Data Intensive Computing

CSE 4/587

Project Title: “Predicting Severity of Car Accidents”

Problem Statement:

“We will analyze ‘US Accidents (2016-2020) Dataset’ to find out ‘accident severity’, and address ‘unsafe road accidents from occurring’ with a data-driven solution”.

Data Sources:

Our data source is from Kaggle; data set name “US Accidents (2016-2021)” which has the records of countrywide traffic accidents from 2016 to 2021.
There is a total of ‘2845342’ rows and ‘47’ columns in the raw dataset.
Link: https://www.kaggle.com/datasets/sobhanmoosavi/us-accidents

Phase – 1:

In this phase we have performed data cleaning and pre-processing. After this we performed Exploratory Data Analysis (EDA), this helped us understood out of 47 features what are features that are majorly affecting the severity of an accident that has occurred.

Phase – 2:

In this phase we applied 5 different algorithms to predict the accident severity and also used visualizations for the result. We also used cross validation on the five model and performed hyperparameters tuning to find the best possible model for predicting the severity.

Phase – 3:

This is the final phase in which we built a web application that helps a user to upload a dataset and understand how severe the accident can be for the given data. The future scope of this project can be an alert app that helps a user to get an alert at an accident-prone area and helps the user to avoid any kind of accident.