XGBoost Predictive Model for TikTok's Claim Classification: EDA, Hypothesis Testing, Logistic Regression, Tree-Based Models

Executive Summary for Tree-Based Classification Models

Overview

The TikTok team wants to develop a machine learning model to assist in the classification of videos as either claims or opinions. Previous investigation into the dataset indicated that video engagement levels were highly indicative of claim status.

Problem

TikTok videos receive a large number of user reports for many different reasons. Not all reported videos can be reviewed by a human moderator. Videos that make claims are much more likely to contain content that violates the TikTok's terms of service. The goal is developing method to identify videos that make claims to prioritize them for review.

Solution

The data team built two tree-based classification models. Both models were used to predict on validation dataset, and final model selection was determined by the model with the best recall score. The final model was then used to score a test dataset to estimate future performance.

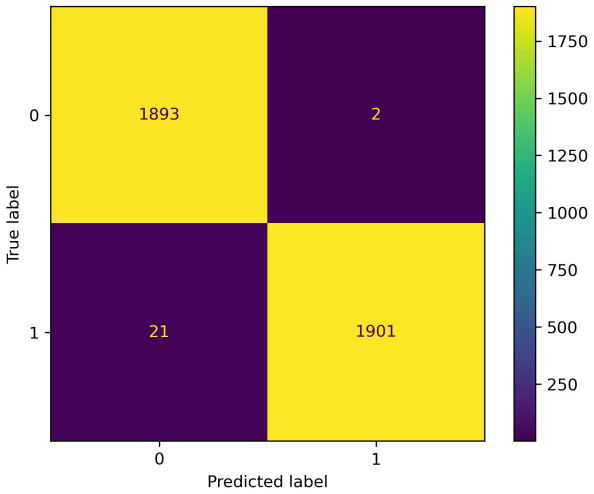
Details

Both random forest and XGBoost models performed exceptionally well. The XGBoost had a better precision and recall score and was selected as the champion.

Performance on the test holdout data yielded near perfect scores, with only 23 misclassified samples out of 3,817.

Subsequent analysis indicated that the primary predictor was 'video view count' -- related to video engagement levels. In conclusion, videos with higher user engagement levels were much more likely to be claims. In fact, no opinion video had more than 10,000 views.

Confusion matrix for the champion XGBoost model on test holdout data shows only 23 misclassified samples out of 3817.



Next Steps

Before deploying the model, further evaluation using additional subsets of user data is recommended. Moreover, the data team suggests monitoring the distributions of video engagement levels to ensure that the model remains robust to fluctuations in its most predictive features.