Title

Subtitle

Project Overview

Details

Key Insights

Image Alt-Text Here

Next Steps

Title

Subtitle

ISSUE / PROBLEM

RESPONSE

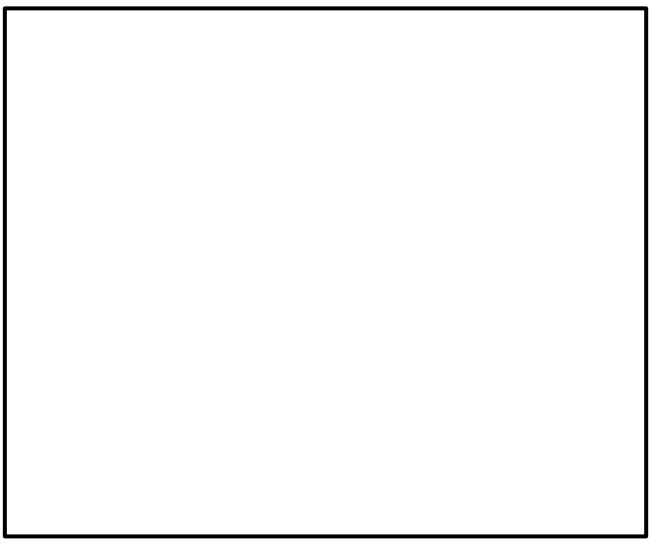


Image Alt-Text Here

IMPACT

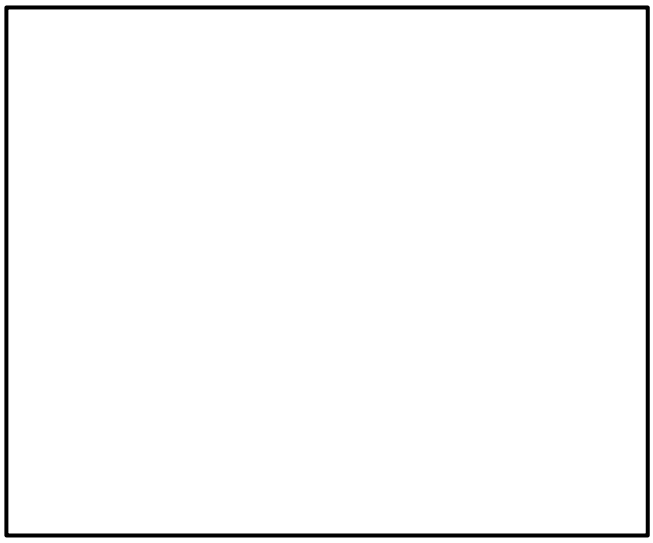


Image Alt-Text Here

KEY INSIGHTS

Title

Subtitle

ISSUE / PROBLEM

RESPONSE

KEY INSIGHTS

IMPACT

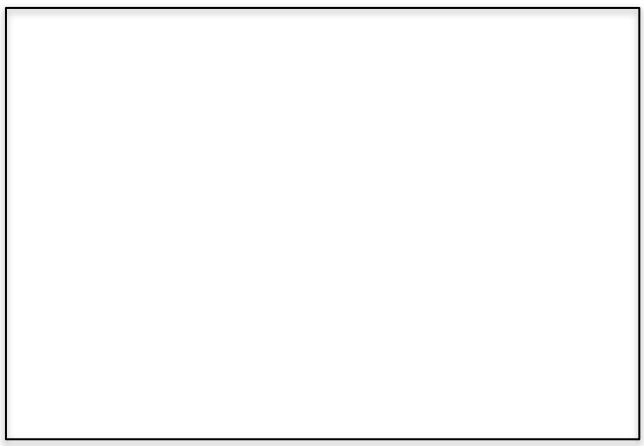


Image Alt-Text Here

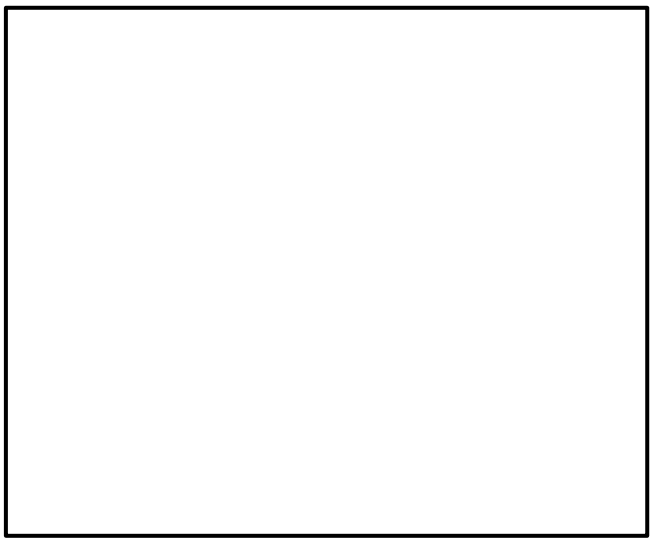
Title

Subtitle

OVERVIEW

PROJECT STATUS

KEY INSIGHTS



NEXT STEPS

Claim Classification Project

Executive Summary - Final Machine Learning Model Outcome

Overview

The TikTok data team seeks to develop an accurate predictive model that determines whether a video contains a claim or an opinion.

Problem

TikTok videos receive a large number of user reports for many different reasons. It's impossible for human moderators to review each reports individually. Therefore Tiktok needs a way to identify videos that make claims to prioritize them for review.

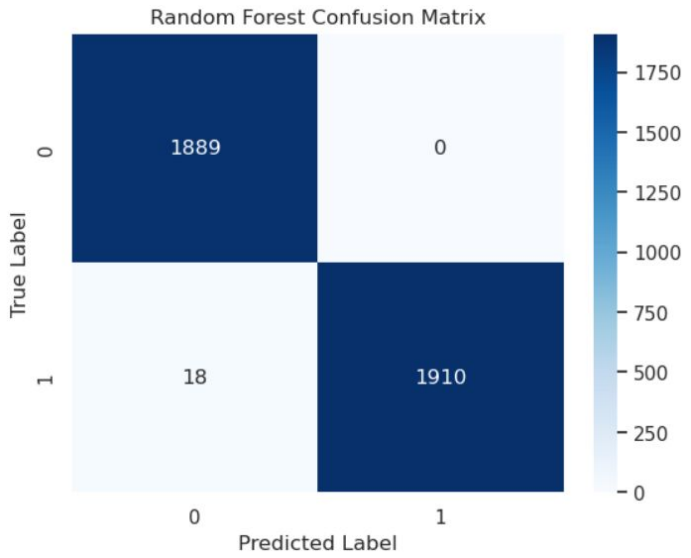
Solution

The data team built a random forest classification model to classify whether a claim or opinion is presented in the video.

Details

We built two models, random forest and XGBoost. They both performed exceptionally well and have similar scores (99% accuracy).

The confusion matrix on the right shows the result of the random forest model on the testing dataset. It predicted opinions perfectly and only misclassified 19 claims videos as opinion.



Subsequent analysis on feature importance shows that the most predictive features are related to video engagement levels such as video view count, like count, share count, and download count. With these results, we can conclude that videos with higher user engagement levels are much more likely to be claims.

Next Steps

The data team recommends further evaluation using more user data to test the model and make sure the model performance doesn't fluctuate much before deploying the model.

Title

Subtitle

Overview

Objective

Results

Next Steps