**Machine Learning**

**Assignment # 2**

In this Assignment you will use the dataset of the 1st assignment, which is now ready to use.

**Required**

1. Summary of training at least three linear regression models or any three models of your choice, all use the same training and test splits, or the same cross-validation method.

2. A paragraph explaining which of your model you recommend as a final model that best fits your needs in terms of accuracy and explainability.

We applied three regression model on our previously prepared data. We tried three regression algorithms which are linear regression, knn regression and SVM regression. The dataset is splitted into 70:30 ratio. All the models go through 5-fold data validation. The end results revealed the SVM regressor as the best model in terms of training score, testing score, RMSE, MAE, R2 score, and cross validation score. Our model crosses the 0.95 score in testing, R2 score and cross validation scores.

3. Summary Key Findings and Insights, which walks your reader through the main drivers of your model and insights from your data derived from your model.

Below are the key metrics of the best model on given data.

Root mean squared error (RMSE): 0.1791641564793874

--------------------------------------------------

Mean absolute error (MAE): 0.13384616887495007

--------------------------------------------------

R2 score: 0.9877361259363776

--------------------------------------------------

Avg Cross Validation Score: 0.9789654662078542

--------------------------------------------------

-----------------------------------------------------------

Time for detection (SVMRegressor): 0.49 seconds...

-----------------------------------------------------------

Training Score Testing Score MAE RMSE R2 Score

0.979594 0.987736 0.133846 0.179164 0.987736

​

4. Suggestions for next steps in analyzing this data, which may include suggesting revisiting this model adding specific data features to achieve a better explanation or a better prediction.

In given scenario, the dataset has 1.8k records. Someone can acquire more data and feed into the model so that model can be trained and tested on more data.