Hotspot Prediction of Severe Traffic Accidents in the Federal District of Brazil

Vinicius Lima* & Dr. Vetria Byrd

***PhD student at Purdue University**
vlima@purdue.edu

PURDUE UNIVERSITY | Polytechnic Institute

IARIA

Background

Vinicius Lima

- Currently a PhD student at Purdue University, where main focus is on Data Visualization.
- Previously, Forensic Scientist at the Civil Police of the Federal District in Brazil.
- 3 years experience with Crime Scene Investigation (Crimes Against Life).
- 6 years experience with Traffic Accident Reconstruction.
- From 2019 – 2021, Head of the Traffic Accident Reconstruction Sector situated at Brasilia's Criminalistics Institute.
- Master in Computer Forensics and Bachelor in Network Engineering from the University of Brasilia.

Current Research

- Today the focus is on identifying patterns in crime data.
- Some previous work consists of patterns in domestic violence case, using machine learning to output pattern of sequence events based on police reports.
- Use of Natural Language Processing (NLP) for information retrieval from crime scene forensic reports.

Background – Traffic Accident Analysis

PURDUE UNIVERSITY | Polytechnic Institute

Motivation for this paper

- When working as a Traffic Accident Reconstructionist there was a sense among the experts that similar accidents were happening at the same spots.
- That led me to an empirical analysis of reports to understand the problem.
- We developed a dashboard tool to visualize statistics and map the main hotspots of accident in the Federal District of Brazil.
- This study clearly pointed the pattern behavior of accident in the main hotspots and made us question if Machine Learning (ML) could be applied to identify these patterns in an automatic and efficient way.

In red, location of the Federal District in Brazil.

Research Questions

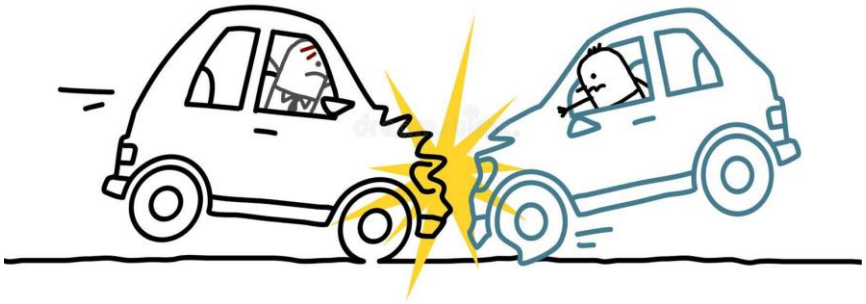
- Moreover, we want to check whether we can use ML to predict the future number of accidents in a region, in other words, predict accident hotspots.
- We also questioned if weather parameters influence hotspot prediction.
- We are mostly interested in the following research questions:

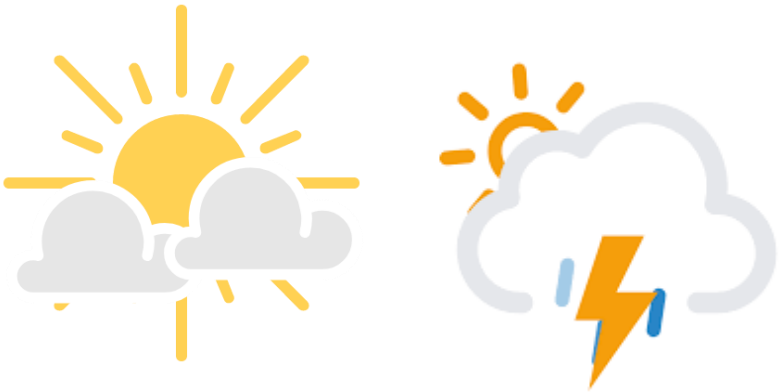
- 1. Is it possible to the predict number of accidents in defined regions (i.e., hotspots) with traffic accidents and weather information?

- 2. If yes, what are the main features involved in the prediction outcomes?

Dataset – Civil Police of the Federal District of Brazil

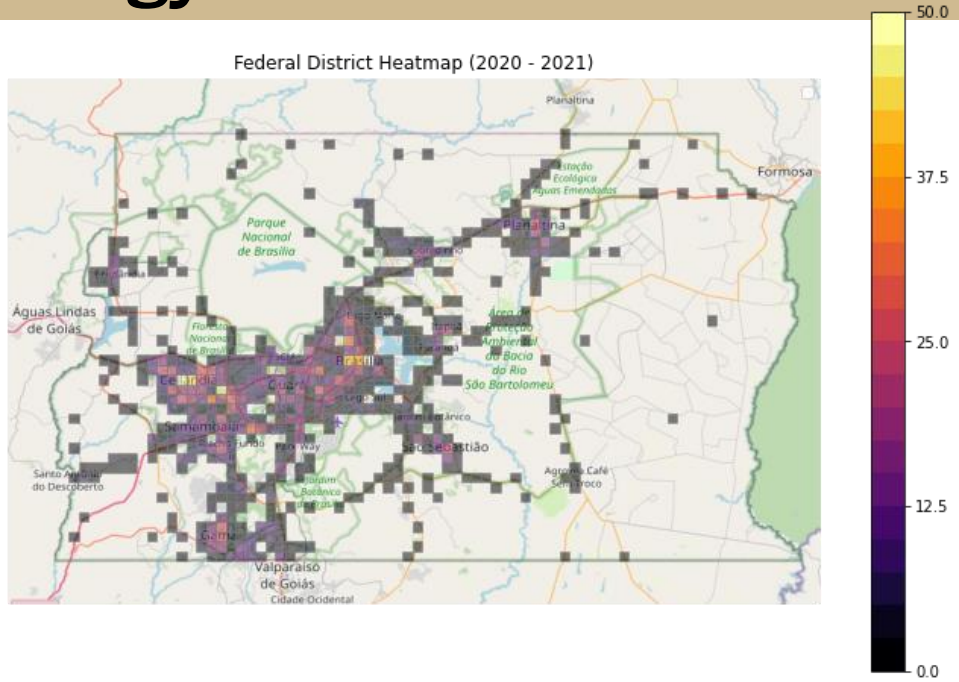
- Type of Accident
  Date **(2020 - 2021)**
  Time
  Latitude
  Longitude
  Speed limit
  Number of vehicles involved
  Type of vehicle
  Average Temperature
  Humidity
  Pressure
  Wind
  Cloud amount
  Solar Insolation
  Rain Precipitation





Data Source: Federal District Civil Police and Meteorology Institute of Brazil

Methodology



- We divided our map into 80 regions, which gives us a grid with dimensions of 4.6 miles by 6.0 miles.
- This value can be set accordingly to the authorities' necessity.
- To each accident at the corresponding accident time, we added the closest weather station information.

Methodology

Longitude, Latitude, Street Speed limit, Temperature, Humidity, Pressure, Wind, Solar Radiation, Rain, Year, Month, Workday, Hour of the day.

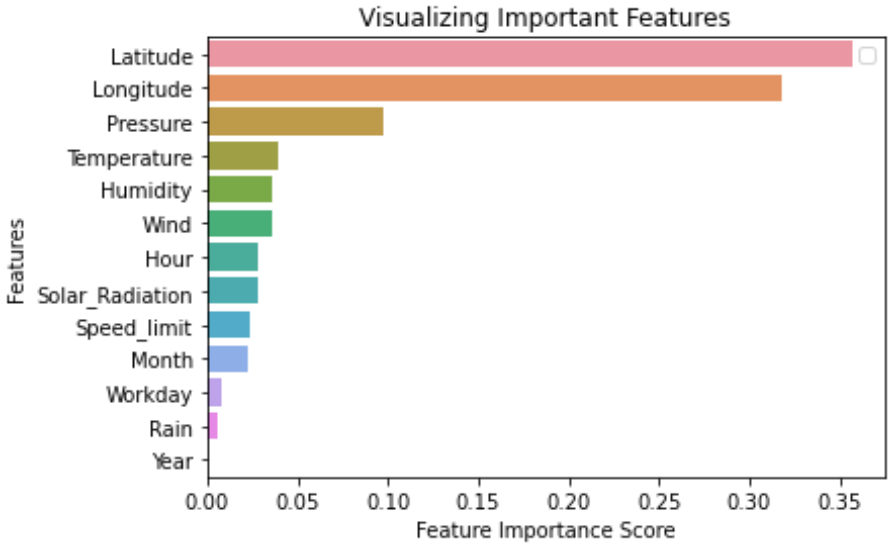
1. **Random Forest**
2. **Multi-layer Perceptron Classification**

Number of accidents in specific regions (grids).

Results and Discussion

Model	Type	Accuracy
Random Forest	Regression	98.8 %
Multilayer Perceptron	Regression	79.6 %
Random Forest	Classification	94.1 %
Multilayer Perceptron	Classification	88.2 %

Results and Discussion



- This figures shows the ranking of features used in the Random Forest algorithm. To each feature, it is assigned the proportion of which each contributes to the outputs.

- Here we observe that weather information does not add much to the prediction as the location of the accidents.
- Based on our findings, we can conclude that if we want to understand more about the causes of the accidents, we need to study deeper the characteristics of the location of the accident rather than looking at other variables, such as the climate conditions.

Conclusions

- For this work, we tested traditional machine learning algorithms, analyzing the performance as regression and classification problems. Random Forest and Multilayer Perceptron outputted high accuracies, with the former outperforming the latter.
- Random Forest model also revealed that location attributes, such as longitude and latitude, were considered more important features to predict the number of accidents than the meteorology parameters, such as rain and temperature, and time parameters, such as hour and day. In other words, the place where the accident happened is a key factor contributing to the identification of hotspots. Public authorities and researchers need thus to focus their studies primarily on the site characteristics, for example, suggesting road infrastructure interventions.
- Moreover, we demonstrated that data from forensic teams can be used not only for traditional one-by-one investigations but also to understand patterns and associations in the big picture, delivering useful information for public policies. This work aligns with the new era of Forensic Intelligence, where data science plays a key role.

Vinicius Lima
PhD stundet
Polytechnic Institute, Purdue University, IN, USA.
https://www.linkedin.com/in/vinnyolima/

polytechnic.purdue.edu

/ TechPurdue

PURDUE UNIVERSITY | Polytechnic Institute