**DS Project on Accidents**

**Introduction –**

As the number of vehicles is increasing the number of accidents occurring has also increased. This Project is to predict the severity of an accident – Low of High. This will be helpful to warn us beforehand about the chances of accident so that it can be avoided.

**Data –**

The data used is Collisions—All Years data and it is provided by by SPD and recorded by Traffic Records.

The collision data set of Seattle City. The data consist of many attributes and observations. The data is labelled.

The data is a collection of records of the accidents that have occurred from 2004 and onwards. This includes all types of collisions. Collisions will display at the intersection or

mid-block of a segment. It is updated weekly.

The data has many Features/attributes 38 in total. But for the study only few are selected for the analysis. These are mostly having information about severity of collision, Type of Address, Road condition, Weather condition, Alertness of driver, Alcohol consumption or not by the driver, and few others.

**Methodology –**

**Data Processing** is needed. Such as

* + Removing missing values – Many rows and many columns have missing values. These values are removed to smooth the process of analysis.

**Exploratory Data analysis**

* + Feature Selection – after studying different attributes of the data few are selected to be used for generation of model to predict the severity of the accident based on selected attributes such as Weather condition, Road condition and type of accidental location.

**Changing categorical values to numerical values** –

Many values are present as categorical values and for generation of model we need these values as numerical values.

**Modelling –**

Decision Tree Classification algorithm is used to predict the severity of the accident based on the address type, weather condition and road type.

**Results –**

With the help of Decision Tree Classification algorithm a model is generated which is able to predict what will be the severity of the accident based on the given conditions. The model has the accuracy of 0.7, which is fairly OK.

**Conclusion –**

The attributes used for the model selection can be modified to increase the accuracy of the model.