

Prediction of YouTube Video Performance Using LLM-Driven Sentiment in Cloud Environments

Nur Shafiqah Binti Mohamad Johari
22119564

Masters of Data Science
Faculty of Computer Science & Information Technology

Supervisor : Dr. Riyaz Ahamed Ariyaluran Habeeb



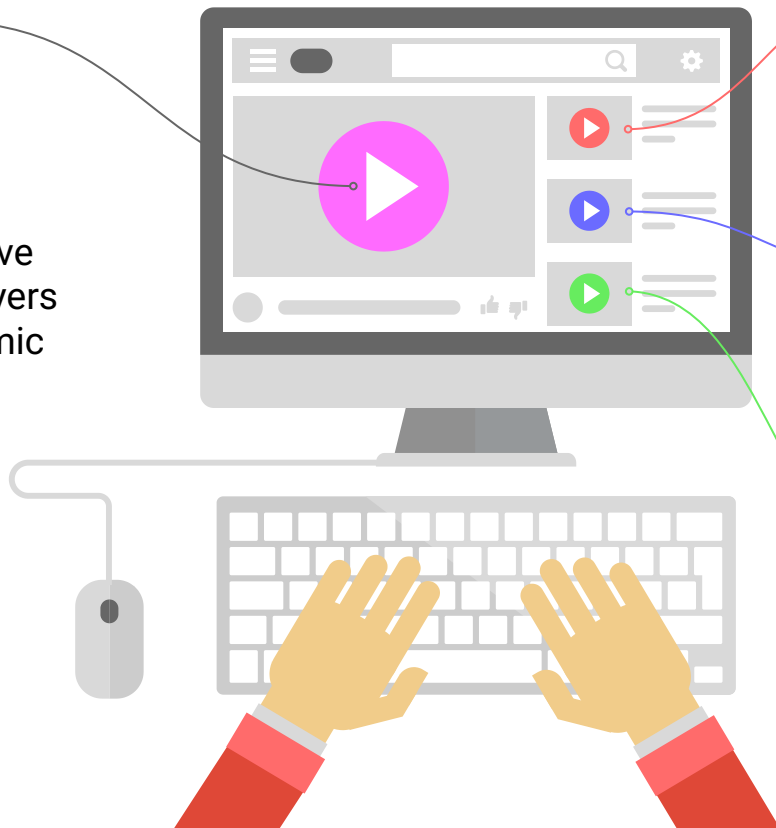
Introduction

\$104.2 B

Content Creation Significant Value

Content creators have become the key players in the digital economic value in 2022

Source : [The Creator Economy Market Size is Growing: How Brands Can Leverage It](#)



\$45 B Youtube's Impact

Youtube alone supported more than 430K jobs in the US in 2023, contributing \$45B to the nation's GDP

Source : [YouTube's Impact in the US](#)

\$36 B Youtube's Ads Revenue

Youtube's advertising revenue reached a record of \$36 billion in 2024, which is an increase of 14.6% over the previous year

Source : <https://www.businessofapps.com/data/youtube-statistics/>

Big Data Solution

Creators face growing pressure to optimize content for engagement, turning to machine learning and cloud-based strategies for support

Problem Statement

Advanced sentiment extraction using LLM as a solution remains limited

Traditional KPIs such as likes, views, comments etc. might not fully capture the dynamics that influence a video's success. A study by Zhou et al. (2015) suggests that advanced feature sentiment extraction are needed to improve a video's performance prediction.

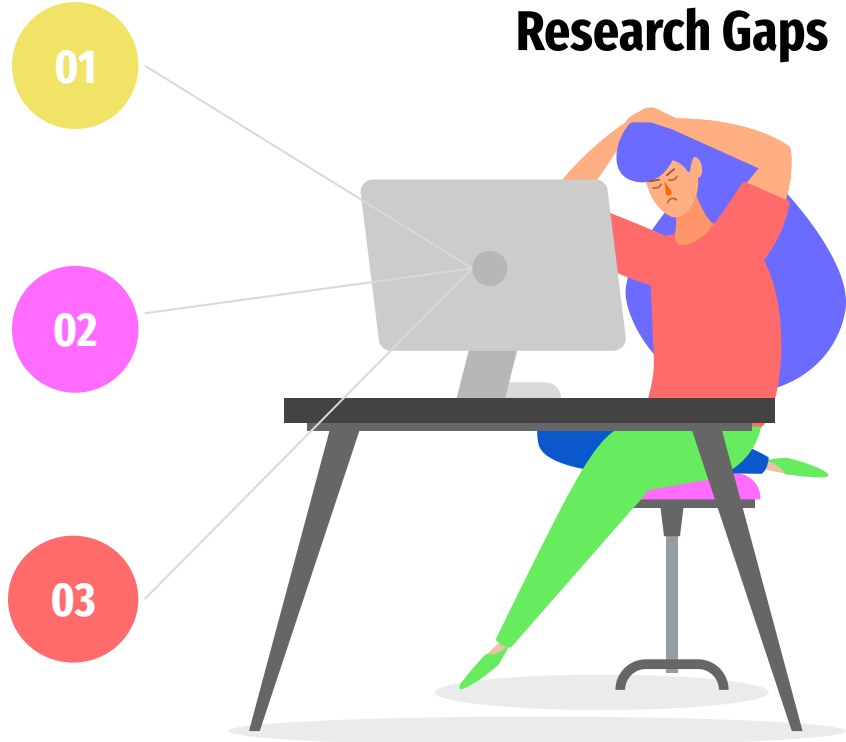
Traditional methods lacks efficiency and scalability, cloud solution is essential for Youtube data

YouTube generates over 500 hours of video per minute, making cloud-based solutions essential for handling data at this scale. However, studies around cloud solution remains limited and fulfillment of gaps are needed (Li et al., 2021)

Model evaluation and deployment pipelines in the cloud are underdeveloped for LLM-based prediction

Deploying machine learning models in cloud environments present significant challenges. A study by Sato et al. (2025) identified common challenges in MLOps adoption, emphasizing the need for integrated deployment strategies to evaluate and deploy ML models in cloud.

Research Gaps



Objectives

01

To investigate which LLMs and ML models best perform in predicting YouTube video performance

02

To develop an efficient cloud-based data pipeline for implementing LLM-enhanced prediction model

03

To evaluate & deploy Youtube video performance predictive model in the cloud

