# Application of Natural Language Processing Techniques for Sentiment Analysis of Social Media

1rd Ahmed Muntasir Hossain
*Connecticut Institute of Technology*
*University of New Haven*
New Haven, CT
ahoss1@unh.newhaven.edu

2nd Mehdi Mekni
*Connecticut Institute of Technology*
*University of New Haven*
New Haven, CT
mmekni@newhaven.edu

***Abstract*—Digital reputation management systems are in high demand for individuals and businesses seeking to enhance and safeguard their online image. However, current systems like Online Social Network Interactions (OSNI) have limitations in effectiveness, accuracy, cost, and scope. To overcome these challenges, sentiment analysis, a powerful natural language processing technique for identifying opinions, emotions, and attitudes in text data, proves invaluable for digital reputation management. This study proposes the development of an open-source, multi-channel, multi-engine sentiment analysis software called Sentiment Analysis of Social Media (SASM). Specifically designed for social media and digital reputation management, SASM collects and analyzes data from platforms like Twitter, Reddit, and Tumblr. Leveraging sentiment analysis engines such as Microsoft Text Analytics (MTA), IBM Watson Natural Language Understanding (IWNLU), and Google Cloud Natural Language API (GCNLA), SASM filters, aggregates, and assesses sentiment trends. Through a case study on major information technology companies, Google, Amazon, and Microsoft, the feasibility and performance of this multi-channel, multi-engine platform, including GCNLA, MTA, and IWNLU, will be evaluated.**

***Index Terms*—Sentiment Analysis, Software Engineering, Social Media Platforms**

## I. Introduction

Social media platforms play a crucial role as communication channels for both companies and customers. Companies use them to promote products, while customers share reviews and inquiries. Analyzing customer-generated content is essential, and one practical method is sentiment analysis, which involves assessing the feelings conveyed in these posts. By utilizing techniques from natural language processing (NLP) and text analysis tools, sentiment analysis identifies opinions and categorizes them as positive, negative, or neutral [1]. This analysis finds applications in tasks like brand monitoring, campaign evaluation, and competitive analysis [2]. Through sentiment analysis, companies can evaluate their online presence and make well-informed decisions. However, challenges remain in effectively collecting and filtering reviews to extract crucial insights.

To address these challenges, companies use sentiment analysis software to assess feedback from websites, emails, and surveys. However, they face limitations in accessing a wide range of reviews, particularly from diverse social media networks. A recent initiative by Samara et al. introduces OSNI Analytics, a system that gathers publicly posted tweets about a company from Twitter and applies Microsoft Text Analytics for sentiment analysis [3]. This project aims to advance Sentiment Analysis of Social Media (SASM) by integrating various platforms like Twitter, Reddit, and Tumblr. This expansion enhances data collection and analysis capabilities. Additionally, the project plans to incorporate different sentiment analysis engines, including Microsoft Text Analytics (MTA), IBM Watson Natural Language Understanding (IWNLU), and Google Cloud Natural Language API (GCNLA). By evaluating and comparing their performance, the project seeks to refine the analysis process. The approach involves collecting data from major IT companies—Google, Amazon, and Microsoft—to showcase how SASM contributes to online reputation management. Future work will concentrate on developing a hybrid model for more dependable and accurate results.

This paper is organized as follows: Section II presents sentiment analysis techniques and applications. Section III outlines the methodology and data management plans. Section IV provides a comprehensive description of SASM, including software requirements engineering, design, and architecture. Section V presents case study results and demonstrates SASM's contributions. Section VI discusses technical choices, methodologies, and future work.

## II. Related Work

In this section, we review relevant studies conducted by scholars, including Haruechaiyasak et al. [4], By et al. [5], and Huang et al. [6]. These studies analyze social media content and its impact on corporate digital reputation and branding. Additionally, we examine their approaches for multi-platform data collection to identify an effective method for our proposed study.

### *A. Impact of Social Media on Corporate Digital Reputation and Branding*

Haruechaiyasak et al. [4] created a software framework named S-Sense to analyze Thai content. They gathered data from Twitter posts and Pantip, a Thai language website, concentrating on mobile services. The S-Sense framework was employed to test analysis modules, study components, and explore the influence of diverse lexicon sets on training models. The authors underscored social media's role in business reputation and branding, stressing the necessity for further research in this field.

Likewise, By et al. [5] explored public feelings toward Rai, an Italian public broadcaster, and compared it to a new private company, La7. They studied over 1000 Facebook posts using the Sentiment and Knowledge Mining System, iSyn Semantic Center, for sentiment analysis. The authors emphasized social media's role in brand perception and marketing, as user feedback on these platforms heavily affects buying choices. Analyzing such posts helps companies grasp trends, gauge customer loyalty, assess new product sentiments, and evaluate marketing success.

These academic works directly connect to the suggested SASM software model. To build on past research, the study will include various text analysis engines in the SASM model and gauge their effectiveness. Also, the model will integrate diverse social media platforms, like Twitter, Reddit, and Tumblr, for more comprehensive content collection and analysis. This broader integration will yield a larger dataset, leading to more meaningful results. Moreover, the potential growth of S-Sense, as advised by Haruechaiyasak et al. [4], aligns with the proposed SASM model features.

### *B. Sentiment Analysis of Content on Multiple Social Media Channels*

Past studies often concentrated on gathering data from just one platform like Twitter and Facebook [5] – [7]. This limited view prevented companies from fully understanding their product's performance. To broaden analysis, Ali et al. [8] expanded their research, collecting data from four platforms: Twitter, Reddit, Instagram, and news forums. They aimed to spot disease outbreaks through social media user sentiments. In crises, people share info on many platforms, allowing sentiment and timing analysis to reveal behaviors and locations [8]. Each platform has its own community and sharing style. Collecting data from various platforms offers an unbiased disease outbreak view. This mirrors our study, which focused on sentiments about tech giants Google, Amazon, and Microsoft. These firms are key tech players. Like Ali et al., our research underscores the importance of gathering data from diverse social media platforms. Hence, we suggest integrating multiple media channels into the SASM software model.

### *C. Advantages of Proprietary Text Analytics Engines*

Chopra et al. [9] conducted a thorough study where they identified and analyzed features of 55 text mining tools designed to aid sentiment and linguistic analysis researchers. The tools were grouped as Proprietary, Open Source, or Online. Among them, 39 were Proprietary, 13 were Open Source, and 3 were Online. Proprietary tools, owned by companies, offered services like text and temporal analysis, content recognition, and text attribution. Most favored tools (42%) utilized natural language processing and/or machine learning algorithms.

Likewise, Atique et al. [10] surveyed recent articles in Sentiment Analysis. Their paper gives an overview of techniques used, assisting researchers in method selection. Their findings showed a focus on machine learning over lexicon-based approaches.

These studies strongly support using proprietary text mining tools like MTA [11], IWNLU [12], and GCNLA [13] in our proposed study. These tools use machine learning for advanced text analysis. Most text mining tools are proprietary [9], and researchers often favor machine learning [10]. These tools offer features like text analytics, classification, and sentiment analysis.

Additionally, Trupthi et al. [7] created an interactive app using Naïve Bayes for real-time Tweet sentiment analysis. The app's dashboard displays sentiment results based on user-defined filters (keywords). Though currently English-only [7], it aims for language expansion. For broader language support in our study, MTA, IWNLU, and GCNLA are chosen. They work with various languages like Chinese, English, French, German, Hindi, Italian, Japanese, Korean, Portuguese, Spanish, and more [11]– [13].

### *D. Evaluation of Multiple Sentiment Analysis Models*

In Praciano et al.'s study [14], they examined spatiotemporal trends in the Brazilian election using sentiment analysis. They combined NLP toolkits TextBlob and OpLexicon with Sentilex. Cutting-edge machine learning algorithms - Support Vector Machine (SVM), Naïve Bayes, Decision trees, and logistic regression - were used to classify sentiment in texts. Algorithm performance was compared using both toolkits, considering metrics like accuracy, precision, recall, and F1 score. Validation was done against the 2014 Brazilian presidential election data in the Superior Electoral Court database. The researchers' model effectively predicted results, with SVM achieving around 90% accuracy.

Unlike prior studies that often used one classifier [1], Praciano et al. [14] compared four machine learning algorithms and two NLP toolkits to find the best classifier. This aligns with our study's goal to enhance SASM's sentiment analysis by testing different text analytics engines. The aim is to assess the integration of these engines and potentially evaluate and compare their performance.

## III. Methodology and Data

In our study, we've developed SASM (Social Media and Digital Reputation Management), an open-source tool for sentiment analysis. SASM is designed for managing online reputation through social media. It gathers information from Twitter, Reddit, and Tumblr. To use these platforms' data, you'll need developer accounts and API keys, which are integrated into the SASM code.

Furthermore, the program requires setup to ensure accurate data collection due to different text length rules on each platform. Once configured, it searches these platforms' databases with specific keywords to find relevant posts. Data is collected concurrently from all three channels, allowing comparisons and trend observations about public sentiment over a set time.

SASM employs filters, aggregation, and analysis techniques to monitor sentiment trends in social media content. It utilizes three sentiment analysis engines and shows results on a user-friendly dashboard. Pie charts illustrate relative engine performance comparisons across platforms.

In our case study, we concentrate on collecting data related to layoffs at Google, Amazon, and Microsoft. We'll gather posts about these companies from three social media channels. Using the three proprietary sentiment analysis engines, we'll evaluate the sentiment of the gathered data. By comparing engine outcomes, we'll identify the most effective combination for sentiment analysis in this context.

### A. Case Study: Big Tech Layoffs

The technology industry has faced notable challenges recently, affecting major companies like Amazon, Microsoft, and Google through layoffs in 2022 and 2023. As of December 2022, TrueUp's tech layoff tracker reported 1,405 layoff rounds impacting 219,959 individuals in tech worldwide [15] [16]. Factors like the COVID-19 pandemic, inflation, and rising interest rates have led to these job reductions [17]. Considering the current economic conditions, it's likely that such layoffs will continue [18].

Given the importance of these layoffs and the wealth of related data, our case study focuses on Google, Amazon, and Microsoft's layoffs. We used our sentiment analysis software, SASM, to gather, analyze, and assess hundreds of posts from Twitter, Reddit, and Tumblr regarding these layoffs. In the next sections, we'll briefly introduce each company under study and the social media platforms—Twitter, Reddit, and Tumblr—that we used.

*1) Google:* In January 2023, Google revealed plans to cut around 12,000 jobs. This was part of a bigger restructure to align their product areas and roles with their core priorities as a company [19]. To aid people in this challenging time, Google pledged for U.S. employees: payment throughout the notification period (minimum 60 days), a severance package starting at 16 weeks' salary plus two weeks for every additional year at Google, and speeding up at least 16 weeks of GSU vesting. They also committed to paying all 2022 bonuses and unused vacation time, offering 6 months of healthcare, job placement services, and immigration assistance for those affected. For non-U.S. employees, support would follow local rules [20].

*2) Amazon:* In November 2022, Amazon cut about 10,000 jobs. These employees were from various departments including Alexa, Amazon's Luna cloud gaming service, hardware, services, human resources, and retail [21]. An internal memo mentioned that the aim of these layoffs was to reduce costs in order to keep investing in Amazon's customer-favored features like broad selection, competitive prices, and quick shipping [22]. According to the same memo, U.S. workers would have a "60-day non-working transitional period with full pay and benefits, plus additional several weeks of severance depending on the length of time with the company, a separation payment, transitional benefits, and external job placement support" [22].

*3) Microsoft:* In January 2023, Microsoft revealed its layoff plan of 10,000 employees in response to economic challenges and rising interest rates. Less than 5% of the total workforce was affected by these job cuts, which concluded in March 2023, the third fiscal quarter of that year [23]. These layoffs impacted workers from HoloLens and Microsoft Edge, as well as major game studios under Microsoft, 343 Industries and Bethesda [24]. Microsoft pledged to handle the process transparently, offering 60 days' notice, generous severance pay, six months of healthcare coverage, ongoing stock vesting, and career transition support for the affected employees [25].

### B. Social Media Data Source

Data was collected from three primary social media channels: Twitter, Reddit, and Tumblr

*1) Twitter:* Twitter is a top platform to gather opinions or sentiments about a topic from a large group [1]. It generates about 500 million Tweets daily, making it one of the biggest social media sites [26]. With its vast user base and real-time data, Twitter is ideal for sentiment analysis. Users span various demographics, making their posts a true reflection of society [27]. The Twitter API V2 offers a recent search feature for Tweets from the past week matching a search query. In this study, the app searches for a specific keyword and collects related recent Tweets [28].

*2) Reddit:* Reddit is a popular platform that emphasizes community engagement [29]. In 2022, it had around 50 million daily users. It ranked tenth among U.S. social networking sites in April 2021 [30], and fifteenth globally [31]. Reddit consists of subreddits, which are dedicated discussion boards for specific topics. Users can join or create subreddits aligned with their interests [29]. Unlike other platforms with data extraction limits, researchers can use the Pushshift Reddit Dataset to access and analyze billions of Reddit submissions and comments. Pushshift, a data archive and analysis platform, has gathered Reddit data since 2015. It offers an API for real-time exploration and aggregation [32]. In this study, user comments from all subreddits are taken within a 30-day period using the API, focusing on a chosen search keyword.

*3) Tumblr:* Tumblr, founded by David Karp in 2007 and currently owned by Automattic, is a microblogging and social networking site [33]. Users can create short-form blogs called tumblogs on this platform to share multimedia content and engage with their audience. While tumblogging is underexplored in current research, it holds promise for understanding sentiments about certain topics. The official Tumblr API has a \tagged endpoint that helps retrieve posts connected to a specific tag [34]. Tags are essential for readers to find posts on a user's blog related to a specific topic [35]. In this study, we collect and analyze Tumblr posts with tags matching the chosen search keyword to gain valuable insights.

## IV. Sentiment Analysis of Social Media (SASM)

SASM is an open-source, cloud-based digital reputation management software solution. The application collects user posts from social media, performs sentiment analysis, and provides meaningful information to the end-user to aid them in evaluating and assessing their corporation's, product's, or service's online reputation.

### A. Requirements Engineering

The SASM system has been meticulously crafted to fulfill distinct requirements, primarily focusing on supporting marketing, reputation, and branding management activities [36]. Figure 1 provides a comprehensive overview of the system's use case diagram, illustrating the key actors and primary processes involved.

In this diagram, two crucial actors have been identified: (1) *User* and (2) *Social-Media API*. The User actor engages with SASM to specify the desired search term, effectively initiating the data retrieval process. On the other hand, the Social-Media API actor plays a pivotal role by providing data feeds and facilitating the retrieval functionalities required by SASM. Incorporating the Social-Media API enables SASM to seamlessly and concurrently integrate various social media channels, ensuring its versatility and adaptability.

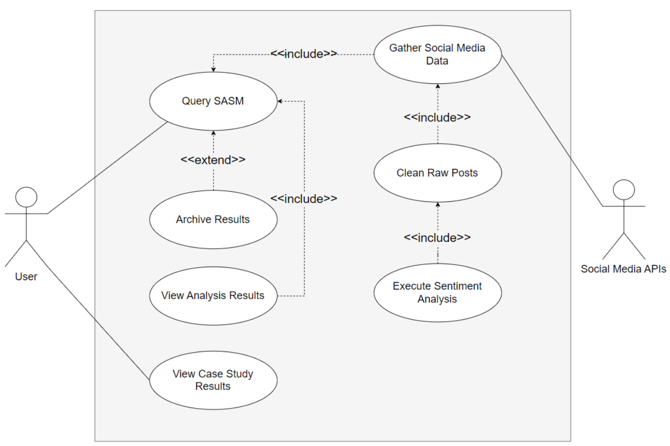



Fig. 1: Sentiment Analysis of Social Media (SASM) Use Case Diagram

According to the software development life cycle and industry best practices, requirements play a vital role in capturing the essential characteristics that a system must possess to fulfill the needs of its stakeholders [37] [38].

The following is a selection of functional requirements:

- **[FR1]** The system shall allow the user to input a search keyword (maximum 512 characters) into the search bar and click on submit. The system will then display relevant graphs comparing and contrasting the public sentiment associated with the keyword on the aforementioned social media platforms;
- **[FR2]** The system shall allow the user to view the results of the searched keyword and configure the charts to extract the necessary information;
- **[FR3]** The system shall allow the user to view the results of the case study and configure the charts to extract the necessary information;

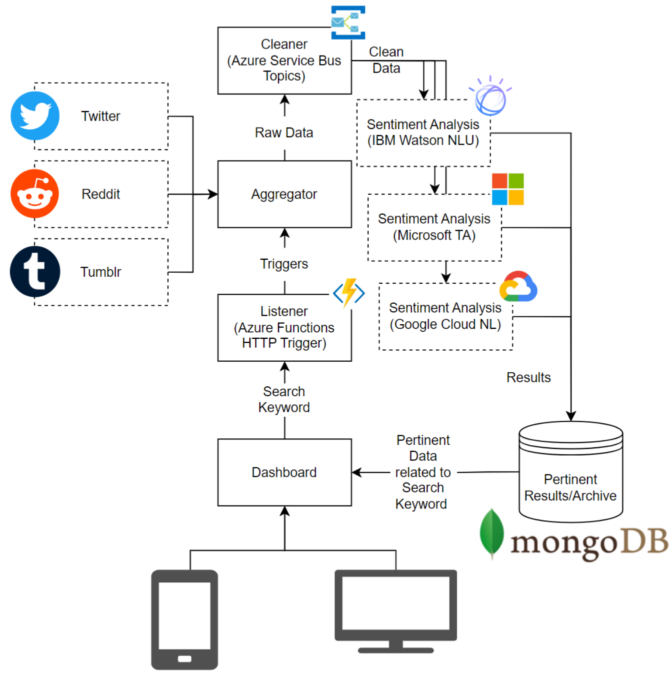



Fig. 2: SASM Software Architectural Overview

- **[FR4]** The system shall allow the user to view the top 10 posts for each social media platform in a tabular format.

The above-listed functional requirements have been analyzed and validated with stakeholders and the following set of quality attributes (non-functional requirements) has been derived:

- **[NFR1]** Availability: The system shall be available 24/7/365;
- **[NFR2]** Operability: The system shall be capable of communicating and retrieving data from common and well-established social-media platforms: Twitter [39], Reddit [40] and Tumblr [35].
- **[NFR3]** Storage: The system shall store the results of each unique searched keyword in separate data stores and therefore create a collection of archived data;
- **[NFR4]** Accessibility: The system shall be integrated into the Laboratory for Applied Software Engineering Research (LASER) website, and it should support all browsers. Additionally, all material presented by the system must meet the Web Content Accessibility Guidelines.

### B. Software Architecture and Design

The SASM system utilizes a client-server architecture, where communication between the client and server is facilitated through a RESTful API. Figure 2 provides an overview of the interactions among the client, server, and the various modules within the server. The subsequent subsections delve into the specific functionalities of each component.

*1) Dashboard:* To initiate a search in the SASM Dashboard, the user first navigates to the Home Page. On this page, they enter a search keyword into the designated input field and then proceed by clicking the submit button. Subsequently, an HTTP POST request is employed to transmit the search keyword to the Listener Module.

*2) Listener Module:* The Listener Module in SASM is integrated with an Azure Functions HTTP Trigger. This HTTP Trigger acts as a mechanism to invoke the Listener Module upon receiving the HTTP POST request that contains the search keyword. Once the Listener Module is invoked, it passes the received search keyword to the Aggregator Module for further processing.

*3) Aggregator Module:* The Aggregator Module plays a crucial role in SASM by handling the task of querying and retrieving data related to the search keyword from the three designated social media platforms: Twitter, Reddit, and Tumblr. To accomplish this, the Aggregator Module utilizes the respective APIs of these platforms. Once the data is collected, it is then published to an Azure Service Bus Topic named *General Cleaner*. This step ensures the organized flow of data to the subsequent cleaning process.

*4) Cleaner Module:* The Cleaner Module plays a critical role in SASM by subscribing to the *General Cleaner* topic and retrieving data from the Azure Service Bus. This data undergoes a comprehensive cleaning process involving multiple steps:

- *Data Preprocessing:* This step involves removing usernames, hashtags, URLs, newlines, punctuation, numbers, and stop words from the text [41]. Stop words, which are common words that do not contribute significant meaning to the specific topic, are eliminated to focus on more relevant and infrequent words that provide context [42].
- *Tokenization:* Tokenization is the process of breaking down the text into individual words or tokens [41]. This step enables further analysis and processing at the word level.
- *Lemmatization:* Lemmatization involves identifying and collecting the inflected forms of a word to represent them as a single unit, known as the word's lemma or its vocabulary form. By converting words to their base form or root, lemmatization helps in standardizing and unifying word variations [43].

*5) Sentiment Analysis Module:* The Sentiment Analysis module plays a crucial role in SASM by receiving a list of cleaned posts. Each post is then passed through individual sentiment analysis engines, namely MTA, GCNLA, and IWNLU, utilizing their respective APIs. These engines provide sentiment values associated with each post, which are converted to numerical scores ranging from -1 to +1:

- Values close to -1 indicate a negative sentiment.
- Values close to 0 represent a neutral sentiment.
- Values close to +1 indicate a positive sentiment.

The resulting sentiment analysis data is stored in a cloud-based MongoDB Atlas database [44]. To visualize and explore these results, the SASM dashboard retrieves the data and presents it through various informative infographics. Figures 3 and 4 showcase examples of the dashboard displaying the sentiment analysis outcomes.

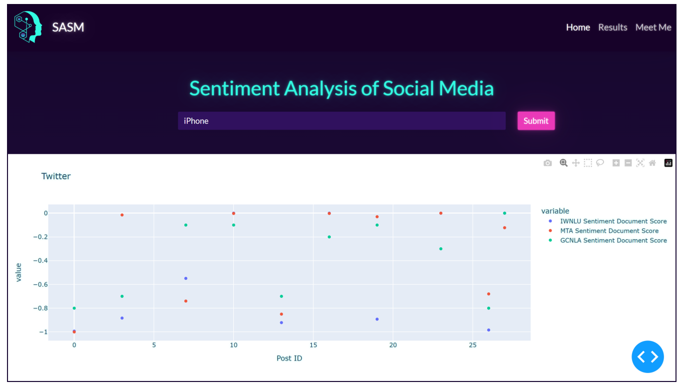



Fig. 3: Home Page: Sentiment Analysis of Posts on Twitter

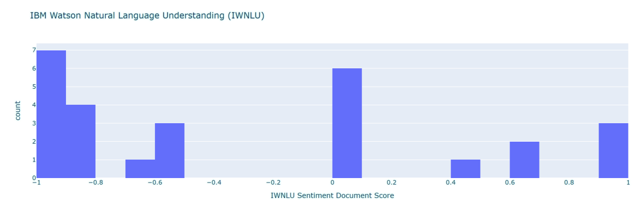



Fig. 4: Frequency Distribution of the Sentiment Value for Posts analyzed by IWNLU

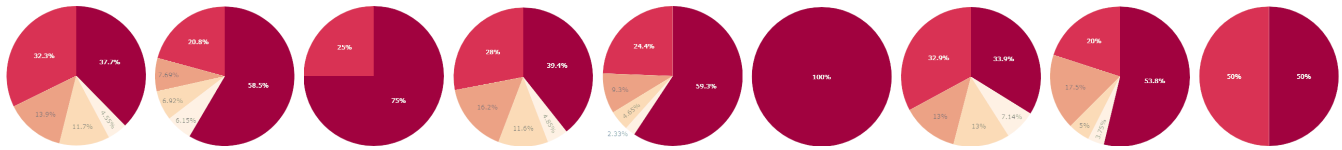
By leveraging sentiment analysis engines, storing the results in a reliable database, and presenting them through an intuitive dashboard, SASM provides valuable insights into the sentiment of social media posts, empowering users to make informed decisions and gain a comprehensive understanding of public opinion.

In Figure 3, a scatter plot showcases the sentiment values of recent Twitter posts retrieved using the search keyword ”iPhone.” The SASM dashboard also generates similar scatter plots for Reddit and Tumblr, providing a distinct understanding of sentiment on each social media platform. These plots offer insights into the overall sentiment of users on a specific platform regarding the search keyword. Additionally, users can compare the sentiment values detected by each sentiment analysis engine (IWNLU, GCNLA, and MTA) for individual posts, enabling a comprehensive analysis.

Lastly, Figure 4 presents a frequency distribution plot for the sentiment values generated by the IWNLU engine. In the SASM dashboard, users can also access frequency distribution plots for GCNLA and MTA. The plot illustrates the cumulative count of posts across all social media platforms, providing an overview of sentiment distribution.

## V. Results

To compare and assess the performance of the three sentiment analysis engines - GCNLA, MTA, and IWNLU - we conducted a case study on the major IT companies' layoffs in 2022-2023: Google, Amazon, and Microsoft. Using a Powershell Task Scheduler, we gathered posts over a 3-day span from January 31st to February 2nd, 2023. SASM



(a) Goog-Twi (b) Goog-Red (c) Goog-Tum (d) Amzn-Twi (e) Amzn-Red (f) Amzn-Red (g) Msft-Twi (h) Msft-Red (i) Msft-Tum

Fig. 5: Sentiment Analysis of big Tech layoffs social media posts collected from Twitter, Reddit, and Tumblr

was used to fetch posts with the keywords ”google layoffs,” ”amazon layoffs,” and ”microsoft layoffs.” A total of 1607 posts were collected, but API limits prevented more collection. Specifically, we gathered 653 ”Google” and ”layoffs” posts, 501 ”Amazon” and ”layoffs” posts, and 453 ”Microsoft” and ”layoffs” posts across the three platforms. For visualization, we employed pie charts. Each chart displays the portion of posts with closely matched sentiment values (+/- 0.1) among specific combinations of sentiment analysis engines. These charts offer insights into how much agreement the sentiment analysis engines have in the given combinations. For instance, Figure 5 shows the percentage of posts collected from each social media platform:

- 38.3% of the posts had sentiment values that were not within +/- 0.1 of each other for all combinations of engines: MTA & GCNLA, MTA & IWNLU, IWNLU & GCNLA;
- 33.7% of the posts had sentiment values that were within +- 0.1 of each other for only MTA and IWNLU engines;
- 12.4% of the posts had sentiment values that were within +- 0.1 of each other for only MTA and GCNLA engines;
- 10.6% of the posts had sentiment values that were within +- 0.1 of each other for only GCNLA and IWNLU engines;
- 5.03% of the posts had sentiment values that were within +- 0.1 of each other for all combinations of engines: MTA & GCNLA, MTA & IWNLU, IWNLU & GCNLA.

The aforementioned process was repeated for each individual social media platform, namely Twitter, Reddit, and Tumblr. The results obtained for each platform are displayed in the second, third, and fourth pie charts in Figure 5, respectively. The purpose of this analysis was to evaluate and analyze trends across different social media platforms. It was observed that, in general, for pairs of sentiment analysis engines, MTA and IWNLU exhibited the highest percentage of posts with sentiment values in agreement (+-0.1 of each other). This trend was followed by MTA and GCNLA, and lastly, IWNLU and GCNLA.

Furthermore, it was interesting to note that this trend was consistent across all search keywords. Figure 5 depict the corresponding pie charts for Amazon and Microsoft layoffs, respectively. The charts demonstrate that the decreasing order of sentiment analysis engine pairs, in terms of the number of posts with similar sentiment values, remained consistent: MTA & IWNLU, MTA & GCNLA, and IWNLU & GCNLA.

## VI. Discussion & Conclusion

The research project encountered several limitations during the data collection and analysis process. One significant limitation was the MTA engine's inability to analyze more than 10 posts at a time, which restricted the total number of posts that could be analyzed in each iteration. Additionally, Tumblr, as a data source, proved to be unreliable due to its image-based nature and the limited availability of posts related to technology and layoffs. Moreover, the inability to verify the sentiment analysis engine's results necessitated reliance on proprietary tools for analysis. The results could only be visually verified by observing the proximity of the dots in the scatter plots presented on the Home Page. Another limitation was the exclusion of spatiotemporal analysis from the project scope. While Twitter provided spatiotemporal data, Tumblr and Reddit did not offer this feature. Collecting spatiotemporal data from all three channels was not feasible, considering the data collection from multiple sources. Furthermore, the application had limitations in search accuracy, occasionally retrieving posts unrelated to the search query.

To validate the effectiveness of SASM, we conducted a case study focusing on the topic of layoffs at Google, Amazon, and Microsoft, utilizing data from Twitter, Reddit, and Tumblr. The primary objective was to assess the feasibility of developing a platform that encompasses multiple channels and engines. Additionally, we aimed to compare, contrast, and evaluate the performance of three sentiment analysis engines: GCNLA, IWNLU, and MTA.

Future research should focus on developing improved applications capable of effectively collecting data from multiple languages. Efforts should be made to overcome the limitations of API calls to enhance the quality and quantity of data collected for sentiment analysis. Additionally, conducting a spatiotemporal analysis of the collected data would provide valuable insights for companies to assess their digital presence across different locations and allocate resources accordingly. Lastly, considering the use of different and diverse social media platforms such as Meta should be explored for data collection purposes.

## References

[1] A. J. Nair, G. Veena, and A. Vinayak, “Comparative study of twitter sentiment on covid-19 tweets,” in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, 2021, pp. 1773–1778.

[2] R. Panikar, R. Bhavsar, and B. Pawar, “Sentiment analysis: a cognitive perspective,” in *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, vol. 1. IEEE, 2022, pp. 1258–1262.

[3] F. Samara, S. Ondieki, A. M. Hossain, and M. Mekni, "Online social network interactions (osni): a novel online reputation management solution," in *2021 International Conference on Engineering and Emerging Technologies (ICEET)*. IEEE, 2021, pp. 1–6.
[4] C. Haruechaiyasak, A. Kongthon, P. Palingoon, and K. Trakultaweekoon, "S-sense: A sentiment analysis framework for social media monitoring applications," *Information Technology Journal*, vol. 14, no. 1, pp. 11–22, 2018.
[5] L. Yue, W. Chen, X. Li, W. Zuo, and M. Yin, "A survey of sentiment analysis in social media," *Knowledge and Information Systems*, vol. 60, pp. 617–663, 2019.
[6] Y.-P. Huang, N. Hlongwane, and L.-J. Kao, "Using sentiment analysis to determine users' likes on twitter," in *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*. IEEE, 2018, pp. 1068–1073.
[7] M. Trupthi, S. Pabboju, and G. Narasimha, "Sentiment analysis on twitter using streaming api," in *2017 IEEE 7th international advance computing conference (IACC)*. IEEE, 2017, pp. 915–919.
[8] K. Ali, H. Dong, A. Bouguettaya, A. Erradi, and R. Hadjidj, "Sentiment analysis as a service: a social media based sentiment analysis framework," in *2017 IEEE international conference on web services (ICWS)*. IEEE, 2017, pp. 660–667.
[9] A. Kaur and D. Chopra, "Comparison of text mining tools," in *2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*. IEEE, 2016, pp. 186–192.
[10] H. T. Phan, N. T. Nguyen, and D. Hwang, "Sentiment analysis for social media: A survey," *Journal of Computer Science and Cybernetics*, vol. 37, no. 4, pp. 403–428, 2021.
[11] S. Pande, "An overview of sentiment analysis using azure cognitive services," *International Journal of Management IT and Engineering*, vol. 9, no. 10, pp. 123–127, 2019.
[12] S. Vergara, M. El-Khouly, M. El Tantawi, S. Marla, L. Sri *et al.*, *Building cognitive applications with IBM Watson services: volume 7 natural language understanding*. IBM Redbooks, 2017.
[13] E. Bisong and E. Bisong, "Google automl: Cloud natural language processing," *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*, pp. 599–612, 2019.
[14] B. J. G. Praciano, J. P. C. L. da Costa, J. P. A. Maranhão, F. L. L. de Mendonça, R. T. de Sousa Júnior, and J. B. Prettz, "Spatio-temporal trend analysis of the brazilian elections based on twitter data," in *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2018, pp. 1355–1360.
[15] C. Trueman, "Tech layoffs in 2022: A timeline," *Computer World*, vol. 1, no. 12, pp. 65–166, 2022.
[16] O. Analytica, "Impact of us tech layoffs will be manageable," *Emerald Expert Briefings*, no. oxan-es, 2023.
[17] J. Lindzon, "Who is getting laid off in canada's tech sector? many firms making cuts that appear to affect women, minorities at disproportionate rates, layoffs. fyi data show." *Globe & Mail (Toronto, Canada)*, pp. B8–B8, 2023.
[18] M. N. Hossen, M. A. Mollah, N. S. Lipy, G. M. S. Hossain, and M. S. Rahman, "Factors affecting layoff in high-tech industry: Evidence from usa," *International Journal of Information, Business and Management*, vol. 15, no. 1, pp. 1–13, 2023.
[19] T. Ray, "Why we need to rethink our relationship with big tech," 2023.
[20] R. Savitha and V. B. Harshini, "“a study on massive layoffs with reference to e-commerce companies"–an empirical study," *Journal of Survey in Fisheries Sciences*, vol. 10, no. 4S, pp. 1783–1793, 2023.
[21] S. Visvanathan, "Shiv visvanathan writes on tech layoffs: The real sadness is our entrenched silence to a tragedy." *Money Control*, 2023.
[22] J. Lindzon, "Who is getting laid off in canada's tech sector? many firms making cuts that appear to affect women, minorities at disproportionate rates, layoffs. fyi data show." *Globe & Mail (Toronto, Canada)*, pp. B8–B8, 2023.
[23] P. Gaur, S. Vashistha, and P. Jha, "Twitter sentiment analysis using naive bayes-based machine learning technique," in *Sentiment Analysis and Deep Learning: Proceedings of ICSADL 2022*. Springer, 2023, pp. 367–376.
[24] C. S. Jayaramireddy, S. V. V. S. S. Naraharisetti, M. Nassar, and M. Mekni, "A survey of reinforcement learning toolkits for gaming: Applications, challenges and trends," in *Proceedings of the Future Technologies Conference (FTC) 2022, Volume 1*. Springer, 2022, pp. 165–184.
[25] M. Mekni, C. S. Jayaramireddy, and S. V. V. S. S. Naraharisetti, "Reinforcement learning toolkits for gaming: A comparative qualitative analysis," *Journal of Software Engineering and Applications*, vol. 15, no. 12, pp. 417–435, 2022.
[26] K. L. Pasala, C. S. Jayaramireddy, S. V. V. S. S. Naraharisetti, S. Atilho, B. Greenfield, B. Placzek, M. Nassar, and M. Mekni, "Smart parking system (sps): An intelligent image-processing based parking solution," in *Smart Energy for Smart Transport: Proceedings of the 6th Conference on Sustainable Urban Mobility, CSUM2022, August 31-September 2, 2022, Skiathos Island, Greece*. Springer, 2023, pp. 291–299.
[27] M. Vondráček, I. Baggili, P. Casey, and M. Mekni, "Rise of the metaverse's immersive virtual reality malware and the man-in-the-room attack & defenses," *Computers & Security*, vol. 127, p. 102923, 2023.
[28] J. Roesslein, "tweepy documentation," *Online] http://tweepy. readthedocs. io/en/v3*, vol. 5, p. 724, 2009.
[29] A. Akkaya and N. İlhan, "Sentiment analysis of the coronavirus vaccine on social media," in *2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*. IEEE, 2021, pp. 295–299.
[30] B. Auxier and M. Anderson, "Social media use in 2021," *Pew Research Center*, vol. 1, pp. 1–4, 2021.
[31] S. Kemp, "Digital 2021 april global statshot report. datareportal," 2021.
[32] J. Baumgartner, S. Zannettou, B. Keegan, M. Squire, and J. Blackburn, "The pushshift reddit dataset," in *Proceedings of the international AAAI conference on web and social media*, vol. 14, 2020, pp. 830–839.
[33] X. Yang, S. Feng, D. Wang, and Y. Zhang, "Image-text multimodal emotion classification via multi-view attentional network," *IEEE Transactions on Multimedia*, vol. 23, pp. 4014–4026, 2020.
[34] A. Bussoletti, "“tumblr is dominated by america:" a study of linguistic and cultural differences in tumblr transnational fandom," *The Communication Review*, vol. 26, no. 1, pp. 24–41, 2023.
[35] Tumblr, "Tumblr," Accessed 2023, May, 2023. [Online]. Available: https://www.tumblr.com/
[36] R. A. Khan, M. A. Akbar, S. Rafi, A. O. Almagrabi, and M. Alzahrani, "Evaluation of requirement engineering best practices for secure software development in gsd: An ism analysis," *Authorea Preprints*, 2023.
[37] K. M. Habibullah, G. Gay, and J. Horkoff, "Non-functional requirements for machine learning: Understanding current use and challenges among practitioners," *Requirements Engineering*, pp. 1–34, 2023.
[38] V. Taecharungroj, "“what can chatgpt do?" analyzing early reactions to the innovative ai chatbot on twitter," *Big Data and Cognitive Computing*, vol. 7, no. 1, p. 35, 2023.
[39] Twitter, "Twitter," Accessed 2023, May, 2023. [Online]. Available: https://twitter.com/
[40] Reddit, "Reddit," Accessed 2023, May, 2023. [Online]. Available: https://www.reddit.com/
[41] M. Desai, B. Gachhinakatti, P. Balaganur, Y. Rajeshwari, and L. Rathod, "Automatic text summarization in natural language processing," in *2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNWC)*. IEEE, 2021, pp. 1–5.
[42] N. Hardeniya, *NLTK essentials*. Packt Publishing, 2015.
[43] D. Khyani, B. Siddhartha, N. Niveditha, and B. Divya, "An interpretation of lemmatization and stemming in natural language processing," *Journal of University of Shanghai for Science and Technology*, vol. 22, no. 10, pp. 350–357, 2021.
[44] C. Huang, M. Cahill, A. Fekete, and U. Röhm, "Data consistency properties of document store as a service (dsaas): Using mongodb atlas as an example," in *Performance Evaluation and Benchmarking for the Era of Artificial Intelligence: 10th TPC Technology Conference, TPCTC 2018, Rio de Janeiro, Brazil, August 27–31, 2018, Revised Selected Papers 10*. Springer, 2019, pp. 126–139.