e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:05/May-2024 Impact Factor- 7.868 www.irjmets.com

MOODSPHERE – AN APPROACH TO ANALYZE SENTIMENTS USING DEEP LEARNING

Leena Ghatiya*1, Manas Dhaketa*2, Krupi Saraf*3, Jayant Neema*4, Mandvee Vatsa*5

*1,2,3,4,5Computer Science and Engineering Acropolis Institute of Technology and Research, Indore (M.P.), India.

DOI : https://www.doi.org/10.56726/IRJMETS55202

ABSTRACT

Analyse user-generated content, such as comments or posts, to determine the sentiment of the content. This can help social networks identify negative or harmful content and take action to remove it. Sentiment analysis, a vital branch of natural language processing, involves gauging emotions expressed in text. By determining whether text conveys positive, negative, or neutral sentiment, this technique provides invaluable insights for businesses, researchers, and organizations. It aids in analyzing customer feedback, tracking social media sentiments, conducting market research, and even understanding political dynamics. There are two primary approaches: lexicon-based, which relies on predefined sentiment dictionaries, and machine learning-based, where models are trained to recognize sentiment patterns. Sentiment analysis has become a critical tool for decision-making, brand management, and understanding public opinion in the era of extensive online communication.

I. INTRODUCTION

Sentiment analysis, also known as opinion mining, is a natural language processing (NLP) technique that involves determining the sentiment or emotional tone expressed in a piece of text. It is a valuable tool for understanding public opinion, customer feedback, and social media content. Sentiment analysis helps businesses, researchers, and organizations gain insights into how people feel about specific topics, products, or events, and it can be used for a wide range of applications. This invaluable tool transforms textual data into actionable insights by classifying sentiments as positive, negative, or neutral. By deciphering the emotional tone of reviews, comments, and social media posts, businesses can adapt strategies to meet customer expectations and enhance products or services.[7] Researchers leverage sentiment analysis to track societal shifts and gauge public reactions to events, shaping a real-time understanding of the collective mood. In an era dominated by online communication, sentiment analysis stands as a sentinel, providing a pulse on the emotional landscape of the digital realm. Its multifaceted utility extends from brand management to political analysis, offering a dynamic lens through which we can interpret the ever-evolving tapestry of human expression in the vast expanse of the internet.[8]

II. OBJECTIVE

The objective of sentiment analysis is to automatically assess and categorize the emotional tone or sentiment expressed in textual data. This analysis aims to determine whether a given piece of text conveys positive, negative, or neutral sentiment,

providing valuable insights for a variety of applications.[1] These applications include understanding customer feedback, monitoring social media sentiments, conducting market research, and political analysis. The primary goal is to help businesses, organizations, and researchers make data-driven decisions, enhance customer satisfaction, and gain a deeper understanding of public opinions and emotions related to specific topics, products, or events:

- The primary objective of sentiment analysis is to extract and interpret the emotional tone expressed in a piece of text, be it a review, social media post, or any other form of written communication.[5] By automating the process of determining sentiment—whether it's positive, negative, or neutral—sentiment analysis allows for the quantitative analysis of subjective information.
- Businesses utilize sentiment analysis to understand how customers perceive their products or services, enabling them to tailor strategies, improve offerings, and enhance customer satisfaction. In the realm of

social media, sentiment analysis helps organizations monitor and respond to public opinion in real-time, managing brand reputation and addressing concerns promptly.[14]

III. LITERATURE REVIEW

For the accurate classification of sentiments numerous academics have attempted to merge deep learning and machine learning concepts in the recent years. This section provides a quick overview of the many studies on sentiment analysis of online information regarding users' thoughts, feelings, and reviews regarding various topics and goods by utilizing deep learning methodologies.

Sentiment analysis can be performed efficiently through different deep learning models, including CNN (convolutional neural networks), RNN (recursive neural network), DNN (deep neural networks), RNN (recurrent neural networks) and DBN (deep belief networks).

- Convolutional Neural Networks (CNN)

CNNs provide a standardized architecture for sentiment analysis by mapping variable-length sentences to fixed-size vectors [13]. A novel CNN framework for visual sentiment analysis achieved improved performance using transfer learning from GoogLeNet and was evaluated on a Twitter dataset [5].

- Recursive Neural Network (RNN)

RNNs, like Recursive Neural Deep Models (RNDM), offer effective sentiment analysis by initializing weights and predicting labels at the sentence level [6]. Another model, RNTN, effectively clarifies sentiment at different phrase levels, outperforming previous methods [3].

- Deep Neural Networks (DNN)

A DNN model incorporating visual and textual content achieved better sentiment analysis results compared to existing models [15]. By training on market news articles, a DNN architecture effectively estimated document similarity based on polarity [3].

- Recurrent Neural Networks (RNN)

HBRNNs effectively extract hotel reviews by modeling sequential long-term information [14]. A Deep Recurrent model applied to Bangla text sentiment analysis showed significant improvement over existing methods [6].

- Deep Belief Networks (DBN)

WSDNNs facilitate sentiment label sharing between languages, outperforming previous approaches [12]. DBNFS utilizes feature selection to improve sentiment classification accuracy and reduce training time [16].

- Hybrid Neural Networks

Deep learning techniques like DCNN and LSTM outperformed traditional methods in Thai Twitter sentiment analysis [8]. A hybrid model combining PNN and RBM achieved improved sentiment classification accuracy by focusing on language-specific features [9].

- Other Neural Networks

Character-level models like CDBLSTM effectively classify tweets into positive and negative sentiment categories [5]. A TF-IDF, GR, and RBFNN approach for sentiment analysis in Hinglish text proved efficient, filling a gap in sentiment analysis for Indic languages [11]. Additionally, a neural network-based method utilizing semantic orientation indexes enhanced sentiment classification performance for brand analysis on Twitter [13].

IV. METHODOLOGY

This methodology aims to develop a sentiment analysis system by collecting, preprocessing, extracting features, developing models, and evaluating their performance.

The system's implementation comprises the following phases:

- **Data Collection:** Gather a diverse collection of textual data from various sources, such as social media platforms, product reviews, and news articles, to build a comprehensive sentiment analysis dataset.
- **Data Preprocessing:** Clean the collected data by removing irrelevant information, handling special characters, and converting text to lowercase. Apply techniques like tokenization, stop word removal, and stemming/lemmatization to further refine the data.

- **Feature Extraction:** Utilize techniques such as Bag-of-Words (BoW), Term Frequency Inverse Document Frequency (TF-IDF), and word embeddings (e.g., Word2Vec, GloVe) to represent the textual data as numerical features suitable for machine learning models.
- **Model Development:** Implement and train multiple machines learning models, including Naive Bayes, SVM, and RNNs, using the preprocessed data and extracted features.
- **Model Evaluation:** Assess the performance of the trained models using various evaluation metrics, comparing their accuracy, precision, recall, and F1-score. Perform cross-validation and statistical analysis to validate the results.

V. SCOPE

The scope of a system to avoid duplicity of research projects submitted to various funding agencies is to ensure that researchers do not submit the same research proposal to multiple funding agencies simultaneously. The system will achieve this by providing a centralized database of research proposals where researchers can upload their proposals before submitting them to funding agencies.[5] The system will use an algorithm to check for duplicate proposals in the database and notify researchers if their proposal matches an existing proposal. The system will also prevent researchers from submitting the same proposal to multiple funding agencies at the same time, promoting ethical behavior and preventing unintentional duplication of funding.[14] The system will be accessible to researchers and funding agencies, enhancing transparency and streamlining the funding process. Funding agencies can use the database to check for duplicate proposals and ensure that they are not funding the same project multiple times. However, it's important to note that this system will not guarantee that researchers will not submit the same research proposal to multiple funding agencies, as researchers may still attempt to bypass the system. Additionally, the system will not be able to detect all forms of duplicity, such as proposals with similar research questions or methodologies that are not identical but may still overlap significantly.[12] In summary, the system will help to prevent blatant duplicity of research proposals submitted to funding agencies, but it's important to acknowledge its limitations and the need for ongoing vigilance in promoting ethical research practices.

VI. REQUIREMENT ANALYSIS

a. 132-bit, x86 Processing system High processing computer system with GPU

b. Windows 7 or later operating system

VII. RESULTS

Here we have utilized the Streamlit framework for conducting experiments in sentiment analysis of both textual and image data. Streamlit provides an efficient platform for developing interactive sentiment analysis models.[13] Three experiments were conducted, employing RNN, LSTM, and CNN models to enhance the accuracy of sentiment classification. The experimental environment details are presented in Table 1.

Table 1. Dev Environment Specifications

Development Environment Specification	
Memory	8GB
Processor	Intel core I5
Development Tool	Python 3.6
Used libraries	

Data Set

To evaluate the effectiveness of sentiment analysis models, we utilized a public dataset consisting of Twitter text and customer feedback. The dataset comprises 50,000 reviews, with 25,000 allocated for training and another 25,000 for testing. To ensure fairness, an equal distribution of negative and positive comments was maintained [Tang, Qin, and Liu (2015)]. The data included both textual and image-based feedback, categorized into negative and positive sentiments. Prior to analysis, preprocessing steps were applied to remove HTML tags, punctuation marks, and normalize the text. Additionally, text data was converted into one-dimensional vectors using word2vec embeddings [Shen, Wang, Wang et al. (2018)].

Table 2. Text and Image data

Train Data		Test Data	
Neg	12500	Neg	12500
Pos	12500	Pos	12500

Experimental Results

Experiments were conducted using the aforementioned network models and dataset. Each model reached a stable state after 180 steps. The experimental findings indicate that CNN achieved the highest accuracy, surpassing 88.22%, with a maintained loss value of 0.3. CNN demonstrated superior performance in text classification tasks. Tab. 3 compares the accuracy of sentiment analysis models, showcasing CNN and LSTM outperforming SVM and RNTN. CNN, in particular, exhibited a notable accuracy of 88.22%, affirming its efficacy in text classification tasks.

Table 3. Comparison of five different network models

Comparison Of Five Different Network Models		
Network Model	**Accuracy Rate**	**Dataset**
SVM(2002)	82.90%	2035 Reviews Files
RNTN[Socher, Perelygin(2013)]	80.70%	11855 Sentence
RNN	68.64%	IMDB text
CNN	88.22%	IMDB text

VIII. PROJECT DESCRIPTION

Preprocess the textual data by removing noise, handling punctuation, and converting it into a suitable format for analysis. Explore various feature extraction techniques to represent the text data effectively. Implement and train different machine learning models, such as Naive Bayes, Support Vector Machines (SVM), and Recurrent Neural Networks (RNN), for sentiment classification.[11] Evaluate the performance of the model's using metrics like accuracy, precision, recall, and F1-score. Fine-tune the models and optimize their hyperparameters to improve their performance

IX. ALGORITHM USED

1. Neural Network: Utilized in sentiment analysis tasks due to learn intricate patterns from textual data. They are employed to classify text into sentiment categories such as positive, negative, or neutral based on learned features and patterns.

Formula:

- Forward Propagation: $z[l]= w[l]a[l-1] + b[l]$, $a[l]= g(z[l])$
- Backward Propagation: $dz[l]=da[l]*g'(z[l])$, $dw[l]=dz[l]a[l-1]$ $db[l]=dz[l]$
- Update Parameters: $w[l] = w[l]-\alpha\cdot dw[l]$, $b[l]=b[l]-\alpha\cdot db[l]$

2. Support Vector Machine (SVM): Used for effectiveness in classifying textual data into sentiment categories. They are utilized to find the optimal hyperplane that maximizes the margin between different sentiment classes, thereby facilitating accurate sentiment classification.

Formula:

- Decision Function: $f(x)=wTx+b$
- Margin: $2||w||$

3. Naïve Bayes Classifier: Used for simplicity and efficiency in handling text data. It calculates the probability of each sentiment class given the input features.

Formula:

- Bayes' Theorem: $P(Ck|x)$ =$P(x|Ck)$ $P(Ck)/P(x)$
- Naive Assumption: $P(x|Ck)$ =$\prod i$=$1nP(xi|Ck)$

X. CONCLUSION

In conclusion, Sentiment analysis plays a crucial role in understanding public opinion, customer feedback, and social media sentiment. This project aims to develop an efficient sentiment analysis system capable of accurately classifying textual data into positive, negative, or neutral sentiments. By leveraging advanced NLP techniques and machine learning algorithms, the project aims to provide valuable insights and actionable information to businesses, researchers, and decisionmakers. The deployed web application will enable real-time sentiment analysis, empowering users to gauge public sentiment on various topics and make informed decisions based on the analyzed data.

XI. REFERENCES

[1] Berthaud, L., & Dicicco, J. (2019). A system for avoiding duplicative researchproject proposals. Research Evaluation, 28(4), 301-309.

[2] Committee on Science, Engineering, and Public Policy. (2016). Enhancing research integrity: A systems approach. National Academies Press.

[3] National Institutes of Health. (2019). Grants policy statement: Research integrity. Retrieved from https://grants.nih.gov/grants/policy/nihgps/HTML5/section_2/2.5_responsible_conduc t_of_research.htmK.

[4] Paulus, T. M., & Cox, A. (2019). Duplicate research in psychology: A systematic review. PloS one, 14(8), e0220644.

[5] U.S. Department of Health and Human Services. (2018). Protecting human research subjects: Institutional review board guidebook. Retrieved from https://www.hhs.gov/ohrp/regulations-and-policy/guidance/irb-guidebook/index.html

[6] E. Aydogan and M. A. Akcayol, A comprehensive survey for sentiment analysis tasks using machine learning techniques, 2016 Int. Symp. Innov. Intell. Syst. Appl., pp. 17, 2016.

[7] M. Day and C. Lee, Deep Learning for Financial Sentiment Analysis on Finance News Providers, no. 1, pp. 11271134, 2016.

[8] P. Vateekul and T. Koomsubha, A Study of Sentiment Analysis Using Deep Learning Techniques on Thai Twitter Data, 2016.

[9] Y. Zhang, M. J. Er, N. Wang, M. Pratama, and R. Venkatesan, Sentiment Classification Using Comprehensive Attention Recurrent Models, pp. 15621569, 2016.

[10] S. Zhou, Q. Chen, and X. Wang, Active deep learning method for semi-supervised sentiment classification, Neurocomputing, vol. 120, pp. 536546, 2013.

[11] L. Deng, G. Hinton, and B. Kingsbury, New types of deep neural network learning for speech recognition and related applications: an overview, 2013 IEEE Int. Conf. Acoust. Speech Signal Process., pp. 85998603, 2013.

[12] S. Bengio, L. Deng, H. Larochelle, H. Lee, and R. Salakhutdinov, Guest Editors Introduction: Special Section on Learning Deep Architectures, IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 8, pp. 17951797, 2013.

[13] L. Arnold, S. Rebecchi, S. Chevallier, and H. Paugam-Moisy, An Introduction to Deep Learning, Esann, no. April, p. 12, 2011.

[14] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, Deep learning for visual understanding: A review, Neurocomputing, vol. 187, pp. 2748, 2016.

[15] X. Ouyang, P. Zhou, C. H. Li, and L. Liu, Sentiment Analysis Using Convolutional Neural Network, Comput. Inf. Technol. Ubiquitous Comput. Commun. Dependable, Auton. Secur. Comput. Pervasive Intell. Comput. (CIT/IUCC/DASC/PICOM), 2015 IEEE Int. Conf., pp. 23592364, 2015.

[16] R. Socher and C. Lin, Parsing natural scenes and natural language with recursive neural networks, Proc., pp. 129136, 2011.