TikTok Claims Classification

Data Inspection and Preparation Update

OVERVIEW

The TikTok data team is working on a machine learning model to classify claims in videos. This project aims to streamline the claims review process by determining whether statements made in videos are claims (fact-based) or opinions (subjective beliefs). By analyzing engagement patterns and textual content, this project will help automate content moderation and improve classification accuracy.

PROJECT STATUS

- ✅ Project proposal approved, allowing access to TikTok's user data.
- ✅ Initial data review and structuring completed, focusing on relevant variables.
- ✅ Identified key differences between claims vs. opinions to guide feature engineering.
- 🚀 Next phase: Preparing dataset for machine learning model training.
-

NEXT STEPS

✔ Feature Engineering for Machine Learning – Develop text-based predictors to distinguish claims from opinions.
✔ Train Initial Classification Model – Apply supervised learning algorithms to assess model accuracy.
✔ Enhance Dataset with Additional Features – Explore context-based variables, such as source credibility or engagement trends.
✔ Test and Validate Model Performance – Use cross-validation to refine model accuracy.
✔ Deploy Model for Content Moderation – Implement AI-powered automation to assist TikTok's content review teams.

KEY INSIGHTS

1. Claims vs. Opinions Classification – Claims involve unsourced or unverified statements, while opinions express subjective viewpoints.
2. Data Organization – Structured dataset columns, removed irrelevant variables, and created meaningful new features for classification.
3. Key Variables Identified – Certain words, phrasing, and engagement metrics correlate with claim vs. opinion classifications, helping guide feature selection.