# TikTok Claims Classification Regression Analysis

Executive Summary

## OVERVIEW

We built a logistic regression model to predict the behaviour of TikTok users based on their account type, verified or not verified. The reason behind the creation of this regression model is to assist in the development of a machine learning model aimed to classify the claims for user submissions. During earlier stages of the project it was observed that if a user is verified they are more likely to post an opinion further pushing the development of this logistic regression model.

## PROJECT STATUS

Since there has been an observed relationship between the verified accounts and posting of opinions we decided to select the verified_status variable for this regression model. A logistic regression model was chosen based upon the characteristics of the dataset, we believe it was the best option.

The model achieved a weighted average precision of 67% as well as a recall of 65% with an f1 accuracy of 64%.

## NEXT STEPS

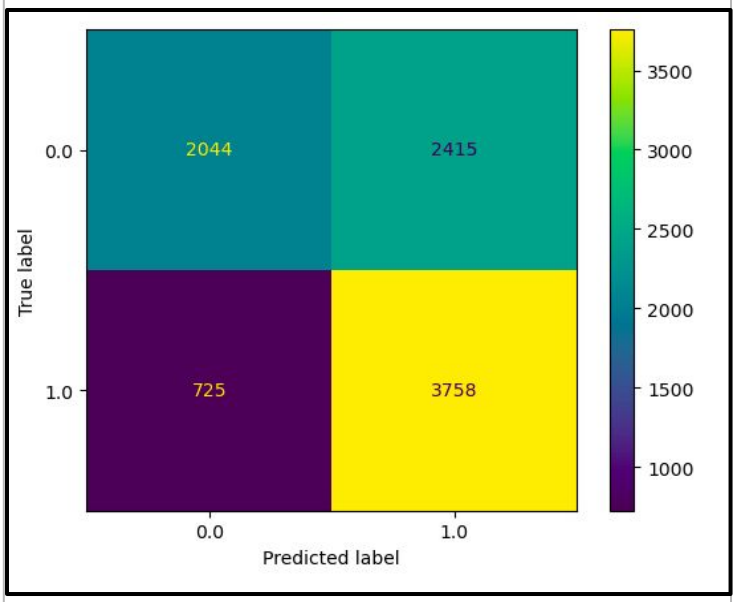
We believe the next steps should be the creation of the classification model that will ultimately predict the status of claims made by TikTok users.

The groundwork has been completed for the creation of the new model, we have a better understanding of user behaviour allowing for the analysis of said model.

## KEY INSIGHTS

Based on the estimated model coefficients it was observed that videos with a longer run time have a higher chance of being posted by a verified user.

Variables other than video length had very small estimated model coefficients, therefore we can assume that the other variables are not associated with verification status




The upper left quadrant: True Negatives
The upper right quadrant: False Positives
The lower left quadrant: False Negatives
The lower right quadrant: True Positives