Classifying Video Claim Status for Tik-Tok

Using Categorical Gradient Boosting Machine Learning

ISSUE / PROBLEM

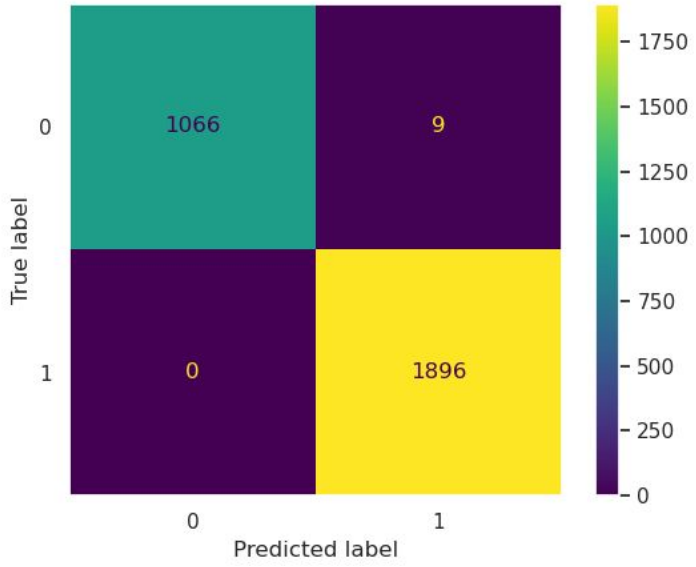
There is a backlog of user reports for classifying Tik-Tok video as “Claims” or “Opinions”. We need to create a classification algorithm to solve this problem.

RESPONSE

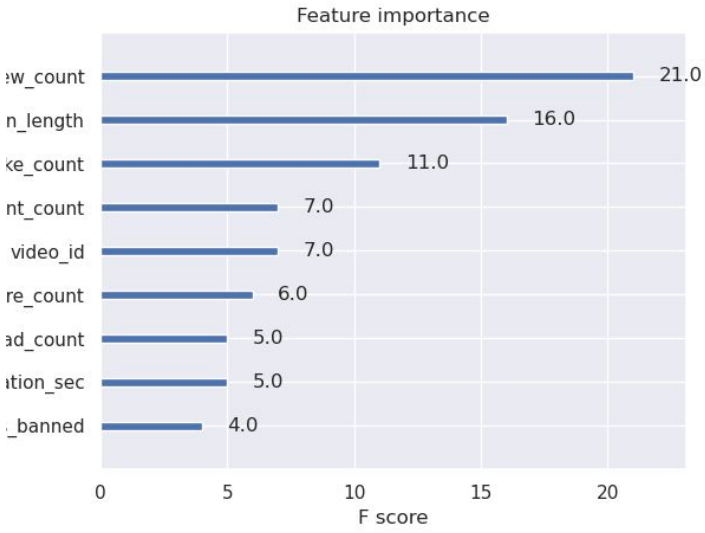
We are using engagement metrics such as video views, comments, shares, download and transcription length to determine the video’s claim status. The transcription length is defined as the number of characters is the transcript of the video.

IMPACT

The model developed is extremely accurate, while avoiding false classification of the video claim status. It was also constructed such that it handles randomness in data and bias.



Confusion matrix for XGBoost machine model.



Feature importance graph fro XGBoost machine model.

KEY INSIGHTS

The model indicate that the features with most predictive power towards video claim status are the video view count, transcription length and the like count. This could suggest that claim status is also related to video’s content.

The metrics used to evaluate the model indicate that it does not predict false positives nor false negatives with high frequency, we see that out of all 2900+ test samples only 9 were classified incorrectly.