Final Claims Classification Model

Milestone 6

Overview

The TikTok data team has developed a machine learning model to classify whether a video contains a **claim** or an **opinion**. This model is intended to help reduce the backlog of user-reported videos and improve the efficiency of moderation efforts. Following previous experiments with logistic regression, the team evaluated advanced tree-based models, ultimately selecting **Random Forest** as the final approach.

Solution

Two high-performing models — Random Forest and XGBoost — were trained and evaluated using a 60/20/20 train/validate/test data split. Based on the evaluation results, Random Forest was selected due to its slightly superior performance. The model was developed using 10 engineered features, primarily centered on video engagement metrics and author attributes.

Key parameters for Random Forest:

- `n_estimators`: 20
- `max_depth`: 5
- `max_features`: 'sqrt'
- `max_samples`: 0.9
- `min_samples_split`: 0.01

Next Steps

While the model performs exceptionally well on test data, the data team recommends:

- Further validation on broader, real-world user data subsets.
- Ongoing monitoring of video engagement feature distributions to ensure model stability.

Results

The Random Forest model demonstrated excellent performance:

- **Accuracy**: 100%
- **Precision**: 0.9995
- **Recall**: 0.9909
- **F1-score**: 0.99+

Top predictive features:

- `video_view_count` (52.1%)
- `video_like_count` (25.0%)
- `video_download_count`, `share_count`, and `comment_count` (combined ~22%)

KEY INSIGHTS

On the test data we got next results: