USING NATURAL LANGUAGE PROCESSING TO ANALYZE SOCIAL MEDIA DATA

Husnara Khan Daljeet Kaur

Department of Computer Science Department of Computer Science

Gyan Ganga College of technology Gyan Ganga College of technology

Jabalpur M.P India-482002 Jabalpur M.P India-482002

Email:husnarakhn@gmail.com Email:daljeetkaurkalsi83@gmail.com

ABSTRACT

Social media sites like Twitter, Facebook, Linkedin, and Instagram have developed into important informational resources. The most well-known branch of artificial intelligence is natural language processing. It is the practise of using texts and other conversational elements to communicate with machines. We use the term "natural language" because we want to speak with a computer or a smart device using human languages rather than programming languages like Java or Python, such as French or Korean. Since the focus is on the interdisciplinary sciences of linguistics, computer science, and psychology, the terms "Computational Linguistics" and "NLP" are occasionally used interchangeably. Social networking platform text analysis is part of this paper. The Python Tweepy package is used to establish the OAuth connection to the Twitter API.

Key word -: social media, social networking, natural language processing, social computing, big data, semantic analysis

# I .INTRODUCTION

The amount of data produced by internet services nowadays is enormous and is growing rapidly every day [1]. Microblogging is utilised on social networking sites and has grown significantly as a medium for communication among Internet users [2]–[3]. Every business, large and small, has joined social networking sites to share their products and look for customer ratings. In order to measure customer happiness and improve their product, the corporation will employ sentiment analysis to understand how customers feel about their products. Especially, the created method to sentiment analysis using, by and large, to look at between any device, public figure, Sports team and so on. After Facebook, Twitter is the second-largest social networking site, producing 21 million tweets each hour and 347,222 tweets per minute [1]. As a result, it opens the door for sentiment analysis and data mining based on user tweets. As sentiment analyses are a type of data mining, they allow for the observation of consumer sentiment towards a variety of topics and goods. It is also the foundation for techniques like machine learning, computational linguistics, biometrics, and natural language processing. Twitter is our platform of choice for sentiment analysis since it allows for options for the sensitivity of articulated disposition. Because Twitter only allows for 140 characters of text, users can use a short message to express their brief ideas. [4]

**Figure 1 – Social Media Logo**



A website needs to exhibit at least seven of the following traits in order to be classified as social media: [6]

A. **Web space** : The website ought to offer consumers free web space so they can post files.

B**. Web address** : Each user receives a special web address, which serves as their online identification. All of their stuff can be shared and posted on this website.

C**. Create profiles**: Users are prompted to submit personal information such as name, address, date of birth, school/college attended, and employment history. The personal information is then mined by the website to link people.

D. **Maintain friendships**: Users are encouraged to share updates about their personal and professional lives. The website then serves as a forum for friends and family to connect.

E. **Post content instantly:** Users are given the resources they need to post content instantly. This information may take the form of text, photographs, music, video, or even likes and dislikes expressed symbolically. The most recent post appears first, keeping the site current.

F. **Allow for discourse:** The ability to remark on posts made by friends and family members is granted to members.

G**. Posts contain a time stamp:** .Because every post has a timestamp, it is simple to track discussions.

**Figure 2- Use of Social Media**



Whenever a new social media platform emerges, it's important to know the difference between a shiny new target and a fast-growing platform that has a chance of sticking around. While no one can tell the future, one way to know if a platform has staying power is to compare its statistics with established social media platforms. Millions of people around the world use smartphones to connect to the Internet and access social media. Social media sites like Facebook, Instagram, WhatsApp, YouTube, Snapchat, Twitter, Linked-in, Tik-Tok, etc. and social networks collect up to trillions of data and marketing agents use this data to promote their market. Well, we analyze this data for this paper and the challenges and difficulties associated with a particular data analysis network. Online media offers a huge pool of buyers ready to interact with brands. However, online media is not always about brands. It is about people who offer their lives to others whom they see as dependent on their normal interests and despise interference, especially when someone is trying to sell them something. So a brand-oriented center can be dangerous. Social media plays an important role for young people. In this article, we will also discuss the impact of social media on young people. In recent years[26], online communication has moved towards user-oriented technologies such as social media (SM), blogs, virtual online communities and the Web. social platforms. These technologies have revolutionized user-generated data, global online communities, and rich content related to human behavior. Because SM sites connect many users to each other, researchers can use the new tool to observe human behavior on an extraordinary scale. Big SM data has provided researchers with unique perspectives to understand individuals on a larger scale and study human behavior patterns on a large scale. Human-generated data and human movement patterns have become important steps in the development of intelligent applications in many fields. At the same time, SM sites provide researchers with extensive user-generated data and human mobility, allowing them to deeply study human behavior in a wide range of applications, from viral marketing to pandemic monitoring. SM sites have become an important source of information for policymakers and consultants, and are one of the most important sources of big data that provide insights into human behavior on a large scale. Understanding human data and preferences is important for developing smart applications and services so that such applications understand people's thoughts and feelings.

Social media analytics is the process of collecting and analyzing data from social networks such as Facebook, Instagram, LinkedIn or Twitter. Part of social media analytics is called social media monitoring or social listening. It is often used by marketers to monitor online chat about products and companies. One author defined it as "the art and science of extracting valuable hidden insights from large volumes of semi-structured and unstructured social media data to enable informed and implicit decisions." [27]

 **Process**: There are three main steps. media analysis. media: data identification, data analysis and data interpretation. To maximize value at each step of the process, analysts can define the question that needs to be answered. Important questions in data analysis are: "Who? What? Where? When? Why? and how?" These questions help determine the right data sources to evaluate, which can influence the type of analysis to be performed. [28]

**Data identification**: Data identification is the process of identifying subsets of available data to focus on for analysis. . Raw data are useful when they have been interpreted. Once data has been analyzed, it can begin to convey a message. Any data that conveys a meaningful message becomes information. At a high level, raw data is translated into a precise message in the following forms: noisy data; important and irrelevant data; filtered data; only important data; knowledge; information that conveys knowledge of a vague message; information that conveys precise message wisdom; information that conveys precise message and its reason. To extract wisdom from raw data, we need to start processing it, refine the data set to include the data we want to focus on, and organize the data to identify the data. In the context of social media analytics, data means identifying "what" content is of interest. In addition to the text of the content, we want to know: who wrote the text? Where was it found or on which social networks did it appear? Are we interested in information on a specific field? When did someone say something on social media?[28]

Attributes of data that need to be considered are as follows:

 **Structure**: Structured data is data organized into a structured repository usually a database so that its elements can be manipulated for more efficient processing and analysis. Unlike structured data, unstructured data is the least structured.[29]

 **Language**: Language becomes important if we want to know the mood of the publication instead of the number of mentions.

 **Region**: It is important to ensure that the data included in the analysis comes only from the region of the world that the analysis focuses on. For example, if the goal is to identify clean water issues in India, we want to ensure that the data collected is only from India.

 **Type of content**: Information content can be text (written text that is easy to read and understand if you know the language), photos (drawings, simple sketches or photos), audio (audio recordings of books, articles, speeches) or discussions) or videos ( recording, streaming).

 **Location**: Social media content is created in many different places, such as news sites and social networks (eg Facebook, Twitter). Depending on the type of project for which data is collected, the site becomes very important.

 **Time**: It is important to collect data that was sent during the time period being analyzed.

**Data ownership**: Is the data private or public? Is the information copyrighted? These are important questions to consider before collecting data.



[[30]](https://en.wikipedia.org/wiki/Social_media_analytics#cite_note-4) Figure 3 Social media analytics process

**Data Analytics**

Data analytics is a set of activities that help transform raw data into insights that in turn lead to a new knowledge base and business value. In other words, data analysis is the step that takes filtered data as input and transforms it into valuable information for analysts. Social media data can be used for many types of analysis, including analysis of posts, sentiment, sentiment factors, geographic, demographic, etc. The data analysis phase begins when we know what problem we want to solve and we know we have enough data to generate an acceptable result. How do we know if we have enough evidence to draw a conclusion? The answer to that question is: we don't know. We won't know until we start analyzing the data. If we analyze whether we found that the information is insufficient, repeat the first step and change the question. If it is believed that the data is sufficient for analysis, we must build a data model.[28] Developing a data model is the process or method by which we organize data elements and standardize how individual data elements relate to each other. This step is important because we want to run a computer program over the data; we need a way to tell the computer which words or topics are important and whether certain words are related to the topic being studied. When analyzing data, it is useful to have several tools at our disposal that allow us to gain a different perspective on the discussions surrounding the topic. The goal is to configure tools to operate at maximum performance for a specific task. For example, when thinking about a word cloud, if we take a large amount of data from computer professionals, say "IT Architect" and build a word cloud, the largest word in the cloud would definitely be "architect". This analysis also applies to the use of tools. Some tools can be good at determining sentiment, while others are able to break down text into a grammatical form that helps us better understand the meaning and usage of different words or phrases. When doing analytics, it's hard to list every step of the analytics journey. It is largely an iterative approach because there is no set way of doing things.[28]

The taxonomy and the insight derived from that analysis are as follows:

**Depth of Analysis**: Simple descriptive statistics based on flowing data, ad hoc analysis of accumulated data or in-depth analysis based on collected data. This dimension of analysis is actually driven by the time required to prepare project deliverables. This can be seen as a broad continuum, with analysis times ranging from a few hours at one end to several months at the other. This analysis can answer the following types of questions: o How many people mentioned Wikipedia in their tweets? o Which politician received the most likes during the debate? o Which competitor garners the most social entrepreneurship mentions?

 **Machine Power**: The amount of CPU needed to process data sets in a reasonable amount of time. Power numbers must match not only the needs of the processor, but also the power of the network needed to retrieve the information. This analysis can be done in real time, near real time, ad hoc research and deep analysis. Real-time analysis on social media is an important tool when trying to understand public perception of a certain issue as it develops, allowing for an immediate response or reorientation. In near-real-time analysis, we expect data to be fed into the tool more slowly than in real-time. Ad hoc analysis is a process that aims to answer one specific question. The result of an ad hoc analysis is usually a report or data summary. Deep analysis refers to analysis that takes a long time and contains a large amount of data, which usually means a high CPU requirement.[28]

 **Area of ​​Analysis**: The area of ​​analysis is broadly classified into external social media and internal social media. Most of the time, when people use the term social media, they mean external social media. This includes content created by popular social media sites such as Twitter, Facebook and LinkedIn. Internal social media includes a corporate social network, which is a private social network used to facilitate communication within companies.[30]

 **Data speed**: Data speed in social media can be divided into two categories: data at rest and data in motion. Mobile data speed measurements can answer the following questions: How does the general population's perception of gamers change during gaming? Does the crowd feel positive about a player who actually loses the game? In these cases, the analysis is done in the order of arrival. The amount of detail produced in this analysis is directly correlated to the complexity of the analytical tool or system. A very complex tool gives more details. Another type of speed-related analysis is data analysis at rest. This analysis is done when the data is fully collected. Doing this analysis can provide insights such as: Which of your company's products are mentioned the most compared to others? What is the relative opinion of your products compared to your competitor's product?[28]

**Information interpretation**

The insights obtained from the analysis can be as diverse as the original question presented in the first phase of the analysis. At this stage, when the recipients of the information are non-technical business users, the way the information is presented becomes important. How can data be effectively rationalized so that it can be used to make good decisions? Data visualization (graphics) is the answer to this question.[31] The best visualizations are those that reveal something new about the underlying patterns and relationships in the data. Discovering patterns and underestimating them is key to the decision-making process. There are mainly three criteria to consider when viewing data.

 • **Comprehensible audience**: before building the visualization, set the goal of conveying a large amount of information in a form that is easy for the information consumer to absorb. It is important to answer "Who is the audience?" and "Can you assume the audience will be familiar with the terminology used?" Specialists have different expectations than the general public; therefore expectations must be taken into account.[33]

 • **Create a clear framework**: the analyst must ensure that the rendering is syntactically and semantically correct. For example, when using an icon, the element should be similar in size, color and location to the thing it represents, which conveys meaning to the viewer.[33]

 • **Tell a story**: Analytical information is complex and difficult to digest, so the purpose of visualization is to understand and rationalize the information. Storytelling helps the viewer understand the information. The visualization should package the information in a structure that is presented as a story and is easy to remember. This is important in many scenarios where the analyst is not the same person as the decision maker.[32] ans and then act intelligently based on learning from SM data.

**II .NLP**

The essence of Natural Language Processing[9] is to make computers understand natural language. However, this is not an easy task. Computers can understand the structured form of data, such as spreadsheets and database tables, but human languages, texts and sounds form an unstructured class of data that is difficult for a computer to understand, creating the need for Natural. Language processing. There is a lot of natural language data in various formats and it would be very easy for computers to understand and process it. We can train models in different ways according to expected returns. Humans have been writing for thousands of years, there is a lot of literature available, and it would be great if we could make computers understand it. But the task will never be easy. There are various challenges floating around such as understanding the correct meaning of a sentence, correct Named Entity Recognition (NER), correctly predicting different parts of a sentence, correlation resolution (the most difficult in my opinion). Computers can't really understand human language. If we input enough data and train the model correctly, it can distinguish and try to classify different parts of speech (noun, verb, adjective, supporter, etc.) based on the data and experience previously input. When it encountered a new word, it would try to make the closest guess, which might sometimes be embarrassing. It is very difficult for a computer to discern the exact meaning of a sentence. For example - fire radiated from the boy like atmosphere. Did the boy have a very motivating personality or did he really radiate fire? As you can see here, parsing English on a computer is difficult. There are several stages in model training. Solving a complex problem in machine learning means building a pipeline. Simply put, it means dividing a complex problem into several small problems, building models for each, and combining those models. Something similar is done in NLP. We can divide the English comprehension process model into several small parts. It would be really cool if a computer could understand that San Pedro is an island in the Central American region of Belize with a population of 16, and the second largest city in Belize. But for a computer to understand it, we have to teach the computer the basic concepts of written language. Let's start by creating an NLP pipeline. It has several steps that will eventually give us the desired output (maybe not in rare cases). Natural Language Processing (NLP) is a branch of artificial intelligence (AI) that deals with the interaction between computers and human languages. NLP is used to analyze, understand and generate natural language text and speech. NLP aims to enable computers to understand and interpret human language in the same way that humans process language.

#### NLP techniques are used in a wide range of applications, including:

* **Speech recognition and transcription[10]:** NLP techniques are used to convert speech to text, which is useful for tasks such as dictation and voice-controlled assistants.
* **Language translation:**NLP techniques are used to translate text from one language to another, which is useful for tasks such as global communication and e-commerce.
* **Text summarization[10]:** NLP techniques are used to summarize long text documents into shorter versions, which is useful for tasks such as news summarization and document indexing.
* **Sentiment analysis**: NLP techniques are used to determine the sentiment or emotion expressed in text, which is useful for tasks such as customer feedback analysis and social media monitoring.
Question answering: NLP techniques are used to answer questions asked in natural language, which is useful for tasks such as chatbots and virtual assistants.
* NLP is a rapidly growing field and it is being used in many industries such as healthcare, education, e-commerce, and customer service. NLP is also used to improve the performance of natural language-based systems like chatbot, virtual assistants, recommendation systems, and more. With the advancement in NLP, it has become possible for computers to understand and process human languages in a way that can be used for various applications such as speech recognition, language translation, question answering, and more.

NLP combines computational linguistics—the rule-based modeling of human language—with statistical, machine learning, and deep learning models. Together, these technologies allow computers to process human language in the form of text or audio data and "understand" its full meaning, including the intentions and feelings of the speaker or writer. NLP controls computer programs that translate text from one language to another, respond to voice commands, and quickly summarize large amounts of text—even in real time. You've probably used NLP in the form of voice-guided GPS systems, digital assistants, speech-to-text software, customer service chatbots, and other consumer conveniences. But NLP is also playing a growing role in business solutions that help streamline business operations, increase employee productivity, and streamline critical business processes.

**NLP Tasks**

Human language is full of ambiguity, which makes it difficult to write software that accurately determines the intended meaning of text or speech data. Homonyms, homophones, sarcasm, idioms, metaphors, exceptions to grammar and usage, variations in sentence structure—these are just some of the irregularities of human language that take years to learn, but programmers must be taught to accurately recognize and understand applications of natural ones. language from the beginning whether these applications are useful. Several NLP tasks break down human text and audio data in ways that help a computer understand what it is eating. Some of these tasks include:

• Speech recognition, also known as speech-to-text, must be performed to reliably convert speech data into text data. Speech recognition is required for all applications that follow voice commands or answer spoken questions. Speech recognition becomes particularly difficult with speech - fast, jumbled words together, with different accents and intonations, different accents and often using incorrect grammar.

• Part marking, also called grammar tagging, is the process of identifying a part of a word or part of a word in a text based on its use and context. Part of speech identifies the word "make" as a verb in "I can make a paper machine" and as a noun in "What brand of car do you have?"

 • Specialization of word meaning refers to choice of word meaning. multiple meanings using semantic analysis to determine the word that makes the most sense in a given context. For example, the word meaning ambiguity helps to differentiate the meaning of the verb "to do" in "grade" (achieve) and "contribute" (to place). • Named Entity Recognition, or NEM, identifies words or phrases as useful entities. NEM does not recognize "Kentucky" as the place or "Fred" as the man's name. • The task of co-reference resolution is to detect if and when two words refer to the same entity. The most common example is to identify the person or thing to which a particular pronoun refers (e.g. 'she' = 'Maria'), but it can also involve identifying a metaphor or language in a text (e.g. the case of "a bear is not an animal but big hairy person)

• Sentiment analysis tries to extract from the text subjective qualities - attitudes, feelings, sarcasm, confusion, doubt.

• Natural language generation is sometimes described as the opposite of speech recognition or speech-to-text; its job is to put structured information into human language.

NLP tools and approaches Python and the Natural Language Toolkit (NLTK) The Python programming language provides a wide range of tools and libraries to attack specific NLP tasks. Many of these can be found in the Natural Language Toolkit, or NLTK, an open source library, programs, and training resources for building NLP programs. NLTK contains libraries for many of the NLP tasks listed above, as well as libraries for subtasks such as sentence analysis, word segmentation, parsing and lemmatization (methods for cutting words down to their roots) and tagging (phrases, sentences, paragraphs). and paragraphs into identifiers that help the computer better understand the text). It also includes libraries for implementing functions such as semantic reasoning, the ability to draw logical conclusions from facts gathered from a text. Statistical NLP, machine learning and deep learning. Early NLP applications were hand-coded rule-based systems that could perform specific NLP tasks, but were not easily scalable to accommodate a seemingly endless stream of exceptions or increasing amounts of text and audio data. Enter statistical NLP, which combines computer algorithms with machine learning and deep learning models to automatically extract, classify and label elements of text and audio data, and then assign a statistical probability to each possible meaning of those elements. Today, deep learning models and learning techniques based on Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) enable NLP systems that "learn" on the fly and extract increasingly accurate meaning from vast amounts of raw, unstructured and unlabeled text. and audio data. NLP use cases Natural language processing is the driving force behind machine intelligence in many modern real-world applications. Here are some examples:

 • Spam detection: You might not think of spam detection as an NLP solution, but the best spam detection techniques use NLP's text classification capabilities to check email language that often indicates spam or phishing. These indicators can include excessive use of financial terms, typical bad grammar, threatening language, inappropriate urgency, misspelled companies, and more. Spam detection is one of the few NLP problems that experts consider "mostly solved" (although you could argue that it's not relevant to your email experience).

 • Machine Translation: Google Translate is an example of widely used NLP technology at work. Truly useful machine translation means more than replacing words in one language with words in another. An effective translation must accurately capture the meaning and tone of the input language and translate it into a text that has the same meaning and desired effect in the output language. Machine translation tools are making great strides in accuracy. A great way to test any machine translation tool is to translate the text into one language and then back to the original. An often-cited classic example: Not long ago, "The spirit is ready, but the flesh is weak" was translated from English into Russian and returned "Vodka is good, but the flesh is rotten."

Virtual agents and chatbots: Virtual agents, such as Apple's Siri and Amazon's Alexa, use speech recognition to recognize patterns in voice commands and natural language generation to respond with appropriate actions or helpful comments. Chatbots perform the same magic in response to typed in-text. The best also learn to recognize contextual cues to people's requests and use them to provide even better answers or choices over time. The next improvement in these programs is question answering, the ability to answer our questions, whether they are expected or not, with appropriate and useful answers in our own words.

• Social Media Sentiment Analysis: NLP has become a key business tool for uncovering hidden information in social media. Sentiment analysis can analyze the language used in messages, replies, reviews and other parts of social media to extract attitudes and feelings about products, campaigns and events – information that companies can use in product design, advertising campaigns and more.

• Text Summarization: Text Summarization uses NLP techniques to digest huge digital texts and create summaries and summaries for registries, research databases or busy readers who do not have time to read the entire text. The best text summarization applications use semantic reasoning and natural language generation (NLG) to add useful context and conclusions to summaries.

**Advantages of Natural Language Processing[9]:**

1. Improves human-computer interaction: NLP enables computers to understand and respond to human language, improving the overall user experience and making it easier for humans to interact with computers.

 2. Automate repetitive tasks: NLP techniques can be used to automate repetitive tasks such as text summing, sentiment analysis and language translation, which can save time and increase efficiency.

3. Enables new applications: NLP enables the development of new applications such as virtual assistants, chatbots and question answering systems that can improve customer service, provide information and more. .

4. Improves decision making: NLP techniques can be used to derive insights from large amounts of unstructured data, such as social media posts and customer feedback, that can improve decision making across industries.

5. Improve accessibility: NLP can be used to make technology easier, for example by providing text-to-speech and speech-to-text capabilities for people with disabilities. Disadvantages of natural language processing[9]:

1. Limited understanding of context: NLP systems have a limited understanding of context, which can lead to misinterpretation or errors in the output.

 2. Requires large amounts of data: NLP systems require large amounts of data to train and improve performance, which can be expensive and time-consuming to collect.

3. Limited ability to understand idioms and sarcasm: NLP systems have a limited ability to understand idioms, sarcasm and other figurative language, which can lead to misinterpretation or errors in the output. .

4. Limited ability to understand emotions: NLP systems have a limited ability to understand emotions and tone of voice, which can lead to misinterpretation or errors in the output.

**III LITERATURE SURVEY**

Social media platforms have proven effective in identifying detailed characteristics of local communities and have played a key role in promoting accountability and transparency in society [12]. A significant body of evidence has shown that British government officials have used social media platforms for prosecution, which has also helped stakeholders and decision-makers to more productively and accurately analyze events that previously seemed unrelated [13]. As confirmed by [1 ], social media has increased public engagement and deeper trust in the context of smart cities. [15] showed, using IDF and Metric-Cluster techniques, that a multidisciplinary collaboration of construction workers and white-collar workers in social media platforms led to the identification of training gaps. The current generation of Internet users has focused on the positive aspects of social media according to the ideology and conceptualization of smart cities and has created a well-functioning management strategy that fosters cooperation and collaboration between citizens and external organizations. This approach reduced the number of projects supported by the state, increasing the responsibility of residents. noted that the city's institutional framework could be strengthened by encouraging citizen participation and participation in decision-making processes. Therefore, it is important to ensure that the citizens of a smart city receive sufficient information and that mutual trust is created between citizens and state institutions. noted that innovative governance and digital media stability must be achieved in all smart cities, especially where political governance is a highly sensitive issue, as emerging social tensions may pose a significant threat to smart city development and sustainability.

In [16], the authors proposed a machine learning algorithm for sentiment analysis with an existing Twitter dataset. The concept of sentiment analysis using the proposed approach, which automatically classified tweets as positive, negative or neutral. They also use specific pole functions and a tree kernel. An ontology-based sentiment analysis model is proposed in [17]. The authors proposed a modeling effort that aims to blend domain ontology with natural language processing techniques to identify the sentiment underlying decisions, providing a description of such polarization. Methodological tests were developed using two separate areas, digital camera and movies [18]. To recognize the polarity of sentiments, sentiment analysis methods have been trained [19]–[20], which can automatically track the sentiments of different documents, blogs, sentences or words. In [21], the authors developed a new method to extract semantic information from research documents and articles using the integration of semantic technology, NLP, and information extraction. In [22], the authors proposed a new method that extracts structured data from emails using data cleaning, deduplication, and data consolidation. In [23], the authors proposed that a supervised learning approach is based on label data that is trained to produce meaningful results. Use Naive Bayes Algorithm, Maximum Entropy and Support Vector Machine to control the learning method which helps to achieve great success in sentiment analysis. In [2 ], the authors showed that they achieved a maximum accuracy of 82.1 degrees using the Naive Bayes algorithm. In [25], the authors used K-Nearest Neighbour classification for sentiment analysis and obtained an accuracy .In [26], the authors presented a survey on sentiment analysis of Twitter data using different techniques. They used various machine learning algorithms such as Maximum Entropy and Vector Machine to analyze the sentiment and show the accuracy of features of different sizes

**IV.METHODOLOGY**

Natural language processing is a subfield of artificial intelligence, perhaps the best known. It is the art of human-machine interaction, including texts and conversational content. We talk about "natural language" because our goal is to communicate with a computer or smart device in human languages ​​like French or Korean, not programming like Java or Python. NLP is sometimes replaced by "computational linguistics" when the focus is on the interdisciplinary fields of linguistics, computer science, and psychology. A data filtering process is used to clean the data of unnecessary characters such as stop words, and the sentence structure is analyzed by linking the text analysis step. Finally, the result analysis step captures the sentences considered important through grammar rules and keyword matching. Twitter was used as a data source to review vital information for those in need. We focused on finding information about missing people. The first step was to understand the context of the tweets. The next step was to categorize the tweets of the missing persons. To achieve this, we searched and classified real-time tweets that focused on missing people. The intercepted tweets were filtered with the corresponding word "none". The Python Tweepy library is used to create an OAuth connection to the Twitter API. Tweets were cleaned and filtered after the search. The filtering process involved removing retweets from the downloaded tweet dataset. This was done because retweets usually contained overlapping information about a particular person or event. Cleaning the data set included removing hyperlinks, unintelligible words and words that could disturb the context of our sentence. The resulting output is saved in a text file, with each tweet separated by a line. Figure 3 shows the final stage of data cleaning and the execution of filtering procedures.

**V. TEXT ANALYSIS**

Analyzing each line of tweets required some kind of semantic analysis for each tweet. Classification carried out using the Natural Language Toolkit (NLTK) library [1]–[3] to distinguish between tweets that discussed human security and those that were not relevant to the security conversation. The classification process includes line segmentation, stop word removal, part of speech (POS) classification, and POS tagging. Line splitting involved splitting the "tweet" part of each tweet block into separate words. Stop word elimination involved removing words that had little or no meaning in the context of each tweet to ensure faster screening. The list of stop words includes "of", "in", "out", "and", "in", "a", "and", "in", "these", "this". Other symbols and characters such as (';', '...', '..', '-', '``', '''', ',', ':', '!', '? ' , '#', '@',) were cleaned up as well. Words containing special characters and their meanings that may have changed were also removed. The words are then marked according to the parts of speech. A custom classifier was implemented by combining some of the speech tags. This is helpful for identifying the context of tweets. Tweets containing the desired information are selected and the names (along with the desired additional information) are aggregated into chunks.

##### Figure 4



##### **REFERENCES**

[1] J. V. Praneeth Sai and B. Balachander, “Sentimental analysis of twitter data using tweepy and textblob,” Int. J. Adv. Sci. Technol., vol. 29, no. 3, pp. 6537–6544, 2020.

[2] A. Shelar and C.-Y. Huang, “Sentiment Analysis of Twitter Data,” in 2018 International Conference on Computational Science and Computational Intelligence (CSCI), 2018, pp. 1301–1302, doi: 10.1109/CSCI46756.2018.00252.

[3] Y. Wang, L. Kung, and T. A. Byrd, “Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations,” Technol. Forecast. Soc. Change, vol. 126, pp. 3–13, Jan. 2018, doi: 10.1016/j.techfore.2015.12.019.

[4] Hiba A. Abu-Alsaad et.al NLP analysis in social media using big data technologies DOI: 0.12928/TELKOMNIKA.v19i6.19339

 TELKOMNIKA (Telecommunication Computing Electronics and Control) · December 2021

[5] <https://www.morganclaypoolpublishers.com/catalog_Orig/samples/9781681736136_sample.pdf>

[6] <https://www.easymedia.in/7-key-characteristics-of-social-media/>

[7] <https://blog.hootsuite.com/types-of-social-media/>

[8] <https://bbrcreative.com/2017/04/understanding-the-features-of-different-social-media-platforms/>

[9] https://www.geeksforgeeks.org/introduction-to-natural-language-processing/

[10 <https://builtin.com/data-science/introduction-nlp>

[11] Charu Virmani International Journal of Computational Intelligence Research ISSN 0973-1873 Volume 13, Number 4 (2017), pp. 621-630

[12] Bellini, E., Bellini, P., Cenni, D., Nesi, P., Pantaleo, G., Paoli, I., et al. (2021). An IoE and big multimedia data approach for urban transport system resilience management in smart cities. Sensors, 21(2), 435. <http://dx.doi.org/10.3390/s21020435>.

[13] Yuan, Y., Lu, Y., Chow, T. E., Ye, C., Alyaqout, A., & Liu, Y. (2020). The missing parts from social media–enabled smart cities: Who, where, when, and what? Annals of the American Association of Geographers, 110(2), 462–475. <http://dx.doi.org/10.1080/> 24694452.2019.1631144.

[14] Sharida, A., Hamdan, A., & Mukhtar, A.-H. (2020). Smart cities: The next urban evolution in delivering a better quality of life. In Toward social internet of things (SIoT): Enabling technologies, architectures and applications (pp. 287–298). Springer, <http://dx.doi.org/10.1007/978-3-030-24513-9_16>.

[15] Hodorog, A., Petri, I., Rezgui, Y., & Hippolyte, J. L. (2021). Building information modelling knowledge harvesting for energy efficiency in the construction industry.Clean Technologies and Environmental Policy, 23(4), 1215–1231. http://dx.doi.org/

10.1007/s10098-020-02000-z.

[16] V. Sahayak, V. Shete, and A. Pathan, “Sentiment analysis on twitter data,” International Journal of Innovative Research in Advanced Engineering (IJIRAE), vol. 2, no. 1, pp. 178–183, 2015.

[17] F. Ceci, A. L. Goncalves, and R. Weber, “A model for sentiment analysis based on ontology and cases,” IEEE Latin America Transactions, vol. 14, no. 11, pp. 4560–4566, 2016.

[18] B. Liu, Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge University Press, 2015.

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

[20] E. Cambria, “Affective computing and sentiment analysis,” IEEE Intelligent Systems, vol. 31, no. 2, pp. 102–107, 2016. [20] R. Upadhyay and A. Fujii, “Semantic knowledge extraction from research documents,” in Computer Science and Information Systems (FedCSIS), 2016 Federated Conference on. IEEE, 2016, pp. 439–445.

[21] S. S. A. Q. Mahlawi, “Structured data extraction from emails,” 2017 International Conference on Networks Advances in Computational Technologies (NetACT), vol. 37, pp. 323–328, 2017.

 [22] S. Rosenthal, N. Farra, and P. Nakov, “Semeval-2017 task 4: Sentiment analysis in twitter,” in Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), 2017, pp. 502–518.

 [23] N. P. M. Vadivukarassi and P. Aruna, “Sentimental analysis of tweets using naive bayes algorithm,” World Applied Sciences Journal, vol. 35, no. 1, pp. 54–59, 2017.

 [24] S. Pal and S. Ghosh, “Sentiment analysis using averaged histogram,” International Journal of Computer Applications, vol. 162, no. 12, 2017.

 [25] V. Kharde, P. Sonawane et al., “Sentiment analysis of twitter data: a survey of techniques,” 2016

[26] Mohammed Ali Al-garadi 2018 International Conference on Computing, Mathematics and Engineering Technologies – iCoMET 2018

[27] Sponder, Marshall; Khan, Gohar F. (2017). Digital analytics for marketing. New York, NY. [ISBN](https://en.wikipedia.org/wiki/ISBN_%28identifier%29) [9781138190672](https://en.wikipedia.org/wiki/Special%3ABookSources/9781138190672). [OCLC](https://en.wikipedia.org/wiki/OCLC_%28identifier%29) [975370877](https://www.worldcat.org/oclc/975370877).

[28] [Jump up to:a](https://en.wikipedia.org/wiki/Social_media_analytics#cite_ref-:0_2-0) [b](https://en.wikipedia.org/wiki/Social_media_analytics#cite_ref-:0_2-1) [c](https://en.wikipedia.org/wiki/Social_media_analytics#cite_ref-:0_2-2) [d](https://en.wikipedia.org/wiki/Social_media_analytics#cite_ref-:0_2-3) [e](https://en.wikipedia.org/wiki/Social_media_analytics#cite_ref-:0_2-4) [f](https://en.wikipedia.org/wiki/Social_media_analytics#cite_ref-:0_2-5) Ganis, Matthew; Kohirkar, Avinash (2015). Social media Analytics: Techniques and insights for Extracting Business Value Out of Social Media. New York: IBM Press. pp. 40–137. [ISBN](https://en.wikipedia.org/wiki/ISBN_%28identifier%29) [978-0-13-389256-7](https://en.wikipedia.org/wiki/Special%3ABookSources/978-0-13-389256-7).

[29]["What is structured data? - Definition from WhatIs.com"](http://whatis.techtarget.com/definition/structured-data). WhatIs.com. Retrieved 2016-12-06.

[30] Ganis, Matthew; Kohirkar, Avinash (2015). Social media Analytics: Techniques and insights for Extracting Business Value Out of Social Media. New York: IBM Press. pp. 247–248. [ISBN](https://en.wikipedia.org/wiki/ISBN_%28identifier%29) [978-0-13-389256-7](https://en.wikipedia.org/wiki/Special%3ABookSources/978-0-13-389256-7).

[31] Kitt, Denise (2012-05-24). ["Enterprise Social Networks Explained"](http://www.crmswitch.com/social-crm/enterprise-social-network/). CRM Switch. Retrieved 2016-11-05.

[32] Steele, Julie (2012-02-15). ["Why data visualization matters"](https://www.oreilly.com/ideas/why-data-visualization-matters). O'Reilly Media. Retrieved 2016-12-11.

[33] [Jump up to:a](https://en.wikipedia.org/wiki/Social_media_analytics#cite_ref-:1_7-0) [b](https://en.wikipedia.org/wiki/Social_media_analytics#cite_ref-:1_7-1) [c](https://en.wikipedia.org/wiki/Social_media_analytics#cite_ref-:1_7-2) ["The Three Elements of Successful Data Visualizations"](https://hbr.org/2013/04/the-three-elements-of-successf). Harvard Business Review. Retrieved 2016-12-11.