Executive Summary: Regression Analysis

TikTok claims classification project

OVERVIEW

The data team is working towards building a machine learning model that can be used to determine whether a video contains a claim or whether it offers an opinion. With a successful prediction model, TikTok can reduce the backlog of user reports and prioritize them more efficiently. The data team has decided to explore how to predict verified status to help them understand how video characteristics relate to verified users.

PROJECT STATUS

Verified_status was selected for this regression model because of the relationship seen between the verified account type and the video content. A logistic regression model was selected because of the data type and distribution.

THE RESULTS

The logistic regression model achieved a precision of 67% and a recall of 65% (weighted averages). It also achieved f1 accuracy of 65%. The predictive power of this model is considered acceptable

NEXT STEPS

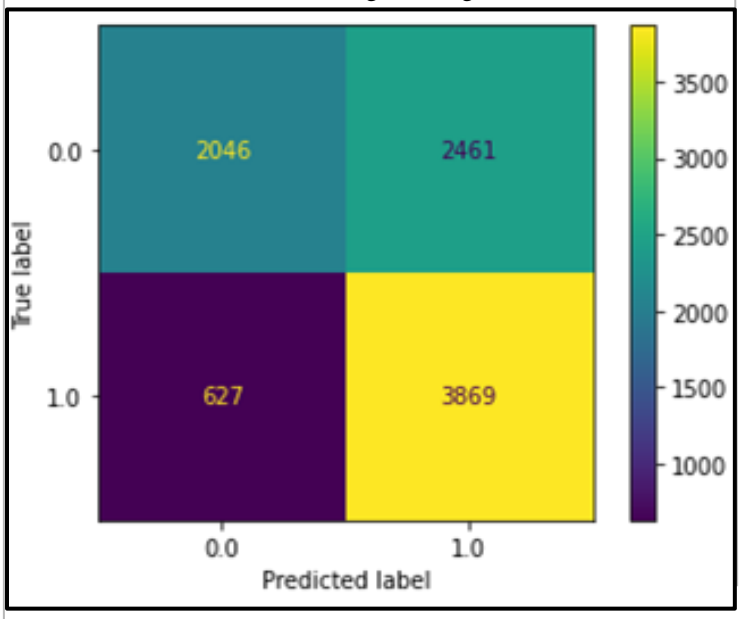
The next step is to construct a classification model that will predict the status of claims made by users.

KEY INSIGHTS

According to the estimated model coefficients derived from the logistic regression analysis, there appears to be a positive association between longer videos and an increased likelihood of users being verified.

Conversely, the model assigns relatively small estimated coefficients to other video features, suggesting that their impact on a user's verified status is minimal.

Confusion matrix for logistic regression model



Upper-left: the number of videos posted by unverified accounts. Upper-right: the number of videos posted by unverified accounts. Lower-left: the number of videos posted by verified accounts. Lower-right: the number of videos posted by verified accounts.