



Enhancing Road Safety through Accident Prediction: A Study Using Machine Learning and Data Analytics

Student name: Ajay Sunil

Student ID number: 2406774

Module: Research Methods and Professional Development

Module code: ACFA7004

Assignment Title: Presentation

Lecturer: Dr Seena Joseph

BACKGROUND

- Global Concern: 1.3M deaths & 50M injuries annually [1].
- Need for Prediction: Reduces accidents, improves road safety [2].
- Data Mining Role: Analyzes factors (road, weather, traffic, driver) [3].
- Techniques Used: Decision Trees, SVM, Naïve Bayes, CNN [4].
- Real-World Impact: Helps traffic management & policy-making [5].
- Future Scope: Integration of GPS & IoT for real-time prediction [6].

PROBLEMS

- High Accident Rates: 1.3M deaths & 50M injuries annually, with India among the most affected [1].
- Ineffective Traditional Methods: Existing statistical models fail to provide accurate real-time predictions [2].
- Complexity of Contributing Factors: Human behavior, road conditions, weather, and vehicle defects interact in unpredictable ways [3].
- Need for Data-Driven Solutions: Machine learning & data mining offer better risk assessment [4].
- Challenges in Model Accuracy: Existing ML models struggle with dynamic, real-time accident data [5]
- Future Need: Integrating IoT & GPS for real-time accident prediction & prevention [6].

AIM & OBJECTIVES

AIM

- To develop a simple yet effective road accident prediction model using Tableau, WEKA, and R, leveraging data visualization, machine learning, and statistical analysis.

OBJECTIVES

- Data Collection & Analysis: Use real-world accident datasets for insights [1].
- Data Visualization in Tableau: Identify trends, high-risk zones, and accident patterns [2].
- Machine Learning in WEKA: Apply classification algorithms (Decision Trees, Naïve Bayes, SVM) to predict accident severity [3].
- Statistical Analysis in R: Perform regression & clustering for deeper insights [4].
- Evaluate Model Performance: Compare accuracy and reliability of different methods [5].

LITERATURE REVIEW

Existing Studies:

- ML models improve accident prediction accuracy [7][8].
- Clustering & classification identify high-risk zones [9][10].
- Deep learning uses driver behavior for accident prediction [11].

How this research uses these findings:

- Tableau: Visualize accident-prone areas & trends [9].
- WEKA: Apply Decision Trees, Naïve Bayes & KNN for classification [7][8].
- R: Perform regression & clustering for accident pattern analysis [9][10].
- Simple & Effective Approach: No deep learning, focus on user-friendly tools.
- Practical Impact: Helps in road safety planning & risk mitigation.

METHODOLOGY – RESEARCH ONION APPROACH

Layer	Description	Relevance to This Research
Research Philosophy	Positivism – Focuses on observable data to identify trends and patterns.	Data-driven approach for predicting road accidents based on past data using Tableau, WEKA, and R.
Research Approach	Deductive Approach – Tests existing theories by applying ML models on real-world accident data.	Applying machine learning models to predict road accidents and assess contributing factors.
Research Strategy	Quantitative Analysis – Analyzes numerical data for accident prediction and risk assessment.	Use of quantitative data to apply classification & clustering models (e.g., SVM, Decision Trees, KNN).
Research Choices	Mono-Method Quantitative – Focuses on a single approach (ML & data analysis) for accident prediction.	Uses quantitative methods (machine learning & statistical analysis) for predictive modeling of accidents.
Time Horizon	Cross-Sectional Study – Analyzes data collected at a single point in time (historical accident data).	Analyzing past accident data from historical datasets to predict future trends and high-risk zones.
Data Collection & Analysis Tools	Tableau: Visualizes accident hotspots. WEKA: Applies ML models (e.g., Decision Trees, Naïve Bayes). R: Performs regression & clustering.	Tableau for data visualization, WEKA for machine learning models, and R for statistical analysis and regression.

DATA COLLECTION & PREPROCESSING

- Data Sources: Government accident records, Kaggle datasets, open traffic data.
- Preprocessing Steps: Data cleaning (removing missing values, duplicates).
- Feature Engineering: Selecting relevant features like road type, weather, speed.
- Balancing Data: Handling imbalanced datasets using oversampling techniques.
- Data Normalization: Standardizing numerical values for accurate ML predictions.

MODEL SELECTION & EVALUATION

Model	Why This Model?	Application in This Research
Decision Trees (WEKA)	Simple, interpretable & effective for classification [1].	Classifies accident severity levels (Fatal, Serious, Minor).
Naïve Bayes (WEKA)	Works well with small datasets, fast prediction [2].	Determines accident likelihood based on historical data.
K-Means Clustering (R)	Groups accident hotspots, identifies patterns [3].	Detects high-risk accident zones & clusters road conditions.
Regression (R)	Predicts trends based on historical accident data [4].	Estimates accident probability based on contributing factors.
Tableau (Visualization)	Enhances data storytelling & trend analysis [5].	Visualizes accident-prone areas, trends & heatmaps.

CRITICAL APPRAISAL OF RESEARCH PROCESS

- Strengths: Simple & effective accident prediction using ML without coding.
- Strengths: Uses well-established tools (Tableau, WEKA, R) for easy implementation.
- Weaknesses: Lack of real-time data processing may limit accuracy.
- Weaknesses: Some ML models require large computational resources.
- Future Scope: Integrating IoT & real-time GPS for dynamic accident prediction.

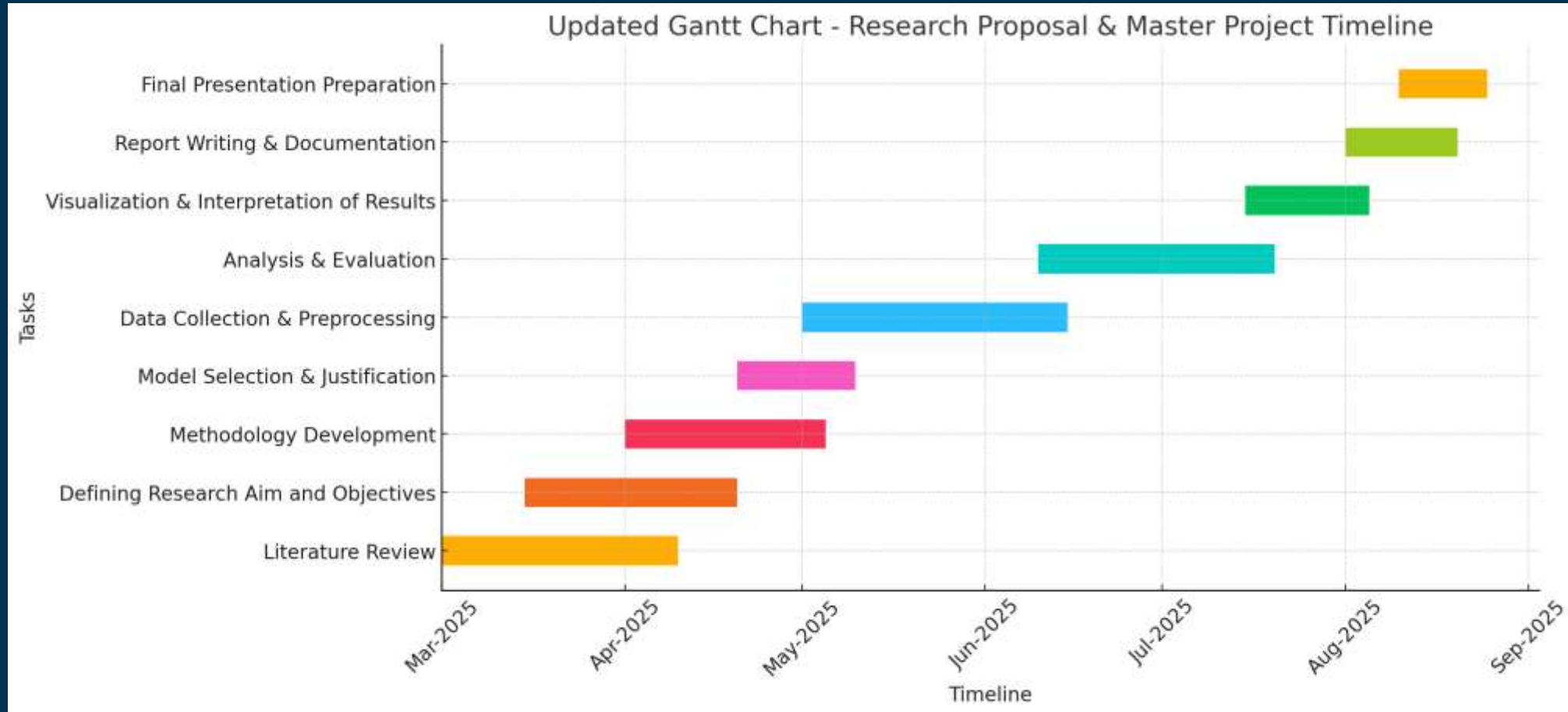
PROJECT IMPACT & LIMITATIONS

- Impact: Helps policymakers & urban planners identify accident-prone areas.
- Impact: Supports real-time road safety measures using ML-based predictions.
- Limitation: Model accuracy depends on data quality & completeness.
- Limitation: Lack of real-time data integration may affect prediction accuracy.
- Limitation: Some ML models require large datasets for better performance.

REAL-WORLD APPLICATIONS & RECOMMENDATIONS

- Application: Traffic authorities can use predictions to reduce accident risks.
- Application: Insurance companies can adjust premiums based on risk zones.
- Application: City planners can design safer road networks based on ML analysis.
- Recommendation: Use live traffic data for real-time accident predictions.
- Recommendation: Improve dataset quality by integrating multiple data sources.

GANTT CHART



REFERENCES

- [1] D. Viswanath, N. R, P. K, and B. R., "A Road Accident Prediction Model Using Data Mining Techniques," in Proceedings of the Fifth International Conference on Computing Methodologies and Communication (ICCMC), IEEE, 2021.
- [2] H. Lazar and Z. Jarir, "Road Traffic Accident Prediction: A Driving Behavior Approach," in Proceedings of the 8th International Conference on Optimization and Applications (ICOA), IEEE, 2022.
- [3] C. P. and S. M., "Road Accident Prediction and Classification using Machine Learning," in Proceedings of the IEEE Mysore Sub Section International Conference (MysuruCon), IEEE, 2022.
- [4] J. S. Reddy, R. M. R., and L. H. Pin, "Road Accident Prediction in Highways using Machine Learning Algorithms," in Proceedings of the IEEE 5th International Conference on Smart Electronics and Communication (ICOSEC), IEEE, 2024.
- [5] T. Augustine and S. Shukla, "Road Accident Prediction using Machine Learning Approaches," in Proceedings of the IEEE 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), IEEE, 2022.
- [6] S. Malik, H. El Sayed, M. A. Khan, and M. J. Khan, "Road Accident Severity Prediction – A Comparative Analysis of Machine Learning Algorithms," in Proceedings of the IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT), IEEE, 2021.
- [7] H. Lazar and Z. Jarir, "Road Traffic Accident Prediction: A Driving Behavior Approach," in Proceedings of the IEEE 8th International Conference on Optimization and Applications (ICOA), IEEE, 2022.
- [8] D. Santos, S. José, Q. Paulo, and B. N. Vítor, "Machine Learning Approaches to Traffic Accident Analysis and Hotspot Prediction," Computers, vol. 10, no. 12, p. 157, 2021.
- [9] A. Hébert, T. Guédon, T. Glatard, and B. Jaumard, "High-resolution road vehicle collision prediction for the city of Montreal," in Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), IEEE, pp. 1804-1813, 2019.
- [10] H. Ren, Y. Song, J. Liu, Y. Hu, and J. Lei, "A deep learning approach to the prediction of short-term traffic accident risk," CoRR, vol. abs/1710.09543, 2017.



Thank You .