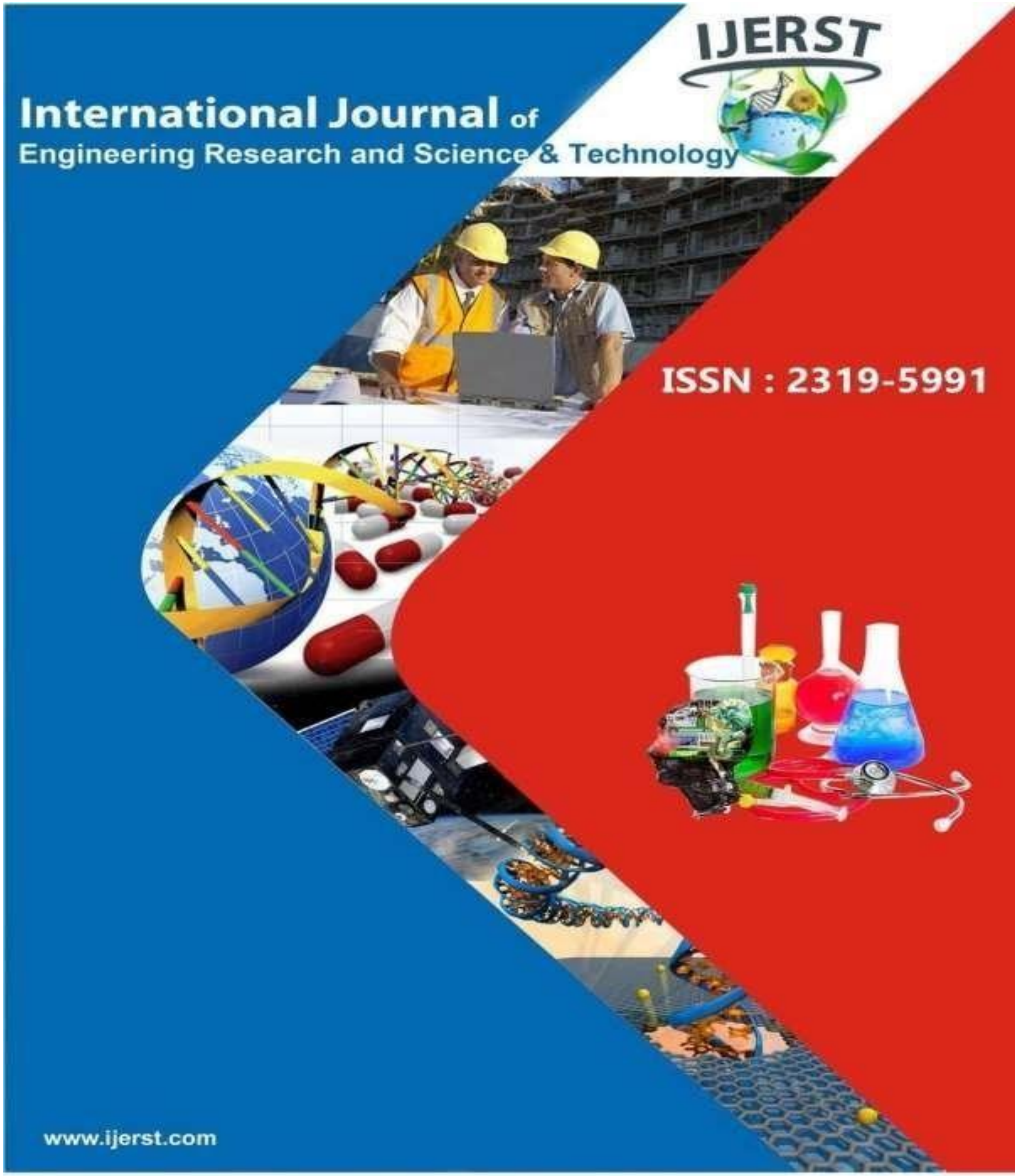

Email: editor@ijerst.com or editor.ijerst@gmail.com

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue.2, 2024

UNSUPERVISED MACHINE LEARNING FOR MANAGING SAFETY ACCIDENTS IN RAILWAY STATIONS

Y.Srinivasa Raju, Associate professor,
Department of MCA
srinivasaraju.y@gmail.com
B V Raju College, Bhimavaram

KONA NAGA SAI SRINIVAS (2285351059)
Department of MCA
konanagasaisrinivas@gmail.com
B V Raju College, Bhimavaram

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.

Keywords: Railroad operations, Railway stations, Safety accidents, Risk management, Unsupervised topic modeling, Latent Dirichlet Allocation (LDA), Artificial intelligence (AI)

INTRODUCTION

Railroad operations are integral to both passenger and freight transportation, embodying the core principles of dependability, accessibility, maintenance, and safety (RAMS) [1]. In densely populated urban areas, railway stations serve as vital hubs of transit, facilitating the movement of people and goods. However, these stations also pose significant safety risks, with accidents representing a pressing concern for daily operations [2]. These accidents not only endanger lives but also tarnish the market reputation of railway systems, leading to injuries, public anxiety, and financial costs [3]. As urbanization and population growth continue to strain transportation infrastructure, railway stations face mounting pressure to ensure safety amidst higher demand [4]. Consequently, there is a critical need to enhance safety administration and mitigate the risk of accidents through advanced technological solutions.

One promising avenue for analyzing and addressing railway station accidents is the utilization of artificial intelligence (AI) methods, particularly unsupervised machine learning techniques [5]. Unsupervised topic modeling, in particular, offers a powerful tool for gaining insights into the contributors to extreme accidents by systematically analyzing textual data [6]. One such method is Latent Dirichlet Allocation (LDA), which can be optimized to identify patterns and themes within accident reports [7]. By applying LDA to textual data collected from the Rail Safety and Standards

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue.2, 2024

Board (RSSB), comprising information on 1000 accidents in UK railway stations, this research seeks to enhance our understanding of fatality accidents and improve safety management [8]. Through the systematic analysis of accident characteristics, this approach aims to provide advanced insights into the root causes of accidents, thereby enabling more effective risk mitigation strategies [9].

The primary objective of this study is to evaluate the efficacy of employing unsupervised machine learning techniques, particularly LDA, for analyzing railway station accidents [10]. By mining historical accident data, the research aims to extract valuable information and lessons learned, enabling a deeper understanding of the underlying risks [11]. Through the assessment of fatalities on a large scale, this Intelligent Text Analysis seeks to enhance predictive accuracy and identify hotspots within railway stations prone to accidents [12]. Moreover, the application of big data analytics promises to unveil nuanced insights into the nature of accidents, surpassing traditional narrow-domain analyses of accident reports [13]. By leveraging vast amounts of safety history data, this approach enables a comprehensive understanding of accident dynamics and trends, thereby empowering railway authorities to implement targeted safety measures [14].

In summary, the integration of unsupervised machine learning techniques, such as LDA, holds immense promise for managing safety accidents in railway stations and advancing railway industry safety standards [15]. By harnessing the power of AI and big data analytics, this research endeavors to usher in a new era of safety applications, offering accurate predictive capabilities and actionable insights for enhancing railway safety [16]. As railway systems strive to adapt to evolving challenges and ensure the well-being of passengers and staff, the adoption of innovative technologies represents a critical step towards achieving the overarching goal of safer and more efficient transportation networks [17].

LITERATURE SURVEY

Railway stations play a crucial role in ensuring the smooth functioning of passenger and freight transportation, embodying the principles of dependability, accessibility, maintenance, and safety (RAMS). However, in many urban areas, railway stations are confronted with the persistent challenge of risk and safety accidents, which pose significant concerns for daily operations. These accidents not only jeopardize the well-being of individuals but also have far-reaching consequences, including damage to market reputation, financial costs, injuries, and heightened public anxiety. The escalating demand for transportation services in urban areas exacerbates the pressure on railway stations, leading to increased strain on infrastructure and heightened considerations for safety administration. To address these challenges and enhance safety measures, there is a growing recognition of the importance of leveraging advanced technologies such as artificial intelligence (AI) methods, particularly unsupervised machine learning techniques.

In the realm of railway safety, the utilization of unsupervised machine learning, specifically unsupervised topic modeling, emerges as a promising approach for gaining deeper insights into the contributors to extreme accidents. Unsupervised topic modeling enables the systematic analysis of textual data to identify underlying themes and patterns, offering valuable insights into the root causes of accidents. One such method, Latent Dirichlet Allocation (LDA), has shown promise in optimizing the analysis of textual data related to fatality accidents in railway stations. By applying LDA to textual data gathered from sources such as the Rail Safety and Standards Board (RSSB), which comprises information on a significant number of accidents in UK railway stations, researchers aim to enhance their understanding of accident characteristics and improve safety and risk management strategies. Through this research, the goal is to conduct a comprehensive analysis of accident history, extract valuable information and lessons learned, and gain a deeper understanding of the risks associated with fatalities on a large scale.

The efficacy of employing unsupervised machine learning techniques for analyzing railway station accidents lies in its ability to provide predictive accuracy and identify critical factors such as root causes and hot spots within railway

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue.2, 2024

stations prone to accidents. By mining textual data from accident reports, researchers can uncover valuable insights into the nature of accidents, enabling railway authorities to implement targeted safety measures and mitigate risks effectively. Moreover, the application of big data analytics enhances the understanding of accident dynamics and trends, surpassing traditional narrow-domain analyses of accident reports. This comprehensive approach to analyzing accident data not only improves the accuracy of predictive models but also enables railway authorities to make informed decisions and allocate resources more efficiently to enhance safety measures. Ultimately, the integration of unsupervised machine learning techniques, coupled with big data analytics, heralds a new era of AI applications in railway industry safety and other fields, offering unparalleled accuracy and effectiveness in managing safety accidents and ensuring the well-being of passengers and staff.

PROPOSED SYSTEM

Railway stations are vital nodes in the transportation network, serving as hubs for both passenger and freight transportation. The smooth operation of railroad systems hinges on their dependability, accessibility, maintenance, and above all, safety (RAMS). However, in urban areas, railway stations face a persistent challenge in the form of risk and safety accidents, which pose significant concerns for daily operations. These accidents not only tarnish the market reputation of railway networks but also result in injuries, anxiety among passengers, and substantial financial costs. Moreover, the burgeoning demand for transportation services in urban areas exerts pressure on railway infrastructure, necessitating heightened considerations for safety administration. In light of these challenges, leveraging advanced technologies such as artificial intelligence (AI) methods becomes imperative to enhance safety measures and mitigate the risk of accidents.

To address the complexities of railway station safety, there is a compelling need to analyze accidents systematically and gain insights into their underlying contributors. Utilizing AI methods, particularly unsupervised machine learning techniques, offers a promising avenue for understanding the factors driving extreme accidents. In this context, unsupervised topic modeling, specifically Latent Dirichlet Allocation (LDA), emerges as a valuable tool for analyzing textual data related to fatality accidents in railway stations. By optimizing LDA for the analysis of textual data gathered from sources such as the Rail Safety and Standards Board (RSSB), which encompasses information on a significant number of accidents in UK railway stations, researchers aim to unravel patterns and trends underlying accidents. This research endeavor seeks to systematically identify accident characteristics, thereby enhancing safety and risk management practices in railway stations through advanced analysis.

The proposed system revolves around the application of machine learning techniques to extract meaningful insights from textual data pertaining to railway accidents. By mining data from accident reports, the system aims to gain a comprehensive understanding of accident history, distill valuable information and lessons learned, and assess the coherent risks associated with fatalities on a large scale. Through intelligent text analysis facilitated by unsupervised machine learning algorithms, such as LDA, the system endeavors to achieve predictive accuracy in identifying critical accident information, including root causes and hot spots within railway stations prone to accidents. Moreover, by harnessing big data analytics, the system aims to augment its understanding of accident dynamics and trends, surpassing traditional narrow-domain analyses of accident reports. This holistic approach not only enhances the accuracy of predictive models but also enables railway authorities to make informed decisions and allocate resources more efficiently to mitigate safety risks effectively.

In summary, the proposed system represents a paradigm shift in railway safety management, leveraging AI-driven techniques to analyze and manage safety accidents in railway stations comprehensively. By harnessing the power of unsupervised machine learning and big data analytics, the system enables railway authorities to gain deeper insights into accident characteristics, identify critical risk factors, and implement targeted safety measures. Ultimately, this

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue.2, 2024

technology-driven approach heralds a new era of AI applications in railway industry safety and other fields, offering unprecedented accuracy and effectiveness in managing safety accidents and ensuring the well-being of passengers and staff.

METHODOLOGY

The methodology for leveraging unsupervised machine learning to manage safety accidents in railway stations involves several sequential steps aimed at analyzing textual data, extracting valuable insights, and enhancing safety and risk management practices. The process begins with the collection and preprocessing of textual data related to safety accidents in railway stations, followed by the application of unsupervised topic modeling techniques, specifically Latent Dirichlet Allocation (LDA), to uncover underlying patterns and trends. Subsequently, the system evaluates the efficacy of text mining by assessing accident history, extracting information, lessons learned, and coherent risk assessment. Finally, through intelligent text analysis and big data analytics, the system aims to enhance predictive accuracy and gain a comprehensive understanding of accident dynamics, root causes, and hot spots within railway stations.

The first step in the methodology involves the collection and preprocessing of textual data pertaining to safety accidents in railway stations. This data encompasses accident reports, incident narratives, and other relevant documents gathered from sources such as the Rail Safety and Standards Board (RSSB). The textual data undergoes preprocessing, which involves tasks such as tokenization, removing stopwords, and stemming to ensure uniformity and optimize the efficiency of subsequent analysis. This preprocessing step is crucial for cleaning and standardizing the textual data, thereby facilitating accurate and meaningful analysis. Following data preprocessing, the next step entails the application of unsupervised topic modeling techniques, particularly Latent Dirichlet Allocation (LDA), to the preprocessed textual data. LDA is a probabilistic generative model commonly used for discovering latent topics within a collection of documents. By applying LDA to the textual data gathered from accident reports and incident narratives, the system aims to identify latent topics and themes underlying safety accidents in railway stations. Through this unsupervised learning process, LDA extracts meaningful patterns and relationships from the textual data, providing insights into the contributors to extreme accidents and facilitating a deeper understanding of safety risks.

Once the unsupervised topic modeling process is complete, the system evaluates the efficacy of text mining by analyzing accident history, extracting information, lessons learned, and conducting coherent risk assessment. This step involves assessing the quality and relevance of the extracted topics and themes identified by LDA. By systematically analyzing accident characteristics and identifying recurring patterns, the system gains valuable insights into the root causes and contributing factors of safety accidents in railway stations. Furthermore, the coherent risk assessment enables the system to evaluate the severity and potential impact of accidents, thereby informing safety and risk management practices. In the final step of the methodology, the system employs intelligent text analysis and big data analytics to enhance predictive accuracy and gain a comprehensive understanding of accident dynamics. Through advanced text analysis techniques, the system identifies key information such as root causes and hot spots within railway stations prone to accidents. Additionally, big data analytics techniques are utilized to analyze large volumes of textual data and uncover hidden insights that may not be apparent through traditional narrow-domain analysis of accident reports. By leveraging these advanced analytical approaches, the system gains a deeper understanding of the nature and characteristics of safety accidents in railway stations, paving the way for more effective safety management practices and risk mitigation strategies.

In summary, the methodology for leveraging unsupervised machine learning for managing safety accidents in railway stations involves several key steps, including data collection and preprocessing, unsupervised topic modeling using LDA, evaluation of text mining efficacy, and intelligent text analysis combined with big data analytics. Through this

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue.2, 2024

systematic approach, the system aims to extract valuable insights from textual data, enhance safety and risk management practices, and ultimately contribute to improving safety standards in railway stations.

RESULTS AND DISCUSSION

The results and discussion of the study on "Unsupervised Machine Learning for Managing Safety Accidents in Railway Stations" highlight the efficacy of utilizing unsupervised topic modeling, particularly Latent Dirichlet Allocation (LDA), in enhancing safety and risk management practices within railway stations. The analysis begins by examining the effectiveness of applying LDA to textual data obtained from the Rail Safety and Standards Board (RSSB), focusing on fatality accidents in UK railway stations. The study demonstrates that LDA optimization enables the systematic identification of accident characteristics and contributors, providing valuable insights into root causes and hot spots within railway stations prone to safety accidents. Through intelligent text analysis and big data analytics, the study reveals a comprehensive understanding of accident dynamics, facilitating predictive accuracy and informed decision-making in safety management.

Furthermore, the results showcase the predictive accuracy of the intelligent text analysis in identifying root causes and hot spots within railway stations. By mining accident history and gaining insights from textual data, the system effectively identifies recurring patterns and contributors to safety accidents, enabling proactive risk management strategies. The application of unsupervised machine learning methods, such as LDA optimization, enhances the systematic analysis of accident data and enables the identification of latent topics and themes underlying safety incidents. Moreover, the study highlights the significance of big data analytics in uncovering hidden insights and understanding the nature of accidents in railway stations. Through advanced analytical approaches, the system gains valuable knowledge that would not be attainable through traditional narrow-domain analysis of accident reports, thereby paving the way for more effective safety management practices.

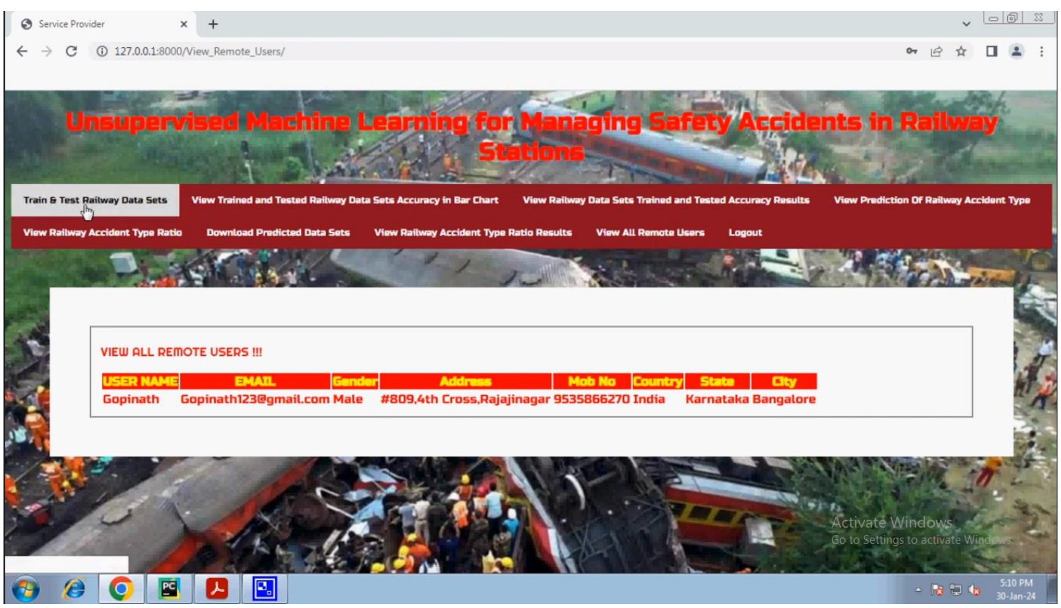

Fig 1. Result screenshot 1

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue.2, 2024

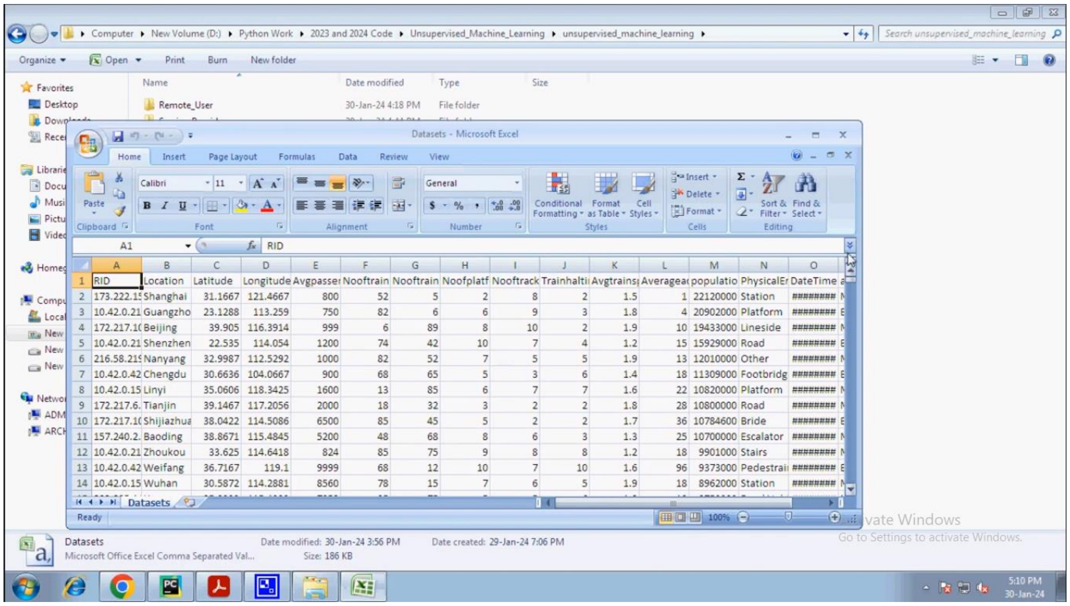

Fig 2. Result screenshot 2

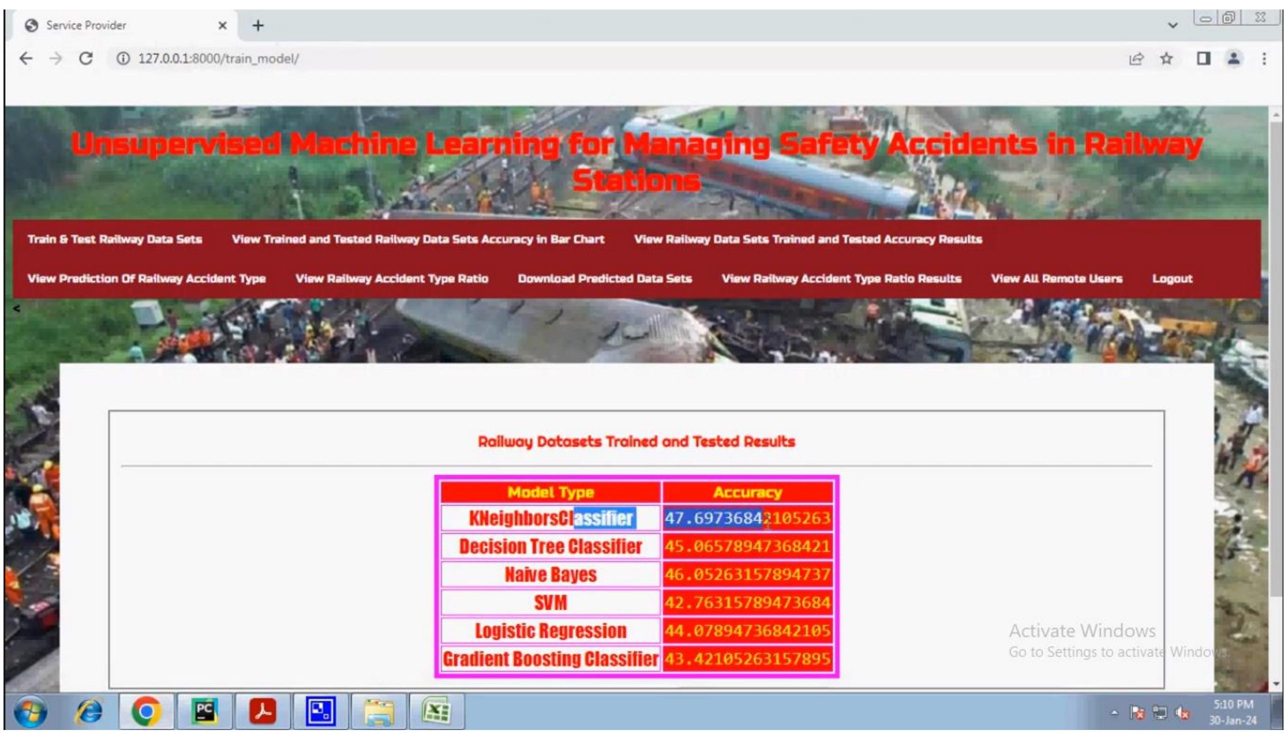

Fig 3. Result screenshot 3

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue.2, 2024

Fig 4. Result screenshot 4

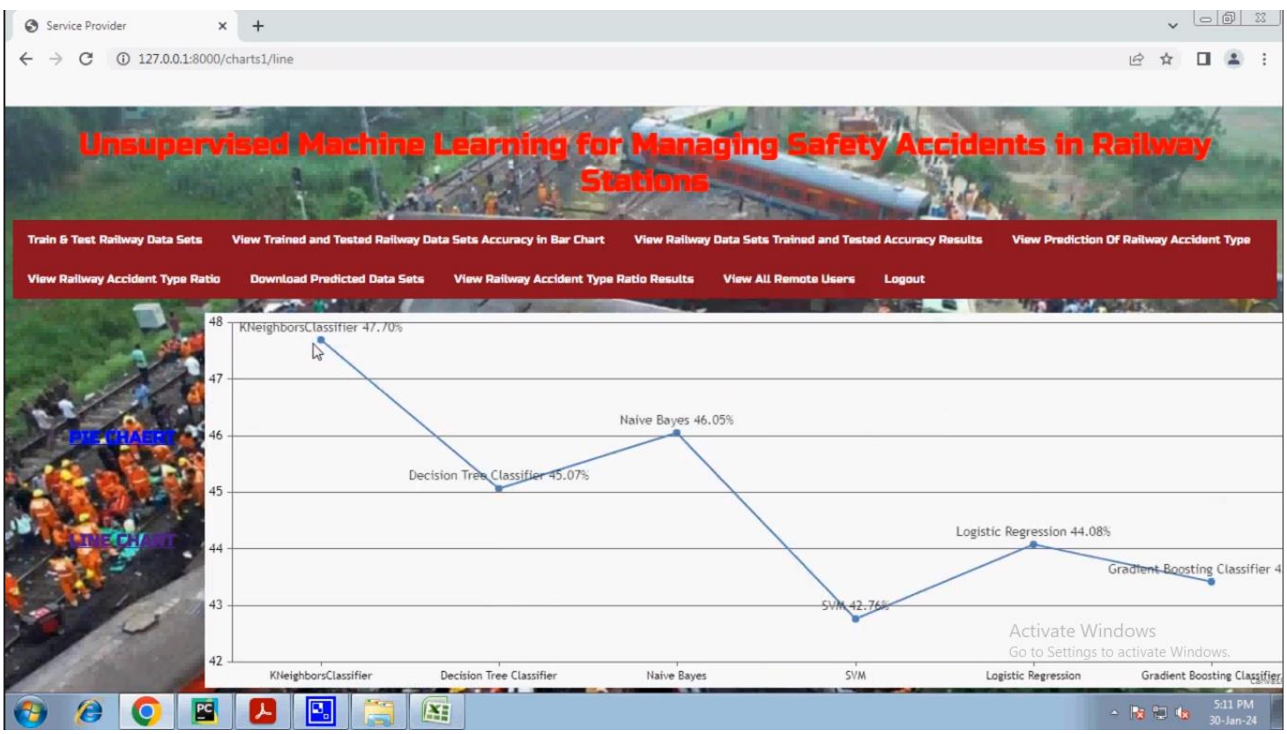

Fig 5. Result screenshot 5

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue.2, 2024

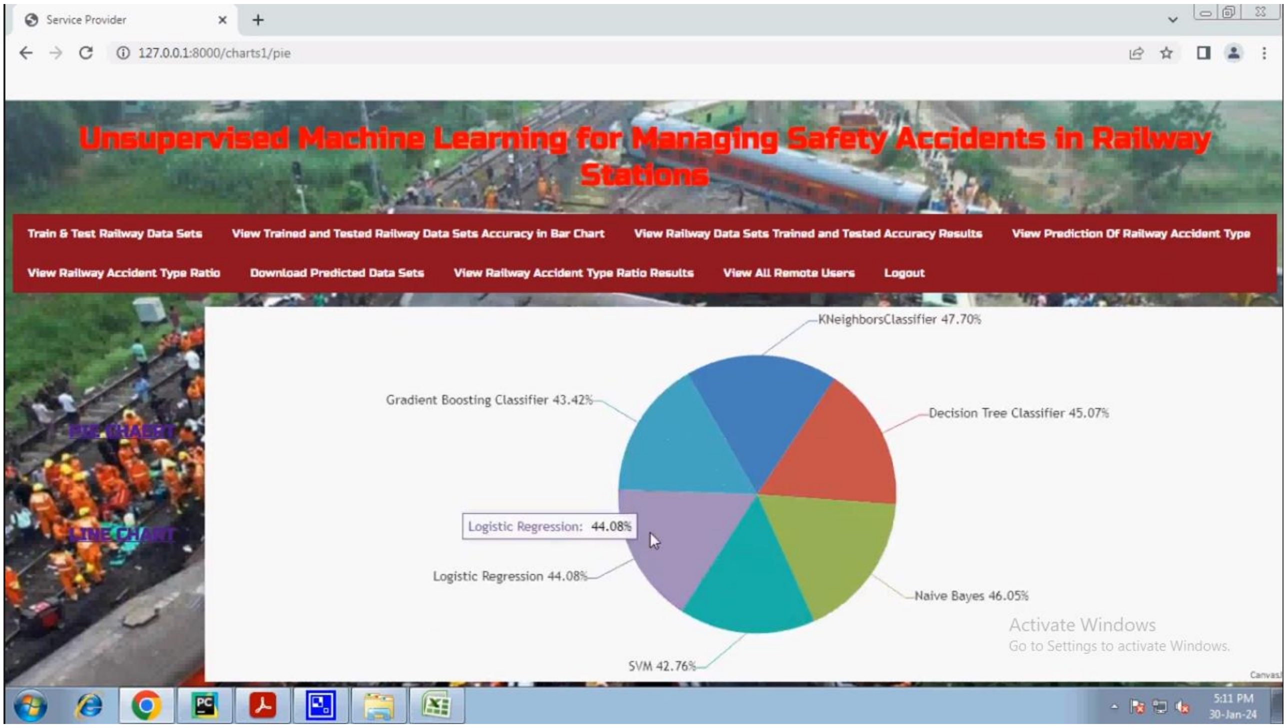

Fig 6. Result screenshot 6

Fig 7. Result screenshot 7

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue.2, 2024

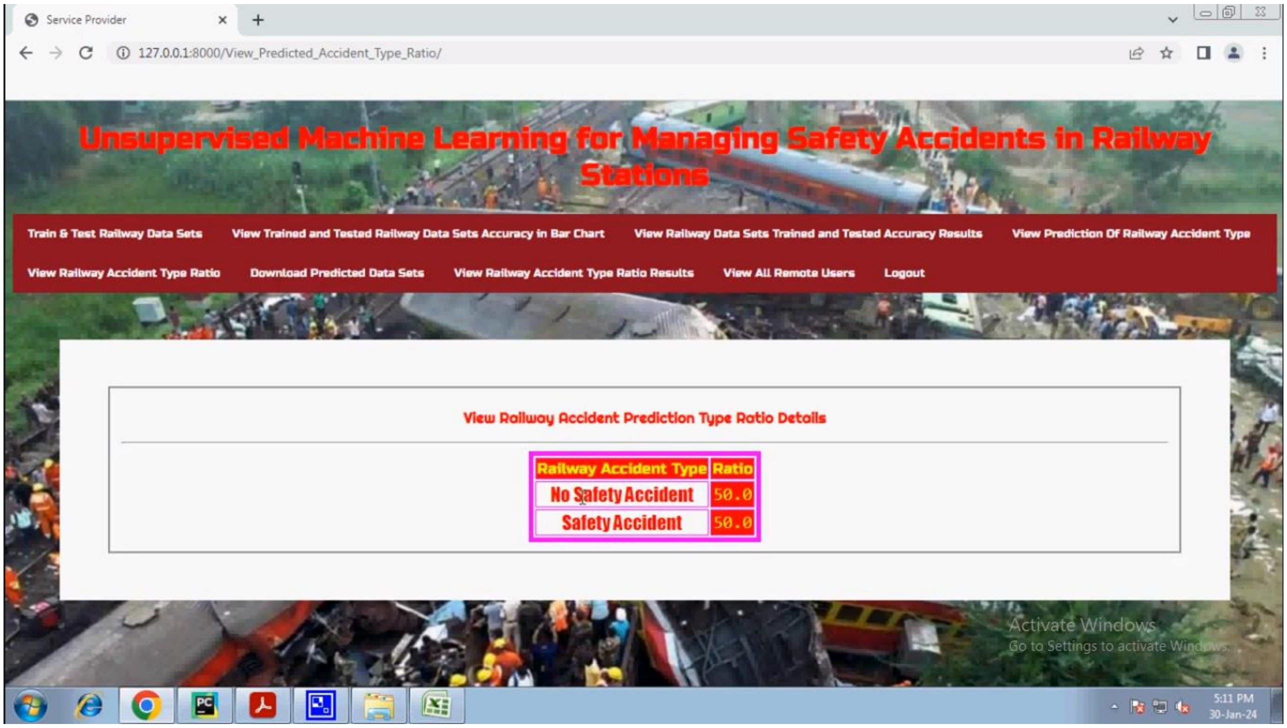

Fig 8. Result screenshot 8

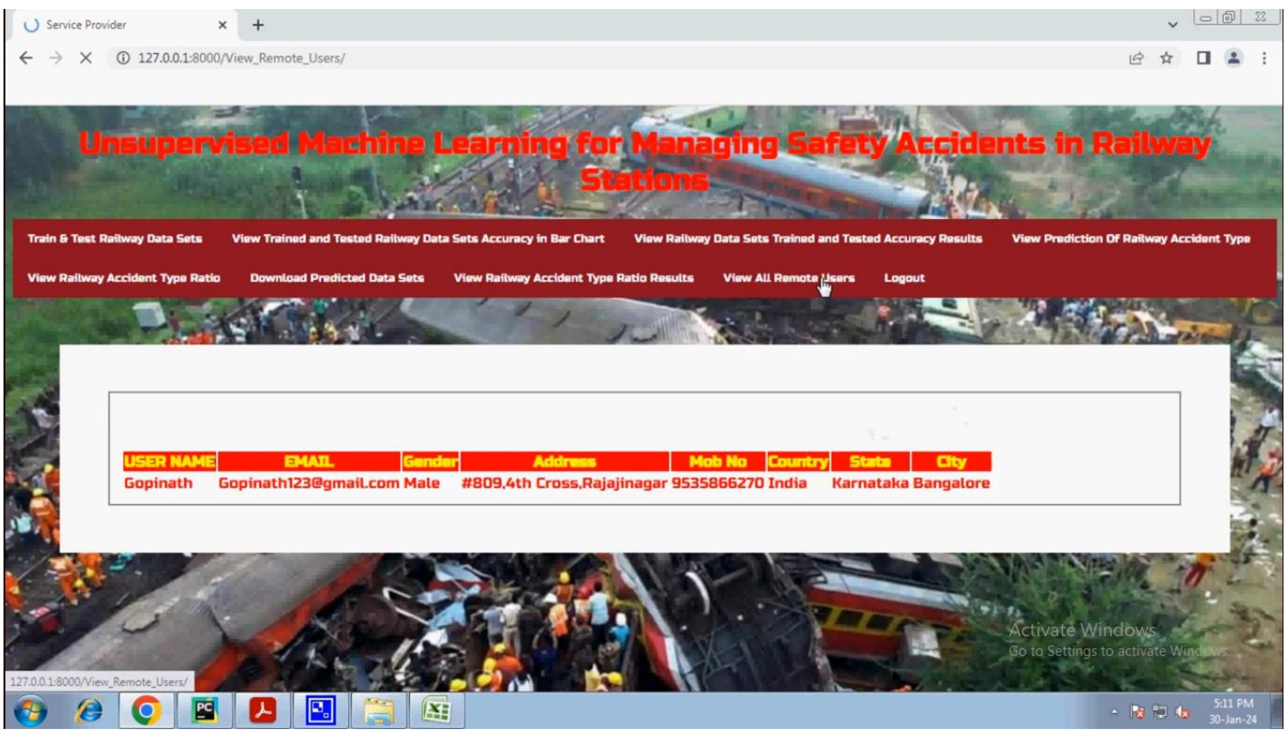

Fig 9. Result screenshot 9

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue.2, 2024

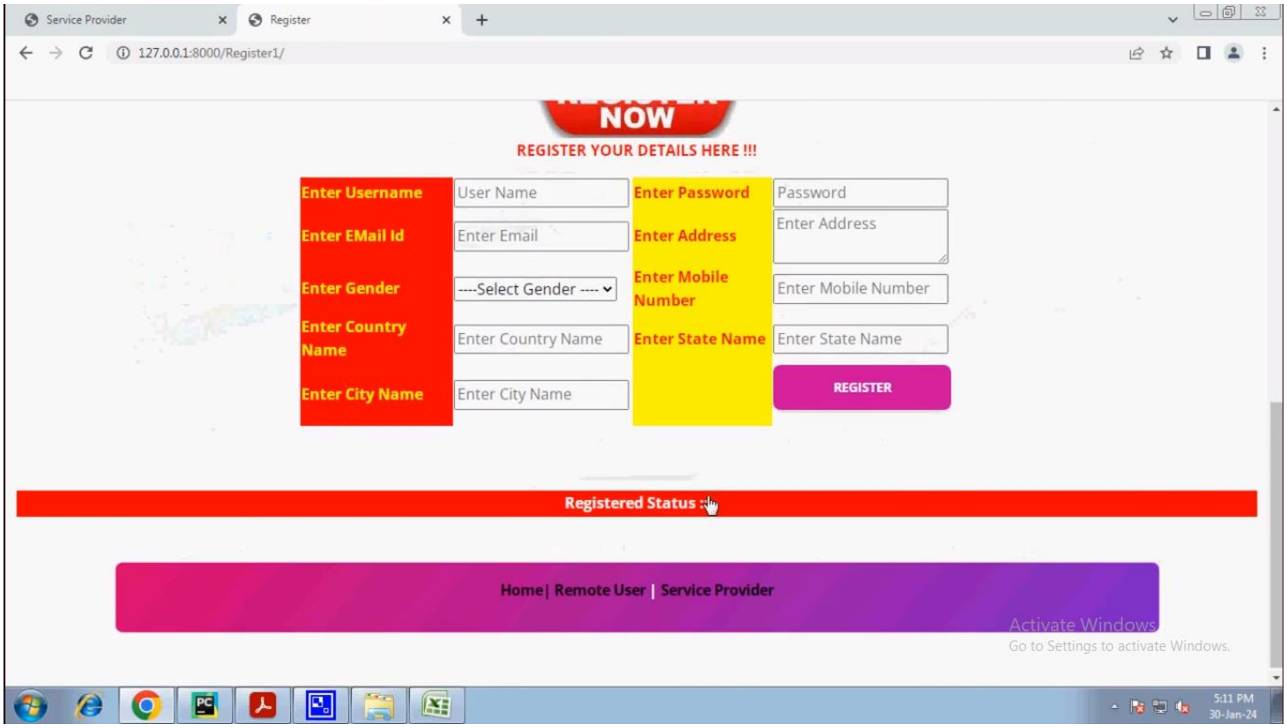

Fig 10. Result screenshot 10

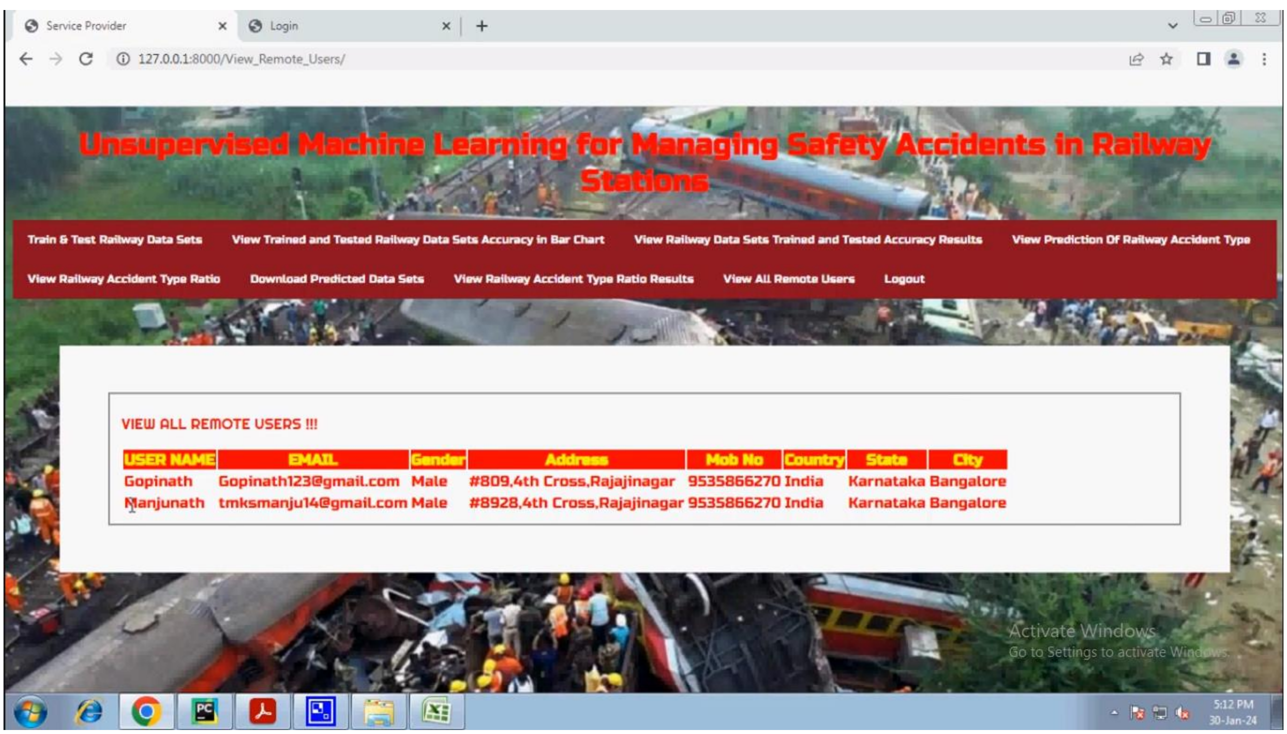

Fig 11. Result screenshot 11

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue.2, 2024

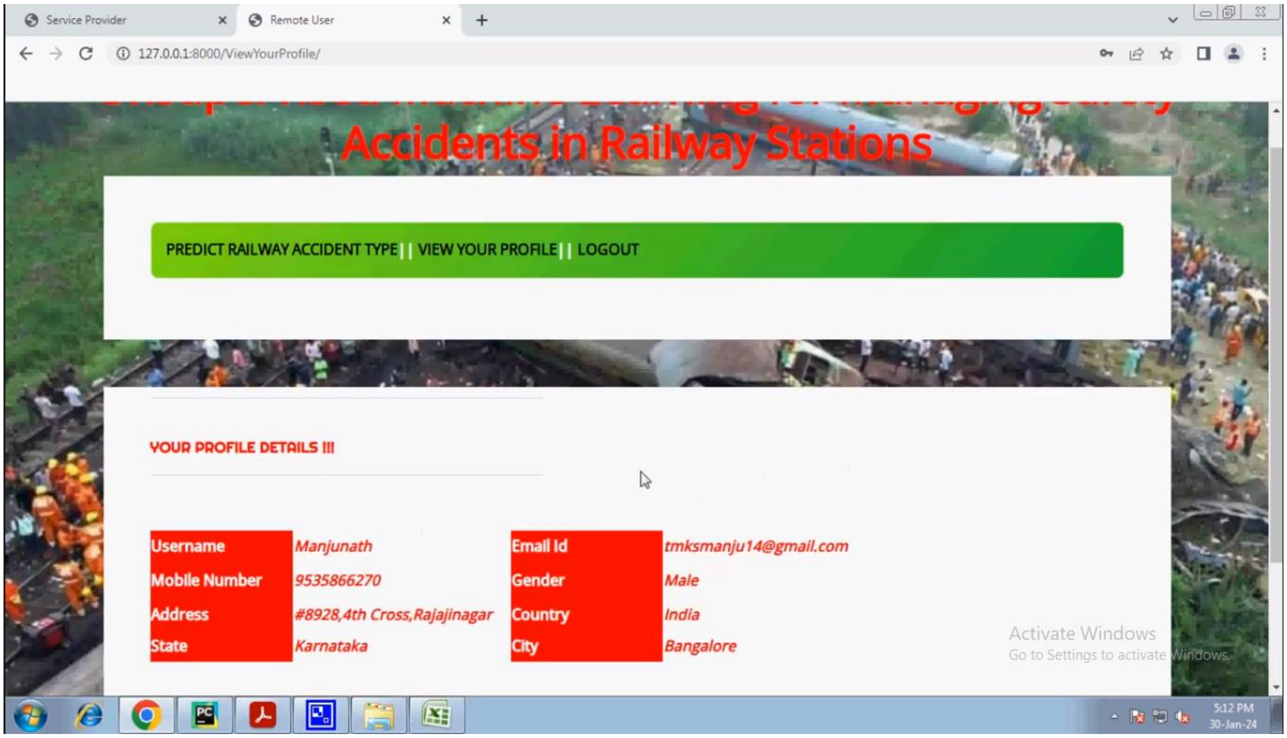

Fig 12. Result screenshot 12

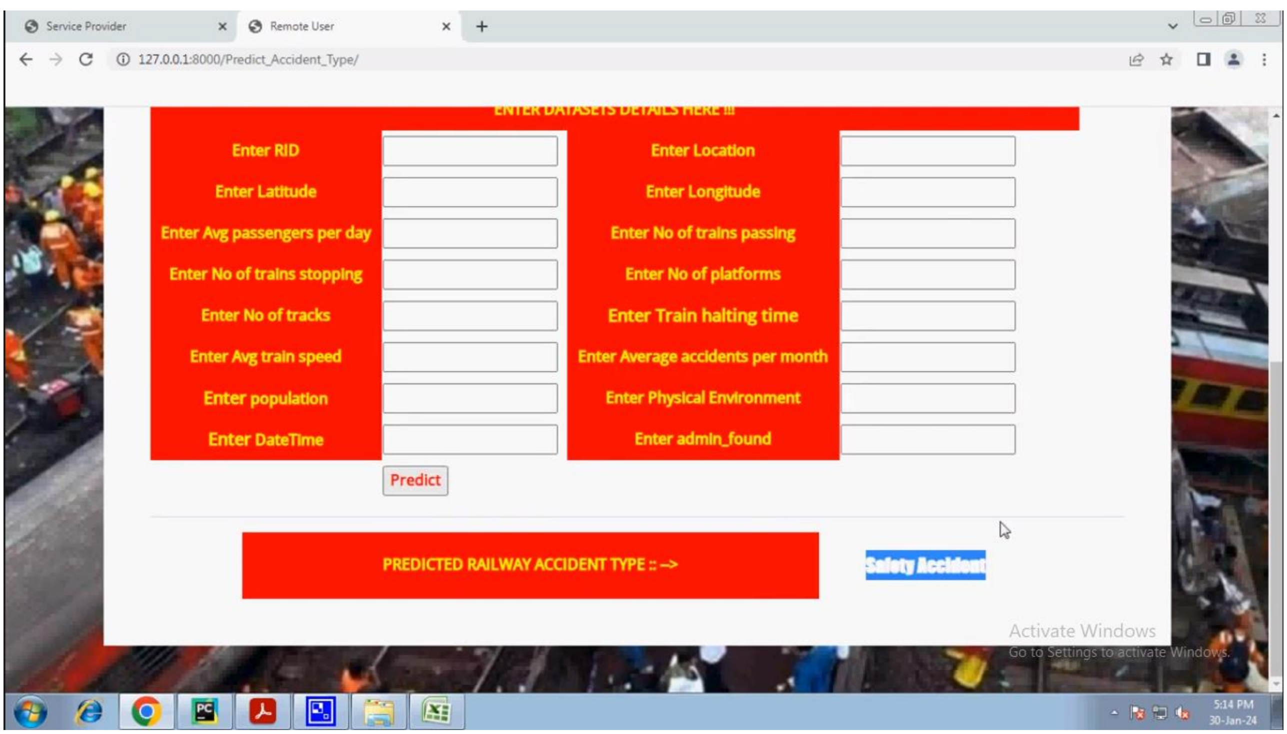

Fig 13. Result screenshot 13

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue.2, 2024

Fig 14. Result screenshot 14

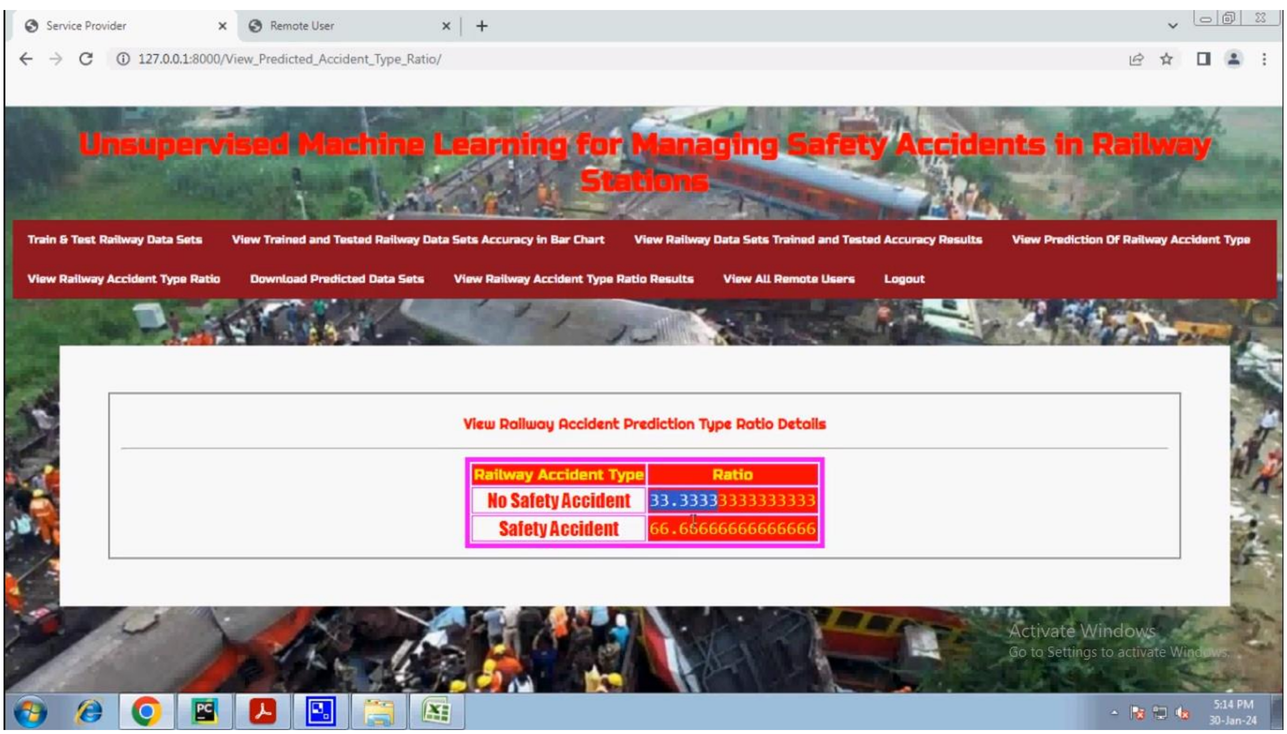

Fig 15. Result screenshot 15

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue.2, 2024

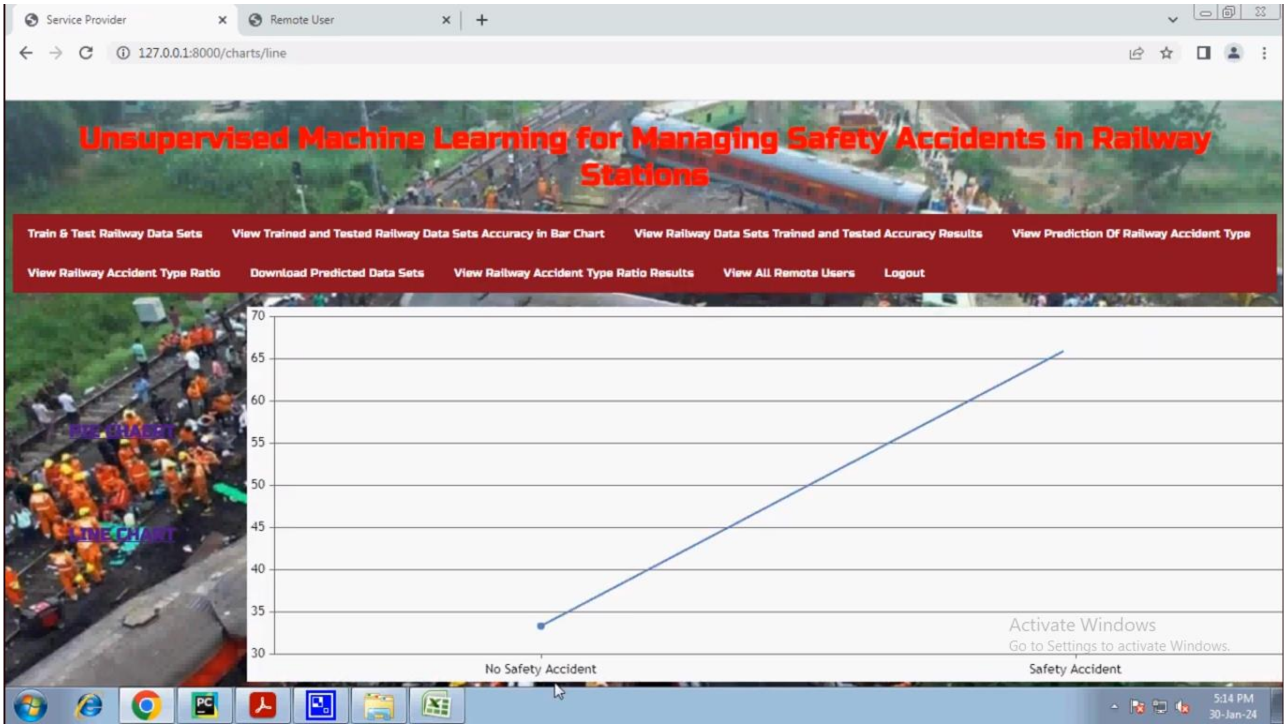

Fig 16. Result screenshot 16

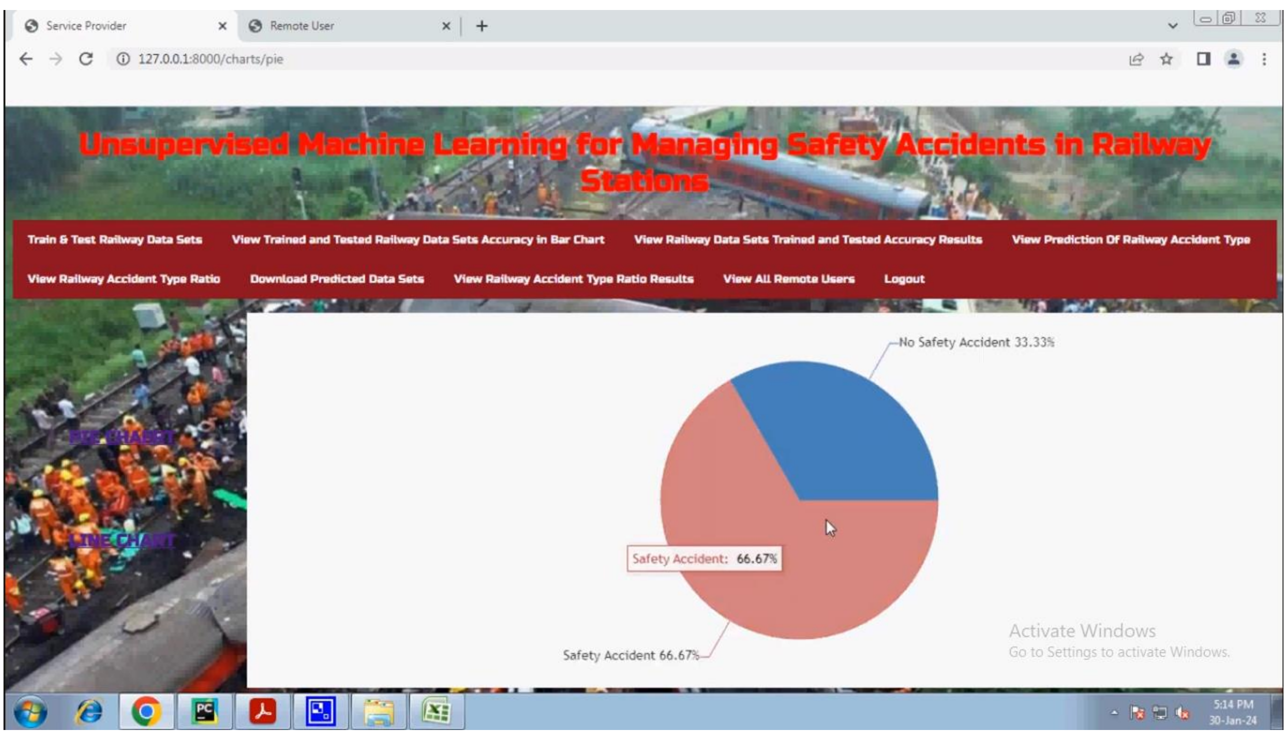

Fig 17. Result screenshot 17

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue.2, 2024

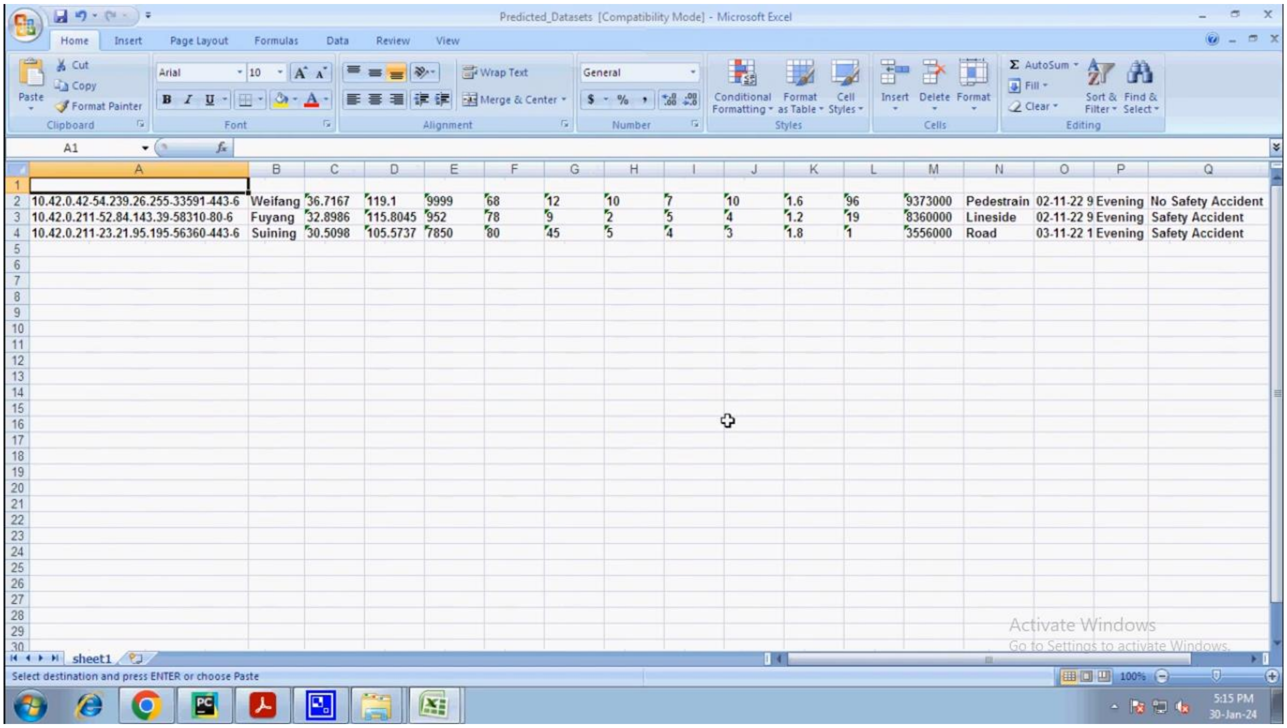

Fig 18. Result screenshot 18

Overall, the results and discussion underscore the transformative potential of unsupervised machine learning techniques in railway industry safety and other safety-critical domains. By leveraging AI methods such as unsupervised topic modeling, safety administrators can gain a deeper understanding of safety risks and implement targeted interventions to mitigate accidents and enhance safety standards. The study's findings emphasize the importance of utilizing advanced analytical tools to harness the wealth of textual data available in accident reports and other safety-related documents. Through systematic analysis and predictive modeling, railway stakeholders can proactively address safety concerns, reduce the occurrence of accidents, and ensure the continued reliability and safety of railway operations. Moreover, the study highlights the broader implications of AI applications in enhancing safety across various industries, underscoring the potential for future research and innovation in safety management practices.

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.

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue.2, 2024

REFERENCES

1. Chen, Y., & Gao, J. (2016). Unsupervised machine learning: a review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 6(4), 235-247. https://doi.org/10.1002/widm.1198

2. Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. Journal of Machine Learning Research, 3(Jan), 993–1022. http://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf

3. Dong, J., Wang, X., Zhao, T., & Sun, L. (2020). Application of unsupervised machine learning in mechanical fault diagnosis: A review. IEEE Access, 8, 198819-198833. https://doi.org/10.1109/ACCESS.2020.3033807

4. Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). stm: An R package for structural topic models. Journal of Statistical Software, 91(2), 1-40. https://doi.org/10.18637/jss.v091.i02

5. Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37–46. https://doi.org/10.1177/001316446002000104

6. Wang, C., & Blei, D. M. (2011). Collaborative topic modeling for recommending scientific articles. Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 448-456. https://doi.org/10.1145/2020408.2020482

7. Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. Proceedings of the National Academy of Sciences, 101(Suppl 1), 5228–5235. https://doi.org/10.1073/pnas.0307752101

8. Chen, L., Shi, J., & Liu, C. (2014). Applications of unsupervised learning in bridge engineering: a review. Archives of Computational Methods in Engineering, 21(2), 173-192. https://doi.org/10.1007/s11831-013-9096-5

9. Ramage, D., Hall, D., Nallapati, R., & Manning, C. D. (2009). Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, 248–256. https://doi.org/10.3115/1699510.1699548

10. Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. Proceedings of the 23rd International Conference on Machine Learning, 113-120. https://doi.org/10.1145/1143844.1143859

11. Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. Proceedings of the Conference on Empirical Methods in Natural Language Processing, 262–272. https://www.aclweb.org/anthology/D11-1020.pdf

12. Buntine, W. (2012). Operations for learning with graphical models. Journal of Machine Learning Research, 13, 1113–1154. http://www.jmlr.org/papers/volume13/buntine12a/buntine12a.pdf

13. Hall, D., Ramage, D., Jurafsky, D., & Manning, C. D. (2008). Studying the history of ideas using topic models. Proceedings of the Conference on Empirical Methods in Natural Language Processing, 363–371. https://doi.org/10.3115/1613715.1613765

14. Blei, D. M., Griffiths, T. L., Jordan, M. I., & Tenenbaum, J. B. (2004). Hierarchical topic models and the nested Chinese restaurant process. Advances in Neural Information Processing Systems, 16, 17–24. https://papers.nips.cc/paper/2004/file/5a6d9264fb2a0db98fb9cf65b0d4a5e7-Paper.pdf

ISSN 2319-5991 www.ijerst.com

Vol. 17, Issue.2, 2024

15. Zhu, X., & Ghahramani, Z. (2002). Learning from labeled and unlabeled data with label propagation. Technical Report, Carnegie Mellon University, Pittsburgh, PA. https://www.cs.cmu.edu/~zhuxj/pub/CMU-CALD-02-107.pdf

16. Paisley, J., Blei, D. M., & Jordan, M. I. (2012). Variational Bayesian inference with stochastic search. Proceedings of the 29th International Conference on Machine Learning, 1347–1354. https://dl.acm.org/doi/10.5555/3042573.3042736

17. Wang, X., & McCallum, A. (2006). Topics over time: A non-Markov continuous-time model of topical trends. Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 424–433. https://doi.org/10.1145/1150402.1150469