Project Proposal: Analyzing Content Strategy for Maximum Engagement

Social media platforms are of the utmost importance in today's society and important in the way businesses market their products. Maximizing engagement is beneficial in using these platforms correctly for larger profit margins. This project explores how different social media content types, hashtags, and platforms influence user engagement. The goal is to help brands and content teams optimize their content strategies based on data, not guesswork. Given the metadata of a social media post (platform, hashtag, content type, region, and raw engagement metrics), can we predict whether the post will have Low, Medium, or High engagement?

Companies always ask: "What type of content should we post, and where?" Companies spend significant time and money producing content for platforms like TikTok, Instagram, YouTube, and Twitter, but often lack insight into what drives engagement across regions or content types. With this project that question could be answered by building tailored strategies based on platform and/or region. These businesses that utilize these strategies can increase ROI by helping marketing teams pre-evaluate content and optimize social strategy.

To deem this project successful, the model will need to predict all three classes fairly well and perform better than naive baselines. As well as, explain why the model works and which factors drive engagement and deliver clear takeaways that a marketing/content team could act on. The model will simulate a real product or tool that could improve content decisions.

This initiative is focused exclusively on helping content strategy and marketing teams make data-informed decisions about which types of posts are most likely to receive high engagement before publishing. This project is focused on predicting engagement level using post metadata, identifying content strategies that perform well, and delivering insights that help teams plan more successful content.

This project faces several key constraints that may limit its success. The dataset lacks important features such as post content and user-level data, which are known to influence engagement. The definition of Engagement_Level may be unclear or inconsistently labeled, and class imbalance could bias the model toward the most common outcomes. Additionally, regional and platform-specific engagement behaviors may not be fully captured, and without a clear deployment or adoption plan, the model's business value may be limited.

Key stakeholders for this project include the marketing and content strategy teams, who will use the model's insights to optimize social media campaigns, and the data science or analytics team, responsible for building and validating the model. Data is sourced from a public social media trends dataset. Once a solution is identified, recommendations will be presented to marketing leadership or campaign managers to guide data-driven content decisions.

To solve the problem of predicting social media engagement levels, key data needed includes the platform, content type, hashtag, and region of each post, along with engagement metrics like views, likes, shares, and comments. These features help the model learn which factors contribute to a post receiving low, medium, or high engagement. The engagement level serves as the target variable for prediction.