The Turkish Journal of Occupational / Environmental Medicine and Safety

Vol:1, No:4, 2016 Web: http://www.turjoem.com ISSN : 2149-4711

Analysis of the relation between the characteristics of workers and occupational accidents using data mining

Muhammet Gul*, Ali Fuat Guneri*, Fatih Yilmaz*, Oguzhan Celebi*

*Yildiz Technical University, Mechanical Faculty,
Department of Industrial Engineering, Besiktas, Istanbul, TURKEY

mgul@yildiz.edu.tr, guneri@yildiz.edu.tr, ftyilmaz@yildiz.edu.tr, oguzhancelebi@gmail.com

ABSTRACT

Occupational accidents that occur in several industries frequently result in huge losses and casualties. The main causes of occupational accidents are often stemmed from safety management problems requiring analysis. Data mining techniques have been commonly applied in many fields. However, these methods are still rarely used in occupational health and safety (OHS) issues. This study aims to draw attention to two points on an occupational accident data including 234 instances from different sectors in Turkey using k-means clustering algorithm of Weka software. First, it seeks how to manage the investment on workers based on the characteristics of workers and occupational accidents with maximizing efficiency. Second, it hypothesizes if there is any relationship between the characteristics of workers and occupational accidents. The results of this study show that the use of data mining techniques in OHS provides improvement policies for reducing the occupational accidents and protecting workers from these accidents.

Keyword: *Occupational accident, data mining, k-means clustering algorithm, Weka*

The Turkish Journal of Occupational / Environmental Medicine and Safety

Vol:1, No:4, 2016 Web: http://www.turjoem.com ISSN : 2149-4711

1. INTRODUCTION

Occupational accidents constitute one of the major problems in Turkey as well as other countries over the world [1,2]. In every year, hundreds of people die due to occupational accidents and more than are injured. This leads to social and economic issues for these people. At the same time, occupational accidents have a strong influence for sustainability of enterprises due to their costs and environmental impacts [3, 4]. According to the statistics of Republic of Turkey Ministry of Labor and Social Security between the years 2004 and 2010, 46.4% of the accidents and 41.1% of the deaths occurred only in mining, metal and construction industries [5]. Therefore, taking preventive measures and suggesting solutions can reduce losses and provide serious improvements in workplace environment. The issue of workplace and worker safety becomes a priority in OHS management policies and requires studies that investigate the causes of accidents and the relation between the characteristics of workers and occupational accidents.

Recently, data mining studies are gained importance with storing the information easily through the developed computer systems. Data mining is a process of uncovering information that is not clear, unknown and implicit. Data mining approaches are applied to the different sectors such as marketing, banking, insurance, telecommunication, healthcare, and manufacturing. However, these methods are still rarely used in OHS management. In the field of OHS, Turkey has legislative, practical and institutional knowledge accumulated over the recent years. Even if having institutional knowledge, Turkey can’t still prevent occupational accidents. So that it is very important to propose improvement policies for reducing the occupational accidents and protecting workers from these accidents.

Amiri et al. [6] applied decision tree and association rules based data mining methods to analyze different types of occupational accidents in the construction industry using 21864 data records. They concluded that in the falls and falling objects accidents, the frequency of accidents at night shift is less than the others, and injury to the head, back, spine and lower extremities are more prevalent. In the hit by vehicle, electric shock, collapse in the excavation and fire or explosion accidents, the frequency of accidents among married and older workers is more than single and young workers. They reached a higher frequency in the evening and especially night shifts as well as during the weekends. The injuries to the head, face and neck are greater than the other accidents in these accidents.

Cheng et al. [7] applied data mining classification and regression tree (CART) methods to the causes and distribution of occupational accidents in Taiwan construction industry using a database of 1542 accident cases during the period 2000–2009. They conclude that the occurrence rules for falls and collapses in both public and private project construction industries serve as key factors to predict the occurrence of occupational injuries. In another study, Cheng et al. [8] applied data mining classification and regression tree (CART) methods to examine the distribution and rules governing the factors of the disasters using 349 cases of major occupational accidents in the petrochemical industry between 2000 and 2010 in Taiwan. They concluded that using equipment and implementing safety management measures can effectively prevent accidents.

Nenonen [9] applied decision tree and association rules based data mining methods to analyze factors related to slipping, stumbling, and falling (SSF) accidents at work from 2006 to 2007 using Finnish national occupational accidents and diseases statistics database. It was concluded that the most important factor influencing whether or not an accident at work is the specific physical activity of movement and the risk of SSF accidents at work depends on the occupation and the age of the worker. Zhang and Jiang [10] applied cluster analysis, association rules, time series and decision tree based data mining methods to analyze fire accident cases. The study was considered as valuable in fire accidents and other work safety accident data analysis and helps policy makers understand the nature and essence of accident.

Hajakbari and Minaei-Bidgoli [11] proposed a new scoring system to classify workplaces and determine their risk levels. The proposed method includes calculating the variables, normalization, clustering, determining the risk level for each cluster, and scoring of each cluster using the Iran's Ministry of Labor data for 2010 and 2011. By the study, 21 workplaces are identified as critical and requiring periodic inspections in 2012.

Rivas et al. [12] applied data-mining techniques (decision rules, Bayesian networks, support vector machines and classification trees) in order to identify the most important causes of accidents in mining and construction sectors.

Sanmiquel et al. [13] applied Bayesian classifiers, decision trees or contingency tables, among other data mining techniques in order to analyse main factors of accidents using a database with 70,000 occupational accidents and fatality reports corresponding to the decade 2003–2012 in the Spanish mining sector. They perform statistical analyses with Weka software and provide behavioural patterns based on certain rules.

As can be seen from the mentioned literature, studies have used data mining techniques to analyze databases of occupational accidents and have obtained important outcomes by such techniques. In this study, a data mining based application related to OHS is presented. A data set with 234 instances that includes characteristics of the workers who had occupational accidents and some way of features of the occupational accidents is used in order to analyse the relation between them. Then the data are subjected to the k-means clustering algorithm of Weka software in data mining analysis to explore potential relationships and accident occurrence rules in occupational accidents.

This study make contributions to the literature with some aspects apart from the current studies. First, there is limited research on the key relations between the characteristics of workers and occupational accidents. So, this study aims at analysing this relationship using k-means clustering data mining algorithm in Weka software. Second, it uses a real data of occupational accidents from a series of sectors in Turkey and the data mining application performed here is the first attempt in order to reduce accidents and suggest improvement policies in OHS environment.

The rest of the study is prepared in the following manner. Section 2 presents related literature on data mining with OHS issues. Section 3 presents an overview of data mining, a brief explanation of k-means clustering and Weka software. Section 4 deals with our data set. Section 5 shows the results of the application case with the data set. At last section, the conclusion and limitation of the study and future recommendations are provided.

2. METHODS

In this section an overview of data mining concept and techniques is given, a brief explanation of k-means clustering is explained and Weka software is introduced with general aspects.

2.1. Data mining

Data mining is an analyzing and parsing process that uses one or more computerized learning technique to select meaningful and useful data in a data chunk. The aim of data mining is to expose evident and similar features in data [14]. Nowadays, very large amount of data is collected and registered with computerized technology to reach success and power. As the amount of data increases, it is so difficult to reach meaningful knowledge. Meaningful data shows the pattern of data chunk. So, data mining has gained importance to reach meaningful knowledge [15]. Data mining is the whole work of separation of reserved, unknown and potentially useful knowledge from the database. The stages of reaching knowledge in a dataset are shown in Figure 1.

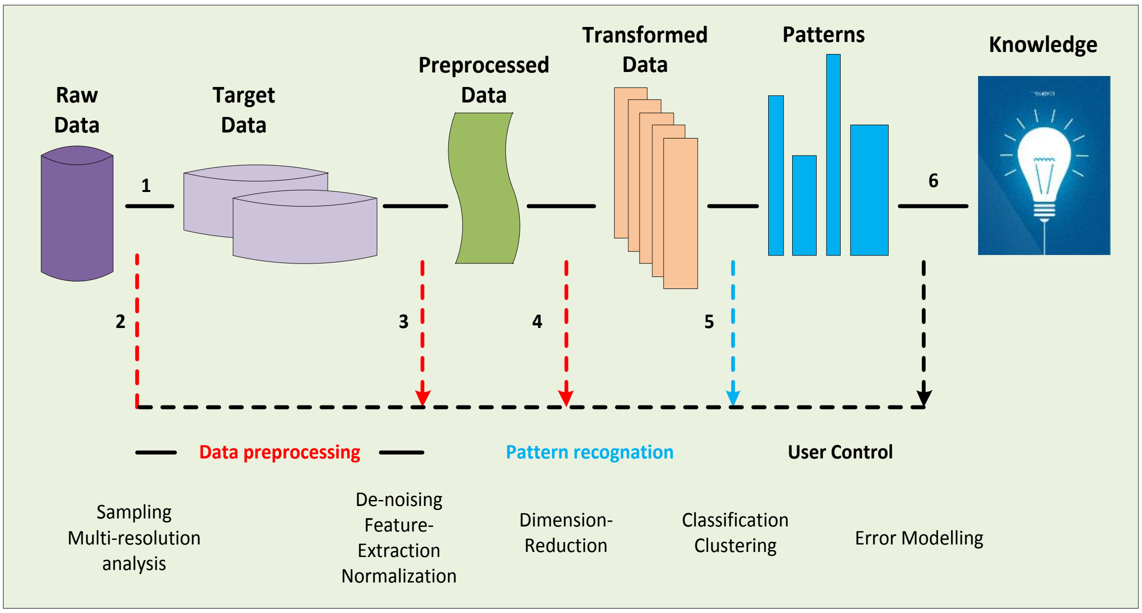


Figure 1

Data mining key steps in an iterative and interactive process

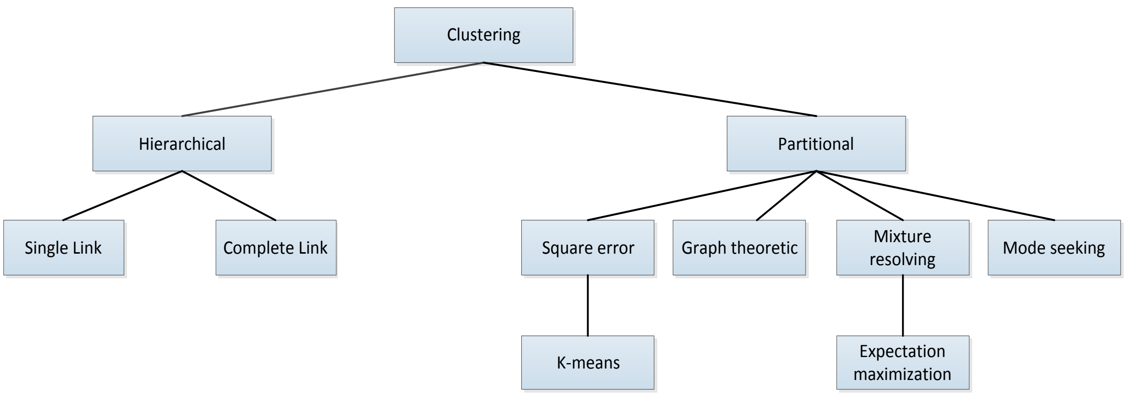
The first four stages remove unimportant data and convert to suitable form for data mining. In other words, these four stages are different forms of data preprocessing, where the data are prepared for mining. The last two stages present the result of data mining [16]. Data mining has many different applications in many different areas such as banking, marketing, insurance, telecommunication, healthcare, manufacturing and safety.

To decide on data mining models is the hardest step to start a project for a specific problem. Data mining approaches should be implemented after defining the problem clearly and collecting all the potential useful data. Moreover, it should be considered about the financial way of the process before the implementation of data mining.

Data mining depends on the applications such that different applications may be required different data mining methods. Functions of data mining are categorized based on the main methods listed as follows: (1) association analysis, (2) characterization, (3) classification, (4) clustering and (5) outlier analysis.

2.2. Clustering

Clustering is the processes of creating a partition so that all the members of each set of the partition are similar according to some metric. A cluster is a set of objects grouped together because of their similarity [11]. Objects are often decomposed into an exhaustive and/or mutually exclusive set of clusters. Clustering according to similarity is a very powerful technique, the key to it being to translate some intuitive measure of similarity into a quantitative measure. When learning is unsupervised then the system has to discover its own classes i.e. the system clusters the data in the database. The system has to discover subsets of related objects in the training set and then it has to find descriptions that describe each of these subsets. There is a number of approaches for forming clusters. One approach is to form rules which dictate membership in the same group based on the level of similarity between members. Another approach is to build set functions that measure some property of partitions as functions of some parameter of the partition. Figure 2 gives a good taxonomy for the different methods for clustering analysis:



The Turkish Journal of Occupational / Environmental Medicine and Safety

Vol:1, No:4, 2016 Web: http://www.turjoem.com ISSN : 2149-4711

Figure 2 Taxonomy for different type of clustering analysis [20]

In general when we make a comparison between hierarchical algorithm and partitional algorithm, the fact is that hierarchical algorithms cannot provide optimal partitions for their criterion which is in contrast with partitional algorithm [17]. However, partitional methods assume given the number of clusters to be found and then look for the optimal partition based on the objective function. Lastly, the most important difference between hierarchical and partitional approach is that hierarchical methods produce a nested series of partitions while partitional methods produce only one. We are going to discuss in more detail with regards to k-means in partitional method, and focuses on the existing approaches that are related to the work in this study.

2.3. K-means clustering algorithm

K-means clustering as the most intuitive and popular clustering algorithm, iteratively is partitioning a dataset into k groups in the vicinity of its initialization such that an objective function defined in terms of the total within-group sum-of-squares is minimized. However, there exists several flaws; not only regarding the sensitivity of the algorithm to initialization, the algorithm's success in finding globally optimal partitions depends on its starting values. Several different initialization results have been proposed for the k-means algorithm and can be easily trapped at a local minimum regarding to the measurement (the sum of squared errors) used in the model [17].

This method initially takes the number of components of the population equal to the final required number of clusters. In this step itself the final required number of clusters is chosen such that the points are mutually farthest apart. Next, it examines each component in the population and assigns it to one of the clusters depending on the minimum distance. The centroid's position is recalculated every time a component is added to the cluster and this continues until all the components are grouped into the final required number of clusters [11, 16, 18].

2.4. Weka

Weka is software that has machine learning algorithms for data mining tasks. These algorithms can be applied directly for a dataset or your own Java code can be invoked. Weka workbench includes set of visualization tools and algorithms which is applied for better decision making through data analysis and predictive modeling. It also has a graphical user interface (GUI) for ease of use. Weka provides tools for classification, clustering, data pre-processing, regression, association rules and visualization. Moreover, it is very suitable for improving new machine learning schemes. As can be seen in Figure 3, it has four main application tools.

The Turkish Journal of Occupational / Environmental Medicine and Safety

Vol:1, No:4, 2016 Web: http://www.turjoem.com ISSN : 2149-4711

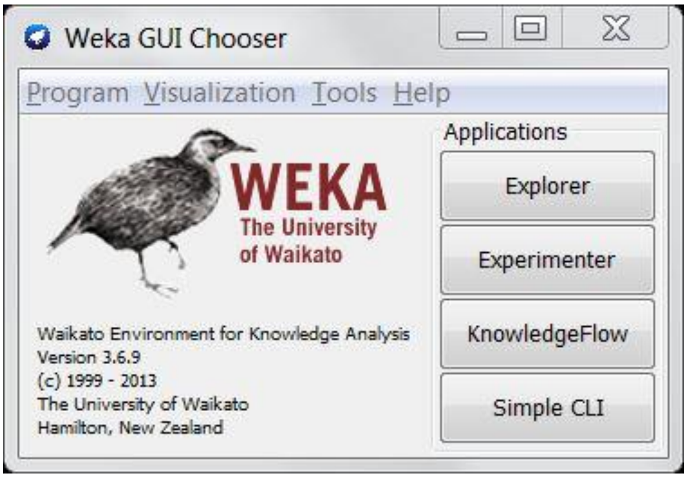


Figure 3 Weka startup screen

By the *Explorer* application it is used for exploring the data with Weka by providing access to all the facilities by the use of menus and forms. Weka *Experimenter* allows one to create, analyze, modify and run large scale experiments. It can be used to answer question such as out of many schemes which is better (if there is). *KnowledgeFlow* application has the same function as that of *Explorer*. It supports incremental learning. It handles data on incremental basis. It uses incremental algorithms to process data. *Simple CLI* stands for command line interface. It just provides all the functionality through command line interface [19].

3. DATA

The data we deal with in this study is collected from different sectors of the field. It is provided some information about data set before giving the application case. First one is about the data set characteristics. In this study, multivariate characteristics are determined. As attribute characteristics, there are two different distinctive attributes identified and these are ‘numerical’ and ‘nominal’. As mentioned hereinabove, in this study clustering methods are used to analyse relations. 234 numbers of instances are characterized and 8 different types of attributes are selected to provide better process and eliminate the missing values. The main steps of the study are shown in Figure 4.

The data set concerns about OHS and occupational accidents in general. Because of privacy, all instances’ names are represented by a code from A1 to A8. Mean and standart deviation values of numeric attributes in this dataset is provided in Table 1. According to attribute type, there are 4 numerical and 4 nominal attributes. Marital status of workers indicates whether he/she is married or single. Scene of accident represents the place of the occurred accident. There are two alternatives; indoor and outdoor. Month of accident refers to the months of the calendar. Day of accident similarly refers to the days of the calendar.

The Turkish Journal of Occupational / Environmental Medicine and Safety

Vol:1, No:4, 2016 Web: http://www.turjoem.com ISSN : 2149-4711

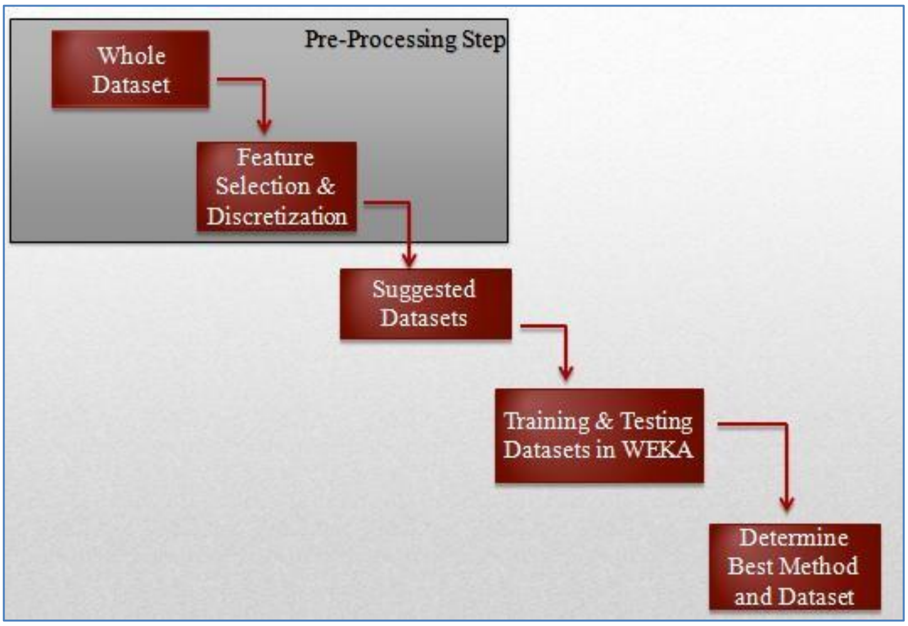


Figure 4 Research process flow chart

Table 1 Attributes and their characteristics

# of attribute	Attribute type	Mean	Standard deviation	Number of distinct
A1	Numeric	1.923	2.736	-
A2	Numeric	29.103	6.608	-
A3	Nominal	-	-	2
A4	Numeric	1.65	1.634	-
A5	Nominal	-	-	2
A6	Numeric	4.316	1.619	-
A7	Nominal	-	-	12
A8	Nominal	-	-	7

A1: Tenure, A2: Age, A3: Marital status, A4: Number of family members in his/her care, A5: Scene of accident, A6: Time of accident, A7: Month of occurence , A8: Day of occurence

4. RESULTS OF THE APPLICATION CASE

In this section, first it is focused on the issue that is 'If it is desired to invest in workers who had an occupational accident, how it should be managed the investment based on the characteristics of workers and occupational accidents with maximizing efficiency'. To do so, data set is examined at Weka and all values of sum of square errors (SSE) are recorded for between 2 to 10 cluster numbers

TURJOEM The Turkish Journal of Occupational / Environmental Medicine and Safety

The Turkish Journal of Occupational / Environmental Medicine and Safety

Vol:1, No:4, 2016 **Web:** http://www.turjoem.com **ISSN : 2149-4711**

and certain range of seeds' values which the range is between 1 to 50. Each case is tested separately. If the meaning of that is considered, minimum SSE values are picked for each cluster number. Thus, the seed values which provide minimum SSE values are determined. In the following, Table 2 shows that which seed value provides the minimum SSE value for each cluster number.

Table 2 Minimum SSE values for each cluster with seed value combination

Number of Cluster (k)	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Seed	49	4	25	8	12	9	9	9	44
SSE	49.85	31.35	21.07	17.62	14.97	12.87	11.26	8.67	7.17
Cluster0	65	10	33	30	17	24	24	24	7
Cluster1	35	25	24	16	24	7	6	6	3
Cluster2	-	65	10	19	9	17	15	15	6
Cluster3	-	-	32	10	16	17	17	17	5
Cluster4	-	-	-	24	24	16	16	16	19
Cluster5	-	-	-	-	10	9	9	9	16
Cluster6	-	-	-	-	-	10	10	5	15
Cluster7	-	-	-	-	-	-	3	3	17
Cluster8	-	-	-	-	-	-	-	6	7
Cluster9	-	-	-	-	-	-	-	-	6

According to Table 2, if requested is the most representative cluster, then the smallest SSE value should be picked. So that, cluster number which provides the smallest SSE value is 10 and seed value is 44. Clusterer output screen is given in Figure 5.

```
Number of iterations: 2
Within cluster sum of squared errors: 7.167360518787783
Missing values globally replaced with mean/mode

Cluster centroids:
                                       Cluster#
Attribute                  Full Data          0          1          2          3          4          5          6          7          8          9
                               (234)       (17)        (7)       (15)       (11)       (44)       (37)       (35)       (39)       (16)       (13)
=================================================================================================================================================
Tenure                        1.9231          1          1          1         10          1          1          1          1          1         10
Age                          29.1026         32         47         32    34.7273    32.3409         21         32         21         32    32.6154
Family Members in His Care    1.6496          0          4          4          4        1.5     0.0811          4     0.1154     1.4063     1.3846
Time of Accident              4.3162     4.4118     4.3571        2.5     4.5455      5.625     5.2703     5.2857     2.4487       2.25     4.4615



Time taken to build model (full training data) : 0.04 seconds

=== Model and evaluation on training set ===

Clustered Instances

 0       17 (  7%)
 1        7 (  3%)
 2       15 (  6%)
 3       11 (  5%)
 4       44 ( 19%)
 5       37 ( 16%)
 6       35 ( 15%)
 7       39 ( 17%)
 8       16 (  7%)
 9       13 (  6%)
```

The Turkish Journal of Occupational / Environmental Medicine and Safety

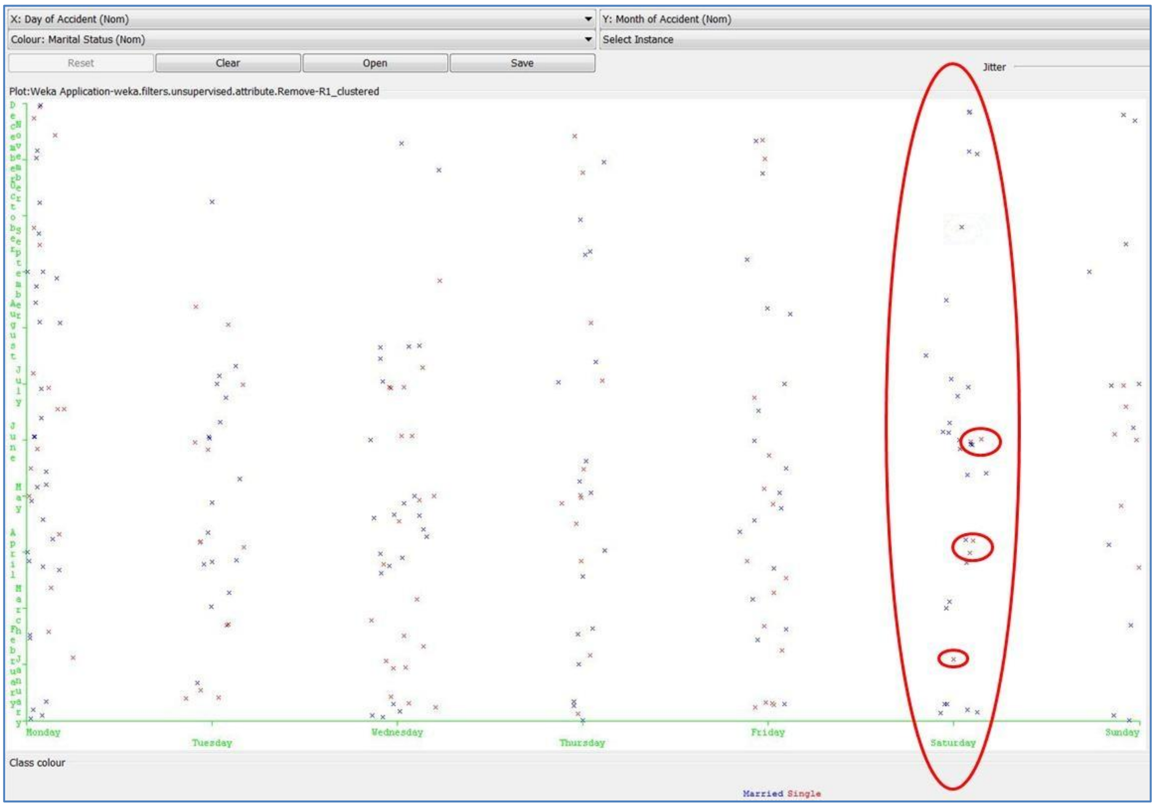
Vol:1, No:4, 2016 Web: http://www.turjoem.com ISSN : 2149-4711

Figure 5 Clusterer output screen in Weka

As shown in Figure 5, Cluster 4 is the largest percentage as %19. It can be invested in the light of the characteristics of Cluster 4. This cluster represents the tenure as 1, age as around 33, family members in his care as around 2, time of accident as around 5 and 6. As can be seen, if it is invested considering classic average method on the characteristics of the workers and occupational accidents, it may be taken wrong decision. The average of full data's characteristics represents tenure as 2, age as 29, family members in his care as 1, and time of accident as around 4. Consequently, there is a significant difference between Cluster 4's and full data's characteristics. It should be provided a better efficiency with clustering and k-means algorithm.

Focusing on the data set with a different point of view, it can be hypothesized if there is any relationship between the characteristics of workers and occupational accidents. All hypothesis are based on visualization of cluster assignments at Weka.

The purpose of first hypothesis (Hypothesis 1) is to put forward to that '*Is there any relationship between marital status and day of occurence?*'. To do so, when it was assigned Day of occurence for x, Month of occurence for y and Color is Marital status, Table 3 shows proportion of married workers who have had an accident for each day. As can be see, proportion of married workers who have had an accident for Saturday is %85. So it can be concluded from Figure 6 and Table 3 that; a single worker has fewer accidents on Saturdays than a worker who is married.



The Turkish Journal of Occupational / Environmental Medicine and Safety

Vol:1, No:4, 2016 Web: http://www.turjoem.com ISSN : 2149-4711

Figure 6 Display of Hypothesis 1 in Weka

Table 3 Proportion of married workers who have had an accident for each day

Days	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Percent	%67	%57	%52	%59	%55	**%85**	%65

The purpose of second hypothesis (Hypothesis 2) is to put forward to that '*Is there any relationship between marital status and time of accident?*'. To do so, when it was assigned Day of occurence for x, Time of accident for y and Color is Marital status, Table 4 shows proportion of married workers who have had an accident at certain times in shifts. As can be see, proportion of married workers who have had an accident at around the last hours is %82. So it can be concluded from Figure 7 and Table 4 that; a married worker has more accidents than a worker who is single at the last hour of shifts.

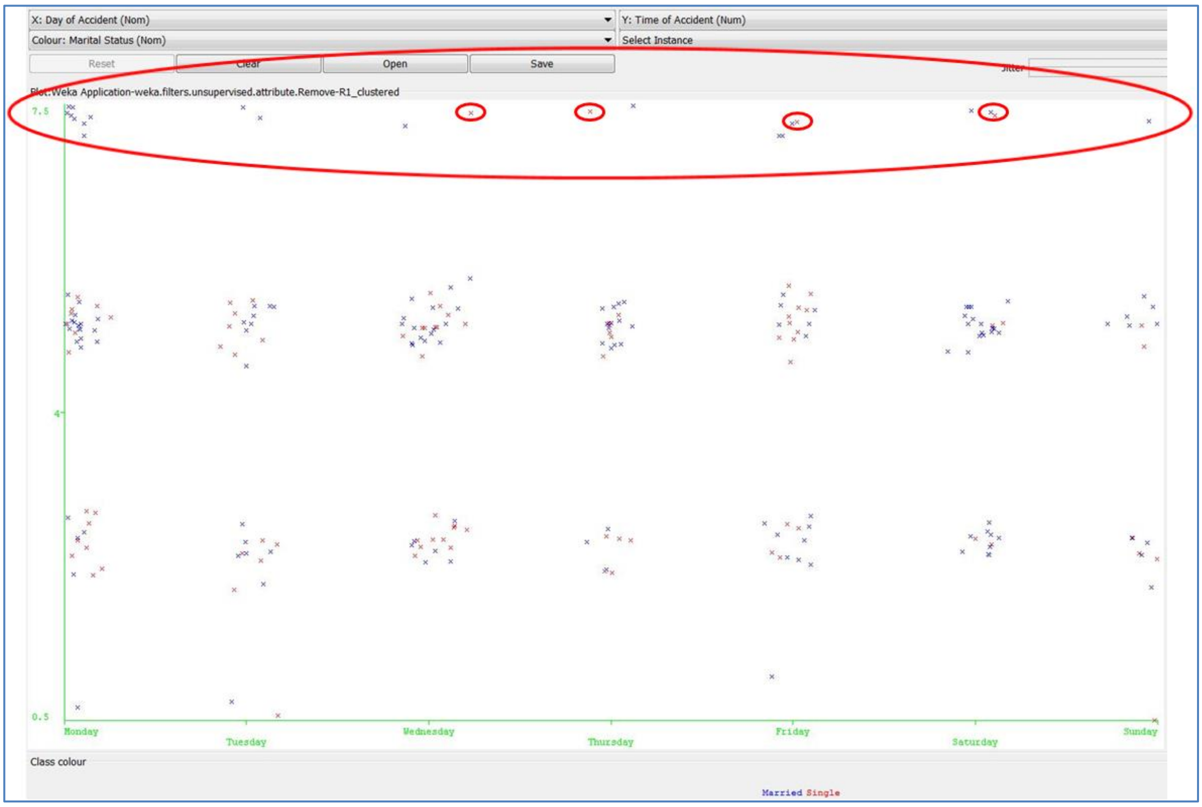


Figure 7 Display of Hypothesis 2 in Weka

The Turkish Journal of Occupational / Environmental Medicine and Safety

Vol:1, No:4, 2016 Web: http://www.turjoem.com ISSN : 2149-4711

Table 4 Proportion of married workers who have had an accident at certain times in shifts

Time of Accident	0.5	2.5	5	7.5
Percent	%60	%52	%65	**%82**

The purpose of third hypothesis (Hypothesis 3) is to put forward to that '*Is there any relationship between marital status and seasons of occurence?*'. To do so, when it was assigned Month of occurence for x, Instance number for y and Color is Marital status, Table 5 shows seasonal proportion of married workers who have had an accident. As can be see, proportion of married workers who have had an accident on autumn is %81. So it can be concluded from Figure 8 and Table 5 that; a married worker has more accidents on autumn than a worker who is single.

Figure 8 Display of Hypothesis 3 in Weka

Table 5 Seasonal proportion of married workers who have had an accident

Season	Spring	Summer	Autumn	Winter
Percent	%65	%60	**%81**	%52

The Turkish Journal of Occupational / Environmental Medicine and Safety

Vol:1, No:4, 2016 Web: http://www.turjoem.com ISSN : 2149-4711

The purpose of fourth hypothesis (Hypothesis 4) is to put forward to that '*Is there any relationship between scene of accident and seasons of occurrence?*'. To do so, when it was assigned Month of occurence for x, Instance number for y and Color is Scene of accident, Tablo 6 shows a worker has more accidents in indoor than a worker who had accidents in outdoor area on autumn. As can be see, proportion of workers who have had an accident in indoor area on autumn is %87. So it can be concluded from Figure 9 and Table 6 that; a worker has more accidents working indoor places of a workplaace on autumn than a worker who had accidents in outdoor area.

Figure 9 Display of Hypothesis 4 in Weka

Table 6 Seasonal proportion of workers who have had an accident in indoor area

Season	Spring	Summer	Autumn	Winter
Percent	%72	%80	**%87**	%46

The purpose of fifth hypothesis (Hypothesis 5) is to put forward to that '*Is there any relationship between number of family members in his/her care and scene of accident?*'. To do so, when it was assigned number of family members in his/her care x, Instance number for y and Color is Scene of accident, Tablo 7 shows proportion of family member in his/her care of worker who have had an accident for certain numbers in indoor. As can be see, number of 1 to 2 family members in the care of a worker who have had an accident for certain numbers in indoor is %78. So It can be concluded from Figure 10 and Table 7 that; a worker with almost 1 to 2 family members in his/her care has more accidents in indoors than the other alternatives

The Turkish Journal of Occupational / Environmental Medicine and Safety

Vol:1, No:4, 2016 **Web:** http://www.turjoem.com **ISSN : 2149-4711**

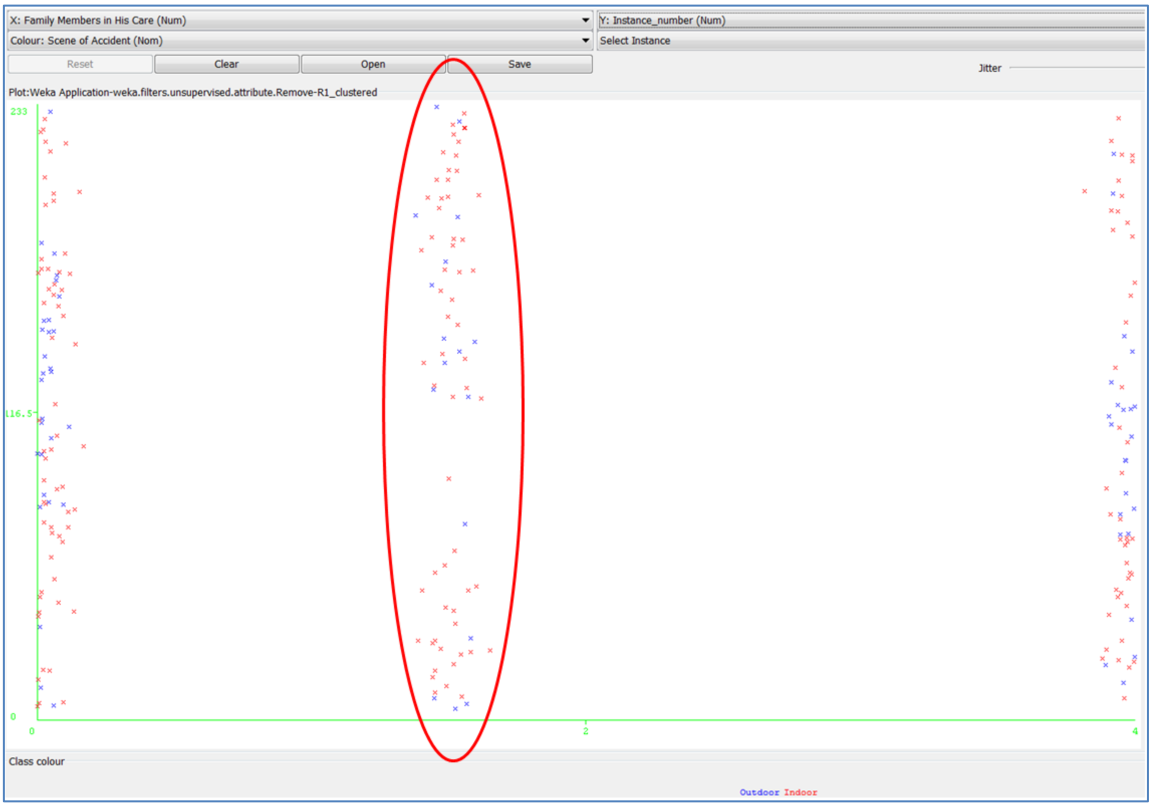


Figure 10 Display of Hypothesis 5 in Weka

Table 7 Proportion of family member in his/her care of worker who have had an accident for certain numbers in indoor

Family Member in His Care	0	1.5	4
Percent	%68	**%78**	%66

5. CONCLUSION

Most data is collected and saved with developed computer systems that are worked based on different algorithms. These computerized systems keep the data and collect them as datasets according to the related areas and factors. These data includes kinds of information even if it does not make any sense before reaching them. Data mining helps to find the unknown and implicit data if it is not clear. It puts the data into familiar data set if it does not have any classes. After the implementation of the data mining techniques, data has meaningful form and potential to use effectively in own data chunk. Data mining techniques have been used in various sectors. However, there are limited studies related with OHS. Most countries and enterprises spend serious time and money to prevent accidents. From this point of view, this study aims to find solutions on OHS issues using data mining. It draws attention to two points on an occupational accident data including 234 instances from different sectors in Turkey using k-means clustering algorithm of Weka software. First, it seeks how to manage the investment on workers based on the characteristics of workers and occupational accidents with maximizing efficiency. Second, it hypothesizes if there is any relationship between the characteristics of workers and occupational accidents. The results of this study show that the use of data mining techniques in OHS provides improvement policies for reducing the occupational accidents and protecting workers from these accidents.

Thinking about data mining, there is no limit to try. Data mining has already become indispensable in many sectors. Data mining boundaries have slightly expanded with this study since it considers problems on OHS management. Weka and other data mining software systems provide many algorithms and get different results in data mining techniques. So that, it can be easily adopted to problems of other untested sectors. This study tries to show future research that gaining a new perspective about the OHS. Serious challenges still exist in data collection and hard to get sense on it.

REFERENCES

1. Unsar S, Sut N. General assessment of the occupational accidents that occurred in Turkey between the years 2000 and 2005. Saf Sci. 2009;47(5):614-619.
2. Yilmaz F. Monitoring and Analysis of Construction Site Accidents by Using Accidents Analysis Management System in Turkey. J Sust Dev. 2015;8(2):57-65.
3. Yilmaz F. Analysis of Occupational Accidents in Construction Sector in Turkey. J Multi Eng Sci Tech. 2014;1(5):421-428.
4. Chong HY, Low TS. Accidents in Malaysian construction industry: statistical data and court cases. Int J Occup Saf Ergon. 2014;20(3):503-513.
5. Ceylan H. Analysis of Occupational Accidents According to the Sectors in Turkey. Gazi Univ J Sci. 2012;25(4):909-918.
6. Amiri M, Ardeshir A, Soltanaghaei E. Analysis of High Risk Occupational Accidents in Construction Industry Using Data-mining Methods. Iran Occup Health. 2014;11(4):31-43.
7. Cheng CW, Leu SS, Cheng YM, Wu TC, Lin CC. Applying data mining techniques to explore factors contributing to occupational injuries in Taiwan's construction industry. Accident Anal Prev. 2012;48:214-222.
8. Cheng CW, Yao HQ, Wu TC. Applying data mining techniques to analyze the causes of major occupational accidents in the petrochemical industry. J Loss Prevent Proc. 2013; 26(6):1269-1278.

9. Nenonen N. Analysing factors related to slipping, stumbling, and falling accidents at work: Application of data mining methods to Finnish occupational accidents and diseases statistics database. Appl Ergon. 2013;44(2):215-224.
10. Zhang D, Jiang K. Application of Data Mining Techniques in the Analysis of Fire Incidents. Procedia Eng. 2012;43:250-256.
11. Hajakbari MS, Minaei-Bidgoli B. A new scoring system for assessing the risk of occupational accidents: A case study using data mining techniques with Iran's Ministry of Labor data. J Loss Prevent Proc. 2014;32:443-453.
12. Rivas T, Paz M, Martín JE, Matías JM, García JF, Taboada J. Explaining and predicting workplace accidents using data-mining techniques. Reliab Eng Syst Safe. 2011;96(7):739-747.
13. Sanmiquel L, Rossell JM, Vintró C. Study of Spanish mining accidents using data mining techniques. Saf Sci. 2015;75:49-55.
14. Roiger R, Geatz M. Data mining: a tutorial-based primer (Vol. 22). New York: Addison Wesley.2003.
15. Chen MS, Han J, Yu PS. Data mining: an overview from a database perspective. IEEE T Knowl Data En. 1996;8(6):866-883.
16. MacQueen J. Some methods for classification and analysis of multivariate observations. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. 1967;1(14):281-297.
17. Pang-Ning T, Steinbach M, Kumar V. Introduction to data mining. In Library of Congress (p. 74).2006.
18. Shaw MJ, Subramaniam C, Tan GW, Welge ME. Knowledge management and data mining for marketing. Decis Support Syst. 2001;31(1):127-137.
19. http://www.cs.waikato.ac.nz/ [20.07.2015]
20. Jain AK, Murty MN, Flynn PJ. Data clustering: a review. ACM Comput Surv. 1999;31(3):264-323.