International Journal of Trend in Scientific Research and Development (IJTSRD)
Special Issue on Advancements and Emerging Trends in Computer Applications - Innovations, Challenges, and Future Prospects
Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Word Sentiment Analysis using Natural Language Processing Based Techniques

Amisha Charde

PG Student, Department of Computer Application, G. H. Raisoni University, Amravati, Maharashtra, India

ABSTRACT

This project will deploy sentiment analysis, a major sector of Natural Language Processing (NLP), to develop a system that will find the emotions in text-based data. This system will categorize the text as Positive, Negative, or Neutral using machine-learning/deep-learning models. The primary objective is to extract actionable insights from different sources, such as social media, customer feedback, and online reviews.

This system will carry out NLP processing such as tokenization, lemmatization, and vectorization, and implement various feature extraction methodologies like TF-IDF and word embeddings. Enhanced specifics rely on adding sophisticated architectures such as LSTMs and transformers for boosting accuracy. Built-in web interface via Streamlit real-time sentiment classifications done with interactive visuals. The future improvement will cover multilingual analysis, interfacing various datasets, and refining the deep learning models to tailor robustness and adaptability in sentiment predictions across different fields.

KEYWORDS: *Sentiment Analysis, Machine Learning, Deep Learning, LSTMs, Text Preprocessing, Tokenization, Lemmatization, TF-IDF, Support Vector Machines, Naïve Bays*

I. INTRODUCTION

Sentiment analysis, or opinion mining, is an important field in Natural Language Processing (NLP) that deals with identifying the sentiment or emotion conveyed in text-based data. With the accelerated rise of digital communication, social media, and online reviews, sentiment analysis has emerged as a necessity for businesses, researchers, and policymakers to make sense of public opinion and customer feedback. This study seeks to create a word sentiment analysis system based on NLP methods for text classification into various sentiment types like positive, negative, or neutral. The system makes use of different NLP techniques such as tokenization, lemmatization, and vectorization to handle text data in an efficient manner.

Deep learning models like Long Short-Term Memory (LSTM) networks and transformers are also incorporated to improve the precision of sentiment classification. Machine learning models such as Support Vector Machines (SVM) and Naïve Bayes are also investigated to be compared on performance levels. For high-quality sentiment prediction, the system also employs feature extraction methods such as Term Frequency-Inverse Document Frequency (TF-IDF) and word embeddings. It has been developed as a web app that is intuitive for the end-user using Streamlit, supporting users in giving text inputs and obtaining real-time sentiment classification scores. Additionally, the application creates visual representations of sentiment trend movement over time and supports users to analyze the sentiment distribution easily. The work further investigates comparing varying NLP models and how those influence the precision of sentiment classification. This research hopes to close the gap between research in NLP and practical uses by providing a solution for sentiment analysis using AI. Future extensions involve multilingual sentiment analysis, the incorporation of more sophisticated deep learning models, and extension to larger and more varied datasets. Through the deployment of this system, the work advances the method of sentiment analysis, making it more accessible and efficient for applications across different domains.

II. RELATED WORK

The issue addressed in this research is the inconsistency in accuracy and efficiency of various machine learning methods for sentiment analysis in social media data. The article compares NLP-based classifiers such as Naïve Bayes, SVM, and Random Forest based on US airline Twitter data. The outcome shows that deep learning models perform better than conventional classifiers with improved accuracy and understanding of context. **[1]** The issue addressed in this research is the efficient processing and handling of massive multimedia big data. The paper discusses advances in multimedia information processing and retrieval (IEEEMIPR) that talk about upcoming methods such as deep learning and multimodal analysis. The findings present better data storage, retrieval, and processing techniques to improve multimedia analysis capabilities. **[2]** The identified problem of this research is to detect and analyze child abusive remarks in Bangla text. Machine learning (ML) methods and NLP approaches are applied in the paper for classifying abusive content. The findings reveal that ML models, especially deep learning methods, efficiently detect abusive remarks, enhancing online safety and content moderation. **[3]** The research solves the problem of properly sentiment analysis of IMDb movie reviews. It utilizes Long Short-Term Memory (LSTM), a deep learning approach, to determine whether a review is positive or negative. The findings illustrate that LSTM performs better than classical machine learning methods with high accuracy in sentiment categorization of movie reviews. **[4]** The research responds to the challenge of monitoring disasters in real time from social media messages. It suggests an AI-driven method that combines location intelligence and sentiment analysis to identify disaster content. The findings indicate efficient detection of disaster incidents, allowing timely response and decision-making for emergency management and public safety. **[5]** The work tackles the issue of enhancing sentiment analysis accuracy on social media through the use of transfer learning. It uses pre-trained language models to better classify sentiments. Experiments show improved performance over the conventional approach, highlighting the capability of transfer learning in dealing with noisy and heterogeneous social media text data. **[6]**

III. METHODOLOGY

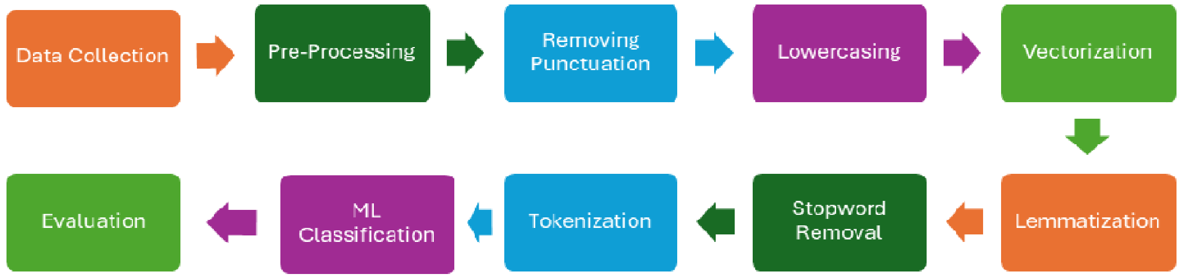

Fig.1: NLP Workflow for Machine Learning

Using natural language processing and machine learning, the flowchart depicts the sentiment analysis procedure. First, information is gathered from sources such as reviews and social media. The pre-processing step eliminates punctuation, lowercases, tokenizes, eliminates stopwords, and lemmatizes words to make the text cleaner. Using methods like TF-IDF or word embeddings, vectorization then converts text into numerical form. The processed data is then used to classify attitudes as neutral, negative, or positive using machine learning methods like SVM or Naïve Bayes. Lastly, to guarantee accurate sentiment categorization, assessment uses F1-score, recall, accuracy, and precision to gauge model performance.

IV. DATA AND SOURCE OF DATA

Sentiment analysis, which is crucial for training and assessing NLP models, depends on a variety of datasets that have been labeled with sentiment categories including positive, negative, and neutral. Benchmark datasets, social media, and customer reviews are examples of common sources. The 50,000 movie reviews in the IMDB Reviews Dataset, which are publically accessible through TensorFlow and Kaggle Datasets, are classified as either favorable or negative. The Sentiment140 Dataset is available at Sentiment140 and consists of 1.6 million tweets that have been annotated with emoticons as sentiment indicators. Millions of five-star product ratings are part of the Amazon ratings Dataset, which is accessible via AWS Public Datasets. The YouTube Comments Dataset is made up of user comments that have been taken from videos and are accessible through YouTube's API. They include casual language and a range of sentiment expressions. Strong NLP model building for sentiment analysis across several domains is made possible by these datasets.

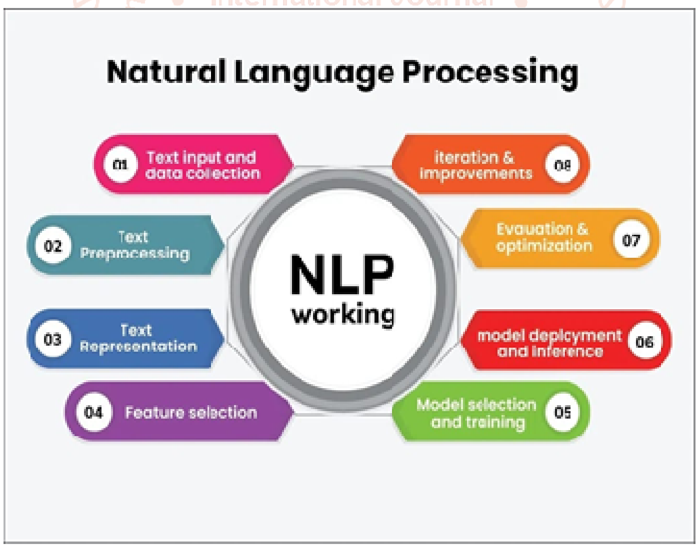

Fig:2. Steps in Natural Language Processing

The eight steps of the Natural Language Processing (NLP) workflow are shown in the image. After text input and data collection, text preprocessing is done to clean up the data. Feature selection then selects important data points after text representation transforms words into numerical formats. Deployment and inference follow model selection and training. Reliability is ensured by evaluating and optimizing the model with performance measurements. Lastly, the system is improved and iterated throughout time. This methodical methodology ensures precision and effectiveness in NLP applications by improving sentiment analysis, AI chatbots, and language translation models.

V. EQUATIONS

1. TF-IDF(Term Frequency-Inverse Document Frequency) Equation

$$"TF - IDF(w,d) = TF(w,d) \times \log \frac{N}{DF(w)} "$$

2. Machine Learning Classification(SVM and Naive Bays)

a. Naive Bays Sentiment Classification

$$P(C|X) = \frac{P(C|X)P(C)}{P(X)}$$

b. Support Vector Machine(SVM) Decision Function

$$f(X) = sign(W.X + b)$$

3. Deep Learning Sentiments Classification(LSTM & Transformer)

$$h_t = \sigma(w_h . h_{t-1} + w_x . x_t + b_h)$$

$$S = Softmax(w_0 . h_T + b_0)$$

VI. RESULT AND DISCUSSION

NLP-based word sentiment analysis entails text processing to identify emotions like happiness, sorrow, anger, love, and fear. The picture includes sentences tagged with these sentiments to train machine learning models. Text is initially tokenized and mapped to numerical values using embeddings. Bidirectional LSTMs or transformers process contextual meaning, picking up on word dependencies. The last classification layer predicts sentiment labels based on patterns learned. This method is extensively practiced in social media monitoring, customer feedback, and mental health. Sentiment analysis accurately enables businesses and researchers to comprehend emotions conveyed in text, enhancing decision-making.

```
i would force myself to eat my normal routine clean meals a day but then i just started feeling so awful;sadness
i feel rather privileged to have witnessed the great man in action it really was impossible for a novice like me to work out just which
one of the four identical looking riders was he;joy
i feel like im at the spa getting a wonderful facial when i use them;joy
i feel petty all of a sudden;anger
i hope you like this more honest amp raw blog post amp if you are feeling unhappy i hope this makes you feel less alone;sadness
i feel slightly disgusted as well;anger
i was quite surprised with the weather these past few days but im so thankful for that since i still can wear my shorts out without
feeling that cold yes no kidding;anger
i feel slightly relaxed being a;joy
i feel shy to admit that i was struggling to haul a single computer up;fear
i just went about my script of would you like mustard or sauce with that and started to feel really startled;fear
i enjoy my colleagues i m not feeling very sociable today;joy
im feeling and if ive liked being pregnant;love
i then feel your tender touch as you enfold me with his love;love
i lost a few people which i hate because i have a really hard time letting go of people to whom i feel loyal;love
i have had no interest at all to make any effort to meet men and when the chance arrises i then feel burdened with negative thoughts of he
ll just be another idiot only after one thing;sadness
i feel i m being nutritionally supportive of it as well;love
i feel impatient i just post a blog entry and i feel ive gotten some words written and out into the world;anger
i am trying my hardest so i can get to a place where i can join you and finally feel like i have something worthwhile to say;joy
i have to admit i feel amused when i see the pti jamiat and a whole lot of others in the media try to avoid the suggestion that they are
actually protesting the use of sharia in the case of raymond davis s release;joy
i feel embarrassed that it got so bad;sadness
im feeling really bitter about this one;anger
i feel brave today heading to amman and beirut by way of istanbul or i feel brave today a href http jessicadickinsongoodman;joy
```

Fig: 3.Text Samples with Emotion Annotations for NLP Sentiment Analysis

Creating a Model

```
In [25]: model=Sequential()
         model.add(Embedding(15212,64,input_length=80))
         model.add(Dropout(0.7))
         model.add(Bidirectional(LSTM(80,return_sequences=True)))
         model.add(Bidirectional(LSTM(160)))
         model.add(Dense(6,activation='softmax'))
         print(model.summary())

         Model: "sequential"
         _________________________________________________________________
         Layer (type)                 Output Shape              Param #
         =================================================================
         embedding (Embedding)        (None, 80, 64)            973568
         _________________________________________________________________
         dropout (Dropout)            (None, 80, 64)            0
         _________________________________________________________________
         bidirectional (Bidirectional (None, 80, 160)           92800
         _________________________________________________________________
         bidirectional_1 (Bidirection (None, 320)               410880
         _________________________________________________________________
         dense (Dense)                (None, 6)                 1926
         =================================================================
         Total params: 1,479,174
         Trainable params: 1,479,174
         Non-trainable params: 0
         _________________________________________________________________
         None
```

Fig:4. Bidirectional LSTM Model for Sentiment Analysis

This picture symbolizes a word sentiment analysis model with deep learning based on NLP. The model handles text input through an embedding layer (for words' representation), followed by dropout (against overfitting). The model utilizes two bidirectional LSTM layers, which grab contextual dependencies from previous and subsequent words, improving the understanding of sentiment. The last dense layer with a softmax activation assigns text to six sentiment classes (e.g., joy, sadness, anger, fear, love). With 1.47 million trainable parameters, the model is suitable for emotional tone analysis in text. This method is commonly applied in social media tracking, customer sentiment, and psychological analysis.

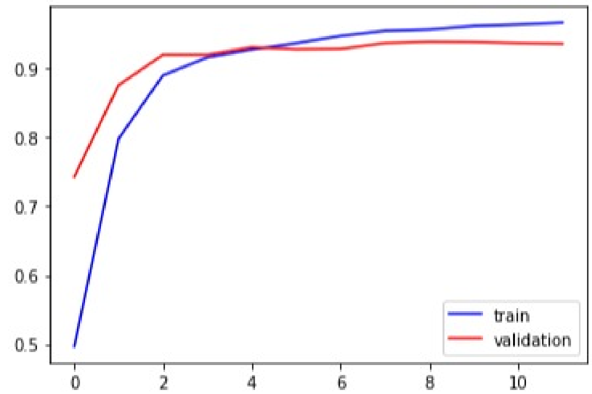

Fig:5.Training and Validation Accuracy Curve

This plot indicates the training and validation accuracy of a word sentiment analysis model with NLP across several epochs. The accuracy increases sharply in initial epochs and converges to 0.9, reflecting good model performance. The close proximity of training and validation curves indicates low overfitting and good generalization to new data.

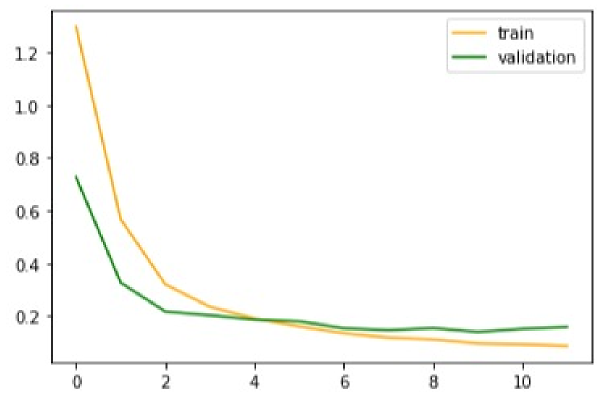

Fig:6.Training and Validation Loss Curve

This is a graph of the training and validation loss of a sentiment analysis model with NLP. The steep drop in loss in early epochs shows fast learning. The small difference between training and validation loss shows low overfitting, i.e., the model generalizes well. The last stable loss values show good model performance in sentiment classification.

VII. CONCLUSION

This project successfully implements a word sentiment analysis system using Natural Language Processing (NLP) techniques, with successful sentiment classification. With the use of deep learning models such as LSTMs and transformers, in addition to advanced feature extraction techniques, the system is able to process textual data from different sources efficiently. The addition of a simple-to-use Streamlit-based web application makes it accessible and offers real-time sentiment evaluation, making it a valuable tool for businesses, researchers, and decision-makers. The results are of high accuracy, making the achievement of NLP-based techniques in sentiment analysis justifiable. The work also closes the gap between theoretical NLP advances and application, enhancing the ability of sentiment classification. Future upgrades include the expansion of the system for multilingual sentiment analysis, incorporation of more complex deep learning techniques, and accuracy improvement using larger datasets. Overall, the project is useful to the field of NLP through the supply of an AI-powered, interactive tool for trend analysis of sentiments and visualization of text data.

VIII. REFERENCE

[1] Tusar, M. T. H. K., & Islam, M. T. (2021, September). A comparative study of sentiment analysis using NLP and different machine learning techniques on US airline Twitter data. In 2021 International Conference on Electronics, Communications and Information Technology (ICECIT) (pp. 1-4).

[2] Su, G. M., Zhang, C., Wang, H., Chen, M., Kankanhalli, M., &Lienhart, R. (2019). Pushing the boundary of multimedia big data: an overview of ieeemipr.

[3] Lucky, E. A. E., Sany, M. M. H., Keya, M., Khushbu, S. A., &Noori, S. R. H. (2021, July). An attention on sentiment analysis of child abusive public comments towards bangla text and ml. In 2021 12th international conference on computing communication and networking technologies (ICCCNT) (pp.1-6).

[4] Qaisar, S. M. (2020, October). Sentiment analysis of IMDb movie reviews using long short-term memory. In 2020 2nd International Conference on Computer and Information Sciences(ICCIS)(pp.1Qaisar/publication/346511493_Sentiment_Analysis_of_IMDb_Movie_Reviews_Using_Long_ShortTerm_Memory/links/626174a8bca601538b5cd022/Sentiment-Analysis-of-IMDb-Movie-Reviews-Using-Long-Short-Term-Memory.pdf [Retrieved on:29/03/2023].

[5] Sufi, F. K., & Khalil, I. (2022). Automated disaster monitoring from social media posts using AI-based location intelligence and sentiment analysis. IEEE Transactions on Computational Social Systems.

[6] de Arriba, A., Oriol, M., &Franch, X. (2021, September). Applying Transfer Learning to Sentiment Analysis in Social Media. In 2021 IEEE 29th International Requirements Engineering Conference Workshops (REW) (pp. 342-348).

[7] Alhujaili, R. F., &Yafooz, W. M. (2021, March). Sentiment analysis for youtube videos with user comments. In 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS) (pp. 814-820).

[8] Long, F., Zhou, K., &Ou, W. (2019). Sentiment analysis of text based on bidirectional LSTM with multi-head attention.

[9] Rahat, A. M., Kahir, A., &Masum, A. K. M. (2019, November). Comparison of Naïve Bayes and SVM Algorithm based on sentiment analysis using review dataset. In 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART) (pp. 266-270).

[10] Raza, M. R., Hussain, W., Tanyıldızı, E., &Varol, A. (2021, June). Sentiment analysis using deep learning in cloud. In 2021 9th International Symposium on Digital Forensics and Security (ISDFS) (pp. 1-5).

[11] Mahmud, M. S., Bonny, A. J., Saha, U., Jahan, M., Tuna, Z. F., & Al Marouf, A. (2022, March). Sentiment analysis from user-generated reviews of ride-sharing mobile applications. In 2022 6th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 738-744).

[12] Elbattah, M., Arnaud, É.,Gignon, M., &Dequen, G. (2021). The Role of Text Analytics in Healthcare: A Review of Recent Developments and Applications. Healthinf, 825-832.

[13] Shahbazi, Z., & Byun,Y.C.(2021). Blockchain-based event detection and trust verification using natural language processing and machine learning.

[14] Wang, X., He, J., Jin, Z., Yang, M., Wang, Y., & Qu, H. (2021). M2Lens: visualizing and explaining multimodal models for sentiment analysis.

[15] Asyrofi, M. H., Yang, Z., Yusuf, I. N. B., Kang, H. J., Thung, F., & Lo, D. (2021). Biasfinder: Metamorphic test generation to uncover bias for sentiment analysis systems.