



Reddit Posts

How can we predict post engagement?

By Mike Verrochi



Analysis of Reddit Posts

We look to see how certain features affect engagement of Reddit posts.

How to create most beneficial title, and what to avoid in title

Data Used

We collected about 12,500 posts from Reddit.

Most recent hot, top, and rising posts.

Outliers were not removed, as we want to look at what makes a post receive few comments as well as many.

Surprisingly, no duplicates.

Posts from 08/2012 - 09/2022

Random Forest Classifier

When taking the only the post title into consideration, we have an accuracy of about 55% on predicting whether or not a post will have above the median amount of comments.

Most Common words to appear in titles:

just, like, new, time

4 out of the top 10 words to appear have to do with the concept of time

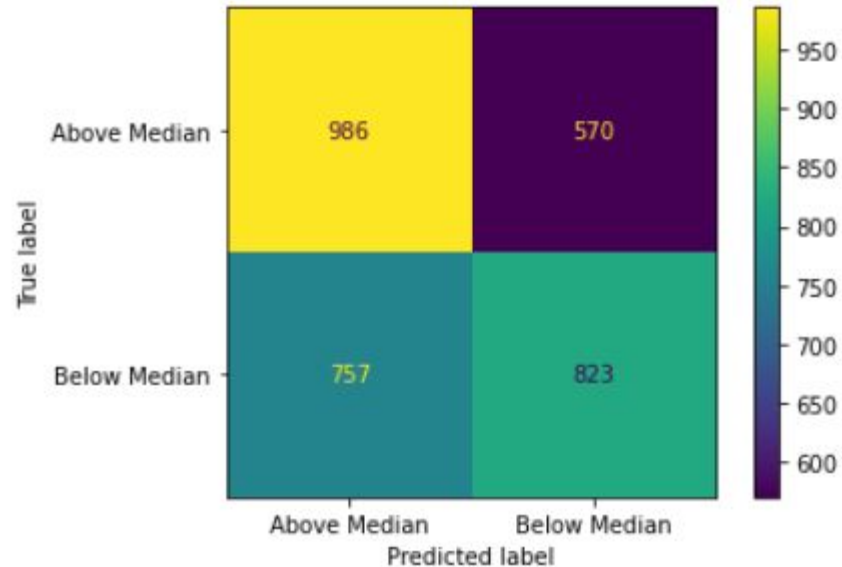
time, today, years, year

Logistic Regression

Confusion Matrix - shows us regression model results on our test data

About 56% accurate

Baseline - about 50%



Conclusions

Analyzing title alone predicts better than our baseline

Time - may be beneficial to analyze further on the concept of time

Recomendations

Be mindful of what title is used to post as it does appear to have an affect on comments received

Having a title/post relating to Time may increase chance of post popularity

Continuing Investigation

Compare more features such as subreddit, over 18, spoiler, and whether or not the post has a video to predict further.

Combine results with title analysis

Thank you for listening!