SOCIAL MEDIA SENTIMENT ANALYSIS

Alli Sunvith Reddy, Atla Tarun Kumar, Annepally Sanjay, Rohith

B. Tech Computer Science- Data Science AI & ML with upgrad, Lovely Professional University Phagwara, Punjab- 144411.

Abhinaya Anand

***Abstract*— The significance of text analysis on social media has grown in relation to comprehending market dynamics, sentiment trends, and user behaviour. We provide a fresh technique to social media analysis using the Streamlit framework and Python in answer to this demand. In order to give comprehensive answers to the many challenges of social media analysis, our system incorporates multiple APIs, such as the VADER Sentiment Analysis for sentiment evaluation, the PRAW for Reddit data extraction, and the Gemini API for automated text production and translation. Through real-time text correction, sentiment analysis, and language translation capabilities, our project enables rapid trend analysis, sentiment monitoring, and cross-lingual communication while offering enterprises, researchers, and analyst's useful data.**

Keywords*— *NLP, Emotions, Reddit, Visual, Interactions.

I. INTRODUCTION

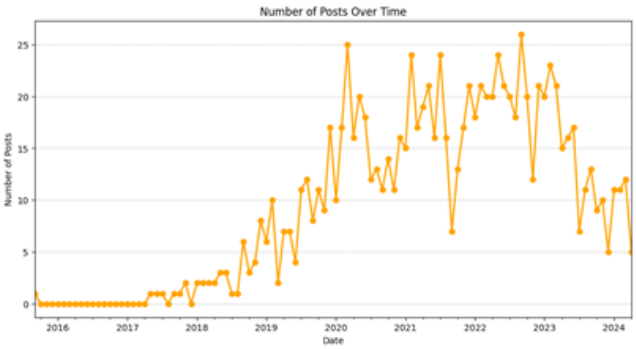
Because they provide insightful information on the beliefs and behaviours of their users, social media platforms are excellent sources of data. Our project's major goal is to utilize Streamlit and Python to the fullest extent possible for social media word analysis. Our aim is to equip academics and analysts with strong instruments for trend research, sentiment monitoring, and multilingual communication by merging APIs such as VADER Sentiment research and PRAW for Reddit data extraction with the Gemini API for text production and translation. This research aims to improve our knowledge of social media dynamics and facilitate wise decision-making for various businesses.

II. METHODOLOGY

Selecting Reddit as the primary data collection platform was the first stage in the project. Reddit's vast user base and variety of material are the reasons we utilize it. We made sure to collect a range of data types from the Reddit API in order to represent the diverse views and behaviours of users on social media platforms. on addition, we created real-time data gathering tools that let us monitor and analyse current discussions and patterns on online discussion boards.

Following data acquisition, we initiated a thorough pre-processing phase. This included sanitizing the data, verifying its formatting, and classifying it in compliance with pertinent standards. The rationale behind this initial processing. Sturdy storage options allowed for efficient data organizing and storage while also making data retrieval and access simple for analytical purposes.

The temporal dynamics of user interactions and sentiments were provided with relevant information by using time series analytical techniques to identify trends over time. To give a comprehensive picture of user attitudes and opinions, two visualizations that were used were pie charts and bar graphs, which display the sentiment distribution across postings.



Word clouds originally provided a visually understandable representation of key topics and were a powerful tool for locating recurring themes and subjects in text data. In addition, the dataset's underlying patterns and linkages were examined using correlation heatmaps, which provided insight into correlations between various variables.

Users were able to study and analyse the insights derived from the vast dataset that was acquired from Reddit by means of interactive dashboards that were designed to facilitate dynamic data exploration.

III. AUTO-CORRECTION

Auto-correction is not only a perk in social media text analysis; it is a must. Because social media conversations are informal and usually unstructured, autocorrection ensures that the data being analyzed is accurate, understandable, and clear. To elaborate further on this project element, three key components could be used: The three key areas of concentration are context comprehension, error detection and repair, and enhanced user engagement.

A) Error Detection and Correction:

Slang and Abbreviation Correction Challenges: Text on social media is full with slang, abbreviations, and unusual terms. In order to overcome this difficulty, a dynamic, ever-changing library of slang and acronyms used in social media contexts had to be created. Our correction algorithms can now comprehend and repair social media language in a more sophisticated and thorough manner thanks to the integration of this database.
The significance of data cleaning In the field of data analysis, the quality of the insights obtained is directly influenced by the cleanliness of the data. By fixing errors at their source, an auto-correction process helps us greatly enhance the quality of the data and guarantees that the analysis done on it is founded on accurate and trustworthy information.

Application of Advanced Algorithms: Our auto-correction feature is based on the application of advanced natural language processing (NLP) algorithms that can recognize grammatical faults as well as slang in addition to spelling mistakes. In order to detect and recommend corrections, this calls for sophisticated pattern recognition algorithms that compare text strings to enormous databases.

B) Contextual Understanding
Adapting to New Linguistic Trends: The constantly evolving language of social media sites presents a serious challenge to static auto-correction. Our method is based on continuous learning, where the system learns to accommodate new usage patterns, slang terms, and language. Because of this, the system remains operational over time, despite changes in the language used on social networking platforms. Beyond Simple Solutions: Our auto-correction engine uses semantic analysis in addition to spelling corrections to make sure each post makes sense. Maintaining the integrity of the data depends on making sure that any repairs are carried out in a way that preserves the original intent and meaning of the user's message.

Our technology increases sentiment analysis by accurately comprehending and correcting text, resulting in increased dependability. To accurately measure the sentiment of social media messages, inaccuracies must be corrected as well as context understood. This is critical for businesses, politicians, and researchers utilizing social media data.

C) User Interaction Enhancement: Auto-correction enables users to enter precise data and enhance their analytical queries more effortlessly, hence boosting platform engagement. For non-native speakers and those unfamiliar with social media writing, in particular, this democratizes access to data. Because our method is constantly evolving due to its auto-correction based on user feedback. After evaluation, user-submitted corrections could be included to the system. By using a community-driven approach, this promotes increasing accuracy and relevance.

Enhancing the Whole User Experience: Auto-correction is a user-facing and backend service that makes the whole experience better. Analytical results are more accurate and pertinent when comments and corrections are made in real time.

Social media text analysis is made much more accessible and of higher quality thanks to auto-correction, which uses a multidimensional method that includes error detection, contextual

comprehension, and user participation. In order to guarantee that the conclusions drawn from social media data are precise and useful for a variety of stakeholders, it tackles the inherent difficulties associated with evaluating informal and unstructured content.

IV. SENTIMENT ANALYSIS

A)Sentiment Analysis Framework:

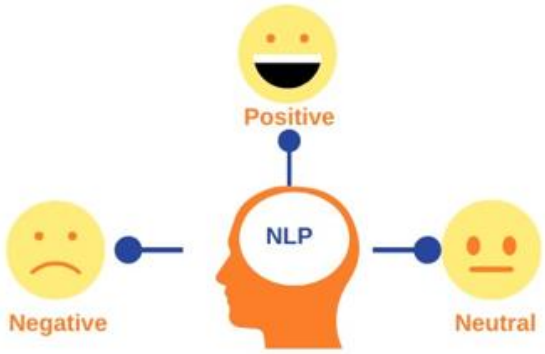
To evaluate textual data sentiment, including context, slang, and sarcasm in social media posts, our technology incorporates cutting-edge Natural Language Processing (NLP) approaches. To precisely identify positive, negative, neutral, and mixed emotions, the algorithm makes use of enormous datasets.

To precisely categorize sentiment on social media sites, we employ machine learning models like Recurrent Neural Networks (RNN) and Support Vector Machines (SVM). We keep adding new data to these models so we can improve accuracy and make adjustments for evolving linguistic usage. Sentiment Scoring and Categorization: We identify sentiment trends with ease by giving each post a sentiment score based on our analysis of text and emojis.

B) Emoji Analysis:

Emoji as Sentiment Indicators: Our investigation examines emoji use as a digital language to describe emotions. Each emoji is mapped to specific sentiments to improve sentiment analysis. Our method produces trustworthy sentiment analysis by taking into account contextual and cultural differences in how emojis are interpreted.

A thorough grasp of the emotional underpinnings of social media conversation is provided by the sentiment analysis portion of our social media text analysis research, which focuses on textual and emoji analysis. Understanding and reacting to public opinion in the digital age requires the kind of detailed analysis that our technology offers. It achieves this through the use of complex algorithms, emoji sentiment interpretation, and machine learning models.



Sentiment Analysis

V. AUTO GENERATION

Integrating Gemini API, a potent tool for natural language creation, made text generation easier. As a useful tool for content production on several platforms, this technology allowed the development of text that was both coherent and contextually appropriate dependent on user input. Whether for articles, social media posts, or other types of communication, users were able to easily create text that was customized to meet their needs for Gemini API. Those who wanted to create interesting and educational content to share with Reddit's broad community found this feature very helpful for creating text posts.

The adaptability of the Gemini API extended beyond a single language, allowing text to be generated in numerous languages. Because it addresses language barriers and improves engagement with worldwide audiences, this feature makes developed material more accessible and has a bigger audience. Users may easily produce text in many languages by utilizing the multilingual features of Gemini API, which would accommodate varying linguistic preferences and broaden the scope of their content creation initiatives.

When the intended text language or another language other than English was used to retrieve Reddit posts, the auto-generation capability provided a smooth means of translating and editing data to the target language. By using this method, the content of the Reddit post that was retrieved may be copied and pasted into the auto-generation tool to create the exact same text in the desired language. Saving time and money, this automated approach maintained the accuracy and

consistency of the generated information by eliminating the need for human translation.

Rather than single subject specifications, along with the use of automatic transmission of assorted material it is required to ensure a diversity that would be amended accordingly to changing information. With this in mind, considering the fact that the usage of machine learning algorithms is growing and the manner of users´ interaction is diversifying the assistant could display a range of voice characteristics, such as tempo, duration or tone.

During that period, there were a lot of newspapers that took the normal slang into an official language in short time. My view on languages other than my mother tongue i.e. independence and with a different fan community, I think, it's just another thing I'm interested in. the listening skills that demonstrated such virtues were what were involved in this nation and by being assimilated by the people became great people who could understand the hidden meaning.

Within the research project, the application of auto-generation capability which subsists the mergence of intelligent technologies and practical applications manifest the benefits of the natural language generation for content creation. Through the use of an API developed by Gemini, people were allowed to create good text in a quick and effective way, which paved the way for communication and connectivity on digital platforms.

In summary, the introduction of auto-generation capabilities has played an essential role in the improvement of the research project through the provision of a thorough and quick technique that is used in text production for different miscellaneous purposes, including Reddit posts. The Gemini API seamlessly facilitated such an integration which enabled users to bypass language barriers, edit produced content and simplify the whole content creation which in the end enhanced the projects impact and efficiency.

VI. WEB APPLICATION

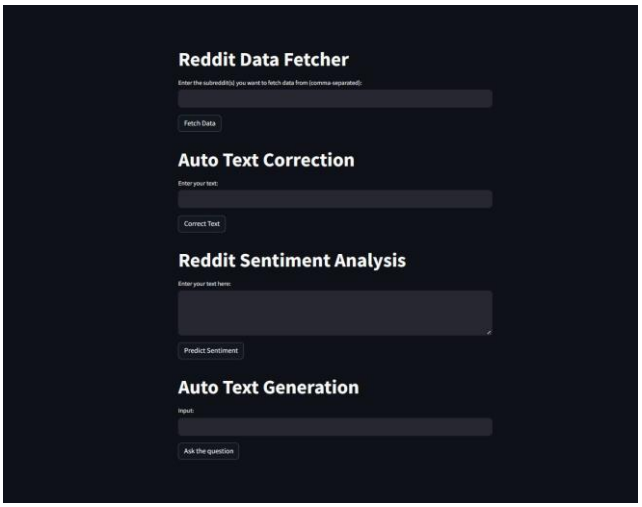
The web application began with choosing Streamlit, which is a Python Library, famously, for its magical abilities of bringing into being easy and interactive web applications. By means of Streamlit's simple and robust tool, we have entered the design and realization process of the platform for text analysis on social media that ultimately offers a user-friendly environment. Web application was designed in a way to provide users with an entertaining and user-friendly experience, making possible the interactions among different data pools, analysis process and dynamic charts appearance.

Architecture of web app was well designed featuring scalability, performance, and reliability in mind. The in-built feature of Streamlit comprising the widgets, layouts and components enabled the users interface that accommodate the needs of diverse activities and individuals. The convenient application details attention to usability and inclusiveness, concordance in the control of the handling, and modification of the data.

Once users had access to the developed web-based application they were shown a welcome interface that could handle various functions for social media text analysis. The home page exhibited how the app selected the key features of the application, which would give the users an insight of the various tools of analysis and visualizations available for them to select. The intuitive and fluid movement between the application's components was made possible by a user-friendly navigation menu. Users were able to swiftly carry out their tasks through this option.
The main goal of the social media web application was realizing such opportunities as Reddit has got. It is known for its numerous users and different content everyone can find something interesting. With the help of PRAW library the application was making an integration with the Reddit and given API for retrieving posts, comments and other data for subsequent analysis. Users could set their preferences and define the relevant data by means of them, for instance, subreddit selection or time period, to adapt the analysis to meet their needs in particular.

Data acquired was then processed and examined utilizing a range of approaches such as sentiment analysis, keyword extraction, and topic

modeling. The Streamlit framework was used to create interactive report widgets and visualization tools for providing analysis results to the public in a way that is clear, concise, and visually appealing. Users will be empowered to access insights gained from analyzing the data either through interaction with charts, graphs, and tables or simply observing the trends, opinions, and patterns revealed.



To boost the comfort of the app, the app included interactive elements like drop down menus, sliders and text inputs that enabled the users to fine tune the analysis settings and explore the data in real time. fortunately, using Streamlit's react programming model ,that allowed the application to update instantly in response to users' moves, turned out to be appropriate for the type of user experience we expected

Furthermore, the web application allowed data analysis, generation, auto-correction, and also translation of languages, which was not the case with the Smartphone app, thus yielding increased utility and versatility. These features employed NLP techniques and external APIs, such as Google Generative AI API, thus users enjoyed the high quality of text handling and
creation.
After users scrutinized the data well and finished the analysis and exploration, they could save their results either save them or export them for more analysis or to share. Streamlit's embedded capabilities to save and export data enabled easy interoperability across all tools and websites, since absolutely anything was possible.

VII. RESULTS

These studies indicate that such a kind of an investigational approach can be applicable for the information made available in an open forum as well as the one that is exclusive for Reddit users. This platform allows for the employment of , among others, that are necessary for the users to check their social impact and reach sound conclusion from the public response.

For instance, auto-correction and auto-suggest make the usage of these tools more convenient while translation allows one speak in different languages. These applications can be employed for the purpose of doing different things basically ranging from reading texts, indexing them, translating to comprehending texts. Hereby, a keen scrutiny of the text you are reading is entirely feasible. This declaration sums up what we cannot avoid nowadays, as almost each business is focused on our habits, personal information and the data we are processing.

VIII. CONCLUSION

In short, the development of the social media web application for text analysis is just the start of using evidence-based information as a matter of fact as at proven via social medias like Reddit amongst many others. Offering as simple interface and detailed analytical tools, users are empowered to analyze data published on social networks while fully capitalizing on its implications for business. Application provides services of data mining and variety of analysis: sentiment analysis; keywords extraction; topic modeling. Thus it is possible to delve into internal minds of users and study people's sentiments and opinions or their groups.

Besides, one of the things that set Cortana apart is the actual details that have been worked into the application like auto-correction, auto-generation, and language translation that, in turn, provides the utility and flexibility which the users with different preferences and prerequisites use. The functions that make use of the most advanced NLP (Natural Language Processing) technology and external APIs will empower the users to make their expressions individualized and energetic. If it is about rephrasing sentences and correcting spelling and grammar issues, or even translate texts from

one language to another, this software features all the necessary tools you may need for creating and editing text in a hassle-free way.

In the coming stage, our app can empower researchers, analysts, and content creators in discovering hidden patterns, detecting trends and making data-driven decisions using social media data. With the souring size and complexity of data, the importance of the accessibility of data and easily understandable data analysis bar would grow higher. From this point on, the software will be regularly updated to a new version, new features added, and lots of new functionalities introduced that address the rapidly changing users' needs, letting the software be at the top among other similar research tools.

REFERENCES

[1] A. Ritter, S. Clark, Mausam, and O. Etzioni, "Named entity recognition in tweets: An experimental study," in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2011.

[2] M. Thelwall, K. Buckley, and G. Paltoglou, "Sentiment strength detection for the social web," Journal of the American Society for Information Science and Technology, vol. 63, no. 1, pp. 163-173, 2012.

[3] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: Bringing order to the web," Stanford InfoLab, Technical Report, 1999.

[4] J. Chang and S. Gerrish, "Bigtable: A distributed storage system for structured data," ACM Transactions on Computer Systems (TOCS), vol. 26, no. 2, pp. 4:1-4:26, 2008.

[5] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL), vol. 1, pp. 655-665, 2014.

[6] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), 2010.

[7] *D. Diakopoulos, "Characterizing debate performance via aggregated Twitter sentiment," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI), pp. 1195-1198, 2010.*

[8] A. H. Kavukcuoglu, Y. Bengio, and G. S. Hinton, "Deep convolutional networks for emotion classification," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 241-248, 2013.

[9] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1631-1642, 2013.

[10] M. J. Paul and M. Dredze, "You are what you tweet: Analyzing Twitter for public health," in Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM), 2011.