ANALYSIS ON SOCIAL MEDIA-A CASE STUDY

PANKAJ KUMAR JAIN SHAILENDRA CHOURASIA

Assistant Professor, CSE, GGCT Assistant Professor, CSE, GGCT

[pankajjain@ggct.co.in](mailto:pankajjain@ggct.co.in) [shailendrachourasia@ggits.org](mailto:shailendrachourasia@ggits.org)

ABSTRACT

As social networks gain more and more traction, the study of social network analysis has grown in significance and interest. The method of examining social structures using networks and graphs is known as social network analysis. The information on social networks is unstructured, so in order to use the important information, it is necessary to extract the organised information. The research that enables the systematic use of such a vast volume of scattered, unstructured information is focused on the extraction of data from social networks. In order to improve the precision of viewing the structured data scattered over the social network, natural language processing is used.The main goal of monitoring is to decipher important information from texts that ignorant social network users have submitted. In order to gather data about many kinds of things, relationships, or events, it analyses natural language text. By means of this research, the Natural Language techniques are being extensively examined. As the foundation of Natural Language Processing, text mining systems are used in this study article to evaluate information from social networks and to study sentiment analysis processing using machine learning.

Key words: Online social networks, text mining, sentimentanalysis, classification.

I.INTRODUCTION

Recent globalization of social media has resulted in a phenomenon that has drawn billions of users. Through social networking sites, individuals can create their own content to share in a variety of media, including information, private words, images, audio, and videos. As a result, social computing is established as a developing field of study and research that covers a variety of subjects, including Web semantic, artificial intelligence, natural language processing, network analysis, and big data analytics [1].

Online word-of-mouth is influencing people's opinions and choices more and more these days, which has caught the attention of many. Social media platforms can be broadly categorized as content-sharing websites, forums, blogs, and microblogs. People communicate with one another through exchanging information, messages, images, videos, and other sorts of content via content sharing websites (including Facebook, Instagram, Foursquare, Flickr, and YouTube). People publish specialized information, queries, or answers in Web user forums (like StackOverflow, CNET forums, and Apple Support). People can publish messages and other content on blogs (like Gizmodo, Mashable, Boing Boing, and many others) and exchange knowledge and ideas. Micro-blogs (like Twitter, Sina Weibo, and Tumblr) are only able to share brief texts for news and opinions.7].

The use of social media has changed greatly in recent years. Ideas can be exchanged between people with the touch of a mouse. One of the most popular social media platforms today is Twitter. Some of the subjects get a lot of interest. Some of these pieces of information circulated around a large group of individuals, and this process is referred to as "going viral." A company must examine these tendencies and then take the required actions to advance the company. The research community places a lot of importance on social media trends.

|  |  |  |
| --- | --- | --- |
| Type | Characteristics | Examples |
| Social Networks | A social networking website allows the user to build a web page and connect with a friend or other acquaintance in order to share user-generated content. | MySpace, Facebook, LinkedIn, Meetup, Google Plus+ |
| Blogs and Blog Comments | A blog is an online journal where the blogger can create the content and display it in reverse chronological order. Blogs are generally maintained by a person or a community. Blog comments are posts by users attached to blogs or online newspaper posts. | Huffington Post, Business Insider, Engadget, and online journals |
| Microblogs | A microblog is similar to a blog but has a limited content. | Twitter, Tumblr, Sina Weibo, Plurk |
| Social News | Social news encourage their community to submit news stories, or to vote on the content and share it. | Digg, Slashdot, Reddit |
| Media Sharing | A website that enables users to capture videos and pictures or upload and share with others. | YouTube, Flickr, Snapchat, Instagram, Vine |

Table 1: Social media platforms and their characteristics

Social networking has become a global phenomenon that has changed how people interact and communicate. Practically every aspect of our lives is influenced, including communication, politics, healthcare, and social connections. sharing online with people who have similar interests your blog entries, emails, photos, and videos. Also, it empowers individuals to aggressively express their ideas or viewpoints. [3].

A subfield of artificial intelligence in computer science is called natural language processing. The main focus of natural language processing is communication between humans and computers. Both Natural Language Processing and Natural Language Understanding are components of NLP.

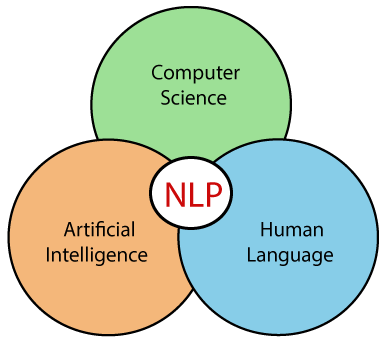


Fig-1- NLP as hybrid concept

This paper is divided into five sections: section 1 provides the introduction, section 2 discusses the aspect of Natural Language Processing, Section 3 discusses about Open-Source NLP Libraries, its challenges when imposed on social network and Section 4 provides the text mining approaches and section 5 concludes the paper.

# II. NATURAL LANGUAGE PROCESSING (NLP)

There are several alternatives available on social media engagement platforms. One of the most significant is through text posts. Natural language processing (NLP) of traditional media, such as printed news and articles, has become a popular research topic over the past 25 years. NLP frequently enables computers to extract meaning from natural language input using knowledge from computer science, artificial intelligence, and linguistics.

NLP for social media text is a new research area that necessitates either adapting current NLP techniques to these types of texts or developing brand-new techniques suitable for information extraction and other social media-related activities. Natural language processing is the practise of manipulating natural language by a computer (NLP).

Any written or spoken language that has human origins is considered to be a natural language. [Wik21] gives the following definition of NLP: In order to study how computers and human language interact, the field of "natural language processing," or NLP, combines linguistics, computer science, and artificial intelligence. Its main goal is to develop methods for teaching computers to process and analyse enormous amounts of natural language data. As a result, a computer can now "understand" the contents of documents, including the subtleties of their language when applied in particular situations. It might be simple to compare different writing styles by counting word frequencies at one extreme. On the other hand, NLP requires "understanding" complete human statements, at least to the point of being able to respond, according to [BKL09].

The first stage in social network data collection is identifying issues and opportunities. The information may be kept as text or it may be dynamically collected online and analysed in real time to satisfy specific requirements. The next stage is the SASM pipeline, which consists of specific NLP tools for data processing and social media analysis. Social media data are large, noisy, unstructured datasets. SASM transforms social media data into messages that are relevant and understandable using social information and knowledge. SASM then analyses the data from social media to produce social media intelligence. Social media intelligence can be made available to users or provided to decision-makers in order to foster better awareness, communication, and decision-making.

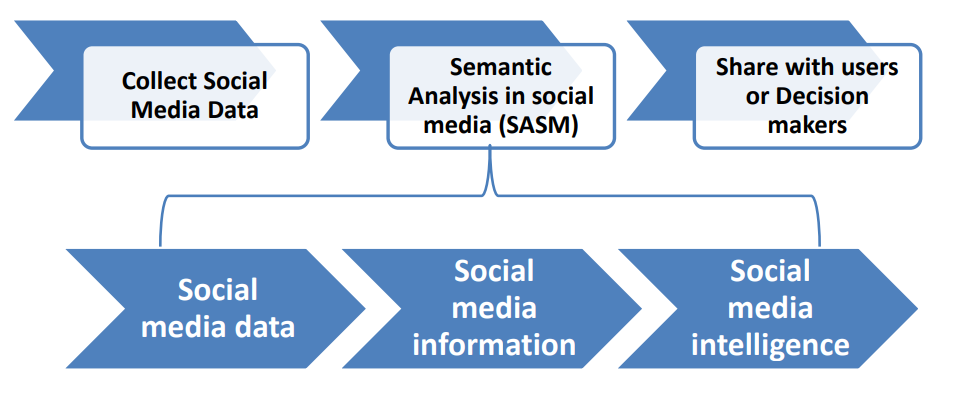


Figure 2: A framework for semantic analysis in social media, where NLP tools transform the data into intelligence

The presentation of analysed data by SASM could be completed by data visualization methods mining has developed across different scientific disciplines such as statistics, computer science, linguistics, and library science. Text mining focus on automated analysis of textual data as a form of natural language, its techniques deal with the unstructured text. Despite the absent of a unifying definition of text mining, there is an agreement on the general process of analysis. Furthermore, text mining is interconnected with Natural Language Processing (NLP) that is related to the analysis of natural languages. Due to the need of using automatic tools for analyzing the textual data and extracting relevant information, software solutions are available for analyzing social media applications. Text mining tools are used to identify and analyze posts, likes, followers in online social networks to explore people’s reactions and behaviour. Moreover, it shows the variation in views and opinions regard different topics. The fundamental process of text mining includes data collection, pre-processing, content analysis, finding and integration.

## Ambiguity and Uncertainty in Language

Ambiguity, generally used in natural language processing, can be referred as the ability of being understood in more than one way. In simple terms, we can say that ambiguity is the capability of being understood in more than one way. Natural language is very ambiguous. NLP has the following types of ambiguities −

### Lexical Ambiguity

The ambiguity of a single word is called lexical ambiguity. For example, treating the word silver as a noun, an adjective, or a verb.

### Syntactic Ambiguity

This kind of ambiguity occurs when a sentence is parsed in different ways. For example, the sentence “The man saw the girl with the telescope”. It is ambiguous whether the man saw the girl carrying a telescope or he saw her through his telescope.

### Semantic Ambiguity

This kind of ambiguity occurs when the meaning of the words themselves can be misinterpreted. In other words, semantic ambiguity happens when a sentence contains an ambiguous word or phrase. For example, the sentence “The car hit the pole while it was moving” is having semantic ambiguity because the interpretations can be “The car, while moving, hit the pole” and “The car hit the pole while the pole was moving”.

### Anaphoric Ambiguity

This kind of ambiguity arises due to the use of anaphora entities in discourse. For example, the horse ran up the hill. It was very steep. It soon got tired. Here, the anaphoric reference of “it” in two situations cause ambiguity.

### Pragmatic ambiguity

Such kind of ambiguity refers to the situation where the context of a phrase gives it multiple interpretations. In simple words, we can say that pragmatic ambiguity arises when the statement is not specific. For example, the sentence “I like you too” can have multiple interpretations like I like you (just like you like me), I like you (just like someone else dose)

NLP approaches that are essential for Social Network monitoring which are Automatic Summarization, Chunking, Part-of-speech tagging, Named Entity Recognition, Named Entity Disambiguation, Fact/Relation Extraction, Word sense Disambiguation and Sentiment Analysis.

1. Automatic Summarization Automatic Summarization is the process of reducing a text document with the help of a computer program in order to create a summary that retains the most significant points of the original document.

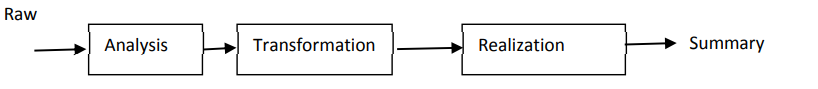


Fig3.-Process of Auto Summarization

1. Chunking - Chunking is the basic technique used for entity detection. Chunking selects a subset of the tokens rather than tokenization that omits whitespaces. The pieces formed in the source text do not overlap as the output of tokenization.
2. Parts-of-Speech Tagging -Parts-of-speech tagging is a piece of software that reads text in some language and assigns parts of speech to each word such as noun, verb, adjective to name a few. Generally computational applications utilize more fine-grained Parts of speech tagging include tags like 'noun-plural'. Dictionaries have category or categories of a particular word which implies that a word may belong to more than one category. For example, ‘Run’ is both a noun and verb. Taggers employ ‘Probabilistic Information’ to solve this ambiguity.
3. Named Entity Recognition- Named Entity Recognition is a subtask of information extraction that seeks to locate and classify Named Entities in text into pre-defined categories such as the names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc.
4. Named Entity Disambiguation -The task of linking the identity of entities available in the text is referred as Named entity disambiguation. However, it is distinctive from named entity extraction as it identifies not the occurrence of names but their reference. It needs a Knowledge Base of entities to which names can be linked.
5. Fact/Relation Extraction- Once named entities have been identified in a text, we can then extract the relations or facts that exist between specified types of named entity. The objective of the fact extraction is to detect and distinguish the semantic relations between entities in text or relations and fill it in a predefined template using the entities.
6. Stemming and lemmatization are 2 popular techniques in NLP. Both normalizes a word but in different ways.
   1. Stemming: It truncates a word to its stem word. For example, the words “friends,” “friendship,” “friendships” will be reduced to “friend.” Stemming may not give us a dictionary, grammatical word for a particular set of words.
   2. Lemmatization: Unlike the stemming technique, lemmatization finds the dictionary word instead of truncating the original word. Lemmatization algorithms extract the correct lemma of each word, so they often require a dictionary of the language to be able to categorize each word correctly.
7. Word Sense Disambiguation This is an open NLP and ontology subject that identifies the correct sense of the word in a sentence where multiple meanings of the word exist. It’s easy for a human to understand the significance of a word based on the basis of its background knowledge of the subject. However, identification the aspect of the word is difficult for a machine to understand. This methodology provides a mechanism to diminish the ambiguities of words in the text. For example: Word Net is a free lexical database in English that contains a large collection of words and senses
8. Sentiment Analysis Sentiment Analysis is an NLP process which identifies, extracts, enumerates the attitude of the user to the information that is provided by the user in a free form text.

# III. OPEN-SOURCE NLP LIBRARIES

# The algorithmic foundation of NLP in practical applications is provided by NLP libraries. To build up or deploy servers and infrastructure, it offers a free API.

# • Apache OpenNLP: This machine learning toolkit is open source and offers natural language text. It offers several services, including part-of-speech tagging, named entity extraction, summarization, searching, information grouping, tokenization, natural language production, feedback analysis, and more. It offers a command-line interface where models can be trained and tested with some predefined models.

# • Natural Language Toolkit (NLTK): This well-known Python toolkit offers modules for parsing, categorising, tokenizing, stemming, sematic reasoning, and other text processing tasks. It offers friendly user interfaces for more than 50 corpora and lexical resources, including WordNet.

# • Stanford NLP is a collection of NLP tools that includes named entity recognition, sentiment analysis, part-of-speech tagging, and more. It offers tools for statistical NLP, deep learning NLP, and rule-based NLP, all of which are widely utilised in business, academia, and government.• MALLET: This Java programme offers Latent Dirichlet Allocation, document classification, clustering, topic modelling, information extraction, and more features.

# • Genism-gensim is a powerful library for semantic modelling that may be used in a variety of applications.

# Stanford's CoreNLP A variety of human language technology technologies make up CoreNLP. It seeks to make it simple and effective to apply linguistic analysis methods to a document.

# • SpaCy is a Python-based open-source NLP library. It allows you to create applications that process and comprehend massive amounts of text because it is specifically intended for use in production environments.

# CHALLENGES IN NLP

Informal language: On social networking sites, people post messages that are noisy, casual, and that lack punctuation, spelling errors, capitalization, and grammatically accurate sentences.

• Part-Of-Speech tags make it harder to extract information from social networks.

• Brief contexts: Social networks, like Twitter, have minimal duration requirements. As a result, the user utilises more acronyms in their posts to provide more detailed information. Due to the posts' briefness and the co-references among the feeds, it is challenging to resolve specified entities.

• Noisy scant contents: Social network posts by users don't always contain pertinent information. Filtering is needed as a preliminary processing step to clean up the input postings stream.

Information about entities: As people typically utilise social networks to express information about their everyday activities, events, or happenings, the knowledge base does not contain information about entities. The methods used for information extraction connect the entities involved in the retrieved information to a knowledge base. A new suit for information extraction from social network posts is required.

• Unknown contents: Not every information on social networks can be relied upon. Contributions from users often contain information that is inconsistent with reliable outside sources. It is challenging to deal with the uncertainty in the retrieved relations and facts.

1. CASE STUDY NLP BASED SOCIAL MEDIA ANALYSIS

The methodology has been developed based on the objective of the study, that is, to explore the recent text mining techniques applied in social networks analysis. The study is to explore the following aspects:

* The text mining techniques that most used in social networking.
* The text mining algorithms that most applied in social networking.
* The tools used to perform the opinion mining / sentiment analysis.

A) .Data Collection -This is the process of gathering and measuring information in a systematic manner, which then enables one to answer relevant questions and evaluate outcomes. It deals with the challenge that updated information can be searched for a couple of days and the previous ones are not found.

We used the Twitter API to accumulate tweets. Appropriate hashtag (#) can be used to filter and gather the specific class of tweets from the enormous collection. These hashtags were elected after investigating the top trendingtags from online assets and Twitter itself (Table-2).

Table-2.Trending Hashtag on Tweeter

|  |  |
| --- | --- |
| Twitter Hashtags | |
| #NewEducationPolicy2020 | #Facebook |
| #Socialmediamarketing | #Onlinebusiness |
| #digitalmarketing | #love |
| #socialmedia | #photography |
| #marketing | #startup |
| #branding | #ecommerce |
| #business | #motivation |
| #marketingdigital | #success |
| #seo | #webdevelopment |
| #onlinemarketing | #fashion |
| #contentmarketing | #content |
| #instagram | #facebookmarketing |

B) Pre-processing - This step refers to the processing of raw data to deliver a podium for data analysis. Preparing text data for future processing is the goal of the pre-processing stage. Pre-processing the data is a crucial step because it determines how effectively the subsequent processes will work. The tweets must be syntactically corrected as needed.

* Eliminate Unicode strings and noise
* Swapping URLs and user mentions
* Convert the text to lower
* Replacing slang and abbreviations
* Removal of Special symbols
* Removing numbers
* Swapping negations with antonyms
* Eliminating punctuation
* Eliminating stop words
* Removal of web links
* Spelling correction

The following steps are carried out during text pre-processing: • Tokenization, which involves dividing the text into sentences and the sentences into words. We delete the punctuation and lowercase the words.

Tokenization is the process of fragmenting a text into smaller components known as tokens. Tokens can be letters, numbers, or commas. Another name for it is word segmentation.

Beds and chairs are examples of certain sorts of furniture.

Bed and Chair

The packages and the details of their installation are as follows −sent\_tokenize package, word\_tokenize package, WordPunctTokenizer package [11].

Stemming - For grammatical reasons, language has many varieties. variations in the sense that both English and other languages have a variety of word forms. Examples include democratic reform, democracy, and similar concepts. It is essential for machine learning projects that the machines recognise how many different words, like the ones stated above, have the same fundamental structure. As a result, when performing text analysis, it is very beneficial to remove the words' root forms from the text. The several stemming packages that are provided by the NLTK module include PorterStemmer, LancasterStemmer, and SnowballStemmer.

Lemmatization applies a morphological analysis to words by taking into account a language's entire lexicon in addition to word reduction. For instance, third-person terms are transformed to first-person.

• Normalization is the process of turning a list of words into a more standardised order. This is helpful when getting text ready for processing later. Other operations can interact with the data and won't have to deal with problems that might jeopardise the process by converting the words to a standard format. For instance, making all terms lowercase will make searching easier.

• Word embeddings, also known as vectorization, is a technique used in natural language processing (NLP) to convert words or phrases from a lexicon into a corresponding vector of real numbers. This vector is then used to identify word predictions, word similarity, and semantics. Vectorization is the process of turning words into numbers.C). Data Analysis

A feature is a sizeable body of information that can be interpreted as a particular that aids in problem-solving. Excellence and magnitude are important qualities since they have an emotional impact on the results produced by the selected model. Feature extraction is an important phase in the sentiment classification process since it involves extracting relevant information from the text data and directly affects the model's performance. Feature extraction increases the accuracy of learned models by using features derived from the input data. The process of extracting features from tweets is the selection of useful terms [9].

•Unigram features: Each word is individually assessed to see if it qualifies as a feature.

• N-gram features, which consider several words simultaneously.

Usage of a group of words with specific positive or negative meanings is known as using an external lexicon.

WORDCLOUD

A Word Cloud is an interactive visualisation that dynamically changes the size and colour of words based on how frequently they are used in social posts in your social wall feed [8].



Fig-4 word cloud of social media (Ref.<https://www.flickr.com/photos/bitsfrombytes/45219986531> )

In order to draw a conclusion, we can put into practise eight elements derived from the social cognition theory as intangible features causing user behaviour, such as trust and distrust, belief and expectation, benefit and risks, positive and negative social impact, influencing study, policy and execution, ethics, trends, etc.

D). sentiment analysis tools

By using classification, it is possible for different statistical categories to be unintentionally labelled into multiple groupings depending on certain characteristics of the data. The basic approach is to assess the emotional intent of the sentence and identify the characteristics of the product using word frequency statistics. By creating a classifier using labelled training data, we proposed a novel way for categorising eight features from the user's behaviour aspect. For these crucial components, our recommended strategy is novel [22].

-

First, a sentiment categorization based on the analysis of public conduct views is established.

- Second, we relate the technology of natural language process to a novel perceptual wisdom and grouping.

Important proof for our approach comes from the ability to identify emotions based on user actions by fusing machine learning with a lexical vocabulary. Many characteristics can be utilised to predict a person's behaviour, as was already mentioned.

In order to analyse user behaviour, we place a strong emphasis on categorising a notion that may be utilised to represent user intentions and feelings surrounding e-word. In user behaviour analysis, we may also think of this as an emotional multi-classification problem. The majority of earlier research on emotion grouping concentrated on positive and negative classification, or binary classification. Nonetheless, it is frequently more appropriate to announce additional specific supporting data regarding various classification methods. Because there are multiple elements that can influence whether users accept a new technology or not, these influencing factors are broken down into different categories. Eventually, a meta-classifier that combines the predictions of various classifiers into a final classification result can be created using a variety of classification techniques.

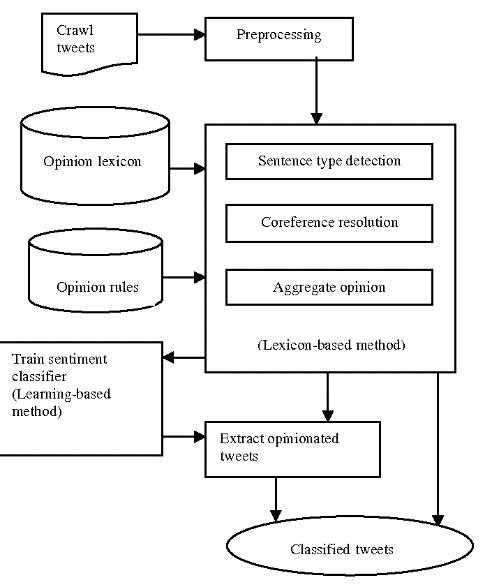


Fig.5: Architecture of sentimental analysis

The development of sentiment analysis can be improved by employing a hybrid technique. The programming language chosen for usage while putting this tactic to the test is Python. After pre-processing, the proposed algorithm consists of three steps. The first step, which is concerned with determining the optimum operating constraints for the classifier, deals with the lexicon-based model. On the other hand, in the second section, the learning-based paradigm is discussed and its implementation is reviewed. It is mentioned as the hybrid model to be examined in the third and final section.

To determine the polarity of a tweet, the lexicon-based method makes use of pre-built lexicon assets that contain divergence of sentiment arguments, such as SentiWordNet (SWN) 3.0 [10]. The lexicon-based method leverages pre-built lexicon assets containing divergence of sentiment arguments, such as SentiWordNet (SWN) 3.0 [10], to formulate a tweet's polarity.

To determine the polarity of a tweet, the lexicon-based method makes use of pre-built lexicon assets that contain divergence of sentiment arguments, such as SentiWordNet (SWN) 3.0 [10]. The lexicon-based method leverages pre-built lexicon assets containing divergence of sentiment arguments, such as SentiWordNet (SWN) 3.0 [10], to formulate a tweet's polarity.

In the combined lexicon-based and machine learning approach, the outcome from the lexicon-based method was used in the machine learning method as a training data to train the two classifiers (Naïve Bayes, SVM classifiers). Then, the same testing data (unlabelled data) was used to inspect the effectiveness of the two classifiers which provided Consequence.

The features of achieving foundational literacy, elasticity, multidisciplinarity, universal acceptance, education, importance of positive belief, understanding, inspiration, and life-threatening thinking, integrities and human & legal values, multilingualism, widespread use of technology, and topographical challenges can all be analysed in different ways [8].

The accuracy score evaluates how well a SA model predicts sentiment discussion after applying a number of machine learning classifiers (Naive Bayes, SVM, DT, and ME). By dividing the total number of events by the true positive and true negative rates, it is calculated. The results confirmed that applying feature selection techniques leads to improved performance.

1. CONCLUSION AND PERSPECTIVES.

In social media platforms, user-generated text material and text-based communications often feature a variety of writing styles, subjects, and languages as well as typos, freeform emojis, and abbreviations. It is more difficult to use social media data for NLP tasks due to the variety of content and language. Emoji removal, hyperlink removal, language detection and translation, and typo correction are just a few of the cleaning approaches mentioned that have been shown to be helpful in priming and pre-processing language of this kind. These techniques produce cleaner output when applied to text before other Natural Language Processing (NLP) techniques like named-entity identification and key phrase extraction.

In addition to text data, social media data also includes statistics that reflect how people interact with and are interested in various kinds of information. Inferences drawn from named-entity recognition (NER) and key phrase extraction NLP approaches, when combined with these statistics, have been proven to be useful for trend analysis, analytics, and identifying correlations and affinities between user involvement with social media.

Even though the current research's findings are reliable, more can be done. First, in the situation of cross-domain, the model's accuracy cannot be guaranteed. Second, training the model is challenging when the information is insufficient, and larger and more varied databases are required to increase the model's accuracy. Finally, category labels must be manually annotated, which involves heavy manual labour.

REFERENCES

[1]. Anna Atefeh Farzindar, Diana Inkpen,” Natural Language Processing for Social Media” Third Edition, Springer Nature Switzerland AG 2022, DOI 10.1007/978-3-031-02175-6

[2]. Hiba A. Abu-Alsaad, Rana Riad K. AL-Taie,” NLP analysis in social media using big data technologies”, TELKOMNIKA Telecommunication, Computing, Electronics and Control Vol. 19, No. 6, December 2021, pp. 1840~1846

[3]. Huda N Alhazmi, Text Mining in Online Social Networks: A Systematic Review ,IJCSNS International Journal of Computer Science and Network Security, VOL.22 No.3, March 2022

[4].Jyotika Singh, Social Media Analysis using Natural Language Processing Techniques, PROC. OF THE 20th PYTHON IN SCIENCE CONF. (SCIPY 2021)

[5]. CharuVirmani , Dr. Anuradha Pillai , Dr. Dimple Juneja,” Extracting Information from Social Network using NLP”, International Journal of Computational Intelligence Research ISSN 0973-1873 Volume 13, Number 4 (2017), pp. 621-630,

[6]. Glen Coppersmith, Ryan Leary, Patrick Crutchley and Alex Fine,” Natural Language Processing of Social Media as Screening for Suicide Risk” Biomedical Informatics Insights Volume 10: 1–11 © The Author(s) 2018 Article reuse guidelines: sagepub.com/journals-permissions DOI: 10.1177/1178222618792860.

[7]. Hui Li, Qi Chen , ZhaomanZhong , Rongrong Gong , Guokai Han ,E-word of mouth sentiment analysis for user behavior studies, Information Processing & Management Volume 59, Issue 1, January 2022, 102784, <https://doi.org/10.1016/j.ipm.2021.102784>

[8].H.S.Hota,DineshK.Sharma,NileshVerma(2021), Lexicon-based sentiment analysis using Twitter data: a case of COVID-19 outbreak in India and abroad, <https://doi.org/10.1016/B978-0-12-824536-1.00015-0>

[9]. Gamal, D., M. Alfonse, E.-S. M. El-Horbaty and A.-B. M. J. P. C. S. Salem (2019). "Implementation of Machine Learning Algorithms in Arabic Sentiment Analysis Using N-Gram Features." 154: 332-340

[10]. A Comparison of Machine Learning Techniques for Sentiment Analysis,DOI: <https://doi.org/10.17762/turcomat.v12i3.999>.

[11]. Aastha Tyagi, A Review Study of Natural Language Processing Techniques for Text Mining, International Journal of Engineering Research & Technology (IJERT), http://www.ijert.org ISSN: 2278-0181 , Vol. 10 Issue 09, September-2021