

# Seattle City Accident Severity Prediction

## Introduction/ Business Understanding

Car accident severity is a big concern across all states in the US, including Washington. According to Washington State Department of Transport (WSDOT), there was a total of 10,315 crashes in Seattle city in 2019. The crashes include 22 fatal car accidents, 190 suspected serious injury collisions, 834 suspected minor injuries, 2612 possible injuries, and 6657 no apparent injuries. This research mainly focuses on Seattle city of Washington for the period 2004-todate. This traffic safety research, which is based on the data obtained from Seattle Department of Transportation Traffic Management Division, aims not only at prevention but also at reducing the severity of car accident. Identifying the more probable attributes that greatly contributes to accident severity was given due emphasis in this research.

## Data Understanding

In this research, the data set used was derived from sample of 194673 collision events (rows) containing 37 attributes (columns) reported by WSDOT. This research aims at some of the key attributes such as weather condition, road condition, light condition, inattentiveness of driver(`INATTENTIONIND`) and driving under influence(`UNDERINFL`) that would enable to predict the severity of car accident in the city. Identifying high areas of accidents and high areas of accident severity can highlight areas of concern.

The research aims at predicting car accident severity in Seattle City considering the response variable, namely severity code – 1 if the accident involves ‘property damage only’ and 2 if the accident involve ‘physical injury’.

## Data Cleaning

The data set from WSDOT is unbalanced and missed certain values. Missing values were encoded as unknown. Some attributes with less relevance and very few values were removed from the data set. In short, dropping nonrelevant attributes, sorting out missing values, handling missing values and balancing the data set were made to accurately predict the severity of car accident while at the same time accurately analyzing accidents can help concerned Seattle City authority to improve the safety of the road and highways and other factors contributing to car accident in the city.

## Exploratory Analysis

After the data cleaning completed, exploratory analysis on existing correlation in the data was made and identification of key attributes that would likely help to accurately predict the car accident severity was also made. In this research, attributes like weather condition, road condition, and light condition were considered for accident severity prediction. The focus was mainly on property damage and physical injury. All irrelevant attributes were dropped, and missing values were retrieved to balance the data set to make it fit into the machine learning algorithm model. The balanced data set was trained on similar instances with respect to severity of car accident. The major factors contributing to car accident include weather condition, light condition and driving under influence. Of all the factors, weather condition comes first in car accident while light condition and driving under influence come second and third respectively in resulting car accident.

## Methodology

The machine learning algorithms used in the research include Decision Tree, K-Nearest Neighbors (KNN), Logistic Regression, Support-Vector Machine (SVM) and Random Forest Classifier.

1. **Logistic Regression**- is used to estimate discrete binary values i.e. 0 and 1 from a set of independent variables. 0 if the incident is property damage and 1 if it is physical injury. LR involves prediction of a dichotomous dependent variables. It helps predict the probability of car accident severity by fitting the data to a logit function.
2. **Decision Tree** – as one of the most popular machine learning algorithms used in data analysis by splitting collision events into two or more similar sets based on the key attributes/ independent variables. It is supervised learning algorithm used for classifying problems.
3. **Support-Vector Machine**- it is a method of classification used in this car accident analysis by plotting the raw data as points in an n-dimensional space, where n is the number of features in the data set. The value of each feature was tied to a particular coordinate, making it is easy to classify the data. The classifiers were used to split the data and plot them on the graph.
4. **KNN**-It is used to solve classification problems in data analysis. It is simple algorithm and stores all available cases and classifies any new cases by taking a majority vote of its k

neighbors. The case is therefore assigned to the class with which it has the most in common. A distance function performs this measurement.

5. **Random Forest Classifier**-This method is used by collectively looking the decision trees. To classify a new object based on its attribute, each tree is classified, and the tree “votes” for that class. The forest chooses the classification having the most votes.

## Results

The data from the cleaned dataset was explored using Decision Tree, K-Nearest Neighbors (KNN), Logistic Regression, Support-Vector Machine (SVM) and Random Forest Classifier to determine reasonable candidate attribute for the model. Th cleaned dataset focused on weather condition, light condition, and road condition to predict accurately the severity of car accident in Seattle City.

The machine learning models used in this research resulted similar outcomes as indicated here below.

### Logistic Regression-

	precision	recall	f1-score	support
1	1.00	1.00	1.00	9847
2	1.00	1.00	1.00	9724
micro avg	1.00	1.00	1.00	19571
macro avg	1.00	1.00	1.00	19571
weighted avg	1.00	1.00	1.00	19571

1.0

### Decision Tree-

	precision	recall	f1-score	support
1	1.00	1.00	1.00	9847
2	1.00	1.00	1.00	9724
micro avg	1.00	1.00	1.00	19571
macro avg	1.00	1.00	1.00	19571

weighted avg	1.00	1.00	1.00	19571
--------------	------	------	------	-------

1.0

### KNN-

	precision	recall	f1-score	support
1	1.00	1.00	1.00	9847
2	1.00	1.00	1.00	9724
micro avg	1.00	1.00	1.00	19571
macro avg	1.00	1.00	1.00	19571
weighted avg	1.00	1.00	1.00	19571

1.0

### Random Forest-

	precision	recall	f1-score	support
1	1.00	1.00	1.00	9847
2	1.00	1.00	1.00	9724
micro avg	1.00	1.00	1.00	19571
macro avg	1.00	1.00	1.00	19571
weighted avg	1.00	1.00	1.00	19571

1.0

### SVM-

	precision	recall	f1-score	support
1	1.00	1.00	1.00	9847
2	1.00	1.00	1.00	9724
micro avg	1.00	1.00	1.00	19571
macro avg	1.00	1.00	1.00	19571
weighted avg	1.00	1.00	1.00	19571

1.0

## **Discussion**

The result shows that the various Machine Learning techniques can significantly vary in terms of performance for the different evaluation metrics. It is also necessary that the more complex the Model, the more prediction accuracy can be achieved. The prediction performance of all models trained with SMOTE was found out to be more realistic than models trained without SMOTE in this study. The study therefore shows potential of machine learning methods for predicting all-cause fatality using collision events dataset. As discussed in the result section of this research, the accuracy is the same for all models used (100 %), The model is overfitted and thus requires to be trained on more data.

## **Conclusion**

Since the response variable has two categories- physical injury and property damage, the machine learning techniques were used to develop the model in this research. The intent was to provide a demonstration of a model that can be used to assess the most important factors contributing to the severity of car accidents in Seattle City. Based on traffic police accident data and techniques used, it was able to the correlation between the key attributes and collision events. The results presented in this research show that the model provided a reasonable statistical fit.

## **Limitations and recommendation**

Other predictors like speeding and inattentive were not included as key attributes to predict car accident. Therefore, future research should consider these attributes for comprehensive assessment of collision events and reduction of fatality resulting from car accident. The WSDOT should introduce an alert system that would notify drivers to drive safely to reduce fatality.

-Ends-