

Project Proposal

TikTok Claims Classification with Machine Learning

1. Introduction

This project proposal outlines the development of a machine learning model to classify user comments on TikTok videos as containing claims or expressing opinions. An effective model will significantly improve TikTok's ability to prioritize user-reported content and address potential moderation issues more efficiently.

2. Project Goals

- Build a reliable machine learning model that effectively classifies user comments as containing claims or expressing opinions.
- Reduce the backlog of user-reported content through efficient prioritization based on model predictions.
- Enhance the overall user experience on TikTok by ensuring timely moderation of flagged content.

3. Audience

This proposal is intended for both technical and non-technical audiences within TikTok.

- The Data Team (Willow Jaffey, Rosie Mae Bradshaw, Orion Rainier): This section provides detailed technical tasks and considerations.
- Cross-Functional Team Members (Mary Joanna Rodgers, Margery Adebawale, Maika Abadi): This section focuses on high-level project goals, milestones, and deliverables.

4. Methodology

The project will follow the PACE workflow, a structured framework for data science projects. This approach ensures a logical progression through data acquisition, exploration, analysis, model construction, and execution.

5. Project Deliverables

- **Project Management Plan:** This document will outline the project timeline, milestones, tasks, and resource allocation. It will be a key reference point for monitoring progress and ensuring adherence to the project schedule.
- **Exploratory Data Analysis (EDA) Report:** This report will summarize the initial exploration of the comment data. It will describe key characteristics of the data, identify potential issues, and provide insights that will inform subsequent stages of the project.
- **Machine Learning Model Documentation:** This document will comprehensively detail the chosen machine learning model, including its architecture, training process, and evaluation metrics. It will also address the model's assumptions and limitations.
- **Presentation Materials:** A clear and concise presentation will be developed to communicate the project's findings to stakeholders. It will showcase the machine learning model, its performance, and potential impact on content moderation efforts.

6. Timeline

The project is expected to be completed within six weeks, following the breakdown of stages outlined in the PACE workflow:

- **Plan (1 week):** Focus on data acquisition, initial exploration, project workflow definition, and model selection.
- **Analyze (2 weeks):** Clean and pre-process the data, train and evaluate the chosen machine learning model.
- **Construct (1 week):** Refine and optimize the model based on evaluation results.
- **Execute (1 week):** Develop a deployment plan and document the final model.

7. Team Roles and Responsibilities

- **Willow Jaffey (Data Science Lead):** Provide guidance on model selection, regression techniques, and hypothesis testing. Oversee the overall project direction.
- **Rosie Mae Bradshaw (Data Science Manager):** Facilitate communication and collaboration within the data team. Manage project resources and timelines.
- **Orion Rainier (Data Scientist):** Perform data acquisition, exploration, cleaning, and preprocessing. Train and evaluate the machine learning model.

- **Mary Joanna Rodgers (Project Management Officer):** Track project progress and ensure adherence to milestones. Assist in creating visuals for presentations.
- **Margery Adebawale (Finance Lead, Americas):** Provide financial considerations and resource allocation if needed.
- **Maika Abadi (Operations Lead):** Offer insights on potential operational impacts of the model's deployment.

9. Follow Up

Upon approval of this proposal, the data team will immediately commence the project activities outlined in the PACE workflow. Regular communication and progress updates will be provided to stakeholders throughout the project lifecycle.