TikTok claims classification project

Phase-5-Executive summary report for TikTok prepared by the TikTok data team

ISSUE / PROBLEM

The main issue is the backlogs on user reports made by users on specifically videos. The project aims to create a classification model to classify reports into claims and opinions. This particular stage explored the relationship between verified status(a strong predictor of claim_status as found in previous stages) and other variables in the dataset.

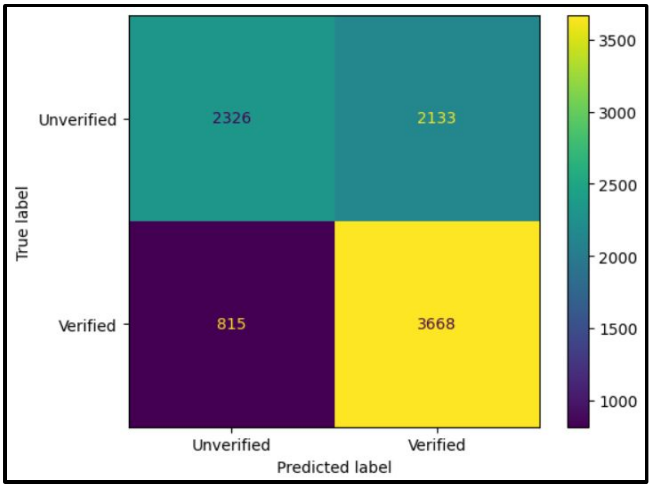
RESPONSE

A logistic regression model was developed with verified_status as the dependent variable and video engagement metrics—comments, views, shares, and downloads—as independent variables. The dataset was balanced, the model was built using necessary libraries, and evaluated with a confusion matrix. This helped identify key predictors of verification status, which can later be used as covariates or interaction terms in the final claim classification model to better distinguish between claims and opinions.

IMPACT

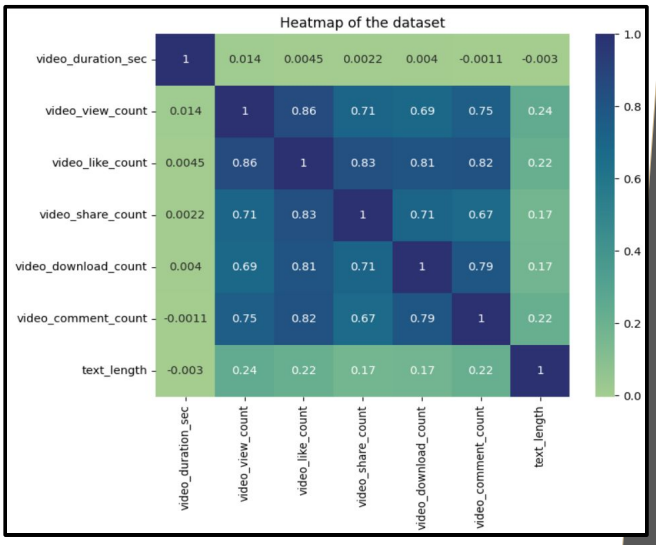
The overall impact of the project includes a reduction in user report backlogs, while this stage plays a crucial role in the final classification model. Understanding and modeling the relationship between verified_status and video engagement metrics ensure that the final model is robust, interpretable, and free from hidden biases.

The logistic regression model achieved 63% precision and 82% recall for the "not verified" class, with an overall accuracy of 67%. The "verified" class had 74% precision and 52% recall, with weighted averages balancing both classes.



The confusion matrix shows the model's predictions vs. actual verification status. The top-left represents correctly classified "Unverified" users, while the bottom-right shows correctly classified "Verified" users. Misclassifications appear in the top-right (false positives) and bottom-left (false negatives). Lighter colors indicate higher values.

The correlation matrix shows strong relationships among engagement metrics, with video_like_count highly correlated with views (0.86), shares (0.83), downloads (0.81), and comments (0.82). Since logistic regression is sensitive to multicollinearity, removing video_like_count helps maintain model stability.



This correlation matrix visualizes the relationships between different video engagement metrics. Darker shades indicate stronger correlations, while lighter shades represent weaker relationships. High correlations are observed between video views, likes, shares, and downloads, suggesting that these metrics often increase together. Video duration and text length show little to no correlation with engagement metrics.

KEY INSIGHTS

- Claim Status Drives Verification - Users with *claim_status_opinion* are far more likely to be verified, linking content type to verification.
- Bans & Reviews Lower Verification Chances - Banned users rarely get verified, while *under review* users also face hurdles.
- Engagement Has Little Impact - Views, shares, downloads, and comments don't significantly affect verification.
- Content & Policy Matter More - Verification is influenced by *claim_status_opinion* and compliance with platform rules.
- Next Step - With sufficient insights into user behavior and verification patterns, the next step is building a classification model to predict whether a user's submission is a claim or an opinion—the final goal outlined by the TikTok team.