iJRASET

International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** XI **Month of publication:** November 2023

DOI: https://doi.org/10.22214/ijraset.2023.56925

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XI Nov 2023- Available at www.ijraset.com

Sentiment Analysis in the Perspective of Natural Language Processing

Sutapa Chakraborty[1], Sohini Sadhukhan[2], Anirban Bhar[3], Moumita Ghosh[4]
[1, 2]B.Tech student, Department of Information Technology, Narula Institute of Technology, Kolkata, India
[3, 4]Assistant Professor, Department of Information Technology, Narula Institute of Technology, Kolkata, India

***Abstract:** Sentiment analysis, which is also called opinion mining, is a way to use natural language processing to figure out how someone feels about something written down. It includes reading the text and putting it into one of three groups: positive, negative, or neutral. In this paper we are going to give an overview about Natural Language Processing (NLP) and its subset Sentimental Analysis and how Natural Language Processing (NLP) is used to function the sentiment analysis. Natural language processing (NLP) is a subfield of artificial intelligence or AI—concerned with giving computers the ability to understand text and spoken words in much the same way human beings can. Right now, this is a very popular and new way of doing things because it works so well. In this paper we will explore the different ways NLP can be used for Sentiment Analysis, the challenges to look for and how it can revolutionize marketing strategies of MNCs and improve customer experiences. Sentiment analysis also known as opinion mining, is the process of analyzing digital text to determine if the emotional tone of the message is positive, negative, or neutral. It's common for businesses to use this method to find out and sort people's thoughts on a product, service, or idea. It is an important business intelligence tool that helps companies improve their products and services.*
***Keywords:** Natural Language Processing, Machine Learning, Amazon, Python, Tokenization.*

I. INTRODUCTION

Sentiment analysis is a popular task in Natural Language Processing (NLP). The goal of using sentiment analysis is to classify the text based on the mood or mentality expressed in the text, which can be positive, negative, or neutral.

Sentiment analysis is the process used to classify whether a block of text is positive, negative, or, neutral. The goal which Sentiment analysis tries to reach is to analyse people's opinions in a way that can help businesses expand. It focuses on factors like not only polarity (positive, negative & neutral) but also on emotions (happy, sad, angry, etc.). It uses various Natural Language Processing algorithms such as Rule-based, Automatic, and Hybrid approaches.

For Example: If we want to analyse if a product is satisfying customer requirements and standards, we can use sentiment analysis to collect statistical data to draw beneficial conclusions. Many MNCs like Amazon, YouTube, Twitter uses personalized surveys to calculate customer interests on visual data to increase the system's algorithm to benefit not only the users as well as its marketing strategy. Net Promoter Score (NPS) surveys are used to gain knowledge on how a customer views a certain product or service. Sentiment Analysis has also gained immense popularity on the premise of tackling real life situations and processing large volumes of NPS responses in a short amount of time. In this paper we aim to represent both domains working as a singular and the methodology following which they are used in the market. Sentiment Analysis also helps in mitigating the challenges faced when it comes to Natural Language Processing and gives us a better understanding of the sentimental array of human emotions.

Natural Language Processing (NLP) is a field of artificial intelligence focusing on the interaction between computers and human language. Sentiment analysis, also known as opinion mining, is a specific application of NLP that involves determining the sentiment or emotion expressed in a piece of text. Here are the key steps and techniques involved in using NLP for sentiment analysis: Data Collection: The first step in sentiment analysis is to gather data. This data can be in the form of text from various sources, such as social media, customer reviews, news articles, or any text data that contains sentiment information.

A. Text Preprocessing

1) *Tokenization:* Split the text into individual words or tokens.
2) *Lowercasing:* Convert all text to lowercase to ensure consistent analysis.
3) *Stopword Removal:* Eliminate common words like "the," "and," "is" that don't carry sentiment.
4) *Stemming or Lemmatization:* Reduce words to their base or root form.

B. Feature Extraction

1) *Bag of Words (BoW):* Create a matrix of word frequencies in the text.
2) *Term Frequency-Inverse Document Frequency (TF-IDF):* Assign weights to words based on their importance in the document relative to the entire corpus.
3) *Word Embeddings (Word2Vec, GloVe):* Represent words as dense vectors, capturing semantic meaning.
4) *Sentiment Labelling:* Annotate the data with sentiment labels, often as positive, negative, or neutral. This can be done manually or with the help of pre-labelled datasets.

C. Model Selection

1) *Machine Learning:* Train a supervised machine learning model like logistic regression, Naive Bayes, or support vector machines using the labelled data.
2) *Deep Learning:* Utilize neural network architectures like Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) for sequence data.
3) *Training the Model:* Use the labelled data to train the selected model. The model learns to predict sentiment based on the features extracted from the text.
4) *Evaluation:* Assess the model's performance using metrics like accuracy, precision, recall, F1-score, and confusion matrix to understand how well it classifies sentiment.
5) *Deployment:* Once the model performs well, deploy it in a real-world application where it can analyze sentiment in new, incoming text data.
6) *Fine-Tuning:* Continuously update and fine-tune the model based on new data and feedback to improve accuracy and adapt to changing language trends.
7) *Post-processing and Visualization:* After sentiment analysis, you can visualize the results, generate reports, or perform additional post-processing to gain insights.
8) *Scaling:* For large-scale sentiment analysis, consider parallel processing and distributed computing to handle big data efficiently.
9) *Handling Negation and Context:* Advanced sentiment analysis models should be capable of handling negations and understanding the context in which sentiments are expressed to avoid misinterpretations.
10) *Domain-Specific Sentiment Analysis:* Fine-tune models for specific industries or domains, as sentiment expressions can vary widely across different contexts.
11) *Ethical Considerations:* Be mindful of potential biases in the data and model. Regularly audit and review your sentiment analysis system to ensure fairness and avoid reinforcing stereotypes.

Sentiment analysis is a valuable tool for understanding public opinion, customer feedback, and social media trends. It has applications in various fields, including customer service, market research, and social listening. The choice of techniques and tools will depend on the specific requirements and data at hand.

II. LITERATURE REVIEW

There have been studies on automated extraction of sentiment from the text. For example, Pang and Lee [1] have used movie review domains to experiment machine learning techniques (Naïve Bayes, maximum entropy classification, and Support Vector Machine (SVM)) in classifying sentiment. They achieved up to 82.9% of accuracy using SVM and unigram model. However, as the performance of sentiment classification is based on the context of documents, the machine learning approaches have difficulties in determining the sentiment of text if sentiment lexicons with contrast sentiment are found in the text.

On the research by Blenn et al [2], a system that worked through a combination of grammatical analysis with traditional word frequency analysis was proposed. Grammatical analysis studied the structure of text, and associated the sentiment lexicons with subject by identifying the relationship between sentiment lexicons and subject. It was a significant improvement in sentiment analysis for short colloquial text as previous approach did not achieve high detection accuracy as it did. It did not need any supervised training, but managed to improve the accuracy of previous work by 40%. In this research, a system is proposed to carry out sentiment analysis on tweets based on specific topic. Several pre-processes steps have been carried out to clean the noise in tweets and present tweets in formal language. To determine the sentiment pf tweets, NLP is implemented to find out the subjective portion of tweets that associates the subject, and classify the sentiment of tweets. The tweets will be labelled as positive, negative or neutral.

The vocabulary of natural languages and the polarity in complex sentences causing several difficulties is highlighted in the paper (Neha Gaur,2017). A new aspect called Product Aspect Ranking is also featured in their paper. Sentiment Tokens and Sentiment Scores are information extracted from original dataset (Neha Gaur,2017). By presenting the data in an easy-to-understand format, such as graphs and charts, people can get a better grasp of the sentiment of a particular product or topic (IJRASET,2014).

There are three steps in sentiment classification: subjectivity classification, semantic association and polarity classification (Wei Yen Chong, 2014).

Sentiment Analysis has gained popularity all over the world for business strategists, stocks, basing algorithms. Various papers have been published to evaluate and highlight its core running issues and solutions to resolve them. The key findings as the benefits over conventional methods are: user friendliness, weekly removal of average, multiple platform levels, light-weight and contains bar graphs which is easy to understand.

III.SENTIMENT ANALYSIS: APPLICATION AND IMPLEMENTATION PHASES

Companies use sentiment analysis to monitor and improve customer support interactions. Chatbots equipped with sentiment analysis can detect frustration or dissatisfaction in customer messages and respond accordingly.

Brands and organizations track social media sentiment to gauge public opinion about their products, services, and campaigns. This information informs marketing and public relations strategies. NLP is used to analyse customer feedback from surveys, reviews, and other sources. It helps businesses understand what customers like or dislike and identify areas for improvement. Sentiment analysis is applied to news articles, social media posts, and financial reports to assess market sentiment. It helps investors and traders make more informed decisions. Sentiment analysis is used to gauge public sentiment about political candidates, policies, and issues. It can help political campaigns and policymakers tailor their messaging. In the healthcare industry, sentiment analysis is used to analyse patient feedback and reviews about hospitals and healthcare providers to improve patient experiences. Companies use sentiment analysis to assess employee feedback and engagement surveys to identify areas where they can enhance the workplace environment. Sentiment analysis helps businesses identify areas where they can improve their products or services based on customer feedback. Companies use NLP to monitor and manage their online brand reputation. They can detect negative sentiment early and take corrective actions. Sentiment analysis enhances recommendation systems by tailoring recommendations to users' preferences and emotional states. NLP is used to filter and moderate user-generated content on online platforms to remove inappropriate or harmful material and maintain a safe environment. Sentiment analysis is applied to legal documents and contracts to identify potential risks and issues. In manufacturing, NLP can be used to analyse customer feedback and warranty claims to identify product defects and quality issues. Sentiment analysis can help in interpreting opinion poll results and surveys, providing a deeper understanding of public sentiment. Digital assistants and smart devices use NLP to detect and respond to users' emotional states for more empathetic interactions. NLP-powered sentiment analysis enables companies to create marketing campaigns that resonate with their target audience's emotions and preferences. During the COVID-19 pandemic, sentiment analysis was used to track public sentiment about the virus, vaccines, and government responses. Content creators and studios use sentiment analysis to gauge audience reactions to movies, TV shows, and other forms of entertainment.

The applications of NLP in sentiment analysis continue to expand as organizations recognize the value of understanding and responding to public sentiment and user emotions. NLP technologies, including advanced neural networks and pre-trained language models, are being increasingly employed to achieve more accurate and nuanced sentiment analysis.

The process of Natural Language understanding comprises of five analytical phases. These Phases are:

A. Morphological Analysis

Morphological analysis is a linguistic and natural language processing (NLP) technique that involves breaking down words into their smallest meaningful units, known as morphemes. Morphemes are the smallest grammatical units in a language and can include prefixes, suffixes, roots, and other affixes. Morphological analysis is essential for understanding the structure and meaning of words in various languages. In NLP, morphological analysis is used to understand the structure of words in a text. It helps in various tasks like stemming (reducing words to their root form), lemmatization (finding the dictionary or base form of a word), and part-of-speech tagging.

B. Syntactic Analysis

Syntactic analysis, also known as parsing, is a fundamental component of Natural Language Processing (NLP) that involves the analysis of the grammatical structure of sentences in a natural language.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XI Nov 2023- Available at www.ijraset.com

It aims to understand how words in a sentence are combined to form phrases, clauses, and sentences, and how these elements relate to each other. Syntactic analysis is crucial for interpreting the meaning of sentences and for various NLP tasks such as machine translation, information extraction, and question answering. The first step in syntactic analysis is tokenization, where the text is divided into individual words or tokens. These tokens are the basic units for syntactic analysis. Constituency parsing involves breaking down a sentence into a hierarchical structure of constituents, such as noun phrases (NP), verb phrases (VP), and prepositional phrases (PP). It represents how words group together to form phrases. Common algorithms for constituency parsing include the Earley parser, CYK parser, and chart parsing. CFGs are used to formalize the syntactic structure of natural language. They consist of a set of rules that describe how words can be combined to form phrases and sentences. Syntactic analysis can be enhanced with machine learning and deep learning techniques, including neural network-based parsers. These approaches often yield state-of-the-art results in parsing tasks.

C. Semantic Analysis

Semantic analysis in Natural Language Processing (NLP) is the process of understanding the meaning of words, phrases, sentences, or documents in a way that can be interpreted by machines. Unlike syntactic analysis, which deals with the grammatical structure of language, semantic analysis focuses on the actual meaning and interpretation of language. It aims to capture the meaning of text and the relationships between words and concepts. SRL aims to identify the roles that words or phrases play in the predicate-argument structure of a sentence. It helps determine who is doing the action, what the action is, and who or what it is done to. For example, in the sentence "John (agent) ate (predicate) an apple (patient)," SRL would label the semantic roles of "agent" and "patient." Semantic parsing is the process of converting natural language into a formal representation of its meaning, often in the form of a logical or structured query language. It is used in question answering systems and dialogue systems. In sentiment analysis, semantic analysis helps identify the sentiment expressed in text, taking into account the meaning and context of words and phrases. Semantic analysis aids in summarizing text by identifying the most important content and extracting key information.

D. Disclosure Integration

In the context of Natural Language Processing (NLP), disclosure integration can refer to the incorporation of disclosure statements or disclaimers within text, particularly in applications where transparency and compliance are essential. This is relevant in various NLP tasks and applications. For example,
Ram won the race.
Mohan ate half of a pizza.
He liked it.
In the above example, “He” can be Ram or Mohan. This creates an ambiguity. The word “He” shows dependency on both sentences. This is known as disclosure integration. It means when an individual sentence relies upon the sentence that comes before it. Like in the above example the third sentence relies upon the sentence before it. Hence the goal of this model is to remove referential ambiguity.

E. Pragmatic Analysis

Pragmatic analysis in Natural Language Processing (NLP) is the study and interpretation of language that goes beyond the literal meanings of words and phrases to consider the intended meaning, context, and the social or communicative aspects of language use. It focuses on how language is used to convey meaning and the various implied, indirect, or context-dependent meanings that may not be immediately evident in the text. Pragmatic analysis is crucial for understanding the real-world context of language use and for building NLP systems that can handle more nuanced and context-aware communication.
The pragmatic analysis means handling the situation in a much more practical or realistic manner than using a theoretical approach. As we know that a sentence can have different meanings in various situations.
For example, the average is 18.

The average is 18. (average may be of sequence)
The average is 18. (average may be of a vehicle)
The average is 18. (average may be of a mathematical term)

We can see that for the same input there can be different perceptions. To interpret the meaning of the sentence we need to understand the situation. To tackle such problems, we use pragmatic analysis. The pragmatic analysis tends to make the understanding of the language much clearer and easier to interpret.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XI Nov 2023- Available at www.ijraset.com

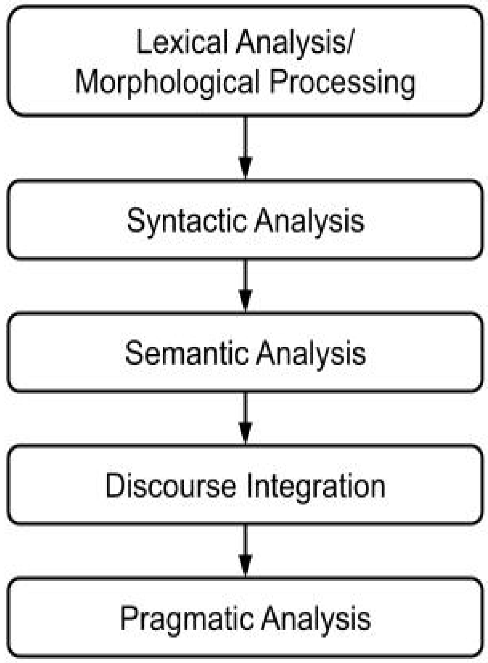

Fig. 1: The Different Implementation Phases of NLP

All these phases have their own desired boundaries, but these boundaries are not completely simple to comprehend. They occasionally follow a proper sequence, or sometimes all at once. When one process enrols in a sequence, this process may request for assistance to another one.

IV. AMAZON ACQUIRING LEVERAGE USING SENTIMENT ANALYSIS

Multinational corporations (MNCs) like Amazon and YouTube use sentiment analysis in various ways to gain insights, improve user experiences, and make data-driven decisions. Here are some examples of how these companies might leverage sentiment analysis:

1) *Customer Feedback and Reviews:* Amazon analyses customer reviews and product ratings to understand customer sentiment about products and services. Sentiment analysis helps identify popular products and issues that need attention.
2) *Customer Support:* Sentiment analysis is employed in Amazon's customer support interactions, such as chatbots and email responses, to understand customer sentiment and provide more personalized and effective support.
3) *Product and Content Recommendations:* Sentiment analysis plays a role in Amazon's recommendation systems. It helps suggest products to users based on their past interactions, reviews, and preferences.
4) *Market Research:* Amazon uses sentiment analysis to monitor and analyse social media and news sentiment related to its brand and products. This can inform marketing and product development strategies.
5) *Content and Advertising Effectiveness:* When running advertising campaigns, Amazon can use sentiment analysis to gauge the effectiveness of ad messaging and its impact on customer sentiment.
6) *Product and Service Improvement:* Sentiment analysis of customer feedback can identify common issues, enabling Amazon to make improvements in its products and services.
7) *Competitor Analysis:* Amazon may employ sentiment analysis to monitor how customers and users feel about competitors' products and services, helping them stay competitive in the market.
8) *Fraud Detection:* Amazon may use sentiment analysis to detect fraudulent activities or deceptive product reviews by identifying patterns of unusual sentiment in reviews and feedback.
9) *Ethical Considerations:* Amazon also uses sentiment analysis to detect and mitigate hate speech, discrimination, and other harmful content to maintain a safer and more inclusive online environment.

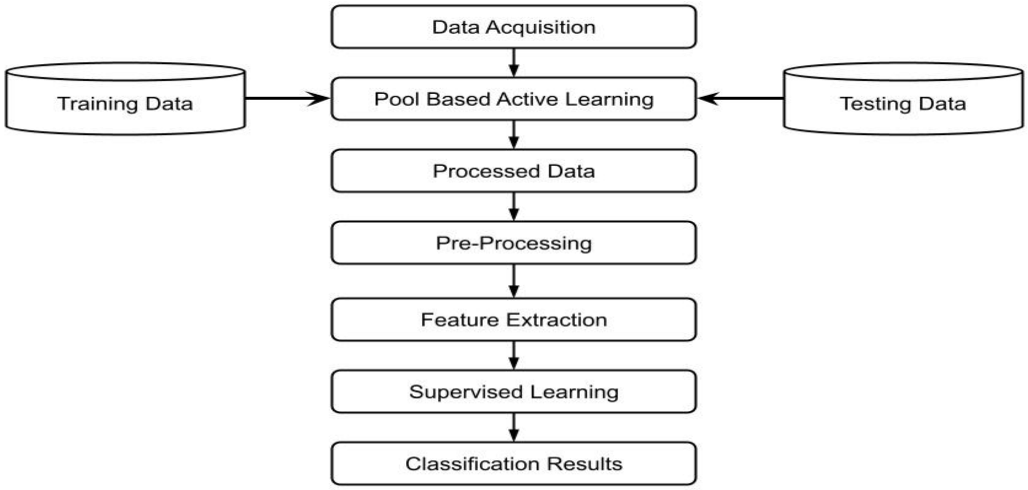

Fig. 2: Sentiment Analysing reviews by Amazon

It's important to note that the specific applications of sentiment analysis in these companies may vary and evolve over time. The goal is to leverage sentiment analysis to better understand customer sentiment, enhance user experiences, and make informed business decisions.

V. FUTURE CONSIDERATIONS BASED ON CHALLENGES

As the field of Natural Language Processing (NLP) and sentiment analysis continues to evolve, several future considerations and trends are emerging. These trends reflect the ongoing advancements and challenges in the application of NLP for sentiment analysis:

1) *Multimodal Sentiment Analysis:* Combining text with other modalities like images, audio, and video for a more comprehensive understanding of sentiment. This is particularly relevant for platforms like social media and video-sharing websites.
2) *Emotion Recognition:* Moving beyond basic positive/negative sentiment to recognize specific emotions such as joy, anger, sadness, and fear. This has applications in understanding user experiences and content creation.
3) *Contextual Analysis:* Developing models that can understand and interpret the context in which sentiments are expressed, as the same words may have different meanings in various contexts.
4) *Cross-Lingual Sentiment Analysis:* Improving the capability to analyse sentiment in multiple languages to cater to a global audience.
5) Real-time Sentiment Analysis: Enhanced capabilities for real-time sentiment analysis to respond promptly to changing trends and user feedback.
6) *Fine-Grained Sentiment Analysis:* Going beyond binary sentiment (positive/negative) to provide more fine-grained sentiment labels, such as strongly positive, weakly positive, strongly negative, etc.
7) *Sentiment Analysis for Conversational AI:* Developing conversational agents that can recognize and respond to user sentiment, making interactions more empathetic and effective.
8) *Ethical and Bias Considerations:* Continuing efforts to address bias in sentiment analysis models and ensure fairness, transparency, and accountability in NLP applications.
9) *Privacy Concerns:* Developing techniques to perform sentiment analysis without violating user privacy, especially in light of stricter data protection regulations.
10) *Healthcare Applications:* Increasing applications in healthcare for analyzing patient sentiments and feedback, mental health monitoring, and sentiment-based interventions.
11) *Sentiment Analysis for Autonomous Vehicles:* Understanding the emotional state of passengers or other road users to enhance safety and user experiences in autonomous vehicles.
12) *Customized and Domain-Specific Models:* The development of specialized sentiment analysis models tailored to specific industries and use cases, such as finance, healthcare, or legal services.

13) *Energy and Sustainability:* Using sentiment analysis to gauge public sentiment about environmental issues and sustainable practices, which can inform corporate social responsibility strategies.
14) *Interactive Sentiment Analysis:* Developing systems that allow users to interact with and customize sentiment analysis models based on their specific needs and preferences.
15) *Continuous Model Training:* Employing continuous learning and adaptation techniques to ensure that sentiment analysis models remain up-to-date and effective in understanding changing language patterns and trends.
16) *Human-AI Collaboration:* Creating systems where AI and human experts work together to achieve more accurate and context-aware sentiment analysis.
17) *Sentiment Analysis for Social Good:* Applying sentiment analysis for societal benefits, such as detecting signs of depression or cyberbullying, and offering support where needed.
18) *Quantifying the Uncertainty:* Developing methods to quantify the uncertainty in sentiment predictions to better understand model confidence and potential errors.

The future of sentiment analysis in NLP is likely to be characterized by increased sophistication, broader applications, and a growing emphasis on ethical and responsible AI. As NLP models become more powerful and nuanced, they will play a crucial role in understanding and responding to human emotions and sentiments across various domains and platforms.

VI. CONCLUSION AND FUTURE DIRECTION

To begin with, mental health professionals can obtain the most reliable data from social media. The data is generated by genuine users, and the practitioners are able to collect authentic information that may offer valuable insights into the mental state of patients afflicted with mental illness due to the anonymity afforded by the Internet.

Patients are permitted to leave comments in the social media platform using the language of their preference. It is advisable to partition and compare such data in order to identify any shared characteristics among patients, regardless of their cultural heritage. For instance, an intriguing inquiry could be whether the fundamental causes of depression, as reported by depressed users in an Arabic-speaking country and the United States, are identical.

Moreover, although English continues to be the preferred language among the majority of Internet users, nuanced variations can be observed in the manner in which individuals compose their comments. Therefore, methods of segregating English-speaking patients according to their cultural and geographical origins and deriving the distinctions between them would be an intriguing area of research. For instance, do users originating from India exhibit identical symptoms to those originating from the United Kingdom? Lastly, the capability of deducing the Socioeconomic Status (SES) of social media users who post will be an additional aspect to consider. Healthcare providers utilize socioeconomic status (SES) as a significant data point in forecasting the initiation of diverse ailments. An element of SES is the educational attainment of the user making the post.

The future of sentiment analysis using NLP seems promising, with possible applications in tailored marketing, mental health monitoring, real-time political analysis, and enhanced customer assistance. These prospective applications are driven by advancements in deep learning and context-aware models, which are expected to broaden the scope of sentiment analysis in the future. Through the use of sentiment analysis, we are able to achieve new heights in terms of user-friendly shopping environments and consumer comfort, thereby making life simpler for everyone. By resolving the challenges that are present in NLP, it is possible to minimize machine buffering, which will result in improved output outcomes and will revolutionize the AI platform.

REFERENCES

[1] Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. arXiv preprint cs/0205070.
[2] Blenn, N., Charalampidou, K., & Doerr, C. (2012). Context-sensitive sentiment classification of short colloquial text. In NETWORKING 2012: 11th International IFIP TC 6 Networking Conference, Prague, Czech Republic, May 21-25, 2012, Proceedings, Part I 11 (pp. 97-108). Springer Berlin Heidelberg.
[3] Wei Yen Chong, Bhawani Selvaretnam, Lay-Ki Soon, 2014 4th International Conference on Artificial Intelligence with Applications in Engineering and Technology, “Natural Language Processing for Sentiment Analysis”.
[4] Dr.K.Sindhura, Journal of Data Acquisition and Processing(2023), “Sentiment Analysis using Natural Language Processing and Machine Learning”.
[5] Drashti Panchal, Mihika Mehta, Aryaman Mishra, Saish Ghole, Mrs. Smita Dandge, Ijraset Journal For Research in Applied Science and Engineering Technology(2013), “Sentiment Analysis Using Natural Language Processing”.
[6] Md. Taufiqul Haque Khan Tusar, Md. Touhidul Islam, 2021 International Conference on Electronics, Communications and Information Technology (ICECIT), “A Comparative Study of Sentiment Analysis Using NLP and Different Machine Learning Techniques on US Airline Twitter Data”.
[7] S.Prathap, SK Moinuddin Ahmad, 2018 IJSRSET | Volume 4 | Issue 8, “Sentiment Analysis On Movie Reviews”.
[8] Xing Fang, Justin Zhan, Journal of Big Data (2015), “Sentiment Analysis using Product Review Data”.