

CIS 545 Project Proposal

Dinesh Jagai

1. Data source

The data is from Kaggle (<https://www.kaggle.com/sobhanmoosavi/us-accidents>) and represents countrywide traffic accidents in the US from (2016 -2019).

2. Project Plan

Investigate how the different features of the data affect the severity of an accident taking place – for example, I shall be investigating the impact of precipitation or other environmental stimuli on accident occurrence and severity. The overall objective is to make a model that predicts the occurrence and severity of an accident given certain environmental features. The baseline models will be trained using ordinary least squares, (OLS), Ridge Regression, and Lasso Regression. Furthermore, other, more advanced, models will look at using elastic net regression, kernel regression, random forest regression, gradient boosting, and bagging trees regression. Lastly, models using deep learning with neural networks, long short-term memory (LSTM) will be used.

In terms of evaluation, I shall be using cross-validation on the training set and partitioning the dataset into training and testing sections. Furthermore, I shall be using Mean Absolute Error (MAE) as my main error metric, as it is commonly used in regression-based models.

3. Why is this project interesting?

Predicting the occurrence of an accident based on the given environmental features is extremely important as it can be used to inform the respective authorities on policies that can be implemented to reduce them from occurring. Furthermore, I believe this project contains most, if not all, of the concepts that we've learned throughout this semester. There is also room to learn new concepts such as mean/regression-based imputations to fill in the missing values for the different features.

4. What challenges and obstacles might you anticipate with this project?

Missing data could potentially lead to erroneous values. Some accidents are just random and as such our models could be predicting erroneous values. Overfitting of the model