

Machine Learning Model Outcomes

Overview

I'm working on developing a machine learning model to classify TikTok videos as either claims or opinions. After analyzing the available data, I found that video engagement levels, such as views, likes, and shares, are strong indicators of whether a video is a claim.

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.

Solution

I built two tree-based classification models. Both models were used to predict on a held-out 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.

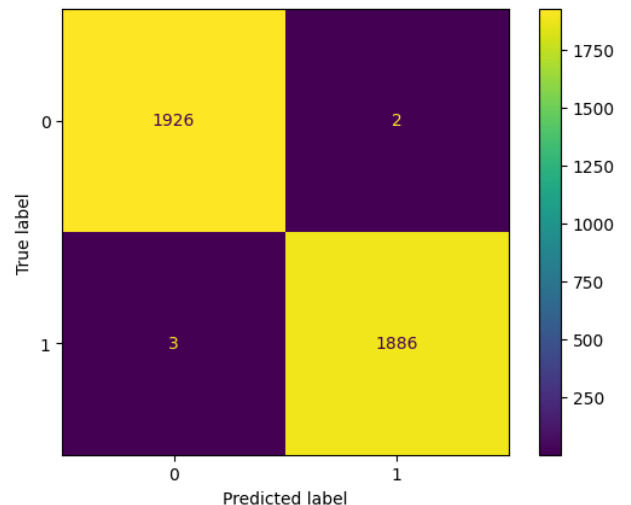
Details

Both model architectures—random forest (RF) and XGBoost—performed exceptionally well. The RF model had a better recall score (0.995) and was selected as champion.

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

Subsequent analysis indicated that, as expected, the primary predictors were all related to video engagement levels, with video view count, like count, share count, and download count accounting for nearly all predictive signal in the data. With these results, we can conclude that 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 RF model on test holdout data shows only five misclassified samples out of 3,817.



Next Steps

I've observed that the model performed exceptionally well on the test holdout data. Before deploying it, I recommend further evaluation using additional subsets of user data to ensure comprehensive accuracy.