

VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY
An Autonomous Institute Affiliated to University of Mumbai
Department of Computer Engineering



Project Report on

Twitter User Analysis

In partial fulfilment of the Fourth Year, Bachelor of Engineering (B.E.) Degree in Computer
Engineering at the University of Mumbai
Academic Year 2023-24

Submitted by

Tamanna Bathija - D17C - 06
Harsha Chelani - D17B - 14
Khusboo Kimtani - D17C - 26
Shreya Kukreja - D17A - 33
Piyush Waghmare - D17A - 71

Project Mentor

Prof. Sujata Khandaskar
(2023-24)

VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY
An Autonomous Institute Affiliated to University of Mumbai
Department of Computer Engineering



Certificate

This is to certify that of **Tamanna Bathija (D17C, 06), Harsha Chelani (D17B, 14), Khusboo Kimtani (D17C, 26), Shreya Kukreja (D17A, 33), Piyush Waghmare (D17A, 71)**

Fourth Year Computer Engineering studying under the University of Mumbai have satisfactorily completed the project on “**Twitter User Analysis** ” as a part of their coursework of PROJECT-II for Semester-VIII under the guidance of their mentor **Prof. Sujata Khandaskar** in the year 2023-24 .

This project report entitled **Twitter User Analysis** by **Tamanna Bathija, Harsha Chelani , Khusboo Kimtani, Shreya Kukreja , Piyush Waghmare** is approved for the degree of **B.E. Computer Engineering.**

Programme Outcomes	Grade
PO1,PO2,PO3,PO4,PO5,PO6,PO7, PO8, PO9, PO10, PO11, PO12 PSO1, PSO2	

Date:

Project Guide:

Project Report Approval For B. E (Computer Engineering)

This project report entitled Twitter User Analysis by Tamanna Bathija, Harsha Chelani , Khusboo Kimtani, Shreya Kukreja , Piyush Waghmare is approved for the degree of B.E. Computer Engineering.

Internal Examiner

External Examiner

Head of the Department

Principal

Date:

Place: Mumbai

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Tamanna Bathija (D17C-6)

Harsha Chelani (D17B-14)

Khusboo Kimtani (D17C-26)

Shreya Kukreja (D17A-33)

Piyush Waghmare (D17A-71)

Date:

ACKNOWLEDGEMENT

We are thankful to our college Vivekanand Education Society's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to Assistant Professor **Prof. Sujata Khandaskar** (Project Guide) for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to Head of the Computer Department **Dr. (Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J. M. Nair** , for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is a great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement several times.

Computer Engineering Department
COURSE OUTCOMES FOR B.E PROJECT

Learners will be to,

Course Outcome	Description of the Course Outcome
CO 1	Able to apply the relevant engineering concepts, knowledge and skills towards the project.
CO2	Able to identify, formulate and interpret the various relevant research papers and to determine the problem.
CO 3	Able to apply the engineering concepts towards designing solutions for the problem.
CO 4	Able to interpret the data and datasets to be utilised.
CO 5	Able to create, select and apply appropriate technologies, techniques, resources and tools for the project.
CO 6	Able to apply ethical, professional policies and principles towards societal, environmental, safety and cultural benefit.
CO 7	Able to function effectively as an individual, and as a member of a team, allocating roles with clear lines of responsibility and accountability.
CO 8	Able to write effective reports, design documents and make effective presentations.
CO 9	Able to apply engineering and management principles to the project as a team member.
CO 10	Able to apply the project domain knowledge to sharpen one's competency.
CO 11	Able to develop a professional, presentational, balanced and structured approach towards project development.
CO 12	Able to adopt skills, languages, environment and platforms for creating innovative solutions for the project.

Index

Chapter No.	Title	Page No.
1	Introduction	11
1.1	Introduction to the project	11
1.2	Motivation for the project	11
1.3	Problem Definition	12
1.4	Existing Systems	12
1.5	Lacuna of the Existing Systems	12
1.6	Relevance of the Project	12
2	Literature Survey	13
A	Brief overview of Literature Survey	13
2.1	Research Papers a. Abstract of the research paper b. Inference drawn from the paper	13
2.2	Patent search	19
2.3	Inference Drawn	19
3.	Requirement Gathering for the proposed System	20
3.1	Introduction to Requirement Gathering	20
3.2	Functional Requirements	20
3.3	Non-Functional Requirements	21
3.4	Hardware, Software, Technology and tools utilised	21
3.5	Constraints	23
4.	Proposed Design	23
4.1	Block diagram of the system	23
4.2	Modular design of the system	23
4.3	Project Scheduling & Tracking using Timeline / Gantt Chart	24

5.	Implementation of the Proposed System	25
5.1	Methodology employed for development	25
5.2	Datasets source and utilisation	26
6.	Testing of the Proposed System	27
6.1	Introduction to testing	27
6.2	Types of tests Considered	27
6.3	Various test case scenarios considered	28
6.4	Inference drawn from the test cases	29
7.	Results and Discussions	31
7.1	Screenshots of User Interface (UI) for the respective module	32
8.	Conclusion	37
8.1	Limitations	37
8.2	Conclusion	37
8.3	Future Scope	37
	References	38
	Appendix	39
1	Paper I	39
a	Paper I	39
b	Plagiarism Report of Paper I	45
c	Project review sheet	45

List of Figures:

Fig no.	Heading	Page no.
4.1	Block Diagram	23
4.2	Modular Diagram	23
4.3	Gantt Chart	24
7.	Results and Discussions	31
7.1	Screenshot for home page	32
7.2	Profilysis	32
7.3	Screenshot for Profile Information	34
7.4	Screenshot for Start of Analysis	34
7.5	Screenshot for Tweets Analysis	34
7.6	Screenshot for Activity Time	35
7.7	Screenshot for Most Tweeted Topics	35
7.8	Screenshot for Most Used Hashtags	35
7.9	Screenshot for Community Mentions	36
a	Paper I	45
b	Plagiarism Report of Paper I	45
c	Project review sheet	46

Abstract

User analysis plays a pivotal role in shaping successful strategies across various domains, ranging from marketing to product development and beyond. By meticulously scrutinizing user behavior, preferences, and demographics, organizations can gain profound insights into their target audience. Understanding users' needs, desires, and pain points enables businesses to tailor their offerings accordingly, enhancing customer satisfaction and loyalty. Moreover, user analysis serves as a compass for innovation, guiding the development of products and services that resonate deeply with consumers. By leveraging user data, organizations can optimize their marketing efforts, crafting personalized campaigns that resonate with specific audience segments. Ultimately, user analysis serves as a cornerstone for informed decision-making, empowering organizations to stay agile, responsive, and ahead of the curve in an ever-evolving landscape.

The Twitter Profile Analyzer is a versatile tool meticulously crafted to empower organizations in delving deeper into the Twitter profiles of pertinent individuals or entities. By harnessing data scraped from Twitter, including tweets, mentions, and hashtags, this system conducts comprehensive analyses spanning mention extraction, subject analysis, hashtag identification, and mention-based community recognition. Through these sophisticated analytics, organizations gain invaluable insights into the interactions, interests, and influence wielded by profile owners on Twitter.

The generated reports furnish organizations with actionable intelligence, facilitating well-informed decisions regarding marketing strategies, partnership opportunities, and engagement tactics. Tailored to the unique features and audience dynamics of each profile, these insights enable organizations to fine-tune their approaches, ensuring maximum efficacy in reaching their objectives. With the Twitter Profile Analyzer at their disposal, organizations can navigate the intricate landscape of Twitter with confidence, leveraging data-driven strategies to achieve their goals and enhance their online presence.

Chapter 1: Introduction

1.1. Introduction:

Twitter(Now X),a global social media giant, has rapidly become an essential platform for real-time communication, sharing information, and fostering online communities. It has not only revolutionized the way we connect and interact but has also emerged as a treasure trove of valuable data for researchers, analysts, and businesses. In this section, we provide an in-depth introduction to our study on Twitter network analysis, shedding light on the significance of this research and the comprehensive exploration of the Twitter ecosystem.

Twitter's Significance:

Twitter stands as one of the most influential social media platforms in the digital landscape. Its unique format of short, concise messages, or tweets, coupled with real-time interactions, has made it a vital source for observing the flow of information and the dynamics of user engagement. This study delves into the intricacies of Twitter, aiming to uncover the mechanisms underpinning the platform's functioning.

1.2. Motivation:

The impetus behind developing the Twitter Profile Analyzer is the increasing significance of social media platforms, especially Twitter, as potent instruments for marketing and communication in the current digital environment. Given that millions of people use Twitter every day, it is now crucial for both individuals and businesses to comprehend the subtleties of each profile.

Organisations can gain important insights into the preferences, interests, and behaviours of their audience by analysing Twitter profiles. Organisations can better connect with their target audience by identifying chances for collaboration, customising their marketing strategies, and optimising their engagement tactics by taking these factors into consideration.

Maintaining and expanding one's online presence requires a grasp of the performance of one's Twitter profile and audience interaction, especially for influencers and content providers. By offering insightful feedback on content, audience interactions, and overall influence, the Analyzer assists users in improving their online presence and strategy.

The main driving force behind the development of the Twitter Profile Analyzer is its capacity to provide businesses and individuals with useful information that they can use to improve their social media strategies, make well-informed decisions, and ultimately increase their influence and success online.

1.3. Problem Definition:

The problem definition of our research stems from the pronounced knowledge gap in understanding user behavior, influential voices, and community dynamics within the Twitter network, despite its immense potential as a real-time social data source. This gap obstructs informed decision-making across a spectrum

of domains, including marketing, public opinion research, and community building. Our research endeavors to address these challenges by leveraging advanced data mining techniques, including network analysis and sentiment analysis, with a particular focus on identifying key influencers, uncovering hidden relationships, and mapping the intricate community structures within the Twitter ecosystem. In doing so, we aspire to contribute to a comprehensive understanding of online dynamics, empowering stakeholders to make data-driven decisions in their respective areas of interest.

1.4. Existing Systems:

Existing systems have paid little attention to the skills and jobs as a combined feature. LinkedIn has become the world's largest online professional networking site with more than 562 million members in over 200 countries and territories worldwide but it only connects the recruiter to the people [12]. Coursera is a platform that provides various courses but does not connect the recruiter to the people [13]. Thus a combined solution is required. ALISON (Advance Learning Interactive Systems Online) is an Irish online education platform for higher education, teaching workplace skills and exploring new interests that provides certificate courses and accredited diploma courses.

1.5. Lacuna of the Existing System:

Here's a comparative analysis highlighting the gaps or lacunas in prevalent approaches compared to this project:

- **User-Friendly Interface:** Many existing systems lack a user-friendly interface, making it challenging for non-technical users to navigate and utilize the platform effectively. This lack of user-friendliness may hinder user adoption and limit the system's overall impact.
- **Transparency in Methodology:** The project meticulously documents each step, from getting access to Twitter data to applying different algorithms. This transparency is often lacking in many proprietary systems.

1.6. Relevance of the Project:

The relevance of our project extends across a broad spectrum of domains, making it a crucial endeavor. By delving into the network analysis of Twitter, we aim to provide valuable insights into information dissemination, user influence, and community structures. This project has significance for marketers who can optimize their strategies by understanding user behavior and content virality. It is equally vital for political analysts and social researchers, as it contributes to a deeper comprehension of public opinion formation and information spread. Additionally, understanding community structures within Twitter benefits those interested in fostering online communities and collaborations. Academic researchers also stand to gain, as our study enriches the knowledge base in social network analysis and data mining. In essence, our research is a quest to unlock the intricate dynamics of the Twitter network, empowering stakeholders in diverse domains to make informed, data-driven decisions.

Chapter 2: Literature Survey

A. Overview of literature survey:

The papers discussed here focus on various job recommendation skills as well as different skills and job portals. These papers are studied to understand how the skills and jobs are dependent on each other. The studies examine how to create the job recommendation system more efficient. Overall, the papers highlight the importance of taking a comprehensive approach to address how these above factors can be used and enhanced for the development of a complete system which can provide both the courses as well as jobs.

2.1. Research Papers :

1. S. Cao and J. He

Label Propagation Algorithm based on Similarity and Weighted

IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 2020

- a) **Abstract:** The research on complex networks has been a hotspot in many fields, among which community structure is a common feature in complex networks. The whole network is composed of many communities, and the connections between nodes in the same community are very tight, while the connections between communities are relatively sparse. In this context, this paper focuses on a SWLPA (Similarity and Weighted Based Label Propagation Algorithm) label propagation algorithm based on similarity and neighbor node weight. This paper first calculates the probability of label propagation according to the similarity between nodes and the weight of neighbor nodes, then filters labels according to the influence of each node and carries out label propagation, and finally gets the result of community division. The experimental result shows that the method effectively improved the accuracy of community division results.
- b) **Inference:** The SWLPA algorithm enhances community detection in complex networks by considering node similarity and neighbor node weights for label propagation. By probabilistically propagating labels and filtering based on node influence, it achieves more accurate community divisions, as demonstrated in experimental results, advancing community structure analysis in complex networks.

2. N. Duhan, A. K. Sharma and K. K. Bhatia

Page Ranking Algorithms: A Survey

IEEE International Advance Computing Conference, Patiala, India, 2019

- a) **Abstract:** Web mining is an active research area in present scenario. Web Mining is defined as the application of data mining techniques on the World Wide Web to find hidden information. This hidden information i. e. knowledge could be contained in content of Web pages or in link structure of WWW or in Web server logs. Based upon the type of knowledge, Web mining is usually divided in three categories: Web content mining, Web structure mining and Web usage mining. An application of Web mining can be seen in the case of search engines. Most of the search engines are ranking their search results in response to users' queries to make their search navigation easier. In this paper, a survey of page ranking algorithms and comparison of some important algorithms in context of performance has been carried out.
- b) **Inference:** The paper surveys page ranking algorithms in the context of web mining, focusing on their performance. It highlights the importance of web mining in extracting hidden knowledge from the vast amount of data on the World Wide Web. It categorizes web mining into web content, structure, and usage mining, emphasizing their significance in various applications such as search engine optimization.

3. Austin P. Logan¹ · Phillip M. LaCasse² · Brian J. Lunday²

Social media analysis of Twitter interactions

in US government publications ,2020

- a) **Abstract:** To effectively utilize social media for social influence, a deep understanding of the target audience is essential. This study used Social Network Analysis to model user interactions on Twitter, identifying discussion topics with Latent Dirichlet Allocation. This led to a multilayer network, revealing influential users and connected groups. PageRank was the most effective measure for individual influence, while community identification methods varied in results. This four-step process is easily replicable and cost-effective for future research.

- b) Inference:** This study demonstrates a systematic approach to leveraging social media for social influence by integrating Social Network Analysis (SNA) with Latent Dirichlet Allocation (LDA) on Twitter data. Through this method, the study successfully identifies discussion topics, influential users, and connected groups within the network. PageRank emerges as a potent measure for individual influence assessment, while community identification methods exhibit variability. The study's replicable four-step process offers a cost-effective framework for future research endeavors in this domain.

4. Quoc Dinh Truong , Quoc Bao Truong , and Taoufiq Dkaki

Graph Methods for Social Network Analysis

**Institut de Recherche en Informatique de Toulouse, Université de Toulouse,
Toulouse, France**

- a) Abstract:** A social network is a complex structure connecting social actors through various relationships, typically depicted as weighted, labeled, and directed graphs. Social Network Analysis (SNA) involves techniques to measure influence and visualize network structures. SNA is widely used in diverse fields, examining problem-solving, organizational interactions, and individual roles. In this paper, we focus on two methods: graph visualization and network analysis through vertex comparison.
- b) Inference:** The paper explores Social Network Analysis (SNA), a method crucial for understanding complex social structures through graph visualization and vertex comparison. It highlights SNA's broad applicability across multiple domains, facilitating insights into problem-solving, organizational dynamics, and individual roles within networks. This emphasis suggests a comprehensive exploration of network structures and dynamics to uncover valuable insights.

5. Omar Y. Adwan, Marwan Al-Tawil, Ammar M. Huneiti, Rawan A. Shahin, Abeer A. Abu Zayed, Razan H. Al-Dibsi “Twitter Sentiment Analysis Approaches : A Survey” iJET, Vol. 15, No. 15, August 2020

- a) **Abstract:** Twitter produces humungous of data in a daily basis. Hadoop platform provide best solution to analyze and process large data sets. TSA started using semantic technologies to generate ontologies representing concepts of a domain. Visualization-based applications use visual or graphic structures, such as images, maps or graphs (individually and in combinations) to represent associations between tweets and or users
- b) **Inference:** The statement implies that Twitter generates massive amounts of data daily, necessitating robust tools like the Hadoop platform for efficient analysis and processing of large datasets. Furthermore, the Transportation Security Administration (TSA) has adopted semantic technologies to create ontologies that represent domain concepts, likely to enhance data understanding and analysis. Additionally, visualization-based applications are employed to represent associations between tweets and users using graphical structures like images, maps, or graphs, which can offer insights into the data in an accessible manner. Overall, these approaches reflect a concerted effort to harness technology for effective data management and analysis in the context of Twitter data.

6. Monica Bhakuli “Evolution and Evaluation: Sarcasm Analysis for Twitter Data Using Sentiment” in Hindawi Journals - 2021

- a) **Abstract:** Sentiment analysis is a challenging interdisciplinary task which includes natural language processing, web mining and machine learning. It has different levels of granularity. Analyses was done on this labeled datasets using various feature extraction technique. The different machine learning techniques trains the dataset with feature vectors and then the semantic analysis offers a large set of synonyms and similarity which provides the polarity of the content.

- b) Inference:** The statement underscores the complexity of sentiment analysis, which encompasses multiple disciplines such as natural language processing, web mining, and machine learning. The process involves analyzing labeled datasets using diverse feature extraction techniques and machine learning algorithms to train on feature vectors. Additionally, semantic analysis aids in determining the polarity of content by providing synonyms and assessing similarity. This interdisciplinary approach highlights the depth and breadth required to accurately analyze sentiment in textual data, reflecting the sophistication needed in this field.

7. Shirley Ann Williams, Melissa Terras “What people study when they study Twitter ” in iJET, Vol. 15, No. 15, August 2020

- a) Abstract:** IWith the growing availability of easily accessible and Low cost mobile technology, a niche area has developed generically as microblogging. The use of microblogs has become a means of real time Commenting on, responding to, and amplifying the impact of current events.
- b) Inference:** The statement suggests that the proliferation of affordable mobile technology has led to the emergence of microblogging as a prominent communication medium. Microblogs provide a platform for individuals to engage in real-time commentary, response, and dissemination of information regarding current events. This trend underscores the increasing role of social media platforms in shaping public discourse and amplifying the impact of ongoing events. The accessibility and immediacy of microblogging platforms enable users to participate in discussions and share opinions rapidly, reflecting the evolving landscape of digital communication.

8. **David Howoldt · Henning Kroll · Peter Neuhäusler · Alexander Feidenheimer**
“Understanding researchers’ Twitter uptake, activity
and popularity” iScientometrics 128, 325–344 (2023)

- a) **Abstract:** It examines whether research productivity as measured by publication output, conference visits and citations affects the probability of researchers taking up Twitter use, being active on Twitter, and increasing their popularity. On the other hand, research quality as measured by citations and conference visits are positively associated with popularity.
- b) **Inference:** The study investigates the relationship between research productivity and quality, and researchers' engagement with Twitter. It suggests that researchers with higher publication output, conference visits, and citations are more likely to adopt Twitter, be active on the platform, and increase their popularity. Furthermore, the findings indicate that research quality, measured by citations and conference visits, positively influences researchers' popularity on Twitter. This implies that active engagement and visibility on Twitter are correlated with both productivity and quality of research, highlighting the platform's role in scholarly communication and dissemination.

2.2. Patent Search :

1. System and method for identifying influential users in a social network

Inventor: David William Woolf, Yuri Finkelstein.

This patent covers techniques for analyzing social network data, including Twitter, to identify influential users based on their interactions, engagement metrics, and network centrality.

2. System and method for determining engagement in a social network" -

Inventors: David William Woolf, Yuri Finkelstein.

This patent describes a system for analyzing engagement metrics on social networks like Twitter, including user interactions such as retweets, likes, and replies, to determine the level of engagement with specific content or users.

2.3. Inference Drawn:

The patents may cover a range of techniques, including sentiment analysis, trend detection, personalized recommendation systems, and identification of influential users. Such inventions could have applications in various fields, including marketing, advertising, social media management, and academic research, where understanding and leveraging social network dynamics are essential.

Overall, these patents reflect efforts to innovate and advance the field of social network analysis, particularly with respect to Twitter, potentially providing valuable insights and tools for understanding and navigating the complex landscape of online social interactions.

Chapter 3: Requirement Gathering for the Proposed System

In this chapter we are going to discuss the resources we have used and how we analysed what the user actually needs and what we can provide. We will also discuss the functional and non-functional requirements and finally the software and hardware used.

3.1. Introduction to Requirement Gathering:

The Requirement Gathering is a process of requirements discovery or generating list of requirements or collecting as many requirements as possible by end users. It is also called as requirements elicitation or requirement capture.

The requirements gathering process consists of six steps :

- Identify the relevant stakeholders
- Establish project goals and objectives
- Elicit requirements from stakeholders
- Document the requirements
- Confirm the requirements
- Prioritise the requirements

3.2. Functional Requirements:

- The system must be capable of scraping data from Twitter, including tweets, mentions, and hashtags, to gather relevant information about the target profiles.
- It should have the ability to identify and extract mentions of the target profiles from tweets and other interactions on Twitter.
- The system must conduct subject analysis to understand the topics in the content associated with the target profiles.
- It should be able to identify and analyze hashtags used in association with the target profiles to gauge their relevance and popularity.
- The system should recognize communities or groups based on mentions and interactions with the target profiles, providing insights into their influence and engagement levels.

3.3.Non-Functional Requirements:

- The system should be scalable to accommodate a growing volume of data and increasing user demands without compromising performance.
- It must adhere to strict security protocols to safeguard the collected data and ensure user privacy and compliance with data protection regulations.
- The analyses performed by the system should be accurate and reliable, providing organizations with trustworthy insights for decision-making.
- The user interface should be intuitive and user-friendly, allowing organizations to navigate and interpret the generated reports with ease.
- The system should be responsive and efficient, capable of processing and analyzing data in a timely manner to meet the needs of organizations for real-time insights and decision-making.

3.4.Hardware, Software, Technology and Tools Utilised:

A. Hardware Requirements:-

- a. **Processor:** Multi-core, modern CPU for faster data processing.
- b. **Memory (RAM):** A minimum of 8GB, though 16GB or more is recommended for smoother database operations and handling large datasets.
- c. **Storage:** Solid State Drive (SSD) with adequate space (depending on the size of your dataset), for faster read/write operations.
- d. **Network:** Stable internet connection for fetching data from Twitter's API and possibly other online resources.

B. Software Requirements:-

For Scraping:

- Selenium Java
- Maven Package Manager
- IntelliJ IDE

For Interactive UI

- React Js
- Tailwind CSS
- Apex Charts Plugin
- React router dom
- Node js and NPM

For Analysis

- streamlit - turns data scripts into shareable web apps
- scikit-learn - predictive data analysis
- nltk - natural language processing tool
- wordcloud - tagging, dictionary

Web Hosting

Netlify for Frontend hosting, Render for Backend

Techniques:-

- **Reactjs :-** React (also known as React.js or ReactJS) is a free and open-source front-end JavaScript library for building user interfaces based on UI components. It is maintained by Meta (formerly Facebook) and a community of individual developers and companies. React can be used as a base in the development of single-page, mobile, or server-rendered applications with frameworks like Next.js.
- **Tailwind CSS :-** Tailwind CSS is a utility-first CSS framework for rapidly building modern websites without ever leaving your HTML. Tailwind CSS works by scanning all of your HTML files, JavaScript components, and any other templates for class names, generating the corresponding styles and then writing them to a static CSS file. It's fast, flexible, and reliable — with zero-runtime.

Tools:-

- **Vscode:-** Visual Studio Code is a streamlined code editor with support for development operations like debugging, task running, and version control. It aims to provide just the tools a developer needs for a quick code-build-debug cycle and leaves more complex workflows to fuller featured IDEs, such as Visual Studio IDE.

3.5. Constraints:

- Internet Access is required.
- Users should be able to interpret the data in the form of visualisation.
- Our system is restricted to analyse for certain twitter accounts only.

Chapter 4: Proposed Design

4.1. Block Diagram of the proposed system:

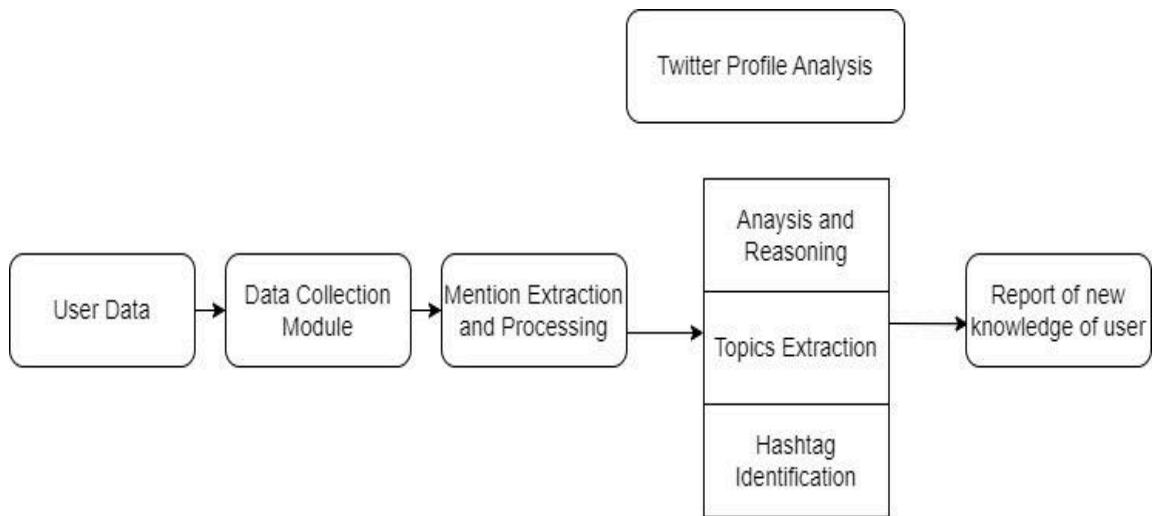


Fig 4.1: Block Diagram

4.2. Modular diagram of the system:

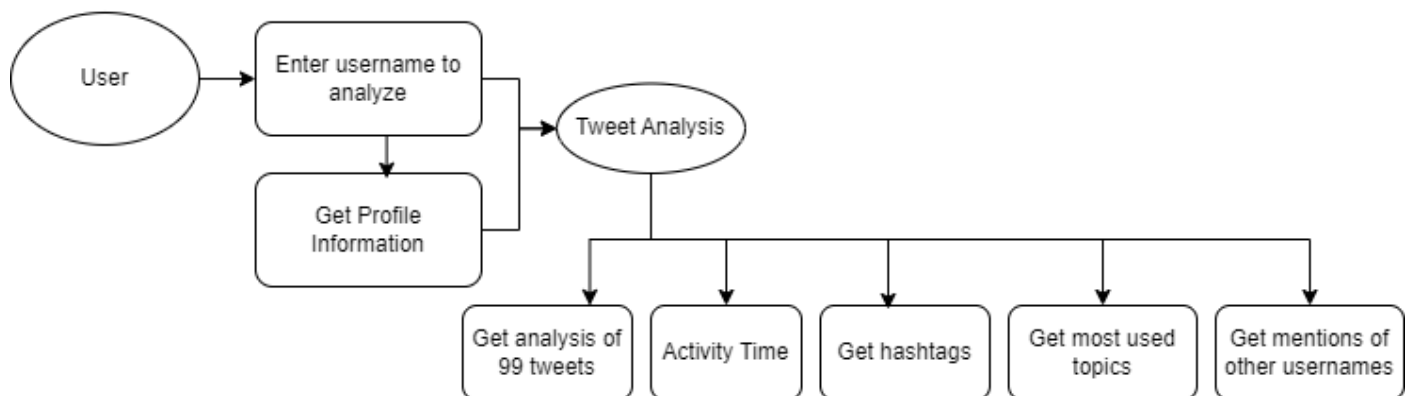


Fig 4.2: Modular Diagram

- Marketing: Twitter user analysis helps businesses identify key demographics, preferences, and behaviors to tailor marketing strategies and content for optimal engagement.
- Brand management: By analyzing Twitter users' interactions and perceptions, companies can gauge brand sentiment, address customer concerns, and maintain a positive brand image.
- Social research: Researchers utilize Twitter user analysis to study public opinion, societal trends, and emerging topics, providing valuable insights into cultural phenomena and societal dynamics.
- Influencer identification: Through analyzing engagement metrics and audience demographics, brands can identify influential Twitter users whose endorsement can enhance their reach and credibility within specific niches or industries.
- Target audience segmentation: Twitter user analysis enables businesses to segment their target audience based on interests, behaviors, and demographics, allowing for more precise targeting and personalized communication strategies.

4.3. Project Scheduling & Tracking using Time line / Gantt Chart:

The Gantt chart of our project where we worked for the whole semester to create this model is shown in a timeline pattern. It is the most important part to think and design the planning of your topic and so we planned our work like the gantt chart shown.

Gantt Chart

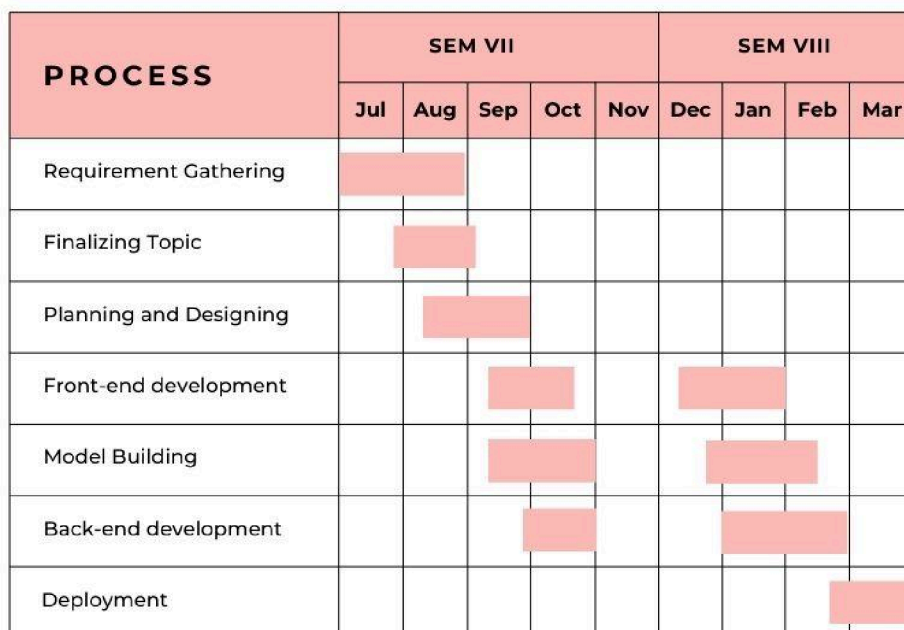


Fig 4.3 : Gantt chart

Chapter 5: Implementation of the Proposed System

5.1.Methodology employed for development:

Our methodology for analyzing Twitter profiles is meticulously crafted to derive meaningful insights from the vast reservoir of data available on the platform. By harnessing advanced data analysis techniques and natural language processing (NLP) algorithms, we aim to unveil valuable information regarding user behavior, content preferences, and interaction patterns. The methodology comprises several pivotal steps, encompassing data collection, preprocessing, topic extraction, time activity analysis, and community detection from mentions. Through rigorous analysis and visualization of the gathered data, our goal is to furnish users with actionable insights to enrich their comprehension of Twitter profiles and refine their engagement strategies. Furthermore, our iterative evaluation and refinement approach ensure the accuracy, relevance, and effectiveness of our analysis methodology in capturing the intricacies of Twitter interactions.

Data Collection:

Initiating the process involves the collection of data from Twitter profiles, encompassing recent tweets and associated metadata such as timestamps, mentions, hashtags, and user interactions.

Preprocessing:

Subsequently, the collected data undergoes preprocessing to cleanse and prepare it for analysis, entailing the elimination of noise such as special characters and URLs, and tokenizing the text into individual words or phrases.

Topic Extraction:

Our methodology employs natural language processing (NLP) techniques, specifically term frequency-inverse document frequency (TF-IDF), to discern recurring themes or topics within the tweets. This procedure categorizes tweets into clusters based on content similarity, elucidating the primary topics of discussion.

Time Activity Analysis:

The time activity analysis endeavors to unveil patterns in user activity over time, scrutinizing tweet timestamps to ascertain peak activity periods, posting frequency, and overall user engagement throughout the day. By visualizing this data, users can pinpoint optimal engagement times and detect irregularities indicative of spam or automated activity.

Community Detection from Mentions:

Community detection from mentions entails the identification of clusters or communities of users based on mentions and interactions within tweets. This process employs graph theory algorithms, such as

Girvan-Newman or Louvain modularity optimization, to pinpoint densely connected user groups. These communities elucidate cohesive groups within the Twitter network, offering insights into user relationships and interaction dynamics.

Application Development:

The project commenced with the installation of React-App, a widely utilized JavaScript library for constructing user interfaces. React provided a sturdy foundation for developing a dynamic and interactive analysis platform.

Integration of Insights Using RAFC (React Functional Component):

To seamlessly integrate insights into the React application, we adopted the RAFC (React Functional Component) architecture. This facilitated the creation of reusable and modular components to exhibit various insights, encompassing information, statistics, topics, and time analysis.

Utilization of React Routes:

React Routes were implemented to streamline navigation within the application. By delineating distinct routes for each analysis section (information, statistics, topics, etc.), users could effortlessly navigate between different insights utilizing a single-page application approach.

Integration of ApexCharts for Data Visualization:

We integrated ApexCharts, a robust and customizable charting library, into the React application to visualize data. By leveraging ApexCharts, a myriad of charts and graphs including line charts, bar charts, and pie charts, were generated to visually represent insights derived from Twitter profile analysis..

l on annotated text data to recognize and classify named entities in new, unseen resumes.

Chapter 6: Testing of the Proposed System

6.1.Introduction to Testing :

Software testing of the current system in Twitter user analysis is a meticulous process aimed at ensuring the platform's functionality, performance, and reliability. Functionality testing scrutinizes every feature and aspect of the system, verifying that data collection from Twitter profiles is accurate, preprocessing steps effectively clean and prepare the data, and analysis algorithms produce meaningful insights. Data integrity testing ensures the consistency and trustworthiness of the data throughout the analysis pipeline, guarding against corruption or loss. Performance testing assesses how well the system handles varying workloads and user interactions, evaluating factors such as data collection speed, preprocessing efficiency, and analysis response times to ensure optimal performance under different conditions.

Usability testing focuses on the user interface and experience, ensuring that users can intuitively navigate the platform and interpret the insights it provides. Security testing is also critical, identifying and addressing potential vulnerabilities to protect user data and maintain the platform's integrity. Together, these testing procedures ensure that the Twitter user analysis system operates reliably, delivering accurate and actionable insights to users.

6.2.Types of tests Considered:

A. Pre testing phase

In the pre-testing phase of the Twitter user analysis system, objectives and requirements are defined, and a dedicated testing team is assembled. A comprehensive test plan is crafted, outlining the testing approach, test cases, and schedule. Test environments are set up, along with necessary tools and prerequisites like sample data sets and access to Twitter APIs. This phase ensures thorough preparation for the testing process, enabling the team to execute tests effectively, identify potential issues early on, and ensure the system meets quality standards prior to deployment.

B. Beta-Testing Phase

Twitter user analysis involves examining various aspects of user behavior and content preferences on the platform. This analysis typically includes studying hashtags, most frequently used words, and 24-hour activity patterns.

1. **Hashtag Analysis:** This involves identifying and analyzing the hashtags used by a Twitter user or within a specific set of tweets. It helps understand the topics of interest to the user or the broader Twitter community, providing insights into ongoing discussions, trends, or events.
2. **Most Used Words:** Analyzing the most frequently used words in a user's tweets or a collection of tweets reveals recurring themes or topics of conversation. This analysis helps uncover the user's interests, preferences, and the type of content they engage with or produce.

3. 24-hour Activity: Examining a user's activity patterns over a 24-hour period helps identify peak tweeting times, posting frequency, and overall engagement levels throughout the day. Understanding these patterns enables users to optimize their content strategy, ensuring they reach their audience when they are most active on the platform.

By conducting Twitter user analysis encompassing hashtags, most used words, and 24-hour activity, researchers, marketers, and individuals gain valuable insights into user behavior, content trends, and optimal engagement strategies on the platform.

6.3.Variou test case scenarios considered:

	Test Cases
Case 1 Home Page	
Case 2: The input field for 'userna me' should only valid accept username	

Case 3:
Display of user profile

Profilysis

@narendramodi

Narendra Modi Verified

Updated: March 17, 2023

Prime Minister of India

Location

India

Language Preference

English

Joined

Sat Jan 10 17:18:56 +0000 2009

URL

<https://t.co/m2qxtsbpwb>

Followers

87,141,613

Following

2,528

Ratio

34,470.57

Tweets

35,850

Listed

29,672

2024 Profilysis™. All Rights Reserved.

Case 4:
Start of Analysis

Get Started With Analysis

Know more about this profile, get powerful insights and trends

Start Analysis

6.4.Inference drawn from the test cases:

Test Cases

Case 1:
Proper Breakdown of Tweets
of the User

Profilysis

Tweets Analysis

We analyzed 99 of the tweets by @narendramodi and here's what we found

Tweets with @mentions

15 / 100

Tweets with #hashtags

4 / 100

Tweets with links

84 / 100

Tweets with media

15 / 100

Retweets

0 / 100

Replies

0 / 100

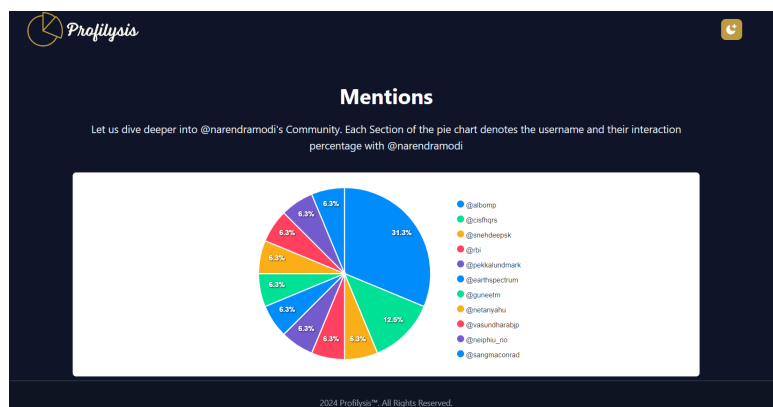
<p>Case 2 :</p> <p>Showing activity time analysis</p>	
---	--

<p>Case 3:</p> <p>Display of most tweeted topics</p>	
--	--

<p>Case 4:</p> <p>Display of most used hashtags</p>	
---	--

Case 5:

Display of Mention
Percentage



Chapter 7: Results and Discussions

7.1.Screenshot of Use Interface(UI) for the system:



Fig 7.1: Screenshot for Home page

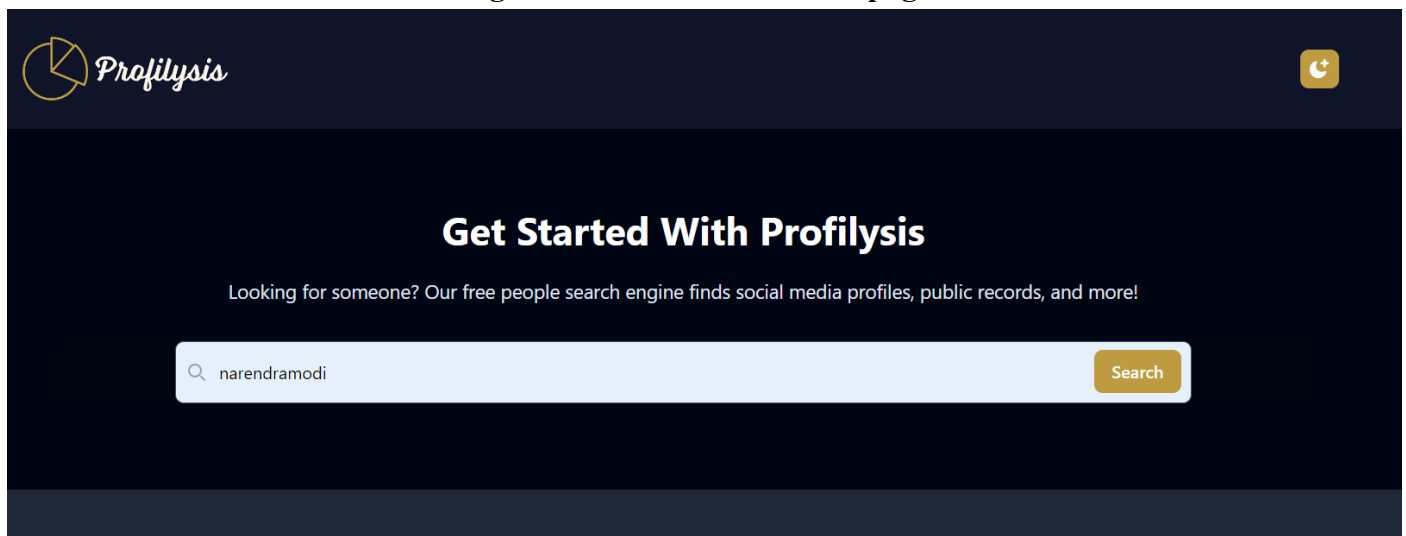





Fig 7.2:Profilysis



Profile Information



@narendramodi
Narendra Modi Verified
Updated: March 17, 2023

Prime Minister of India

Location	India	Followers	87,141,613
Language Preference	English	Following	2,528
Joined	Sat Jan 10 17:18:56 +0000 2009	Ratio	34,470.57
URL	https://t.co/m2qxibpwb	Tweets	35,850
		Listed	29,672

Fig 7.3: Screenshot for profile information

Get Started With Analysis

Know more about this profile, get powerful insights and trends

Start Analysis

Fig 7.4: Screenshot for start of analysis

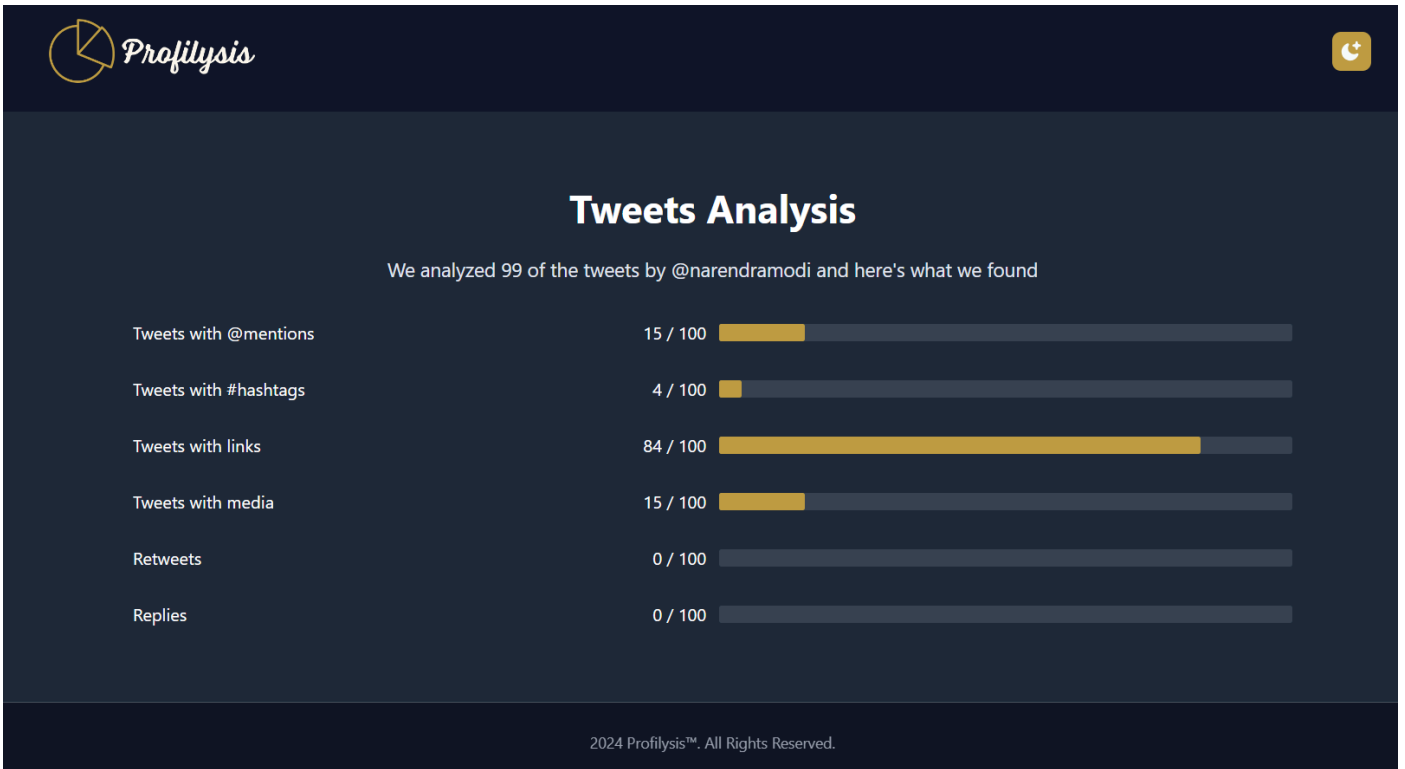


Fig 7.5: Screenshot for Tweets Analysis

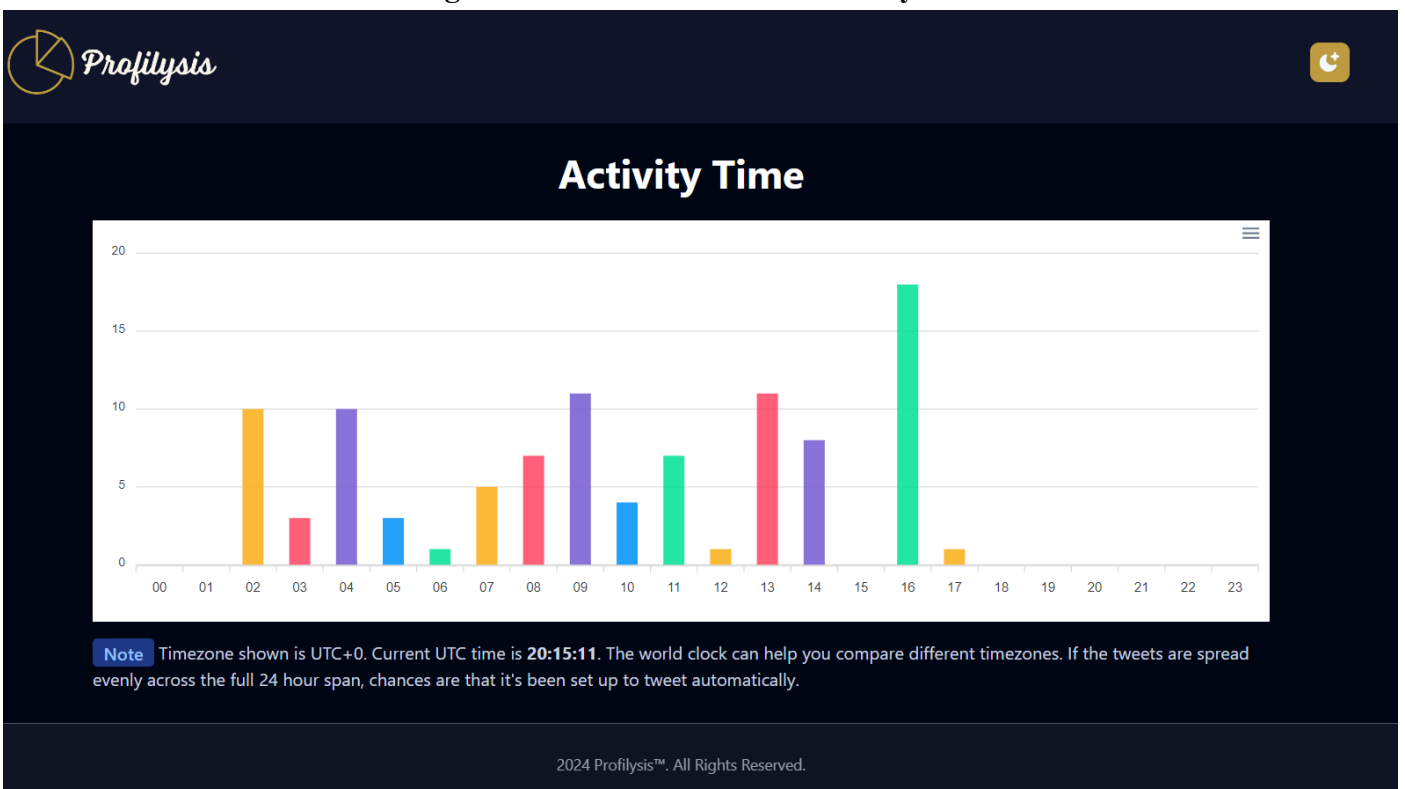


Fig 7.6: Screenshot forActivity time

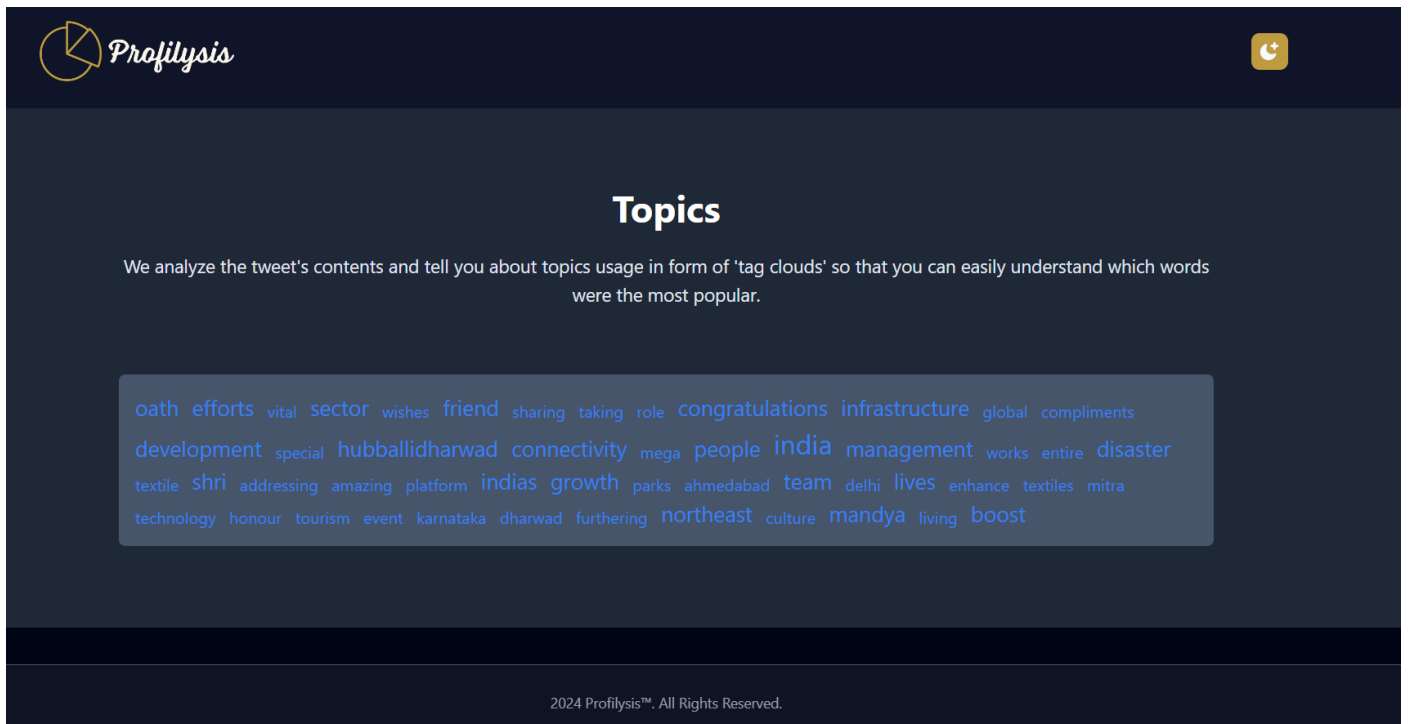


Fig 7.7: Screenshot for Most tweeted topics

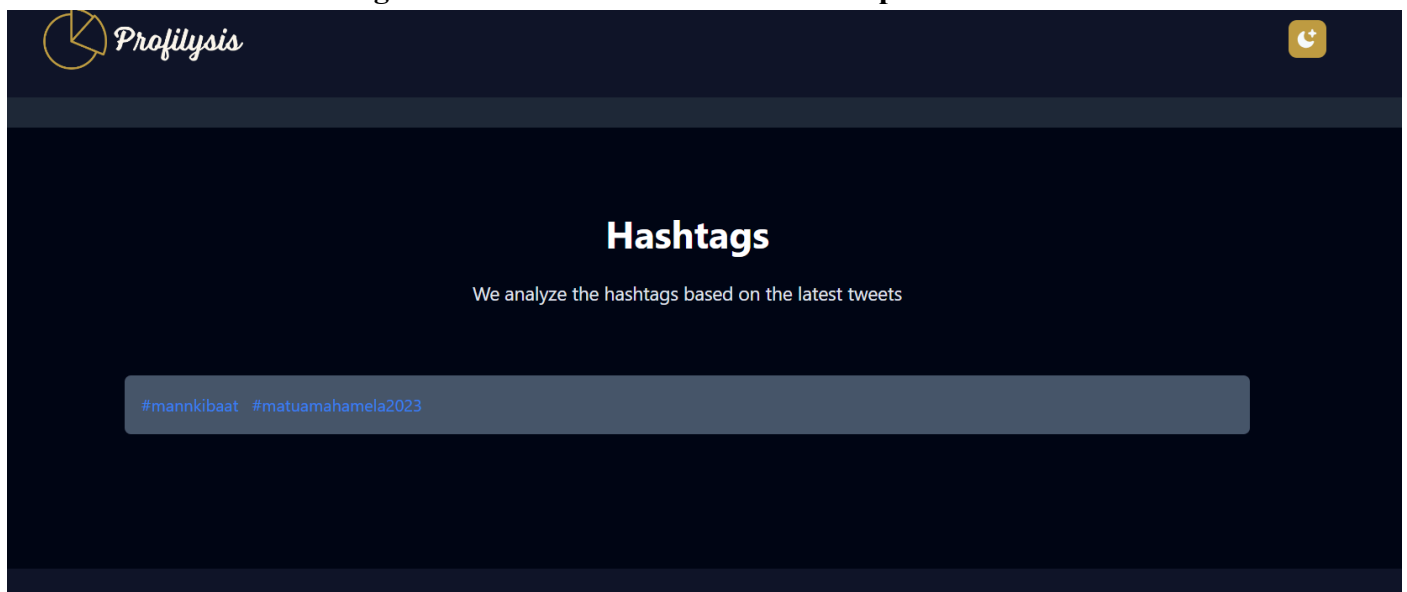
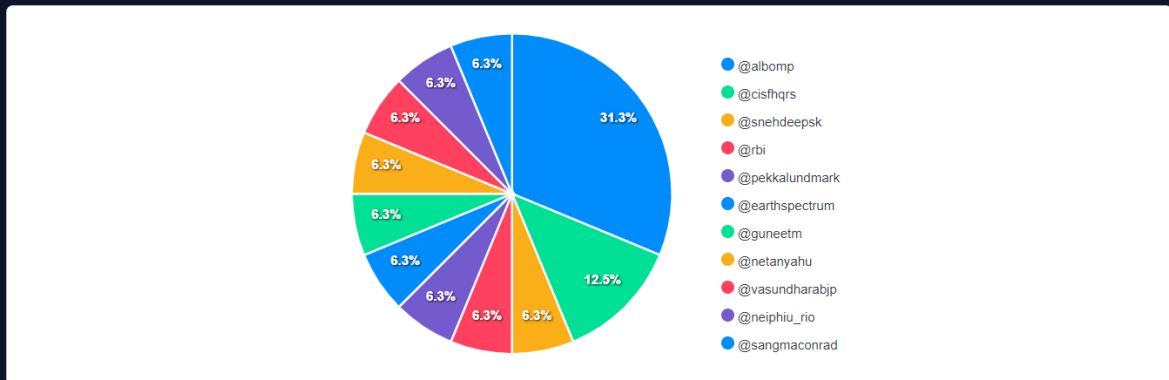


Fig 7.8: Screenshot for Most used hashtags

Mentions

Let us dive deeper into @narendramodi's Community. Each Section of the pie chart denotes the username and their interaction percentage with @narendramodi



2024 Profilysis™. All Rights Reserved.

Fig 7.9: Screenshot for Community Mention

Chapter 8: Conclusion

8.1.Limitations:

- Do not provide further interview services after resume selection to the recruiter.
- Recommendation do not use searching-text analysis.
- Further Interview process is not on the platform.

8.2.Conclusion:

In conclusion, the methodology employed in our project for analysing Twitter profiles represents a systematic and comprehensive approach to extracting valuable insights from social media data. Through the utilisation of advanced data analysis techniques, natural language processing algorithms, and visualisation tools, we have been able to uncover trends, patterns, and relationships within Twitter profiles. By focusing on key aspects such as topic extraction, time activity analysis, and community detection from mentions, we have provided users with actionable insights to enhance their understanding of Twitter dynamics and optimise their engagement strategies. Our iterative approach to evaluation and refinement ensures the accuracy and relevance of the analysis methodology, allowing us to continuously adapt to evolving user needs and platform dynamics. Overall, our methodology offers a powerful framework for unlocking the potential of Twitter data, empowering users to make informed decisions and drive meaningful outcomes in their social media endeavours.

8.3.Future Scope:

- a) Implementation of a video conferencing platform for interview purposes.
- b) Integrating the Startup module.
- c) Enhancing the Courses Recommendation Model by using Deep learning.
- d) Improve accuracy of model via hyper parameterization of Content based recommendation.

References

An Analysis and Research on Sentiments of Twitter Data : <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9725702&isnumber=972538>

Literature review - Twitter as A Tool of Market Intelligence for Businesses
https://aaltodoc.aalto.fi/bitstream/handle/123456789/42417/bachelor_Autio_Heikki_2020.pdf?sequence=1&isAllowed=y

Social network analysis of Twitter interactions: a directed multilayer network approach
<https://link.springer.com/article/10.1007/s13278-023-01063-2>

Using Twitter to engage with customers: a data mining approach
<https://www.emerald.com/insight/content/doi/10.1108/IntR-11-2013-0249/full/html>
Understanding researchers' Twitter uptake, activity and popularity
<https://doi.org/10.1007/s11192-022-04569-2>

Methodological Considerations in
Analysing Twitter Data <https://doi.org/10.1093/jncimonographs/lg t026>
Data Mining And Feature Extraction Techniques For opinion mining analysis in
twitter
<https://www.jatit.org/volumes/Vol100No14/20Vol100No14.pdf>

Appendix

1| Paper I details :-

Twitter Profile Analysis

Tammana Bathija
Department of Computer
Engineering
V.E.S. Institute of Technology
Mumbai, India

Harsha Chelani
Department of Computer
Engineering
V.E.S. Institute of Technology
Mumbai, India

Shreya Kukreja
Department of Computer
Engineering
V.E.S. Institute of Technology
Mumbai, India

Khusboo Kimtani
Department of Computer
Engineering
V.E.S. Institute of Technology
Mumbai, India

Piyush Waghmare
Department of Computer
Engineering
V.E.S. Institute of Technology
Mumbai, India
Mrs. Sujata Khandaskar
Assistant Professor
Department of Computer
Engineering
V.E.S. Institute of Technology

Abstract—An adaptable tool designed to help organisations research and learn more about Twitter profiles of relevant people or entities is the Twitter Profile Analyzer. The system uses information gathered by scraping twitter data , such as tweets, mentions, and hashtags, to perform extensive analyses including mention extraction, subject analysis, hashtag identification, and mention-based community recognition. Organisations can learn a great deal from these analytics about the interactions, interests, and influence of the profile owner on Twitter. The reports that are generated enable organisations to make well-informed decisions about marketing strategies, partnerships, and engagement tactics that are customised to the features and audience dynamics of the profile.

Keywords — Twitter Profile Analyzer, Data Collection, Mention Extraction, Topic Analysis, Hashtag Identification, Community Detection, Insights Generation, Social Media Influence, Engagement Metrics, Collaboration Opportunities, Marketing Strategy Optimization, Social Media Analytics

I. INTRODUCTION

Social media sites like Twitter have developed into effective tools for people and businesses to Mumbai, India interact with their audience, share information, and communicate in the current digital era. Businesses looking to make the most of these platforms must comprehend the intricacies of Twitter profiles, including user interactions, content choices, and community involvement. A complex system called the Twitter Profile Analyzer was created to offer in-depth analysis of the features and functionality of Twitter profiles.

The Analyzer gathers extensive data from Twitter profiles, including tweets, mentions, and hashtags, using straightforward but efficient methods. The system extracts useful information from user interactions, including hashtags that are widely used, subjects that are discussed, and community structures. It does this by using simple data processing techniques.

Organisations can learn more about Twitter profiles, including their audience demographics,

engagement levels, and influence, thanks to the insights that are created. Equipped with this understanding, organisations can make well-informed choices about marketing plans, audience engagement techniques, and collaboration opportunities that are particular to the attributes of the target profile.

The Twitter Profile Analyzer is a useful tool for businesses looking to maximise their online presence and understand the intricacies of Twitter in this age of digital marketing and social media influence.

II.MOTIVATION

The impetus behind developing the Twitter Profile Analyzer is the increasing significance of social media platforms, especially Twitter, as potent instruments for marketing and communication in the current digital environment. Given that millions of people use Twitter every day, it is now crucial for both individuals and businesses to comprehend the subtleties of each profile.

Organisations can gain important insights into the preferences, interests, and behaviours of their audience by analysing Twitter profiles.

Organisations can better connect with their target audience by identifying chances for collaboration, customising their marketing strategies, and optimising their engagement tactics by taking these factors into consideration.

Maintaining and expanding one's online presence requires a grasp of the performance of one's Twitter profile and audience interaction, especially for influencers and content providers. By offering insightful feedback on content, audience interactions, and overall influence, the Analyzer assists users in improving their online presence and strategy.

The main driving force behind the development of the Twitter Profile Analyzer is its capacity to provide businesses and individuals with useful information that they can use to improve their social media strategies, make well-informed decisions, and ultimately increase their influence and success online.

III.EXISTING SYSTEM

Existing systems in Twitter network analysis encompass a diverse range of tools, platforms, and software that researchers and analysts have employed to understand different aspects of the

Twitter ecosystem. These existing systems serve as valuable resources, providing a foundation for comprehending the complexities of Twitter data. Here are some key categories of existing systems:

1. **Social Network Analysis Tools:** These tools offer comprehensive solutions for analysing social networks, including Twitter. Examples include Gephi, NodeXL, and Cytoscape. They allow users to visualise Twitter networks, perform centrality analysis, and detect communities within the network.

2. **Sentiment Analysis Tools:** Given the significance of sentiment analysis in Twitter data, various sentiment analysis tools and APIs are widely used. Tools like VADER and TextBlob aid in assessing the sentiment of tweets and understanding public opinion on the platform.

3. **Data Collection and Extraction Tools:** Twitter network analysis begins with data collection. Tools like Tweepy and software libraries for accessing the Twitter API are essential for gathering Twitter data. Researchers can extract tweets, user profiles, and interaction data using these tools.

4. **Analytics Dashboards:** Some existing systems offer user-friendly analytics dashboards that provide real-time insights into Twitter data. These dashboards are particularly valuable for businesses and marketers, allowing them to monitor trends, track user engagement, and evaluate key performance indicators.

5. **Custom-Built Solutions:** In some cases, researchers develop custom software tailored to their specific research needs. These solutions can offer flexibility in data collection and analysis, allowing for a more targeted approach

IV.PROPOSED SYSTEM-

Our proposed system aims to enhance the existing Twitter profile analysis project built using React by introducing several new features and improvements across different sections. Firstly, we plan to improve the user interface to provide a more intuitive and visually appealing experience for users navigating through the analysis sections. Additionally, we intend to implement real-time data analysis capabilities to ensure that users receive the most up-to-date insights into Twitter profiles, including recent activity and engagement metrics. Furthermore, we aim to expand the statistical

analysis section to include more comprehensive metrics beyond follower count and followers-to-following ratio, such as engagement rates, tweet frequency, and user interaction trends. Moreover, we aim to enhance the time analysis section to provide more detailed insights into the user's activity patterns throughout the day, including identifying peak activity times and frequency of interactions. Expanding the tweet analysis section to provide more in-depth analysis of the content within the user's tweets is also part of our plan. Additionally, we aim to introduce customizable reports, allowing users to tailor the generated reports according to their specific preferences and requirements. Exploring opportunities for integration with third-party analytics tools and platforms to provide users with additional insights and functionality is also on our agenda. Lastly, ensuring data privacy and security by complying with regulations and implementing robust security measures to protect users' sensitive information is a key consideration for the proposed system. Overall, our goal is to provide users with a comprehensive and customizable platform for analysing Twitter profiles, empowering them to make informed decisions and optimise their social media strategies effectively.

V.METHODOLOGY

The methodology employed in our project for analysing Twitter profiles is designed to extract meaningful insights from the vast amount of data available on the platform. By leveraging advanced data analysis techniques and natural language processing (NLP) algorithms, we aim to uncover valuable information about user behaviour, content preferences, and interaction patterns. This methodology involves several key steps, including data collection, preprocessing, topic extraction, time activity analysis, and community detection from mentions. Through rigorous analysis and visualisation of the collected data, we strive to provide users with actionable insights to enhance their understanding of Twitter profiles and optimise their engagement strategies. Additionally, our iterative approach to evaluation and refinement ensures that the analysis methodology remains accurate, relevant, and effective in capturing the dynamics of Twitter interactions. Overall, our methodology offers a systematic and comprehensive framework for uncovering valuable insights from Twitter data, empowering users to

make informed decisions and drive meaningful outcomes in their social media endeavours.

1. **Data Collection:** The first step involves collecting data from Twitter profiles, including recent tweets and associated metadata such as timestamps, mentions, hashtags, and user interactions.
2. **Preprocessing:** The collected data undergoes preprocessing to clean and prepare it for analysis. This includes removing noise, such as special characters and URLs, and tokenizing the text into individual words or phrases.
3. **Topic Extraction:** Topic extraction involves employing natural language processing (NLP) techniques, term frequency-inverse document frequency (TF-IDF), to identify recurring themes or topics within the tweets. This process categorises tweets into clusters based on similarity of content, revealing the main topics of discussion.
4. **Term Frequency - Inverse Document Frequency (TF-IDF)** is a widely used statistical method in natural language processing and information retrieval. It measures how important a term is within a document relative to a collection of documents. Words within a tweets are transformed into importance numbers by a text vectorization process. As its name implies, TF-IDF vectorizes/scores a word by multiplying the word's Term Frequency (TF) with the Inverse Document Frequency (IDF).

Term Frequency: TF of a term or word is the number of times the term appears in a document compared to the total number of words in the tweets.

$TF = \text{number of times the term appears in the tweet} / \text{total number of terms in the tweets}$

Inverse Document Frequency: IDF of a term reflects the proportion of tweets in the corpus that contain the term. Words unique to a small percentage of tweets receive higher importance values than words common across all tweets.

$IDF = \log(\text{number of the tweets in the corpus} / \text{number of tweets in the corpus contain the term})$ Then,

$TF-IDF = TF * IDF$

5. **Time Activity Analysis:** Time activity analysis aims to identify patterns in user activity over time. This involves analysing the timestamps of tweets to determine peak activity periods, frequency of posting, and overall user engagement throughout the day. By visualising this data, users can identify optimal times for engagement and detect anomalies that may indicate spam or automated activity.
6. **Community Detection from Mentions:** Community detection from mentions involves identifying clusters or communities of users based on mentions and interactions within tweets. This process utilises graph theory algorithms, such as Girvan-Newman or Louvain modularity optimization, to identify densely connected groups of users. These communities represent cohesive groups within the Twitter network, providing insights into user relationships and interaction patterns.
7. **Installation of React-App:** The project began with the installation of React-App, a popular JavaScript library for building user interfaces. React provided a robust foundation for developing a dynamic and interactive analysis platform.
8. **Integration of Insights Using RAFC (React Functional Component):** To integrate insights into the React application, RAFC (React Functional Component) architecture was employed. RAFC allowed for the creation of reusable and modular components to display various insights, such as information, statistics, topics, and time analysis.
9. **Utilisation of React Routes:** React Routes were implemented to facilitate navigation within the application. By defining different routes for each section of analysis (information, statistics, topics, etc.), users could seamlessly navigate between different insights using a single-page application approach.
10. **Integration of ApexCharts for Data Visualization:** ApexCharts, a powerful and customizable charting library, was integrated into the React application to visualise data. By utilising ApexCharts,

various charts and graphs, such as line charts, bar charts, and pie charts, were created to visually represent insights derived from Twitter profile analysis.

11.

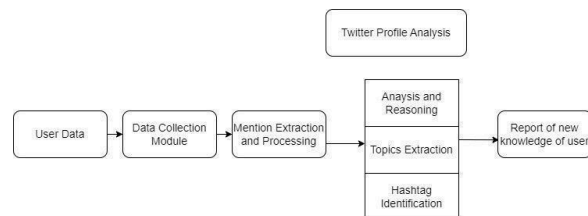


Fig 1. Block Diagram

Methodology Employed For Topic Analysis:

1. **Data Collection:** Utilise the scraping data code to fetch the latest 99 tweets from the user's timeline.
2. **Preprocessing:** Tokenize the text of each tweet and remove stop words. Algorithm: NLTK (Natural Language Toolkit) for tokenization and stopword removal.
3. **Vectorization:** Convert the preprocessed tweets into numerical representations using TF-IDF (Term Frequency-Inverse Document Frequency).
4. **Visualisation:** i) Visualise the extracted topics using word clouds. ii) For a word cloud, the size of each hashtag is proportional to its frequency of occurrence.

Methodology Employed For Activity Time Analysis:

1. **Data Collection:** Utilise the scraping data code to fetch the latest 99 tweets from the user's timeline.
2. **Timestamp Extraction:** Extract the timestamp of each tweet. Twitter provides timestamps in UTC by default.
3. **Activity Time Analysis:** i) Convert each tweet's timestamp to the hour of the day in UTC. ii) Count the frequency of tweets for each hour of the day.
4. **Visualisation:** i) Visualise the distribution of tweet activity over a 24-hour period using a histogram. ii) X-axis represents the hours of the day (0 to 23), and the Y-axis represents the frequency of tweets.
5. **Insights Generation:** Analyse the visualisation to identify patterns in the user's tweeting behaviour. ii) Determine

the peak hours of activity and any notable trends or irregularities.

Methodology Employed For Hashtag Analysis:

1. Data Collection: Utilise the scraping data code to fetch the latest 99 tweets from the user's timeline.
2. Hashtag Extraction: Extract hashtags from each tweet. Hashtags are typically denoted by the '#' symbol followed by the tag itself.
3. Frequency Count: Count the frequency of each unique hashtag across all the tweets.
4. Visualisation: Visualise the frequency distribution of hashtags using word cloud..

Methodology Employed For Community Detection using mentions:

1. Data Collection: Utilise the scraping data code to fetch the latest 99 tweets from the user's timeline.
2. Mention Extraction: Extract mentions of other users from each tweet. Mentions are typically denoted by the '@' symbol followed by the username.
3. Community Building: i) Count the frequency of each unique mention across all the tweets to identify the mentioned users. ii) Group the mentioned users into a community based on the frequency of their mentions. Those with higher frequencies are likely to be more closely associated with the user.
4. Visualisation: i) Visualise the distribution of mentioned users using a pie chart. ii) Each slice of the pie represents a mentioned user, and the size of the slice corresponds to the frequency of their mentions

VI.RESULTS AND DISCUSSIONS

Twitter id : @narendramodi

Updated: March 17, 2023

Location: India

Language Preference: English

Joined: Sat Jan 10 17:18:56 +0000 2009

URL: <https://t.co/m2qxixbpwb>

Followers: 87,141,613

Following: 2,528

Ratio: 34,470.57

Tweets: 35,850

Listed: 29,672

Topics: oath, efforts, vital, sector, wishes, friend, sharing, taking, role, congratulations, infrastructure, global, compliments, development, special, hubballidharwad, connectivity, mega, people, india, management, works, entire, disaster, textile, shri, addressing, amazing, platform, indias, growth, parks, ahmedabad, team, delhi, lives, enhance, textiles, mitra, technology, honour, tourism, event, karnataka, dharwad, furthering, northeast, culture, mandya, living, boost

Hashtags: #mannkibaat, #matuamahamela2023

Mentions: @albomp, @cisfhqrs, @rbi, @snehdeepesk, @pekkalundmark, @earthspectrum, @guneetm, @netanyahu, @vasundharabjp, @neiphiu_rio, @sangmaconrad

VII. CONCLUSION

In conclusion, the methodology employed in our project for analysing Twitter profiles represents a systematic and comprehensive approach to extracting valuable insights from social media data. Through the utilisation of advanced data analysis techniques, natural language processing algorithms, and visualisation tools, we have been able to uncover trends, patterns, and relationships within Twitter profiles. By focusing on key aspects such as topic extraction, time activity analysis, and community detection from mentions, we have provided users with actionable insights to enhance their understanding of Twitter dynamics and optimise their engagement strategies. Our iterative approach to evaluation and refinement ensures the accuracy and relevance of the analysis methodology, allowing us to continuously adapt to evolving user needs and platform dynamics. Overall, our methodology offers a powerful framework for unlocking the potential of Twitter data, empowering users to make informed decisions and drive meaningful outcomes in their social media endeavours.

VIII.REFERENCES

1. An Analysis and Research on Sentiments of Twitter Data : <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9725702&isnumber=972538>

2. Literature review - Twitter as A Tool of Market Intelligence for Businesses
https://aaltodoc.aalto.fi/bitstream/handle/123456789/42417/bachelor_Autio_Heikki_2020.pdf?sequence=1&isAllowed=y
3. Social network analysis of Twitter interactions: a directed multilayer network approach
<https://link.springer.com/article/10.1007/s13278-023-01063-2>
4. Using Twitter to engage with customers: a data mining approach
<https://www.emerald.com/insight/content/doi/10.1108/IntR-11-2013-0249/full/html>
5. Understanding researchers' Twitter uptake, activity and popularity
<https://doi.org/10.1007/s11192-022-04569-2>
6. Methodological Considerations in Analysing Twitter Data
<https://doi.org/10.1093/jncimonographs/lgt026>
7. Data Mining And Feature Extraction Techniques For opinion mining analysis in twitter
<https://www.jatit.org/volumes/Vol100No14/20Vol100No14.pdf>

Title of the Paper:		Twitter Profile Analysis			
Criteria:	Continuity	Text structure	Understanding and Illustrations	Explanatory Power	Detailing
Points out of 100%:	91%	90%	93%	84%	92%
Unique Contents: 95%			Paper Accepted: Yes		

a. Paper I :-

c. Project review sheet ;

Project review sheet 1:

Inhouse/ Industry_Innovation/Research:

Sustainable Goal:

Title of Project: TWITTER USER ANALYSIS.

Group Members: Jammana Bathya (D17C/06), Khushboo Kuntani (D17C/26), Harsha Chelani (D17B/14), Shreya Kukreja (D17A/82), Pooja Waghmare (D17A/71)

Class: D17 A/B/C

Group No.: 42

Project Evaluation Sheet 2023 - 24

Engineering Concepts & Knowledge	Interpretation of Problem & Analysis	Design / Prototype	Interpretation of Data & Dataset	Modern Tool Usage	Societal Benefit, Safety Consideration	Environment Friendly	Ethics	Team work	Presentation Skills	Applied Engg&Mgmt principles	Life-long learning	Professional Skills	Innovative Approach	Research Paper	Total Marks
(5)	(5)	(5)	(3)	(5)	(2)	(2)	(2)	(2)	(2)	(3)	(3)	(3)	(3)	(5)	(50)
4	3	3	2	3	2	2	2	1	2	2	2	2	2	—	33

Comments: Implementation details are not need to Explained in detail
case study based / usecase based applicability of need to show
proper individual efforts are required, data set?

Name & Signature Reviewer 1: Yug Chhaya 10/02/24

Engineering Concepts & Knowledge	Interpretation of Problem & Analysis	Design / Prototype	Interpretation of Data & Dataset	Modern Tool Usage	Societal Benefit, Safety Consideration	Environment Friendly	Ethics	Team work	Presentation Skills	Applied Engg&Mgmt principles	Life-long learning	Professional Skills	Innovative Approach	Research Paper	Total Marks
(5)	(5)	(5)	(3)	(5)	(2)	(2)	(2)	(2)	(2)	(3)	(3)	(3)	(3)	(5)	(50)
4	3	3	2	3	2	2	2	1	2	3	3	2	2	—	33

Comments: Research paper

Date: 10th february, 2024

Name & Signature Reviewer 2: 10/02/24

Project review sheet 2

Industry/Innovation/Research:

Sustainable Goal:

Project Evaluation Sheet 2023 - 24

Class: D17 A/B/C

Group No.: 42

Title of Project: Twitter User Analysis

Group Members: Tamanna Batlija (D17C16) Khushti Kintani (D17C16) Harsita Chelani (D17B14) Shreya Kulkarni (D17A17) Pooja Waghmare (D17A17)

Engineering Concepts & Knowledge	Interpretation of Problem & Analysis	Design / Prototype	Interpretation of Data & Dataset	Modern Tool Usage	Societal Benefit, Safety Consideration	Environment Friendly	Ethics	Team work	Presentation Skills	Applied Engg&Mgmt principles	Life-long learning	Professional Skills	Innovative Approach	Research Paper	Total Marks
(5)	(5)	(5)	(3)	(5)	(2)	(2)	(2)	(2)	(2)	(3)	(3)	(3)	(3)	(5)	(50)
3	4	4	2	4	2	2	2	1	1	2	2	2	2	02	35

Comments: Group work/contribution is missing. In-depth analysis and processing should be shown.

Name & Signature Reviewer1

Engineering Concepts & Knowledge	Interpretation of Problem & Analysis	Design / Prototype	Interpretation of Data & Dataset	Modern Tool Usage	Societal Benefit, Safety Consideration	Environment Friendly	Ethics	Team work	Presentation Skills	Applied Engg&Mgmt principles	Life-long learning	Professional Skills	Innovative Approach	Research Paper	Total Marks
(5)	(5)	(5)	(3)	(5)	(2)	(2)	(2)	(2)	(2)	(3)	(3)	(3)	(3)	(5)	(50)
3	4	4	2	4	2	2	2	1	2	2	2	2	2	03	37

Comments:

Date: 9th March, 2024

Shijata Khandasakar
Name & Signature Reviewer 2

