

Machine Learning Model Outcomes

Executive summary report for TikTok prepared by the TikTok data team

Overview

The TikTok data team is working on creating a machine learning model to categorize videos as either claims or opinions. Earlier exploration of the existing data indicated that the levels of video engagement strongly correlate with claim status. The team expresses confidence that the upcoming model will fulfill all performance requirements.

Problem

TikTok videos attract numerous user reports for various reasons, and due to the volume, it's not feasible for every reported video to undergo human moderation. However, videos that make claims (as opposed to expressing opinions) are more likely to contain content that violates the platform's terms of service. TikTok is actively seeking a method to identify videos that make claims, aiming to prioritize them for review.

Solution

The data team constructed two tree-based classification models. Both models were applied to predict outcomes on a reserved validation dataset, and the ultimate model selection was based on the one with the highest recall score. The chosen model was subsequently employed to assess a test dataset, providing an estimate of its performance in future scenarios.

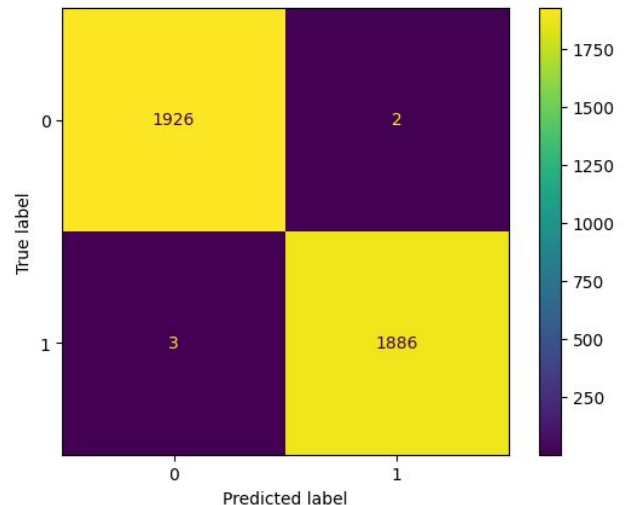
Details

Both model architectures, Random Forest (RF), and XGBoost demonstrated exceptional performance. The RF model, boasting a superior recall score of 0.995, was selected as the champion model.

Evaluation on the test holdout data produced near-perfect scores, with only five misclassified samples out of 3,817.

Further analysis revealed that, as anticipated, the primary predictors were closely linked to video engagement levels. Variables such as video view count, like count, share count, and download count contributed to nearly all predictive signals in the data. These results lead to the conclusion that videos with higher user engagement levels were significantly more likely to be claims. Interestingly, 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

As previously highlighted, the model demonstrated outstanding performance on the test holdout data. Before proceeding with deployment, the data team suggests conducting additional evaluations using different subsets of user data. Additionally, it is recommended to closely monitor the distributions of video engagement levels to ensure the model maintains its robustness in the face of variations in its most predictive features.