REAL TIME INSIGHTS FROM SOCIAL MEDIA

Mini Project

BACHELOR OF TECHNOLOGY

(Computer Science and Engineering)

Submitted by:

Sawayam Kumar (2820061)

Ashish Godia (2820075)

Paras Gakhar (2820321)

Prabhjot Singh (2820322)

Srishti Sharma (2820323)

Submitted to:

Dr. B.K Verma

Professor

Department: ET

Affiliated to

KURUKSHETRA UNIVERSITY KURUKSHETRA, INDIA

Acknowledgement

We are highly grateful to the **Dr. B.K Verma**, *Panipat Institute of Engineering and Technology, Samalkha,* for providing this opportunity to carry out the mini project work **Real Time Insights from Social Media**.

The constant guidance and encouragement received from Dr. B.K Verma has been of great help in carrying out the project work and is acknowledged with reverential thanks. We would like to express a deep sense of gratitude and thanks profusely to him without his wise counsel and able guidance, it would have been impossible to complete the project in this manner. We express gratitude to other faculty members of computer science and engineering department of PIET for their intellectual support throughout the course of this work.

Finally, we are indebted to all whosoever have contributed in this report work.

Project Report

Real Time Insights from Social Media

INTRODUCTION

The purpose of this project report is to provide an overview of the project "Real-Time Insights from Social Media." The project aimed to develop a system that leverages social media data to extract valuable insights in real-time. This report outlines the objectives, methodology, key findings, and future recommendations of the project.

Fear of missing out, curiosity, self-esteem, speed, it's like social media has changed our basic human needs. These baits are keeping us hooked and engaged. And Twitter is a master at this game.

Twitter has this huge influence on the world because of the type of its users as Elon Musk's tweets keep Wall Street on its toes; Trump's tweets have the potential of starting wars —Data from Twitter-storms is available in near real-time. This means we can learn about the big waves of thoughts and moods around the world as they arise. So of course, we are not going to miss the chance to analyze this treasure trove.

Real Time Analytics:

Real-time analytics is the discipline that applies logic and mathematics to data to provide insights for making better decisions quickly. For some use cases, real time simply means the analytics is completed within a few seconds or minutes after the arrival of new data. On-demand real-time analytics waits for users or systems to request a query and then delivers the analytic results. Continuous real-time analytics is more proactive and alerts users or triggers responses as events happen.

Social Media Insights:

Social media insights are defined as consumer insights from social media channels only. Generally, a consumer insight is newly found valuable information that helps businesses gain a deeper understanding of their audience.

Tracking and analysing a brand’s social interactions online provides business owners with customer knowledge, also known as social insights. Companies need to analyse human behaviours in terms of consumer habits, interests, and needs to create effective digital marketing campaigns.

We can’t consider all data points as insights. In order for that to happen, this piece of information point needs to be new, unforeseen, and relevant to your business.

OBJECTIVES

Objectives of project are as follows:

- Develop a system to monitor social media platforms in real-time.
- Extract relevant data from social media posts, including text, images, and metadata.
- Analyze the collected data to identify trends, sentiments, and key influencers.
- Provide actionable insights to stakeholders based on the analysis.

Methodology

In this Project, you have use pre-downloaded datasets to understand the nuts and bolts of Twitter Data.

In particular, we have done a thorough analysis of a hot-trend. We followed below given steps for getting insights and achieve our objectives from our downloaded dataset.

- Local and global thought patterns
- Prettifying the output
- Finding common trends
- Exploring the hot trend
- Digging deeper
- Frequency analysis
- Activity around the trend
- Visualize the data
- Analysing used languages
- Final Conclusion

1. Local and Global Thought Pattern:

While we might not be twitter fans, we have to admit that it has huge influence on the world. Twitter data is not only fold in terms of insights, but Twitter-storms are available for analysis in near real-time. This means we can learn about the big waves of thoughts and moods around the world as they arise. As any place filled with riches, Twitter has *security guards* blocking us from laying our hands on the data right away. Some authentication steps are needed to call their APIs for data collection. Our data is ready for usage in the datasets folder.

2. Prettifying the Output:

Our data was hard to read and not understandable. So we need to prettify our output. It refers to improving the appearance or formatting of the text or information that is being displayed or presented. It involves making the output more visually appealing, organized, and easier to read for human users.

After this step our data become more aesthetically pleasing and user-friendly.

3. Finding common trends:

From the above step, we got following observations :

- We have an array of trend objects having: the name of the trending topic, the query parameter that can be used to search for the topic on Twitter-Search, the search URL and the volume of tweets for the last 24 hours, if available. (The trends get updated every 5 mins.)
- At query time ***#BeratKandili, #GoodFriday*** and ***#WeLoveTheEarth*** were trending world wide.
- *"tweet volume"* tell us that *#WeLoveTheEarth* was the most popular among the three.
- Results are not sorted by *"tweet_volume"*.
- There are some trends which are unique to the US.

4. Exploring the hot trend:

From the intersection in the above step output we can see that, out of the two sets of trends (each of size 50), we have 11 overlapping topics. In particular, there is one common trend that sounds very interesting: ***#WeLoveTheEarth*** — so good to see that *Twitteratis* are unanimously talking about loving Mother Earth!

If we query Twitter's search API with this hashtag as query parameter, we get back actual tweets related to it. We have the response from the search API stored in the datasets folder as *'WeLoveTheEarth.json'*.

5. Digging deeper:

Printing the first two tweet items makes us realize that there's a lot more to a tweet than what we normally think of as a tweet — there is a lot more than just a short text. But instead of getting overwhelmed by all the information in a tweet object we focus on a few interesting fields and try to find some hidden insights there.

6. Frequency analysis:

Just from the first few results of the last extraction, we deduce that:

- We are talking about a song about loving the Earth.
- A lot of big artists are the forces behind this Twitter wave, especially Lil Dicky.
- Ed Sheeran was some cute koala in the song — "EdSheeranTheKoala" hashtag!

Observing the first 10 items of the interesting fields gave us a sense of the data. We then take a closer look by doing a simple, but very useful, exercise — computing frequency distributions. Starting simple with frequencies is generally a good approach.It helps in getting ideas about how to proceed further.

7. Activity around the trend:

Based on the last frequency distributions we further build-up on our deductions:

- We can more safely say that this was a music video about Earth (hashtag 'EarthMusicVideo') by Lil Dicky.
- DiCaprio is not a music artist, but he was involved as well (Leo is an environmentalist so not a surprise to see his name pop up here).
- We can also say that the video was released on a Friday, very likely on April 19th.

We have been able to extract so many insights.Then we further analyze the data to find patterns in the activity around the tweets — did all retweets occur around a particular tweet?

If a tweet has been retweeted, the *'retweeted status'* field gives many interesting details about the original tweet itself and its author.

We measure a tweet's popularity by analyzing the ***retweetcount*** and ***favoritecount*** fields. But we also extract the number of followers of the tweeter — we have a lot of celebs in the picture, so we can tell if their advocating for **#WeLoveTheEarth** influenced a significant proportion of their followers.

The retweet_count gives us the total number of times the original tweet was retweeted. It should be the same in both the original tweet and all the next retweets. Tinkering around with some sample tweets and the official documentation are the way to get your head around the many fields.

8. Visualize the Data:

In this step we manipulate the data further and visualize it in a better and richer way. So that it can be easy to understand. We created a table that include tweet ,username , number of followers, number of retweets. And use that table for further analysis.

9. Analysing used languages:

Our table from above step tells us that:

- Lil Dicky's followers reacted the most — 42.4% of his followers liked his first tweet.
- Even if celebrities like Katy Perry and Ellen have a huuge Twitter following, their followers hardly reacted, e.g., only 0.0098% of Katy's followers liked her tweet.
- While Leo got the most likes and retweets in terms of counts, his first tweet was only liked by 2.19% of his followers.

The large differences in reactions could be explained by the fact that this was Lil Dicky's music video. Leo still got more traction than Katy or Ellen because he played some major role in this initiative.

10. Conclusions from language distribution:

- Most of the tweets were in English.
- Polish, Italian and Spanish were the next runner-ups.
- There were a lot of tweets with a language alien to Twitter (lang = 'und').

11. Final Conclusions:

- A music video about loving the mother Earth was made by Lil Dicky
- Leo DiCaprio, being an encironmentalist was also inolved in it
- Most of the related tweets were in English, so, the language of the video was English.
- The video was published on April 19th, which was Good Friday

Future Recommendations Based on the project findings, the following recommendations are made for future improvements:

a) Real-Time Monitoring: Enhance the system to provide continuous real-time monitoring of social media platforms, ensuring that insights are up-to-date and relevant.

b) Advanced Sentiment Analysis: Implement more sophisticated sentiment analysis techniques, such as emotion detection and sarcasm identification, to improve the accuracy of sentiment analysis.

c) Network Analysis: Explore network analysis techniques to understand the relationships between social media users and identify influential communities or clusters within the network.

d) Integration with CRM Systems: Integrate the insights from social media into Customer Relationship Management (CRM) systems to provide a comprehensive view of customer behaviour and preferences.

e) Predictive Analytics: Develop predictive analytics models based on historical social media data to anticipate trends, customer behaviour, and potential crises.