A comprehensive study of US traffic accident evaluation and prediction

Yaxin Deng, Kefei Wang, Minxing Guo, Yunpeng Cheng, Wenshan Zheng

Introduction

➢ **What is the problem** ① Current traffic accident prediction is limited by the heterogeneity of traffic accident data and scale. ② Existing models are designed on city and town levels and are limited in explanatory powers due to the dynamic nature of traffic accident data. ③ Objectives of this project is to evaluate US traffic accidents based on various environmental and geographical conditions

➢ **Why is it important** A more accurate method of occurrence and severity prediction and granting the accessibility of an interactive risk prediction platform to the public can allow interested parties to take on precautionary measures and actions to reduce the risk.

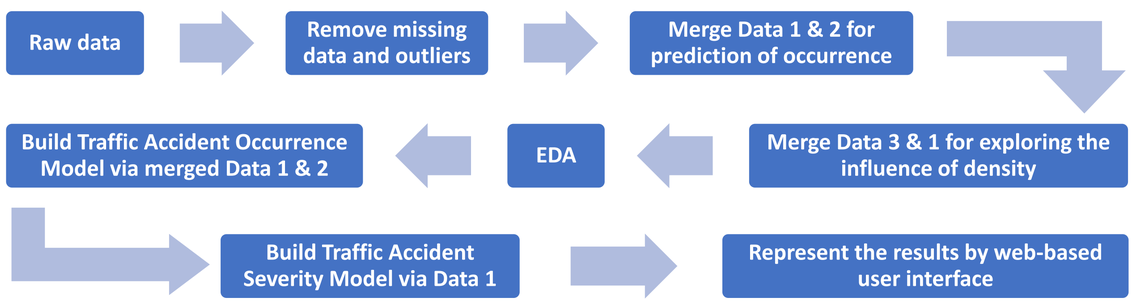
Data Source

➢ **How data was collected** ① All data are directly sourced from the web and government databases. ② The timeframe for the datasets used are from 2016 to 2021.

- **Data 1**: The US country-wide traffic accident dataset
  Includes features such as time, location, weather condition, etc.
  1.5 mil Records | **47 Attributes** | **542 MB**
- **Data 2**: Large-Scale Traffic and Weather Events Dataset
  Includes features such as traffic type, time, location, weather condition, etc.
  4 mil Records | **19 Attributes** | **1.2 GB**
- **Data 3**: Population density by city and by zip code
  Includes population density and location information
  61K Records | **17 Attributes** | **10 MB**

Approach

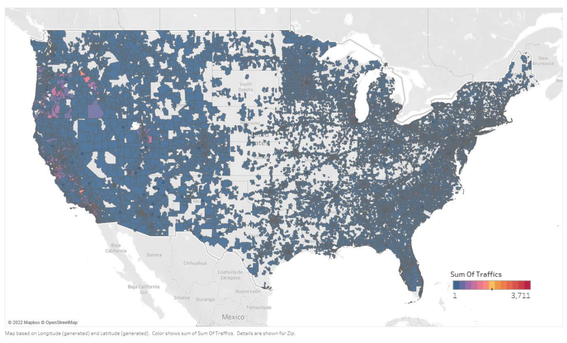
➢ **What was the Team's Approach and How does it Work**

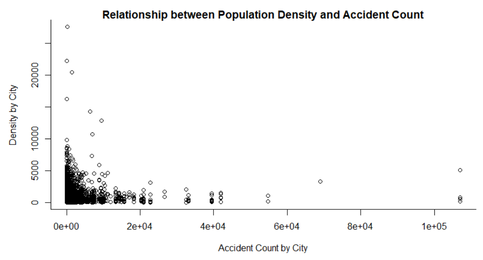


➢ **Why does it solve the Problem**

- Our method allows for drivers to analyze and forecast trip risk based on traffic accident occurrence and severity levels with a user-friendly GUI
- Our method combines comprehensive, up to date traffic accident data and dynamic weather data with the predictive analysis of machine learning (classification, decision trees, random forest, regression)

➢ **Visualization:** The relationship between population density and traffic accident



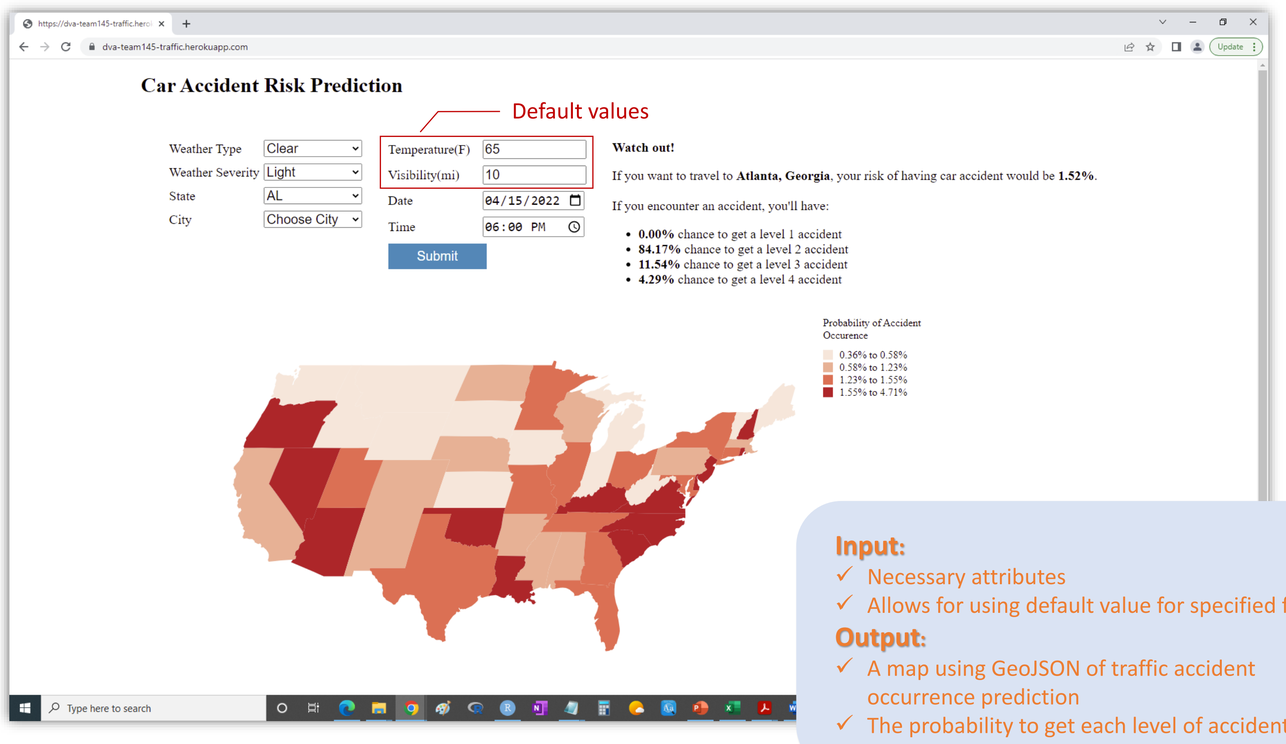


- Most accidents are congregated along the west and east coast where population density is higher. However, the scatter plot did not suggest an obvious positive linear relationship between the two.

Interactive Web-accessible Platform

➢ We build a web-based user interface using python Flask and JavaScript, specifically the D3 library, and published it via the Heroku app.

➢ The standard default value is the most prevalent data point of all for that specific category.



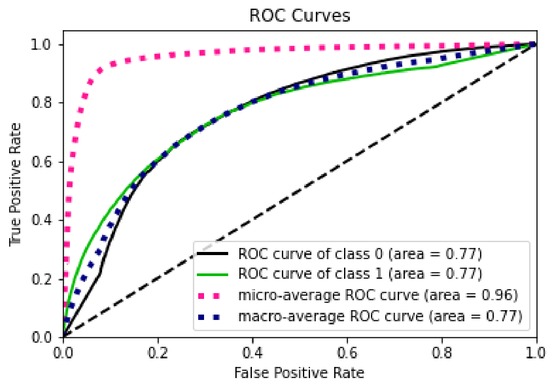
Input:
- ✓ Necessary attributes
- ✓ Allows for using default value for specified field

Output:
- ✓ A map using GeoJSON of traffic accident occurrence prediction
- ✓ The probability to get each level of accidents

Experiments and Results

➢ **How was the Project Evaluated**

- **Traffic Accident Occurrence Prediction** Random Forest Model trained on 20%, validated on 16%, and tested on 64% of dataset
- **Traffic Accident Severity Prediction** Four models were built: Logistic Regression, Decision Tree Classifier, Linear SVC, and Random Forest Classifier
  - Best performance is the Decision Tree Classifier model with max depth = 10
  - $Numerical\ Colum = Severity + Temperature + Humidity + Pressure + Visibility + Wind\_Speed + Hour + Day\_of\_Week + Month$
  - $Categorical\ Column = City + Weather\_Condition + Civil\_Twilight$



Model	Logistic Regression	Decision Tree Classifier	Linear SVC	Random Forest Classifier
Accuracy	0.934997	0.937548	0.9351	0.840936
F1-score	0.903863	0.917452	0.903808	0.877676
Precision	0.892018	0.911884	0.906701	0.934114
Roc_Auc	0.776316	0.834185	-	0.906396

➢ **What are the results**

- Traffic Accident Occurrence Prediction achieved an accuracy of 91.6% and an AUC of 76.6%
- Traffic Accident Severity Prediction achieved an accuracy of 93.7% and AUC of 83.4%

➢ **How does the Team's Method Compare to Other Methods**

- The team's method is unique in its predictions of traffic occurrence and severity considers more comprehensive and dynamic attributes for the entire country

➢ **What's new**

- Traffic accident occurrence prediction using comprehensive dataset that consists of US country-wide traffic incidents including accidents, constructions, and congestions
- Utilization of more complex traffic condition and weather dynamic for traffic accident severity prediction
- Introduction of US population density into traffic accident analysis for the first time