Machine Learning Model Outcomes

Executive summary report of Model Evaluations

Overview

The data team was tasked at making a ML model that can classify videos into claims and opinions. Out of all the predictors, video engagement levels proved to be more indicative of the status and thus the model was trained on them. The random forest model is able to show a f1 score of 0.99

Problem

Everyday several videos get reported by users. Videos that are claim are prioritized to be reviewed by human moderator. The task is to identify video that are claims for easy review.

Solution

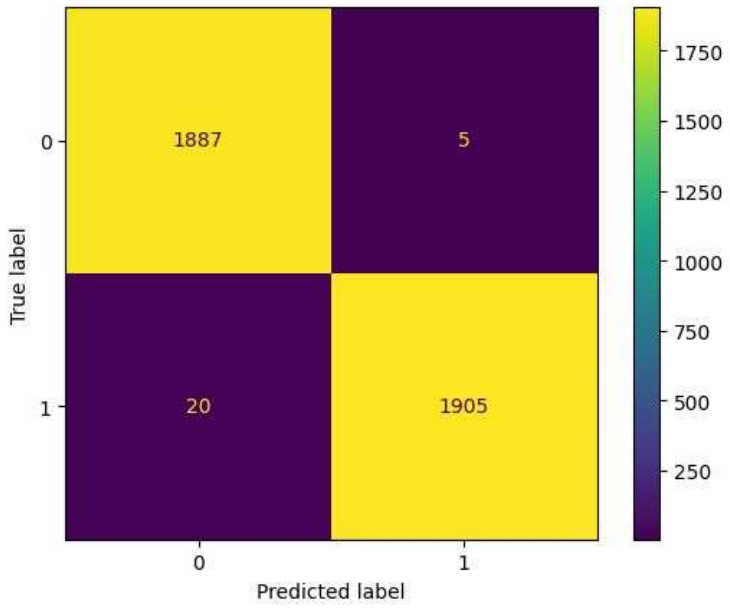
The data team built random forest and XGBoost models. Both models are worked well on validation sets and testing sets. The random forest model turned out to be better than XGBoost.

Details

Both the random forest and XGBoost models perform well on the testing sets but based on higher f1 score random forest is selected as champion model.

The random forest model had around 25 misclassifications compared to 30 for XGBoost.

In both models, the primary indicators were engagement levels of videos like views, shares, downloads etc.



Next Steps

The model performed well and can be used for deployment. Some more testing is recommended using different fluctuation levels of these features.