DA Data Solutions

https://github.com/schahmatist

https://github.com/Auustiino

Overview

Using data science techniques to examine traffic accident data

Key Question:

Is it possible to use raw data and machine learning techniques to make significant inferences about the causes of car accidents?

Goals

Build a model to infer if a person is innocent in causing a car accident

Determine what information is useful in making this inference

Data Source

data.cityofchicago.org

-Data from the City of Chicago

-Data is recent up to 4 / 13 / 2022

Data understanding

Crashes

-Traffic Control Device
-Device condition
-Roadway Surface Condition
-Lighting Condition

Vehicles

-First Contact Point
-Vehicle Defect
- Number of Passengers

People

-Driver Vision
-Age
-Maneuver

Target

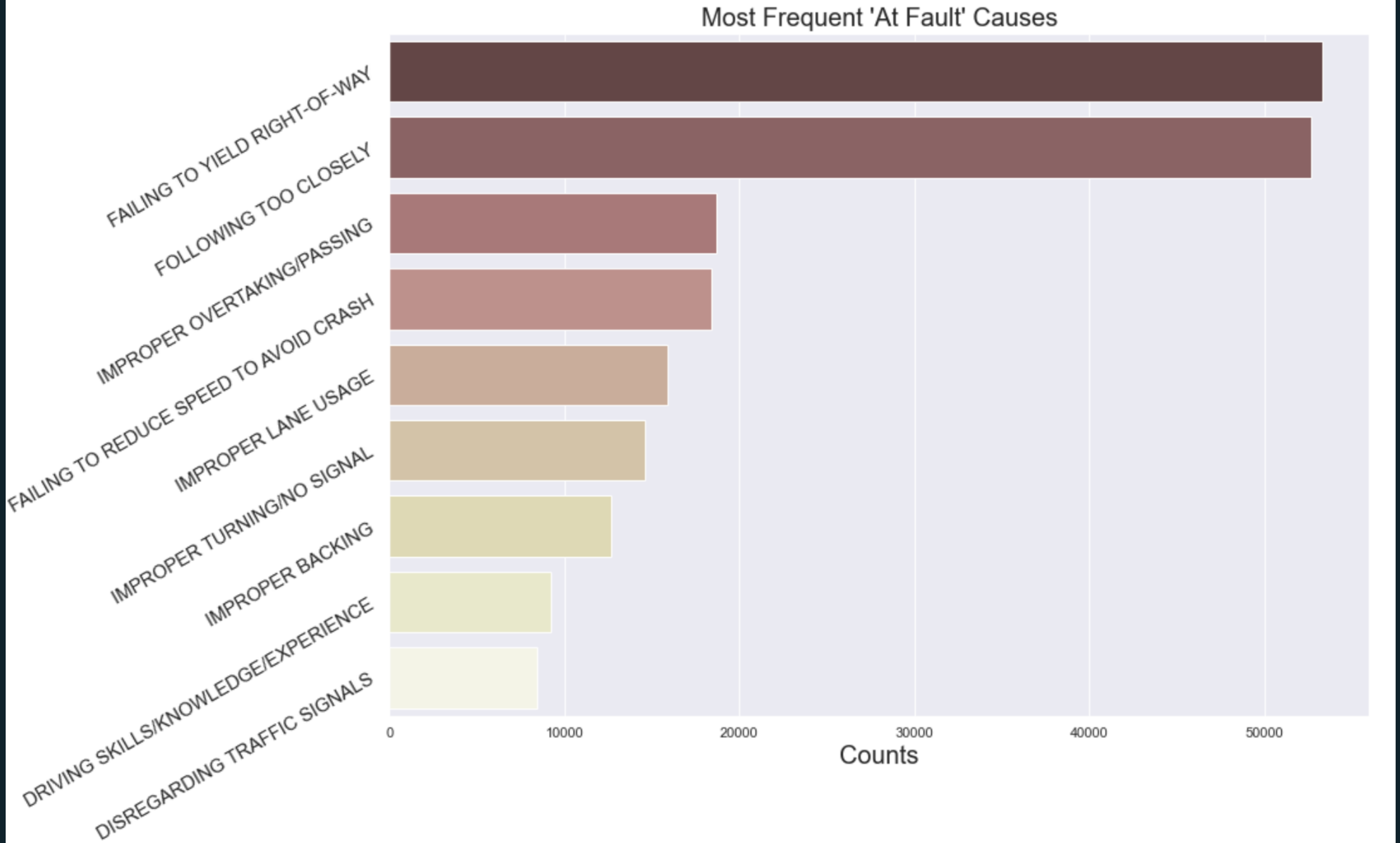
What is Fault?

At Fault

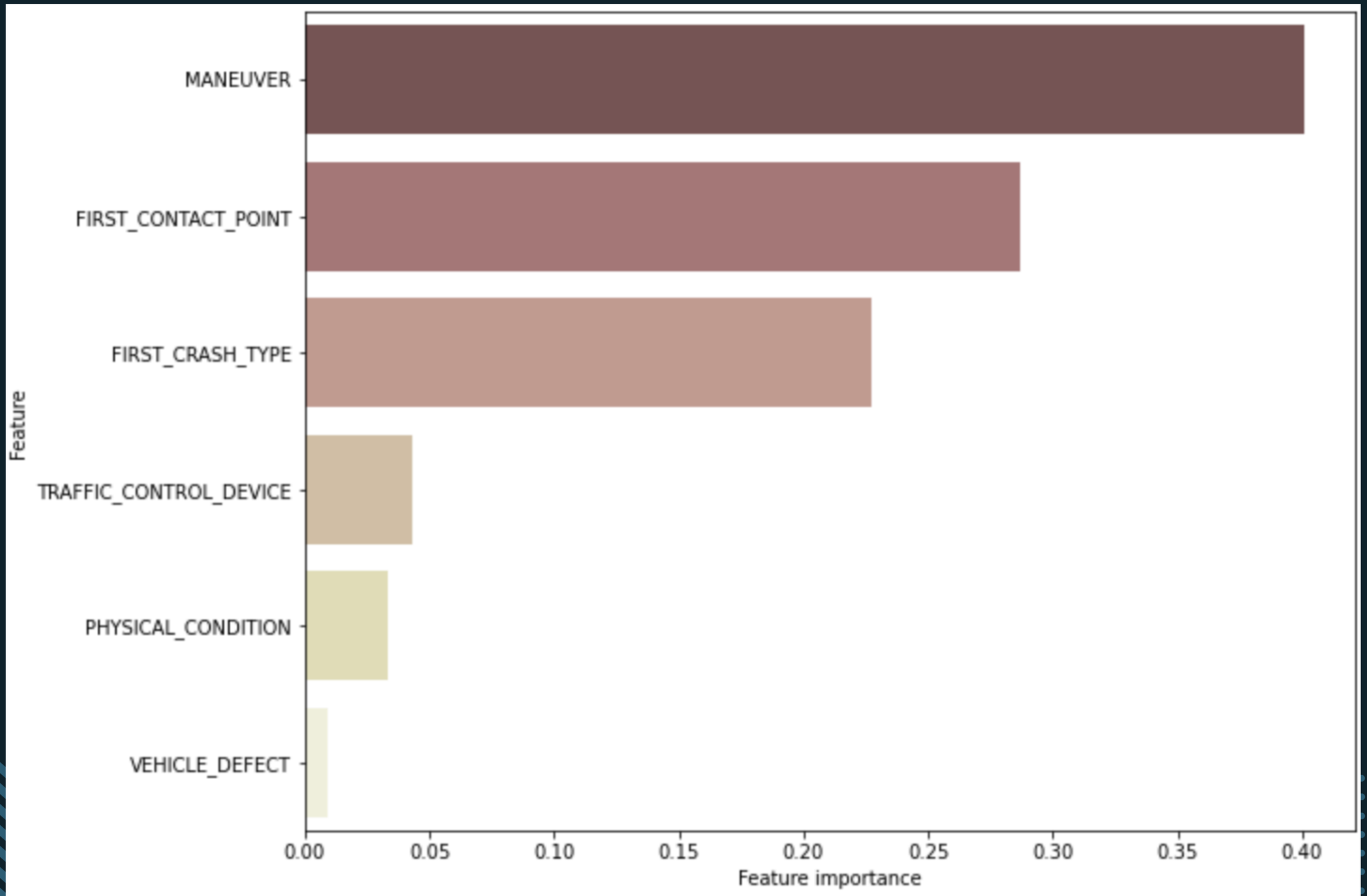
Under the Influence of Alcohol / drugs
Exceeding Authorized speed limit
Following too closely

Not at Fault

Road Engineering / Surface /Marking defects
Weather
Animal


Most Frequent 'At Fault' Causes
FAILING TO YIELD RIGHT-OF-WAY
FOLLOWING TOO CLOSELY
IMPROPER OVERTAKING/PASSING
FAILING TO REDUCE SPEED TO AVOID CRASH
IMPROPER LANE USAGE
IMPROPER TURNING/NO SIGNAL
IMPROPER BACKING
DRIVING SKILLS/KNOWLEDGE/EXPERIENCE
DISREGARDING TRAFFIC SIGNALS
0
10000
20000
30000
40000
50000
Counts

Significant Predictors

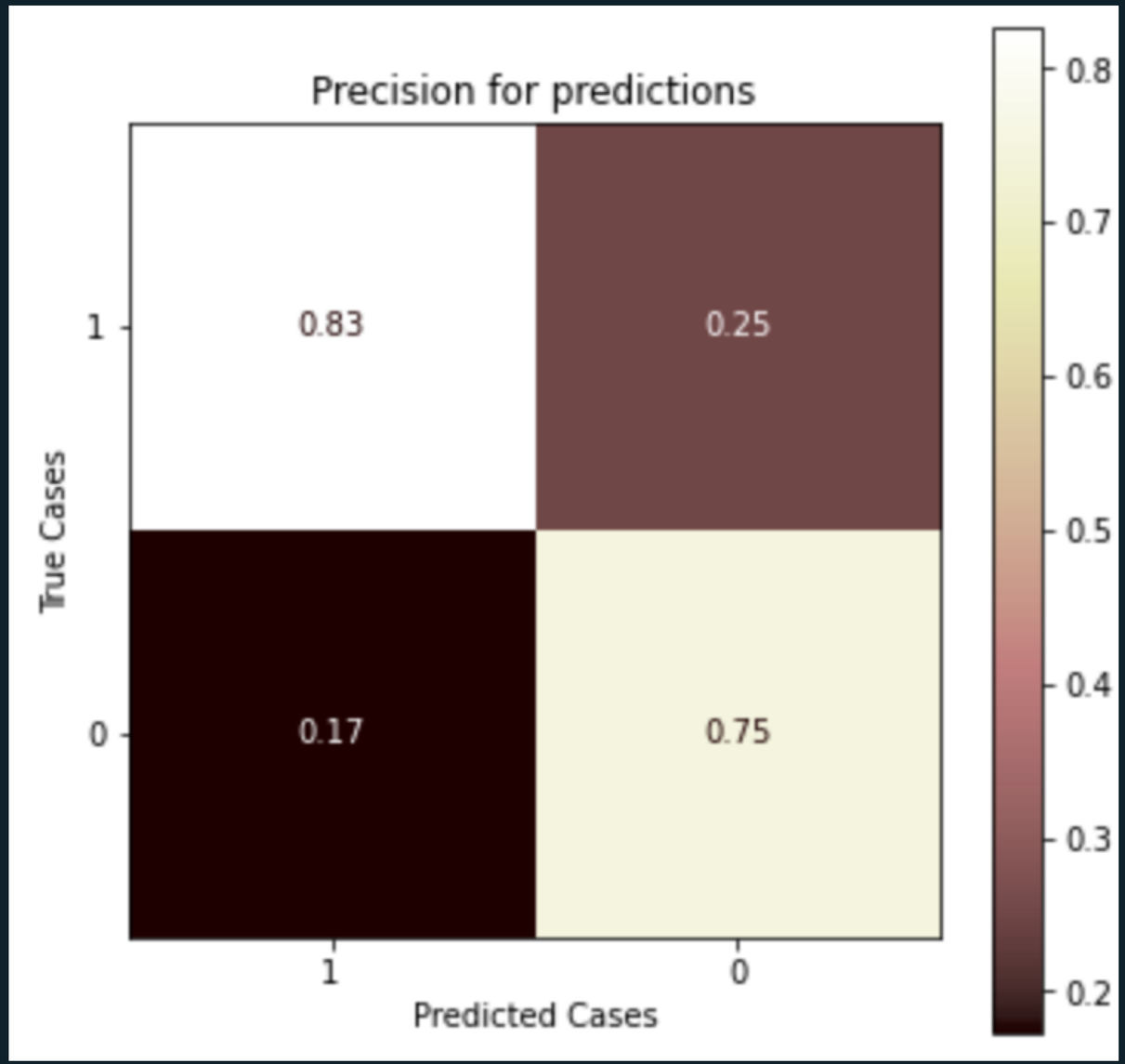


Model

XGBoost

Precision : 83%

Out of 100 people who our model claim are innocent, 83 of them will be innocent

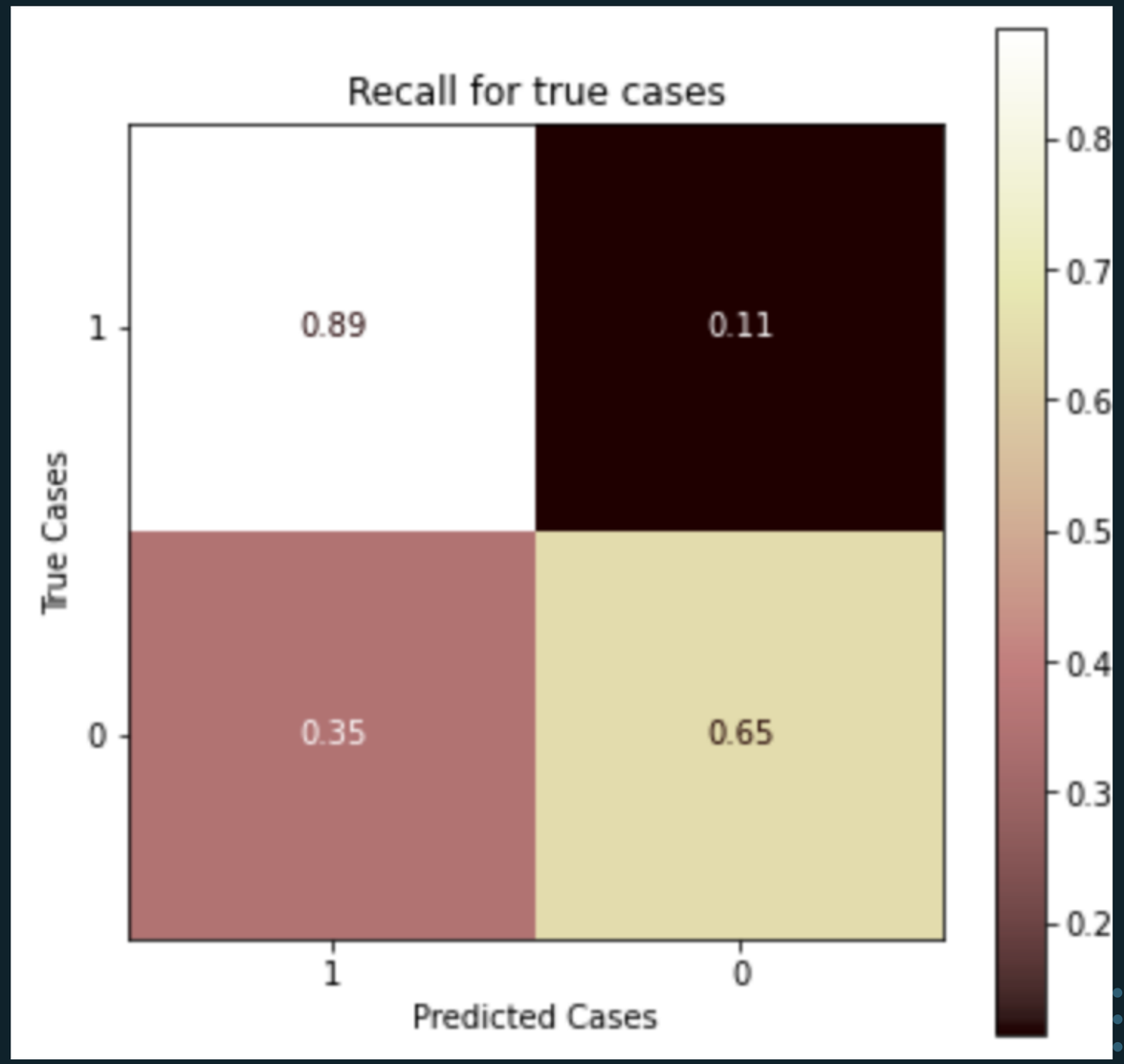


Model

XGBoost

Recall : 89%

Out of 100 True innocent cases our model correctly identifies 89 of them



Insight

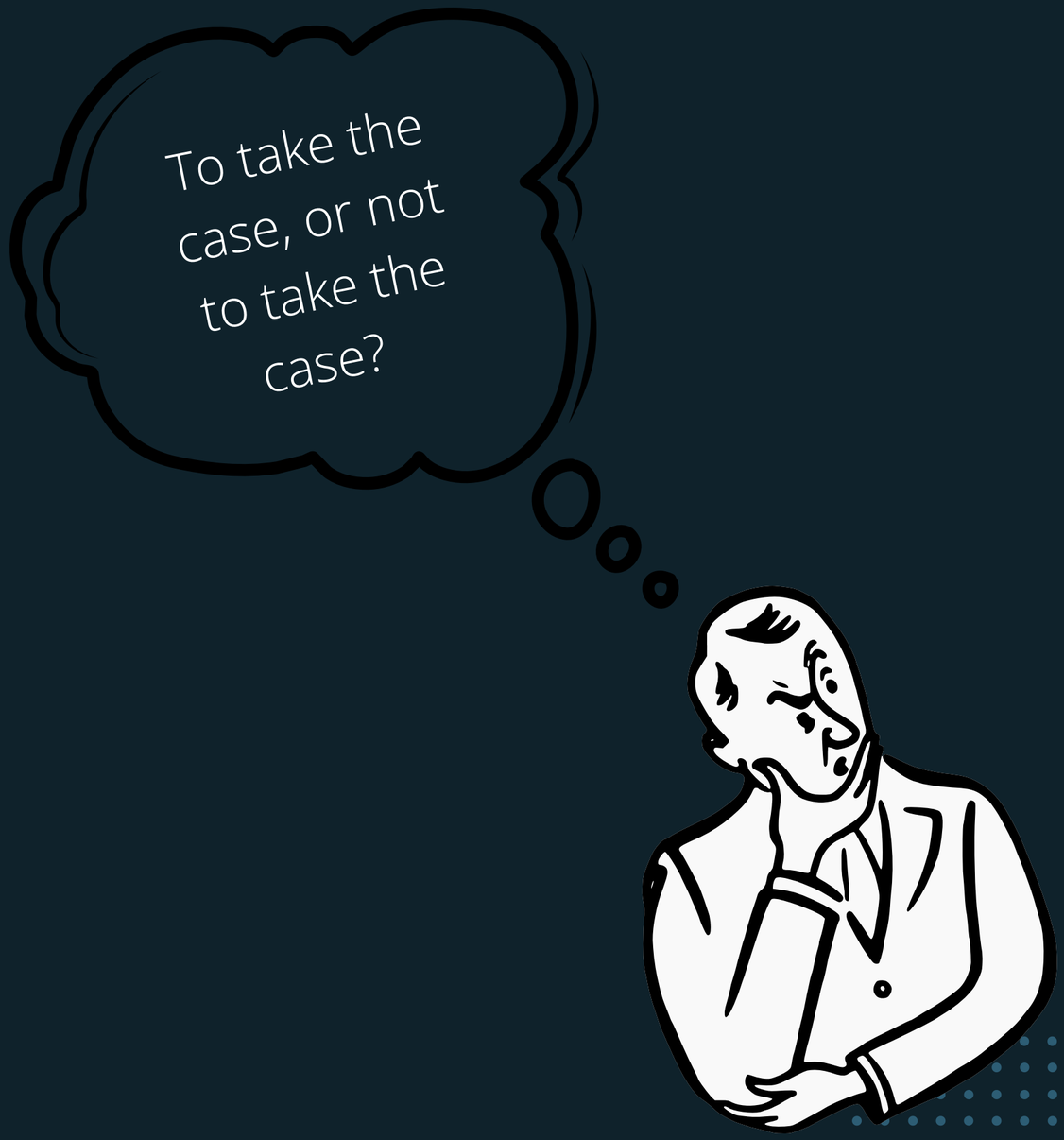
Who cares?

<u>Defense Attorneys</u>

- Can easily determine probability of client innocence

<u>Insurance Companies</u>

- Can easily determine when not to increase rates



Case Example

Defendant has been accused of causing a car accident due to following the car in front too closely.

Our model is able to predict with an 83% percent accuracy whether not this is statistically probable

Next Steps and Improvements

- Given more data, and time to perform feature cleaning and engineering, model accuracy could be improved
- Analayse the feature selection and cleaning process and reevaluate biases

Thank You

https://github.com/schahmatist

https://github.com/Auustiino