International Journal for Quality research
UDK – 638.124.8
Short Scientific Paper (1.03)

Jelena Ruso[1)]

Vera Stojanović [1)]

1) Faculty of organization science, University of Belgrade, Serbia, jelena.ruso@fon.rs, vera.stojanovic11@ gmail.com

OCCUPATIONAL HEALTH AND SAFETY USING DATA MINING

Abstract: *Large amounts of the data gathered in organizations through business operations won't have utility value unless they are used in a proper way. With growing amount of data, the issue of their storage, processing and analysis is becoming more complex. The proper data usage and analysis should provide guidance, solutions and the basis for predictions with the objective of improving and initiating future smart decisions based on the acquired results. Data mining is the tool which exactly enables discovering of emerging patterns and important business information. This work presents the example of Data Mining implementation in the field of workplace health, safety and welfare at HIP- Petrohemija, in Pančevo, as well as various approaches of data analysis and processing by various authors in this field.*

Key words: *Injuries at work, security, data mining, data*

1. INTRODUCTION

In order to use the records on accidents or injuries at work in terms of finding patterns in the available data, DM[1] offers various tools and techniques. Extracting interesting and potentially useful information contained in large databases and their analysis is used for making clusters which help extract the most common injuries and employee groups (with common attributes), upon which our decision-making is based.

The present approach to safety management highlights management functions, guidelines, national and international quality standards and principles in setting up systems in safety management in organizations. These approaches could present a step forward in safety management, but could also be insufficient for effective problem solving in safety management (Santos-Reyes, 2008).

2. OCCUPATIONAL HEALTH AND SAFETY AND INJURIES

Recent trends (new technologies, machines, computers and constant changes in terms of innovations) in organizations can affect workers' health in various ways, increasing the risk of stress related illnesses, cardiovascular diseases, muscular and psychological disorders, increasing the exposure to dangerous substances and violence at workplace, or by affecting professional health services and training programmes. A lot needs to be learnt about the nature of changes in work organization, and the ways they affect workers' health and safety. As long as the available evidence is

[1] Data Mining

insufficient, they, as such, prove the increasing risk of work-related illnesses (Landsbergis, 2003).

The authors of "NORA Organization of Work Team Members" hold that changes in organization of work could also have immediate influence on the physical hazard exposure level at job. For example the workers with multiple jobs and extended shifts could be in danger of overrunning permittable level of exposure to industrial chemicals.

However, as Figure 1 illustrates, conceptually, the present discussion accepts broader definition of organization of work which includes human resources and labour market characteristics (Steven, 2002).

The researchers put forth various conclusions, one of them stating that middleaged workers, 28-47 years of age, had the highest rate of accidents at work. However, another researcher draw the conclusion that the attention decreases with years of age for workers over 25 (Root, 1981). These different interpretations of the accidents at work and age relation have a contradicting explanation. Older workers have lower rates of of accidents, being more experienced, mature and aware of workplace hazards. The other researcher calims that, on the contrary, older workers have a higher rate of accidents due to workplace inattentiveness decreased reflex functions such as sight and hearing. On the other hand, younger workers (under 28) have increased injury ratings due to negligence and not taking workplace hazards seriously. The author of the contrasting conclusion consider younger workers to have lower rates of accidents because of superior reflexes and lesser exposure to hazardous tasks which require longer experience.

Until 1930 it was considered that the main cause of industrial incidents were unsafe work conditions and physical hazards, heavy equipment, trenches, mechanical explosions, ionizing radiation, inflammability, corrosion, reactivity, rapid movement of vehicles, uneven surfaces, etc. However, Heinrich stated that 88% of all accidents are caused by human error (Nouri, 2008).

3. CASE STUDY OF OCCUPATIONAL INJURIES IN TAIWAN CONSTRUCTION INDUSTRY

Recent studies provide the explanation for the weather conditions impact on occupational injuries in Taiwan construction industry. The authors of the aforementioned study examined the ways of creating the data base on injuries and thus reducing the risk of being injured and act preventively for the future by means of the injury prevention strategy. This article scrutinizes 309 workplace accidents with fatal injuries in construction industry during the period from 1999-2004. The factors which contribute to workplace injuries are classified into four categories: individual factors, task factors, management factors and environmental factors. Each factor is divided into several levels upon which the database is created. After the research, it was noticed that the most common occupational injuries were caused by the age factor as well as the season of works on the project (Summer 07:00h-11:00h, the most prevalent on the six day after the rainfall) on both sites. Workers 45-54 years of age, in civil engineering are more exposed to injury, with experience of over 365 days; whereas in structural engineering with experience of 30-180 days. The investment size (more than NT $ 500,000,000) is a potential risk factor according to data on civil engineering; whereas between NT$50,000,000–NT$500,000,000 in structural engineering, as well as 81,43% fatality in relation to the scope of project (with cost estimate of more than NT $ 500,000,000) in civil engineering; and between NT$50,000,000–NT$500,000,000 in structural engineering. Besides that, weather conditions have the

obvious impact on occupational injuries in the northern Taiwan and Taipei County. (Days since the beginning of precipitation, the amount of precipitation, days after the precipitation) (Chia-Wen Liao, 2008).

"Data mining (DM) is the process of designing various questionnaires and extracting useful information, patterns and trends previously unknown, contained in large databases" (Krulj, 2003). "Simply said, the data refer to ore extraction or knowledge mining from large amounts of data"(Han, 2001).

4. CASE STUDY: THE ANALYSIS OF OCCUPATIONAL INJURIES IN HIP PETROHEMIJA, PANČEVO

This work presents occupational injuries during 2010 and 2011 which occurred in the previously mentioned plants, grouped according to the following parameters: gender, workplace, date, time, place of occurrence and severity of occupational injury. Table 1 shows the attributes in order as well as the type of the data used.

Table 1. Attribute names and the type of data used

Attribute names	Type of data
Gender	Binary
Workplace	Text
Age	Numerical
Date	Date
Place of occurrence	Text
Severity of injury	Text
Source of injury	Text
Cause of injury	Text

In order to determine the relations and importance of the above-mentioned data, the programme Clementine has been used and the following graph and statistics were produced (Figure 1).

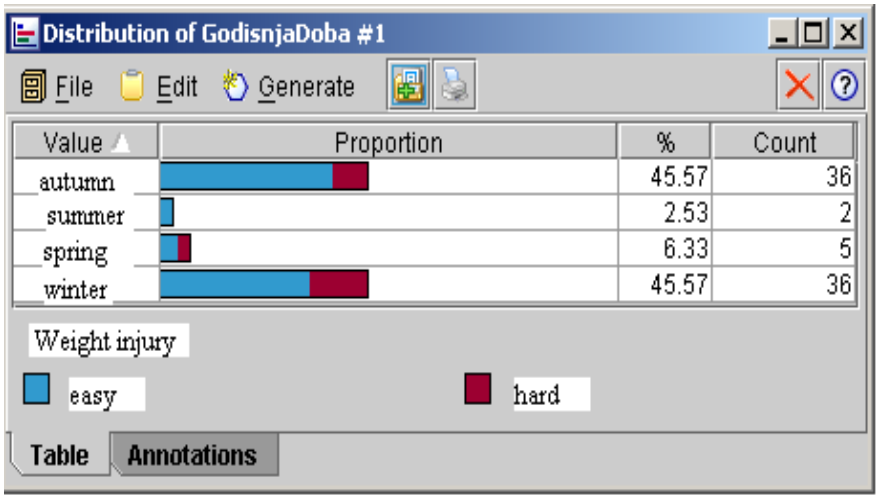

Figure 1. Distribution of the severity of occupational injuries and seasons

Based on the statistical overview a conclusion can be drawn that in 2010 and 2011 out of a total of 79 occupational injuries in HIP- Petrohemija Pančevo, 82, 28% are male, and 17, 72% are women.

Since thre are more male employees, greater occurance of injuries among men is logical. The above figure shows that 77, 22%

of injuries are light, and 22,78% are heavy.

The studies have shown that injuries of women are most common in the Laboratory in Petrohemija and Petroplast, whereas injuries among man were most common in Mechanical Maintenance Unit, HDPE, Ethylene, LDPE. A decision can be made to increase the number of trainings and after training questionnaires. To employ the supervisors to enforce protective equipment use, as well as a surveillance system (cameras) at the site of the greatest injury occurrence in order to determine the causes and reasons for injuries with more certainty, give ergonomic assessment and make appropriate interventions. The Laboratory is the critical workplace because of the exposure to chemicals, and therefore the protective equipment is essential for safety and prevention. The number of the injured is grouped according to the age; whereas up to 30 years of age is considered to be the younger age, 30-50 middle age, and over 50 years of age, older population. The data analysis have shown that the injuries were prevalent in middle (54,43%) and older (39,24%) age structures. The places of occurrence of injuries are various, but some of them are common such as arriving to work and leaving after the work. The ratio between the shifts (morning 06:00h-14:00h , afternoon 14:00h-22:00h, night shift 22:00h-06:00) age and gender was analyzed, arriving at a conclusion that the number of occupational injuries is greater in the morning shift from 06:00h-14:00h with 87,34% and night shift 12,66%. However, women got injured mostly during the morning shift, whereas men had the highest number of injuries recorded in the morning and night shifts.

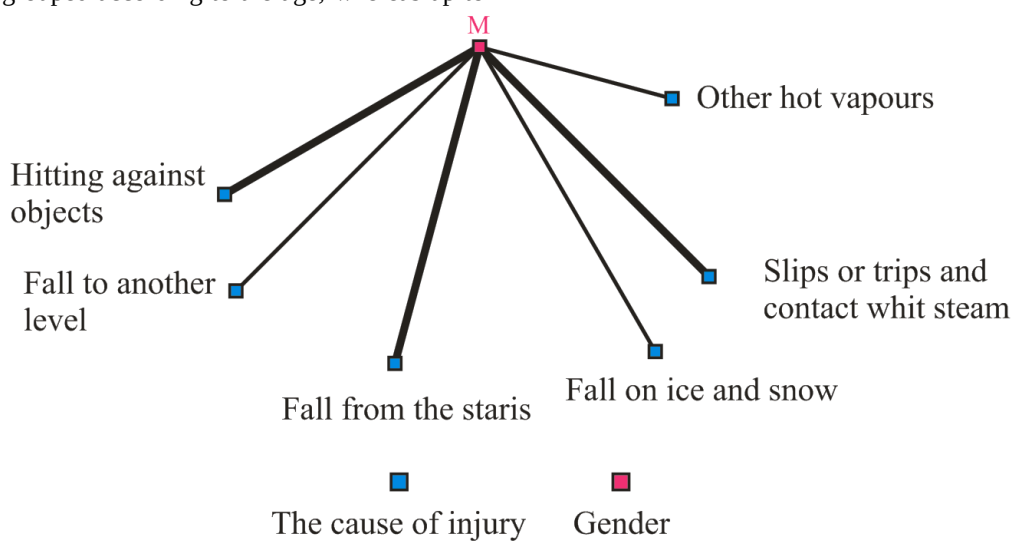

Figure 2. Network: the cause of injury and gender

More serious bodily injuries were most prevalent in the morning, first shift, in all age structures, considering both light and serious bodily injuries. From a psychological point of view, people are drowsy in the morning and in the night shift, their ability to concentrate is reduced, which can lead to inattentiveness and occurrence of injuries. It frequently happens that due to a colleague's absence from work, another worker has to fill in because of insufficient staff in plants such as LDPE, HDPE, MM Unit. Under these circumstances the risk of incident occurrence is much higher. The management's decision has to be providing sufficient number of replacement workers in

order to prevent fatigue and decreased attentiveness.

The causes of injuries were considered initially among male workers because these injuries outnumber injuries of women. The most prevalent causes are: falls (to another level, from the stairs, on ice and snow), hitting against objects, slips or trips and contact with steam and other hot vapours. Based on a detailed scrutiny and report on each injury, one of the main initiators is failure to use protective equipment especially in terms of steam burns and hitting against other objects. Also, the falls to another level occur due to the lack of safety fences on some floors, as well as inadequate footwear which causes frequent trips. Falls on ice occur outside the factory premises when arriving to or leaving work, due to inadequate footwear or workers' inattentiveness.

The major causes of occupational injuries in 2010 and 2011 among men are as follows:

- Hitting against other objects,
- Fall from the stairs,
- Slips, trips and stumbles,
- Contact with steam and other hot vapours,
- Falls on ice and snow.

The major types of injuries among workers are:

- Foot and ankle distortions and dislocations,
- Knee distortions and dislocations,
- External head injuries,
- Shin bruise,
- External injuries to the hand and arm,
- Superficial injury of ankle and foot.

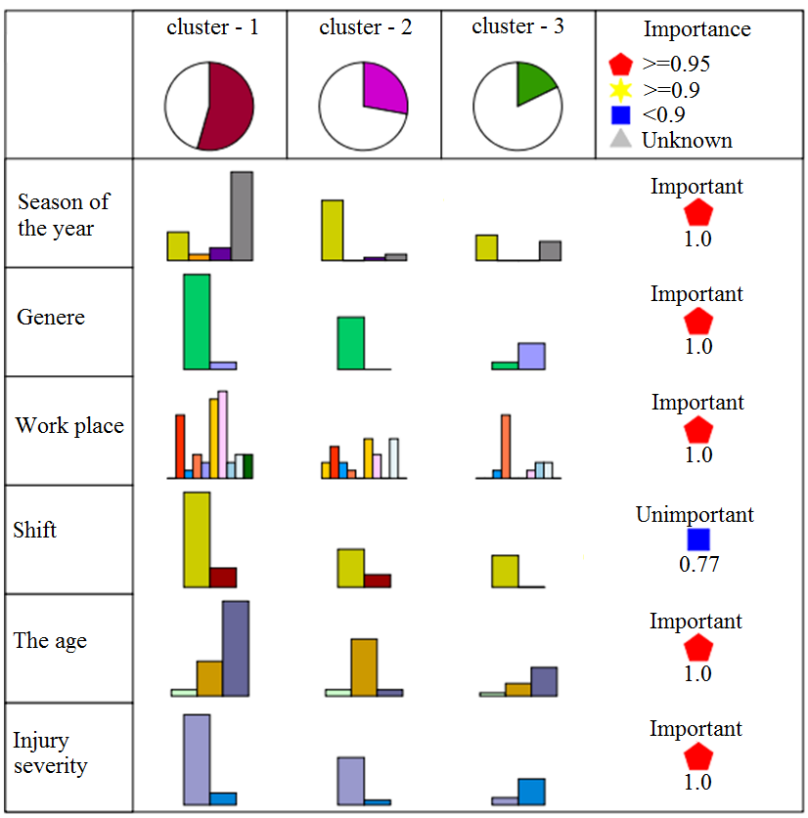

Figure 3. K-means algorithm

According to Mingers and Rosenhead (2004) a good way to assess and prevent a risk is to select an *ad hoc* panel of experts in order to spur communication and and hold briefings during which members could spread their knowledge and information about these processes. A panel would be composed of 7 participants, 2 academics whose studies mostly focused on risk assessment, 2 company managers, 2 factory supervisors and the representative of a healthcare institution's administration. This number of participants, which at first glance seems large, can support an organization with insufficiently specified problems, enable the exchange of knowledge among various actors and reach a high concensus degree.

The DM algorithm used in this work is *K-means*. The number of clusters is 3 (three). After the analysis has been carried out, the following graph is produced. (Figure 4) Three groups of clusters are distinguished; inputs (shift, season, gender, age, place of occurrence and severity of injury) and the importance of differentiation. The distance between clusters is given in Table 2.

Table 2. The distance between clusters

2	0.874	1	0.874	1	1.125
3	1.125	3	1.314	2	1.314

The following conclusions can be drawn from these data:

Cluster 1: Occupational injuries are most common in winter in the morning shift among male employees, among older adults. The places of occurrence of injuries are HDPE and MM, and the severity of injury is light.

Cluster 2: Occupational injuries are most common in autumn in the morning shift and they occur exclusively among men and among middle aged workers. The places of occurrence of injuries are the LDPE and MM, and the severity of injury is light.

Cluster 3: Occupational injuries are most common in autumn in the morning shift among women employees and older adults. The place of occurrence of injuries is the Laboratory, and the severity of injury is serious.

Certain similarities as well as differences among clusters can be observed. The clusters should be analyzed and the appropriate strategy for each of them should be identified and thus the safety should be increased and work related injuries reduced. Being guided by the remark that *similar problems have similar solutions* (Suknović, 2010), some of the attributes will have the same or similar strategies, which will be considered further.

Cluster 1. strategy: The number of injuries prevalent during winter, in this cluster, is explained by increased precipitation, ice formation and slipperiness. Considering the studies which prove that work related injuries are more common in older population because attentiveness gradually reduces with age (Root, 1981) we assume that the same applies to HIP Petrohemija Pančevo. In that case, workers over 50 years of age with higher occupational injury rate (as seen in this cluster) need to be placed in less hazardous positions or they should be replaced by older and middle-aged adults between 30 and 50. Signs need to be placed where the injuries may occur due to slipperiness (stairs, metal ladders, etc.). LDPE and MM are workplaces which require workers to be in constant motion in and outside the plants, and to use the stairs and ladders, so the cause of injuries in these units is reasonable. The severity of injury in this cluster is "light" because the employees conscientiously wear protective equipment and therefore the injuries are not serious. Predominance of occupational injuries in the morning shift in each of the three clusters is justified by drowsiness, occurrence of snow and ice (after nighttime precipitation), busier traffic during the morning shift and therefore the greater number of accidents. Considering precipitation, workers who will prepare and clean the premises should be provided in order to eliminate problems in the morning

shift and therefore the injuries as well. During the first half an hour at job, workers should be given the opportunity for adaptation to the work environment (the time for morning coffee, breakfast, talking to colleagues, becoming alert). Thereby inertia would be overcome and concentration and alertness would increase. The gender that suffered injuries is male because most employees in these plants are men, and therefore it is impossible to take any measures in relation to this issue.

Cluster 2. strategy: Strategies and measures from Cluster 1 that can be repeated and reused when the same attributes are in question (the morning shift, the season of occurrence of injuries is autumn, the gender is male, the severity of injuries is light). There are differences in the strategy used for employees' age and workplace. It is observed that MM plant recurs (the employees from this unit often work with machines, tools and they are mobile in other plants as well) and incidence is greater in the high density polyethylene plant- HDPE. High density polyethylene is processed using various techniques: blow modeling, injection molding, extrusion, pressing, by which most diverse products are made - from household items and technical parts to natural gas and water pipes. The conclusion can be drawn that a considerable number of machines and tools is being used in this company and therefore protective equipment and experience is crucial when occupational injuries are in question. From the interview with the manager of this unit, we acquired the information that there is a problem of insufficient number of replacement workers in this organizational unit, i.e. it frequently happens that due to a colleague's absence from work (health and family issues), another worker has to stay working a double shift because of insufficient staff in plants such as HDPE. Therefore a sufficient number of workers need to be employed or hired from other similar plants, in case of overstaffing issues.

Cluster 3 strategy: Strategies and measures from Cluster 1 and 2 that can be repeated and reused are: age (Cluster 1.), shift and season. Cluster 3. has the greatest number of differences and specificities because in this cluster women are injured, the type of injury is serious, and the place of occurrence of injury is the Lab. In cases of more serious injuries, the managers claim that protective equipment in the Lab (respiratory mask, goggles or gloves) is rarely used, and therefore the injuries are more serious. The focus of service of Occupational health and safety should be taking action to improve and increase control when it comes to the habit of using the equipment in the Lab. It is necessary to employ a person who will periodically visit the Lab or to install security cameras which will remind the employees about their duties and responsibilities regarding protective equipment. Hiring or job swaps are needed between men and women in the Lab because women are more often absent from work (maternity leave, oversensitivity or children's illnesses) and thus insufficient staff, overload and increased number of more serious injuries. The exception which can be singled out is serious injuries among women in the Lab. Considering the fact this unit employs mostly women, the gender aspect can be excluded.

Having defined actions, for each cluster, to be taken when a worker is grouped in any of these clusters, we can say *there is sufficient knowledge to make decisions* (Suknović, 2010). Using these three clusters we can consider the steps to reduce the injury rates. All plants included in these clusters are of key importance for business operations of Petrohemija Pančevo. If injuries occur, the employees are dissatisfied, the number of workers decreases due to sick leave, and therefore the productivity is reduced, which altogether can impact the final product. Of considerable importance is to implement the aforementioned actions and strategies in each of the three clusters in order to reduce

occupational injuries and increase workers' safety.

A measure which should be taken to provide a clearer insight into occupational injuries, according to the article 49, paragraph 2, of the Law on Occupational Health and Safety, is that the employer is liable to keep records of jobs with higher risk rates, the records of employees in workplaces with higher risk rates, and of medical exams of these employees, occupational injuries, professional diseases, work related illnesses, records of personal protective equipment checks and inspection, etc. (Ljajić, 2007).

The algorithm for generating association rules should be applied to the presented data to find out if other regularities which clustering algorithm couldn't mine exist. Association rule is one of the most important and a well tested data mining technique. It aims to extract interesting correlations, frequent patterns, associations or casual structures among sets of items in the transaction databases or other data repositories. Association rules are widely used in various areas such as telecommunication networks, market and risk management, inventory control etc. (Kotsiantis, 2006).

Apriori algorithm is one of the most commonly used for finding *association rules*. This algorithm is chosen because of its speed of implementation.

5. CONCLUSION

This work shows the utility value of *K-means* algorithm as well as of various virtualization techniques of occupational injury analysis. The possibility for predicting and discovering criteria which influence the analyzed occurrences, alone, offers great advantages to this study. One of them is finding target groups among employees prone to occupational injuries. Such information could provide a substantial benefit for the employer and the work safety and health protection service, because based on the reports on the last 2 years it's easier to predict the occurrence of occupational injuries in HIP-Petrohemija Pančevo in the following years (Strategy 1, Strategy 2 and Strategy 3).

REFERENCES:

Bevilacqua, M., & Ciarapica, F. E. (2012). Analysis of injury events with fuzzy cognitive maps. *Journal of Loss Prevention in the Process Industries*, 25(4), 677–685.

Han, J., & Kamber, M. (2001). *Data mining - Concepts and Techniques.* San Francisco: Morgan Kaufmann Publish- ers.

Krulj, D. (2003). *Primena data mining algoritama u statističkoj obradi podataka.* Paper presented at Simpozijum o operacionim istraživanjima SYM-O-PIS 2003, Herceg Novi.

Landsbergis, P. (2003). The changing Organization. *Journal of Occupational & Environmental Medicine*, 45 (1), 61-72.

Liao, C.-W., & Perng, Y.-H. (2008). Data mining for occupational injuries in the Taiwan construction industry. *Safety Science*, 46(7), 1091–1102.

Ljajić, R. (2007). *Ministarstvo rada i socijalne politike.* Retrieved from http://www.minrzs.gov.rs

Nouri, A. A. (2008). The evaluation of safety behaviors in a gas treatment company in Iran. *Journal of Loss Prevention in the Process Industries*, 21(3), 319-325.

Root, N. (1981). Injuries at work are fewer among older employees. *Law Journal Library*, 104, 30-34.

Santos-Reyes, J., & Beard, A. N. (2008). A systemic approach to managing safety. *Journal of Loss Prevention in the Process Industries*, 21(1), 15-28.

Steven, L. S. (2002). *The Changign Organization of Work and the Safety and Health of Working People.* Retrieved from http://www.cdc.gov/niosh/docs/2002-116/pdfs/2002-116.pdf