore, the main objective of this article is to explore data from the annual digital database of occupational accidents in Spain between 2003 and 2015 and to identify the strongest variables associated with both national and international construction workers. For this purpose, the association rule mining technique is applied, allowing intuitive knowledge expressed as linguistic statements (the meaning of the categorised variables in the domain) to be obtained, which would be useful for corrective and/or preventive actions.

After this introduction, the remainder of the paper is structured as follows. Section 2 is devoted to introducing some preliminary concepts concerning the factors contributing to occupational accidents and proposals using data mining in this research domain. Section 3 outlines the methodology in three steps. First, the data collection and selection of data is explained, the selection of variables is then detailed and finally the association rule mining technique is introduced. Section 4 presents and discusses the results obtained while 5 presents the conclusions and guidelines for future research.

## 2. Background

Apart from the inherently dangerous nature of construction work previously mentioned, some aspects contributing to occupational accidents have been analysed in the literature. In addition, data mining techniques are successfully applied in this scenario. In this regard, many research documents have focused on identifying the factors influencing the incidence of accidents in the construction sector and data mining approaches have been used to identify groups of data or relationships among them. Section 2.1 highlights some distinguishing factors in occupational accidents and section 2.2 introduces some previous studies that have applied data mining techniques in the construction context.

### *2.1 Factors contributing to occupational accidents*

In this section, some distinguishing factors are analysed and summarised in order to assist us in the decision-making process concerning the variables to be considered in our proposal. Usually, these factors are classified into categories that group together aspects of a similar nature, as for example: personnel, company, accident and project. Firstly, the personnel category includes attributes such as gender, age, nationality, work experience. Secondly, the company category includes attributes such as company size and code of activities. Thirdly, the accident category includes hour of day, hour of work, day of week, activity, deviation and place of accident. Finally, the project category contains financial, budgetary and duration of project factors. Table 1 presents documents and distinguished authors in this research domain, and the factors that they consider relevant in their studies.

| Research documents | Personnel | Company | Accident | Project |
|---|---|---|---|---|
| Chau et al. [37] | x | | x | |
| Gibb et al. [30] | x | x | x | |
| López et al. [28] | x | x | x | |
| Liao and Perng [38] | x | | x | |
| Chi et al. [39] | x | x | x | |
| Fernández-Muñiz et al. [40] | x | x | | |
| Cheng et al. [34] | x | x | x | |
| Rivas et al. [41] | x | | x | |
| López-Arquillos et al. [29] | x | x | | |
| Martínez-Rojas et al. [35] | x | x | | x |
| López-Arquillos et al. [42] | x | x | | |
| Stoilkovska et al. [43] | x | | | |
| Kang et al. [44] | | | x | x |
| Marín Ruiz et al. [45] | x | x | x | x |

| | | | | |
|---|---|---|---|---|
| Mohammadi et al. [1] | x | x | | x |
| Shao et al. [46] | | | x | |

Table 1: Variables grouped by categories

As can be seen in Table 1, personnel and accident categories have been identified as important when analysing cause-effect relationships in occupational accidents. Similarly, the company category has also been widely considered in the literature. However, it seems that factors related to the project category are less relevant or they are difficult to obtain. As a general conclusion, an analysis of existing literature reveals that factors tend to be similar in different countries [47].

Finally, some proposals identify other kinds of factors influencing safety performance on construction projects, although these are beyond the scope of this article. These include, for example, financial aspects, work pressure and culture [1, 26].

### *2.2. Data Mining*

Data mining explores knowledge from a large data set and transforms it into an understandable structure [48]. There are different approaches based on the objective to be achieved, as for example, to discover groups of data (e.g., cluster analysis), unusual data (e.g., anomaly detection), and relations among variables (e.g., association rule mining).

As mentioned in the Introduction section, the process of association rule analysis consists of the exploration of large amounts of data based on certain terms or variables with the aim of identifying patterns (or rules) that are hidden in the mass of data. This method has been successfully used in a variety of research domains, such as market basket analysis [49], customer relationship [50], mining sector [51] and medical [52]. Similarly, the association-rule technique has also been applied to diverse problems in the construction management domain. Examples include energy sustainability [53, 54], post project reviews [55], construction defects [56] and building performance [57]. This section focuses on the review of proposals applying data mining to occupational safety analysis.

- Some studies carried out an analysis of occupational accident cases in Taiwan’s construction industry that had occurred in different periods between 1999-2009 [38, 47].
- Rivas et al. [41] evaluate diverse data-mining techniques (such as Bayesian networks, decision rules, classification trees, logistic regression, and support vector machines) to identify the major causes of accidents and develop predictive models.
- Ayhan and Tokdemir [32] propose a prediction model to prevent incidents on construction sites by analysing previous incidents.

The following proposals focus on the application of association rule mining in occupational safety.

- Cheng et al. [34] decided to use the association rule method of data mining due to the large number of factors involved and the complexity of the relationships between them. They perform an analysis of 1347 accidents in the Taiwan construction industry during the period 2000-2007.
- Amiri et al. [58] use multiple-correspondence analysis, decision tree, ensembles of decision trees and association rules methods to analyse a database of construction accidents in Iran between 2007 and 2011.
- Li et al. [59] use association rules to find a relationship between the contributing factors and non-helmet use behaviour.
- Shin et al. [60] discover intuitive knowledge expressed as association rules from a database of 98,189 serious injury and fatal accidents that occurred on Korean building construction sites in the period 2006-2010.

Most authors revealed that these techniques are more useful than classical statistical techniques in predicting and identifying the factors underlying accidents/incidents because they allow large amounts of data to be managed efficiently. As a result, these studies proposed broad recommendations such as improving inspection plans, training for workers, adherence to safety work procedures and the promotion of safety management.

## 3. Methodology

This section outlines the methodology that has been applied in this paper. Firstly, the data collection process and the selection of data are explained. Secondly, the selection of variables is described and, finally, association rule mining is introduced.

### *3.1. Data Collection and Selection of Data*

"Accident at work" is defined for the European Statistics on Accidents at Work (ESAW) as a discrete occurrence in the course of work that leads to physical or mental harm [7]. The phrase "in the course of work" means "while engaged in an occupational activity or during the time spent at work". Spanish Legislation defines it as "Any bodily injury suffered by a worker during or as a consequence of the work he/she performs for others" [61].

Since 2003 in Spain, the Ministry of Work, Migration and Social Security [62] must be notified of all accidents resulting in one or more days off work, which is compulsory by Spanish Law [63]. The notifications must be sent through the electronic system "DELT@" and involves the completion of an official workplace incident notification form. The Ministry of Work, Migration and Social Security [62] provided us with the anonymised data of all workplace accidents in Spain during the period 2003-2015. Each accident is identified by 58 attributes using the methodology from the third edition of ESAW [12]. Some examples of these attributes are worker age and nationality, day of the accident, etc. These attributes will be named as variables in our mining process.

The initial study population is comprised of 5,495,609 instances of occupational accidents recorded during the period mentioned. Since this study only considers accidents occurring during construction activities, the first objective is to reduce the dataset to those activities. Figure 1 shows the flowchart of the methodology that is explained below.

Figure 1: The flowchart of the proposed methodology

Data pre-processing is a critical step in the analysis process, and it has a direct impact on the success of data mining techniques. This step includes cleaning incomplete and noisy data, filtering desired data, reducing the number of variables and categorising variables. For this purpose, the KoNstanz Information MinEr (KNIME) software [64] has been applied because it allows large amounts of data to be managed and different filters to be applied in an easy, intuitive way.

The different filters that have been applied in order to obtain construction accidents are detailed below:

- To split the data contained in the variable related to the date of an accident since the format is day/month/year in the same cell. Specifically, extracting the year into a separated field is necessary due to a change of codification in the variable concerning occupation beginning in 2011.
- To filter construction accident data based on the worker's occupation at the time of the accident. The occupation of the workers is stated using the National Occupation Code (CNO-94). In this step, the change in codification of this variable beginning in 2011 has been taken into account. After this process, 1,525,865 accidents are retrieved, which represents 27.77% of the total. In the following section, in Table 2, the selected occupations are provided.
- To standardise the occupation variable due to the aforementioned codification change.

As a consequence of this filtering process, our study focuses on a sample of 1,525,865 accidents taken from the total number of accidents reported in Spain between 2003 and 2015.

Next, the dataset is split into national and international accidents and the association rule technique is applied. As a result of this process, 1,280,495 (83.92%) and 245,370 (16.08%) accidents were found corresponding to national and international workers, respectively.

The next section outlines the selected variables and its categorization.

### *3.2. Selection of Variables*

As mentioned in the previous section, in Spain, all accidents must be notified through the electronic system DELT@. This process involves the completion of an official workplace incident notification form that contains 58 variables. As a first step, it will be necessary to select those that are of interest for the study.

The selection process considers two main criteria, such as relevant published results on this topic and our previous experience and analysis. For this purpose, the variables in Section 2 that have been identified by the authors as relevant in contributing to occupational accidents were presented. An analysis was also performed that was designed to gain an overall understanding of the variables in our datasets regarding occupational accidents in national (N) and international (I) workers.

The variables identified as relevant in the Literature (see Table 1) were grouped into four categories: personnel, company, accident and project. In our proposal, variables from all categories are included, except the project category, as this kind of information does not appear in the aforementioned notification form. Other variables that could represent an interest for this study were initially identified. These include, for example employment status, type of employment contract, contractor or subcontractor or habitual work. However, after a statistical analysis to explore the behaviour of all variables, these were discarded because they would not provide relevant information for the analysis.

The variables considered were categorized into ranges (or groups), mostly pre-defined by the ESAW system [12]. Next, the selected variables are detailed along with their classes or categories. Firstly, Table 2 details the variables defining characteristics of the worker and company involved in the accident: age of the injured worker at the time of the accident, occupation, experience (in months) of the injured worker and the number of employees in the company.

| Description | Categorization of Variables |
|---|---|
| **Age:** age of the victim at the time of the accident | **A1** (below 20 years), **A2** (21-30 years), **A3** (31-40 years), **A4** (41-50 years), **A5** (51-60 years), **A6** (over 60 years) |
| **Occupation:** victim's occupation at the time of the accident | **O1** (Managers and team leaders), **O2** (Heads of workshop and heads of workers of finishing works), **O3** (Heads of Painters), **O4** (Reinforced concrete workers), **O5** (Bricklayers and related works), **O6** (Carpenters), **O7** (Other structural construction workers), **O8** (Plasterers), **O9** (Plumbers and pipe fitters), **O10** (Floor layers and tile setters), **O11** (Painters and related works), **O12** (Building structure cleaners), **O13** (Air |

| | |
|---|---|
| | conditioning and refrigeration mechanics), **O14** (Other installers), **O15** (Electricians), **O16** (Construction Labourer) |
| **Length of service in the company** | **L1** (Less than 1 month), **L2** (1-2 months), **L3** (2-6 months), **L4** (6 months-1 year), **L5** (1-3 years), **L6** (3-5 years), **L7** (5-10 years), **L8** (10-20 years), **L9** (more than 20 years) |
| **Size of enterprise:** number of employees working at the local unit of the workplace | **S1** (1-9 employees), **S2** (10-49 employees), **S3** (50-249 employees), **S4** (250-499 employees), **S5** (500 employees or more) |

Table 2: Categories of worker variables

Secondly, variables related to the accident itself are presented in Table 3. Once the data pre-processing and filtering step and the selection of variables have been detailed, the next section focuses on explaining association rule mining.

| Description | Categorization of Variables |
|---|---|
| **Workplace:** place of the accident | **WA1** (Habitual Workplace), **WA2** (Non-habitual workplace), **WA3** (On the way from worksite to worksite), **WA4** (Going to the worksite or back home) |
| **Day of the week:** day of the week when the accident occurred | **DW1** (Monday), **DW2** (Tuesday), **DW3** (Wednesday), **DW4** (Thursday), **DW5** (Friday), **DW6** (Saturday), **DW7** (Sunday) |
| **Hour of the day:** hour of the day when the accident occurred | **HD1** (Early morning, from 0:00 a.m. until 7:59 a.m.), **HD2** (First 2 hours of the morning, from 8:00 a.m. until 9:59 a.m.), **HD3** (Midmorning, from 10:00 a.m. until 11:59 a.m.), **HD4** (Late morning, from 12:00 p.m. until 13:59 p.m.), **HD5** (Afternoon, from 14:00 p.m. until 15:59 p.m.), **HD6** (Evening, from 16:00 p.m. until 19:59 p.m.), **HD7** (Night, from 20:00 p.m. until 23:59 p.m.) |
| **Hour of the work:** hour of the work when the accident occurred | **HW1** (First working hour), **HW2** (Second working hour), … |

| | |
|---|---|
| **Risk Assessment:** If there is a risk assessment of the work in which the accident occurred | **R1** (Yes), **R2** (No) |
| **Working Process:** main type of work or task (general activity) being performed by the victim at the time of the accident. | **WP0** (No information), **WP1** (Production, manufacturing, processing, storing - All types - Not specified), **WP2** (Excavation, Construction, Repair, Demolition - Not specified), **WP3** (Agricultural type work, forestry, horticulture, fish farming, work with live animals - Not specified), **WP4** (Service provided to enterprise and/or to the general public; intellectual activity - Not specified), **WP5** (Other work related to tasks coded under 10, 20, 30 and 40 - Not specified), **WP6** (Movement, sport, artistic activity - Not specified), **WP9** (Other Working Processes not listed in the above classification) |
| **Physical Activity**: the activity being performed by the injured worker just before the accident | **PA0** (No information), **PA1** (Operating machine - Not specified), **PA2** (Working with hand-held tools - Not specified), **PA3** (Driving/being on board a means of transport or handling equipment - Not specified), **PA4** (Handling of objects - Not specified), **PA5** (Carrying by hand - Not specified), **PA6** (Movement - Not specified), **PA7** (Presence - Not specified), **PA8** (Other Specific Physical Activities not listed in this classification) |
| **Deviation**: a description of the way in which the circumstances of the accident differed from normal practice | **D0** (No information), **D1** (Deviation due to electrical problems, explosion, fire - Not specified), **D2** (Deviation by overflow, overturn, leak, flow, vaporisation, emission - Not specified), **D3** (Breakage, bursting, splitting, slipping, fall, collapse of Material Agent - Not specified), **D4** (Loss of control (total or partial) of machine, means of transport or handling equipment, hand-held tool, object, animal - Not specified), **D5** (Slipping - Stumbling and falling - Fall of persons - Not specified), **D6** (Body movement without any physical stress (generally leading to an external injury) - Not specified), **D7** (Body movement under or with physical stress (generally leading to an internal injury) - Not specified), **D8** (Shock, fright, violence, aggression, threat, presence - Not specified), **D9** (Other Deviations not listed above in this classification.) |

| | |
|---|---|
| **Contact:** the contact that injured the victim | **FC1** (Contact with electrical voltage, temperature, hazardous substances), **FC2** (Drowned, buried, enveloped), **FC3** (Horizontal or vertical impact with or against a stationary object (the victim is in motion)), **FC4** (Struck by object in motion, collision with), **FC5** (Contact with sharp, pointed, rough, coarse Material Agent), **FC6** (Trapped, crushed, etc.), **FC7** (Physical or mental stress), **FC8** (Biting, kicking, etc.), **FC9** (Other Contacts - Modes of Injury not listed) |
| **Type of injury:** physical consequences for the victim | **TI1** (Unknown injury), **TI2** (Wounds and superficial injuries), **TI3** (Bone Fractures), **TI4** (Dislocations, sprains and strains), **TI5** (Traumatic amputations), **TI6** (Concussions and internal injuries), **TI7** (Burns, scalds and frostbite), **TI8** (Poisonings and infections), **TI9** (Drownings and asphyxiations), **TI10** (Effects of sound, vibration and pressure), **TI11** (Effects of temperature extremes, light and radiation), **TI12** (Traumatic shocks), **TI13** (Multiple injuries), **TI14** (Heart attacks, strokes and other non-traumatic pathologies), **TI15** (Other specified injuries not included under other headings) |

Table 3: Categories of accident variables

### *3.3. Association rule mining*

As mentioned before, the association rules are widely used to study relationships of variables from large databases in depth, and to explore potential associations which occur mutually in a given data set.

Generally, given a set of items I and two itemsets A, B being disjoint subsets of I, and given a multiset of transactions T, each transaction being also a subset of I, a standard association rule is expressed in the form of $A \Rightarrow B$, where A is the antecedent and B is the consequent. Such rule means that every transaction in T containing A, contains B. For example, the following is a simple association rule related with male gender (M) extracted from the construction accident database:

$$\{Age < 40, Gender = M\} \Rightarrow \{Accident = Fatal\}, \quad (1)$$

which is also commonly expressed as a logical rule in the form Age < 40, Gender = M $\Rightarrow$ Accident = Fatal. This example rule indicates that the worker who is less than 40 years old and the gender is Male, is more likely to suffer a fatal accident.

Typical quality measures for association rules are support, confidence, and lift [60]. Support (S), defined as

$$S(A \Rightarrow B) = P(A \cup B), \qquad (2)$$

where P means probability, represents the probability that both itemsets A and B occur simultaneously in a transaction. Support is symmetric. Therefore, the support of rule A ⇒ B is equivalent to the support of B ⇒ A.

Confidence (C), defined as

$$C(A \Rightarrow B) = P(B \mid A) = \frac{P(A \cup B)}{P(A)}, \qquad (3)$$

represents the conditional probability that B is in a transaction where A is. It is not symmetric. Therefore, the confidence of the rule A⇒ B may be different from the confidence of the rule B ⇒A.

Lift (L), defined as

$$L(A \Rightarrow B) = \frac{Confidence}{P(B)} = \frac{P(B|A)}{P(B)} = \frac{P(A \cup B)}{P(A)P(B)}, \qquad (4)$$

measures how many times more often A and B occur together in a transaction than expected if their occurrences were statistically independent.

Following, how to interpret this measure is detailed:

- L = 1 indicates no correlation between antecedent and consequent.
- L > 1 indicates positive correlation between antecedent and consequent.
- L < 1 indicates negative correlation between antecedent and consequent.

The Apriori algorithm [65] is one of the most commonly used method for the mining of association rules. This algorithm divides a rule mining process into two steps: firstly, the database is analysed to find all the itemsets with support values above the predefined minimum; secondly, a rule is generated if it satisfies the predefined minimum confidence. An implementation of the Apriori algorithm in the R programming language [66] has been considered.

## 4. Results

This section analyses the national and international datasets from two different perspectives. On the one hand, statistical results to gain an overall understanding of variables in occupational construction accidents are explored. In addition, the Apriori algorithm is applied to examine the relationships of variables in the form of association rules.

Both the statistical and association rules results are obtained by an experiment implemented in the R programming language and a free software environment [66]. This environment provides the required infrastructure to create and manipulate input datasets for the mining algorithms and for analysing the resulting itemsets and rules.

As described in section 3, this study analyses a total of 1,525,865 occupational construction accidents in accordance with the filtering process shown in Figure 1.

In order to discover the rules, minimum thresholds for support (S) and confidence (C) measures need to be specified. Numbers of association rules generated are inversely proportional to the threshold S and threshold C. Therefore, it depends on the user to establish the threshold values for pruning large numbers of association rules [36]. A comprehensive analysis considering a wide range of values both for the support and the confidence measures was carried out. Then, according to this analysis, in this study, the threshold S and C values have been empirically fixed to 4% and 80%, respectively. For the lift measure (L), no limit has been established initially, but rules with a higher lift (greater than one) values will be considered since they are stronger and more interesting.

Notice that only rules that meet all three thresholds are accepted as valid association rules. As a result, a total of 59 and 45 rules have been obtained for the national and international datasets, respectively. In Figure 2, the scatter plot of the two datasets displays values for support and confidence in x-axis and y-axis, respectively. Additionally, the lift measure is represented on the right of each plot by colour coding the points, with the darkest being the highest value of the lift measure. Similar results are obtained for both datasets. As can be observed, most rules have a confidence value higher than 84% and a support value between 4% and 8%. In addition, the positive correlation between the antecedent and the consequent can be guaranteed since the lift value is higher than 2 in most rules.

(a) 59 rules (b) 45 rules

Figure 2: Scatter plot for national (a) and international (b) datasets

After visualising the rules from a general point of view taking into consideration only the measures, the antecedents and consequents are explored. For this purpose, Figure 3 shows parallel coordinates for visualising multivariate data, such as association rules. The y-axis represents the variables that appear in the antecedent and the consequent while the x-axis represents the position of such variables in the antecedent. The arrow is used to indicate the consequent item. The width of the arrows represents support, and the intensity of the colour represents confidence. Some rules from each dataset (national and international) which represent examples of rules with the most relevant consequents have been selected as example to illustrate the meaning of Figure 3. Concretely a total of eight rules (Rule1, Rule2, ... Rule8) have been selected and highlighted with different colour dotted lines. These rules have also been represented in the Table 4 providing more detailed information of the rules. For example, Rule 3 corresponds to a rule in the national dataset represented in Fig.3 (a) with S = 0.04, C = 0.84 and L = 2.81 that would be interpreted as: The deviation "Body movement under or with physical stress (generally leading to an internal injury) (D7)" is most likely to have happened when the worker was carrying by hand (PA5) and he/she suffered physical or mental stress (FC7). In addition, in this rule example, there was a risk evaluation (R1).

(a) National (b)International

Figure 3: Parallel coordinates plot for national (a) and international (b) datasets

Note: A2= 21-30 years, A3= 31-40 years, D4= Loss of control of machine, means of transport or handling, D7= Body movement under or with physical stress, DW1= Monday, FC5= Contact with sharp, pointed, rough, coarse Material Agent, FC7= Contact: Physical or mental stress, HD2= First hour of the morning, from 8:00 a.m. until 9:59 a.m, HD3= Midmorning, from 10:00 a.m. until 11:59 a.m, HW2= Second working hour, HW4= fourth working hour, O5= Bricklayers and related works, O16= Construction Labourer, PA2= Working with hand-held tools, PA4= Handling of objects, PA5= Carrying by hand, R1= Risk Assessment, R2= No Risk Assessment, S1= 1-9 employees, S1= 10-49 employees, TI2= Wounds and superficial injuries, TI4= Dislocations, sprains and strains, WP2= Excavation, Construction, Repair, Demolition

| Figure | Antecedent | Consequent | Support | Confidence | Lift | Rule |
|---|---|---|---|---|---|---|
| Fig. 3(a) | HD3,D7,TI4 | FC7 | 0.05 | 0.86 | 2.51 | *Rule*1 |
| Fig. 3(a) | HW2,R1,WP2 | HD2 | 0.05 | 0.87 | 3.64 | *Rule*2 |
| Fig. 3(a) | R1,PA5,FC7 | D7 | 0.04 | 0.84 | 2.81 | *Rule*3 |
| Fig. 3(a) | HW4,TI4 | HD3 | 0.05 | 0.85 | 3.05 | *Rule*4 |
| Fig. 3(b) | D7,TI4,A3 | FC7 | 0.06 | 0.87 | 2.83 | *Rule*5 |
| Fig. 3(b) | HW2,WP2 | HD2 | 0.10 | 0.85 | 3.63 | *Rule*6 |
| Fig. 3(b) | WP2,PA5,FC7 | D7 | 0.05 | 0.86 | 3.06 | *Rule*7 |
| Fig. 3(b) | HW4,R1,WP2 | HD3 | 0.05 | 0.84 | 3.90 | *Rule*8 |

Table 4: Selected rules and their corresponding measures from Figure 3 for both national (Fig.3(a)) and international Fig.3(b)) datasets.

Note: A3= 31-40 years, D7= Body movement under or with physical stress, FC7= Contact: Physical or mental stress, HD2= First hour of the morning, from 8:00 a.m. until 9:59 a.m, HD3= Midmorning, from 10:00 a.m. until 11:59 a.m, HW2= Second working hour, HW4= fourth working hour, PA5= Carrying by hand, R1= Risk Assessment, TI4= Dislocations, sprains and strains, WP2= Excavation, Construction, Repair, Demolition

These two previous graphs allow the most frequent variables to be identified in a quick and intuitive way, both in the antecedent and in the consequent part. Regarding the consequents, as can be observed for both national and international datasets, the most frequent are related to the time of the accident, specifically, the first hour of the morning (HD2) and midmorning, to be precise (HD3). Another frequent consequent is related to the type of injury, specifically, wounds and superficial injuries (TI2). Finally, most occupational accidents involve a worker in movement under or with physical stress (D7 and FC7). As can be observed, all these variables corresponding to the consequent part refer to the accident itself and there are no general variables regarding the company or the workers themselves. However, with respect to antecedents, there are some differences between the national and international datasets, which will be explained later in detail.

Once the rules obtained have been visualised from a general perspective, in the next step the rules are analysed and summarised in depth. To do this, rules with higher values for the three measures (S, C and L) are presented. In addition, for the sake of simplicity, rules that contain at least three variables in the antecedent are presented since rules that contain fewer variables are represented in those that contain more variables.

Figure 4 summarises the rules where the contact variable acts as a consequent both in national and international datasets, specifically, FC7 (Physical over exertion or mental stress). Notice that the contact variable describes the precise way in which the departure from normal practice resulted in an accident. Figure 4 is divided into two parts. The variables that act as antecedents are illustrated on the left both for the national and international workers and on the right the consequent is represented, which is common for both the national and international datasets. Above each rule, the three measures are detailed for each one: support (S), confidence (C) and lift (L).

In order to facilitate an understanding of these figures, an example is provided below. This example corresponds to the first rule in Figure 4, where antecedents are equal for both datasets although the measures obtained are different. This rule should be interpreted as the consequent FC7 (Physical over exertion or mental stress) in a national and international worker is more likely to have happened when the following antecedents are present:

- workers made an unusual "Body movement under or with physical stress" (D7)
- during the "Excavation, Construction, Repair, Demolition" working process (WP2)
- causing them the following types of injury: "Dislocations, sprains and strains" (TI4)
- and there was a risk evaluation of the work in which the accident occurred (R1)

As mentioned before, this rule appears in both datasets, but the measures obtained differ slightly. As can be observed in Figure 4, the confidence is 86% for the national dataset and 87% for the international dataset. The Lift measure is higher in the international (2.85) than in the national (2.5) dataset.

Figure 4: Association rules for *FC*7 as consequent

Note: D7= Body movement under or with physical stress, FC7= Contact: Physical or mental stress, O5= Bricklayers and related works, PA4= Handling of objects, R1= Risk Assessment, S2= 10-49 employees, TI4= Dislocations, sprains and strains, WP2= Excavation, Construction, Repair, Demolition

When comparing national and international workers, similar variables appear in the antecedent part. There is a strong relation between working process, deviation and type of injury variables in both cases. However, in the national dataset, there is a variable, the size of the company, that differs from the international dataset: companies with between 10 and 49 employees (S2). This result is consistent with many studies that conclude that small construction companies have a higher risk of injury than large construction companies [67, 68].

Another interesting result is the difference regarding the occupation variable. In the national dataset, the occupation is "bricklayers and related works" (O5) while in the international dataset it is “construction labourer” (O16). The “construction labourer” is the worker who frequently has less training and qualifications [15, 69]. These results are also in accordance with studies carried out in other countries stating that international workers usually work in lower-paid and lower-skilled jobs and work in conditions that are less safe [70, 24, 71]. Generally, employers cannot provide appropriate safety measures and training to international workers, thereby exposing them to higher risks in the workplace compared to local workers [15]. These ideas are also reinforced based on others studies that highlight the importance of improving training programs for these workers [72, 23].

In the international dataset, a variable regarding Physical activity appears in the second rule as an antecedent, specifically, “Handling of objects” (PA4). This variable provides more information in relation to the activity that the worker was doing when the accident occurred. Notice that most variables refer to the accident itself rather than personal information about the worker or the company. Finally, as can be observed, these measures indicate a strong and interesting relationship between the antecedent and the consequent.

Figure 5 illustrates the rules where the type of injury (TI2) is the consequent. In this case, only one relevant rule is obtained for the national dataset while in the international dataset three rules have been identified. As can be seen in the figure, national workers suffer “wounds and superficial injuries” when they are in contact with “sharp, pointed, rough, material Agents” (FC5) notwithstanding the existence of a risk evaluation (R1). In the international dataset additional variables appear as antecedents, such as: the working process “Excavation, Construction, Repair, Demolition” (WP2), physical activity “‘Working with hand-held tools” (PA2), deviation “loss of control” (D4) and size of company “10-49 employees” (S2). Similar to the previous rule, the values for the measures indicate a strong relation between the antecedent and the consequent.

Figure 5: Association rules for *TI*2 as consequent

Note: D4= Loss of control (total or partial) of machine, means of transport or handling equipment, hand-held tool, object, animal, FC5= Contact with sharp, pointed, rough, coarse Material Agent, PA2= Working with hand-held tools, R1= Risk Assessment, S2= 10-49 employees, WP2= Excavation, Construction, Repair, Demolition

Figure 6 shows the rules where the first hour in the morning (HD2) is the consequent. There is a difference between the two datasets in the type of injury variable. National workers are more likely to suffer “Dislocations, sprains and strains” (TI4) while international workers suffer “wounds and superficial injuries” (TI2). In both datasets, accidents are more frequent during “Excavation, Construction, Repair, Demolition” (WP2) activities and during the second hour of work (HW2). Similar to the first rule, in the international dataset, the occupation variable (O16) appears as an antecedent in the international dataset. It is interesting that the accident occurs in the second hour of work, when the worker still has no signs of fatigue. It can also be noted that the day of week variable does not appear as an antecedent in any case. These two variables have been identified as relevant in other studies in the literature [73, 46].

Figure 6: Association rules for *HD*2 as consequent

Note: FC7= Contact: Physical or mental stress, HW2= Second working hour, O16= Construction Labourer, R1= Risk Assessment, TI2= Wounds and superficial injuries, TI4= Dislocations, sprains and strains, WP2= Excavation, Construction, Repair, Demolition

## 5. Conclusions

The construction sector is one of the most hazardous industries, presenting a huge number of accidents in the workplace. For this reason, industrialised countries have become aware of this situation over time and have developed policies to attempt to deal with this issue. In addition, the construction research community has reflected this concern through various lines of research, such as identifying factors that influence accidents, exploring differences in occupational injury rates, managing a safety climate, etc. Another research line that has motivated researchers is the analysis of occupational accidents among migrant and local workers. Globalisation has increased cultural diversity, and this can influence the attitudes, beliefs and behaviour of construction workers. After analysing the literature, it is observed that most studies focus on fatal rather than of non-fatal accidents. Additionally, there are no studies that analyse this interesting issue in Spain.

Therefore, the aim of this paper is to explore the role that national culture may play in occupational safety in the construction sector in Spain. To do this, all workplace accidents between 2003 and 2015 in Spain that have been notified through an official electronic system have been collected. After data pre-processing and filtering to make the data ready for analysis, the relevant variables based on reference results in the literature, our previous experience and statistical analysis have been selected.

To address this objective, a data mining technique based on association rules that is useful in identifying relations in a large amount of data has been applied. Specifically, in construction accidents from two datasets (national and international) where, in contrast with traditional methods, association rules were identified automatically after analysing the large amounts of data and validated by using interesting measures. The results of this research represent an advance in the Spanish construction domain in terms of understanding and managing information on workplace accidents, specifically among national and international workers.

From a general perspective, the association rules obtained from both datasets present similar behaviour in spite of the difference in the number of accidents (83.92% and 16.08% for national and international, respectively). In addition, the metrics to evaluate the rules support the proposition that results are promising and acceptable. Most rules obtain a value between 1.5 and 2.5 for lift measure, which indicates a positive correlation between antecedent and consequent variables. On the other hand, the confidence measure shows a strong association since the value is closer to 1. Finally, support is suitable for the datasets, which is diverse both in the variables and in the number of accidents for the national and international datasets.

From a deeper perspective, interesting results can be observed. As mentioned before, in spite of the difference in the number of accidents in both datasets, the frequent variables in the antecedent and the consequent are very similar. Nevertheless, some differences can be observed when analysing rules that share the same consequent from both datasets. For example, the national workers who frequently suffer an accident are better trained and qualified than international workers. This issue needs to be explored in depth in Spain, given the large number of research studies carried out in other countries and the contradictory results they present. This kind of study will allow action plans to be designed to minimise accident rates in general, and for international workers in particular.

An interesting outcome is that most of the variables that appear in the antecedent part refer to the accident itself. The information on the worker or the company does not seem to be so significant with regard to the accident. In our proposal, unlike traditional techniques, a large number of variables have been considered to extract associations using the association rule technique. Meanwhile, other studies focus on a smaller number of variables, mostly related to the worker. Therefore, the results of our study are promising since they allow us to define measures in relation to the work, irrespective of the company employing the worker itself.

As a general conclusion, a positive safety climate can motivate workers to comply with safety regulations and use safe work procedures. For this to happen in a multicultural environment, language and cultural barriers must be eliminated from the entire production chain to ensure that health and safety information is correctly transmitted. In addition, international workers, who are usually emotionally vulnerable, should feel that they are part of the occupational safety and health programs. This requires that leadership competencies must be defined for multicultural safety contexts so that managers can detect whether workplace risks could be made worse by the presence of international workers. This will provide them the opportunity to define focused and appropriate preventive measures. In this type of proactive safety culture, all employees will share a vision of safety and thus improve safety.

Finally, despite the legislative and economic efforts focused on reducing the number of accidents in small and medium-sized enterprises, it is still necessary to analyse the health and safety education and training provided to their workers, especially with regard to international workers. Concerning further work, the authors would suggest two main lines. On the one hand, the application of different algorithms to explore and to extract useful safety information for the construction sector considering the nationality of workers. For example, the use of clustering algorithms that allow variables to be grouped in different categories. On the other hand, conducting a qualitative study to explore training and education in the context of construction companies will be considered.

## References

[1] Amir Mohammadi, Mehdi Tavakolan, and Yahya Khosravi. Factors influencing safety performance on construction projects: A review. Safety Science, 109:382 – 397, 2018. ISSN 0925-7535.

[2] Martínez-Aires, María D., López-Alonso, M., and Martínez-Rojas, M., Building information modeling and safety management: A systematic review. Safety Science, 101:11 – 18, 2018. ISSN 0925-7535.

[3] Merchán, M. D. C. R., Gómez-de-Gabriel, J. M., Fernández-Madrigal, J. A., & López-Arquillos, A. (2020). Improving the prevention of fall from height in

construction sites through the combination of technologies. International journal of occupational safety and ergonomics, 1-18.

[4] Loosemore M. and Malouf. N., Safety training and positive safety attitude formation in the Australian construction industry. Safety Science, 113:233 – 243, 2019. ISSN 0925-7535.

[5] Sadia Yasmeen, Hong Liu, Yuxin Wu, and Baizhan Li. Physiological responses of acclimatized construction workers during different work patterns in a hot and humid subtropical area of china. Journal of Building Engineering, 30:101281, 2020. ISSN 2352-7102. DOI: https://doi.org/ 10.1016/j.jobe.2020.101281. URL http://www.sciencedirect.com/ science/article/pii/S2352710219318455.

[6] Martínez-Rojas, M., Marín, N., and Amparo Vila, M., The role of information technologies to address data handling in construction project management. Journal of Computing in Civil Engineering, 30 (4):04015064, 2016.

[7] Eurostat. European Statistics on Accidents at Work. Glosary: Accident at work, 2018. URL https://ec.europa.eu/eurostat/ statistics-explained/index.php/Glossary:Accident_at_work.

[8] Simukonda, W., Manu, P., Mahamadu, A. M., & Dziekonski, K. (2020). Occupational safety and health management in developing countries: a study of construction companies in Malawi. International journal of occupational safety and ergonomics, 26(2), 303-318. [9] Martínez-Rojas, M., Martín Antolín, R., Salguero-Caparrós, F., and Rubio-Romero, J.C., Management of construction safety and health plans based on automated content analysis. Automation in Construction, 120:103362, 2020. ISSN 0926-5805. DOI: https://doi.org/10.1016/j.autcon.2020.103362.

[10] Starren, A., Hornikx, J., and Luijters, K., Occupational safety in multicultural teams and organizations: A research agenda. Safety Science, 52:43 – 49, 2013. ISSN 0925-7535. Selected Papers from the 1st International Safety and Crisis Management in the Construction, Tourism and SME Sectors.

[11] International Labor Organization. ILO global estimates on inter- national migrant workers labour migration branch conditions of work and equality department of statistics results and methodology, 2019. URL https://www.ilo.org/global/publications/books/ WCMS_652001/langen/index.htm.

[12] Eurostat. European Statistics on Accidents at Work. Methodology-2001 Edition, 2013.

[13] Spanish Statistical Office, 2019. URL https://www.ine.es.

[14] SEOPAN. Asociación de empresas constructoras y concesionarias de infraestructuras. informe anual, 2018. URL https://seopan.es/.

[15] Ching-Wu Cheng and Tsung-Chih Wu. An investigation and analysis of major accidents involving foreign workers in Taiwan's manufacture and construction industries. Safety Science, 57:223 – 235, 2013. ISSN 0925-7535.

[16] Bong,A., Chao, TC., and Lee J., Risks of death from industrial accidents of foreign workers in singapore. Annals of the Academy of Medicine Singapore, (5(2)):138–142, 1976.

[17] Trong-Neng Wu, Saou-Hsing Liou, Chao-Chun Hsu, Show-Lin Chao, Shu-Fen Liou, Kquei-Nu Ko, Wen-Yu Yeh, and Po-Ya Chang. Epidemiologic study of occupational injuries among foreign and native workers in Taiwan. American journal of industrial medicine, 31(5):623–630, 1997.

[18] Bontempo, R. N., Bottom, W.P., and Weber, E.,. Cross-cultural differences in risk perception: a model-based approach. Risk analysis, 17(4):479–488, 1997.

[19] Schenker, M.B., A global perspective of migration and occupational health. American journal of industrial medicine, 53(4):329–337, 2010.

[20] Griffin M.A., and Neal, A., Perceptions of safety at work: A framework for linking safety climate to safety performance, knowledge, and motivation. Journal of occupational health psychology, 5(3):347, 2000.

[21] Mohamed, S., Ali, T.H., and Tam, W.Y.V., National culture and safe work behaviour of construction workers in Pakistan. Safety Science, 47 (1):29 – 35, 2009. ISSN 0925-7535.

[22] Guldenmund, F., Cleal B., and Mearns, K., An exploratory study of migrant workers and safety in three European countries. Safety Science, 52:92 – 99, 2013. ISSN 0925-7535.

[23] Albert PC Chan, Arshad Ali Javed, Sainan Lyu, Carol KH Hon, and Francis KW Wong. Strategies for improving safety and health of ethnic minority construction workers. Journal of Construction Engineering and Management, 142(9), 2016.

[24] Serdar Korkmaz and Dal Jae Park. Comparison of safety perception between foreign and local workers in the construction industry in republic of korea. Safety and Health at Work, 9(1):53 – 58, 2018. ISSN 2093-7911.

[25] T Sterud, T Tynes, I Sivesind Mehlum, KB Veiersted, B Bergbom, A Airila, Bo Johansson, M Brendler-Lindqvist, K Hviid, and M-A Fly- vholm. A systematic review

of working conditions and occupational health among immigrants in Europe and canada. BMC public health, 18 (1):770, 2018.

[26] Mohammad Tanvi Newaz, Peter Davis, Marcus Jefferies, and Manikam Pillay. Using a psychological contract of safety to predict safety climate on construction sites. Journal of Safety Research, 68:9 – 19, 2019. ISSN 0022-4375.

[27] José A García-Arroyo and Amparo Osca Segovia. Occupational accidents in immigrant workers in Spain: The complex role of culture. Safety science, 121:507–515, 2020.

[28] Miguel A. Camino López, Dale O. Ritzel, Ignacio Fontaneda, and Oscar J. González Alcántara. Construction industry accidents in Spain. Journal of Safety Research, 39(5):497 – 507, 2008. ISSN 0022-4375.

[29] Antonio López-Arquillos, Juan Carlos Rubio-Romero, and Alistair Gibb. Analysis of construction accidents in Spain, 2003-2008. Journal of Safety Research, 43(5):381 – 388, 2012. ISSN 0022-4375.

[30] Alistair GF Gibb, Roger Haslam, Diane E Gyi, Sophie Hide, and Roy Duff. what causes accidents. Institution of Civil Engineers/Thomas Telford, 2006.

[31] María D. Martínez-Aires, María C. Rubio-Gámez, and Alistair Gibb. Prevention through design: The effect of european directives on construction workplace accidents. Safety Science, 48(2):248 – 258, 2010. ISSN 0925-7535.

[32] Bilal Umut Ayhan and Onur Behzat Tokdemir. Predicting the outcome of construction incidents. Safety Science, 113:91 – 104, 2019. ISSN 0925-7535.

[33] Amir Mohammadi and Mehdi Tavakolan. Modeling the effects of production pressure on safety performance in construction projects using system dynamics. Journal of Safety Research, 71:273 – 284, 2019. ISSN 0022-4375.

[34] Ching-Wu Cheng, Chen-Chung Lin, and Sou-Sen Leu. Use of association rules to explore cause–effect relationships in occupational accidents in the taiwan construction industry. Safety Science, 48(4):436 – 444, 2010. ISSN 0925-7535.

[35] M. Martínez-Rojas, N. Marín, C. Molina, and M. Vila. Cost analysis in construction projects using fuzzy olap cubes. In 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2015.

[36] Abhishek Verma, Sudha Das Khan, J. Maiti, and O.B. Krishna. Identifying patterns of safety related incidents in a steel plant using association rule mining of incident investigation reports. Safety Science, 70:89 – 98, 2014. ISSN 0925-7535.

[37] Nearkasen Chau, G´erome C. Gauchard, Christian Siegfried, Lahoucine Benamghar, Jean-Louis Dangelzer, Martine Fran¸cais, R´egis Jacquin, Alain Sourdot, Philippe P. Perrin, and Jean-Marie Mur. Relationships of job, age, and life conditions with the causes and severity of occupational injuries in construction workers. International Archives of Occupational and Environmental Health, 77(1):60–66, Jan 2004. ISSN 1432-1246.

[38] Chia-Wen Liao and Yeng-Horng Perng. Data mining for occupational injuries in the Taiwan construction industry. Safety Science, 46(7):1091– 1102, 2008. ISSN 0925-7535.

[39] Chia-Fen Chi, Chong-Cheng Yang, and Zheng-Lun Chen. In-depth accident analysis of electrical fatalities in the construction industry. International Journal of Industrial Ergonomics, 39(4):635 – 644, 2009. ISSN 0169-8141. Special issue: Felicitating Colin G. Drury.

[40] Beatriz Fernández-Muñiz, José Manuel Montes-Peón, and Camilo José Vázquez-Ordás. Relation between occupational safety management and firm performance. Safety Science, 47(7):980 – 991, 2009. ISSN 0925- 7535.

[41] T. Rivas, M. Paz, J.E. Martín, J.M. Matías, J.F. García, and J. Taboada. Explaining and predicting workplace accidents using data-mining techniques. Reliability Engineering & System Safety, 96(7):739– 747, 2011. ISSN 0951-8320.

[42] Antonio López-Arquillos, Juan Carlos Rubio-Romero, and Alistair Gibb. Accident data study of concrete construction companies' similarities and differences between qualified and non-qualified workers in Spain. International journal of occupational safety and ergonomics, 21(4):486–492, 2015.

[43] Biljana Blaževska Stoilkovska, Valentina Žileska Pančovska, and Goran Mijoski. Relationship of safety climate perceptions and job satisfaction among employees in the construction industry: the moderating role of age. International journal of occupational safety and ergonomics, 21(4): 440–447, 2015.

[44] Youngcheol Kang, Sohaib Siddiqui, Sung Joon Suk, Seokho Chi, and Changwan Kim. Trends of fall accidents in the U.S. construction industry. Journal of Construction Engineering and Management, 143(8): 04017043, 2017.

[45] Nicolás Marín Ruiz, María Martínez-Rojas, Carlos Molina Fernández, Jose Manuel Soto-Hidalgo, Juan Carlos Rubio-Romero, and María Amparo Vila Miranda. Flexible management of essential construction tasks using fuzzy OLAP cubes. In Fuzzy Hybrid

Computing in Construction Engineering and Management: Theory and Applications, pages 357–388. Emerald Publishing Limited, 2018.

[46] Bo Shao, Zhigen Hu, Quan Liu, Shu Chen, and Wenqin He. Fatal accident patterns of building construction activities in china. Safety Science, 111:253 – 263, 2019. ISSN 0925-7535.

[47] Ching-Wu Cheng, Sou-Sen Leu, Ying-Mei Cheng, Tsung-Chih Wu, and Chen-Chung Lin. Applying data mining techniques to explore factors contributing to occupational injuries in Taiwan’s construction industry. Accident Analysis & Prevention, 48:214 – 222, 2012. ISSN 0001-4575. Intelligent Speed Adaptation + Construction Projects.

[48] Ian H Witten, Eibe Frank, Mark A Hall, and Christopher J Pal. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2016.

[49] Mauricio A. Valle, Gonzalo A. Ruz, and Rodrigo Morrás. Market basket analysis: Complementing association rules with minimum spanning trees. Expert Systems with Applications, 97:146 – 162, 2018. ISSN 0957- 4174.

[50] E.W.T. Ngai, Li Xiu, and D.C.K. Chau. Application of data mining techniques in customer relationship management: A literature review and classification. Expert Systems with Applications, 36(2, Part 2):2592 – 2602, 2009. ISSN 0957-4174.

[51] Lluis Sanmiquel, Josep M. Rossell, and Carla Vintró. Study of Spanish mining accidents using data mining techniques. Safety Science, 75:49 – 55, 2015. ISSN 0925-7535.

[52] Jesmin Nahar, Tasadduq Imam, Kevin S. Tickle, and Yi-Ping Phoebe Chen. Association rule mining to detect factors which contribute to heart disease in males and females. Expert Systems with Applications, 40(4):1086 – 1093, 2013. ISSN 0957-4174.

[53] Byeongjoon Noh, Juntae Son, Hansaem Park, and Seongju Chang. In-depth analysis of energy efficiency related factors in commercial buildings using data cube and association rule mining. Sustainability, 9(11): 2119, 2017.

[54] Cheng Fan and Fu Xiao. Mining gradual patterns in big building operational data for building energy efficiency enhancement. Energy Procedia, 143:119 – 124, 2017. ISSN 1876-6102. Leveraging Energy Technologies and Policy Options for Low Carbon Cities.

[55] N. Ur-Rahman and J.A. Harding. Textual data mining for industrial knowledge management and text classification: A business-oriented approach. Expert Systems with Applications, 39(5):4729 – 4739, 2012. ISSN 0957-4174.

[56] Yusi Cheng, Wen der Yu, and Qiming Li. Ga-based multi-level association rule mining approach for defect analysis in the construction industry. Automation in Construction, 51:78 – 91, 2015. ISSN 0926-5805.

[57] Fu Xiao and Cheng Fan. Data mining in building automation system for improving building operational performance. Energy and Buildings, 75:109 – 118, 2014. ISSN 0378-7788.

[58] Mehran Amiri, Abdollah Ardeshir, Mohammad Hossein Fazel Zarandi, and Elahe Soltanaghaei. Pattern extraction for high-risk accidents in the construction industry: a data-mining approach. International journal of injury control and safety promotion, 23(3):264–276, 2016.

[59] Heng Li, Xiaoying Li, Xiaochun Luo, and Joanna Siebert. Investigation of the causality patterns of non-helmet use behavior of construction workers. Automation in Construction, 80:95 – 103, 2017. ISSN 0926- 5805.

[60] Dong-Pil Shin, Young-Jun Park, Jongwon Seo, and Dong-Eun Lee. Association rules mined from construction accident data. KSCE Journal of Civil Engineering, 22(4):1027–1039, Apr 2018. ISSN 1976-3808.

[61] Boletín Oficial del Estado. Real decreto legislativo 8/2015, de 30 de octubre, official state bulletin, 2015. URL https://www.boe.es/buscar/ act.php?id=BOE-A-2015-11724.

[62] Labour, Migrations and & Social security Ministry, 2019. URL http: //www.mitramiss.gob.es/en/index.htm.

[63] Boletín Oficial del Estado. Orden tas/2926/2002, de 19 de noviembre, por la que se establecen nuevos modelos para la notificación de los accidentes de trabajo y se posibilita su transmisión por procedimiento electrónico., 2002. URL https://www.boe.es/buscar/doc.php?id= BOE-A-2002-22650.

[64] KNIME. Knime analytics platform, 2018. URL https://www.knime. com/.

[65] Rakesh Agrawal, Ramakrishnan Srikant, et al. Fast algorithms for mining association rules. In Proc. 20th int. conf. very large data bases, VLDB, volume 1215, pages 487–499, 1994.

[66] R. R project for statistical computing, 2018. URL https://www. r-project.org/.

[67] Ching-Wu Cheng, Sou-Sen Leu, Chen-Chung Lin, and Chihhao Fan. Characteristic analysis of occupational accidents at small construction enterprises. Safety Science, 48(6):698 – 707, 2010. ISSN 0925-7535.

[68] Bozena Hola and Mariusz Szóstak. An occupational profile of people injured in accidents at work in the polish construction industry. Procedia Engineering, 208:43 – 51, 2017. ISSN 1877-7058. Innovative Solutions in Construction Engineering and Management. Flexibility in Sustainable Construction.

[69] CM Tam, SX Zeng, and ZM Deng. Identifying elements of poor construction safety management in china. Safety science, 42(7):569–586, 2004.

[70] Emily Q Ahonen, Fernando G Benavides, and Joan Benach. Immigrant populations, work and health - a systematic literature review. Scandinavian journal of work, environment & health, pages 96–104, 2007.

[71] Alessandro Messeri, Marco Morabito, Michela Bonafede, Marcella Bugani, Miriam Levi, Alberto Baldasseroni, Alessandra Binazzi, Bernardo Gozzini, Simone Orlandini, Lars Nybo, et al. Heat stress perception among native and migrant workers in Italian industries-case studies from the construction and agricultural sectors. International journal of environmental research and public health, 16(7):1090, 2019.

[72] Maria J Brunette. Development of educational and training materials on safety and health: Targeting hispanic workers in the construction industry. Family & community health, 28(3):253–266, 2005.

[73] Young Ai Kim, Boong Yeol Ryoo, Yong-Su Kim, and Woon Chan Huh. Major accident factors for effective safety management of highway construction projects. Journal of Construction Engineering and Management, 139(6):628–640, 2013. DOI: 10.1061/(ASCE)CO.1943-7862. 0000640.