IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025
https://doi.org/10.62647/ijitce.2025.v13.i2.pp584-597

ENHANCING RAILWAY STATION SAFETY THROUGH UNSUPERVISED MACHINE LEARNING

[1]kunigiri Ajay Kumar,MCA Student, Department of MCA
[2] H. Ateeq Ahmed, M.Tech, (Ph.D), Assistant Professor, Department of MCA
[12]Dr KV Subba Reddy Institute of Technology, Dupadu, Kurnool

ABSTRACT

Railroad operations must be reliable, accessible, maintained, and safe (RAMS) for both passenger and freight transit. Railway station safety and risk incidents are a major safety concern for day-to-day operations in many metropolitan settings. Additionally, the incidents cause harm to the market's brand in addition to expenses and injuries to individuals. Higher demand is putting pressure on these stations, using up infrastructure and raising safety administration concerns. It is recommended to employ unsupervised topic modelling to better understand the factors that contribute to these extreme incidents in order to analyse them and use technology, such as artificial intelligence techniques, to improve safety. Latent Dirichlet Allocation (LDA) for fatality accidents at railway stations is optimised using textual data collected by RSSB, which includes 1000 incidents in UK railway stations. This study offers advanced analysis and explains how to improve safety and risk management in the stations by applying the machine learning topic technique for systematic spot accident characteristics. Through information mining, lessons learnt, and a thorough understanding of the danger posed by evaluating deaths in accidents on a broad and long-lasting scale, the study assesses the effectiveness of text. Predictive accuracy for important accident data, such the underlying reasons and the hot spots at train stations, is provided by this intelligent text analysis. Additionally, the advancement of big data analytics leads to a better understanding of the nature of accidents than would be feasible with a large safety history or with a restricted domain examination of accident reports. High precision and a new, advantageous era of AI applications in railway sector safety and other safety-related domains are provided by this technology.

1. INTRODUCTION

Compared to other modes of public transit, trains have historically been thought to be safer. However, a number of overlapping issues, including station operation, design, and customer behaviours, might put people at train stations at danger. There are possible hazards when the stations are operating because of the steadily rising demand, the highly crowded society, the layout and design complexity of some stations, and more. Additionally, the railway industry's top priority and one of the system's most important components is passenger, human, and public safety. The Reliability, Availability, Maintainability, and Safety (RAMS) standard, EN 50126, was implemented by the European Union in 1999. aiming to maintain a high standard of

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025
https://doi.org/10.62647/ijitce.2025.v13.i2.pp584-597

safety in railway operations and prevent accidents. The principles of RAMS analysis result in increased safety and acceptable risk reduction. That has been a pressing issue, though, and reports continue to indicate that a number of people are murdered annually in train stations, with some incidents resulting in injuries or fatalities. For instance, In 2016, there were 420 accidents in Japan, including being hit by a train, which claimed 202 lives. Of the 420 incidents, 179 (with 24 fatalities) involved falling off a platform and subsequent injuries or fatalities due to collisions with trains [1]. According to reports, the majority of passenger injuries in the UK in 2019–20 are caused by incidents that happen at stations. The best Major injuries result from slips, trips, and falls, of which there were around 200 [2]. These incidents have a major influence on lowering the number of injuries on station platforms and on providing a high-quality, dependable, and secure travel environment for all passengers, employees, and members of the public. Even in cases when there are no fatalities or serious injuries, accidents can nevertheless create delays, expenses, worry and panic among the public, disruptions in business operations, and harm to the industry's brand. Additionally, it is essential to take into account the risks of both railway incidents and station hazards when providing or investing in any control safety measures. This includes identifying numerous factors that contribute to accidents by having a thorough understanding of the underlying causes of accidents while taking into account all available technology.

In order to introduce a smart approach that is expected to develop the safety level in the future, the risk management process, and the means to collect data in the railway stations, the research aims to analyse a collection case of accidents between 01/01/2000 and 17/04/2020 data. RSSBS has collected this data and consented to use it for research. It is a difficult task to analyse large amounts of data that have been recorded in various formats. These days, it is difficult to find precise information in the vast amount of digital data that includes photographs, videos, Web content, and other sources; it is like trying to find a needle in a haystack. Therefore, there is a real need for a strong tool to help organise, explore, and comprehend these enormous volumes of data [3], [4]. In order to extract useful features from the vast quantity of safety data in the stations, including textual data, several pre-processing methods and algorithms are needed. In order to find useful characteristics, such as the underlying cause of accidents, the study looks at modelling. It also looks at factors, which are collections of words or phrases that explain and summarise the information covered in accident reports in a short amount of time with highly accurate results. Strong, intelligent approaches, topic modelling techniques are widely used in natural language processing for semantic mining and topic recognition in unstructured data. As a result, the LDA model—one of the most well-known probabilistic unsupervised learning techniques—has been proposed in this study to identify the implicit subjects in a set of situations. [5]. This paper proposed a clever analysis using topic modelling techniques, which can be very helpful and

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025
https://doi.org/10.62647/ijitce.2025.v13.i2.pp584-597

effective to semantic mining and latent discovery context documents and datasets, given the growing use of new technologies, the revolution of data, the development of technology, and the use of AI in many fields. The unstructured textual data is selected since the other data sources (numerical and images-videos) have been studied using AI techniques that encompass supervised learning [6], [7].

Investigating topic modelling ways to hazards and safety accident subjects in the stations is what motivates us. In order to contribute to the future of smart safety and risk management in the stations, this study offers the technique of subject modelling based on LDA with additional models for advanced analytics. By using the models, we look into railway safety incidents that result in fatalities.

This research presents a novel approach to investigate the effective use of textual sources of data from railway station accident reports to identify the underlying causes of accidents and make a comparison between the textual and potential reasons. Whereas the fully automated procedure that can receive text input and produce results is not yet complete [8]. By using this approach, it is anticipated that problems will be resolved, including helping the decision-maker in real time and extracting the most important information from non-experts, better identifying the accident's specifics, designing a smart safety system with expertise, and making efficient use of safety history records. A These findings may contribute to a more methodical and intelligent investigation of safety and risk management. Our method captures important textual information on accidents and their causes using a cutting-edge LDA algorithm. This is how the remainder of the paper is organised: Related work in deep learning text categorisation and accident analysis has been described in Section II. Section III provides evaluation criteria and a detailed description of the methodology that has been employed. Details of our implementations are given in Section IV, and the outcomes are reported in Section V. Lastly, the conclusion is presented in Section VI.

2. LITERATURE SURVEY

S. Terabe, T. Kato, H. Yaginuma, N. Kang, and K. Tanaka, ‘‘Risk assessment model for railway passengers on a crowded platform The purpose of this study is to develop a risk assessment measure to provide an understanding of the safety of railway station platforms. We estimated the number of accidents on a platform in a year. It was influenced by the factors such as the design, equipment, and the profile of the station users. Consequently, 16 factors were defined, such as the platform design and passenger movement. Poisson regression and negative binomial regression models were employed to estimate and analyze the number of accidents from a station database containing 158 platforms from 52 stations in Japan. The results show that the number of accidents is related to the length of the narrow part of a platform, the width of the gap between the platform and train, the curvature of the platform, passenger flow crossing, the number of trains passing and stopping, and

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025
https://doi.org/10.62647/ijitce.2025.v13.i2.pp584-597

the audio and visual announcements concerning approaching trains. We expect that this result will allow railway companies to identify weaknesses in station safety and subsequently set priorities for investments in safety. Furthermore, administrative authorities can evaluate the safety performances of railway companies, and consider subsidies for investments in safety. M. Gethers and D. Poshyvanyk, ‘‘Using relational topic models to capture coupling among classes in object-oriented software systems Coupling metrics capture the degree of interaction and relationships among source code elements in software systems. A vast majority of existing coupling metrics rely on structural information, which captures interactions such as usage relations between classes and methods or execute after associations. However, these metrics lack the ability to identify conceptual dependencies, which, for instance, specify underlying relationships encoded by developers in identifiers and comments of source code classes. We propose a new coupling metric for object-oriented software systems, namely Relational Topic based Coupling (RTC) of classes, which uses Relational Topic Models (RTM), generative probabilistic model, to capture latent topics in source code classes and relationships among them. A case study on thirteen open source software systems is performed to compare the new measure with existing structural and conceptual coupling metrics. The case study demonstrates that proposed metric not only captures new dimensions of coupling, which are not covered by the existing coupling metrics, but also can be used to effectively support impact analysis. D. M. Blei, A. Y. Ng, and M. I. Jordan, ‘‘Latent Dirichlet allocation We describe latent Dirichlet allocation (LDA), a generative probabilistic model for collections of discrete data such as text corpora. LDA is a three-level hierarchical Bayesian model, in which each item of a collection is modeled as a finite mixture over an underlying set of topics. Each topic is, in turn, modeled as an infinite mixture over an underlying set of topic probabilities. In the context of text modeling, the topic probabilities provide an explicit representation of a document. We present efficient approximate inference techniques based on variational methods and an EM algorithm for empirical Bayes parameter estimation. We report results in document modeling, text classification, and collaborative filtering, comparing to a mixture of unigrams model and the probabilistic LSI model. H. Alawad, S. Kaewunruen, and M. An, ‘‘A deep learning approach towards railway safety risk assessment Railway stations are essential aspects of railway systems, and they play a vital role in public daily life. Various types of AI technology have been utilised in many fields to ensure the safety of people and their assets. In this paper, we propose a novel framework that uses computer vision and pattern recognition to perform risk management in railway systems in which a convolutional neural network (CNN) is applied as a supervised machine learning model to identify risks. However, risk management in railway stations is challenging because stations feature dynamic and complex conditions. Despite extensive efforts by industry associations

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025
https://doi.org/10.62647/ijitce.2025.v13.i2.pp584-597

and researchers to reduce the number of accidents and injuries in this field, such incidents still occur. The proposed model offers a beneficial method for obtaining more accurate motion data, and it detects adverse conditions as soon as possible by capturing fall, slip and trip (FST) events in the stations that represent high-risk outcomes. The framework of the presented method is generalisable to a wide range of locations and to additional types of risks. H. Alawad, S. Kaewunruen, and M. An, ‘‘Learning from accidents: Machine learning for safety at railway stations In railway systems, station safety is a critical aspect of the overall structure, and yet, accidents at stations still occur. It is time to learn from these errors and improve conventional methods by utilising the latest technology, such as machine learning (ML), to analyse accidents and enhance safety systems. ML has been employed in many fields, including engineering systems, and it interacts with us throughout our daily lives. Thus, we must consider the available technology in general and ML in particular in the context of safety in the railway industry. This paper explores the employment of the decision tree (DT) method in safety classification and the analysis of accidents at railway stations to predict the traits of passengers affected by accidents. The critical contribution of this study is the presentation of ML and an explanation of how this technique is applied for ensuring safety, utilizing automated processes, and gaining benefits from this powerful technology. To apply and explore this method, a case study has been selected that focuses on the fatalities caused by accidents at railway stations. An analysis of some of these fatal accidents as reported by the Rail Safety and Standards Board (RSSB) is performed and presented in this paper to provide a broader summary of the application of supervised ML for improving safety at railway stations. Finally, this research shows the vast potential of the innovative application of ML in safety analysis for the railway industry.

3. EXISTING SYSTEM

Despite the scatter of applying such method and the differences in terms been using in the literature, there is a shortage of such applications in the railway industry. Moreover, the NLP has been implemented to detect defects in the requirements documents of a railway signaling manufacturer [13].Also, for translating terms of the contract into technical specifications in the railway sector [14]. Additionally, identifying the significant factors contributing to railway accidents, the taxonomy framework was proposed using (Self-Organizing Maps – SOM), to classify human, technology, and organization factors in railway accidents [15].Likewise, association rules mining has been used to identify potential causal relationships between factors in railway accidents [16]. In the field of the machine learning and risk, safety accident, and occupational safety, there are many ML algorithms been used such as SVM, ANN, extreme learning machine (ELM), and decision tree (DT) [7], [17]. Scholars have been conducted the topic modeling in, where such method has been proved as one of the most powerful methods in data mining [18] many fields and applied in various areas such as software

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025
https://doi.org/10.62647/ijitce.2025.v13.i2.pp584-597

engineering [19], [4], [20], medical and health [21], [22], [23], [24] and linguistic science [25], [26], etc., Furthermore, from the literature It has been utilized this technique in for predictions some areas such as occupational accident [17], construction [8], [27], [28] and aviation [29], [30], [31]. For Understand occupational construction incidents in the construction and for construction injury prediction the method been conducted [32], [33], for analyzing the factors associated with occupational falls [34], for steel factory occupational incidents [35] and Cybersecurity and Data Science [36]. Moreover, From 156 construction safety accidents reports in urban rail transport in china risks information, relationships and factors been extracting and identified for safety risk analysis [37]. From the literature it has been seen that,there is no perfect model for all text classifications issues and also the process of extracting information from text is an incremental [38], [11]. In the railway sector, a semi-automated method has been examined for classifying unstructured text-based close call reports which show high accuracy. Moreover, for future expectations, it has been reported that such technology could be compulsory for safety management in railway [11]. Applying text analyzing methods in railway safety expected to solve issues such as time-consuming analysis and incomplete analysis. Additionally, some advantages have been proved, automated process, high productivity with quality and effective system for supervision safety in the railway system. Moreover, For the prevention of railway accidents, machine learning methods have been conducted. Many methods used for data mining including machine learning, information extraction (IE), natural language processing (NLP), and information retrieval (IR). For instance, to improve the identification of secondary crashes, a text mining approach (classification) based on machine learning been applied to distinguish secondary crashes based on crash narratives, which appear satisfactory performance and has great potential for identifying secondary crashes [39]. Such methods are powerful for railway safety, which aid decision-maker, investigate the causes of the accident, the relevant factors, and their correlations [40].It has been proved that text mining has several areas of future work development and advances for safety engineering railway [41]. Text mining with probabilistic modeling and k-means clustering is helpful for the knowledge of causes factors to rail accidents. From that application analysis for reports about major railroad accidents in the United States and the Transportation Safety Board of Canada, the study has been designating out that the factors of lane defects, wheel defects, level crossing accidents and switching accidents can lead to the many of recurring accidents [42]. Text mining is used to understand the characteristics of rail accidents and enhance safety engineers, and more to provide a worth amount of information with more detail. An accident reports data for 11 years in the U.S. are analyzed by the combination of text analysis with ensemble methods has been used to better understand the contributors and characteristics of these accidents, yet and more research is needed [41]. Also, from the U.S, railroad equipment accidents report are

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025
https://doi.org/10.62647/ijitce.2025.v13.i2.pp584-597

used to identify themes using a comparison text mining methods (Latent Semantic Analysis(LSA)and Latent Dirichlet Allocation(LDA)) [43]. Additionally, to identify the main factors associated with injury severity, data mining methods such as an ordered probit model, association rules, and classification and regression tree (CART) algorithms have been conducted. In the context of deep learning, Data From 2001 to 2016 rail accidents reports in the U.S. examined to extract the relationships between rail road accidents' causes and their correspondent descriptions. Thus for automatic understanding of domain specific texts and analyze railway accident narratives, deep learning has been conducted, which bestowed an accurately classify accident causes, notice important differences in accident reporting and beneficial to safety engineers [53].Also text mining conducted to diagnose and predict failures of switches [54]. For high-speed railways, fault diagnosis of vehicle onboard equipment, the prior LDA model was introduced for fault feature extraction [55] and for fault feature extraction the Bayesian network (BN) is also used [56]. For automatic classification of passenger complaints text and eigenvalue extraction, the term frequency-inverse document frequency algorithm been used with Naive Bayesian classifier [57].

Disadvantages

- The system never implemented ML algorithms been used such as SVM, ANN, extreme learning machine (ELM), and decision tree (DT) which are more accurate and efficient.
- The system didn't implement Self-Organizing Maps–SOM model to classify human, technology, and organization factors in railway accidents.

4. PROPOSED SYSTEM

This research presents a novel approach to investigate the effective use of textual sources of data from railway station accident reports to identify the underlying causes of accidents and make a comparison between the textual and potential reasons. where the fully automated procedure that can receive text input and produce unfinished results is located. By using this approach, it is anticipated that problems will be resolved, including helping the decision-maker in real time and extracting the most important information from non-experts, better identifying the accident's specifics, designing a smart safety system with expertise, and making efficient use of safety history records. A These findings may contribute to a more methodical and intelligent investigation of safety and risk management. Our method captures important textual information on accidents and their causes using a cutting-edge LDA algorithm.

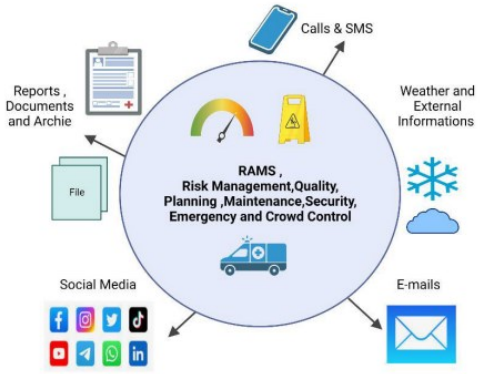
Advantages

- A decision assistance tool known as a DT uses a tree-like structure of choices and their probable results [40], [53]. Numerous machine learning (ML) techniques are available for safety assessments. To be more precise, we teach a DT to categorise the incidents that happened in the stations as well as the trends that emerged from them.
- Strong, useful information may be found in the textual data, including

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025
https://doi.org/10.62647/ijitce.2025.v13.i2.pp584-597

the time, location, description of the accidents, and victim's age range. For further mining to capture precise timings, the accident times were split into the halves of the day.

5. ARCHITECTURE



6. ALGORITHIM

Gradient boosting

Among other machine learning applications, gradient boosting is utilised for classification and regression problems. An ensemble of weak prediction models, usually decision trees, is what it provides as a prediction model. [1] [2] The resultant technique, known as gradient-boosted trees, often performs better than random forest when a decision tree is the weak learner. Like other boosting techniques, a gradient-boosted trees model is constructed step-by-step; however, it goes one step further by permitting optimisation of an arbitrary differentiable loss function.

K-Nearest Neighbors (KNN)

- The technique is straightforward yet incredibly effective; it classifies using a similarity metric.
- It doesn't "learn" until the test example is provided. Non-parametric slow learning
- We identify the K-nearest neighbours of fresh data from the training data whenever we have new data to categorise.

Logistic regression Classifiers

The relationship between a collection of independent (explanatory) factors and a categorical dependent variable is examined using logistic regression analysis. When the dependent variable simply has two values, like 0 and 1 or Yes and No, the term logistic regression is applied. When the dependent variable contains three or more distinct values, such as married, single, divorced, or widowed, the technique is sometimes referred to as multinomial logistic regression. While the dependent variable's data type differs from multiple regression's, the procedure's practical application is comparable.

When it comes to categorical-response variable analysis, logistic regression and discriminant analysis are competitors. Compared to discriminant analysis, many statisticians believe that logistic regression is more flexible and appropriate for modelling the majority of scenarios. This is due to the fact that, unlike discriminant analysis, logistic regression does not presume that the independent variables are regularly distributed.

Both binary and multinomial logistic regression are calculated by this program for

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025
https://doi.org/10.62647/ijitce.2025.v13.i2.pp584-597

both category and numerical independent variables. Along with the regression equation, it provides information on likelihood, deviance, odds ratios, confidence limits, and quality of fit. It does a thorough residual analysis that includes diagnostic residual plots and reports. In order to find the optimal regression model with the fewest independent variables, it might conduct an independent variable subset selection search. It offers ROC curves and confidence intervals on expected values to assist in identifying the optimal classification cutoff point. By automatically identifying rows that are not utilised throughout the study, it enables you to confirm your findings.

Naïve Bayes

The supervised learning technique known as the "naive bayes approach" is predicated on the straightforward premise that the existence or lack of a certain class characteristic has no bearing on the existence or nonexistence of any other feature.

However, it seems sturdy and effective in spite of this. It performs similarly to other methods of guided learning. Numerous explanations have been put forth in the literature. We emphasise a representation bias-based explanation in this lesson. Along with logistic regression, linear discriminant analysis, and linear SVM (support vector machine), the naive bayes classifier is a linear classifier. The technique used to estimate the classifier's parameters (the learning bias) makes a difference.

Although the Naive Bayes classifier is commonly used in research, practitioners who wish to get findings that are useful do not utilise it as often. On the one hand, the researchers discovered that it is very simple to build and apply, that estimating its parameters is simple, that learning occurs quickly even on extremely big databases, and that, when compared to other methods, its accuracy is rather excellent. The end users, however, do not comprehend the value of such a strategy and do not receive a model that is simple to read and implement.

As a consequence, we display the learning process's outcomes in a fresh way. Both the deployment and comprehension of the classifier are simplified. We discuss several theoretical facets of the naive bayes classifier in the first section of this lesson. Next, we use Tanagra to apply the method on a dataset. We contrast the outcomes (the model's parameters) with those from other linear techniques including logistic regression, linear discriminant analysis, and linear support vector machines. We see that the outcomes are quite reliable. This helps to explain why the strategy performs well when compared to others. We employ a variety of tools (Weka 3.6.0, R 2.9.2, Knime 2.1.1, Orange 2.0b, and RapidMiner 4.6.0) on the same dataset in the second section. Above all, we make an effort to comprehend the outcomes.

Random Forest

Random forests, also known as random decision forests, are ensemble learning techniques that build a large number of decision trees during training for tasks like regression and classification. The class chosen by the majority of trees is the

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025
https://doi.org/10.62647/ijitce.2025.v13.i2.pp584-597

random forest's output for classification problems. The mean or average forecast of each individual tree is given back for regression tasks. The tendency of decision trees to overfit to their training set is compensated for by random decision forests. Although random forests are less accurate than gradient enhanced trees, they often perform better than choice trees. However, their performance may be impacted by data peculiarities.

Tin Kam Ho[1] developed the first algorithm for random decision forests in 1995 by utilising the random subspace technique, which in Ho's definition is a means of putting Eugene Kleinberg's "stochastic discrimination" approach to classification into practice.

Leo Breiman and Adele Cutler created an algorithm extension and filed for a trademark in 2006 for "Random Forests" (owned by Minitab, Inc. as of 2019). The extension builds a set of decision trees with controlled variance by combining Breiman's "bagging" concept with random feature selection, which was initially proposed by Ho[1] and then separately by Amit and Geman[13].

Businesses commonly employ random forests as "blackbox" models since they need little configuration and produce accurate forecasts across a variety of inputs.

SVM

The goal of a discriminant machine learning approach in classification problems is to identify a discriminant function that can accurately predict labels for newly acquired instances based on an independent and identically distributed (iid) training dataset. A discriminant classification function takes a data point x and assigns it to one of the several classes that are part of the classification job, in contrast to generative machine learning techniques that call for calculations of conditional probability distributions. Discriminant techniques are less effective than generative approaches, which are mostly employed when prediction entails the identification of outliers. However, they need less training data and processing resources, particularly when dealing with a multidimensional feature space and when just posterior probabilities are required. Finding the equation for a multidimensional surface that optimally divides the various classes in the feature space is the geometric equivalent of learning a classifier.

SVM is a discriminant approach that, unlike genetic algorithms (GAs) or perceptrons, which are both often used for classification in machine learning, always returns the same optimal hyperplane value since it solves the convex optimisation issue analytically. The initialisation and termination criteria have a significant impact on the solutions for perceptrons. While the perceptron and GA classifier models are distinct every time training is started, training yields uniquely specified SVM model parameters for a given training set for a certain kernel that converts the data from the input space to the feature space. The sole goal of GAs and perceptrons is to reduce training error, which will result in several hyperplanes satisfying this criterion.

International Journal of Information Technology & Computer Engineering
ISSN 2347-3657
Volume 13, Issue 2, 2025
https://doi.org/10.62647/ijitce.2025.v13.i2.pp584-597

7. IMPLEMENTATION

Modules

Service Provider

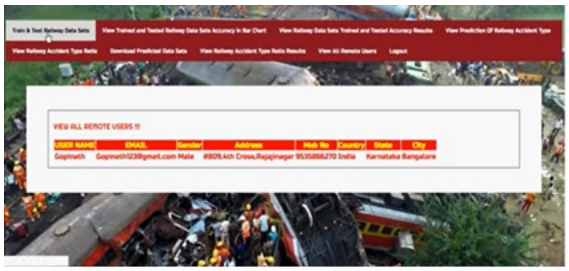
In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Train & Test Railway Data Sets, View Trained and Tested Railway Data Sets Accuracy in Bar Chart, View Railway Data Sets Trained and Tested Accuracy Results, View Prediction Of Railway Accident Type, View Railway Accident Type Ratio, Download Predicted Data Sets, View Railway Accident Type Ratio Results, View All Remote Users.

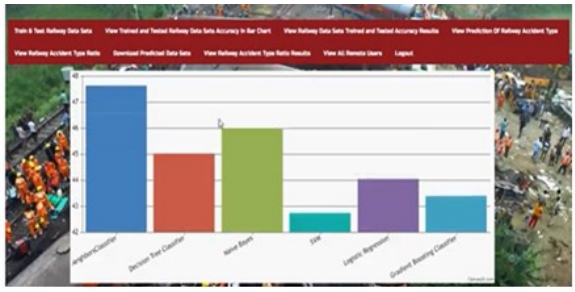
View and Authorize Users

The administrator may see a list of all registered users in this module. Here, the administrator may see the user's information, like name, email, and address, and they can also grant the user permissions.

Remote User

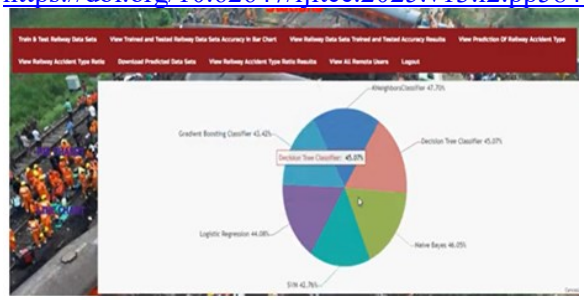
Additionally, there are n users in this module. The user must first register before proceeding. The user's information will be entered into the database when they register. Once his registration is complete, he must use his password and authorised user name to log in. The user can perform the following after successfully logging in: REGISTER AND LOGIN, PREDICT RAILWAY ACCIDENT TYPE, and VIEW YOUR PROFILE.





8. SCREEN SHOTS

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025
https://doi.org/10.62647/ijitce.2025.v13.i2.pp584-597



9. CONCLUSION

In many domains, including the safety and risk management of train stations for text mining, topic models are crucial. A set of terms that appear in statistically significant methods is called a topic in topic modelling. Voice recordings, investigative reports, risk document evaluations, and more can all be considered texts.

This study presents several examples of how unsupervised machine learning topic modelling may support risk management, safety accident investigation, and industry-based accident recording and documentation. The platforms are the hot spot in the stations, according to the proposed model and the explanation of the accident's underlying reasons. The results show that falls, being hit by trains, and electric shock are the four primary reasons why accidents happen at the station. Furthermore, it appears that the dangers are higher on certain days of the week and at night.

Increased safety text mining allows for the acquisition of knowledge across a broad range of time periods, which improves RAMS efficiency and gives all stakeholders a comprehensive viewpoint.

Using unsupervised machine learning is beneficial for safety since it can solve problems, uncover hidden patterns, and address a variety of issues, including:

• Text data in unstructured formats and from a variety of viewpoints
• Capture the relationships, causations, and more for ranking risks and related information;
• Prioritise risks and measures implementations;
• Smartly label, cluster, centroids, sampling, and associated coordinates;
• Power for discovery, handling missing values, and identifying safety and risk kyes from data
• Support the safety review process and the process of learning from the extensive and lengthy experience.
• Scale and weighted configuration options are available for use in risk assessment.

Even though this paper emphasises the innovative use of unsupervised machine learning in railway accident classification and root cause analyses, future research on large-scale data topics pertaining to station diversity, size, safety cultures, and other factors must be prioritised with additional unsupervised machine learning algorithm techniques. Lastly, this study improves safety, but it also highlights the value of

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025
https://doi.org/10.62647/ijitce.2025.v13.i2.pp584-597

textual data and recommends rethinking the data collection process to be more thorough.

REFERENCES

[1] S. Terabe, T. Kato, H. Yaginuma, N. Kang, and K. Tanaka, ‘‘Risk assessment model for railway passengers on a crowded platform,’’ *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2673, no. 1, pp. 524–531, Jan. 2019, doi: 10.1177/0361198118821925.

[2] *Annual Health and Safety Report 19/2020*, RSSB, London, U.K., 2020.

[3] D. M. Blei, ‘‘Probabilistic topic models,’’ *Commun. ACM*, vol. 55, no. 4, pp. 77–84, Apr. 2012, doi: 10.1145/2133806.2133826.

[4] M. Gethers and D. Poshyvanyk, ‘‘Using relational topic models to capture coupling among classes in object-oriented software systems,’’ in *Proc. IEEE Int. Conf. Softw. Maintenance*, Sep. 2010, pp. 1–10, doi: 10.1109/ICSM.2010.5609687.

[5] D. M. Blei, A. Y. Ng, and M. I. Jordan, ‘‘Latent Dirichlet allocation,’’ *J. Mach. Learn. Res.*, vol. 3, nos. 4–5, pp. 993–1022, Mar. 2003, doi: 10.1016/B978-0-12-411519-4.00006-9.

[6] H. Alawad, S. Kaewunruen, and M. An, ‘‘A deep learning approach towards railway safety risk assessment,’’ *IEEE Access*, vol. 8, pp. 102811–102832, 2020, doi: 10.1109/ACCESS.2020.2997946.

[7] H. Alawad, S. Kaewunruen, and M. An, ‘‘Learning from accidents: Machine learning for safety at railway stations,’’ *IEEE Access*, vol. 8, pp. 633–648, 2020, doi: 10.1109/ACCESS.2019.2962072.

[8] A. J.-P. Tixier, M. R. Hallowell, B. Rajagopalan, and D. Bowman, ‘‘Automated content analysis for construction safety: A natural language processing system to extract precursors and outcomes from unstructured injury reports,’’ *Autom. Construct.*, vol. 62, pp. 45–56, Feb. 2016, doi: 10.1016/j.autcon.2015.11.001.

[9] J. Sido and M. Konopik, ‘‘Deep learning for text data on mobile devices,’’ in *Proc. Int. Conf. Appl. Electron.*, Sep. 2019, pp. 1–4, doi: 10.23919/AE.2019.8867025.

[10] A. Serna and S. Gasparovic, ‘‘Transport analysis approach based on big data and text mining analysis from social media,’’ *Transp. Res. Proc.*, vol. 33, pp. 291–298, Jan. 2018, doi: 10.1016/j.trpro.2018.10.105.

[11] P. Hughes, D. Shipp, M. Figueres-Esteban, and C. van Gulijk, ‘‘From free-text to structured safety management: Introduction of a semi automated classification method of railway hazard reports to elements on a bow-tie diagram,’’ *Saf. Sci.*, vol. 110, pp. 11–19, Dec. 2018, doi: 10.1016/j.ssci.2018.03.011.

[12] A. Chanen, ‘‘Deep learning for extracting word-level meaning from safety report narratives,’’ in *Proc. Integr. Commun. Navigat. Surveill. (ICNS)*, Apr. 2016, pp. 5D2-1–5D2-15,doi: 10.1109/ICNSURV.2016.7486358.

[13] A. Ferrari, G. Gori, B. Rosadini, I. Trotta, S. Bacherini, A. Fantechi, and S. Gnesi, ‘‘Detecting requirements defects with NLP patterns: An industrial experience in the railway domain,’’ *Empirical Softw. Eng.*, vol. 23, no. 6, pp. 3684–3733, Dec. 2018, doi: 10.1007/s10664-018- 9596-7.

ISSN 2347–3657
Volume 13, Issue 2, 2025
https://doi.org/10.62647/ijitce.2025.v13.i2.pp584-597
[14] G. Fantoni, E. Coli, F. Chiarello, R. Apreda, F. Dell'Orletta, and G. Pratelli, ''Text mining tool for translating terms of contract into technical specifications: Development and application in the railway sector,'' *Comput. Ind.*, vol. 124, Jan. 2021, Art. no. 103357, doi: 10.1016/j.compind.2020.103357.
[15] G. Yu, W. Zheng, L. Wang, and Z. Zhang, ''Identification of significant factors contributing to multi-attribute railway accidents dataset (MARA-D) using SOM data mining,'' in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 170–175, doi: 10.1109/ITSC.2018.8569336.