JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 07 Issue 02
2024
Research Article
https://doi.org/10.56979/702/2024

Sentiment Analysis of Social Media Data: Understanding Public Perception

Shumaila Mughal[1*], Arfan Jaffar[1], and M Waleed Arif[1]

[1]Department of Computer Science, Superior University Gold Campus, Raiwind Road, Lahore, Pakistan.
[*]Corresponding Author: Shumaila Mughal. Email: shumailamughal05@gmail.com

Received: March 28, 2024 **Accepted:** July 27, 2024 **Published:** September 01, 2024

Abstract: This paper provides a sentiment analysis model that combines dimensionality reduction, part-of-speech tagging, and natural language processing (NLP) for social media data. The model uses machine learning methods (Naive Bayes, Support Vector Machine, and K-Nearest Neighbor) to categorize sentiment as positive, negative, or neutral properly. The model's performance was assessed using two datasets and compared to other sentiment analysis algorithms that were already in use. The outcomes show increased performance and offer perceptions of the public's thoughts on various topics. This work addresses the problem of language-specific models and advances the creation of accurate sentiment analysis models. In contrast to conventional polls, the study's conclusions present a novel viewpoint on public opinion and offer suggestions for improving the platform so that users can access additional options and conveniences. The proposed model has potential applications in social media monitoring, market research, and political analysis. Future work can extend the model to accommodate multiple languages and explore the use of deep learning techniques. By providing a more accurate and efficient sentiment analysis tool, this research contributes to the growing field of social media analytic and its practical applications.

Keywords: Natural Language Processing (NLP); Opinion Mining; Sentiment Analysis; Text Classification.

1. Introduction

Sentiment analysis and user-generated data across several social media platforms have significantly increased in this sophisticated environment. Using social media to gain knowledge has seemed to resemble dominating automation. The number of people using social media is rising so quickly that in 2019 there were up to 2.77 billion SM (social media users) globally, and by 2022 there will be 4.59 billion. Concerning this rise, a significant amount of text, audio, video, and photo-based content is exchanged and published. The study of people's thoughts and feelings is known as sentiment analysis, and it has become increasingly popular over time [1].

We examine the content, sentiment expressions, and psychological content of Twitter tweets. We investigate how social media features, such as the presence or absence of photographs, influence the various linguistic ideas that appear in tweets. We look at how Twitter tweets' sentiment and searches on Google and Wikipedia relate to each other, building a model that connects search to social media. We also look at the relationship between a dictionary of important COVID terms included in Twitter tweets and the likelihood that messages sent on social media will be tweeted and enhance the reputation of the sender within social media [2]. Public opinion about different statements or remarks made after an election, particularly the 2024 presidential election, will be the subject of analysis in this study. The rationale behind selecting this subject is that Indonesians have been debating it extensively. Finding out how to get Twitter data relating to related topics—specifically, popular opinion regarding presidential candidates following the 2024 presidential debate—is the aim of this study to determine the degree of accuracy that the Naive Bayes Classifier algorithm offers for sentiment analysis as well as the sentiment analysis capabilities of the Naive Bayes Classifier algorithm. More specifically, to better comprehend complicated text analysis systems, text mining makes use of technologies including batch processing, natural language processing,

information extraction, and data retrieval. To uncover potentially hazardous behavior and other security risks, text mining was first employed to give intelligence to governments and security services. Text analysis and technology from outside fields like computer science, human resources, machine learning, and statistics are used in text mining to boost productivity. Since then, allied areas have made substantial use of correction as a technique since it is vital for performing practical research [3]. As part of the modelling process, classification involves assigning labels or classes to pieces of data based on preexisting attributes [4].

Three different levels of sentiment analysis have been identified: sentence-level analysis, document-level analysis, and feature-level analysis. Through sentiment analysis of multimodal data, users' opinions and attitudes on most topics are made understandable by the piece of data that is readily available online. Compared to other things like political elections, movie box office predictions, and public book readings, and many more, multimodal sentiment analysis provides unique advantages. The datasets were painstakingly vetted by hand-annotating social media information. The Kappa score was used to confirm the dataset's quality, yielding an astounding 0.97 agreement.

The rising interest of scientists and engineers in identifying sentiment patterns in the social media domain is the driving force behind this study. Sentiment analysis helps identify emotional tones like grief, happiness, anger, excitement, excellence, surprise, and more by automatically identifying feelings conveyed in textual data [5]. Tweets, which are 140-character messages, may be posted and read on this social networking platform. Users may debate with billions of other users on this platform, promote their research, and share their thoughts and opinions in succinct, direct conversations. There isn't always a mutual relationship between all users in the Twitter network. It is either a directed or unguided connection in this case. Due to the massive amount of data, this microblogging platform provides, including user information, the number of followers and followers in the network, and tweet messages, most studies have investigated and assessed the various interpretation techniques to obtain the most recent innovations [6].

Sentiment analysis, often referred to as opinion mining, is textual context-specific processing that locates and extracts subjective data from the source material. Monitoring online discussions, helps organizations enhance the human element of their brand, products, or organization Sentiment analysis has become more relevant in the contemporary setting due to social media's rapid growth [7].

Social media sites like Facebook, Instagram, and Twitter have grown in importance in today's world as a result of the quick and massive advancement of information technology. These platforms have had rapid growth and a significant influence on people's everyday lives in the past several years. They are used by a large number of users not only to socialize and share their lives with new people, but also to communicate their thoughts and feelings about a range of goods, services, and companies through postings and comments [8]. Emails, chat transcripts from customers, and user evaluations of goods and services may all fall under this category. For organizations to make well-informed judgements, it would be imperative to identify the patterns in this massive volume of data that may aid in making critical decisions. Companies are competing with one another these days to gather, examine, and use patterns that might provide insight into the opinions of their stakeholders and consumers. This involves examining how society speaks about and feels about new technical advancements and items to understand how people behave and react to them. The objective is to use customer experience data and online sentiment analysis to better comprehend people's needs and emotions. Real-time data is collected, processed, and prepossessed in several ways to do this.

1.1. Machine Learning

The development of computer programs that mimic human intelligence is known as artificial intelligence (AI). Initially, researchers working on ID recognized that intelligent machines needed to be able to learn from their surroundings and adapt to new input regularly. This means that machine learning is a branch of AI that deals with computer algorithms that can learn from data and knowledge without human intervention and improve themselves over time [9]. According to reports, machine learning has deep roots and interacts with numerous other fields, including statistics, computer science, information theory, and many more. This, among other reasons, explains why it has grown rapidly over the past decade and continues to pique the interest of many researchers. Nowadays, there is an abundance of data available, which allows for the training of machine learning models. Thanks to the increased processing power of modern computer processors, several areas, like contextual machine translation, picture face and

object recognition, and many more, have made tremendous strides, producing high-quality results at significantly faster speeds. But for now, there isn't a system that can do everything that needs to be done in just one task. So, for example, it's not feasible to use a single machine learning model to translate across all languages or to identify thousands of distinct items. We found that machine learning models work well when broken down into smaller sections, like those between two languages or a handful of objects. This suggests that systems for identifying objects among thousands of categories or translating between dozens of languages could be developed soon. Among other things, machine learning algorithms can learn probability and statistical data from a data set and use that information to create an equation-based set of rules for solving mathematical problems. There are a variety of methodologies and methods of machine learning available, depending on the purpose that each system aims to accomplish. To help you see them better, we'll highlight a couple of these tactics and provide some examples of how they work: Keep in mind that several ways and processes might be used to achieve some of the examples given below. First, categorization:

This indicates that the algorithm gives a way to classify the input according to the available options. Binary and multi-focal classification issues exist. Machine learning-based complaint categorization for an online retailer D. Rajendra Kumar K. and S. Natarajan Feature extraction and classification can be applied to a wide variety of tasks, including sentiment analysis, map imagery, and fraudulent identification. Logistic Regression, Support Vector Machines (SVM), the Naive Bayes family, k-nearest neighbors, and random forests are among the classification methods. The second technique is clustering, which groups similar things into multiple clusters defined by shared characteristics. When making decisions based on the outcomes of objects grouped under the same cluster, clustering algorithms like K-Means and Singular-Value Decomposition (SVD) find extensive usage in recommendation system design and targeted marketing. 3. Dimensionality Reduction: This technique mapped the data from a high-dimensional format to a lower-dimensional format while preserving rich information and significant features. Data management, noise exclusion, and valuable feature extraction are all helped along with computational time and cost savings. Some of the most popular techniques include Eigen faces for face and image recognition and principal component analysis (PCA) for text mining. The learning strategy is a common denominator among the aforementioned and unnamed machine learning techniques; furthermore, the data used for training, both as input and output and for other purposes, varies between the various methods. The three main types of educational strategies will be covered: 1. whether one of them could converse with the other while simultaneously conversing with other individuals is not made apparent in the story. Social media is full of current, raw information, as well as cutting-edge tools, such as machine learning (ML) and artificial intelligence (AI), which enable data to be processed and transformed into knowledge that is relevant to the general public. Through an examination of relevant compositions from 2015 to 2022, the paper will provide a clear grasp of the use of sentiment analysis on social media platforms. This technique uses natural language processing (NLP) to abstract and alter information from social media platforms, classifying it as either good, bad, or neutral. For over 15 years, a large number of academics have been studying sentiment analysis and publishing articles in journals. This field is currently rapidly expanding. Three different levels of sentiment analysis have been identified: sentence-level analysis, document-level analysis, and feature-level analysis. Through sentiment analysis of multi-modal data, the internet provides a plethora of information that facilitates understanding users' opinions and attitudes on a wide range of topics. Local In supervised learning, the goal is to produce an equation for the datasets that includes the input variables and the predicted result. Open sans before an algorithm processes the data vectors, the user supplies training samples that include feature values and their correct class. The next step is to apply an algorithm to unlabeled data to label it. It is clear from the problem statement that this is a classification problem requiring a supervised solution. Load-based learning: Contrarily, unsupervised algorithms differ from their supervised counterparts in that they do not require developer input regarding targets. These algorithms examine the structure of the input data and determine which patterns in the output data to augment with new data based on the data's similarities and differences. Using unsupervised approaches, clustering techniques are implemented. Semi-supervised learning combines supervised and unsupervised learning and involves controlled learning using a set of labels. The semi-supervised learning training approach makes use of both large amounts of unlabeled data and a small amount of labeled data. The shortcomings of the first two learning approaches inspired the

development of semi-supervised learning. The problem with supervised learning is its expensive approach to data categorization, while the unsupervised format has a very limited range of applications. In this case, the semi-supervised learning method involves classifying and grouping the unlabeled data by making predictions about the unlabeled items based on the labeled data to further the use of sentiment analysis We have all the information on any good, service, location, or event that enables sentiment analysis research, and the majority of the information or data required came from social media. Marketing professionals are increasingly seeking modern approaches that provide an approximated label presentation when evaluating label equity due to the growth of big data analytic in the modern era. Activities rely on traditional methods of data collection and inquiry, such as questionnaires and one-on-one or preliminary interviews which are notably effective. One of the papers they submitted is a computational model that integrates sentiment and subject organization to extract significant insights from customer understanding of social media. A computational approach that integrates subject and sentiment organization to extract potent insights from customers' insight into social media is one of the studies they present. Its dummy develops a narrative genetic algorithm that amplifies a set of tweets in meaningful logical groupings, which function as necessary circumstances while attempting to penetrate the dominant topic in a massive information structure. They utilized the Uber matrix, which is based on information obtained from Twitter, to comprehend the reasoning behind their dummy. Its conclusion is client-available and creates awareness for two fundamental label equity dimensions: label meaning and label consciousness. Social media has dispersed a massive amount of data. Every day, billions of individuals exchange messages and tweets, discussing their solar day. Social media platforms provide the finest chance to investigate human features, spread information, and circulate effects at a level that is hard to understand or could even be thought to be impossible because of this enormous endeavor. The newest diamond cache for data mining and predictive analytics is thought to be this social media data. Even if we are aware that social media data is different from traditional data, learning about the essential research techniques and distinctive qualities of this kind of data is still exciting. In this effort, I'm trying to look at bias in information or data gathering. Application Programming Interface (API) ambush, which operates on algorithms, information/data, and substantiate results, is the most recent kind of information unfairness. This kind of work enables us to comprehend how various information kinds, concepts, and social media information/data features may be extensively taken into account and validated, as well as how to create arbitration processes to manage potentially negative outcomes. This kind of work aids in our comprehension of the many information forms, concepts, and attributes of social media information/data that may be taken into account and verified, as well as how arbitration procedures might be used to regulate negative outcomes. It is possible to comprehend the different goals of social media data and information by carefully considering the process and findings of this endeavor. A 2017 article examined public opinion on a new school food program aimed at preventing juvenile obesity. It identified characteristics related to those opinions and acknowledged potential regional and gender differences in the U.S. Between February 9, 2010, and December 31, 2015, it gathered 14.317 relevant tweets from 11,715 clients about the approval of the national policy. They use belief-mining techniques to separate tweets into groups that are good, negative, and neutral. They can also control text analysis to get knowledge about the characteristics of a belief turn of phrase, such as earmark, receptacle, origin, and goal. These days, a lot of research is being done to investigate machine learning methods and language processing from posts on social media and other online platforms to gauge public opinion on significant problems. According to a 2018 study by Gurrajala and Matthews that examined 6,000 tweets about transportation and air quality, people who used public transportation were more likely to tweet negatively about the quality of the air, while people who drove their cars were more likely to tweet positively. The majority of prior sentiment analysis research has demonstrated that people's sentiments vary depending on the experiences and events they go through. There aren't many studies that demonstrate how attitudes towards eco-friendly transportation and air quality have changed because COVID-19 is such a new occurrence. With the help of this review, we can say that our study delves further into this subject and aids in comparing and analyzing the shifts in people's opinions on these subjects expressed through tweets between before and after the epidemic.

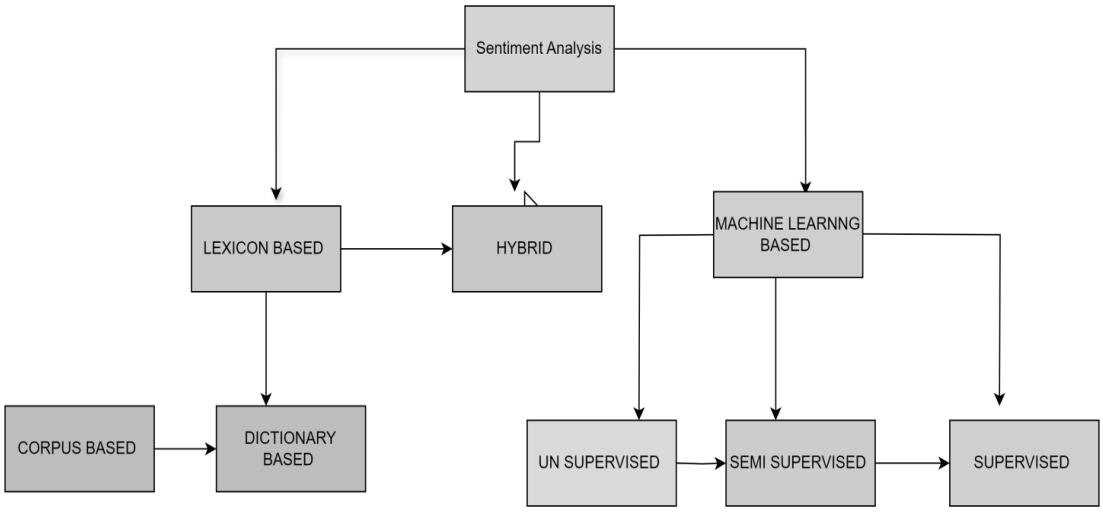

Figure 1. Sentiment analysis using Twitter data: a comparative application of lexicon- and machine-learning-based approach

2. Literature Review

This social media data is said to be the most recent gem cache for predictive analytics and data mining. We are aware that social media data differs from traditional data, and it is fascinating to learn about its unique characteristics and critical research methodology. Wu, Y. Liu [2022] [9]. Researchers and scientists are always trying to create a sentiment analysis model that is both accurate and useful. As they develop an accurate sentiment analysis algorithm that is not limited to English, they encounter several challenges. To further the use of sentiment analysis. We have all the information on any item, service, location, or event that allows sentiment analysis research; the majority of the information or data required came from social media. The primary point of contention with current operations is that they rely on traditional methods of data collection and inquiry, such as questionnaires and one-on-one or preliminary interviews, which have a significant amount of sway. A computational approach that integrates subject and sentiment organization to extract potent insights from customers' insight into social media is one of the studies they present. T. Zhao, [2013] [1].

Sentiment analysis has the potential to become a platform that can compete with advanced polling methodologies for political election prediction, as evidenced by the more reliable data analyses from Twitter, which has a 94% correlation with polling data. Joyce, Brandon, and Jing Deng. [2017] [10]. Sentiment analysis may be used to examine global events, including sporting events, disasters, and events. Mahtab, S. Arafin, [11, 12].

Using the six-step principles for performing a systematic literature review in management, a systematic review was conducted. First, the research question has to be defined. Next, ascertain which attribute is necessary for the research. Proceed by gathering possibly relevant books and choosing appropriate reading. Next, we combine pertinent data from the literature, and the last stage is to present the review's findings. Durach, Christian F [2017] [13].

The review made use of five reliable and respectable internet resources that produced work in the fields of computer science and information. "Sentiment analysis, social media, Facebook, Twitter" is the search string term utilized for all five internet databases. 407 items in total were found through the database search. 34 articles from Emerald Insight, 244 from Science Direct, 24 from the Association for Computing Machinery (ACM), 54 from Scopus, and 51 from IEEE were found. (NLP) is information extraction (IE), which identifies and extracts structured information from unstructured text automatically. This entails turning unstructured text into structured data that may be utilized for building knowledge graphs, populating databases, and making data-driven decisions, among other things. Named entity recognition (NER), relation extraction, and event extraction are the main tasks in information extraction. Firstly, sentiment analysis has been the topic of several studies in the past. Sentiment analysis of user-generated data from various social networking websites, such as Facebook, Twitter, Amazon, and others, is the focus of current research in this field. The majority of sentiment analysis research relies on machine learning

algorithms, whose primary goals are to determine the polarity of a text and determine if a particular text is in favor of or against something. This chapter will provide us with an overview of some of the research projects that have deepened our understanding of the subject. O. Grljevic, Z. Bosnjak [2020] [15].

The profusion of user-generated content and the growth of online platforms have attracted a lot of interest in sentiment analysis in social media data in recent years. An extensive overview of current research and methodology in this field is given in this part, which also highlights the many methods, strategies, and instruments used in sentiment analysis. There is also a discussion of each approach's advantages, disadvantages, and uses. Md M. Rahman, M.N. Islam [2022] [16].

Their primary goal was to categories text according to general attitude rather than merely subject, for as assigning a favorable or negative rating to a movie review. They use a database of movie reviews to test machine learning algorithms, and the findings show that these algorithms perform better than those created by humans. They employ support vector machines, maximum entropy, and Nave-Bayes machine learning techniques. They also conclude—after looking at several variables—that categorization of emotion is extremely difficult.

3. Objectives

With Twitter data from the 2010 us midterm elections, 96% of the results were accurate. Sentiment analysis for real-time data was proposed during the 2012 selection, and it was tested across several domains with the main goal of identifying genuine or fabricated political events. The machine learning technique classifies text data by training by using machine learning algorithms, such as Nb. Using linguistic and syntactic characteristics from the same or distinct domains, these methods may categories text into predetermined groups. There are three types of machine learning: semi-supervised, unsupervised, and supervised. Unlike unsupervised learning, which makes use of unlabeled training materials, supervised learning necessitates labeled training documents. A combination of supervised and unsupervised learning approaches is known as semi-supervised learning. We examined several studies that applied machine learning techniques to sentiment analysis. An NLP application that is thought to have connections to several intriguing topics is at the center of the current project and application, which is centered on the identification of "sentiment" in written text input. For example, recommendation systems and data about different amounts of evaluations on specific things could benefit greatly by extracting mood and emotion from passages. The current project and application are made to be accessible to a wide range of users, even those without any programming background. This is a generic term for a "machine learning model that understands the sentiment of a sentence" that describes a somewhat broad application. in particular, this thesis's application aims to do two things: (1) make it easy for users to analyses their sentences through the use of a graphical user interface, and (2) use various online sources to test groups of opinions and comments or specific keywords.

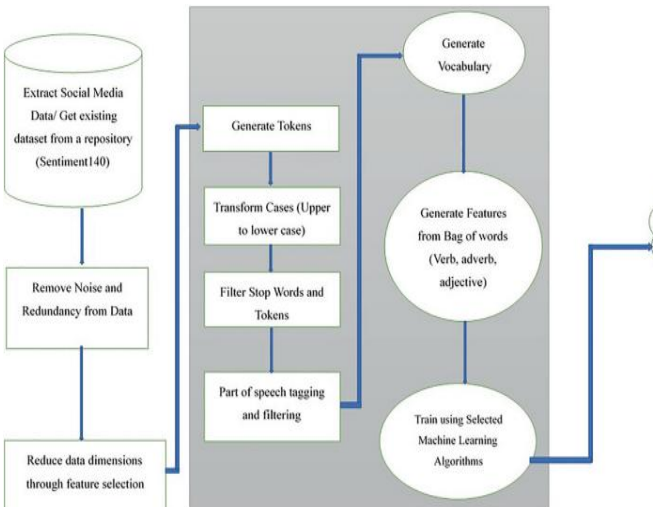

Figure 2. Block Diagram

Find how to present the results of the analysis in a manner that will allow a regular end-user to easily understand it. Visualizations therefore assist investigators in gaining an independent understanding

of trends or patterns of the sentiment of social media. Problem Statement Volume and Variety: for instance, social media yields a huge volume of information in textual, image, and video formats. Manual analysis of such a large data set and the nature of the data is not feasible because of the great amount of information. Social media content is typically lexical in ways that are much more a matter of degree than of kind, which complicates the identification of sentiment.

3.1. Solution

Natural Language Processing (NLP): Integrate text mining and data mining tools to processors as well as analyses social media data gathering. In addition to the ability to perform analyses on big data sets, NLP algorithms can also deal with vast amounts of text. Integrate models for sentiment analysis that have been fine-tuned for analysis on social media platforms. These models can help resolve the polarity of the text, where polarity indicates the positive, negative, or neutral stance of a social media post. Whenever possible, fine-tune models by using labeled social media posts to increase the model's sentiment analysis reliability. These models can input and analyses data and pass through varying scenarios as their main characteristic is patch ability. Forecast to integrate actual-time monitoring tools that will track the social media data as per the generation. It facilitates quick review in cases that are still undergoing investigation and analysis. Create algorithms that will say how important the post is from the social media platform to improve the sentiment analysis. NLP, or natural language processing, is a sub field of artificial intelligence that focuses on helping robots comprehend human language. Humans utilize natural language, which may be either text or voice, to communicate with one another. NLP can facilitate natural communication between humans and robots.

Sentiment analysis involves a procedure called text classification. It is the division of people's thoughts or expressions into several emotions. Feelings include Happy, Sad, Review Ratings, Positive, Neutral, and Negative. Sentiment analysis is a useful tool for analyzing people's opinions on a variety of consumer-focused companies and products.

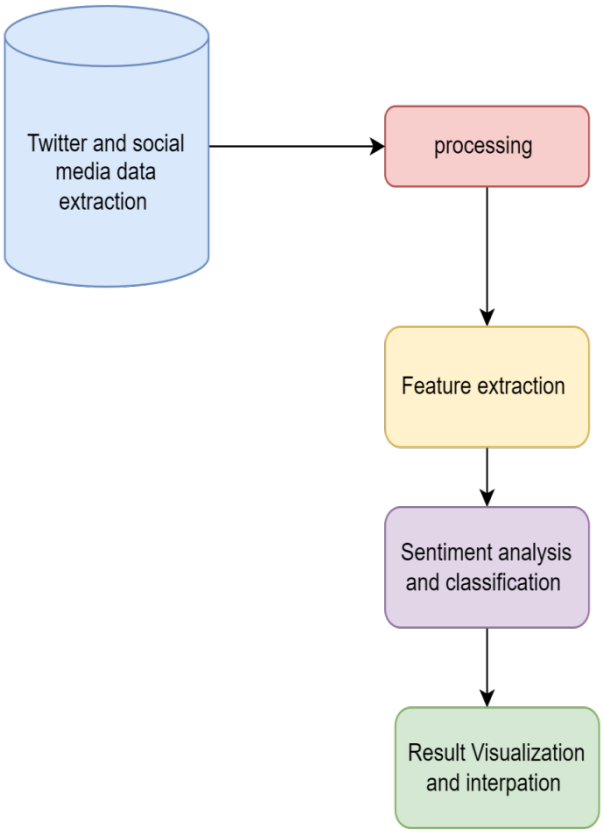

Figure 3. Showing Methods/Approach Results.

Contextual Understanding: Design machine learning methodologies that accomplish sentiment analysis by taking into account the context of the posts on the social media platform.

Human Oversight: Introduce human elements in the context of analysis, especially because certain results obtained through the usage of artificial intelligence tools may be dubious.

Visualization Tools: Interpretation of the delivered sentiment analysis must also be presented in a format that is easily understandable by everyone.

4. Implementation

In this section, we explain the results of applying the sentiment analysis framework described in Section III to text data samples from social media platforms. The process involved several key steps: Data gathering, data preparation, and data feature extraction, model learning procedure, real-time analysis and assessment. Determine the sentiment scores. The classify tweet with scores function accepts as inputs the p reprocessed tweet, log-likelihood values, and log priors. It then determines the sentiment scores for each category. The sentiment forecast is then given together with the individual sentiment scores, based on whatever sentiment score was obtained first. To categorize text based on computed scores, sentiment analysis jobs frequently employ this method.

0	stgnews	Bridger Palmer	Pine View High teacher wins Best in State awar...	ST. GEORGE — Kaitlyn Larson, a first-year teac...	https://www.stgeorgeutah.com/news/archive/2024...	2024-07-12T23:45:25+00:00	positive	Business
1	Zimbabwe Mail	Staff Reporter	Businesses Face Financial Strain Amid Liquidit...	Harare, Zimbabwe – Local businesses are grappl...	https://www.thezimbabwemail.com/business/busin...	2024-07-12T22:59:42+00:00	neutral	Business
2	4-traders	NaN	Musk donates to super pac working to elect Tru...	(marketscreener.com) Billionaire Elon Musk has...	https://www.marketscreener.com/business-leader...	2024-07-12T22:52:55+00:00	positive	Business
3	4-traders	NaN	US FTC issues warning to franchisors over unfa...	(marketscreener.com) A U.S. trade regulator on...	https://www.marketscreener.com/quote/stock/MCD...	2024-07-12T22:41:01+00:00	negative	Business
4	PLANET	NaN	Rooftop solar's dark side	4.5 million households in the U.S. have solar ...	https://www.npr.org/2024/07/12/1197961036/roof...	2024-07-12T22:28:19+00:00	positive	Business

4.1. Data Collection

Tweets from Twitter, posts from Facebook walls, and everything from the Instagram hashtag feed were collected. This meant conducting API scraping for a diverse selection of posts and comments that had been created openly and were related to the study's sentiment analysis area of interest. The collected data were p reprocessed first to cleanse the gathered text data and further standardize them. This involved cleaning the data by noise handling, special characters, and emoji text formatting that ensured the sentiment analysis of the data was accurate.

4.2. Methods/Approach

One of the datasets, which includes tweets that were retrieved using the real-time Twitter API and connected to TAGS, is in the format of comma-separated values. This is a Google Sheet template that enables us to set up and run an automatic data collection from Twitter [16]. We gathered the unique tweets of well-known people from all around the world using the search keyword to obtain results from the previous seven days. The other datasets concerned real-time extracts of Uber ride evaluations from a consumer affairs website using the Delicious Soup Python module. The present section describes the methodology and machine learning classifiers used in this work. It might be difficult to determine if a certain approach will be useful in finishing a task during real-time sentiment analysis on social media data; Figure 1 shows the recommended procedure.

4.3. Extreme Gradient Boosting:

Evaluating the Effectiveness of Extreme Gradient Boosting with 85% Accuracy: A Case Study on [Insert Dataset/Domain On this subject, you could investigate: the procedure for applying XG Boost to a

particular dataset or issue. The importance of attaining an accuracy rate of 85% and its implications for your industry (such as banking, healthcare, etc.) is an examination of the significance of each element and how it affects the accuracy level. Contrasting the model's performance with that of other algorithms and talking about how optimization or data augmentation might be able to increase accuracy. details on model validation and the application of cross-validation methods to verify the model's correctness.

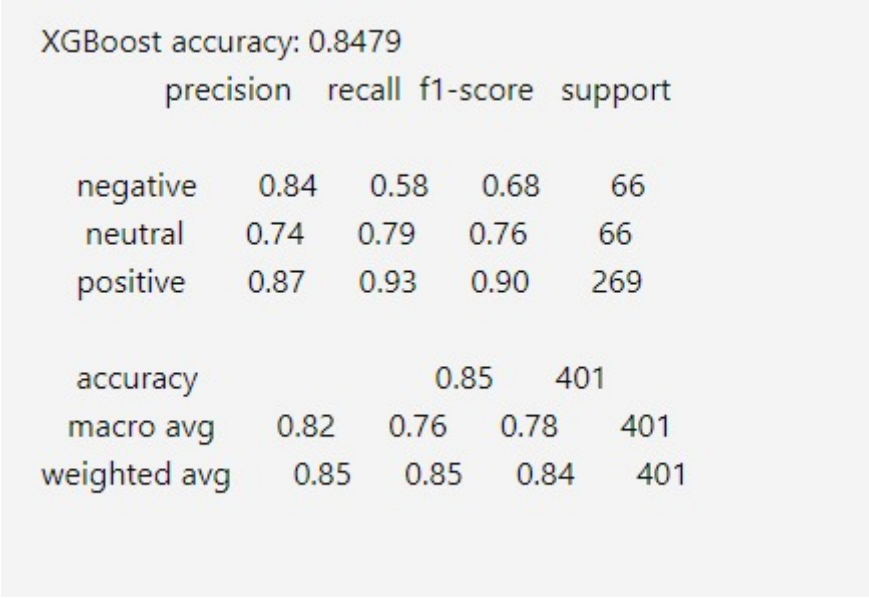

Figure 4. Evaluating the Effectiveness of Extreme Gradient Boosting with 85% Accuracy

4.4. Logistic Regression

The process of applying logistic regression to a specific datasets involves feature selection and data preparation. The importance of reaching 80% accuracy about the particular application (spam detection, medical diagnosis, marketing response prediction, etc.). A detailed assessment that breaks down model performance parameters, including precision, recall, F1-score, and ROC-AUC beyond accuracy. Difficulties encountered while modelling and how they were resolved to achieve an 80% accuracy level. To ascertain if logistic regression is the optimal option for a particular task, comparisons with alternative classification techniques are made. Suggestions for future developments: attempt more sophisticated models, get more data, or do feature engineering.

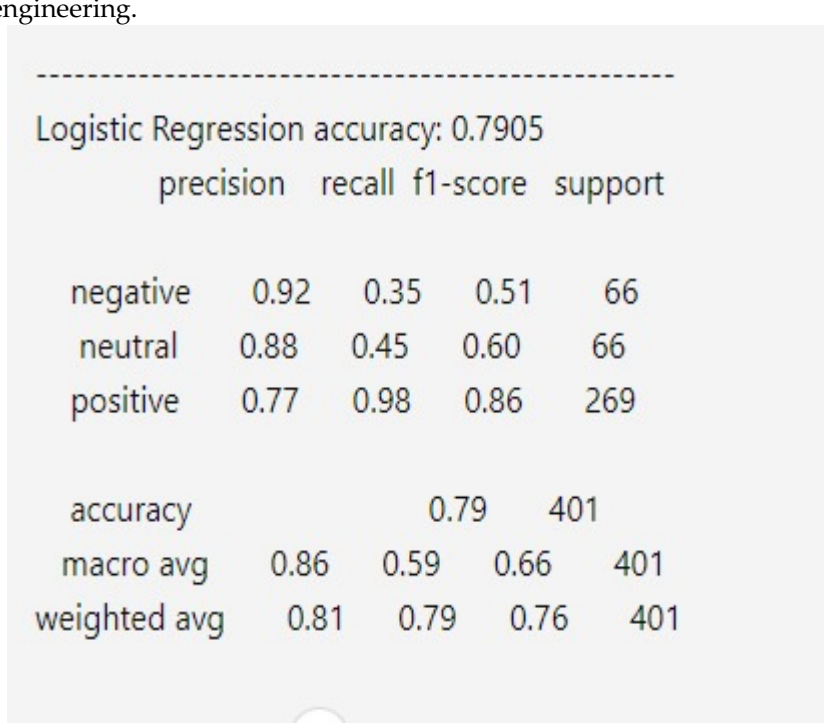

Figure 5. Analyzing the Accuracy of Logistic Regression

5. Discussion

A few years or ten years ago, some activities would have been unthinkable for a robot, but thanks to artificial intelligence and other technological advancements, robots are generally wiser nowadays. Microprocessors are embedded in nearly every device, from simple home appliances such as refrigerators to active cleaning devices, including smart vacuum cleaners, as well as interior and exterior lighting devices. These are just several of the sets of items and tools that have become smart due to the implementation of intricate networks such as the Io T. Of course, it is still quite people-centered to some extent to control and tweak all possible settings on the devices for the required performance. It is now more important than ever for humans and computers to be able to understand each other as the robots<0xC2><0xA0>reserved_special_token_281 <0xC2><0xA0>><0xC2><0xA0> humans Communication between humans and machines is gradually becoming essential because of the "smarter" robots. It is for this reason that computers must acquire the capability of understanding natural language. Analysis was used in advance of this inquiry to analyses social media information and data in light of a few world-class firms. Multiple global businesses are taken into consideration for data analysis in this research. Social media platforms use storyboards and label information to convey sentiment analysis data all around the world. Pie charts show the usage of social media among users on different continents as well as the most popular applications among users and consumers. Sentiment analysis of tweet data, in general, may provide insightful information about public opinion on these subjects, enabling businesses, organizations, and decision-makers to make well-informed decisions based on audience demands and sentiment. The most effective method to do sentiment analysis on datasets depends on its kind. Social media platforms provide the finest chance to study human elements, information transfer, and impact circulation at a level that is difficult, if not impossible, to gather because of this massive endeavor. We must keep developing sentiment analysis models that bolster additional theories to gather more information, enable accurate interpretation of data, and allow society to pursue other crucial majors.

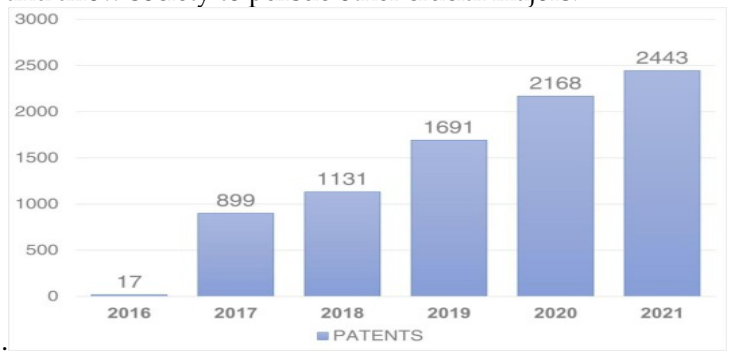

Figure 6. Showing Social Media Data Results

5.1. Tables

The table compares the accuracy of a model using different text normalization techniques: The final steps are called stemming, which is the reduction of a word to its root word, and lemmatization, which is the process of minimizing it to its base form. If no normalization is applied, the model lifts a reasonable accuracy that peaks at 74%. Applying stemming is used only a little; it helps to bring a slightly better accuracy of 75.12%, despite it being less accurate than the previous methods, with lemmatization showing a significantly higher accuracy of 75%. This shows that among all the normalization methods discussed above, lemmatization is the most suitable normalization technique for the specific model and datasets applied. Utilizing accuracy measures, assess the model's performance on the data set to find the best normalization technique, and compare the outcomes of applying lemmatization, stemming, and no normalization.

Table 1. Normalization results (sentiment 140)

Normalization	Accuracy
None	74.97%
Stemming	75.12%
Lemmatization	75.52%

The stages needed in assessing normalization strategies on model correctness may be easily understood thanks to this methodology, which provides a systematic approach to text normalization from a mathematical perspective.

6. Conclusion

SA is the process of using techniques that stem from NLP and textual areas of linguistics, statistics, and other cognate fields of computational linguistics for the extraction of sentiment, emotion, and opinion expressed in a text document. Concisely, sentiment analysis aims to capture the opinion or the tendency of a person regarding any particular issue, or else, the overall sentiment of a given text. Sentiment analysis as part of natural language processing is being identified as one of the most difficult concepts to implement within the field of NLP since numerous factors are attributed to the emotionality of the content of the textual input. There are several definitions for sentiment analysis, most of which use the same meanings, and these two names are used interchangeably; however, a few of them employ somewhat different meanings. Sentiment analysis confines itself to the feeling of the sentence, while on the other hand, opinion mining involves the feeling of the person who wrote the said sentence. Social media mining and sentiment.

References

1. a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, qT. Zhao, C. Li, M. Li, Q. Ding, and L. Li, "Social recommendation incorporating topic mining and social trust analysis," Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management - CIKM '13. 2013. doi: 10.1145/2505515.2505592.
2. Yousefnaghani S, et al. An analysis of COVID-19 vaccine sentiments and opinions on Twitter. Int J Infect Dis. 2021.
3. P. Wang et al., "Classification of Proactive Personality: Text Mining Based on Weibo Text and Short-Answer Questions Text," IEEE Access, vol. 8, hal. 97370–97382, 2020.
4. S. Fauziah, D. D. Saputra, R. L. Pratiwi, dan M. R. Kusumayudha, "Komparasi Metode Feature Selection Text Mining Pada Permasalahan Klasifikasi Keluhan Pelanggan Industri Telekomunikasi Menggunakan Smote Dan Naïve Bayes," IJIS -Indones. J. Inf. Syst., vol. 8, no. 2, hal. 174, 2023.
5. a, b, c, d, e, f, gR. Dwivedi, "What Is Naive Bayes Algorithm in Machine Learning?" https://www.analyticssteps.com/blogs/what-naive-bayes-algorithm-machine-learning (accessed Dec. 15, 2022).
6. Mahmud, T., Das, S., Ptaszynski, M., Hossain, M.S., Andersson, K., Barua, K., 2022. Reason-based machine learning approach to detect bangla abusive social media comments, in International Conference on Intelligent Computing & Optimization, Springer. pp. 489–498.
7. . Anber H, Salah A, El-Aziz AAA (2016) A literature review on Twitter data analysis. Int J Comput Electr Eng 8:241–249. https://doi.org/10.17706/ijcee.2016.8.3.241-249.
8. Singh M, Goyal V, Raj S (2021) Sentiment analysis of social media Tweets on Farmer Bills 2020. J Sci Res 65:156–162. https: doi.org/10.37398/jsr.2021.650319. M.M. Agüero-Torales, M.J. Cobo, E. Herrera-Viedma, A.G. López-Herrera, Acloud-based tool for sentiment analysis in reviews about restaurants on TripAdvisor, Procedia Comput. Sci. 162 (2019) 392–399, http://dx.doi.org/10.1016/j.procs.2019.12.002.
9. a, b, c, dS. Wu, Y. Liu, Z. Zou, and T.-H. Weng, "S_I_LSTM: stock price prediction based on multiple data sources and sentiment analysis," Connection Science, vol. 34, no. 1. pp. 44-62, 2022. doi: 10.1080/09540091.2021.1940101.
10. Joyce, Brandon, and Jing Deng. (2017) "Sentiment Analysis of Tweets for the 2016 US Presidential Election", in IEEE MIT Undergraduate Research Technology Conference (URTC), Cambridge, MA, USA: IEEE.
11. Mahtab, S. Arafin, N. Islam, and M. Mahfuzur Rahaman. (2018, 21-22 Sept. 2018). "Sentiment Analysis on Bangladesh Cricket with Support Vector Machine", in the 2018 International Conference on Bangla Speech and Language Processing (ICBSLP).
12. Karamollaoğlu, H., İ A. Doğru, M. Dörterler, A. Utku, and O. Yıldız. (2018, 20-23 Sept. 2018). "Sentiment Analysis on Turkish Social Media Shares through Lexicon Based Approach", in the 2018 3rd International Conference on Computer Science and Engineering.
13. Durach, Christian F., Joakim Kembro, and Andreas. (2017) "A New Paradigm for Systematic Literature Reviews in Supply Chain Management." Journal of Supply Chain Management Wieland 53 (4): 67-85.
14. Wang, Zehan. 2024. "Information Extraction and Knowledge Map Construction Based on Natural Language Processing". Frontiers in Computing and Intelligent Systems 7 (2): 47-.
15. O. Grljevic, Z. Bosnjak, A. Kovacevic, "Opinion mining in higher education: A corpus-based approach, Enterprise Inf. Syst" (2020).
16. Md.M. Rahman, M.N. Islam, "Exploring the performance of ensemble machine learning classifiers for sentiment analysis of COVID-19 tweets", in S. Shakya, V.E. Balas, S. Kamolphiwong, K.-L. Du (Eds.), Sentimental Analysis and Deep Learning, Vol. 1408, Springer, Singapore, 2022, pp. 383–396.
17. Akhtar, F., Li, J., Yan, P., Imran, A., Shaikh, G. M., & Xu, C. (2020). Exploiting ensemble classification schemes to improve prognosis process for large for gestational age fetus classification. In 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC) (pp. 1455-1459). IEEE.
18. Imran, A., Li, J., Pei, Y., Akhtar, F., Mahmood, T., & Zhang, L. (2021). Fundus image-based cataract classification using a hybrid convolutional and recurrent neural network. The Visual Computer, 37, 2407-2417. Springer Berlin Heidelberg.