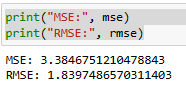
**Project Report-2**

Modelling Procedure for Goal-1:-

* The goal 1 of the project would focus on predicting the negative emotion words from the selected features of Linguistic, psychological, grammatical category of words. The prediction will give us how well does any type of word when tallied with emotional words shows the emotion quotient of every group.
* The second sub analysis of the goal is taking selected features such as anger, anxiety, sadness tallying them with the negative emotions, how highly they and closely related they are under negative emotion words.
* **The Approach :-**

1. Main analysis – It is to check how overall features react with negemo words, so initially, date comments section are dropped and divide the target feature into X and y feature set; y being the target feature i.e. negemo, that will help predict in the first sub analysis that how many other psychological related words are closely related to negative emotion words and I would assume rest to be positive emotion words.
2. Normalization of the features are done using standard scalar methods. Scaling the values helps the values to perform better on the ML algorithms and improve interpretability of data. It is done on the independent feature columns.
3. For prediction of the whole independent and target variable – Decision tree classifier is used, by splitting the X and y into train test split.
4. While conducting the decision tree classification the scaled value are taken into consideration, the training data of both X and y, are fit and the predictions are taken into considerations. The accuracy score is predicted.
5. Calculation of accuracy score defines the performance of all the tallied features with the target feature column. Hence after predictions I got an accuracy score of 0.7726.
6. Doing the Linear Regression :- As to compare using the RMSE and MSE values and how efficient the model was predicted, The values obtained are shown below



1. For the analysis of the second sub-feature where all the feature columns are taken irrespective of psychological emotion words or not and is compared with negative emotion words.
2. Linear Regression is performed getting an accuracy score of 0.67143 and the model behaves well hence proving that 67% of the data are close to negative emotional words.