Approach

1. First of all, I have converted all the severity data into a grading from 1 to 4,this converted the text data into numerical. For this, I have created a function and with the help of loop the data is converted.
2. Then, I’ve added the severity grade data into the Data frame of training data.
3. By analyzing the data I’ve found that some columns like, ‘Safety Score’.
4. Removing all the unnecessary data columns and do the required data pre-processing. Same done for the test dataset also.
5. Then, I split the data into dependent and independent variables.
6. Using ‘Decision tree classifier’ and ADA Boost Classifier, result is predicted on the test dataset.
7. Then the predicted values are converted into the text values of severity.
8. After adding the ‘Accident ID’ columns, the data is converted into the ‘csv’ file.

Tools Used

1. Numpy.
2. Dataframes and other tools from Pandas.
3. Pyplot from Matplotlib.
4. Random Forest Classifier from Scikit learn.
5. Loops and If-else Statements from python.
6. Ms-Excel for final editing of Submission File.