Safety Science 130 (2020) 104873

Contents lists available at ScienceDirect

Safety Science

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

Using machine learning and keyword analysis to analyze incidents and reduce risk in oil sands operations

Daniel Kurian[a], Fereshteh Sattari[b], Lianne Lefsrud[b], Yongsheng Ma[a,*]

[a] Department of Mechanical Engineering, University of Alberta, 9211 – 116 St, Edmonton, Alberta T6G 1H9, Canada
[b] Department of Chemical and Materials Engineering, University of Alberta, 9211 – 116 St, Edmonton, Alberta T6G 1H9, Canada

ARTICLE INFO

Keywords:
Machine learning
Keyword analysis
Risk matrix
Trend analysis
Prevention and mitigation strategies
Leading indicators

ABSTRACT

Many companies maintain large databases of incident reports. Incidents that have severe consequences are analyzed in detail to prevent recurrence, while minor incidents are typically just stored without any further evaluation. Especially with common incidents and those with lesser consequences, details that are necessary to understand the cause of the incident might be missing or recorded inconsistently. We argue that incidents can be reported more accurately and analyzed to provide learning value to companies maintaining databases to better prevent and mitigate risks, lower the cost of losses, and improve safety culture. The aim of this research is to apply machine learning and keyword analysis to create a digitalized system for efficiently reporting incidents that can be used to generate a risk matrix, trend report, prevention and mitigation strategies, and leading indicators for every incident report that is inputted.

During this research project, 15,000 incident reports were analyzed to build a customized library. The customized library included the labels used in machine learning, the keywords from the incident database, and a list of statements used to accurately describe incidents. The labels and keywords were matched to the statements in a logical manner and output results were also programmed to match the statements using a company's safety guidelines, standard operating procedures, and asset management systems. The basic structure for generating outputs was demonstrated using a large incident database provided by collaborators of the project and anonymized sample inputs. Three incident report case studies are also processed and presented using the proposed methodology, delivering risk matrix, trend analysis, prevention and mitigation strategies, leading indicators that can be used by workers and companies to increase hazard awareness and improve safety performance.

1. Introduction

Risk, as defined by the International Organization for Standardization [ISO] (2018), is uncertainty of all types and sizes, both internal and external, which can affect an organization as it attempts to achieve its goals. According to the Project Management Institute [PMI] (2004), risk management processes need to be tailored specifically to each project. As such, organizations work to manage risk by identifying, analyzing, and evaluating risk, and then taking appropriate courses of action – planning responses, implementing changes, and continual monitoring. Generally, the process of identifying hazards and estimating risk is considered qualitative risk management and should be conducted first to identify and prioritize risks requiring detailed quantitative analysis.

An effective method of identifying hazards and estimating risk is to analyze historical data (Patriarca et al., 2018). In the oil and gas industry, such historical data can be found in incident reports (Nordlöf et al., 2015; Laberge et al., 2014). Incident reports contain many instances of past shortcomings or failures, which can be used as learning experiences to prevent similar incidents from reoccurring. Companies can use this knowledge to train their workers, and workers can study specific cases to identify hazards and to learn appropriate responses and countermeasures.

In Alberta, incident reports are required to contain the location of the incident, time and date, name of the employer involved, contact information of the site contact, and a general description of the incident (Government of Alberta, 2019). To build rapport, some companies add further details to incident investigation. Some measures might include root cause analysis, hazard and operability (HAZOP) studies, and basic risk ranking procedures such as risk matrices (Nordlöf et al., 2015; Pasquini et al., 2017). The risk matrix is a tool used to provide an estimation of the frequency and possible consequences of the incident (on

* Corresponding author.
E-mail addresses: dkurian@ualberta.ca (D. Kurian), sattari@ualberta.ca (F. Sattari), lianne.lefsrud@ualberta.ca (L. Lefsrud), yongsheng.ma@ualberta.ca (Y. Ma).

https://doi.org/10.1016/j.ssci.2020.104873
Received 6 February 2020; Received in revised form 27 May 2020; Accepted 6 June 2020

two axes), identify the relative severity of the risk (mapping zones of low, medium, high, etc.), and to determine what course of action must be taken to prevent or mitigate future incidents of that type (Albery et al., 2016).

A risk matrix is easy to implement, maintain, understand, and explain – due to these benefits, the tool is commonly used by companies to assess risk (Thomas et al., 2013); however, there are many drawbacks to using a risk matrix for risk analysis, including human bias and inconsistencies when reporting (Goerlandt and Reniers, 2016; Duijm, 2015). To strengthen this existing system, we applied a supervised machine learning approach to accurately analyze and evaluate risk in incident reports in previous research.

Artificial Intelligence and Machine Learning (AI/ML) hold great promise for enhancing process safety management by visualizing data and recognizing patterns across big datasets in real-time, determining the most effective leading indicators, especially how they may relate to low-frequency high-consequence events, and prioritizing improvements to safety processes. AI/ML have already been applied to established process safety tools like bowties (Khakzad et al., 2013), process hazard assessments (PHAs) and layers of protection analysis (LOPAs) (Xu et al., 2018), and hazard and operability studies (HAZOPs) (Zhao et al., 2009).

However, for many reasons, implementing AI/ML into companies' legacy process safety management systems has been slow. These databases create an overwhelming amount of (often dirty) data that are rarely analyzed in detail and effectively leveraged.

There are several reasons for this. First, operators tend to only analyze incidents with severe consequences to prevent recurrence, while minor incidents are only stored without any further evaluation. Yet, high-frequency and low-consequence incidents often display leading indicators that are overlooked but would be useful to predict high-consequence incidents (Aven, 2011; Steen and Aven, 2011).

Second, while detailed data is used to create HAZOPs, PHAs, LOPAs, and bowties, there are issues with the data itself. This data is often 'dirty' or incomplete, fragmented across data sources, proprietary with little incentive for sharing with others, or has uncertain or contested ownership (Dong et al., 2017; Ransbotham et al., 2017).

Third, leading operators have invested in developing internal AI/ML skills through training or hiring, but many operators outsource their AI/ML services. Yet, operators are surprised by AI/ML researchers' and suppliers' requirement for large datasets to allow their algorithms to learn, which results in operators perceiving AI/ML as a high-effort, low-payoff venture (Ransbotham et al., 2017).

To address these barriers, researchers and consultants often aggregate data across operators to create more complete, consistent, and larger datasets to enhance algorithm training and 'detectability' of leading indicators (Kurian et al., 2020). Yet, cross-organizational aggregation and collaboration introduces other barriers such as: differences in representativeness, context, and content that makes the data incommensurate (Zuboff, 2015; Kellogg et al., 2020) and model overfitting that can lead to inaccurate predictions when the model is used on different or more general data (Bengio et al., 2017).

We have recognized and begun to address these barriers in our previous research: a total of 15,000 incidents were manually classified: descriptive labels, actual and potential risk scores, and consequence labels (environment, finance, health/safety, and reputation) were applied to each incident. The incident reports were then divided into training and test data, and the machine learning algorithm used the training data to predict labels for the test data. The result of this research was a machine learning algorithm that could apply labels to incidents with 75–90% accuracy (depending on the label), and the outputs were used to develop risk matrices and to analyze trends in incidents.

The machine learning used in previous research was an attempt to remove human bias, and this method allowed for consistent reporting of incidents. However, many different variables (mentioned earlier) had to be manually analyzed and it was difficult to improve the accuracy beyond a certain level. Some incident reports lacked the detail required for classification and it was impossible to completely remove bias as using a supervised learning model implies manual training.

We continue to address these barriers with this research, by using machine learning to attach a basic label to describe an incident report. Furthermore, this research applies additional keyword analysis to increase the accuracy of machine learning classification. This research provides significant changes to the current system of incident reporting. By having the user select options from a standardized list that allow for detailed analysis of risk, the user is required to accurately describe the risk involved in an incident. Additionally, due to the increased efficiency in reporting incidents, it becomes possible to provide practical outputs beyond typical risk evaluation: prevention and mitigation strategies, such as leading indicators to increase the awareness of hazards in the workplace. This information can be used to predict incidents and to train workers to prevent/mitigate the risk from incidents that might occur in the future.

2. Methodology

The objectives of this research are to:

- Strengthen the current incident reporting system by creating a customized library using artificial intelligence, machine learning, and statistics.
- Support the design of more sensitive risk prevention and mitigation strategies, as well as leading factors; and
- Enhance organizations learning from incidents and create opportunities to reduce losses.

Fig. 1 shows an overview of the methodology used in this research. The process of reporting incidents is expanded to a three-step procedure with an intermediate step for user input. For the first step, a company is required to provide data pertaining to past incidents and safety requirements. The second step involves designing a customized library for analyzing future incidents that are reported. The final step provides a detailed analysis and suggestions that can be used to prevent incidents from occurring or to minimize the damage caused by such events.

Fig. 2 provides a more detailed description of the steps involved in designing the customized library and delivering outputs (with the corresponding steps coloured similarly). For this research, several collaborating companies provided access to their incident databases containing incident reports from 2013 to 2017, inclusive. The methods described in this research are applied to the data supplied from one of the participating companies including a total of approximately 15,000 incident reports. A customized library is generated from this data and output results are programmed given the input data.

2.1. Input data

To design a customized library, there was a requirement for input data from the companies participating in the research. This data included an incident database containing incidents for the past five years, standard operating practices, safety procedures, and guidelines to develop proper responses to incidents and hazards, and asset management systems to better understand the different systems and equipment involved in incident reporting. Input data was stored securely and used to design a customized library of keywords for a company.

15,000 incident reports were selected from the provided incident reports, analyzed, and used to generate output results. These incident reports were used to train a machine learning algorithm to predict class labels for new incident reports that will be inputted. By using these class labels in conjunction with keyword analysis, it was possible to develop outputs for any incident report that shares similarities to other incidents in the incident database.

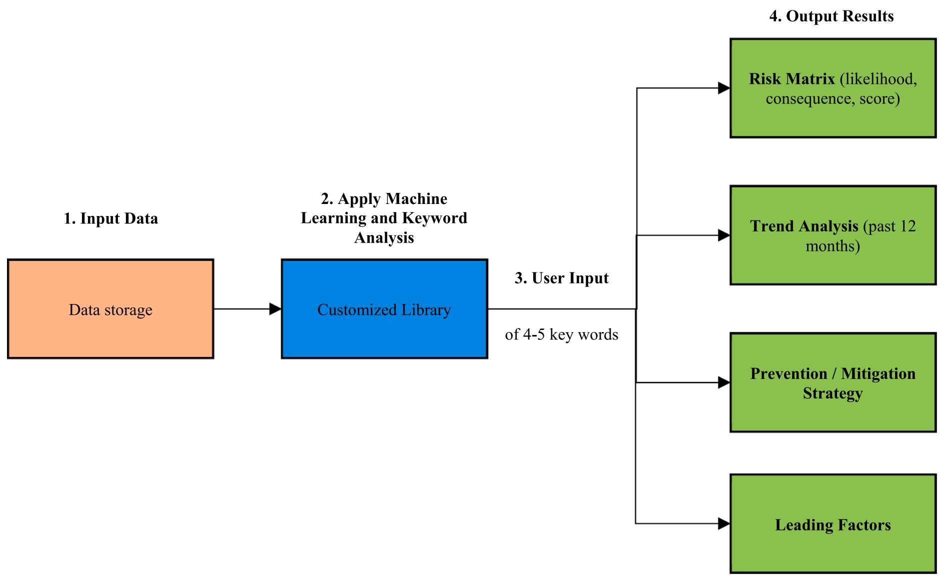

Fig. 1. Overview of methodology.

2.2. Apply Machine learning and keyword analysis

Applying machine learning and keyword analysis to incidents reports is a multi-step process. For this research, a supervised machine learning algorithm was used to classify incident reports. Depending on the data being analyzed and the selected classifier, the total computational time of supervised machine learning algorithms can be very small compared to other approaches (Singh et al., 2016). Supervised machine

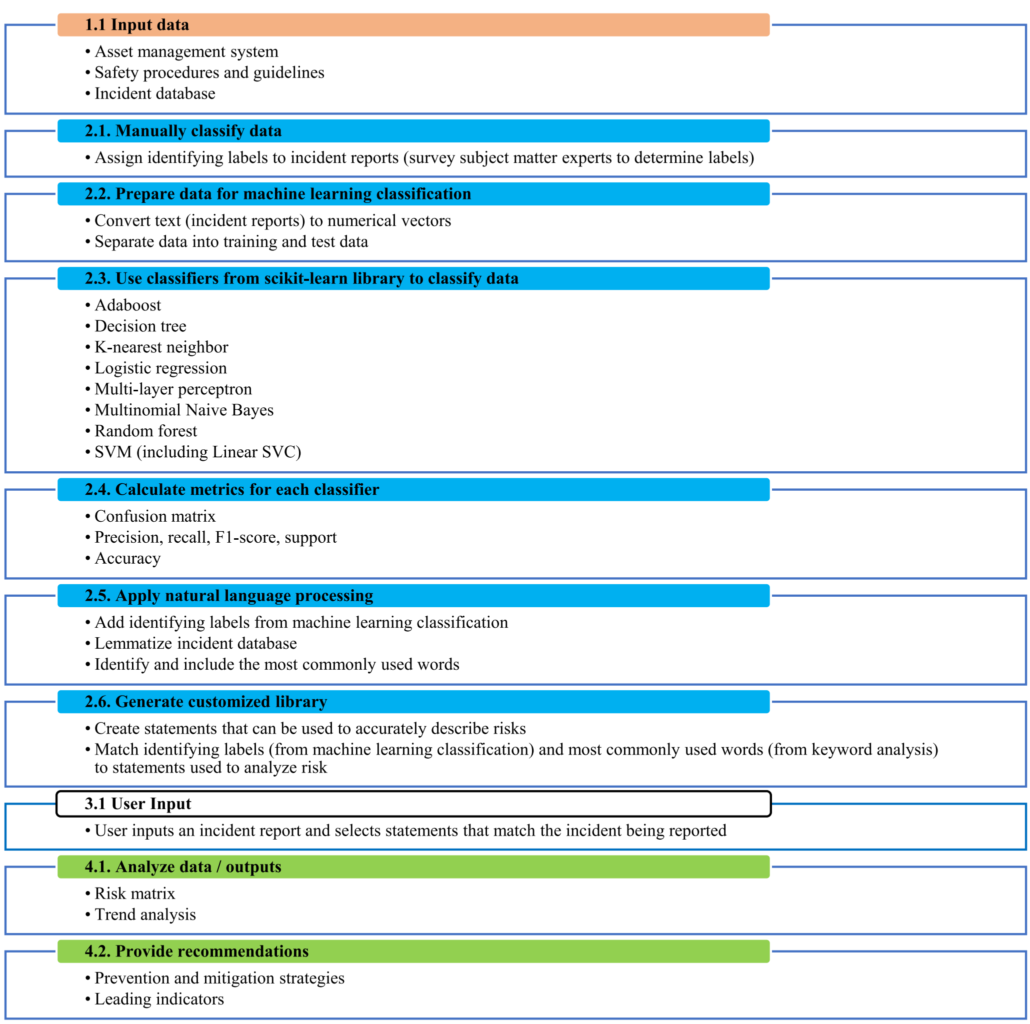

Fig. 2. Detailed description of methodology.

learning operates by using predictor features to forecast class labels – it aims to categorize data by utilizing prior information (Kotsiantis, 2007). The first step to implementing supervised machine learning towards the classification of incident reports is to manually classify incident reports by labelling them with consistent identifiers (key descriptors, immediate and latent causes, contributing factors). By interviewing university professors and industry experts from participating companies, the following labels were selected to identify incidents: communication, health/safety, leak/spill, miscellaneous, operation, uncategorized, and vehicle. The label of "uncategorized" was assigned to incident reports that could not be classified.

Once the incident reports were manually classified, the data in the incident database was prepared for machine learning classification. The TfidfVectorizer feature was used from Python's scikit-learn library to transform each incident report into a numerical vector, and thus, the incident database is transformed into a matrix (Imani et al., 2018). Alternatively, the incident database can be viewed as a dictionary with the individual incident reports being documents and the words found in the incident reports being terms.

The occurrence of each term is counted, and weights are applied by comparing how often a term is found in a document versus the entire dictionary. The result is the transformation of text to a numerical vector. These manually classified incidents were then separated into training and test data sets, containing 70% and 30% of the data, respectively (Ng).[1] The numerical vectors of the incident reports in the training set were expressed graphically, and a classifier was used to generate decision boundaries used to classify data. The numerical vectors representing the incident reports are considered sparse matrices – matrices in which most of the numbers are 0.

The reason for this sparsity is because of the way that the dictionary was built using the TfidfVectorizer – every word (term) found in the incident database is added to the dictionary sequentially. For every word (term) found in an incident report (document), a count is applied to the position of the word in the incident database (dictionary). Subsequently, the terms in the dictionary that are not found in the document are assigned values of 0. Given the massive number of terms compiled in the dictionary, the vectors used to represent each incident report, and thus, the matrix used to represent the incident database, will be sparse.

A number of classifiers from the scikit-learn library that are compatible with sparse matrices were used to classify the incident reports: Adaboost classifier, decision tree classifier, k-nearest neighbors, logistic regression, multi-layer perceptron classifier, multinomial Naïve Bayes classifier, random forest classifier, and support vector machine classifier (including linear support vector classifier). The supervised machine learning algorithm then attempted to identify features in the incident report that could be used to connect it to a given label, and metrics were calculated for different classifiers to identify the most suitable classifier for the data.

Previous research discovered that the most accurate classifier for categorizing incident reports from Alberta's oil and gas industry was the Linear Support Vector Classifier (Linear SVC), boasting accuracies close to 90% when predicting labels (Kurian et al. 2020). The metrics used were the confusion matrix, classification report, and accuracy score (Garreta et al., 2017). The confusion matrix was used to demonstrate how a classifier makes predictions for labels and requires the true and predicted classifications of the model. The confusion matrix is calculated by counting the number of true positives, true negatives, false positives, and false negatives. In a confusion matrix, the true label can be found on the y-axis and the predicted label on the x-axis. The classification report delivers precision, recall, F1-score, and support with inputs of the actual and predicted labels. These metrics can be described as follows:

$$Precision = \frac{true\ \ positives}{truepositives + falsepositives}$$

$$Recall = \frac{true\ \ positives}{true\ \ positives + false\ \ negatives}$$

$$F1 - score = \frac{precision + recall}{2}$$

Values for precision, recall, and F1-score will be between 0 and 1, where values closer to 1 represent a more robust model. Support is the count of true occurrences for each label. Finally, the accuracy score is the percentage of predicted labels that the model correctly identifies:

$$Accuracy = \frac{true\ \ positives + true\ \ negatives}{total}$$

After determining accuracies from the machine learning classification, Natural Language Processing (NLP) was used to analyze keywords. NLP allows computers to interact with humans by processing and analyzing natural language data (Srinivasa-Desikan, 2018). Aside from the scikit-learn library which was used to convert incident reports into numerical vectors, there are two Python libraries that are commonly used for NLP: spaCy and Natural Language Toolkit (NLTK). The primary difference between these two libraries is that spaCy adopts an object-oriented approach while NLTK is used as a string processing library. Consequently, spaCy is more efficient when working with words, while NLTK performs better than spaCy when analyzing sentences (Malhotra, 2018). As such, spaCy was selected for keyword analysis in this research. SpaCy has many features that can be used to pre-process text data – it comes with tokenization and lemmatization features which were used to transform the words in the incident database to their canonical form (Srinivasa-Desikan, 2018). For instance, the words "run," "running," and "ran" would all be reverted to "run."

Keyword analysis was completed by lemmatizing all the words found in the incident database. A counter was then used to identify and tally the lemmatized words, and these words were then arranged from most frequent to least frequent. The keywords that could be used to classify incidents were then selected to include in the customized library (stop words, punctuation, names of individuals, etc. were removed).

The customized library was created with two variables: the identifying *labels* used to train the machine learning algorithm and the *keywords* identified using the spaCy library. The labels and keywords stored in the customized library were then matched to *statements* that could be used to analyze and evaluate risk. These statements were used to encompass varying levels of risk and restrict a user to select an option that could be used to accurately analyze the risk in an incident. The purpose of using both machine learning and a "manual" keyword approach was to increase accuracy and ensure that the generated statements could accurately describe any incident. To some extent, the keyword analysis was also used as a buffer to compensate for misclassification by the machine learning algorithm.

2.3. User input

The labels and keywords found in the customized library were used to generate a list of statements. These statements were rule-based outputs developed in accordance to the inputted standard operating procedures, safety guidelines, and asset management systems provided by the company. When a user enters an incident into the system, there is a prompt to select statements applicable to the incident being reported. A list of statements is generated from which the user can select those relevant to the incident being reported. Parameters can be assigned to generate a specific number of statements and to restrict the maximum number of statements that can be selected. There was also an attempt to attach priority to the statements most likely to match the

[1] The sensitivity of the results to alternative ratios of training-test data (50–50, 60–40) changed the overall accuracy by < 1%.

Table 1
Confusion matrix for the Linear SVC when predicting the identifying label (Kurian et al., 2020).

		Predicted						
		Comm.	Health/S	Leak/Spill	Misc.	Operation	Uncat.	Vehicle
Actual	Comm.	62	2	0	1	9	0	0
	Health/S	0	709	4	0	110	0	20
	Leak/Spill	0	2	241	0	48	0	4
	Misc.	0	1	1	43	10	1	0
	Oper.	2	68	14	5	2571	4	51
	Uncat.	0	16	0	4	38	43	2
	Vehicle	0	31	7	2	91	1	547

incident – the statements generated by the supervised machine learning algorithm appear first followed by the statements selected by the keyword analysis. In practice, the statements selected by the keyword analysis should have a wider range of selection and should include the statements selected by the machine learning algorithm. This is because machine learning predicts a single label to match the incident whereas the keyword analysis identifies every word that is common in the incident report and the customized library. In the case where both the machine learning algorithm and keyword analysis yields the same results, the duplicates are removed, and the statements selected by the machine learning algorithm retain priority in the listing. Finally, there is also a feedback loop that is designed in the user input stage. When a user selects statements to match the incident report, this information is recorded and used to improve machine learning accuracy for future incident report classification.

2.4. Output results

When an incident report is inputted, four outputs are delivered (based on statements selected by the user): a risk matrix for the incident, trends of similar incidents, prevention and mitigation strategies to reduce the risk of the incident in the future, and leading indicators that can be identified by workers prior to the recurrence of a similar event.

A risk matrix is generated by calculating frequency and consequence (Ni et al., 2010). Frequency is a prediction of how likely it is for an incident to occur within a given time period. With access to a company's incident database, the actual count of incidents was used to calculate frequency as opposed to predicting the frequency of an incident. In Alberta's oil and gas industry, consequences can be categorized into four types: impact to worker health/safety, environmental damage, financial loss, and harm to a company's reputation (Muhlbauer, 2004). Based on the statement selected by a user, each incident is categorized into one or more of these consequence categories.

Another practical output that was delivered was trend reports. Trends were calculated by analyzing the statements selected by the user and the date of the incident. The total count of the selected statements was plotted by month to show incident trends and identify where and when improvements are needed and where safety measures are excelling.

Based on the inputted standard operating practices, safety, and asset management system, specific prevention and mitigation strategies were assigned to each of the statements that were selected. Additional statements were also programmed for specific groups of statements that were commonly selected together. This same process was applied to identify leading indicators for specific incidents. This type of output is based entirely on the input of an incident report and provides actionable information to users as they enter incident reports and to companies as they seek to reduce risk in their work sites.

2.5. Summary of methodology

To summarize the methodology, the first step is to input data, the second step is to process data, the third step is to input new incident reports, and the fourth step is to provide outputs. There is a feedback loop between steps 1 and 3 where new incident reports that are inputted will be analyzed and then added to the existing database.

Supervised machine learning was implemented in previous research to complete basic risk analysis and evaluation of incident reports in the form of risk matrix outputs, and this research was integrated into the current methods of analysis. For example, the trend reports currently generated are based on the outputs of machine learning from our past research (Kurian et al. 2020); however, the system is designed to create updated trend reports as new incident reports are inputted. As such, current analysis is based on the incident database that has already been provided by companies. In the future, as incident reports are added using the proposed methodology, outputs will become more specific to the newly inputted incident data as accuracy continues to improve.

3. Results and discussion

Identifying labels were used to manually classify 15,000 incident reports: communication, health/safety, leak/spill, miscellaneous, operation, uncategorized, and vehicle. As suggested by the previous study (Kurian et al., 2020), supervised machine learning was used with the Linear SVC classifier to predict labels for incidents since it provides the highest accuracy. Table 1 (Kurian et al., 2020) displays the confusion matrix for the Linear SVC classifier. The actual (manually classified) labels are shown on the y-axis while the predicted (machine learning classified) labels are shown on the x-axis. The main diagonal of this matrix demonstrates the number of true labels that the classifier accurately predicted.

Table 2 (Kurian et al., 2020) is the classification report for the Linear SVC when predicting the identifying labels for incident reports. From here, it can be seen how accurately each label is predicted by the supervised machine learning algorithm. The overall accuracy of the Linear SVC when predicting the identifying label is ~ 88.48%. Furthermore, by adding user selection, the accuracy would increase to an even greater extent. F1-scores (average of precision and recall) closer to 1 signify better model accuracy while support is the number of true

Table 2
Classification report for the Linear SVC when predicting the identifying label (Kurian et al., 2020).

Identifying Label	Precision	Recall	F1-score	Support
Comm.	0.97	0.84	0.90	74
Health/S	0.86	0.84	0.85	843
Leak/Spill	0.90	0.82	0.86	295
Misc.	0.78	0.77	0.77	56
Oper.	0.89	0.95	0.92	2715
Uncat.	0.88	0.42	0.57	103
Vehicle	0.88	0.81	0.84	679

occurrences of each label. [2]A low F1-score and support means the model requires more exposure to predictor features to become more accurate.

Prior research (Kurian et al., 2020) applied the supervised machine learning only to the incident database classification. The supervised machine learning is now used to predict labels for any new incident report that will be reflect on the interface of the system when the user enters. This function will be further developed for more useful inquiries with more advanced prediction capability but not in the scope of this paper.

Table 3 matches statements that can be used to accurately describe incidents to the identifying labels used in supervised machine learning classification. When the user inputs an incident report, the machine learning algorithm will predict a class label for the incident report. The statements found corresponding to the predicted label will then be made available for user selection. Note that some labels have only minor differences in syntax (e.g. 6 types of statements pertaining to equipment describing different types of risk and severity of consequences). These labels can play a strong role when determining outputs; further, selecting a specific statement from this list can help to distinguish a minor incident from a major incident.

Table 4 shows how statements are matched to keywords. It is important to remember that the spaCy library lemmatizes the words in the incident reports. This means that keywords can be inputted in their canonical form without having to account for variations of a word (i.e. verb tense, singular vs plural, etc.). Here, the 15,000 incident reports were analyzed, and words found in the incident reports that could be used to classify incidents were matched with corresponding statements. One point to note is that abbreviations are also considered as keywords – the spaCy library ignores words that it does not recognize when lemmatizing the incident reports. Some common abbreviations found in the incident reports are: HT (haul truck), MOP (maximum allowable pressure), SOL (safe operating limit), QA (quality assurance), ROW (right of way), and STF (slip/trip/fall).

To summarize, both the *labels* used in supervised machine learning and the *keywords* found in the incident reports are assigned to *statements*. When an incident report is inputted into the system, the machine learning algorithm predicts a label to describe the incident and the incident report is lemmatized for keyword analysis. A list of statements is then generated based on the predicted label and matching keywords. The user is required to select statements that accurately describe the incident.

The statements are also matched to practical outputs that can be used by industry. Table 5 demonstrates how statements are categorized into the consequence categories used to generate a risk matrix. Several of these statements can fall into multiple categories.

Our prior research generated a consequence scale to assign a numerical value denoting the severity of the risk, found in Fig. 3 (Kurian et al. 2020). This consequence scale was created using the average values of consequences taken from the risk matrices of several companies collaborating with this research.

Table 5 uses the consequence scale (from Fig. 3) to assign a severity rating to each statement and frequency scores were determined by using the tally of keywords in the incident database. Using the consequence and frequency scores, a risk score is generated. If multiple statements are selected, every category pertaining to the selected statements are represented on the risk matrix with their corresponding risk scores. If multiple selected statements have different consequence or frequency ratings within the same risk category, the greatest consequence value is selected to be represented on the risk matrix (along with its corresponding frequency).

By counting the selected statements, and taking into account the date of the incident, it is also possible to plot trends of specific incident types by month. It is also possible to design prevention and mitigation strategies to match statements and combinations of statements. A similar process can also be used to identify leading indicators. These suggestions can be designed using a company's safety guidelines and procedures and asset management systems.

Table 3
Statements generated for user selection based on labels selected by supervised machine learning algorithm.

Statements	Identifying Label
Equipment (Damage - cost < $1m)	Operation
Equipment (Failure - cost < $1m)	Operation
Equipment (General - cost < $1m)	Operation
Equipment (Damage - cost > $1m)	Operation
Equipment (Failure - cost > $1m)	Operation
Equipment (General - cost > $1m)	Operation
Fatality	Health/Safety, Vehicle
Fire (Damage)	Vehicle
Fire (Injury)	Health/Safety
Incorrect Operations	Communication, Operation
Injury	Health/Safety, Vehicle
Laceration/abrasion	Health/Safety
Leak/spill	Health/Safety, Leak/Spill
Minor Injury	Health/Safety, Vehicle
Miscellaneous	Miscellaneous
Miscommunication	Communication, Operation
Missing Equipment	Miscellaneous
Near Miss	Health/Safety, Vehicle
No Treatment Injury	Health/Safety, Vehicle
Property Damage	Miscellaneous
Quality Assurance	Communication
Severe Injury	Health/Safety
Slip/trip/fall	Health/Safety, Weather
Snow/ice	Weather
Sprain/strain	Health/Safety
Vehicle (heavy equipment)	Vehicle
Vehicle (light vehicle)	Vehicle
Vehicle collision (no injury)	Vehicle
Vehicle collision (with injury)	Vehicle
Weather	Weather
Wildlife	Miscellaneous

To illustrate our methodology, we present case studies with different consequences. We have received inputs from companies and generated the customized library. We assume that a user is inputting new incident reports, make assumptions about the statements selected by the user, and review the outputs created using the proposed methodology to demonstrate its practicality.

3.1. Case Study 1

The following sample incident taken from a company database is presented to demonstrate how this methodology is used to produce results: "*Hose-Traceability. Heat number on the elbows do not match with the heat number on the documents. Followed up with vendor to get appropriate heat numbers as they showed something different.*"

Given this incident report, the user will have to select from the following statements: Equipment (Damage - Cost < $1M), Equipment (Failure - Cost < $1M), Equipment (General - Cost < $1M), Equipment (Damage - Cost > $1M), Equipment (Failure - Cost > $1M), Equipment (General - Cost > $1M), Incorrect Operations, Miscommunication, Quality Assurance.

It was assumed that the user selects: Miscommunication, Quality Assurance, and Equipment (General – Cost < $1M).

With these statements, it was determined that the consequence is a financial risk with a consequence score of 5. By looking at the total number of occurrences of similar incidents, frequency was assigned a score of 3. This can be seen in the risk matrix found in Fig. 4. The sample incident provided is a low consequence incident that occurs

[2] Linear SVC uses the one-against-rest approach and the SVC uses the one-against-one approach (Milgram et al., 2006).

Table 4
Statements generated for user selection based on keywords found in incident report.

Statements	Keywords
Equipment (Damage - cost < $1m)	damage, defective, equipment, exchanger, filter, hose, maintenance, not working, pump, seal, sump, valve, working
Equipment (Failure - cost < $1m)	defective, equipment, exchanger, failure, filter, hose, maintenance, not working, pump, seal, sump, valve, working
Equipment (General - cost < $1m)	defective, design, equipment, exchanger, filter, hose, maintenance, missing, not working, pump, seal, SOL, sump, trip, valve, venting, working
Equipment (Damage - cost > $1m)	damage, defective, equipment, exchanger, filter, hose, maintenance, not working, pump, seal, sump, valve, working
Equipment (Failure - cost > $1m)	defective, equipment, exchanger, failure, filter, hose, maintenance, not working, pump, seal, sump, valve, working
Equipment (General - cost > $1m)	defective, design, equipment, exchanger, filter, hose, maintenance, missing, not working, pump, seal, SOL, sump, trip, valve, venting, working
Fatality	fatality, fire, h2s, vehicle
Fire (Damage)	alarm, burn, burnt, fire, flame
Fire (Injury)	alarm, burn, burnt, fire, flame
Incorrect Operations	adequate, allowable, engineering, exceed, exceeded, improper, incorrect, incorrect, operations, knowledge, less, management, missing, missing, sign, missing, tag, MOP, performance, skill, SOL, unacceptable, unauthorized, verbal, wrong
Injury	abrasion, fall, finger, fire, h2s, illness, injure, injury, laceration, rest, slip, sprain, stf, strain, trip, vehicle
Laceration/abrasion	abrasion, bruise, cut, finger, laceration, paper, cut, papercut
Leak/spill	contaminate, drain, overflow, spill, leak, smell, seal
Major leak/spill	contaminate, drain, overflow, spill, leak, smell, seal
Minor Injury	abrasion, aid, fall, finger, fire, first, illness, injure, injury, laceration, slip, stf, treatment, trip, vehicle
Miscellaneous	missing, missing, equipment, theft
Miscommunication	communicate, communication, incorrect, management, miscommunicate, miscommunication, missing, missing, tags, operation, order, unacceptable, vendor, wrong, performance, less, adequate, verbal, skill
Missing Equipment	missing
Near Miss	miss, near, near, miss
No Treatment Injury	no, treatment, stf, treatment
Property Damage	drain, fire, leak, odor, odour, smell
Quality Assurance	assurance, document, documentation, incorrect, order, qa, quality, vendor, wrong
Severe Injury	fall, fire, h2s, illness, slip, sprain, stf, strain, trip, vehicle, rest, injury, injure, finger, disability
Slip/trip/fall	fall, fell, ice, injure, injury, oil, slip, snow, stf, trip, water
Snow/ice	ice, nature, poor, weather, snow, weather
Sprain/strain	back, finger, ice, injure, injury, lift, oil, slip, snow, treatment, water
Vehicle (heavy equipment)	accident, bulldozer, collision, dozer, haul, truck, ht, loader, loader, ROW, vehicle, zoom, boom, zoomboom
Vehicle (light vehicle)	accident, bus, car, collision, light, vehicle, lv, ROW, truck, vehicle
Vehicle collision (no injury)	accident, bulldozer, bus, collision, crane, dozer, excavator, fork, lift, forklift, haul, truck, ht, loader, truck, vehicle, zoom, boom, zoomboom
Vehicle collision (with injury)	accident, bulldozer, bus, collision, crane, dozer, excavator, fork, lift, forklift, haul, truck, ht, loader, truck, vehicle, zoom, boom, zoomboom
Weather	hail, ice, nature, poor, weather, rain, sleet, snow, weather, wind
Wildlife	animal, bird, fish, fox, wildlife, wolf

Table 5
Statements categorized by consequence type.

Statement	Consequence
Leak/spill, Major leak/spill, Wildlife	Environment
Equipment (Damage - cost < $1m), Equipment (Failure - cost < $1m), Equipment (General - cost < $1m), Equipment (Damage - cost > $1m), Equipment (Failure - cost > $1m), Equipment (General - cost > $1m), Fire (Damage), Incorrect Operations, Leak/spill, Major leak/spill, Miscellaneous, Miscommunication, Missing Equipment, Property Damage, Quality Assurance, Vehicle (heavy equipment), Vehicle (light vehicle), Vehicle collision (no injury), Weather, Wildlife	Finance
Fatality, Fire (Injury), Injury, Laceration/abrasion, Minor Injury, Near Miss, No Treatment Injury, Severe Injury, Slip/trip/fall, Snow/ice, Sprain/strain, Vehicle collision (with injury), Weather	Health/Safety
Fatality, Severe Injury, Wildlife	Reputation

somewhat frequently. In most cases, missing quality assurance documents are simply inconvenient and may result in minor financial losses. Companies might decide to implement prevention methods or to encourage workers to be more methodical when filing such documents.

With respect to the formatting of the risk matrix, the axis labels (from least to greatest, 1–5 frequency scale and 5–1 consequence scale) was determined based on industry practice. The risk matrix is also given a gradient effect to show low impact risks as green and high impact risks as red.

A trend report, shown in Fig. 5, is generated by creating a histogram of similar incidents by month. The algorithm counts the number of occurrences of incidents with the same statements selected and displays the trends of these incidents for the past year. In this example, it can be observed that it is common for incidents of this type to occur during the middle of the year. It might also be worth investigating why such incidents occurred frequently in December, but were quite rare in January.

Based on the selected statements, it is also possible to design prevention and mitigation strategies. These suggestions can be modified in the future and tailored to match the safety guidelines and procedures of different companies. For this incident, the program has been designed to provide the following prevention and mitigation strategies: (1) verify that instructions are clearly received, (2) clarify any doubts with the individual(s) assigning the task, (3) ensure that the task at hand is logical without blindly completing the assigned work, (4) verify part number before and after the order, and (5) ensure that all documents are properly handled and stored. The leading indicators are identified as: (1) poor filing system and (2) inadequate training.

3.2. Case Study 2

Next, we present a different incident with multiple consequences to demonstrate the versatility of the methodology: "*Icy road conditions. Employee truck and 3rd party vehicle made driver side contact. Employee complained about minor whiplash.*"

This incident prompts the user to select from the following

Degree of Severity	5	4	3	2	1
Health/Safety	Minor injuries or illnesses that do not require first aid treatment or may require basic first aid treatment	One or more injuries or illnesses requiring medical treatment or resulting in restricted work.	One or more injuries or illnesses resulting in lost time	Single fatality or one or more long term disabilities	Multiple fatalities
Environmental	Inconsequential or no adverse effects, clean up confined to site or close proximity	Minor adverse effects, local emergency response, 0-6 months clean up	Medium adverse effects, local emergency response, short to medium term effects, 7-12 months clean up	Medium to significant adverse effects, intermediate emergency response, 1-4 years clean up	Off property impact requiring remediation taking 5 years or more. Major emergency response with significant adverse effects.
Reputation	No media coverage. Single stakeholder involvement with concerns addressed in the normal course of businesses. Temporary side road closure.	Local media coverage. Multiple stakeholders involved with concerns addressed in the normal course of business. Secondary road closure lasting < 24 hours	Extended local media coverage or one-time national media coverage. Key stakeholder involvement. Extended secondary road closure > 24 hours	National media coverage. Involves multiple key stakeholders. Operations interrupted. Major road closure < 24 hours.	International media coverage. Multiple key stakeholders involved. Operations shutdown and/or potential of future operations being prevented. Extended closure of major road.
Financial	Cost < \$1M	\$1M < Cost < \$10M	\$10M < Cost < \$100M	\$100M < Cost < \$500M	Cost > \$500M

Fig. 3. Consequence scale used to analyze the severity of incidents (Kurian et al. 2020).

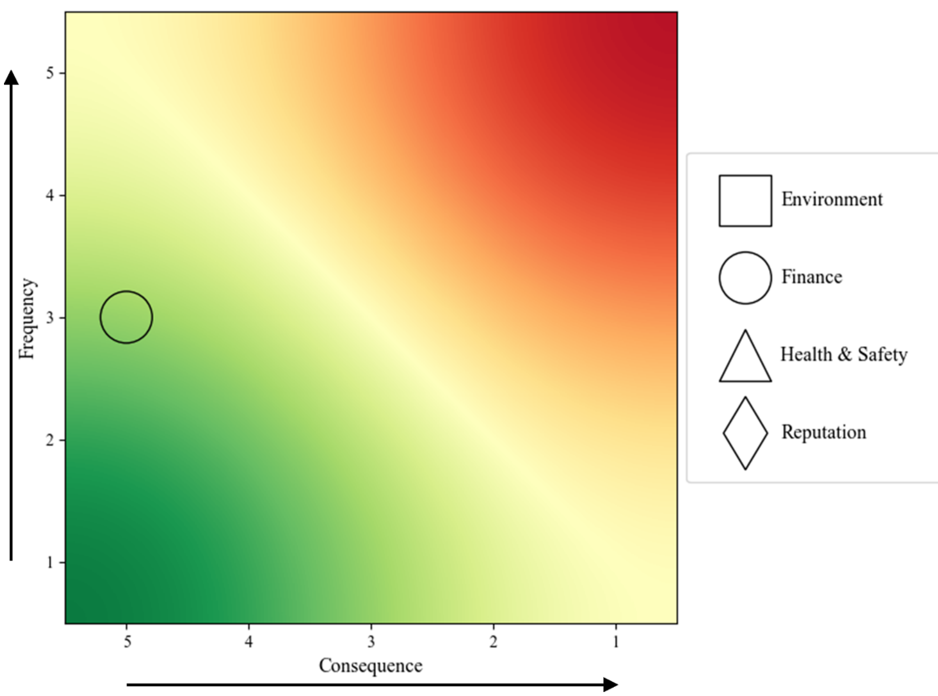

Fig. 4. Risk matrix for Case Study 1.

statements: Fatality, Fire (Damage), Fire (Injury), Injury, Minor Injury, Near Miss, No Treatment Injury, Severe Injury, Vehicle (Heavy Equipment), Vehicle (Light Vehicle), Vehicle Collision (No Injury), Vehicle Collision (With Injury).

Here, it was assumed that the user selects: Injury, Vehicle (Light Vehicle), and Vehicle Collision (With Injury).

Given these selections, it was determined that the consequence is a Health/Safety risk with a consequence score of 3 and a Financial risk with a consequence score of 5. By looking at the total number of occurrences of similar incidents, the frequency was assigned a score of 2 – the reports of vehicle collisions with injury are much fewer in number than those of vehicle collisions with no injury. The risk matrix for this incident can be seen in Fig. 6. The incident can be considered a low-to-medium risk where the health and safety component of the incident has more ramifications than the financial loss.

Fig. 7 shows the trends for vehicular incidents with injuries. As expected, vehicular incidents involving collisions occur more frequently in winter. It might be beneficial for the company to identify factors pertaining to the cause of similar incidents, such as the geographic location, time of day, existing traffic signs, driving conditions, etc., in order to determine methods for prevention and mitigation.

Based on the statements selected, the algorithm has been designed to provide the following prevention and mitigation strategies: (1) drive at a speed suitable to road conditions, (2) ensure that vehicle is properly equipped for winter weather (e.g. winter tires, first aid kit, etc.), (3) pay attention to other vehicles on the road (e.g. make sure other drivers are not distracted, maintain safe following distance, check blind spots), (4) make sure that the seat is properly adjusted to provide ample neck and lumbar support, and (5) provide training for workers to drive in winter road conditions. The leading indicators are identified as: (1) winter weather and (2) poor traction.

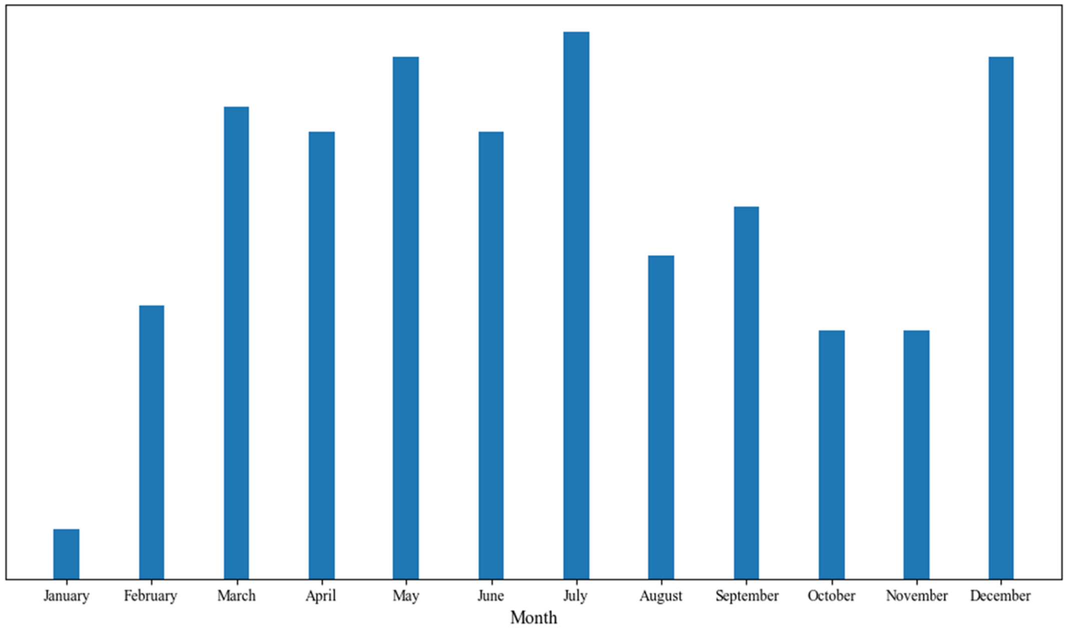

Fig. 5. Example of trend analysis for Case Study 1 indicating number of incidents per month.

3.3. Case Study 3

The final case study is a very frequent incident found in the database: "*Worker slipped and fell in the parking lot. Employee took a shortcut between the middle and largest (left most) park after parking her vehicle in the middle parking lot.*"

The resulting statements are given for the user to select from: Fatality, Fire (Injury), Injury, Laceration / Abrasion, Leak / Spill, Major Leak / Spill, Minor Injury, Near Miss, No Treatment Injury, Severe Injury, Slip / Trip / Fall, Sprain / Strain, Vehicle (Heavy Equipment), Vehicle (Light Vehicle), Vehicle Collision (No Injury), Vehicle Collision (With Injury).

It was assumed that Minor Injury and Slip / Trip / Fall were selected.

With these selections, the risk has a health and safety consequence score of 4 and a frequency of 5. Such incidents are very common, especially in the winter season. The risk matrix for the incident is displayed in Fig. 8.

An example of a trend report for slip/trip/fall incidents can be seen in Fig. 9. Slip/trip/fall incidents are common, particularly due to snow or ice in winter. There are also other contributing factors in other seasons that might result in slippery surfaces. It might be beneficial for a company to impress upon workers the importance of proper footwear, such as cleats, and requesting signage or countermeasures (e.g. salt or gravel) at the source of a tripping hazard.

Based on the statements selected to describe the incident, the algorithm provides the following prevention and mitigation strategies: (1) walk slowly and carefully on snow and ice, (2) wear proper and

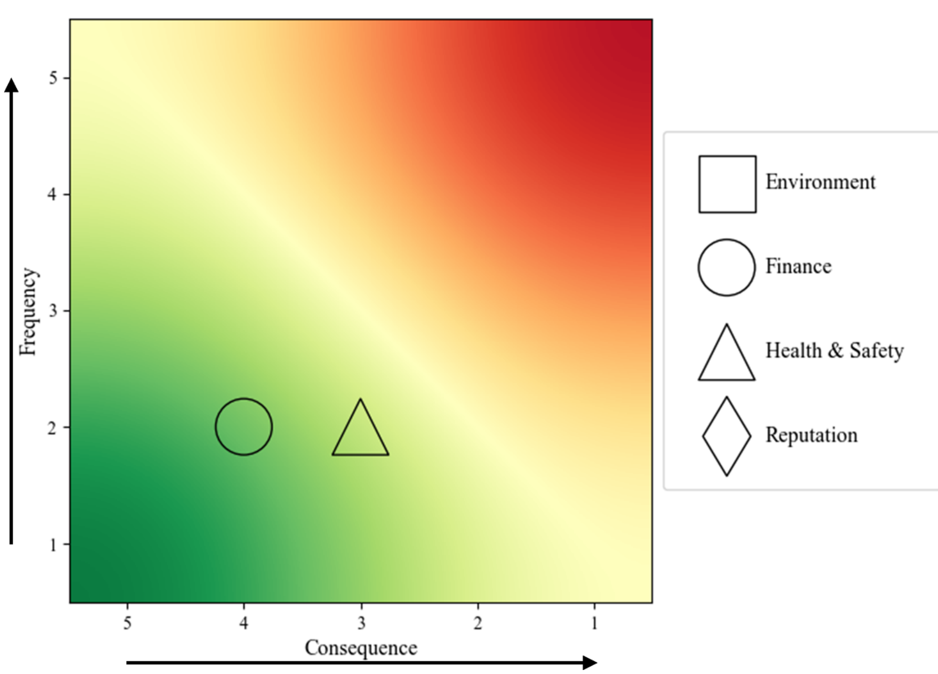

Fig. 6. Risk matrix for Case Study 2.

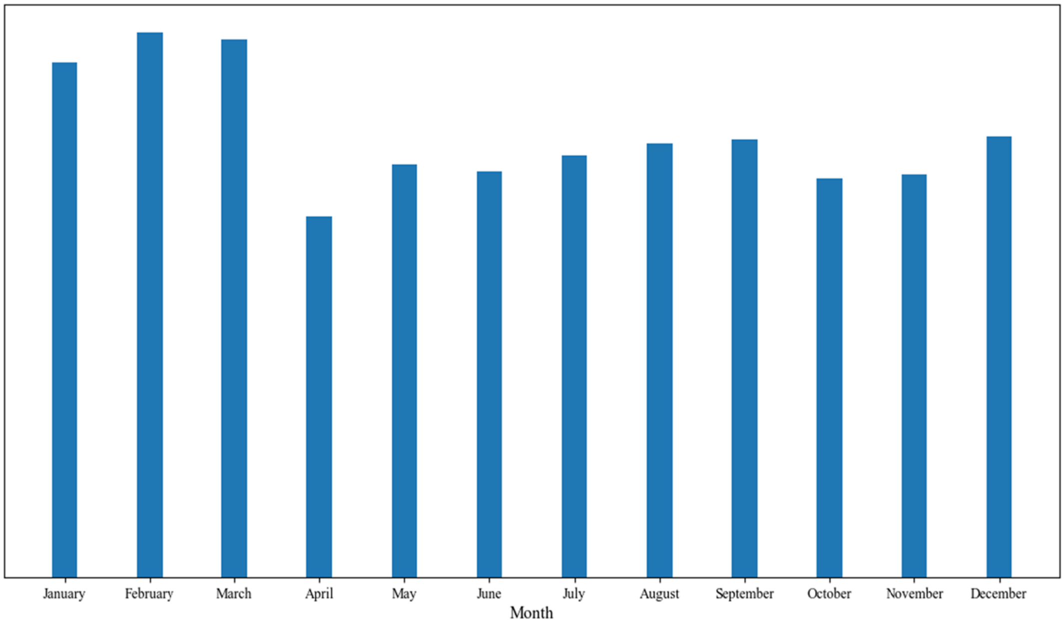

Fig. 7. Example of trend analysis for Case Study 2 indicating number of incidents per month.

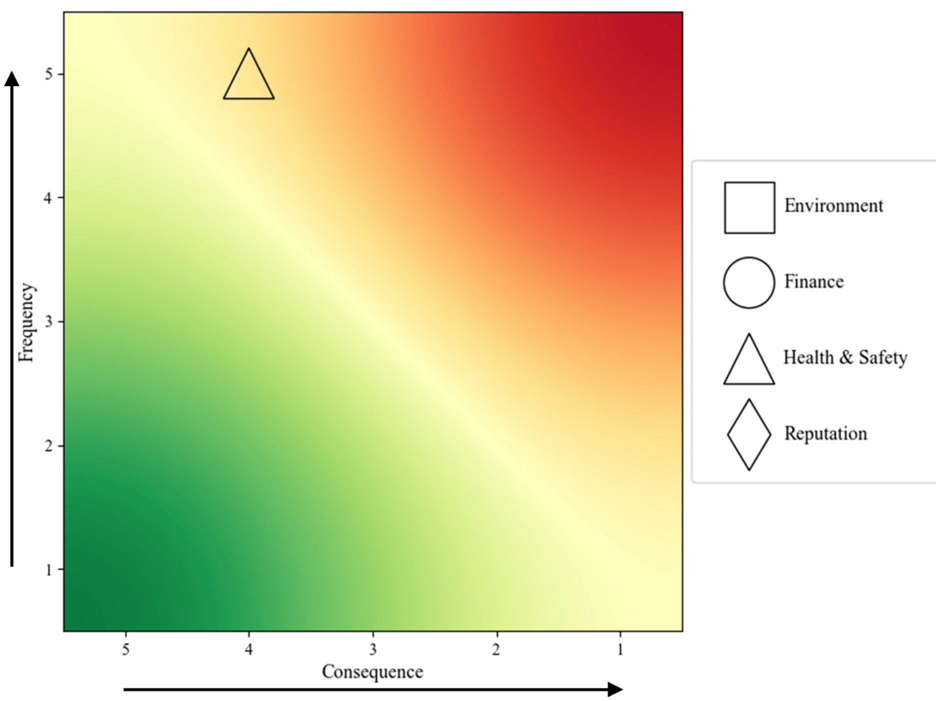

Fig. 8. Risk matrix for Case Study 3.

weather-appropriate footwear outdoors, and (3) ensure that proper authorities are notified to apply salt/gravel to regular walkways and parking lots. The leading indicators are identified as: (1) snow and/or ice and (2) lack of salt/gravel.

As demonstrated by the case studies, applying such methods to incident reporting makes it possible to improve safety in a company at a foundational level. Having access to such information can allow companies to enhance their safety culture by providing timely prevention and mitigation strategies; giving feedback on safety performance versus historical trends; building a reputation with their employees for protecting occupational safety, process safety, and the environment; and reducing financial losses by focusing on higher priority risks that require attention.

4. Summary and discussion

We began with the observation that AI/ML hold great promise for enhancing companies' safety management systems. Yet, the adoption of AI/ML tools has been slow given an overwhelming quantity of dirty, incomplete, and fragmented data; a multitude of low-consequence incidents of unknown analytical value; and operators' resulting perception that this is a high-effort, low-payoff venture (Ransbotham et al., 2017). We have addressed these barriers by aggregated data across operators and using keyword analysis to create more complete, 'consistent', and larger datasets to enhance algorithm training, avoid overfitting, and more sensitive 'detectability' of leading indicators (Kurian et al., 2020).

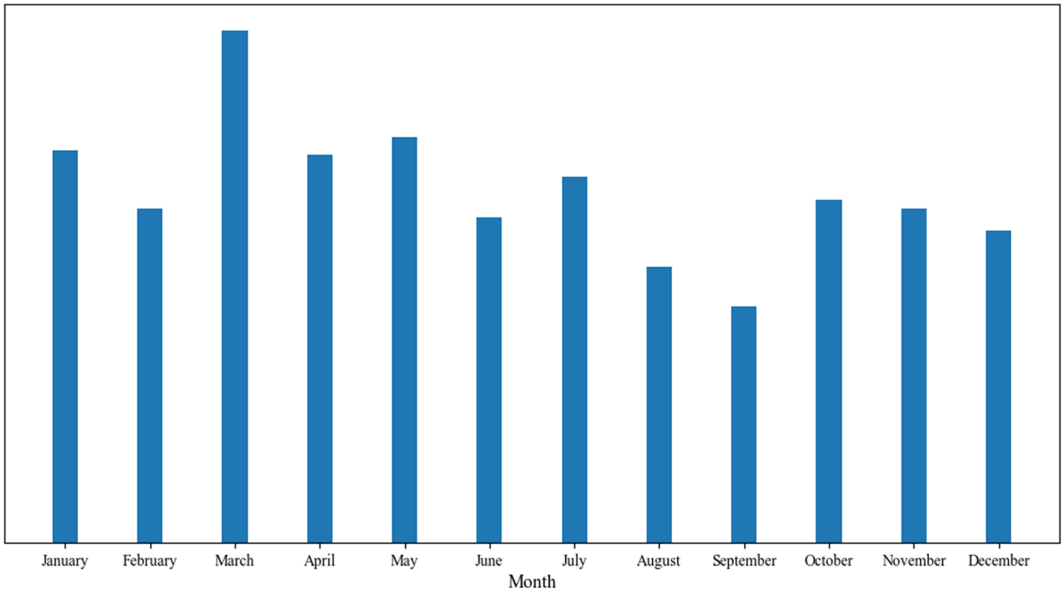

Fig. 9. Example of trend analysis for Case Study 3 indicating number of incidents per month.

Our objectives for expanding on this research was to improve current incident reporting systems, provide practical and tailored outputs to prevent and mitigate risk, and to create opportunities to reduce losses due to incidents. By implementing a system that incorporates machine learning and keyword analysis with an intermediate step for user input, it was possible to accomplish these goals. With the methodology used in this research, we analyzed incident reports and generated a framework for evaluating and reducing risk.

The analysis of incident reports used a supervised machine learning algorithm to predict identifying labels incidents for incidents. Next, the spaCy library from Python was used to lemmatize the incident reports. The resulting words were tallied and the most common words, along with the identifying labels used for machine learning analysis, were used to generate a customized library. The words stored in the customized library were assigned to statements used to accurately describe risk involved in incidents. When a new incident report is inputted, this system runs the machine learning algorithm to predict an identifying label for the incident report, based on the newly expanded text corpora. The incident report is then lemmatized and cross referenced with the words stored in the customized library. Statements corresponding to the words in the customized library are then provided to the user. When the user selects statements matching the incident that occurred, a series of output results are provided. These outputs include a risk matrix, trends of similar incidents within the past year, suggested prevention and mitigation strategies, and any leading indicators that could be identified to prevent future occurrences of similar events. In this manner, the system is constantly learning from newly inputted incident data. In this manner, the system is constantly learning from newly inputted incident data. Likewise, experts can examine trends and revisit suggested prevention and mitigation strategies, to continually refine these.

Three case studies of incident report inputs were analyzed using the proposed methodology. These incidents included quality assurance, vehicular, and slip/trip/fall -type incidents. By analyzing the trends in these incidents, it was surprising to see the low number of quality assurance-type incidents in January (in comparison to the rest of the year). As expected, the highest number of vehicular and slip/trip/fall incidents occurred during the winter months. By analyzing the factors resulting in the trends of such incidents, it is possible for companies to develop plans for future incidents. For example, a company could attempt to identify the reason for the low number of quality assurance incidents in January and attempt to reproduce these results for the remainder of the year. Additional effort could also be focused on reducing vehicular and slip/trip/fall incidents to protect worker safety.

5. Conclusion and future works

This research proposed new methods to report and analyze incident reports using artificial intelligence, machine learning, and keyword analysis. The reports generated by implementing this methodology can allow a company to better focus its efforts on preventing those incidents that are causing the greatest losses and to identify strengths within their existing systems. The method described in this paper directly contributes to reporting and analyzing incidents, as well as providing outputs consistently that can be used to prevent/mitigate risk, i.e. risk matrix, trend report, prevention/mitigation strategies, leading indicators. The outputs of this research can be further used as inputs for other established practices, e.g. bowtie analysis, root cause analysis, fault tree analysis, HAZOP (hazard and operability) studies.

Furthermore, the methodology described in this research can be applied by any industry seeking to reduce risk using incident reports. There is much potential for future use and implementation – many of the variables used in this study can be easily modified to match the varying needs of different companies. The machine learning approach can be enhanced by tuning parameters with additional algorithms (e.g., Whale Optimization, Cuckoo Search, Particle Swarm, etc.) which are expected to substantially improve predictions (Dizangian et al., 2017)).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Albery, S., Borys, D., Tepe, S., 2016. Advantages for risk assessment: Evaluating learnings from question sets inspired by the FRAM and the risk matrix in a manufacturing environment. Safety Sci. 89, 180–189. https://doi.org/10.1016/j.ssci.2016.06.005.

Aven, T., 2011. On some recent definitions and analysis frameworks for risk, vulnerability, and resilience. Risk Anal. An Int. J. 31 (4), 515–522.

Bengio, Y., Goodfellow, I., Courville, A., 2017. Deep learning. MIT press.

Dong, C., Dong, X., Gehman, J., Lefsrud, L., 2017. Using BP neural networks to prioritize risk management approaches for China's unconventional shale gas industry. Sustainability 9 (6), 979.

Dizangian, Babak & Hooshyari, Ali. (2017). Comparing the particle swarm, whale, water cycle, and cuckoo search algorithms in optimization of unconstrained problems.

Duijm, N.J., 2015. Recommendations on the use and design of risk matrices. Saf. Sci. 76, 21–31. https://doi.org/10.1016/j.ssci.2015.02.014.
Garreta, R., Hauck, T., Hackeling, G., 2017. Scikit-learn: machine learning simplified. Packt Publishing, Birmingham, UK.
Goerlandt, F., Reniers, G., 2016. On the assessment of uncertainty in risk diagrams. Saf. Sci. 84, 67–77. https://doi.org/10.1016/j.ssci.2015.12.001.
Government of Alberta, 2019. Reporting and investigating injuries and incidents – OHS information for employers, prime contractors and workers.
Imani, A., Forman, J.E., Amir, W., 2018. A Clustering Analysis of Codes of Conduct and Ethics in the Practice. of Chemistry.
International Organization for Standardization [ISO], 2018. Risk Management - Guidelines (ISO 31000:2018E).
Kellogg, K.C., Valentine, M.A., Christin, A., 2020. Algorithms at work: The new contested terrain of control. Academy Manage. Ann. 14 (1), 366–410.
Kotsiantis, S. B., 2007. Supervised machine learning: a review of classification techniques. https://doi.org/10.1115/1.1559160.
Khakzad, N., Khan, F., Amyotte, P., 2013. Dynamic safety analysis of process systems by mapping bow-tie into Bayesian network. Process Saf. Environ. Prot. 91 (1–2), 46–53.
Kurian, D., Ma, Y., Lefsrud, L., Sattari, F., 2020. Seeing the Forest and the Trees: Using Machine Learning to Categorize and Analyze Incident Reports for Alberta Oil Sands Operators. J. Loss Prev. Process Ind. 64, 104069. https://doi.org/10.1016/j.jlp.2020.104069.
Laberge, M., Maceachen, E., Calvet, B., 2014. Why are occupational health and safety training approaches not effective? Understanding young worker learning processes using an ergonomic lens. Saf. Sci. 68, 250–257. https://doi.org/10.1016/j.ssci.2014.04.012.
Malhotra, A., 2018. Introduction to Libraries of NLP in Python - NLTK vs. spaCy. Retrieved from https://medium.com/@akankshamalhotra24/introduction-to-libraries-of-nlp-in-python-nltk-vs-spacy-42d7b2f128f2.
Milgram, J., Cheriet, M., & Sabourin, R., 2006. "One Against One" or "One Against All": Which One is Better for Handwriting Recognition with SVMs? Tenth International Workshop on Frontiers in Handwriting Recognition, Université de Rennes 1, Oct 2006, La Baule (France). inria00103955. Retrieved from https://hal.inria.fr/inria-00103955.
Muhlbauer, W. K. (2004). Pipeline Risk Management Manual - Ideas, Techniques, and Resources (Third Edit). 200 Wheeler Road, Burlington, MA, USA: Gulf Professional Publishing (an imprint of Elsevier).
Ng, A. (n.d.). Machine Learning. Retrieved from https://www.coursera.org/learn/machine-learning/.
Ni, H., Chen, A., Chen, N., 2010. Some extensions on risk matrix approach. Saf. Sci. 48 (10), 1269–1278. https://doi.org/10.1016/j.ssci.2010.04.005.
Nordlöf, H., Wiitavaara, B., Winblad, U., Wijk, K., Westerling, R., 2015. Safety culture and reasons for risk-taking at a large steel-manufacturing company: Investigating the worker perspective. Saf. Sci. 73, 126–135. https://doi.org/10.1016/j.ssci.2014.11.020.
Pasquini, A., Pozzi, S., Save, L., Sujan, M.-A., 2017. Requisites for Successful Incident Reporting in Resilient Organisations. Resilience Eng. Pract. 237–256. https://doi.org/10.1201/9781317065265-17.
Patriarca, R., Bergström, J., Gravio, G.D., Costantino, F., 2018. Resilience engineering: Current status of the research and future challenges. Saf. Sci. 102, 79–100. https://doi.org/10.1016/j.ssci.2017.10.005.
Project Management Institute. (2004). A guide to the project management body of knowledge (PMBOK guide). Newtown Square, Pa: Project Management Institute.
Ransbotham, S., Kiron, D., Gerbert, P., Reeves, M., 2017. Reshaping business with artificial intelligence: Closing the gap between ambition and action. MIT Sloan Manage. Rev. 59 (1).
Singh, A., Thakur, N., & Sharma, A., 2016. A review of supervised machine learning algorithms.
Srinivasa-Desikan, B., 2018. Natural language processing and computational linguistics a practical guide to text analysis with Python, Gensim, spaCy, and Keras. Birmingham: Packt.
Steen, R., Aven, T., 2011. A risk perspective suitable for resilience engineering. Saf. Sci. 49 (2), 292–297.
Thomas, P., Bratvold, R. B., & Bickel, J. E., 2013. The Risk of Using Risk Matrices Decision Analysis View project A Generalized Sampling Approach for Multilinear Utility Functions Given Partial Preference Information View project The Risk of Using Risk Matrices. Article in SPE Economics and Management, (April 2015). https://doi.org/10.2118/166269-MS.
Xu, Q., Xu, K., Li, L., Yao, X., 2018. Safety assessment of petrochemical enterprise using the cloud model, PHA–LOPA and the bow-tie model. R. Soc. Open Sci. 5 (7), 180212.
Zhao, J., Cui, L., Zhao, L., Qiu, T., Chen, B., 2009. Learning HAZOP expert system by case-based reasoning and ontology. Comput. Chem. Eng. 33 (1), 371–378.
Zuboff, S., 2015. Big other: surveillance capitalism and the prospects of an information civilization. J. Info. Technol. 30 (1), 75–89.