SOLUTION APPROACH FOR ENGAGEMENT SCORE PREDICTION

Problem statement: Engagement score prediction

Solution: XGBoost model following by the Feature Engineering, Feature Selection , Model Building

Steps followed to achieve the solution of predicting engagement score for video are as follows

1. **Data analysis phase**
    - Finding out the number of numerical and categorical features
    - Finding out the number of descrete and continuous features from numerical features
    - Finding out the relation between each of independent features with dependent features(engagement_score) by drawing the heatmap countplot bar plot etc...

2. **Feature engineering**
    - Encoding the categorical features
    - Standardizing the numerical features using log normal distribution

3. **Feature scaling**
    - Scaling down all necessary independent features to same unit using minmaxScaler

4. **Feature selection**
    - feature selection and checking whether selected features are relevant using ‘lasso regression’

5. **Build XGBoost model on train dataset**
    - Defining the model
    - Training the model by setting the parameters
    - Quick sanitize on model performance analysis
    - Repeating the above 3 steps till the high accuracy and less possible error

6. **Model predictions on test dataset**
    - Scaling down the test data by using scaler instance which is fitted on trained data
    - Predicting the engagement score for unseen/un-trained test data set

7. **Storing the predicted results to csv file**