TikTok Claims Classification Project – Machine Learning Model

Executive Summary Report V

Milestone 6

ISSUE / PROBLEM

Videos that make claims, rather than simply expressing opinions, are more likely to contain content that breaches TikTok's terms of service. Therefore, TikTok needs a system to differentiate between videos that make claims and those that express opinions.

IMPACT

Given the model's strong performance on test data, the team recommends additional evaluations using diverse user data subsets before deployment. Additionally, it is advisable to monitor shifts in video engagement trends to maintain the model's effectiveness as the key predictors fluctuate.

RESPONSE

Two classification models based on tree algorithms were developed and tested. Both models were validated on a separate dataset, with the model demonstrating the highest recall score being chosen. This selected model was subsequently used to assess a test dataset, providing an estimate of its real-world performance.

KEY INSIGHTS

- Two model structures, Random Forest and XGBoost, are tested and both showed excellent results. The Random Forest model achieved a superior recall score of 99% and was thus chosen as the final model. Testing on the holdout data resulted in near-perfect accuracy, with only a few misclassified instances.
- Further analysis confirmed that the leading predictors were engagement-related metrics, such as video views, likes, shares and downloads, which contributed most significantly to the model's predictive power.
- The results suggest that videos with higher engagement are more likely to be categorized as claims.

	model	precision	recall	F1	accuracy
	Random Forest	0.997356	0.998412	0.997884	0.997904

Negative (0) = Opinion

Positive (1) = Claim