



PREDICTING SEVERITY OF ACCIDENT

Introduction

- ▶ **Background**

- ▶ Accidents occur due to some environmental factor, road nature, or certain point on road. Also drivers carelessness and relaxation during driving is the key factor for accident. There should be some mechanism by which government alerts drivers in particular weather, on certain point of road to be careful or change route of travelling

- ▶ **Problem**

- ▶ Problem is how to predict chances of accident severity using past data.

- ▶ **Interest**

- ▶ People travelling on daily basis particularly out of the city will take it as critical alert. Secondly helps government officials in taking care of roads.

Data Acquisition and Feature Selection

- ▶ Data Source

- ▶ Data downloaded from an online source.
- ▶ http://data-seattlecitygis.opendata.arcgis.com/datasets/5b5c745e0f1f48e7a53acec63a0022ab_0.csv

- ▶ Feature Selection

- ▶ From a big data source only those features are selected which can easily be categorized and are correlated with forecasted knowledge.
 - ▶ SEVERITYCODE (Target Data)
 - ▶ WEATHER
 - ▶ ROADCOND
 - ▶ LIGHTCOND
 - ▶ JUNCTIONTYPE

Target Data

- ▶ Our target data was initially composed of 5 categories, which were reduced to two after eliminating the values whose frequency is less than 0.5% in overall data. Secondly values that were similar to each other were placed in one category. Value 0 and 3 were insignificant in overall data, while 2b was much similar to 2.
 - ▶ 1. Property Damage Only. Collision
 - ▶ 2. Injury Collision
 - ▶ 2b. Serious Injury Collision
 - ▶ 3. Fatality
 - ▶ 0. Unknown

Predictive Modeling

- ▶ Logistic Regression and Decision tree is used since our target data is based on two values and independent variables are also converted into binary values using one-hot encoding.
- ▶ For accuracy measurement we used following metrics.
 - ▶ Jaccard_similarity
 - ▶ F1_score
 - ▶ Accuracy_score
- ▶ At the end we used confusion metrics to have a clear image of our overall data distribution.

Results

We have below result of our accuracy metrics.

	Jaccard_similarity	f1_score	accuracy_score
Logistic Regression	0.585	0.619	0.585
Decision Tree	0.586	0.597	0.586

Observation

Results of all accuracy measures shows almost same values for both models. This shows that there is almost 60% accuracy in the selected data and the target data. Initially I have chosen only weather, road condition and light as independent variables but I observed that with the addition of fourth variable i.e. Junction type we have got some better results.

Conclusion

With the results and data, we have for analysis, it is observed that accident count due to normal weather and normal road condition with normal light status are more as compared to other reasons. This shows that in these circumstances, drivers may show more carelessness and relaxed while driving and face accident. For other reason behind accidents are darkness and wet road with rainy weather. When we put junction types into analysis, we observed that in all these weathers, accidents occurrences are more at some junctions as compared to others. So, drivers should be more careful while arriving at these points to avoid accidents. Or drivers can change their route, if feasible, to avoid certain junctions.