INTERNATIONAL JOURNAL OF SUSTAINABLE CONSTRUCTION ENGINEERING AND TECHNOLOGY
VOL. 14 NO. 4 (2023) 40-48

IJSCET

http://publisher.uthm.edu.my/ojs/index.php/ijscet

ISSN : 2180-3242 e-ISSN : 2600-7959

International Journal of Sustainable Construction Engineering and Technology

Prediction of Safety Performance by Using Machine Learning Algorithms: Evidence from Indian Construction Project Sites

Rajaprasad SVS[1*], Rambabu Mukkamala[1]

[1]National Institute of Construction Management and Research,
Shamirpet, Hyderabad, 500101, Telangana State, INDIA

*Corresponding Author

DOI: https://doi.org/10.30880/ijscet.2023.14.04.004
Received 21 November 2022; Accepted 15 August 2023; Available online 15 October 2023

Abstract: The construction industry in India happens to be the second most contributor to its gross domestic product (GDP) but high rates of accidents and fatalities have tarnished the image of the industry in India. To enhance the importance and alertness among the stakeholders in construction project sites, the present study proposes a framework for predicting safety performance. In this retrospective study, the data pertaining to the 69 construction project sites across India from January, 2021, to July, 2022 was analysed. The data analysis was conducted in two phases, in the first phase of the study the efficiency of project sites was computed by implementing data envelopment analysis (DEA). In the second phase, the results of the first phase are utilized to predict the safety performance of construction sites by applying four machine learning (ML) algorithms. In the first phase of the study, three input and three output variables were considered to compute the efficiency of the project sites. Results of four ML classifiers revealed that the random forest classifier with high recall percentage of 95.0 is considered the best in predicting the safety performance. Finally, the results indicate that the ML classifiers enable a good accuracy level in predicting the safety performance of project sites. Among the four ML classifiers, notably the Random Forest Classifier enables identifying the inefficient project sites and advising the site management to implement control measures. Finally, a safety performance prediction tool was developed to understand the results.

Keywords: Occupational health and safety**, e**fficiency, machine learning, prediction

1. Introduction

The contribution of the construction industry in India is 6% of the national GDP. The safety performance of construction project sites is in declining trend even after implementing the safety systems. Construction activity was increased several folds due to infrastructure development and Government policies on growth of urbanization. Owing to this OHS matters are rising and causing concern to all the stakeholders. Construction works progress with time and the OHS measures implementation are required in parallel to avoid the accidents. In India, the safety performance in construction industry is not up to the desired level due to employing migrant workers, lack of commitment from management and enforcement from authorities (Berger, 2000). Fewer construction organizations are achieving the better safety performance due to client involvement in following up the implementation of safety programs and organizing the safety training to employees (Hinze & Gambatese, 2003). Safety performance of a construction organization is directly depending on the management commitment in introducing the OHS practices and reviewing the sufficiency of existing practices periodically (De Silva N & Wimalaratne, 2012). Communicating the safety information and instructions is multi directional and it is useful to predict the safety performance.

In the past, many studies have been conducted to explore the impact of safety initiatives on safety performance in the construction industry. Studies conducted in Asian countries reveal that lack of safety training, dearth in qualified

**Corresponding author: rajaprasad@nicmar.ac.in*

safety professionals, poor record keeping and insufficient allocation budget to fulfill the safety activities are the determinants that effect the performance (Alkilani et al., 2013). The safety performance in Oman is low due to lack of understanding on cost of accidents, effect of thermal stress and OHS regulations (Umar & Egbu, 2018). Worker and organizational factors have a considerable influence on safety performance compared to environmental factors in construction organizations in Iran (Mehdi & Shadiya, 2019). In the past, researchers conducted studies to examine the impact of safety climate on performance but there is no evidence of agreement achieved yet (Zohar,2010). The safety performance in Malaysian in construction industry is impacted by illiterate workforce, lack of awareness and failure to conduct safety trainings regularly (Mohd Nasrun et al., 2016). The organization culture will have predictive power and impact on safety performance. Imparting safety training to employees in construction sites is part of OHS management system in order to accomplish better safety performance. Safety trainings are vital in guiding the employees to adhere to the safe practices.

Recording and maintaining safety statistics by the organizations are instrumental in launching proactive safety initiatives. In India, the safety performance is evaluated in terms of frequency and severity rate and these indicators fail to reflect the overall safety scenario (Wanberg,2013). The insufficient information about the safety performance is a hurdle due to which the construction organizations are focusing the implementation of proactive indicators as tools to represent predictors for safety performance. The accident statistics are useful to gauge the safety climate of an organization which in turn predict the safety performance (Cooper & Phillips, 1994). Management commitment, near-miss reporting, worker participation in safety activities, safety auditing safety risk assessment are leading indicators for predicting future safety. The factors that influence safety performance in construction sites are climate, culture, attitude, budget and employee behavior (Dinesh & Junwu, 2020).

Safety performance can be evaluated based on consequences/ negative indicators, compliance based or leading indicators. It’s not simple to correlate both the indicators due to complexity in construction project sites (Sevar &D, Salahaddin, 2019). Safety performance indicators can be considered as filters through which reality is depicted and acknowledged. In the developing countries, the data pertaining to the injury and severity rates are not recorded at the organization level, so the data cannot be adopted to benchmark the safety performance of project sites (Raheem & Hinze, 2012). Lagging indicators are adopted to gauge the safety performance compared to leading indicators, which aims in minimizing the accidents. Lagging indicators still cannot be excluded from performance measurement since they provide useful facts about the enhancement of safety at the workplace (Swuste et al., 2016). Accident information in construction sites indicate that there is desperate need to minimize the prevalence of injuries and require a mechanism for evaluating the safety performance.

In light of existing literature, the importance of retrospective data in conducting the safety performance analysis is highlighted. There is no common conclusion in the past studies regarding the metrics for safety measurement and treating the measurement data successfully. The aim of the study is to evaluate the safety performance construction project sites using the DEA based on retrospective data; and to test and compare four ML classifiers to develop a predictive model.

2. Methodology

This study was performed in two phases by collecting the data from 69 Indian construction project sites between January, 2021, to July, 2022. Data envelopment analysis (DEA) and four ML classifiers are adopted to predict the safety performance.

2.1 Frame Work of the Study

The study was conducted in different stages and framework of the study is shown in Figure 1.

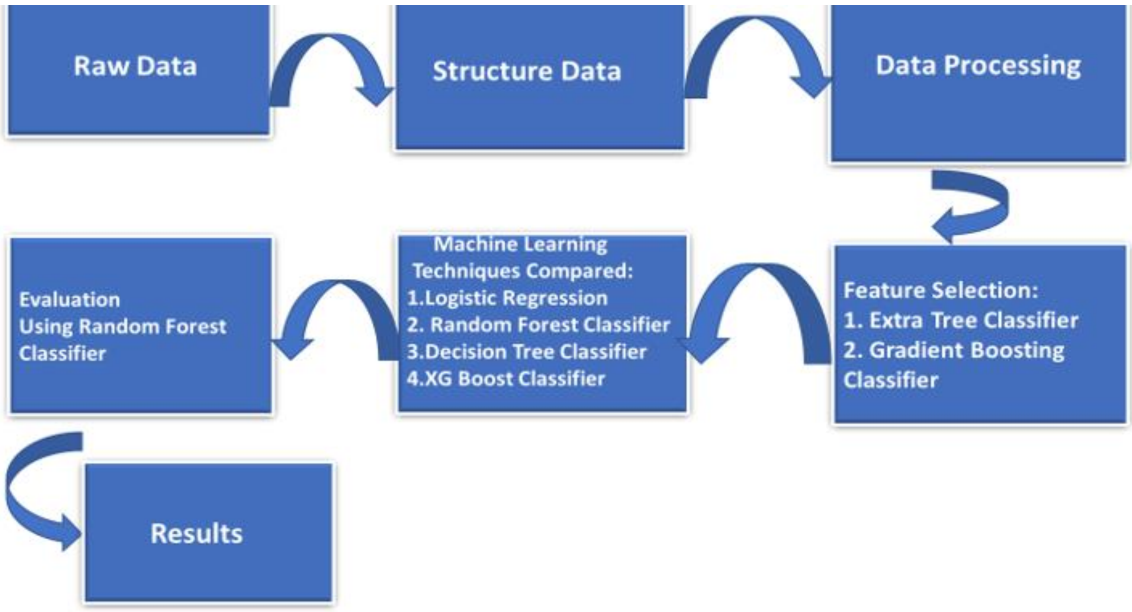

Fig.1 - Frame work of the study

2.2 Data Collection

The data required for the study was collected by directly contacting the officials of the safety department of project sites and briefed the objectives. The inputs considered for the analysis are number employees working at the site, safety team, percentage of expenditure towards safety activities of project cost and; the outputs are the injuries/ accidents, number of working days lost and the cost of damages. Many of the site safety managers are not positive to disclose the information regarding input/ outputs. Few project sites though furnished the data with a condition to utilize the data for the present study only. The information was collected from 69 project sites across India between the period January, 2021, to July, 2022. Owing to negative factors involved in the outputs, the data was normalized before conducting the analysis.

2.3 Data Envelopment Analysis

DEA is a mathematical programming technique that has applied in real world situations for computing the performance of identical units. DEA is a methodology based on extension of linear programming and effectively implemented for It was originally developed for performance measurement and successfully employed for evaluating the relative performance of firms that utilize the number of similar inputs to produce the number of similar outputs. DEA has been implemented to evaluate the safety performance in coal mines in China and construction and allied industries in India (Lei & Ri-jia, 2008; Beriha, 2011). The organizations are named as decision-making units (DMUs) and DEA is appropriate method computing the relative performance of group of organizations. The inputs are converted into outputs in a DMU whose performance is evaluated. DEA is a linear programming-based tool for measuring the relative efficiency of each unit. The Charnes, Cooper and Rhodes (CCR) model is computes the efficiency of DMUs by calculating the ratio of weighted sum of its outputs to the weighted sum of its inputs.

2.4 Data Pre-processing

Most real-world data set for machine learning are very likely to be missing, inconsistent, erroneous data and presence of outliers because of their heterogeneity. Pre-processing refers to the modifications exercised to data set prior to presenting it to the algorithm. It's a tool applied to transform the raw data into a clean data set. In this study, partial data relating to the inputs and outputs effecting the safety performance were found from 20 project sites were not considered for the analysis.

2.5 Feature Selection

In this study, the feature selection is important to predict the safety performance. Feature importance technique is applied to finalize the importance variables by utilizing a trained supervised classifier. The feature selection technique was implemented with Extra Tree Classifier and recursive feature elimination (RFE) using Gradient Boosting Classifier.

2.6 ML Classification Algorithms

Four ML algorithms, including logistic regression (LR), random forest (RF), decision tree (DT) and extreme gradient boosting (XGBoost) were applied in this study for predicting the project sites safety performance. The hyperparameters involved in models were optimized by using the grid search method. The grid search method is effective for refining the variables during the training stage and improve efficiency of the ML classifiers (Afrash et al., 2022).

2.6.1 LR Classifier

This is a ML method that is applied to answer classification problems and applied to predict the possibility of a binary dependent variable. The dependent variable in logistic regression is a categorical with data coded as one (efficient) or zero (inefficient). The objective of method is to determine a link between variables and the possibility of precise result and; adopts sigmoid function instead of a linear function.

2.6.2 RF Classifier

Random forests classifier was a method of combining several decision trees using bagging. The two vital points incorporated with random forest model are the consistency of estimators that are randomly generated to assure convergence. RF classifier adhere definite procedure for tree growing, tree combination, self-testing and post-processing, resilient to overfitting and balanced in the presence of outliers than other ML algorithms (Zurada et al., 2011; Yeşilkanat, 2020).

2.6.3 DT Classifier

DT is a flow- chart like structure, where each internal node represents a test on a variable, each branch denotes the result, and each leaf node represents a class (Qing-yun, 2016). The benefits of this classifier are its power to categorize quantitative and qualitative parameters, better understanding through tree structure, and fulfil better depiction of classification of the dataset (Afrash et al., 2022).

2.6.4 XGBoost Classifier

XGBoost classifier succeeds the other ML algorithms in respect of accuracy, training speed, normality assumption of the input variables, elucidating and require minimum tuning of variables. It's also a regression method and the association between the input and output variables need not be linear invariably. Owing to high scalability of this classifier, the time required is less compared to other ML methods and uses less memory (Shibaprasad & Shankar, 2012).

2.7 Performance Metrics of Classifiers

The efficacy and performance of the four ML classifiers are evaluated with regard to accuracy (total number points classified exactly), recall (to ascertain the completeness of the results), precision (usefulness of the results), specificity (part of true negatives classified exactly), f -measure (likelihood that a positive prediction is true), receiver operating characteristic (ROC) curve to compare and evaluate different classification algorithms. The values of the performance metrics are evaluated by using the "equations 1 to 4".

$$Accuracy = \frac{(\mathrm{TP} + \mathrm{TN})}{(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN})} * 100 \tag{1}$$

$$Recall = \frac{TP}{(TP + FN)} * 100 \tag{2}$$

$$Precision = \frac{TP}{(TP + FP)} * 100 \tag{3}$$

$$f - measure = \frac{2 * Precision * Recall}{(Precision + Recall)} * 100 \tag{4}$$

True Positive (TP) is construed as the model predicted positive class and it is True, False Positive (FP) is interpreted as the model predicted positive class but it is False, False Negative (FN) is read as the model predicted negative class but it is False and True Negative (TN) is interpreted as the model predicted negative class and it is True.

3. Results

In the first phase of the analysis, the safety performance of 69 project sites was computed by applying the DEA constant return scale model. DEA considers a DMU as the unit for transforming the inputs into outputs. The information relating to the inputs and outputs considered in the study were discussed in detail in data collection section. The safety performance of project sites with efficiency score of one are treated as efficient and any score below one is treated as inefficient. The input and output data were run by using the DEA OS software. The descriptive statistics of the results of DEA are shown in Figure 2.

	Employees	Team	Expenditure	Accidents	Mandays	Damages	Efficiency
count	69.000000	69.000000	69.000000	69.000000	69.000000	69.000000	69.000000
mean	433.202899	3.913043	0.651725	0.629029	0.714246	0.733667	0.057971
std	116.294072	1.039501	0.085535	0.126394	0.101045	0.094751	0.235401
min	195.000000	2.000000	0.498000	0.364000	0.488000	0.486000	0.000000
25%	350.000000	3.000000	0.588000	0.534000	0.644000	0.688000	0.000000
50%	415.000000	4.000000	0.644000	0.636000	0.712000	0.733000	0.000000
75%	510.000000	5.000000	0.712000	0.712000	0.786000	0.789000	0.000000
max	715.000000	6.000000	0.875000	0.988000	0.966000	0.964000	1.000000

Fig.2 - Descriptive statistics of DEA results

3.1 Data Pre-processing

The data was analysed for missing values and outliers and the results show that the data is free from errors. In the next step, the data was analysed for selection of variables influencing the safety performance of the project sites.

3.2 Feature Selection by Feature Importance Technique

Feature Importance is a method to compute a score for the six independent variables used in the study. A higher score of a variable will have a positive effect on the model that is being used to predict. The variables were selected by a feature importance technique with the help of Extra Tree Classifier and recursive feature elimination (RFE) using Gradient Boosting Classifier. The results of the variable selection were shown in Table 1 and all the six independent variables are considered for further analysis basing on the results of feature importance.

Table 1 - Results of feature selection

Variable	Extra Tree Classifier	Gradient Boosting Classifier
Employees	0.1808237	True
Team	0.1532362	True
Expenditure	0.1735054	True
Accidents	0.1790167	True
Man days	0.1748305	True
Damages	0.1411733	True

3.3 ML Classifiers

The data was analysed by using four ML classifiers by considering 51 project sites information (75%) for training and the rest 18 (25%) for testing. The results of performance metrics of the classifier are shown in Table 2. The best predictive model is identified based on recall percentage. In safety performance prediction, the six variables are directly related to expenditure incurred for true or false in confusion matrix. Owing to this, RF is associated with highest Recall percentage which in turn specify that this model is best predict the low occurrence like accidents, man days lost and damages. According to Table 2, the confusion matrix (Figure 3) and ROC curve (Figure 4) for RF classifier with accuracy 95.0%, precision 100%, recall 95.0%, f1 score 97.5% and area under curve (AUC) score 96.0% achieved the best performance in predicting the safety performance of construction project sites. To overcome the imbalance of dependent variable, random over-sampling techniques were used in the analysis.

Table 2 - Performance metrics of ML classifiers

ML classifier	Accuracy (%)	Precision (%)	Recall (%)	f1- score (%)	AUC score (%)
LR	88.0	82.8	73.6	86.0	94.0
RF	95.0	100	95.0	97.5	98.0
DT	86.0	78.0	66.8	88.0	86.0
XGBoost	88.0	86.0	81.2	90.0	88.0

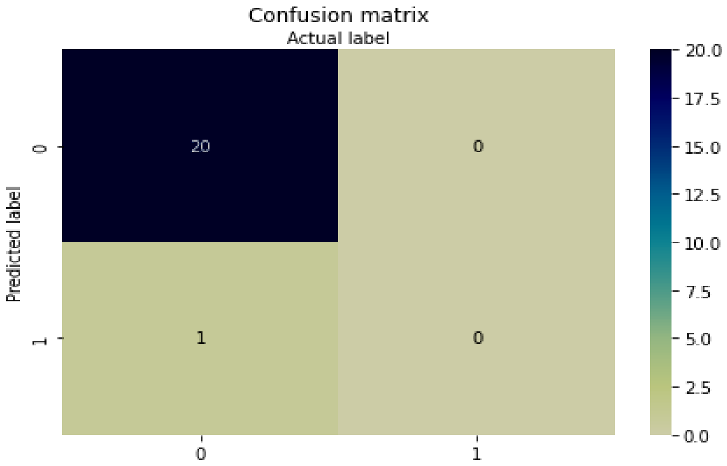

Fig. 3 - Confusion matrix of RF classifier

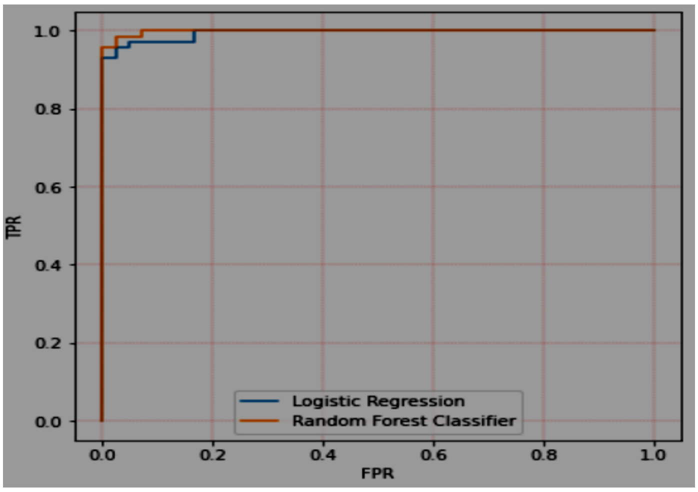

Fig. 4 - ROC curve of RF and LR classifier

3.4 Safety Performance Prediction Tool

The most important and final step in building a ML classifier is developing a tool to understand and interpret the results. In practice there are many frameworks available for developing a ML classifier model on the web. In this study, a new framework known as Streamlit was implemented to develop the ML model. The safety performance of the construction organizations by using the RF classifier code is run on the Python Charm integrated development environment. The user interface safety performance prediction is developed using new Framework known as Streamlet and the logic is applied by using the Python Programming. The developed safety prediction tool is depicted in Figure 5. To interpret the safety performance, if the result approximate to 0, then the organization is “Not efficient” and in case the result approximate to 1, then the organization is “Efficient”. The prediction tool has been tested by using random data from the actual data set used in the analysis and compared with the results generated by the tool. Based on the comparison, the accuracy of the result is more than 90%.

Fig. 5 - Safety performance prediction tool

4. Discussion

In the analysis, four ML classification algorithms namely LR, RF, DT and XGBoost were trained using the variables that affect the safety performance. The past studies related to the parameters influencing the safety performance by applying the ML classifiers was analysed. The indicators included in the reviewed studies are number of accidents, length of service for each injured worker, the type of construction, injury occurrence, loss time injury, hazard assessment, weighted safety inspection score (Jafaria et al., 2019; Shayboun et al., 2020; Ahmed et al., 2021). In the present study, Extra Tree Classifier and RFE using Gradient Boosting Classifier are applied for feature selection and the six variables considered in the study are important for analysis using the ML classifiers. The study proved that the RF classifier increase analytical accuracy and safety performance prediction efficiency. Therefore, four ML classifiers were trained using selected variables. Finally, the RF classifier with 95% accuracy, 100% precision, 95% recall, 97.5% f-measure and auc score 96.0% outperformed from the results of other algorithms. RF classifier outperformed the Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naïve Bayes (NB) and DT in predicting the occupational risk injury severity with higher accuracy and f1-score (Khairuddin et al., 2022).

The performance metric recall is considered as an opt measure for selecting the leading indicators influencing safety performance measurement. For instance, in safety performance analysis, if the damages due to a mishap (True True) is predicted as no damages (True False), the outcome can lead to incurring huge amount for the organization. Due to this, recall is considered as a performance metric to ascertain the best model. The classifiers KNN and NB classifiers are considered as best models based on recall percentage in identifying the leading indicators effecting the safety performance (Jafaria et al., 2019). Recall is the best measure to predict the low frequency class, for instance the number accidents, safety team and, man days lost.

The f1- score in the present study is high in case of RF classifier compared to other three ML classifiers. The high f1- score is an indication of robustness of the model (Chadyiwa et al., 2022). The RF model yielded best results in predicting the factors influencing occupational injuries (Sarkar et al., 2017). The performance of ML algorithms will be better when the number of observations is nearly equal in each class (Witten et al., 2016). The results DEA indicate that a total of 15 "efficient" and 54 "inefficient" project sites were noticed and due to this imbalance, the performance of algorithms will be affected. To overcome the imbalance, random over-sampling techniques were used to reduce the negative impact (R Core Team, 2018). The predicted model by applying the ML classifiers is proven efficient in evaluating the safety leading indicators which are useful to construction organizations to predict safety risk (Clive et al., 2018). RF and stochastic gradient tree boosting were applied to predict construction injury attributes namely type of injuries and parts of the body affected (Tixier et al., 2016). The DT classifier and ML techniques are useful in predicting the safety behaviour classification (Goh & Chua, 2013). The results of the ML models in the present study exhibited satisfactorily in predicting the safety performance of project sites, still there are certain deficiencies that must be addressed. First, due to the analysis of past data, certainly the data consists of duplicate and insignificant values. Second, the data is collected from various construction sites involved in variety of works restricts in drawing the conclusions applicable to the industry as a whole.

5. Conclusion

In the past several studies were conducted to gauge the safety performance of construction organizations. In this study, DEA was applied to the data to classify the sites as efficient or inefficient and the results of DEA are used to develop a model the safety performance by adopting the ML classifiers. These models are useful to the management of construction organizations in allocating sufficient budget for safety activities, minimizing the man days lost and accidents. The dataset was collected from the construction project sites across India. All the six variables are having significant impact on safety performance. After the analysis, Random Forest (RF) classifier yielded the best, outperforming the other ML algorithms with a Recall of 95%.

In the past studies, which have applied statistical tools but this analysis has adopted the ML approaches to establish a safety performance prediction model. The results are useful to the stakeholders of construction organizations to implement accident prevention controls to enhance the safety performance, and guide the site management to organize the safety training programs, effective implementation of engineering controls and providing the healthy work environment which in turn minimize the accident costs and increase the morale of employees.

As application of ML in the domain of construction safety performance modelling is gaining momentum and the analysis are useful in integrating with the optimization/statistical techniques which provide base for future research. However, it should be noted that the analysis is considered six independent variables to develop a model, this is mainly due to the reason that the lack of record keeping and enforcement from the authorities. In future, similar studies is needed to validate the RF model by considering the greater number of features and project sites and the results are useful to generalize its application to the construction industry.

Acknowledgement

The authors would like to thank and acknowledge the National Institute of Construction Management and Research, Shamirpet, Hyderabad, India for all kinds of support given.

References

Afrash, M.R. Kazemi -Arpanahi, H. & Nopour, R. (2022). Designing an Intelligent Decision Support System for Early Prediction of Intubation Need among COVID -19 Hospitalized Patients. *J Environ Health Sustain Dev*, 7 (3). pp. 1698 -707.

Ahmed, O.O. Anuoluwapo, O.A. & Lukumon, O.O. (2021). Machine learning predictions for lost time injuries in power transmission and distribution projects. *Machine Learning with Applications*,100158. pp.1-12.

Alkilani, S.Z. Jupp, J.& Sawhney A. (2013). Issues of construction health and safety in developing countries: A case of Jordan. *Journal of Construction Economics and Buildings*, 13(3). pp. 141-56.

Berger, J (2000). Safety Coordination and Quality in Construction. *Proceedings of International Conference of CIB Working Commission 99 and Task Group 36*. Milan, Italy. June 22-23, 2000. pp. 51-60.

Beriha, G.S. Patnaik, B. & Mahapatra, S.S. (2011). Safety Performance Evaluation of Indian Organizations Using Data Envelopment Analysis. *Benchmarking: An International Journal*, 18(2). pp.197-220.

Chadyiwa, M. Kagura, J. & Stewart, A. (2022). Investigating Machine Learning Applications in the Prediction of Occupational Injuries in South African National Parks. *Mach. Learn. Knowl. Extr*, 4. pp. 768–778.

Clive, Q.X.P. Chalani Udhyami, U. & Yang Miang, G. (2018). Safety leading indicators for construction sites: A machine learning approach. *Automation in Construction*, 93. pp.375-86.

Cooper, M.D. & Phillips, R. (1994). Reducing accidents using goal setting and feedback: A field study. *J. Occupational & Organizational Psychology*, 67. pp. 219–40.

De Silva, N. & Wimalaratne, P.L.I. (2012). OSH management framework for workers at construction sites in Sri Lanka. *Engineering, Construction and Architectural Management*, 19(4). pp. 369-92.

Dinesh, S. & Junwu, W. (2020). Prospective Safety Performance in Construction Industries in Nepal. *Jordan Journal of Civil Engineering*, 14(4). pp.458-75.

Goh, Y.M. & Chua, D. (2013). Neural network analysis of construction safety management systems: a case study in Singapore. *Constr. Manag. Econ*, 31(5). pp. 460–70.

Hinze, J. & Gambatese, J. (2003). Factors influencing safety performance of specialty contractors. *Journal of Construction Engineering and Management*, 129(2). pp. 159-64.

Jafaria, P. Mohameda, E. Pereiraa, E. Kanga, S. & AbouRizka, S. (2019). Leading Safety Indicators: Application of Machine Learning for Safety Performance Measurement. *36th International Symposium on Automation and Robotics in Construction*. Banff, Canada. 2019. pp. 501-506.

Khairuddin, M.Z.F. Lu Hui, P. Hasikin, K. Abd Razak, N.A. Lai, K.W. & Mohd Saudi Ibrahim, S.S. (2022). Occupational Injury Risk Mitigation: Machine Learning Approach and Feature Optimization for Smart Workplace Surveillance. *Int. J. Environ. Res. Public Health*, 13962. pp.2-19.

Lei, T. & Ri-jia, D.V. (2018). Efficiency Assessment of Coal Mine Safety Input by Data Envelopment Analysis. *Journal of China University of Mining and Technology*, 18(1). pp. 88-92.

Mehdi, P. & Shadiya, M.S.B. (2019). Determination of Factors Affecting Safety Performance in Iran Construction Industries. *Singapore Journal of Scientific Research*, 9 (3). pp.105-112.

Mohd Nasrun, M.N. Siti Halipah, I. Rohaida, A. Nor Azalina, R. & Fazlin Mohamad, B. (2016). Factor Affecting Safety Performance Construction Industry. *International Review of Management and Marketing*, 6(S8). pp.280-85.

Qing-yun, D. Chun-ping, Z. & Hao, W. (2016). Research of Decision Tree Classification Algorithm in Data Mining. *International Journal of Database Theory and Application*, 9(5). pp.1-8.

R Core Team. (2018). *A language and environment for statistical computing* [Internet], Vienna: R Foundation for Statistical Computing; 2018. Available from online at https://www.R-project.org/.

Raheem, A. & Hinze, J. (2012). Reasons for the Poor Implementation of Worker Safety in the Construction Industry of Pakistan: A Contractor's Prospective. *CIB W099 International Conference, Modeling and Building Health and Safety*. Singapore. 2012. pp. 54–63.

Sarkar, S. Pateshwari, V. & Maiti, J. (2017). Predictive model for incident occurrences in steel plant in India. In Proceedings of the 2017 8th International Conference on Computing, Communication and Networking Technologies, Delhi, India. 3–7 July 2017. pp. 1–5.

Sevar, D. & Salahaddin, N. A. (2019). Comparative Study of Safety Leading and Lagging Indicators Measuring Project Safety Performance. *Advances in Science, Technology and Engineering Systems Journal*, 4(6). pp. 306-312.

Shayboun, M. Kifokeris, D. & Koch, C. (2020). Machine Learning for Analysis of Occupational Accidents Registration Data. *Proceedings of the 36th Annual ARCOM Conference*, UK.7-8 September 2020. pp. 485-494.

Shibaprasad, B. & Shankar, C. (2012). Application of XGBoost algorithm as a predictive tool in a CNC turning process. *Reports in Mechanical Engineering*, 2(1). pp. 190 – 201.

Swuste, P. Theunissen, J. Schmitz, P. Reniers, G. & Blokland, P. (2016). Process safety indicators, a review of literature. *Journal of Loss Prevention in the Process Industries*, 40. pp.162–73.

Tixier, A.J.P. Hallowell, M.R. Rajagopalan, B. & Bowman, D. (2016). Application of machine learning to construction injury prediction. *Autom. Constr*, 69. pp.102–114.

Umar, T.& Egbu, C. (2018). Causes of construction accidents in Oman. *Middle East Journal of Management*, 5(1). pp-21-33.

Wanberg, J. Harper, C. Hallowell, M.R. Rajendran, S. (2013). Relationship between construction safety and quality performance. *Journal of Construction Engineering and Management*, 139 (10). pp. 04013003.

Witten, I.H. Frank, E. Hall, M.A. & Pal, C.J. (2016). *Data mining: Practical machine learning tools and techniques*. Massachusetts. Morgan Kaufmann Publishers.

Yeşilkanat, C.M. (2020). Spatio-temporal estimation of the daily cases of COVID -19 in worldwide using random forest machine learning algorithm. *Chaos Solitons Fractals*, 140. pp.1 -8.

Zohar, D. (2010). Thirty years of safety climate research: Reflections and future directions. *Accid. Anal. Prev*, 42(5). pp.1517–22.

Zurada, J.M, Levitan, A.S. & Guan, J. (2011). Non-conventional approaches to property value assessment. *Journal of Applied Business Research*, 22(3). pp.1-14.