Utilizing Unsupervised Machine Learning for Safety Accident Management in Railway Stations

[1] B.Ganga Bhavani, [2] V.Likitha, [3] M.Nishitha, [4] MS Karthika
[1,2,3]UG Scholar, Department of Computer Science and Engineering, St. Martin's Engineering College, Secunderabad, Telangana, India, 500100
[4]Assistant Professor, Department of Computer Science and Engineering, St. Martin's Engineering College, Secunderabad, Telangana, India, 500100
karthikamscse@gmail.com

Abstract:

The application of unsupervised machine learning techniques to improve safety and accident management in railway stations. With the increasing complexity of station environments and passenger volumes, traditional safety management systems often struggle to detect and respond to risks in real time. This research focuses on leveraging clustering and anomaly detection algorithms to identify unusual patterns in station data, such as crowd density fluctuations, equipment malfunctions, and unanticipated passenger behaviors. By processing data from surveillance systems, sensors, and historical safety records, the system can proactively detect safety risks and trigger alerts before incidents occur. By leveraging clustering, anomaly detection, and dimensionality reduction methods, the study proposes an. automated system capable of identifying unusual patterns in real-time data collected from various sensors, cameras, and other monitoring systems in the railway environment.

Keywords: Unsupervised machine learning, safety accident management, railway stations, anomaly detection, clustering algorithms, pattern recognition, predictive maintenance, passenger safety, risk assessment, surveillance data analysis, real-time monitoring, sensor data processing, fault detection

1. INTRODUCTION

Trains as public transportation has been considered as safer than other means. However, passengers on trains stations sometimes face many risks because of many overlapping factors such as station operation, design, and passenger behaviours. Due to the gradually increasing demand and the heavily congested society and the state of some

AI approaches which cover supervised learning, so the unstructured textual data is targeted. Hence, our motivation is to investigate the topic modelling approaches to risks and safety accident subjects in the stations. This work provides the method of topic modelling based on LDA with other models for advanced analytics, aiming to make contributions in the future of smart safety and risk management in the stations. Through applying the models, we investigate the safety accidents for fatality accident in the railway. This paper establishes an innovative method in the area to studies how the textual source of data of railway station accident reports could be efficiently used to extract the root causes of accidents and establish an analysis between the textual and the possible cause. where the full automated process that has ability to get the input of text and provide outputs not yet ready and public safety is the main concern of the railway industry and one of the critical parts of the system. European Union put into practice Reliability, Availability, Maintainability and Safety (RAMS)as a standard in 1999 known as EN 50126. Aiming to prevent railway accidents and ensure a high level of safety in railway operations.

The RAMS analyses concepts lead to minimizing the risks to acceptable levels and rise safety levels. However, that have been an urgent issue and still, the reports show several people are killed every year in the railway station, some accidents lead to injuries or fatalities. For example, In Japan in 2016, 420 accidents occurred that included being struck by a train, which resulted in 202 deaths. This including of those 420 accidents, 179 (resulting in 24 fatalities) included falling from a platform and following injury or death as a consequence of hitting with a train. In the UK, 2019/20, it has been reported that Most passenger injuries occur from accidents in stations. Greatest Major injuries are the outcome of slips, trips and falls, of which there were approximately 200 [2] play significant impact in reducing injuries on station platforms and provide quality, reliable and safe travel environment for all passengers, worker and public. Even if some accident does not result in deaths or injuries, such accidents cause delay, cost, fear and anxiety among the people,

2. LITERATURE SURVEY

Railway stations are critical nodes in transportation networks, prone to various safety challenges such as accidents and incidents. Traditional methods of managing safety rely heavily on reactive measures, which may not adequately prevent accidents or mitigate their impact. In recent years, the application of unsupervised machine learning (ML) techniques has emerged as a promising approach to enhancing safety management in railway stations. This literature review explores the current state of research and applications of unsupervised ML in identifying and managing safety accidents within railway station environments. Key studies highlight the effectiveness of anomaly detection algorithms, clustering techniques, and pattern recognition models in analysing large volumes of heterogeneous data sources, including sensor data, video surveillance feeds, and historical incident reports.

The review discusses various unsupervised ML algorithms employed in safety accident management, such as k means clustering, Gaussian mixture models, and autoencoders. These algorithms enable railway operators to uncover hidden patterns and anomalies indicative of safety risks, facilitating timely intervention and proactive safety measures. Moreover, the review addresses challenges and limitations associated with the application of unsupervised ML in railway safety, including data quality issues, interpretability of results, and the need for domain-specific expertise in model development and validation. Ethical considerations related to data privacy, algorithm transparency, and stakeholder engagement are also critically examined. Finally, the literature review identifies gaps in current research and suggests future research directions to advance the field. These include the integration of real-time data streams, development of hybrid supervised unsupervised ML models, enhancement of algorithm scalability, and the importance of interdisciplinary collaboration between data scientists, transportation engineers, and safety experts.

In conclusion, while unsupervised ML shows considerable promise in transforming railway station safety management, further research and technological advancements are essential to fully harness its potential and address existing challenges in ensuring safe and efficient railway operations.

Chen et al. propose a robust system employing clustering techniques to analyze accident data effectively. Kumar and Kaur delve into the application of various unsupervised machine learning algorithms for accident prediction and prevention, showcasing promising results in real-world scenarios. Liang et al. focus on anomaly detection and risk assessment, presenting a framework that combines clustering and outlier detection methods to identify safety related anomalies. improvement in railway stations, while Park et al. propose an integrated approach combining unsupervised learning with IoT devices for safety monitoring. Journal of Engineering Sciences .Zhou and Liu emphasize data-driven safety management using unsupervised learning techniques, while Yang et al. delve into anomaly detection in surveillance videos. These studies collectively underscore the potential of unsupervised machine learning in enhancing safety measures and accident management in railway stations.Through a review of these works, it is evident that automation plays a crucial role in cyber threat identification and profiling. Ongoing research in AI-driven cybersecurity solutions, integration of predictive analytics, and advancements in secure data-sharing mechanisms will be instrumental in strengthening cyber resilience. The development of dynamic and adaptive threat intelligence models remains a top priority for cybersecurity researchers, ensuring organizations remain well-equipped to handle evolving cyber threats. 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 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. Applying text analysing 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 Such methods are powerful for railway safety, which aid decision-maker, investigate the causes of the accident, the relevant factors, and their correlations .It has been proved that text mining has several areas of future work development and advances for safety engineering railway .Text mining with probabilistic modelling 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 . 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 analysed by the combination of text analysis with ensemble methods has been used to better understand the contributorsand characteristics of these accidents, yet and more research is needed. Also, from the U.S, railroad equipment accidents report are used to identify themes using a comparison text mining methods (Latent Semantic Analysis (LSA)and Latent Dirichlet Allocation(LDA)) . 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. Using the U.S accidents highway railroad grade crossings database for the period 2007–2013, where Some factors have been discussed such the train speed, age, gender and the time. In recent years, the revolution of big data is opportunities in the railway industry, and that is opening up for safety analysis depends on data, so, the approach to proactively identify high-risk scenarios been recommended such as applying the Natural Language Processing (NLP) analysis .From Big Data Application Case A Supervision System has been introduced as a significant role tool in railway safety supervision system. Applying Text Mining Methods in Railway Safety from accident and fault analysis reports been conducted . Also, As well as big data and natural language is an opportunity should be to use for processing for Analysing Railway Safety, NLP framework for analysing accident data been explained using investigation reports of railway accidents.

3. PROPOSED METHODOLOGY

This paper establishes an innovative method in the area to studies how the textual source of data of railway station accident reports could be efficiently used to extract the root causes of accidents and establish an analysis between the textual and the possible cause. where the full automated process that has ability to get the input of text and provide outputs not yet ready. Applying this method expected to come overcome issues such as aid the decision-maker in real time and extract the key information to be understandable from non-experts, better identify the details of the accident in-depth, design expert smart safety system and effective usage of the safety history records. A Such results could support in the analysis of safety and risk management to be systematic and smarter. Our approach uses state-of-the-art LDA algorithm to capture the critical texts information of accidents and their causes.

Predictive Maintenance and Anomaly Detection

Predict and prevent accidents caused by equipment failure. Sensors are installed on trains and infrastructure (e.g., tracks, signals) to continuously collect data on operational parameters like temperature, vibration, pressure, speed, and wear. Machine Learning Models (e.g., Random Forest, Neural Networks, Support Vector Machines) are used to process this data and identify patterns that indicate potential failures. For example, excessive vibration could indicate a failure in the train's braking system

Real-time Safety Monitoring and Incident Detection

Computer Vision (e.g., Convolutional Neural Networks - CNNs) Used to analyze live video feeds for hazardous conditions (e.g., workers not wearing safety gear, workers entering restricted areas, machine malfunctions). Sensor Fusion (e.g., Kalman Filters, Sensor Fusion Algorithms) Combining data from multiple sensor sources (e.g., wearable sensors, environmental sensors) to identify accidents in real-time. Natural Language Processing (NLP): Analyzing text-based data from incident reports, work logs, and communication tools to detect patterns or keywords related to safety hazards (e.g., employee complaints or machinery issues).

Real-time Decision Systems (Automated Alerts)

Rule-based Expert Systems: Based on predefined rules, these systems can trigger alarms or send alerts when certain conditions are met (e.g., an equipment malfunction is detected or a worker is in a dangerous zone).Reinforcement Learning (RL) RL models can be used to improve decision-making over time by learning optimal responses to certain environmental conditions or safety incidents.

Machine Learning Models for Accident Prevention

The heart of the system lies in the predictive and classification models used for accident management. Some key models include .Predictive Models (Accident Prevention).Supervised Learning (e.g., Logistic Regression, Decision Trees, Random Forests, XG Boost) To predict the likelihood of an accident based on historical data (e.g., trends indicating increased risk factors such as high temperatures, malfunctioning machinery, or worker fatigue). Time Series Forecasting (e.g., ARIMA, LSTM) To predict potential accidents based on past incident patterns over time. Anomaly Detection (e.g., Isolation Forests, Autoencoders) Detect outliers or abnormal conditions that could lead to accidents, such as unexpected behavior in equipment or workers (e.g., sudden motion detected in a robotic system).

Post-Incident Analysis

After an accident, the system can analyze the root cause of the incident, and learn from it to prevent future occurrences. Incident Categorization Using ML clustering algorithms (e.g., K-means, DBSCAN), incidents can be categorized and compared, helping to identify recurring patterns. Root Cause Analysis the system can help determine what caused the accident (e.g., faulty equipment, human

Enhanced Decision-making for Security Teams

The automation of cyber threat intelligence reduces the workload for security analysts, allowing them to focus on high-priority incidents. By providing actionable insights and predictive analytics, the system helps organizations make informed security decisions and strengthen their defense mechanisms against emerging cyber threats.By implementing this automated approach, cybersecurity teams can improve threat detection, reduce manual workload, and enhance their ability to respond to emerging cyber threats in real-time. The use of AI and ML ensures that organizations remain ahead of adversaries, reducing the impact of cyberattacks and improving overall security prepare

error, environmental factors), and this can feed back into the system to improve safety protocols. Continuous Learning as more data is collected over time, the system's machine learning models can be retrained to adapt to new patterns and improve prediction accuracy.

4. EXPERIMENTAL ANALYSIS

Unsupervised machine learning techniques have been increasingly applied to enhance safety and manage accidents in railway stations. These methods analyze complex datasets to identify patterns and anomalies without predefined labels, offering valuable insights for accident prevention and response.

One notable application is the use of clustering algorithms to analyze safety incidents. Researchers Gupta and Sharma explored proactive safety improvement approaches in railway stations, while Park et al. proposed integrating unsupervised learning with IoT devices for safety monitoring. These studies highlight the potential of unsupervised machine learning in enhancing safety measures and accident management in railway stations. While these applications demonstrate the promise of unsupervised machine learning in railway safety management, challenges remain. The complexity of railway operations and the need for high accuracy necessitate continuous research and adaptation of these models to ensure their effectiveness in real-world scenarios. Another significant approach involves topic modeling, such as Latent Dirichlet Allocation (LDA), to systematically detect accident characteristics from textual data. A study utilizing LDA analyzed 1,000 accident reports from UK railway stations, revealing that platforms are common hotspots for incidents, with primary causes including falls and collisions. This analysis provides a deeper understanding of risk factors, aiding in the development of targeted safety measures.

Additionally, anomaly detection techniques have been employed to identify unusual patterns in surveillance videos and sensor data, facilitating early intervention and accident prevention. These methods can detect deviations from normal behavior, such as unauthorized access or equipment malfunctions, enabling prompt responses to potential safety threats.

Figure 1: Home Page

Figure 2: User Login Page

Figure 3: User Login Page 2

Figure 4: Browser Dataset

Figure 5: Railway Accident Ratio Result

Figure 6: Railway Accident Ratio Type

5. CONCLUSION

Topic models have an important role in many fields and in such case of safety and risk management in the railway stations for texts mining. In Topic modeling, a topic is a list of words that occur in statistically significant methods. A text can be voice records investigation reports, or reviews risk documents and so on. This research displays various cases for the power of unsupervised machine learning topic modeling in promoting risk management, safety accidents investigation and restructuring accidents recording and documentation on the industry based level. The description of the root causes accident, the suggested model, it has been showing that the platforms are the hot point in the stations. The outcomes reveal the station's accidents to be occurring owing to four main causes: falls, struck by trains, electric shock. Moreover, the night time and days of the week seems to contact to the risks are significant. With increased safety text mining, knowledge is gained on a wide scale and different periods resulting in greater efficiency RAMS and providing the creation of a holistic perspective for all stakeholders. Application of the unsupervised machine learning technique is useful for safety since, which is solving, exploring hidden patterns and deal with many challenges such as: Text data from many perspectives and in unstructured forms Power for discovery, dealing with missing values, and spot safety and risk kyes from data Smart labelling ,clustering, centroids, sampling, and associated coordinates Capture the relationships, causations, more for ranking risks and related information Prioritization risks and measures implementations. Aid the process of safety review and learning from the long and massive experience. Can be used the scale and weighted as configuration options which can be used for assessing risks

Although this paper highlights the innovative of unsupervised machine learning in accidents classification of railway accidents and root cause analyses, it is a necessity to focus on expanded research on the huge data topics concerning the diversity of the station's locations, size and safety cultures and other factors with further techniques of unsupervised machine learning algorithms in the future. Finally, this research enhances safety, but it raises the importance of data in text form and suggests redesigning the way of gathering data to be more comprehensive. Through clustering techniques such as K-Means and DBSCAN, railway authorities can group safety incidents, recognize high-risk areas, and proactively implement preventive measures. Additionally, anomaly detection models enable the identification of rare but critical safety hazards, such as fire outbreaks, overcrowding and derailments, improving overall passenger safety.

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 semiautomated 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.

[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,'' Compute. 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. In tell. Transp. Syst.(ITSC), Nov. 2018, pp. 170–175, doi: 10.1109/ITSC.2018.8569336.

[16] Y. Wang, W. Zheng, H. Dong, and P. Gao, ''Factors correlation mining on railway accidents using association rule learning algorithm,'' in Proc. IEEE 23rd Int. Conf. In tell. Transp. Syst., vol. 22, Sep. 2020, pp. 1–6, doi:10.1109/ITSC45102.2020.9294317.

[17] S. Sarkar, S. Vinay, R. Raj, J. Maiti, and P. Mitra, ''Application of optimized machine learning techniques for prediction of occupational accidents,'' Compute. Oper. Res., vol. 106, pp. 210–224, Jun. 2019, Doi: 10.1016/j.cor.2018.02.021.

[18] H. Jelodar, Y Wang, Cyan, X. Feng, X. Jiang, Y. Li, and L. Zhao, ''Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey,'' Multimedia Tools Appl., vol. 78, no. 11, pp. 15169–15211, Jun. 2019, doi: 10.1007/s11042-018-6894-4.

[19] S. W. Thomas, ''Mining software repositories using topic models,'' in Proc. 33rd Int. Conf. Softw. Eng. (ICSE), May 2011, pp. 1138–1139, doi:10.1145/1985793.1986020.

[20] H. U. Asuncion, A. U. Asuncion, and R. N. Taylor, ''Software traceability with topic modeling,'' in Proc. ACM/IEEE 32nd Int. Conf. Softw. Eng., vol. 1, May 2010, pp. 95–104, doi: 10.1145/1806799.1806817.