Logistic Regression on Author's Verified Status

Determining verification status using engagement metrics

ISSUE / PROBLEM

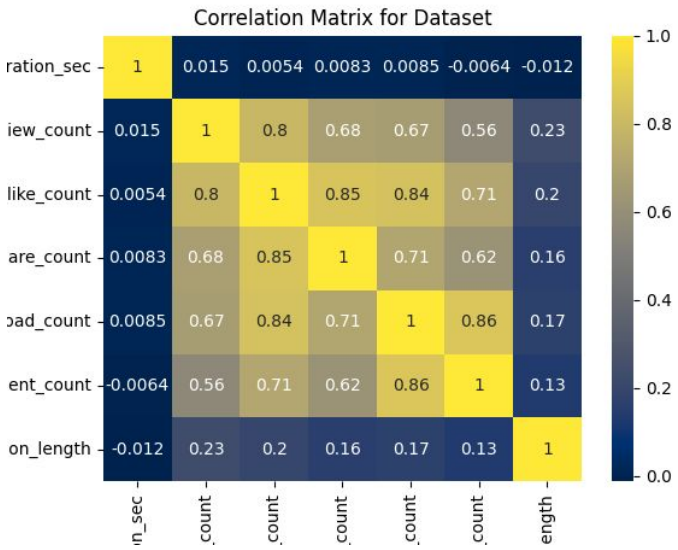
Tik-Tok needs to develop a model for determining the author verification status to eliminate the bottleneck on the classification process.

RESPONSE

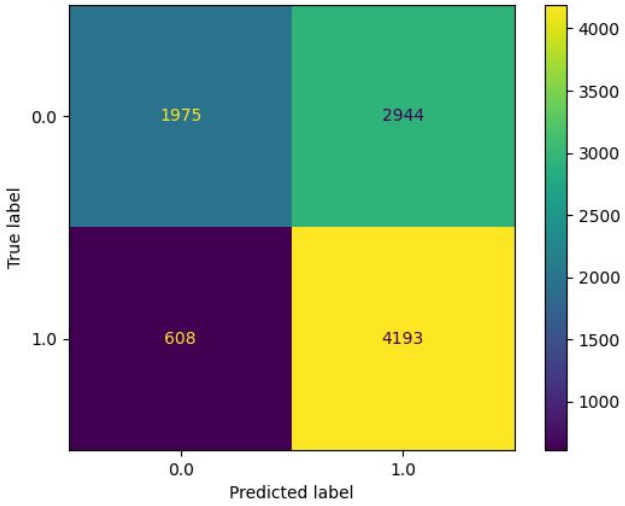
We developed a logistic regression model which considers multiple engagement metrics such as: views, comments, shares, likes, and downloads. In addition the model also consider categorical variables such as author ban status and and video claim status.

IMPACT

The model has a high recall value which means it predicts little false negatives.
However, the model does not perform very well with false positives as the precision metric is 59%.



Correlation matrix for numeric engagement metrics for tik-tok videos.



Confusion matrix for logistic regression model.

KEY INSIGHTS

The precision value tells us that the model is able to correctly classify the author verification status 59% of the time. The recall value indicates that this model will correctly find all instances of the positive class 87% of the time. The recall value compares the predicted positive instances (not verified) to the real values of the data.