

USING NATURAL LANGUAGE PROCESSING TO ANALYZE SOCIAL MEDIA DATA

Abstract

Social media sites like Twitter, Facebook, LinkedIn, and Instagram have developed into important informational resources. Natural language processing is the area of artificial intelligence that is most well-known. The act of conversing with machines through texts and other conversational elements. We refer to this as "natural language" because we prefer to communicate with computers and other smart devices using human languages like French or Korean rather than programming languages like Python or Java. Since linguistics, computer science, and psychology are the three interdisciplinary fields that are highlighted, 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.

Keyword: Social networking, big data, semantic analysis, natural language processing, social computing

Autors

Husnara Khan

Department of Computer Science,
Gyan Ganga College of technology,
Jabalpur M.P India.

Email:husnarakhn@gmail.com

Daljeet Kaur

Department of Computer Science,
Gyan Ganga College of technology,
Jabalpur M.P India.

Email:daljeetkaurkalsi83@gmail.com

I. INTRODUCTION

The amount of data produced by internet services is currently enormous and growing rapidly every day [1]. Microblogging, a popular method of communication among Internet users, is used by social networking sites [2]–[3]. Both big and small businesses have accounts on social networking sites where they can share their products and look for customer reviews. The company will use sentiment analysis to determine how customers feel about its products in order to gauge customer satisfaction and improve the product. In particular, the developed method for sentiment analysis is used frequently to compare any device, famous person, sports team, and more. Twitter is the second-largest social networking site after Facebook, generating 347,222 tweets per minute and 21 million tweets per hour [1]. As a result, it makes user-generated tweet sentiment analysis and data mining possible. As a subset of data mining, sentiment analyses enable the observation of consumer sentiment toward a range of subjects and goods. It also serves as the basis for methods like natural language processing, computational linguistics, biometrics, and machine learning. Since Twitter offers options for the sensitivity of articulated disposition, it is our preferred platform for sentiment analysis. Twitter only permits 140 characters of text per message, so users can use a short message to convey their succinct thoughts. [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]

1. **Web Space:** In order for users to upload files, the website needs to provide them with free web space.
2. **Web Address:** Each user is assigned a unique web address that acts as their online ID. They allow sharing and posting of all of their content on this website.
3. **Create Profiles:** When creating a profile, users are asked for personal information like their name, address, date of birth, the school or college they attended, and their employment history. The website then uses the personal data to link individuals.
4. **Keep Friendships Strong:** Users are urged to post updates on their personal and professional lives. The website then acts as a platform for connecting with friends and family.

5. **Instant Content Posting:** It is a possible thanks to the resources provided to users. Text, images, music, video, or even likes and dislikes expressed symbolically can all be used to convey information. The most recent post is displayed first to keep the website up to date.
6. **Allow for Discourse:** The ability to remark on posts made by friends and family members is granted to members.
7. **Posts Contain a Time Stamp:** .Because every post has a timestamp, it is simple to track discussions.

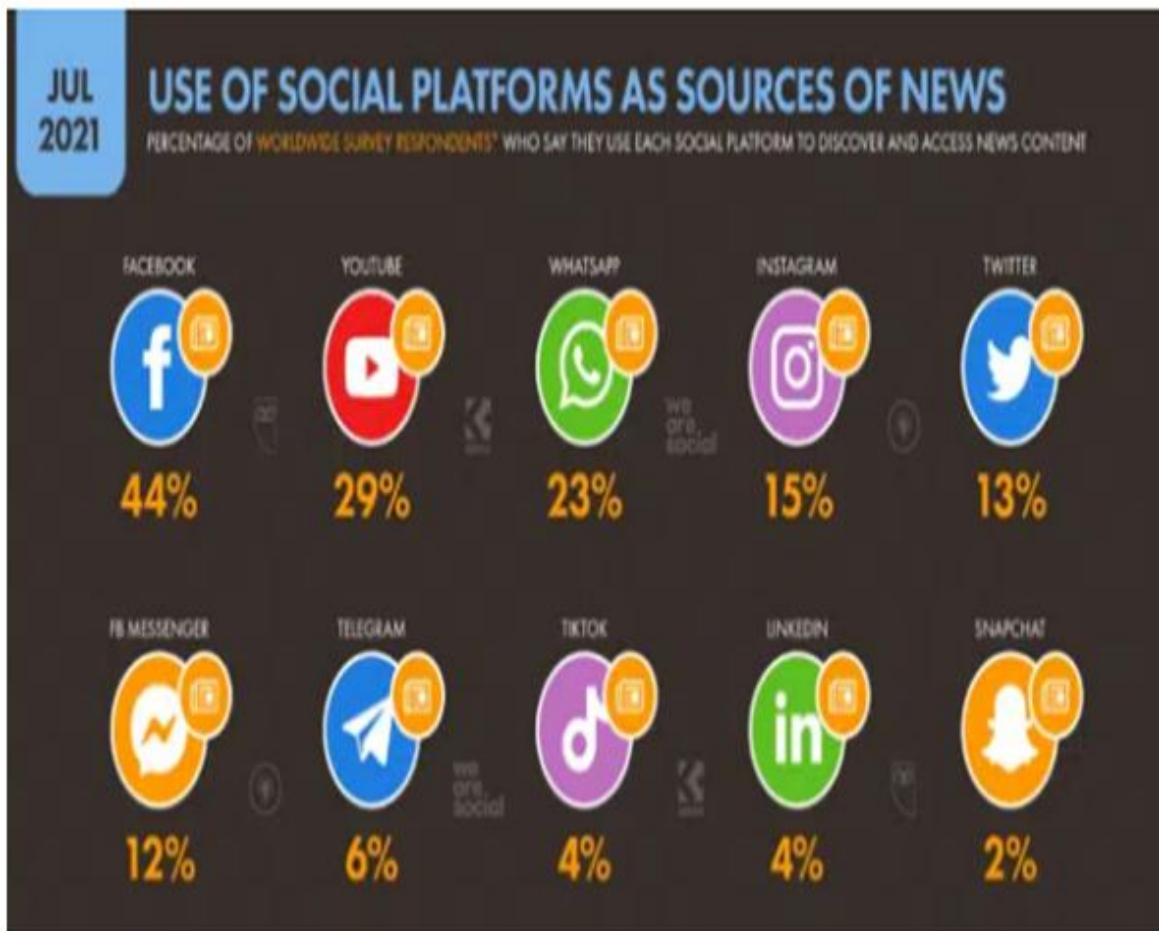


Figure 2: Social Media Usage

Knowing the difference between a shiny new target and a quickly expanding platform with longevity is crucial whenever a new social media platform appears. While no one can predict the future, comparing a platform's statistics to those of well-established social media platforms can help determine whether it will survive. Smartphones are used by millions of people worldwide to connect to the Internet and access social media. Social media platforms like Facebook, Instagram, WhatsApp, YouTube, Snapchat, Twitter, Linked-in, Tik-Tok, and others gather trillions of pieces of data, which marketing professionals use to advance their client's brand. In this paper, we analyze this data as well as the problems and difficulties that a specific data analysis network faces. It is about individuals who give up their lives for those they perceive as being dependent on their regular interests and despise interference, particularly when someone is attempting to sell them something. Consequently, a brand-

focused center could be harmful. Young people are heavily reliant on social media. We will also talk about how social media affects young people in this article. Online communication has shifted in recent years [26] to user-oriented technologies like social media. (SM), blogs, virtual online communities, and social networking sites on the Internet. User-generated data, international online communities, and rich content about human behavior have all been transformed by these technologies. Because SM sites connect many users to each other, researchers can use the new tool to observe human behavior on an extraordinary scale. Researchers now have unique perspectives to study human behavior patterns on a large scale and comprehend individuals on a larger scale thanks to big SM data. Human movement patterns and data generated by humans have gained importance in the development of intelligent applications in many fields. At the same time, With the abundance of user-generated data and human mobility that SM sites offer, researchers can thoroughly examine how people behave in a variety of contexts, from pandemic monitoring to viral marketing. As one of the most significant sources of big data that offer insights into widespread human behavior, SM sites have emerged as a crucial informational resource for consultants and policymakers. In order to create smart applications and services that can comprehend people's thoughts and feelings, it is crucial to understand human data and preferences.

Social media analytics is the process of collecting and analysing data from social networks like Facebook, Instagram, LinkedIn, or Twitter. Social media analytics includes social media monitoring, commonly referred to as social listening. It is frequently used by marketers to monitor internet talk of products and companies. One author described it as "the art and science of extracting valuable hidden insights from large volumes of semi-structured and unstructured social media data" in order to "enable informed and implicit decisions." [27].

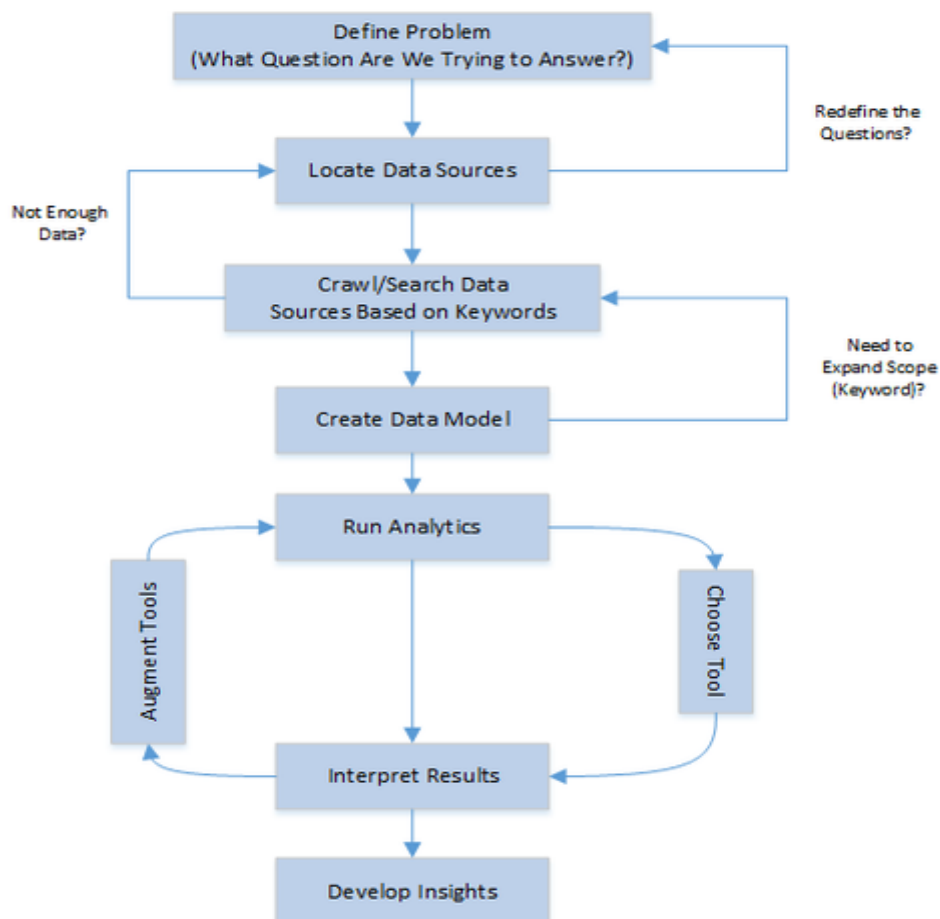
There are three main steps in the procedure. media evaluation, identification analysis, and interpretation of data through media. Analysts are able to specify the question that needs to be answered in order to maximize value at each stage of the process. The following questions are crucial when analyzing data: "Who, What, Where, When, Why, and How?" These queries assist in identifying the appropriate data sources to consider, which may affect the kind of analysis to be carried out. [28]

Data identification is the process of choosing which portions of the available data to concentrate on for analysis. When raw data have been interpreted, they are useful. Data can start to communicate a message after analysis. Information is any data that conveys a valuable message.

At a high level, noisy data, important and irrelevant data, filtered data, only important data, knowledge, information that conveys knowledge of a vague message, wisdom, information that conveys precise message, and information that conveys precise message and its reason are all ways that raw data is translated into a precise message. We must begin processing the raw data, narrow the data set to include the data we want to concentrate on, and organize the data to distinguish the data in order to extract wisdom from it. Data in the context of social media analytics refers to the determination of "what" content is relevant. Along with the content's text, we also need to know who wrote it. Where did it turn up or which social media platforms did it use? Do we want details about a particular subject? When was something posted on social media? [28]

The following data attributes need to be taken into account:

1. **Structure:** Structured data is data that has been put into a database or other structured repository so that its components can be changed for faster processing and analysis. Unstructured data is the least structured, in contrast to structured data. [29]
2. **Language:** If we want to understand the tone of the publication rather than just the number of mentions, language becomes crucial.
3. **Region:** It is crucial to make sure that the analysis's data only comes from the area of the world that it is focused on. For instance, we want to make sure that the data collected is only from India if the objective is to identify clean water issues in India.
4. **Type of Content:** Text (written text that is simple to read and understand if you know the language), photos (drawings, simple sketches, or photos), audio (audio recordings of books, articles, speeches, or discussions), and videos (recorded, streaming) can all be used as information content.
5. **Location:** Many different places, including news websites and social networks (like Facebook and Twitter), are used to create social media content. The site becomes very important depending on the type of project for which data is collected.
6. **Time:** It is crucial to gather information sent during the time frame being studied.
7. **Data Ownership:** Is the data private or public? Is the information copyrighted? The answers to these queries should be taken into account before data collection.



^[30]**Figure 3:** Process of social media analytics

Data Analytics: The process of turning raw data into insights that produce new knowledge and economic value is known as data analytics. Data analysis, in other words, is the process that takes filtered data as input and turns it into useful information for analysts.. Social media data can be used for many types of analysis including post analysis, sentiment, and sentiment-related factors, geographic, demographic. When we are certain that we have enough data to produce a reliable result and that we are aware of the problem we are trying to solve, the data analysis phase can begin. How can we tell if the evidence is strong enough to support a conclusion? We have no idea, is our response to that query. Until we begin analyzing the data, we won't know. Repeat the first step and change the question if we determine after analysis that the information is insufficient. If it is believed that the data is We need to create an analysis-ready data model.[28] The method or process by which we organize data elements and standardize how various data elements relate to one another is called developing a data model. This step is crucial because we want to run a computer program over the data; we need a way to communicate to the computer what terms or topics are crucial and whether specific terms are connected to the topic being studied. When analyzing data, it is useful to have a variety of tools available to us that allow us to gain a different perspective on the discussions surrounding the topic. The objective is to set up tools to perform at their peak level for a particular task. When considering a word cloud, for instance, the largest word in the cloud would unquestionably be "architect" if we took a lot of data from computer professionals, say "IT Architect," and created one. The use of tools is covered by this analysis as well. Some tools are adept at detecting emotion, while others can deconstruct text into grammatical structures that enable us to better comprehend the meaning and application of various words and phrases. When doing analytics, It is challenging to outline each step of the analytics process. Due to the lack of a predetermined procedure, it is largely an iterative process. [28]

Following is a taxonomy and insight based on that analysis:

Depth of Analysis: Flowing data-based simple descriptive statistics, ad hoc analysis of accumulated data, or in-depth analysis based on collected data. The time actually drives this aspect of analysis required to prepare project deliverables. With analysis times ranging from a few hours at one end to several months at the other, this can be thought of as a broad continuum. The following categories of queries can be addressed by this analysis: How many tweets included references to Wikipedia? o Which candidate won the debate with the most likes? Which rival receives the most mentions for social entrepreneurship?

Machine Power: The number of CPUs required to process data sets in a timely manner. Power specifications must take into account both the processor's requirements and the power of the network required to retrieve the data. Deep analysis, ad hoc research, and real-time or nearly real-time analysis are all options for this analysis. Real-time social media analysis is a crucial tool for determining how the public feels about a particular issues develops, allowing for an immediate response or reorientation. We anticipate that data will enter the tool more slowly than in real-time analysis than during real-time analysis. Ad hoc analysis is a process that aims to answer one specific question. A report or data summary is frequently the outcome of an ad hoc analysis. Deep analysis is an analysis that requires a lot of time and data, which usually means a powerful computer. [28]

Area of Analysis: Internal social media and external social media are two broad categories for the analysis area. The majority of the time, when people refer to social media, they are referring to external social media. Content produced by well-known social media platforms like Twitter, Facebook, and LinkedIn is included in this. An example of internal social media is a corporate social network, a private social network used to promote communication within businesses. [30]

Data Rate: There are two types of data speed in social media: data at rest and data in action. Mobile data speed measurements can answer the following questions: How does the general population's perception of gamers change during gaming? Does the crowd support a player even when they lose the game? In these instances, the analysis is done in the order of arrival. The complexity of the analytical tool or system is directly correlated with the level of detail generated in this analysis. A very intricate tool gives more details. Another type of speed-related analysis is data analysis. This evaluation is done when all of the info has been gathered. Doing this analysis can reveal information like: Which of your company's goods receives the most mentions in comparison to other products? What is the general consensus regarding your goods in comparison to those of your rivals? [28]

Interpretation of Information: As varied as the initial question posed in the first stage of the analysis, the conclusions drawn from it can be. At this point, When non-technical business users are the information's intended audience, the presentation of the information becomes important. How can data be effectively rationalized so it can be utilized to make good choices? Graphical representations of data serve as the answer to this query.[31] The most effective visualizations are those that make a novel discovery regarding the underlying patterns and connections in the data. The decision-making process relies heavily on patterns and underestimating them. When viewing data, there are primarily three factors to take into account.

- **Comprehensible Audience:** Establish the objective before creating the visualization of conveying a large amount of information in a digestible format for information consumers. The question "Who is the audience?" must be answered. "Can you make a guess at the audience will be familiar with the terminology used?" Specialists have Expectations must be considered because they differ from those of the general public [33]
- **Create Clear Framework for Different Expectations:** The analyst must guarantee that the rendering is both syntactically and semantically accurate. for illustrations rather than the general populace, when using an icon, the element should be similar in size, color and location to the thing it represents, which conveys the viewer meaning.[33]
- **Tell a Tale:** The goal of visualization is to comprehend and make sense of analytical information because it is intricate and challenging to understand. The use of stories to convey information aids comprehension. The information should be packaged in the visualization in a way that makes sense as a story and is simple to recall. This is crucial in many situations where the analyst and decision-maker are different people [32] and then take informed action after learning from SM data.

II. NLP

Making computers understand natural language is the core goal of natural language processing [9]. This is not a simple task, though. Natural is required because human languages, texts, and sounds constitute an unstructured class of data that is challenging for a computer to comprehend. Computers can understand structured forms of data, such as spreadsheets and database tables. the process of language. The amount of natural language data available in a variety of formats makes it very simple for computers to comprehend and process. Models can be trained in different ways according to expected returns. It would be great if we could make computers understand the literature that humans have been producing for thousands of years. However, the task will never be simple. It can be difficult to comprehend a sentence's intended meaning or perform accurate Named Entity Recognition (NER), correctly predicting different parts of a sentence, correlation resolution (which, in my opinion, is the most difficult). It's difficult for computers to comprehend human language. If we train the model with sufficient data input correctly, Based on the data, it can identify and attempt to categorize various parts of speech (nouns, verbs, adjectives, supporters, etc.) and experience previously input. When it encountered a new word, it would try to make the closest guess, which might sometimes be embarrassing. A computer has a very difficult time understanding the precise meaning of a sentence. For instance - The boy like environment radiated fire. Did the young man truly exude fire, or did he just have a very inspiring personality? As you can see, it's challenging for computers to parse english. There are several stages in model training. Building a pipeline is necessary for machine learning to solve complex problems. It simply means creating models for each small problem after breaking a complex problem into several smaller ones. In NLP, a similar process is used. The English comprehension process model can be broken down into a number of smaller components. The second largest city in Belize and an island in Central America with a population of 16 people, San Pedro, would be really cool if a computer could comprehend. But in order for a computer to comprehend it, we must first teach it the fundamentals of written language. Firstly, let's build an NLP pipeline. It has a number of steps that, in the end, will produce the desired result (although not always). NLP stands for Natural Language Processing, a subfield of artificial

NLP methods are used in a range of situations, such as:

- **Speech Transcription and Recognition [10]:** NLP techniques are used to translate speech to text, which is helpful for activities like dictation and voice-activated assistants.
- **Language Translation:** Text can be translated into another language using NLP techniques, which is helpful for e-commerce and other international communication jobs.
- **Text Summarising[10]:** NLP techniques are used to condense lengthy text documents into manageable chunks, which is helpful for tasks like document indexing and news summaries.
- **Sentiment Analysis:** To determine the sentiment or emotion expressed in text, NLP techniques are applied. This is beneficial for activities like monitoring social media and analysing customer comments. Natural language processing (NLP) techniques

are used to respond to questions in natural language, which is important for chatbots and virtual assistants, among other applications.

- NLP is a fast expanding field that is employed in a variety of sectors, including customer service, e-commerce, healthcare, and education. NLP is also used to enhance the functionality of systems that employ natural language, such as chatbots, virtual assistants, recommendation engines, and more.
- The development of NLP has made it feasible for computers to comprehend and process human languages in a way that may be utilised for a variety of applications, including speech recognition, language

NLP combines statistical, machine learning, and deep learning models with computational linguistics—the rule-based modeling of human language. Together, these technologies enable computers to process text or audio data that contains human language and "understand" it in its entirety, including the speaker or writer's intentions and feelings. Computer programs that translate text between languages, react to voice commands, and quickly sum up large amounts of text—even in real time—are all controlled by NLP. You've probably used NLP in the form of speech-to-text software, chatbots for customer service, voice-guided GPS systems, digital assistants, and other consumer conveniences. However, NLP is also taking on a bigger part in business solutions that help to streamline crucial business processes, boost employee productivity, and streamline business operations.

NLP Work: Because human language is so ambiguous, it is challenging to create software that can correctly ascertain the intended meaning of text or speech data. Programmers must be trained to accurately recognize and comprehend the applications of natural language irregularities, which include homophones, sarcasm, idioms, metaphors, exceptions to grammar and usage, and variations in sentence structure. whether these applications are useful from the start. A number of NLP tasks dissect human text and audio data in a way that enables a computer to comprehend what it is consuming. Some of these jobs consist of:

- **Speech Recognition**, also referred to as speech-to-text, must be performed to reliably convert speech text data into data from. Speech synthesis is necessary for all applications that follow verbal instructions or respond to questions. Speech synthesis becomes particularly difficult with speech - fast, jumbled words together, with different accents and intonations, different accents and frequent grammatical errors.
- **Part Marking**, the process of identifying a word or part of a word in a text based on its use and context is known as, also known as grammar tagging. The word "make" is classified as a verb in the sentence "I can make a paper machine" and as a noun in the sentence "What brand of car do you have?"
- **Specialization of Word Meaning**, refers to choice of word meaning. multiple meanings using ,semantic analysis to identify the word that fits the context the best. The word meaning ambiguity, for instance, can be used to distinguish between the meaning of the verb "to do" in the words "grade" (achieve) and "contribute" (to place). • NEM, or named entity recognition, recognizes certain words or phrases as useful entities. NEM is unable to identify either Kentucky as the location or Fred as the person. • Co-reference resolution's task is to determine whether and when two words refer to the same thing. The most typical example is determining who or what a specific pronoun refers to (for example, "she" = "Maria"), but it can also involve

determining a metaphor or language used in a text (for instance, "a bear is not an animal").

- **Sentiment Analysis** looks for intangible elements like attitudes, feelings, sarcasm, confusion, and doubt in the text.
- Its job is to translate structured data into human language; this process is sometimes referred to as the opposite of speech recognition or speech-to-text.

NLP techniques and tools Natural Language Toolkit (NLTK) and Python To tackle particular NLP tasks, a variety of tools and libraries are available in the Python programming language. The Natural Language Toolkit, also known as NLTK, is a collection of open source programs, libraries, and training materials used to create NLP programs. Libraries for subtasks like sentence analysis, word segmentation, parsing, lemmatization (methods for breaking words down to their roots), and tagging (phrases, sentences, paragraphs) are all included in NLTK, which includes libraries for many of the NLP tasks mentioned above. and paragraphs into symbols that aid a computer's comprehension of the text). Libraries for implementing features like semantic reasoning, the capacity to draw logical conclusions from information gathered from, are also included.. Deep learning, statistical NLP, and machine learning. Early NLP applications were manually programmed rule-based systems that were capable of performing particular NLP tasks, but they were not easily scalable to handle an apparently never-ending stream of exceptions or growing amounts of text and audio data. Enter statistical NLP, which uses computer algorithms, machine learning, and deep learning models to automatically extract, classify, and label elements of text and audio data. It then gives each potential meaning of those elements a statistical probability. Currently, deep learning models and learning methods based on convolutional neural networks (CNN) and recurrent neural networks (RNN) allow natural language processing (NLP) systems to "learn" on the fly and extract ever-greater amounts of meaningful information from massive amounts of unlabeled, unstructured, raw text and audio data. NLP examples In many contemporary real-world applications, machine intelligence is powered by natural language processing.

Here are a few instances:

- **Spam Identification:** Although you might not think of NLP as a solution for spam detection, the most effective methods check emails for language that frequently denotes spam or phishing. The overuse of financial terms, common poor grammar, threatening language, inappropriate urgency, misspelled companies, and other factors are examples of these indicators. One of the few NLP issues that experts consider "mostly solved" is spam detection, though you could argue that it has little to do with how you use email.
- **Machine Translation:** One application of NLP that is frequently used is Google Translate. More than simply swapping out words from one language for another is required for machine translation to be truly useful. The intent and tone of the original text must be faithfully translated into an output text that conveys the same meaning and achieves the desired effect. The accuracy of machine translation tools is improving dramatically. Translating the text into one language and then back to the original is a great way to test any machine translation software. The phrase "The spirit is ready, but the flesh is weak" was recently translated from English into Russian, and the result was "Vodka is good, but the flesh is rotten," which is a commonly used classic example. Virtual agents and chatbots: Virtual agents, like Apple's Siri and

Amazon's Alexa, recognize patterns in voice commands using speech recognition, and they use natural language generation to provide appropriate responses or helpful comments. In response to text input, chatbots work the same magic. The best also develop the ability to recognize contextual cues in relation to people's requests and use them over time to offer even better options or answers. The next advancement in these programs is question answering, which allows us to provide appropriate and helpful answers in our own words to any of our questions, regardless of whether they are expected or not.

- **Social Media Sentiment Analysis:** NLP has emerged as a crucial business tool for locating hidden data in social media. Language can be analyzed using sentiment analysis. used in messages, replies, reviews and other parts of social media to extract attitudes and feelings about products, campaigns and events, information that businesses can use in everything from product design to advertising campaigns.
- **Text Summarization:** It processes massive digital texts using NLP techniques to produce summaries and summaries for registries, research databases, or busy readers who don't have time to read the entire text. The best text summarization software uses natural language generation (NLG) and semantic reasoning to give summaries relevant context and conclusions.

Advantages of Automatic Language Recognition [9]

1. NLP makes it possible for computers to comprehend and respond to human language, which enhances user experience overall and makes it simpler for people to connect with computers.
2. Automate repetitive operations: By automating time-consuming tasks like text summarization, sentiment analysis, and language translation, NLP approaches can enhance productivity and save time.
3. Facilitates the development of new applications: NLP facilitates the creation of new apps like virtual assistants, chatbots, and question-answering programmes that can enhance customer service and offer information.
4. Enhances decision-making: NLP methods have the capacity to glean insights from huge amounts of unstructured data, such as social media posts and consumer reviews, which can enhance decision-making across industries.
5. Make things more accessible: 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. Contextual knowledge is limited by NLP systems, which might result in incorrect interpretations or inaccuracies in the output.
2. Needs a lot of data: NLP systems need a lot of data to be trained and perform better, but gathering this data may be expensive and time-consuming.
3. Limited understanding of idioms and sarcasm: NLP systems have a limited understanding of idioms, sarcasm, and other figurative language, which might result in output that is misinterpreted or contains errors.
4. Limited emotional understanding: NLP systems have difficulty deciphering facial expressions and voice tones, which can cause misinterpretations or inaccuracies in the output.

III. LITERATURE SURVEY

Social media channels have proven effective at recognizing detailed characteristics locally based communities and have played a crucial part in fostering social responsibility and openness [12]. The use of social media by British government officials for prosecution has been demonstrated by a sizable body of evidence, which has benefited 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 greater participation and trust within the context of smart cities [15] showed, using Utilizing IDF and Metric-Cluster techniques are a multidisciplinary collaboration of construction workers and white-collar workers in social media platforms led to the identification of training gaps. Internet users in this generation has focused regarding the advantages of social media according to the philosophy and conceptualization of shrewd cities and has created a properly running management a plan that encourages interaction and coordination between locals and outside groups. This strategy resulted in fewer projects. supported by the state, increasing the responsibility of residents. noted that by promoting citizen involvement and participation in decision-making processes, the institutional framework of the city may be strengthened. Consequently, it is crucial to make sure that the citizens of a smart city receive sufficient information between citizens and state institutions, there is mutual trust Noting that emerging social tensions may seriously jeopardize the growth and sustainability of smart cities, it was noted that innovative governance and digital media stability must be achieved in all smart cities, particularly where political governance is a highly sensitive issue.

With an existing Twitter dataset, the authors of [16] proposed a machine learning algorithm for sentiment analysis. The proposed method for sentiment analysis automatically categorized tweets as positive, negative, or neutral. In addition, 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 using a domain ontology and natural language processing techniques, one can describe this polarization by identifying the sentiment that underlies decisions. Methodological tests were created using the two distinct mediums of movies and digital cameras. [18]. To recognize the polarity of sentiments, sentiment analysis methods have been trained [19]–[20], which can track the emotions expressed in various documents, blogs, sentences, or words automatically. By combining semantic technology, NLP, and information extraction, the authors of [21] created a new technique for extracting semantic information from research documents and articles. Data cleaning, deduplication, and data consolidation are used in the authors' new method for extracting structured data from emails that they described in [22]. A supervised learning approach is based on label data that is trained to produce relevant results, according to the authors' proposal in [23]. Utilize the Naive Bayes Algorithm, Maximum Entropy, and Support Vector Machine to manage the learning strategy that greatly aids in the success of sentiment analysis. The authors of [2] demonstrated that they used the Naive Bayes algorithm to achieve a maximum accuracy of 82.1 degrees. The accuracy was achieved in [25] by the authors using K-Nearest Neighbour classification for sentiment analysis. The authors of [26] presented a study on sentiment analysis of Twitter data using various methods. To analyze the sentiment and demonstrate the accuracy of features of various sizes, they used a variety of machine learning algorithms, including Maximum Entropy and Vector Machine.

IV. METHODOLOGY

The most well-known branch of artificial intelligence is natural language processing. It is the practice of interacting with machines through texts and other conversational content. Our goal is to communicate with a computer or smart device in human languages like French or Korean, not programming languages like Java or Python, which is why we talk about "natural language." When the emphasis is on the interdisciplinary fields of linguistics, computer science, and psychology, NLP is occasionally replaced by "computational linguistics." The data is cleaned of extraneous characters like stop words using a data filtering process, and the sentence structure is analyzed by linking the text analysis step. Last but not least, the result analysis step records the sentences deemed significant through grammar rules and keyword matching. Twitter was used as a data source to review vital information for those in times of need, our attention was on finding information on people who are missing. Understanding the context of the tweets was the first step. The tweets from the missing people were then categorized. To do this, we looked for and categorized current tweets about missing people. The intercepted Twitter was filtered using the corresponding word "none". The Python Tweepy library is used to create an OAuth connection to the Twitter API. Tweets were filtered and cleaned following the search. The During the filtering process, retweets from the downloaded tweet dataset. This was done because retweets usually contained overlapping details about a specific person or event. Link removal, words that were unintelligible, and words that might change the meaning of our sentence were all part of cleaning the data set. Each tweet in the output is separated by a line and saved in a text file. The execution of filtering procedures and the conclusion of data cleaning are shown in Figure 3.

V. TEXT ANALYSIS

It was necessary to perform some sort of semantic analysis on each line of the tweets. Classification carried out separating tweets that discussed human security from those that were unrelated to the security conversation by using the Natural Language Toolkit (NLTK) library [1]-[3]. Line segmentation, stop word removal, part of speech (POS) classification, and POS tagging are all parts of the classification process. Each tweet block was split into its individual words using line splitting. In order to ensure faster processing, stop words were eliminated, which involved removing words with little to no meaning in the context of each tweet. screening. Stop words on the list include "of," "in," "out," "and," "in," "a," "and," "in," "these," and "this." Other motifs and characters such as (';', '...', '!', '-', '\', '"', ',', ':', '!', '?', '#', '@,') were cleaned up as well. Additionally deleted were words with special characters and words whose definitions might have changed. After that, the words are labeled with the various parts of speech. Some of the speech tags were combined to create a unique classifier. This aids in understanding the context of tweets. The names of the selected tweets with the desired information are then combined into chunks with the desired extra information.

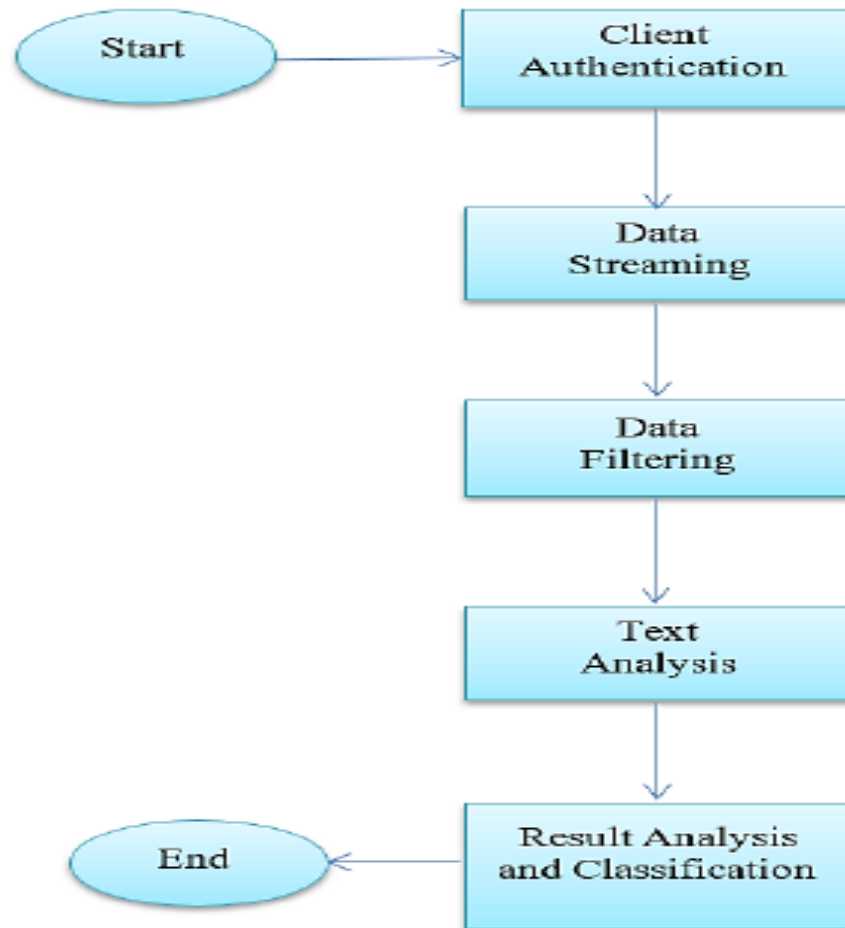


Figure 4

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-Asaad 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 9781138190672. OCLC 975370877.*
- [28] Jump up to:^{a b c d e f} *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 978-0-13-389256-7.*
- [29] *"What is structured data? - Definition from WhatIs.com". 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 978-0-13-389256-7.*
- [31] *Kitt, Denise (2012-05-24). "Enterprise Social Networks Explained". CRM Switch. Retrieved 2016-11-05.*
- [32] *Steele, Julie (2012-02-15). "Why data visualization matters". O'Reilly Media. Retrieved 2016-12-11.*
- [33] Jump up to:^{a b c} *"The Three Elements of Successful Data Visualizations". Harvard Business Review. Retrieved 2016-12-11.*