TikTok Engagement & Sentiment Analysis

A Data-driven approach to understanding user interaction

Introduction

- TikTok is a platform driven by engagement—likes, comments, shares, and views.
- This project aims to **predict engagement levels** and explore **sentiment in captions** using machine learning.
- **Key Questions:**
    1. What factors contribute to high engagement?
    2. How does sentiment in video captions affect engagement?

LIVE

Dataset overview

- Data of 1200 TikTok users
- Key features:
  - User-related: Followers, likes, verification status
  - Content-related: Hashtags, mentions, video duration
  - Engagement metrics: views, comments, likes, share
- Target variable: Engagement score – categorised into Low, Medium, High group

$$\text{Engagement score} = \frac{\text{Likes} + \text{Shares} + \text{Comments}}{\text{Views}} * 100\%$$

LIVE

Engagement Prediction - Approach

Data preprocessing
Feature Engineering
Model selection

- Baseline: Logistic regression
- Advanced: Random forest/ HGBoost (permutation importance ranking)
- Evaluation: F1-score

Engagement Prediction - Findings

Most influential factors:

- Play and like contribute the most to engagement, with a significant impact on the permutation score.
- Hashtags: Positive effect on engagement, though their influence is relatively small.
- Video duration: Slight impact on engagement, with longer videos showing a small positive effect.

Model performance

- Logistic regression: Best-performing model with an F1-score of 0.9.
- Other models: Random Forest (0.75) and HistGradientBoosting (0.86) show lower performance compared to Logistic Regression.

Sentiment Analysis

- User bios feature common terms like ***follow, instagram, youtube, love*** showing a cross-platform engagement.
- Trending hashtags include ***viral, tiktok, for you, funny*** showing interest in discoverable, humorous content.
- Positive sentiments like ***thank, funny meme, amazing*** dominate, while negative ones include ***everything, crazy.***
- Sentiment shows limited correlation with engagement metrics, suggesting other factors drive interaction.

A. B.

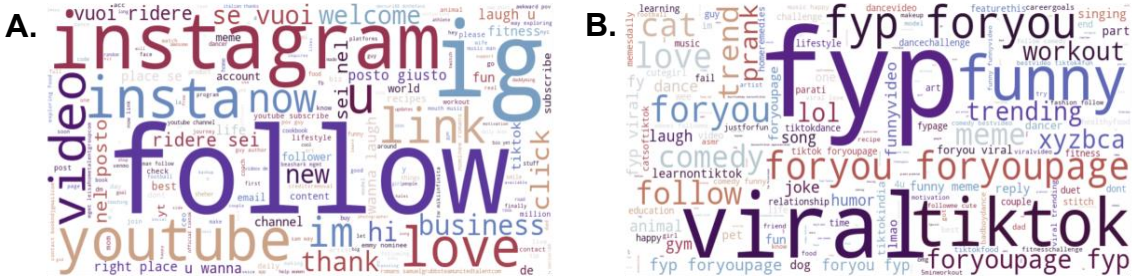

Fig 1. User bio (A) and hashtag (B) popularity breakdown

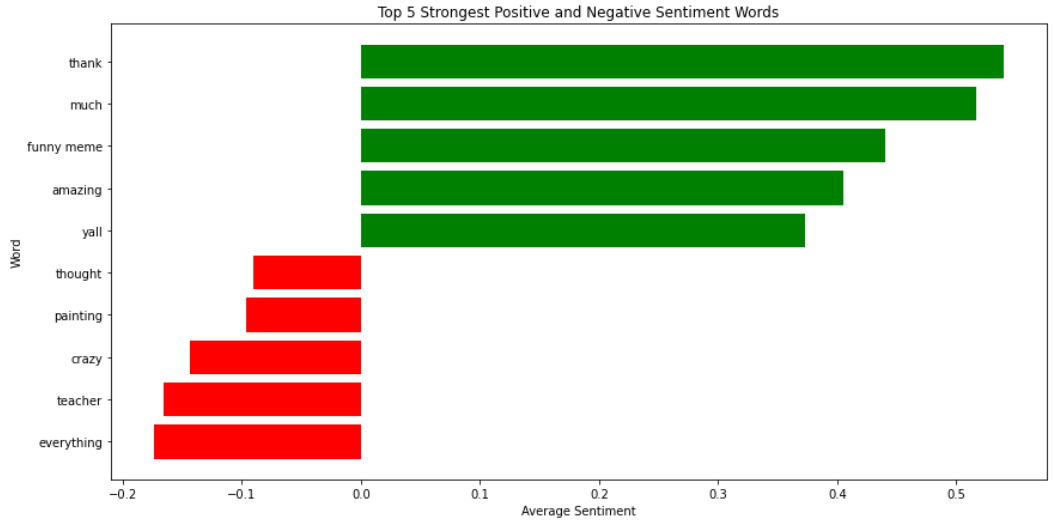

Fig 2. Top 5 positive and negative sentiment words

Conclusion & Business Impact

Insights for TikTok strategy

- Play and like are the most significant drivers of user engagement, suggesting that focusing on content that encourages interactions can have a strong positive impact.
- Short videos perform better: Encourage creators to focus on videos that are 15-30 seconds long, as this length seems to maximise plays and engagement without losing viewers' attention.
- Simple model (Logistic regression) can provide valuable insights into predicting engagement.

Next steps:

- Optimise models: Refine predictive models and integrate new features for better engagement forecasting.
- A/B testing: Implement tests to validate key engagement factors like play/like actions and hashtags.
- Real-time insights: Deploy real-time analytics and automated content recommendations to drive higher engagement.

LIVE

My name is
Mia Nguyen

mia.lhnguyen@gmail.com
LinkedIn
GitHub
0412275323

100K
5678
16,4k

@mia_nguyen
About me

Why I am a good fit?

- Leverage statistical model and machine learning to uncover insights and drive strategic decisions
- Proficient in SQL, Python R, enabling efficient data analysis and automation
- Experience in A/B testing, predictive analytics and data visualisation to optimise performance and engagement.
- Passionate about transforming data into actionable strategies to enhance system efficiency and user experience.

Thanks For Watching!