Accident Analysis and Prevention 118 (2018) 77–85

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Factors influencing unsafe behaviors: A supervised learning approach

Yang Miang Goh[a,*], Chalani U. Ubeynarayana[a], Karen Le Xin Wong[b], Brian H.W. Guo[c,e]

[a] Safety and Resilience Research Unit (SaRRU), Dept. of Building, School of Design and Environment, National Univ. of Singapore, 4 Architecture Dr., Singapore, 117566, Singapore

[b] Formerly Dept. of Building, School of Design and Environment, National Univ. of Singapore, 4 Architecture Dr., Singapore, 117566, Singapore

[c] Department of Civil and Natural Resources Engineering, University of Canterbury, Private Bag 4800, Christchurch, 8140, New Zealand

[e] Formerly Safety and Resilience Research Unit (SaRRU), Dept. of Building, School of Design and Environment, National Univ. of Singapore, Singapore

ARTICLE INFO

Keywords:
Theory of reasoned action
Safety behavior
Construction safety
Working at height
Machine learning

ABSTRACT

Despite its potential, the use of machine learning in safety studies had been limited. Considering machine learning's advantage in predictive accuracy, this study used a supervised learning approach to evaluate the relative importance of different cognitive factors within the Theory of Reasoned Action (TRA) in influencing safety behavior. Data were collected from 80 workers in a tunnel construction project using a TRA-based questionnaire. At the same time, behavior-based safety (BBS) observation data, % unsafe behavior, was collected. Subsequently, with the TRA cognitive factors as the input attributes, six widely-used machine learning algorithms and logistic regression were used to develop models to predict % unsafe behavior. The receiver operating characteristic (ROC) curves show that decision tree provides the best prediction. It was found that intention and social norms have the biggest influence on whether a worker was observed to work safely or not. Thus, managers aiming to improve safety behaviors need to pay specific attention to social norms in the worksite. The study also showed that a TRA survey can be used to extend a BBS to facilitate more effective interventions. Lastly, the study showed that machine learning algorithms provide an alternative approach for analyzing the relationship between the cognitive factors and behavioral data.

1. Introduction

Understanding and managing unsafe behavior had always been an important aspect of construction safety management. In the seminal Domino Theory (Heinrich, 1931), unsafe behavior was identified as a key cause of accidents. According to Heinrich, among the direct causes of accidents, 88% are unsafe behavior, 10% are unsafe conditions, and 2% are unpreventable. However, unlike Heinrich (1931), who was focused on the individual's contribution to the unsafe behavior, current views are better reflected in systemic incident causation models (Chua and Goh, 2004) such as the Swiss cheese model (Reason, 1997) and Loss Causation Model (Bird et al., 2003). In these systemic models, unsafe behaviors are active failures influenced by underlying organizational and cultural issues. It is now common knowledge that frontline workers are not solely responsible for unsafe behaviors and managers are expected to implement measures to promote safe behaviors.

A number of safety behavior models were developed and tested over the past decades (Cui et al., 2013; Fang et al., 2015; Griffin and Neal, 2000; Guo et al., 2016; Seo, 2005). These studies provided important insights into safety behavior shaping mechanisms and behavior change interventions. In addition, existing models of behavior and/or motivation in the area of psychology, particularly theory of planned behavior (TPB) (Ajzen, 1991) and theory of reasoned action (TRA) (Fishbein and Ajzen, 2010)), have often been adopted to explain and predict safety behavior (Bakar et al., 2017; Fang et al., 2016; Goh and Binte Sa'adon, 2015; Johnson and Hall, 2005; Quick et al., 2008). The usual analysis approach to test behavioral models is through traditional statistical modelling techniques, such as linear regression, logistic regression, or structural equation modelling and model validation evaluated using goodness-of-fit tests and residual examination (Breiman, 2001).

In recent years there had been a growing interest in applying machine learning techniques in construction safety research (Ciarapica and Giacchetta, 2009; e.g. Goh and Binte Sa'adon, 2015; Goh and Chua, 2013; Patel and Jha, 2014a,b; Tixier et al., 2016). As a subset of artificial intelligence, machine learning can be defined as an algorithmic approach that learn from data without relying on rule-based programing (Alpaydin, 2010). In fact, machine learning and traditional statistical modelling are concerned with the same question, that is, what can be learned from data? Even though Breiman (2001) suggested that machine learning can be used as a more accurate and informative

* Corresponding author.
E-mail address: bdggym@nus.edu.sg (Y.M. Goh).

https://doi.org/10.1016/j.aap.2018.06.002
Received 26 September 2017; Received in revised form 1 June 2018; Accepted 2 June 2018
Available online 18 June 2018
0001-4575/

alternative to data modelling on smaller data sets, a vast majority of safety behavior studies adopted traditional statistical modelling approaches to test the relationships between variables (Fogarty and Shaw, 2010; Guo et al., 2016; Johnson and Hall, 2005; Poulter et al., 2008; Quick et al., 2008). The lack of adoption can be attributed to the fact that machine learning is relative new.

Considering the context, this paper applies machine learning to analyze data collected based on the theory of reasoned action (TRA) and observed safety behavior. In specific, the objectives of this paper are to (1) evaluate the relative importance of different TRA cognitive factors in influencing observed safety behavior, and (2) evaluate the effectiveness of six different machine learning algorithms in analyzing the cognitive and behavioral data. The first objective is concerned with developing a better understanding of predictive power of different cognitive factors, while the second is linked to the purpose of identifying better-performing algorithms so as to reduce future effort spent on selecting suitable algorithms for analysis of cognitive and behavioral data.

2. Literature review

2.1. TRA and its applications to safety behavior

In the 1970s, Fishbein and Ajzen developed the theory of reasoned action (TRA), with an attempt to "identify a relatively small set of variables that can account for a substantial proportion of the variance in any given behavior" (Fishbein 2008) (p.834). The early version of TRA posited that behavior is a function of behavioral intentions that are determined by attitudes and subjective norms. Subsequently, the theory included perceived behavioral control as an additional factor. In recent years, TRA had been updated as shown in Fig. 1 (Fishbein, 2008). It suggests that social human behavior can be predicted from an individual's intention and that effects of intention are moderated by actual control (e.g., skills, abilities, and environmental factors). The intention is determined by attitude towards the behavior, perceived norm, and perceived behavioral control.

There are three beliefs underlying the three determinants of intention, including behavioral beliefs (BB), normative beliefs (NB), and control beliefs (CB). According to the TRA, belief is defined as the subjective probability that an object has a certain attribute (Fishbein and Ajzen, 2010). A behavioral belief is the subjective probability that the behavior will produce a given outcome. Attitude, as a result of BB, is "*a latent disposition or tendency to respond with some degree of favorableness or unfavorableness to a psychological object*" (Fishbein and Ajzen, 2010) (p. 76). In the model, perceived norm (PN) is defined as perceived social pressure to conduct a given behavior. PN consists of injective (known as subjective in the TPB) and descriptive norms which capture the desires and the actions of important referent persons, respectively. PN is determined by normative beliefs (NB) which are beliefs that a particular person or group thinks *I should or should not perform a given behavior*. Perceived behavioral control (PBC), as a result of control beliefs, is another significant predictor of intention. PBC is defined as "*the extent to which people believe that they are capable of performing a given behavior, that they have control over its performance*" (Fishbein and Ajzen, 2010) (p. 154). Given positive attitude and PN to perform a given behavior, the greater the PBC, the stronger should be the intention to perform the behavior.

The TRA has been consistently shown to accurately predict behavioral intention and behavior in a wide range of domains, such as health-related behavior (Blank and Hennessy, 2012; Chassin et al., 1981; Finlay et al., 1999; Jemmott, 2012), environment protection behavior (Jones, 1989), and voting (Singh et al., 1995). The TRA has also been used as a useful framework to design behavior change interventions (Abraham and Michie, 2008; Ajzen and Albarracín, 2007; Gielen and Sleet, 2003; Jemmott, 2012). In addition, both TRA and TPB have been applied to safety behavior research. For example, Johnson and Hall (2005) applied TPB to safe-lifting among 136 materials management employees at a heavy manufacturing company. Using structural equation modelling and factor analysis, the study found that perceived behavioral control and intention were the strongest predictors of safe-lifting behavior. Similarly, Poulter et al. (2008) applied and tested the usefulness of TPB in predicting truck driving behavior by conducting path analysis. More recently, Fogarty and Shaw (2010) demonstrated the usefulness of TPB to understand violation behaviors in aircraft maintenance. One common characteristic of these studies is that traditional statistical modelling approaches were used to test the usefulness of TPB or TRA in predicting different behaviors.

In the construction industry, however, the application of TRA and TPB to safety behavior has been limited. In order to explore ways to reduce unsafe behaviors during work-at-heights, Goh and Binte Sa'adon (2015) investigated the cognitive factors influencing scaffolders' decision to anchor safety harnesses. The authors adopted TPB (Ajzen, 1991) in their study and they realized that among the constructs highlighted in TPB, namely, attitude, subjective norms, perceived behavioral control, and intention, subjective norm had the greatest influence on a worker's decision to anchor his or her harness. The study also highlighted the problems in using linear regression to analyze the relationship between cognitive factors and safety behaviors. It was discovered that machine learning techniques, more specifically artificial neural network and decision tree, were shown to produce more accurate predictions. Nevertheless, the study by Goh and Binte Sa'adon (2015) was exploratory in nature and several recommendations for further studies were provided. In addition, Fang et al. (2016) developed a cognitive model of construction workers' unsafe behaviors in part based on TPB. The model can help better understand the causal mechanisms of unsafe behaviors on site and therefore develop targeted

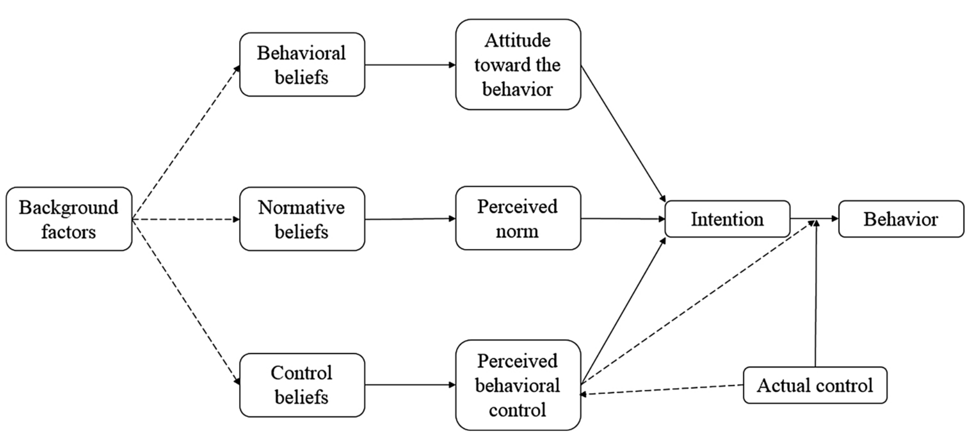

Fig. 1. Schematic presentation of the reasoned action model.

Table 1
Supervised machine learning algorithms.

Type of algorithms	Description
Support vector machine	Support vector machine algorithm is established on creating hyperplanes based on training data and then optimizes the hyperplanes, which are the basis for classifying data points among classes. Support vectors are the data points nearest to the hyperplane and a hyperplane is a line derived from a function which classifies a set of data into classes.
Random Forest	A random forest classifier is an ensemble of decision trees. Each tree in the forest predicts their final class label. The collection of trees then voted for the most popular class as the final class label.
K-nearest neighbor	K nearest neighbor classifier is a simple algorithm that stores all available class labels and classifies new class labels based on a distance vector when there is no prior knowledge about the underlying distribution of the data.
Naïve Bayes	Naïve Bayes algorithm uses Bayes rule with strong independent assumptions between features. It simplifies the calculation of probabilities by assuming that the probabilities of each feature belonging to a given class label are independent of all other features.
Artificial neural network	Neural networks are information processing algorithms which operates similar to the neurons in human brain. The neurons in the neural network are connected by links and, these neurons contain functions and are connected to other neurons. The network learns and adjusts the weight assigned to the links to improve its performance.
Decision Tree	A decision tree is a hierarchical model in which input space is divided into local regions by a sequence of recursive splits in a smaller number of steps. A decision tree consists of internal decision nodes and terminal leaves.

interventions.

2.2. Machine learning in construction safety research

The use of machine learning techniques in construction safety management research has received growing interest in recent years (e.g. Goh and Binte Sa'adon, 2015; Goh and Chua, 2013; Liao and Perng, 2008; Patel and Jha, 2014a, b; Tixier et al., 2016). Mitchell (1997) also provided a formal definition: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E".

Past studies had demonstrated the potential of machine learning techniques in predicting and evaluating safety-related data. For example, Arciszewski et al. (1995) applied a multi-step machine learning process to transform input (accident reports) into output (decision rules). Likewise, Tixier et al. (2016) applied Random Forest (RF) and Stochastic Gradient Tree Boosting (SGTB) to injury reports to predict a range of target variables such as injury type, body part and injury severity. The outcome variables that Tixier et al. (2016) used included 78 accident precursors abstracted from the injury reports. Similarly, Goh and Chua (2013) focused on injury severity, but they used safety management system audit scores to train an ANN to predict injury severity. ANN was also used in Patel and Jha (2014a, b), Patel and Jha (2014a, b) and Goh and Binte Sa'adon (2015) to predict safety behavior or safety culture variables. As reported earlier, in addition to ANN, Goh and Binte Sa'adon (2015) used DT and it was shown that both ANN and DT provided better prediction performance than linear regression. Yang et al. (2016) applied a semi-supervised pattern-recognition algorithm (i.e., support vector machine) to detect near-miss falls of ironworkers. Results demonstrated that the approach achieved a 75.8% recall and 87.5 accuracy in near miss detection. In addition, effort was made by Akhavian and Behzadan (2016) to recognize and classify construction workers' activities. In the study, five different machine learning algorithms were trained. Results suggested that neural networks outperformed other classifiers, with an accuracy up to 97%.

3. Data and methods

3.1. Machine learning

Machine learning is a branch of artificial intelligence that gives computers the ability to learn without being explicitly programmed. It deals with the development and applications of algorithms that can learn from data and make predications. Machine learning has its origins from statistics (Alpaydin, 2010). There are two cultures in the use of statistical modelling to learn from data: the data modelling culture and algorithmic modelling culture (Breiman, 2001). As the latter, machine learning shifts focus from data models to the algorithms, with emphasis on predictive accuracy.

This study adopts machine learning for predicting unsafe behavior for the following reasons. First of all, machine learning outperforms traditional statistical modelling such as logistic regression with respect to predictive accuracy. Breiman (2001) provided an example to support the claim, where both logistic regression and random forests were used to identify the most important variables that can predict the survival or non-survival of 155 hepatitis patients. Results suggested that the random forests predictive error rate is 12.3%, almost a 30% reduction from the logistic regression error. Secondly, compared to statistical modelling, machine learning often produces much more (useful) information about underlying mechanism in data (Breiman, 2001). Last but not least, in general machine learning is spared from most of assumptions that must be met in statistical modelling, such as homoscedasticity, liner relation between independent and dependent variables, and normal distribution.

There is a wide range of machine learning algorithms developed over the years and they can be classified into supervised learning, unsupervised, and semi-supervised learning or reinforcement learning (Russell and Norvig, 2010). Supervised learning was applied in this study considering the fact that there has been a clear set of target variables (i.e., unsafe behavior) and corresponding attributes (i.e., cognitive factors in the TRA model). Established algorithms commonly used in supervised learning include support vector machine (SVM), Random Forest (RF), K-nearest neighbor (KNN), naïve Bayes (NB), artificial neural network (ANN) and Decision Tree (DT) (Kotsiantis, 2007). These six algorithms were used in this study. Since these are well-established machine learning algorithms, readers are referred to relevant textbooks for descriptions of the algorithms (e.g. Russell and Norvig, 2010; Witten, 2011). A brief description of each of the learning algorithms is presented in Table 1

3.2. Data collection

Data were collected from a Mass Rapid Transit (MRT) tunneling project in Singapore. The contract value of the project is over S$ 255 million and a tunnel boring machine (TBM) was used in the construction. The total length of the tunnel is 1.65 km. The main contractor (the company thereafter) takes great pride in its high safety standards and it aims for zero incidents, zero pollution, and zero ill-health. A behavior-based safety (BBS) program was implemented in this project, which was strongly supported by the company's top management. The BBS program consists of four key steps: (1) BBS checklist development, (2) baseline observations, (3) goal setting and intervention, and (4) review and continuous improvement. These four steps form a BBS cycle. A BBS taskforce was established to oversee the BBS implementation, consisting of stakeholders such as site engineers, project managers,

Table 2
List of behavior checklist items for working at height.

No.	Positive Behaviors
1	Worker does not use damaged ladder.
2	Worker keeps three points of contact with the ladder at all times.
3	Worker does not carry any items while climbing up or down the ladder.
4	Worker does not work on scaffolds, which are tagged unsafe, and/or not in safe condition.
5	Workers employ safety harness and anchored at firm anchoring point.
6	Safety harness is not loosely worn.
7	Scaffold component shall not be removed or altered without approval from scaffold supervisor.
8	Workers shall not discard any article from height when the location is not adequately cordon off.
9	Workman working near the opening shall ensure that no activity is ongoing below unless safety measures are in place.

construction manager and supervisors from various subcontractors. A set of unsafe behavior checklists was developed by the BBS taskforce. The checklists cover nine categories of safety behavior, including lifting operation, excavation, working at height, work platform & access, manual handling, hot work (welding and gas cutting), plant and equipment, traffic management, and personal protective equipment (PPE). A random sample of 80 workers took part in the BBS program, a sizeable number that can serve as a good representation of the given population of the construction project.

Based on the observation data that the site had been collecting, working at height has the highest % unsafe behavior. Thus, it was decided that this study will focus on work at height observation data. The work at height behavior checklist (see Table 2) was adopted from the BBS program. Five safety supervisors who were trained as observers were recruited to assist in this study. Observations of unsafe behavior were conducted before the questionnaire survey. This is critical as administrating the survey first may convey a message that workers' unsafe behavior will be observed and therefore negatively influence the accuracy of data collected. The observation procedure was in accordance to the BBS program, except that observers recorded the working ID of the worker under observation. This is necessary to match the observation to the same worker's questionnaire response. Care was taken to assure the workers that the deviation from their usual practice of not identifying the workers was purely for research purpose only. Each observer monitored only one worker at each time and each observation took approximately thirty minutes.

Unsafe behavior was measured using the percentage rating scale, as suggested by Duff et al. (1994). Unsafe actions were computed and analyzed based on Eq. (1):

$$\%\ unsafe\ \ behavior = \frac{Total\ \ number\ \ of\ \ observations\ \ with\ \ unsafe\ \ behavior}{Total\ \ number\ \ of\ \ observations} \qquad (1)$$

All 80 workers who participated in the BBS program were invited to complete the questionnaire survey. The questionnaire was translated into different languages so as to make sure that migrant workers from different countries understand the questions. All 80 workers completed and returned the questionnaire. Survey participants include six main trade works: general works (41.3%), plumbing & sanitary trade (28.8%), metal works (15%), ceiling works (6.3%), rigger & signalman (6.25%), and lifting works (2.5%). The age of respondents ranged from 22 to 50 years old. The respondents had 1 to 17 years (mean = 5.8 years) of experience. 72.5% of the workers had more than three years of experience. 60% of the respondents are Bangladeshi workers, followed by Indians (31.25%), Chinese (3.75%) and Thai (5%) workers.

3.3. TRA survey instrument

The TRA questionnaire was designed to measure key constructs of TRA (see Fig. 1), except for actual control. Fishbein and Ajzen (2010, p.21) define actual controls as the "relevant skills and abilities as well as barriers to and facilitators of behavioral performance". As the study uses a survey instrument to collect data, perceived behavioral control and actual control will become very similar, as both will be based on the perception of the respondents. In addition, in Singapore, where the data was collected, there is a regulated safety training courses that all workers in the same trade must attend. The site which was studied had very strict processes to ensure that the workers have attended all the mandatory training courses before they can enter the site. Thus, the survey instrument did not include items to measure the "actual control" construct.

Items were adapted from Goh and Binte Sa'adon (2015) and Fishbein and Ajzen (2010). The survey questionnaire in this study used a 7-point Likert Scale, with "1" being "Strongly Agree" and "7" being "Strongly Disagree". Details of the items are provided below.

- Behavior belief was measured using, "I believe that I will not get into an accident".
- Normative belief was measured by eight items: (1) Most of my fellow workers think I should work safely; (2) My supervisor(s) think I should work safely; (3) The Workplace Safety & Health Officer (WSHO) thinks I should work safely; (4) My manager(s) thinks I should work safely; (5) When it comes to safety, I would want to do what my fellow workers think I should do; (6) When it comes to safety, I would want to do what my Supervisor(s) think I should do; (7) When it comes to safety, I would want to do what my WSHO thinks I should do; (8) When it comes to safety, I would want to do what my manager(s) think I should do. The Cronbach's Alpha is 0.731, which indicates an acceptable level of internal consistency.
- Attitude is measured by "Following safety rules all the time is: (1) represents "good", (2) and (3) are between good and neutral, (4) is "neutral", (5) and (6) are between neutral and bad, and (7) represents "bad".
- Perceived norm was measured by four items: (1) my fellow workers work safely all the time; (2) my supervisor(s) work safely all the time; (3) my WSHO works safely all the time; (4) my manager(s) work safely all the time. The Cronbach's Alpha is 0.655, which indicates an acceptable level of internal consistency.
- Perceived belief control was measured using six items: (1) I have the necessary training(s) to work safely; (2) I have the necessary equipment to work safely; (3) I am given enough time to work safely; (4) I will work safely if I have the necessary training(s); (5) I will work safely if I have the necessary equipment; and (6) I will work safely if enough time was given. The Cronbach's Alpha is 0.612, which indicates an acceptable level of internal consistency.
- The intention was measured using a single item: "I want to work safely all the time"

Some constructs were measured using single-item measures. Compared to multiple-item measure, single-item measures have advantages in the context of this study. For example, single-item measures are easier to understand by less-educated migrant workers, who are common in Singapore. In addition, they can be completed more quickly, which is important due to the dynamic and competitive construction environment. In terms of reliability and validity, single-item measures can perform comparably with the multiple-item measure in all statistical analyses (Dolbier et al., 2005; Nagy, 2002; Wanous et al., 1997). For constructs measured using multiple items, the mean of the ratings for the items were used to represent the construct.

3.4. Data analysis

According to the theory of reasoned action, attitude, perceived norm (PN) and perceived behavioral control (PBC) can affect the intention of the workers. Further, it can be seen from the model in Fig. 1, PBC and Actual Control moderate the relationship between intention and the behaviour of workers. Therefore, in order to predict how these variables affect the behavior of the workers, abovementioned data mining techniques were applied to the dataset. The attributes are the behavioral belief, normative belief, attitude, perceived norm, perceived behavioral control, and intention. The target variable, % unsafe behavior, ranges from 0% to 36.4%; it was categorized into two classes namely Class 1, and Class 2. Class 1 represents safer workers with relatively low % unsafe behavior, and this was determined to be less than 10%. Class 2 represents relatively unsafe workers with % unsafe behavior greater than 10%. A 10% interval was chosen based on the consideration that this study used human observers whose observations may not be 100% correct. Categorizing workers who demonstrated 100% safe behavior into a "safe" class is not realistic, while any percentage of unsafe behavior less than 80% is too low to be considered as a "safe" class.

The data set consists of 80 observations and, after categorizing the unsafe behavior into Class 1 and Class 2, there were 63 observations for Class 1 and, 17 observations for Class 2. Clearly, there is an imbalance between the two classes. The ratio of majority class samples to minority class samples is 63:17, and this imbalance of the two classes will lead the classifiers to classify Class 1 more accurately while misclassifying many samples of Class 2. To address the issue of imbalance class, Synthetic Minority Over-sampling Technique (SMOTE) (Chawla et al., 2002) balancing technique was applied. SMOTE is an over sampling method in which the over samples are created by creating synthetic examples instead of over sampling with replacement of the minority class. Then, a new set of balanced data is generated by producing synthetic samples of the minority class. The minority class samples are over-sampled by generating synthetic samples along the line segments (a length calculated using Euclidean distance between two points data space) joining the K nearest neighbors of the minority class. In our algorithm, we used $K = 5$ (i.e. five nearest neighbors) to create synthetic samples. As the next step, the difference between the feature vector (an n-dimensional vector used to represent the variables of the problem being studied). Then that measured difference is multiplied by a random number between 0 and 1, and added to the feature vector. This approach forces the decision region of the minority class to generate the synthetic samples in the direction of the nearest neighbors.

After balancing the data set, six machine learning algorithms namely, decision tree (DT), random forest (RF), support vector machine (SVM), k-nearest neighbor (KNN), naïve Bayes (NB) and, artificial neural network (ANN) were implemented. In addition, logistic regression (LR) was also implemented for comparison.

In accordance to Hair et al. (2006) (p227), variance inflation factor (VIF) was calculated to test if there is multicollinearity between independent variables. The results are presented in Table 3. As can be observed, the VIF values are very much lower than the common cutoff threshold of 10 (Hair et al., 2006) (p227). Thus, we can assume that there is no significant multicollinearity between the independent variables.

Model parameters were selected based on a series of trials (see Table 4 for examples of the trials) in order to improve the predictive ability of the algorithms.

The models were evaluated using the receiver operating characteristic curves (ROC curves) (Bradley, 1997). ROC analysis uses the true positive rate (TPR) (see Eqs. (2) and (3)) and false positive rate (FPR) to evaluate the performance of a classifier model. True positives (TP) are samples correctly labelled as positives. False positives (FP) refer to negative samples incorrectly labelled as positive. True negatives (TN) correspond to negatives correctly labelled as negative. Finally, false negatives (FN) refer to positive examples incorrectly labelled as negative. ROC curves are useful for evaluating the performance of classifiers when the classification task is binary (Davis and Goadrich, 2006) and also, ROC curve is independent of the proportion of Class 1 and Class 2. X axis of the ROC curve plot represents the false positive rate, and the Y axis represents the true positive rate (see the Eqs. (2) and (3)). ROC curves can be thought of as representing the family of best decision boundaries for relative costs of TP and FP. They are useful in evaluating and visualizing the performance of the classifiers. The coordinate point (0,100) in the ROC curve is the best point where all the positive samples are classified correctly and no negative samples are misclassified.

Table 3
VIF of independent variables.

Independent Variable	VIF
Behavioural Belief	1.055906
Normative Belief	2.437455
Attitude	1.130690
Subjective Norm	1.559869
PBC	1.596069
Intention	1.411049

Table 4
Selection of model parameters.

Trials	Decision Tree Classifier		Random Forest Classifier		K Nearest Neighbor Classifier	
	Min samples leaf [a]	*AUC value*	*N estimator* [b]	*AUC value*	*N neighbors* [c]	*AUC value*
1	1	0.976	4	0.912	4	0.928
2	2	0.968	5	0.920	5	0.928
3	3	0.952	6	0.937	6	0.920
4	4	0.928	8	0.920	8	0.912
5	5	0.936	10	0.920	10	0.896
6	6	0.920	20	0.920	15	0.865

Note: Figures in Bold are chosen as the optimized parameters.
[a] Min samples leaf assigns the minimum number of samples required to be at a leaf node.
[b] N estimator refers to the number of trees in the forest.
[c] N neighbors refers to the default number of neighbors to use for k-neighbors queries.

$$True\ \ positive\ \ rate = \frac{TP}{(TP + FN)} \tag{2}$$

$$False\ \ positive\ \ rate = \frac{FP}{(FP + TN)} \tag{3}$$

The Area under the Curve (AUC) of the ROC curves were used as a performance metric for the learning algorithms. The AUC score lies in between 0 and 1. The threshold value is considered as 0.5 and it is the line y = x in ROC curve which represents random guessing of the classes (Fawcett, 2006). The higher the AUC indicates a better classification. AUC is a useful metric for classifier performance as it is independent of the decision criterion selected and prior probabilities. AUC value can be interpreted as the probability that a classifier is able to distinguish a randomly chosen positive samples from a randomly chosen negative samples.

The models are further evaluated using a leave-one-out cross-validation technique. Cross-validation methods (Airola et al., 2011) estimate how accurately a predictive model performs and avoid overfitting of the data and also, cross validation techniques ensure that every sample in the data set has the same chance of being in the training set and testing set. In this experiment leave-one-out cross-validation technique (Kohavi, 1995) is used to validate the models. Leave-one-out cross-validation technique creates a training set by taking all the samples except one, and the testing set is the sample left out. This process is repeated for the entire N samples in the data set. The leave-one-out cross-validation technique efficiently uses all the data to provide accurate predictions. The algorithms used in this study were primarily derived from Python 2.7 (Python Software Foundation 2016) and scikit-learn library version 0.18.

In the second experiment, the data set is fitted into a CART (Classification and Regression Trees) algorithm in Python and, a decision tree is plotted. A decision tree is a model that predicts the class of a target variable in a classification problem, based on simple decision rules acquired from the features of the data set. The final decision tree can explain why a specific prediction is made. The algorithm looks at the features in a data set, determines the most important feature that corresponds to a classification, and using it as the root node. A tree is created by progressively splitting the training set. The decision tree's splits are based on purity (or impurity) of a node, which is an estimation of how well the two classes separated (Timofeev, 2004). Each decision nodes have one incoming branch and two outgoing branches. Leaf nodes are the final nodes, and no further splitting is done.

The impurity measure used in CART is the Gini Index (Du and Zhan, 2002). Gini index is an impurity-based criterion that measures the divergences between the probabilities of the target variable values. Intuitively, the Gini Index or Gini Impurity measures the probability of misclassification. The algorithm grows the decision tree, until each leaf node converges to the lowest impurity with lowest Gini index. The following equations from Eqs. (4)–(7) are used to calculate the Gini index.

$$D = Total\ \ number\ \ of\ \ data\ \ entries \tag{4}$$

$$D_i = Number\ \ of\ \ data\ \ entries\ \ for\ \ class\ \ i \tag{5}$$

$$p_i = \frac{D_i}{D}\ ratio\ \ of\ \ instances\ \ of\ \ class\ \ i \tag{6}$$

$$k = number\ \ of\ \ classes \tag{7}$$

$$Gini\ \ index = 1 - \sum_{i=1}^{k} P_i^2 \tag{8}$$

4. Results and findings

As indicated earlier, the classification was conducted using six well-known classification algorithms and logistic regression. The AUC of the different classifiers are presented in Table 5.

The best AUC value of 0.976 was achieved using the DT classifier. The second best classifier was ANN, and NB classifier produced the lowest AUC value of 0.865. It is noted that LR (also known as a statistical method), performed poorer than five of the other machine learning techniques. The following discusses DT, the best performing classifier.

Fig. 2 illustrates the decision tree. The highest Gini index value of 0.5, belongs to the root node (Intention) and it contains all the training samples. It can also be seen that the tree was fully grown until the leaf nodes, which have minimum impurity with Gini index 0. The root node has the value less than 1.0303, which is based on the Likert scale used in the questionnaire and less than 1.0303 implies a strong intention to perform safe behaviors. The root node contains all the samples (63 observations) of Class 1 (and 62 samples of Class 2. One sample from Class 2 was used as the test set to evaluate the model.

Table 5
AUC values for the seven classifiers.

Classifier	AUC
Decision Tree (DT)	0.976
Artificial Neural Network (ANN)	0.944
Random Forest DT (RF)	0.937
K-Nearest Neighbor (KNN)	0.929
Support Vector Machine (SVM)	0.928
Logistic (Regression)	0.920
Naïve Bayes (NB)	0.865

Generally, variables that are closer to the root node are better at classifying workers with high or low % unsafe behavior. As such, intention is the most important factor to predict % unsafe behavior, which is followed by normative belief, perceived norm, and perceived behavior control. By focusing on features that are more important, i.e. intention and normative belief, the following rules can be derived:

- when a worker does not have strong intention to work safely, he is more likely to display unsafe behavior ($p = \frac{56}{57} = 0.98$);
- when a worker has strong intention to work safely and the normative belief is strong, it is likely that the worker will display safe behavior ($p = \frac{56}{57} = 0.98$); and
- when a worker has strong intention to work safely, but has weak normative belief, the worker is more likely to display unsafe behavior ($p = \frac{11}{17} = 0.65$).

5. Discussion

5.1. Predictors of unsafe behavior

This paper aims to evaluate the relative importance of different cognitive factors within the theory of reasoned action in influencing unsafe behavior. It is suggested that intention is the most important variable in affecting workers' safety behaviors. This is consistent with previous studies (Poulter et al., 2008; Quick et al., 2008; Zhou et al., 2016), in which intention is statistically significantly related to target behaviors. The link between intention and behavior is also supported by systematic reviews and meta-analyses of empirical findings (Armitage and Conner, 2001). Note that inconsistency between intention and safety behavior was also evidenced by the decision tree. There is a 65% chance that workers intending to behave safely, but was observed to work unsafely because of weak normative belief, i.e. they believe that their colleagues do not think they should work safely. The result shows that workers' psychosocial environments can exert strong influence on their actions. The finding is supported by past studies in construction industry (Choudhry and Fang, 2008; Guo et al., 2015, 2016; Mullen, 2004). Individuals comply with perceived social pressure mainly because of reward power, coercive power, legitimate power, expert power, and referent power (Fishbein and Ajzen, 2010). The concept of normative belief is also related to the construct of safety culture, which is essentially the shared values and beliefs in relation to safety (Cooper, 2000). Past decades have seen a large number of studies that support the proposition that safety climate (as a proxy to safety culture) is a leading indicator of safety behavior (Zohar, 2010).

More often than not, unsafe behavior is explained as an outcome of "unsafe or lazy attitude" or "lack of safety knowledge or skills". This is where workers are sent for counselling or re-training when they are found to contravene safety rules. However, this study suggests attitude and perceived behavioral control are not strong predictors of safety behavior.

5.2. Managerial implications

Behavior change intervention strategies could be designed by placing emphasis on the two key predictors (i.e., intention and normative belief) identified in this study. First and foremost, interventions can be devised to change workers' intention of performing unsafe behavior. Managers can change workers' negative intention by persuasive communication. Persuasive and supportive message can be communicated to the target workers to urge them to work safely. For the message to be comprehended and accepted, it is critical for managers and supervisors to demonstrate commitment to safety in a consistent manner, so that workers believe that safe behaviors are desirable. Another possible way to change intention is to promote safety through the use of emotional

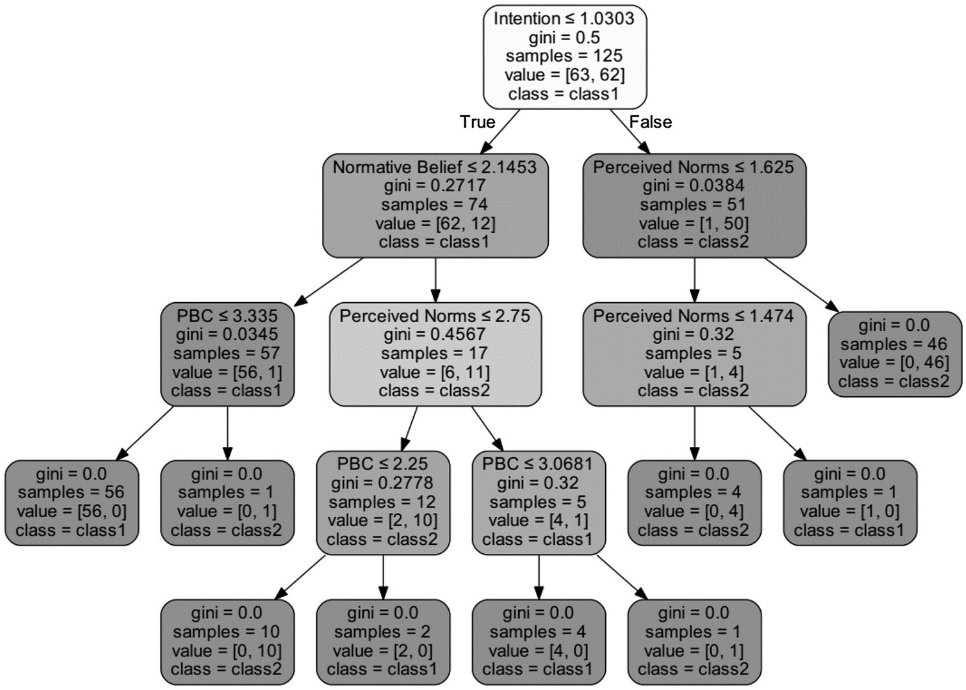

Fig. 2. Decision tree diagram.

appeal pertaining to the theme of family, which can potentially reduce the effect of weak normative belief, a moderator of the effects of safe intentions.

Even though attitude and perceived behavioral controls are not strong predictors of safety behavior, counselling and training are still important interventions. However, these interventions should attempt to reinforce beliefs that the organization place great significance on safety and safety behavior will be rewarded socially. Special attention should be directed towards managing negative effects of production pressure and peer pressure on site. Managers must emphasize that even under production pressure, taking a risky behavior and shortcut are still unacceptable. Such messages must be reinforced by the words and actions of managers and supervisors on the ground on a daily basis.

In addition, a BBS program can be extended with a TRA survey to derive useful inputs for selection and design of interventions to improve safety behavior. BBS program provides a systematic approach to improve safety-related behavior, but the focus had traditionally been on the observation and feedback process. The TRA survey provides a richer theoretical framework to facilitate identification of possible interventions. Coupling BBS and TRA can overcome some of the limitations of the BBS approach, which include the tendency to over-focus on the individual workers' behavior and neglecting the importance of social norms influenced by managers, supervisors and peer workers.

5.3. Methodological implications

This paper adopted a machine learning approach to study safety behavior in the construction industry. A literature review presented earlier revealed that traditional statistical modelling dominates the safety behavior research in almost all industries. A vast majority of past studies in safety behavior used traditional data modelling, such as liner regression and structural equation modelling, to enhance the understanding of safety behavior by estimating the relationships between input variables **x** (independent variables) and the response variables **y** (dependent variables). A data modelling method assumes that data are generated by a given stochastic data model and the model is validated using goodness-of-fit tests and residual examination (Breiman, 2001). It is true that previous safety behavior studies using statistical modelling contributed a lot to the field of research in terms of safety behavior shaping mechanism. However, this by no means suggests that data modelling is the only method to safety behavior, not to mention that focusing on data models may lead to irrelevant theory and questionable scientific conclusions (Breiman, 2001). As another culture of statistical modelling, algorithmic modelling is a better approach to complex statistical problems (Breiman, 2001). Considering that nowadays a wide range of digital technologies have been developed and implemented in construction projects, we can expect that a large amount of data on safety behavior over terabytes can be more easily collected in the near future. The traditional data modelling approach may not be applicable in many situations because there may not be a suitable "stochastic data model" and machine learning is a promising alternative.

The focus of this paper has been on using collected data to predict safety behavior rather than testing the relationships proposed in the TRA. One strength of this paper is that it compared numerous machine learning algorithms based on the AUC values. The performances of all the seven classifiers were good, but decision tree performed exceptionally well (AUC = 0.976) and artificial neural network was second (0.944). In contrast, logistic regression was ranked second last.

Thus, this paper suggests that future studies on safety-related behaviors should consider the use of neural network and decision tree. Decision tree has an added advantage because it provides a transparent model, where non-experts can easily use to assess the likelihood of unsafe behavior and understand the relative importance of different cognitive factors. This paper made no attempt to refute the usefulness of the logistic regression, since it also produced a high value of AUC (0.92). Instead, the key point is that a machine learning approach is a good alternative to traditional statistical approaches. In fact, machine Learning algorithms have several advantages over traditional statistical methods like regression. Machine learning algorithms consider potential interactions comprehensively and they do not need a predefined hypothesis, which makes it less likely to overlook the predictor variables. Predictive models using machine learning algorithms would therefore better facilitate recognition of unsafe behaviors among construction workers. Moreover machine learning algorithms can easily incorporate new data to update and optimize the algorithms with minimal error than traditional statistical methods.

6. Conclusions

Management of safety behavior is an important aspect of construction safety management because most accidents are related to behavioral issues. To reduce the likelihood of accidents, managers need to design interventions to make safe behavior a norm in the worksites. Based on the Theory of Reasoned Action (TRA), previously known as Theory of Planned Behavior (TPB), behaviors are influenced by intention, which in turn is influenced by attitude, perceived norm and perceived behavioral control (through actual control). These cognitive factors are further influenced by different types of beliefs. The TRA provides a framework to evaluate the cognitive factors influencing unsafe behaviors and facilitate design of interventions.

This study evaluated the relative importance of different cognitive factors within the TRA in influencing unsafe behavior. Compared to the most previous safety behavior studies adopting traditional statistical modelling (e.g., liner regression and structural equation modelling), this study adopted a supervised machine learning approach due to its potential in predictive accuracy and freedom from most assumptions constrained on traditional statistical modelling.

In addition, instead of using self-reported safety behavior as the target variable, observation data obtained from the established behavior-based safety (BBS) program of the site was used. Six supervised machine learning techniques were evaluated and they include support vector machine (SVM), Random Forest (RF), K-nearest neighbor (KNN), naïve Bayes (NB), artificial neural network (ANN) and Decision Tree (DT). In addition, logistic regression (LR) was also implemented. The data set was balanced using Synthetic Minority Over-sampling Technique (SMOTE). Performance of the seven algorithms were evaluated using the area under the receiver operating characteristic (ROC) curves through a leave-one-out cross-validation process. DT was identified as the best performing algorithm. Based on the DT developed in this study, it was established that even though a worker may intend to work safely, beliefs about social norms within the worksite is an important cognitive factor (as compared to the other factors in the TRA) that may cause the worker to work unsafely. Thus, managers aiming to improve safety behaviors need to develop interventions to influence intentions and social norms. The study also showed that the TRA survey can be used to extend a BBS program. Furthermore, the findings suggest that decision tree and artificial neural network can be used to analyze the relationship between the different cognitive factors and behavioral data.

7. Limitations and future study

This study has a number of limitations. First, as the sample was from one single site, it is important to further validate the findings using data from other sites. The second limitation is the modest sample size. However, using the Synthetic Minority Over-sampling Technique (SMOTE), sample size increased from 80 to 125, which is comparable with previous machine learning studies, such as Breiman (2001) which applied random forests on a sample size of 81 with 4682 variables. Third, % unsafe behavior was measured based on observations, which is influenced by the accuracy of the observers. Nevertheless, it is believed that this is an improvement as compared to self-reported unsafe behavior items used in Goh and Binte Sa'adon (2015). The fourth limitation is that some constructs were measured using single-item. The choice of using single-item measures was made based on the consideration of situational constraints such as less-educated workers and considerable production pressure. A concern was that a questionnaire that contains redundant items could contribute to survey fatigue. Despite the fact that single-item measures can be used as a psychometrically sound instrument as discussed earlier, the reliability and validity need further tests as conducted by Dolbier et al. (2005). Last, it must be noted that the machine learning approach is a data-driven approach as opposed to a theoretical approach. In this sense, the original relationships between the TRA constructs were not strictly adhered to. Instead, a wide range of potential interactions were tested by the machine learning algorithms and the emphasis is on predictive performance of the algorithms. This "brute force" method is a limitation from the theoretical perspective.

Future efforts can be made to administer the questionnaire to the whole construction industry so as to obtain a larger sample size. Concurrently, based on the proposition that social norms is a key influencing factor, future research should identify and evaluate suitable interventions to change safety-related social norms on construction sites.

Acknowledgement

This study is funded by Singapore Ministry of Education (MOE) Tier 1 grant (Grant No.: R-296-000-161-112).

References

Abraham, C., Michie, S., 2008. A taxonomy of behavior change techniques used in interventions. Health Psychol. 27 (3), 379.

Airola, A., Pahikkala, T., Waegeman, W., De Baets, B., Salakoski, T., 2011. An experimental comparison of cross-validation techniques for estimating the area under the ROC curve. Comput. Stat. Data Anal. 55 (4), 1828–1844.

Ajzen, I., 1991. The theory of planned behavior. Organ. Behav. Hum. Decis. Process. 50 (2), 179–211.

Ajzen, I., Albarracín, D., 2007. Prediction and change of health behavior: applying the reasoned action approach. In: Ajzen, I., Albarracín, D., Hornik, R.C. (Eds.), Predicting and Changing Behavior: A Reasoned Action Approach. Lawrence Erlbaum Associates, Inc., Publishers, London.

Akhavian, R., Behzadan, A.H., 2016. Smartphone-based construction workers' activity recognition and classification. Autom. Constr. 71, 198–209.

Alpaydin, E., 2010. Introduction to Machine Learning. MIT press, London.

Arciszewski, T., Michalski, R.S., Dybala, T., 1995. STAR methodology-based learning about construction accidents and their prevention. Autom. Constr. 4 (1), 75–85.

Armitage, C.J., Conner, M., 2001. Efficacy of the theory of planned behaviour: a meta-analytic review. Br. J. Soc. Psychol. 40 (4), 471–499.

Bakar, E.A., Isa, N.S., Osman, S., 2017. Application of theory of planned behavior in the motor vehicle repair and service industry. Saf. Sci. 98, 70–76.

Bird, F.E., Germain, G.L., Clark, M.D., 2003. Practical Loss Control Leadership. Det Norske Veritas (U.S.A.), Inc., Duluth, Georgia.

Blank, M.B., Hennessy, M., 2012. A reasoned action approach to HIV prevention for persons with serious mental illness. ANNALS Am. Acad. Polit. Soc. Sci. 640 (1), 173–188.

Bradley, A.P., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognit. 30 (7), 1145–1159.

Breiman, L., 2001. Statistical modeling: the two cultures (with comments and a rejoinder by the author). Statist. Sci. 16 (3), 199–231.

Chassin, L., Corty, E., Presson, C.C., Olshavsky, R.W., Bensenberg, M., Sherman, S.J., 1981. Predicting adolescents' intentions to smoke cigarettes. J. Health Soc. Behav. 445–455.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357.

Choudhry, R.M., Fang, D., 2008. Why operatives engage in unsafe work behavior: investigating factors on construction sites. Saf. Sci. 46 (4), 566–584.

Chua, D.K.H., Goh, Y.M., 2004. Incident causation model for improving feedback of safety knowledge. J. Constr. Eng. Manag. - Am. Soc. Civ. Eng. 130 (4), 542–551.

Ciarapica, F.E., Giacchetta, G., 2009. Classification and prediction of occupational injury risk using soft computing techniques: an Italian study. Saf. Sci. 47 (1), 36–49.

Cooper, M.D., 2000. Towards a model of safety culture. Saf. Sci. 36 (2), 111–136.

Cui, L., Fan, D., Fu, G., Zhu, C.J., 2013. An integrative model of organizational safety behavior. J. Saf. Res. 45, 37–46.

Davis, J., Goadrich, M., 2006. The relationship between precision-recall and ROC curves. In: Proc., Proceedings of the 23rd International Conference on Machine Learning. ACM. pp. 233–240.

Dolbier, C.L., Webster, J.A., McCalister, K.T., Mallon, M.W., Steinhardt, M.A., 2005. Reliability and validity of a single-item measure of job satisfaction. Am. J. Health Promot. 19 (3), 194–198.

Du, W., Zhan, Z., 2002. Building decision tree classifier on private data. In: Proc., Proceedings of the IEEE International Conference on Privacy, Security and Data Mining-Volume 14. Australian Computer Society, Inc.. pp. 1–8.

Duff, A., Robertson, I., Phillips, R., Cooper, M., 1994. Improving safety by the modification of behaviour. Constr. Manag. Econ. 12 (1), 67–78.

Fang, D., Wu, C., Wu, H., 2015. Impact of the supervisor on worker safety behavior in construction projects. J. Manag. Eng. 31 (6), 04015001.

Fang, D., Zhao, C., Zhang, M., 2016. A cognitive model of construction workers' unsafe behaviors. J. Constr. Eng. Manag. 142 (9), 04016039.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognit. Lett. 27 (8), 861–874.

Finlay, K.A., Trafimow, D., Moroi, E., 1999. The importance of subjective norms on intentions to perform health behaviors. J. Appl. Soc. Psychol. 29 (11), 2381–2393.

Fishbein, M., 2008. A reasoned action approach to health promotion. Med. Decis. Making 28 (6), 834–844.
Fishbein, M., Ajzen, I., 2010. Predicting and Changing Behavior: The Reasoned Action Approach. Taylor & Francis Group, New York.
Fogarty, G.J., Shaw, A., 2010. Safety climate and the theory of planned behavior: towards the prediction of unsafe behavior. Accid. Anal. Prev. 42 (5), 1455–1459.
Gielen, A.C., Sleet, D., 2003. Application of behavior-change theories and methods to injury prevention. Epidemiol. Rev. 25 (1), 65–76.
Goh, Y.M., Binte Sa'adon, N.F., 2015. Cognitive factors influencing safety behavior at height: a multimethod exploratory study. J. Constr. Eng. Manag. 141 (6), 04015003.
Goh, Y.M., Chua, D., 2013. Neural network analysis of construction safety management systems: a case study in Singapore. Constr. Manag. Econ. 31 (5), 460–470.
Griffin, M.A., Neal, A., 2000. Perceptions of safety at work: a framework for linking safety climate to safety performance, knowledge, and motivation. J. Occup. Health Psychol. 5 (3), 347.
Guo, B.H.W., Yiu, T.W., González, V.A., 2015. Identifying behaviour patterns of construction safety using system archetypes. Accid. Anal. Prev. 80, 125–141.
Guo, B.H.W., Yiu, T.W., González, V.A., 2016. Predicting safety behavior in the construction industry: development and test of an integrative model. Saf. Sci. 84, 1–11.
Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., Tatham, R.L., 2006. Multivariate Data Analysis. Prentice hall Upper Saddle River, NJ, New Jersey.
Heinrich, H.W., 1931. Industrial Accident Prevention: A Scientific Approach. McGraw-Hill book Company, Incorporated.
Jemmott, J.B., 2012. The reasoned action approach in HIV risk-reduction strategies for adolescents. Ann.Am. Acad. Polit. Soc. Sci. 640 (1), 150–172.
Johnson, S.E., Hall, A., 2005. The prediction of safe lifting behavior: an application of the theory of planned behavior. J. Saf. Res. 36 (1), 63–73.
Jones, P., 1989. Understanding paper recycling in an institutionally supportive setting: an application of the theory of reasoned action. J. Environ. Syst. 19 (4), 307–321.
Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proc., Ijcai, Stanford. CA. pp. 1137–1145.
Kotsiantis, S.B., 2007. Supervised Machine Learning: a Review of Classification Techniques. Informatica.
Liao, C.-W., Perng, Y.-H., 2008. Data mining for occupational injuries in the Taiwan construction industry. Saf. Sci. 46 (7), 1091–1102.
Mitchell, T.M., 1997. Machine Learning. McGraw-Hill, New York.
Mullen, J., 2004. Investigating factors that influence individual safety behavior at work. J. Saf. Res. 35 (3), 275–285.
Nagy, M.S., 2002. Using a single-item approach to measure facet job satisfaction. J. Occup. Organ. Psychol. 75 (1), 77–86.
Patel, D., Jha, K., 2014a. Neural network approach for safety climate prediction. J. Manag. Eng. 31 (6), 05014027.
Patel, D., Jha, K., 2014b. Neural network model for the prediction of safe work behavior in construction projects. J. Constr. Eng. Manag. 141 (1), 04014066.
Poulter, D.R., Chapman, P., Bibby, P.A., Clarke, D.D., Crundall, D., 2008. An application of the theory of planned behaviour to truck driving behaviour and compliance with regulations. Accid. Anal. Prev. 40 (6), 2058–2064.
Quick, B.L., Stephenson, M.T., Witte, K., Vaught, C., Booth-Butterfield, S., Patel, D., 2008. An examination of antecedents to coal miners' hearing protection behaviors: a test of the theory of planned behavior. J. Saf. Res. 39 (3), 329–338.
Reason, J.T., 1997. Managing the Risks of Organizational Accidents. Ashgate Aldershot.
Russell, S.J., Norvig, P., 2010. Artificial Intelligence: a Modern Approach. Prentice Hall, Upper Saddle River, N.J.
Seo, D.-C., 2005. An explicative model of unsafe work behavior. Saf. Sci. 43 (3), 187–211.
Singh, K., Leong, S.M., Tan, C.T., Wong, K.C., 1995. A theory of reasoned action perspective of voting behavior: model and empirical test. Psychol. Market. 12 (1), 37–51.
Timofeev, R., 2004. Classification and Regression Trees (CART) Theory and Applications. Humboldt University, Berlin.
Tixier, A.J.P., Hallowell, M.R., Rajagopalan, B., Bowman, D., 2016. Application of machine learning to construction injury prediction. Autom. Constr. 69, 102–114.
Wanous, J.P., Reichers, A.E., Hudy, M.J., 1997. Overall Job Satisfaction: How Good Are Single-Item MeasUres? American Psychological Association.
Witten, I.H., 2011. Data Mining : Practical Machine Learning Tools and Techniques. Morgan Kaufmann, Burlington, MA.
Yang, K., Ahn, C.R., Vuran, M.C., Aria, S.S., 2016. Semi-supervised near-miss fall detection for ironworkers with a wearable inertial measurement unit. Autom. Constr. 68, 194–202.
Zhou, H., Romero, S.B., Qin, X., 2016. An extension of the theory of planned behavior to predict pedestrians' violating crossing behavior using structural equation modeling. Accid. Anal. Prev. 95, 417–424.
Zohar, D., 2010. Thirty years of safety climate research: reflections and future directions. Accid. Anal. Prev. 42 (5), 1517–1522.