Predictive model for incident occurrences in steel plant in India

Conference Paper · July 2017
DOI: 10.1109/ICCCNT.2017.8204077

3 authors:

Sobhan Sarkar
Indian Institute of Technology Kharagpur
20 PUBLICATIONS **26** CITATIONS
SEE PROFILE

Vishal Pateshwari
Indian Institute of Technology Kharagpur
2 PUBLICATIONS **3** CITATIONS
SEE PROFILE

Jhareswar Maiti
Indian Institute of Technology Kharagpur
78 PUBLICATIONS **1,265** CITATIONS
SEE PROFILE

Predictive model for Incident Occurrences in Steel Plant in India

Sobhan Sarkar
Research Scholar
Dept. of Industrial & Systems Engineering, IIT Kharagpur
Email: sobhan.sarkar@gmail.com

Vishal Pateshwari
M. Sc.
Dept. of Geology & Geophysics, IIT Kharagpur
Email: vishal.pateshwari@gmail.com

J Maiti
Professor
Dept. of Industrial & Systems Engineering, IIT Kharagpur
Email: jhareswar.maiti@gmail.com

***Abstract*—Steel industry is considered to be an economic sector with higher number of accidents. Workers in this industry are exposed to a wide variety of hazards during working hours. Thus, the database maintained in industry varies in terms of the types of data indicating nature of accidents, causes of accidents, date and time-stamp etc. The objective of this study is to give predictive solution of accident occurrences in a steel industry based on free-text data or narratives logged in the database from previous incidents report. Utilizing the database comprised of 9488 observations of occupational accidents during the year 2010 to 2013 in an integrated steel manufacturing plant in India, text mining technique has been used, and its outputs have been fed into three classifiers namely Support Vector Machines, Random Forest, and Maximum Entropy which are tested for the evaluation of best fit model that could predict the injury occurrences in workplace. Furthermore, with the help of the same three classifiers, the causes of injuries are also predicted. Maximum Entropy and Random Forest classifiers are found out to be the best of all other classifiers in terms of higher area under curve (AUC) value in binary and multi-class prediction model, respectively.**

***Keywords*—Occupational injury, Steel industry, Prediction model, Binary & multi-class classifiers, AUC value.**

I. Introduction

The generic term accident implies an unwanted, unfortunate incident occurring unintentionally and causes injury, damage, loss, mishap, or casualty. As per the statistics from National Programme for Control and Treatment of Occupational Diseases in 2009, 100 million occupational injuries are reported in India resulting in 0.1 million deaths. In addition, it is also reported that 45,000 fatal injuries (45% of the total deaths due to injuries in the world) and 17 million non-fatal injuries (17% of the world) at work take place in each year [1]. There are various factors present behind an accident to take place. In India, a total of 1433, 1383, and 1417 occupational fatal accidents took place in 2011, 2012, and 2013, respectively [2]. Among these accidents, Indian steel plants share a total of 29, 52, and 31 fatal accidents respectively [2]. Thus, there is an urgent need for the management of steel plant to take initiatives to curve down the number of accidents by improved safety practices. Usually, there is a large number of factors responsible for an accident to happen. So, a good prediction model is necessary that can capture the synergistic effect of all these factors and predict the occurrence of accidents/incidents so that preventive measures could be initiated well in advance.

In this study, we proposed a predictive model that can capture the challenging aspect of the concurrent problem of the steel industry facing the occurrence of incidents in their workplace. The main aim of our study is two-fold. First, we aim to predict the occurrence of injury or non-injury cases from only free-text data logged by safety professional as a safety observation data. Second, we also aim to capture the causes of accidents if injury report is predicted from the narratives. In our study, the database of 9488 observations obtained from steel industry with proper classification (either injury and non-injury classes, or 11 primary cause classes) from the year 2010 to 2013 have been used to develop our model for both binary class and multi-class prediction of accidents. The classifiers, namely Support Vector Machines (SVM), Random Forest (RF), and Maximum Entropy (MaxEnt) have been used for the both purposes, i.e., for binary and multi-class classification.

The remaining of the paper is organized as follows. In Section II, a brief description of related works is presented. In Section III, the proposed methodology is discussed in short. A practical industrial application as a case study is presented to show the usefulness of our proposed methodology in Section IV. Results and discussion have been illustrated in Section V, and in Section VI, conclusions with future works are discussed.

II. RELATED WORKS

In occupational accident research domain, many researchers used different techniques to encounter different occupational accident problems. A study using neuro-fuzzy networks has been carried out to predict the consequences of 190,116 cases of injuries that have occurred over a span of five years [3]. With the help of text mining as well as fuzzy Bayes approach, Marucci et al. developed coding classifications to injury text data for a U.S. insurer [4]. Wellman et al. implemented fuzzy Bayesian model to examine the accuracy of data mining algorithm to classify injury narratives [5]. Using data from the Ministry of Labour and Immigration in 2008, Rivas et al. applied decision rules, Bayesian networks (BN), support vector machines (SVM) and classification trees in mining and civil construction companies [6]. In another study on accident data retrieved from the construction industry in Singapore, Goh and Chua used neural network analysis for prediction of the severity of accidents with considerable accuracy [7]. Using Negative Binomial Regression (NBR), Chi-Squared Automatic Interaction Detection (CHAID), Exhaustive CHAID, Classification And Regression Trees (CART), Quick, Unbiased,

Efficient Statistical Tree (QUEST), Artificial Neural Network (ANN) and Neuro-Fuzzy Systems (FIS), Bevilacqua et al. [8] obtained a classification of input data on the basis of their significance and/or influence on the risk level of injuries. Zurada et al. [9] implemented neural networks, logistic regression, decision trees, memory-based reasoning, and the ensemble model for classification of industrial jobs on the risk of work-related low back disorders. Snchez et al. [10] classified the workers who encountered a work-related accident in the last 12 months according to their working conditions using semi-parametric principal component analysis (SPPCA), multivariate adaptive regression splines (MARS) and SVM.

Narrative text is one of the key resources for the investigation of the accident. Bulzacchelli et al. [11] studied on fatal injuries in maintenance and service department of a particular US manufacturing industry by using narrative text from 1984 to 1997. The typical causes of injury were found out to be the electrocution, or struck by the object, or being caught in between parts of equipment. Likewise, Bunn et al. [12] used a multivariate logistic regression analysis for a relationship between the outcomes and the variables. Of which, slip trip falls (STF) is one of the major issues frequently occurred in the industry. Narrative text of STF incidents could help the decision makers to take decisions the main factors behind STF [13]–[15]. Some of the studies found that the rate of STF is higher for older male/female workers than the younger workers [16]. Smith et al. [17] also used narrative text to study 9826 ladder-related injuries in order to find out the causes behind the injury cases. Narrative text is found out to be the potential resource for analysis that leads to identify 17% more features than injury codes alone. Using free text description of the injury, Dement et al. [18] investigated the injuries originated from nail gun activities among the construction workers in order to identify the causes of the injuries.

Narratives could provide much more details of the incidents, and also identify the factors contributing to the injury. So, detailed information in terms of narratives is necessary to build a model that could support the safety management system (SMS) of any industry [19], [20]. But Vallmuur et al. [21] reviewed the main aspects of machine learning using injury text data for last 20 years. Their study revealed the fact that with the increase in computational technique over time, the utility and the applications of narratives has augmented tremendously. The use of human- machine learning approach could reduce the human intervention much more. Tixier et al. [22] tried to develop a system that could overcome the problem by decoding the unstructured reports from accident database. The system developed by them could scan the unstructured injury database with 101 attributes, and produce 95% accuracy. Some of the authors used computerized coding of injury narratives to investigate the accuracy of a fuzzy Bayesian model [23]. Their model could accurately classify the cause-of-injury codes from narratives. Similar kind of approach by using Bayesian method was done by Lehto et al. [24] for classifying the injury narratives into cause groups. With the help of Bayesian model, many researchers attempted to evaluate the causation of workers compensation claims [25], [26]. Adbat et al. developed the BN-based probabilistic model to extract recurrent scenarios from 143 accounts to serious occupational accidents with movement disturbances [15]. For mining free text component of workers compensation claims the data of 1992 to 2002, Brooks used SAS Text Miner that could offer more detailed analyses regarding the problem of claims [27].

From the in-depth survey of literature in accident analysis domain, though not elucidated in this paper in details, it reveals that there have been a very few works in extracting classified knowledge of injuries from narrative or text description of injuries. This kind of study is still in its infancy, particularly from steel industry point of view and in general, from accident/injury research point of view. Further, these reported studies failed to classify the knowledge from a predictive analytics point of view. The present study attempts to bridge the gaps of using text report by exploiting text mining algorithm as powerful tool for extraction of useful information from unstructured free-text and deploys prediction algorithms for prediction of occurrence of injury and causes of injury. Thus, this study has much more potential to contribute to the occupational accident research through predicting occupational injury cases and their corresponding causes from narratives by machine learning approaches.

III. METHODOLOGY

This paper describes the methodology of the text mining, and others machine learning algorithms like SVM, RF, and MaxEnt. For the detailed understanding, interested readers are requested to go through text mining [28], SVM [29], [30], [35], Max Entropy [31], and Random Forest [32], [36].

A. Text mining

Text mining is a process of retrieving useful, valuable information from the text in the form of interesting patterns and trends by statistical pattern learning techniques. Initially, from our database, free text is extracted with classification details. The free text is basically the character strings, which is analyzed by term creation and term filtering process [34].

1) Term creation: The most important terms are obtained in this step. This task has been performed through the steps of tokenization, removal of stop words, stemming or lemmatization operations in sequence. Open-source tool R- statistical software was used for the analysis. Outputs were manually crosschecked in each stage.

2) Term document matrix and weighted term document matrix: In this step, generated terms are pruned based on their frequencies of occurrence. So, it is needed to remove not only the terms which do not occur frequently but also those which follow constant distribution of occurrence in the database considered. In this case, term frequency-inverse document frequency (tf-idf) technique of information retrieval process is used. After filtering the words, a term document matrix (TDM) is created, and then by multiplying by the tf-idf with the frequency of words across documents, weighted term document matrix (WTDM) is formed.

B. Classification algorithms

1) Support Vector Machine (SVM): SVM was first conceptualized by Vapnik [33]. Here, selected number of sample vectors (support vectors) is used for parameter estimation using the least square estimation technique. Usually, the support vectors are much less in number than the total number of

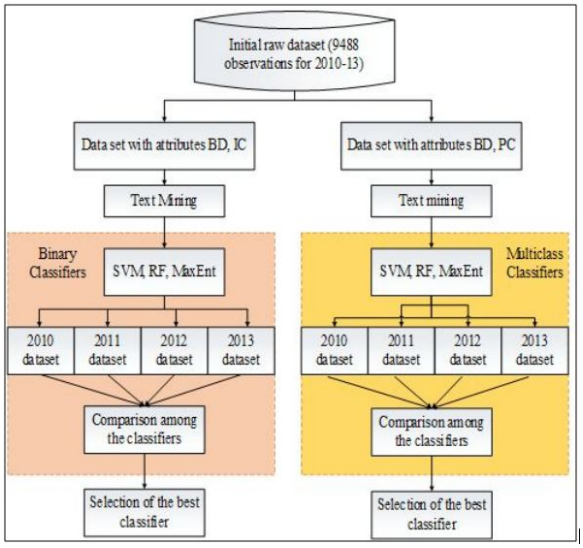

Fig. 1: The proposed methodology.

sample. The Radial Basis Function (RBF) has been used in this study as the kernel function.

2) Maximum Entropy (MaxEnt): In generic text classification domain, MaxEnt is an effective technique. The method can be characterized by their entropy values. High entropy methods make fewer distinctions when classifying documents but do not take full advantage of the signal coming from the training data set, whereas, low entropy methods build many distinctions and can suffer from over interpretation of the training data [31].

3) Random Forest (RF): This method generates many classification trees, called random forest. When it classifies a new object from an input vector, it is made as the input vector to all the trees. Thus, each of the trees gives a definite classification for the new object, and accordingly all possible classes get votes from the collection of all trees in the forest. Finally, based on the maximum number of votes received by the input object for a given class, only that particular class is assigned to the new object. It runs efficiently on the large data set, and it does not overfit.

In Fig. 1, our proposed methodology is displayed in which BD, IC and PC represent brief description of incidents (free-text/ narratives of incident of accident), incident category (injury or non-injury), and primary causes (11 classes of causes, in total), respectively.

IV. CASE STUDY

A. Problem statement and objective

In this section, we have presented a case study carried out in an integrated steel plant in India. The company has been experiencing a large number of incidents in recent past. The descriptive analysis of four years (from 2010 to 2013) data reveals the severity of incidents as property damage (25%), injury (35%), and near misses (40%). Based on the information and data, management has been trying to build a predictive analytics-based decision support system that could provide them a better predictive solution for minimizing the incident/accident occurrences. So, in order to provide them a feasible solution, this study experiments various predictive algorithms using outputs of text mining approach applied on companys incident database. Based on a comparative analysis of the performance measures, the best fit model in both binary and multi-class prediction cases have been proposed to be used by the company.

B. Data collection and preparation

Data was collected from the electronic database of a steel making company from 2010 to 2013. In this study, only text or narratives was considered for analyzing 9488 incidents in total, out of which 2506 are injury reports, and rest are non-injury reports. There are three attributes are selected for this study. First attribute is the brief description (BD) of incident of accident cases which is a portion of free-text or narratives of incidents taken place. The second attribute is incident category (IC) that is categorized into two classes, i.e., injury or non-injury cases. The last attribute is primary causes (PC) which is categorized into 11 classes of primary causes of accidents.

As this text data are huge in size, text mining approach was employed to extract useful information or patterns of occurrence of accidents by finding some important keywords for each event through data pre-processing technique. Then, TDM and WTDM matrices were formed which were further analyzed using three classification algorithms and their prediction performances were compared. We used 10-fold cross validation (C.V.) technique to check the performance of classifiers in each fold to ensure about the overfitting nature of the dataset.

V. RESULTS & DISCUSSION

A. For binary classification prediction

In binary class prediction model, we have experimented the prediction of injury or injury cases with the three classifiers, i.e., SVM, RF, and MaxEnt across the four years. In this case, 10-fold C.V. is used. The average performance results (from 10 folds C.V.) of all three binary classifiers are shown in Table I. As per accuracy, RF, and MaxEnt have similar value i.e., 0.88 while SVM has average accuracy little low i.e., 0.87. According to precision, MaxEnt outperforms others with substantial large average value i.e., 0.84. Similarly, for recall, MaxEnt has minimum average value (0.80). As per F1-score, MaxEnt has the highest average value equal to 0.82. Thus, in our study, to select a classifier performing better than others, we used area under curve (AUC) values. The higher the AUC value, the better the classifier is in its power of prediction than others. Therefore, in the case of binary classification, MaxEnt classifier is preferred over others due to higher value in AUC i.e., 0.87. It implies that the preferred MaxEnt algorithm is predicting injury and non-injury cases better than other classifiers (refer to Table I, and Fig. 2).

B. For multi-class classification prediction

For multi-class prediction model, we used one-vs-all approach in prediction of 11 causes of injury by three classifiers namely, SVM, RF, and MaxEnt. For the multi-class classification, RF has highest average accuracy equal to 0.92 which is higher than either SVM (0.90) or MaxEnt (0.90). As per

TABLE I: COMPARISON OF BINARY CLASSIFIERS (10-FOLD C.V.).

Classifier	Performance parameter	2010	2011	2012	2013	Average
SVM	Accuracy	0.9	0.86	0.86	0.85	0.87
	Precision	0.44	0.7	0.7	0.7	0.64
	Recall	0.92	0.94	0.96	0.94	0.94
	F1- score	0.6	0.8	0.81	0.8	0.75
	AUC	0.72	0.84	0.84	0.83	0.81
RF	Accuracy	0.9	0.88	0.86	0.88	0.88
	Precision	0.56	0.83	0.81	0.84	0.76
	Recall	0.8	0.86	0.87	0.88	0.85
	F1- score	0.66	0.85	0.84	0.86	0.8
	AUC	0.77	0.87	0.86	0.87	0.84
Max Ent	Accuracy	0.89	0.87	0.89	0.87	0.88
	Precision	0.77	0.86	0.87	0.86	0.84
	Recall	0.66	0.82	0.88	0.85	0.8
	F1- score	0.71	0.84	0.87	0.85	0.82
	AUC	0.84	0.87	0.89	0.87	0.87

TABLE II: COMPARISON OF MULTI-CLASS CLASSIFIERS (10-FOLD C.V.).

Classifier	Performance parameter	2010	2011	2012	2013	Average
SVM	Accuracy	0.92	0.91	0.88	0.9	0.9
	Precision	0.43	0.45	0.48	0.51	0.46
	Recall	0.74	0.7	0.66	0.71	0.7
	F1- score	0.49	0.48	0.48	0.53	0.49
	AUC	0.67	0.68	0.7	0.69	0.69
RF	Accuracy	0.93	0.92	0.91	0.92	0.92
	Precision	0.58	0.52	0.54	0.51	0.54
	Recall	0.74	0.67	0.69	0.66	0.69
	F1- score	0.62	0.54	0.55	0.53	0.56
	AUC	0.76	0.72	0.71	0.7	0.72
Max Ent	Accuracy	0.9	0.89	0.9	0.9	0.9
	Precision	0.47	0.47	0.45	0.46	0.46
	Recall	0.44	0.43	0.45	0.45	0.44
	F1- score	0.44	0.43	0.44	0.45	0.44
	AUC	0.7	0.65	0.65	0.67	0.67

precision, RF has the highest average value equal to 0.54. Similarly, for recall measure, MaxEnt has the lowest value i.e., 0.44. To get the trade-off between precision and recall value, we used F1-score. And, in the measure of F1-score, RF has the highest values i.e., 0.56. Finally, to select the best classifiers in multi-class prediction model, highest AUC value is considered. Accordingly, RF outperforms other classifiers with the average AUC value of 0.72 (refer to Table II, and Fig. 3).

In both the cases of binary and multi-class prediction, in SVM, cost function, gamma function, and kernel are taken as 1.00, 0.002, and radial basis function, respectively. In RF algorithm, the number of trees is set as 500. In MaxEnt algorithm, the regularizer is set as 0.

VI. CONCLUSION

Prediction and analysis of incident occurrences in steel industry demand some focused research through data mining approaches that incorporate various machine learning techniques to provide good prediction based on previously stored large database. In order to do this, three different prediction algorithms have been tested using a companys large dataset through text mining approach. The experimental results show that MaxEnt and RF outperform others in binary and multi-class classification problems, respectively. Thus, from the evidence of free-text safety observation report, our proposed approach has much more potential to predict the occurrence of

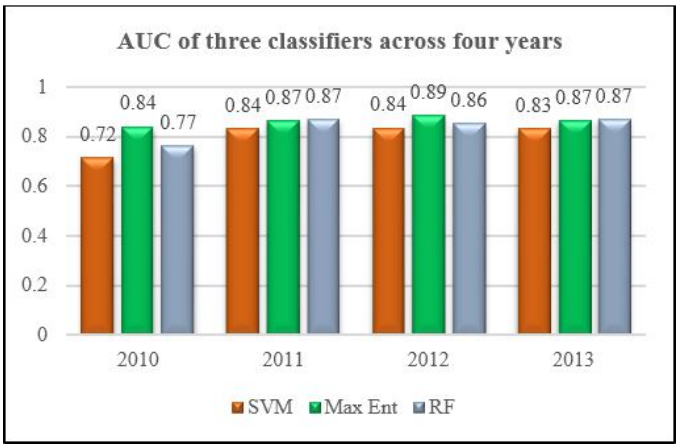

Fig. 2: Comparison of AUC values all four binary classifiers across the four years.

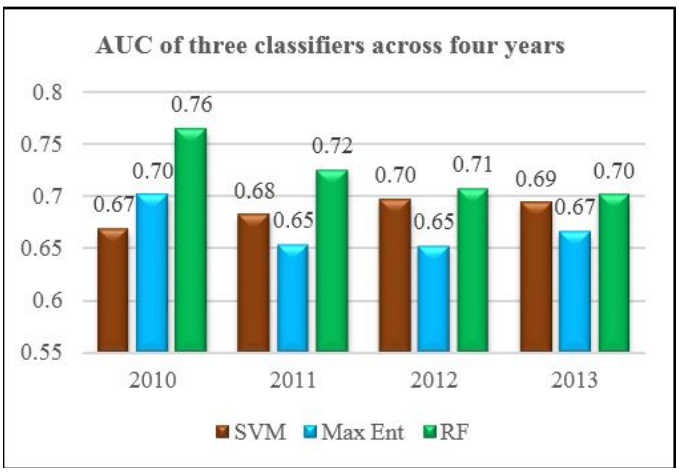

Fig. 3: Comparison of AUC values all four multi-class classifiers across the four years.

injury cases and their probable causes thus facilitating better decision making by the respective organization. As future study, this model should be tested with other models or in other application domain with a view to obtain better insight of prediction of occurrence of incidents based on text mining. In addition, a decision support system (DSS) could be made automated and real time from the text retrieval phase to final decision-making scenario.

ACKNOWLEDGMENT

The authors are thankful to the safety management system of the steel industry for providing the valuable data for our research and their kind support in this project in various aspects from the beginning to the end of the project.

REFERENCES

[1] NL Wanger, NL Wagner, P. Jayachandra, Learning courses in occupation medicine and good practice, Indian J Occup Environ Med, Vol. 9, pp. 57 - 61, 2005.

[2] Retrieved from: www.indiastat.com/table/crimeandlaw/6/ industrialaccidents.

[3] F. E. Ciarapica, G. Giacchetta, Classification and prediction of occupational injury risk using soft computing techniques: an Italian study, Safety science, 47(1), pp. 36 - 49, Elsevier, 2009.

[4] H. R. Marucci, M. R. Lehto, H. L. Corns, Computer classification of injury narratives using a Fuzzy Bayes approach: improving the model, In Human Interface and the Management of Information. Methods, Techniques and Tools in Information Design (pp. 500 - 506). Springer Berlin Heidelberg, 2007.

[5] H. M. Wellman, M. R. Lehto, G. S. Sorock, G. S. Smith, Computerized coding of injury narrative data from the National Health Interview Survey, Accident Analysis & Prevention, 36(2), pp. 165 - 171, 2004.

[6] T. Rivas, M. Paz, J. E. Martn, J. M. Matas, J. F. Garca, J. Taboada, Explaining and predicting workplace accidents using data-mining techniques, Reliability Engineering & System Safety, 96(7), pp. 739 - 747, 2011.

[7] Y. M. Goh, D. Chua, Neural network analysis of construction safety management systems: a case study in Singapore, Construction Management and Economics, 31(5), pp. 460 - 470, 2013.

[8] M. Bevilacqua, F. E. Ciarapica, G. Giacchetta, Data mining for occupational injury risk: a case study, International Journal of Reliability, Quality and Safety Engineering, 17(04), pp. 351 - 380, 2010.

[9] J. Zurada, W. Karwowski, W. Marras, Classification of jobs with risk of low back disorders by applying data mining techniques, Occupational Ergonomics, 4(4), pp. 291 - 305, 2004.

[10] A. S. Snchez, P. R. Fernndez, F. S. Lasheras, F. J. de Cos Juez, P. G. Nieto, Prediction of work-related accidents according to working conditions using support vector machines, Applied Mathematics and Computation, 218(7), pp. 3539-3552, 2011.

[11] M.T. Bulzacchelli, J.S. Vernick et al., Circumstances of fatal lockout/tagout related injuries in manufacturing, American Journal of Industrial Medicine, 51 (10), pp. 728 - 734, 2008.

[12] T.L Bunn, S. Slavova et al, Narrative text analysis of Kentucky tractor fatality reports, Accident Analysis and Prevention, 40 (2), pp. 419 - 425, 2008.

[13] H. Shibuya, B. Cleal, P. Kines, Hazard scenarios of truck drivers, occupational accidents on and around trucks during loading and unloading, Accident Analysis & Prevention, 42(1), pp. 19 - 29, 2010.

[14] M. Amiri, A. Ardeshir, M. Hossein, F. Zarandi, E. Soltanaghaei, Pattern extraction for high-risk accidents in the construction industry: a data-mining approach, International journal of injury control and safety promotion, (ahead-of-print), pp. 1 - 13, 2015.

[15] F. Abdat, S. Leclercq, X. Cuny, C. Tissot, Extracting recurrent scenarios from narrative texts using a Bayesian network: Application to serious occupational accidents with movement disturbance, Accident Analysis & Prevention, 70, pp. 155 - 166, 2014.

[16] K. Kemmlert, L. Lundholm, Slips, trips and falls in different work groupswith reference to age and from a preventive perspective, Applied Ergonomics, 32 (2), pp. 149 - 153, 2001.

[17] G.S. Smith, R.A. Timmons et al., Work-related ladder fall fractures: identification and diagnosis validation using narrative text, Accident Analysis and Prevention, 38 (5), pp. 973 - 980, 2006.

[18] J.M, Dement, H. Lipscomb et al., Nail gun injuries among construction workers, Applied Occupational and Environmental Hygiene, 18 (5), pp. 374 - 383, 2003.

[19] J. Glazner, J. Bondy et al., Factors contributing to construction injury at Denver International Airport, American Journal of Industrial Medicine, 47 (1), pp. 27 - 36, 2005.

[20] H.J. Lipscomb, J. Glazner et al., Analysis of text from injury reports improves understanding of construction falls, Journal of Occupational and Environmental Medicine/American College of Occupational and Environmental Medicine, 46 (11), pp. 1166 - 1173, 2004.

[21] K. Vallmuur, H. R. Marucci-Wellman, J. A. Taylor, M. Lehto, H. L. Corns, G. S. Smith, Harnessing information from injury narratives in the big dataera: understanding and applying machine learning for injury surveillance, Injury Prevention, 2016.

[22] A. J. P. Tixier, M. R. Hallowell, B. Rajagopalan, D. Bowman. Automated content analysis for construction safety: A natural language processing system to extract precursors and outcomes from unstructured injury reports, Automation in Construction, 62, pp. 45 - 56, 2016.

[23] H.M. Wellman, M.R. Lehto et al., Computerized coding of injury narrative data from the National Health Interview Survey, Accident Analysis and Prevention, 36 (2), pp. 165 - 171, 2004

[24] M. Lehto, H. Marucci-Wellman, H. Corns, Bayesian methods: a useful tool for classifying injury narratives into cause groups, Injury Prevention, 15(4), pp. 259 - 265, 2009.

[25] S. J. Bertke, A. R. Meyers, S. J. Wurzelbacher, J. Bell, M. L. Lampl, D. Robins, Development and evaluation of a Nave Bayesian model for coding causation of workers compensation claims, Journal of safety research, 43(5), pp. 327 - 332, 2012.

[26] J. A. Taylor, A. V. Lacovara, G. S. Smith, R. Pandian, M. Lehto, Near-miss narratives from the fire service: A Bayesian analysis, Accident Analysis & Prevention, 62, pp. 119-129, 2014.

[27] B. Brook, Shifting the focus of strategic occupational injury prevention: mining free-text, workers compensation claims data, Safety Science, 46 (1), pp. 1 - 21, 2008.

[28] J. Bentham, D. J. Hand, Data mining from a patient safety database: the lessons learned, Data Mining and Knowledge Discovery, 24(1), pp. 195 - 217, 2012.

[29] H. A. Farfani, F. Behnamfar, A. Fathollahi, Dynamic analysis of soil-structure interaction using the neural networks and the support vector machines, Expert Systems with Applications, 42(22), pp. 8971 - 8981, 2015.

[30] H. Hong, B. Pradhan, C. Xu, D. T. Bui, Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines, Catena, 133, pp. 266-281, 2015.

[31] K. Nigam, J. Lafferty, A. McCallum, Using maximum entropy for text classification, In IJCAI-99 workshop on machine learning for information filtering, Vol. 1, pp. 61 - 67, 1999.

[32] Retrieved from: www.stat.berkeley.edu/∼breiman//RandomForests/cc _home.htm.

[33] V. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.

[34] S. Sarkar, S. Vinay, & J. Maiti, "Text mining based safety risk assessment and prediction of occupational accidents in a steel plant", In Computational Techniques in Information and Communication Technologies (ICCTICT), pp. 439 - 444, IEEE, 2016.

[35] S. Sarkar, S. Vinay, V. Pateshwari, & J. Maiti, "Study of optimized SVM for incident prediction of a steel plant in India", In India Conference (INDICON), pp. 1 - 6, IEEE, 2016.

[36] S. Sarkar, A. Patel, S. Madaan, & J. Maiti, "Prediction of occupational accidents using decision tree approach", In India Conference (INDICON), IEEE Annual, pp. 1 - 6, IEEE, 2016.