Safety Science 117 (2019) 257–262

Contents lists available at ScienceDirect

Safety Science

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

Evaluating machine learning performance in predicting injury severity in agribusiness industries

Fatemeh Davoudi Kakhki[a,*], Steven A. Freeman[b], Gretchen A. Mosher[b]

[a] Department of Technology, San Jose State University, San Jose, CA, USA
[b] Depaertment of Agricultural & Biosystems Engineering, Iowa State University, Ames, IA, USA

ARTICLE INFO

Keywords:
Injury severity classification
Injury severity prediction
Machine learning

ABSTRACT

Although machine learning methods have been used as an outcome prediction tool in many fields, their utilization in predicting incident outcome in occupational safety is relatively new. This study tests the performance of machine learning techniques in modeling and predicting occupational incidents severity with respect to accessible information of injured workers in agribusiness industries using workers' compensation claims. More than 33,000 incidents within agribusiness industries in the Midwest of the United States for 2008–2016 were analyzed. The total cost of incidents was extracted and classified from workers' compensation claims. Supervised machine learning algorithms for classification (support vector machines with linear, quadratic, and RBF kernels, Boosted Trees, and Naïve Bayes) were applied. The models can predict injury severity classification based on injured body part, body group, nature of injury, nature group, cause of injury, cause group, and age and tenure of injured workers with the accuracy rate of 92–98%. The results emphasize the significance of quantitative analysis of empirical injury data in safety science, and contribute to enhanced understanding of injury patterns using predictive modeling along with safety experts' perspectives with regulatory or managerial viewpoints. The predictive models obtained from this study can be used to augment the experience of safety professionals in agribusiness industries to improve safety intervention efforts.

1. Introduction

Occupational incidents can affect workers' life, both in and out of work, and impose a considerable economic burden on employers, employees, insurance companies, medical care systems, and society (Suárez Sánchez et al., 2011). According to an estimation by the International Labor Organization (Organization, 2008), nearly 337 million occupational incidents are reported per year globally. Occupational injuries and incidents are caused through multiple factors that contribute to the occurrence of an incident (Sarkar et al., 2018). Considering the enormous human capital and financial losses from injuries, researchers have continually sought ways to gain a better understanding of factors that affect the occurrence and severity of incidents, and to improve the accuracy of predicting the likelihood of future injuries (Lord and Mannering, 2010).

A valuable and informative source of injury data with detailed information about the incident, the injured and the cost of injury is workers' compensation claims data. They provide useful details of workplace incidents such as injury cause and nature, injured body part, demographics of injured workers, and injury narratives (Wurzelbacher et al., 2016; Utterback et al., 2012). Using injury data, occupational incident analysis focuses on identifying prevalent causes of incidents to design proper prevention measures (Jacinto et al., 2009). Due to the significance of occupational injury management from the engineering and economic points of view in industry (Bevilacqua et al., 2008), it is necessary to learn from past incidents to plan measures that reduce the likelihood of future incidents (Field et al., 2014).

A review of literature, presented in Section 2 of this work, shows that machine learning (ML) methods were widely used in classifying and predictive modeling of future events in various fields including occupational injury analysis. However, there is no literature on evaluating the performance of ML techniques in classifying and predicting the severity of occupational incidents in agribusiness industries in the United States. The aim of this study is to apply, validate and compare the performance of ML methods in accurately classifying severity of occupational injury outcomes in various agribusiness industries in the Midwest of the United States using to a data set with over 33,000 workers' compensation claims in.

This study contributes to the rare current literature on analysis of non-farm agricultural-related occupational injuries by evaluating the

* Corresponding author.
E-mail addresses: fatemeh.davoudi@sjsu.edu (F. Davoudi Kakhki), sfreeman@iastate.edu (S.A. Freeman), gamosher@iastate.edu (G.A. Mosher).

https://doi.org/10.1016/j.ssci.2019.04.026
Received 12 July 2018; Received in revised form 5 March 2019; Accepted 18 April 2019

performance of ML methods, and specifically support vector machines (SVMs), for classification purposes in severity outcome prediction of such incidents. In addition, interpreting the results from boosted trees (BT) and Naïve Bayes (NB) models could provide important insights about the characteristics and severity outcomes of potential future occupational incidents, which aids safety practitioners and planners in agriculture-related industries who are interested in revising safety measures and reducing occupational incident rates.

The second section of the paper includes a literature review of ML methods and their application for classification and prediction purposes. Section 3 includes a discussion of the workers' compensation claims data set that the modeling was based on, followed by a brief description of methodology. Section 4 presents, assesses, and compares the results from each classification prediction ML model. Conclusions and discussion about the performance of the prediction methods, the effect of classification task on injury severity prediction and the applications of the proposed analyses in the safety field complete the paper.

2. Literature review

The typical tool used to study workplace incidents consists of generally descriptive statistical analysis (Matias et al., 2008). To identify the hidden patterns in data, more advanced analyses can be done using large data sets (Chen and Luo, 2016). Comparing traditional statistical models to machine learning methods, the latter outperforms in predicting future events (Sarkar et al., 2018). Generally, the machine learning algorithms that are applied for classification and regression purposes are decision trees, artificial neural networks, extreme learning machines, Bayesian networks, and SVMs (Witten et al., 2016).

In classification and predictive modeling, ML algorithms are used for predicting the class or category of an observation based on the information extracted from a data set consisting of training data points. The models are then validated on a set of new data called testing data. ML classification and prediction models are preferred over parametric models since the latter does not show optimal performance in catching the relationship between the independent variables and the target variable of interest in the analysis (Tixier et al., 2016). Classifying and predictive modeling of future events has been done successfully in various fields such as engineering, management, healthcare, and medicine (Oztekin et al., 2018; Kavakiotis et al., 2017; Kotsampasakou et al., 2017; Crown, 2015; Zheng et al., 2018; Chou et al., 2014; Aviad and Roy, 2011). However, ML techniques have been used on a limited basis in the analysis of occupational injuries (Sears et al., 2014; Bevilacqua et al., 2008). A review of literature shows that different ML methods were applied in a limited way in analysis, classification, and prediction of occupational injuries, mostly in the construction industry (Chokor et al., Chen and Luo, 2016; Yi et al., 2016; Tixier et al., 2016; Leu and Chang, 2013; Rivas et al., 2011), mining industry (Sanmiquel et al., 2015; He et al., 2010), and crash severity analysis in transportation (Lord and Mannering, 2010; Ding et al., 2018; Delen et al., 2017; Alikhani et al., 2013; Yu and Abdel-Aty, 2013; Li et al., 2011).

2.1. Support vector machines

SVMs with nonlinear kernels have been successfully used for classification purposes due to their ability to map data in a higher dimensional feature space where classes are easily separable (Olson et al., 2012). Kecman (2005) stated that SVM algorithms have shown comparable or higher accuracy in classification and regression problems, in comparison with other statistical and ML methods. SVM predictive models yield results with high accuracy in binary classification problems (Gangsar and Tiwari, 2017). The reason lies in the SVM algorithm, which identifies an optimal boundary able to separate the two classes of the target variable (Mwangi et al., 2015). The three popular kernel functions for SVM are linear, radial basis function (RBF), and polynomial of degree *d*.

2.2. Gradient boosting trees

Decision trees (DT) have gained popularity because they are transparent and easily interpretable powerful classification algorithm (Olson et al., 2012). The main reason to use DT is to interpret quantitative and qualitative patterns in the data to explore hidden information (Sarkar et al., 2018). Among different classification DT algorithms, boosting was considered as one of the most important advances in ML over the last 20 years since it can turn an ensemble of weak classifiers into strong classifiers (Hastie et al., 2001). Boosting is an ensemble approach that combines many base models to create predictions (Freund et al., 1999). While building BT, a sequence of very small trees is grown such that a successive tree focuses on the attributes of the training set that were missed in the preceding tree (Hastie et al., 2001). In addition to successfully classifying and predicting injury severity out- comes (Sut and Simsek, 2011), BT models can contribute to identifying factors that affect incident severity from the input variables (Zheng et al., 2018). In addition, BT models can be used in computing the extent of each variable's importance in predicting the severity classification of an injury (Saha et al., 2015).

2.3. Naïve Bayes

Bayesian classifiers are based around the Bayes rule that uses conditional probabilities for classification of a categorical target variable based on the input variables (Troussa et al., 2013). The NB classifier is one of the most widely-used classifiers in machine learning. NB assumes that variables are conditionally independent and, despite being a simplistic method, it reports the best performance in various classification tasks (Moreira et al.,). In other words, the NB algorithm reduces the complexity of Bayesian classifiers by making a conditional independence assumption that dramatically decreases the number of parameters to be estimated from the original $2(2^n - 1)$ to just 2^n when modeling $P(X|Y)$, where X is the independent variable, Y is the categorical response variable, and n is the number of independent variables used in the analysis (Mitchell). In addition, NB is one the fastest classifiers for prediction and classification purposes on large-scale data sets that can handle both categorical and continuous data (Bhowmik, 2015). Therefore, NB has been proven to be a simple and effective ML classifier in text classification studies (Liu et al., 2013). NB models usually have high accuracy when the response variable has two classes such as in injury severity with non-severe and severe classes (Marucci-Wellman et al., 2017).

3. Material and methods

In this research, a set of data was available from a leading insurance company located in the Midwest of the United States with 34 variables in 33,458 rows (claims). After cleaning the data, 16 variables were selected for the analysis. The 16 columns that were not used in the analysis included either irrelevant or repeated information. Two columns of insurance account number for the employers who filed the claims and jurisdiction state of the claims were irrelevant factors to the severity of the incident. The other columns repeated information since they included the numerical representation of the categorical variables selected in the study. For instance, for the variable *Type of Injury*, various numerical codes exist such as 06 for medical injuries, 09 for total permanent injury, etc., Also, *Age* and *Tenure* variables were calculated and added to the dataset since the original data included the date of birth, the date of employment, and the date of incident for the injured workers. Thus, the unique relevant selected categorical variables are *Market*, *Sex*, *Accident State*, *Class description*, *Occupation*, *Type of Injury*, *Injury Cause*, *injury Cause Group*, *Injury Nature*, *Injury Nature Group*, *Body Part Group*, *Body Part Injured*, and continuous variables are

Age, *Tenure*, and *Class Code* (which is the numerical representative of Class Description). It is worth mentioning that the variable *Market* refers to non-farm agribusinesses such as grain handling, feed milling operations, refine fuels, fertilizer blending and distribution, grain milling plants, seed conditioning, mushroom farming, dairy processing, and fruit and vegetable processing.

In predictive modeling, data is preferably divided into training and test data sets. The reason is that the model that is fitted on the training set should be applied on the test set for an evaluation of the overall model performance. The selection of data points' assignment to train or test set is done through stratified re-sampling method. Stratified sampling is a well-known sampling technique in data mining that can adequately capture the characteristics of the data (Shields and Teferra, 2015; May et al., 2010). Stratified sampling is conducted in two steps: first, the whole data set is split into non-overlapping subgroups (where data points for any individual stratum are in close statistical agreement); second, data points are re-sampled randomly from the strata into the two subgroups data called training and testing sets (Rezk et al., 2017; Sahoo et al., 2012).

3.1. Severity classification

Binary classification is the most popular classification task in ML modeling (Sokolova and Lapalme, 2009). A typical workers' compensation claim is monetary value, called "total incurred" amount, which consists of three main cost categories: medical costs, indemnity costs, and other expenses. The dependent variable in this study was the severity outcome of an incident based on its workers' compensation claim monetary loss. Thus, a new variable, "severity", was added to the data set with two main levels. Claims with total incurred cost between $0 and $10,000 were classified as non-severe (NS). Claims with total incurred value above $10,000 were classified as severe incidents (S). Based on the new severity classification, 87.6% of claims were representative of non-severe incidents followed by 12.29% classified as severe incidents out of 33,458 total claims.

3.2. Predictor selection

To determine the dependency of two categorical variables, the chi-square statistical test of independence was used. First, chi-square statistics plus its relevant *P-value* was calculated between each attribute (predictor) and the target variable. If the target variable was independent of the input variable, the predictor variable was discarded. Otherwise, the input variable was counted as an important predictor of the target variable. The result of chi-square analysis is shown in descending order in Table 1. All attributes selected for the study showed a significantly high chi- square statistics and thus were as independent variables in the modeling stage.

As shown in Table 1, the type of injury was the most important factor in determining the severity of an incident. The least important variable, yet statistically significant, was the injury cause. The chi-square test showed that the agribusiness industries and the workers' specific occupation class code were also predictors of the incident outcome. However, they were not included in the modeling because the models were built with and without those two variables, yet model performance did not change. To reduce the complexity and volume of the proposed models, agribusiness industries and the workers' specific occupation class code were not used for the final modeling phase and results that were presented in this work.

Table 1
Variable importance using chi-square test.

Independent variable	chi-square	*p*-value
Injury	15989.72	0.00
Nature of injury	3070.40	0.00
Injured body part(s)	2056.01	0.00
Injury cause group	1056.91	0.00
Age	777.63	0.00
Injured body group	513.72	0.00
Tenure	210.80	0.00
Injury nature group	166.29	0.00
Cause of injury	25.96	0.00

3.3. Partitioning data

Data for this analysis was divided into two parts: training set, and testing set, using stratified sampling method. The training set includes 70% of the data points. This set is used to fit the model of interest and estimate model parameters. The model fitted to the training set is, then applied to the testing set, which includes 30% of data points that have not been used in the training data points and is used to assess the overall error of the final model. The decision about the usefulness of a predictive model is made against the test set.

3.4. Classification and prediction modeling

The methodology in this research was predictive modeling via several ML classification methods. Three classifiers (1) SVMs with linear, quadratic and RBF kernels, (2) Boosted Trees (BT), and (3) Naïve Bayes (NB) were applied for classifying and predicting occupational incident severity outcomes in agribusiness industries. The reason for performing SVMs with different kernel functions was the importance of kernel function selection in improving the accuracy of prediction classifiers. The response variable was binary severity class of the incident, and the input variables were those from Table 1 that were selected as the main predictors of incident severity, using chi-square test, in the current data set. The analyses were done using JMP Pro statistical software version (JMP®, Version < 13.2 > . SAS Institute Inc., Cary, NC, 1989–2007), and MATLAB 2017a (The MathWorks, Inc., Natick, Massachusetts, United States.).

3.5. Model quantitative performance metrics

To compare classification models, various performance metrics gained from a confusion matrix are used typically. The confusion matrix for a binary classifier is shown in Table 2. The confusion matrix, which has the form of a contingency table, shows how the observations are spread over actual classes (rows) and predicted classes (Guns et al., 2012). In classification methods, confusion matrix is the basis of the predictability power of the model. A confusion matrix shows the correct and incorrect number of cases classified under a defined target. It is used to calculate the accuracy of the prediction. The matrix has four kinds of instances. True positive (TP) and false positive (FP) are instances of correct and incorrect classifications per actual class, respectively. True negative (TN) and false negative (FN) are instances of correct and incorrect rejections per actual class, respectively (Labatut and Cherifi, 2011). In this study, the binary confusion matrix for each ML model was used for calculating the model performance quantitative measures. In the following, the metrics for model performance evaluation are presented and described (Sokolova and Lapalme, 2009; Guns et al., 2012; Mathew, 2016; Shreve et al., 2011).

Table 2
Confusion matrix for binary classification.

Actual class	Predicted class	
	NS (Negative)	S (Positive)
NS (Negative)	TN	FP
S (Positive)	FN	TP

Recall or *sensitivity* ($\frac{TP}{(FN+TP)}$) shows the effectiveness of a classifier in identifying positive labels (Sokolova and Lapalme, 2009). *Specificity* ($\frac{TN}{(TN+FP)}$) shows how effectively a classifier recognizes negative labels (Sokolova and Lapalme, 2009). *Precision* ($\frac{TP}{(TP+FP)}$) evaluates class agreement of the data labels with the positive labels defined by the classifier (Sokolova and Lapalme, 2009). *F-score* ($\frac{2(Precision * Recall)}{(Precision + Recall)}$) is a weighted average of the recall and precision (Guns et al., 2012; Shreve et al., 2011). Overall accuracy ($\frac{TN+TP}{Total}$) shows how often the classifier is correct in overall while overall error rate shows how often the classifier is wrong in overall ($1 - Overall\ \ Accuracy$) (Mathew, 2016).

4. Results

The ML models were used to classify the binary severe/non-severe response using the input variables from Table 1. In this section, the performance of the ML models on the training, testing, and overall data sets is discussed. The quantitative measures of model performance were gained from the confusion matrices, which included the frequency of the binary response in actual and predicted classes. The model performance metrics are also explained. A discussion of the information gained from BT and NB models regarding the factors influential on predicting the injury severity outcomes, completes this section.

4.1. Analysis and model evaluation

Data was split into training set (70%) that includes 23,421 incidents and testing set (30%) that has 10,037 incidents records. Assigning data points to the training and testing data sets was done using stratified resampling. The models that were built using the training data were then used on the testing data to evaluate their performance. Table 3 includes the results of models in classifying severe and non-severe injuries in actual versus predicted relevant categories.

Results form Table 3 were used to calculate the numerical values for recall, specificity, precision, *F*-score, overall accuracy, and overall error (misclassification) metrics per model. The main purpose of comparing model performance is determining the accuracy differences among all model types to choose the best model (Oztekin et al., 2018). The prediction results on test data sets are presented in Table 4.

Positive and negative classes in this study were considered as S, and NS respectively. This was used in interpreting recall and specificity values. Recall value showed the models' performance in classifying the S cases while specificity revealed the models' ability in classifying the NS cases correctly. All models were capable of classifying NS injuries with high accuracy between 94.28% and 99.64%. This was expected due to the high frequency of NS cases in the original data set. Regarding the recall values, SVM (RBF) had the highest overall classifying power of 89.55% compared to all others that had a recall value between 64.02% and 80.83%, considering both training and testing datasets. This is important since the proportion of S cases was only 12.29% of all the data points.

Table 3
Confusion matrix for all models (train vs test data).

Model	Actual class	Predicted (train)		Predicted (test)	
		NS	S	NS	S
SVM (linear)	NS	20,100	443	8,620	183
	S	1,010	1,868	44	790
SVM (quadratic)	NS	20,104	439	8,638	165
	S	920	1,958	389	845
SVM (RFB)	NS	20,465	78	8,775	28
	S	412	2,466	129	1,105
BT	NS	20,095	448	8,622	181
	S	1,002	1,876	792	1,234
NB	NS	19,392	1,151	8,274	529
	S	658	2,220	274	960

Another metric used in this study was *F*-score. To evaluate the performance of a classifier, the *F*-score is one of the most useful measures since it is the harmonic mean of precision and recall (Mathew, 2016). Overall, SVM classifiers showed a higher *F*-score compared to BT and NB with values of 0.72, 0.75, and 0.92 for linear, quadratic and RBF kernels respectively. Considering *F*-score as a weighted measure of performance between recall and precision, SVM (RBF) showed better performance in predicting the incident severity classification.

Based on the performance metrics for all models, SVM with RBF kernel outperformed the linear and quadratic SVM models, as well as the BT and NB models, indicating the best performance. In addition, all SVM models, regardless of the kernel function, showed equal or higher values of *F*-score, overall accuracy, and lower overall misclassification rate compared to BT and NB.

4.2. Application in safety

Based on the analysis in this study, BT and NB models did not show the best performance in predicting incident severity outcomes. Yet, they had a considerable predictive accuracy and could provide useful information about the most important factors in predicting the severity of occupational incidents in agribusiness industries. The results from the BT models indicated that, on average, the most significant variable in prediction of injury severity level was the type of injury (61.14%). Cause of injury, injured body part (s), and nature of injury were statistically important variables as well. The least significant variables in estimating the severity outcomes were age and tenure of the injured workers. Considering layouts for all of the trees, the factors with the highest contribution to severe injuries were identified. All permanent partial disabilities were predicted as severe. The causes of injury that contributed significantly to the severity of occupational injuries in this analysis included injuries caused by repetitive motions, twisting, pushing and pulling, lifting, strain, slip on ice or snow, falling from ladder or elevation, using tools or machinery, objects being lifted or handled, and falling or flying objects. The injured body parts and groups that contributed to the severity of occupational injuries in upper extremities and lower extremities were predicted to occur in shoulder (s) and knee, lower leg, ankle, wrist, elbow, skull, soft tissues, hip, and abdomen including groin. The most significant natures of injuries predicted to result in severe injuries included concussion, dislocation, carpal tunnel syndrome, hernia, rupture, fracture, strain or tear, multiple injuries, and respiratory disorders.

According to the NB predictive model, permanent partial disabilities, in addition to temporary total or temporary partial disabilities, had higher probability of ending severe compared to other types of injuries. In addition, cause groups of injury with the highest contribution to severe incidents were strain or injury by, and fall, slip, or trip injury groups. Considering cause of injury, injured body part (s) and groups, and injury nature, results from NB models agreed with the BT identifying the same factors and levels as the most important predictors of severe incidents.

These results are significant in practicing safety analytics since they show high predictive and accurate classification power. In addition, the proposed models can assist safety practitioner for the purpose of classifying potential incidents, identifying the link between underlying causes of injuries, and planning relevant strategies to remove such causes to reduce or eliminate sources of injury at work places.

Another application of the proposed modeling is that safety practitioners can prioritise to invest available resources on preventing those incidents that have a higher probability of turning into severe injuries and impose much higher burden on all involved parties including the injured workers and employers. For instance, the analysis showed that type of injury can help to predict the severity of the incidents up to 61%. To see how important the injury type was in the severity of the incident, the following information was extracted from the NB model.

Table 4
Model performance on test data.

Model	Recall	Specificity	Precision	*F*-score	Overall accuracy	Overall error
SVM (Linear)	64.02%	97.92%	91.19%	0.72	93.75%	6.25%
SVM (Quadratic)	68.48%	98.13%	83.66%	0.75	94.48%	5.52%
SVM (RBF)	89.55%	99.68%	97.53%	0.93	98.44%	1.56%
BT	64.18%	97.94%	81.40%	0.72	93.79%	6.21%
NB	77.80%	93.99%	64.47%	0.70	92.00%	8.00%

This was a scenario for an injured 41-year-old worker with 5.5 years at work with an incident caused by slip or trip in the pelvis in the lower extremities body group. When the injury was a medical type, it had 90% chance of ending as a non-severe injury while the same injury resulting in permanent partial disability had 97% chance of becoming a severe injury with high cost.

Another example of the applications of the modeling results in safety management of workplaces is the NB model that can provide information about the dominant causes of incidents that have a higher probability of resulting in severe permanent partial disability injuries. Those injuries caused by object being lifted, or handled, stepping on sharp object, temperature extremes, dusts, gases, fumes, or vapors, and fall from elevation (different levels) have a higher probability of turning into severe permanent partial disability injuries.

5. Conclusions

Many studies have applied regression and classification models in prediction of injury outcomes in various fields including medical, mining, and construction sectors. To the authors' knowledge, this study is the first to apply three machine learning methods in classifying injury severity outcomes in agroindustry. This study incorporated a large data set with a large number of demographic information and injury specific details in prediction of occupational injury severity in agribusiness industries. The relative variable importance helped in providing insights about the information with higher value in predicting injury severity level.

Considering all metrics of model evaluation, support vector machines with RBF kernel outperformed all other models in the current data set and can be proposed as a superior method for injury severity outcome prediction based on information from workers' compensation claims data. The high prediction power of the support vector machines classifiers indicate that they supersede simpler models such as boosting trees and Naïve Bayes in correctly and specifically classifying a target variable and justifies the choice of machine learning algorithms over parametric models. In addition, the results indicated that the machine learning models were able to predict the severity of injury outcomes with the highest accuracy rate of 98.44% on the test data with 99.68% accuracy in classifying non-severe and 89.55% accuracy in classifying severe outcomes successfully. This suggest that injury severity is not random and underlying patterns and trends can be revealed and discussed using powerful machine learning models. According to this study, the authors suggest that occupational injuries should be studied empirically and quantitatively in addition to being qualitatively approached through expert opinions with regulatory or managerial perspectives only. This provides the ground for applying quantitative modeling techniques in addressing safety concerns prior to, or along with, risk planning and management. Machine learning models can be used in complementing the experts' opinions including data-driven decision-making for safety practitioners and risk analysts in safety management field. For instance, the machine learning models built based on prior data for a specific industry can be used on new injury data from similar industries as a useful platform for providing safety practitioners with actionable feedback to plan more effective intervention efforts in a given workplace.

Even though the support vector machines showed the highest predictive and classification power in this study, they have been criticized for performing as a black-box which cannot be directly used to identify the relationships between the input variables and the outcomes (Li et al., 2011). To extract the detailed information about the relationship between the input variables and the outcomes, sensitivity analysis is used as a valuable method to evaluate the relationship between the inputs and outputs. A future study can focus on sensitivity analysis to provide insight from SVMs models to improve the application of analytics in safety field.

It should be noted that the results of the study are conditional on the occurrence of the incident. In other words, the machine learning models used the information from the workers' compensation claims that are filed for a past occupational injury as the predictors of injury severity. Predicting the severity of injury outcomes is valuable if done prior to the occurrence of incidents based on workers' medical records, injury history, and work environment. Since a consistent database linking the medical records and injury history of the workforce in specific sectors is currently unavailable, future research can focus on gathering such data to identify the important predictors of injury severity based on workers' specific medical history combined with the workplace environmental factors. Another area for future research would be to combine workers' medical data and workers' compensation data to model the interactive factors of injury, injury severity and costs, to predict days away from work as the result of the injury. Another important direction for future research is to apply the same machine learning techniques in other industries to validate the techniques and results of this study.

References

Alikhani, M., Nedaie, A., Ahmadvand, A., 2013. Presentation of clustering-classification heuristic method for improvement accuracy in classification of severity of road accidents in iran. Saf. Sci. 142–150.

Aviad, B., Roy, G., 2011. Classification by clustering decision tree-like classifier based on adjusted clusters. Expert Syst. Appl. 8220–8228.

Bevilacqua, M., Ciarapica, F.E., Giacchetta, G., 2008. Industrial and occupational ergonomics in the petrochemical process industry: a regression tree approach. Accid. Anal. Prev. 1468–1479.

Bhowmik, T.K., 2015. Naive bayes vs logistic regression: theory, implementation and experimental validation. Inteligen- cia Artificial 14–30.

Chen, H., Luo, X., 2016. Severity prediction models for falling risk for workers at height. Procedia Eng. 439–445.

Chokor, A., Naganathan, H., Chong, W.K., Asmar, M.E., Analyzing arizona osha injury reports using unsupervised machine learning, Procedia Engineering 1588–1593.

Chou, J.-S., Cheng, M.-Y., Wu, Y.-W., Pham, A.-D., 2014. Optimizing parameters of support vector machines using fast messy genetic algorithm for dispute classification. Expert Syst. Appl. 3955–3964.

Crown, W.H., 2015. Potential application of machine learning in health outcomes research and some statistical cautions. Value in Health 134–140.

Delen, D., Tomak, L., Topuz, K., Eryarsoy, E., 2017. Investigating injury severity risk factors in automobile crashes with predictive analytics and sensitivity analysis methods. J. Trans. Health 118–131.

Ding, C., Chen, P., Jiao, J., 2018. Non-linear effects of the built environment on automobile-involved pedestrian crash frequency: a machine learning approach. Accid. Anal. Prev. 116–126.

Field, W.E., Heber, D.J., Riedel, S.M., Wettschurack, S.W., Roberts, M.J., Grafft, L.J., 2014. Worker hazards associated with the use of grain vacuum systems. J. Agric. Safety Health 147–163.

Freund, Y., Schapire, R., Abe, N., 1999. A short introduction to boosting. J. Japanese Soc. Artif. Intell. 771–780.

Gangsar, P., Tiwari, R., 2017. Comparative investigation of vibration and current monitoring for prediction of mechanical and electrical faults in induction motor based on multiclass-support vector machine algorithms. Mech. Syst. Sig. Process. 464–481.

Guns, R., Lioma, C., Larsen, B., 2012. The tipping point: F-score as a function of the number of retrieved items. Inf. Process. Manage. 1171–1180.

Hastie, T., Tibshirani, R., Friedman, J., 2001. The elements of statistical learning. The Elements of Statistical Learning, Spring Series in Statistics. Spring New York Inc, New York.
He, X., Chen, W., Nie, B., Zhang, M., 2010. Classification technique for danger classes of coal and gas outburst in deep coal mines. Saf. Sci. 173–178.
Jacinto, C., Canoa, M., Guedes Soaresa, C., 2009. Workplace and organizational factors in accident analysis within the food industry. Saf. Sci. 626–635.
Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., Chouvarda, I., 2017. Machine learning and data mining methods in diabetes research. Comput. Struct. Biotechnol. J. 104–116.
Kecman, V., 2005. Support vector machines- an introduction. In: Support Vector Machines: Theory and Applications. Springer, pp. 1–47.
Kotsampasakou, E., Montanari, F., Ecker, G.F., 2017. Predicting drug-induced liver injury: the importance of data curation. Toxicology 139–145.
Labatut, V., Cherifi, H., 2011. Accuracy measures for the comparison of classifiers. In: Proceedings of the fifth international conference on information technology, pp. 3790.
Leu, S.-S., Chang, C.-M., 2013. Bayesian-network-based safety risk assessment for steel construction projects. Accid. Anal. Prev. 122–133.
Li, Z., Liu, P., Wang, W., Xu, C., 2011. Using support vector machine models for crash injury severity analysis. Accid. Anal. Prev. 478–486.
Liu, B., Blasch, E., Chen, Y., Shen, D., Chen, G., 2013. Scalable sentiment classification for big data analysis using naive bayes classifier. In: IEEE International Conference on Big Data, pp. 99–104.
Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transp. Res. Part A 291–305.
Marucci-Wellman, H.R., Corns, H.L., Lehto, M.R., 2017. Classifying injury narratives of large administrative databases for surveillance—a practical approach combining machine learning ensembles and human review. Accid. Anal. Prev. 359–371.
Mathew, D., 2016. Data mining and machine learning algorithm for workers' compensation early severity predict ion.
Matias, J.M., Rivas, T., Martin, J.E., Taboada, J., 2008. A machine learning methodology for the analysis of workplace accidents. Int. J. Computer Math. 559–578.
May, R.J., Maier, H.R., Dandy, G.C., 2010. Data splitting for artificial neural networks using som-based stratified sampling. Neural Networks 283–294.
Mitchell, T.M. Generative and discriminative classifiers: Naïve Bayes and logistic regression 1–17.
Moreira, M.W, Rodrigues, J.J., Oliveira, A.M., Saleem, K., Neto, A. Performance evaluation of predictive classifiers for pregnancy care, IEEE.
Mwangi, B., Wu, M.-J., Bauer, I.E., Modi, H., Zeni, C.P., Zunta-Soares, G.B., Soares, J.C., 2015. Predictive classification of pediatric bipolar disorder using atlas-based diffusion weighted imaging and support vector machines. Psychiatry Res.: Neuroimag. 265–271.
Olson, D.L., Delen, D., Meng, Y., 2012. Comparative analysis of data mining methods for bankruptcy prediction. Decis. Support Syst. 464–473.
I. L. Organization, 2008. Promoting safe and healthy jobs, the ilo global program on safety, health and the environment.
Oztekin, A., Al-Ebbini, L., Sevkli, Z., Delen, D., 2018. A decision analytic approach to predicting quality of life for lung transplant recipients: a hybrid genetic algorithm-based methodology. Eur. J. Oper. Res. 639–651.
Rezk, E., Awan, Z., Islam, F., Jaoua, A., Al Maadeed, S., Zhang, N., Rajpoot, N., 2017. Conceptual data sampling for breast cancer histology image classification. Comput. Biol. Med. 59–67.
Rivas, T., Paz, M., Martin, J.E., Martias, J.M., Garcia, J.F., Taboada, J., 2011. Explaining and predicting workplace accidents using data-mining techniques. Reliab. Eng. Syst. Saf. 739–747.
Saha, D., Alluri, P., Gan, A., 2015. Prioritizing highway safety manual's crash prediction variables using boosted regression trees. Accid. Anal. Prev. 133–144.
Sahoo, A.K., Zuo, M.J., Tiwari, M.K., 2012. A data clustering algorithm for stratified data partitioning in artificial neural network. Expert Syst. Appl. 7004–7014.
Sanmiquel, L., Rossell, J.M., Vintro, C., 2015. Study of Spanish mining accidents using data mining techniques. Saf. Sci. 49–55.
Sarkar, S., Vinay, S., Raj, R., Maiti, J., Mitra, P., 2018. Application of optimized machine learning techniques for prediction of occupational accidents. Comput. Oper. Res. 1–15.
Sears, J.M., Blanar, L., Bowman, S.M., 2014. Predicting work-related disability and medical cost outcomes: a comparison of injury severity scoring methods. Injury 16–22.
Shields, M.D., Teferra, K., Hapij, A., Daddazio, R.P., 2015. Refined stratified sampling for efficient monte carlo based uncer- tainty quantification. Reliability Eng. System Safety 310–325.
Shreve, J., Schneider, H., Soysal, O., 2011. A methodology for comparing classification methods through the assessment of model stability and validity in variable selection. Decis. Support Syst. 247–257.
Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. Inf. Process. Manage. 427–437.
Suárez Sánchez, A., Riesgo Fernánde, P., Sánchez Lasheras, F., de Cos Juez, F.J., García Nieto, P.J., 2011. Prediction of work-related accidents according to working conditions using support vector machines. Appl. Math. Comput. 3539–3552.
Sut, N., Simsek, O., 2011. Comparison of regression tree data mining methods for prediction of mortality in head injury. Expert Syst. Appl. 15534–15539.
Tixier, A.J.-P., Hallowell, M.R., Rajagopalan, B., Bowman, D., 2016. Application of machine learning to construction injury prediction. Autom. Constr. 102–114.
Troussa, C., Virvou, M., Espinosa, K., Llaguno, K., Caro, J., 2013. Sentiment analysis of facebook statuses using naive bayes classifier for language learning. IISA, 10-12-2013. IEEE.
Utterback, D.F., Schnorr, T.M., Silverstein, B.A., Spieler, E.A., Leamon, T.B., Amick III, B.C., 2012. Occupational health and safety surveillance and research using workers' compensation data. J. Occup. Environ. Med. 171–176.
Witten, I.H., Frank, E., Hall, M.A., Pal, C.J., 2016. Data mining: practical machine learning tools and techniques. Data Mining: Practical machine learning tools and techniques, Morgan Kaufman.
Wurzelbacher, S.J., Al-Tarawneh, I.S., Meyers, A.R., Timothy Bushnell, P., Lampl, M.P., Robins, D.C., Schnorr, T.M., 2016. Development of methods for using workers' compensation data for surveillance and prevention of occu – pational injuries among state-insured private employers in ohio. Am. J. Ind. Med. 1087–1104.
Yi, W., Chan, A.P., Wang, X., Wang, J., 2016. Development of an early-warning system for site work in hot and humid environemnts: a case study. Autom. Constr. 101–113.
Yu, R., Abdel-Aty, M., 2013. Utilizing support vector machine in real-time crash risk evaluation. Accid. Anal. Prev. 252–259.
Zheng, Z., Lu, P., Lantz, B., 2018. Commercial truck crash injury severity analysis using gradient boosting data mining model. J. Saf. Res. 115–124.