

TikTok Claim Classification Project

Claim Classification

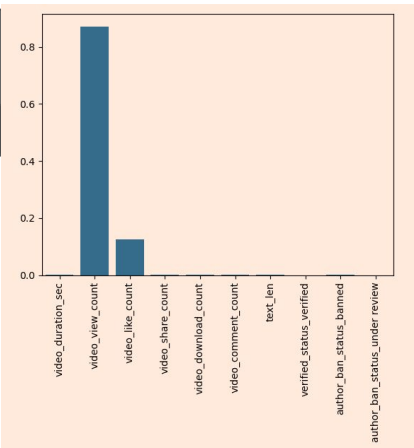
Project Overview

TikTok has tasked us with creating a classification model that will distinguish between real claims and opinions for videos based of some video metrics.

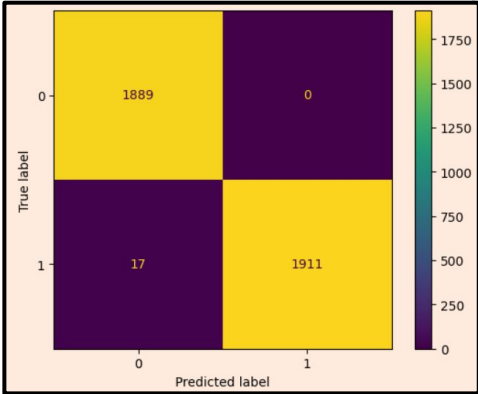
Key Insights

The model we have built achieves a 99.5% accuracy on test data with 100% recall. While these sound good, it makes its predictions based of mostly view count and like count which will end up discriminating against popular videos and will miss predicting less popular videos as breaking T&C's. To fix this, we will either need additional data with more varied distribution of claims across view counts or build another model that doesn't use these features.

Details



Feature importance for final model



Confusion Matrix for test data

Image on the right shows confusion matrix on test data having 0 false negatives and image on left shows importance of views and likes combined being almost 1.0

Next Steps

Create a model that is less reliant on engagement metrics and deploy it to bring down videos that break Terms and Conditions faster.