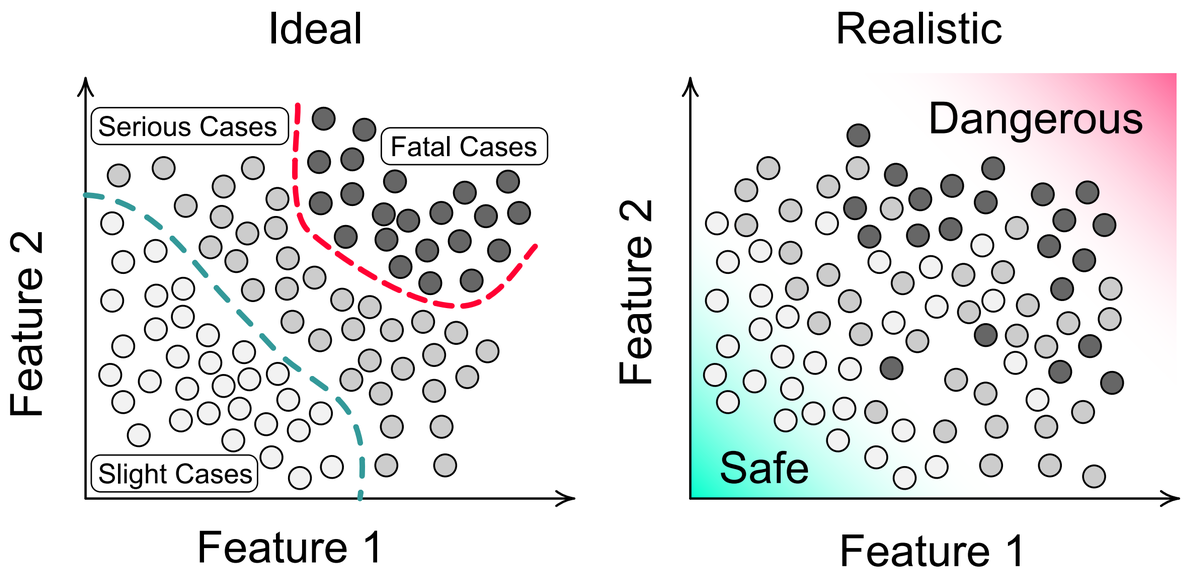
Understanding the Road Accident in UK

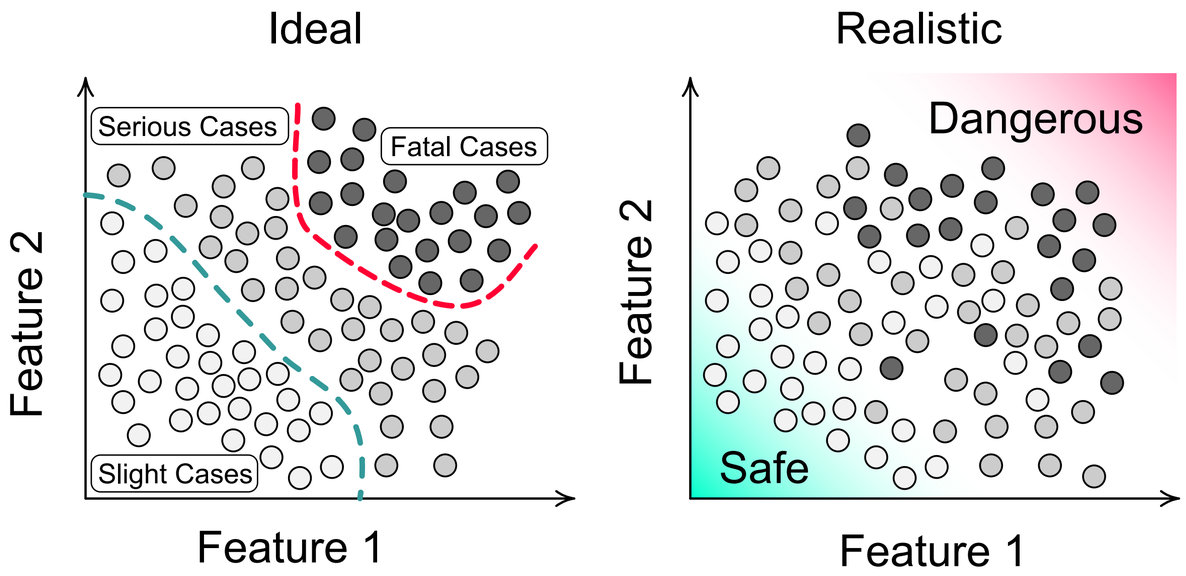
Yushi 2020-10

The Motivation



Road accident might not “just random”

The Motivation

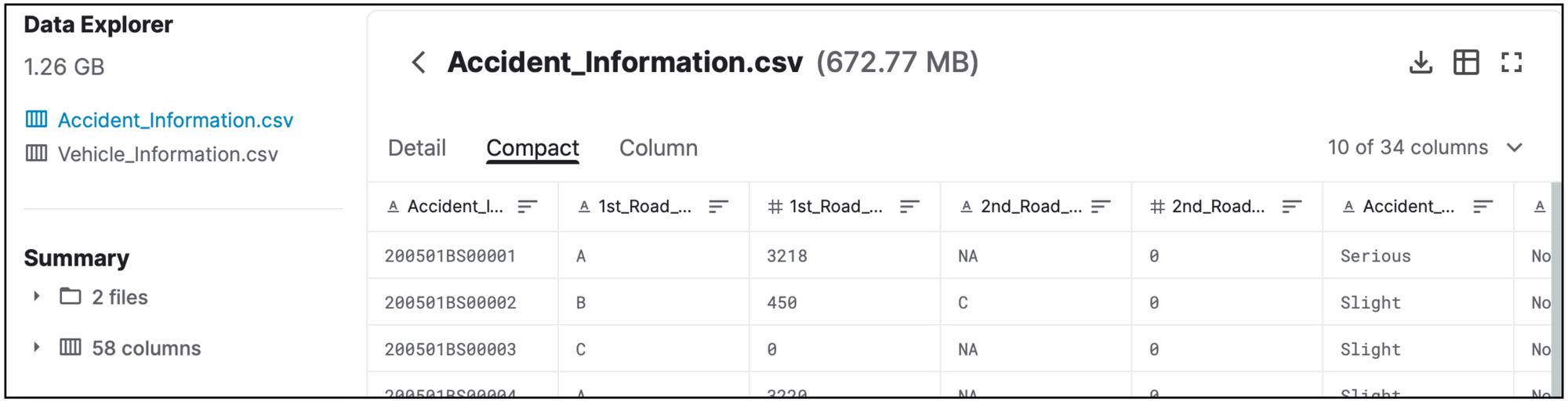


Effective model for accident severity

= Predicting if a travel is safe or not

= save lives

The Dataset



Data Explorer

1.26 GB

Accident_Information.csv

Vehicle_Information.csv

Summary

2 files

58 columns

Accident_Information.csv (672.77 MB)

Detail | Compact | Column

10 of 34 columns

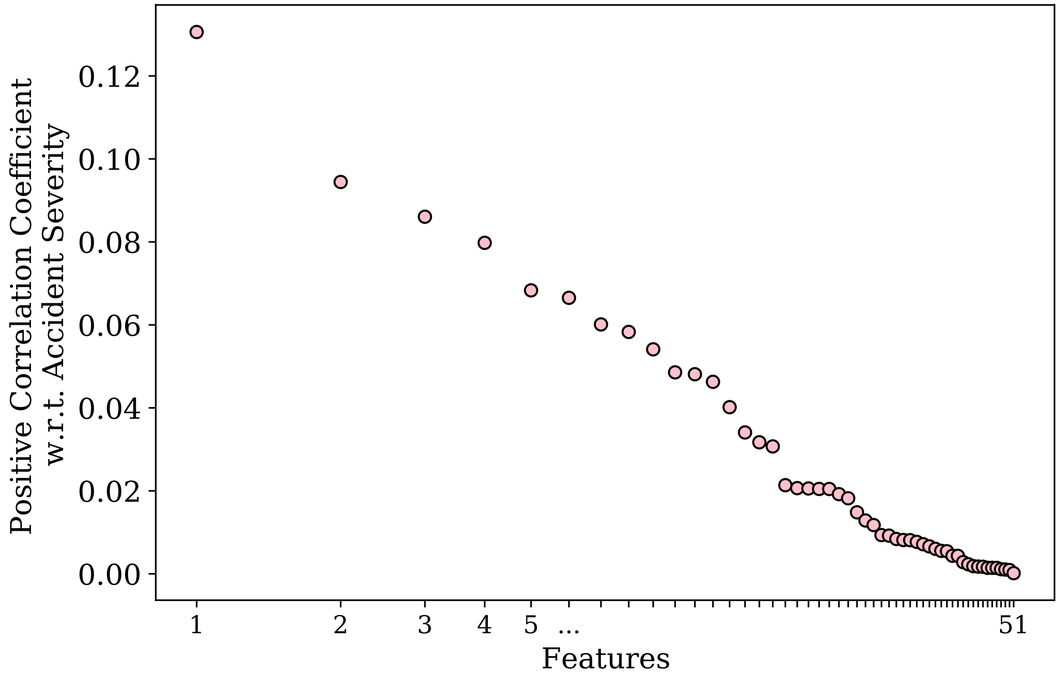
Accident_I...	1st_Road_...	1st_Road_...	2nd_Road_...	2nd_Road...	Accident_...	A
200501BS00001	A	3218	NA	0	Serious	No
200501BS00002	B	450	C	0	Slight	No
200501BS00003	C	0	NA	0	Slight	No
200501BS00004	A	3220	NA	0	Slight	No

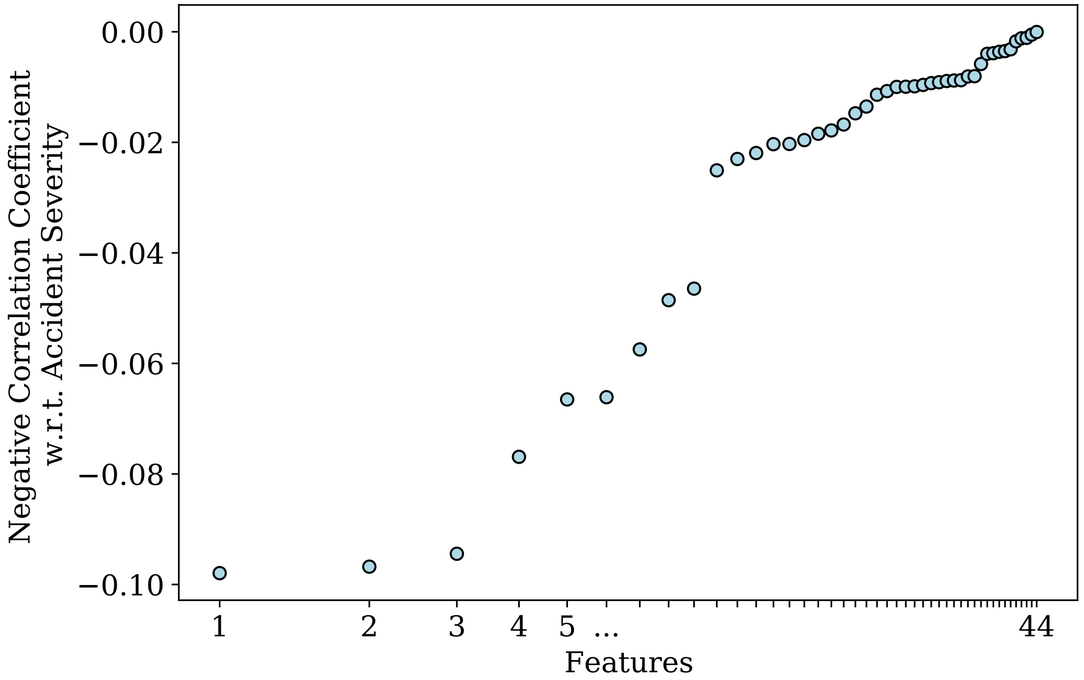
Source: Kaggle

Number of samples: 1,793,224

Number of features: 55

The Correlation Analysis





Non-trivial correlations discovered

17 correlated features were selected for model building

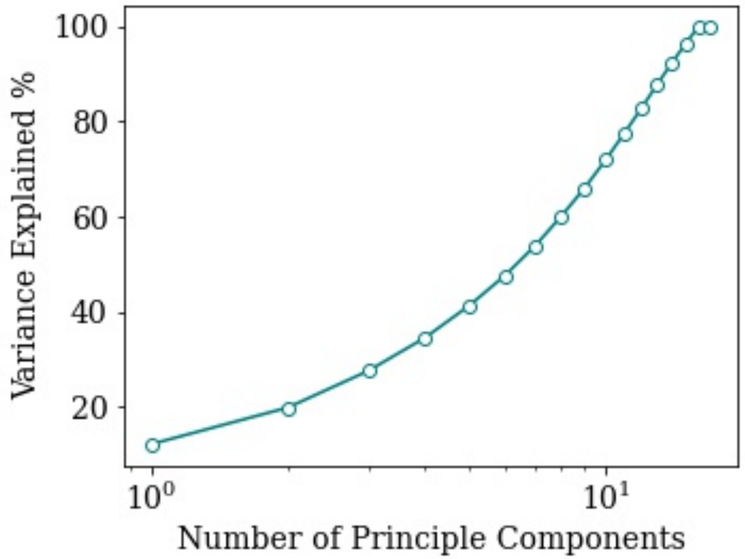
The Correlation Analysis

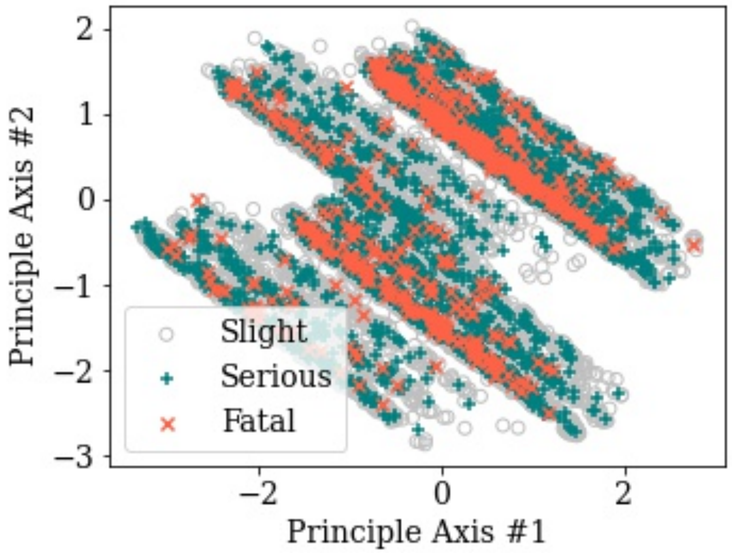
Table 1: Selected correlation values of different features with respect to the accident severity. The positive correlation values indicate the dangerous nature of the corresponding features, and the features with a negative correlation values can be considered as safe-ensuring.

Feature Name	Correlation Value
Vehicle Type: Motorcycle over 500cc	13.06 %
Urban or Rural Area: Rural	9.44 %
Speed limit	8.61 %
Junction Detail: Not at junction or within 20 metres	7.98 %
Sex of Driver: Male	6.65 %
Vehicle Leaving Carriageway: Did not leave carriageway	-9.79 %
Vehicle Type: Car	-9.68 %
Urban or Rural Area: Urban	-9.44 %
Vehicle Manoeuvre: Waiting to go - held up	-7.69 %
Sex of Driver: Female	-6.65 %

Some highly correlated features favouring/rejecting severe accidents

The PCA Analysis

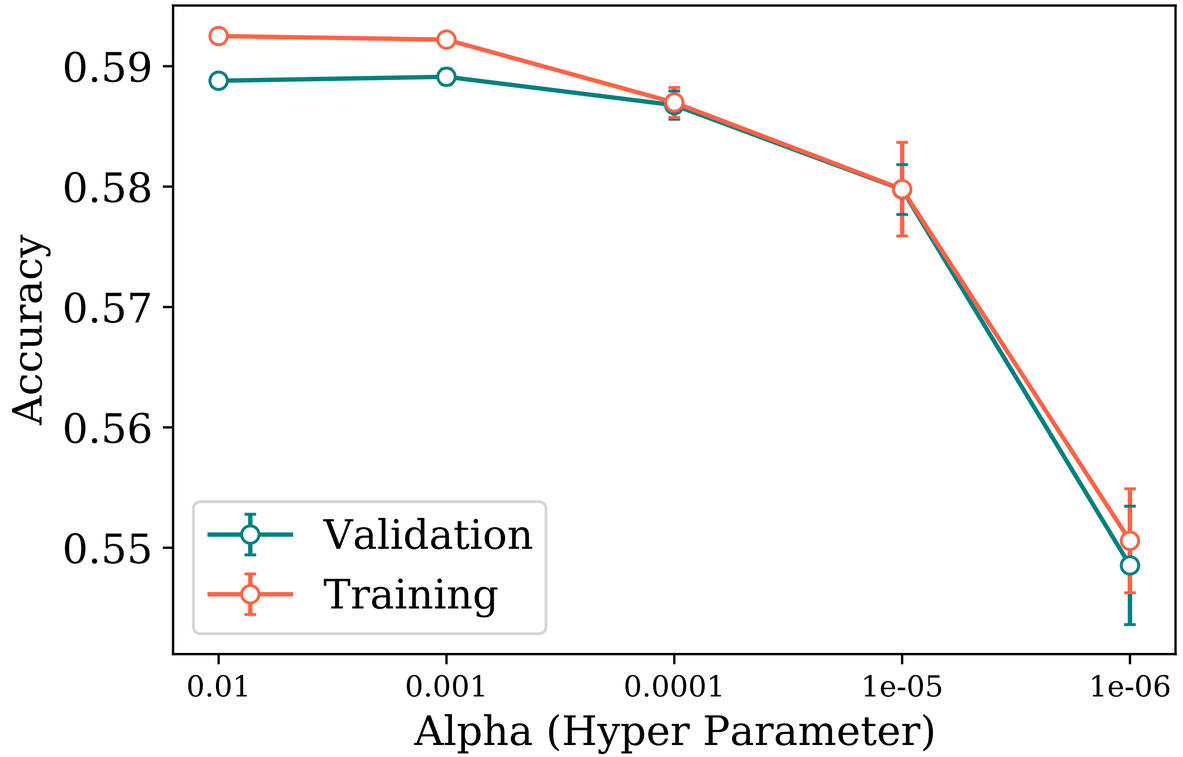




The selected results can not be further compressed

No obvious structure found in the first 2 principle axes

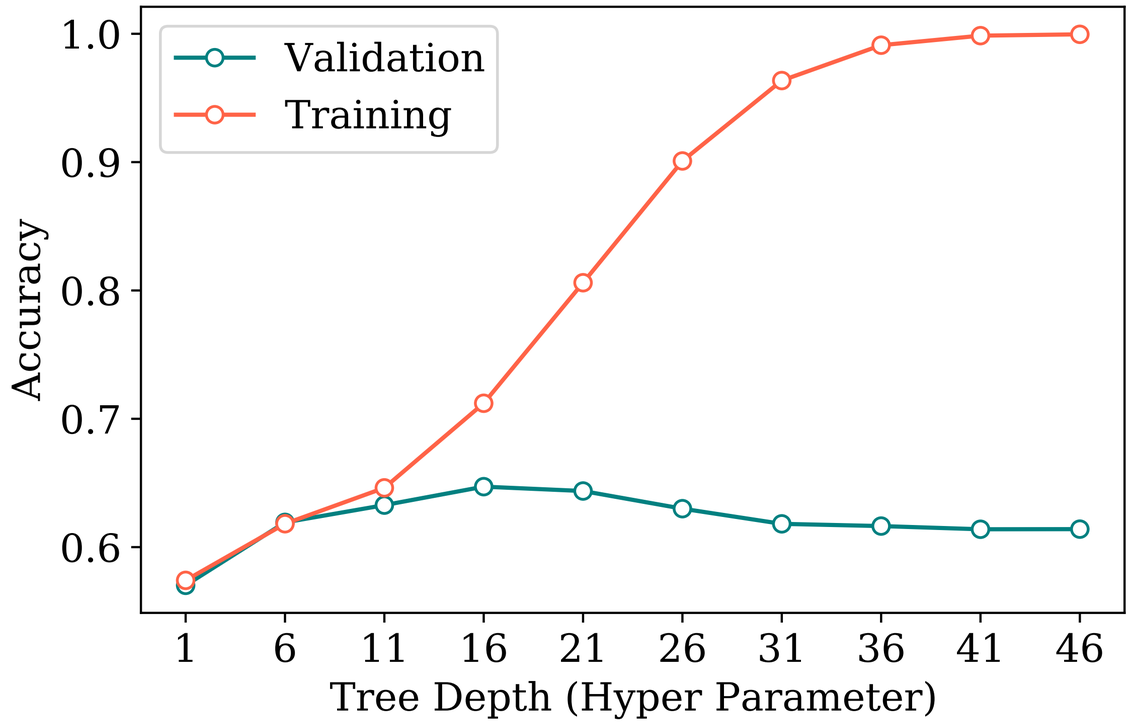
The SVM Modelling



No overfitting

Result accuracy is not great

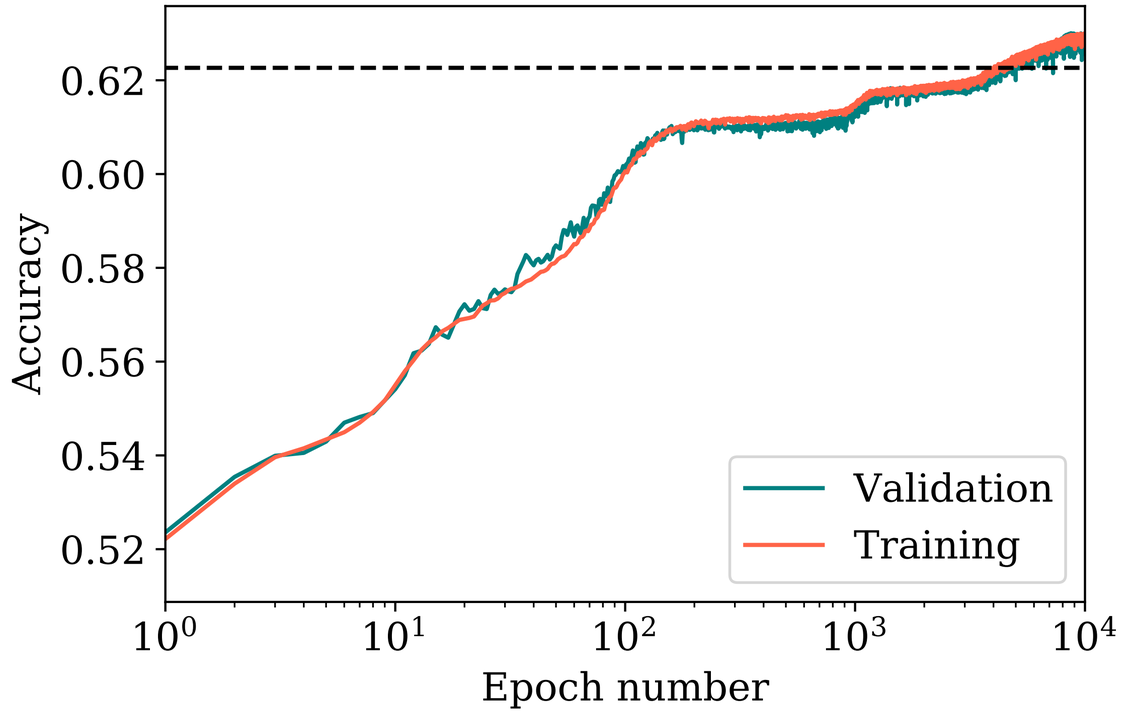
The Decision Tree Modelling



Overfitting at large depth values

Better accuracy comparing with SVM

The Neural Network



No overfitting

Better accuracy expected with longer training

Conclusion

- Non-trivial correlation can be discovered for road accidents
- Initial modelling not great, only achieved 63% accuracy
- The neural network is promising with more training epochs