AllState Insurance Claims Severity – Kaggle Project

For this project, as instructed by the professor, I downloaded and created a subset of the train.csv and test.csv data from the Kaggle project page. I ran my learning algorithm on this subset.

Algorithm Used

Since this was a regression problem, not a classification problem (where we can easily use decision trees), I used **Linear Regression –** which the simple most Machine Learning algorithm when it comes to regression.

Library Used

I used Scikit Learn (SK Learn) library to train and test on the given data. I also utilized some features of the NumPy library for data manipulation.

Steps for the project:

I went through 5 steps for project:

1. Reducing training and testing data to subsets.
2. Cleaning or making the data fit for applying algorithm.
3. Applying learning algorithm.
4. Calculating MAE (Mean Absolute Error)
5. Storing the results

Cleaning the Data:

Since the data had strings and linear regression can only be applied to integer values, I changed the string values into distinct integers.

Mean Absolute Error:

The MAE for the algorithm was: **1232.78**