Classification of TikTok Videos
Machine Learning Project

Overview

The main objective of this project is to develop a machine learning model to predict whether videos reported by users presented claims or opinions to improve triaging process of videos for further review by human moderators.

Tools Used

* Programming Language: **Python**
* Data Science Libraries/Packages: **pandas, numpy, matplotlib, seaborn, plotly, statsmodels, scikit-learn, xgboost**
* Statistical Methods: **Descriptive statistics, Two-sample proportions z-test**
* Machine learning algorithms: **Random forest, Gradient boosting (XGBoost)**

Dataset Used

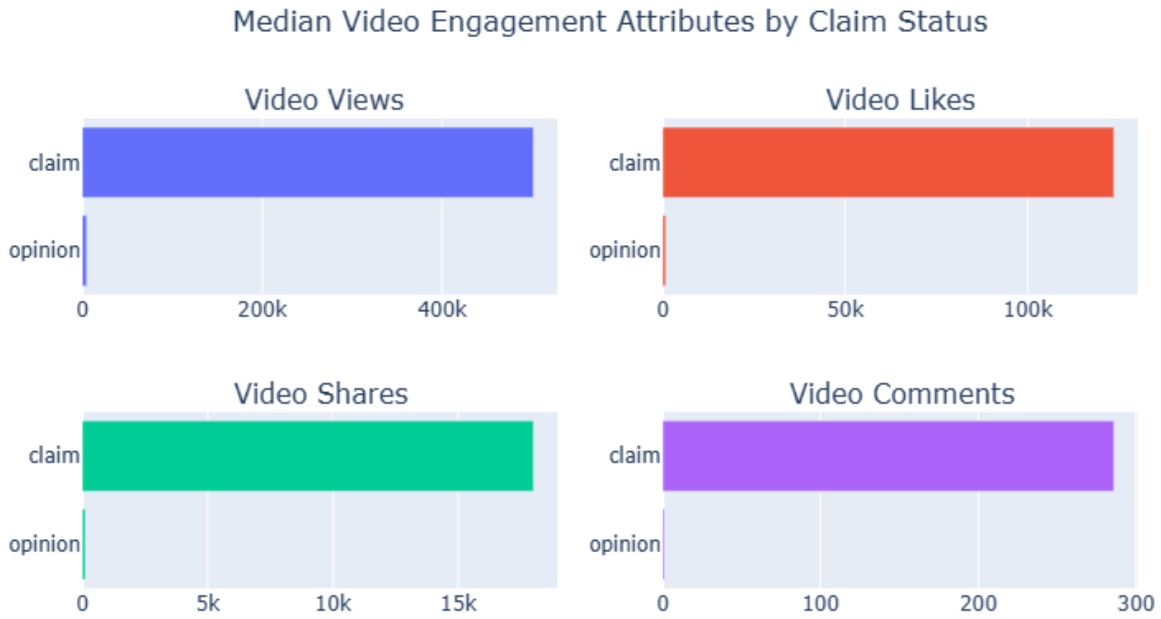
The dataset [tiktok_dataset.csv](data/tiktok_dataset.csv) contains data of reported videos where each video is labeled as having a claim or opinion

Project Background

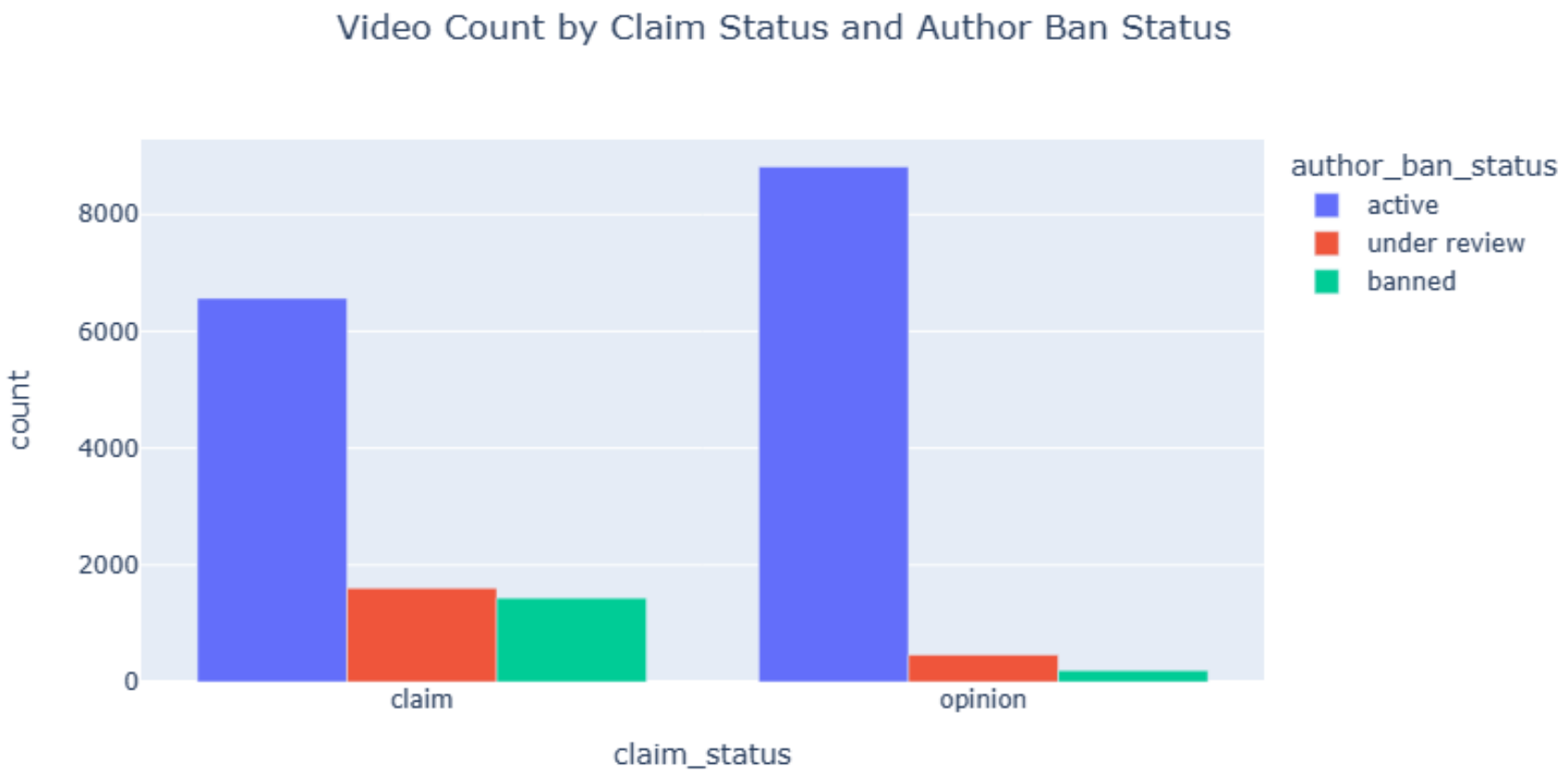
TikTok is the leading social platform for short-form videos. TikTok users can report videos and comments having user claims. These reported videos may have content that violates the company's terms of use policy and needs to be reviewed by human moderators. The successful development of a predictive model will help reduce the backlog of user reports and process them more efficiently.

Exploratory Data Analysis

A dataset having reported videos labeled as `claim` or `opinion` was provided for the project. Cleaned data has 19,084 entries with 12 attributes. Exploratory data analysis show that claim videos tend to have higher video engagement (views, likes, shares, comments, downloads) compared to opinion videos as shown in the figure below.



There is also a statistically significant difference in proportion of claim videos between videos posted by verified and unverified accounts, where verified accounts are more likely to post opinion videos. In addition to, authors that post claim videos were also shown to have a higher likelihood of getting banned as shown in the figure below, suggesting that claim videos are more likely to have content that violate TikTok's terms of use policy.



Along with video engagement attributes, video duration, verified status, and author ban status, features were engineered from the video transcription text were selected as features to be used for building the video claims classification model.

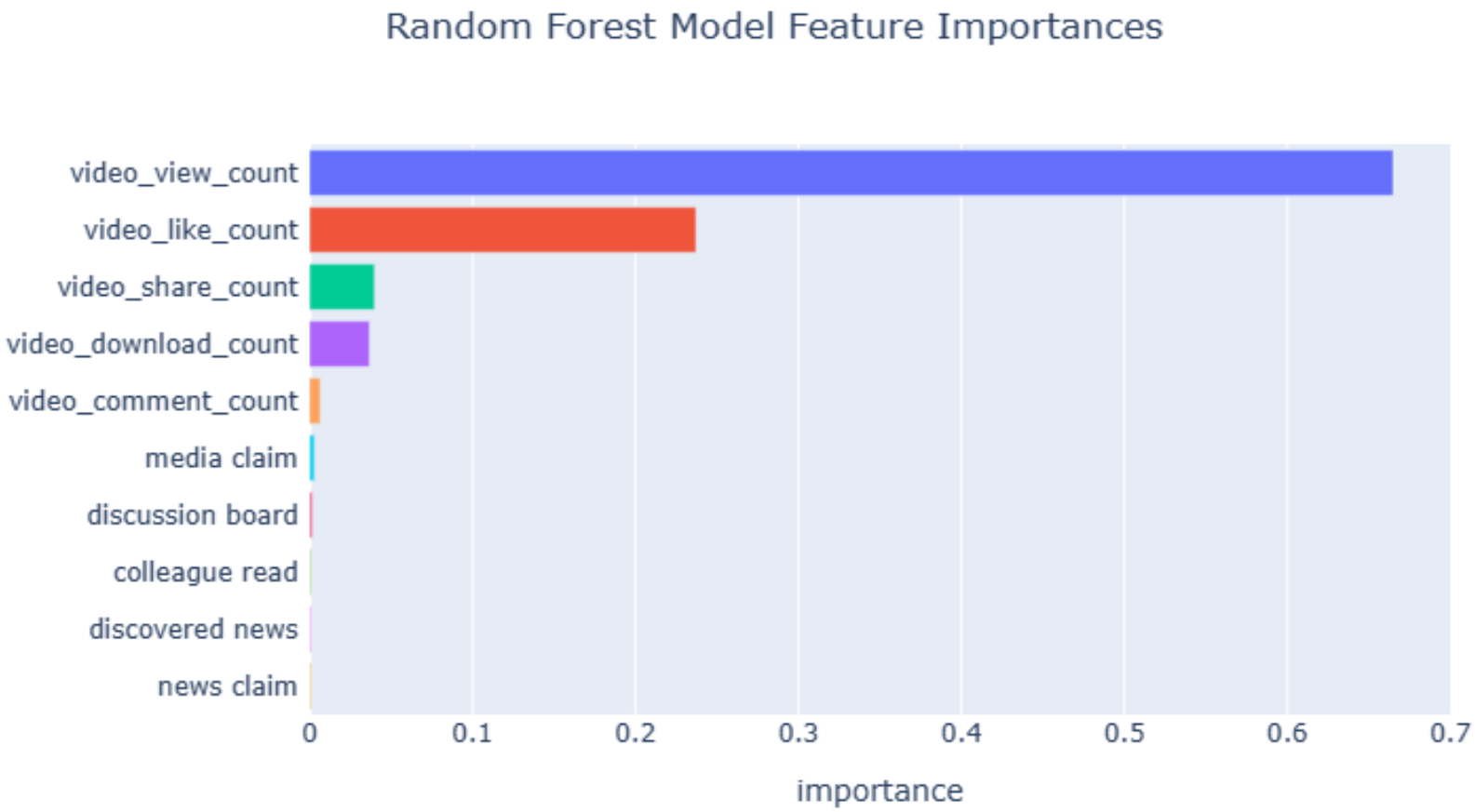
Machine Learning Model Development

A successful predictive model will help streamline the review process of reported videos, but incorrect predictions come with consequences. False positive predictions will prioritize videos that are less likely to violate terms of use policy, decreasing the efficiency of the review process. On the other hand, false negatives will fail to identify videos that are more likely to violate terms of use policy and need to be prioritized for human review. False negative predictions have more negative impact to the business, so recall score will be used as an evaluation metric for the machine learning model to penalize false negative predictions.

Random forest and XGBoost models were constructed and evaluated on validation data. The random forest model performed better; therefore, it was selected as the champion model to be evaluated using test data to assess real-world performance. The constructed model predicted the test data almost perfectly, with performance metrics shown in the table below.

Metric	Score
Accuracy	0.9987
Precision	0.9989
Recall	0.9984
F1	0.9987

The top features with highest importances were also obtained and plotted in a bar chart shown below.



Unsurprisingly, the top five most important features are the video engagement attributes. While a random forest model does not provide detailed information on how these features were used to make predictions, earlier EDA has shown that claim videos have higher video engagement. The next five features were the vectorized transcription text. While these engineered features may have helped in classifying claims, these features have very low importances compared to the top five.

Conclusion

The model performs very well on test data, implying excellent real-world performance. The model is ready to be incorporated in the video review process as an automation tool for initial screening of reported videos. While the model accurately recognizes patterns differentiating claim from opinion videos, it is recommended to periodically re-evaluate the real-world performance of the model as these patterns may change over time.

Finally, a recommended next step is to streamline the model development pipeline so that the model can be easily updated regularly with new data. This will enable the developed model to capture new trends that may emerge over time.