Tiktok Claims Classification Project

Executive Summary - Regression Analysis

OVERVIEW

The data team is tasked to create a machine learning model that can categorize user videos into claims vs opinions categories. We observed that accounts that are verified are more likely to post opinion videos compared to claims. Therefore in this part the team focused on creating a regression model that predicts account type.

PROJECT STATUS

The regression model was selected as task was categorizing the account type. Video engagement stats along with ban status were predictors.

Overall the model was able to achieve an accuracy of 65%. With a f1 score of 61%. The model is currently on the lower end on performance scale.

NEXT STEPS

The next step is to make a classification model that will predict the claim status of the video.

KEY INSIGHTS

The number of verified account types were significantly less therefore upscaling was required.

Video engagement stats had a high correlation with eachother. The like stats had > 0.8 correlation value therefore was dropped from model training.

Longer videos have a high chance of being from a verified account.