

The background image shows a person's arm and hand in a blue and red plaid shirt, gesturing while sitting at a desk. A laptop is open in front of them, displaying a dashboard with various charts and graphs. A smartphone is lying on a notebook on the desk. A thick, curved blue line with a slight gradient and a drop shadow sweeps across the image from the top left towards the bottom right, partially obscuring the background.

IBM Applied Science Capstone Project

Accident Severity Prediction

INTRODUCTION

Traffic accident can has caused over 1 million death in 2016-2019, and mainly cause teenager with age 17-30 years.

Why we need to predict accident severity ???

Predicting accident severity can be used to measure in advance about exact accident place, equipment, and personal staff thus can reduce the possibility of accident and save huge amount of lives each year

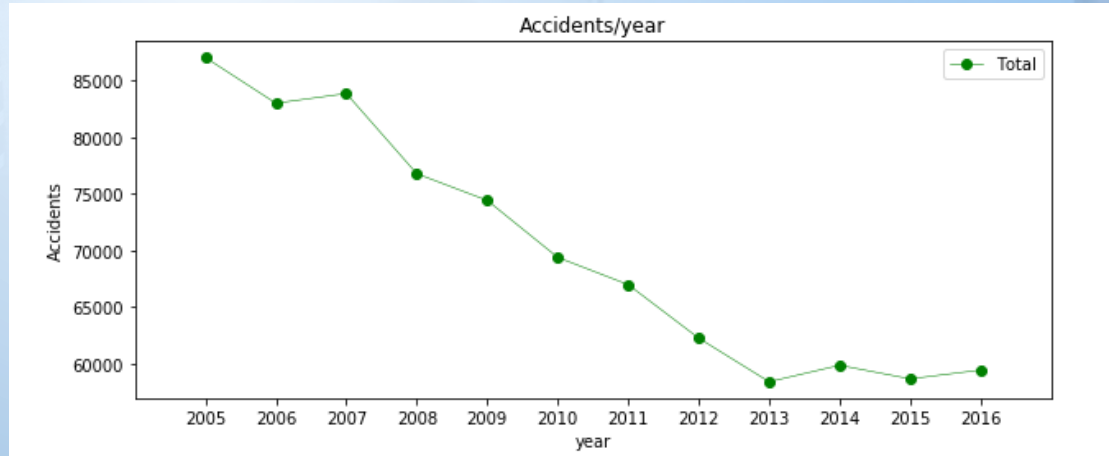
This project could be precursor to development of technology that can help reduce accidents and also improve road driver safety.

DATA

In this project we use data from Kaggle and use total 28 features selected. Here is 0 and 1 accident severity. 1 mean high severity, there are highly injured and even death. Based on the result, the measurement balanced enough.

```
Accidents classified:  
0      471695  
1      368290  
Name: sev, dtype: int64
```

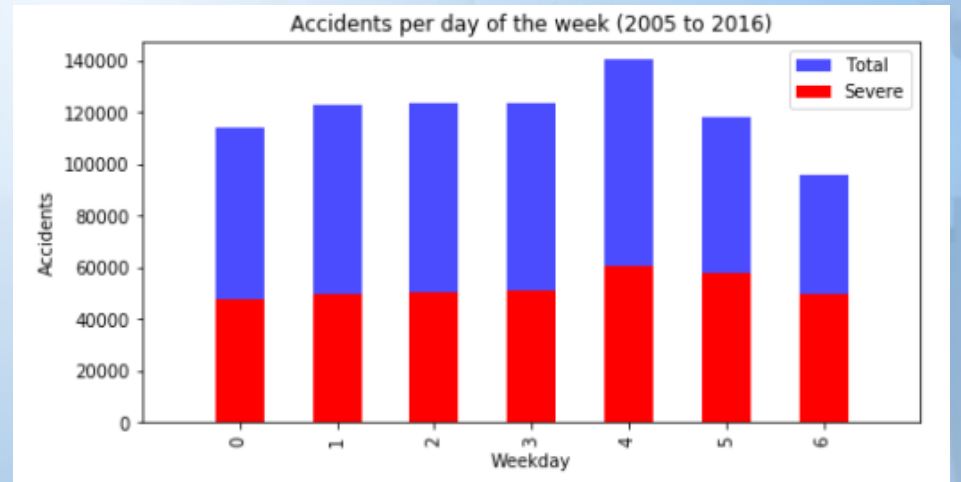
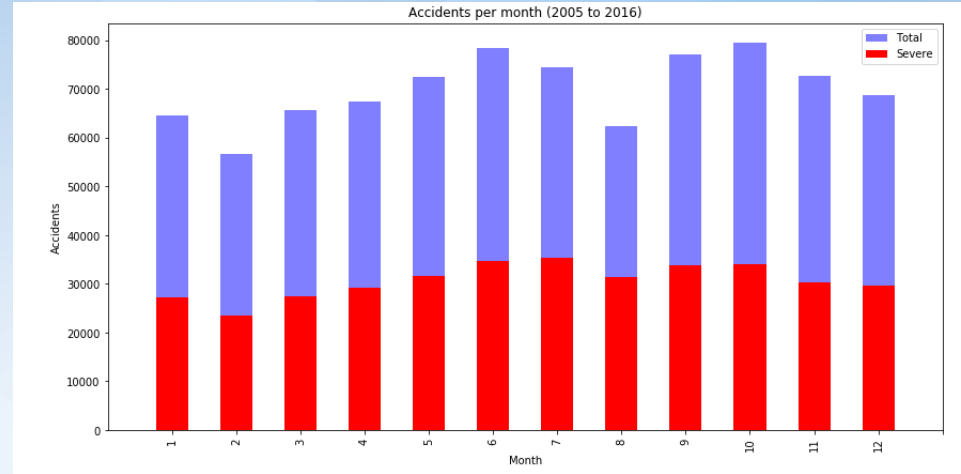
Based on the data, start from 2005, number of accident decreased and trend become stable



DATA

Highest accidents per month occur in October and June.

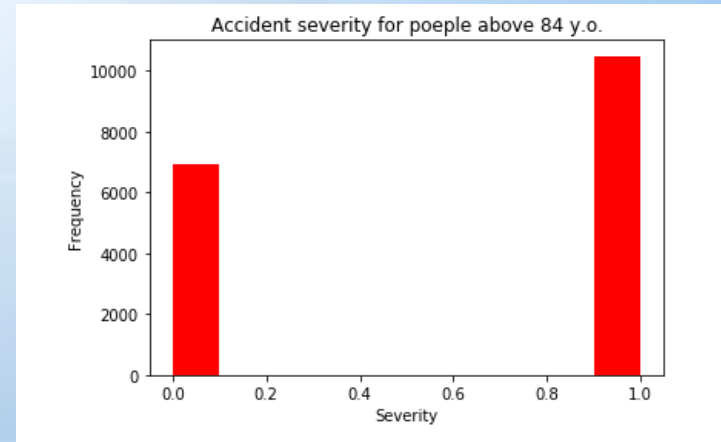
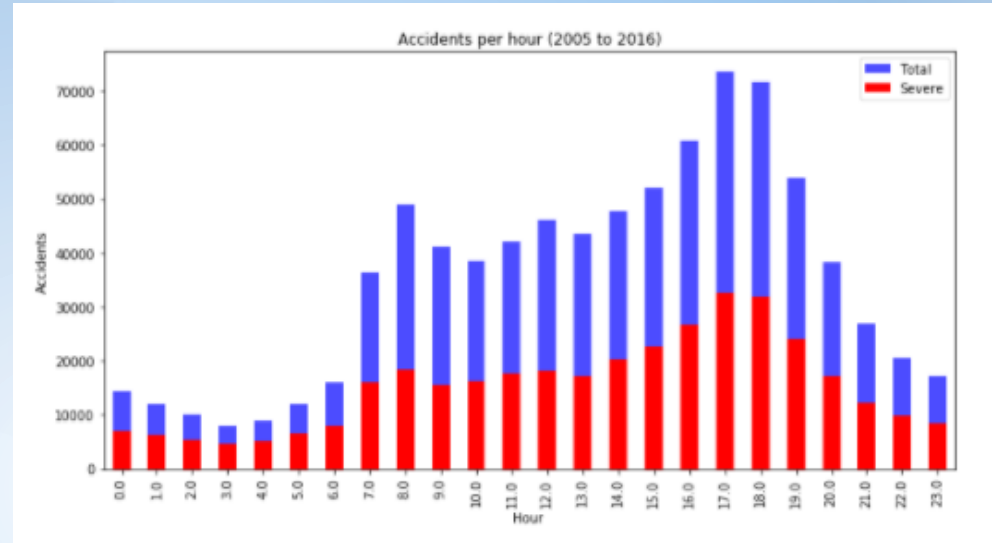
Highest accidents per day occur in Thursday



DATA

Highest accidents per hour occur at 5 pm

People above 84 year gives more likely highly injured (high severity)



MODEL DEVELOPMENT

In this project, two methods used to determine the predicted value: Random Forest and Logistic Regression.

With total 10 decision trees for Random Forest method. Total 4 models of RF used with different number of decision trees and max depth of decision tree model.

And $C = 0.001$ is used for Logistic Regression Method

RESULT

Algorithm	Jaccard	f1-score	Precision
Random Forest	0.72	0.72	0.72
Logistic Regression	0.66	0.65	0.67

Random forest so higher Jaccard Factor, F1-score, and Precision than Logistic Regression Method

We need to consider again different method (SVM, KNN, etc)

CONCLUSION

We need to add some other models to precisely predict traffic accident severity. We need to improve accuracy of the model too

In the future, this model still need improvement, and we need some additional features to increase prediction accuracy



THANK YOU