PREDICTING CHICAGO CAR ACCIDENTS & INJURIES

By Titus Mutuku

Phase 3 Final Project

2023

MORINGA
Discover · Grow · Transform

MORINGA
Discover · Grow · Transform

PROBLEM

- **Four-day average results in 1000+ car accidents:**
  - On average, over 1000 car accidents occur within a span of four days.
- **Total individuals involved exceed 2000:**
  - The combined count of drivers, passengers, pedestrians, and cyclists involved in these accidents surpasses 2000 people.
- **45% experience minor to fatal injuries:**
  - Approximately 45% of individuals involved in these accidents are expected to sustain injuries ranging from minor to fatal.
- **Objective: Predict factors causing accidents leading to injuries:**
  - The goal is to predict the specific factors or causes of accidents that result in injuries.

Significance of the Project

- Reduce Injuries/fatalities
- Make the streets safer for drivers, passengers, pedestrians and cyclist
- Alleviate work for Emergency Responders and Officers
- Reduce Traffic
- No accidents, no congestion
- Shorter travel times
- Safety
- Safe environment for the citizens of Chicago
- Happy Drivers

THE DATA

- Traffic Crashes – Vehicles- https://data.cityofchicago.org/d/68nd-jvt3
- Traffic Crashes – People- https://data.cityofchicago.org/d/u6pd-qa9d
- Traffic Crashes – Crashes- https://data.cityofchicago.org/Transportation/Traffic-Crashes-Crashes/85ca-t3if

Approach

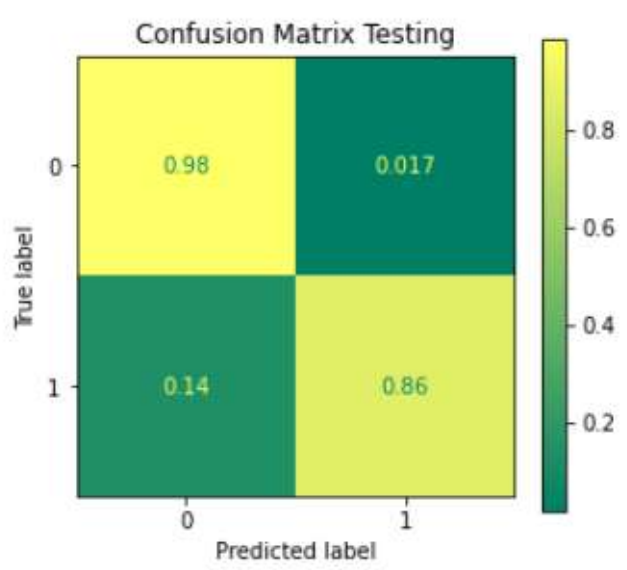
- **Data Acquisition and Observation:**
  - Start by obtaining the necessary data and carefully observe its values.
- **Data Cleaning:**
  - In the scrubbing phase, eliminate columns that are irrelevant or redundant, and address any missing values by replacing them.
- **Exploratory Data Analysis:**
  - Explore the data by generating visualizations to gain insights, and consider converting data types if needed.
- **Modeling:**
  - Move on to the modeling stage where various types of models are created, and the relationships between different variables are thoroughly understood.
- **Interpretation:**
  - Finally, interpret the model's outcomes, providing a comprehensive explanation of what insights the model conveys.

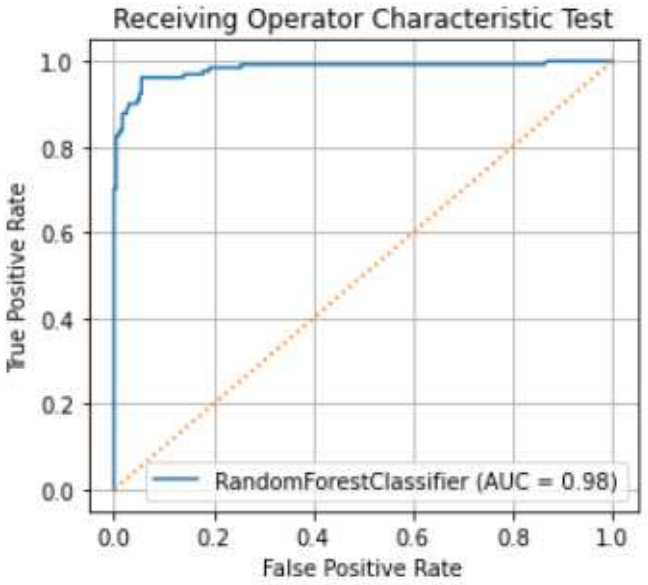
MODEL

The Random Forest Classifier, with an accuracy rate of 89%, stood out among other models considered. Its performance either matched or surpassed alternatives.

Random Forest CLASSIFICATION REPORT TESTING

	precision	recall	f1-score	support
0	0.93	0.98	0.95	235
1	0.97	0.86	0.91	131
accuracy			0.94	366
macro avg	0.95	0.92	0.93	366
weighted avg	0.94	0.94	0.94	366





Conclusion and Recommendations

- Based on my classification model, the type of accidents that appears to occur most frequently is drivers colliding with pedestrians or cyclists on the street.
- Analyzing the data indicates that pedestrians typically do not wear contrasting or reflective clothing, and the majority of accidents take place in the afternoon or during rush hour.
- Furthermore, my findings reveal that a significant number of accidents occur in speed limit zones labeled between 30-40 mph.

Recommendations:

- Pedestrians and cyclists should heighten their awareness and wear contrasting colors, particularly during rush hour.
- To enhance safety, the city can implement measures such as increasing the number of traffic signs and creating lanes that are more accommodating for pedestrians and cyclists.
- Additionally, considering the higher frequency of accidents during the afternoon and rush hour, the city might consider lowering speed limits during these times or implementing increased patrols in zones with speed limits ranging from 30 to 40 mph.

Thank You!
Any questions?