

Introduction

The increase in the population and the increase in sophistications have led to the increase in the use of private transports. Though the environmental concerns are creating awareness among individuals letting them take a public transport for commute, the car accident cases still prevail and there seems to be no significant declines. The car accidents severity is very crucial as it helps us understand better the damages involved with it.

Data Source

The dataset that is considered for the project is the one provided in the class on the car accident severity. The dataset provides various parameters related to the accident severity. I would like to analyze the accident severity based on weather, road condition, light condition and address type. The various machine learning algorithm will be used and best one will be chosen.

Data Cleaning

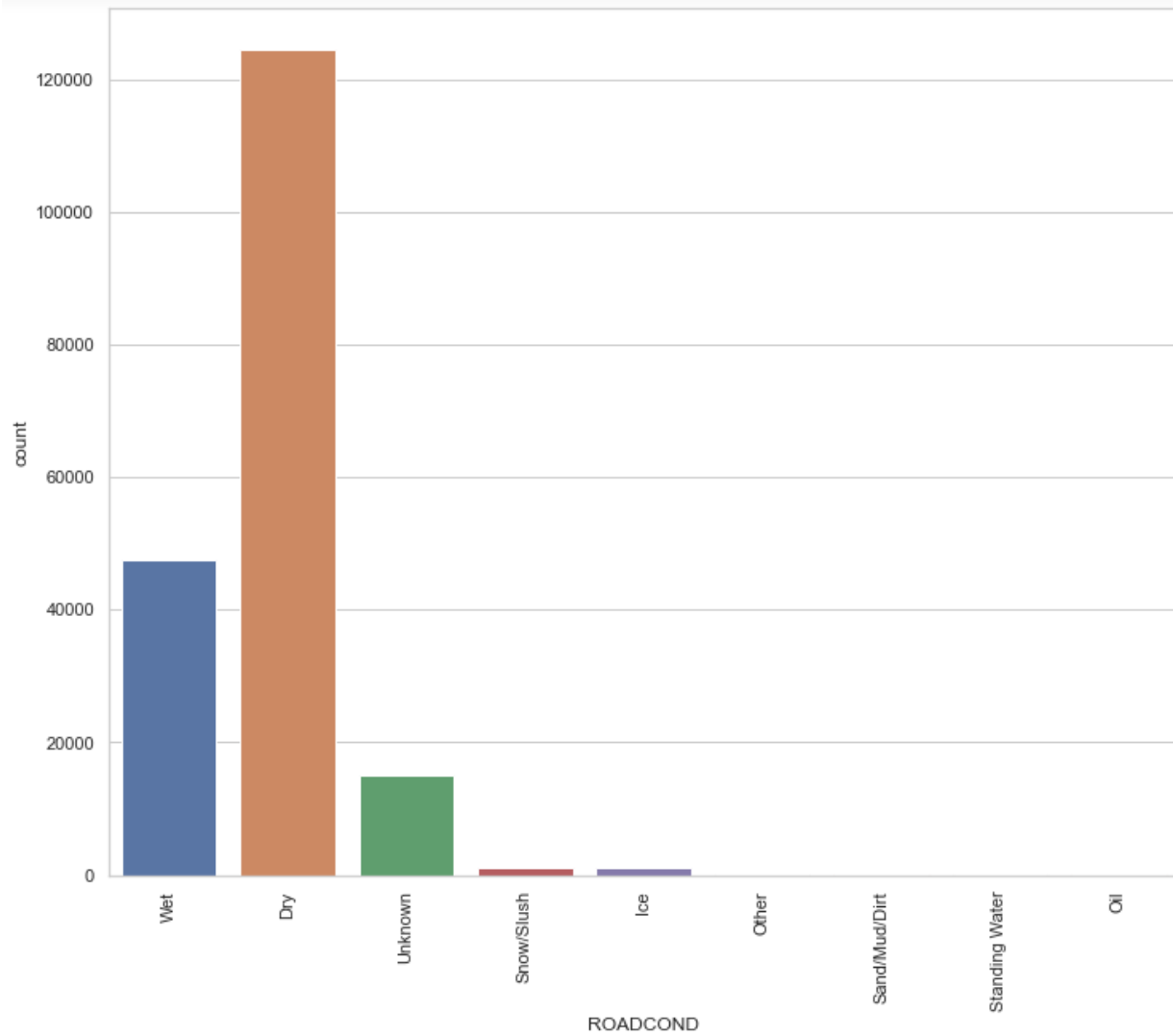
The data was cleaned by replacing all the NaN in the columns with desired values.

Feature Extraction

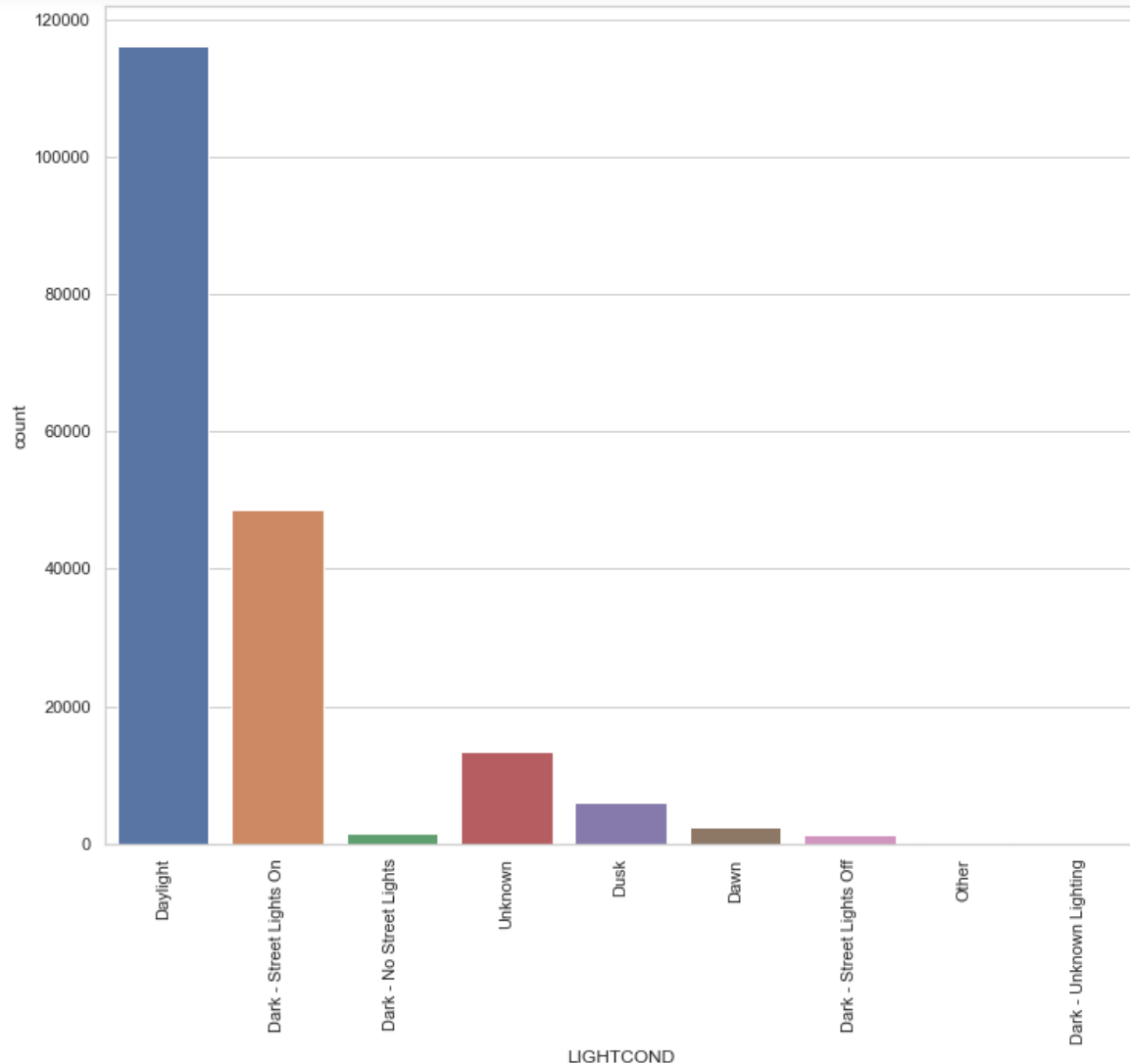
The unwanted rows and columns were removed. The categorical variables were converted into numerical values for further processing.

Data Analysis

The data of light condition and road condition leading to the accidents were analyzed.



The maximum accidents happened in dry road conditions.



The maximum accidents occurred in daylight conditions.

Predictive modelling

The support vector machine, k-nearest neighbor, decision tree and logistic regression were used to analyze the model. Out of these the KNN was best.

KNN Jaccard index: 0.48

KNN F1-score: 0.60

DT Jaccard index: 0.50

DT F1-score: 0.58

SVM Jaccard index: 0.50

SVM F1-score: 0.58

LR Jaccard index: 0.50

LR F1-score: 0.58

The KNN has the highest F-score. However, the range of F-score is not good for the models.

Conclusion

It is considerable to use the KNN for predicting this model. The accidents were high when during daylight and dry road.

Future Direction

Test-train split ratio can be changed and data can be further analyzed.