Decision Support System for Prediction of Occupational Accident: A Case study from a Steel Plant

Sobhan Sarkar[1], Mainak Chain[2], Sohit Nayak[3], and J. Maiti[4]

[1] Department of Industrial & Systems Engineering, IIT Kharagpur, Kharagpur, India
sobhan.sarkar@gmail.com@gmail.com,
[2] Department of Metallurgical & Materials Engineering, IIT Kharagpur, Kharagpur, India
mainakchain21@gmail.com,
[3] Department of Metallurgical & Materials Engineering, IIT Kharagpur, Kharagpur, India
sohit.nayak@gmail.com,
[4] Department of Industrial & Systems Engineering, IIT Kharagpur, Kharagpur, India
jhareswar.maiti@gmail.com

Abstract. Decision Support System (DSS) is a powerful tool which helps decision makers take unbiased and insightful decisions from the historical data. In the domain of occupational accident analysis, decision making should be effective, insightful, unbiased and more importantly prompt. In order to obtain such decision, development of DSS is necessary. In the present study, an attempt has been made to build such DSS for accident analysis in an integrated steel plant. Two classifiers i.e., support vector machine (SVM), and random forest (RF) have been used. RF has been found to be better classifier in terms of accuracy i.e., 99.34%. The developed DSS has full potential in making insightful decisions and can be used in other domains like manufacturing, construction, etc.

Keywords: Occupational accident analysis, Decision Support System, SVM, RF, Steel industry

1 Introduction

1.1 Decision support system

Decision support system (DSS) is a kind of computer-based information system that can help decision makers utilize data, models and other knowledge on computer to solve semi-structural and non-structural problems, which cannot be measured or modeled.This aspect of semi-structured problems requires human intervention, and therefore, solution to semi-structured problems are often achieved by allowing a decision maker to select and evaluate practical solutions from a finite set of alternatives. The history of the implementation of such systems traces back to the mid-1960s. Ferguson and Jones reported the first experimental study using a computer aided decision system in 1969 [1]. The aim of DSS is to help decision makers improve decision-making effectiveness and efficiency by combining information resources and analytical tools.

1.2 Importance of DSS in occupational safety

Occupational safety of workers in industry is one of the prime concerns since a huge number of accidents reported at workplace globally in each year. Carelessness of the workers, insufficient safety training and education, unawareness of costs of accidents, erroneous series of human operations and inadequate worksite environment remain the key risk factors for occupational accidents. The occupational accidents such as injuries, fatalities, material and/or environmental damages are followed by economical loss. Common causes of occupational accidents are observed to be high elevation, toxic, flammable and explosive materials, fire, moving machinery, dangerous gases, work on/close to haphazard established heavy structures, misuse or failure of equipment, poor ergonomics, untidiness, poor illumination, exposure to general hazards including electricity, and inadequate protective clothing. Hence, safety managers try to find alternative solutions to minimize potential hazards by addressing occupational accidents. Conventional data analysis approaches using machine learning techniques provide satisfactory results but at the cost of time and reliability of results [2–10]. DSS deals with numerous application domains. Therefore, it is not also surprising that the uses of DSS are also diverse. In a nutshell, the aim of DSS is to provide with better decisions and/or provide a better decision making process. It automates the tedious tasks, thus allowing a decision maker to explore the problem more thoroughly. This additional exploration capacity is made possible by the virtue of DSS, which improves understanding of the problem by the decision maker or others in the organization and thus able to come up with a more viable solution. This system can enhance the process of decision making, can increase the organizational control and can improve the personal efficiency and interpersonal communication. Nevertheless, it quantitatively and qualitatively aids to the satisfaction of the decision maker as well. In addition, the decisions taken have the effects of reduction in costs and risks, increasing revenue and improving assets usage efficiency.

1.3 Applications of DSS

There are numerous applications of DSS in and around the industrial, business and research sectors. A self predicting system was made aiming to assess accidental risks in building project sites. The results from the same was astonishingly significant in the field of risk assessment. In order to mitigate the risks in maintenance activities and make safety certain, another support system is developed. Yet another proposition of a web-based decision support system (WDSS) to predict potential risks at shipyards [11] has been made that aims to provide a set of preventive policies to avoid risks at the site. The clinical field is not left untouched by the revolutionary impact of support systems. A clinical decision system to suggest appropriate rehabilitation for injured workers [12] and a WDSS suggesting automatic diet prescription for the patient [13] has been built. Large is it's impact in environmental field as well. A knowledge-based support system for risk assessment of groundwater pollution [14] and measuring it's impact on groundwater quality.Decision support systems has made their impact on occupational, industrial, clinical, environmental and technical realms to a great extent.

1.4 Research issues

Based on the review of literature on the application of DSS mentioned above, some of the research issues have been identified. They are given below:

(i) In most of the cases of decision making, information and overall understanding of the system play an important role. For human, it is a very difficult task to derive decisions from the information lying within a huge amount of data with a stipulated time
(ii) Decision making often demands a substantial amount of human labor which consequently consumes time and manpower
(iii) Decisions taken by humans might be biased and inaccurate in some specific conditions, and in most of the time, they depend largely on human judgments or experience

1.5 Contributions

Realizing the need for research on some of the issues as identified above, the present work attempts to contribute in the following ways:

(i) A DSS has been developed in the domain of occupational accident analysis.
(ii) Decisions from the accident data for the prediction of incident outcomes i.e., injury, near miss, and property damage have been made automated.
(iii) To the best of authors' knowledge, development of DSS for occupational accident analysis can be regarded as an initiative in the steel industry.

The rest of this paper is organized in three sections as follows. In Section 2, development of DSS has been discussed along with the methods or models embedded in the system. Section 3 presents a case study of an integrated steel plant. In Section 4, results obtained from the DSS are presented and discussed. Finally, conclusions are drawn with recommendations for further research in Section 5.

2 Methodology

In this section, a short description of DSS building and the models i.e., algorithms embedded in the system has been provided in the following section. The overall proposed DSS structure has been depicted in Fig. 1

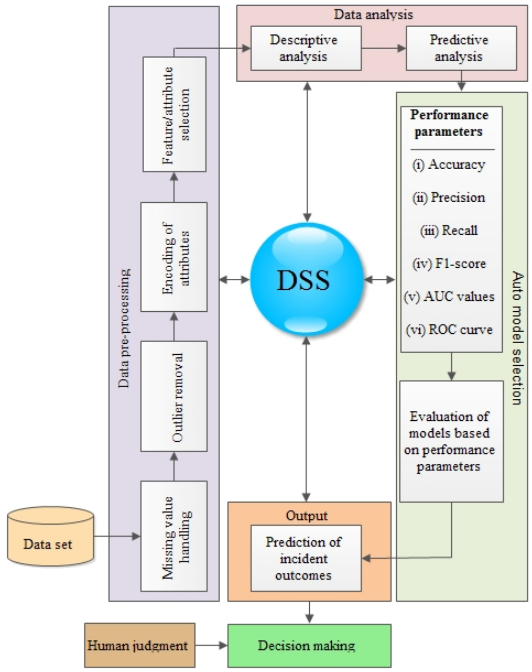


Fig. 1: Overall proposed Decision Support System (DSS) structure.

2.1 DSS building

The DSS framework is built entirely on Python, an open-source, interactive, object-oriented, and high-level programming language. It is based on a Graphical User Interface (GUI), which includes visual aids such as graphs. It consists of a series of back-end programming codes of machine learning algorithms intended for the analysis of accident data. The structure of DSS with Pyhton libraries is depicted in Fig. 2.

2.2 Methods used

(i) **Random Forest:** Random forest (RF) is a popular classification algorithm, which can handle non-linear data efficiently along with linear ones. It belongs to the broader class of decision tree family called *ensemble learning* algorithm. Ensemble learning involves the combination of several classifiers or models to solve a single prediction problem. It works by generating multiple classifiers/models which learn and make predictions independently. Those predictions are then combined into a single (mega) prediction that should be as good or better than the prediction made by any one of the classifiers. The forest chooses the classification having the most votes (over all the trees in the forest) and in case of regression, it takes the average

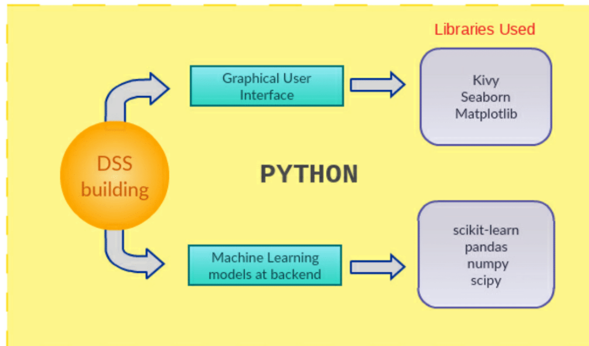


Fig. 2: Building of complete DSS using Python libraries

of outputs by different trees. One of the major benefits of this algorithm is the ability of handling large amount of data with higher dimensionality. The detailed description of the algorithm is kept beyond the scope of the paper. Interested readers may refer to [15–17].

(ii) **Support Vector Machines:** Support Vector Machine (SVM) is a supervised learning algorithm, which can be used for both classification or regression types of problems. In this algorithm, each data item is plotted as a point in n-dimensional space (where n is number of features) with the value of each feature being equal to the value of a particular coordinate. Then, classification task is performed by finding the hyper-plane that differentiates the two classes distinctively. Support vectors are simply the co-ordinates of individual observation. SVM is a frontier which best segregates the two classes (hyper-plane/ line). In SVM, it is easy to have a linear hyper-plane between these two classes. SVM uses kernels which are basically the functions taking low dimensional input space and transforming it to a higher dimensional space i.e., it converts non-separable data to separable data. It is found to be the useful method for non-linear separation problem. For understanding the mathematics behind this algorithm, one may refer to [18–20].

3 Case Study

In this study, to validate the proposed model, data were retrieved from the integrated steel plant in India, and the results of prediction of incident outcomes were obtained from the classifiers, namely random forest and SVM embedded in the DSS. A short description on data set and its pre-processing technique has been provided below.

Table 1: A short description of data set used.

Attributes	Description	Data type
Date of Incident	Date on which the incident took place	Text
Month	Month of the occurrence of incident	Catagorical
Division	Division in which the incident occurred	Catagorical
Department	Department involved in the incident	Catagorical
Incident outcomes (injury, nearmiss and property damage)	Outcomes of the incident	Catagorical
Injury types	Types of injury occurred to victim(s)	Catagorical
Primary causes	Cause of the occurrence of the incident	Catagorical
Brief Description of Incident	Description of how the incident happened	Text
Status	Current status of the incident	Catagorical
Event	Events which led to the incident	Text
Working Condition (single or in group)	Condition in which the victim was working Catagorical	
Machine Condition (working or not working state)	Condition of the machine when the incident occurred Catagorical	
Observation Type	Type of obseravtion	Catagorical
Employee Type	Contractor/Employee	Catagorical
Serious Process Incident Score	Score corresponding to the seriousness of the incident	Numerical
Injury Potential Score injury to the victim	Score corresponding to the seriousness of Numerical	
Equipment Damage Score	Score corresponding to the damage to equipment	Numerical
Safety Standards time of occurrence	Standard of safety at the incident scenario at the Catagorical	
Incident Type	Type of incident	Catagorical
Combined_SOP	Combination of SOP Adequacy, SOP Compliance, SOP Availability, SOP Not Availability	Catagorical

3.1 Data set

To validate the results from the developed DSS, accident data collected from an integrated steel plant were used. The data set consists of 9478 incident reports describing accident consequences like injury, near miss, and property damage. The attributes of the data set used in this study are listed in Table 1 with description and types.

3.2 Data pre-processing

Once the data were obtained, they were pre-processed since the quality of results of the analysis is directly proportional to quality of data. Missing data, and other inconsistencies were removed from the raw data set. Thereafter, the data were coded for easy handling during and after the analyses.

4 Results and Discussion

In this secton, the results of the two classification algorithms i.e., SVM and RF are discussed. Based on their performance in terms of classification accuracy, the classifiers are compared and automatically better one is selected in DSS platform, which is later used for the prediction of incident outcomes.

The analytical platform for this study is DSS, which is built by Python. There were seven basic libraries used for the building of DSS structure. They are 'Kivy', 'Seaborn', 'Matplotlib', 'Scikit-learn', 'Pandas', 'numpy', and 'Scipy'. Using these libraries, the 'Homepage' of the DSS is so developed that it is able to import the data set from the drive. Once the data is loaded, features or attributes of the data set can be selected manually. In addition, for feature selection, algorithms including Chi-square, and RF are also embedded into the system. Indeed, there are two basic operations are included as data pre-processing tasks in DSS. First is the descriptive analysis and second is the predictive analysis. Once the features are selected for analysis, descriptive analyses were performed for each of them. For example, in Fig. 3, graphs from descriptive analysis are displayed which explains the frequency of primary cause happened in each of the department with employee types.

In the window of predictive analysis in DSS, two basic classifiers i.e., SVM and RF are included. Using 10-fold cross validation, classification accuracy of both SVM and RF algorithm is calculated, which produce 70.56% and 99.34% accuracy, respectively. A provision has been made automatically to select the best one from a set of classifiers. As a result, in this case, RF has been selected. Apart from accuracy, F1-score, AUC values, and ROC curve have been shown in Fig. 4(a,b). It is noteworthy to mention that provisions are so made in DSS that comparison among classifiers can also be performed using different performance parameters like F1-score, AUC values.

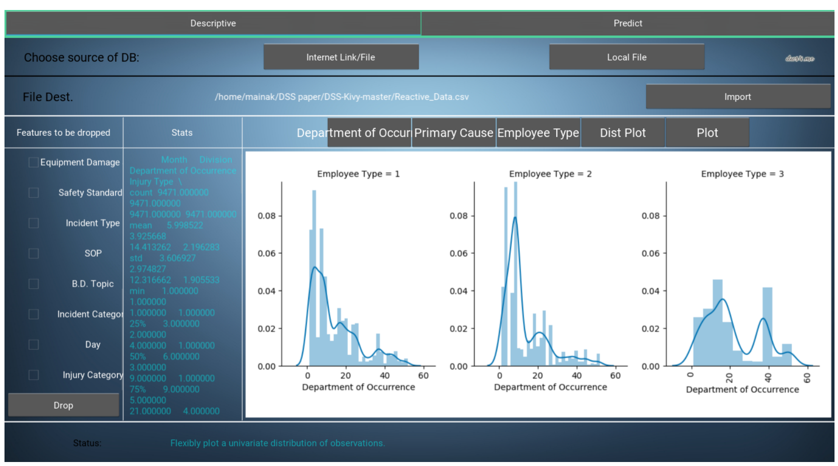
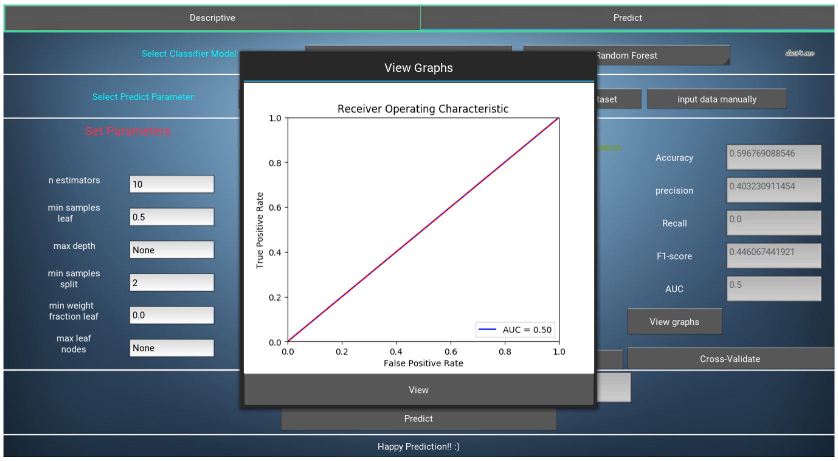


Fig. 3: Display of results window of DSS showing the frequency of 'Primary cause' to accident across various 'Department of Occurrence' considering different employee types.



(a)

(b)

Fig. 4: (a) Display of results window of DSS showing ROC curve for incident outcome prediction; (b) Display of results window of DSS showing accuracy, F1-score, and AUC values.

5 Conclusions

The present study attempts to develop a DSS for occupational accident analysis and prevention. The developed DSS can reduce a substantial amount of human labor during the task of data pre-processing including missing value handling, outlier or inconsistency removal, data analysis through automatic selection of model (i.e., classifier) based on a certain performance parameter usually set by the user. The developed DSS can make an unbiased estimation of results in terms of classification accuracy of incident outcomes i.e., injury, near miss, and property damage from the historical accident data. Out of the two classifiers used in this study i.e., SVM, and RF, the later one performs better with accuracy 99.34% obtained from 10-fold cross validation. Therefore, the initial attempt to build such a DSS for accident analysis in steel industry has full potential to help decision makers take more prudent and insightful decisions.

As future scopes of the present study, one may opt for automating the data analysis process entirely from import of data to the export of results. Another important task, which can be made DSS more strengthened, is to use optimization algorithms for determining optimal value of parameter of the classifiers. Rule-based analysis, which is very important for occupational accident analysis, can be embedded also in the DSS structure.

References

1. Ferguson, R.L., Jones, C.H.: A computer aided decision system. Management Science 15(10), B–550 (1969)
2. Sarkar, S., Vinay, S., Maiti, J.: Text mining based safety risk assessment and prediction of occupational accidents in a steel plant. In: Computational Techniques in Information and Communication Technologies (ICCTICT), 2016 International Conference on. pp. 439–444. IEEE (2016)
3. Krishna, O.B., Maiti, J., Ray, P.K., Samanta, B., Mandal, S., Sarkar, S.: Measurement and modeling of job stress of electric overhead traveling crane operators. Safety and health at work 6(4), 279–288 (2015)
4. Gautam, S., Maiti, J., Syamsundar, A., Sarkar, S.: Segmented point process models for work system safety analysis. Safety science 95, 15–27 (2017)
5. Sarkar, S., Patel, A., Madaan, S., Maiti, J.: Prediction of occupational accidents using decision tree approach. In: India Conference (INDICON), 2016 IEEE Annual. pp. 1–6. IEEE (2016)
6. Sarkar, S., Vinay, S., Pateshwari, V., Maiti, J.: Study of optimized svm for incident prediction of a steel plant in india. In: India Conference (INDICON), 2016 IEEE Annual. pp. 1–6. IEEE (2016)
7. Sarkar, S., Lohani, A., Maiti, J.: Genetic algorithm-based association rule mining approach towards rule generation of occupational accidents. In: International Conference on Computational Intelligence, Communications, and Business Analytics. pp. 517–530. Springer (2017)
8. Sarkar, S., Verma, A., Maiti, J.: Prediction of occupational incidents using proactive and reactive data: A data mining approach. In: Industrial Safety Management, pp. 65–79. Springer (2018)

9. Verma, A., Chatterjee, S., Sarkar, S., Maiti, J.: Data-driven mapping between proactive and reactive measures of occupational safety performance. In: Industrial Safety Management, pp. 53–63. Springer (2018)
10. Sarkar, S., Pateswari, V., Maiti, J.: Predictive model for incident occurrences in steel plant in india. In: 8-th ICCCNT, 2017. pp. 1–5. IEEE (2017)
11. Cebi, S., Akyuz, E., Sahin, Y.: Developing web based decision support system for evaluation occupational risks at shipyards. Brodogradnja 68(1), 17–30 (2017)
12. Gross, D.P., Zhang, J., Steenstra, I., Barnsley, S., Haws, C., Amell, T., McIntosh, G., Cooper, J., Zaiane, O.: Development of a computer-based clinical decision support tool for selecting appropriate rehabilitation interventions for injured workers. Journal of occupational rehabilitation 23(4), 597–609 (2013)
13. Caballero-Ruiz, E., García-Sáez, G., Rigla, M., Villaplana, M., Pons, B., Hernando, M.E.: A web-based clinical decision support system for gestational diabetes: Automatic diet prescription and detection of insulin needs. International Journal of Medical Informatics 102, 35–49 (2017)
14. Uricchio, V.F., Giordano, R., Lopez, N.: A fuzzy knowledge-based decision support system for groundwater pollution risk evaluation. Journal of environmental management 73(3), 189–197 (2004)
15. Breiman, L.: Random forests. Machine learning 45(1), 5–32 (2001)
16. Wang, B., Gao, L., Juan, Z.: Travel mode detection using gps data and socioeconomic attributes based on a random forest classifier. IEEE Transactions on Intelligent Transportation Systems (2017)
17. Ahmed, M., Rasool, A.G., Afzal, H., Siddiqi, I.: Improving handwriting based gender classification using ensemble classifiers. Expert Systems with Applications 85, 158–168 (2017)
18. Poria, S., Peng, H., Hussain, A., Howard, N., Cambria, E.: Ensemble application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis. Neurocomputing (2017)
19. Dumais, S., et al.: Using svms for text categorization. IEEE Intelligent Systems 13(4), 21–23 (1998)
20. Chen, T., Lu, S.: Subcategory-aware feature selection and svm optimization for automatic aerial image-based oil spill inspection. IEEE Transactions on Geoscience and Remote Sensing (2017)