

ROAD ACCIDENT SEVERITY PREDICTION

IBM – Coursera

Data Science Specialization Capstone

Olajide ILELABOYE – September 2020



ROAD ACCIDENT CHALLENGES

- Traffic-accident is one major daily encounter by road users' in urban areas.
- Road accidents result to fatalities, injuries or just property damage.
- External factors like weather, road and visibility conditions sometimes reveal clues about severity of such accidents.
- The goal is to provide early-warnings by exploring available traffic data and build a machine learning model.
- Road users and Traffic administrators or controller may learn few tips from this presentation for proper planning.

DATA ACQUISITION AND CLEANING

- Traffic data collision csv by GISWEB provided by Coursera.
- Total of 194,673 rows and 39 features in raw dataset.
- Non-related features and those with "No Applicable" data were dropped.
- Clean data 7 features out which 4-features are use to create model for this project.
- The final dataset split into train set (70%) and test set (30%).
- The train dataset is used to train the model, and the test set is used to test the accuracy using Jaccard, F1-score and Log Loss

IMPACT OF WEATHER CONDITION ON ROAD ACCIDENT

Clear	111135
Raining	33145
Overcast	27714
Unknown	15091
Snowing	907
Other	832
Fog/Smog/Smoke	569
Sleet/Hail/Freezing Rain	113
Blowing Sand/Dirt	56
Severe Crosswind	25
Partly Cloudy	5

Name: WEATHER, dtype: int64

IMPACT OF ROAD CONDITION ON ROAD ACCIDENT

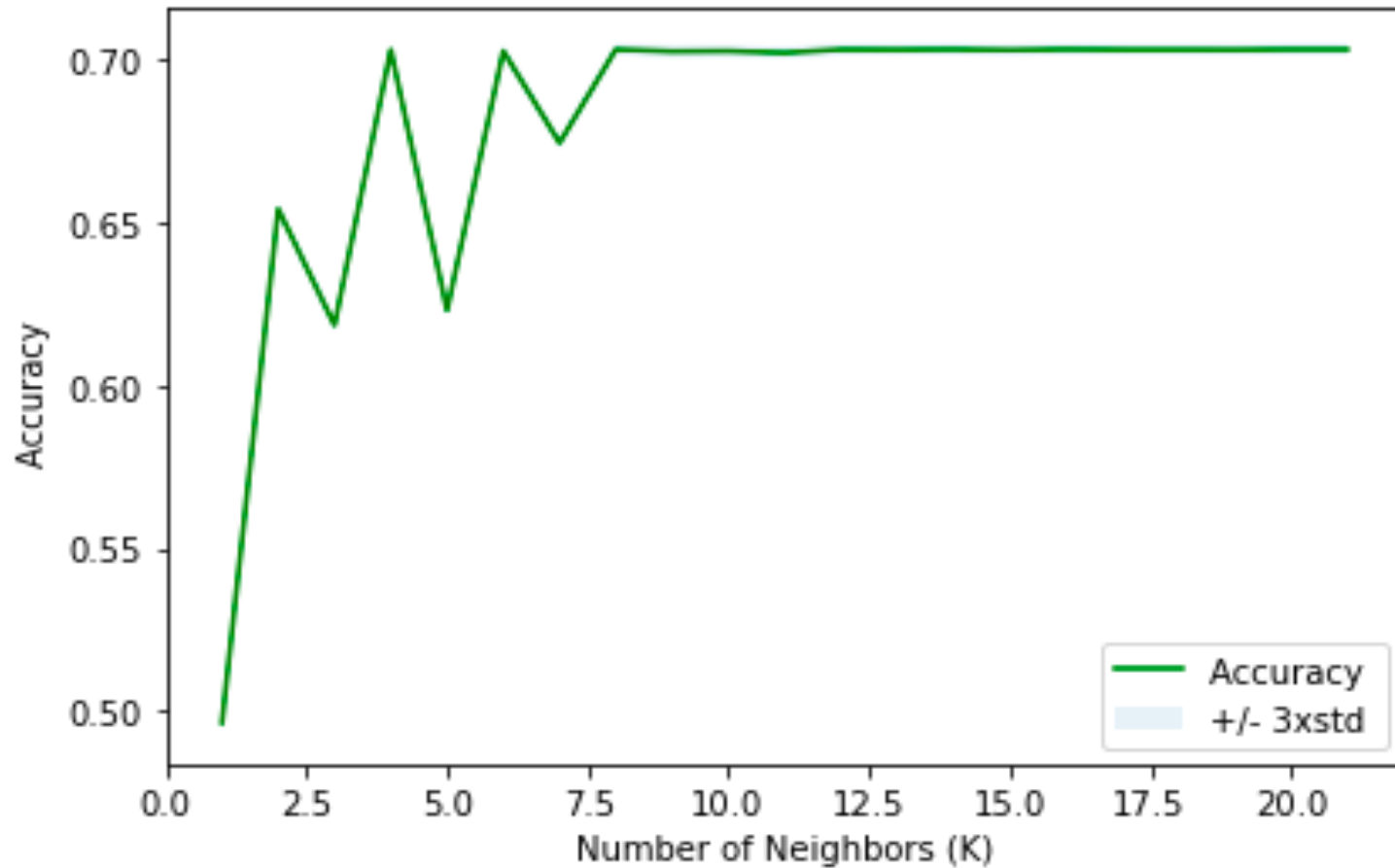
Dry	124510
Wet	47474
Unknown	15078
Ice	1209
Snow/Slush	1004
Other	132
Standing Water	115
Sand/Mud/Dirt	75
Oil	64

Name: ROADCOND, dtype: int64

IMPACT OF LIGHTING CONDITION ON ROAD ACCIDENT

Daylight	116137
Dark - Street Lights On	48507
Unknown	13473
Dusk	5902
Dawn	2502
Dark - No Street Lights	1537
Dark - Street Lights Off	1199
Other	235
Dark - Unknown Lighting	11
Name: LIGHTCOND, dtype: int64	

SEVERITY OF ACCIDENT USING CLASSIFICATION MODEL



KNN:

Accuracy of using K-nearest neighbor in prediction of accident severity is optimum at $K = 14$

RESULTS AND EVALUTION FROM USING CLASSIFICATION MODEL

The table shows the frequency table of the true value of severity and the forecasted severity using different classification algorithms

Algorithm	Jaccard	F1-score	LogLoss
KNN	0.703	0.581	NA
Decision Tree	0.703	0.413	NA
Logistic Regression	0.703	0.581	0.601

FORECAST RESULTS FROM TEST DATA

As seen below, severity of some of the car accident are most likely to cause “Property Damage” frequent in severity code category (1)

	y_test	KNN_yhat	DT_yhat	LR_yhat
1	41083	58392	58402.0	58402.0
2	17319	10	NaN	NaN



CONCLUSION AND FUTURE DIRECTION

- Certain weather and road conditions may impact decision about travelling or on road usage which may result in property damage (1) or injury (2)
- Accuracy of the models has room for improvement
- To include speed value in the future data collection for road accident for more refined data analysis will be a great idea.