SENTIMENT ANALYSIS

Mrs Snehal Shah[1], Miss Akshata Bhat[2], Miss Sumitra Singh[3], Miss Arya Chavan[4], Master Aryan Singh[5]

[1]Electronics And Telecommunication, Shree L R Tiwari College of Engineering Mumbai, India.

[2,3,4,5]Electronics and Computer Science Shree L R Tiwari College of Engineering, Mumbai, India.

DOI: https://www.doi.org/10.58257/IJPREMS33384

ABSTRACT

The process of locating and categorizing viewpoints or feelings represented in the source text is known as sentiment analysis. Social media information like blog posts, status updates, and tweets generate a tonne of sentiment-rich data. Understanding the opinions of the masses can be enormously aided by sentiment analysis of this user-generated data. Twitter sentiment analysis is more challenging than generic sentiment analysis because of the prevalence of misspellings and slang terms.

Twitter allows a character count of up to 140 characters. The knowledge base approach and Machine learning approach are the two strategies used for analyzing sentiments from the text. In this study, we try to analyze Twitter posts on electronic devices, like computers and smartphones, using the machine learning technique. By doing sentiment analysis in a specific domain, it is possible to identify the effect of domain information in sentiment classification. We introduce a new feature vector for extracting people's thoughts about products and classifying the tweets as positive or negative.

1. INTRODUCTION

In the era of digital connectivity, social media platforms such as Twitter have evolved into indispensable repositories of up-to-the-minute information and perspectives. Numerous individuals worldwide express their opinions on a wide range of topics regularly, analyzing this massive amount of data essential for determining public opinion, measuring market dynamics, evaluating brand reputations, and more. It is within this context that sentiment analysis emerges as a vital tool. [3]

Sentiment analysis, a discipline within natural language processing (NLP), entails computationally discerning and cat-egorizing opinions expressed in text as positive, negative, or neutral. [2] Utilizing machine learning and NLP techniques, sentiment analysis algorithms can sift through vast amounts of textual data, extracting valuable insights without the risk of copyright infringement [1]

One popular platform for conducting such analyses is Google Colab, a cloud-based Jupyter Notebook environment that allows for seamless execution of Python code. For ma- chine learning applications, such as sentiment analysis, GoogleColab is especially well-suited because it provides free GPU and TPU access. [3]

Setting up sentiment analysis of tweets in Google Colab involves importing relevant libraries, such as TensorFlow, NLTK (Natural Language Toolkit), or Scikit-learn, and loading the Twitter dataset. [2]Sentiment analysis typically involves the following steps after the environment is ready: data pre- processing, feature extraction, model training, evaluation, and prediction. [4]

Data preprocessing encompasses several tasks aimed at en- hancing the quality of textual data. In addition to cleaning and tokenizing the text, stop words are eliminated, and stemming and lemmatization techniques are employed to standardize the language. [3]Subsequently, feature extraction transforms the textual data into numerical features, facilitating comprehension by machine learning algorithms. [2]

Labeled Twitter datasets can be utilized to train machine learning models using preprocessed data and extracted features. Numerous techniques are accessible, such as support vector machines, logistic regression, and more sophisticated deep learning models like recurrent neural networks (RNNs) or transformer-based designs like BERT (Bidirectional Encoder Representations from Transformers). [3]

In summary, employing sentiment analysis on tweets via Google Colab presents an effective avenue for understanding public sentiment and trends on one of the most widely used social media platforms globally. Professionals and researchers can gain important insights that apply to a broad spectrum of fields, from politics and marketing to customer service and beyond, by utilizing the potential of machine learning and natural language processing. [4]

2. RESEARCH METHODOLOGY

In order to perform sentiment analysis, we are required to collect data from the desired source (here Twitter). This data undergoes various steps of pre-processing which makes it moremachine sensible than its previous form. [2]
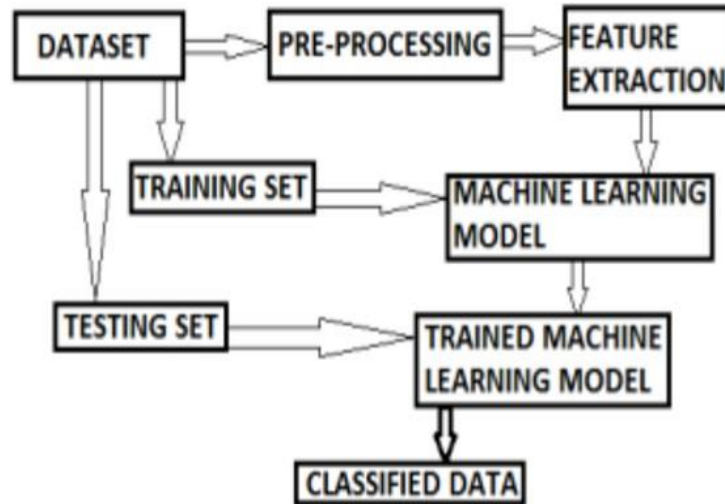

Fig. 1. Block Diagram

Dataset: The machine learning process will start with thisset of data.

Pre-processing: This step involves cleaning the data, han-dling missing values, and normalizing the data. [4]

Feature Extraction: Selecting the most crucial features from the dataset to train the model is known as feature selection.

Training Set: This dataset will serve as the training set forthe machine learning model.

Machine Learning Model: Using the training set is intendedto train this particular model. [3]

Testing Set: We will use this subset of the dataset to evaluatethe trained machine-learning model.

Trained Machine Learning Model: This corresponds to themodel trained using the training set.

Classified Data: This is the trained machine learningmodel's output after being tested on the testing set. [3]

A. Data Processing

- The method has been trained using the Kaggle dataset, which is displayed below:

- A short Python code and its result, showing a sample of data from a Twitter-training CSV file.

- The code imports the pandas library, reads the twit- ter training.csv file, and prints the shape and a random sample of the DataFrame. [2]

- The Excel file included four columns, and the review function only returned good and negative tweets.

- Positive and negative tweets were up to 56364, and the table printed was texts and their sentiments.

- The data then gets divided into three categories: positive, negative, and irrelevant. [2]

- The positive data is then assigned a value of 0, the negative a value of 1, and the irrelevant a value of -1.

- Converting the text input into a language that a machine learning model can comprehend is now the main compo- nent of Python sentiment analysis. Convert the text into an array of vector embeddings. Word embeddings are a beautiful way of representing the relationship between thewords in the text. [3]

- To accomplish this, assign a unique number to each one of these distinctive words before replacing it with the assigned number. [3]

- First, retrieve all the text data from the dataset. [1]

- Before starting with the Python sentiment analysis project, tokenize all the words in the text using Tokenizer. Tokenization breaks down all the words/sentences of a text into small parts called tokens. [2]

- The fit on texts() method associates the words with the provided integers. The tokenizer.word index attribute stores this association as a dictionary.

- Replace the words with their assigned numbers using the text to sequence() method. [3]

- Each of the sentences in the dataset is not of equal length. The sentences have been shortened to equal length using this technique. [2]

INTERNATIONAL JOURNAL OF PROGRESSIVE
RESEARCH IN ENGINEERING MANAGEMENT
AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 04, April 2024, pp: 1542-1547

e-ISSN :
2583-1062

Impact
Factor:
5.725

B. Building the text classifier

• In sentiment analysis projects, machine learning models commonly employ LSTM (Long Short Term Memory) layers in their architecture. This setup typically consists of an embedding layer, an LSTM layer, and a Dense layer at the end. A Dropout mechanism is frequently added between the LSTM layers to combat overfitting. [4]

• Neural networks with Long Short-Term Memory (LSTM) architecture are especially well-suited for sequential data analysis tasks like text processing. In contrast to tradi- tional RNNs, LSTMs are effective at identifying long- term dependencies in the data, which overcomes the drawback of ordinary RNNs, which struggle to retain information over long sequences. [4]

• Dropout serves as a regularization technique to prevent overfitting in neural networks. During training, Dropout randomly drops neurons to strengthen the model. It is parameterized with a value between 0 and 1, representing the probability of dropping neurons. By preventing the model from being unduly dependent on particular neu- rons and characteristics, this method helps the model to generalize more effectively. [3]

C. Train the model

• Use the entire dataset to train the sentiment analysis model over ten epochs using a batch size of 32 and a validation split of 20%.

• The Python sentiment analysis model achieved an accu- racy of 52.62% on the training set and 55.15% on the test set.

• Let's plot these metrics using the matplotlib. [3]

3. RESULT

A. Data Processing

• Dataset: The Kaggle dataset was utilized, containing Twitter data.

• Data Cleaning: The data was cleaned, handling missing values and normalizing it.

• Data Categorization: The data was categorized into pos- itive, negative, and irrelevant tweets.

• Text Preprocessing: Text data was tokenized and con- verted into sequences for machine learning model com- prehension.

• Word Embeddings: Words were represented as vectors us- ing word embeddings to capture the relationship between them.

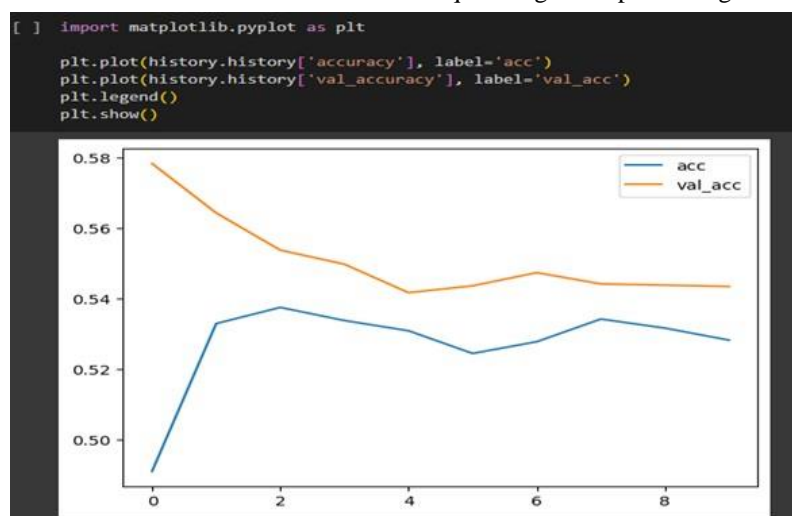• Equal Length Sentences: Sentences were standardized to equal length for processing efficiency.

Fig. 2. Grph I.

B. Building the text classifier

• Model Architecture: A LSTM-based neural network ar- chitecture was employed for sentiment analysis.

• LSTM Layers: LSTM layers were utilized for sequential data analysis, especially effective for text processing due to their ability to capture long-term dependencies.

• Dropout Regularization: Dropout layers were added to prevent overfitting by randomly dropping neurons during training.
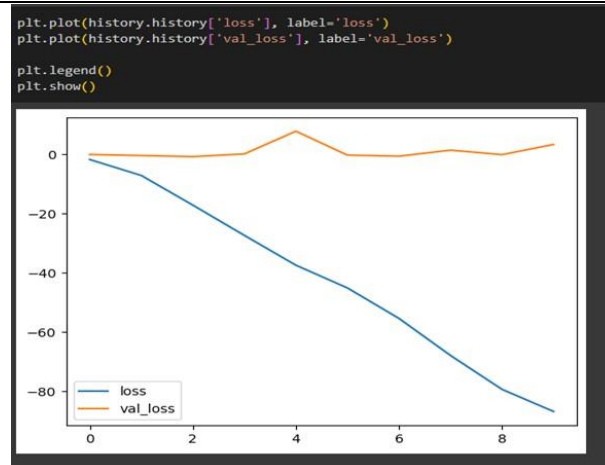
Fig. 3. Graph II

C. Trained model

- Training Process: The model was trained over ten epochs using the entire dataset with a batch size of 32 and a validation split of 20%.

- Performance Metrics: The sentiment analysis model achieved an accuracy of 52.62% on the training set and 55.15% on the test set.

- Visualization: The training and testing accuracies were plotted using matplotlib for visualization.

- Overall, the machine learning process involved preprocessing the data, building a text classifier using LSTM layers with dropout regularization, and training the model to achieve a reasonable accuracy for sentiment analysis on Twitter data.

- Testing

- Define a function that takes a tweet as input and gives output with its prediction label

- The model predicted the text as a positive or negative tweet and gave output in that format.

Fig. 4. Test Result

4. ADVANTAGES AND DISADVANTAGES

A. Advantages

Market Research: Analyzing sentiment in market research involves gauging public opinion, preferences, and trends related to products, services, or brands. It helps businesses understand consumer sentiment towards their offerings and competitors.

Customer Support: Sentiment analysis in customer support involves analyzing the sentiment of customer inquiries, feed- back, and complaints. It enables businesses to prioritize and effectively respond to client complaints, resulting in higher customer satisfaction and retention.

Brand Management: Sentiment analysis assists in monitor- ing brand sentiment across various channels to understand how customers perceive a brand. Positive sentiment suggests brand loyalty and satisfaction, but negative sentiment may indicate concerns that require solving.

Political Analysis: Sentiment analysis in a political analysis refers to the study of public sentiment toward political indi- viduals, policies, and events. It assists political organizations in comprehending voter opinion, identifying critical concerns, and tailoring communications efforts appropriately.

Healthcare: In healthcare, sentiment analysis entails assess- ing patient comments, reviews, or attitudes about healthcare professionals, treatments, or medical technologies. It helps healthcare organizations improve patient experience, identify areas for improvement, and address concerns.

Competitive Analysis: Sentiment analysis is a technique to evaluate attitudes regarding competing products, services, or brands. It aids businesses in assessing their standing against competitors and pinpointing opportunities for gaining a com- petitive edge.

B. Disadvantages

Inaccuracy in Tone Detection: Sentiment analysis may struggle to effectively discern the text's tone, such as sarcasm or irony, resulting in sentiment categorization errors.

Handling Slang and Informal Language: Sentiment analysis often faces difficulties in understanding and interpreting slang, informal language, or cultural nuances, which are prevalent in online communication.

Limited to Text-Based Data: Sentiment analysis is primarily designed for text-based data, limiting its applicability to other forms of data such as audio or video.

Difficulty in Handling Multilingual Content: Sentiment analysis may struggle to accurately analyze sentiment in multilingual content due to variations in language structure, sentiment expression, and cultural differences.

Dynamic and Evolving Language Use: Language is con- stantly evolving, with new words, phrases, and expressions emerging over time, posing challenges for sentiment analysis models to stay updated and relevant.

Emotional Complexity: Sentiment analysis faces challenges in capturing the complexity of human emotions, which can be nuanced and multifaceted, leading to oversimplification or misinterpretation of sentiment.

5. CONCLUSION

In conclusion, sentiment analysis of tweets represents a pivotal tool for decoding the collective emotions and opinions coursing through the vast expanse of social media. This real- time analytical approach offers invaluable insights into public sentiment, allowing businesses to gauge customer reactions, researchers to monitor social trends, and policymakers to stay attuned to the pulse of public opinion. Despite the undeniable benefits, challenges persist, with the nuanced nature of human expression posing difficulties in accurately deciphering sar- casm and subtle contextual cues within the brevity of tweets. Moreover, continued research is required to guarantee the equitable and appropriate usage of the technology they are using due to ethical issues and potential biases in sentiment analysis algorithms. The juxtaposition of machine learning with sentiment analysis portends a new era of flexibility and predictive power. As advancements in technology unfold, the synergy between these fields enhances the precision and scope of sentiment analysis. Sentiment analysis is a vital tool that can navigate the complex terrain of human emotions in the quick-changing world of social media, especially on sites like Twitter, thanks to the dynamic interaction between changing language patterns and advanced algorithms.

6. FUTURE SCOPE

Multimodal Sentiment Analysis: It extends sentiment analy- sis beyond text to include other modalities like images, videos, and audio, enabling a more comprehensive understanding of sentiment expressed across different media formats.

Cross-Lingual Sentiment Analysis: This aims to analyze sentiment in text written in multiple languages, requiring mod-els to understand and interpret sentiment expressions across different linguistic structures and cultural contexts.

Emotion Recognition: Emotion recognition involves iden- tifying and categorizing emotions expressed through various forms of communication, such as speech or writing, to under- stand individuals' emotional states.

Context-aware Sentiment Analysis: This considers the con- text in which sentiment is expressed, including factors such as the speaker's identity, relationship with the audience, and situational context, for more accurate sentiment analysis.

Industry-Specific Application: Sentiment analysis is tailored to specific industries (e.g., retail, healthcare, finance) to ad- dress unique challenges and requirements, such as analyzing customer feedback, monitoring brand sentiment, or predicting market trends.

Real-time and Dynamic Analysis: Real-time and dynamic sentiment analysis enables businesses to promptly respond to emerging trends, customer feedback, or events as they unfold.

ACKNOWLEDGMENT

7. REFERENCES

[1] Nisha Rathee, Nikita Joshi, Jaspreet Kaur "Sentiment Analysis us- ing Machine Learning Techniques on Python" 978-1-5386-2842- 3/18/$31.00 ©2018 IEEE .

[2] J. Neethu M S, Rajasree R "Sentiment Analysis in Twitter using Machine Learning Techniques" IEEE - 31661, 4th ICCCNT 2013 July 4 - 6, 2013, Tiruchengode, India

[3] Bhumika Gupta, Monika Negi, Kanika Vishwakarma, Goldi Rawat, Priyanka Badhani "Study of Twitter Sentiment Analysis using Ma- chine Learning Algorithms on Python" International Journal of Computer Applications (0975 – 8887) Volume 165 – No.9, May 2017

[4] Hanan Alasmari1 "Sentimental Visualization: Semantic Analysis of Online Product Reviews Using Python and Tableau" IEEE ON BIG DATA VO., XX, NO., X, DECEMBER 2020