Executive Summary: Regression Analysis

TikTok Claims Classification Project

Project Overview

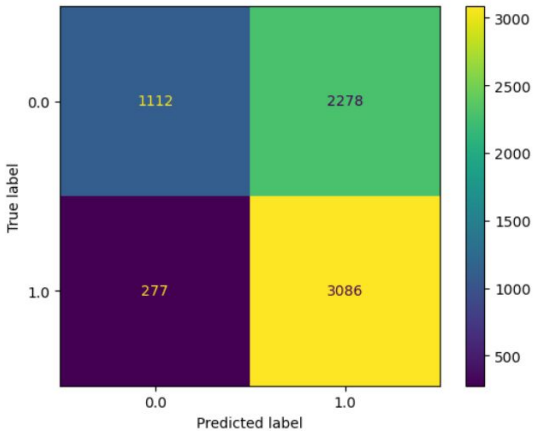
This project aims to build a logistic regression model that can predict the verified status of the user. This will assist the work on the larger project scope which is to predict whether a video is a claim or an opinion, as it has been observed that verified users are much more likely to post opinions.

Key Insights

- A logistic regression model was built to predict the verified status of the user given information about the video content.
- Video like count and video share count variables were found to be highly correlated. Therefore, only one of these variables - video like count - was used in the model.
- The model evaluation metrics were calculated as:
  - 69% precision
  - 62% recall
  - 59% f1-score.

Details

- The coefficient of regression of the variables were very small, therefore they do not seem to have a large effect on the verified status.



The matrix shows the number of true labels vs. the number of predicted labels (where label could be verified or unverified).

Next Steps

- The next step would be to construct the final model, i.e., to classify content into claims vs. opinions. This is the larger scope of the project, and the team now has all the necessary information to start building the model.