

**Experiment 8**

**Date of Performance : 20/04/23** **Date of Submission: 27/04/23**

**SAP ID: 60004190057** **Name : Junaid Girkar**

**Div: A** **Batch : A2**

**Aim of Experiment**

Apply analytics to social media activity (Using FB, Twitter, Instagram or any social media dataset)

**Theory / Algorithm / Conceptual Description**

**Social Media Analytics**

Social media analytics is the ability to gather and find meaning in data gathered from social channels to support business decisions — and measure the performance of actions based on those decisions through social media.

Social media analytics is broader than metrics such as likes, follows, retweets, previews, clicks, and impressions gathered from individual channels. It also differs from reporting offered by services that support marketing campaigns such as LinkedIn or Google Analytics.

Social media analytics uses specifically designed software platforms that work similarly to web search tools. Data about keywords or topics is retrieved through search queries or web ‘crawlers’ that span channels. Fragments of text are returned, loaded into a database, categorized and analysed to derive meaningful insights.

Social media analytics includes the concept of social listening. Listening is monitoring social channels for problems and opportunities. Social media analytics tools typically incorporate listening into more comprehensive reporting that involves listening and performance analysis.

Social media analytics helps companies address these experiences and use them to:

1. Spot trends related to offerings and brands.
2. Understand conversations — what is being said and how it is being received.
3. Derive customer sentiment towards products and services.
4. Gauge response to social media and other communications.
5. Identify high-value features for a product or service.
6. Uncover what competitors are saying and its effectiveness.
7. Map how third-party partners and channels may affect performance.

**Key capabilities of effective social media analytics**

*Natural language processing* and machine learning technologies identify entities and relationships in unstructured data — information not pre-formatted to work with data analytics. Virtually all social media content is unstructured. These technologies are critical to deriving meaningful insights.

*Segmentation* is a fundamental need in social media analytics. It categorizes social media participants by geography, age, gender, marital status, parental status and other demographics. It can help identify influencers in those categories. Messages, initiatives and responses can be better tuned and targeted by understanding who is interacting on key topics.

*Behavior analysis* is used to understand the concerns of social media participants by assigning behavioral types such as user, recommender, prospective user and detractor. Understanding these roles helps develop targeted messages and responses to meet, change or deflect their perceptions.

*Sentiment analysis* measures the tone and intent of social media comments. It typically involves natural language processing technologies to help understand entities and relationships to reveal positive, negative, neutral or ambivalent attributes.

*Share of voice* analyzes prevalence and intensity in conversations regarding brand, products, services, reputation and more. It helps determine key issues and important topics. It also helps classify discussions as positive, negative, neutral or ambivalent.

*Clustering analysis* can uncover hidden conversations and unexpected insights. It makes associations between keywords or phrases that appear together frequently and derives new topics, issues and opportunities. The people that make baking soda, for example, discovered new uses and opportunities using clustering analysis.

*Dashboards and visualization charts*, graphs, tables and other presentation tools summarize and share social media analytics findings — a critical capability for communicating and acting on what has been learned. They also enable users to grasp meaning and insights more quickly and look deeper into specific findings without advanced technical skills.

**FIFA Tweets Dataset ⚽🏆**

Football is one of the most loved sports worldwide. The FIFA World Cup, a global football sporting event that takes place every four years, is in Qatar this year. This dataset contains 30,000 tweets from the first day of the FIFA World Cup 2022.

*Data Collection*

The dataset was created using the Snscrape and the cardiffnlp/twitter-roberta-base-sentiment-latest model in Hugging Face Hub.

*Data Preprocessing*

The dataset includes tweets in English containing the hashtag #WorldCup2022. For data preprocessing, we used a tokenizer for the cardiffnlp/twitter-roberta-base-sentiment-latest model and the following function:

def preprocess(text):

    new\_text = []

    for t in text.split(" "):

        t = '@user' if t.startswith('@') and len(t) > 1 else t

        t = 'http' if t.startswith('http') else t

        new\_text.append(t)

    return " ".join(new\_text)

*Data Storage*

The collected tweets have been consolidated into a single dataset & shared as a Comma Separated Values file, "fifa\_world\_cup\_2022\_tweets.csv".

*Content*

The dataset contains as following columns:

1. Date Created
2. Number of Likes
3. Source of Tweet
4. Tweet
5. Sentiment

*License*

License: CC0 1.0 Universal (CC0 1.0) - Public Domain Dedication

No Copyright - You can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission.

|  |
| --- |
| Text, table  Description automatically generated with medium confidenceGraphical user interface, table  Description automatically generated with medium confidenceChart, histogram  Description automatically generatedText  Description automatically generatedChart, bar chart  Description automatically generatedChart, bar chart  Description automatically generatedText, letter  Description automatically generatedText, letter  Description automatically generatedChart  Description automatically generated with low confidenceChart, box and whisker chart  Description automatically generated |

**Conclusion**

Sentiment analysis is a powerful tool that allows generalization of user feedback for various demographics with multiple levels of granularity. The most common implementation over social media data allows companies and events to evaluate user feedback and consumer sentiment towards products and policies without conducting explicit surveys. It also removes bias that is inherent with organizational surveying by directly reading user interactions in a social web of communication.