Leading Safety Indicators: Application of Machine Learning for Safety Performance Measurement

P. Jafari[a], E. Mohamed[a], E. Pereira[a], S. Kang[a], and S. AbouRizk[a]

[a]Department of Civil and Environmental Engineering, University of Alberta, Canada
E-mail: parinaz@ualberta.ca, ehmohame@ualberta.ca, estacio@ualberta.ca, sckang@ualberta.ca, abourizk@ualberta.ca

Abstract –

Proactive approaches designed to prevent incidents before they occur are essential for achieving effective safety management. Emerging as an important component of proactive safety management, leading indicators are used to assess and control safety performance. With the aim of reducing the number or severity of worksite accidents, methods capable of predicting future safety performance using leading safety indicators have been developed. However, these methods have been developed for a specific set of leading indicators. This has substantially limited their application in practice, as leading indicators with the greatest impact on safety performance vary considerably between organizations and projects. An approach for predicting accident occurrence on construction sites that can be applied to any combination of leading indicators is proposed to address these limitations. Data used to develop the proposed approach were collected by a construction company from eight construction projects over a period of two years. Feature selection techniques were used to filter the original factors into the most critical subset, which were then used as inputs. Various supervised learning algorithms, namely support vector machine (SVM), logistic regression, and random forest, were then tested to determine which algorithm(s) yielded the highest prediction accuracy. The results demonstrate that the proposed procedure can be used for early recognition of potentially hazardous project characteristics and site conditions regardless of the number or type of leading indicators available within an organization. Research in this area is expected to facilitate the implementation of targeted safety management controls and to improve safety performance.

Keywords –

Safety Leading Indicators; Safety Management; Safety Performance; Machine Learning

1 Introduction

Safety leading indicators are an essential component of proactive safety management, providing valuable information regarding organizational safety performance [1]. Leading indicators, defined by Grabowski et al. [2], Hinze et al. [3], and Kjellén [4], are conditions, events, or measures antecedent to undesirable events that can be used to predict the occurrence of accidents, near misses, or any undesirable safety state. Measuring leading indicators allows practitioners to define a threshold value for metrics below which corrective actions should be taken, with the aim of reinstating the performance above the required level [5]. Accordingly, leading safety indicators have emerged as a more effective alternative than traditional lagging indicators, which are measured after the occurrence of an accident [5].

Safety professionals have contended that the careful selection, measurement, and mitigation of leading indicators can result in real improvements in practice [3]. Practical improvements are attributed to the ability of leading indicators to be linked causally and proactively with safety outcomes in terms of accidents [6]. Criteria for selecting useful leading indicators have been defined [7] and include validity, reliability, sensitivity, representativeness, openness to bias, and cost-effectiveness. Øien et al. [8] have expanded on this, indicating that successful safety indicators should also be measurable on a numerical scale, updated regularly, and reflect selected determinants of overall safety.

Because of the potential interactive effect associated with combining and analyzing various leading indicators, a comprehensive set of leading indicators is believed to provide the best predictive result [9]. Indeed, many researchers, including Garza et al. [10] and Hinze et al. [9], recommend that companies should use a combination of leading indicators to assess safety performance rather than depending on a single indicator. Furthermore, companies should consider indicators outside of the safety department that have been found to reliably predict safety performance, such as those associated with project performance. While construction companies collect a liberal amount of project-performance data to track the overall performance of their projects, they struggle to make use of these data for safety performance assessment purposes.

A number of leading indicator-based methods capable of predicting safety performance have been reported in the literature. However, these methods have been developed using only a specific set of leading indicators. Accordingly, many of these methods are dependent on certain dataset types or formats, making it difficult for practitioners to generalize the application of these methods in practice.

This study proposes an approach capable of proactively assessing safety performance regardless of the number or type of leading indicators available to a specific construction organization or project. The first part of this work is published in [11] where the authors showed that project performance data collected by different departments can be used as inputs for creating machine learning algorithms for safety leading indicators. Using a machine-learning approach, the proposed methodology can be applied to any dataset. In this particular case, the dataset is collected for non-safety associated purposes (e.g., cost, quality, and schedule) to predict the occurrence (i.e., yes/no) of an accident. The proposed approach was used to assess the safety performance of an organization based on data collected from eight projects over a period of two years., Various measures not traditionally associated with safety performance (i.e., quality, cost, and schedule) were included in the analysis to demonstrate the adaptability of the proposed approach. The various machine-learning algorithms were compared.

2 Literature Review

While several theoretical safety leading indicator studies have been conducted [3, 5, 12], few construction companies have successfully implemented programs capable of monitoring safety leading indicators in practice. In cases where companies have implemented such systems, there is little information available in the literature detailing which specific leading indicators have been applied by these companies [9]. Several factors identified by the Construction Industry Institute (CII) and other academic researchers may explain why successful leading-indicator monitoring programs have yet to be widely used in practice (i.e., knowledge gaps).

2.1 Predictive Models in the Safety Domain

Previous investigators have suggested that personal protective equipment (PPE) and proactive detection-based strategies are not sufficient for facilitating a zero injury state [13] and that predictive models of safety performance may play a key role in bridging this gap [14]. Predicting and understanding anticipated changes in safety performance can assist organizations in developing accident mitigation strategies more efficiently and effectively. Indeed, Schultz [13] has demonstrated the ability of one company to reduce injury rates through the application of advanced and predictive analytics together with a growing safety dataset. A number of predictive models have been developed with the aim of proactively forecasting where and when workplace injuries will occur. Ghodrati et al. [14] built various models to predict construction safety outcomes at the macro level. They demonstrated that the number of companies in any liable earnings category predicted safety outcome as well as the number of claims and entitlement claims in all constructed models. They also found a positive and significant relationship between the number of companies in various liable earnings categories and the severity of associated occupational injuries. In another study, Esmaeili et al. [15] used principal component analysis and a linear prediction model to test the validity of risk attributes for predicting safety outcomes. Although their method was able to predict safety outcomes effectively, it was limited to small-scale projects. A Bayesian-network hybrid model has been proposed by Xia et al. [16], to holistically explore safety risk factors in construction projects and predict the probabilities of project safety states.

2.1.1 Machine Learning-Based Models

Machine-learning algorithms function by learning from historical data in a manner that easily complements expert opinion. Although machine learning has been used widely in construction research for more than two decades, its application in the field of construction safety remains limited [17]. Poh et al. [18] presented a machine-learning approach to develop leading indicators that classify construction sites by their safety risk. A machine-learning approach using occupational accident data was also used by Sarkar et al. [19] to predict occupational accident outcomes, such as injuries, near misses, and property damage. Although the results of the studies mentioned above demonstrate considerable promise, the specificity of the datasets that the researchers used to develop these methods renders the widespread application of these methods difficult in practice for organizations with dissimilar data; in particular, generalization is difficult for datasets related to project performance rather than safety performance.

3 Research Methodology

Figure 1illustrates the research methodology. This methodology is a standard data mining process, as

Table 1. Collected variables

Performance Category	Variable/Feature
General	Project ID, Contract Type, Report Date, Contract Change Order (CCO), Outstanding CCO, CCO Submitted to Date, Request for Information (RFI), RFI Submitted to Date, Open RFI
Cost	Original Budget, Re-Baseline Budget, Approved Changes, Revised Contract Value, Pending Changes, Forecast at Completion, Earned Value, Incurred Value, Outstanding Change %, Work Order CPI
Schedule	Work Order % Complete, Work Order SPI, Work Order HPI
Quality	Non-Conformance Report (NCR), Open NCR, Field Surveillance Report (FSR), FSR Submitted to Date, Open FSR.
Safety	(0-1) Years of Experience Direct Hours, (1-2) Years of Experience Direct Hours, (2-3) Years of Experience Direct Hours, (3-4) Years of Experience Direct Hours, +4 Years of Experience Direct Hours, Foreman Hours, Shift Hours, Exposure Hours, Accident

defined by Witten et al. [20]. First, data are collected and studied to determine if the dataset is suitable for further processing. Second, raw data are cleaned and prepared (i.e., removal of missing data and outliers) for input into machine-learning algorithms. Data cleaning and preparation attempts to address any empty fields in records, data entry errors, and instances where data have been collected in an ad hoc manner [21]. This step ensures that the machine-learning algorithms produce an ideal model with improved performance. Finally, various prediction models are developed, and their performances are evaluated.

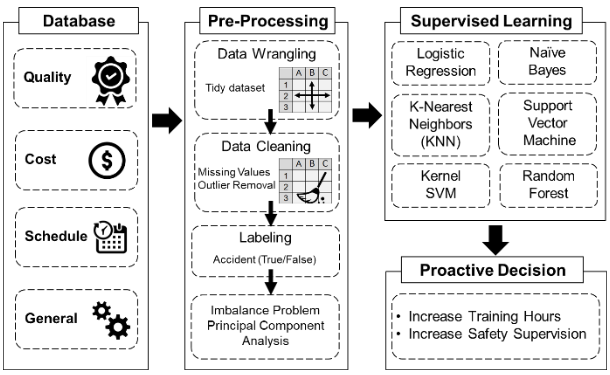

Figure 1. Research methodology

3.1 Data Collection

The dataset used in this research was obtained from a large construction company in Alberta, Canada, with a broad range of construction expertise including building, industrial, and infrastructure projects. The dataset, which encompassed eight industrial projects completed over two consecutive years (2016 to 2017), initially contained 123 biweekly records of each project's performance- (e.g., cost, schedule, and quality performance) and safety-related data (e.g., direct worker hours, foreman hours, shift hours, and accident occurrence). Biweekly accident records were associated with biweekly project performance- and safety-related data. More details about the definition of each of these features and integration of data from different departments is provided in [11]. A description of the variables collected is presented in Table 1.

3.2 Data Cleaning and Preparation

Features and variables that were not useful for this work were first removed. For example, three variables related to the budget and contract value (i.e., original budget, re-baseline budget, revised contract value) all provided similar information. Here, only one variable (re-baseline budget) was considered and kept. Second, missing values were removed or substituted by values that allowed for the development of the model while, at the same time, not adversely affecting model performance [20]. In this particular dataset, the value "Years of Experience Direct Hours" was missing in over 50% of the total records. Due to the limited number of data points, associated columns were removed. The features listed in Table 1 were filtered, resulting in a dataset of 23 features.

3.2.1 Data Labeling

Two labels were selected: (1) True, indicating an accident occurred in the worksite or (2) False, indicating no accident occurred.

3.2.2 Class Imbalance Problem

Machine-learning algorithms perform best when the number of observations is approximately equal in each class [20]. After data cleaning and feature selection, a total of 79 “no accident” and 23 “accident” cases were recorded. Random over-sampling techniques [22] were used to overcome the class imbalance and to reduce the negative impact this may have on algorithm performance.

3.2.3 Principal Component Analysis (PCA)

In machine learning, Principal Component Analysis (PCA) is useful for dimension reduction when analyzing high-dimensional datasets, reducing the number and increasing the independence of predictors [23]. Each principal component is a normalized principal component or linear combination of original variables. Based on the concept that principal components are orthogonal to each other (and correlation coefficients are all zero), PCA allows for the removal of multicollinearity in the features. In PCA, the first principal component is the one that captures the maximum variance of the data set. In other words, it determines the direction of the highest variability in the data.

Scatter plots and correlations between the independent variables are depicted in Figure 2. Of note, the highest correlation observed was between Shift Hours and Foreman Hours, (r=0.93; Figure 2). The results of PCA showed that PC1 explained 59% of the variability in the data set and that the first four principal components explained 90% of the variability in the dataset.

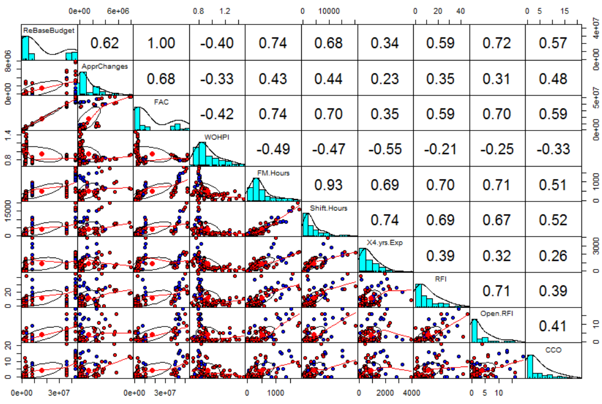

Figure 2. Correlation between independent variables

3.3 Model Generation

After the dataset was cleaned, various machine-learning models were generated and tested. In the current study, classification models were developed using the statistical package *R* [22]. While a large variety of classification models are available in the literature, this study selected models that (1) are widely used for investigating construction problems; (2) have been used in previous studies to predict safety performance and accident occurrence; and/or (3) are useful for modeling complex relationships. The models chosen include k-nearest neighbours (k-NN), logistic regression (LR), random forest (RF), support vector machine (SVM), Kernel support vector machine (KSVM), and Naïve Bayes.

4 Results

Model error is estimated using cross-validation or split validation, and often repeated several times [20]. In this study, the classification models were trained and validated using a 70/30 split validation approach, where the dataset was randomly divided into two parts used for (70%) training and (30%) testing purposes.

Models were evaluated and chosen based on several performance metrics derived from the confusion matrix (Table 2)[24]. These metrics include the following: *Recall*, which describes the probability of correctly predicting “True” relative to the total number of actual “True” and “False False” (FF) in the dataset [Eq. (1)]; *Precision* [Eq. (2)]; and *Accuracy*, which describe the probability correct predictions relative to the total number of predictions [Eq. (3)] [20]. These measures are commonly applied as performance indicators by researchers due to their straightforward interpretation.

Table 2. Description of confusion matrix

	Predicted False	Predicted True
Actual False	True False (TF)	False True (FT)
Actual True	False False (FF)	True True (TT)

$$Recall = TT/(TT + FF) \quad (1)$$

$$Precision = TT/(TT + FT) \quad (2)$$

$$Accuracy = (TT + TF)/(TT + FT + FF + TF) \quad (3)$$

Table 3. Summary of performance metrics of models

Table 3. Model performance

Model	Recall (%)	Accuracy (%)	Precision (%)
LR	71.4	93.5	100
KNN	85.7	87.0	66.6
SVM	57.1	87.0	80.0
Kernel SVM	28.5	80.6	66.6
NB	85.7	87.0	66.6
RF	57.1	87.0	80.0

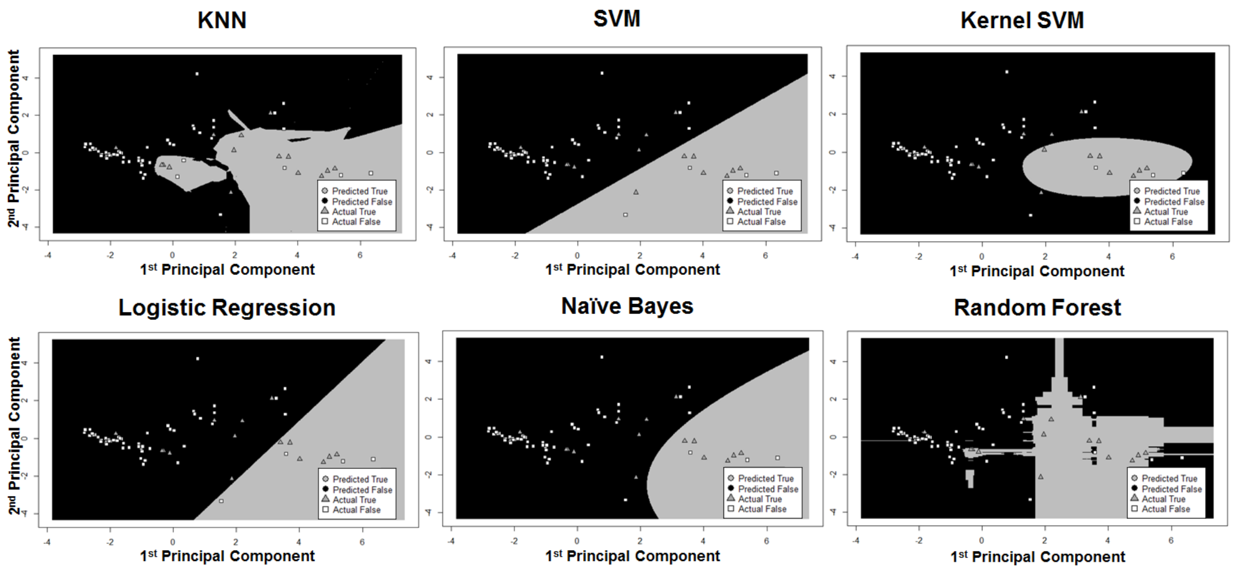

Figure 3. Hyperplanes and decision boundaries of different classifiers

Although *Accuracy* offers a simple measure of the overall performance of a model, it (1) does not consider the type of prediction errors being made and (2) does not consider the distribution (i.e., relative frequencies) of the classes.

Recall is considered the most suitable metric for selecting a model when there is a high cost associated with True/ False. For instance, in accident prediction, if an accident (True True) is predicted as no accident (True False), the consequence can lead to enormous costs for the company. For this reason, *Recall* is used to identify the best model in this study. Both KNN and Naïve Bayes classifiers were associated with the greatest *Recall*, indicating that these models were able to best predict the low-frequency class (in this case, accident occurrence).

Figure 3 visualizes the decision boundaries and the performance of the mentioned classification methods on a randomly selected subset of data. A dimension reduction algorithm (PCA) was applied to reduce the dimensions of the feature vector to two dimensions. As mentioned, the first PCA captures the highest amount of variance in the feature vector, and the second PCA is a component orthogonal to the first one. In Figure 3, as indicated in the legend, the gray and black areas represent proportions predicted by the classifier to be True (i.e., accident) or False (i.e., no accident), respectively. Gray triangles and white squares represent data points that are actually True or False, respectively.

5 Conclusion

While many research studies have been conducted investigating the impact of safety leading indicators on safety performance of construction sites, their use remains limited by certain practical challenges. Due to the approach in which the models were designed in particular, the implementation of prediction methods is often limited to a specific set of leading indicators.

This research study used a machine-learning approach to develop a model that can forecast safety performance from leading indicators regardless of the number or type of leading indicators available. First, the dataset is cleaned and pre-processed. Then, feature selection techniques are applied to reduce the dimensionality of the dataset and to identify the important features affecting safety performance. Finally, machine-learning algorithms are applied for training and validating purposes.

The method was tested using a dataset provided by a large construction company in North America. Out of 23 features, only 10 were found to have a notable influence on safety performance. After evaluation, Naïve Bayes (NB) and K-Nearest Neighbours (KNN) were found to be the best performing algorithms, achieving a *Recall* of 0.857.

Unlike previous studies, which have used theoretical or statistical approaches, this study has used a machine-learning approache to develop a safety performance prediction metric. The results of this study can be used by safety practitioners to (1) improve safety practices or (2) guide the allocation of resources towards monitoring indicators with the most influence on safety performance, saving organizations time and money. For future research, the authors suggest further segmenting the variables of safety (safety hours) with respect to weather conditions, overtime, or other impacting factors.

6 Acknowledgments

This work was supported. through a Collaborative Research and Development (CRD) Grant (CRDPJ 492657) from the National Science and Engineering Research Council (NSERC) of Canada.

References

[1] Reiman T. and Pietikäinen E. Leading indicators of system safety: Monitoring and driving the organizational safety potential. *Safety Science*, 50(10): 1993–2000, 2011.

[2] Grabowski M., Ayyalasomayajula P., Merrick J., Harrald J.R. and Roberts K. Leading indicators of safety in virtual organizations. *Safety Science*, 45 (10): 1013–1043, 2007.

[3] Hinze J., Hallowell M.R., Gibbons B., Green L., Trickel S. and Wulf D. *Measuring safety performance with active safety leading indicators* RS248-1. Construction Industry Institute, Austin, Texas, 2012.

[4] Kjellén U. The safety measurement problem revisited. *Safety Science*, 47(4): 486–489, 2009.

[5] Akroush N.S. and El-adaway I.H. Utilizing construction leading safety indicators: Case study of Tennessee. *Journal of Management in Engineering*, 33(5): 06017002, 2017.

[6] Guo B.H.W. and Yiu T.W. Developing leading indicators to monitor the safety conditions of construction projects. *Journal of Management in Engineering*, 32(1): 04015016, 2016.

[7] Hale A. Why safety performance indicators? *Safety Science*, 47(4): 479–480, 2009.

[8] Øien K., Utne I.B., and Herrera I.A. Building safety indicators: Part 1 – Theoretical foundation. *Safety Science*, 49(2): 148–161, 2011.

[9] Hinze J., Thurman S. and Wehle A. Leading indicators of construction safety performance. *Safety Science*. 51(1): 23–28, 2013.

[10] de la Garza J.M., Hancher D.E. and Decker L. Analysis of safety indicators in construction. *Journal of Construction Engineering and Management*, 124(4): 312–314, 1998.

[11] Mohamed E., Jafari P., Kang SC., Pereira E. and AbouRizk S. Leading indicators for safety management: understanding the impact of project performance data on safety performance. *In Proc., Canadian Society for Civil Engineering, 2019, Montreal, Canada*

[12] Hallowell M.R., Hinze J.W., Baud K.C., and Wehle A. Proactive construction safety control: Measuring, monitoring, and responding to safety leading indicators. *Journal of Construction Engineering and Management*, 139(10): 04013010, 2013.

[13] Schultz G. Using advanced analytics to predict and prevent workplace injuries. Online: https://ohsonline.com/Articles/2012/07/01/Using-Advanced-Analytics-to-Predict-and-Prevent-Workplace-Injuries.aspx, Accessed: 01/30/2019.

[14] Ghodrati N., Yiu T.W., Wilkinson S. and Shahbazpour M. A new approach to predict safety outcomes in the construction industry. *Safety Science*, 109(2018): 86–94, 2018.

[15] Esmaeili B., Hallowell M.R. and Rajagopalan B. Attribute-based safety risk assessment. II: Predicting safety outcomes using generalized linear models. *Journal of Construction Engineering and Management*, 141(8):04015022, 2015.

[16] Xia N., Zou P.X.W., Liu X., Wang X., and Zhu R. A hybrid BN-HFACS model for predicting safety performance in construction projects. *Safety Science*, 101: 332–343, 2018.

[17] TixierA.J.-P., Hallowell M.R., Rajagopalan B. and Bowman D. Application of machine learning to construction injury prediction. *Automation in Construction*, 69(2016): 102–114, 2016.

[18] Poh C.Q.X., Ubeynarayana C.U. and Goh Y.M. Safety leading indicators for construction sites: A machine learning approach. *Automation in Construction*, 93 (2018): 375–386, 2018.

[19] Sarkar S., Vinay S., Raj R., Maiti J. and Mitra P. Application of optimized machine learning techniques for prediction of occupational accidents. *Computers and Operations Research* (2018): 1-15.

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

[21] Soibelman L. and Kim H. Data preparation process for construction knowledge generation through knowledge discovery in databases. *Journal of Computing in Civil Engineering*, 16(1) 39–48, 2002.

[22] R Core Team. R: A language and environment for statistical computing. Online: https://www.r-project.org/, Accessed: 01/30/2019.

[23] Abdi H. and Williams L.J. Principal component analysis. *Computational Statistics*, 2(2010): 433–459.

[24] Kuhn M. and Johnson K. *Applied predictive modeling*. Springer, New York, NY, 2013.