Final Project Report: DADS5001

Project Name:
YouTube Influencer Analysis in the Tourism Sector

Objective:
Our goal is to develop a tool that aids entrepreneurs of all sizes in selecting influencers who provide value within a constrained budget. This tool extends beyond traditional metrics like views and subscriber counts, incorporating diverse data to support more strategic decisions. We extract YouTube data, including channel details and specific video content from each influencer. Using large language models (LLMs), we classify video titles related to tourism, analyze sentiment in video comments, and present our insights through Dash, an advanced data visualization tool.

Problem Statement:
Entrepreneurs looking to engage influencers for promotional video content in tourism often face the challenge of choosing the right influencer amidst a saturated market, especially within limited budget constraints. The decision-making process can be overwhelming without comprehensive data analysis tools that summarize and visualize influencer data to enhance decision-making efficiency.

Data Collection:
We employ Python scripts in Jupyter Notebook for data collection and transformation tasks. Data is fetched using the *requests* library to execute REST API calls to the YouTube API. The data is temporarily stored in *pandas* DataFrames for processing and subsequently exported to CSV files for persistence.

Data Storage:
Data is stored in CSV files.

Results of Data Analysis and Data Presentation:
Our analysis results are presented through two Dash-based dashboards:

1. **YouTuber's Channel Stats Analysis Dashboard:**
   This dashboard displays statistics for each YouTuber's channel, such as total subscribers, views, published videos, and engagement metrics (likes and views). It evaluates the success of past influencer campaigns and examines the frequency of video uploads to assess how regularly influencers engage their audience.
2. **Video Comment Sentiment Analysis Dashboard:**
   Post-influencer engagement, this dashboard provides insights into the public reception of promotional videos. It analyzes sentiments expressed in video comments to determine if the promotion was received positively or negatively. For negative feedback, the dashboard highlights the most critical comments, enabling quick identification of issues for prompt resolution. This feature aids entrepreneurs in engaging with their audience effectively by addressing concerns and ensuring continuous improvement in customer satisfaction.

Idea of using LLM models

1. **Facebook/bart-large-mnli**
   https://huggingface.co/facebook/bart-large-mnli

   Task: Text classification
   Application: We use this model to classify video titles as related or unrelated to tourism.
   This classification helps identify relevant video content for further analysis in the tourism
   sector and filters out non-tourism content for our YouTuber's Channel Stats Analysis
   Dashboard.

2. **Helsinki-NLP/opus-mt-th-en**
   https://huggingface.co/Helsinki-NLP/opus-mt-th-en

   Task: Language translation (Thai to English)
   Application: Many video titles in our dataset are in Thai, which is incompatible with the
   Facebook/bart-large-mnli model designed for English text. Therefore, we translate these titles
   from Thai to English prior to classification, ensuring accurate processing and analysis.

3. **Cardiffnlp/twitter-xlm-roberta-base-sentiment**
   https://huggingface.co/cardiffnlp/twitter-xlm-roberta-base-sentiment

   Task: Sentiment analysis
   Application: This model evaluates the sentiment of user comments on YouTube videos,
   categorizing them as positive, neutral, or negative, along with a confidence score. The results
   are crucial for our Video Comment Sentiment Analysis Dashboard, where they help visualize
   and analyze public sentiment towards promotional content, facilitating rapid response
   strategies to feedback.

Project members :

6610422007 : Chayaphon Sornthananon
6610422013 : Bunrawat Charoenyuennan

Source Code :
https://github.com/chayaphon/DADS5001/tree/main/Project

Subject :   DADS5001 Data Analytics and Data Science Tools and Programming
Semester :   02/2023
Submission : 11/05/2024
Lecuturer :   Asst. Prof. Dr. Ekarat Rattagan