**SENTIMENTAL ANALYSIS OF TWEETS USING**

**DEEP LEARNING ALGORITHMS**

A PROJECT REPORT

***Submitted by***

      BL.EN.U4CSE20028 Chadalavada Sai Vivek

      BL.EN.U4CSE 20086 Likesh.K

      BL.EN.U4CSE20151 Srujan Sai

BL.EN.U4CSE20196 Jayanth.Veeranki

**BACHELOR OF TECHNOLOGY**

IN

COMPUTER SCIENCE AND ENGINEERING



AMRITA SCHOOL OF COMPUTING, BENGALURU

 AMRITA VISHWA VIDYAPEETHAM

BENGALURU 560 035

APRIL 2024

**AMRITA VISHWA VIDYAPEETHAM**

**AMRITA SCHOOL OF COMPUTING, BENGALURU, 560035**



**BONAFIDE CERTIFICATE**

This is to certify that the project report entitled **“SENTIMENTAL ANALYSIS OF TWEETS USING DEEP LEARNING ALGORITHMS ”**submitted by

BL.EN.U4CSE20028 Chadalavada Sai Vivek

BL.EN.U4CSE 20086 Likesh.K

BL.EN.U4CSE20151 Srujan Sai

BL.EN.U4CSE20196 Jayanth.Veeranki

in partial fulfillment of the requirements as part of **Bachelor of Technology** in “**COMPUTER SCIENCE** **AND** **ENGINEERING”** is a bonafide record of the work carried out under my guidance and supervision at Amrita School of Computing, Bengaluru.

*signature*

*Guide name* Mr. Vishwas HN

*Designation* Assistant Professor (Sr. Gr.)

Dept. of CSE, School of Computing Dept. of CSE, School of Computing

This project report was evaluated by us on ………

*<signature> <signature> <signature>*

 Internal Examiner 1       Internal Examiner 2 External Examiner

**ACKNOWLEDGEMENTS**

The satisfaction that accompanies successful completion of any task would be incomplete without mention of people who made it possible, and whose constant encouragement and guidance have been source of inspiration throughout the course of this project work.

We offer our sincere pranams at the lotus feet of **“AMMA”,** **MATA AMRITANANDAMAYI DEVI** who showered her blessing upon us throughout the course of this project work.

We owe our gratitude to **Prof. Manoj P.**, Director, Amrita Vishwa Vidyapeetham Bengaluru Campus. We would like to place our heartfelt gratitude to **Dr.** **Gopalakrishnan E.A.,** Principal, Amrita School of Computing, Bengaluru for his valuable support and inspiration.

It is a great pleasure to express our gratitude and indebtedness to our project guide**Mr. Vishwas H N, Assistant Professor (Sr. Gr.),** Department of Computer Science and Engineering, Amrita School of Computing, Bengaluru for her/his valuable guidance, encouragement, moral support, and affection throughout the project work.

We would like to thank express our gratitude to project panel members for their suggestions, encouragement, and moral support during the process of project work and all faculty members for their academic support. Finally, we are forever grateful to our parents, who have loved, supported and encouraged us in all our endeavors.

**ABSTRACT**

Social media platforms are rich sources of user-generated material that are updated in real time like twitter and may offer important insights into public opinion. Taking advantage of the dynamic nature of user-generated information in real-time on social media sites like Twitter, this research uses a strict approach that includes data collection, sentiment categorization, feature extraction, and preprocessing. Sentiment analysis is guaranteed to have a representative sample when Twitter datasets are used. We combine classical machine learning models with state-of-the-art deep learning models, such as CNN, RNN, FNN, and Bidirectional Long Short-Term Memory (Bi-LSTM). We assess these models' performance using a comparison analysis to determine how accurate they are in yielding meaningful results. By offering a thorough evaluation of deep learning and machine learning models, this study advances the area of Twitter sentiment analysis by assisting in the discovery of more effective techniques for precise sentiment categorization.

**TABLE OF CONTENTS**

**Page No.**

ACKNOWLEDGEMENTS i

ABSTRACT ii

LIST OF FIGURES v or vi

LIST OF TABLES vii

CHAPTER 1- INTRODUCTION 1

CHAPTER 2 – LITERATURE SURVEY 2

CHAPTER 3 – SYSTEM REQUIREMENTS AND ANALYSIS 5

* 1. SOFTWARE REQUIREMENTS 5
  2. HARDWARE REQUIREMENTS

CHAPTER 4 – SYSTEM DESIGN 6

CHAPTER 5 – SYSTEM IMPLEMENTATION 10

* 1. Modules used with description

CHAPTER 6 – RESULTS AND ANALYSIS 13

CHAPTER 7 – CONCLUSION AND FUTURE ENHANCEMENT 16

REFERENCES 17

**LIST OF FIGURES**

Fig. 4.1 Basic Architecture 6

Fig 5.1 Collection of Data 14

Fig 5.2 Collection of Data In dataFrame 14

Fig 5.3 Html Tag removal Code 15

Fig 5.4 Url Removal Code 15

Fig 5.5 Replace emojis to text abbreviation. 16

Fig 5.6 Replacing all the Abbrevations 17

Fig. 6.1 Model’s Training accuracies 18

**CHAPTER - 1**

**INTRODUCTION**

With millions of active users, the microblogging site Twitter has become a bustling hub for real-time exchanges of ideas, attitudes, and opinions. Its vast data repository offers invaluable insights into public sentiment, making sentiment analysis a pivotal tool in discerning trends and understanding user perspectives. By delving into the emotional tone of tweets, sentiment analysis, a facet of natural language processing, enables us to glean nuanced insights from this rich tapestry of user-generated content.

In today's world, sentiment analysis on Twitter finds application across diverse sectors, from marketing and brand management to healthcare and politics. Companies leverage sentiment analysis to gauge consumer sentiment, tailor marketing strategies, and cultivate brand reputation. Similarly, healthcare professionals use sentiment patterns for public health monitoring and crisis management, while political analysts gauge public reaction to legislative policies and events.

Our research endeavor employs a diverse array of algorithms, each offering distinct advantages in analyzing Twitter sentiment. Recurrent Neural Networks (RNN) excel in capturing sequential dependencies and long-term dependencies in temporal data, making them suitable for analyzing the temporal dynamics of sentiment in Twitter data streams. Bidirectional Long Short-Term Memory (BiLSTM) networks, by incorporating information from both past and future states, enhance the model's ability to capture context and long-range dependencies, thereby improving sentiment analysis accuracy. Feedforward Neural Networks (FNN), with their simplicity and ease of implementation, provide a baseline model for sentiment analysis tasks, offering reliable performance with minimal computational overhead. Convolutional Neural Networks (CNN) excel in extracting hierarchical features from text data, enabling them to capture subtle nuances in sentiment expressed within tweets.

In addition to categorizing tweets, our study delves into the sentiment of each word within the tweet, providing granular insights into the underlying emotions expressed. Furthermore, we ascertain the overall sentiment of the tweet, offering a holistic perspective on user sentiment.

By leveraging the strengths of these algorithms, our research endeavors to refine the classification of positive and negative tweets, thereby furnishing a robust framework for extracting actionable insights from the vast expanse of Twitter data. Through this multifaceted approach, we aim to unravel the intricate fabric of public sentiment, empowering stakeholders with the knowledge to navigate and respond effectively to the dynamic discourse on Twitter.

**CHAPTER - 2**

**LITERATURE REVIEW**

One of the categorization methods utilized in the study is the Bidirectional Long Short-Term Memory (BiLSTM) algorithm. Recurrent neural networks (RNNs) like BiLSTM, which have a loop and a hidden layer, have the capacity to recall everything that has happened in the past. The RNN's loops enable the network to retain data. Additionally, it offers descriptive information like word clouds and the frequency of tweets that are favorable and negative[1]. It gives a specific Data Science Trajectory (DST) for the investigation. The following steps are part of the DST: Data Source Exploration, Data Acquisition, Data Value Exploration, Data Preparation, Data Result Exploration, Modeling, Evaluation, and Production Exploration.[2]. The approach of analyzing consumer sentiment in online reviews is mentioned. In particular, it employs pre-trained language models like Electra, XLNet, and BERT to classify and comprehend feelings in customer reviews. The pre-processing techniques, such as text cleaning and normalization, tokenization, and vectorization were used.[3]. The study describes how to analyze tweets that mention the Sustainable Development Goals (SDGs) in terms of sentiment using three Natural Language Processing (NLP) approaches. Vader (Valence Aware Dictionary and Sentiment Reasoner), Text Blob, and BERT (Bidirectional Encoder Representations from Transformers) are the approaches.[4]. The study also highlights the challenges faced by existing sentiment analysis models, such as being trained on structured datasets and open source, leading to outdated tools that do not provide valuable information.[5]. The development of a supervised model for live Twitter sentiment analysis using the TextBlob library in Python and the Streamlit framework. The paper presents a detailed description of the emotional analysis cycle, categorizing Twitter's unstructured information into positive or negative sentiments using various methods such as Twitter knowledge-based strategies and machine learning strategies . [6] By experimenting with different hyperparameters in the deep learning model, the study reports training accuracy of 86.33%. The study provides a thorough process for predicting election outcomes through sentiment analysis of tweets about the 2019 Indian Lok-Sabha election. The authors gathered tweets from multiple platforms, such as GitHub and Twitter, and preprocessed the data by converting the text to lowercase and eliminating stopwords, hashtags, and punctuations. [7] Five distinct machine learning models were trained using the labeled dataset. These models were each integrated with two feature extraction methods, TF-IDF and Bag-of-words. They discovered that the combination of Decision Tree and tf-idf had the highest accuracy with 86.3% in forecasting election results. The study also addresses the drawbacks of utilizing Twitter as a source for election outcome prediction. This study uses sentiment analysis of twitter to compare malaysian private hospitals and fulfil the need for input. Administrative process, cost, communication, knowledge, and service are its five main focal points. [8] The system classifies and displays sentiment in bilingual Twitter reviews using text mining and the Nave Bayes machine learning algorithm. Users can contrast private medical facilities in various states. Usability testing had an average score of 95.42% while functional testing had accuracy scores of 77.13% for English and 77.96% for Bahasa Melayu. The COVID-19 epidemic caused huge changes in people's daily lives, leading to a big shift towards social media as the main form of communication. In a thorough investigation, 26 million tweets were examined to understand how the epidemic affected people' actions and viewpoints. The analysis showed that users' concerns about healthcare services and the pandemic's financial effects were becoming more vocal. It is interesting to note that the general tone of COVID-19-related talks showed signs of improvement, demonstrating users' adaptability and resilience in the face of the current global health crisis. [9] The paper discusses the methodology used to analyze customer sentiments in reviews. Specifically, it uses models such as XLNet, Electra and BERT to categorize and understand customer reviews using sentiments. The authors also discuss the pre-processing methods used to prepare the sentiment analysis data, such as text cleaning and normalization, tokenization, and vectorization. The paper shows highest accuracy with fine tuning with 92.3%.[10] The use of three NLP techniques for analysis of sentiments of tweets pertaining to Sustainable Development Goals (SDGs) is described in the paper. The techniques are VADER, Text Blob, BERT (Bidirectional Encoder Representations from Transformers). The paper discusses the limited focus on sentiment analysis towards the Sustainable[11] Development Goals (SDGs). While sentiment analysis has been applied to various topics, including climate change, there has been limited research on sentiment analysis towards the SDGs. The paper focuses on using Twitter sentiment analysis to analyze hotel data. The authors provide a framework for extracting and analyzing data from Twitter that mixes supervised and unsupervised methods. The proposed approach uses machine learning techniques for sentiment analysis, Twitter API keys for data collecting, and content removal for undesired content. The authors had used SVM model for classification of positive and negative tweets for the hotels.[12] The authors create a new model to get the best results for their movie review research. In order to comprehend emotional analysis, the writers emphasize the importance of word occurrence and usage in both positive and negative settings. They place a strong emphasis on using a sizable dataset—five hundred thousand movie reviews—to enhance forecast performance and overcome the shortcomings of earlier studies.[13] The Soft Voting Ensemble (SVE) approach, which combines the output of separate classifiers to improve accuracy, is presented in this study. The suggested ensemble method performed better than any other classifier, with the following results: 89.9% overall accuracy, 90.0% overall accuracy, 90.0% recall, and 90.0% F1-score. The significance of machine learning models in managing unstructured and extensive text data is also covered by the writers, especially when it comes to emotional analysis. [14] Instead of depending just on human evaluations, the project attempts to automate text sentiment analysis in order to efficiently retrieve viewpoints and emotional opinions about particular issues. The study applies machine learning methodologies and deep neural network (RNN-LSTM), the two primary methods of sentiment analysis, to three Twitter datasets: IMDB, Amazon, and Airline. The authors compare the deep learning models and machine learning models and find that RNN-LSTM are performing better.

**CHAPTER – 3**

**SYSTEM SPECIFICATIONS**

**3.1 Software requirements**

* CPU / GPU machine for computation: A powerful CPU or GPU is essential for fast processing, especially with large datasets and complex machine learning models.
* Python environment: Python is a versatile language with extensive libraries for data analysis, machine learning, and web development, making it ideal for this project.
* streamlit: Streamlit is a user-friendly library for creating interactive web applications, allowing easy deployment of your sentiment analysis tool.
* TensorFlow: TensorFlow is a popular deep learning framework that provides tools for building and training neural networks, suitable for sentiment analysis tasks.
* scikit-learn: scikit-learn offers a wide range of tools for data preprocessing, machine learning model building, and evaluation, essential for sentiment analysis.
* Pandas: Pandas is a powerful library for data manipulation and analysis, which is crucial for preparing and processing Twitter data.
* Tkinter: Tkinter is a standard GUI toolkit for Python, allowing you to create interactive graphical interfaces for your sentiment analysis application.
  1. Hardware Requirements
* CPU / GPU: Depending on the size of the dataset and complexity of the analysis, a decent CPU or GPU is recommended for faster processing.
* Memory (RAM): Sufficient RAM is essential for handling large datasets and running machine learning models efficiently.
* Storage: Adequate storage space is needed for storing datasets and model outputs.
* Network Connectivity: A stable internet connection is required for accessing the Twitter API and fetching tweets.

**CHAPTER - 4**

**SYSTEM DESIGN**

A diagram of a software company

Description automatically generated

**Fig.1: Low level Design with description**

**A. Bidirectional LSTM (Bi-LSTM):**

Description: Bidirectional LSTM (Long Short-Term Memory) networks are a type of recurrent neural network (RNN) architecture designed to better capture long-term dependencies in sequential data. Unlike traditional RNNs, LSTMs are well-suited for tasks involving sequences, such as sentiment analysis on Twitter data.

How it helps in Sentiment Analysis:

Contextual Understanding: Twitter data often consists of brief, fragmented text sequences with complex emotions and nuanced sentiments. Bi-LSTM's bidirectional processing allows it to capture both past and future context, enabling it to understand the sentiment of a tweet within its broader context.

Long-Term Dependency: By retaining both the current input and previous sequences, Bi-LSTM can capture long-term dependencies in the data, crucial for understanding changing sentiment patterns over time.

Memory Factor: The internal memory of the Bi-LSTM helps in remembering past sequences along with current input, enabling it to capture context rather than just individual words.

Benefits for Project:

Improved Context Understanding: By understanding the context of each tweet, Bi-LSTM can more accurately classify tweets into positive, negative, or neutral sentiments.

Effective Handling of Time Dependencies: Bi-LSTM is well-suited for capturing time dependencies in Twitter data, making it adept at identifying changing sentiment patterns over time.

Better Handling of Fragmented Data: Since tweets are often brief and fragmented, Bi-LSTM's ability to process input in both directions helps in capturing the overall sentiment across multiple tweets.

**B. Feed Forward Neural Network (FNN) Model:**

Description: Feed Forward Neural Networks (FNNs) are a class of artificial neural networks where connections between nodes do not form cycles. They are the simplest form of neural networks, where data moves in only one direction—from the input nodes, through the hidden nodes (if any), to the output nodes.

How it helps in Sentiment Analysis:

Supervised Classification: FNNs are used as a supervised classification method for sentiment analysis. They learn to classify tweets into positive, negative, or neutral sentiments based on the features extracted from the data.

Polarity Determination: The primary goal of the FNN model in sentiment analysis on Twitter is to accurately determine the sentiment polarity of the content.

Benefits for Your Project:

Precise Sentiment Polarity Determination: FNNs are effective in precisely determining the sentiment polarity of tweets, helping in better understanding the sentiment conveyed in each tweet.

Supervised Learning: FNNs can be trained using supervised learning techniques, making them suitable for sentiment analysis tasks where labeled data is available.

Scalability and Efficiency: FNNs are computationally efficient and scalable, making them suitable for processing large volumes of Twitter data efficiently.

**C. Convolutional Neural Network (CNN):**

Description: Convolutional Neural Networks (CNNs) are a type of deep neural network primarily used to analyze visual imagery. However, they have been successfully applied to various natural language processing (NLP) tasks, including sentiment analysis, by treating text data as images.

How it helps in Sentiment Analysis:

Contextual Details Extraction: CNNs are used in sentiment analysis on Twitter to extract contextual details, local patterns, and hierarchical characteristics from tweets.

Word Embeddings Utilization: CNNs can utilize word embeddings to represent words in tweets as dense vectors, capturing semantic relationships between words.

Handling Variable-Length Input: CNNs are capable of handling variable-length input, making them suitable for sentiment prediction on dynamic and varied Twitter data.

Benefits for Your Project:

Effective Context Extraction: CNNs are adept at extracting contextual details and local patterns from tweets, helping in better understanding the sentiment expressed in each tweet.

Handling Variable-Length Input: Since tweets can vary in length, CNNs' ability to handle variable-length input makes them suitable for sentiment analysis on Twitter data.

Parallel Processing: CNNs can process multiple tweets simultaneously, making them computationally efficient for analyzing large volumes of Twitter data.

**D. Gated Recurrent Networks (GRU):**

Description: Gated Recurrent Units (GRUs) are a type of recurrent neural network (RNN) architecture similar to LSTM but with fewer parameters. GRUs are designed to capture long-term dependencies in sequential data while addressing the vanishing gradient problem.

How it helps in Sentiment Analysis:

Long-Term Dependency Capture: The GRU model is capable of capturing long-term dependencies in the input sequence, essential for understanding the context of a tweet and accurately classifying its sentiment.

Vanishing Gradient Issue Addressing: GRUs are designed to address the vanishing gradient issue that arises with conventional RNNs, making them more effective for tasks involving long sequences.

Benefits for Your Project:

Accurate Sentiment Classification: GRUs can accurately classify the sentiment of tweets by capturing long-term dependencies in the data, helping in better understanding the context of each tweet.

Addressing Vanishing Gradient Issue: Since tweets can be lengthy, GRUs' ability to address the vanishing gradient issue makes them more effective for sentiment analysis tasks involving long sequences.

Efficient Training: GRUs are computationally efficient to train, making them suitable for processing large volumes of Twitter data.

**E. Recurrent Neural Network (RNN):**

Description: Recurrent Neural Networks (RNNs) are a class of artificial neural networks where connections between nodes form directed cycles. RNNs are particularly effective for tasks involving sequences, such as time series analysis and natural language processing (NLP).

How it helps in Sentiment Analysis:

Sequence Information Capture: RNNs are particularly useful when capturing sequences of information, such as in time series analysis or next-word prediction.

Context Capture: RNNs' internal memory allows them to remember past sequences alongside current input, enabling them to capture context rather than just individual words.

Benefits for Your Project:

Contextual Understanding: RNNs are effective at capturing the context of each tweet, enabling them to accurately classify tweets into positive, negative, or neutral sentiments.

Sequence Information Capture: Since tweets are sequential data, RNNs are well-suited for sentiment analysis tasks involving sequences of text.

Efficient Training: RNNs are computationally efficient and scalable, making them suitable for processing large volumes of Twitter data efficiently.

**CHAPTER – 5**

**SYSTEM IMPLEMENTATION**

**5.1 Modules used with description:**

**A.Collection of Data:**

It is a crucial part where a well selected dataset is necessary for training. By putting together a varied and representative collection of text samples, the Deep learning model will be more equipped to identify patterns and correlations in the textual data, which will improve its capacity for generalization and precise prediction-making.

**A screenshot of a computer

Description automatically generated**

Fig5.1: Collection of Data

A screenshot of a computer

Description automatically generated

Fig5.2: Collection of Data In dataFrame

**B.Preprocessing:**

In Twitter sentiment analysis, preprocessing is crucial for text cleaning and categorization. It also helps models identify sentiment more precisely by we had removed stopwords so that it will eliminate noise from the tweets, removing null values ,standardizing formats, and improving comprehension of user expressions. We had also removed punctuation from the data.

**C.Html Tags Removal:**

We had imported an library named re(regular expression) to remove the html tags and urls by creating function named “remove\_html\_tags” and it takes a regular expression(<.\*?>) and uses it to extract HTML tags from the input text. It is important to remove them from the tweets in order to extract meaningful content from raw text because HTML tags can introduce noise and not contribute to sentiment analysis.

**D.URLS Removal:**

We had use the re library for removing urls also in the tweets it takes the text and uses a regular expression(https?://\S+|www\.\S+) to remove URLs. Though they might not include sentiment-bearing information, URLs are often found in tweets. Removing them enhances the accuracy of sentiment analysis by concentrating the study on the tweets actual textual content.

A screenshot of a computer

Description automatically generated

Fig5.3: Html Tag removal Code

A screen shot of a computer

Description automatically generated

Fig5.4:Url Removal Code

**E. Replacement of Emojis:**

Emojis are converted into meaningful textual representations in this preprocessing stage. We developed a custom method to swap out emojis for their appropriate textual representations in our sentiment analysis text. Emojis are visual symbols that are used to express feelings and emotions. The language of emojis links certain emojis to the sentiments they represent. In this way, ":)" and ":-)" become "smile," ":(" and ":-(" become "sad," and so on. It helps the model to properly collect and analyse the emotional context through the translation of emojis into English.

A screen shot of a computer screen

Description automatically generated

Fig5.5:Replace emojis to text abbreviation.

**F. Replacement of Abbrevations:**

We included a file called "slang" in our sentiment analysis study, which is a mapping of acronyms and short phrases to their long-form equivalents. As an example, "As Far As I Know" is substituted for "AFAIK," "Be Right Back" for "BRB,",etc.This preprocessing stage helps to enhance the comprehension and interpretation of textual data, particularly when working with informal language or material from social media that contains a variety of acronyms.

A screenshot of a computer program

Description automatically generated

Fig5.6:Replacing all the Abbrevations

**G. Tokenization:**

Tokenization is a important step in NLP that divides a raw text into separate parts called tokens. Words, sentences, or other significant language units can serve as these tokens. Tokenization is an essential step in preparing textual input for different NLP tasks A screenshot of a computer program

Description automatically generated

Fig5.7:Tokenization Code.

**H. Stemming:**

Stemming is a linguistic normalization technique used in NLP to reduce words to their simple form. This process reduces the dimensionality of the data and improves text analysis by clustering similar words together.

We had also used Lemmatization for better outcome from the model.

**I.TF-IDF Vectorizer:**

We had used TF-IDF features using the TF-IDF Vectorizer from scikit-learn, with the ability to modify settings like the maximum features, n-gram range, and elimination of English stop words. Textual data is converted into numerical vectors using TF-IDF vectorization, where each vector represents a document. Word significance is taken into account by TF-IDF not just in relation to their frequency inside a document, but also within the context of the entire dataset.

**J. Recurrent Neural Network (RNN):**

Recurrent Neural Networks (RNNs) are a class of artificial neural networks where connections between nodes form directed cycles. RNNs are particularly effective for tasks involving sequences, such as time series analysis and natural language processing (NLP).

**K. Gated Recurrent Networks (GRU):**

Gated Recurrent Units (GRUs) are a type of recurrent neural network (RNN) architecture similar to LSTM but with fewer parameters. GRUs are designed to capture long-term dependencies in sequential data while addressing the vanishing gradient problem.

**L. Convolutional Neural Network (CNN):**

Convolutional Neural Networks (CNNs) are a type of deep neural network primarily used to analyze visual imagery. However, they have been successfully applied to various natural language processing (NLP) tasks, including sentiment analysis, by treating text data as images.

**M. *Feed Forward Neural Network (FNN) Model:***

Feed Forward Neural Networks (FNNs) are a class of artificial neural networks where connections between nodes do not form cycles. They are the simplest form of neural networks, where data moves in only one direction—from the input nodes, through the hidden nodes (if any), to the output nodes.

**N. *Bidirectional LSTM (Bi-LSTM):***

Bidirectional LSTM (Long Short-Term Memory) networks are a type of recurrent neural network (RNN) architecture designed to better capture long-term dependencies in sequential data. Unlike traditional RNNs, LSTMs are well-suited for tasks involving sequences, such as sentiment analysis on Twitter data.

**CHAPTER – 6**

**RESULTS AND ANALYSIS**

A graph of blue rectangular bars

Description automatically generated

Fig. 3. Comparison of models

Fig3.show the accuracy of the trained models in order from the plot we observe the following models and their corresponding accuracies:

CNN (Convolutional Neural Network):

Accuracy: Just below 80%

Interpretation: CNN performs reasonably well, but it doesn’t surpass the 80% accuracy mark. It’s a solid choice for sentiment analysis on Twitter data.

*RNN (Recurrent Neural Network):*

Accuracy: Around 60%

Interpretation: RNN lags behind other models significantly. Its performance is suboptimal for this specific task.

*GRU (Gated Recurrent Unit):*

Accuracy: Slightly higher than CNN, nearing 80%

Interpretation: GRU shows similar performance to CNN. While it’s a viable option, it doesn’t offer a substantial improvement.

Bidirectional LSTM (Long Short-Term Memory):

Accuracy: Just above 80%

Interpretation: Bidirectional LSTM outperforms other models with a slight margin. It’s the top performer in this comparison.

*FNN (Feedforward Neural Network):*

Accuracy: Comparable to CNN and GRU

Interpretation: FNN’s accuracy aligns with CNN and GRU, but it doesn’t provide any significant advantage over them.

From the analysis, we can conclude that, The Bidirectional Long Short-Term Memory (Bi-LSTM) architecture proved to be the most accurate model among the examined configurations once it was trained. This research highlights the effectiveness of Bi-LSTM as a reliable option for tasks like this on social media datasets by demonstrating its ability to capture complex patterns within the Twitter sentiment data. For practitioners looking for the top model settings for Twitter sentiment analysis applications, the finding provide insightful information.

1. EVALUATION
2. Example Tweet for the testing:

“ Sorry, Jonas! You know we wanna make cool skins but pay-to-win isn't acceptable or intentional. 🫤 😭We didn't catch this, but appreciate the community helped us find this. We’re fixing this bug now and will push out a fix soon. 🙏”

-- <https://twitter.com/Preeti_Riot/status/1785783792197906678>

A screenshot of a computer

Description automatically generated

Fig4: Result of sentiment analysis

Based Fig4 on the sentiment analysis results for the provided tweet, here are the predictions from each model:

LSTM Model: Sentiment is Good with 50.24% confidence.

Bi-LSTM Model: Sentiment is Bad with 55.32% confidence.

GRU Model: Sentiment is Bad with 49.79% confidence.

FNN Model: Sentiment is Good with 50.24% confidence.

CNN Model: Sentiment is Good with 69.04% confidence.

RNN Model: Sentiment is Good with 55.99% confidence.

The CNN model predicts the sentiment as Good with the highest confidence of 69.04%. The RNN model also predicts the sentiment as Good, but with a slightly lower confidence of 55.99%. The LSTM and FNN models predict the sentiment as Good, both with a confidence of 50.24%. The GRU model predicts the sentiment as Bad with a confidence of 49.79%. The Bi-LSTM model predicts the sentiment as Bad with the lowest confidence of 8.32%.

**A screenshot of a computer

Description automatically generated**

Fig5: sentiment analysis of every word in the tweet

Based Fig5 on the sentiment analysis results for word Sorry from the provided tweet, here are the predictions from each model:

LSTM Model: Sentiment is Good with 50% confidence.

Bi-LSTM Model: Sentiment is Bad with 55.32% confidence.

GRU Model: Sentiment is Bad with 49% confidence.

FNN Model: Sentiment is Good with 50% confidence.

CNN Model: Sentiment is Good with 69.% confidence.

RNN Model: Sentiment is Good with 56% confidence.

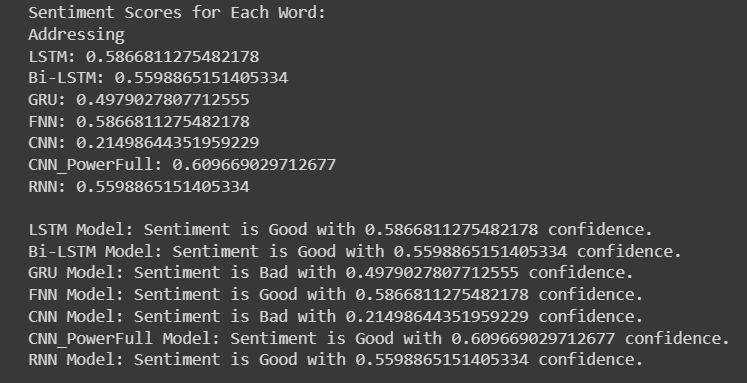


Fig6. For the word Addressing sentiment analysis

Based Fig6 on the sentiment analysis results for the word addressing, here are the predictions from each model:

LSTM Model: Sentiment is Good with 58.66% confidence.

Bi-LSTM Model: Sentiment is Bad with 55.98% confidence.

GRU Model: Sentiment is Bad with 49.79% confidence.

FNN Model: Sentiment is Good with 58.66% confidence.

CNN Model: Sentiment is Good with 21.04% confidence.

CNN Powerful Model Version: Sentiment is Good with 69.99% confidence

RNN Model: Sentiment is Good with 55.99% confidence.

The CNN model predicts the sentiment as Good with the highest confidence of 69.04%. The RNN model also predicts the sentiment as Good, but with a slightly lower confidence of 55.99%. The LSTM and FNN models predict the sentiment as Good, both with a confidence of 50.24%. The GRU model predicts the sentiment as Bad with a confidence of 49.79%. The Bi-LSTM model predicts the sentiment as Bad with the lowest confidence of 8.32%.

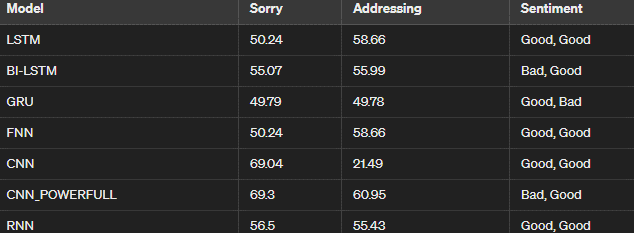


Table1: with confidence rate of words Sorry and addressing with there predicted sentiment

From table1: Here's a comparative analysis for the words "Sorry" and "Addressing" based on the sentiment scores provided across different models:

The sentiment analysis for the words "Sorry" and "Addressing" reveals intriguing insights into their perceived emotional connotations. Across various deep learning models, "Sorry" tends to evoke sentiments closer to neutrality or mild positivity, with scores hovering around the mid-range. The LSTM and FNN models consistently assign similar sentiment scores to both words, indicating a tendency towards neutrality or slight positivity. Conversely, the BI-LSTM model assigns a higher sentiment score to "Sorry" compared to "Addressing," suggesting a nuanced perception of these words' emotional impact. Interestingly, the sentiment analysis varies notably across models for the word "Addressing."

While LSTM, BI-LSTM, and RNN models attribute similar sentiment scores to both words, GRU and CNN models portray a divergence in sentiment, with "Addressing" eliciting a lower sentiment score compared to "Sorry."

This disparity underscores the contextual nuances captured by different model architectures, reflecting the complexity of language interpretation in sentiment analysis tasks. Furthermore, the CNN\_POWERFULL model presents contrasting sentiment assessments for the two words, with "Sorry" being perceived more positively than "Addressing." This discrepancy highlights the model's sensitivity to subtle linguistic cues and its ability to discern varying emotional nuances within textual data.

In summary, while some models exhibit consistency in sentiment assignments for "Sorry" and "Addressing," others demonstrate divergent interpretations, emphasizing the intricate interplay between linguistic context and computational modeling techniques in sentiment analysis.

**CHAPTER – 8**

**CONCLUSION AND FUTURE SCOPE**

This work extends the area of sentiment analysis by offering insightful information on the effectiveness of various machine learning and Deep Learning techniques for determining the sentiment present in Twitter data. The use of preprocessing methods, including lemmatization, stemming, and text cleaning, was critical in improving the models sentiment recognition performance. Additionally, the use of TF-IDF vectorization enhanced feature representation even more, making the model training easier to comprehend the textual data it was based on. The findings show that deep learning models is performing better than machine learning models, especially Bidirectional LSTM. Preprocessing methods like feature engineering and text cleaning also made a big difference in the model's performance. The study emphasizes how crucial it is to use the right preprocessing techniques and algorithms for efficient sentiment analysis of social media data. In order to improve predictions even further, future research may investigate ensemble techniques and sophisticated neural networks.

**REFERENCES**

Follow IEEE format and if websites are there put it at the end and also books.

1. N. O. Aljehane, "A New Approach to Sentiment Analysis on Twitter Data with LSTM," 2023 3rd International Conference on Computing and Information Technology (ICCIT), Tabuk, Saudi Arabia, 2023, pp. 657-663, doi: 10.1109/ICCIT58132.2023.10273876.
2. C. Susmitha, L. Nikhil, L. Akhil, M. Kavitha, V. S. N. Reddy and K. Shailaja, "Sentimental Analysis on Twitter Data using Supervised Algorithms," 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2023, pp. 921-925, doi: 10.1109/ICCMC56507.2023.10084278.
3. G. Prema Arokia Mary, M. S. Hema, R. Maheshprabhu and M. Nageswara Guptha, "Sentimental Analysis of Twitter Data using Machine Learning Algorithms," 2021 International Conference on Forensics, Analytics, Big Data, Security (FABS), Bengaluru, India, 2021, pp. 1-5, doi: 10.1109/FABS52071.2021.9702681.Republic of, 2022, pp. 1-3, doi: 10.1109/ICEIC54506.2022.9748487.
4. S. Raheja and A. Asthana, "Sentimental Analysis of Twitter Comments on Covid-19," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2021, pp. 704-708, doi: 10.1109/Confluence51648.2021.9377048.
5. K. S. Madhu, B. C. Reddy, C. Damarukanadhan, M. Polireddy and N. Ravinder, "Real Time Sentimental Analysis on Twitter," 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2021, pp. 1030-1034, doi: 10.1109/ICICT50816.2021.9358772.
6. B. M, S. S, R. M, S. K. R and S. R, "A detailed study on sentimental analysis using Twitter data with an Improved deep learning model," 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 2021, pp. 408-413, doi: 10.1109/I-SMAC52330.2021.9640850.
7. P. Khurana Batra, A. Saxena, Shruti and C. Goel, "Election Result Prediction Using Twitter Sentiments Analysis," 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC), Waknaghat, India, 2020, pp. 182-185, doi: 10.1109/PDGC50313.2020.9315789.Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2021.
8. Abu Samah, Khyrina Airin Fariza & Azharludin, Nur & Riza, Lala & Jono, Dato Dr Mohd & Moketar, nor aiza. (2023). Classification and visualization: Twitter sentiment analysis of Malaysia’s private hospitals. IAES International Journal of Artificial Intelligence (IJ-AI). 12. 1793. 10.11591/ijai.v12.i4.pp1793-1802.
9. Alshamrani, S., Abusnaina, A., Abuhamad, M., Lee, A., Nyang, D., Mohaisen, D. (2020). An Analysis of Users Engagement on Twitter During the COVID-19 Pandemic: Topical Trends and Sentiments. In: Chellappan, S., Choo, KK.R., Phan, N. (eds) Computational Data and Social Networks. CSoNet 2020. Lecture Notes in Computer Science(), vol 12575. Springer, Cham. <https://doi.org/10.1007/978-3-030-66046-8_7>
10. S. I. . Khan, S. V. . Athawale, M. P. . Borawake, and M. Y. . Naniwadekar, “Sentiment Analysis of Customer Reviews using Pre-trained Language Models”, Int J Intell Syst Appl Eng, vol. 11, no. 7s, pp. 614–620, Jul. 2023.
11. Rosenberg, Emelie & Tarazona, Carlota & Mallor, Fermin & Eivazi, Hamidreza & Pastor-Escuredo, David & Nerini, Francesco & Vinuesa, Ricardo. (2023). Sentiment analysis on Twitter data towards climate action. 10.21203/rs.3.rs-2434092/v1.
12. A. Ritter, S. Clark, Mausam, O. Etzioni, Named entity recognition in tweets: an experimental study, Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP'11, Association for Computational Linguistics, Stroudsburg,PA, USA, 2011, pp. 1524–1534.
13. B.J. Jansen, M. Zhang, K. Sobel, A. Chowdury, Twitter power: Tweets as electronic word of mouth, Journal of the American Society for Information Science and Technology 60 (11) (2009) 2169–2188.
14. J.-M. Xu, K.-S. Jun, X. Zhu, A. Bellmore, Learning from bullying traces in social media,HLT-NAACL, , The Association for Computational Linguistics, 2012. 656–666.
15. M. Cheong, V.C. Lee, A microblogging-based approach to terrorism informatics: exploration and chronicling civilian sentiment and response to terrorism events via twitter, Information Systems Frontiers 13 (1) (2011) 45–59.
16. P.H.C. Guerra, A. Veloso, W. Meira, V. Almeida Jr, From bias to opinion: a transferlearning approach to real-time sentiment analysis, Proceedings of the 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, San Diego, CA, 2011.
17. N.A. Diakopoulos, D.A. Shamma, Characterizing debate performance via aggregated twitter sentiment, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI'10, ACM, New York, NY, USA, 2010, pp. 1195–1198.
18. P.D. Turney, Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews, Proceedings of the 40th Annual Meeting onAssociation for Computational Linguistics, ACL 02, Association for Computational Linguistics, Stroudsburg, PA, USA, 2002, pp. 417–424.
19. B. Pang, L. Lee, S. Vaithyanathan, Thumbs up? Sentiment classification using machine learning techniques, Proceedings of EMNLP, 2002, pp. 79–86.
20. M. Hu, B. Liu, Mining and summarizing customer reviews, Proceedings of the TenthACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'04, ACM, New York, NY, USA, 2004, pp. 168–177.
21. B. He, C. Macdonald, J. He, I. Ounis, An effective statistical approach to blog

post opinion retrieval, Proceedings of the 17th ACM Conference on Information and Knowledge Management, CIKM'08, ACM, New York, NY, USA, 2008,pp. 1063–1072.

1. P. Melville, W. Gryc, R.D. Lawrence, Sentiment analysis of blogs by combining lexical knowledge with text classification, Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'09, ACM, New York, NY, USA, 2009. 1275–1284.
2. A. Balahur, R. Steinberger, M. Kabadjov, V. Zavarella, E. van der Goot, M. Halkia, B. Pouliquen, J. Belyaeva, Sentiment analysis in the news, in: N. C. C. Chair), K.Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, D. Tapias
3. (Eds.), Proceedings of the Seventh International Conference on Language Reoources and Evaluation (LREC'10), European Language Resources Association(ELRA), Valletta, Malta, 2010.
4. P. Ekman, Emotion in the Human Face, vol. 2Cambridge University Press, 1982.
5. B. Liu, Sentiment analysis and opinion mining, Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publishers, 2012.
6. B. Liu, Web data mining: exploring hyperlinks, contents, and usage data, DataCentric Systems and Applications, Springer-Verlag New York, Inc., Secaucus, NJ,USA, 2006.
7. B. Liu, Sentiment Analysis and Subjectivity, Taylor and Francis Group, Boca, 2010.
8. A. Agarwal, P. Bhattacharyya, Sentiment analysis: a new approach for effective use of linguistic knowledge and exploiting similarities in a set of documents to be classified, Proceedings of the International Conference on Natural Language Processing (ICON), 2005