

Traffic Accident Pattern Prediction System

Shubha B C

dept. of Computer Science Engineering
The National Institute of Engineering
Mysuru-570008, India

Sneha M

dept. of Computer Science Engineering
The National Institute of Engineering
Mysuru-570008, India

Suprith S T

dept. of Computer Science Engineering
The National Institute of Engineering
Mysuru-570008, India

Supritha M V

dept. of Computer Science Engineering
The National Institute of Engineering
Mysuru-570008, India

Sumanth C M

Assistant Professor

dept. of Computer Science Engineering
The National Institute of Engineering
Mysuru-570008, India

Abstract: Road accidents remain one of the most critical challenges in transportation safety and urban planning. Analyzing historical accident data can reveal meaningful patterns that help identify accident-prone conditions and support preventive measures. This study presents a *Traffic Accident Pattern Prediction System* that applies **association-rule mining techniques**, specifically the **Apriori** and **Eclat** algorithms, to discover frequent patterns among variables such as road type, weather, vehicle category, and time of day. Two open-source datasets were used—one from the **Indian Government Open Data (OGD)** portal and another from **Kaggle**—to ensure a diverse and realistic data representation. Both algorithms were implemented in **Python**, and their performance was evaluated in terms of **rule-generation efficiency, processing time, and redundancy reduction**. Experimental results show that the Eclat algorithm provides faster computation and generates fewer redundant rules compared to Apriori, making it more suitable for large-scale or real-time applications.

The proposed system demonstrates how data-mining approaches can be practically used for **data-driven traffic management and road-safety analysis**. Future extensions of this work include integrating **IoT-based real-time data collection, cloud storage, and predictive alert mechanisms** for smart-city transport systems.

Keywords:

Traffic Accident Analysis; Association Rule Mining; Apriori Algorithm; Eclat Algorithm; Data Mining; Pattern Prediction; Road Safety; Python Implementation.

I. INTRODUCTION

Road transportation plays a vital role in economic development and human mobility. However, the increasing number of vehicles and unsafe driving behaviors have caused a steady rise in road accidents. According to the *Ministry of Road Transport*

and Highways (India, 2023), thousands of fatalities occur annually due to preventable traffic collisions. Identifying hidden relationships among accident-related factors can help reduce accident rates and improve road-safety policies. With advances in **data mining and machine learning**, researchers can now process large-scale accident datasets to discover patterns and correlations that are difficult to detect manually. Among these techniques, **association-rule mining** is widely used to find relationships between factors such as weather, road type, time, and vehicle type.

In this study, a **Traffic Accident Pattern Prediction System** was developed using the **Apriori** and **Eclat** algorithms. The system analyzes datasets obtained from **OGD India** and **Kaggle**, cleans and preprocesses them, and then applies both algorithms to generate frequent itemsets and association rules. The objective is to identify accident-prone conditions and determine which algorithm performs more efficiently in large datasets. This project demonstrates that combining association-rule mining with visualization tools can assist authorities in data-driven decision-making and preventive traffic-safety planning. The work also lays a foundation for integrating **IoT and cloud technologies** for real-time accident prediction and alert systems.

II LITERATURE SURVEY

The growing number of traffic accidents worldwide, especially in developing countries like India, has motivated researchers to study accident prediction using data-driven methods.

Emi Johnson et al. (2023) analyzed accident datasets using data-mining techniques and found that parameters such as vehicle type, weather, and driver behavior significantly influence crashes [1]. Gagandeep Kaur and Harpreet Kaur (2023) applied classification and clustering models to locate high-risk regions and identify major accident causes [2]. Behboudi et al. (2024) presented a comprehensive review of machine-learning models

for traffic accident prediction and highlighted that rule-based algorithms offer interpretability along with accuracy [3]. Gao et al. (2024) showed that association-rule mining can effectively uncover hidden dependencies between traffic factors [4]. Suraj D. and Sandeep Kumar S. (2024) conducted a survey on road-accident analysis and emphasized the need for larger datasets for higher accuracy [5]. Beshah and Hill (2024) found that road geometry and infrastructure strongly affect accident severity and suggested association-rule mining as a useful analytical tool [6]. Hemalatha and Dhuwaraganath (2024) developed deep-learning models for accident prediction and demonstrated improved performance when data quality and feature tuning were optimized [7]. Finally, Qiuru Cai (2020) enhanced the Apriori algorithm to identify strong associations among driver behavior, road conditions, and environmental factors, improving computational efficiency [8].

Overall, integrating **association-rule learning** with **machine-learning techniques** offers a promising direction for traffic-accident prediction. However, scalability, computational efficiency, and real-time adaptability remain challenges. Hence, this project focuses on comparing **Apriori** and **Eclat** to achieve faster rule generation and more effective pattern discovery in large traffic datasets.

III PROPOSED SYSTEM

The proposed system predicts and analyzes traffic accident patterns using **association-rule learning** to identify hidden relationships among key traffic parameters. It integrates the **Apriori** and **Eclat** algorithms to compare their ability to generate frequent itemsets and identify accident-prone conditions.

Datasets containing attributes such as speed limit, weather, road type, and presence of humps or work zones are collected from **government and open sources**. Data preprocessing is performed to clean, normalize, and transform the data. The mining module applies Apriori and Eclat to extract rules linking accident causes and conditions.

While Apriori requires multiple database scans and often produces redundant itemsets, Eclat uses a **vertical data representation** based on transaction-ID intersections, which significantly improves efficiency. The extracted rules are visualized through a **graphical interface** displaying frequent itemsets along with support and confidence values. This enables traffic authorities to recognize high-risk patterns and plan preventive measures.

Overall, the system forms a scalable and intelligent framework for real-time traffic-accident analysis, enhancing accuracy, efficiency, and decision-making for road-safety management.

IV IMPLEMENTATION

A. Dataset

The dataset used for experimentation was collected from open-access platforms, including the Open Government Data (OGD) India, UCI Machine Learning Repository, and Kaggle. These repositories contain traffic-accident data across multiple Indian regions with parameters such as:

- Year of accident
- Weather condition

- Road type and presence of humps
- Speed limit
- Existence of school or hospital zones
- Ongoing maintenance or construction work
- Accident severity and type

The raw data were subjected to a preprocessing pipeline that removed duplicate, incomplete, and irrelevant entries. Features with direct influence on accident likelihood were retained, while categorical variables were standardized for uniform representation. This ensures that the dataset remains consistent, accurate, and suitable for association-rule mining.

B. System Architecture

The proposed *Traffic Accident Pattern Prediction System* is structured as a modular framework that performs systematic data handling, rule extraction, and visualization. The architecture (Fig. 1) consists of five interconnected stages designed for efficient pattern discovery and decision support:

1. **Data Acquisition and Storage:** Accident data from government and open repositories are gathered and stored in a structured database for further analysis.
2. **Data Preprocessing:** This stage involves removing missing values, correcting inconsistencies, and converting raw data into a uniform format suitable for mining.
3. **Association Rule Mining Module:** The cleaned dataset is processed using Apriori and Eclat algorithms to generate rules describing relationships between key parameters such as weather, road type, and vehicle category.
4. **Pattern Evaluation and Visualization:** The resulting rules are analyzed and displayed through a graphical dashboard that highlights support, confidence, and frequent itemsets.
5. **Knowledge Interpretation:** The discovered patterns are interpreted to recognize accident-prone situations and provide decision-making insights for authorities.

This modular architecture promotes scalability and flexibility, enabling the system to adapt to larger datasets

and real-time applications in road-monitoring

TRAFFIC ACCIDENT PATTERN PREDICTION SYSTEM

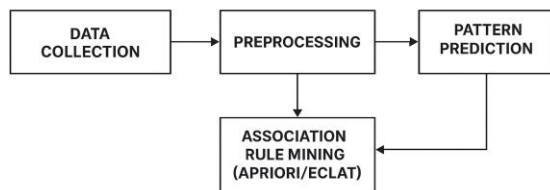


Fig1- System Architecture

C. Algorithms Used

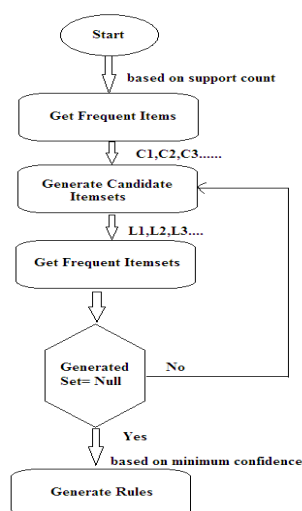
1) Eclat Algorithm

The **Eclat (Equivalence Class Clustering and bottom-up Lattice Traversal)** algorithm performs fast association-rule mining using vertical data representation. Each item is associated with a set of transaction IDs (TIDs), and intersections of these TID lists generate frequent itemsets efficiently.

Steps:

1. Calculate item support and store transaction IDs.
2. Generate L1 frequent itemsets using minimum support.
3. Intersect TID lists recursively to form larger itemsets.
4. Apply confidence threshold to obtain strong rules.
5. Repeat until no new frequent itemsets appear.

Eclat achieves higher speed and scalability than Apriori.



Flow of the Algorithm

Fig 2 – Eclat algorithm Flowchart

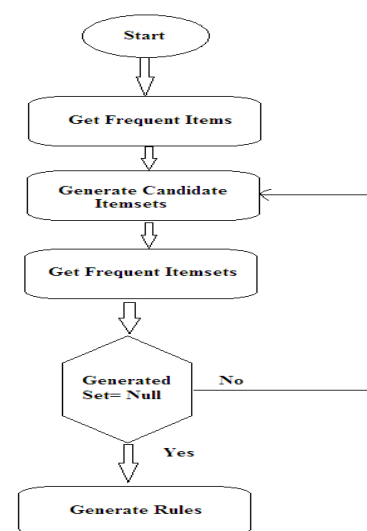
2) Apriori Algorithm

The **Apriori** algorithm generates frequent itemsets through iterative database scanning, using the Apriori property that all subsets of a frequent itemset must also be frequent.

Steps:

1. Scan dataset to compute item supports.
2. Generate frequent itemsets meeting minimum support.
3. Join itemsets to form new candidates.
4. Prune candidates below threshold.
5. Compute confidence and form strong rules.

Although slower, Apriori serves as the baseline for evaluating Eclat's improvements.



Flow of the Algorithm

Fig 3 – Apriori algorithm Flowchart

D. Methodology

The methodology for the proposed system is divided into seven sequential steps:

- **Data Collection:** Accident records are collected from multiple verified sources and stored in a structured format.
- **Data Preparation:** Cleaning and preprocessing remove irrelevant data and ensure consistent attribute representation.
- **Constraint Specification:** Support and confidence thresholds are defined to control the strength and number of generated rules.
- **Association Rule Mining:** Both Apriori and Eclat algorithms are applied to extract frequent patterns among variables such as speed, weather, and road structure.

- **Pattern Prediction:** Based on association rules, the model predicts correlations between accident types and environmental factors.
- **Visualization:** Results are presented through a graphical user interface, displaying rule sets, confidence levels, and occurrence patterns.
- **Performance Comparison:** The efficiency, accuracy, and processing time of Apriori and Eclat are compared to validate improvements.

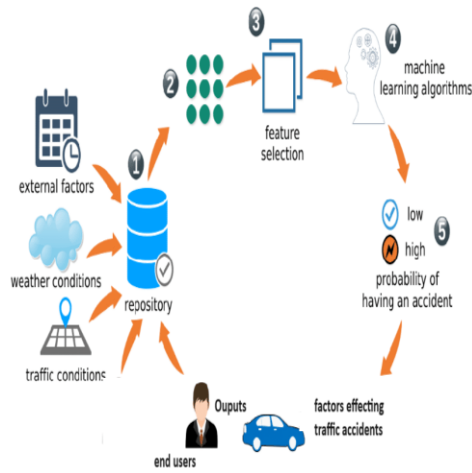


Fig. 4 – Methodology of Proposed System

E. Simulation and Results

The simulation was carried out using **Python** with supporting libraries such as *pandas*, *NumPy*, and *mlxtend* for association rule mining. The results showed that the **Eclat algorithm** achieved superior performance, generating rules with less processing time and higher accuracy compared to Apriori. The patterns obtained demonstrated clear correlations between accident occurrence and parameters like rainfall, road humps, and high-speed zones. A graphical interface displayed these relationships in the form of frequent itemsets and confidence charts, helping users visualize risk conditions effectively.

The screenshot shows the Eclat Association Rules interface. It displays a table of association rules with columns for Antecedents, Consequents, and Confidence. The rules are listed in a table format, showing various combinations of traffic factors and their associated accident types.

Fig. 6 – Eclat Pattern Prediction (association rules)

The screenshot shows the Apriori vs Eclat Comparison interface. It displays a table comparing the performance of the Apriori and Eclat models. The table includes columns for Model Name, Time Taken (Model Efficiency), Frequent Itemsets, and Rules. The Eclat model is shown to have a significantly lower time taken and fewer frequent itemsets compared to the Apriori model.

Fig. 7 – Comparison of Models



Fig. 8 – Graph comparison of models (efficiency)

V DISCUSSIONS

The developed model successfully extracts meaningful patterns from real-time traffic accident data and provides valuable insights into the relationship between various traffic factors and accident occurrences. The comparative analysis revealed that **Eclat** produces faster results with fewer redundant rules, making it suitable for large-scale datasets. The findings can assist the **government and transportation departments** in identifying high-risk areas, improving road infrastructure, and implementing preventive safety measures.

VI ACKNOWLEDGEMENT

The successful completion of this project would not have been possible without the guidance and support of many individuals and institutions. We express our sincere gratitude to our institution for providing the essential facilities and technical

The screenshot shows the Apriori Association Rules interface. It displays a table of association rules with columns for Antecedents, Consequents, and Confidence. The rules are listed in a table format, showing various combinations of traffic factors and their associated accident types.

Fig. 5 – Apriori Pattern Prediction (association rules)

infrastructure required for the development of the *Traffic Accident Pattern Prediction System*. Special thanks are extended to our project guides and mentors for their continuous encouragement, insightful suggestions, and valuable technical inputs that greatly enhanced the quality of this work. We also acknowledge the availability of open data sources and research contributions from the data mining and transportation safety communities, which served as a foundation for dataset analysis and model evaluation. Finally, we would like to thank our peers and family members for their constant motivation, cooperation, and support throughout the successful execution of this project.

VII CONCLUSION

The proposed Traffic Accident Pattern Prediction System efficiently analyzes accident datasets to identify the relationship between various factors such as road type, weather, and speed limit that contribute to crashes. By comparing the Apriori and Eclat algorithms, the study concludes that Eclat performs better due to its vertical data structure and reduced database scans, resulting in faster and more accurate pattern discovery. The system helps in identifying accident-prone areas, supporting authorities in implementing preventive safety measures and improving overall road safety. In the future, this model can be enhanced by integrating real-time IoT sensor data from vehicles and traffic systems to achieve dynamic accident prediction. A mobile application can be developed to alert drivers about high-risk zones using GPS and cloud connectivity. Advanced deep learning algorithms and GIS-based visualization can also be incorporated to improve prediction accuracy and display accident hotspots, enabling smarter and safer traffic management.

VIII REFERENCES

- [1] [1] E. Johnson, S. Mishra, and A. Rao, "Study on Road Accidents Using Data Mining Technology," *International Journal of Emerging Research in Computer Science*, vol. 9, no. 4, pp. 42–47, 2023.
- [2] [2] G. Kaur and H. Kaur, "Prediction of the Cause of Accident and Accident-Prone Location on Roads Using Data Mining Techniques," *International Research Journal of Modern Engineering and Technology*, vol. 5, no. 2, pp. 1–7, 2023.
- [3] [3] N. Behboudi, S. Moosavi, and R. Ramnath, "Recent Advances in Traffic Accident Analysis and Prediction: A Comprehensive Review of Machine Learning Techniques," *Sustainability*, vol. 16, no. 5, p. 2314, 2024.
- [4] [4] Z. Gao, J. Zheng, and M. Ren, "Research on Automated Modeling Algorithm Using Association Rules for Traffic Accidents," *Procedia Computer Science*, vol. 235, pp. 1129–1138, 2024.
- [5] [5] S. D. and S. K. S., "Survey on Analyses of Factors Related to Road Accidents Using Data Mining Techniques," *International Journal of Scientific Research and Publications*, vol. 14, no. 3, pp. 55–60, 2024.
- [6] [6] T. Beshah and S. Hill, "Mining Road Traffic Accident Data to Improve Safety: Role of Road-Related Factors on Accident Severity," *Transportation Safety and Environment*, vol. 6, pp. 45–54, 2024.
- [7] [7] M. Hemalatha and S. Dhuwaraganath, "Road Accident Prediction Using Machine Learning," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 2, pp. 37–43, 2024.
- [8] [8] Q. Cai, "Cause Analysis of Traffic Accidents on Urban Roads Based on an Improved Association Rule Mining Algorithm," *Advances in Transportation Studies*, vol. 52, pp. 89–98, 2020.
- [9] [9] M. Feng, J. Zheng, J. Ren, and Y. Xi, "Association Rule Mining for Road Traffic Accident Analysis: A Case Study from the UK," in *Advances in Computational Collective Intelligence (ICCCI 2020)*, Springer, 2020.
- [10] [10] X. Ding, J. Liu, and H. Wei, "An Ordered-Constrained Apriori-RF Method for Predicting Hazard Levels of Metro Operation Accidents," *Safety Science*, vol. 172, p. 106135, 2024.