IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 5, July 2014.

www.ijiset.com

ISSN 2348 – 7968

Human Incidents Analysis by Knowledge Discovery Method in a Steel Maker Company

Farzad Gerami[1]

[1]Industrial Engineering Department, Naghshe Jahan Institute of Higher Education, Isfahan, Iran

Abstract

One of the strategic industries in all countries is the steel industry. Due to the nature of this industry, there is good ground for the human accidents, but many of these incidents are unfortunate and irreparable. This can cause serious damage to human capital, labor productivity, production rates, devaluation of stock. The innovation of this research is using data mining and knowledge discovery methods to predict patterns of workplace accidents in the steel industries. This study presents a method for the analysis of workplace accidents in the steel industry and also develops a model to predict events in the future. Using this method, management can predict future events, including the time of the incident, type of incident, location of incident, event shift, injure type, education of injured person … and to take appropriate action to prevent serious damage. In this research, using Weka (an advanced software for data mining technique) and related algorithms for linear regression, injury to workers is predicted. The research findings include knowledge extraction and prediction in connection with incidents that occurred during the years 2011 to 2013 (2396 incident has been recorded) in Isfahan-Mobarakeh Steel Complex (one of the largest steel producers in the Middle East). The quantitative results (are reflected in the tables in the text) show that the use of these techniques can be an effective program to prediction and prevent accidents in the steel industry.

Keywords: *Human Incidents Analysis, Knowledge Discovery, steel industries*

1. Introduction

Every year tens of millions of workers worldwide are victims of events that led to the death or disability of a large number of them are. According to statistics published in the advanced industrial countries, the annual one in ten workers is suffered accident and therefore of these accidents, five percent the national work day is wasted. So Data Analysis and Prediction of occupational accidents In order to plan for adopt safety standards and prevention of occupational accidents is important. Due to the large volume of data available and the limitations of statistical methods in data analysis, data mining in recent years has attracted the attention of many. Data mining is the process of managing and extracting knowledge from large databases. This paper aims to use the Weka software and linear regression algorithms to predict the type of injury is Mobarakeh Steel Workers.

Nenonen (2013) applies methods of data mining (decision tree and association rules) to the Finnish national occupational accidents and diseases statistics database to analyse factors related to slipping, stumbling, and falling (SSF) accidents at work from 2006 to 2007. The results did not reveal anything unexpected though. [1]

Cheng et al. (2013), Using descriptive statistics, elucidated the factor distribution of these major occupational accidents. The data mining classification and regression tree (CART) was used to examine the distribution and rules governing the factors of the disasters. This study found that for equipment such as pipelines and control valves, devising high-quality safety and protective devices/maintenance/renewal plans and pipeline setups/design plans can effectively prevent accidents such as fires, explosions, and poisoning caused by material leakage, as well as employees being caught in rolled up in machinery. [2]

Sanchez-Pi et al. (2014) propose an information fusion model for an intelligent oil environment in which employees are alerted about possible risk situations while there are moving around their working place. The layered architecture, implements a reasoning engine capable of intelligently filtering the context profile of the employee (role, location) for the feature selection of an inter-transaction mining process. Depending on the employee contextual information he will receive intelligent alerts based on the prediction model that use his role and his current location. [3]

Sari and Selcuk (2009), a methodology to develop a model that included the number of days lost in a coal mine accident in Turkey are suggested. Finally, a simple prediction model to predict the risk levels expected to be created the analysis techniques used in time series analysis. [4]

Ronza et al. (2003), the sequence of events in the port area by drawing a tree obtained from the previous events were prediction. Analysis of 82 incidents in the port area that is selected from a database with the aim of identifying the sequence of events took place. The processed data is done

ISSN 2348 – 7968

by tree accident and possible scenarios are determined. By tree accident and shaping details of events leading to the accident had been collected from various sources, determined was the sequence of events. [5]

Konstandinidou et al. (2006) reported an analysis of all accidents in the chemical industry in Egypt in the period from spring 1997 to 2003 is presented. This incident was conducted to analyze the important parameters and its results on a database of carefully designed. Evaluation of various departments, the reporting system for chemical incidents identified. [6]

Kifle et al. (2014) investigated the prevalence of work related injuries and associated risk factors among production workers in iron and steel industries in Addis Ababa, Ethiopia. In this research data was collected by face to face interview using pre-tested and structured questionnaire, review of records and by check lists. The data were entered and analyzed using SPSS version 16 and bivariate and multivariate logistic regression analysis were used to identify the associated risk factors of injuries. The results showed that work stress, non-use of PPE, consuming alcohol during working days, and excessive noise were observed as modifiable risk factors. [7]

Carlos et al. (2011) a general model for detection and prediction of workplace accidents in Spain has designed. They aim to identify the real risks are records of historical accidents and finally and the risk of dependence are summarized in a table. And the help of a mathematical analysis - statistical criteria required to evaluate and prioritize calculated. [8]

Berentsen and Holmboe (2004) presented a classification and reporting system for incidents/accidents in an oil and gas company. [9]

Sumit et al. (2006) classified NRC database information into different and by using data mining techniques such as association rules interesting patterns of occupational accidents in the petrochemical industry in America according to features such as the type of equipment involved, the type of chemical released the cause of the accident said.[10]

Chang and Tsai (2014) explored the long-term changes of occupational injury patterns from macro-perspective. The results showed that, even within an industry, the injury pattern and epidemiology vary according to contextual factors such as longitudinal business cycle and cross-sectional industrial structure. [11]

Meel et al. (2007) Bayesian methods for predicting repeat accidents cause event, the equipment involved and therefore have used the incident. This study analyzes the dynamics of work-related accidents in the NRC database is used Bayesian method to model the rate of events with events in the petrochemical and chemical companies, and the field due to an incident (such as equipment failure, employee error, etc.) can be designed.[12]

Sanmiquel et al. (2014) analyzed the influence that occupational safety management had on the accidents that took place in Spanish mining of industrial and ornamental stone during the period 2007–2008. [13]

Cheng et al. (2012) explored the causes and distribution of occupational accidents in the Taiwan construction industry by analyzing such a database using the data mining method known as classification and regression tree. The results of this study show 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. [14]

This study aimed to develop a methodology for knowledge discovery from the data repository is related to events occurring in the steel industry. So that the extracted patterns in the data, we can predict things such as the time of the accident, location of accident, type of accident, injury type and event shift. This prediction model can enable management to take appropriate measures to prevent serious accidents.

In Section 2, an overview of data mining techniques will be discussed. Section 3 Methods of data collection, modeling, analysis and design result. Section 4 contains the findings and finally in Section 5 will be made the conclusions.

2. Data mining

According to the University, MIT, new knowledge of data mining is one of the ten students in developing the next decade will deal with the technological revolution. Thus; in recent years the world has seen a tremendous expansion. Data mining is the process of discovering hidden knowledge that will deal with a wide range of specialized fields described below, explain, predict and control the phenomena, Today has wide applications in various fields, Today, the boundaries and limitations for the application of this knowledge in a way that is not considered and working areas of the particles to the ocean floor as deep space. [15]

Data mining is an interdisciplinary field in various domains such as databases, statistics, machine learning and other related fields with combines. [16]

2.1 Data mining methods

The main goal of data mining is predicted to be more accurate to say: "Data mining to identify the correct patterns, novel, useful and understandable data contained in a database is not accessible using conventional

ISSN 2348 – 7968

processing". Figure 1 a schematic view shows the data mining process. [16]

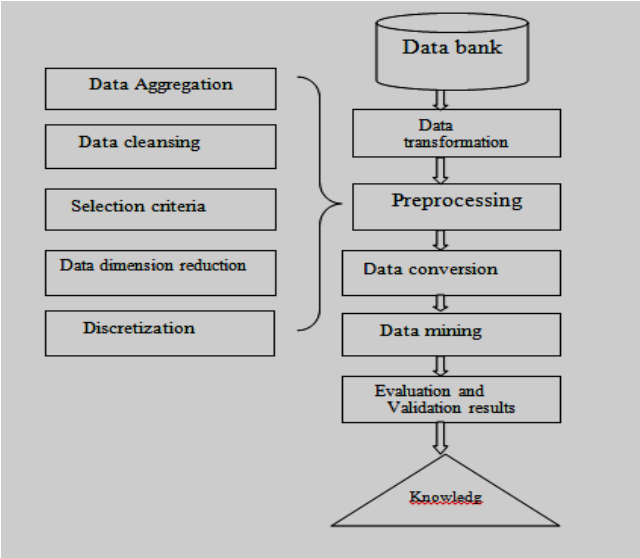

Figure 1: Steps in data mining

Process of data mining techniques to achieve the above two categories, the methods can be divided into descriptive and predictive approaches. The purpose of description is to find patterns in data to be interpreted by humans. Predictive methods to predict the future behavior of the data are used. In order to predict the use of a variable or field in the database to predict unknown or future values of other variables of interest. Data mining functions are shown in the following figure. [16]

Classifications, the process of finding models that diagnosis classes or concepts data cannot handle other objects to predict. Classification learning is a function of a data item into one of the predefined categories is mapped. Training and test data are divided into two parts. Training data to learn the rules used by the system and test data to verify the accuracy of classification are used to prevent over-fitting. [16]

While classification, class labels (i.e., discrete and no way) to predict, prediction methods, functions to model continuous quantity. [16]

•Regression: linear, non-linear

•neural networks, support vector machines

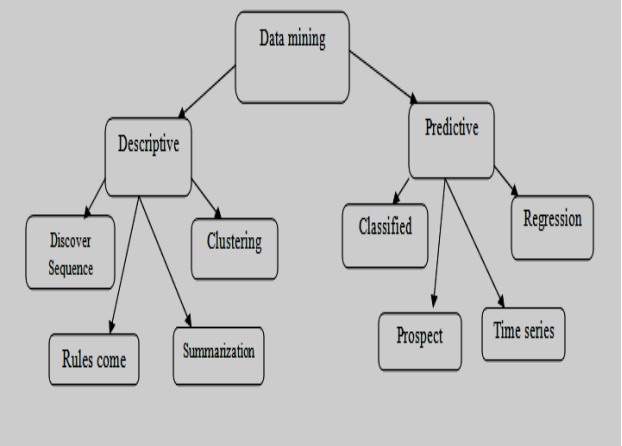

Figure 2: Data mining functions

The best regression model that can be associated with multiple input variables to the output variable. This model makes the data item to a real valued prediction variable is mapped. Traditionally modeled using statistical techniques such as linear and logistic regression and developed. Both classification and regression is used for forecasting. Both classification and regression is used for forecasting. Its simplest form, the model is a linear relationship between input and output variables are linearly connected. [16]

Formula No. 1 (the formula), y and x are the input variables and the output variable and is dependent on a and b are the regression coefficients.

$$y = a + x_1 * b_1 + x_2 * b_2 + \cdots + x_n * b_n$$

2.2 Data Mining Software

To date, commercial software, and training many data for data mining in various fields of science and technology the world has to offer. Each of them according to the types of data that can be explored, are focused on particular algorithms. And a detailed comparison and scientific of these tools have several different aspects of such diverse types and formats of the input data, the volume of possible data processing, algorithm implementation, evaluation methods, results, methods of visualization, data preprocessing methods, interfaces user-friendly, adaptable platforms for performance, price and availability of the software is done. Including traditional applications of data mining in Weka, Clementine, Raoid Miner, Orange, Matlab, R, etc. are software. The Weka software with extensive possibilities, allowing comparison of different methods out there, good guide, Kara graphic interface, compatibility with other Windows applications is presented. [17]

ISSN 2348 – 7968

Software Weka, free and open source software that has been developed at the University of Waikato in New Zealand, and the name of the phrase "Waikato Environment for knowledge Analysis" has been extracted. The software was first written in 1997 as its modernizing. Weka, natural SEO is the name of the bird does not fly in New Zealand, is found. The system is written in Java and based on sweeping the GNU General Public License and has been published. Weka runs on almost any platform and operating system Linux, Windows, Mac, and even on a personal digital receptionist, has been tested. Software Weka, different learning algorithms and provides implementation can easily apply them to their data sets. Also, this software includes a diverse set of tools to convert data sets, such algorithms are discretization. In this environment, a data set can be preprocessed to make it into a learning plan, and analyzed the classification results and its usefulness (all tasks without possible programmatic writing no piece). [17]

In the present study, I will first describe the method used. Following the acquisition and data analysis software, as well as the material covered. In the next section modeling is discussed. Then use the model, the extraction knowledge or predictions is performed and at the end will be presented conclusions.

3. Methodology

Implementing data mining process is based on three main steps: data preparation, modeling and forecasting. Time and type of activity required performing the first step of the contingency and the structure of the data set and compatibility with the purposes of research is based. But the second step of the research virtually uniform in all runs so the standard methodology of data mining, most research has done in this area have used a variation of this method.

3.1 Obtaining and analyzing research data

Data collection for this study was provided by the community it represents, including 2396 cases and 7 features that ranged from 2011 to 2013 are recorded in Mobarakeh Steel Complex. The implementation of this phase of the data mining process in the pre-processing steps included conversion of data and organizes it in code read the data in the Excel spreadsheet for the Weka software. Data Preparation for Data Mining means just having the data is not. But also clean the data and convert it to a more appropriate data may be required. In order to pursue the aim of investigating changes in the characteristics of each sample was necessary. Data files from the Explorer panel are Weka software. According to Figure 3, features include Unity, Toll type, Zone, Toll place, Month, P Havades (it means accidents), Tahsilat (it means educations).

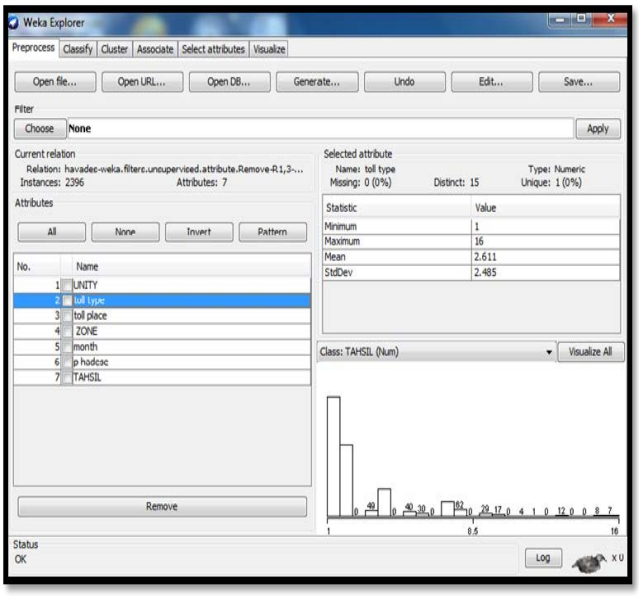

Figure 3: Load Data Explorer panel

3.2 Modeling

With the completion of the data preprocessing stage, two-stage model with the following steps: determining the function of data mining research is needed to solve the problem and select the appropriate tool to advance the work begins.

Since the ultimate goal of this the research is to predict injury Mobarakeh Steel Complex Workers, data-mining functioning in accordance with this objective, "anticipates" would be. The creation of a prediction model, the choice of techniques to predict, select the data set, we set the target variable, is over. As expressed in a subset of 2396 samples of data as data sets were selected models.

Target variable, or dependent variable in this the research "injury" Mobarakeh Steel Complex Workers, and the independent variables include the six-character company type, location, time of incident, the moon and education workers.

As mentioned, the most common tools used to predict "regression" is. Linear regression algorithm is an algorithm used in this the research, which was conducted according Thlsl (according to data of Figure 4) the accuracy of this research is to model data sets.

ISSN 2348 – 7968

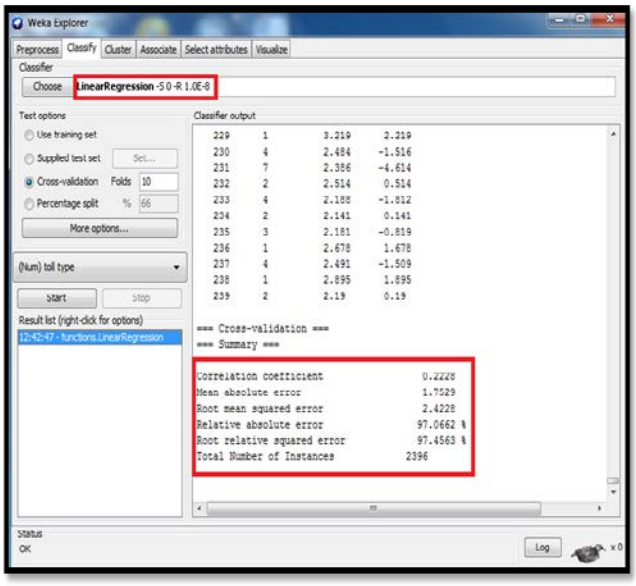

Figure 4: Linear regression model selection

Factors in the model are as follows:

toll type =

−0.2766 * UNITY +
0.1054 * toll place +
−0.0757 * ZONE +
−0.0319 * month +
2.9133

Figure 5: Factors Model

4. Research findings (prediction using model)

For predicting tolls type of matrix I have used two-dimensional in this matrix, the variables month event are changing and variables location, company workers, accident and workers are educated.

For example, in this matrix, worker from the personnel of the company 2 steel complex (company) with degree guidance school that in August between the hours of 4 to 8 pm at the steel marker zone, at the hip and leg has suffered tolls. Now the software should predict the code type of toll (Table number 1 is marked with red color).

According to Table 1 predict tolls type code 2 means "wounds, cuts, scratches and external bleeding" is. Due to the variables in the problem and predicted that the software did all the code matrix planar Table 2 were obtained, codes 2 and 3, the "wounds, cuts, scratches and external bleeding" and "foreign body entrance in the tissues "is.

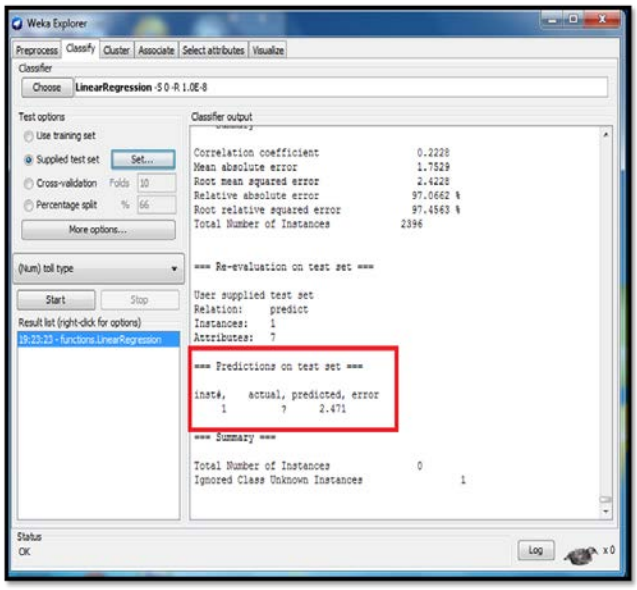

Figure 6: Forecast Model

The following matrix in two dimensions predictions for the type of event due to the month and the Is shown.

Table 1: Matrix model predictions

Mont Zone	1	2	3	4	5	6	7	8	9	10	11	12
1	2.6744	2.642	2.61	2.5778	2.5477	2.5155	2.4883	2.4511	2.4199	2.3887	2.3555	2.3233
2	2.5999	2.5667	2.5335	2.5003	2.4671	2.4339	2.4007	2.3675	2.3343	2.3111	2.2779	2.2447
3	2.5223	2.4891	2.4559	2.4227	2.3955	2.3633	2.3311	2.2999	2.2667	2.2336	2.2004	2.1672
4	2.4447	2.4115	2.3883	2.3551	2.3119	2.2888	2.2566	2.2244	2.1912	2.16	2.1288	2.0966
5	2.3	2.	2.3	2.	2.2	2.2	2.	2.	2.	2.08	2.05	2.02

ISSN 2348 – 7968

	72	34	08	276	44	12	18	148	116	4	2	
6	2.296	2.264	2.232	2.2	2.168	2.136	2.104	2.072	2.04	2.008	1.977	1.945
7	2.22	2.188	2.156	2.124	2.092	2.061	2.029	1.997	1.965	1.933	1.901	1.869
8	2.145	2.113	2.081	2.049	2.017	1.985	1.953	1.921	1.889	1.857	1.825	1.825
9	2.069	2.037	2.005	1.973	1.941	1.909	1.877	1.845	1.813	1.781	1.75	1.718
10	1.993	1.961	1.929	1.897	1.865	1.834	1.802	1.77	1.738	1.706	1.674	1.642

Table number 2, explanatory the coding accidents types that are used in this study.

Table 2: Coding of injury

CODE	Toll Type
1	Contusion and torsion and…
2	Wounds, cuts, scratches, external bleeding
3	foreign body entrance in the tissues
4	foreign body entrance in the eye or eye irritation with foreign materials
5	Bone fractures
6	Scrap Stuff
7	Burn Miscellaneous
8	Damages from heavy lifting
9	Poisoning effects of chemicals into the body tissues
10	Blow to the skull
11	Electrical scurry eyes from welding
12	Asphyxiation caused by gas, water, pressure, or blockage of the respiratory tract, and…
13	Mutilation
14	Multiple tolls
15	Waste such as electric shock, electrocution, and…
16	Hearing loss or sudden deafness caused by acoustic trauma

5. Conclusion

The numerical results show that the use of these techniques, provide a functional program to predict and prevent accidents in the steel industry. In this research we have used data from real workers Isfahan-Mobarakeh Steel Complex. The survey was extracted using a data warehouse. After collecting and preparing data using Weka Software and linear regression model to predict a worker's injury was discussed, according to the accuracy of the available data, Software outputs indicate that the model is made of accuracy. The innovation of this research is using data mining and knowledge discovery methods to predict patterns of workplace accidents in the steel industry. Future research is recommended to the other functions of data mining for field work accidents, as well as run. The implementation of this work in other industries such as oil and gas, petrochemical and automobile can also be attractive. Researchers can also investigate the effectiveness of this technique to reduce accidents in the steel industries.

In the end, we appreciate the efforts and cooperation of the respected staff of safety, health and fire departments in Isfahan-Mobarakeh Steel Company.

References

[1] Nenonen, N. 2013. Analysing factors related to slipping, stumbling, and falling accidents at work: Application of data mining methods to Finnish occupational accidents and diseases statistics database. Applied Ergonomics, 44, 215-224.

[2] Cheng, C. W., Yao, H. Q., Wu, T. C. 2013. Applying data mining techniques to analyze the causes of major occupational accidents in the petrochemical industry. Journal of Loss Prevention in the Process Industries, 26, 1269-1287.

[3] Sanchez-Pi, N., Marti, L., Molina, J. M., Garcia, A. C. B. 2014. High-Level Information Fusion for Risk and Accidents Prevention in Pervasive Oil Industry Environments. Communications in Computer and Information Science, 430, 202-213.

[4] Sari, M., Selcuk, A. 2009. Stochastic modeling of accident risks associated with an underground coal mine in turkey. Safety Science, 47, 78-87.

ISSN 2348 – 7968

[5] Ronza, A., Felez, S., Darbra, R. M., Carol, S., et al. 2003. Predicting the frequency of accidents in port areas by developing event trees from historical analysis. Journal of Loss Prevention in the Process Industries, 16, 551–560.
[6] Konstandinidou, M., Nivolianitou, Z., Markatos, N., Kiranoudis, C. 2006. Statistical analysis of incidents reported in the Greek Petrochemical Industry for the period. Journal of Hazardous Materials, 135, 1–9.
[7] Kifle, M., Engdaw, D., Alemu, K., et al. 2014. Work related injuries and associated risk factors among iron and steel industries workers in Addis Ababa, Ethiopia. Safety Science, 63, 211-216.
[8] Carlos, j., Rubio, E., Garcia, A., Cano, F. 2011. Occupational accidents model based on risk–injury affinity groups. Safety Science, 49, 306-314.
[9] Berentsen, R., Holmboe, R. H. 2004. Incidents/accidents classification and reporting in Statoil. Journal of Hazardous Materials, 111, 155–159.
[10] Sumit, A., Nir, K., Tretter, J., et al. 2006. Harnessing data mining to explore incident databases. Hazardous Material, 130, 33-41.
[11] Chang, D. S., Tsai, Y. C. 2014. Investigating the long-term change of injury pattern on severity, accident types and sources of injury in Taiwan's manufacturing sector between 1996 and 2012. Safety Science, 68, 231-242.
[12] Meel, A., O'Neill, L. M., Leivin, J. H., et al. 2007. Operational risk assessment of chemical industries by exploiting accident databases. Loss Prevention, 20, 113_127.
[13] Sanmiquel, L., Rossell, J. M., Vintro, C., Freijo, M. 2014. Influence of occupational safety management on the incidence rate of occupational accidents in the Spanish industrial and ornamental stone mining. Work: A Journal of Prevention, Assessment and Rehabilitation, 1-8.
[14] Cheng, C. W., Leo, S. S., Cheng, Y. M., et al. 2012. Applying data mining techniques to explore factors contributing to occupational injuries in Taiwan's construction industry. Accident Analysis & Prevention, 48, 214-222.
[15] Shahrabi, j. 2013. Data Mining. Amir Kabir Publication, 147-187.
[16] Ghazanfari, M., Alizadeh, S., Teimurpur, B. 2013. Data Mining and Knowledge Discovery. Science and Technology Publication.
[17] Gerami, F., Rocky, K. 2010. Safety Engineering with Functional Approach. Isfahan Jahad-Daneshgahi press.

Author Biography: I have PhD degree in industrial engineering and now I am teaching in university. The number of my articles is 15 (ISI(3), other journals (9), international conferences (3)) and also I published two books (1-safety engineering (functional approach), 2-fire engineering). My current research interests are healthcare planning and modeling, safety engineering, using mathematics technique in safety.