Sentiment Analysis of Social Media: A Case Study on Big Tech Layoffs

Mehdi Mekni, Ahmed Muntasir Hossain
Department of Electrical and Computer Engineering and Computer Science
Connecticut Institute of Technology
University *of* New Haven
West Haven, United States
e-mail: {mmekni | ahoss1}@newhaven.edu

***Abstract*—Digital reputation management systems are essential for maintaining and improving online reputations. However, current systems like Online Social Network Interactions face critical issues, such as limited effectiveness, high costs, and inaccuracy. Sentiment analysis, a natural language processing technique, can enhance digital reputation management by extracting opinions, emotions, and attitudes from textual data. We propose developing Sentiment Analysis of Social Media, an open-source, multi-channel, multi-engine sentiment analysis software. SASM collects data from Twitter, Reddit, and Tumblr, filtering and analyzing trends using Microsoft Text Analytics, IBM Watson Natural Language Understanding, and Google Cloud Natural Language API. A case study on Google, Amazon, and Microsoft will validate the system and evaluate the performance of the three engines. SASM offers a unique approach by providing reliable sentiment analysis, leveraging multiple engines, and sourcing diverse social media content, enabling companies to manage their digital reputation effectively and affordably.**

Keywords-Sentiment Analysis, Natural Language Processing, Social Media Platforms.

I. Introduction

Social media platforms serve as communication channels between companies and customers. Companies promote their products, and customers post reviews and ask questions. Analyzing customer-posted content is crucial, and sentiment analysis is effective due to the large data volume. Sentiment analysis uses Natural Language Processing (NLP) and text analysis tools to determine the sentiment of a text, categorized as positive, negative, or neutral [1]. Major applications include brand monitoring, campaign monitoring, and competitive analysis [2]. This helps companies evaluate their community presence and make informed decisions. However, companies struggle to collect and filter reviews effectively.

Sentiment analysis software analyzes direct reviews (website, emails, surveys), limiting exposure to social media data. Recently, Samara et al. [3] developed Online Social Network Interactions (OSNI) Analytics to collect and analyze Tweets about companies using Microsoft Text Analytics. The engine categorizes Tweets as negative, mixed, neutral, or positive and returns the results to the company. This research aims to integrate multiple social media channels (Twitter, Reddit, Tumblr) into a Sentiment Analysis of Social Media (SASM), enhancing its ability to analyze social media content. Additionally, the project plans to expand SASM's sentiment analysis capabilities by including Microsoft Text Analytics (MTA), IBM Watson Natural Language Understanding (IWNLU), and Google Cloud Natural Language API (GCNLA). The research question associated with our project is the following "*R1: How do MTA, IWNLU, and GCNLA differ in their sentiment analysis of social media content, and which engine provides the most accurate and nuanced sentiment classification when applied to data from various social media channels?*"

The goal of this project is to investigate and confirm the technical feasibility of integrating multiple sentiment analysis engines with social media channels, while also comparing their performance using a real-world case study. Data will be collected about Google, Amazon, and Microsoft, which are major IT contributors and recently experienced mass layoffs, sparking social media discussions. The project will use the Software Development Life Cycle (SDLC). Each social media platform offers a free connector for academia and research, allowing data retrieval. The data will be processed, categorized, and evaluated using sentiment analysis engines. Future work includes combining results from multiple engines to develop a hybrid model for more reliable and accurate results.

The rest of the paper is organized as follows: Section II presents sentiment analysis techniques and applications. Section III outlines the methodology, the studied use case and the associated 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. Sections VI and VII discuss technical choices, methodologies, and future work.

II. Related Work

In this section, relevant studies of scholars, such as Haruechaiyasak et al. [4], By et al. [5], Baron and Mekni [6], who analyzed social media content and its impact on corporate digital reputation and branding were reviewed. Moreover, their approaches for multi-platform data collection were analyzed to determine an effective method for the proposed study. The rest of the literature review is organized as follows: Impact of Social Media on Corporate Digital Reputation and Branding, Sentiment Analysis of Content on Multiple Social Media Channels, Advantages of Proprietary Text Analytics Engines, and Evaluation of Multiple Sentiment Analysis Models.

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

Haruechaiyasak et al. [4] developed a software framework, S-Sense, to analyze Thai content. They collect data from

 ISBN: 978-1-68558-212-8

Twitter and Pantip, a Thai language forum, about mobile services. S-Sense evaluates analysis modules and components and studies the impact of using different lexicon sets for model training. The authors highlight recent advancements in software frameworks for brand monitoring, campaign monitoring, and competitive analysis, emphasizing the significance of social media on corporate digital reputation and branding.

Similarly, By et al. [5] investigated over 1000 Facebook posts to gauge public sentiment towards Rai, an Italian public broadcasting service, compared to the private company La7. They use the Sentiment and Knowledge Mining System, iSyn Semantic Center, for data collection and sentiment analysis. The study underscores the role of social media in brand imaging and marketing, as user reviews and ratings on social platforms significantly influence purchase decisions. The research supports analyzing social media content by showing that Facebook posts indicate a positive sentiment towards La7, aligning with observations from Osservatorio di Pavia and Auditel. These institutions' data is used to validate the sentiment analysis results from iSyn Semantic Center.

These scholarly works relate directly to the application of the SASM software model. To improve upon previous research, the proposed study incorporates different text analytics engines in the SASM model and evaluates their performance. It also integrates multiple social media channels (Twitter, Reddit, and Tumblr) to enhance its ability to analyze social media content, resulting in a more statistically significant dataset. Lastly, Haruechaiyasak et al. [4] suggest expanding the S-Sense domain, which aligns with the proposed enhancements for the SASM model.

B. Sentiment Analysis of Content on Multiple Social Media Channels

In several previous studies, data is collected from a single platform, primarily Twitter and Facebook [5], [7]. This limits companies from obtaining a holistic analysis of their products. To address this, Ali et al.[8] collect data from four platforms: Twitter, Reddit, Instagram, and news forums. The research aims to identify disease outbreak locations using social media sentiment. During emergencies, people share information across platforms, and analyzing sentiment and spatiotemporal data reveals people's behavior and geographic locations [8]. Each platform has a unique community and information-sharing method. Collecting data from multiple platforms provides an unbiased overview of disease outbreaks. This aligns with the proposed study, which aims to identify sentiment towards three major technology companies: Google, Amazon, and Microsoft. Their research underscores the importance of multi-platform data collection. Thus, this paper suggests integrating several media channels into the SASM software model.

C. Evaluation of Multiple Sentiment Analysis Models

Praciano et al.'s [9] research analyzes spatiotemporal trends in the Brazilian election using NLP toolkits, TextBlob and OpLexicon combined with Sentilex, for sentiment analysis. They apply machine learning algorithms—Support Vector Machine (SVM), Naïve Bayes, Decision trees, and logistic regression—to classify text sentiment. They compare the performance of these algorithms using metrics like accuracy, precision, recall, and F1 score, validating results with 2014 Brazilian presidential election data from the Superior Electoral Court database. Their model effectively predicts election results, with SVM achieving the highest accuracy of about 90%.

This section reviewed relevant studies on the impact of social media on corporate digital reputation and branding. Existing works developed tools and frameworks for data collection and sentiment analysis across various platforms, emphasizing the importance of multi-platform analysis in understanding corporate branding. However, these studies often rely on single platforms, limiting the breadth of insights into brand sentiment. To address these limitations, the proposed study aims to incorporate multiple sentiment analysis engines and analyze data from Twitter, Reddit, and Tumblr, thus providing a more comprehensive dataset. Our project builds on these findings to enhance the SASM software model for more effective sentiment analysis.

III. METHODOLOGY AND DATA

In this study, we created an open-source, multi-channel, multi-engine sentiment analysis software for social media and digital reputation management purposes: SASM. The application collects data/posts from three different social media channels, Twitter, Reddit, and Tumblr. Initially, developer accounts would have to be set up, and API keys for the respective platforms would have to be requested. This would provide the model access to the social media platforms. These platforms' API keys would then be added to the code for the SASM model. Moreover, the program would need to be configured to ensure appropriate data collection as each platform has its own constraints regarding the length of their texts. After configuration, the program will be able to search all these platforms' databases for the particular keyword inputted and return relevant posts. The data will be collected from each of these social media channels simultaneously to compare and observe the trends in the opinions of people across a defined period. The software model then filters, aggregates, and analyzes trends in the sentiment of content posted on social media while leveraging the three sentiment analysis engines. The results of the analysis are then displayed on a dashboard. The performance of each engine on each of the social media platforms would be compared and their relative performance would be determined using pie-charts.

The proposed case study intends to collect data about layoffs from Google, Amazon, and Microsoft. The posts would be collected from the three social media channels. The three proprietary sentiment analysis engines would be used to determine the sentiment of the data collected. The relative agreement of these engines would be evaluated to determine the ideal combination.

A. Case Study

The technology industry has been hit hard by layoffs in 2022 and 2023, with some of the biggest names in the field, such as Amazon, Microsoft, and Google, all experiencing workforce reductions. Through the first week of December 2022, 219,959 people were affected by 1,405 rounds of layoffs at tech companies worldwide, according to TrueUp's tech layoff tracker [10]. Several factors contributed to the layoffs, including the COVID-19 pandemic, inflation, and increasing interest rates [11]. With the current state of the economy, we will likely continue to see these types of workforce reductions in the future [12]. Due to the relevance of the layoffs and the vast data surrounding the conversation, we collected data on Google, Amazon, and Microsoft layoffs for our case study. We used SASM to collect, analyze, and assess several hundred posts from Twitter, Reddit, and Tumblr related to the above-mentioned layoffs. In the following subsections, we will briefly introduce each company we studied as well as the social media platforms (Twitter, Reddit, and Tumblr) we adopted.

1) Google: In January 2023, Google announced a plan to lay off approximately 12000 employees. The layoffs were part of a larger restructuring effort to ensure that their product areas and roles align with their highest priorities as a company [13]. To support people during this difficult decision, according to a company statement, Google promised to pay United States (U.S.) employees during the full notification period (minimum 60 days), offer a severance package starting at 16 weeks salary plus two weeks for every additional year at Google, and accelerate at least 16 weeks of GSU vesting. Additionally, the company stated that they would pay all 2022 bonuses & remaining vacation time, and offer 6 months of healthcare, job placement services, and immigration support for those affected. Outside the U.S., employees would be supported in line with local practices and regulations [14].

2) Amazon: In November 2022, Amazon laid off approximately 10,000 employees. The impacted employees were working on Alexa and Amazon's Luna cloud gaming service. Additionally, Amazon's hardware and services, human resources, and retail teams were affected during the layoff [15]. According to a company memo, the job cuts were intended "to lower their cost to serve so that they could continue investing in the wide selection, low prices, and fast shipping that Amazon customers love" [16]. Furthermore, the memo stated that U.S. workers will be getting 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" [16].

3) Microsoft: In January 2023, Microsoft announced to layoff of 10,000 employees to address the turbulent economic times and rising interest rates. Less than 5% of the company's entire workforce was affected by the job cuts, which ended in the third fiscal quarter of that year, March 2023 [17]. The affected employees were working on HoloLens and Microsoft Edge; moreover, two major game studios under Microsoft, 343 Industries and Bethesda, were significantly impacted by the process [18]. Microsoft stated that they intend to be thoughtful and transparent throughout the whole process and provide 60 days' notice before termination, above-market severance pay, six months of healthcare coverage, ongoing stock vesting, and career transition support to impacted employees [19].

B. Social Media Data Source

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

1) Twitter: According to [1], Twitter is considered one of the best social media platforms through which the opinions/sentiments of a large group of people towards a particular topic can be obtained. Daily, Twitter sends out approximately 500 million Tweets causing it to become one of the largest social media platforms [20]. Due to its massive user base and high volume of real-time, publicly available data; Twitter is an ideal social media platform for sentiment analysis. Furthermore, Twitter users range from all age groups, socio-economic status, and demographics allowing the sentiment extracted from their posts to be an accurate representation of society [21]. The Twitter API V2 provides a recent search endpoint that returns Tweets from the last seven days that match a search query. In this study, the application provides the specified search keyword in the query and pulls all recent Tweets containing the keyword.

2) Reddit: Reddit is a popular social media platform, where the emphasis is on community rather than a single user [22]. In 2022, there was an average of 50 million daily active users. Recent Reddit statistics in April 2021 show that it was the tenth most-used social networking site in the United States [23] and the 15th most-used social platform worldwide [24]. It comprises a large number of communities known as subreddits. Each subreddit has a discussion board for a certain topic. Depending on their interests, users can create or subscribe to a variety of subreddits [22].

Social media platforms, such as Twitter, Instagram, and Meta have API restrictions that prevent applications from extracting data. However, using the Pushshift Reddit Dataset, social media researchers can easily query and analyze billions of submissions and comments on Reddit. "Pushshift is a social media data collection, analysis, and archiving platform that since 2015 has collected Reddit data and made it available to researchers". The Pushshift Reddit dataset is updated in real-time and provides an API to search, aggregate and perform exploratory analysis on the entirety of the dataset [25]. In this study, user comments are pulled from all subreddits within 30 days of the API call containing a specified search keyword.

3) Tumblr: Tumblr is a microblogging and social networking website founded by David Karp in 2007 that is currently owned by Automattic. The service allows users to post multimedia content to form short-form blogs known as tumblogs [26]. Tumblogging has not been used in current research and would be beneficial in obtaining a holistic understanding of the sentiment of people towards a particular topic.

The official Tumblr API provides a \tagged endpoint that allows the application to retrieve posts with a specific tag

[27]. Tags make it easier for readers to find posts about a specific topic on a user's blog [28]. In this study, Tumblr posts containing tags with the specified search keyword were retrieved.

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 designed to meet specific requirements that aim to support marketing, reputation and branding management activities [29]. Figure 1 details the system use case diagram actors and main processes. Two key actors have been identified; (1) User and (2) Social-Media API. While the first actor interacts with SASM to specify the search term, the second actor interacts with SASM providing data feeds and retrieval functionalities. The goal behind considering the Social-Media API is to allow SASM to integrate independently and concurrently different social media channels.

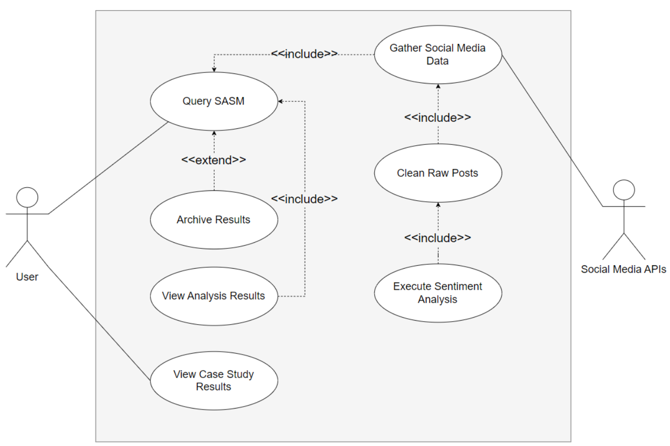

Figure 1: Sentiment Analysis of Social Media (SASM) Use Case Diagram

According to the software development life cycle and best practices, Requirements describe the characteristics that a system must have to meet the needs of the stakeholders. These requirements are typically divided into functional and non-functional requirements. Functional Requirements [FR] describe how software must behave and what are its features and functions [30]. Non-Functional Requirements [NFR] describe the general characteristics of a system [31] They are also known as quality attributes.

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;
- **[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 [32], Reddit [33] and Tumblr [34].
- **[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.

1) Scenarios: Figure 1 illustrates the use case diagram that outlines the necessary actions to fulfil the system requirements. Each user can take multiple paths within this use case. A scenario represents a specific path that a user takes while interacting with the system. It portrays a practical example of how the system is used by one or more users, outlining the steps, events, and actions that occur during the interaction. Usage scenarios can range from highly detailed, describing precisely how the user interacts with the interface, to moderately high-level, outlining the essential actions without specifying how they are performed. This section provides a detailed description of the usage scenarios depicted in the SASM system's system's use case diagram.

Figure 2 detail the process of inputting a search term and triggering the Listener. Figures 3 and 4 illustrate the process of gathering social media posts and performing data cleaning to prepare for the sentiment analysis process, detailed in Figure 5. Figure 6 depicts a scenario where the user inputs a previously used search term, resulting in an archive data store containing both the current and past searches. Lastly, Figure 7 outlines the visualization of the sentiment analysis results for the search term, and Figure 8 specifies the visualization of the sentiment analysis results for the case study.

Use case Name: Query SASM
Preconditions: The user must navigate to the Home page
Main Sequence:
1) User must input a search term/phrase
2) User must click on the submit button

Outcome:
1) Listener is triggered to Gather Social Media data

Figure 2: Query SASM Scenario

Use case Name: Gather Social Media Data
Preconditions: The user must query SASM
Main Sequence:
1) Listener retrieves the search term
2) Aggregator searches for posts related to the search term on Twitter, Reddit, and Tumblr
3) Aggregator retrieves posts and forwards them to the Cleaning service

Outcome:
1) Data for the search term is gathered.

Figure 3: Gather Social Media Data Scenario

Use case Name: Clean Raw Posts
Preconditions: A list of raw posts from Twitter, Reddit, and Tumblr, related to the search term, is available for processing
Main Sequence:
1) Cleaner receives a list of raw posts and removes unneeded stopwords
2) Cleaned posts is sent to Microsoft Text Analytics service, Google Cloud Natural Language service, and IBM Watson Natural Language Understanding service to obtain the sentiment value

Outcome:
1) Gathered Social Media Data is cleaned

Figure 4: Clean Raw Posts Scenario

Use case Name: Execute Sentiment Analysis
Preconditions: List of clean posts has been sent to the sentiment analysis engines
Main Sequence:
1) Engines produces a sentiment value for each of the posts
2) Sentiment values of each post are appended to their corresponding documents in a MongoDB collection

Outcome:
1) Each document in the MongoDB collection consists of the full post, cleaned post, social media platform, and the respective sentiment value from each engine.

Figure 5: Execute Sentiment Analysis Scenario

Use case Name: Archive Results
Preconditions: User specified a search term which was previously inputted
Main Sequence:
1) Results from the current search are stored in the pre-existing collection creating an archived data store
2) Results displayed to the user consist of the entire collection

Outcome:
1) An archived data collection is created for the specified search term

Figure 6: Archive Results Scenario

Use case Name: View Analysis Results
Preconditions: The user has navigated to the Home page, inputted a search term and clicked on the submit button
Main Sequence:
1) User is shown all analysis result reports that pertain to their search term
2) User filters their search for specific information, i.e. social media platforms, sentiment analysis engines, and more
3) Analysis results from the filter are displayed

Outcome:
1) User has viewed the gathered analysis results

Figure 7: View Analysis Results Scenario

Use case Name: View Case Study Results
Preconditions: User navigated to the Results page
Main Sequence:
1) User is shown all analysis result reports that pertain to the case study
2) User filters their search for specific information, i.e. information technology companies, social media platforms, sentiment analysis engines, and more
3) Analysis results from the filter are displayed

Outcome:
1) User has viewed the case study analysis results

Figure 8: View Case Study Results Scenario

B. Software Architecture and Design

SASM is implemented using a client-server architecture. The client and server communicate using a RESTful API. The interactions between the client, server, and all the modules contained in the server are depicted in Figure 9. The following subsections detail their functionalities:

1) Dashboard: A user navigates to the Home Page in the SASM Dashboard, provides a search keyword in the input field, and clicks on the submit button. The search keyword is sent to the Listener Module via a HTTP POST request.

2) Listener Module: The Listener Module is bounded to an Azure Functions HTTP Trigger. The HTTP Trigger invokes the Listener Module when the HTTP POST request containing the search keyword is received. The search keyword is then passed off to the Aggregator Module.

3) Aggregator Module: The Aggregator Module is responsible for querying and retrieving data containing the search keyword from the three social media platforms, Twitter, Reddit, and Tumblr using their respective APIs. The collected data is published to an Azure Service Bus Topic: *General Cleaner*.

4) Cleaner Module: The Cleaner Module subscribes to the *General Cleaner* topic and fetches data from the Service Bus. The data is then cleaned as follows:

- *Data Preprocessing:* Process of removing usernames, hashtags, URLs, newlines, punctuation, numbers, and stop words [35]. Stop words are ubiquitous terms in writing such as 'the', 'and', 'I', and so on that do not give insights into the specific topic of a document. During data preprocessing, these stop words are removed from the text to identify words that are more infrequent and potentially more pertinent to the context [36]. Additionally, HTML links (URLs), usernames, hashtags, and embedding are removed as they do not add value to the data [35]. Data preprocessing is an integral step in ensuring the text is

 ISBN: 978-1-68558-212-8

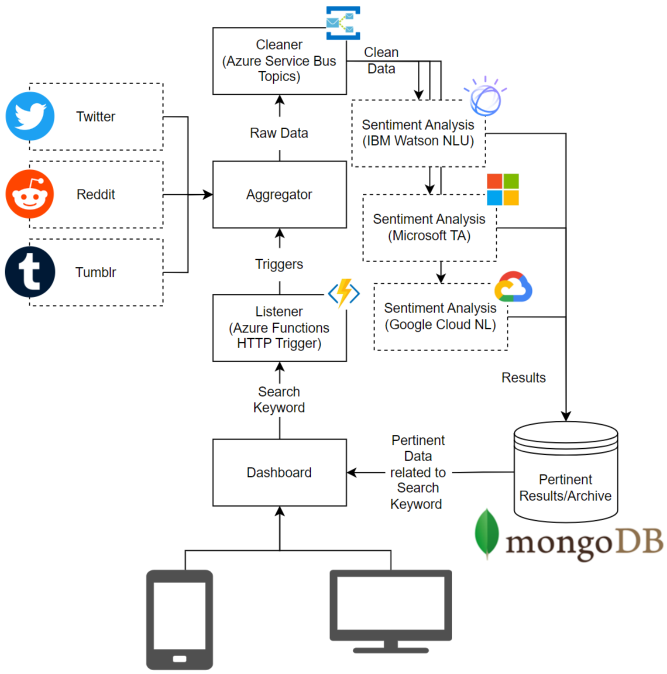

Figure 9: SASM Software Architectural Overview

cleaned and relevant data is extracted and formatted for analysis.

- *Tokenization:* Process of splitting text to individual words [35].
- *Lemmatization:* Process of collecting the inflected elements of a word so that they may be identified as a single unit, called the word's lemma or its vocabulary form. The algorithm converts words to their basic form or root [37]. In Python programming, these tasks can be carried out by NLTK built-in library functions [38].

5) Sentiment Analysis Module: The Sentiment Analysis module receives a list of clean posts. Each post is sent to the individual sentiment analysis engines, MTA, GCNLA, and IWNLU, using their respective APIs. The engines then return the associated sentiment value which is converted to a number between -1 to +1:

- Values near negative 1 have a negative sentiment.
- Values near 0 have a neutral sentiment.
- Values near positive 1 have a positive sentiment.

The results are stored in a cloud-based MongoDB Atlas database [39], and the SASM dashboard retrieves the data and displays it using several infographics. The results from the dashboard can be seen in Figures 10, 11, and 12.

In Figure 10, a scatter plot displays the sentiment values for the most recent posts that were retrieved from Twitter containing the search keyword, *iPhone*. In the SASM dashboard, similar scatter plots for Reddit and Tumblr are generated as well to provide an individual and unique understanding of each social media platform. The plots provides insight into the overall sentiment of people on a particular social media platform towards the search keyword. Moreover, for a specific

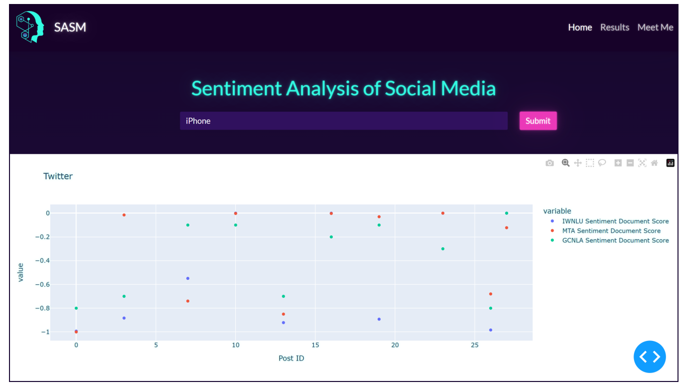

Figure 10: Home Page: Sentiment Analysis of Posts on Twitter

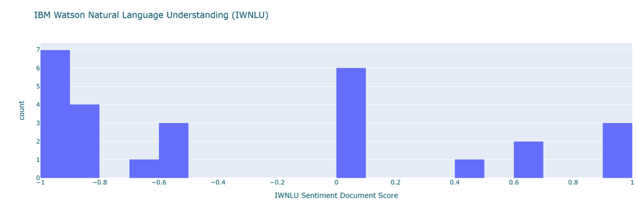

Figure 11: Frequency Distribution of the Sentiment Value for Posts analyzed by IWNLU

post, it allows the user to compare the sentiment value detected by each sentiment analysis engine: IWNLU, GCNLA, and MTA.

Figure 11 displays the frequency distribution of the sentiment values generated by the IWNLU engine. In the SASM dashboard, frequency distribution plots for GCNLA and MTA are also available for users' to view. The count of the posts is cumulative and across all social media platforms.

Lastly, Figure 12 outlines the data collected in a tabular form. It consists of the post ID, full text of the post, social media platform, and the corresponding sentiment values from each engine. The user may filter the table based on a specific social media platform (Twitter, Reddit, and Tumblr) using the dropdown and display those particular posts.

Figure 12: Results for All Social Media Platforms in Tabular Form

V. Results

To compare, contrast, and evaluate the performance of the three sentiment analysis engines, GCNLA, MTA, and IWNLU, we used a relevant case study. Our study focused on the mass layoffs in 2022-2023 from three major information technology companies - Google, Amazon, and Microsoft. Using a Powershell Task Scheduler, posts were collected for 3 days, starting on January 31^{st} 2023 to February 2^{nd}, 2023. We used SASM to query and retrieve posts containing the following keywords: *google layoffs*, *amazon layoffs*, and *microsoft layoffs*. A total of 1607 posts were collected for the above mentioned keywords, but due to API limitations, more data could not be collected. 653 posts containing the keywords "google" and "layoffs"; 501 posts containing the keywords "amazon" and "layoffs"; and 453 posts containing the keywords "microsoft" and "layoffs" were collected from the three social media platforms. The collected data was then visualized using pie charts.

Each chart displays the percentage of posts that have sentiment values in agreement with each other (+- 0.1) for specific combinations of sentiment analysis engines. For instance, Figure 13 consists of 4 pie charts for the Google Layoffs collection. The first pie chart shows the percentage of posts collected from all social media platforms:

- 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 process was repeated for each social media platform, Twitter, Reddit, and Tumblr. Their results are displayed in the second, third, and fourth pie charts in Figure 13, respectively. This was done to evaluate and analyze trends across different social media platforms and it was observed that generally for pairs of sentiment analysis engines, MTA and IWNLU had the greatest percentage of posts where the sentiment values were in agreement (+-0.1 of each other). This was followed by MTA and GCNLA, and lastly, IWNLU and GCNLA.

It was also interesting to observe that the trend was apparent across all the search keywords. Figure 14 and Figure 15 display the corresponding pie charts for Amazon and Microsoft layoffs. It can be seen that the decreasing order of sentiment analysis engine pairs for the number of posts with similar sentiment values continue to be MTA & IWNLU, MTA & GCNLA, and IWNLU & GCNLA.

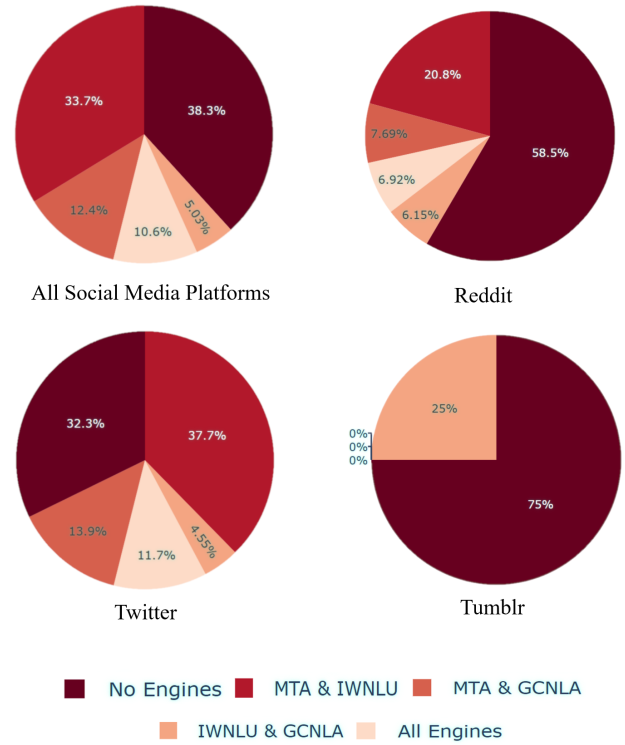

Figure 13: Google Layoffs

VI. Discussion

The research project faced several limitations in the data collection and analysis process. One major limitation was that the MTA engine could not analyze more than 10 posts at a time limiting the total number of posts that could be analyzed in each iteration. Additionally, Tumblr, which was included as a data source, was found to be an unreliable source of information due to it being image-based and the limitation in the number of posts that were available related to technology and layoffs. Furthermore, we were unable to verify the sentiment analysis engine's results and had to trust proprietary tools for analysis. We were only able to visually verify the results by observing the proximity of the dots in the scatter plots presented in the Home Page.

Another limitation was the inability to include spatiotemporal analysis in the scope of the project. This was due to the fact that while Twitter returned spatiotemporal data; Tumblr and Reddit did not. Since data was being collected from three different channels, extracting spatiotemporal data for all three was not feasible. Moreover, the application had a limitation in search accuracy, as it sometimes returned posts that were not related to the search query.

The project also encountered limitations in collecting data from other languages due to the occurrence of Unicode errors, which resulted in the application ignoring those posts. This highlights the need for a better application that takes multi-language into account for effective data collection.

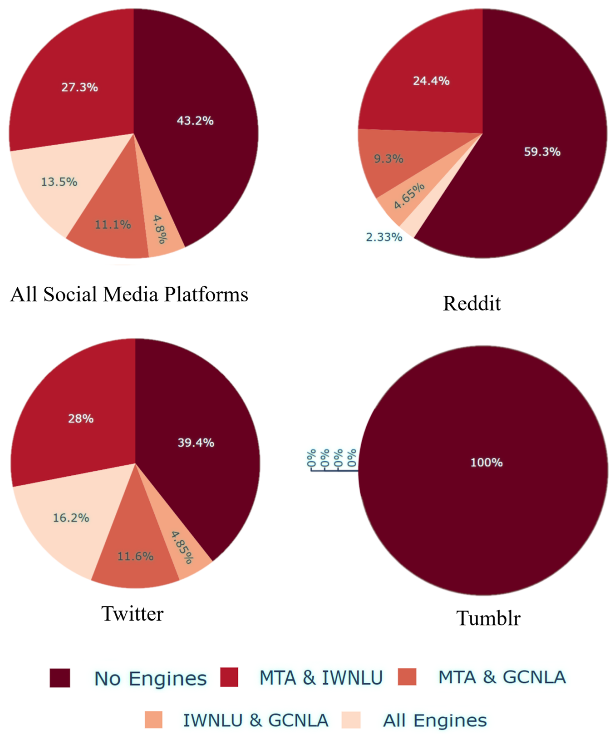

Figure 14: Amazon Layoffs

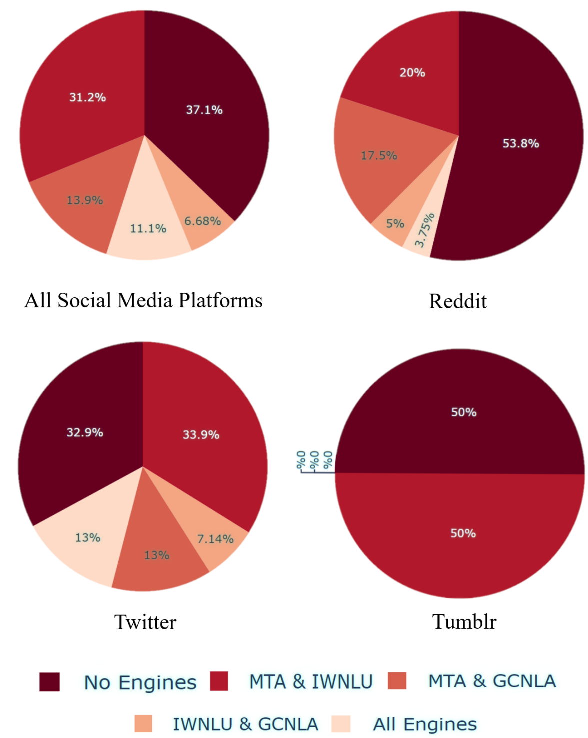

Figure 15: Microsoft Layoffs

Moreover, when the data is limited, these platforms might underperform because their models are typically trained on large, varied corpora and require diverse inputs to generalize well. A small data set may lack the nuance or range of expressions needed to test the models' abilities to handle different tones, contexts, or subtleties in sentiment.

In addition, sentiment analysis engines also have their limitations in analyzing data from different languages, with some engines supporting more languages than others. The 512-character search keyword limitation also poses a challenge in effectively searching for relevant data, as it may not be possible to include all relevant keywords within the character limit.

Moreover, the amount of data that can be pulled from the social media platform is limited by the API calls, which can also affect the quality of the data collected. There is also a limitation in the amount of data that can be passed to the sentiment analysis engines, which can affect the accuracy of the analysis.

VII. Conclusion and Future Work

In this paper, we introduced Sentiment Analysis of Social Media, an open-source, multi-channel, multi-engine sentiment analysis software for social media and digital reputation management purposes. SASM collects user posts from three social media platforms, performs sentiment analysis using three analytic engines, 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. To verify and validate the effectiveness of SASM, a case study was conducted to collect data about layoffs from Google, Amazon, and Microsoft from Twitter, Reddit, and Tumblr. The goal was to study the feasibility of developing a multi-channel & multi-engine platform and potentially comparing, contrasting, and evaluating the performance of the three sentiment analysis engines: GCNLA, IWNLU, and MTA.

Future research should focus on developing better applications that can effectively collect 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. Moreover, additional efforts will focus on expanding data collection to ensure a more comprehensive dataset. Larger datasets will support more fine-grained comparisons between different sentiment analysis engines will be essential to evaluate their performance across various contexts and use cases, allowing for more accurate and nuanced insights. Furthermore, performing a spatiotemporal analysis of the collected data would be interesting as it would allow companies to evaluate their digital presence in different locations and dedicate resources accordingly. Lastly, using a different and diverse social media platform such as Meta should be considered for data collection.

References

[1] A. J. Nair, G. Veena, and A. Vinayak, "Comparative study of twitter sentiment on covid-19 tweets," in *2021 5th Interna-*

tional 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)*, IEEE, vol. 1, 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 sensing," in *Proceedings of the IJCNLP 2013 Workshop on Natural Language Processing for Social Media (SocialNLP)*, 2013, pp. 6–13.

[5] F. Neri, C. Aliprandi, F. Capeci, and M. Cuadros, "Sentiment analysis on social media," in *2012 IEEE/ACM international conference on advances in social networks analysis and mining*, IEEE, 2012, pp. 919–926.

[6] C. R. Barone and M. Mekni, "Advancing continuous authentication using smart real-time user activity fingerprinting," in *2023 IEEE Third International Conference on Signal, Control and Communication (SCC)*, IEEE, 2023, pp. 01–06.

[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] B. J. G. Praciano *et al.*, "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.

[10] J. J. Watkins, ""i'm especially sorry to those impacted": Conventional visions of leadership in the 2022/2023 tech layoffs," 2024.

[11] Z. Saba, "Layoffs and corporate performance: Evidence based on the us tech industry," *Journal of Economics and Finance*, pp. 1–24, 2024.

[12] K. Vedantam, "Tech layoffs: Us companies with job cuts in 2022 and 2023," [Online], Crunchbase News, Apr. 2023, [Online]. Available: https://news.crunchbase.com/startups/tech-layoffs/.

[13] A. Satariano and N. Grant, "Google parent alphabet to cut 12,000 jobs," *The New York Times*, 2023, Online; accessed 20-Jan-2023.

[14] S. Pichai, *A difficult decision to set us up for the future*, Google, [Online]. Available: https://blog.google/inside-google/message-ceo/january-update/, Jan. 2023.

[15] J. Peters, "Amazon confirms cuts to hardware and services teams," *The Verge*, Nov. 2022, [Online]. Available: https://www.theverge.com/2022/11/16/23462439/amazon-layoffs-cuts-hardware-services-teams.

[16] M. Clark, "Amazon begins another round of job cuts as it lays off more than 18,000 people," *The Verge*, Jan. 2023, [Online].

[17] C. Thorbecke and H. Ziady, "Microsoft is laying off 10,000 employees," *CNN*, Jan. 2023, [Online].

[18] T. Warren, "Microsoft announces big layoffs that will affect 10,000 employees," *The Verge*, Jan. 2023, [Online]. Available: https://www.theverge.com/2023/1/18/23560315/microsoft-job-cuts-layoffs-2023-tech.

[19] Q.ai, "What microsoft's recent layoffs mean for the company and investors," *Forbes*, Jan. 2023, [Online].

[20] T. V. S. Krishna *et al.*, "A novel ensemble approach for twitter sentiment classification with ml and lstm algorithms for real-time tweets analysis," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 34, no. 3, pp. 1904–1914, 2024.

[21] Y. Lheureux, "Predictive insights: Leveraging twitter sentiments and machine learning for environmental, social and governance controversy prediction," *Journal of Computational Social Science*, vol. 7, no. 1, pp. 23–44, 2024.

[22] S. Nyawa, D. Tchuente, and S. Fosso-Wamba, "Covid-19 vaccine hesitancy: A social media analysis using deep learning," *Annals of operations research*, vol. 339, no. 1, pp. 477–515, 2024.

[23] J. Gottfried, "Americans' social media use," *Pew Research Center*, 2024.

[24] E. A. Radday, M. Mekni, L. Page, A. Sula, and L. Brown, "Evolving cybersecurity education: An analysis of the gen-cyber teacher academy's progression from 2022 to 2023 and beyond," *Journal of Computing Sciences in Colleges*, vol. 39, no. 8, pp. 113–127, 2024.

[25] S. Rani, K. Ahmed, and S. Subramani, "From posts to knowledge: Annotating a pandemic-era reddit dataset to navigate mental health narratives," *Applied Sciences*, vol. 14, no. 4, p. 1547, 2024.

[26] Q. Shi, W. Xu, and Z. Miao, "Image-text multimodal classification via cross-attention contextual transformer with modality-collaborative learning," *Journal of Electronic Imaging*, vol. 33, no. 4, pp. 043 042–043 042, 2024.

[27] Tumblr, *Tumblr api*, https://www.tumblr.com/docs/en/api/v2, [Online; accessed 22-Oct-2024].

[28] *Tagging your posts – help center*, https://help.tumblr.com/hc/en-us/articles/226161387-Tagging-your-posts, [Online], n.d.

[29] W. D. Schindel, "Requirements statements are transfer functions: An insight from model-based systems engineering," *INSIGHT*, vol. 27, no. 5, pp. 27–34, 2024.

[30] J. J. Shah and M. T. Rogers, "Functional requirements and conceptual design of the feature-based modelling system," *Computer-Aided Engineering Journal*, vol. 5, no. 1, pp. 9–15, 1988.

[31] M. Glinz, "On non-functional requirements," in *15th IEEE International Requirements Engineering Conference (RE 2007)*, IEEE, 2007, pp. 21–26. DOI: 10.1109/RE.2007.45.

[32] Twitter, *Twitter*, https://twitter.com/, [Online; accessed 22-Oct-2024], 2023.

[33] Reddit Inc, *Reddit inc*, https://www.reddit.com/, [Online; accessed October 2023], 2023.

[34] Tumblr, *Tumblr*, https://tumblr.com/, [Online], 2023.

[35] M. Mekni *et al.*, "Smart detection for heart health," in *Proceedings of the 2024 10th International Conference on Computing and Artificial Intelligence*, 2024, pp. 117–122.

[36] M. Mekni and D. Haynes, "Smart community health: A comprehensive community resource recommendation platform," in *HEALTHINF 2020 - 13th International Conference on Health Informatics, Proceedings; Part of 13th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2020*, SciTePress, vol. 11, 2020, pp. 614–624.

[37] D. Khyani and B. S. Siddhartha, "An interpretation of lemmatization and stemming in natural language processing," *Shanghai Ligong Daxue Xuebao/Journal of University of Shanghai for Science and Technology*, vol. 22, pp. 350–357, 2021.

[38] NLTK, *Nltk :: Natural language toolkit*, https://www.nltk.org/index.html, Accessed: 2024-10-22.

[39] MongoDB Inc, *Mongodb atlas*, https://www.mongodb.com/, [Online; accessed 22-Oct-2024], 2023.