Predicting severity of a car accident

IBM CAPSTONE PROJECT

General information

- Each year many car accidents take place
- There are costs associated to them:
  - Human cost
  - Material cost
  - Business cost
- Predicating chance and severity of an accident would be beneficial for everybody
- This project could be interesting for car and GPS navigation software companies

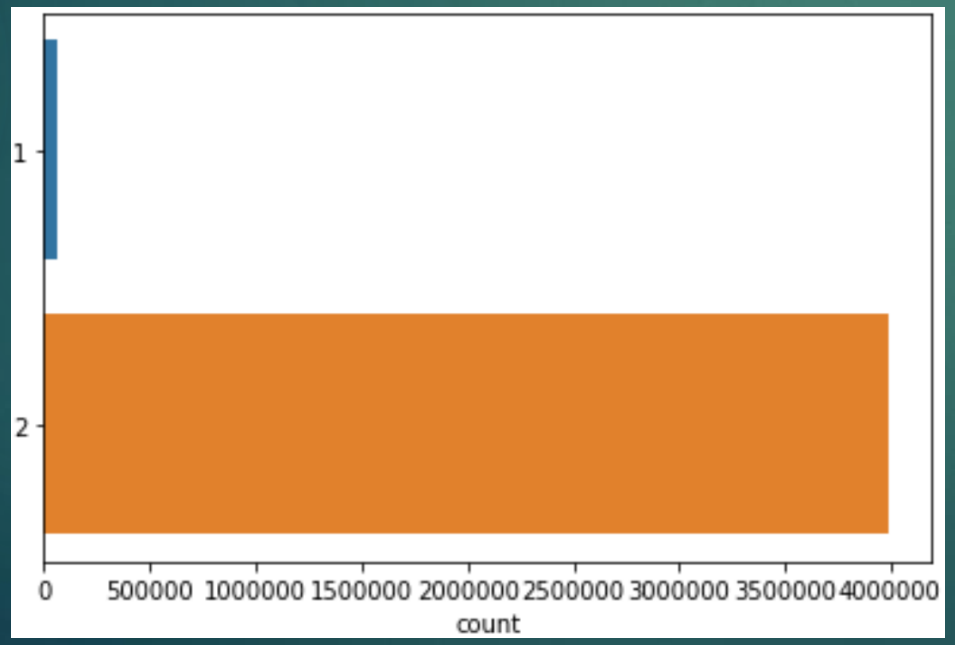
- The aim of this project is to predict **severity** of a car accident

Data

- Data used is made available by Transport Canada and Statistics Canada
- Data contains three levels:
  - Collision level of data
  - Vehicle level of data
  - Person level of data
- After cleaning data set size is 4m records and 12 features

Target variable

- Collision severity:
  - 1: Collision producing at least one fatality
  - 2: Collision producing non-fatal injury
- There are many more non-fatal accidents



Feature selection

- Initially there were 22 columns (21 features and 1 target variable)
- Features not relevant to answer business question were dropped
  - Collison level: year of collision
  - Vehicle level: vehicle id
  - Person level: all fields

Model choice

- The aim of the project is classifying an accident into one of two categories -> **Classification**

- The model to be used is **Decision Tree**

Model development

- Data split into two data sets: training and test

```
# Split data sets (X and y) into train and test sets
X_trainset, X_testset, y_trainset, y_testset = train_test_split(X, y, test_size=0.3, random_state=3)
```

- The model was trained based on training set

```
# Train Decision Tree algorithm
accidentTree.fit(X_trainset,y_trainset)
```

- After testing accuracy of the model by comparing prediction with actual values from test set

```
# Calculate evaluation
from sklearn import metrics
import matplotlib.pyplot as plt
print("DecisionTrees's Accuracy: ", metrics.accuracy_score(y_testset, predTree))
```

```
DecisionTrees's Accuracy:  0.9835718789899772
```

Conclusion

- Analysis of relationship between severity of an accident and
  - Collision level data
  - Vehicle level data

- This model can be very useful for GPS software companies
  - Ex 1: showing a warning when chance of an fatal accident is high
  - Ex 2: taking into account chance of a fatal accident when calculating route