© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1718068

Machine Learning Based Road Crash Data Analysis and Prediction

HARSHITHA M [1], NISARGA H[2], NIKHITHA P[3], PALLAVI M [4], SUPRITHA SHREE B A[5]
[1]*Asst Professor, Department of Computer Science & Engineering Mysuru Royal Institute of Technology, Mandya, Karnataka, India.*
[2,3,4,5] *Student, Department of Computer Science & Engineering, Mysuru Royal Institute of Technology, Mandya, Karnataka, India.*

Abstract- Road crashes have emerged as a major global health issue, particularly impacting vulnerable road users such as pedestrians, cyclists, and two-wheeler riders in developing nations. Existing traffic systems rely on classical learning models that are inefficient, less accurate, and limited to manual record-keeping without performing intelligent analysis. This paper presents a web-based real-time application integrated with data mining and unsupervised machine learning classification algorithms to analyze traffic crash data and predict the environmental, behavioral, and situational factors contributing to accidents. The proposed system automates pattern discovery and parameter tuning to discover hidden traffic associations, providing data-driven insights that assist traffic departments in implementing preventive road safety measures.

Index Terms: Machine Learning, Association Rule Mining, Apriori Algorithm, Traffic Safety, Road Crash Prediction, Digital Platforms.

I. INTRODUCTION

Digital data analytics has become a fundamental pillar in modern urban infrastructure optimization. However, the exponential growth of urban expressways and rapid vehicle usage have significantly intensified global traffic safety challenges. Identifying and analyzing the intersecting causes that contribute to accidents has become critical to public health and law enforcement. Traditional forensic and empirical reporting mechanisms focus heavily on historical record storage but lack dynamic, real-time prediction and adaptability. Forensic cryptic traffic patterns—such as the correlations between sudden environmental shifts, road geometry constraints, and crash severity—frequently go undetected using manual analysis.

To overcome these gaps, modern intelligent transport frameworks must integrate multi-layered data mining architectures. This project introduces a scalable, browser-based solution leveraging Python and association rule learning to systematically extract interpretative accident patterns from dynamic datasets.

II. PROBLEM STATEMENT

Discovering the exact structural associations between traffic accidents and their underlying root causes is essential to minimizing crash occurrences. While modern traffic departments utilize software to log and archive thousands of incident records, these tools are strictly limited to raw data collection and storage without performing predictive analysis. Minimizing road fatalities remains an elusive task because there is currently no automated system equipped to dynamically compute or classify accident risk parameters simultaneously. Intruding external risks—including dangerous combinations of unlit road sections, poor weather, hidden humps, and shifting speed restrictions—co-occur without statistical synthesis. Consequently, authorities cannot apply targeted, proactive countermeasures, leading to an avoidable, continuous rise in regional traffic incidents.

III. OBJECTIVES

The core objective of this project is to architect a secure, automated, and real-time analytical application capable of mining high-volume accident configurations. The systematic objectives are defined as follows

- To collect, cleanse, and preprocess multi-source traffic accident datasets to eliminate missing or irregular values.
- To apply unsupervised learning algorithms (specifically the Apriori principle) to extract latent parameters like speed limits, weather conditions, proximity to schools, and humps to map correlation variables against specific

© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1718068

accident outcomes.

- To automate traffic visualization parameters based on data density & eliminating manual configuration.

IV. METHODOLOGY

1.Presentation and Interface Layer

The graphical user interface is constructed utilizing a responsive web grid (HTML, CSS, JavaScript, and Bootstrap) ensuring cross-platform capability. Entry points are partitioned securely by authorization levels. Unauthenticated public entities can view general home, structural details, and macro-level city pattern summaries. Authorized Traffic In-Charges utilize localized login screens verified against persistent database constraints to interact with operational dataset upload modules..

2. Relational Schema Data Layer

Backend state tracking and incident catalogs are managed dynamically inside a relational MySQL Server environment. Incoming multi-factor records (inclusive of categorical classifications like *Hit and Run*, *Over Speed*, and *Drunk and Drive*) map cleanly across explicitly defined foreign-key hierarchies linking members, localized road indexes, and incident severities.

3. Machine Learning Rule Mining Engine

The core analytical processing depends on executing serialized association pipelines via Flask micro-routing controllers. When an operator initiates a road analysis request, the application pulls categorical data vectors, applies dummy field indexing isolates accident consequence flags from situational parameters, and runs a bottom-up, breadth-first pruning routine to output top-tier association outputs filtered by metric parameters like Lift and Confidence.

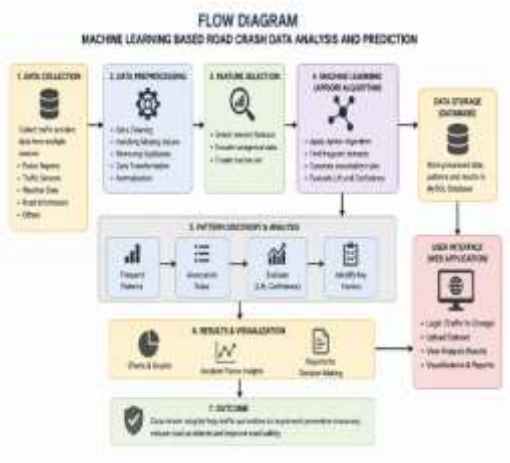
V. SYSTEM ARCHITECTURE

The application runs on a clean Three-Tier Architecture that enforces structural separation between data display, operational rules, and storage backends:

Presentation Layer: Collects operational entries, processes dashboard inputs, and presents graphical charts or analytical tables cleanly to users.

Business Logic Layer: The core routing mechanism. It handles data integrity checks, evaluates active session permissions, manages computational matrices, and feeds data tensors directly into prediction components.

Data Layer: Centralizes relational persistence. It isolates database insertion, updating, and execution loops from the frontend client to protect administrative configurations, credentials, and dataset assets.



VI. KEY TECHNOLOGIES

A. Languages: Python (v3.13.5 backend processing), JavaScript, HTML5, CSS3.

B. Frameworks & Environment: Flask Micro-framework, Visual Studio Code IDE.

C. Data Libraries: Pandas (structured DataFrames), NumPy (matrix operations), Scikit-Learn (predictive utilities), Mlxtend (Apriori and rule mining modules).

D. Database Management: MySQL Server utilizing SQLYog GUI management instrumentation.

© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1718068

VII. TOOLS USED

A. Development & Web Environment
- IDE: Visual Studio Code.
- Languages: Python (v3.13.5) and Front-End (HTML, CSS, JavaScript, Bootstrap).
- Framework: Flask Web Framework.

B. Database Management
- Database: MySQL Server.
- GUI Manager: SQLYog GUI Tool.

C. Libraries & Modeling
- Data Science: pandas and numpy.
- Machine Learning: sklearn and mlxtend (for Apriori Association Rules).
- Model Saving: pickle and joblib.
- Database Driver: pymysql.

VIII. APPLICATIONS

- Predictive Law Enforcement Optimization
- Civil Infrastructure Auditing
- Public Risk Awareness Platforms

IX. CHALLENGES

- Categorical Data Sparsity
- Fragmented System Architecture
- Offline Deployment Constraints

X. FUTURE SCOPE

- Intelligent Push Alert Infrastructures
- Cross-Admin Interaction Channels
- Deep Neural Network Integration

XI. CONCLUSION

This project presents a machine learning-based road crash data analysis and prediction system designed to identify the major factors contributing to traffic accidents. By integrating data mining techniques and association rule mining algorithms, the system enables efficient analysis of accident patterns and supports intelligent decision-making for traffic safety management. The proposed web-based platform improves data handling, automates pattern discovery, and provides meaningful insights that can help authorities implement preventive measures to reduce road accidents and enhance public safety.

REFERENCES

[1] Agrawal, R., Imieliński, T., & Swami, A. "Mining Association Rules Between Sets of Items in Large Databases." *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, pp. 207-216, 1993.

[2] Agrawal, R., & Srikant, R. "Fast Algorithms for Mining Association Rules in Large Databases." *Proceedings of the 20th International Conference on Very Large Data Bases*, pp. 487-499, 1994.

[3] Chang, L.Y., & Wang, H.W. "Analysis of traffic injury severity: An application of non-parametric classification tree techniques."

Accident Analysis and Prevention, 38(5), pp. 1019-1027, 2006.

[4] Breiman, L. "Random Forests." *Machine Learning*, Vol. 45, pp. 5-32, 2001.

[5] Hipp, J., Güntzer, U., & Nakhaeizadeh, G. "Algorithms for Association Rule Mining — A General Survey and Comparison." *SIGKDD Explorations*, 2, pp. 58-64, 2000.

[6] Zhang, H., et al. "In-Memory Big Data Management and Processing: A Survey." *IEEE Transactions on Knowledge and Data Engineering*, Vol. 27, No. 7, pp. 1920-1948, 2015.