TikTok Claims Classification Project – Regression Analysis

Executive Summary Report IV

Milestone 5

ISSUE / PROBLEM

Earlier analysis suggested that verified users are more likely to share opinions in their posts. To better classify claims and opinions, a model that can predict the behavior of verified accounts is essential. A regression model is developed to predict whether a user is verified based on their video features.

IMPACT

The logistic regression model achieved a precision of 63% and a recall of 82%, though its overall accuracy of 67% was slightly lower than ideal. These results have provided valuable insights into video characteristics. The next phase involves building a classification model to predict the nature of claims made by users. This step will provide further insights into the behaviors of verified users, setting the stage for the final project objective.

RESPONSE

The logistic regression model focuses on predicting verified status due to the observed connection between account verification and video content. A logistic regression approach was selected because it aligns well with the type and distribution of the data.

KEY INSIGHTS

- Some variables showed strong correlations, posing potential multicollinearity challenges for the logistic regression model.
- The coefficients indicate that an opinion video significantly increases the likelihood of the user being verified and every additional second of video length slightly decreases the likelihood of the user being verified.
- The model demonstrated adequate performance. The analysis highlighted that opinion videos are more likely to be posted by verified users and that banned users are less likely to be verified, whereas other features had less significant associations with verified status.

	Precision	Recall	F1-Score	Accuracy
LogReg	63%	82%	71%	67%

Negative (0) = Unverified user

Positive (1) = Verified user