

Data Science Capstone Project Report

Introduction and Business Problem:-

In an effort to reduce the frequency of car collisions in a community, an algorithm must be developed to predict the severity of an accident given the current weather, road and visibility conditions. When conditions are bad, so for this purpose this model will alert drivers to remind them to be more careful on these type of roads.

Data Used In the Capstone:-

The data is from Seattle city road accidents data. Our predictor or target variable will be 'SEVERITYCODE' because it is used measure the severity of an accident from 0 to 5 within the dataset. Attributes used to weigh the severity of an accident are 'WEATHER', 'ROADCOND' and 'LIGHTCOND'.

Methodology Followed:-

For implementing the solution, I have used Jupyter Notebook to preprocess data and build Machine Learning models. Regarding coding, I have used Python and its popular packages such as Pandas, NumPy and Sklearn. Also, I have used matplotlib and seaborn libraries to plot different visualizations for better understanding of the features. I have done preprocessing of data than I have used different machine learning models to predict the value of severity using the values of other data. I have used the below mentioned models to predict the output.

- Decision Tree Classifier
- Random Forest Classifier
- Naïve Baye's Classifier
- Logistic Regression

Results:-

By using the three machine learning model I have predicted the result and from that I found that the KNN is best model for predicting these data.

Conclusion:-

From the dataset provided in this capstone from weather, road, and light conditions of Seattle city. I can conclude that particular conditions of weather and road, we can predict the severity of accident very easily.