

# Capstone project

## **CAR ACCIDENT SEVERITY**



Trung Kien NGUYEN – August 2020



# Summary

- Business Problem
- Data
- Methodology
- Result
- Discussion
- Conclusion



# Business Problem

- A lot of accidents occur in **critical condition**.
- The police hope to know the **possibility** of getting into a car accident and how severe it would be.



*Require supervised machine learning to predict the severity based on historical traffic collision.*



# Data

## Required data

- Accident location
- Road conditions
- Weather conditions
- Junction
- The current date and current time
- Car speeding
- Number of people involved
- Light conditions
- Number of vehicles involved in

Data source: Seattle Department of Transportation (SDOT)



# Methodology

- Clean data by fill and drop the missing data.
- Data visualization and do statistics to see the influences of road condition, weather condition, etc. to severity.
- Split time to day of week and hour of day to see the connection between the time and severity.
- Balance data.
- Encode the categorical data.



# Result

	F1-Score	Jaccard score
KNN	0.64	0.47
Decision Tree	0.67	0.51
Logistic Regression	0.62	0.47

The most appropriate model for this problem is decision tree with highest score.



# Discussion

- The data is not highly correlated and highly imbalanced.
- With the imbalanced data set, it is easy to reach more than 70% accuracy.
- The model with balanced data is quite good (0.67 score)



# Conclusion

- Most of the algorithms are biased towards most frequent class.
- Need to balance the data set to give better results.
- Accident severity can be predicted, but no one feature that influences the accident severity.





Thank you for your attention