24-009

An Integrated Method to Classify Ground-Fall Accidents and to Estimate Ground-Fall Trends in U.S. Mines Using Machine Learning Algorithms

Gamal Rashed
CDC/NIOSH/PMRD

Yuting Xue
CDC/NIOSH/PMRD

Connor Brown
CDC/NIOSH/PMRD

Zoheir Khademian
CDC/NIOSH/PMRD

Khaled Mohamed
CDC/NIOSH/PMRD

ABSTRACT

Ground falls in U.S. underground coal mines can lead to significant consequences, including loss of life, injuries, damaged equipment, and production stoppage. Improving the safety of the workplace is of utmost importance for mine workers and the U.S. economy. The Mine Safety and Health Administration (MSHA) accident/injury/illness dataset provides short narratives for reported incidents, including ground-falls. The main objective of this study is to develop a framework that includes: 1) utilizing machine learning algorithms to categorize ground-fall incidents from narratives based on the main cause of the occurrence and 2) demonstrating an example of a user-friendly visualization to display injury/fatality trends from narratives in U.S. coal mines between 1983 and 2021. The developed framework was tested on a subset of the data and achieved an average F1-score of 96% in categorizing the incidents. The outcome will help identify areas requiring additional research and innovative solutions to reduce severe occupational hazards.

INTRODUCTION

Accidents due to ground-fall failures in coal mines can potentially have severe consequences, including both fatal and non-fatal injuries, damage of equipment, impaired ventilation, and production delay/stoppage. Improving the safety of the workplace in U.S. coal mines is of utmost importance for mine workers, mine operators, and the U.S. economy. Between 2010 and 2019, the ground-fall incidents in U.S. mines resulted in 46 fatalities, 33 permanent disabilities, 3,082 injuries, 119,520 non-fatal days lost, and 12,433 days of restricted work activities (Rashed et al. 2022). In MSHA dataset, the accident/injury/illness datafile combines five different types of ground fall accidents (roof fall, rib fall, face fall, rock outburst, and highwall failure) into two categories called "fall of roof, back, or brow from in-place" and "fall of face, rib, pillar, side, or highwall." The MSHA dataset provides short narratives for ground-fall incidents; however, it does not classify them based on the root cause of the incident. MSHA fatality reports include the root cause of incidents. However, they are not considered in this study. From a prevention perspective, ascertaining root causes of incidents, often provided in the narratives, is necessary for the research to address mitigation strategies and identifying potential gaps in scientific research. Note that accidents and incidents are used interchangeably in this study.

Rashed et al. 2022 manually classified thousands of ground-fall narratives into five categories using a subset of the MSHA dataset. However, this process was time-consuming and tedious, which is why in this study the authors raised a question about the capability of utilizing machine learning algorithms to perform the text classification task. Machine learning is a branch of artificial intelligence that can be used to perform tasks automatically based on training data without explicit instructions. In this study, text classification models using supervised machine learning are used to categorize text into organized groups. With supervised learning, the goal is to learn the mapping function from input to output given many examples in the form of input-output pairs. The input-output pairs are known as the training data, and the outputs are specifically known as labels or ground truth. Machine learning models that can automatically apply labels for classification are known as classifiers. In this paper, various machine learning models were trained on manually classified ground-fall data from 2010 to 2019. The size of the trained dataset is about 10% of the entire MSHA dataset between 1983 and 2021. The best model was selected to predict the ground-fall classification for the entire MSHA dataset. Additionally, an interactive/dynamic Dashboard was built to display ground-fall trends in U.S. coal mines between 1983–2021. Several measures of potential ground-fall risk are included in the Dashboard, such as the number of injuries, fatalities, and lost workdays as an index to show the relative level of risk among different operations and conditions in U.S. coal mines. These metrics can be used to determine both the problem areas and the progress in preventing work-related injuries.

MSHA DATABASE

The dataset used in this study is the accident/injury/illness data which can be found and/or downloaded from CDC website (NIOSH mining data, 2023). For a detailed definition of a mine accident, see 30 CFR § 50.2 (MSHA, 1986). The accident/injury/illness data includes 60 variables (column names) and 711,960 reported events/incidents (rows) between 1983 and 2021. Under Part 50 of the U.S. Code of Federal Regulations, mine operators and independent contractors are required to file MSHA Form 7000-1 for reportable incidents within 10 working days after the accident or injury, or 10 working days following the illness diagnosis (NIOSH, 2016). MSHA defines "reportable" as accidents, occupational injuries, or occupational illnesses including all incidents that require medical treatment or result in death, loss of consciousness, inability to perform all job duties on any workday following the injury, or temporary assignment or transfer to another job. By regulation, certain accidents and injuries are immediately reportable to MSHA withing 15 minutes of their occurrence (NIOSH, 2016).

Among the immediately reportable incidents are a death or an injury of an individual that has a potential to cause death, an unplanned roof or rib fall in active working zone that impairs ventilation, and an unplanned roof fall at or above anchorage horizon in active workings where roof bolts are in use (MSHA, 1986). MSHA classifies the reported incidents into twenty-eight different categories. Figure 1 shows the classification and the number of reported incidents in U.S. coal and metal/nonmetal mines, both surface and underground between 1983 and 2021.

GROUND-FALL INCIDENTS AND NARRATIVES

The ground-fall-related incidents can be classified into two groups: the first group is fall of roof, back, or brow (from in place), while the second group is fall of face, rib, pillar, side, or highwall (from in place). The category called "falling, rolling, or sliding rock or material of any kind" is not considered a ground-fall category, which is why it is not included in the analysis. About 11% of the reported incidents to MSHA occurred due to ground falls; see Figure 1. In this study, the authors focused only on the ground-fall incidents that occurred in coal mines and excluded those from metal and nonmetal mines because most of the ground-fall incidents occurred in coal mines (Rashed et al. 2022). Classifying ground-fall incidents based on the main cause and determining injuries and fatalities associated with each category can help identify areas where additional research is needed and where innovative solutions need to be developed to reduce these potential occupational hazards. Table 1 shows examples of some ground-fall narratives and the classifications associated with them. It would be time consuming to investigate every narrative associated with every reported incident, which is why machine learning models were explored in this study to conduct text classification for ground-fall narratives.

Ground-fall narratives in the MSHA dataset are considered to be unstructured data, meaning they do not have a predefined format or organization that makes it more difficult sometimes to be processed by a human or machine learning algorithm. Additionally, some ground-fall narratives are unclear and even human-based classification is difficult. Table 2 shows examples of these unclear ground-fall narratives. The narratives are shown as they exist in the MSHA dataset without modifying or editing. It is recommended that future narratives follow a certain structure or a template and use key phrases to distinguish between the groups.

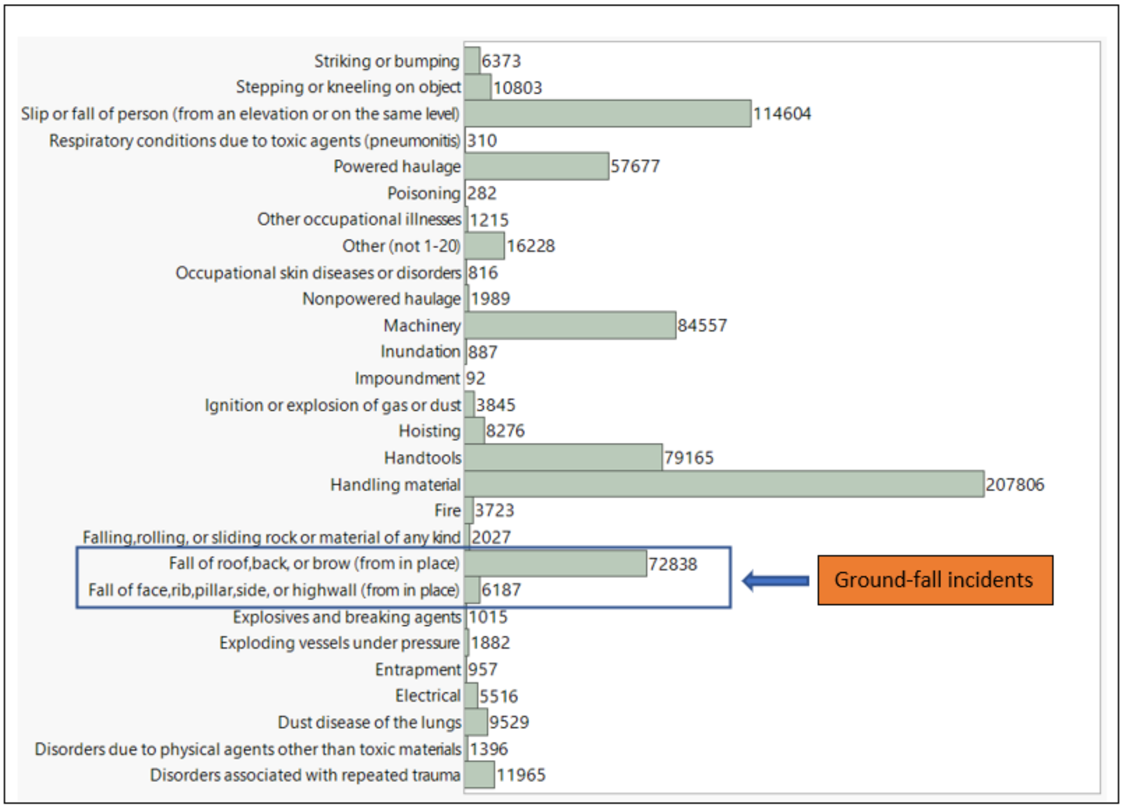


Figure 1. MSHA classification for reported accidents between 1983–2021 for surface and underground U.S. mines; ground-fall incidents are contained in the blue box

Table 1. Examples of ground-fall narratives and their classification

Ground-fall narrative	Classification
"Employee was roof bolting in the face area when a small piece of rock fell from the roof hitting employee in the left forearm. reportable only due to stitches to left forearm."	Roof fall
"Employee was shoveling the ribs in the face as part of the clean-up cycle, when a piece of rock sloughed off the rib near the roof and struck the employee in the right ankle."	Rib fall
"A rock burst occurred in the main drift while employees were constructing a tunnel liner, seven miners were injured."	Rock outburst
"Employee using excavator to clean highwall and boulder hit cab shattering right window getting glass in right eye."	Highwall failure

Table 2. Examples of unclear ground-fall narratives in MSHA dataset

"A piece of rock struck him on the lower part of his left leg, which resulted in a fracture of the small bone in that part of the leg."
"Having advanced a slider tube with co-worker, he was struck on the right foot by a piece of rock measuring 2 ft x 3 ft x 6 in thick. this resulted in fracture of three areas of his right metat tarsal on the right foot. injured began losing time on 3/8/10."
"Injured employee was struck by a rock measuring 20"l x10"w x 3" thick causing a contusion to the left side of his face."
"An unplanned ground fall occurred at 3156, north 148, panel c107 in the underground mine."
"Fall was found at 1 west mains intake #8 entry. fall was 40' l x 19' w x 5' thick.8 left dimensions were 25' l x 19'w x 5' thick."

TEXT CLASSIFICATION USING MACHINE LEARNING

Text classification is a task in natural language processing (NLP) that involves automatically assigning predefined categories or labels to text documents. In this study, the text documents are the ground-fall narratives while the predefined categories are the five ground-fall categories (roof fall, rib fall, face fall, highwall failure, and outburst). The supervised machine learning model to conduct text classification typically involves five main steps shown in Figure 3. Several Python packages were used to implement the machine learning model for text classification, such as Pandas, Numpy, Sklearn (Pedregosa et al. 2011).

PREPROCESSING THE MSHA DATA

To improve the performance of the machine learning models and to make the models more interpretable and robust, it is important to preprocess the raw data before feeding it to the machine learning models. Preprocessing the data includes performing multiple tasks, such as data cleaning, data reduction, and data transformation. Before implementing the machine learning models on MSHA data, a preprocessing step was conducted; for instance, the MSHA dataset was filtered to obtain only ground-fall incidents, see Figure 1, such that the other reported incidents were excluded from the analysis; the variable names in the MSHA dataset were shortened to facilitate working with the data using the Pandas library; Handling the null values was conducted, for example a few narratives were empty, they were excluded from the analysis. The data types of all variables were checked to avoid unexpected errors when conducting mathematical or statistical operations. The data types for some variables were transformed to facilitate working with the machine learning models. For example, the data type of the manually classified narratives was transformed from categorical data type to numerical data type, because machine learning models use mathematical equations and categorical data is not accepted. As shown in Figure 2 the length of the ground-fall narratives varies with a mean of 40 words and a maximum of 92 words. Hence, manually processing the MSHA narratives to extract some information or to conduct a classification would be

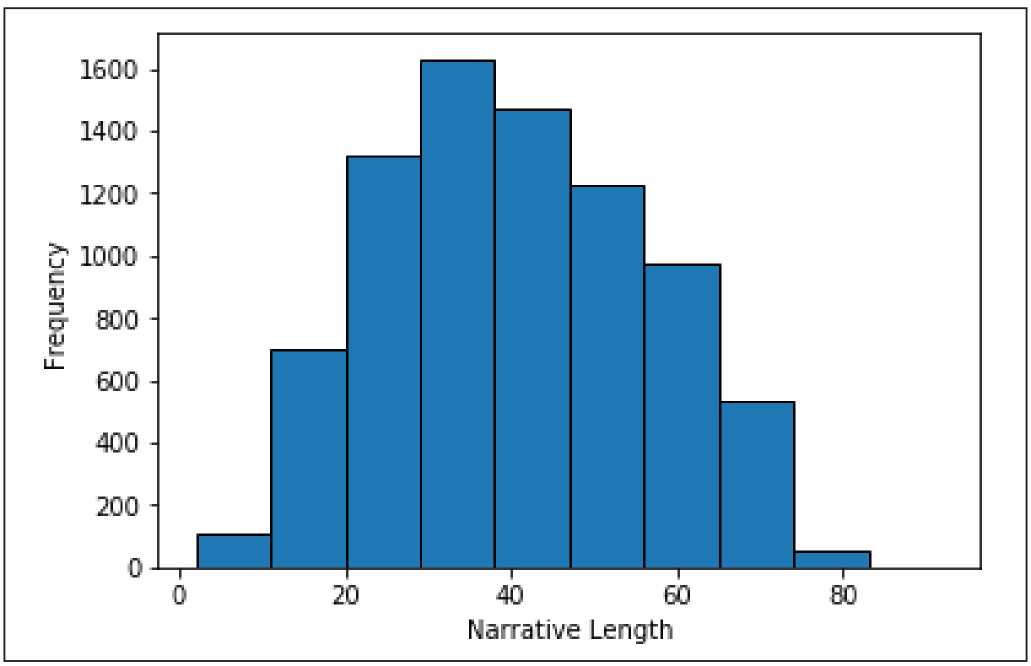


Figure 2. Ground-fall narrative length (number of words) in MSHA dataset between 2010 and 2019

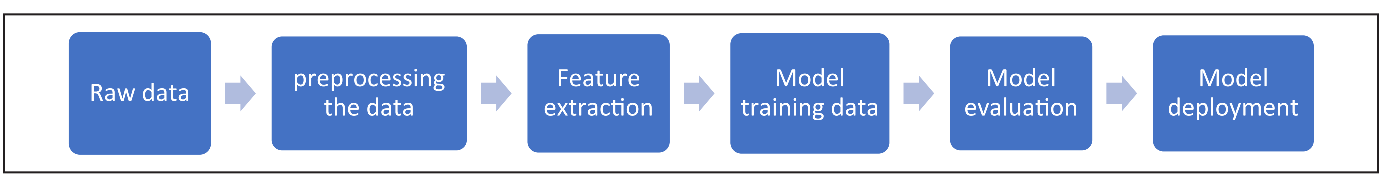


Figure 3. Steps associated with supervised learning for text classification

time-consuming especially if thousands of these narratives will be processed. Machine learning models are heavily used in many engineering disciplines and applications. In this study, the authors explored the capability of machine learning to handle large data and to conduct a supervised machine learning to categorize ground-fall narratives.

FEATURE EXTRACTION

During feature extraction, the ground-fall narratives are transformed into a numerical format that machine learning algorithms can process. This step is commonly known as text vectorization. A common method to vectorize the text (ground-fall narratives) is the Term Frequency-Inverse Document Frequency (TF-IDF). Higher TF-IDF indicates that a term is more discriminative for a particular document. The TF measures the frequency of a term in a document, representing how often the term appears in a document relative to the total number of words in that document. The IDF measures the uniqueness of a term across a collection of documents by penalizing terms that occur frequently in the entire document collection and assigning a higher weight to terms that appear in a smaller number of documents. Terms that appear in many documents have a lower IDF score, while terms that appear in few documents have a higher IDF score.

MODEL TRAINING DATA

The authors manually classified 8,013 ground-fall incidents into five categories (Rashed et al. 2022). Figure 4 shows the number of ground-fall incidents associated with each ground-fall category, these classified incidents have been used to train and test the machine learning models. The 8,013 ground-fall events were split into a training dataset with 70% of the data (5,609 cases) and a testing dataset with the remaining 30% data (2,404 cases). The TfidfVectorizer in the sklearn library was used to vectorize the ground-fall narratives in the training and testing dataset separately. The range of n-value for different n-grams to be extracted using the TfidfVectorizer is (1, 2), which means that single and double words were extracted from the narratives. The maximum number of features used in this study was 2,000, the maximum feature parameter controls the dimensions of the TF-IDF matrix and directly affects the memory and computational requirements of the algorithm.

Figure 4 shows that the dataset is imbalanced; an imbalanced dataset refers to a situation where the distribution of the target classes is not uniform, resulting in one or more classes having significantly fewer samples than others. The roof fall cases are the dominant category in the dataset. Imbalanced data would potentially impact the performance of the machine learning model, such that the model would be biased toward the dominant category because it is more common in the dataset; consequently, the machine learning model would perform poorly on minority classes due to their limited representation in the dataset. Since there are very limited cases for the category of "face," "highwall," "outburst," and 'rib fall" in the dataset, oversampling these minority classes was used to balance the class distribution. The RandomOverSampler within the Imbalanced-Learn library was used for oversampling the minority classes in this study (Lemaˆıtre et al., 2017). The next step is to train a classifier using the labeled training data and the extracted

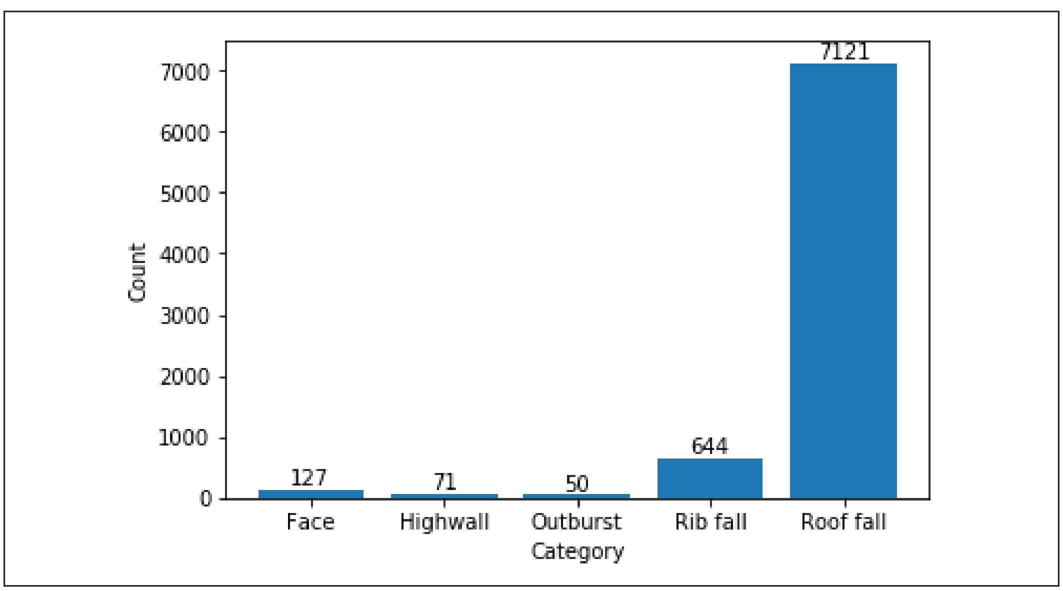


Figure 4. Ground-fall classifications associated in the 2010-2019 dataset

features. During the training process, the method of grid search with cross-validation was used to tune the model hyper-parameters. There are various hyper-parameters for each machine learning algorithm, and they control the training process and affect the performance. The grid search cross-validation method allows the search of the best hyper-parameters by exploring different combinations of the provided hyper-parameter values. The cross-validation method involves randomly splitting the training dataset into K groups (5, by default). For each iteration, one group of data is used as the validation dataset, while the remaining groups are taken as the training dataset. The classifier with one group of hyper-parameters is fitted with the training data and is evaluated with the testing data to obtain an evaluation score. After K iterations, the average evaluation score is used to quantify the performance of the hyper-parameter combination. After comparing the average evaluation score for all the hyper-parameter combinations, the best hyper-parameters can be selected and the classifier with the best performance can be used as the trained classifier.

PERFORMANCE EVALUATION

The performance of the trained model is assessed with the testing data using evaluation metrics. Table 3 summarizes multiple methods that can be used to evaluate the performance of machine learning models, such as the confusion matrix, precision, recall, and F1-score. A confusion matrix for binary classification provides a summary of the predicted and the true labels of the narratives, showing the counts of true positives, true negatives, false positives, and false negatives. The F1-score combines precision and recall into a single value, providing a balanced measure of a model's performance. The F1-score is particularly useful when dealing with imbalanced datasets.

RESULTS AND DISCUSSION

The training dataset was used to train various machine learning models and the testing dataset was used to evaluate the performance of these models. Figure 5 shows the results of the F1-score for some of the machine learning models used in this study to classify ground-fall narratives into five categories. The model performance and confusion matrix for only three models-Multinomial Naïve Bayes, Random Forest, and Logistic Regression-are compared in Table 4. The LabelEncorder within the scikit-learn library was used to encode the five ground-fall categories in a numerical format (Pedregosa et al. 2011). After encoding, 'roof fall' is represented with '4', '3' for 'rib fall', '2' for 'outburst', '1' for 'highwall', and '0' for 'face'. The leading diagonal of the confusion matrix includes the correct prediction for the ground-fall narratives. Generally, the model performance is better as the count on the leading diagonal increases and the off-diagonal decreases.

Table 4 shows that the logistic regression model has the highest F1-scores for each category and overall. In general the higher the F1-score the better the model. As shown in Table 4, the logistic regression model successfully classified 98% of the roof-fall incidents, 85% of the rib-fall incidents, 73% of the rock-outburst incidents, 78% of the highwall-failure incidents, and 72% of the face-fall incidents. According to Figure 4, about 88% of the incidents were manually classified as "roof fall." Hence, if no machine learning model was used and the null hypothesis was assumed such that all ground-fall incidents were classified as "roof fall," the success rate would be 100% for roof

Table 3. Performance evaluation for a machine learning model

<table>
<tr><td>Confusion matrix</td><td>
<table>
<tr><td></td><td></td><td colspan="2">Predicted condition</td></tr>
<tr><td></td><td>Total population
= P + N</td><td>Positive (PP)</td><td>Negative (PN)</td></tr>
<tr><td rowspan="2">Actual condition</td><td>Positive (P)</td><td>True positive (TP)</td><td>False negative (FN)</td></tr>
<tr><td>Negative (N)</td><td>False positive (FP)</td><td>True negative (TN)</td></tr>
</table>
</td></tr>
<tr><td>Precision</td><td>TP/(TP+FP)</td></tr>
<tr><td>Recall</td><td>TP/(TP+FN)</td></tr>
<tr><td>F1_Score</td><td>2*(Precision*Recall)/(Precision + Recall)</td></tr>
</table>

fall categories and zero for all other categories which is not accepted. Therefore, text classification using the supervised machine learning model is a promising technique in analyzing large data in mining application. The authors considered only one predictor to classify the ground-fall incidents, better results could be obtained if more predictors were considered in the analysis. Additionally, the ground-fall narratives were extracted from incident reports and some words were misspelled which could affect the performance of the machine learning models, future work would explore the effect of misspelled words on the model performance and would also consider more than one predictor in the

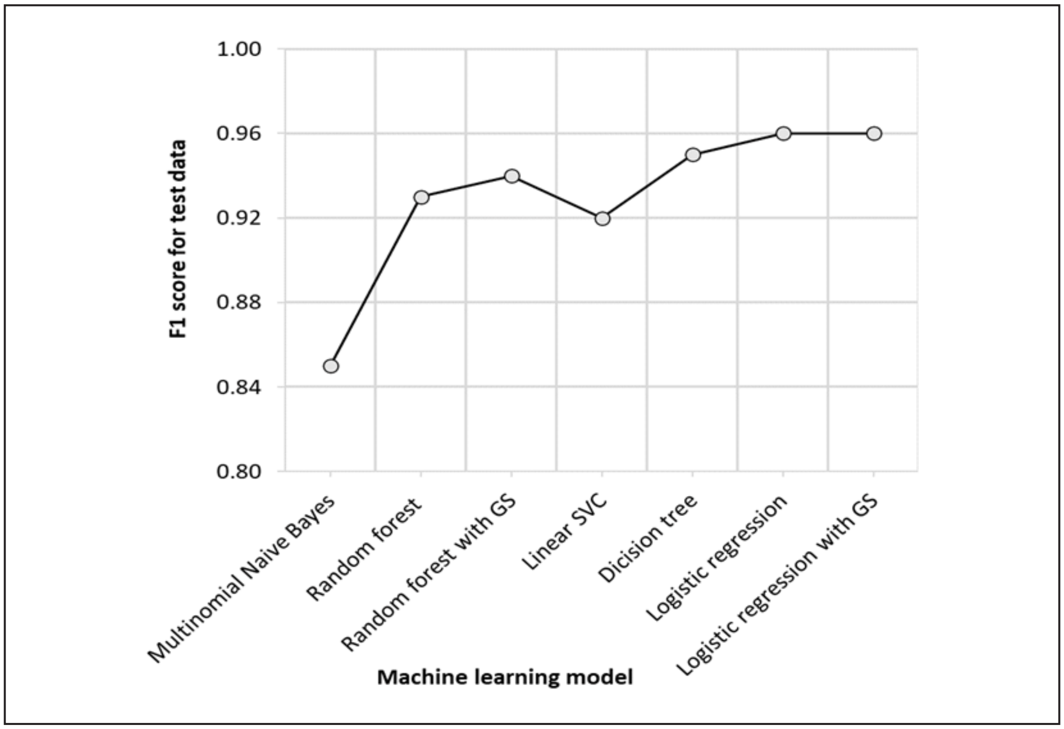
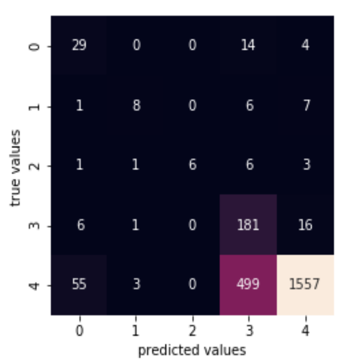
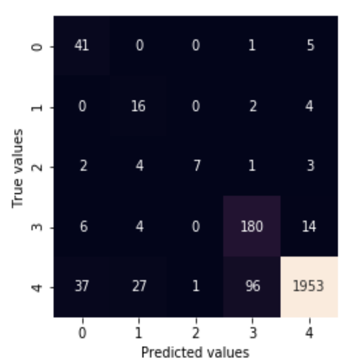
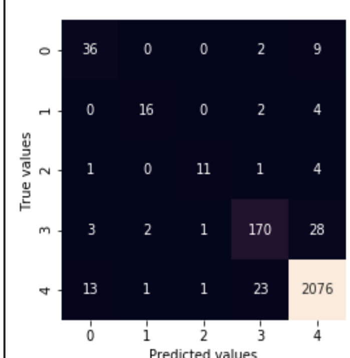


Figure 5. F1-score for various machine learning models to classify ground-fall narratives

Table 4. Model performance for three different machine learning algorithms

	Multinomial Naïve Bayes	**Random forest**	**Logistic regression**
Confusion matrix			
F1_Score for 0	0.42	0.62	0.72
F1_Score for 1	0.46	0.44	0.78
F1_Score for 2	0.52	0.56	0.73
F1_Score for 3	0.40	0.74	0.85
F1_Score for 4	0.84	0.95	0.98
Overall F1_Score	0.74	0.91	0.96

imports (boilerplate)

```
import dash
import dash_html_components as html
import dash_core_components as dcc
```

app instantiation

```
app = dash.Dash(__name__)
```

app layout: a list of HTML and/or interactive components

```
app.layout = html.Div([
    dcc.Dropdown()
    dcc.Graph()
    ...
])
```

callback functions

```
@app.callback()
    ...
@app.callback()
    ...
```

running the app

```
if __name__ == '__main__':
    app.run_server()
```

Figure 6. The general structure of a Dash app to generate a Dashboard

analysis. The model with the best performance (the logistic regression model) was selected such that it can be deployed to classify new and unseen text data (narratives) in the MSHA dataset. The logistic regression model was utilized to classify/predict the ground-fall incidents in the entire MSHA dataset between 1983 and 2021. The result of the logistic regression model is shown in the Dashboard section in Figure 9.

The authors used the general outlines in Figure 6 to demonstrate visualization of ground-fall trends in U.S. coal mines between 1983 and 2021, named in this paper as the Dashboard. The Plotly Dash using Python was used to generate an interactive and dynamic Dashboard (Plotly Technologies Inc., 2015). Other programming languages, such as R can be used to achieve the same goal. Figure 6 shows the general structure of a Dash app. In the imports part, the required packages are imported. The app instantiation is a straightforward way to create the app. The app layout is used to lay out various containers, dropdown menus, and figures. The callback functions link different elements of the layout and make the Dashboard interactive and dynamic.

A benefit example of the Dashboard is the potential to identify areas where additional research is needed and where innovative solutions may need to be developed to reduce these potentially severe occupational hazards. Hence, the Dashboard would be a useful tool to identify ground control related health and safety gaps to areas that experience more injuries/fatalities. Additionally, the use of these tools may complement surveillance statistics efforts to track trends such as reduction in ground fall incidents in U.S. coal mines.

Figure 7 shows the general layout for the prototype Dashboard that was developed for internal use only. This Dashboard, developed to identify ground fall trends in the U.S. coal mines, is composed of four tabs: Project Info, Mining Operations, Mining Methods, and Ground-fall Classification. The Project Info tab explains the goal and the outcome of the Dashboard. It also shows three examples of ground-fall incidents (roof fall, rib fall, and outburst); only one example is shown in Figure 7.

The Mining Operations tab is divided into two panels; the left panel includes three dropdown menus and a slider bar to filter the ground-fall incident database. In the first dropdown menu, the user of the Dashboard can select the mining operation which would be one of the following: surface, underground, or both surface and underground. In the second dropdown menu, the user can select the degree of injury due to ground-fall incidents, examples of the degree of injury are: fatal, days away from work only, and no injury. Note that when the degree of injury is "no injury," that means a ground fall incident occurred without injury such that the rock fall blocked the ventilation or was above the anchorage horizon, which is why it had to be reported to MSHA using the mine the MSHA Form 7000-1 (MSHA, 1986). In the third dropdown menu, the user can filter the ground-fall incidents based on the state, and the slider bar is used to filter the data based on the time period of interest. The Dashboard is dynamic and interactive such that the plots in the right panel of Figure 8 would

Figure 7. The layout of the developed Dashboard for ground-fall trends in U.S. coal mines

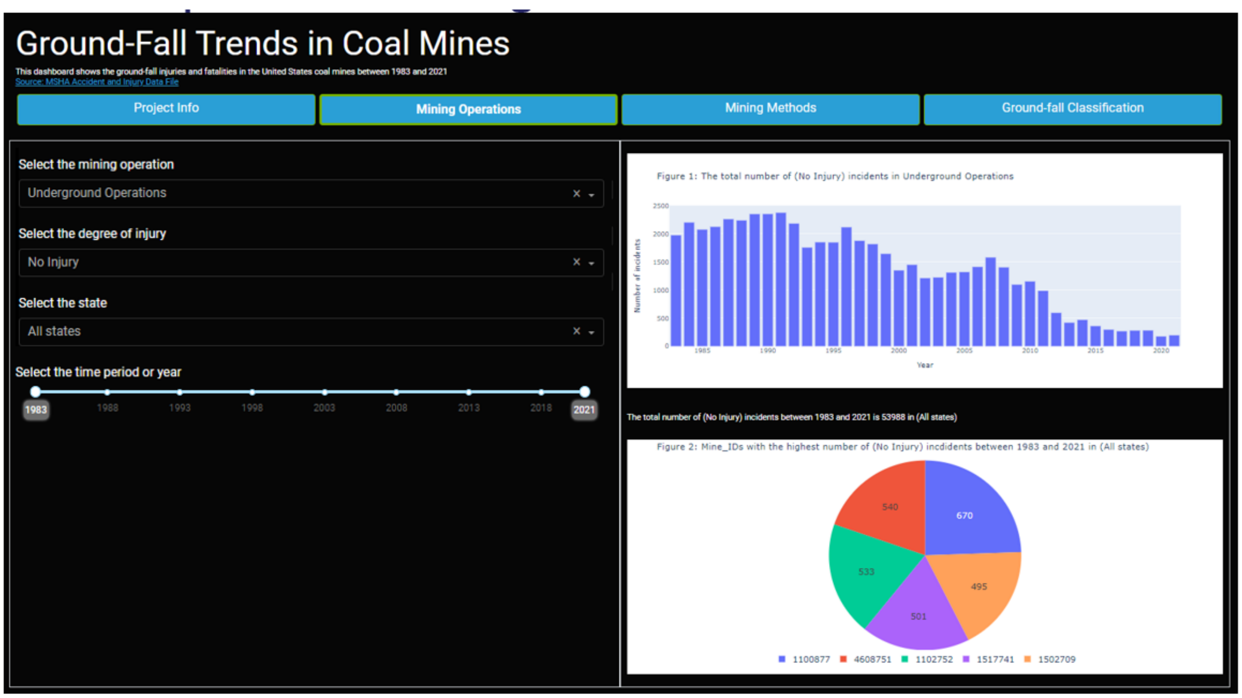


Figure 8. The mining operation tab in the ground-fall Dashboard

change directly based on the selected input. The bar chart shows the number of ground-fall incidents for the selected input in the left panel. The pie chart illustrates the mine ID with the highest number of reportable incidents based on the selection criteria in the left panel. Researchers may begin to further investigate these mine records to determine contributing factors to incident rates such as length of operations, location, and geological & mining conditions In the third tab, Mining Methods, the user would select between room-and-pillar mining or longwall mining methods, and the rest of the dropdown menus are the same with different mining methods.

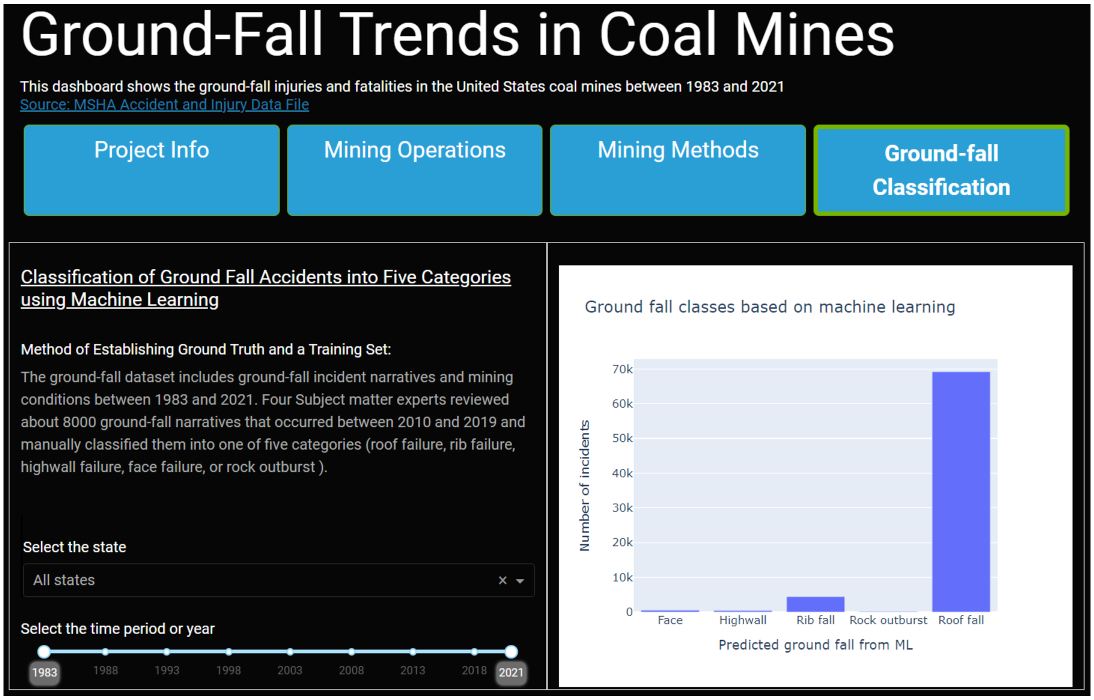


Figure 9. The ground-fall classification tab of the developed Dashboard

The fourth tab, Ground-fall Classification, illustrates the prediction of ground-fall incidents in U.S. coal mines using the machine learning algorithm. The ground-fall classes shown on the right-hand side of Figure 9 were predicted based on the best model which is logistic regression model; the average F1-score for that model was 96% for the testing dataset. The predicted ground falls can be filtered based on the state or the time period as shown on the left-hand side of Figure 9.

CONCLUSIONS

This study provides a new method to analyze and classify massive text dataset in mining application using machine learning models. These machine learning models accurately classified ground-fall-related incidents into five categories with an average F1-score of 96%. The outcome of this effort may enhance research efforts by providing additional tools researchers can leverage and ultimately aim to improve safety in U.S. mines and protect equipment from ground-fall incidents. This study demonstrates that tools can be developed that improves the knowledge gained from evaluating MSHA narratives using machine learning with the ultimate goal being to reduce the amount of time it takes to sort through the narratives.

LIMITATION OF THE STUDY

The work completed in this study was from an exploratory research perspective to evaluate the usefulness of using machine learning being applied to MSHA data. The work is not intended to suggest improvements or changes to data collection methods utilized by MSHA. Nor does the study suggest that immediate ground control improvements and safety considerations can be taken away from the examples presented. The study solely demonstrated that machine learning techniques can be applied to MSHA narratives, specific to ground falls, and shows promise to be successful at conducting classification and efficiently visualizing trends. More work needs to be done to explore the capability of the machine learning related to the other categories not analyzed from Figure 1. At this time, the Dashboard demonstrated in this study is only intended for internal use for visualization purposes. There are some limitations in MSHA dataset, for example not all incidents and injuries are reported to MSHA and this could make the MSHA data biased toward specific categories of ground falls. Also, there is no requirements to complete the narrative section in MSHA Form 7000-1. Hence, some narratives could be “blank” (The authors found four blank narratives out of 8,017 in the trained dataset) or they could be “unclear enough” to be useful in the analysis. The authors gave some examples of unclear narratives in Table 2.

ACKNOWLEDGEMENTS

The authors acknowledge the Data Science Upscaling (DSU) program for all the support and the unlimited resources that the authors received during the DSU. We also are grateful to Dr. Mayer Antoine and Dr. Jeff Colombe for advising us during the training.

DISCLAIMER

The findings and conclusions in this study are those of the authors and do not necessarily represent the official position of the National Institute for Occupational Safety and Health (NIOSH), Centers for Disease Control and Prevention (CDC). Mention of any company or product does not constitute endorsement by NIOSH.

REFERENCES

Lemaître G, Nogueira F, Aridas CK (2017) Imbalanced-learn: a python tollbox to tackle the curse of imbalanced datasets in machine learning. J Mach Learn Res 18:1–5.

Pedregosa F, Varoquaus G, Gramfort A, et al (2011) Scikit-learn: Machine Learning in Python. J Mach Learn Res 12:2825–2830.

Plotly Technologies Inc. (2015) Collaborative data science. plot.ly. Accessed 9 Apr 2023.Rashed G, Xue Y, Khademian Z, Sears M (2022) Ground-Fall Accident Trends in Mining: 2010 to 2019. In: Proceedings of the 2022 International Conference on Ground Control in Mining.

NIOSH mining data (2023): www.cdc.gov/niosh/mining/data/default.html.

NIOSH (2016). The MUG (MSHA data user's guide), Version 5.2. esp.cdc.gov/sites/niosh/DLO/OMSHR/Documents/SST/MUG/MUG.pdf.

MSHA (1986). Report on 30 CFR part 50. U.S. Department of Labor, Mine Safety and Health Administration. PC-7014. December 1986.