Machine Learning Model Outcomes

Executive summary report for TikTok prepared by the
TikTok data team

Overview

The TikTok data team seeks to develop a machine learning model to assist in the classification of videos as either claims or opinions. Previous investigation into the available data revealed that video engagement levels were highly indicative of claim status. The team is confident that the resulting model will meet all performance requirements.

Problem

TikTok videos receive a large number of user reports for many different reasons. Not all reported videos can undergo review by a human moderator. Videos that make claims (as opposed to opinions) are much more likely to contain content that violates the platform's terms of service. TikTok seeks a way to identify videos that make claims to prioritize them for review.

Solution

The data team developed two tree-based classification models, choosing the one with the highest recall score for final selection. The chosen model was then tested to estimate future performance.

Details

The random forest (RF) model, with a recall score of 0.995, marginally outperformed the XGBoost model and was selected as the champion model.

RF architecture misclassified only five out of 3,817 samples on the test dataset.

Analysis showed that the main predictors were video engagement metrics like view count, like count, share count, and download count. Videos with higher engagement were more likely to be claims, and no opinion video exceeded 10,000 views.

Confusion matrix for the champion RF model on test holdout data
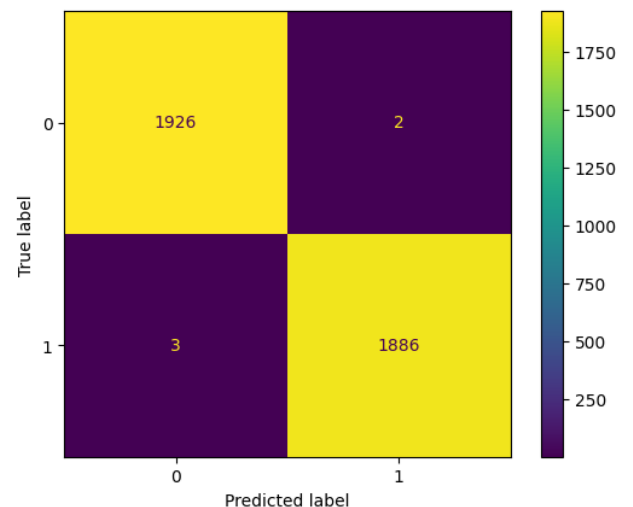
Next Steps

The model showed excellent performance on the test data. Before deployment, the data team advises further evaluation with more user data subsets and recommends monitoring video engagement metrics to maintain model robustness against feature fluctuations.