Predicting Traffic Accident Severity using KNN

By: Jasmine Ly

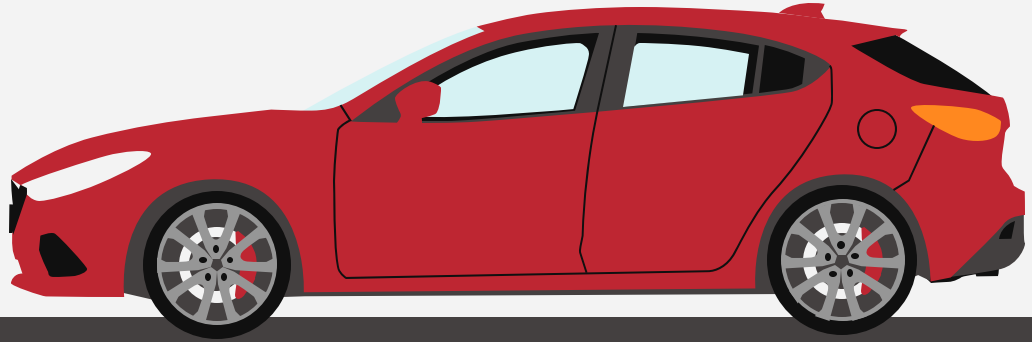


Table of Contents

Introduction

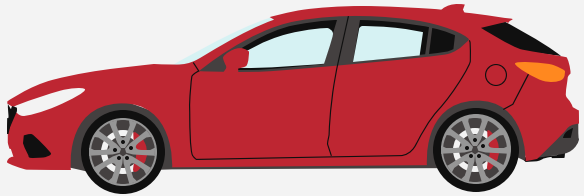
Objective: Develop a KNN model to predict traffic accident severity using key features (distance, weather conditions, and visibility)

Dataset: US Accidents (2016–2023) from Kaggle.

- Weather-related and numerical features

Why?

- Identify patterns in severity across accidents in the US
- Better understanding will allow for creating better prevention and response



Methodology

Features Used

- Distance (mi)
- Wind Speed (mph)
- Precipitation (in)
- Visibility (mi)
- Temperature (F)
- Humidity (%)

Target Variable: Severity

Process

- Quantitative, predicts accident severity
- Used 5% of dataset for the model to prevent errors
- Split the data 80/20 training/testing
- **for** loop tested different k values and created an elbow graph of the results for analysis
- Highest accuracy was 0.67 for k=59

Challenges

Extremely Large Data Set

- Original dataset contains 7.7 million records and 3.06 GB in size
- This lead to extremely long processing times and memory errors

Data Quality Issues

- Columns with mixed data types created preprocessing challenges
- Missing/incomplete data needed cleaning, which could potentially lead to a loss of valuable information.

Feature Selection

- Due to scale of project, we had to narrow down to only 6 factors
- Broadly applicable across regions

Lessons Learned

Limitations of KNN

- While KNN is a straightforward model that is easy to implement, it struggles to compute large multidimensional datasets due to its simplicity

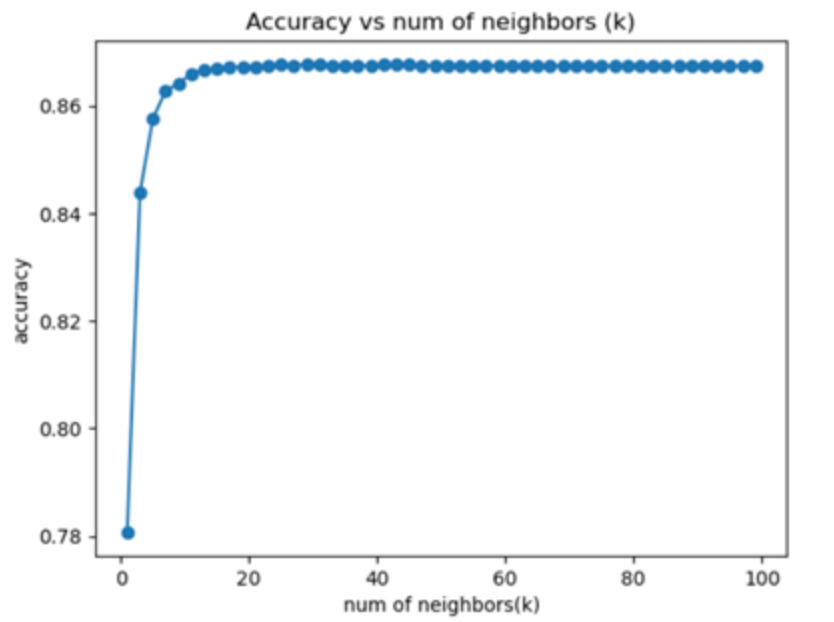
Importance of Data Quality

- Data cleaning and preprocessing are critical steps of building a reliable model.

Value of Hyperparameter Tuning

- Experimenting with different values of K and plotting the results showed how tuning hyperparameters can optimize model performance.

Results



Trained on 7.7 million records

Performance Measures

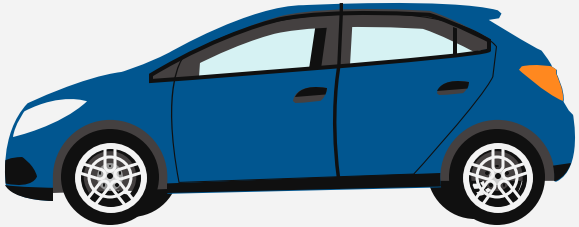
- Accuracy: 86.7%
- Precision: 79%
- Recall: 86.7%

Largest Mutual Coefficients

- Distance: 0.123
- Visibility: 0.225
- Wind Speed: 0.018

Conclusion

- **Project Objective**: Developed a K-Nearest Neighbors (KNN) model to predict the severity of traffic accidents based on critical features like time of distance, weather conditions, and visibility.
- **Key Processes:**
  - Data Preprocessing: Cleaned and prepared the dataset, including handling missing values and encoding categorical features.
  - Exploratory Data Analysis (EDA): Identified key patterns and relationships in the data.
  - Model Tuning: Optimized hyperparameters to improve accuracy and performance.
  - Evaluation: Assessed the model using relevant metrics to validate predictions.
- **Results & Impact**:
  - Identified factors correlating with high accident severity.
  - Demonstrated the potential to inform road safety policies and improve emergency response.
  - Showcased how predictive analytics can enhance public safety.



Thank you