

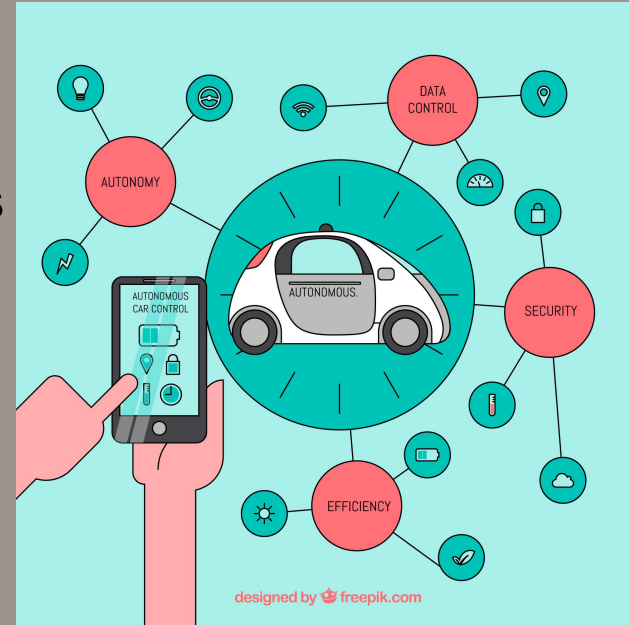
Project 4: Machine Learning Model for Car Crash Data

By: Kevin Carney, Paris Lee, Samantha Schutz, and Joy Weishan
October 3, 2024



Purpose/Inspiration

- Automation impacting every industry
 - Car manufacturing no exception
 - Is this a safe use of AI?
- National Highway Traffic Safety Administration mandated reporting of vehicles with these automated systems
 - Automation in vehicles still heavily in development
- Anyone using public roads possibly impacted
 - Many of our group work within healthcare
 - One of our group works for GM
- Dataset has not been used and a good practice for real world application



Design Concepts/Data Cleaning

- More neutral web background used with greys, as it most appropriately fits the theme
- Tableau visualizations had a wider array of colors, as visualization complexities created a need for more colors to be utilized
- Multiple notebooks required for machine learning and 2 different Tableau dashboards
- Many null values within dataset requiring dropped rows of data
- Many conversions required to use data in a machine learning context



Research Questions



1. Will certain weather conditions cause accidents resulting in whether someone gets injured or not? Is this predictable?
2. Will different vehicle conditions impact the result of an accident?
3. When do these accidents happen the most?
4. What are other environmental conditions that impact the outcome of an accident?

Live Demo



Limitations/Biases

- Not all states had data
 - Majority of the data came from California (San Francisco), Arizona, and Texas
- Machine learning portion: outcome heavily imbalanced toward “No Injuries” vs “Injuries”
 - In addition to the imbalance of the outcome, data had little variance in terms of conditions also making predictive modelling effectively nearly impossible
- Many columns dropped for ease of data manipulation and visualization
- Many ‘null’ values to deal with (filled with most sensible alternative value or dropped altogether)
- Unknown if other vehicle involved and to what capacity
- No standardized reporting methods across manufacturers
- Possible for multiple reportings on same vehicle

Conclusions/Future Work

- Dataset proved to be difficult to make a truly meaningful predictive model from due to imbalances of outcome and varying conditions
- 6 million car crashes a year – there are not many related to automation at this point
- Pull out more data and keep more columns to look further into the type of automated system and to see if different updates reduced number of accidents
- Needs to be better reporting to determine cause of accident
 - Other vehicle involved?
 - Other unusual circumstances pulled out from “Narrative” column
- Column added on if automation system being used appropriately
- Comparing different versions of systems to determine if there were improvements made

Works Cited

Image on slide 1 & 2: designed by freepik.com

Image on slide 4: designed by [vectorstock.com/20383246](https://www.vectorstock.com/20383246)

“Park Assist Technology: A History ” by Yamuna Bindu Nov 11, 2021
<https://blog.getmyparking.com/2021/11/11/park-assist-technology-a-history/>

“Why August is 'the most dangerous driving month' of the year” Published Monday, July 30, 2018. CTV News.

Xpert Learning Assistant:

- [Xpert Learning Assistant](#) - Used for debugging certain codes.

chatGPT:

- chatGPT- Used for debugging certain codes.

“NHTSA Standing General Order on Crash Reporting Dataset Source” DOT of US government. June 2022.

“Using Machine Learning Models to Forecast Severity Level of Traffic Crashes by R Studio and ArcGIS” by Bara’ W. Al-Mistarehi, Ahmad H. Alomari, Rana Imam, Mohammad Mashaqba on April 19,2022.