Car accident severity Predict

1. Introduction

In order to reduce the frequency of car collisions in the community, an optimal model must be developed to predict the severity of the accident under the current weather, road and visibility conditions. When the conditions are bad, the model can warn the driver to remind them to be more careful.

Accidents always happen , but if the main causes of accidents are determined, advance warning or mitigating methods can be performed. For example, certain intersections may be more susceptible to accidents due to heavy usage or the way they are constructed. As a result, better streetlights can be added (only protected left and right turns) or traffic personnel can be used to direct the cars. If it is determined that accidents occur most of a time a driver is speeding, has a high blood alcohol level, or was not paying attention, the data can be used as evidence for enacting harsher laws and regulations. In addition, the data can be advertised to the public to show them the consequences of driving under these conditions. This will hopefully dissuade people in the future. Finally, there are also uncontrollable factors such as weather and road conditions. If certain patterns are discovered to cause many accidents, local government can know when to send alerts to the public to drive more cautiously or even avoid the roads entirely.

The target audience of this analysis is the Seattle government and transportation department. It should identify key causes of accidents and allow them to identify trends for when accidents can be prevented. This will reduce the number of accidents and injuries for the city.

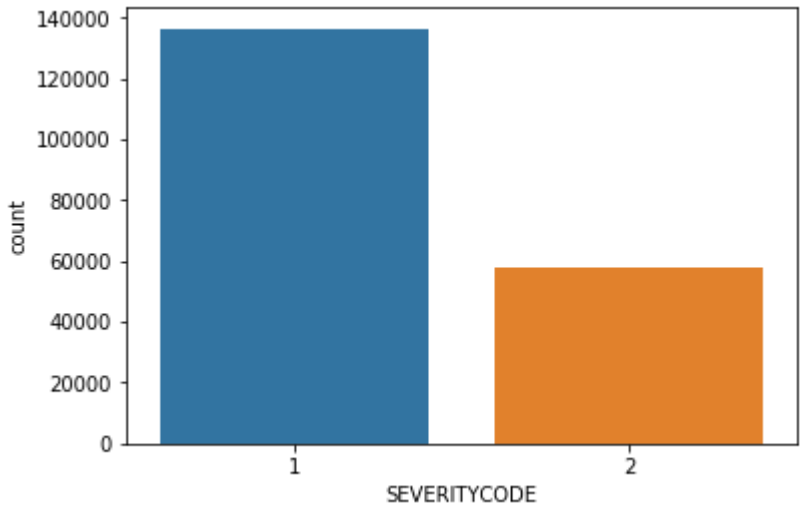
2. Methodology

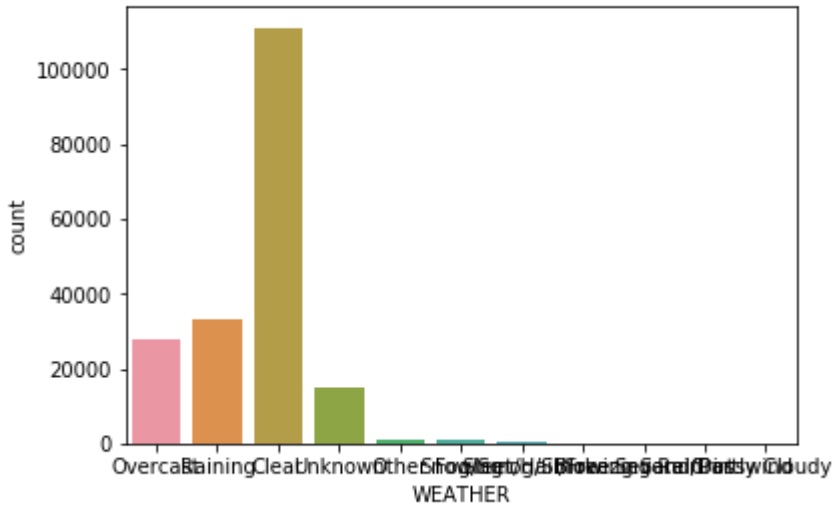
2.1 Data sources

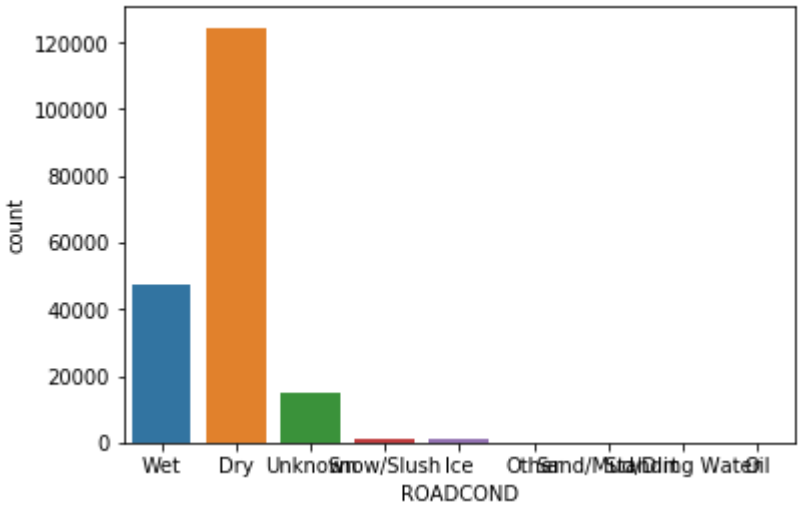
The car accident severity report in Seattle can be found online. There are 194673 cases accidents with 38 kinds of different features.

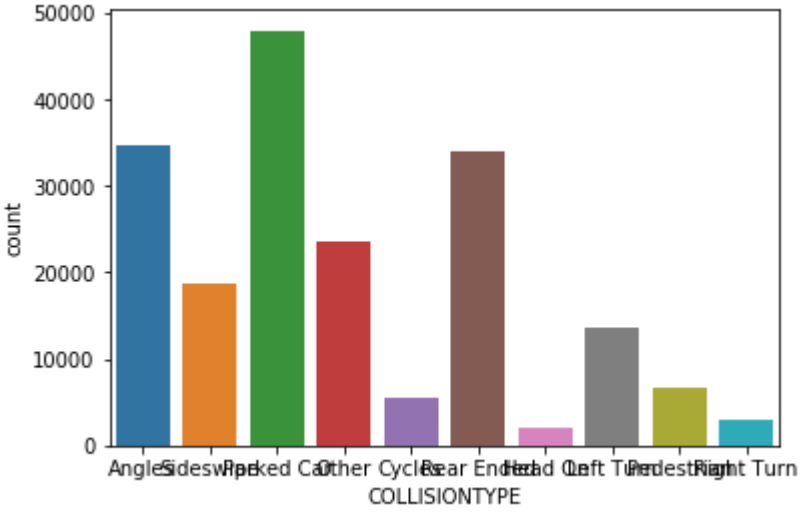
2.2 Feature selection and Data cleaning

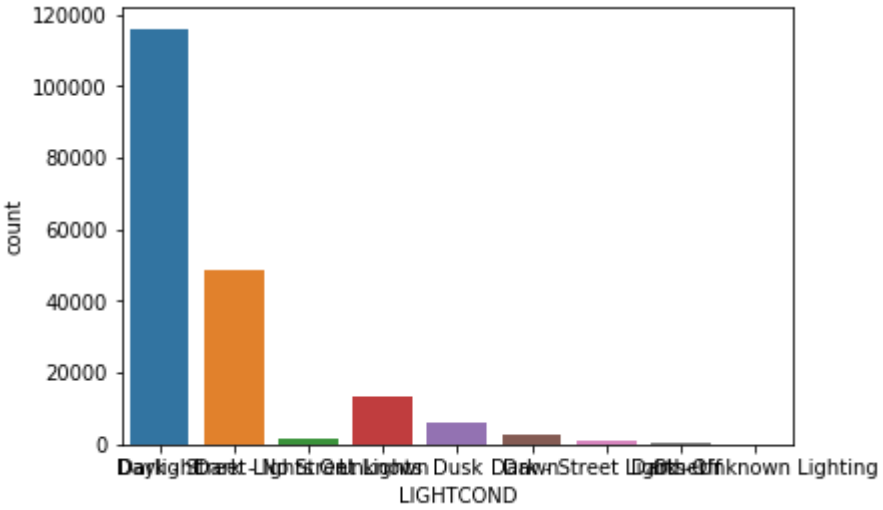
The label we want to predict using this dataset is SEVERITYCODE, 1 means prop damage, 2 means injury. We will use the WEATHER, ROADCOND, COLLISIONTYPE, LIGHTCOND and VEHCOUNT to predict the SEVERITYCODE. We drew value count figures on these features and label.

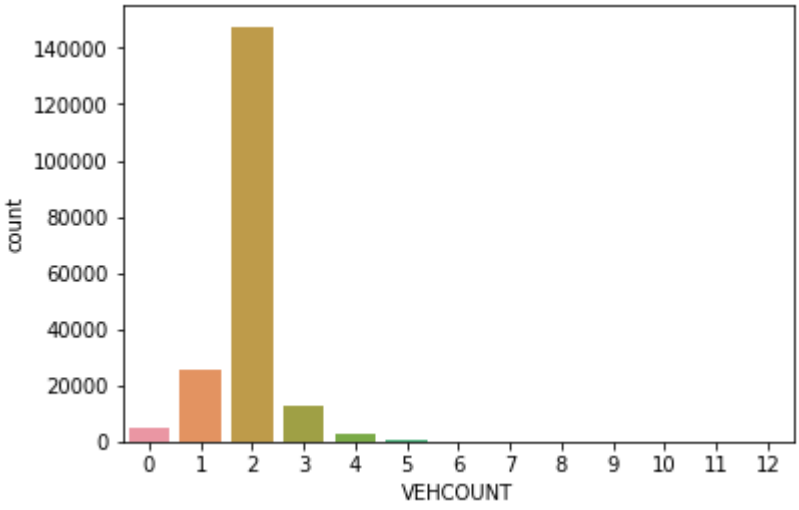

140000
120000
100000
80000
60000
40000
20000
0
count
1
2
SEVERITYCODE


100000
80000
60000
40000
20000
0
count
WEATHER


120000
100000
80000
60000
40000
20000
0
count
Wet
Dry
Unknown
Snow/Slush
Ice
Oil
ROADCOND







Then we convert these category features into numeric features. And drop all the row with NAN values. Here is our data frame after preparation.

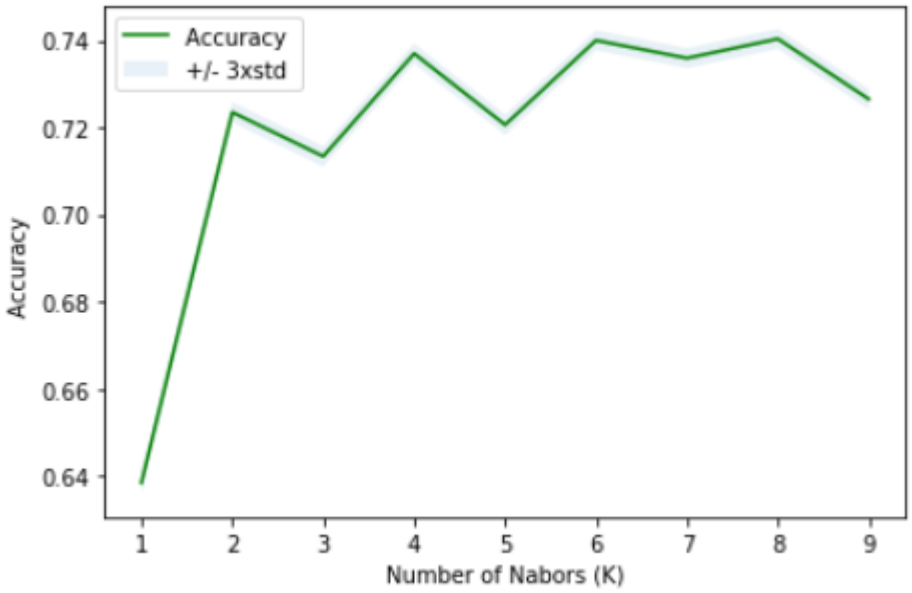
	SEVERITYCODE	WEATHER	ROADCOND	COLLISIONTYPE	LIGHTCOND	VEHCOUNT
0	2	3	2	2	1	2
1	1	2	2	5	2	2
2	1	3	1	1	1	3
3	1	1	1	4	1	3
4	2	2	2	2	1	2

2.3 Model Development

Cause this is a classification case to predict the severity is prop damage or injury. We will use K-neighbors, Decision Tree, Support Vector Machine and Logistic Regression to do the classification.

3. Results

For the KNN classifier, we got the following plot.



Method	Accuracy	Jaccard Score	F1-Score	Log Loss
KNN	0.74	0.74	0.70	\
Decision Tree	0.75	0.75	0.71	\
SVM	0.75	0.75	0.71	\
Logistics Regression	0.72	0.72	0.66	0.58

4. Discussion

From the figure we can see that this model can perform best when the k=8. Also, from the accuracy table, we know that for this classification case, both Decision Tree and Support Vector Machine can perform well and give us an accuracy of

75%.

5. Conclusion

In this study, I analyzed the relationship between car accident features and the severity. I identified weather, road condition , collision type, light condition and vehicle count that affect the accident severity. I built classification models to predict the accident severity. These models can be very useful in helping road management in several ways. For example, drivers should reduce driving when the weather condition is bad, pay more attention to the road with less light, etc.