Predicting Verified User Status Using Logistic Regression

Overview

The TikTok data analytics team developed a logistic regression model to explore how video characteristics relate to whether a user is verified. By analyzing engagement metrics such as likes, comments, and video duration, the model aims to identify patterns associated with verified users. This analysis supports TikTok's broader initiative to understand how verified accounts differ in content and engagement behavior compared to non-verified users.

Objective

The primary goal of this project was to determine which video-level features best predict a user's verification status. The model used the following variables:

- **Dependent Variable:** (1 = verified, 0 = not verified)
- **Independent Variables:**

Results

By identifying these relationships, the team can better inform future modeling efforts and content strategies related to user credibility and classification.

Metric	Score
Accuracy	**0.63**
Precision	**0.58**
Recall	**0.88**
F1-Score	**0.70**

Next Steps

1. **Model Refinement:** Explore advanced algorithms (e.g., Random Forest, XGBoost) to improve prediction accuracy and handle non-linear relationships.
2. **Feature Expansion:** Incorporate additional user-level features (e.g., follower count, posting frequency) for deeper insights.
3. **Operational Use:** Apply engagement-based insights to help optimize content strategies, verification criteria, and future classification models distinguishing between claims and opinions.