Unsupervised Machine Learning for Managing Safety Accidents in Railway Stations

ABSTRACT

For both passenger and freight transportation, railroad operations must be dependable, accessible, maintained, and safe (RAMS). In many urban areas, railway stations risk and safety accidents represent an essential safety concern for daily operations. Moreover, the accidents lead to damage to market reputation, including injuries and anxiety among the people and costs. This stations under pressure caused by higher demand which consuming infrastructure and raised the safety administration consideration. To analysing these accidents and utilising the technology such AI methods to enhance safety, it is suggested to use unsupervised topic modelling for better understand the contributors to these extreme accidents. It is conducted to optimise Latent Dirichlet Allocation (LDA) for fatality accidents in the railway stations from textual data gathered RSSB including 1000 accidents in the UK railway station. This research describes using the machine learning topic method for systematic spot accident characteristics to enhance safety and risk management in the stations and provides advanced analysing. The study evaluates the efficacy of text by mining from accident history, gaining information, lesson learned and deeply coherent of the risk caused by assessing fatalities accidents for large and enduring scale. This Intelligent Text Analysis presents predictive accuracy for valuable accident information such as root causes and the hot spots in the railway stations. Further, the big data analytics ’ improvement results in an understanding of the accidents’ nature in ways not possible if a considerable amount of safety history and not through narrow domain analysis of the accident reports. This technology renders stand with high accuracy and a beneficial and extensive new era of AI applications in railway industry safety and other fields for safety applications.

**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 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 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.

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 [44]. 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 [45], so, the approach to proactively identify high-risk scenarios been recommended such as applying the Natural Language Processing (NLP) analysis [46].

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 was conducted [47]. Also, As well as big data and natural language is an opportunity should be to use for processing for Analyzing Railway Safety, NLP framework for analyzing

accident data been explained using investigation reports of railway accidents [48].Moreover, for Fault Diagnosis in Railway System, classification of maintenance text been proposed using (LDA) algorithm [49], and to improve the fault diagnosis performance [50]. In China railway, for prediction passenger capacity, the social network text data have been used with a combination of text mining and deep learning which show a good accuracy rate [51]. Also from the Chinese Railway, natural language processing has been applied for extraction and analysis of risk factors from accident reports [52].

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

Proposed System

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

**Advantages**

* A DT is a determination support tool that applies a treelike pattern of decisions and their likely outcomes [40], [53]. There are many possible (ML) approaches towards safety analysis. More exactly, we train a DT to classify the accidents and the patterns that occurred in these accidents in the stations.
* The textual data have strong key information which can be used such as the time, description of the accidents, location and the range age of the victim. The time of accidents occurred been divided as the Parts of the Day for more mining to capture accurate times.

**SYSTEM REQUIREMENTS**

➢ **H/W System Configuration:-**

➢ Processor - Pentium –IV

➢ RAM - 4 GB (min)

➢ Hard Disk - 20 GB

➢ Key Board - Standard Windows Keyboard

➢ Mouse - Two or Three Button Mouse

➢ Monitor - SVGA

**SOFTWARE REQUIREMENTS:**

* **Operating system :** Windows 7 Ultimate.
* **Coding Language :** Python.
* **Front-End :** Python.
* **Back-End :** Django-ORM
* **Designing :** Html, css, javascript.
* **Data Base :** MySQL (WAMP Server).