UNRAVELING RAILWAY SAFETY TEXT ANALYTICS ON RAILWAY ACCIDENTS

[1]Dr.R.Rajani, [2]Dr.V.Gayatri

[1]Professor, [2]Associate Professor, [1,2]Department of Computer Science & Engineering, Geethanjali Institute of Science and Technology, Gangavaram, Andhra Pradesh, India

ABSTRACT

Railway operations must prioritize safety (RAMS) for both passenger and freight transportation. Urban railway stations face safety challenges that can harm operations, reputation, cause injuries, and incur costs. With rising demand and pressure on infrastructure, safety administration becomes critical. To address extreme accidents, Unsupervised Topic Modeling, particularly optimized Latent Dirichlet Allocation (LDA), is suggested. Analyzing textual data from 1000 Australia railway station accidents, this method systematically identifies contributors to accidents, optimizes safety, and improves risk management. Machine learning is used to spot accident characteristics, offering advanced analysis and evaluating accident history. Intelligent Text Analysis ensures predictive accuracy for vital accident information. Big data analytics enhances our understanding of accidents, going beyond narrow domain analysis. This technology stands out for its high accuracy, ushering in a new era of AI applications in railway industry safety and beyond.

INTRODUCTION

In my work on using unsupervised machine learning to improve safety at railway stations, I take a step-by-step approach. First, I gather lots of data about accidents that have happened in the past, like where they occurred and what time of day they happened. Then, I clean up this data to make sure it's accurate and easy to understand.

Next, I look for patterns in the data using special computer programs. These programs help me see if there are any common factors or trends that might explain why accidents happen. For example, they might show that accidents tend to occur more often in certain areas of the station or during busy times of the day.

Once I've identified these patterns, I use them to come up with ideas for making the station safer. This could involve things like adding more lights in dark areas, putting up signs to warn people about potential hazards, or changing the layout of the station to make it easier for people to get around safely.

Overall, my goal is to use data and technology to prevent accidents and keep everyone who uses the railway station safe. It's all about using what we know to make things better for everyone.

RAILWAY SYSTEM

Trains as public transportation have 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 station's layout and complexity in design, there exist a set of potential risks during the operation of the stations. Furthermore, Passenger, people 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. Greatest Major injuries are the outcome of slips, trips and falls, of which there were approximately 200 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, interruption in the operations and damage the industry reputation. Also, to provide or invest any control safety measurements the stations it is crucial to Considering the risks associated with the railway incidents and risks in the station and identification of many factors related to the accident by a comprehensive knowledge of the root cause of accidents considering all the possible technology.

The objective of this research is to analysis a collection case of accidents between 01/01/2012 and 17/04/2022 data to introduce a smart method, which expected to develop the safety level future, the risk management process, and the way to collect data in the railway stations. Analysing an extensive amount of data recorded in a different form are a challenging job. Nowadays, it is hard to obtain for specific information in such mix digitization big data in including Web, video, images and other sources, it is research of a needle in a haystack. Thus, a powerful tool for assistance manage, search and understand these vast amounts of information is needed indeed. Many pre-processing techniques and algorithms are required to obtain valuable characteristics from an enormous amount of safety data in the stations including textual. The study covers the topic modelling to identify useful characteristics such the root cause of the accidents and also exploring the factors which are multiple groups of words or phrases that explain and summarize the content covered by an accident's reports reducing time with high accuracy of outcomes. Topic modelling techniques are robust smart methods that extensively applied in natural language processing to topic detection and semantic mining from unstructured documents. Consequently, It has been suggested in this work the LDA model which is one of the best-known probabilistic unsupervised learning methods that marks the topics implicit in collection of contexts . Since increasing of applying new technologies and the revolution of data, the development of technology and utilising AI in many fields it suggested in this paper a smart analysis utilising the topic modelling techniques which can be very useful and effective to semantic mining and latent discovery context documents and datasets.

The other source of data been conducted utilising 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 handling 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.

MOTIVATION

Working on a project to make railway stations safer with machine learning is really about helping people. I'm using smart computer programs that can look at a lot of information and find patterns or unusual things without being told what to look for. This is cool because it can help us see problems or dangers we didn't know about before.

Imagine all the busy times at train stations, with people everywhere, trains coming and going. It's easy for accidents to happen. My goal is to use technology to spot those risky spots or moments before anything bad happens. This way, we can do something about it, like put up more signs, change how people move around, or even just keep a closer eye on the busiest places.

By doing this, I'm not just playing with tech; I'm actually working on something that could keep people from getting hurt. It's about making sure everyone can go about their day, take their trains, and get back home safely. It's a big challenge, but it's really rewarding because it's all about looking after each other.

LITERATURE SURVEY

Author: Hamad Alawad, SakdiratKaewunruen, and Min An.

Title: "Learning From Accidents: Machine Learning for Safety at RailwayStations"
Project Conduct Year:2019

Description: Safety is paramount in railway systems, yet accidents persist despite conventional safety measures. Leveraging machine learning (ML) presents a promising avenue for enhancing safety analysis and prediction in this context. This paper explores the application of the decision tree (DT) method in classifying safety incisdents and analyzing accidents at railway stations to predict passenger traits associated with accidents. A case study focusing on fatalities at railway stations is conducted using data from the Rail Safety and Standards Board (RSSB). Findings demonstrate the potential of supervised ML in improving safety at railway stations by identifying trends and patterns in accidents. The study underscores the importance of integrating ML into railway safety systems and provides recommendations for its implementation. This research highlights the innovative application of ML in safety analysis for the railway industry, offering insights into its transformative potential.

Author: G. Yu, W. Zheng, L. Wang, and Z. Zhang,
Title: Identification of significant factors contributing to multi-attribute railway accident dataset (MARA-D) using SOM data mining,'
Project Conduct Year:2018

Description: Although a lot of labor and financial forces have been put into safety work, railway accidents continue to be the major concern in China. The aim of this study is to identify the significant factors contributing to railway accidents and enable stakeholders to fully learn from accidents. The Cognitive Reliability and Error Analysis Method - Railway Accidents (CREAM-RAs) taxonomy framework was proposed to classify human, technology, and organization factors in railway accidents. To establish a Multi-attribute Railway Accidents Dataset (MARA-D), 392 railway accident reports were collected and collated under the CREAM-RAs framework. The data mining technique (Self-Organizing Maps - SOM) was adopted to convert MARA-D into 2-dimensional maps. The key accident causes were dug out and risk information was transmitted to various related railway departments. Thus, the relevant measures were raised to improve safety and promote health management of railways.

Author:S. Sarkar, S. Vinay, R. Raj, J. Maiti, and P. Mitra,
Title: "Application of optimized machine learning techniques for prediction of occupational accidents,"
Project Conduct Year:2019

Description: The utilization of machine learning (ML) to predict accidents in occupational safety, an area where ML application is relatively new. The research highlights the importance of optimizing parameters and extracting decision-making rules from accident data. Two popular ML algorithms, Support Vector Machine (SVM) and Artificial Neural Network (ANN), are employed, optimized using Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) to improve accuracy and robustness. Results show that PSO-based SVM performs best. Additionally, decision-making rules are extracted by combining PSO-based SVM with the decision tree C5.0 algorithm, revealing nine useful rules for identifying accident root causes. A case study from a steel plant validates the methodology's effectiveness, contributing to improving safety management practices in the occupational safety domain through ML techniques.

Author: ShuhratjonHidirov, Hakan Guler
Title:" Reliability, availability and maintainability analyses for railway infrastructure management"
Project Conducted Year: 2016

Description: The importance of reliability, availability, and maintainability (RAM) in railway transportation, as outlined in the EN 50126 standard established by the European Union in 1999. RAM analysis is crucial for preventing accidents and ensuring a high level of safety in railway operations. The study proposes a model framework comprising several stages and applies RAM analyses to railway infrastructure management, particularly for high-speed railways in Uzbekistan between Tashkent and Sirdaryo train stations. Various RAM analysis techniques are employed to assess track geometry and different railway track components, aiming to achieve reliability, availability, and maintainability.

METHODOLOGY

TOPIC MODELLING

Topic modeling is a powerful technique used to analyze large volumes of text data and extract meaningful insights without the need for manual reading. At its core, topic modeling aims to uncover common themes or topics present within a corpus of text documents. Imagine each document as a mixture of different topics, with each topic represented by a set of words. The goal of topic modeling algorithms like Latent Dirichlet Allocation (LDA) and Latent Semantic Analysis (LSA) is to uncover these underlying topics based on the words found in the documents. This process involves several key steps, including text preprocessing to clean and standardize the data, choosing an appropriate topic modeling algorithm, training the model on the text data to identify topics, and interpreting the results to understand the main themes present in the data.

The applications of topic modeling are diverse and span across various domains. For instance, in content organization, topic modeling can help organize large collections of documents such as news articles or research papers into different categories based on their content. This makes it easier for users to navigate through the data and find relevant information. Moreover, topic modeling can be used for insight generation in business settings. By uncovering hidden themes within customer feedback or market data, businesses can gain valuable insights into customer preferences, market trends, and public sentiment. This, in turn, enables them to make data-driven decisions and tailor their products or services to better meet the needs of their customers.

Furthermore, topic modeling has significant applications in knowledge discovery and scientific research. Researchers can leverage topic modeling techniques to sift through vast amounts of scientific literature and identify relevant studies on specific topics or areas of interest. This helps researchers stay up-to-date with the latest developments in their field and make novel discoveries by uncovering connections between disparate pieces of information. Overall, topic modeling serves as a valuable tool for unlocking the wealth of knowledge hidden within text data and enables us to make informed decisions across various domains in today's data-driven world. Whether it's analyzing customer feedback, categorizing news articles, or advancing scientific research, topic modeling empowers us to extract meaningful insights and derive actionable intelligence from text data.

Topic modeling is a transformative method revolutionizing the analysis of textual data. It offers an automated approach to uncovering latent themes or topics within large corpora, providing invaluable insights into the underlying structure and content of documents. At its essence, topic modeling treats each document as a mixture of various topics, with each topic characterized by a distribution of words. By employing sophisticated algorithms such as Latent Dirichlet Allocation (LDA) or Latent Semantic Analysis (LSA), topic modeling extracts these latent topics from the text data, facilitating a deeper understanding of the underlying themes without the need for exhaustive manual reading.

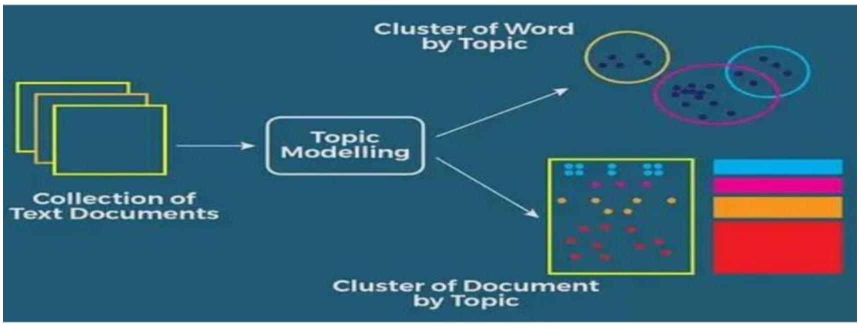


Fig.1. Topic modeling

The process of topic modeling entails several crucial steps, beginning with text preprocessing to cleanse and prepare the data for analysis. This involves tasks like tokenization, removal of stopwords, and stemming or lemmatization to standardize the text. Subsequently, an appropriate topic modeling algorithm is selected based on the nature of the data and the desired outcome. The chosen algorithm is then trained on the preprocessed text data to identify the underlying topics, with the model learning the probability distribution of words for each topic and the probability distribution of topics for each document. Finally, the results are interpreted to discern the main themes present in the corpus, allowing for meaningful insights to be extracted and actionable intelligence to be derived.

DATA SET

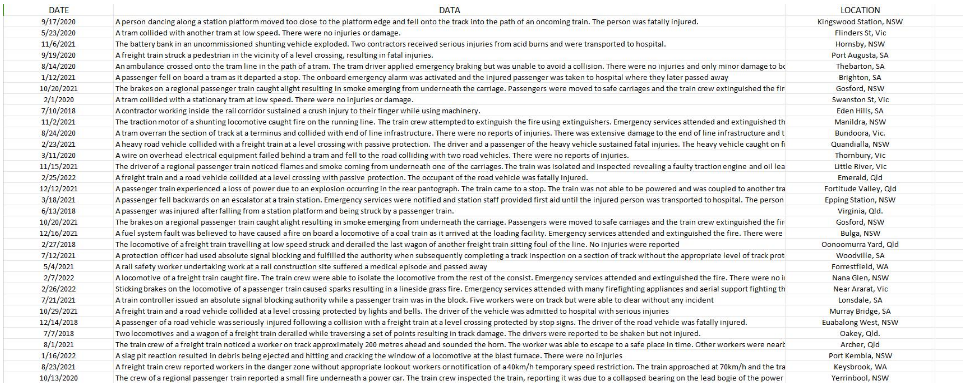
A railway dataset is a structured collection of information related to railway operations, safety incidents, maintenance activities, passenger feedback, and other relevant aspects of railway infrastructure and services. It typically contains a variety of data types, including textual reports, numerical data, images, and possibly sensor data or geospatial information.

The quality and relevance of a railway dataset are crucial for ensuring the effectiveness of analysis and decision-making within the railway industry. A comprehensive dataset should encompass a wide range of factors that impact railway safety and performance. The dataset should also be properly labeled and cleaned to ensure that the data is usable and consistent.

The data set we have chosen to use is collected from the source ONRSR data set, this data is mainly used for research purposes

https://www.onrsr.com.au/publications/corporate-publications/rail-safety-report

Based on the data we have it is displayed in this way



DATE	DATA	LOCATION
9/17/2020	A person dancing along a station platform moved too close to the platform edge and fell onto the track into the path of an oncoming train. The person was fatally injured.	Kingswood Station, NSW
5/23/2020	A tram collided with another tram at low speed. There were no injuries or damage.	Flinders St, Vic
11/6/2021	The battery bank in an uncommissioned shunting vehicle exploded. Two contractors received serious injuries from acid burns and were transported to hospital.	Hornsby, NSW
9/19/2020	A freight train struck a pedestrian in the vicinity of a level crossing, resulting in fatal injuries.	Port Augusta, SA
8/14/2020	An ambulance crossed onto the tram line in the path of a tram. The tram driver applied emergency braking but was unable to avoid a collision. There were no injuries and only minor damage to bc	Thebarton, SA
1/12/2021	A passenger fell on board a tram as it departed a stop. The onboard emergency alarm was activated and the injured passenger was taken to hospital where they later passed away	Brighton, SA
10/20/2021	The brakes on a regional passenger train caught alight resulting in smoke emerging from underneath the carriage. Passengers were moved to safe carriages and the train crew extinguished the fir	Gosford, NSW
2/1/2020	A tram collided with a stationary tram at low speed. There were no injuries or damage.	Swanston St, Vic
7/10/2018	A contractor working inside the rail corridor sustained a crush injury to their finger while using machinery.	Eden Hills, SA
11/2/2021	The traction motor of a shunting locomotive caught fire on the running line. The train crew attempted to extinguish the fire using extinguishers. Emergency services attended and extinguished th	Manildra, NSW
8/24/2020	A tram overran the section of track at a terminus and collided with end of line infrastructure. There were no reports of injuries. There was extensive damage to the end of line infrastructure and t	Bundoora, Vic.
2/23/2021	A heavy road vehicle collided with a freight train at a level crossing with passive protection. The driver and a passenger of the heavy vehicle sustained fatal injuries. The heavy vehicle caught on fi	Quandialla, NSW
3/11/2020	A wire on overhead electrical equipment failed behind a tram and fell to the road colliding with two road vehicles. There were no reports of injuries.	Thornbury, Vic
11/15/2021	The driver of a regional passenger train noticed flames and smoke coming from underneath one of the carriages. The train was isolated and inspected revealing a faulty traction engine and oil lea	Little River, Vic
2/25/2022	A freight train and a road vehicle collided at a level crossing with passive protection. The occupant of the road vehicle was fatally injured.	Emerald, Qld
12/12/2021	A passenger train experienced a loss of power due to an explosion occurring in the rear pantograph. The train came to a stop. The train was not able to be powered and was coupled to another tra	Fortitude Valley, Qld
3/18/2021	A passenger fell backwards on an escalator at a train station. Emergency services were notified and station staff provided first aid until the injured person was transported to hospital. The person	Epping Station, NSW
6/13/2018	A passenger was injured after falling from a station platform and being struck by a passenger train.	Virginia, Qld.
10/20/2021	The brakes on a regional passenger train caught alight resulting in smoke emerging from underneath the carriage. Passengers were moved to safe carriages and the train crew extinguished the fir	Gosford, NSW
12/16/2021	A fuel system fault was believed to have caused a fire on board a locomotive of a coal train as it arrived at the loading facility. Emergency services attended and extinguished the fire. There were	Bulga, NSW
2/27/2018	The locomotive of a freight train travelling at low speed struck and derailed the last wagon of another freight train sitting foul of the line. No injuries were reported	Oonoomurra Yard, Qld
7/12/2021	A protection officer had used absolute signal blocking and fulfilled the authority when subsequently completing a track inspection on a section of track without the appropriate level of track prot	Woodville, SA
5/4/2021	A rail safety worker undertaking work at a rail construction site suffered a medical episode and passed away	Forrestfield, WA
2/7/2022	A locomotive of a freight train caught fire. The train crew were able to isolate the locomotive from the rest of the consist. Emergency services attended and extinguished the fire. There were no i	Nana Glen, NSW
2/26/2022	Sticking brakes on the locomotive of a passenger train caused sparks resulting in a lineside grass fire. Emergency services attended with many firefighting appliances and aerial support fighting th	Near Ararat, Vic
7/21/2021	A train controller issued an absolute signal blocking authority while a passenger train was in the block. Five workers were on track but were able to clear without any incident	Lonsdale, SA
10/29/2021	A freight train and a road vehicle collided at a level crossing protected by lights and bells. The driver of the vehicle was admitted to hospital with serious injuries	Murray Bridge, SA
12/14/2018	A passenger of a road vehicle was seriously injured following a collision with a freight train at a level crossing protected by stop signs. The driver of the road vehicle was fatally injured.	Euabalong West, NSW
7/7/2018	Two locomotives and a wagon of a freight train derailed while traversing a set of points resulting in track damage. The drivers were reported to be shaken but not injured.	Oakey, Qld.
8/1/2021	The train crew of a freight train noticed a worker on track approximately 200 metres ahead and sounded the horn. The worker was able to escape to a safe place in time. Other workers were neart	Archer, Qld
1/16/2022	A slag pit reaction resulted in debris being ejected and hitting and cracking the window of a locomotive at the blast furnace. There were no injuries	Port Kembla, NSW
8/23/2021	A freight train crew reported workers in the danger zone without appropriate lookout workers or notification of a 40km/h temporary speed restriction. The train approached at 70km/h and the tra	Keysbrook, WA
10/13/2020	The crew of a regional passenger train reported a small fire underneath a power car. The train crew inspected the train, reporting it was due to a collapsed bearing on the lead bogie of the power	Yerrinbool, NSW

Fig.2. Data set

The above table provides the information about the railway accidents DataSet,The Railway Safety Incident Dataset is a comprehensive collection of reports detailing accidents, safety issues, and close calls within the railway sector. Each entry in the dataset includes the date of the incident, a description of what happened, and where it occurred. Covering incidents from different parts of Australia over several years, this dataset provides valuable information for understanding trends and challenges in railway safety.

It covers a wide range of incidents, from minor mishaps to more serious accidents, giving us insight into the various risks and issues faced by railway operators. With this dataset, we can analyze patterns, identify common causes of incidents, and work towards improving safety measures in the railway industry.

EXISTING SYSTEM

In the realm of machine learning for risk assessment and safety accident analysis, the utilization of Decision Trees, Random Forests, Support Vector Machines, and Artificial Neural Networks plays a crucial role in enhancing safety protocols and preventing potential hazards. Each of these algorithms brings unique advantages and challenges to the table, impacting their suitability for specific applications within the railway safety domain.

Decision Trees offer a straightforward approach to modeling decision-making processes based on input features, making them particularly useful for interpreting and understanding the logic behind classification outcomes. However, their susceptibility to overfitting and instability with small variations in the data can hinder their effectiveness in accurately capturing complex relationships within railway incident reports.

Random Forests address the limitations of Decision Trees by aggregating multiple trees to improve predictive accuracy and robustness. While Random Forests mitigate the overfitting issue, they introduce computational complexity and reduced interpretability due to the ensemble nature of the model. Balancing the trade-off between model complexity and performance becomes crucial in railway safety applications where transparency and efficiency are paramount.

Support Vector Machines excel in classifying data by identifying optimal decision boundaries in high-dimensional spaces, making them well-suited for scenarios with complex feature interactions. However, SVMs may encounter challenges in scalability and kernel selection, requiring careful parameter tuning and computational resources to achieve optimal performance.

Artificial Neural Networks offer unparalleled flexibility in capturing intricate patterns and nonlinear relationships within text data, making them ideal for processing unstructured incident reports. Nevertheless, their resource-intensive nature and susceptibility to overfitting demand rigorous validation and regularization techniques to ensure reliable performance in real-world railway safety applications.

PROPOSED SYSTEM

In this project we proposed an innovative approach to analyze the Railway Accidents by Using the Latent Dirichlet Allocation (LDA) stands out as a powerful tool for unraveling the intricacies hidden within the unstructured text data of railway safety incident reports. LDA brings a host of advantages to the table, making it a prime candidate for our analytical needs.

First and foremost, LDA excels at uncovering latent topics or themes buried within a corpus of documents. By scrutinizing the distribution of words across these documents, LDA adeptly discerns underlying patterns, revealing common issues or occurrences prevalent in railway safety incidents. This capability offers invaluable insights into the fundamental causes and trends underpinning these incidents, guiding us towards more informed decision-making and preventive measures.

Moreover, LDA facilitates dimensionality reduction by representing each document as a distribution over topics. This transformation simplifies the complexity of the data while retaining crucial semantic information. Consequently, LDA streamlines the analysis and visualization of extensive text data, empowering stakeholders to grasp the underlying structure more effectively.

Another compelling advantage of LDA lies in its interpretability. By generating topics as distributions over words, LDA renders them easily understandable to domain experts. Each topic encapsulates a collection of words frequently associated with a particular theme or concept, enabling stakeholders to pinpoint common issues, trends, and patterns within railway safety incidents. This interpretability fosters targeted interventions to mitigate risks and enhance safety measures.Additionally, LDA demonstrates commendable scalability, capable of efficiently handling large datasets. With appropriate implementation strategies, LDA processes extensive document collections within reasonable timeframes. This scalability is particularly advantageous given the voluminous nature of text data typically encountered in railway safety incident reports.

Furthermore, LDA seamlessly integrates with other machine learning techniques and algorithms, bolstering the analytical capabilities of our proposed system. For instance, LDA-derived topic distributions serve as valuable features for downstream tasks such as classification, clustering, or anomaly detection, enriching the predictive prowess of the system.

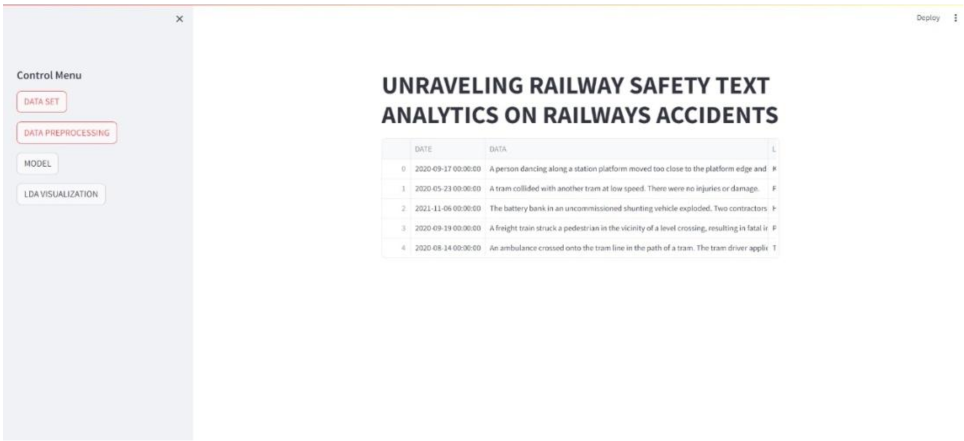
RESULT

EVALUATION CRITERIA

In our project focused on enhancing railway safety through text analytics on accident reports, we embark on a systematic approach to uncover valuable insights. Using advanced techniques, we analyze accident reports to identify underlying patterns and contributing factors. We employ methods like Latent Dirichlet Allocation (LDA) for topic modeling, enabling us to discern latent topics within the reports. By leveraging unsupervised learning algorithms, we group similar incidents and detect anomalies, facilitating proactive safety measures. Our system architecture integrates various components, ensuring efficient processing of accident data.

Through stakeholder engagement and continuous improvement, we aim to translate our findings into actionable recommendations, ultimately contributing to a safer railway environment.

The implementation of the System will be explained clearly with the help of the continuous screenshots. Here, in the first step of our project is to collect Data, the collected Data will shown on the User Interface(UI). The first five Data set samples will be shown.



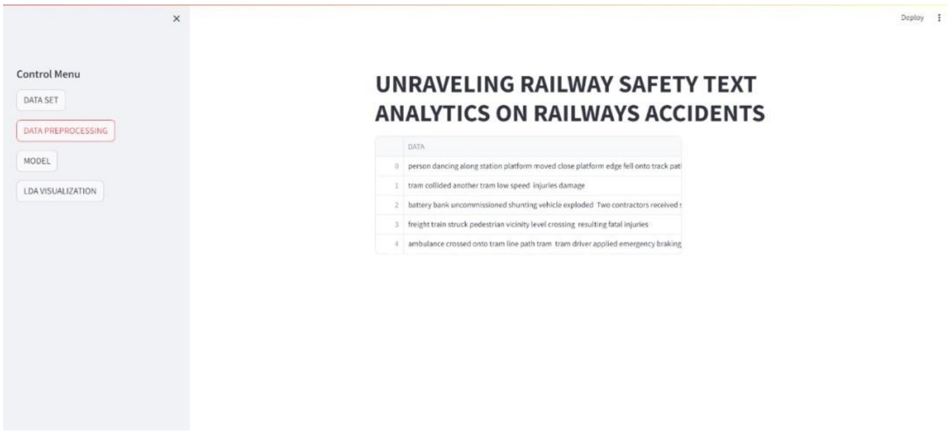
UNRAVELING RAILWAY SAFETY TEXT ANALYTICS ON RAILWAYS ACCIDENTS

	DATE	DATA	L
0	2020-09-17 00:00:00	A person dancing along a station platform moved too close to the platform edge and	K
1	2020-05-23 00:00:00	A tram collided with another tram at low speed. There were no injuries or damage.	F
2	2021-11-06 00:00:00	The battery bank in an uncommissioned shunting vehicle exploded. Two contractors	H
3	2020-09-19 00:00:00	A freight train struck a pedestrian in the vicinity of a level crossing, resulting in fatal ir	F
4	2020-08-14 00:00:00	An ambulance crossed onto the tram line in the path of a tram. The tram driver applic	T

Fig.3. Data set

Description: Fig 3 The dataset for railway accidents encompasses various parameters such as time, date, location, and other relevant factors. These parameters provide crucial insights into the circumstances surrounding each incident, enabling a comprehensive analysis of railway safety. By examining the time and date of accidents, we can identify patterns and trends in accident occurrence, including peak times or seasons with higher incident rates. The location data allows us to pinpoint areas with a higher frequency of accidents, highlighting potential safety hazards or infrastructure issues that require attention. Additionally, other parameters such as weather conditions, train speed, and type of incident provide further context for understanding the causes and mitigating risks associated with railway accidents. Overall, the dataset's comprehensive coverage of diverse parameters enables a thorough examination of railway safety and informs evidence-based strategies for accident prevention and response.

After the Data collection the next step is to Pre-process the data i.e Cleaning of Data, removing of Unwanted data.

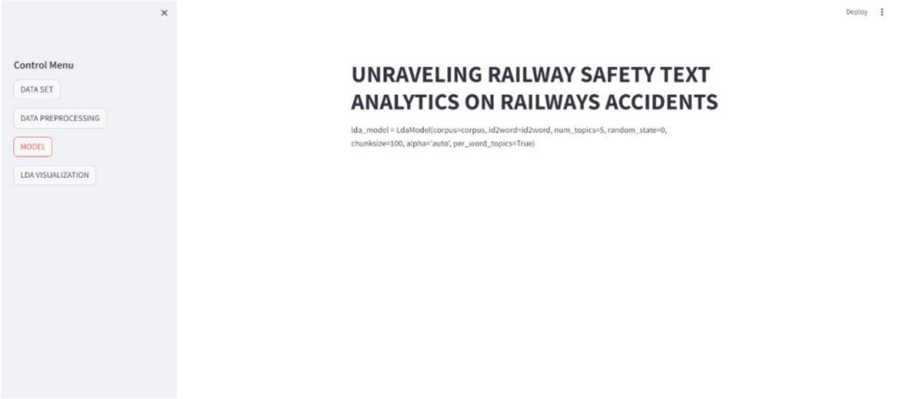


UNRAVELING RAILWAY SAFETY TEXT ANALYTICS ON RAILWAYS ACCIDENTS

	DATA
0	person dancing along station platform moved close platform edge fell onto track pat
1	tram collided another tram low speed injuries damage
2	battery bank uncommissioned shunting vehicle exploded Two contractors received s
3	freight train struck pedestrian vicinity level crossing resulting fatal injuries
4	ambulance crossed onto tram line path tram tram driver applied emergency braking

Fig 4 Pre-processing

Description: Figure 4 depicts the pre-processing stage within our text analytics framework dedicated to enhancing railway safety. This critical phase is pivotal in transforming raw accident reports into a structured format primed for analysis. Initially, data collection sources raw accident reports from diverse channels, including railway authorities and safety regulators. Subsequently, the gathered text undergoes meticulous cleaning and standardization, purging it of noise, irrelevant details, and formatting discrepancies. The cleaned text is then segmented into tokens through tokenization, streamlining subsequent analysis. Common yet non-essential words, termed stopwords, are eliminated to mitigate data clutter. Additionally, stemming or lemmatization techniques are applied to normalize words, enhancing consistency across the dataset. Following this, the text is vectorized, translating it into numerical representations conducive to quantitative analysis, with relevant features such as time, date, location, and incident specifics meticulously selected for inclusion. This preparatory phase lays the foundation for subsequent analyses, including topic modeling, clustering, and anomaly detection, essential for deriving actionable insights to bolster railway safety.



UNRAVELING RAILWAY SAFETY TEXT ANALYTICS ON RAILWAYS ACCIDENTS

```
lda_model = LdaModel(corpus=corpus, id2word=id2word, num_topics=5, random_state=0,
chunksize=100, alpha='auto', per_word_topics=True)
```

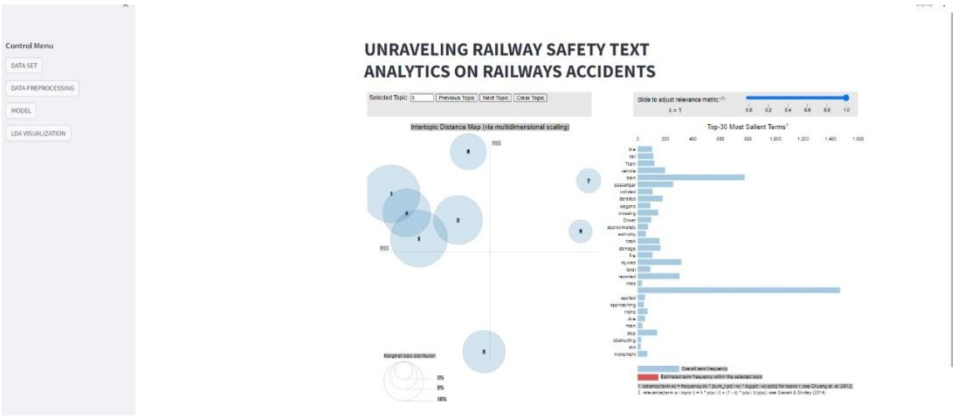
Fig 5 LDA model

Description :

In Figure 5, we present the model of Latent Dirichlet Allocation (LDA), a pivotal component within our text analytics framework aimed at bolstering railway safety. LDA is a powerful probabilistic model used for topic modeling, enabling us to discern latent topics within accident reports. The LDA model operates on the principle that each document in the dataset is a mixture of various topics, with each topic characterized by a distribution of words.

The graphical representation showcases the interconnectedness between documents, topics, and words. Documents are depicted as circles, with arrows indicating their probabilistic connections to different topics represented as rectangles. Each topic, in turn, exhibits associations with specific words, illustrated by lines connecting topics to individual words.

By leveraging the LDA model, we can uncover hidden themes and patterns within accident reports, facilitating a deeper understanding of the underlying causes of railway accidents. These insights empower stakeholders to implement targeted interventions and preventive measures, ultimately fostering a safer railway environment for all.



.

Fig 6 Intertopic Distance map

Description:

The intertopic distance map illustrates the proximity or similarity between topics in multidimensional space. Each topic is represented as a point on the map, and the distance between points reflects the degree of similarity between topics. Topics that are closely related or share common themes are positioned closer together, while topics with distinct characteristics are further apart.

By examining the intertopic distance map, stakeholders can gain valuable insights into the thematic structure of the accident reports. Patterns of topic clustering and dispersion can highlight recurring themes, dominant topics, and areas of divergence within the dataset. This visualization aids in the interpretation and exploration of the results obtained from the LDA analysis, facilitating a deeper understanding of the underlying factors contributing to railway accidents

CONCLUSION

In this project our research emphasizes the pivotal role of AI-driven technologies, particularly unsupervised topic modeling using LDA, in advancing railway safety through comprehensive accident analysis. By extracting valuable insights from accident reports, we've identified critical factors contributing to severe accidents, empowering proactive risk mitigation strategies. The integration of intelligent text analysis into safety protocols not only enhances accuracy but also lays the groundwork for innovative AI applications across industries. Our study marks a significant step forward in leveraging technology to create safer transportation systems and underscores the potential of AI to revolutionize safety management practices globally.

FUTURE ENHANCEMENTS

Enhanced Visualization:

- Collaborating with data visualization experts can lead to the creation of visually appealing and intuitive visualizations that effectively communicate insights from the accident data.
- Interactive visualizations, such as heatmaps, network graphs, or time-series plots, can allow users to explore accident trends, identify patterns, and gain deeper insights into safety issues.
- Implementing features like zooming, filtering, and drill-down capabilities can enhance the user experience and enable stakeholders to interact with the data in meaningful ways.

Feedback Mechanism:

- Designing and implementing a feedback mechanism within the system enables users to provide input on the relevance, accuracy, and usability of the topic models generated by the system.
- Feedback can be collected through various channels, such as surveys, feedback forms, or direct user interactions within the application interface.
- Analyzing user feedback and iteratively refining the topic models based on real-world usage can lead to continuous improvement and optimization of the system over time.
- Additionally, integrating machine learning techniques, such as sentiment analysis or text clustering, can automatically analyze and categorize user feedback, allowing for more efficient processing and action.

References

[1] S. Terabe, T. Kato, H. Yaginuma, N. Kang, and K. Tanaka, ‘‘Risk assessment model for railway passengers on a crowded platform,’’ , Jan. 2019.

[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.

[4] M. Gethers and D. Poshyvanyk, ‘‘Using relational topic models to capture coupling among classes in object-oriented software systems 2010.

[5] D. M. Blei, A. Y. Ng, and M. I. Jordan, ‘‘Latent Dirichlet allocation,’’ Mar. 2003,

[6] H. Alawad, S. Kaewunruen, and M. An, "A deep learning approach towards railway safety risk assessment,’’ 2020.

[7] H. Alawad, S. Kaewunruen, and M. An, ‘‘Learning from accidents: Machine learning for safety at railway stations,’’ 2020

[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,’’Feb 2016.

[9] J. Sido and M. Konopik, ‘‘Deep learning for text data on mobile devices,’’ in Proc. Int. Conf. Appl. Electron., Sep. 2019.

[10] A. Serna and S. Gasparovic, ‘‘Transport analysis approach based on big data and text mining analysis from social media,’’jan 2018.

[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,’’Dec 2018.

[12] A. Chanen, ‘‘Deep learning for extracting word-level meaning from safety report narratives,’’ in Proc. Integr. Commun. Navigat. Surveill. (ICNS), Apr. 2016.

[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,’’Dec 2018.