PROJECT OVERVIEW

-BY ANSHU KUMAR

PROJECT TITLE

Traffic Accidents Prediction and Risk Analysis

THE CRITICAL CHALLENGE

Road traffic accidents remain a leading cause of injuries and fatalities worldwide. Rapid urbanization and increasing vehicle numbers have intensified the urgency for effective safety interventions. Traditional reactive approaches are no longer sufficient—we need predictive intelligence to prevent accidents before they occur.

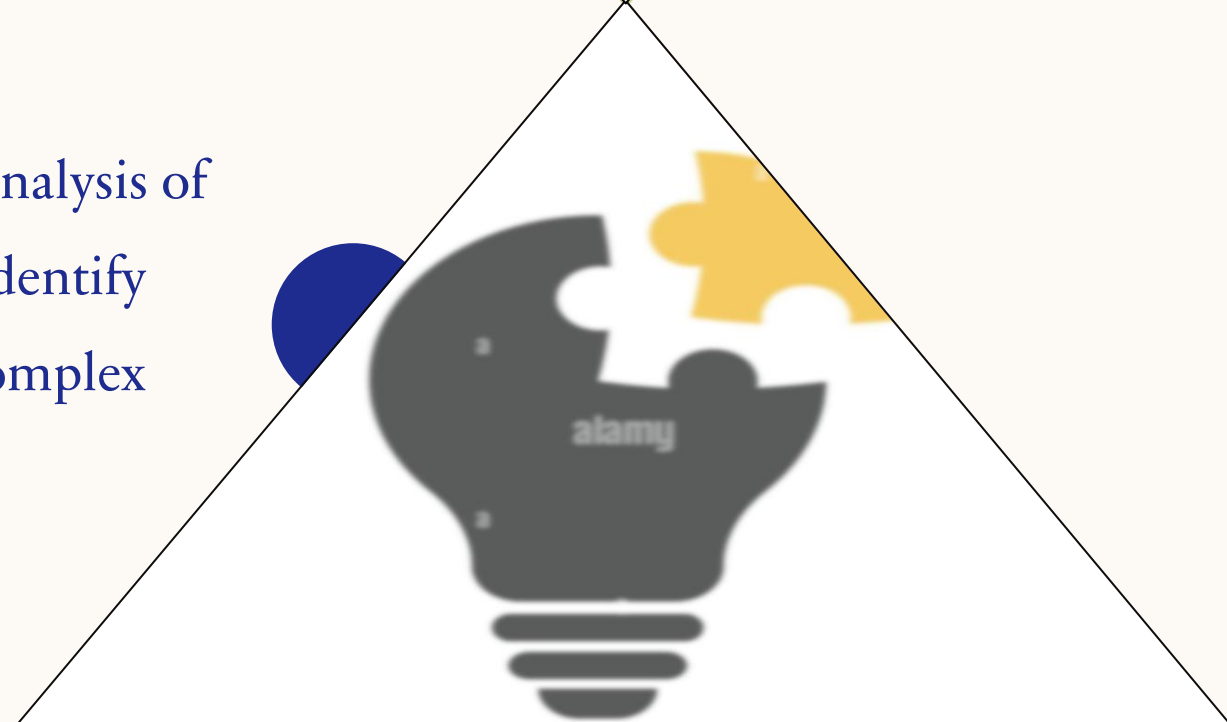
Using historical traffic-accident records, this project builds predictive models to estimate whether an accident will occur and to predict accident severity (Low / Moderate / High). The work combines exploratory data analysis, feature engineering, cleaning, machine learning classification to produce actionable insights for road-safety planning.

BUSINESS CASE

Enable transportation authorities and planners to implement proactive safety measures, optimize resource allocation, and reduce accident-related economic losses through predictive analytics.

PROBLEM

Current traffic safety approaches rely on reactive analysis of past incidents. We need predictive capabilities to identify risks before accidents occur and understand the complex interplay of contributing factors.



DATASET

- The dataset has 840 records and 14 features covering weather, road type, time of day, traffic density, speed limit, vehicle type, driver age/experience, alcohol use and more — so it contains the primary risk drivers for accidents.
- Accident severity distribution in the data: Low is the majority (~478), Moderate (~241), High (~79). This class imbalance must be handled in modeling.

DATASET DOWNLOAD LINK:

- https://www.kaggle.com/datasets/denkuznetz/traffic-accident-prediction?resource=download

OBJECTIVES

1. Clean and validate the dataset (handle missing values, correct/cap outliers, and standardize categories).
2. Perform EDA to quantify how weather, road type, time of day, traffic density, driving under influence, road condition, vehicle type, speed limit, and driver demographics influence accident occurrence and severity.
3. Build two predictive tasks:
   - Binary classification: Predict occurrence (accident vs no accident) if such label is present/meaningful.
   - Multi-class classification: Predict Accident_Severity (Low / Moderate / High).
4. Produce risk maps/profiles by grouping by road type, time-of-day, and weather (heatmaps / aggregated risk scores).

METHODOLOGY

Data cleaning

- Drop or impute rows with missing target.
- Impute categorical and numeric features.
- Treat outliers.

Exploratory Data Analysis

- Univariate distributions and bar charts for categorical variables.

Feature engineering

- Scaling.
- Encoding.

Modeling

- Baseline: Logistic Regression for quick interpretability.
- Stronger models: Random Forest, XGBoost for accuracy and feature importance

Evaluation

- Accuracy, Precision, Recall, F1, ROC-AUC.

EXPECTED OUTCOMES

- A predictive model capable of classifying accident severity (Low, Moderate, High) with high accuracy.
- Clear visual insights showing how risk varies by time of day, traffic density, and road type.
- Identification of key factors contributing to severe accidents.

CONCLUSION

The study demonstrates how data analytics and machine learning can effectively predict and analyze traffic accident risks. By identifying patterns related to driver behavior, environmental conditions, and road types, the model helps pinpoint critical risk factors that lead to severe accidents.

These insights can be used to:

- Implement targeted road safety policies (e.g., better lighting on mountain roads or night driving restrictions).
- Guide infrastructure improvements, such as drainage in high-rainfall areas.
- Promote awareness campaigns for risky driving behaviors like speeding and drunk driving.

THANK YOU

Anshu Kumar
9784377627
Email: singhanshukumar224@gmail.com
GitHub: https://github.com/anshu5159
LinkedIn: linkedin.com/in/singhanshukumar