

ACCIDENT PREDICTION USING SDOT COLLISION DATASET

INTRODUCTION

Using the dataset provided from SDot Traffic Management Department, we have developed a simple model using KNN and SVM for predicting the type of accident that has taken place under various weather condition, light condition, etc.

Predicting the type of accident that will take place, is a form of knowing which type of accident have higher probability of occurring on a certain day, and help the Traffic management department prepare accordingly beforehand. Data Science and predictive analysis can be utilised efficiently to predict the type of collision that might take place on a certain day, based on the characteristics of the day.

Data

Data Understanding

For the predictive we are going to use the dataset provided to us by the Seattle Traffic Management Department, that is weekly updated database of collisions or traffic accidents that occurred in the city of Seattle.

The data set contains a total of 38 columns detailing everything from exact location of the accident to weather conditions and the type of collision that takes place. Out of all this data we create another data frame that helps us determine the type of collisions that takes place. We choose the attributes, like- the key of incident, address type, severity code, collision type, number of people involved, number of vehicles involved, the type of junction where the accident takes place, speeding information, information of if driver was intoxicated, weather, road condition, light condition, and day of the week to make a prediction.

Data Preparation

Before modelling our data, we have to make the data prepared by cleaning the data, and normalizing it. We begin preparing our data by first removing nan values and changing the types of various attributes. Also attributes like collision type with string values have to label encoded into float form.