

Approach

I have read the data from given files and performed EDA by using different methods.

There is no Null values and no outliers.

Based on the EDA created new fields.

1. Video_eng – average engagement score for each video
2. cu_eng - average engagement score for each video for each user & category
3. u_eng - average engagement score for each user
4. C_eng - average engagement score for each category
5. Age Groups – Binning of age field

After that I have removed all the unnecessary fields. I have also tried with and without follower and view field as per heatmap readings.

I have used one hot encoding for the field gender and profession.

I have tried with following 3 approach.

1. Random forest
2. XGBoost
3. Deep neural network

My final submission is result of deep neural network with 2 hidden layers with nodes 32.