

Executive Summary

Milestone # 05

➤ ISSUE / PROBLEM

TikTok team si adamant to pick out ideas and classify them as claims or opinions. The team had previously performed exploratory data analysis and hypothesis testing to gain key insights about data..

➤ RESPONSE

To know about how different variables are associated with whether a user is verified, we are building a logistic regression model using `verified_status` as the outcome variable.

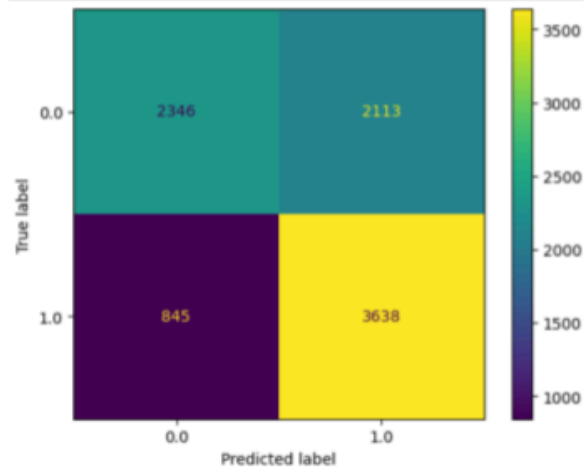
One of the model assumptions for logistic regression is no severe multicollinearity among the features. To build a logistic regression model that meets this assumption, we are excluding `video_like_count`

➤ IMPACT

This logistic regression will help us understand the association and correlation of multiple variables with `verified_status`

➤ ANALYSIS

Built and Evaluated Logistic Regression Model



- The upper-left quadrant displays the number of true negatives: the number of videos posted by unverified accounts that the model accurately classified as so.
- The upper-right quadrant displays the number of false positives: the number of videos posted by unverified accounts that the model misclassified as posted by verified accounts.
- The lower-left quadrant displays the number of false negatives: the number of videos posted by verified accounts that the model misclassified as posted by unverified accounts.
- The lower-right quadrant displays the number of true positives: the number of videos posted by verified accounts that the model accurately classified as so.

➤ KEY INSIGHTS

The classification report above shows that the logistic regression model achieved

- A precision of 61%
- A recall of 84%
- An accuracy of 65%

The logistic regression model had not great, but acceptable predictive power: a precision is less than ideal, but a recall is very good. Overall accuracy is towards the lower end of what would typically be considered acceptable.

Overall, This model might not be ideal for the scenario.