Synthesis of Comments to Social Media Posts for Business Applications

Peter Adebowale Olujimi[1], Abejide Ade-Ibijola[2]
Department of Applied Information Systems-College of Business and Economics,
University of Johannesburg, Johannesburg, South Africa[1]
Research Group on Data-Artificial Intelligence-and-Innovations
for Digital Transformation-Johannesburg Business School,
University of Johannesburg, Johannesburg, South Africa[2]

***Abstract*—Responding to enormous comments on social media platforms is one major challenge facing businesses in recent times, especially when dealing with irate consumers. Customers have increasingly adopted social networks as a platform for expressing their concerns and posting comments on business pages, posing a great challenge for customer support agents and digital marketers alike. Analyzing and responding manually to these enormous comments is a time-consuming task, necessitating the adoption of Artificial Intelligence (AI) tool that can complete the task swiftly — *automatic comprehension of social media posts for comment generation*. In this paper, we present algorithms and a tool for the automatic comprehension of customer tweets and generation of responses to these tweets. This was done in two-fold: using existing Natural Language Processing (NLP) libraries to preprocess and tokenize these tweets, and secondly, using rule-based algorithms to find *a matching response* to each customer, based on the array of extracted tokens from the customer's tweet. This was built into a tool called `Comment-Synthesizer`. This tool takes unfiltered tweets as input, preprocesses the tweets, and matches the tweet with predefined responses using a rule-based algorithm with a success rate of 76%. This tool, if implemented in a desktop automation application, can be used to respond automatically to a large volume of customers' social media comments/posts.**

***Keywords*—Natural language comprehension; social media; natural language processing; customer engagements; artificial intelligence; comment generation**

I. Introduction

In recent years, social media has become an indispensable resource for companies to incorporate into their marketing plans [1]. Social media is described as a collection of Web applications that allow people to produce and generate content through the innovative use of Internet technology [2]. The widespread use of social networks by customers influences businesses that must join to connect with customers [3]. Businesses are embracing social media for an expanding number of marketing tasks, including image, market analysis, demand planning, provision of services, and commerce. Additionally, businesses have found that social networking websites, in particular, are crucial for drawing customers to their products and brands [4]. As a result of the recent explosion of interest in social networks like Facebook, Twitter, and LinkedIn, tons of businesses are posting information to influence consumers about their products/services [5]. The number of posts and comments received every day gives a surprising indication of the enormous influence of social networks, with Facebook accounting for one billion active users, YouTube with over four billion views, and nearly 140 million active users on Twitter [6], [7], [8]. As a means of communication, the usage of social networks is becoming significantly important, particularly because it enables firms to cultivate relationships with their clients [9]. Social networks can be used to distribute well-known e-Commerce solutions, but businesses that do not engage with their customers lose their potential [10].

In the early days of social media, PR firms would keep an eye on customer comments on a company's website to find and handle irate clients, but with the proliferation and volume use of social media platforms, this is no longer nearly adequate [6]. When a business adopts a Social Media strategy, the volume of postings increases quickly, requiring considerable human work for manual analysis [11]. Consequently, many businesses designate a human expert to constantly update their pages with relevant content, monitor comments, and respond to them on time. However, responding manually to these enormous comments is a time-consuming task, necessitating the adoption of Artificial Intelligence (AI) tool that can complete the task swiftly — *automatic comprehension of social media posts for comment generation*.

A long-standing aim of AI is the creation of programs that can understand and generate human language [12]. Natural Language Processing (NLP), an aspect of AI that deals with the interaction between machines and human language and has been around for more than four decades [13], is one of such programs. Machines are programmed to successfully process large natural language data. NLP makes it feasible for Intelligent machines to analyse the complexities of human expression in order to obtain information from numerous types of text, such as blogging, evaluation of products, as well as the numerous daily posts on Twitter, Facebook, and news feeds [14]. The digital representation and understanding of human language have recently attracted significant interest in this domain. Its applications have spread to numerous domains such as email spam surveillance, translation software, and retrieval of information, summarization, medicine, and question answering [15].

In this paper a set of rules was used to develop an algorithm that automatically comprehends social media comments and responds to them in real-time. This involves performing lexical analysis on the raw tweets with several preprocessing techniques, such as eliminating Twitter handles, URLs, whitespaces, etc. In addition, we performed sentiment analysis to identify the underlying emotions in the dataset, and then we defined rule-based algorithms to find *a matching response* to each customer, based on the array of extracted tokens from the customer's tweet. Fig. 1 depicts the processes.

The following is a summary of the significant contributions made in this paper.

1) leverage the Python Natural Language Toolkit (NLTK) for the analysis of the dataset. A tool that supports NLP data structures and algorithms in Python,
2) design a Natural Language Generation (NLG) algorithm system that uses a rule-based approach to automatically generate social media comments,
3) create `Comment-Synthesizer`, a software tool that deploys the techniques described in (3) above, and
4) present deployment of the tool, as well as results demonstrating the social media comments generated.

The remaining sections of this paper are organized as follows. Section II provides background and discusses the related work. Section III elaborates on the design concept used in this work. In Section IV, we demonstrate the implementation of the `Comment-Synthesizer` tool and the results obtained from the synthesized comments. In Section V, the results are evaluated, while Section VI provides a summary of the work and the further study.

II. Background and Related Work

This section examines works related to the problem domain and presents the various terms and techniques for solving the problem. It provides a definition of the terms and the motivation behind this work.

A. Social Media Marketing in the Digital Age

Social media is a dynamic and fast-moving domain across the globe. In the digital age, leveraging social media to advertise to a targeted demographic of potential customers is an effective strategy [16]. This allows a business to remain connected with customers and respond to their queries without delay. Every day, people around the world use the Web, social networks, smartphone applications, and other electronic communication systems [17]. Since the emergence of social networks, many businesses have realized that they can add value to a firm's entire business plan, particularly in terms of improving brand perception.

Businesses are leveraging alternative marketing strategies that are more cost-effective and beneficial for consumer engagement than conventional ways, as they recognize the potential of the Internet as a crucial component of their platform. They are now supporting open communication practices within the firm, such as tweeting, texting, and other methods, to promote branding outside the company [18]. Its no secret that platforms like Facebook, Instagram, and Twitter are essential tools for interacting with customers and spreading the word about a firm's products and services. Given how frequently consumers use social media for several hours each day, marketers have adopted it as a powerful marketing tool [19]. According to the statistics, the average internet user spends approximately 144 minutes every day exploring social networks. As a result, large corporations and their strategists employ social media to create brand images [20].

B. Social Media Automation

Digital marketing has metamorphosed significantly over the past two decades. Businesses are now expected to improve their strategies to earn customers' loyalty. The social web statistics indicate that automation of digital marketing applications, especially social media, is the key as this enables firms to make effective use of their time by concentrating on the more important side of the business. Social media automation can be described as an approach to remodelling business social media campaigns by utilising automation tools, such as Hootsuite, Buffer, Social Pilot, and the like [21]). Automating social media posts reduces the time spent on maintaining and growing brand accounts. Therefore, time and resources could be allocated to other areas of the marketing budget and to meet strategic goals [22]. A significant advantage of automation is the propensity to organize the company's entire social media strategy in advance. Content or posts can be scheduled on Twitter, Facebook, Instagram, and other social networks, which are then automatically published [23].

C. Challenges of Digital Marketing

One of the most direct channels of engagement between a business and a target audience is through social media marketing — bidirectional communication. As consumers spend more time than ever online, the ability to effectively engage with them, respond to their comments, and solve their problems in real time is crucial [24]. Marketing experts have less control over the message on social networks due to its conversational nature than they do with more conventional means of marketing communication [3]. There is more exposure to digitization and social media from several consumers. Social media power is a fundamental challenge for any digital marketer. The management of customer queries is one of those challenges.

A key goal of marketers is customer engagement, a study in 2012 found that customer engagement was reinforced by 78 per cent of marketing reports when they use social media to promote their products [25]. Customer expectations must be managed to avoid harmful or negative posts. The use of negative post-reactions in social media marketing is one facet of digital marketing that can be particularly destructive to marketing campaigns. There is not much a marketer can do to prevent unfavorable comments, images, or videos from being posted online by dissatisfied consumers or competitors in the same business [26].

D. Artificial Intelligence and Automation

Artificial Intelligence is the ability to make machines execute intelligent tasks like humans. It uses intelligence to perform automated tasks. A more in-depth definition of AI describes it as the flexibility in a machine's ability to learn from experience, incorporate that knowledge into its operation, and use that performance to achieve predetermined ends [27]. As automation becomes more complex, the demand for AI is growing due to its ability to solve complex problems even with little human expertise and resources in a short timeframe [28]. With abstraction to phase out human efforts in routine tasks,

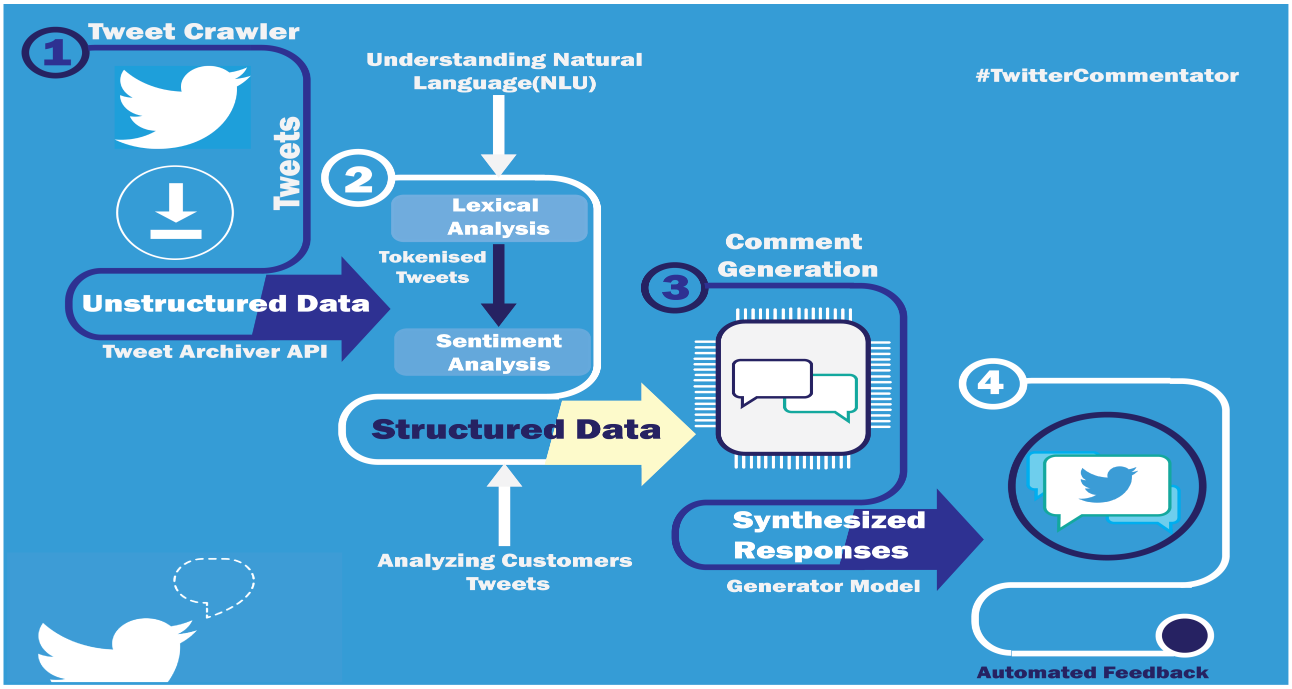

Fig. 1. Processes of social media comment generation

advancing machine learning (ML) will continuously activate paradigm shifts in multiple sectors of the technology industry. Among the components of AI is NLP, which is concerned with the interaction of machines and humans through natural language. As AI is penetrating the business arena, it has led to a fundamental change in the concept and application [29].

The application of AI is used significantly in different business domains, including healthcare, finance, manufacturing, law, education, and more. Doctors can now diagnose diseases faster than before with the assistance of ML. The application of AI in social media marketing results in spending less time on routine administrative tasks and providing customers with satisfaction. The adoption of AI can be cost effective and complement customer engagement [12]. Applications such as chatbot assist patients with invoicing and assist clients with appointment scheduling. In pedagogy, AI can provide auto-grading, assist students with learning by catering to their individual needs and ensuring that they stay on course. Artificial intelligence in the legal field has made it easier for lawyers to precisely and efficiently evaluate thousands of large legal documents, which is generally a difficult undertaking. Industrial robots have made manufacturing considerably easier and more efficient than it was only a few years ago [28].

E. Motivation

There is an increase in the number of comments available on a business social media platform every single day, especially on Twitter. It is projected that there will be 3.43 billion monthly active users on social media platforms worldwide in 2023. This figure represents almost one-third of the total population of the world [20]. Thus, it becomes a problem for a human expert to deal with a large volume of questions in different real-life data handling applications. Social media comments are important data in social media marketing, as those comments can make or break a business. However, all posts must be evaluated to effectively utilize the data acquired in the social media posts [11]. Against this background, there is the necessity for the creation of an algorithm that can complete the task swiftly.

F. The Gap

Unprecedented social media marketing growth has led to a constant influx of customers posting a massive volume of comments on business social media platforms that may be too much for a human expert to handle. When customers do not receive a prompt response to their inquiries, there is an underlying unhappiness that exists. Therefore, it is necessary to create a novel strategy to solve this issue without involving humans in easy-to-automate tasks.

G. Why Use Rule-Based Techniques?

The rule-based system has advanced several NLP systems. Rule-based logic is at the centre of most automated processes. A rule-based system encrypts expert human knowledge in a rather narrow domain using an automated system [30]. The foundation of rule-based systems is solid language comprehension and the syntactic insights gained with an NLP system can be applied to a similar task in another system [31]. One of the simple methods to represent the human intellect in AI is to transform it into plain linguistic phrases utilizing the format IF-THEN rules [32]. The knowledge expert system is an example of a rule-based technique that uses predefined rules to make deductions and draw conclusions to perform automated tasks.

A rule-based approach consists primarily of a collection of rules, a statement of facts, and a termination criterion [30]. Despite the fact that a rule-based method is neither AI nor ML, it may be used to power certain components of AI and

ML, as it essentially follows the rules specified by humans. Although an AI can determine the action to take itself, the rule-based approach does exactly as humans instructed. ML, on the other hand, is concerned with the possibility of computer programs accessing data and using it to learn for themselves. The objective is to provide machines with the ability to learn independently of human intervention and modify their behavior as necessary. Therefore, it is essential to use an approach that is most suitable for a specific field of study as this will influence the specified result. ML components in this case lack enough data to form a proper generalization. The benefit of the rule-based system is that it offers a detailed analysis of social media comments using morphological and syntactic information, while ML approaches face some challenges with specific NLP tasks, such as detecting well-formed and poorly formed inputs [31].

H. NLP Tools and Techniques

NLP deals with computer-natural language communication. Algorithmic systems can integrate language understanding and language generation through the use of NLP metrics [15]. NLP techniques enable enormous collections of social media posts to be automatically analysed and useful information extracted. Table I illustrates some examples of NLP tools and techniques.

TABLE I. SOME EXAMPLES OF NLP TOOLS AND TECHNIQUES

Tools	Techniques
AllenNLP: Premised on PyTorch tools and frameworks to use deep learning techniques for applications in NLP research and industry [33].	**Text Classification:** NLP techniques to classify vast volumes of unstructured text [34].
Keras: A library for neural networks that is highly modular, which is built on top of TensorFlow [35].	**Summarisation:** Using an NLP system to synthesize complex text by generating a concise summary [36].
NLP Architect: A Python module for evaluating deep learning architectures and techniques for optimizing NLP and NLU neural networks [37].	**Sentiment Analysis:** The analysis of data to assess its positivity, neutrality, or negativity [38].
SpaCy: Open-source NLP library that specializes in syntactic analysis, word-to-vector transformations, and conversational UIs [35].	**Keyword Extraction:** Automatically extracting the most useful textual data employing AI and machine learning algorithms [39].
NLTK: The Natural Language Toolkit is an accessible NLP resource for tokenising, stemming, labeling parts of speech, parsing, and logic [40].	**Named Entity Recognition:** Tags and extracts named identities from the using NLP algorithms for further analysis [41].

I. Related Work

According to our understanding and considerable research, no effort has been attempted to automatically comprehend and generate comments for social media posts using rule-based technique. Nevertheless, there are numerous similar works in NLP, particularly in the area of tweet message analysis on social media sites like Twitter, and various works in automatic comprehension and summarisation, language generation, and synthesis of things. Several of these interrelated concepts will be covered in this section.

1) *Twitter Immediate Response Using NLP Techniques:* Ahmad et al [42] proposed implementing NLP algorithms to analyze and evaluate various preprocessing, data augmentation, classification, and fusion approaches to provide instant response on water quality from social media posts.

2) *Social Media Posts as Programming Feedback:* Kabaso and Ade-Ibijola [43] presented a Context-Free-Grammar for automatically generating educational feedback on Twitter for novices in programming.

3) *Comment Generation Based on User Profile:* Zeng et al [44] proposed that each user's personalized comments be generated automatically based on the data in their profile. The work employed a large-scale and high-quality Chinese dataset as the foundation for creating a link between user profiles and the comments made on social media platforms.

4) *Social Media Profile Synthesis:* Ade-Ibijola [45] presented novel approaches to create imaginary social media profiles on an automatic basis utilizing probabilistic CFG.

5) *Financial Chat Analyzer (FINCHAN):* IMs with FX options obtained from Bloomberg Terminals were analyzed by [13] using a formal grammar-based technique. The work presented a novel algorithm to produce descriptive overviews of these financial chats by cleaning and retrieving terms from the conversations. The conversation token syntax was then specified using a CFG.

6) *Analysis of Tweets for Identifying Trends:* Novel algorithms were developed by [46] to identify trends in tweets related to natural disasters, such as earthquakes.

J. Definition of Terms

Definition 1: Symbol, Alphabet, and String [47]: A single element is regarded as a symbol. Any finite collection of symbols is represented by an alphabet Σ, while a string consists entirely of zero or more symbol combinations.

Definition 2: Tokenization: The process of dividing raw text into small parts is called tokenization. Raw text is divided into words and sentences and is referred to as tokens. These tokens are helpful in the creation of NLP contexts and models. Tokenizing complete sentences or individual words is possible using this method. The term "word tokenization" refers to the act of splitting a string of words into constituent lexemes. Sentence tokenization is the term used for the process of tokenizing a sentence [48].

Definition 3: Regular Languages and regular expressions [47], [13]: Languages that can be accepted by finite automata are regular languages that can be expressed by formulas known as regular expressions that use the operations of union, concatenation, and the Kleene star. An algebraic notation to describe regular language is called regular expressions (RE). The following terms apply to a regular language L over some alphabet Σ:

1. Ø denotes an empty set in regular language.

2. λ is a collection of all conceivable null strings in regular language.

3. For each $a \in \Sigma$, a denotes the regular language $\{a\}$.

4. Given L_1 and L_2 are R, then the following is true: L_1 and L_2 is R that combines both languages. The combined language, identified as L_1, L_2, is considered a regular language. L_1^* is another common language that symbolizes the Kleene closure feature and is also referred to as the Kleene star.

III. Design

The rules for generating comments for social media posts are described in this section. Our test case focuses on Twitter. Twitter is a social networking and microblogging platform that was launched in 2006 by a small group of people, including Jack Dorsey [49]. Every second, over 6,000 tweets are sent, equating to about 350,000 messages sent in a single minute, 500 million tweets produced in a single day, and 200 billion tweets sent in a single year [50]. To increase user and consumer participation, marketers employ Twitter and other social networking platforms in their campaigns [51]. A tweet has a structure that allows content to be shared among its audience with a 280-character limit [52]. Twitter can facilitate brand engagement with consumers by monitoring and responding to online consumer opinions [53]. In this paper, the structure of a tweet is described using the following attributes:

1) **Tweet Text:** it is a short message containing 280 characters or less that a Twitter user shares with a specific audience. For example, text, pictures, graphics and videos,
2) **Screen Name:** a Twitter handle or username that is assigned when a user creates their first Twitter account. This is followed by the "@" symbol, which your followers use when responding, mentioning, and sending direct messages (DM). For instance, @steveleo,
3) **Full Name:** it is a unique identifier that is different from a Twitter username. For instance, Jay Paul,
4) **Date:** the time and date when a tweet was published on Twitter,
5) **Tweet ID:** Twitter's unique identifier for each user's tweet,
6) **Location:** the location of the Twitter user,
7) **Retweet:** sharing a tweet from another account with your followers,
8) **Type of App:** the device that a user uses to send Tweets,
9) **Hashtag:** a topic or narrative preceded by the sign (#) symbol. Users can use hashtags to explore topics of interest, and
10) **Url:** the address to a user's profile on Twitter.

A. Tweet Extraction

To begin with, we first crawl tweets of a keyword associated with the hashtag "loadshedding". In South Africa, the country's electricity company, Eskom, frequently uses the term "load shedding" to prevent blackouts. As the primary source of electricity in South Africa, Eskom is tasked with ensuring the reliability of the country's power supply to best support the nation's economy and society. Their main responsibilities in the country of South Africa include the generation, transmission, and distribution of electricity. Loadshedding is a well-planned response to unplanned events that is carried out with the objective of preventing the complete collapse of the energy grid [54]. The primary goal of load shedding is to ensure that the grid does not fail completely. Consumers are frequently notified of imminent load shedding, which typically takes place in a sequence of stages: 1, 2, 3, 4 etc. The tweets were extracted using Tweet Archiver API that spans a period of 12 weeks, which are then stored in Google sheets for further preprocessing.

B. Tweet Preprocessing

It is crucial to conduct a lexical analysis to parse the tweet text. Tweets are unstructured data that frequently contains graphics and abbreviations and must be converted into a format that machines can understand and analyse. Tokenization is the initial stage in any NLP pipeline. Algorithm 1 shows the tweet handling method using regex to search for the instance of tweet data in the text. A tokenizer as presented in Algorithm 2 was designed to separate the dataset into distinct pieces of information. The dataset was first preprocessed using the NLTK tool, which involves the following cleaning process: removing Twitter handles, whitespace, emojis, URLs from the tweet, hashtag and punctuation removal, converting text to lowercase, stop word removal, tokenizing the string with regular expressions into a list of words, and reducing words to their stems, i.e., buying, buys become buy. During the tokenization process, some stop words, such as "the" and "a", will be eliminated because these words do not contribute significantly to the information obtained. An illustration of a tweet preprocessing technique can be seen in Fig. 2.

Algorithm 1 : Tweet Handling

```
Function text_handle_remove(input_text):
    using regex to search for the instance of tweet data in the text
    regex.sub(r'https?://[^\s\n\r]+', ", tweet)
    regex.sub(r'@[^\s\n\r]+', ", tweet)
    re.sub(r'#', ", tweet)
    return new_tweet_text
    end
End Function
```

C. Sentiment Analysis

The comprehension of massive amounts of unstructured data (tweets) can be facilitated for businesses using sentiment analysis. We are primarily interested in gaining a sense of the consumer perspective with respect to the dataset and the polarity of that perspective, that is, whether it is positive, negative, or neutral. The dataset was augmented with labels to facilitate the computation of opinion scores and their classification. The results shows a slightly high positive score (35.6%) compared to the negative score (35.2%), as shown in Fig. 3. It is observable that loadshedding has a negative impact on the consumers, and they viewed it as incompetence and maladministration. In addition, we utilised a word cloud as a

	Tweet Text	Tweet_No_url_cleaned	Tokenized_tweet
1	Please Fix #Eskom 😭 RT @elonmusk: Hate to say ...	Please Fix Eskom RT Hate to say it but we ne...	[please, fix, eskom, rt, hate, say, need, incr...
2	South Africa's state-owned power utility #Esko...	South Africas stateowned power utility Eskom h...	[south, africa, stateowned, power, utility, es...
3	If indeed we are serious about the township ec...	If indeed we are serious about the township ec...	[indeed, serious, township, economy, please, e...
4	To avoid impacting South Africa's electricity ...	To avoid impacting South Africas electricity s...	[avoid, impact, south, africa, electricity, su...
5	VIDEO | The Koeberg Power Station's Unit 2 has...	VIDEO The Koeberg Power Stations Unit 2 has b...	[video, koeberg, power, station, unit, 2, offl...
...	...	...	...
4037	The National Energy Regulator of South Africa ...	The National Energy Regulator of South Africa ...	[national, energy, regulator, south, africa, n...
4038	#Eskom South Africans quality for 100units of ...	Eskom South Africans quality for 100units of f...	[eskom, south, african, quality, 100units, fre...
4039	There you have it #Eskom RT @Eskom_SA: #EskomG...	There you have it Eskom RT EskomGauteng Media...	[eskom, rt, eskomgauteng, mediastatement, elec...
4040	[WATCH] A heavy police presence in #Diepkloof ...	WATCH A heavy police presence in Diepkloof for...	[watch, heavy, police, presence, diepkloof, fo...
4041	Don't you love how #Eskom is being sabotaged a...	Dont you love how Eskom is being sabotaged and...	[dont, love, eskom, sabotage, publicly, loot, ...

Fig. 2. Raw tweets before and after tokenization

Algorithm 2 : Tweet Tokenization

```
Function tokenize_text(input_text):
    tokenized_tweet=tweet_tokenizer.tokenize(tweet)
    stemmer = PorterStemmer()
    intialise english stopwords
    clean_tweet list
    for word in tokenized_tweet:
        if (word not in stopwords_english and
                word not in string.punctuation):
        wordnet_lemmatizer.lemmatize(word,
pos='verb')
        wordnet_lemmatizer.lemmatize(lemmatizer_word,
pos='noun')
        add lemmatized_word to clean_tweet list
    return tweets_clean
    end
End Function
```

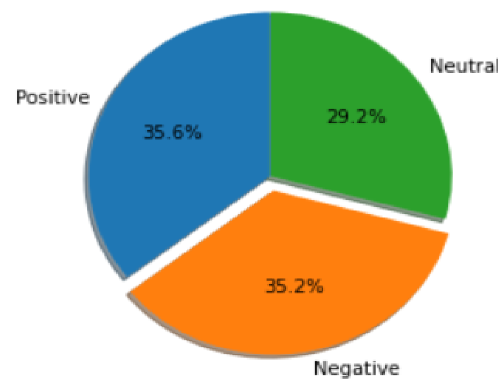

Fig. 3. Distribution of sentiments in the tweet

means of disseminating crucial information and developing an appealing visualisation strategy for the purpose of highlighting important textual data points, which helped monotonous data stand out as seen in Fig. 4.

IV. Implementation and Results

This section presents a new tool for implementing the techniques described in this article and the results obtained. The name of this tool is called `Comment-Synthesizer`.

A. Implementation details

We discuss the components of our technique for synthesizing comments and generating automated responses. First, we generated a list of intents and responses that spans the entire category of customer posts in the dataset — preprocessed Tweets. Based on the analysis of the patterns observed in the dataset, the intent of the users was classified into 19 groups, each class representing a distinct user response. We tagged each post response based on the human expert's estimation using the process below:

1) Create a list of intents and responses.
2) Use WordNet to find similar words with a tag of intent.
3) Combine words that are similar in meaning.
4) Use Regular Expressions (RegEx) to find similar patterns.

The tag in each intent is determined by the categorization of the dataset. For example, a question concerning power outages was tagged "outages" while inquiries from customers regarding

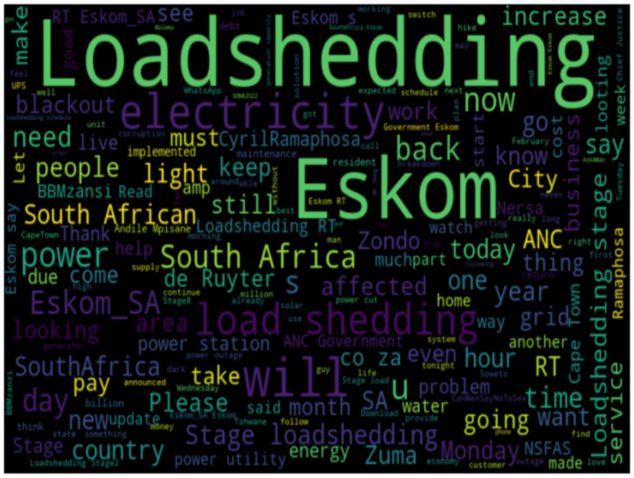

Fig. 4. Visualization of the frequent words in the tweet

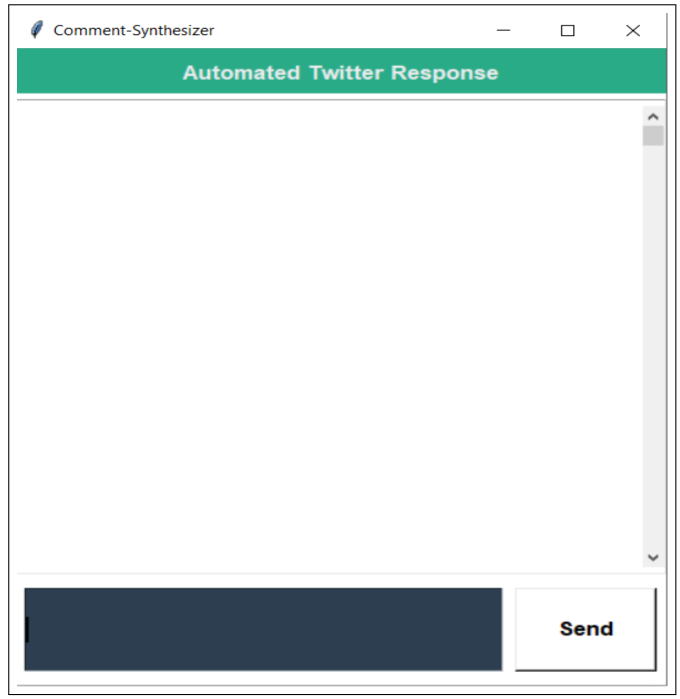

Fig. 5. The `Comment-Synthesizer` user interface

unavailable lights were tagged "lights out", etc. In addition to the tags that were entered manually during the process, the WordNet library was utilized to locate words that were related to those tags. A RegEx was used for word combinations to determine the right number of data points within the dataset. This involves querying similar tag information and retrieving the assigned tag when relevant data are discovered during the process. The full dataset and algorithm implementation were subsequently embedded into a desktop interface. Table II illustrates the function of RegEx to find matching patterns within the dataset while some implementation aspects of the tool are indicated by the steps of Algorithm 3.

TABLE II. REGEX FOR TWEET PATTERN MATCHING

Token	**Symbol**
Newline	\n
Alphabet	[a-zA-Z]
Number	[0-9]
Carriage return	\r
Tab	\t
Number and alphabet	[^a-zA-Z0-9 \n\.]
BackSpace character	\b

B. Desktop Application

The tool was implemented and integrated into a desktop application using the Python Tkinter library. This application shows a simple interface that allows users to type in their tweet and receive an automated response based on that tweet. We integrated the algorithm into a Python script using Tkinter to develop the interface and then made it feasible to employ interface formatting to create a user experience that will represent a functional chatbot application as displayed in Fig. 5.

C. Results

We have tested `Comment-Synthesizer` on tweets that were extracted from Twitter using the Tweet Archiver API between December 2021 and March 2022. The dataset was extracted using keywords associated with #Loadshedding, #EskomLoadshedding, and #Eskom_SA. In total, the dataset comprises 4,041 Tweets, which are then stored in a Google Sheet for further preprocessing. After preprocessing approaches, such as removing Twitter handles, emojis, URLs, punctuation, and whitespace, we divided the dataset into training and test datasets. There are 2,733 tweets in the training set and 1,308 tweets in the test set. The tool worked successfully and produced Twitter responses that were, for the most part, accurate. Fig. 6 and Fig. 7 shows the program interface during testing execution.

V. EVALUATION OF THE TOOL

The performance of a model can be explained using evaluation metrics. The ability of assessment metrics to differentiate between different model outputs is an essential characteristic of these metrics. The development of an algorithm and a tool that achieves a high level of accuracy when applied to data that are not part of the sample dataset is the primary objective of this research. As a result, it is essential to perform an accuracy check on this model before attempting to compute anticipated values. The evaluation of `Comment-Synthesizer` is based on the following metrics: recall, precision, and accuracy while predicting syntactically correct or incorrect responses in the dataset. In the evaluation, the confusion matrix classification was used, and the tool performed well and generated accurate tweet responses for the most part — about 76% accuracy. The classification of the tool performance is as shown below:

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{449}{449+163} = 0.73$$

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{449}{449+153} = 0.75$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{449+543}{449+543+153+163} = 0.76$$

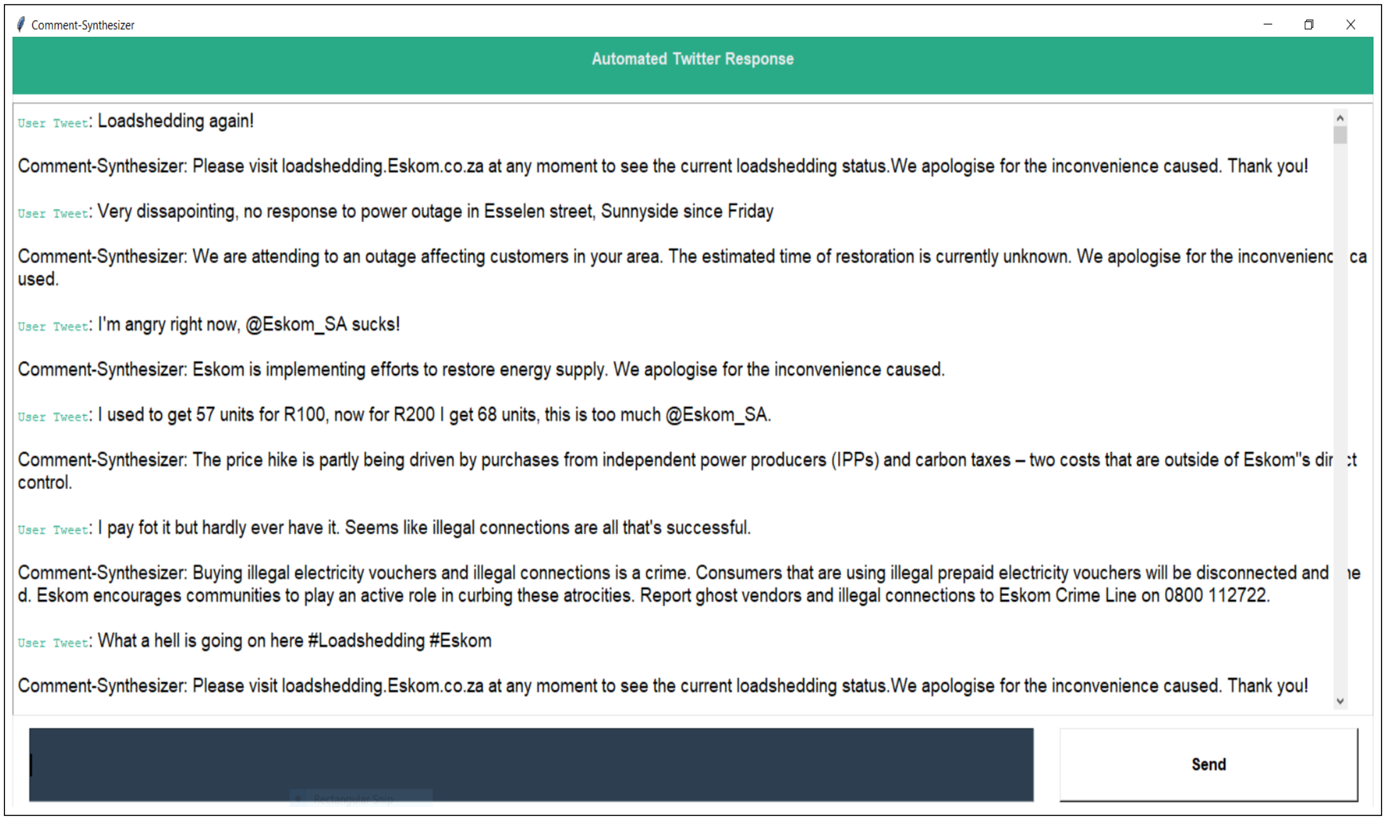

Fig. 6. The `Comment-Synthesizer` tool displays the users' tweet and the responses generated automatically (1/2)

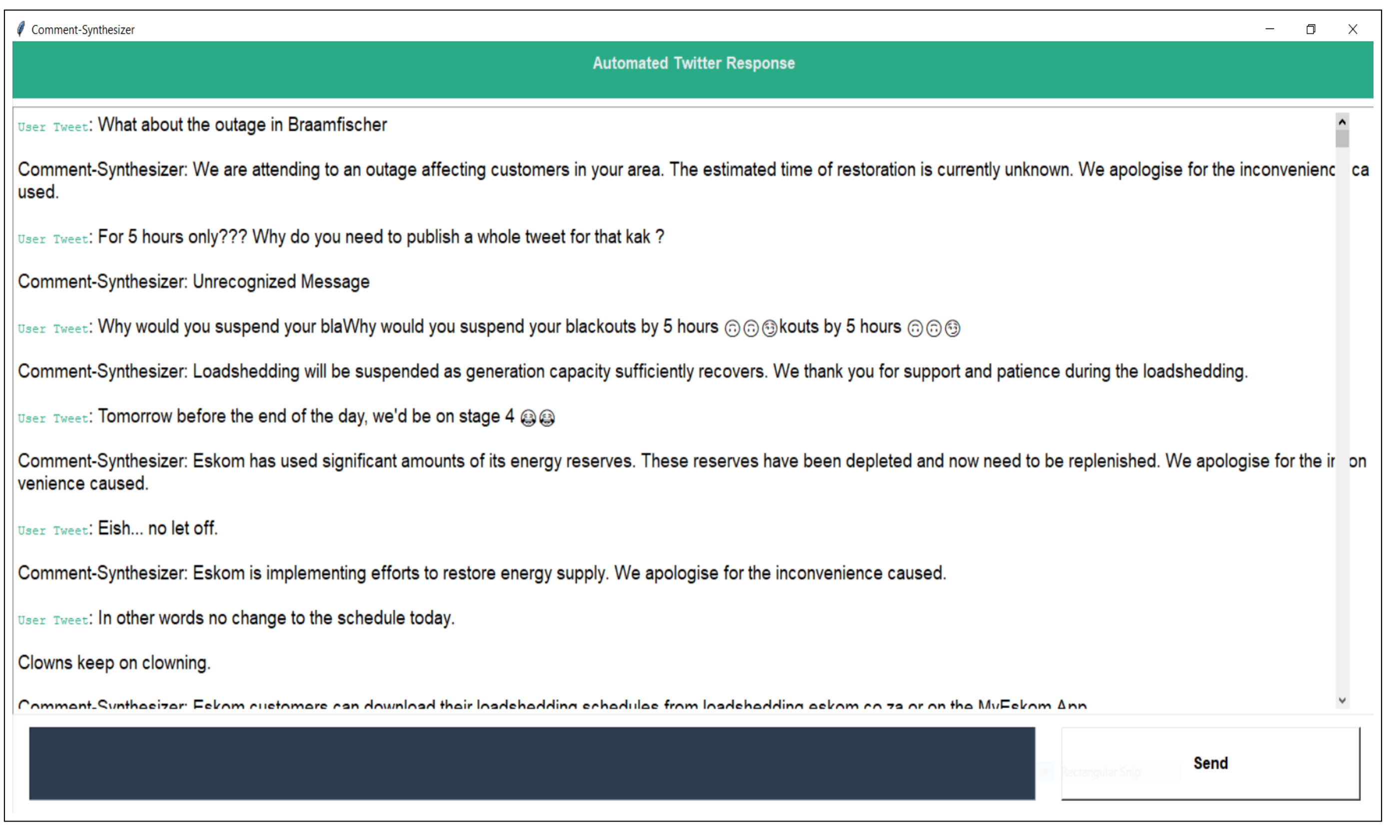

Fig. 7. The `Comment-Synthesizer` tool displays the users' tweet and the responses generated automatically (2/2)

Algorithm 3 : Comment-Synthesizer Algorithm

```
Function get_tweet_synonyms(tag)
    define the list of intents, response and tag
    preprocess tweet
    for each_tag in tag
        get synonyms of tag from Wordnet.synsets
            for each_word in synonyms
                lemmatize each_word
        add user customized synonyms
    return processed_synonym
    end
End Function
Function create_pool_of_intent(tag)
    for each_synonmys in tags_of_intent:
        add searchable regex function(".*\\b")
        Add regex compile action()
    return new_synonyms
    end
End Function
Function algorithm_implementation(user_tweet)
    for each_word in user_tweet
        convert each_word to lower_case
        remove tweet_handle, hash_tag in each_word
    produce user_tweet
    for intent in dictionary_of_intent:
        regex search for user_tweet_pattern
    if user_tweet_pattern is found:
        produce the appropriate tag and response
    else
        produce unrelated message
    return appropriate response to the tag found
    end
End Function
```

where TRUE POSITIVES are represented by the letters TP, TRUE NEGATIVES by the letters TN, FALSE POSITIVES by the letters FP, and FALSE NEGATIVES by the letters FN.

VI. Conclusion and Future Work

The design of a software program called `Comment-Synthesizer` is described, which is designed to automatically comprehend and generate comments on social media posts. This entails performing lexical analysis on raw tweets using various preprocessing techniques, such as removing Twitter handles, URLs, whitespaces, etc. In addition, we performed sentiment analysis to identify the underlying emotions in the dataset, and then we defined rule-based algorithms to find *a matching response* to each customer, based on the array of extracted tokens from the customer's tweet. The Python library Tkinter was used to implement and build the tool into a desktop application. We demonstrate the deployment of the tool with the results of the social media comments it generated. During the testing, the `Comment-Synthesizer` algorithm predicted with 76 percent accuracy for the most part. This tool, if implemented in a desktop automation application, can be used to respond automatically to a large volume of customers' social media comments/posts, thereby improving customer experience. In the future, we will increase the classification of datasets to enhance the functionality of the `Comment-Synthesizer` tool.

Acknowledgment

This research was partially supported by the Global Excellence Stature (GES) awards and National Research Fund (NRF) with grant number - 119041.

References

[1] C. Castronovo and L. Huang, "Social media in an alternative marketing communication model," *Journal of marketing development and competitiveness*, vol. 6, no. 1, pp. 117–134, 2012.

[2] A. M. Kaplan and M. Haenlein, "Users of the world, unite! the challenges and opportunities of social media," *Business horizons*, vol. 53, no. 1, pp. 59–68, 2010.

[3] J. B. Thomas, C. O. Peters, E. G. Howell, and K. Robbins, "Social media and negative word of mouth: strategies for handing unexpecting comments," *Atlantic Marketing Journal*, vol. 1, no. 2, p. 7, 2012.

[4] M. Monika and S. Sharma, "Exploring the influence of social media promotions on brand engagement, trust, perceived value and perceived risk of online shopping behavior," 2019.

[5] K. Bontcheva, G. Gorrell, and B. Wessels, "Social media and information overload: Survey results," *arXiv preprint arXiv:1306.0813*, 2013.

[6] W. Fan and M. D. Gordon, "The power of social media analytics," *Communications of the ACM*, vol. 57, no. 6, pp. 74–81, 2014. [Online]. Available: doi.org/10.1145/2602574.

[7] L. Yu, S. Asur, and B. A. Huberman, "What trends in chinese social media," *arXiv preprint arXiv:1107.3522*, 2011.

[8] X. Jin, C. X. Lin, J. Luo, and J. Han, "Socialspamguard: A data mining-based spam detection system for social media networks," *Proceedings of the VLDB Endowment*, vol. 4, no. 12, pp. 1458–1461, 2011.

[9] L. Zailskaitė-Jakštė, A. Ostreika, A. Jakštas, E. Stanevičienė, and R. Damaševičius, "Brand communication in social media: The use of image colours in popular posts," in *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. IEEE, 2017, pp. 1373–1378.

[10] M. J. Culnan, P. J. McHugh, and J. I. Zubillaga, "How large us companies can use twitter and other social media to gain business value." *MIS Quarterly Executive*, vol. 9, no. 4, 2010.

[11] J. M. Schwaiger, M. Lang, C. Ritter, and F. Johannsen, "Assessing the accuracy of sentiment analysis of social media posts at small and medium-sized enterprises in southern germany," 2016.

[12] R. Akerkar, *Artificial intelligence for business*. Springer, 2019. [Online]. Available: doi.org/10.1007/978-3-319-97436-1.

[13] A. Ade-Ibijola, "Finchan: A grammar-based tool for automatic comprehension of financial instant messages," in *Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists*, 2016, pp. 1–10. [Online]. Available: https://doi.org/10.1145/2987491.2987518.

[14] J. Kietzmann, J. Paschen, and E. Treen, "Artificial intelligence in advertising: How marketers can leverage artificial intelligence along the consumer journey," *Journal of Advertising Research*, vol. 58, no. 3, pp. 263–267, 2018.

[15] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: State of the art, current trends and challenges," *Multimedia Tools and Applications*, pp. 1–32, 2022. [Online]. Available: https://doi.org/10.1007/s11042-022-13428-4.

[16] S. Kabaso and A. Ade-Ibijola, "Sell-bot: An intelligent tool for advertisement synthesis on social media," in *The Disruptive Fourth Industrial Revolution*. Springer, 2020, pp. 155–178.

[17] A. T. Stephen, "The role of digital and social media marketing in consumer behavior," *Current opinión in Psychology*, vol. 10, pp. 17–21, 2016. [Online]. Available: https://doi.org/10.1016/j.copsyc.2015.10.016.

[18] A. Adamson, "Companies should encourage social networking among employees," *Social Media, Forbes. com, available at http://www.forbes. com/2009/06/02/charles-schwab-spyfacebook-leadership-cmo-net*, 2009.

[19] G. Appel, L. Grewal, R. Hadi, and A. T. Stephen, "The future of social media in marketing," *Journal of the Academy of Marketing Science*, vol. 48, no. 1, pp. 79–95, 2020. [Online]. Available: https://doi.org/10.1007/s11747-019-00695-1.

[20] P. H. Tankovska, "Social media - statistics & facts," 2021.

[21] H. Pham, "Top 12 Social Media Automation Tools You Should Know About," 2021.

[22] C. Samanta, "Social Media Automation – A Social Media Marketing Trend," 2020.

[23] A. Sonnenberg, "What Is Social Media Automation? What Isn't Social Media Automation? Why Do You Need Social Media Automation?" 2021.

[24] L. Hailey, "The importance of social media marketing today," *Retrieved June*, vol. 26, no. 2012, pp. 233–240, 2010.

[25] C. Nguyen, N. Nguyen, and A. Duong, "The relationships of social media marketing, consumer engagement and purchase intention," *Test Eng. Manag*, vol. 2, pp. 24 653–24 666, 2020.

[26] C. M. Cheung, M. K. Lee, and D. R. Thadani, "The impact of positive electronic word-of-mouth on consumer online purchasing decision," in *World Summit on Knowledge Society*. Springer, 2009, pp. 501–510.

[27] A. Kaplan and M. Haenlein, "Siri, siri, in my hand: Who's the fairest in the land? on the interpretations, illustrations, and implications of artificial intelligence," *Business Horizons*, vol. 62, no. 1, pp. 15–25, 2019.

[28] S. S. Shekhar, "Artificial intelligence in automation," *Artificial Intelligence*, vol. 3085, no. 06, pp. 14–17, 2019.

[29] M. S. S. E. Namaki, "How Companies are Applying AI to the Business Strategy Formulation," *Scholedge International Journal of Business Policy & Governance ISSN 2394-3351*, vol. 5, no. 8, p. 77, 2018.

[30] C. Grosan and A. Abraham, *Intelligent systems*. Springer, 2011. [Online]. Available: https://doi.org/10.1007/978-3-642-21004-4.

[31] K. Shaalan, "Rule-based Approach in Arabic Natural Language Processing," *the International Journal on Information and Communication Technologies (IJICT)*, vol. 3, no. 3, p. 11–19, 2010.

[32] S. Vashishtha and S. Susan, "Fuzzy rule based unsupervised sentiment analysis from social media posts," *Expert Systems with Applications*, vol. 138, no. July, 2019. [Online]. Available: doi.org/10.1016/j.eswa.2019.112834.

[33] M. Gardner, J. Grus, M. Neumann, O. Tafjord, P. Dasigi, N. Liu, M. Peters, M. Schmitz, and L. Zettlemoyer, "Allennlp: A deep semantic natural language processing platform," *arXiv preprint arXiv:1803.07640*, 2018.

[34] M. Thangaraj and M. Sivakami, "Text classification techniques: A literature review," *Interdisciplinary Journal of Information, Knowledge, and Management*, vol. 13, p. 117, 2018. [Online]. Available: https://doi.org/10.28945/4066.

[35] B. Srinivasa-Desikan, *Natural Language Processing and Computational Linguistics: A practical guide to text analysis with Python, Gensim, spaCy, and Keras*. Packt Publishing Ltd, 2018.

[36] S. Dohare, H. Karnick, and V. Gupta, "Text summarization using abstract meaning representation," *arXiv preprint arXiv:1706.01678*, 2017.

[37] J. Mamou, O. Pereg, M. Wasserblat, A. Eirew, Y. Green, S. Guskin, P. Izsak, and D. Korat, "Term set expansion based nlp architect by intel ai lab," *arXiv preprint arXiv:1808.08953*, 2018.

[38] J. Serrano-Guerrero, J. A. Olivas, F. P. Romero, and E. Herrera-Viedma, "Sentiment analysis: A review and comparative analysis of web services," *Information Sciences*, vol. 311, pp. 18–38, 2015.

[39] L. Marujo, W. Ling, I. Trancoso, C. Dyer, A. W. Black, A. Gershman, D. M. de Matos, J. P. Neto, and J. G. Carbonell, "Automatic keyword extraction on twitter," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, 2015, pp. 637–643.

[40] S. Bird, E. Klein, and E. Loper, *Natural language processing with Python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc.", 2009.

[41] C. Li, J. Weng, Q. He, Y. Yao, A. Datta, A. Sun, and B.-S. Lee, "Twiner: named entity recognition in targeted twitter stream," in *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, 2012, pp. 721–730.

[42] K. Ahmad, M. A. Ayub, K. Ahmad, J. Khan, N. Ahmad, and A. Al-Fuqaha, "Merit-based fusion of nlp techniques for instant feedback on water quality from twitter text," *arXiv preprint arXiv:2202.04462*, 2022.

[43] S. Kabaso and A. Ade-Ibijola, "Synthesis of social media messages and tweets as feedback medium in introductory programming," in *Annual Conference of the Southern African Computer Lecturers' Association*. Springer, 2019, pp. 3–16.

[44] W. Zeng, A. Abuduweili, L. Li, and P. Yang, "Automatic generation of personalized comment based on user profile," *arXiv preprint arXiv:1907.10371*, 2019.

[45] A. Ade-Ibijola, "Synthesis of social media profiles using a probabilistic context-free grammar," in *2017 Pattern Recognition Association of South Africa and Robotics and Mechatronics (PRASA-RobMech)*. IEEE, 2017, pp. 104–109. [Online]. Available: doi.org/10.1109/robomech.2017.8261131.

[46] T. Sakaki, M. Okazaki, and Y. Matsuo, "Tweet analysis for real-time event detection and earthquake reporting system development," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 4, pp. 919–931, 2013.

[47] J. C. Martin, *Introduction to Languages and the Theory of Computation*. McGraw-Hill Higher Education, 2011.

[48] D. Ramachandran and R. Parvathi, "Analysis of twitter specific preprocessing technique for tweets," *Procedia Computer Science*, vol. 165, pp. 245–251, 2019.

[49] K. T. Taka, M. Paora, and R. Stacey, "An analysis of te reo mäori tweets," pp. 1–17, 2015.

[50] LiveStats. (2022) Twitter usage statistics. [Online]. Available: https://www.internetlivestats.com/twitter-statistics.

[51] L. D. Hollebeek, M. S. Glynn, and R. J. Brodie, "Consumer brand engagement in social media: Conceptualization, scale development and validation," *Journal of Interactive Marketing*, vol. 28, pp. 149–165, 2014.

[52] Twitter. (2022) What to tweet. [Online]. Available: https://business.twitter.com/en/basics/what-to-tweet.html.

[53] A. I. Canhoto and M. Clark, "Customer service 140 characters at a time: The users' perspective," *Journal of Marketing Management*, vol. 29, pp. 522–544, 2013.

[54] Eskom. (2022) What is load shedding? [Online]. Available: https://loadshedding.eskom.co.za/LoadShedding/Description.