Flipkart

GRID 4.0

2022 Campus Challenge

Extract Trends from social media data

Team Name: Trinetra

Institute Name: VNR Vignana Jyothi Institute of Engineering & Technology, Hyderabad

Team members details

Team Name			
	Trinetra		
Institute Name			
	VNR Vignana Jyothi Institute of Engineering & Technology		
Team Members >	1 (Leader)	2	3
Name	Gautham Mallipeddi	Srinivasa Deepesh Kommineni	Tapan Ganesh Naram
Batch	2020-2024	2020-2024	2020-2024

Glossary

- Sentiment analysis: The process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular topic, product, etc. is positive, negative, or neutral.

- Web scraping: Web scraping is an automatic method to obtain large amounts of data from websites. Most of this data is unstructured data in an HTML format which is then converted into structured data in a spreadsheet or a database so that it can be used in various applications.

- Engagement Count: The total number of posts related to a particular Hashtag.

- Rising: The percentage increase in search of the particular product.

Use-cases

The only actor in this application is the user who wants to find trending products in a category.

- P0: The most important task of the user is to specify the category in which he wants to find trending products. Examples are: "Men shirts","Women pants"
- P1: After completing the execution, the user will be provided with the product details (product name, product category, Product image, Links from which the trend is derived and a graph that shows the trend graph of the product) sorted in the decreasing order of their trend score. The user can then use these trends in further activities.

These trends can further be used to:

- Suggest a particular product on Flipkart.
- Apply discounts in effectively.

Solution statement/ Proposed approach

- Whenever a product is trending on any social media platform, it will be one of the most searched topics on Google. Therefore, we first extract the most searched terms/keywords in a particular category from Google trends.
- We search the social media platform - YouTube for the trending topics.
- Comments are scraped from the videos of trending topics.
- Since YouTube is filled with spam comments, we remove the spam comments.
- Comments on a YouTube video can be divided into two types:
    1. Comments about the YouTuber- Example: "Your videos are great."
    2. Comments regarding the product- Example: "A great product."
- Using Machine Learning, we obtain the comments that are related to the product only.
- Sentiment analysis is performed on the comments to verify if the product is trending in a positive manner or not.
- If the number of positive comments for a video are greater than 75%, we then calculate the engagement count for a particular hashtag on Instagram using it's respective post count.
- The trend score for a particular keyword can be calculated by the formula:

$$Trend\ score = \frac{(Engagement\ count)/3 + Percentage\ of\ positive\ comments\ on\ YouTube + Ri\sin g)}{3}$$

- Finally,product details are displayed with their trend score scored in descending order. The following product details are displayed: Product name, Category mapping, Image link of the product, YouTube links from which the trends are derived, trend score, trend graph.

Limitations

- Though web scraping is a tool that can be used to automate the collection of data, if the number of trending products are very large it will be a time consuming process to get the comments of all videos that are based on a particular topic.

- Since web scraping depends on the structure of the data in the website, if a website changes its structure, the code will need to be changed.

- Web scraping of Instagram won't give us detailed insights like what instagram provides for creator mode.

- Since web scraping is a time consuming process, all videos for a particular product couldn't be examined properly. Hence, there maybe some inaccurate results.

Future Scope

- Using image-processing, we can examine posts of various influencers and determine the trends related to the required fields using the detailed study of engagement on the posts.

- For this project, we have manually entered search terms like 'Mens shirt', 'Womens shirt', etc to analyse the trends. In the future, we can develop a Machine Learning model that automatically determines the search words.