**Engagement Score Prediction**

1. A brief on the approach used to solve the problem:

For solving the given problem statement, I followed these steps:

1. Importing the relevant libraries
2. Data Inspection
3. Data cleaning
4. Exploratory Data Analysis(EDA)
5. Building Model

After Importing the required libraries, I read the given dataset to inspect how many rows and columns are present in the train data and test data. And also verified whether any null values are present. Later divided the categorical and numerical features in the data.

As part of third step, I checked for missing values to impute them with mean. But there are no missing values in the dataset and I followed the next step.

In fourth step, I analyzed the data with categorical features and done the data visualization with different plots and graphs. I come to know some variables are most important to generate the engagement score.

Later in Fifth step, Started building model, But as per the dataset and given problem statement, I come to know that it is regression problem and thought of using Linear Regression model to predict the engagement scores. Finally, after doing the Labelencoding, I have taken the 40% of validation data and used the Linear regression model to predict the engagement scores. I have used RSME as accuracy metric to check the model performance. With 0.75 of RMSE , I predicted the engagement scores with row\_id and written the same into submission.csv.

1. Which Data-preprocessing / Feature Engineering ideas really worked? How did you discover them?

Techniques of data preprocessing :

**Vectorization:** Dataset which contains different categories of classes then we have to do one-hot encoding to convert that categorical values into integer representations, which can be handled by our model. Used **one-hot encoding in this project** to convert categorical values to integer values.

**Handling Missing Values**

1. What does your final model look like? How did you reach it?

My Final model will have the predicted values of Engagement scores with row id. Also Calculated the accuracy of the model using Root Mean Square Error (RMSE) with 0.752 score. RMSE is a standard way to measure the error of a model in predicting quantitative data.

Analyticsvidhya is a very good platform to learn many Models/techniques in Machine Learning. And this platform helped me alot to reach out this model with previous conducted hackathons as well as trained us with good exercise before participating in JOB-A-THON.

I sincerely thank Analyticsvidhya.com for helping me in all the way to reach successful participation in JOB-A-THON.

Thanking You,

D. Ramya Krishna.