EVALUATING SENTIMENT ANALYSIS STRATEGIES FOR MACHINE LEARNING AND LEXICON-BASED MODEL

M. P. Sudha Assistant Professor Department of Computer Science, Dwaraka Doss Goverdhan Doss Vaishnav College, Arumbakkam, Chennai E-Mail: mpsudhacsdgvc@gmail.com
Dr.M.Ramesh Kumar Associate Professor and Head PG and Research Department of Computer Science, Government Arts College (Autonomous), Nandanam, Chennai. E-Mail:,
proframeshkumar@gmail.com

Abstract:
The most recent and broadest data technology available on the internet. The internet accumulates massive data from social media networking sites like Facebook, Twitter, and others. More than eighty percent of the data generated is exclusively in text format due to an abrupt surge in digital data. Sentiment analysis is crucial in various fields such as political science, marketing, financial analysis, social media monitoring, and consumer feedback analysis. The research project aims to construct a classifier that can use sentiment analysis between Lexicon-Based approaches like Dictionary-Based and Corpus-Based approaches and Machine Learning Algorithms, including Decision Trees, Support Vector Machines (SVMs), and Neural Networks. To gauge overall sentiment, we used both qualitative and quantitative methods to categorize it as "positive," "neutral," or "negative." We used Machine Learning and an Unsupervised Lexicon-based technique to analyse sentiment and evaluate the model's performance in this study. This method utilizes the different datasets and language processing techniques to generate summarized results. The results of this study add to our understanding of sentiment analysis and provide insight into the success of Lexicon-based and Machine-Learning techniques in this field. Additionally, this research should be valuable to those of us who are relatively new to NLP in determining the best approach for their NLP analysis.

Keywords:
Machine Learning methodologies, Lexicon based approach, Dictionary-based, Corpus-based, Decision trees, Support Vector Machines (SVMs), Neural Networks, *and Natural Language Processing*

1. INTRODUCTION

Sentiment analysis is a Natural Language Processing technique (NLP) that helps to identify the sentiments like positive, negative, and neutral, which are expressed in a document. Nowadays, companies have expansive volumes of content information like emails, chat transcripts, social media comments, and surveys. Ultimately, Machine Learning and Lexicon-Based methods are more adept at focusing on the complexities of human language. They are currently serving as the most popular strategy for sentiment analysis in figure 1. A Machine Learning Technique for sentiment analysis involves developing a model with Machine Learning algorithms based on training data. Sentiment analysis can be organized as a classification issue, with two classes: Supervised learning and Unsupervised learning. The most prevalent strategy is Supervised learning, which uses sentiment labels to train models on large datasets of attached text, such as movie reviews or social media posts. Frequently used algorithms include Decision Trees, Support Vector Machine (SVM), as well as more advanced methods such as Recurrent Neural Networks (RNNs). These models learn to identify sentiment by monitoring trends in word usage and environment. Once trained, these models can assess new text and predict the sentiment with high accuracy. Unsupervised algorithms, such as Clustering, can also be used to investigate and detect sentiment patterns in the absence of labelled data, however, they are less frequent. Overall, Machine Learning improves the scalability and accuracy of sentiment analysis, making it an important tool for organizations and scholars seeking to comprehend public opinion and emotions represented in the text.

Lexicon-based algorithms make use of different methods for evaluating text sentiment are provided by Dictionary-Based and Corpus-Based approaches to sentiment analysis in Figure1. The Dictionary-Based method matches and scores individual words based on their emotional meanings to determine

the overall sentiment of a text using prepared lists of terms labelled with sentiment. Although this approach is simple, it might have trouble with statements with complex meanings and context.
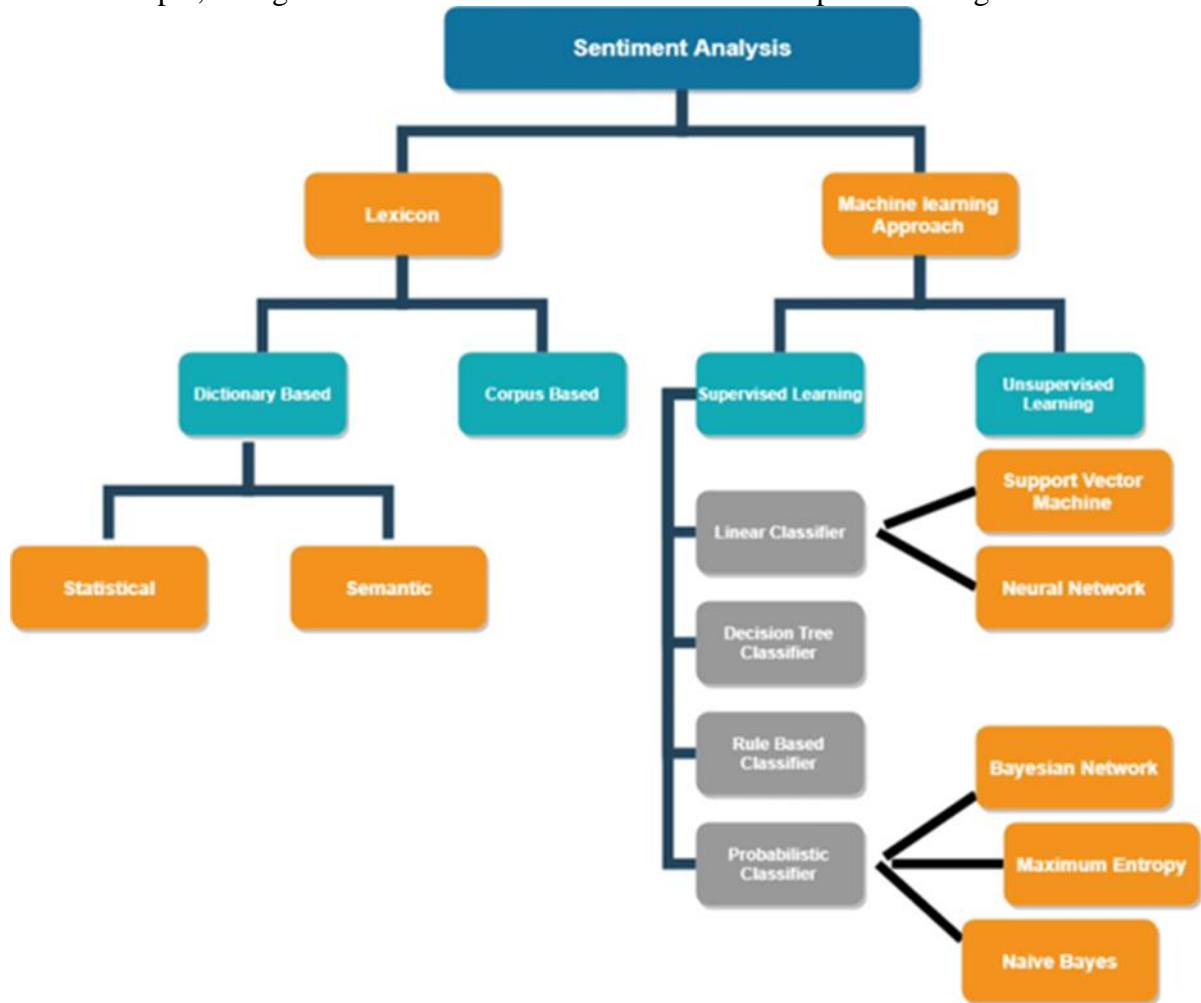

Fig.1. Types of Sentiment Analysis

The Corpus-Based approach uses machine learning models that have been trained on sizable annotated datasets to recognize and predict sentiment based on semantic complexities and cultural patterns. Although this method demands a significant amount of data and computer power, it is more adaptable and can handle complicated moods and a wider range of language use. Dictionary-Based approaches are straightforward and simple to use, but through contextual analysis, Corpus-Based approaches enable a deeper comprehension of sentiment.

The organization of this article is as follows. Section 2 discusses about the various Machine Learning and Lexicon-Based techniques for sentiment analysis of relevant literature. Section 3 explores a table that compares the sentiment analysis of customer evaluations which are produced and analyzed by various researchers. Finally, section 4 offers the research work's concluding observations based on its discoveries.
.

2. LITERATURE SURVEY

A literature review provides an overview of the scholarly works published by experts on a specific subject. Its goal is to acquaint the reader with the concepts and knowledge developed on the subject, including their pros and cons. The literature review should be guided by a central idea, such as the study's objective, the problem being addressed, or the thesis being argued. It should not just be a list of summaries or a comprehensive inventory of resources.

2.1 MACHINE LEARNING METHODOLOGIES FOR SENTIMENT ANALYSIS

Social media users are helped in understanding reviews that are collected from different social media platforms such as Facebook, Instagram, and so on. The several research articles listed below state that sentiment analysis approaches and techniques are used in different review datasets to identify the top ones. A study conducted by Bimantara, M. D., & Zufria, I. in [1] titled "Text Mining Sentiment

Analysis on Mobile Banking Application Reviews". This paper finds favorable reviews as the main goal to capitalize on the advantages of the BRImo app, while addressing negative reviews is the goal in addition to addressing any deficiencies that can affect its competitiveness. Finally, the TF-IDF method, in combination with the NLP approach, was utilized for the sentiment analysis. The SVM algorithm was trained on the training data to perform the analysis effectively. Shoeb, Md, and Jawed Ahmed [2]. "Sentiment analysis and classification of tweets using data mining."Authors' thoughts by applying data mining techniques to sentiment analysis on Twitter data, a widely used social networking platform. The findings indicate that the Decision Tree classifier outperformed K-NN and NaiveBayes in terms of both accuracy and precision, making it the most suitable classifier for sentiment analysis of Twitter data.

Alhardi, anas, and Ahmed alkadasI.[3] title is "Sentiment analysis on hotel reviews using machine learning techniques" The authors examine the body of research on sentiment analysis, with a particular emphasis on the development of machine learning techniques applied to this discipline. This covers both more sophisticated approaches like deep learning and neural networks, as well as more conventional ones like Naive Bayes and Support Vector Machines. Ultimately, the new model outperformed the old one by correctly predicting the target variable with the highest test accuracy score of 0.8493. Campos, Diogo, Rodrigo Rocha Silva, and Jorge Bernardino [4] by "Text Mining in Hotel Reviews: Impact of Words Restriction in Text Classification." The practice of gathering knowledge or patterns of interest from unstructured text sources is known as text mining. Hotels employ reviews from guests to confirm that their own amenities or services are meeting guests' expectations. According to our preliminary findings, Naïve Bayes is the optimal method to utilize as it requires less time and memory to identify the correct class compared to other algorithms.

Bachtiar, Fitra Abdurrachman, Wirdhayanti Paulina, and Alfi Nur Rusydi [5]. "Text Mining for Aspect-Based Sentiment Analysis on Customer Review: this article performs sentiment analysis at the aspect level of the review's text, the data set was compiled from the websites Agoda.com, Expedia, Pegi-Pegi, Booking.Com, and TripAdvisor. Analysis of sentiment classification results in SVM class Accuracy, Precision, Recall, and F1-Score being higher than Naive Bayes, demonstrating that SVM is more successful. F. Neri, C. Aliprandi, F. Capeci, M. Cuadros and T [6]. By, "Sentiment Analysis on Social Media," The study discusses and compares several sentiment analysis techniques, such as rule-based systems, supervised learning, and hybrid methods. It also emphasizes the need to understand the context and the need for advanced algorithms to efficiently manage the complexities of social media data. Machine learning's successful application in this context proves its utility for monitoring public sentiment and managing brand perception in real time.

The study by A. P. Jain and P. Dandannavar[7] "Application of machine learning techniques to sentiment analysis," The research presents a sentiment analysis methodology that leverages Naïve Bayes and Decision trees machine learning algorithms. The analysis demonstrates the efficacy of these techniques in interpreting sentiments conveyed in user-generated content on Twitter. in the paper's suggested framework; the Decision tree method performs very well, with 100% accuracy, precision, recall, and F1-Score. L.Mandloi and R. Patel [8] by "Twitter Sentiments Analysis Using Machine Learning Methods," The fundamental purpose of this study is to evaluate social media platforms such as Twitter have emerged as a valuable source of data for sentiment analysis. Twitter, for example, limits users to 280 characters for each tweet, making sentiment analysis more manageable. The study indicates that, while all three technique likes SVM, Naïve Bayes Classification and Maximum Entropy Classification have advantages and disadvantages, selecting the optimal classifier is determined by the sentiment analysis task's specific criteria. Poornima and K. S. Priya [9] title is "A Comparative Sentiment Analysis of Sentence Embedding Using Machine Learning Techniques,". This study compares the performance of Multinomial Naive Bayes, SVM, and Logistic Regression algorithms for the classification of Twitter texts. The results show that Logistic regression performs better when compared to other supervised machine learning algorithms for Twitter sentiment analysis. Wankhade, M., Rao, A.C.S. & Kulkarni, C. present [10] "A survey on sentiment analysis methods, applications, and challenges". This article presents the study reviews, compares, and investigates methodologies to understand the benefits of sentiment analysis In this field, supervised machine learning techniques are frequently the most widely used approaches because of their ease of use and high accuracy. NB and

SVM classification algorithms are frequently used in sentiment Analysis and serve as standards by which recently suggested methods can be evaluated

Table 1: An overview of the Machine Learning Algorithms technique summarized results

Paper Ref. No.	Authors	Datasets	Methods Used	Results & Accuracy
1	Bimantara, M. D., & Zufria, I.	Google Play Store for the BRImo mobile banking application	Support Vector Machine (SVM),TF-IDF method	The model achieved an accuracy rate of 92%, a Precision Score of approximately 92%, a Recall Score of 100%, and an F1 Score of around 95%, indicating the effectiveness of the sentiment analysis approach used in the research
2	Shoeb, Md, and Jawed Ahmed	Twitter data	Decision Tree, K-NN, and NaiveBayes	Decision Tree - 95.96%, K-NN - 90.00%, and NaiveBayes - 67.08%
3.	Alhardi, anas, and Ahmed alkadasI	The Trip Advisor hotel reviews dataset from Kaggle	Decision Tree, Logistic Regression, and Naive Bayes	Hybrid model high accuracy
4	Campos, Diogo, Rodrigo Rocha Silva, and Jorge Bernardino	Sentiment analysis with hotel reviews from Kaggle	Naïve Bayes, Decision Tree, Random Forest, Support Vector Machine (SVM)	SVM ("Linear") yields higher results in terms of accuracy, Naïve Bayes is faster and requires fewer resources when it comes to text classification
5	Bachtiar, Fitra Abdurrachman, Wirdhayanti Paulina, and Alfi Nur Rusydi	Agoda.com, Expedia, Pegi-Pegi, Booking.Com, and TripAdvisor	Support Vector Machine, Naïve Bayes, TF-IDF, Stratified k-Fold Cross Validation	The SVM class of Accuracy, Precision, Recall, and F1-Score outperforms Naive Bayes in the sentiment classifications even Displaying SVM's greater effectiveness
6	F. Neri, C. Aliprandi, F. Capeci, M. Cuadros and T	Facebook	Bayesian method and K-Means algorithm	The study focused on how social media affects customer preferences, attitudes, and behaviours.
7	A. P. Jain and P. Dandannavar	Twitter data	Multinomial Naive Bayes and Decision trees	The Decision tree algorithm exhibited exceptional performance, achieving 100% accuracy, precision, recall, and

				F1-Score in sentiment analysis tasks
8	L. Mandloi and R. Patel	NLTK's Twitter	Naïve Bayes(NBand Support Vector Machine (SVM)and Logistic Regression	emphasizing their special qualities and aptitude for handling text data.
9	Poornima and K. S. Priya	Twitter	Support Vector Machine (SVM), Multinomial Naive Bayes, and Logistic Regression	The algorithms were trained and evaluated with unigram models, with Naive Bayes, MaxEnt, and SVM achieving accuracies of around 81.0%, 80.4%, and 82.9%, respectively
10	Mayur Wankhade, Annavarapu Chandra Sekhara Rao and Chaitanya Kulkarni	Social media, Weblog, Electronic Commerce website	Machine learning approaches, Lexicon based approaches	NB and SVM classification algorithms are frequently used in sentiment analysis

Table 1 presents findings from multiple studies that have used different datasets and techniques for sentiment analysis. These studies examined sentiment in data from sources like Twitter, Facebook, the Google Play Store, and hotel review websites. They utilized algorithms like Support Vector Machine (SVM), Naïve Bayes, Decision Tree, Logistic Regression, and others. While the accuracy rates of the models varied, some achieved high recall, precision, and F1 ratings. Overall, the studies demonstrated the effectiveness of the sentiment analysis techniques. Being employed, the SVM and Decision Tree algorithms have shown remarkable results in certain instances

2.2 A LITERATURE SURVEY USING LEXICON BASED APPROACHES

A basic method for comprehending sentiment in text is provided by vocabulary-based sentiment analysis. Despite its shortcomings in context, ambiguity, and domain adaptability, it is nevertheless useful, especially when paired with more sophisticated methods. This section provides an overview of the earlier research on lexicon-based sentiment Analysis conducted by various researchers. and have discovered numerous valuable conclusions that can be summarized. Ms. Minu Choudhary and Mr. Prashant Kumar Choudhary's [11] idea of research is "Sentiment Analysis of Text Reviewing Algorithm using Data Mining " The writers talk about how sentiment analysis has developed over time, moving from simple lexicon-based methods to more complex machine learning and deep learning strategies. They pointing out that while lexicon-based approaches are easy to understand and straightforward, they frequently lack accuracy because they are unable to account for language contexts and complexity. The findings show that for sentiment classification, the SVM algorithm performs better than the other techniques in terms of efficiency and accuracy.

Bonta, Venkateswarlu, Nandhini Kumaresh, and Naulegari Janardhan [12] presented "A comprehensive study on lexicon-based approaches for sentiment analysis." This essay places a lot of importance on the difficulties of locating opinion content on many websites, it can be challenging for the average reader to evaluate the quality of reviews and accurately combine multiple viewpoints. Almosawi, Mrtdaa Mohammed, and Salma A. Mahmood[13] by "Lexicon-based approach for sentiment analysis to student feedback." The research shows that textual materials can be analyzed for their positive or negative polarity.In educational institutions, it's important to gauge the satisfaction of students with the performance of their lecturers Educational institutions seeking to improve instruction quality through student assessments will benefit from this knowledge. S. Zahoor and R. Rohilla [14]

by "Twitter Sentiment Analysis Using Lexical or Rule Based Approach: A Case Study," This study effectively suggested feelings based on unprocessed data, such as tweets gathered from Twitter for a range of occasions, including political campaigns and movie premieres. Overall, the paper's findings demonstrate how well the Lexical approach—which is aided by the TextBlob and VADER libraries—works for assessing opinions shared on Twitter about various events. Mowlaei, M., Abadeh, M., & Keshavarz, H [15] by A Lexicon Generation Method for Aspect-Based Opinion Mining. The study's primary focus is on aspect-based sentiment analysis, which involves identifying the polarity-based features in reviews. The performance of the suggested lexicon generation approach is assessed in the study through a comparative analysis with other notable lexicons. The F-measure is used to quantify the outcomes and evaluate how well the classifiers predict sentiment polarity in various domains.

Umar, Mahmood, Hauwa Ibrahim Binji, and Anas Tukur Balarabe [16] by"Corpus-based Approaches for Sentiment Analysis: A Review." The research acknowledges limitations of Corpus-Based sentiment analysis, including data sparsity and challenges with domain adaptation, which can impact its effectiveness in certain situations. The article presents the effectiveness, flexibility, and ongoing enhancements in corpus-based sentiment analysis techniques, supported by various tools, unique methodologies, and evaluation criteria. Shaik, Thanveer, Xiaohui Tao, Christopher Dann, Haoran Xie, Yan Li, and Linda Galligan [17] by "Sentiment analysis and opinion mining on educational data: A survey" The study emphasizes the importance of using both qualitative and quantitative data to comprehend student feedback and enhance teaching methods through sentiment analysis. Azeema Sadia, Fariha Khan, and Fatima Bashir [18] by "An Overview of Lexicon-Based Approach For Sentiment Analysis "This study compares our binary class classification and multiclass sentiment analysis approaches, offering a comprehensive overview of the latest developments in lexicon-based sentiment analysis and their limitations.

Another paper presents a structured approach to sentiment analysis, emphasizing the importance of preprocessing and classification techniques, as well as the limitations of lexicon-based methods reliant on predefined dictionaries.Tandiapa, Solagratia Saron, and Gladly Caren Rorimpandey [19] by "Analysis Sentiment Ulasan Pengguna Pada Aplikasi Threads Dengan Metode Lexicon Based dan Naive Bayes Classifier." We will analyse Android users' sentiments about LinkedIn, one of the apps available on the Google Play Store, using reviews from the comments section The sentiment analysis's findings show that the majority of user sentiments are positive. Alqurashi, Tahani [20]. "Arabic sentiment analysis for Twitter data: a systematic literature review". This systematic review's primary goals are to identify potential new research directions and look into trends in the issues that were reported. This makes it extremely difficult for academics to build Arabic text classifier models for use in SA.

Table 2: A Survey on the experimental outcomes of Lexicon-Based Techniques

Ref. No.	Authors	Datasets	Methods Used	Results & Accuracy
11	Ms. Minu Choudhary and Mr. Prashant Kumar Choudhary	phone reviews collected from Twitter	Dictionary-Based	A Lexicon-Based approach was used for sentiment analysis, a standard method for identifying sentiments from text data.
12	Bonta, Venkateswarlu, Nandhini Kumaresh, and Naulegari Janardhan	Rotten Tomatoes movie reviews	NLTK, TextBlob, and VADER are used in Dictionary Based approaches	VADER is a more effective tool for sentiment analysis than TextBlob in terms of sentiment classification
13	Almosawi, Mrtdaa Mohammed, and Salma A. Mahmood	Their collected feedback	Dictionary-Based	The suggested approach has demonstrated its reliability in evaluating emotions in educational settings, achieving

				an impressive 98% accuracy rate.
14	S. Zahoor and R. Rohilla	Twitter data	NLTK, TextBlob, and VADER are used in Dictionary-based approaches	Sentiments from unlabelled Twitter data can be effectively extracted using this method, which eliminates the necessity for labelled input and output parameters
15	Mowlaei, M., Abadeh, M., & Keshavarz, H	SemEval-2014,user-generated ratings from the restaurant and laptop sectors.	Frequency Based Sentiment Analysis (FBSA) algorithm	SentiWordNet and AFINN, two more popular lexicons for sentiment analysis, were compared to the suggested lexicon. The outcomes showed that although the suggested lexicon did not obtain the maximum accuracy on the restaurant dataset,
16	Umar, Mahmood, Hauwa Ibrahim Binji, and Anas Tukur Balarabe	Social media and News Agency	Corpus Based approach	The results demonstrate that these methods can effortlessly handle various datasets, showcasing their practical utility
17	Shaik, Thanveer, Xiaohui Tao, Christopher Dann, Haoran Xie, Yan Li, and Linda Galligan..	six Moodle datasets from postgraduate programs at the Hellenic Open University	Lexicon-Based and Corpus-Based approaches	With the least human interaction, these algorithms enable the sentiment-oriented classification of student responses.
18	Azeema Sadia, Fariha Khan and Fatima Bashir	Twitter data	Machine Learning, Lexicon Learning, Hybrid Based approaches	binary class classification achieved an accuracy of 85.5%
19	Tandiapa, Solagratia Saron, and Gladly Caren Rorimpandey..	Google Play Store reviews of the LinkedIn application	Lexicon-Based, utilizing the VADER	--
20	Alqurashi, Tahani	Use of Twitter data	The Corpus-based Approach, The Lexicon-based Approach, Hybrid based	MSA can produce satisfactory results while converting Arabic to English.

Table 2 finds the analysis of Sentiment analysis methods using lexicon-based and corpus-based approaches that were applied to various datasets, such as Twitter data, movie reviews, and user-generated ratings. Different tools and algorithms, including NLTK, TextBlob, VADER, and machine learning, were used to achieve high accuracy in sentiment classification. The results demonstrate the practical utility of these methods in handling diverse datasets with minimal human interaction

3. Comparative Analysis of Sentiments in Text Data.
Figure 2 depicts the two techniques visually categorizing various sentiment research techniques. The sentiment analysis techniques used with various datasets and tools are covered in both Tables. A

comparison of several Machine Learning Algorithms (such as SVM and Decision Tree) across multiple data sources is shown in Table 1, which also notes variations in accuracy but highlights specific algorithms with high recall, precision, and F1 scores. Table 2 presents the efficacy of Lexicon-Based and Corpus-Based Techniques, showcasing the high accuracy of sentiment categorization with low manual work of tools like NLTK, TextBlob, and VADER. All things considered, these tables demonstrate the efficacy and usefulness of various sentiment analysis methods in managing heterogeneous data with favourable outcomes. Research results suggest that Supervised Machine Learning techniques, like Support Vector Machines (SVM) and Naive Bayes, indicate better precision.
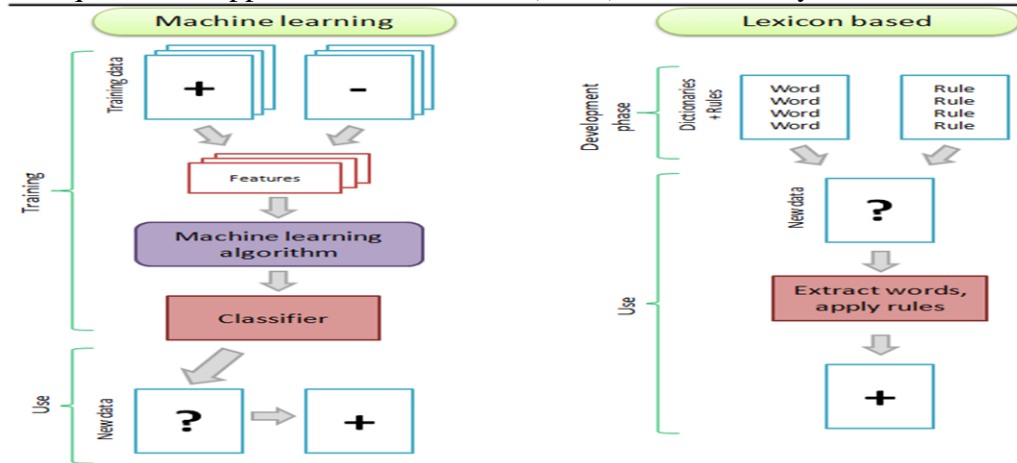

Figure 2: The Report of sentiment Analysis

3.1 Support Vector Machine

SVM algorithms, such as Support Vector Machines, are powerful tools used for a wide range of tasks including regression, outlier identification, and linear or nonlinear classification. They are widely applied in text classification, picture classification, handwriting recognition, spam detection, face detection, gene expression analysis, and anomaly detection due to their ability to handle high-dimensional data and nonlinear relationships. SVMs excel in identifying the largest possible separation hyperplane between multiple classes in the target feature. The primary objective of the SVM algorithm is to find the best hyperplane in an N-dimensional space that can be used to separate data points into different classes in the feature space. The hyperplane aims to maximize the distance between the nearest points of different classes. The dimension of the hyperplane is determined by the number of features. For instance, if there are only two input features, the hyperplane is essentially a line. With three input features, the hyperplane becomes a 2-D plane. Visualizing the hyperplane becomes challenging when there are more than three features.

3.2 Naive Bayes

The Naïve Bayes classifier is an application of the Naïve Bayes method that is often used for text classification. Text classification involves data with high dimensionality, where each word represents a feature. This method is commonly used for rating categorization, sentiment analysis, spam filtering, and other similar tasks. One of the key advantages of Naïve Bayes is its speed, especially when dealing with a large number of data dimensions. This makes it a quick and efficient classification algorithm, allowing for the rapid creation of Machine-Learning models with fast prediction capabilities. The Bayes Theorem calculates the likelihood of an event based on the likelihood of an earlier event.

The following equation represents the mathematical formulation of Bayes' theorem:

$$P(A/B) = P(B)P(B/A)P(A)$$

Essentially, given that event B is true, we are attempting to determine the probability of an event
i) A Evidence is another word for Event.
B. $P(A)$ is the prior probability of A or the probability of an event occurring before evidence is observed. An attribute value of an unidentified instance (in this case, an event
ii) B serves as the proof.
The probability of evidence, or marginal probability, is $P(B)$.

iii)P(A|B) represents the likelihood of an event after evidence is observed, or the a posteriori probability of B.

The likelihood that a hypothesis will materialize in light of the data is represented by the probability P(B|A).

4. CONCLUSIONS

This research work discusses two main subcategories of sentiment analysis namely Machine Learning (ML) and Lexicon-Based approaches. Research comparing these methods indicates that while Lexicon-Based approaches may not always be as accurate as Machine Learning Methods, the performance difference can be minimal. For instance, in one study, Lexicon-Based sentiment analysis achieved an accuracy of 0.95, while a Machine Learning-Based SVM model achieved 0.986. In another study, VADER, a Lexicon-Based tool, outperformed the AFINN Lexicon, with 88.7% and 86.0% accuracy, respectively. However, Lexicon-Based methods may struggle to understand sarcasm, context, and subtle emotional cues. These findings suggest that while Machine Learning Techniques generally offer higher precision, Lexicon-Based Methods remain competitive and practical, especially given their low need for extensive labelled training data. The findings of this study are useful for NLP beginners, as they will assist them in selecting the best sentiment analysis approach based on their specific needs and data characteristics. We have explored various Machine Learning methods that could be utilized with sentiment analysis datasets in the future.

REFERENCES

[1] Bimantara, M. D., & Zufria, I., "Text Mining Sentiment Analysis on Mobile Banking Application Reviews using TF-IDF Method with Natural Language Processing Approach", J*ournal of Information and Visualization*, Vol. *5, Issue* 1, pp. 115-123,2024.

[2] Shoeb, Md, and Jawed Ahmed. "Sentiment analysis and classification of tweets using data mining." *International Research Journal of Engineering and Technology (IRJET)* 4, no. 12, 2017.

[3] Anis, Sarah, Sally Saad, and Mostafa Aref. "Sentiment analysis of hotel reviews using machine learning techniques." *In Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2020*, pp. 227-234, 2021.

[4] Campos, Diogo, Rodrigo Rocha Silva, and Jorge Bernardino. "Text Mining in Hotel Reviews: Impact of Words Restriction in Text Classification." *In KDIR*, pp. 442-449. 2019.

[5] Bachtiar, Fitra Abdurrachman, Wirdhayanti Paulina, and Alfi Nur Rusydi. "Text Mining for Aspect Based Sentiment Analysis on Customer Review: A Case Study in the Hotel Industry." In *IICST*, pp. 105-112, 2020.

[6] F. Neri, C. Aliprandi, F. Capeci, M. Cuadros and T. By, "Sentiment Analysis on Social Media," IEEE*/ACM International Conference on Advances in Social Networks Analysis and Mining, Istanbul, Turke*y, pp. 919-926, 2012. doi: 10.1109/ASONAM.2012.164.
URL: https://ieeexplore.ieee.org/document/6425642

[7] A. P. Jain and P. Dandannavar, "Application of machine learning techniques to sentiment analysis," 2*nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, Bangalore, India, pp. 628-632, 2016. doi: 10.1109/ICATCCT.2016.7912076.
URL: https://ieeexplore.ieee.org/document/7912076

[8] L. Mandloi and R. Patel, "Twitter Sentiments Analysis Using Machine Learning Methods," *2020 International Conference for Emerging Technology (INCET), Belgaum, India* , pp. 1-5, 2020 .doi: 10.1109/INCET49848.2020.9154183.
URL:https://ieeexplore.ieee.org/document/9154183

[9] A. Poornima and K. S. Priya, "A Comparative Sentiment Analysis of Sentence Embedding Using Machine Learning Techniques," 6*th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, pp. 493-496,2020. doi: 10.1109/ICACCS48705.2020.9074312.
URL:https://ieeexplore.ieee.org/document/9074312

[10] Wankhade, M., Rao, A.C.S. & Kulkarni, C. A survey on sentiment analysis methods, applications, and challenges. *Artif Intell Rev* **55**, pp. 5731–5780,2022. URL:https://doi.org/10.1007/s10462-022-10144-1

[11] M. Choudhary and P. K. Choudhary, "Sentiment Analysis of Text Reviewing Algorithm using Data Mining," *International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Tirunelveli, India, 2018, pp. 532-538, 2018. doi: 10.1109/ICSSIT.2018.8748599.
URL:https://ieeexplore.ieee.org/document/8748599

[12]   Bonta, Venkateswarlu, Nandhini Kumaresh, and Naulegari Janardhan. "A comprehensive study on lexicon based approaches for sentiment analysis." *Asian Journal of Computer Science and Technology* 8, no. S2 ,pp.1-6,2019.

[13] Almosawi, Mrtdaa Mohammed, and Salma A. Mahmood. "Lexicon-based approach for sentiment analysis to student feedback." *Webology (ISSN: 1735-188X)* 19, no. 1,2022.

[14] S. Zahoor and R. Rohilla, "Twitter Sentiment Analysis Using Lexical or Rule Based Approach: A Case Study," 8*th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Noida, India, 2020, pp. 537-542,2019. doi: 10.1109/ICRITO48877.2020.9197910.URL:https://ieeexplore.ieee.org/document/9197910

[15] Mowlaei, M., Abadeh, M., & Keshavarz, H. "A Lexicon Generation Method for Aspect-Based Opinion Mining". *IEEE 22nd International Conference on Intelligent Engineering, pp. 000107-000112,2018.*

[16] Umar, Mahmood, Hauwa Ibrahim Binji, and Anas Tukur Balarabe. "Corpus-based Approaches for Sentiment Analysis: A Review." *Asian Journal of Research in Computer Science* 17, no. 7, pp.95-102,2024.

[17] Shaik, Thanveer, Xiaohui Tao, Christopher Dann, Haoran Xie, Yan Li, and Linda Galligan. "Sentiment analysis and opinion mining on educational data: A survey." *Natural Language Processing Journal 2, p.* 100003,2023.

[18] Sadia, Azeema, Fariha Khan, and Fatima Bashir., "An overview of lexicon-based approach for sentiment analysis", *3rd International Electrical Engineering Conference* pp. 1-6. 2018.

[19] Tandiapa, Solagratia Saron, and Gladly Caren Rorimpandey. "Analisis Sentimen Ulasan Pengguna Pada Aplikasi Threads Dengan Metode Lexicon Based dan Naive Bayes Classifier." *Jurnal Cahaya Mandalika ISSN 2721-4796 (online)* 3, no. 1, pp.339-352,2023.

[20] Alqurashi, Tahani. "Arabic sentiment analysis for twitter data: a systematic literature review." *Engineering, Technology & Applied Science Research* 13, no. 2 ,pp. 10292-10300,2023.