International Journal of Engineering, Science and Mathematics
Vol. 10 Issue 2, February 2021,
ISSN: 2320-0294 Impact Factor: 6.765
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

INTEGRATING AI-BASED SENTIMENT ANALYSIS WITH SOCIAL MEDIA DATA FOR ENHANCED MARKETING INSIGHTS

Dheeraj Varun Kumar Reddy Buddula, Software Engineer, Anthem Inc

Hari Hara Sudheer Patchipulusu, Software Engineer, Iheartmedia

Achuthananda Reddy Polu, SDE3, Goldman Sachs

Bhumeka Narra, Sr Software Developer, Statefarm

Navya Vattikonda, Business Intelligence Engineer, International Medical Group Inc

Anuj Kumar Gupta, Oracle ERP Senior Business Analyst ,Genesis Alkali

***Abstract*—Sentimental analysis is a crucial method for opinion mining in light of the ever-expanding reach of social media and the Internet. Recently research work describing the performance of some common methods for determining an individual's emotional state, ranging from basic lexicon-and rule-based approaches to deep learning algorithms. In this study, an Artificial Neural Network (ANN) model for sentiment classification on a Twitter dataset with various data preprocessing, feature extraction using TF-IDF and word embeddings as well as performance evaluation. It is observed that the ANN model surpasses Naïve Bayes (95.98% accuracy) and K-Nearest Neighbors (88.80% accuracy) on all evaluation metrics, accomplishing an improvement in accuracy (97.59%), precision (97.79%), recall (98.64%), and F1 score (97.45%). ANN model is fairly effective in capturing complex sentiment patterns using TFIDF and embedding features, has high classification accuracy and overfitting is modest, therefore leading to strong generalization to new data. These findings validate ANN as a suitable tool for decision-making based on sentiments to enhance marketing, customer engagement and brand reputation management.**

***Keywords*—Sentiment Analysis, emotions, Machine Learning, Social Media Analytics, Twitter dataset.**

I. INTRODUCTION

In the digital era, Modern communication cannot be imagined without social media since people use Twitter, Facebook and Instagram to share their opinions, premonitions, emotions and experiences [1]. Social media is a goldmine for sentiment-rific data containing the public's perception on different subjects, with the amount of posts generated daily running into millions. More and more, businesses, policymakers and researchers are reading this data to measure consumer interest, measure trends, and make decisions. However, social media content is huge and unstructured, so automated sentiment extraction and analysis methods are needed. The sentiment that expresses emotions and opinion that is in the text [2], It has a significant role in understanding consumer attitude. Sentiment analysis is becoming an essential component of public opinion interpretation, product reception evaluation and drawing more customers[3]. Sentiment analysis is a computational technique that uses methods to label the material as good or bad depending on which that can make the said solutions using data-driven techniques. Previous methods have employed traditional sentiment analysis methods, such as lexicon-based and linguistic approaches, to extract sentiments from textual data [4]. However, most of these methods suffer from complex language constructs such as sarcasm, context change and domain specific terminologies.

In marketing, AI-based sentiment analysis in marketing provides businesses an edge to look out for emerging consumer trends and evaluates brand perception and how well an ad copy performs over the net. The integration of AI and ML has really helped sentiment analysis for information gathered via social media[5]. The deep learning & NLP derived AI sentiment analysis technique is much more expedient & scalable as compared to traditional sentiment analysis[6]. When AI trained models are run with large datasets, they can sense the fluctuation of the sentiment variations, know the context, and make sharper predictions. In addition, the widely used methods to perform sentiment classification are ML methods and techniques. The effort to use both training and test data sets to predict the polarity of feelings using ML.

A. Significance and Contributions of the Study

This study is significant because it has the potential for sentiment analysis in the data collected from social media by combining deep NLP techniques with the learning machine models for higher classification accuracy. The proposed

International Journal of Engineering, Science and Mathematics
Vol. 10 Issue 2, February 2021,
ISSN: 2320-0294 Impact Factor: 6.765
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

approach is robust to large dataset sizes through effective preprocessing methods, feature extraction, and selection methods. Sentiment trends give businesses an idea what markets show interest, what marketing strategies are more useful, how to monitor and improve brands perception, and to increase customer engagement. Additionally, this study advances the field by illustrating the impact of techniques on sentiment classification, providing a foundation for future advancements in social media analytics. Key contributions of paper are:

- Create efficient ML and DL models using the Twitter dataset for sentiment analysis.
- improves data representation by implementing feature extraction methods including TF-IDF and word embeddings, noise reduction, tokenization, and stemming.
- Utilizes feature selection methods to enhance model efficiency and adaptability for large-scale social media datasets.
- Integrates advanced NLP techniques with machine learning models (ANN, NB, KNN) to improve sentiment analysis accuracy.
- Evaluate the model efficiency with accuracy, ROC, precision, recall, and f1-score

B. Organization of the Paper

The research is organized as follows: The pertinent literature on incorporating sentiment analysis based on artificial intelligence is reviewed in Section II. In Section III, the study's methodologies, dataset, and strategies are explained. Section IV presents the experimental results and evaluates the recommendation system's effectiveness. Section V wraps up the study by outlining the main conclusions and talking about potential avenues for further research.

II. Literature Review

This section present recent advancements and reviews the body of literature exploring the use of sentiment analysis applied to social data and its role in improving predictive capabilities.

Guo and Li (2019) proposed the TSS model incorporates a new baseline correlation method that reduces computation burden, speeds up decision-making, and doesn't require previous data knowledge while demonstrating reasonable prediction accuracy. Without using previous TSS or market data, the obtained TSS forecasts the future stock market trend 15 times (30 working hours) ahead of time with an accuracy of 67.22% using the suggested baseline criteria. The accuracy of TSS in predicting the future market's upward tendency is 97.87% when utilizing logistic regression and linear discriminant analysis [7].

Saputra and Nurhadryani (2018) aim to enhance performance by taking into account the denial, Indonesian punctuation and sentence structure used to indicate a finished or paused tone. During the 2018 West Java election campaign, Twitter provided the data used in this study. Four steps make up the model-building methodology: feature extraction, classification, assessment, and data preparation. The accuracy of the model used in this research increased by 3%, from 60.15% to 63.7%[8].

Amjad et al. (2017) concentrate on the evaluation of the sentiment of news tweets from Pakistan's major news sources in Urdu. Over the course of 10 months, Twitter data was compiled to develop an Urdu sentiment lexicon. Additionally, create an algorithm that uses the text's cumulative sentiment score to categorize Urdu content into positive, negative, or neutral classifications. The accuracy of the sentiment analysis algorithm is 77%. Additionally, perspective analysis was performed, and the accuracy of the estimation of bias in government-related news reporting via tweets was 77.45% [9].

Juneja and Ojha (2017) Determine the most accurate machine learning classifier and predict the Delhi Corporation Elections' result by utilizing several supervised machine learning classifiers to categorize Twitter data into positive and negative sentiments. The study highlights the effectiveness of several classifiers on the Twitter dataset of political parties. The most accurate sentiment predictors, according to experiments, are the Multinomial Naïve Bayes classifier (78%)[10].

Woldemariam (2016) utilizes the RNTN model in the Stanford core NLP package and the lexicon-based sentiment prediction technique in the Apache-Hadoop framework. and the latter is trained on the Sentiment Treebank, which has 215,154 words and is tagged using Amazon Turk. The overall performance test shows that RNTN outperforms the lexicon- based system by 9.88% on variable length positive, negative, and neutral statements. Additionally, it was discovered that the Lexicon-based F1-score values are 0.16 higher than the RNTN's[11].

Fan et al. (2016) aim to accomplish three subtasks: text sentiment prediction, sentiment word extraction, and sentiment word polarity detection. Analyze the effectiveness of vector representations over different text sources and the quality of domain-dependent vectors. Utilizing simple vector- based features, get an F1 of 85.77%, a recall of 85.20%, and an accuracy of 86.35% for text sentiment analysis of APP evaluations[12].

Table I summarizes comparative analyses of sentiment analysis methods, datasets, performance metrics, and limitations, highlighting advancements in lexicon-based, machine-learning, and domain-specific approaches across various languages and datasets.

TABLE I. SUMMARY OF RECENT ADVANCES IN SENTIMENT ANALYSIS IN SOCIAL MEDIA USING MACHINE LEARNING

Reference	Methodology	Dataset	Performance	Limitations and Future Work
Pradha, 2019	Data preprocessing (stemming, lemmatization, spelling correction), sentiment weight scoring, SVM, DL, and NB models	1,314,000 Twitter data (Google Now vs. Amazon Alexa)	SVM performed best with 90.3% accuracy	No precise cleaning methods were used in previous studies; scope for improving text preprocessing.
Saputra and Nurhadryani, 2018	Considered negation, sentence structure in Indonesian, and punctuation	Twitter data from the 2018 West Java election campaign	Improved accuracy by 3%, from 60.15% to 63.7%	Could explore more advanced NLP techniques for Indonesian language processing

	in sentiment analysis			
Amjad et al. 2017	Sentiment lexicon-based classification of Urdu news tweets; bias estimation in news reporting	News tweets (Urdu) over 10 months	Sentiment analysis accuracy: 77%; Bias estimation accuracy: 77.45%	Focused on news sentiment only, lacks generalizability to other domains. Future work could refine lexicon-based methods and incorporate deep learning.
Han and Kim, 2017	Morphological sentence pattern model for sentiment analysis	IMDb, Rotten Tomatoes, Metacritic, YouTube, Twitter (movie reviews)	Accuracy: 91.2%	Potentialto enhance model generalization across different text genres
Woldemariam, 2016	Recursive neural tensor networks (RNTN), Apache-Hadoop framework, lexicon-based sentiment prediction, Stanford CoreNLP, and sentiment treebank	Sentiment Treebank containing 215,154 words (labeled by Amazon Turk)	In terms of accuracy, RNTN beat the lexicon-based model by 9.88%	Lexicon-based model had higher F1- score; scope for hybrid approaches
Fan et al., 2016	Sentiment words extraction, polarity detection, text sentiment prediction, domain-dependent vector representations	APP review dataset	Accuracy: 86.35%, F1-score: 85.77%, Recall: 85.20%	Potential for improving vector-based feature representation for better sentiment analysis

III. Methodology

The methodology for sentiment analysis with social media data begins with data collection from platforms like Twitter, where a large set of user-generated content is gathered. After that, data preprocessing is performed, including noise removal, tokenization, stemming, and part-of-speech tagging, to clean and structure the text. Next, features like word embeddings and TF-IDF are extracted, Word2Vec or BERT are applied on textual data to transform text into numerical representation. Feature selection is another crucial step that improves model performance and makes possible scalability issues. After that, the dataset is separated into training and testing sets. Machine learning models like ANN, NB, and KNN are trained and validated using the training set. Lastly, the accuracy, precision, recall, F1 score, and ROC analysis are used to assess the classification performance. The results will prove to be insightful to marketers in terms of the customer sentiment trends helped businesses get better at engagement strategies, brand perception and product positioning. In Figure 1, research illustrates the overall workflow of the methodology on the process of sentiment analysis.

International Journal of Engineering, Science and Mathematics
Vol. 10 Issue 2, February 2021,
ISSN: 2320-0294 Impact Factor: 6.765
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Fig. 1. Data Flow Diagram for sentiment analysis

The following provides a quick explanation of each phase in a data flow diagram:

A. *Data Collection*

The present study utilizes the Twitter dataset to analyze sentiment classification. The dataset consists of 1.6 million tweets, with a balanced subset of 16,011 records, including 8,001 positive and 8,010 negative samples. The visual insights aid in understanding dataset composition, improving feature selection, and enhancing sentiment analysis performance. Sentiment class distribution is shown in Figure 2.

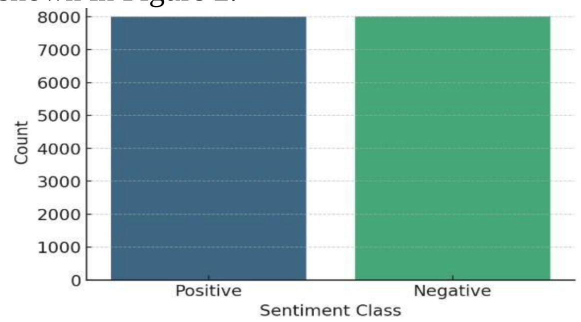

Fig. 2. Sentiment Distribution Analysis in Twitter Data

Figure 2 bar chart depicts the emotion distribution in a Twitter dataset with 16,011 entries, categorized into Positive (dark blue) and Negative (green) sentiment classes. Both classes are nearly equally distributed, each containing around 8,000 samples, ensuring dataset balance. This balance enhances the robustness of sentiment analysis models by providing equal representation of both sentiment categories.

Figure 3 shows a right-skewed distribution of tweet sentiment, with a mode around 0.25, indicating more positive than negative tweets. The sentiment range spans from -1 to 1, with a few outliers on the negative side. This provides a quick insight into the sentiment distribution, highlighting overall positivity and some extreme negatives.

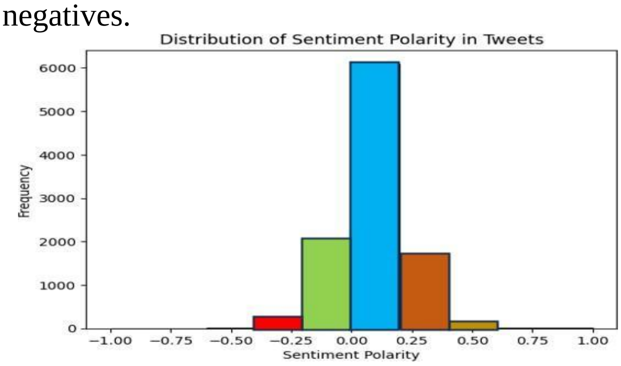

Fig. 3. Histogram for Distribution of sentiment Polarity

B. *Data Preprocessing*

The process of data preparation is the process of transforming raw data into an understandable, structured format suitable for analysis and machine learning. Various NLP tasks, including using part-of-speech tagging, tokenization, and word stemming, are included in this module.[13]. Key terms of pre-processing steps are as follows:

- **Garbage Removal:** In this phase, customized regular expressions are used to eliminate the text of undesired characters (non-ASCII characters), including online addresses, URLs, and internet connections.
- **Tokenization:** A stream of text can be tokenized by breaking it up into words, phrases, symbols, or other important elements called tokenization. The Apache Lucene package's Ling Pipe Tokenizer is used to tokenize the text while maintaining punctuation. At first, employed String Tokenizer; however, because of its intrinsic restrictions.
- **Stemming:** The stemming method is used to remove a word's prefix and suffix. It can also mean identifying a word's root and stem and eliminating them[14].

C. Feature Optimization

The feature vector size is notably big when features are extracted using bag-of-words data structures since the feature vector contains all keywords with any associated sentiment value. When used for a bigger dataset, this approach presents serious scaling issues. In order to address this issue, that must optimize the feature vector by decreasing its size without sacrificing accuracy.

D. Data Splitting

Data splitting involves dividing data into several groups with the aim of training and testing models. Two sets of data, called training and testing sets, were used in this investigation. Between the two data sets, the training set has 80% and the testing set contains 20%. The data splitting ratio is 80:20.

E. Classification with Artificial neural network

One branch of AI that draws influence from the human brain is neural networks[15]. Layered neurons in an ANN process information by responding dynamically to external stimuli[16]. An ANN is a kind of supervised deep learning; it uses a multi-layer structure, with several nodes or neurons per layer, to simulate the way the brain works [17]. Each neuron in the model receives input from all of the neurons in the layer below it and sends out signals to all of the neurons in the layer above it. Before feeding it into a function like Sigmoid or Ru, each neuron in the model has its inputs multiplied by their own weights and a bias. Neurons in the subsequent layer get the result of the applied function as input. Neurons at the last layer provide the ultimate forecast. What each neuron produces. Using hidden layer neurons yielded the best results in terms of accuracy. In Equation (1)

$$f(b + \sum xiwi) \quad (1)$$

- **False Positive (FP):** shows the quantity of data that are misclassified as the occurrence of any unwanted event.
- **True Negative (TN):** displays the quantity of accurately categorized records as typical.
- **False Negative (FN):** displays the quantity of records with inaccurate classifications as typical.

Accuracy: Accuracy is the proportion of accurate predictions to all of the testing dataset's predictions. Equation (2) is used to compute the accuracy:

$$Accuracy = \frac{True\ Positive + True\ Negative}{Totl\ number\ of\ transections} \quad (2)$$

Precision: The ratio of anticipated positive observations to the total number of positive observations is known as precision. Here's how it's calculated in Equation (3):

$$Precision = \frac{True\ Positive}{True\ Positiv + False\ Positive} \quad (3)$$

Recall: Recall is the proportion of correctly predicted positive observations to all of the class's observations. This is how it is calculated in Equation (4):

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{4}$$

F1-Score: The weighted average of recall and accuracy is called the f-score. a more significant metric than accuracy in cases when the data's class distribution is not uniform. This is how it is computed Equation (5):

$$F1 - Score = \frac{2*Precision*Recall}{(Precision+Recall)} \tag{5}$$

Loss: Loss is a measure of the model's prediction error, representing how far the predicted values deviate from the actual values. It is used to update model weights during training.

- xi are the inputs from the neurons in the preceding layer
- wi are the weights for that neuron
- b is the bias term

F. Model Evaluation

This paper evaluates the performance of sentiment analysis by utilizing various metrics, which helps to analyze the effectiveness of models on the Twitter dataset for sentiment analysis. The models are first assessed using confusion matrices in terms of TP, FP, TN, and FN, from which:

- **True Positive (TP):** shows the quantity of data that were accurately categorized as the existence of any unwanted occurrence.

ROC: The ROC is a visual plot that illustrates a classifier's performance. The ROC enables us to examine their model's performance overall potential thresholds.

IV. Result Analysis And Discussion

This research was conducted and assessed utilizing a machine running Windows 10 Professional 64-bit including a Core TM-i5 8250U CPU running at 1.8 GHz and 12 GB of RAM. Although the ML and DL models used for sentiment analysis in social media are implemented using Python 3. As indicated in Table II, the experiment results of the suggested ANN model on the Twitter dataset are presented in this section throughout the performance matrix.

TABLE II. Results of ANN Model for the Sentiment Analysis on the Twitter Dataset

Performance Matrix	Artificial Neural Network (ANN)
Accuracy	97.59
Precision	97.79
Recall	98.64
F1-score	97.45

International Journal of Engineering, Science and Mathematics
Vol. 10 Issue 2, February 2021,
ISSN: 2320-0294 Impact Factor: 6.765
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

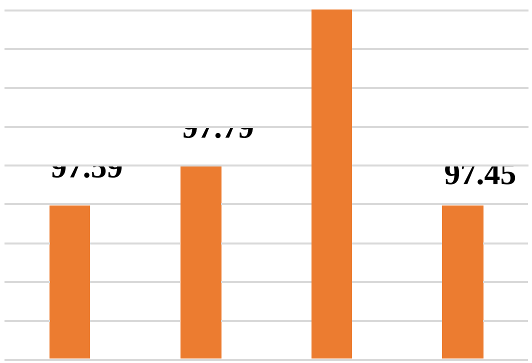

Fig. 4. Bar graph of ANN model Performance

The results of the ANN show in Figure 4, the bar chart and Table II illustrate the ANN model's excellent sentiment analysis performance on the Twitter dataset. It produces a high F1 score of 97.45%, recall of 98.64%, precision of 97.79%, and accuracy of 97.59%. The recall metric stands out most as it evaluates the model's accuracy of picking out the positives, thus making ANN a sound option for sentiment analysis in marketing.

Fig. 5. Training and Validation Accuracy of ANN model

Figure 5 shows the trends in the training and validation accuracy of the ANN model across 25 epochs. The accuracy for both training and validation increases steadily in the initial epochs, showing rapid improvement before gradually stabilizing. The training accuracy and validation accuracy roughly match, suggesting little overfitting. The model achieves near-perfect accuracy towards the final epochs, suggesting effective learning and generalization. The graph highlights that the ANN model successfully converges while maintaining consistency between training and validation performance.

In Figure 6, the graph shows the ANN model's training and validation losses, which are 25 epochs. Each of these two losses reduces in the initial epochs, showing that learning is taking place effectively. While such variance is normally observed, the overall losses are comparatively minimal, with validation losses comparatively close to those of training losses. This shows that the proposed model yields good generalization performance and brings little overfitting to the sentiment analysis tasks.

Fig. 6. Training and Validation loss of ANN model

International Journal of Engineering, Science and Mathematics
Vol. 10 Issue 2, February 2021,
ISSN: 2320-0294 Impact Factor: 6.765
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

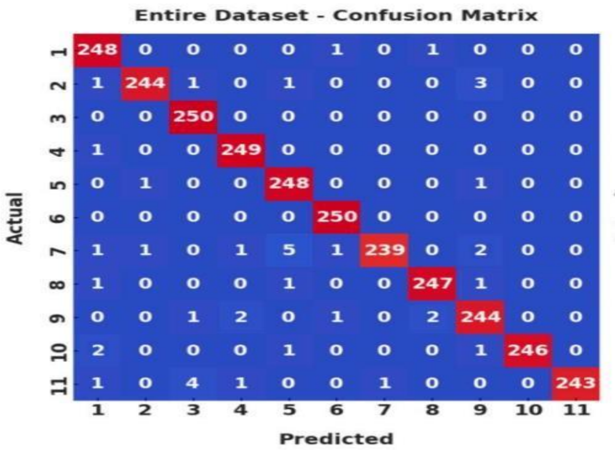

Fig. 7. Confusion matrix for ANN model

Figure 7 shows the confusion matrix demonstrates the fact of high accuracy of sentiment analysis provided by the ANN model where most of the predictions are positioned correctly along the diagonal (248, 244, 250). There are a few misclassifications that show that the model is accurate for all 11 classes of images. This explains the ANN's better scores, including a 97.59% accuracy and 97.45 % F1 score, making it effective in multi-class sentiment analysis of data from social media.

A. Comparative Analysis of the Models

There is a comparative analysis in this section of three AI- based learning algorithms, including ANN, NB [18], and KNN[19], for analyzing sentiment in social media Twitter data across a performance matrix.

TABLE III. COMPARISON BETWEEN ANN WITH DIFFERENT MODELS ON VARIOUS EVOLUTION MATRICES.

Model	Accuracy	Precision	Recall	F1-score
ANN	97.59	97.79	98.64	97.45
NB[18]	95.98	96.22	95.98	95.97
KNN[19]	88.80	83.00	75.00	79.00

Table III evaluates different ML models to use sentiment analysis to gain deeper marketing insights from social media data. In terms of performance, the ANN model is superior to others by generating 97.59% accuracy, 97.79% precision, 98.64% recall and achieves 97.45% F1-score which establishes it as the most dependable model. Naive Bayes implements 95.98% accuracy alongside consistent metrics following KNN's moderate performance reaching 88.80% accuracy coupled with 83.00% precision alongside 75.00% recall. ANN proves itself as the best tool for extracting actionable insights which strengthens marketers' capability to construct data-driven strategies.

V. Conclusion And Future Scope

The vast amount of information on the internet has necessitated the development of systems that classify and display data in a way that is both comprehensible and practical. In this regard, sentiment analysis—which uses polarity to categorize texts—has received a lot of interest lately. The study introduces an improved framework for sentiment analysis dealing with large imbalanced Twitter datasets while using ML and DL approaches to interpret unstructured social media data. This study presents an ANN- based sentiment analysis model for social media data, demonstrating superior performance with a 97.59% accuracy, outperforming Naïve Bayes and KNN in all evaluation metrics. The findings highlight ANN's effectiveness in extracting valuable sentiment insights, enabling businesses to refine marketing strategies and enhance customer engagement. However, limitations exist, including the computational complexity of ANN models, the need for extensive labeled data, and potential biases in sentiment classification due to imbalanced datasets. Future research can explore the integration of transformer-based architectures like BERT and GPT for enhanced contextual understanding, domain adaptation techniques to improve model generalization across diverse datasets, and real-time sentiment tracking applications for dynamic market analysis.

References

[1] C. A. Iglesias and A. Moreno, "Sentiment analysis for social media," *Appl. Sci.*, vol. 9, no. 23, pp. 1–4, 2019, doi: 10.3390/app9235037.

[2] B. Liu, "Sentiment Analysis and Opinion Mining," *Synth. Lect. Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, May 2012, doi: 10.2200/S00416ED1V01Y201204HLT016.

[3] Z. Li, Y. Fan, B. Jiang, T. Lei, and W. Liu, "A survey on sentiment analysis and opinion mining for social multimedia," *Multimed. Tools Appl.*, 2019, doi: 10.1007/s11042-018-6445-z.

[4] G. Xu, Z. Yu, H. Yao, F. Li, Y. Meng, and X. Wu, "Chinese Text Sentiment Analysis Based on Extended Sentiment Dictionary," *IEEE Access*, 2019, doi: 10.1109/ACCESS.2019.2907772.

[5] A. Hasan, S. Moin, A. Karim, and S. Shamshirband, "Machine Learning-Based Sentiment Analysis for Twitter Accounts," *Math. Comput. Appl.*, 2018, doi: 10.3390/mca23010011.

[6] V. Kolluri, "An In-Depth Exploration of Unveiling Vulnerabilities: Exploring Risks in AI Models and Algorithms," *Int. J. Res. Anal. Rev.*, vol. 1, no. 3, 2014.

[7] X. Guo and J. Li, "A Novel Twitter Sentiment Analysis Model with Baseline Correlation for Financial Market Prediction with Improved Efficiency," in *2019 6th International Conference on Social Networks Analysis, Management and Security, SNAMS*

2019, 2019. doi: 10.1109/SNAMS.2019.8931720.

[8] F. T. Saputra and Y. Nurhadryani, "Analysis of Indonesian sentiments using Indonesian sentiment lexicon by considering Denial," in *2018 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2018*, 2019. doi: 10.1109/ICACSIS.2018.8618142.

[9] K. Amjad, M. Ishtiaq, S. Firdous, and M. A. Mehmood, "Exploring Twitter news biases using urdu-based sentiment lexicon," in *2017 International Conference on Open Source Systems & Technologies (ICOSST)*, IEEE, Dec. 2017, pp. 48–53. doi: 10.1109/ICOSST.2017.8279004.

[10] P. Juneja and U. Ojha, "Casting online votes: To predict offline results using sentiment analysis by machine learning classifiers," in *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, IEEE, Jul. 2017, pp. 1–6. doi: 10.1109/ICCCNT.2017.8203996.

[11] Y. Woldemariam, "Sentiment analysis in a cross-media analysis framework," in *2016 IEEE International Conference on Big Data Analysis (ICBDA)*, IEEE, Mar. 2016, pp. 1–5. doi: 10.1109/ICBDA.2016.7509790.

[12] X. Fan, X. Li, F. Du, X. Li, and M. Wei, "Apply word vectors for sentiment analysis of APP reviews," in *2016 3rd international conference on systems and informatics (ICSAI)*, 2016, pp. 1062– 1066.

[13] F. Iqbal *et al.*, "A Hybrid Framework for Sentiment Analysis Using Genetic Algorithm Based Feature Reduction," *IEEE Access*, vol. 7, pp. 14637–14652, 2019, doi: 10.1109/ACCESS.2019.2892852.

[14] V. V. Kumar, A. Sahoo, and F. W. Liou, "Cyber-enabled Product Lifecycle Management: A Multi-agent Framework," *Procedia Manuf.*, vol. 39, pp. 123–131, 2019, doi: 10.1016/j.promfg.2020.01.247.

[15] J. Q. Gandhi Krishna, "Implementation Problems Facing Network Function Virtualization and Solutions," *IARIA*, pp. 70–76, 2018.

[16] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, and H. Tang, "Predicting Diabetes Mellitus With Machine Learning Techniques," *Front. Genet.*, 2018, doi: 10.3389/fgene.2018.00515.

[17] Y. H. Rajarshi Tarafdar, "Finding majority for integer elements," *J. Comput. Sci. Coll.*, vol. 33, no. 5, pp. 187–191, 2018.

[18] D. Gamal, M. Alfonse, E.-S. M.El-Horbaty, and A.-B. M.Salem, "Twitter Benchmark Dataset for Arabic Sentiment Analysis," *Int. J. Mod. Educ. Comput. Sci.*, vol. 11, no. 1, pp. 33–38, Jan. 2019, doi: 10.5815/ijmecs.2019.01.04.

[19] M. Rezwanul, A. Ali, and A. Rahman, "Sentiment Analysis on Twitter Data using KNN and SVM," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 6, pp. 19–25, 2017, doi: 10.14569/ijacsa.2017.080603.

[20] Routhu, K., Bodepudi, V., Jha, K. M., & Chinta, P.

C.R. (2020). A Deep Learning Architectures for Enhancing Cyber Security Protocols in Big Data Integrated ERP Systems. *Available at SSRN 5102662*.