International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

ANALYSIS OF ROAD TRAFFIC FATAL ACCIDENTS USING DATA MINING TECHNIQUES

[1]JALLURI VIDYA SAI MOUNIKA,[2]DR.V.BHASKAR MURTHY

[1]MCA Student,B V Raju College, Bhimavaram,Andhra Pradesh,India

[2]Professor,Department Of MCA,B V Raju College,Bhimavaram,Andhra Pradesh,India

ABSTRACT

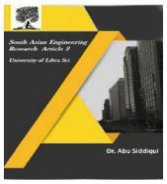
Ensuring roadway traffic safety is a critical concern for both transportation authorities and the general public. To provide effective safe driving recommendations, a thorough analysis of traffic data is essential to identify key factors contributing to fatal accidents. In this study, we utilize statistical analysis and data mining techniques on the FARS Fatal Accident dataset to explore these relationships. Specifically, we examine the impact of various attributes, including collision manner, weather conditions, road surface conditions, lighting conditions, and driver intoxication, on fatal accident rates. We employ the Apriori algorithm to uncover association rules, use the Naïve Bayes classifier to develop a classification model, and apply K-means clustering to identify patterns in the data. Based on these analyses, we propose actionable safety recommendations derived from statistical insights, discovered patterns, and predictive models.

Index Terms—Traffic safety, fatal accidents, association rules, classification, clustering.

I.INTRODUCTION

Millions of vehicles travel on roadways every day, making traffic accidents an ever-present risk. Some of these accidents result in fatalities, leading to tragic losses of life. As responsible drivers, we all strive to avoid accidents and ensure road safety. To gain valuable insights into safer driving practices, data mining techniques can be applied to traffic accident datasets to identify significant patterns and risk factors. Data mining encompasses various techniques and algorithms designed to uncover relationships within large datasets. Over the past decades, it has become a crucial tool in information technology [2]. One widely used approach is association rule mining, which helps identify significant correlations between data attributes in extensive databases and plays a key role in frequent itemset mining [1]. A classical method for association rule mining is the Apriori algorithm, which identifies frequent itemsets—an approach we utilize to analyze roadway traffic data. Another essential data mining technique is classification, which involves building a model (classifier) from a training dataset to predict the class labels of new, unseen records. The Naïve Bayes classifier, a fundamental probability-based method, is based on Bayes' theorem and assumes independence between variables, making it an efficient and widely used classification technique.

For this study, we used the Fatality Analysis Reporting System (FARS) dataset, which includes records of all fatal accidents on public roads in 2007 as reported to the National Highway Traffic Safety Administration (NHTSA) [9]. The dataset, obtained from California Polytechnic State University, originates from FARS and contains 37,248 records and 55 attributes. A detailed description of the dataset can be



found in the FARS Analytic Reference Guide (1975–2007) [11].

II.LITERATURE SURVEY

1. Divya Bansal and Lekha Bhambhu(2013) This study explores the application of the Apriori algorithm in data mining to analyze crime patterns, particularly those related to violence against women. The authors emphasize how frequent pattern mining can uncover hidden correlations in large datasets, providing law enforcement with valuable insights. The Apriori algorithm identifies strong association rules by finding frequent itemsets, which can be applied to crime data to detect high-risk areas, trends in criminal activity, and factors influencing crime rates. While the study focuses on crimes against women, its methodology is applicable to various domains, including traffic accident analysis. The research highlights the significance of association rule mining in predictive analysis and policy-making, offering a structured approach to crime prevention. The study also discusses challenges such as data sparsity and computational complexity. Its findings contribute to the broader field of data-driven decision-making, demonstrating how Apriori can enhance crime analysis and safety strategies.

2. Amira A. El Tayeb, Vikas Pareek, and Abdelaziz Araar (2015) This research investigates the use of association rule mining for analyzing traffic accidents in Dubai. The authors apply data mining techniques to a large traffic accident dataset to identify key factors contributing to accidents. The study employs various association rule mining algorithms, including the Apriori algorithm, to discover relationships between different variables such as weather conditions, time of day, road conditions, and driver behaviors. The findings indicate that certain conditions, such as wet roads and nighttime driving, significantly increase the likelihood of accidents. The paper also emphasizes the importance of data preprocessing to remove inconsistencies and improve analysis accuracy. The study contributes to traffic safety by offering insights into accident-prone conditions, which can help authorities implement preventive measures. Overall, this research demonstrates how association rule mining can enhance traffic accident analysis, providing a framework for developing data-driven policies to reduce roadway fatalities and improve transportation safety.

3. WilliamM.Evanco(1999) This study examines the potential benefits of Mayday systems in reducing vehicular crash fatalities, particularly in rural areas. A Mayday system is an automated crash notification system that alerts emergency services immediately after a severe accident occurs. The research analyzes accident response times and their impact on survival rates, showing that reducing the delay in medical response significantly increases the chances of saving lives. Using statistical models, the author estimates that implementing Mayday systems could prevent a significant number of fatalities each year, particularly in remote locations where emergency response times tend to be longer. The study highlights the technological and logistical challenges of widespread Mayday system adoption, such as infrastructure development and cost concerns. The findings support the integration of intelligent crash reporting systems into modern vehicles, emphasizing how real-time data transmission can play a crucial role in emergency response and roadway safety.

4. K. Jayasudha and C. Chandrasekar (2009) This paper provides a comprehensive overview of data mining techniques applied to road traffic accident analysis. The authors discuss the role of classification, clustering, and association rule mining in uncovering hidden patterns in accident datasets. The study highlights how decision trees, neural networks, and K-means clustering can be used to predict accident hotspots and determine risk factors associated with severe crashes. The paper also discusses the challenges in traffic accident analysis, such as data quality, missing values, and computational complexity. The authors emphasize the importance of integrating real-time data into traffic monitoring systems to enhance safety measures. The study serves as a foundational reference for researchers working on traffic accident prediction models and highlights the potential of big data analytics in road safety improvements. The findings stress the importance of data-driven policies in reducing accident rates and improving urban traffic management.

5. S. Krishnaveni and M. Hemalatha (2011) This research provides a detailed analysis of traffic accidents using data mining techniques, focusing on the predictive modeling of accident occurrence. The authors use classification algorithms such as decision trees, Naïve Bayes, and support vector machines (SVMs) to develop models that can predict accident severity based on environmental and behavioral factors. The study identifies key accident predictors, including road conditions, time of day, vehicle speed, and driver behavior. The authors also discuss the challenges of handling imbalanced datasets, a common issue in accident analysis where severe accidents are less frequent than minor ones. To address this, they explore oversampling techniques and feature selection methods to improve model accuracy. The study highlights the potential of machine learning algorithms in traffic safety applications, demonstrating how predictive analytics can assist in risk assessment and decision-making for transportation authorities. The research underscores the importance of integrating AI and data mining in traffic safety management.

III.PROPOSED METHODOLOGY

A. Data Preparation

Before constructing models, we conducted data preprocessing to ensure data quality. Records with missing values (denoted as 99 in the dataset) were removed. Numerical values were converted into categorical (nominal) values based on the dataset's data dictionary. The fatal rate was computed and categorized into two groups: high and low fatality risk.

Additionally, some new variables were derived from existing attributes. Examples include:

Fatal Rate: This metric represents the proportion of fatalities in an accident, calculated as:

$$\text{Fatal Rate} = \frac{\text{Fatalities}}{\text{Total Persons Involved}}$$

Records with missing time-related values were removed, and early morning hours (past midnight) were adjusted by adding 24 hours for consistency.

B. Modeling

To uncover key insights, we first performed statistical analysis on the dataset. We then applied the following data mining techniques:

- Association Rule Mining (using the Apriori algorithm) to identify relationships among variables.
- Clustering (using K-Means) to categorize states based on their population and fatal accident rates.
- Classification (using Naïve Bayes) to determine whether a region has a high or low risk of fatal accidents.

C. Result Analysis

Our analysis provided insights into:

1. Association Rules – Discovered relationships among accident attributes.
2. Clustering of U.S. States – Grouped states based on accident frequency and population.
3. Risk Classification – Identified high-risk and low-risk accident zones.

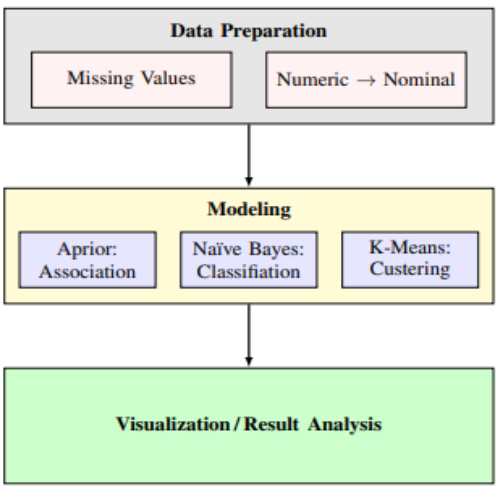


Fig. 1. Work flow

IV.CONCLUSION

Statistical analysis, association rule mining, and classification results indicate that environmental factors such as road surface, weather, and lighting conditions have a relatively weaker impact on fatal accident rates. In contrast, human factors, such as alcohol consumption and collision type, show a significantly stronger influence on accident severity.

Clustering analysis reveals that certain states or regions experience higher fatality rates compared to others. This suggests that drivers should exercise extra caution when traveling through high-risk areas.

Through this study, we recognize that data availability is a crucial limitation in making definitive conclusions. Additional datasets, such as non-fatal accident records, real-time weather conditions, and vehicle mileage data, could enhance analysis and lead to more precise safety recommendations. Expanding the dataset would allow for more comprehensive testing and stronger insights, ultimately improving roadway safety measures.

V.REFERENCES

[1] D. Bansal and L. Bhambhu, “Execution of Apriori algorithm of data mining directed towards tumultuous crimes concerning women,” International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, no. 9, Sep. 2013.

[2] A. A. El Tayeb, V. Pareek, and A. Araar, “Applying association rules mining algorithms for traffic accidents in Dubai,” International Journal of Soft Computing and Engineering, Sep. 2015.

[3] W. M. Evanco, “The potential impact of rural Mayday systems on vehicular crash fatalities,” Accident Analysis & Prevention, vol. 31, no. 5, pp. 455–462, Sep. 1999.

[4] K. Jayasudha and C. Chandrasekar, “An overview of data mining in road traffic and accident analysis,” Journal of Computer Applications, vol. 2, no. 4, pp. 32–37, 2009.

[5] S. Krishnaveni and M. Hemalatha, "A perspective analysis of traffic accidents using data mining techniques," International Journal of Computer Applications, vol. 23, no. 7, pp. 40–48, Jun. 2011.

[6] S. Kumar and D. Toshniwal, "Analyzing road accident data using association rule mining," in Proceedings of the International Conference on Computing, Communication and Security, pp. 1–6, 2015.

[7] E. M. Ossiander and P. Cummings, "Freeway speed limits and traffic fatalities in Washington state," Accident Analysis & Prevention, vol. 34, no. 1, pp. 13–18, 2002.

[8] K. M. A. Solaiman, M. M. Rahman, and N. Shahriar, "AVRA Bangladesh: Collection, analysis & visualization of road accident data in Bangladesh," in Proceedings of the International Conference on Informatics, Electronics & Vision, pp. 1–6, IEEE, 2013.

[9] Trac Integrated SCM & Project Management, "Fatal Accidents Dataset," [Online]. Available: https://wiki.csc.calpoly.edu/datasets/wiki/HighwayAccidents.

[10] U.S. Census Bureau, "Population Estimates," [Online]. Available: http://www.census.gov/popest/data/historical/2000s/vintage_2007/, 2007.

[11] U.S. Department of Transportation, FARS Analytic Reference Guide, 1975 to 2009, 2010. [Online]. Available: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/811352.