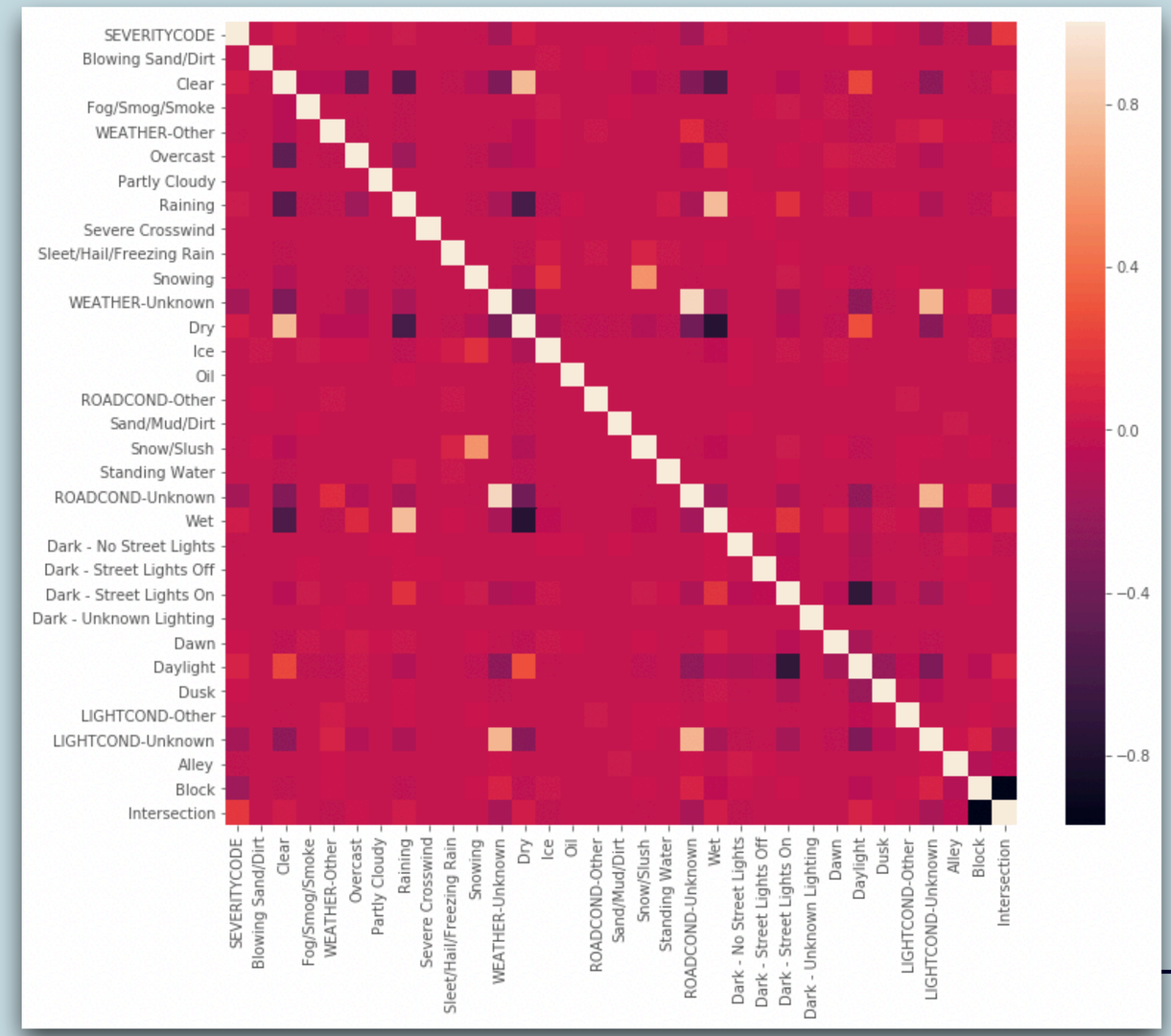
IBM Data Science Capstone Project

Car Accident Severity Risk

CHANTAL BEATRICE BAMIGBELE, PHD

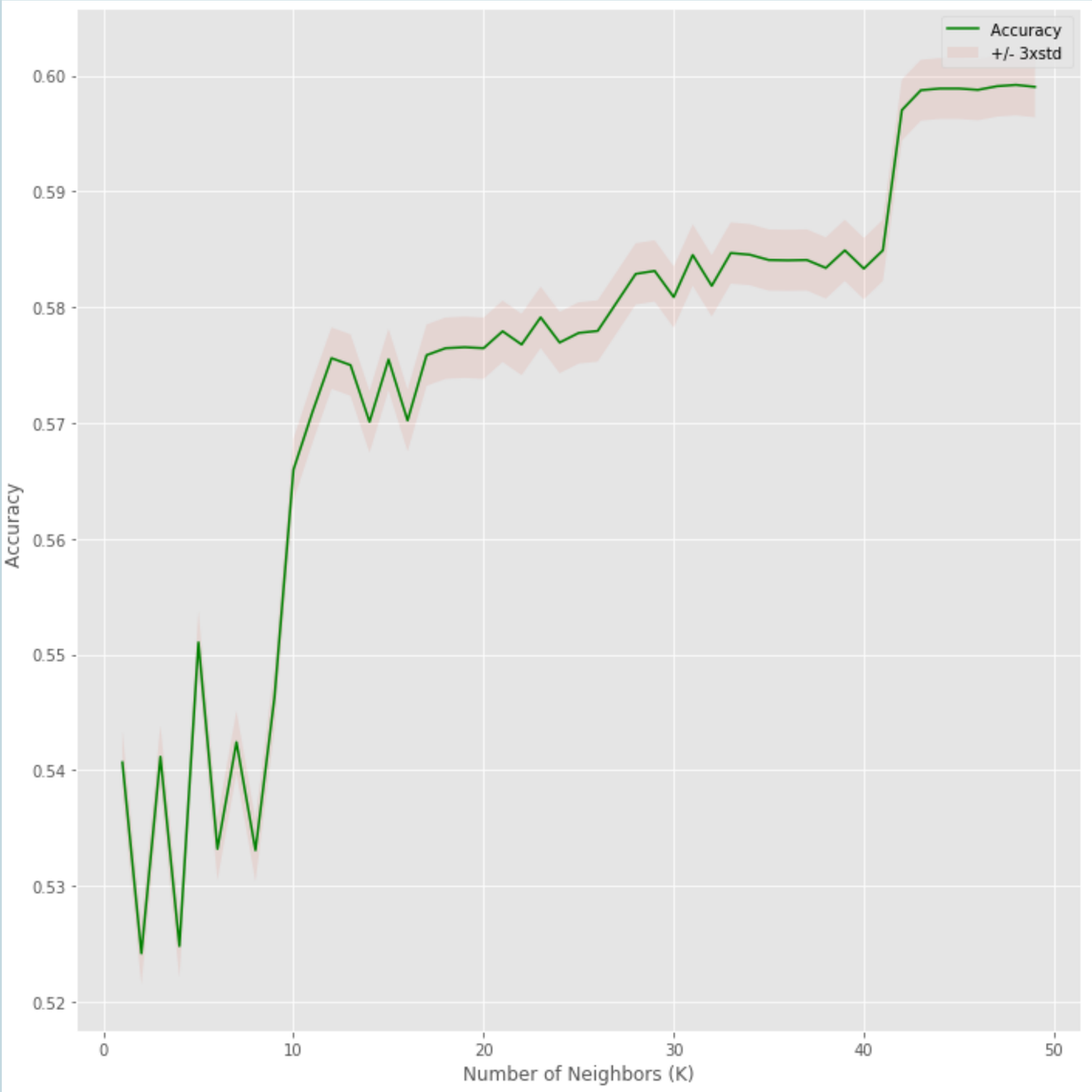
9TH NOV 2020

Data exploration and Visualisation



- Using the Pearson correlation and the p-value the most important variables were narrowed down to the represented one.
- The correlation between the variables and SEVERITYCODE was statistically significant, however there was almost no linear relationship (close to 0) meaning that they can't affect the SEVERITYCODE alone, probably in combination.

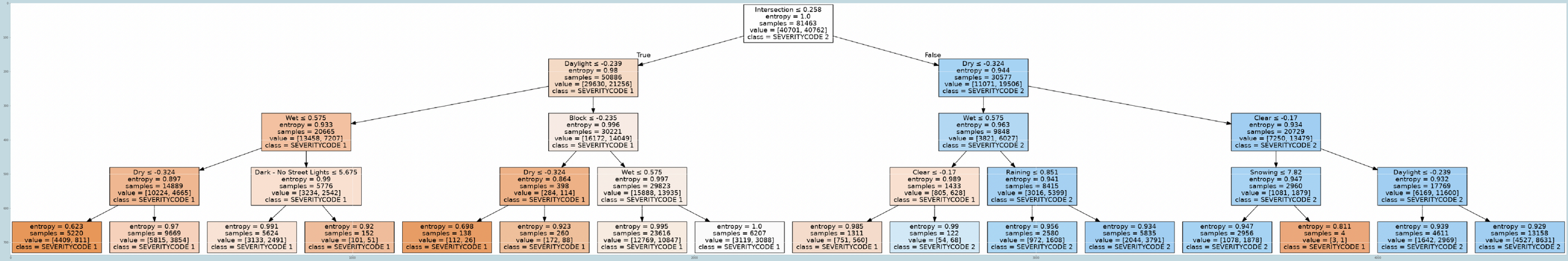
KNN MODEL



	precision	recall	f1-score
1	0.58	0.69	0.63
2	0.62	0.51	0.56

The best accuracy was 0.59 with k=48.

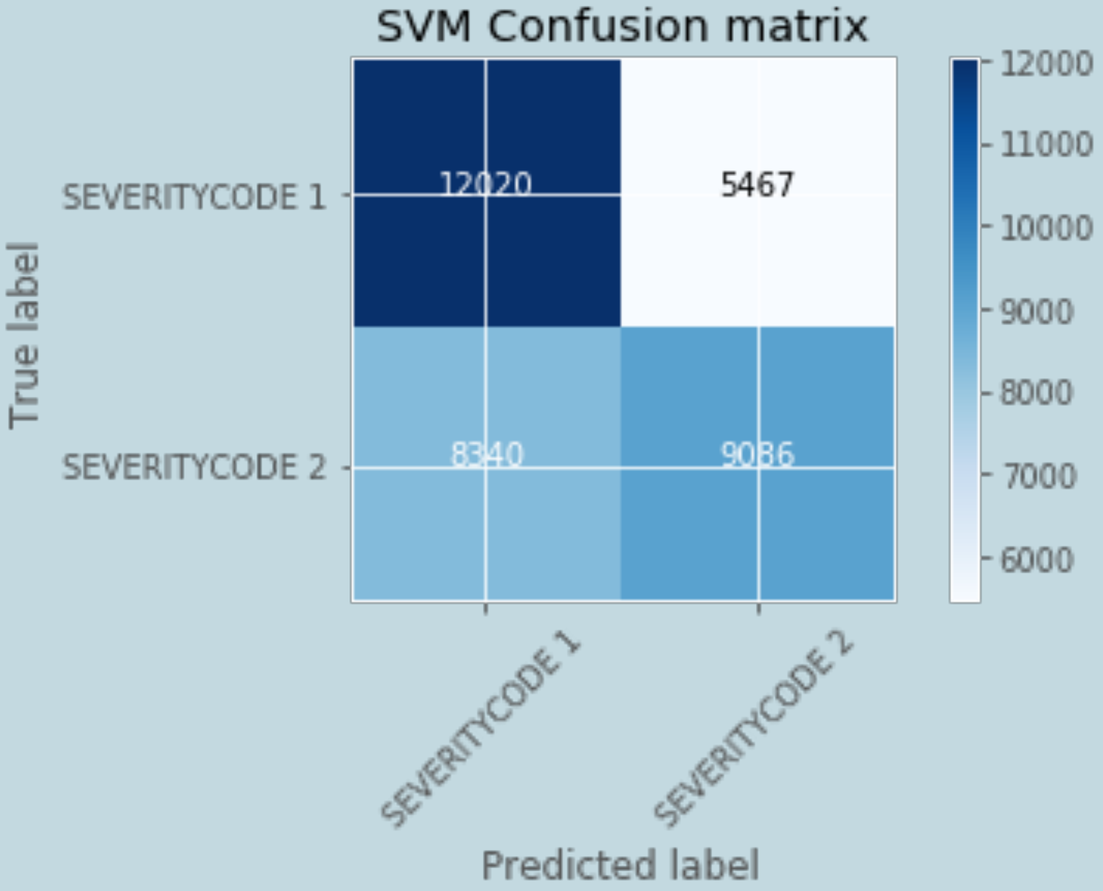
DECISION TREE MODEL



	precision	recall	f1-score
1	0.58	0.69	0.63
2	0.62	0.51	0.56

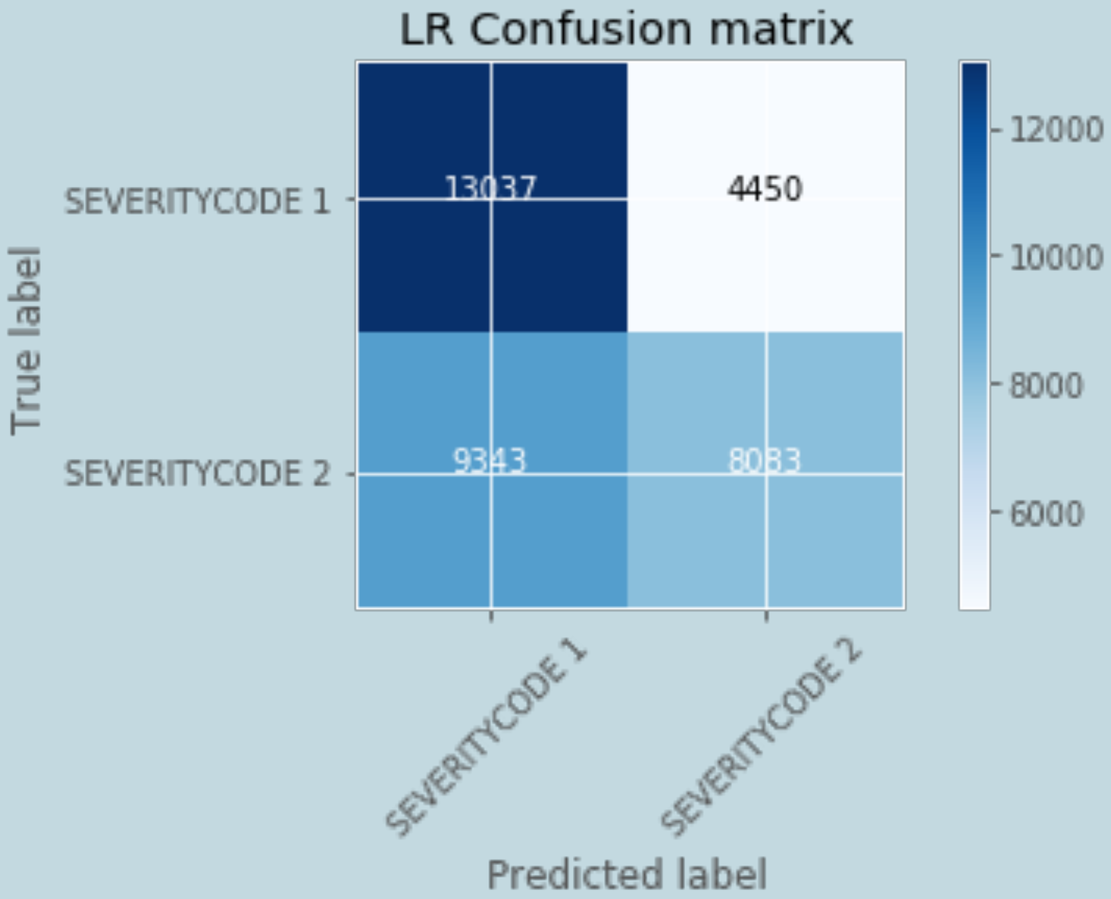
Decision Tree's Accuracy was 0.60.

SUPPORT VECTOR MACHINE



	precision	recall	f1-score
1	0.59	0.69	0.64
2	0.62	0.52	0.57

LOGISTIC REGRESSION



	Precision	recall	f1-score
1	0.58	0.75	0.65
2	0.64	0.46	0.54

SUMMARY

	Algorithm	Jaccard	F1-score	LogLoss
1	KNN	0.599032	0.595649	NA
2	Decistion Tree	0.605562	0.597444	NA
3	SVM	0.604531	0.601777	NA
4	Logistic Regression	0.604932	0.596914	0.654374

- Were able to build models to predict the severity risk of accidents.
- However, models accuracy can still be improved.