Analyze traffic accident data to identify patterns related to road conditions, weather, and time of day. Visualize accident hotspots and contributing factors

Introduction

The study aims to analyze and predict traffic accident severity using machine learning techniques applied to a comprehensive dataset. Factors such as weather conditions, road types, and time of day are examined to understand their influence on accident occurrence and severity. By leveraging predictive modeling, this research seeks to provide insights that can enhance road safety measures and accident prevention strategies.

Data Overview

The dataset comprises detailed records of traffic accidents, including:

- **ID**: Unique identifier for each accident record
- **Source**: Data source providing the accident information
- **Severity**: Severity of the accident, categorized from 1 (minor) to 4 (fatal)
- **Start_Time, End_Time**: Timestamps indicating the start and end times of each accident
- **Start_Lat, Start_Lng, End_Lat, End_Lng**: Geographic coordinates of accident locations
- **Distance(mi)**: Distance covered by the accident
- **Weather Conditions**: Various weather conditions prevailing at the time of the accident
- **Road Characteristics**: Information about road types, traffic signals, and other relevant road features

Models Trained

Logistic Regression Model

The Logistic Regression model was trained to predict accident severity based on factors such as weather conditions, road types, and time of day. It achieved an accuracy of 82% in classifying accident severity levels. This model provides valuable insights into the likelihood of accidents occurring under different conditions, aiding in the identification of high-risk scenarios.

XGBoost Classifier

To address class imbalance and improve predictive performance, an XGBoost Classifier was employed. This ensemble learning model further enhanced the accuracy of predicting accident severity by effectively capturing complex relationships between various features in the dataset.

The XGBoost Classifier achieved an accuracy of 85%, demonstrating its robustness in handling imbalanced data and improving predictive outcomes.

Visualizations and Analysis

Accidents by Year

Analysis of accident trends over the years reveals patterns and changes in accident frequencies, highlighting long-term trends in road safety.

Accidents by Month

Examining accidents by month identifies seasonal variations in accident rates, influenced by factors such as weather conditions and holiday periods.

Accidents by Day of the Week

Differentiating accident rates between weekdays and weekends provides insights into varying traffic patterns and risk factors associated with different days.

Top 10 Weather Conditions with Most Accidents

Identifying the most prevalent weather conditions during accidents helps prioritize weather-specific safety measures and guidelines for drivers.

Number of Accidents by Severity

Categorizing accidents by severity levels (minor to fatal) offers a distribution overview, crucial for allocating emergency response resources effectively.

Number of Accidents by State

Geographic analysis pinpoints states with the highest accident rates, guiding localized safety interventions and policy measures.

Number of Accidents by Visibility

Correlating accident rates with visibility conditions underscores visibility's impact on road safety, influencing driving precautions and infrastructure planning.

Number of Accidents by Temperature

Analyzing accidents across temperature ranges reveals temperature-related driving challenges and informs seasonal road maintenance strategies.

Number of Accidents by Day and Night

Comparing accident rates during daylight versus nighttime hours highlights visibility and traffic density factors affecting accident occurrence.

Accidents by Simplified Weather Condition and Severity

Simplifying weather conditions and their relationship with accident severity clarifies weather-related risks and mitigation strategies.

Heatmap of Accidents by Simplified Weather Condition and Severity

Visual representation of weather condition impacts on accident severity aids in pinpointing high-risk weather scenarios and prioritizing safety interventions.

Number of Accidents by Road Condition

Analyzing accidents based on road conditions (e.g., wet, dry, icy) identifies road-specific risk factors and guides infrastructure improvements.

Accidents by Hour of Day

Hourly accident distribution profiles peak accident times, influencing traffic management strategies and driver awareness campaigns.

Distribution of Accidents by Wind Speed

Impact analysis of wind speed on accident rates informs wind-related safety protocols for drivers and road maintenance crews.

Distribution of Accident Severity

Distribution analysis of accident severity levels provides insights into the frequency and impact of different accident types on road safety.

Number of Accidents by County

Geospatial analysis of accident distribution within counties identifies localized accident hotspots, guiding targeted safety measures and traffic management initiatives.

Correlation Matrix of Numeric Variables

Quantifying relationships between numeric variables (e.g., distance, visibility) elucidates key factors contributing to accident severity, supporting data-driven policy decisions.

Number of Accidents by Season

Seasonal accident analysis highlights seasonal variations in accident rates, influencing seasonal traffic safety campaigns and resource allocation.

Conclusion

This project demonstrates the application of machine learning models in predicting traffic accident severity based on diverse factors such as weather conditions, road types, and time of day. The Logistic Regression and XGBoost Classifier models achieved accuracies of 82% and 85%, respectively, in predicting accident severity levels. The insights derived from this study can inform targeted interventions and policies aimed at reducing accident rates and improving road safety across different conditions and regions.