LEVERAGING MACHINE LEARNING TECHNIQUES FOR ANALYZING AND IDENTIFYING SENTIMENT IN SOCIAL MEDIA POSTS

Kaneez Fatma[1], Dipti Ranjan Tiwari[2]

[1]Master of Technology, Computer Science and Engineering, Lucknow Institute of Technology, Lucknow, India
[2]Assistant Professor, Department of Computer Science and Engineering, Lucknow Institute of Technology, Lucknow, India

---***---

Abstract - *Sentiment analysis is a field that examines people's opinions on various topics, from products to political and social events. It has gained widespread attention as it helps stakeholders make informed decisions based on public sentiment. Opinion mining is a key technique used to extract insights from platforms like search engines, blogs, Twitter, and social networks. However, manually analyzing large volumes of tweets, which are often in unstructured text form, can be challenging. To overcome this, researchers use computational techniques like the Bag-of-Words (BoW) model, which identifies sentiment-bearing words through machine learning. In this study, a lexicon-based method was employed to automatically detect sentiments in tweets collected from Twitter. Researchers applied three machine learning algorithms: Naive Bayes (NB), Maximum Entropy (ME), and Support Vector Machines (SVM), to assess their effectiveness in classifying tweets by sentiment. The results showed that both NB with Laplace smoothing and SVM were reliable classifiers, especially when using specific features like unigrams or Part-of-Speech (POS) tags. Overall, sentiment analysis is a valuable tool for understanding public opinions shared on platforms like Twitter, allowing stakeholders to gauge public reactions to various subjects.*

Key Words: Sentiment Analysis, Social Media, Natural Language Processing (NLP), Deep Learning, Transformers, Supervised Learning, Text Classification.

1.SENTIMENT ANALYSIS IN SOCIAL MEDIA AND ONLINE POSTS

Sentiment analysis, also referred to as opinion mining, is a pivotal domain within natural language processing (NLP) that entails discerning the emotional undertones within a body of text. In the realm of social media and online discourse, it holds substantial significance in comprehending the sentiments articulated by users across various platforms such as Twitter, Facebook, and online forums. Given the proliferation of user-generated content, organizations, scholars, and governmental bodies are increasingly dependent on sentiment analysis to assess public perceptions pertaining to products, services, political occurrences, and societal matters. Nonetheless, owing to the informal and colloquial lexicon utilized in these posts—frequently incorporating slang, emoticons, and ambivalent sentiments—the task presents distinctive complexities. Machine learning methodologies have been extensively embraced to refine the precision and scalability of sentiment classification, rendering it an indispensable instrument for scrutinizing extensive quantities of online material.

Figure-1: Social Media Sentiment

2.INTRODUCTION

Sentiment identification plays a crucial role in various aspects such as understanding public opinions, customer feedback, and emerging trends. By delving into how people perceive products, services, events, or societal issues, businesses can gain valuable insights that drive actionable decisions. For instance, when a company analyzes customer feedback, it can pinpoint specific areas for improvement, leading to enhanced product quality and increased customer satisfaction. This, in turn, fosters brand loyalty and strengthens the company's market position.

In the realm of public opinion, sentiment analysis serves as a powerful tool for governments, organizations, and media outlets. By monitoring reactions to policies, news events, or social movements, stakeholders can gauge public sentiment and make informed decisions. For example, during a political campaign, tracking sentiment trends can offer valuable insights into voter behavior and preferences, helping candidates tailor their strategies for maximum impact.

Moreover, sentiment analysis enables industries to anticipate consumer behavior, predict market shifts, and even forecast political outcomes. By leveraging sentiment data, businesses can stay ahead of the competition and make strategic decisions that align with evolving trends. This proactive approach not only enhances decision-making processes but also supports long-term planning efforts, giving organizations a competitive edge in today's dynamic landscape. Ultimately, sentiment identification remains a cornerstone for informed decision-making and strategic planning across various sectors.

3.AMBIGUOUS LANGUAGE, SLANG, AND MIXED SENTIMENTS

Sentiment analysis encounters various challenges due to the intricate nature of human language. One significant hurdle lies in deciphering sarcasm, irony, and humor, where words' literal meanings diverge from the intended emotions, thus posing a challenge for algorithms to accurately capture true sentiments. For instance, a statement like "Oh, great, another Monday!" may seem positive on the surface but actually conveys a sense of dread or dislike. This complexity makes it tough for sentiment analysis tools to discern the underlying sentiment accurately.

The prevalence of slang, abbreviations, and informal language, particularly on social media platforms, adds another layer of difficulty. These linguistic nuances often vary across different regions and communities, making it challenging for algorithms to interpret sentiments accurately. For example, expressions like "lit" or "fam" may have different connotations based on cultural context, further complicating sentiment analysis processes.

Another key challenge arises from mixed sentiments, where a single post contains both positive and negative emotions. This juxtaposition can confuse sentiment analysis models, as understanding the overall sentiment requires nuanced interpretation of the context. Additionally, the context dependency of sentiment analysis presents a significant obstacle. The sentiment associated with a word or phrase can shift based on the surrounding text, necessitating a comprehensive understanding of the entire message to accurately gauge emotions.

The rise of multilingual and code-switching posts further complicates sentiment analysis tasks. Users often blend multiple languages in their communications, making it challenging for algorithms to accurately identify sentiments across diverse linguistic landscapes. Overall, navigating these challenges requires sophisticated algorithms and a deep understanding of the intricacies of human language to ensure accurate sentiment analysis in an increasingly complex digital world.

4.SENTIMENT ANALYSIS TECHNIQUES

Sentiment analysis techniques have evolved significantly over time, transitioning from traditional methods to more advanced machine learning approaches. Traditional methods, such as lexicon-based techniques, rely on predefined lists of words categorized as positive, negative, or neutral. For instance, a lexicon-based approach might classify words like "happy" as positive and "sad" as negative based on their inherent meanings. However, these methods face challenges when it comes to capturing context and dealing with ambiguity. This can result in lower accuracy when analyzing sentiments in more complex texts.

In contrast, machine learning approaches have revolutionized sentiment analysis, especially with the advent of supervised learning techniques. For example, Support Vector Machines (SVM), Naive Bayes, and deep learning models like neural networks have shown remarkable success in automatically learning patterns from large labeled datasets. These algorithms can adapt to various contexts, including slang and nuanced expressions that may pose challenges for traditional methods. By leveraging word embeddings and transformers like BERT, machine learning models can grasp intricate relationships between words, enhancing their understanding of sentiment nuances.

Machine learning approaches offer more accurate and scalable solutions for sentiment analysis compared to traditional methods. While they do require substantial amounts of data and computational resources, their ability to capture complex patterns and adapt to diverse contexts makes them invaluable in today's data-driven world. As sentiment analysis continues to play a crucial role in understanding customer feedback, social media trends, and market sentiments, the advancements in machine learning are set to further revolutionize how we interpret and analyze sentiments in textual data.

5.CURRENT MACHINE LEARNING MODELS USED FOR SENTIMENT ANALYSIS

Current machine learning models used for sentiment analysis encompass a diverse array of techniques, ranging from traditional algorithms to cutting-edge deep learning methodologies. Among the traditional algorithms, Support Vector Machines (SVM) stand out for their proficiency in establishing decision boundaries that effectively segregate positive and negative sentiments within textual data. For instance, in a sentiment analysis task analyzing customer reviews of a product, SVM can accurately classify sentiments based on the specific language used to convey satisfaction or dissatisfaction.

In contrast, the Random Forest algorithm, categorized as an ensemble method, showcases its prowess by amalgamating multiple decision trees to enhance prediction accuracy and mitigate overfitting issues. Consider a scenario where market sentiment towards a particular stock needs to be analyzed using a vast dataset of financial news articles; Random Forest would excel in discerning nuanced sentiments and preventing the model from memorizing noise in the data.

Delving into the realm of deep learning, neural networks have revolutionized sentiment analysis by delving deep into the intricate patterns embedded within textual content.

Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have emerged as stalwarts in this domain. For example, in a sentiment analysis task analyzing social media posts, CNNs can effectively capture spatial hierarchies of words to discern sentiment nuances in varying contexts.

RNNs, particularly Long Short-Term Memory (LSTM) networks, shine in understanding the sequential nature of language, making them indispensable for tasks involving context or lengthy sentences. Imagine a scenario where sentiment analysis is conducted on movie reviews; LSTM networks would excel in grasping the sentiment evolution throughout a review, capturing the nuances in how opinions change over the course of the text. Recently, transformer models like BERT have taken the sentiment analysis landscape by storm, leveraging attention mechanisms to grasp the semantic relationships between words in the context of a whole sentence. By understanding complex language structures, sarcasm, and contextual dependencies, BERT and its counterparts outperform traditional methods. In a sentiment analysis task analyzing customer service interactions, BERT could accurately interpret the underlying sentiment even in cases of subtle nuances or implicit meanings within the conversation.

6.COMPARISON OF SUPERVISED VS. UNSUPERVISED APPROACHES IN SENTIMENT IDENTIFICATION

Supervised and unsupervised approaches in sentiment identification represent two distinct methodologies, each with its own advantages and limitations.

6.1.Supervised Approaches

Supervised sentiment identification is a crucial process that heavily relies on labeled datasets. These datasets contain text samples that are meticulously annotated with the corresponding sentiment, be it positive, negative, or neutral. To effectively train machine learning algorithms like Support Vector Machines (SVM), Random Forest, and neural networks, this labeled data is utilized. These algorithms learn intricate patterns and associations between various features (such as words or phrases) and their respective sentiments. For instance, in the case of sentiment analysis for product reviews, a labeled dataset may consist of reviews labeled as positive, negative, or neutral based on the overall sentiment expressed towards the product.

When these machine learning models are trained on such datasets, they acquire the ability to predict sentiment in new, unseen text. For example, consider a scenario where a sentiment identification model has been trained on a dataset of movie reviews. The model can then be used to analyze newly released movie reviews and predict whether the sentiment expressed is positive, negative, or neutral. This predictive capability is derived from the knowledge and patterns the model has learned during its training phase.

In essence, the process of supervised sentiment identification involves leveraging labeled datasets to train machine learning algorithms, enabling them to predict sentiment in unseen text based on the patterns and associations learned during training. This methodology has applications in various domains such as customer feedback analysis, social media sentiment tracking, and opinion mining. The accuracy and effectiveness of these sentiment identification models heavily depend on the quality and diversity of the labeled datasets used for training. As the field of natural language processing continues to advance, so too will the capabilities of supervised sentiment identification models in understanding and analyzing human emotions and opinions expressed through text.

6.1.1.Advantages

Supervised models have shown remarkable accuracy in scenarios where extensive and well-annotated datasets are accessible. This is primarily due to their ability to discern intricate patterns from the provided training data. For instance, models like BERT and LSTM excel in capturing contextual nuances and word dependencies, thereby enhancing sentiment analysis, especially in handling complex sentence structures. Through supervised learning, these models can be fine-tuned to cater to specific tasks or domains, such as distinguishing between product reviews and movie reviews.

Furthermore, supervised learning leverages cutting-edge algorithms like deep learning and transformers, which have revolutionized the field of sentiment analysis. These advanced algorithms play a pivotal role in achieving superior performance outcomes in sentiment prediction tasks. By incorporating deep learning techniques, supervised models can delve deeper into the intricacies of language patterns and semantic meanings, enabling them to deliver more accurate sentiment analysis results.

The utilization of supervised learning methodologies, coupled with state-of-the-art algorithms, has significantly elevated the effectiveness and precision of sentiment prediction tasks. Through continuous advancements in machine learning techniques, the realm of sentiment analysis continues to evolve, paving the way for more sophisticated and precise models in the future.

6.1.2.Limitations

The effectiveness of supervised approaches heavily depends on the quality and size of the labeled dataset. Labeled data is often scarce and expensive to create. Supervised models can overfit to the training data, meaning they may perform well on known examples but poorly on unseen data, especially if the training set lacks diversity. Training large models, especially deep learning architectures, requires significant computational resources and time.

Supervised learning algorithms, such as support vector machines and neural networks, rely on labeled data to learn patterns and make predictions. For instance, in image classification tasks, a supervised model needs a dataset with images labeled according to their respective classes (e.g., cat, dog, car). Without sufficient labeled data, the model might struggle to generalize to new, unseen images accurately.

Moreover, overfitting is a common challenge in supervised learning. Imagine a scenario where a model is trained on a dataset that predominantly consists of images of white cats. If the model memorizes specific features unique to white cats, it may fail to correctly classify a black cat in a real-world scenario. This highlights the importance of diverse and representative training data to prevent overfitting.

When dealing with large-scale datasets, particularly in deep learning, the computational demands can be immense. Deep neural networks with multiple layers require substantial computational resources for training and inference. For example, training a deep learning model on a dataset of high-resolution images can take days or even weeks on powerful hardware like GPUs. The challenges associated with supervised approaches underscore the critical role of high-quality labeled data, the risk of overfitting, and the resource-intensive nature of training complex models. Balancing these factors is essential for developing accurate and robust supervised learning systems.

6.2.Unsupervised Approaches

Unsupervised sentiment identification, by contrast, does not require labeled data. Instead, these methods attempt to infer sentiment from patterns or relationships in the text itself. Lexicon-based methods, clustering, and topic modeling are common unsupervised techniques. For instance, in a lexicon-based approach, the system uses predefined lists of positive and negative words to determine the overall sentiment of a text by analyzing word frequencies and co-occurrences.

In lexicon-based methods, the system assigns a score to each word based on its sentiment, such as +1 for positive words and -1 for negative words. By summing up these scores across all words in a text, the system can calculate an overall sentiment score. For example, if a text contains words like "happy," "joyful," and "exciting," the sentiment score would likely be positive. On the other hand, words like "sad," "angry," and "disappointing" would contribute to a negative sentiment score. Clustering techniques group similar texts together based on their content, allowing the system to identify common sentiment patterns. For instance, if a group of texts frequently mentions terms related to happiness and satisfaction, the system may infer a positive sentiment. Conversely, if another group of texts contains words associated with sadness and frustration, a negative sentiment can be deduced.

Topic modeling is another unsupervised technique that involves identifying underlying themes or topics in a collection of texts. By analyzing the distribution of words across different topics, sentiment can be inferred based on the prevalent themes. For example, if a topic is dominated by words like "love," "laughter," and "celebration," it suggests a positive sentiment, while a topic characterized by words like "loss," "grief," and "regret" indicates a negative sentiment. Unsupervised sentiment identification methods offer a valuable approach to analyzing sentiment in text data without the need for labeled training data. By leveraging techniques like lexicon-based analysis, clustering, and topic modeling, these methods can effectively extract sentiment patterns and relationships from unstructured text, providing valuable insights for various applications.

6.2.1.Advantages

Unsupervised approaches offer a cost-effective and efficient solution for processing vast amounts of text data without the need for manual labeling. This means that large datasets can be analyzed without the time-consuming task of annotating each piece of information. One practical example of this is in social media monitoring, where unsupervised methods can sift through massive amounts of user-generated content to identify trends and sentiments without the need for human intervention.

These approaches are not limited by domain-specific labeled data, making them adaptable to various industries and fields. For instance, in the healthcare sector, unsupervised methods can be utilized to analyze patient feedback and reviews to improve services and identify areas for enhancement without the constraints of predefined labels.

Another advantage of unsupervised methods is their simplicity and ease of implementation. Lexicon-based techniques, for example, rely on predefined dictionaries or word lists to categorize text, making them more accessible to users with limited technical expertise. This straightforward approach requires fewer computational resources compared to complex supervised models, making it a practical choice for organizations with limited computing capabilities.

Unsupervised methods often offer a higher level of interpretability compared to deep learning models. By utilizing clear rules or clustering techniques, these methods provide transparency in decision-making processes. This transparency is crucial in applications such as fraud detection, where understanding how a system reaches a conclusion is essential for validation and compliance purposes. Unsupervised approaches present a flexible, cost-effective, and interpretable solution for text analysis across various domains. Their versatility, simplicity, and transparency make them a valuable tool in extracting insights from unstructured data without the need for extensive manual intervention.

6.2.2.Limitations

Unsupervised approaches generally perform worse than supervised models because they rely on broad generalizations (e.g., lexicons) and cannot learn domain-specific patterns or context. This means that unsupervised methods struggle to achieve the same level of accuracy and precision as supervised models, particularly in tasks that require nuanced understanding or context-specific insights. For instance, when analyzing sentiment in text, unsupervised approaches may falter when faced with complex sentences that involve sarcasm, irony, or subtle contextual cues. Consider a scenario where a positive word is used in a sarcastic manner; an unsupervised model might misclassify the sentiment due to its inability to grasp the underlying tone.

Lxicon-based methods, which rely on predefined dictionaries of words and their associated sentiments, often fall short in capturing the full sentiment spectrum. These systems may struggle with sentences that convey mixed emotions or subtle nuances, leading to inaccurate sentiment analysis outcomes. For example, a lexicon-based system might misinterpret a sentence that expresses both joy and sorrow simultaneously. Lexicon-based systems are static in nature and do not adapt to evolving language trends or real-time expressions. This lack of adaptability means that they might miss out on capturing the latest slang terms, emerging expressions, or shifts in sentiment that are prevalent in dynamic text streams. As a result, these systems may not be as effective in analyzing contemporary language usage or capturing the ever-changing nuances of sentiment in online conversations.

While unsupervised approaches and lexicon-based methods have their own strengths, they both have limitations that hinder their performance in sentiment analysis tasks requiring a deep understanding of context and emotion. To overcome these challenges, a more sophisticated blend of supervised learning techniques, natural language processing algorithms, and contextual analysis may be necessary to achieve more accurate and nuanced sentiment classification in text data.

7.SentiWordNet

SentiWordNet is a valuable lexical resource that has been specifically designed to facilitate opinion mining and sentiment analysis. It serves as a reliable tool for individuals who require assistance in carrying out these tasks, with its primary objective being to provide support and guidance in such endeavors. The approach used by SentiWordNet for generating term-level opinion polarity is semi-automatic, which ensures that the results are accurate and reliable. This polarity information is made available to users, enabling them to conduct their analyses with greater ease and efficiency.

One example of how SentiWordNet can be used is in the analysis of product reviews. Imagine a company wanting to understand customer sentiments towards a new product they launched. By utilizing SentiWordNet, they can quickly assess whether the reviews are positive, negative, or neutral, allowing them to make data-driven decisions on potential product improvements.

The underlying source of information used to generate this polarity data is the WordNet database, which contains an extensive collection of English words along with their respective relationships. By leveraging this rich database, SentiWordNet provides users with a highly effective solution for opinion mining and sentiment analysis, allowing them to make informed decisions based on reliable insights.

SentiWordNet acts as a guiding light for individuals and businesses seeking to delve into the realm of opinion mining and sentiment analysis. Its robust methodology and reliance on the WordNet database ensure that users have access to accurate and valuable information, empowering them to navigate the complexities of analyzing sentiments with confidence and efficiency. Through its user-friendly interface and comprehensive data, SentiWordNet stands as a beacon of support for those embarking on the journey of understanding and interpreting opinions in the vast landscape of text data.

8.WordNet

WordNet, developed at Princeton University, is a robust lexical database designed to delve into the intricate semantic relationships among English language terms. The core aim of this initiative was to delve deeper into the interconnectedness of words and their meanings within our language. This recognition of the significance of such relationships propelled the creation of WordNet, an indispensable tool for individuals seeking to explore the complexities and nuances of English vocabulary. The process of developing WordNet was no small feat, requiring extensive research into the myriad ways in which words can be linked. This encompassed not only scrutinizing their definitions but also exploring synonyms, antonyms, hypernyms (more general terms), hyponyms (more specific terms), and meronyms (words denoting parts of larger entities).

By meticulously constructing a comprehensive database of interconnected words and concepts, WordNet has furnished linguists and researchers with a potent instrument for delving into the subtleties of English vocabulary. One of the key advantages of WordNet lies in its capacity to enhance users' comprehension of the relationships between different words. For instance, consider the words "big" and "large." While they may appear synonymous at first glance, a closer examination reveals nuanced distinctions in their usage and connotations. WordNet facilitates this exploration by showcasing how these words are interconnected through their meanings and usage contexts.

WordNet's utility extends beyond mere word relationships, as it also sheds light on broader semantic connections within the English language. For example, the relationships between animals and their habitats or between tools and their functions can be elucidated through WordNet's expansive network of terms. This intricate web of connections serves as a testament to the depth and complexity of language, offering a rich tapestry for exploration and analysis. In essence, WordNet stands as a testament to the power of linguistic research and the profound insights that can be gleaned from a thorough examination of word relationships. Its enduring legacy lies in its ability to unravel the intricacies of language and provide a platform for continued exploration and discovery in the realm of semantics.

9.FOR TWITTER DATASET

In our comprehensive investigation, we delved into a diverse array of factors, each wielding a substantial influence on the outcomes of sentiment analysis. The focal point of our research revolved around the utilization of N-gram features, specifically honing in on unigrams (n = 1) and bigrams (n = 2). These particular N-gram features, such as unigrams and bigrams, are commonly leveraged in various text classification tasks, with sentiment analysis being a prominent application. To elucidate, unigrams with a single token and bigrams consisting of two adjacent tokens were meticulously scrutinized. By employing these features, we sought to ascertain the interplay between them, experimenting with boolean attributes to gauge their interconnectedness.

For instance, when considering unigrams with n = 1, we observed that these individual tokens could provide valuable insights into the sentiment expressed within a text. In contrast, bigrams with n = 2 offered a more nuanced understanding by capturing the relationship between adjacent pairs of words. By toggling the boolean qualities associated with these N-gram features, we aimed to discern the impact of their presence or absence on the sentiment analysis process. This dynamic approach allowed us to fine-tune our methodology and extract nuanced sentiments from the text data. The boolean values linked to each n-gram feature played a pivotal role in our study. These values could be activated or deactivated based on user preferences, ensuring flexibility in the analysis process. However, it is crucial to note that the boolean value is only deemed true when the requisite n-gram is detected within the tweet under consideration. This stringent criterion helped us maintain the integrity of our analysis and enhance the accuracy of our findings.

In Table 5.1, we meticulously documented the accuracy ratings obtained from various classifiers, offering a detailed overview of the characteristics we employed in our study. This tabulated data served as a valuable reference point, enabling us to compare and contrast the performance of different classifiers and the impact of varied characteristics on sentiment analysis outcomes. By presenting this information in a structured format, we were able to glean valuable insights and draw meaningful conclusions from our research endeavors.

Table 1: Emotion dataset accuracy utilising several characteristics

Serial Number	Features	# of features	MNB classifier
1	Unigram	4635	95.0%
2	Bigram	17628	71.23%
3	Unigram+Bigram	35356	95.3%
4	POS	12443	92.9%
5	Adjective	1503	84.5%

Upon delving into the realm of data collection, we stumbled upon the remarkable utility of harnessing hashtags to automatically gather tweets, a method that far surpasses the manual annotation of individual tweets. This realization dawned on us as we meticulously compared the dataset constructed through automated hashtag collection with the one painstakingly annotated by hand. Our thorough examination led us to a profound discovery - the inherent precision with which authors express their sentiments through their own words far exceeds what annotators can discern solely from textual analysis.

To illustrate this point further, consider a scenario where a user tweets, "I am thrilled to announce my new project!" The enthusiasm and excitement exuded in this tweet may be crystal clear to the author, but nuances like these might easily elude the comprehension of annotators relying solely on text. This stark contrast in precision underscores the limitations of conventional annotation methods, which often fall short in capturing the true essence of a writer's intentions.

As a consequence, misinterpretations and inaccuracies abound when annotators attempt to decipher the objectives and underlying emotions embedded within a tweet. Such misreadings can lead to erroneous conclusions and flawed analyses, ultimately distorting the original message intended by the author. In essence, our exploration into the realm of tweet collection methods has shed light on the critical role that author-generated content plays in preserving the genuine voice and intent behind each tweet.

Table 2: MNB classifier's F1 rating for the unigram feature

S.No	Class label	Precision(%)	Recall(%)	F1 score(%)
1	anger	99.48	97.98	98.72
2	Fear	94.95	96.72	95.82
3	Joy	92.19	96.42	94.25
4	Love	95.90	95.90	95.9
5	Sad	86.89	98.35	92.26
6	surprise	99.5	97.59	98.53

10.CONCLUSION

This study constitutes an academic exploration into the realm of sentiment analysis, with a particular emphasis on the utilization of lexical resources and machine learning algorithms to categorize the emotional undertones of tweets and text messages - both exemplars of unstructured data sources. Given the profusion of subjective information accessible online, Sentiment Analysis holds manifold applications in diverse sectors such as online advertising and market research. Within knowledge management, opinion data assumes a pivotal role as it frequently influences pivotal decisions. Our investigation seeks to delve into the complexities associated with Sentiment Analysis, while scrutinizing the methodologies devised to tackle them. Unearthing the underlying sentiments from social media data poses a formidable challenge owing to its sheer volume and diversity; hence, we opted for public stream tweets for our analysis utilizing lexicon-based techniques and machine learning methodologies for Sentiment Analysis.

To refine our lexicon while devising our Twitter Sentiment Analysis framework, we subjected the SentiWordNet lexicon to rigorous scrutiny. Nonetheless, we ascertained that the precision was profoundly context-dependent despite encompassing phrases alongside their emotional ratings since pivotal terms particular to a locale were still absent. Hence, the formulation of a lexicon tailored to the test corpus would be advisable for classification endeavors. Our model surpasses SentiWordNet on our dataset by integrating Google search engine ratings via pointwise mutual information computation, thereby enabling the handling of one of Sentiment Analysis' most significant challenges - unforeseen transitions from positive to negative polarities.

REFERENCE

1. Alqahtani, A., Khan, S. B., Alqahtani, J., AlYami, S., & Alfayez, F. (2023). Sentiment analysis of semantically interoperable social media platforms using computational intelligence techniques. Applied Sciences, 13(13), 7599. https://doi.org/10.3390/app13137599

2. Aremu, A. O., & Muhammad, I. (2024). Sentiment analysis in social media: A case study of hike in university school fees in selected Nigerian universities. Journal of Informatics and Web Engineering, 3(2), 98–104. https://doi.org/10.33093/jiwe.2024.3.2.7

3. Dandash, M., & Asadpour, M. (2024). Personality Analysis for Social Media Users using Arabic language and its Effect on Sentiment Analysis. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2407.06314

4. El-Badaoui, M., Gherabi, N., & Quanouni, F. (2024). TED talks comments sentiment classification using machine learning algorithms. Revue D Intelligence Artificielle, 38(3), 885–892. https://doi.org/10.18280/ria.380315

5. Fithriasari, K., Jannah, S. Z., & Reyhana, Z. (2020). Deep learning for social media sentiment analysis. MATEMATIKA, 99–111. https://doi.org/10.11113/matematika.v36.n2.1226

6. Furqan, M., & Nasir, A. F. A. (2024). Big Data approach to sentiment Analysis in Machine Learning-Based Microblogs: Perspectives of religious Moderation Public Policy in Indonesia. Journal of Applied Engineering and Technological Science (JAETS), 5(2), 955–965. https://doi.org/10.37385/jaets.v5i2.4498

7. Hassan, F., Qureshi, N. A., Khan, M. Z., Khan, M. A., Soomro, A. S., Imroz, A., & Marri, H. B. (2023). Performance evolution for sentiment classification using machine learning algorithm. Journal of Applied Research in Technology & Engineering, 4(2), 97–110. https://doi.org/10.4995/jarte.2023.19306

8. Janjua, S. H., Siddiqui, G. F., Sindhu, M. A., & Rashid, U. (2021). Multi-level aspect based sentiment classification of Twitter data: using hybrid approach in deep learning. PeerJ Computer Science, 7, e433. https://doi.org/10.7717/peerj-cs.433

9. Joshi, S., & Deshpande, D. (2018). Twitter Sentiment Analysis System. In International Journal of Computer Applications (Vols. 180–180, Issue No.47, pp. 35–36). https://www.ijcaonline.org/archives/volume180/number47/joshi-2018-ijca-917319.pdf

10. Kimani, J., Karanjah, A., & Kihara, P. (2024). Sentiment Classification of Safaricom PLC social media sentiments on X(Formerly Twitter). Asian Journal of Probability and Statistics, 26(6), 31–40. https://doi.org/10.9734/ajpas/2024/v26i6622

11. Li, C., & Li, F. (2023). Emotion recognition of social media users based on deep learning. PeerJ Computer Science, 9, e1414. https://doi.org/10.7717/peerj-cs.1414

12. Mapping vaccine sentiment: analyzing spanish-language social media posts and survey-based public opinion. (n.d.). JMIR Preprints. https://preprints.jmir.org/preprint/63223

13. Omuya, E. O., Okeyo, G., & Kimwele, M. (2022a). Sentiment analysis on social media tweets using dimensionality reduction and natural language processing. Engineering Reports, 5(3). https://doi.org/10.1002/eng2.12579

14. Omuya, E. O., Okeyo, G., & Kimwele, M. (2022b). Sentiment analysis on social media tweets using dimensionality reduction and natural language processing. Engineering Reports, 5(3). https://doi.org/10.1002/eng2.12579

15. Parveen, N., Chakrabarti, P., Hung, B. T., & Shaik, A. (2023). Twitter sentiment analysis using hybrid gated attention recurrent network. Journal of Big Data, 10(1). https://doi.org/10.1186/s40537-023-00726-3

16. Pratama, J. A., Suprijadi, Y., & Zulhanif, Z. (2017). The Analisis sentimen sosial media Twitter dengan algoritma machine learning menggunakan software R. Jurnal Fourier, 6(2), 85. https://doi.org/10.14421/fourier.2017.62.85-89

17. Pyate, N. P. G., & Srinivasan, N. B. B. (2024). Social Media Platforms: Investigate sentiment analysis for transforming business decisions in car segments. Deleted Journal, 1(2), 70–85. https://doi.org/10.62569/fijc.v1i2.24

18. Twitter Sentiment Analysis using Machine Learning and Optimization Techniques. (2018). In International Journal of Computer Applications (Vols. 179–179, Issue No.19, pp. 5–6). https://www.ijcaonline.org/archives/volume179/number19/bansal-2018-ijca-916321.pdf

19. Vidyashree, K. P., & Rajendra, A. B. (2023). An improvised sentiment analysis model on Twitter data using Stochastic Gradient Descent (SGD) optimization algorithm in Stochastic Gate Neural Network (SGNN). SN Computer Science, 4(2). https://doi.org/10.1007/s42979-022-01607-x

20. Widyayulianto, M. R., Susanty, M., & Irawan, A. (2022). Sentiment analysis terhadap tulisan mengenai Universitas Pertamina di media sosial Twitter. Petir, 15(2), 276–286. https://doi.org/10.33322/petir.v15i2.1197