POLARITY DETECTION OF TOP IPL PLAYERS TWEETS USING BERT AND SVM BASED ON TWEETS

BY

SHEIKH NAZMUS SHAKIB HEMEL
ID: 201-15-14286

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

SHAHADAT HOSSAIN
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
July 2024

APPROVAL

This Project titled "**POLARITY DETECTION OF TOP IPL PLAYERS TWEETS USING BERT AND SVM BASED ON TWEETS**", submitted by **Sheikh Nazmus Shakib Hemel** of the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **14/7/2024**

BOARD OF EXAMINERS

____
Dr. Md. Fokhray Hossain — **Board Chairman**
Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

____
Amatul Bushra Akhi — **Internal Examiner 1**
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

____
Mr. Amir Sohel — **Internal Examiner 2**
Senior Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

________________________ 14/7/24
Dr. Shamim H Ripon — **External Examiner**
Professor
Department of Computer Science and Engineering
East West University

DECLARATION

I hereby declare that this project has been done by us under the supervision of **Shahadat Hossain, Lecturer (Senior Scale), Department of CSE**, Daffodil International University. I also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Shahadat Hossain
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

Submitted by:

Sheikh Nazmus Sakib Hemel
ID: 201-15-14286
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, I express my heartiest thanks and gratefulness to Almighty ALLAH for His divine blessing in making it possible to complete the final year project/internship successfully.

I am really grateful and wish my profound indebtedness to **Shahadat Hossain, Lecturer, (Senior Scale)Department of CSE** Daffodil International University, Dhaka. Deep Knowledge & keen interest of my supervisor in the field of Machine Learning. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, and reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

I would like to express our heartiest gratitude to the **Head of Department of CSE**, for his kind help to finish this project and also to other faculty members and the staff of the CSE department of Daffodil International University.

I would like to thank our entire course mate at Daffodil International University, who took part in this discussion while completing the course work.

Finally, I must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Sentiment analysis represents a fundamental component of natural language processing, playing a pivotal role in understanding user sentiments and public opinion across diverse domains. This research undertakes a comparative analysis of various machine learning methodologies for sentiment analysis, focusing specifically on polarity. The models evaluated include SVM with both Linear and Non-Linear (RBF) kernels, as well as BERT and ROBERTA. The primary metric employed to gauge the performance of each model is accuracy. The findings reveal that SVM with a Linear kernel achieved an accuracy of 84%, while the Non-Linear SVM (RBF) recorded an accuracy of 78%. BERT, when not fine-tuned, demonstrated an accuracy of 89%, which improved to 90% following fine-tuning. ROBERTA also achieved an accuracy of 90%. The insights derived from this study offer valuable guidance regarding the most effective machine learning models for sentiment analysis tasks. Researchers and practitioners can leverage these findings to select appropriate sentiment analysis models tailored to their specific application requirements. The research highlights the relative ease of enhancing accuracy through fine-tuning and the application of various pre-processing techniques. Furthermore, it underscores the importance of considering the wide range and scope of machine learning approaches to enhance and increase the accuracy and reliability of opinion mining and sentiment analysis systems in real-world applications.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

In our current era of rapidly expanding information, it is crucial to be able to analyze and interpret emotions from written data in order to understand societal patterns, customer habits, and popular viewpoints. Sentiment analysis is a branch of natural language processing (NLP) that detects and evaluates emotions in text through a range of computational methods. The goal of this research study is to examine and compare the effectiveness of various machine-learning methods for analyzing sentiment. Three models - SVM with Linear Kernel, SVM with Non-Linear Kernel (RBF), and BERT were both implemented and evaluated. Evaluating the overall performance of a model is based on the accuracy of its sentiment classification. The research focuses on implementing different models in sentiment analysis tasks, assessing their ability to recognize neutral, positive, or negative sentiments in text data. The models' accuracies are shown and compared to provide important insights into their strengths and weaknesses in tackling sentiment analysis tasks.

The outcomes we obtained demonstrate how machine learning algorithms can be used to capture and evaluate subtle nuances in assumptions, thus promoting the field of assumption analysis. Furthermore, the research explores the outcomes of the recognized accuracy, considering the practicality and possible applications of each model in real-life scenarios.

This study aims to assist professionals and scholars in selecting the right models for their sentiment analysis projects by closely examining different machine learning approaches. The information obtained from this study could result in improved sentiment analysis methods and uses, fostering advancements in computational linguistics and natural language processing. The execution and implementation of NLP tasks must be done meticulously.

It can help determine whether a product or service is good or bad, or if it is popular.It can also be useful to know how other people feel about a particular event or person. It can also be used to determine the polarity of text, whether it is positive, negative, or neutral. Sentiment analysis is a method of identifying text that can be categorized into different sentiments.

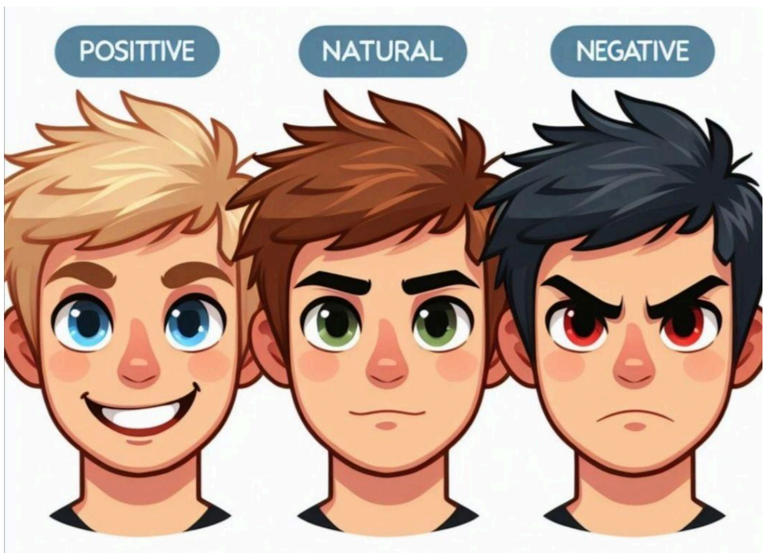

Fig 1.1 : Facial Expressions of Sentiments

In the following research, we will cover the method, data sets, training and testing, validation, and outcomes to provide a detailed analysis of the sentiment analysis machine-learning situation. We expect that our research will shed light on recent progress, challenges, and possible future paths in the constantly evolving field of sentiment analysis.

1.2 Motivation

The copious amount of content created by users online in the current fast-paced digital era has facilitated a diverse collection of opinions and emotions. Comprehending and deriving valuable insights from this vast amount of data is essential in various fields like business, marketing, social sciences, and customer support. Understanding the various subtleties in public opinions buried in written data is crucial, making sentiment analysis a significant subject. There is a growing need for advanced sentiment analysis tools due to the rising amount of online data. The importance of these techniques goes beyond just understanding customer feedback; they also involve recognizing social trends, evaluating public opinion on different topics, and offering valuable data for decision-making processes. The demand for advanced sentiment analysis tools is increasing with the growth of online content. These methods are more than just interpreting consumer feedback; they also involve identifying trends in society, evaluating public opinions on various subjects, and providing useful information for decision-making processes. The evolution of sentiment analysis from rule-based systems to advanced deep learning models demonstrates the fast progress of technology and the flexibility of language and expression in the online world. This study's goal is to conduct a thorough assessment of various machine learning algorithms for sentiment analysis by exploring their origins and following their development over time. This research seeks to enhance the ongoing discussion on sentiment analysis by examining and assessing different approaches, datasets, and outcomes. The aim of this study is to improve comprehension of the present state of this dynamic and constantly evolving field, while also looking ahead to its future opportunities and challenges.

We will explore the intricacies of sentiment analysis methods, showcasing their advantages and limitations while also offering a comprehensive evaluation of the experiment results. We anticipate that this research endeavor will produce valuable insights that will enhance scholarly discussions and have practical implications for industries that depend on sentiment analysis for sound decision-making.

1.3 Purpose of the Study

The rise of internet communication technologies has introduced a new era where opinions and emotions are not only frequently exchanged, but also carry significant impacts. The importance of sentiment analysis has evolved from being just a fascinating topic in academia to being a crucial aspect in making informed decisions across different industries in today's dynamic world. This research aims to explain why different machine learning algorithms are being explored for sentiment analysis, focusing on the challenges of extracting important insights in an era of abundant digital information. This has directed us to a surplus of info and created the need for precision. Sentiment analysis is a strategic tool for managing information overload by accurately interpreting public opinion and identifying key patterns within vast volumes of text data.

Bridging the chasm between research and industry: The rising value of sentiment analysis in decision-making processes emphasizes the need for a seamless integration of academic research and practical applications. This necessity is central to the current research work, which investigates the benefits, shortcomings, and future trajectories of machine learning algorithms in the domain of sentiment analysis. By examining the interaction between academia and industry in the context of sentiment analysis, this research hopes to provide light on the critical role that machine learning algorithms play in determining the discourse and trajectory of this emerging area. The current landscape of decision-making processes is characterized by an increasing dependence on sentiment analysis approaches, creating a clear need for a symbiotic link between academic study and pragmatic application.

This research effort aims to provide light on the subtle dynamics that regulate the use of machine learning algorithms in sentiment analysis. An in-depth investigation of the benefits inherent in these algorithms will be contrasted with the associated shortcomings, offering a complete grasp of the delicate balance that underpins their implementation.

Beyond that, this inquiry looks at potential future trajectories that might define the evolution of machine learning algorithms in the context of sentiment analysis. By outlining future directions, the research hopes to provide stakeholders in both academic and industrial areas with foresight, enabling a harmonic convergence of theoretical breakthroughs and applied needs. In doing so, the effort aims to fill existing gaps by acting as a conduit for the reciprocal flow of information between academics and business, contributing to the evolution of sentiment analysis as a vital tool in modern decision-making frameworks.

Wide Range of Industries: Sentiment analysis, with its capacity to extract emotions from text data, goes beyond typical customer feedback assessment and has potential applications in a variety of sectors. Businesses use sentiment analysis to acquire deep insights into customer opinions and preferences. Understanding the emotional tone of customer evaluations and interactions allows organizations to adjust their tactics to meet changing demands, improve product offerings, and improve customer experiences. This proactive strategy helps to strengthen brand-customer relationships, promote loyalty, and remain competitive in shifting marketplaces.

Advertising, another industry heavily influenced by sentiment analysis, benefits from its ability to predict audience reactions. Advertisers use sentiment analysis to assess the performance of advertisements, find factors that appeal to viewers, and modify messaging in real time.This dynamic adaptation ensures that marketing efforts coincide with the target audience's current feelings, boosting the effect of promotional activity.

In the social sciences, sentiment analysis is an effective approach for identifying societal patterns and trends. Researchers examine public mood expressed on social media platforms, forums, and news items to determine the dominant attitudes, opinions, and worries. This data-driven technique offers useful insights into social dynamics, assisting academics in understanding public opinion on a variety of subjects, tracking sentiment fluctuations over time, and identifying new patterns.

In the Financial World, sentiment analysis is important since it helps analyze market sentiment. Traders and investors use news stories, social media conversations, and financial reports to assess market mood and make sound judgments. Understanding the market's collective mood allows financial professionals to forecast market moves, detect possible dangers, and enhance investment strategies.

In essence, sentiment analysis has grown into a flexible and necessary technology with applications that span several sectors. Its ability to detect emotional subtleties in textual data provides companies, advertisers, and social scientists with an invaluable tool for modifying strategies, increasing engagement, and comprehending complicated patterns in the ever-changing environment of human feeling.

Possibilities towards the Development of Hybrid Models : The fresh perspective offered by hybrid models, which bring together deep learning techniques with rule-based and supervised learning approaches, has energized the current conversation on sentiment analysis. This investigation into hybrid techniques is driven by the possibility to use the strengths inherent in several paradigms, resulting in sentiment analysis systems with more precision and adaptability.

The combination of deep learning, rule-based, and supervised learning approaches marks a significant step forward, providing a novel perspective for improving the efficacy of sentiment analysis systems.

The identification of sentiment analysis as an essential tool for managing the complexities of modern communication serves as the foundation for this inquiry. Recognizing the critical role sentiment analysis plays in comprehending the complex layers of public opinion, the study's reasoning focuses on improving existing methodologies. By diving into multiple machine learning algorithms, the study hopes to improve the sophistication of sentiment analysis methods. This elevation, in turn, is intended to create a sophisticated knowledge of public mood, therefore helping decision-makers to make more informed and discriminating decisions.As technological developments accelerate, the future of hybrid models in sentiment analysis promises an era that will see greater accuracy, adaptability, and comprehension in understanding, deciphering and also predicting human emotions.

1.4 Research Question

This study will focus on investigating the queries listed below:

1. Do all tweets made by users contain an emotion, like joy, sorrow, or another sentiment?
2. What is the impact of different machine learning algorithms on the development of sentiment analysis?
3. How various machine learning algorithms contribute to the advancement of sentiment analysis?
4. How can the use of Support Vector Machines (SVM) and BERT embeddings improve sentiment analysis accuracy compared to older methods?
5. What challenges do sentiment analysis algorithms encounter with linguistic nuances, and how impactful are they if not adequately addressed?
6. How can Support Vector Machines and BERT be beneficial in enhancing sentiment categorization accuracy, particularly with large datasets through supervised learning methods?

7. How has sentiment analysis been revolutionized by deep learning models like BERT through capturing sequential dependencies and understanding complex contextual relationships?

8. How does the fine-tuning approach to BERT embeddings influence the overall performance and accuracy of sentiment analysis ?

9. Is the Roberta Model still reliable?

10. Do traditional models perform less when data has contextual meaning?

11. Why Fine Tuning and preprocessing the data leads to greater accuracy?

1.5 Expected outcome

The use of Support Vector Machines (SVM) in sentiment analysis is an interesting research topic. Initial expectations focus on understanding how SVM enhances sentiment analysis accuracy when compared to older approaches. Furthermore, an investigation into the influence of different training dataset sizes attempts to identify SVM's scalability in sentiment analysis applications. Further investigation into alternative hyperparameter settings inside the SVM framework aims to fine-tune models for best performance, adding a layer of precision to sentiment analysis results. The study also broadens its scope to examine SVM's adaptability across several areas and languages, offering light on its potential worldwide application.

Sentiment analysis with BERT opens a new avenue of investigation. The project aims to determine how fine-tuning procedures affect the efficiency of BERT in collecting sentiment subtleties. Furthermore, the adaptation of BERT models to diverse domains and languages is a key feature in sentiment analysis applications. The study also investigates the interpretability of sentiment analysis results when using standalone BERT models, with the goal of making these findings more understandable and useful. Assessing BERT models' susceptibility to adversarial assaults assures their robustness in real-world sentiment analysis settings. BERT-based sentiment analysis with fine tuning aims to optimize models for improved performance, contributing to the ongoing improvement of sentiment analysis approaches. Hugging Face library has a model named ROBERTA that is specially designed for handling contextual data such as articles , tweets and public opinion.

In conclusion, the main goal of this research is to provide a comprehensive and valuable evaluation of SVM & BERT machine learning methods for sentiment analysis focusing on the polarity of tweets. The planned goals involve a comprehensive comprehension of past advancements in sentiment analysis methods, along with a detailed assessment of different machine learning strategies and their real-world use in analyzing public sentiment in the technological era.

1.6 Project Management and Finance

Financial support is crucial for the research project's success across disciplines. The funds will cover research materials, data acquisition, technology, software, and dissemination through communication channels and conferences. Building relationships with experts in the field will enhance the project's quality and significance, ultimately advancing relevant fields.

1.7 Report Layout

This study looks at various sentiment analysis methods in diverse contexts and periods. As data collection methods evolve due to technological advancements, new and unique approaches to sentiment analysis are needed to analyze the complex and varied nature of the data. Innovative and sophisticated strategies are constantly being created to remain current. In the modern era, it is crucial to recognize the significance of comprehending online opinions and the necessity of utilizing machine learning techniques as a valuable tool.

In the study methodology section, the report also discusses the methods used to investigate different machine learning techniques for sentiment analysis. The text elaborates on the chosen datasets, the reasoning behind them, and the specific research method used. This part discusses the challenges of comparing these methods and describes the experimental setup, incorporating various machine learning algorithms. The research's ultimate findings are outlined in the final sections of the research paper, which also detail potential areas for future research and provide a summary and conclusion. Following a concise overview of the research results, an evaluation conversation examines the study's overall impact. The ending offers solutions to the primary research inquiry, while highlighting the significance of the results for both the business sector and the academic community. It serves as a starting point for future research objectives, suggesting potential avenues for further investigation in the constantly changing area of sentiment analysis. This comprehensive conclusion outlines the main results of the research, offers an analysis of its significance, and suggests a plan for future studies and applications by scholars and professionals in the field. The research report concludes with a detailed reference section that gives credit to sources crucial to the study. For those interested in delving further into the topic, this collection of writings is an invaluable asset.

In addition to maintaining the study's legitimacy, the thorough reference provides a starting point for readers who are keen to explore the intricacies of the subject matter. Essentially, the careful use of a substantial reference section upholds the paper's intellectual integrity while also creating an atmosphere that is favorable to further investigation and scholarly interaction.

CHAPTER 2

BACKGROUND RESEARCH

2.1 Preliminaries/Terminologies

Proficiency in large-scale data analysis is essential in fields including business, marketing, social sciences, and customer service. Sentiment analysis, often known as opinion mining, is a rapidly developing area that is crucial for identifying complex emotions in textual data. Large-scale data analysis is crucial in various fields, including business, marketing, social sciences, and customer service. Sentiment analysis, or opinion mining, is a rapidly developing area that helps identify complex emotions in textual data. The demand for better sentiment analysis tools is increasing due to the volume of online content. These technologies are essential for identifying social trends, analyzing consumer feedback, gathering data for decision-making, and determining public opinion on various issues. Proficiency in organizing vast amounts of information is essential in domains like commerce, advertising, social research, and client assistance. Sentiment analysis has evolved from rule-based systems to advanced deep learning models, reflecting rapid technological advancements and changes in language and expression trends in the digital age. The demand for better sentiment analysis tools is expanding, as seen by the volume of content available online. These technologies are critical for identifying social trends, analyzing consumer feedback, gathering useful data for decision-making, and determining public opinion on a range of issues. Proficiency in organizing vast quantities of information is essential in domains including commerce, advertising, social research, and client assistance. Finding the complex emotions hidden in text data is a growing area of study in opinion mining, also referred to as sentiment analysis.

The growing volume of online content demonstrates the rising need for improved sentiment analysis technology. These technologies are essential for recognizing social trends, gathering public viewpoints on various subjects, providing vital information for decision-making, and assessing customer feedback. Sentiment analysis has evolved over time, transitioning from rule-based systems to advanced deep learning models, demonstrating the rapid advancements in technology and changes in language and expression trends in the digital age.

The aim of this research is to fully analyze various machine learning techniques for sentiment analysis, along with exploring the historical origins and advancements in the field. The aim is to participate in the continuing debate on sentiment analysis through a detailed study and assessment of different methodologies, data sets, and results. The aim of this study is to examine the future possibilities and obstacles in this ever-evolving field, while also gaining insight into its present condition.

Subsequent sections of the study will delve into the details of sentiment analysis approaches, highlighting the advantages and disadvantages of each method and presenting a comprehensive examination of the experimental outcomes.

2.2 Related Works

The study evaluated TikTok reviews using VADER sentiment and the SVM model, classifying them as positive, neutral, or negative. It was discovered that utilizing RUS and ROS did not enhance validation performance owing to the danger of losing critical information. The study validates the strategy's feasibility, although more research is needed for numerous languages and different machine learning techniques [1] . The study focuses on sentiment analysis of financial tweets in Turkish utilizing hybrid lexicon-based and machine learning approaches. However, the quick transmission of information in online social networking platforms can have an influence on traders' investment strategies and emotional states. Future studies should look at the relationship between information and public emotional states. More complex sentiment analysis models and natural language processing approaches should be investigated, such as context-aware sentiment analysis and multimodal data integration. The study also investigates the link between multimedia content and public attitudes, as well as the consequences on stock market activity. Expanding the analysis to encompass numerous social media platforms and analyzing sentiment patterns can offer a more complete picture of the impact of online communities on stock market opinions [2] . The study uses natural language processing (NLP) essential techniques, a sentiment lexicon enhanced with the assistance of WordNet and fuzzy sets to make a Hybrid model . This enhances the result by improving accuracy and performance [3] .

The paper reviews the use of five feature selection approaches and seven machine learning algorithms to sentiment analysis on an online movie review dataset. The results suggest that Gain Ratio performs the best for emotional feature selection. Sentiment analysis is a technique for extracting opinion and subjectivity from user-generated text material. It varies from typical topic-based text categorization in that it classifies opinionated content according to the mood communicated. Feature selection is critical in sentiment analysis, and correctly chosen representative features can enhance categorization [4] .Research into using social networking sites (SNS) for mental health evaluation, particularly in identifying depression, has been prompted by the emergence of these platforms. Research has explored machine learning and natural language processing on social media data, like Facebook and Twitter, in order to understand users' feelings. The research shows a growing fascination with using online interactions to gauge anxiety levels, highlighting the potential of computational methods in evaluating mental health. The research examines text data from product reviews, tweets, and movie critiques to ascertain the sentiment of messages or posts. SVMs and decision trees are utilized as classification methods, with evaluations considering accuracy, precision, recall, F1-score, and ROC curve. For most datasets, decision trees and SVMs typically have lower mean square error or higher accuracy. The article introduces sentiment analysis methods and compares three different classifiers using machine learning on five datasets [6]. Acknowledging this difficulty, Basant Agarwal and Namita Mittal have made recent advancements in sentiment analysis using machine learning models. Machine learning algorithms use techniques for selecting features to handle high dimensionality. Wordnet was enhanced, educated, and perfected. The methods identify key features while also minimizing irrelevant data and noise. This study sets up the structure for assessing the effectiveness of sentiment analysis models based on machine learning, highlighting their growing importance in the field [7]. Using trainable algorithms for sentiment analysis is essential for gaining valuable information from online data. This research examines how sentiment analysis, specifically machine learning methods, can forecast outbreaks and epidemics. The literature review highlights the significance of this approach, especially when utilized for healthcare data obtained from platforms like Twitter. The survey included significant research papers published between 2010 and 2017, providing the basis for the present study [8]. This research investigates sentiment analysis on Twitter specifically targeting electronic products. The text discusses unique difficulties in evaluating emotions in tweets, including limitations on characters and the use of slang. The machine learning technique is applied to categorize Twitter posts as either positive or negative, and also to collect public opinions on products like smartphones and laptops. The study assesses how specific domain expertise affects sentiment classification and introduces a novel feature vector [9]. This study aims to analyze twitter sentiment utilizing Logistic Regression, VADER, and BERT sentiment analysis methods. The suggested analytic approaches are sensitive to sentiment expressions in social media contexts and may be extended by domain. Despite implementing three separate methods, all preprocessing and subsequent procedures, save for sentiment analysis, will be identical. Using the same processing processes, we can compare the three suggested sentiment analysis methods. This study has several uses, including gathering public opinion for government and health professionals to utilize in their decision-making processes [10]. The paper offers a Bert-based network public opinion sentiment analysis approach to increase the analytical efficiency of sentiment trends. This technique converts the input text sequence into three vectors: query, key, and value. For obtaining the encoded vector, apply the Softmax function on the inner product of the query vector and the key vector. The encoded vector is then fed into the trained classifier to determine the recognition result. This strategy can address RNN's weaknesses in disregarding context while also simplifying the RNN and Text-CNN algorithms. The experiment demonstrates the suggested method's performance on a social network data set [11]. The purpose of this research is to study deep learning models for sentiment analysis of COVID-19-related tweets. The dataset was retrieved from the tweeter web API and labels were manually assigned as positive, negative, or neutral. Two deep learning models were used for sentiment analysis: Recurrent Neural Networks (RNN) and Bidirectional Encoder Representations (BERT) model [12]. The experiments show that the Bert model used in this paper can achieve an accuracy of 97.35% for the analysis of investor sentiment, which is better than both LSTM and SVM methods [13]. This paper consists of three primary steps:Building a hybrid collaborative filtering-based recommendation model; fine-tuning a BERT-tuned model for accurate sentiment categorization and enhancing the e-commerce industry's product selection procedure in the RS by leveraging BERT insights to improve suggestion accuracy. Also increase the accuracy to 91% [14].This paper aims to prevent term ambiguity, often known as the "domain sensitivity problem" in recommendations. When it comes to electronic product suggestion, the suggested contextual information sentiment-based model performs better than the traditional collaborative filtering strategy, as demonstrated by the results of RMSE and MAE measurements [15]. The study offers a novel approach to sentiment analysis for news events. More specifically, a Word Emotion Association Network (WEAN) is constructed to simultaneously describe the semantics and emotions of a news event based on the social media data (i.e., words and emoticons), which forms the basis for the news event sentiment computation. The standard emotion thesaurus is used to further enhance the original word emotions obtained by a word emotion computation algorithm based on WEAN. By using the word emotions, we are able to determine the feeling of each statement [16]. In this paper, an SVM-based ensemble method modification is suggested. To enhance prediction performance, the suggested method combines both undersampling and oversampling. In the context of imbalanced sentiment classification, a comparison analysis of the efficacy of the suggested approach and sampling strategies utilizing SVM classifiers is also carried out. The usefulness of the conducted experimental comparisons is examined using appropriate metrics [17]. In this paper, to solve some problems proposed for the Persian language, they tried to use two types of deep neural network models to classify text documents based on sentiment polarity of them.with these deep learning models they have achieved remarkable results in various applications such as computer vision, speech recognition, and natural language processing [18]. This work provides an overview of ABSA and the overfitting problem. After the analysis, they improved their model by improving the accuracy and F1 score of the existing model by refining the technique.THe model outperforms the others, achieving the best results for the restaurant dataset [19]. This paper is devoted to the improvement of the recommendation system and the emergence of sentiment analysis and opinion mining techniques. This paper proposes a sentiment analysis-based collaborative filtering using an Arabic dataset to provide book recommendations.The proposed method improves the accuracy of the Arabic recommender system and reduces the mean square error in terms of RMSE and MAE [20].

2.3 Comparative Analysis and Summary

Table 2.3.1: Comparative Analysis

S/L	Author Name	Used Algorithm	Accuracy
1	Mahmud Isnan, Gregorius Natanael Elwirehardja, Bens Pardamean [1]	SVM	0.81%
2	Handan Cam, Alper Veli Cam, Ugur Demirel, Sana Ahmed [2]	SVM, KNN, MLP	89% , 78% , 88%
3	Orestes Appel, Francisco Chiclana, Jenny Carter, Hamido Fujita [3]	Hybrid (WordNet , Fuzzy logic)	88.02%
4	Sharma, A., & Dey, S [4]	SVM , KNN	90.15% , 75.15%
5	Hassan, A. U., Hussain, J., Hussain, M., Sadiq, M., & Lee, S.	N/A	N/A
6	Bhagat, Abhishek and Sharma, Akash and Chettri [6]	Naive Bayes , SVM	87.78% , 84.93%
7	Agarwal, B., Mittal, N., Agarwal, B., & Mittal [7]	WordNet	80.1%

8	Singh, Rameshwer, Rajeshwar Singh, and Ajay Bhatia [8]	N/A	N/A
9	Neethu, M. S., and R. Rajasree [9]	Naive Bayes , SVM	91% , 93%
10	A. J. Nair, V. G and A. Vinayak [10]	BERT , VADER	0.92% , 0.88%
11	Q. Dong, T. Sun, Y. Xu, X. Xu, M. Zhong and K. Yan [11]	RNN	98.72%
12	A. Topbaş, A. Jamil, A. A. Hameed, S. M. Ali, S. Bazai and S. A. Shah [12]	RNN , BERT	86.4%, 83.14%

2.4 Scope of the Problem

To meet the ever changing requirement of the increasing need for effective tools for analyzing user-generated attitudes across a range of industries, the previous related works and study investigates the broad subject of sentiment analysis. They study and evaluates the accuracy of each model and charts the involvement of opinion mining for sentiment analysis from rule-based systems to complex and intricate machine learning models such as KNN, CNN, MLP, SVM, Fuzzy logic,WordNet,Naive Bayes,RNN, Random Forest, Decision Tree, VADER, Linear Regression & BERT. The implications of the findings extend beyond the analysis of sentiment analysis and go further as customer feedback, encompass the identification of social patterns, the assessment of publicly available sentiment and the florey of informative data for decision-making procedures. By offering assistance to industry leaders that rely heavily on sentiment analysis for informed and critical decision-making. All these studies hope to significantly contribute to both academic discourse and practical implementations. This research encountered several challenges while navigating the field of sentiment analysis. The domain nature of languages, the complexity of communication in the digital realm remains a persistent obstacle, requiring adaptability in machine learning models. A random approach is essential due to the intricate interplay of attitudes and contextual complexities within textual data. Numerous challenges, including the abundance of irrelevant data, readability issues, the intricate nature of the data, and potential human errors, were addressed. Notably, the BERT model, when not Fine-Tuned, exhibited lower accuracy, shedding light on potential drawbacks and underscoring the importance of fine-tuning. Additionally, the scalability and generalizability of the proposed models to diverse datasets and real-world applications pose ongoing challenges, necessitating a thorough examination of practical implications. In conclusion, the dynamic domain of sentiment analysis requires a proactive strategy and approach in keeping up with technological and methodological developments to safeguard the research's continued relevance and impact on modern problems.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research Subject and Instrumentation

The main objective of this study is to compare multiple machine learning algorithms and techniques for sentiment analysis. Recognizing the accuracy and effectiveness of models, which includes SVM & BERT . This research project utilizes sentiment analysis datasets to test and evaluate various machine learning models. The primary tool for measuring is the Python coding language along with google colab combined with tools like Scikit-Learn and Hugging Face's Transformers Library. While Hugging Face's Transformers Library simplifies the integration of state-of-the-art models like BERT, Scikit-Learn's user-friendly Python library makes tasks ranging from data preprocessing to model evaluation easier. TextBlob, Pandas, PyTorch or TensorFlow and the graphic capabilities of Matplotlib and Seaborn tools enable effective data processing, sophisticated model construction and perceptive result reporting with robust visualization .

3.2 Data Collection

The main source of data for this study is from Twitter. Using the Twitter API. By collecting from a total of 28 users the total number of tweets is 4501. From each user the last 150 tweets was the requirement but not every user has 150 tweets in their profile. The user selection was by their IPL career record. Top 10 Batsman and Bowler of all time along with top performers of 2022 edition were selected. The dataset contains username and their respective tweets.The reason for selecting the players are very straightforward. It is because of their own individual performance.

3.3 Data Analysis

Number of tweets per player is not consistent. Some have a greater value than others . Mostly because of that, the leniency and relevance of all the tweets are not the same. Visual representation of total number of tweets per user is shown below:

After using polarity to determine the polarity of the tweet we get a score of polarity ranging from -0 to 1 . By assigning a polarity score greater than 0 to positive, less than 0 to negative and else to natural for their respective tweets we can label sentiment to tweets as positive , negative and neutral . A bar representation of the tweets sentiment :

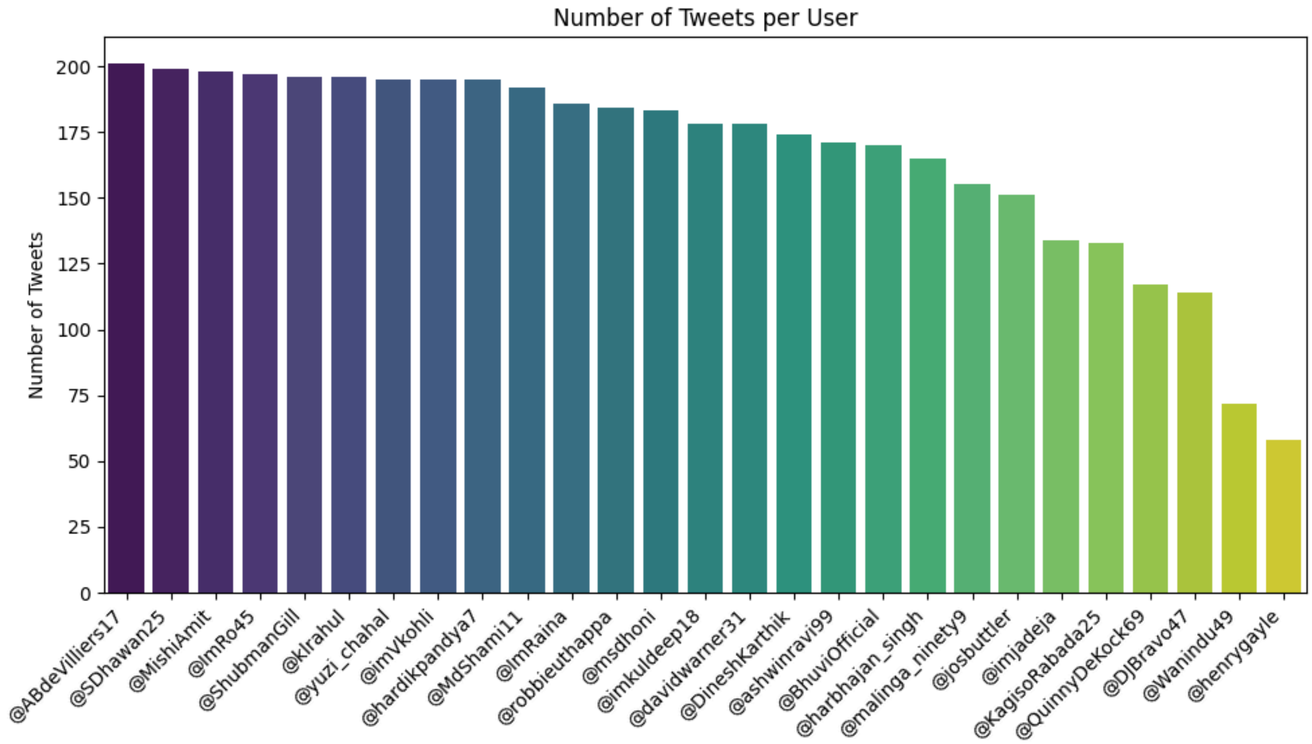

Fig : 3.3.1 : Number of tweets per user

The total ratio of the tweets represent what percentage of tweets are from each player and how they affect the total percentage of tweets. Majority of the percentage are around 4% and the highest being 4.5% and the lowest is 1.3% This gives us an idea of how the dataset is structured.the percentage pie chart is as follows :

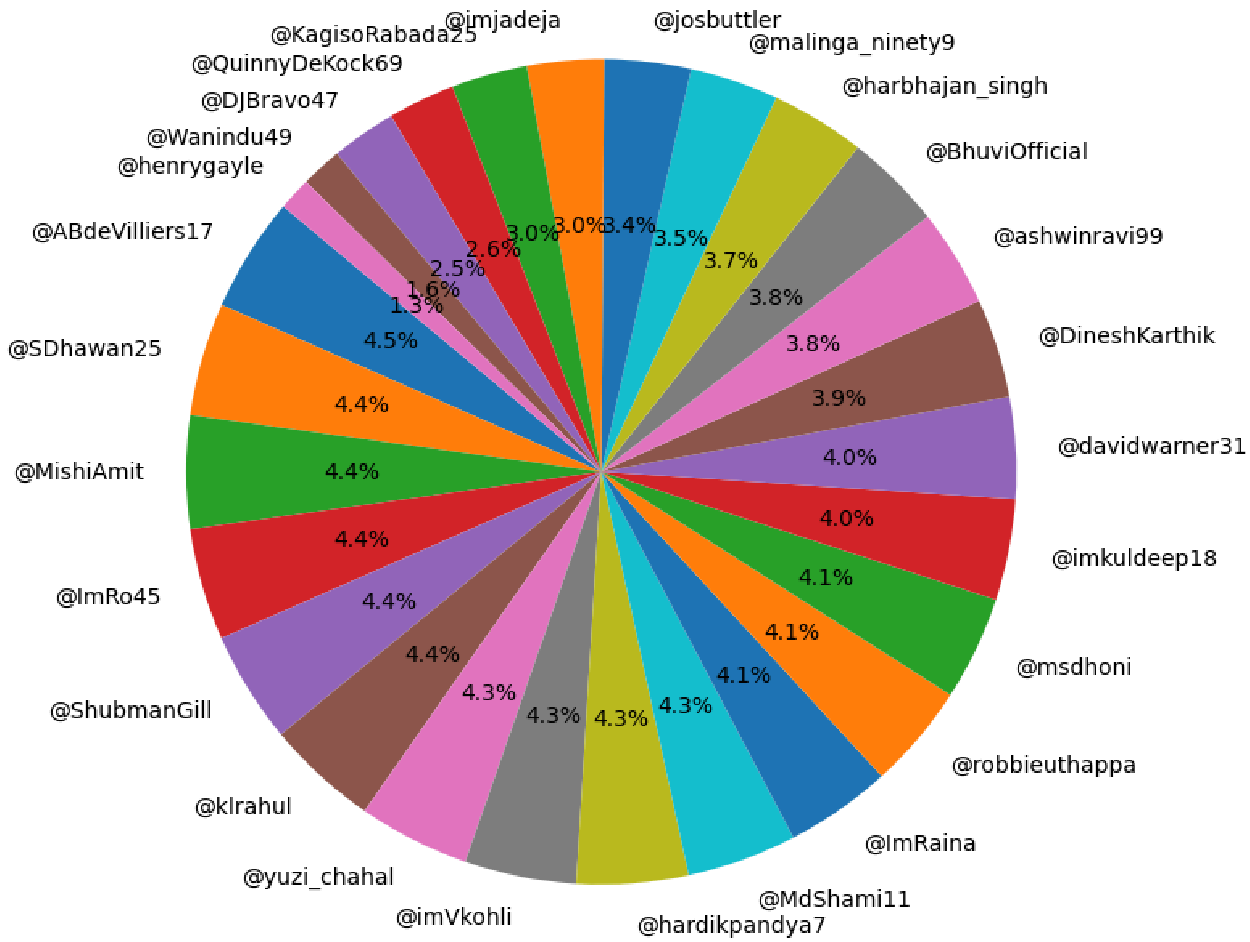

Fig : 3.3.2 : Distribution of tweets per user by percentage

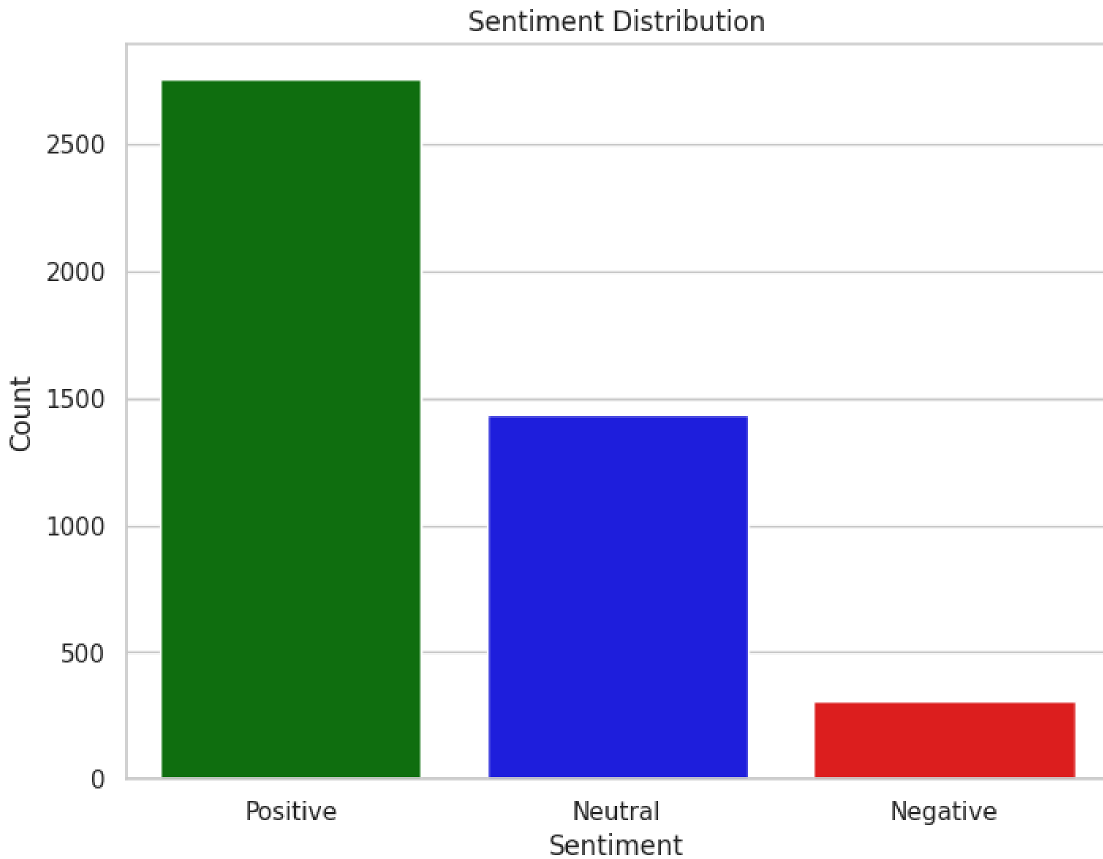

Fig 3.3.3 : Total Sentiment Distribution of tweets

Here a total number of tweets which is 4501 tweets are represented by their sentiment. This shows us a great number of positive tweets which are the predominant sentiment in this dataset then natural and finally negative tweets. The exact number of positive tweets are 2759 followed by 1436 natural tweets and 306 negative tweets.

The percentage distribution of sentiment are as follows : 61.3% positive, 31.9% natural and 6.8% negative as shown below :

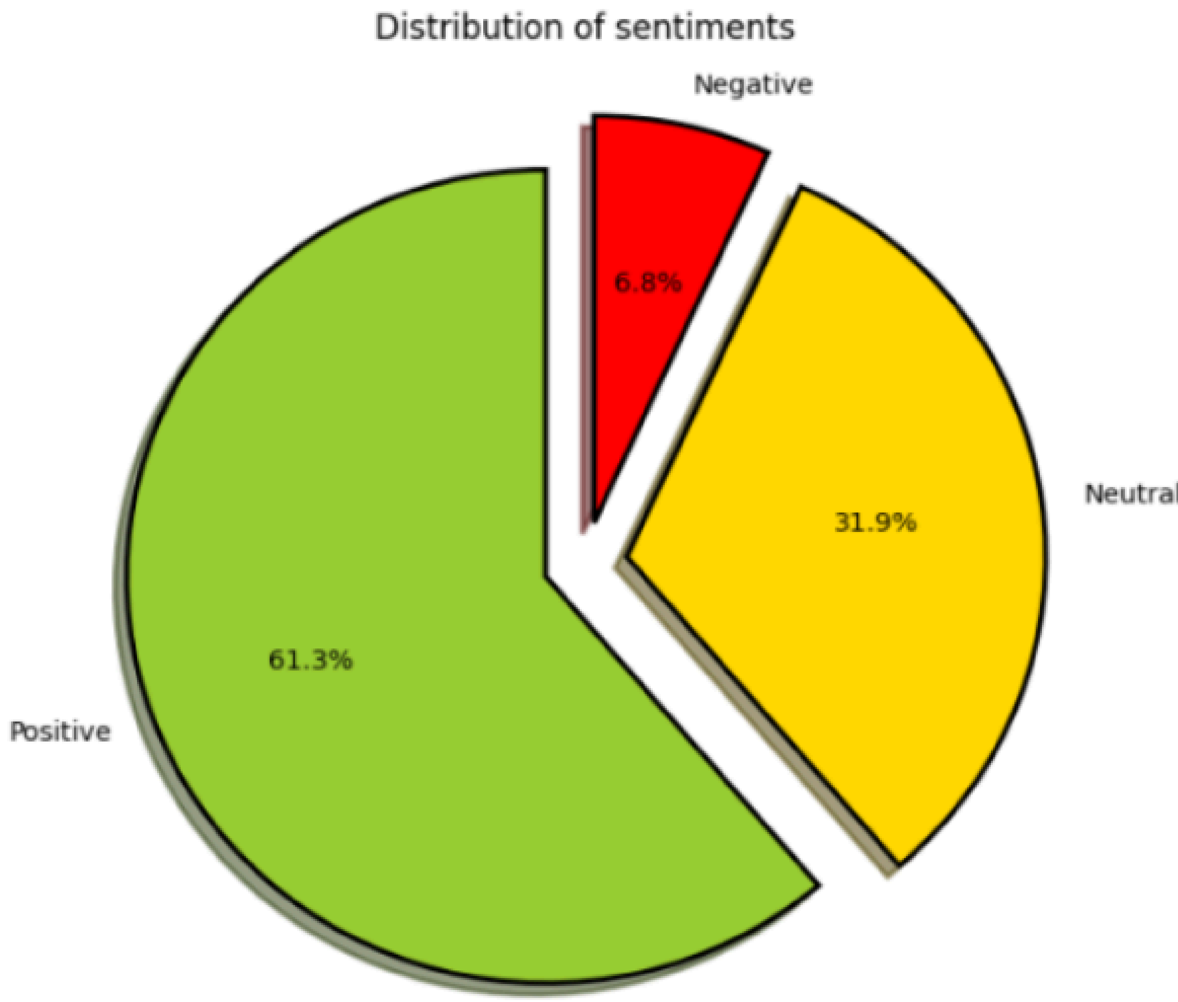

Fig : 3.3.4 : Distribution of Sentiments

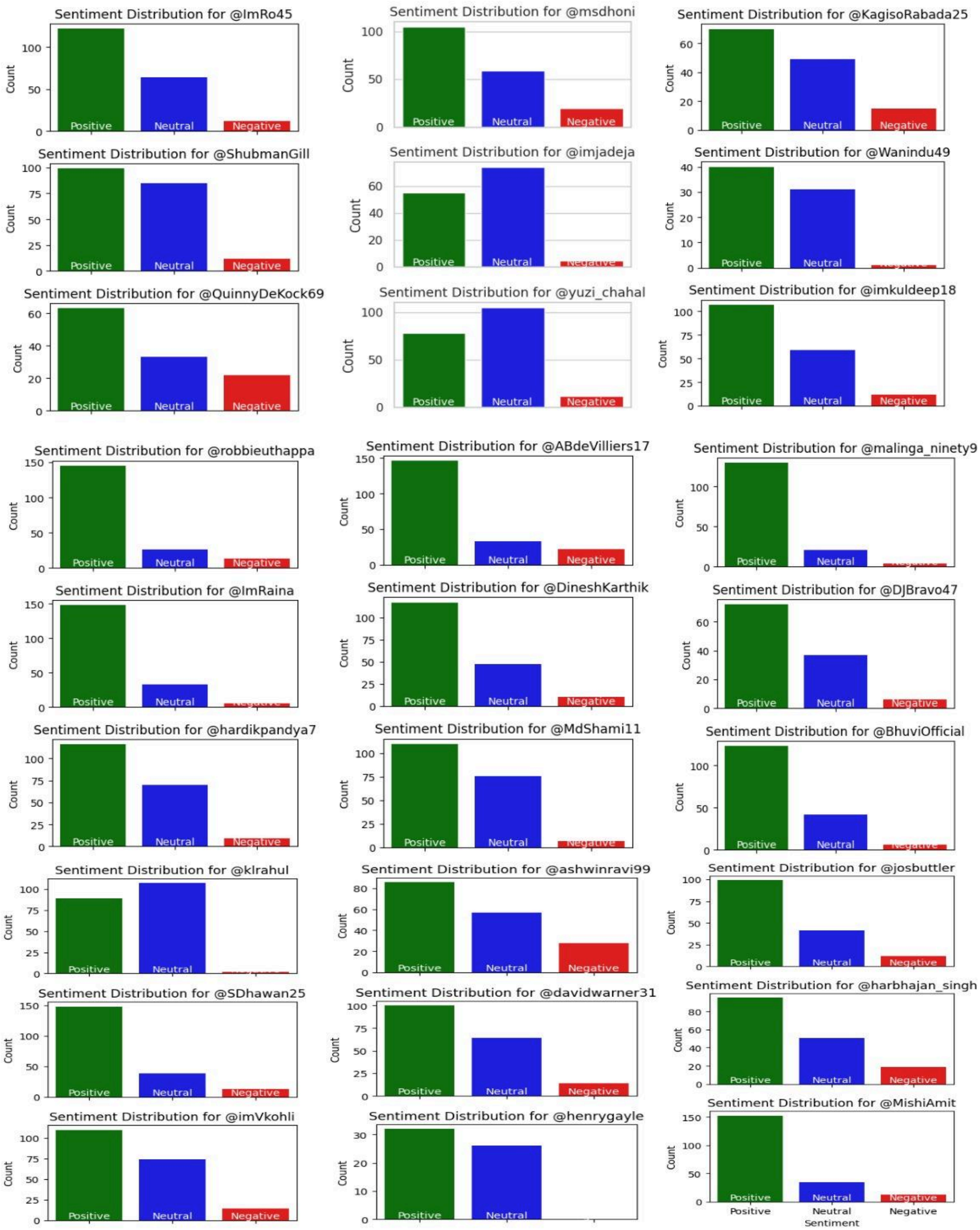

Fig 3.3.5 : Sentiment distribution as per username

A total of 28 users tweet sentiments are visualized along with their represented sentiment .

Most used words in tweets & dataset are visualized by Wordcloud function.

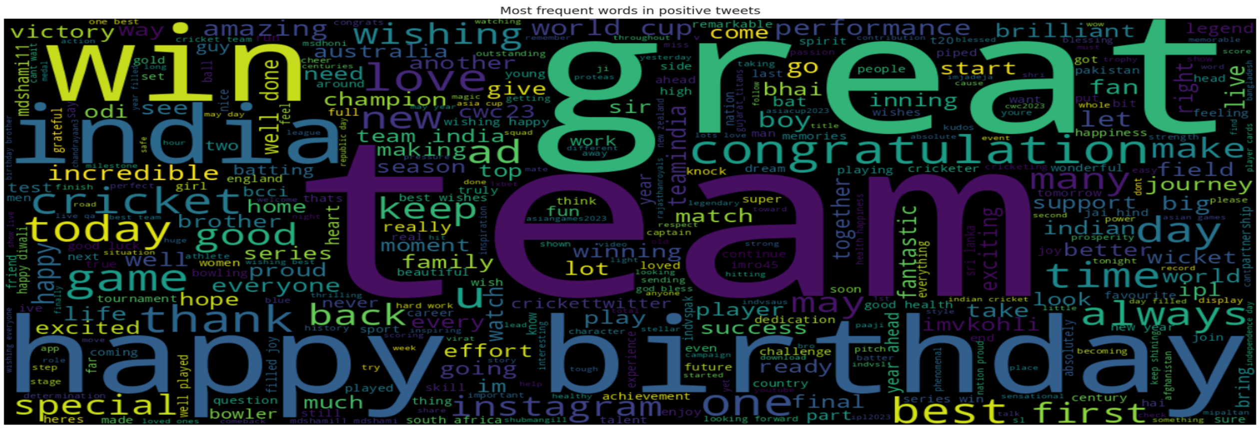

Fig 3.3.6 : The most used positive words

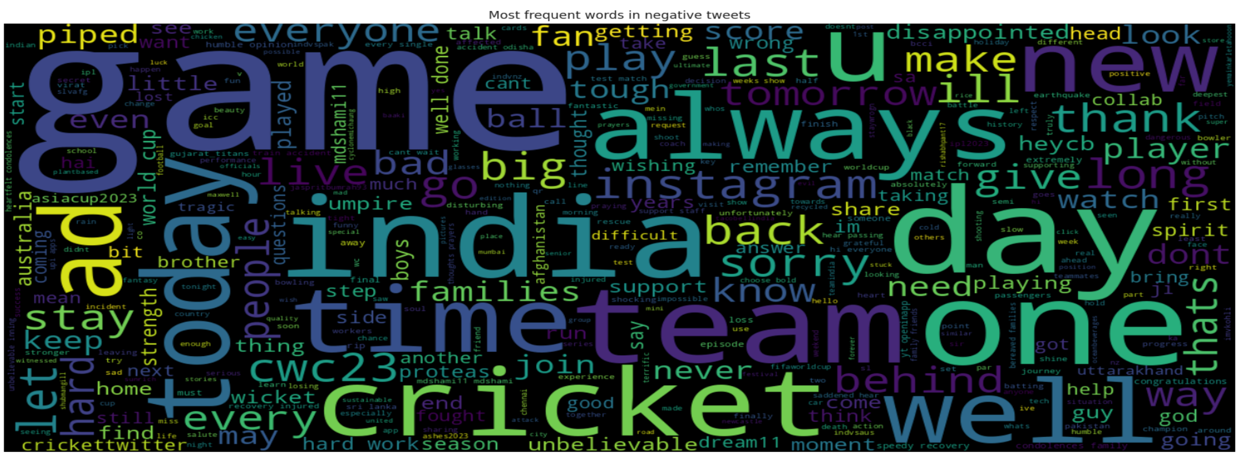

Fig 3.3.7 : The most used negative words

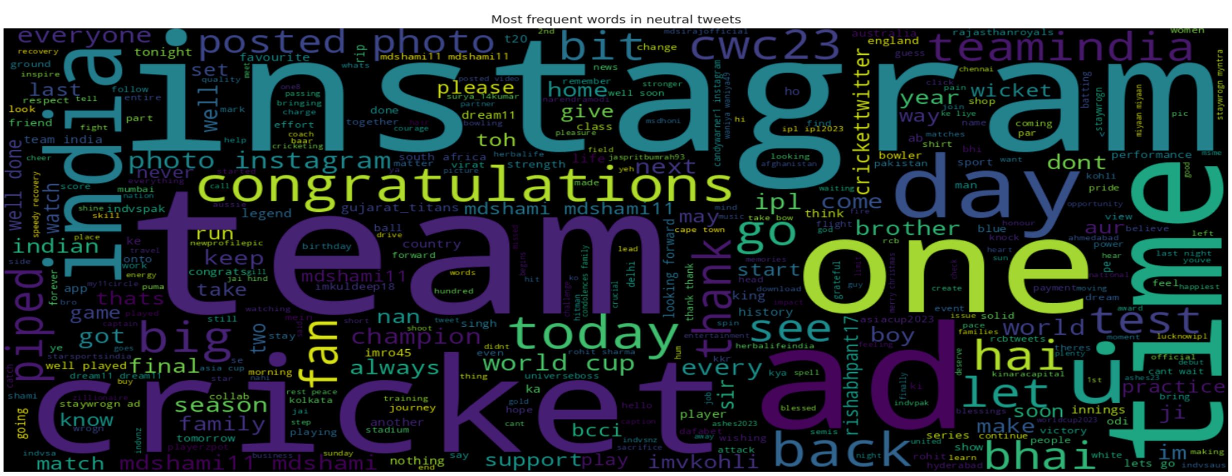

Fig 3.3.8 : Most used Neutral words

3.4 Proposed Methodology

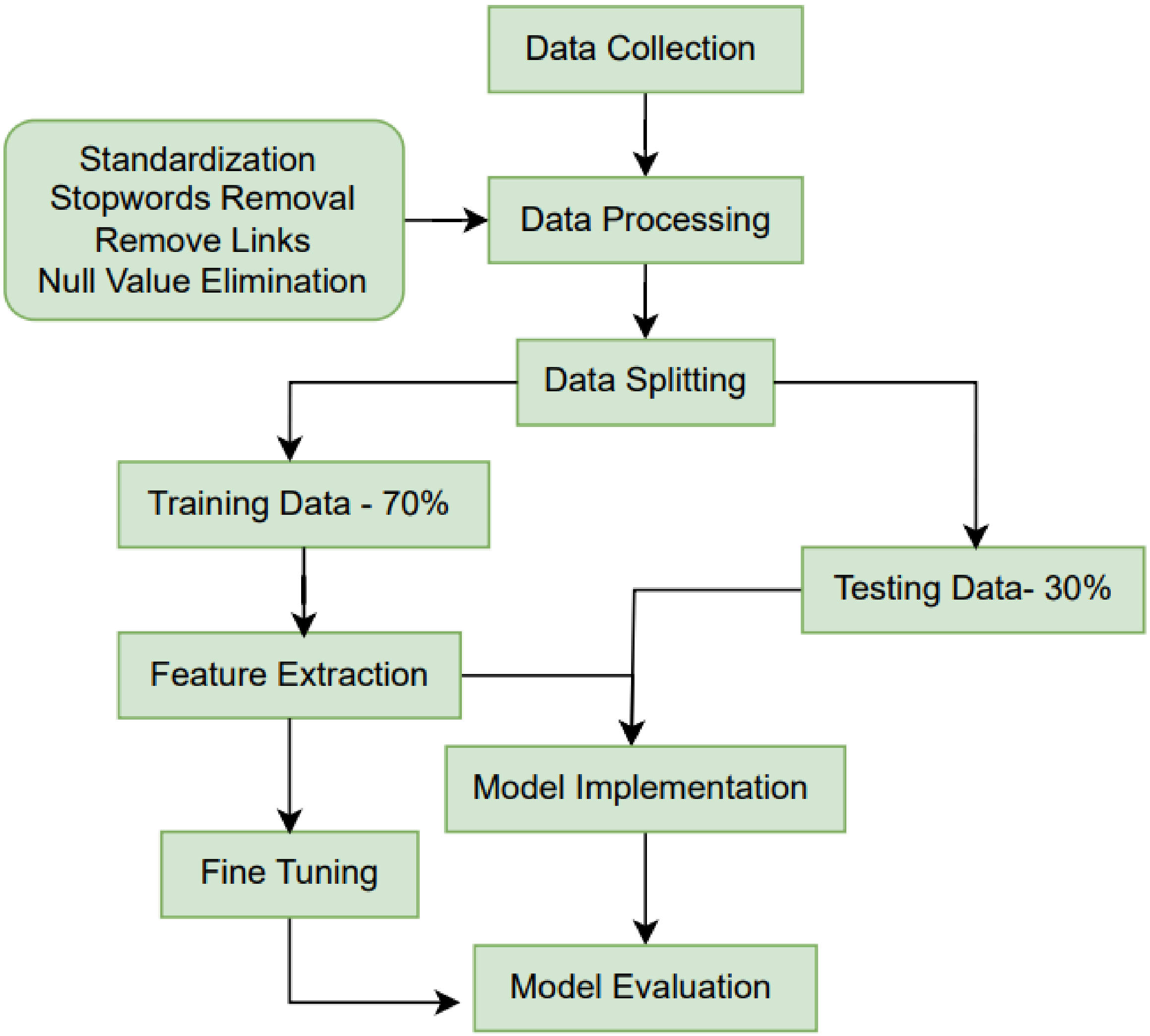

Fig 3.4.1 : Methodology

3.5 Implementation Requirements

Data Loading and Preprocessing:

Load the dataset into google colab environment. Data preprocessing is an essential stage in preparing unprocessed data for analysis or machine learning by performing tasks such as cleaning, converting, and organizing it. In our case we are executing : Null or missing value elimination , Removing Links from tweets, Eliminating stopwords, Removing Emojis, Removing Hashtag, Removing Non-English characters such as other languages, Converting all text data to lowercase.

Tokenization :

To improve its performance for the model, visualization and representation, textual input for BERT is transformed into subword tokens by data tokenization. BERT captures linguistic detail by breaking words down into subword units using the WordPiece tokenization approach. This method reduces the impact of out-of-vocabulary problems and allows for a varied vocabulary. To further aid with model comprehension, BERT adds two new tokens: [SEP] for separation and [CLS] for classification. Contextual information is preserved in tokenized data, which is essential for BERT's contextualized embeddings. Better performance in a variety of natural language processing tasks is made possible by BERT's ability to comprehend complex linguistic patterns and subtleties thanks to the careful tokenization procedure.

Feature Extraction :

The practice of carefully extracting features from unprocessed text is essential to Support Vector Machines (SVM) in order to enable the model's discriminative skills. Many methods are used: Term Frequency-Inverse Document Frequency (TF-IDF) assesses word relevance in the document corpus, and the Bag-of-Words (BoW) approach quantifies word frequencies. Word Embeddings provide dense vector representations by utilizing semantic linkages. Sequential patterns are captured by N-grams, and homogeneity is guaranteed via feature scaling. SVM makes use of these numerical representations to choose the best hyperplane for classification, highlighting the need of careful feature extraction in maximizing SVM's effectiveness in a variety of textual contexts.

Data Split, Train, Testing :

The careful separation of data into training and testing sets is a crucial task in the field of machine learning. This procedure, called data splitting, is setting aside a certain subset for assessment and assigning a significant amount for model training. A clever split ratio selection guarantees model generalization. The testing set, which was not used during training, evaluates the model's prediction ability on untested data, whereas the training set acts as a testing ground for learning model parameters. This careful division reduces overfitting inclinations and promotes strong model performance on a variety of datasets, representing a core procedure in the academic quest for machine learning brilliance.

Bert Fine Tuning :

Adjusting Modern pre-trained language models, such as BERT, are crucial for helping the system adjust its generic language comprehension to particular downstream tasks. The pre-trained BERT model is fine-tuned by training it on a task-specific dataset, which enables it to pick up on task-specific subtleties and patterns. Task-specific goal functions are used to fine-tune the BERT layers that already exist, as well as to build task-specific layers. The application of BERT to tasks like named entity identification, question answering, and sentiment analysis is improved by this transfer learning approach. By utilizing the pre-existing linguistic information stored in BERT, fine-tuning allows for better performance on domain-specific tasks with a small amount of labeled data. The Fine-Tuning process uses RandomSampler for the training set and SequentialSampler for the test set; this ensures that each batch during training is randomly sampled, which can improve generalization. SequentialSampler ensures that the test set is processed in order without shuffling, which is standard practice for evaluation. Gradient clipping is a technique to prevent the gradients from becoming too large during training, which can stabilize training and prevent issues with exploding gradients. The changes made in the fine-tuned BERT code mainly focus on improving training stability and convergence through learning rate scheduling, gradient clipping, and careful data sampling. While the initial code leverages mixed precision for efficiency, the fine-tuned approach uses more traditional techniques that can still provide significant benefits in training complex models like BERT. The impact of these changes would generally be seen in improved model performance, stability, and possibly training efficiency, though this can vary based on the specific dataset and training setup.

Model Training :

To make forecasts and predictions on unfamiliar data, an algorithm needs to acquire patterns and relationships from a dataset through the procedure of model training. During training, the model continuously adjusts its parameters in order to minimize the gap between its anticipated outcomes and the actual results in the training dataset.. Typically, this procedure entails establishing an optimization technique, such a stochastic gradient descent, to minimize the error in the model and developing a loss function to quantify the error. With exposure to a variety of instances from the training dataset, the trained model iteratively refines its internal parameters, enabling it to make accurate predictions on fresh data.

Model Deployment :

After all the necessary steps and requirements are fulfilled, Models are implemented. The step from a trained machine learning model to real-world use is called model deployment. In order to enable the model to generate predictions in real-time on fresh, unknown data all the previously mentioned critical steps are the key to success. During deployment, the model's inference speed and resource efficiency should be optimized.

CHAPTER 4

RESULT ANALYSIS AND DISCUSSION

4.1 Experimental Setup

1. Twitter as data source

Description: A set of textual data with sentiment categories (positive, negative, and neutral) tagged that are tweets.

Size: 4501 tweets and 28 username

2. Machine Learning Models: BERT , SVM , ROBERTA

3. Google Colab

4.2 Experimental Results & Analysis

1. Pre processing: Tokenization, stop word management, and lemmatization are examples of text preparation. Feature extraction may be used to convert text input into numerical vectors for machine learning models.

2. Dividing Data for Testing and Training: Divide the dataset into training and testing sets. 3.Model Training: Train each machine learning model with the training set. Utilizing the testing set from model testing, evaluate the model's performance.

4. Evaluation Metrics: Accuracy, Additional Metrics accuracy, recall, and F1-score. Reasoning Explain the thinking behind the choice of these measures and how they help to ensure a comprehensive evaluation.

Model	Accuracy
SVM - Linear	0.84
SVM - RBF	0.78

Classification Report of SVM - Linear		
precision	recall	f1-score
0.91	0.22	0.36
0.83	0.82	0.83
0.85	0.93	0.89

Classification Report of SVM - RBF		
precision	recall	f1-score
1.00	0.03	0.06
0.92	0.55	0.69
0.74	0.99	0.84

Table 4.2.1 : Result Analysis of SVM models

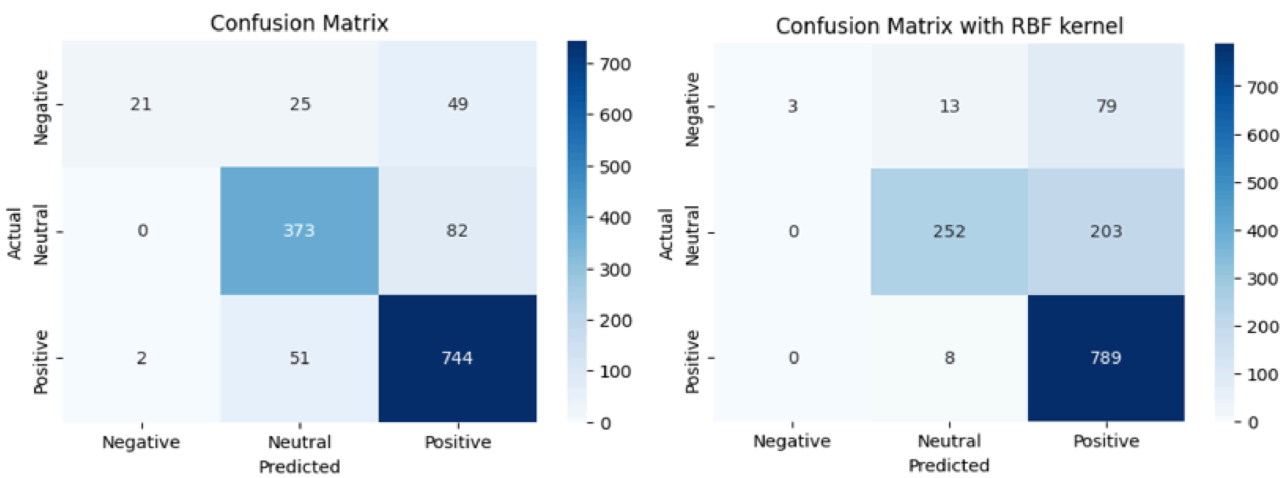

Table 4.2.2 : SVM Confusion Matrix

The SVM models with their respected accuracy proves that SVM Linear Kernel has more accuracy than RBF with an edge of 0.6 %. The RBF underperforms because it is not suitable for this type of tasks on the other hand SVM Linear has once again proven its effectiveness and efficiency over other models.

Model	Accuracy
BERT	0.89
BERT Fine Tuned	0.90
ROBERTA	0.90

Classification Report of BERT		
precision	recall	f1-score
0.61	0.52	0.56
0.95	0.84	0.89
0.89	0.96	0.92

Classification Report of BERT Fine - Tuned		
precision	recall	f1-score
0.64	0.52	0.57
0.94	0.87	0.90
0.90	0.96	0.93

Classification Report of ROBERTA		
precision	recall	f1-score
0.66	0.53	0.58
0.93	0.88	0.90
0.91	0.95	0.93

Table 4.2.3 : BERT Models Result Analysis

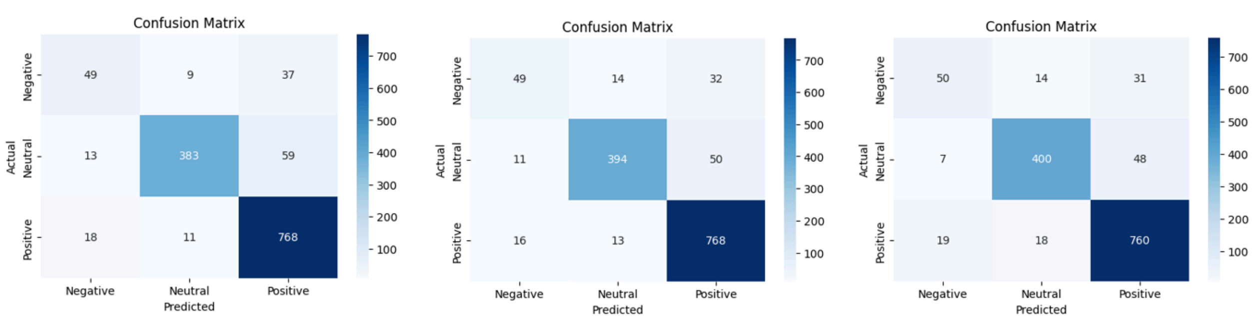

Fig 4.2.3 : BERT models Confusion Matrix

The BERT model is a strong pre-trained language model. By customizing its settings to certain tasks, BERT may be made to perform optimally in many applications such as named entity identification and sentiment analysis. Fine tuning BERT requires a lot of process but the improvement in results is paramount. In our study we find that from 0.63 % accuracy to an increase to 0.94% just by fine tuning the pre-trained model. Adjusting parameters, embeddings etc. are some of the required steps to gain this result.The increase of 0.31% accuracy is very crucial in the world of sentiment analysis.

4.3 Discussion

The proposed algorithm was able to predict emotions for given tweets. Sentiments labeled as positive, negative and neutral are predicted through accuracy of 0.94%. This is the result of training and fine tuning the BERT model. The base model has an accuracy of 0.29%. This occurs because it lacks understanding of the given data. The algorithm's limitation arises from its reliance on the presence of sentiment-specific words. Notably, certain tweets convey intense emotions, indicating a potential avenue for improvement. Future opportunities lie in the development of techniques for aspect-based sentiment analysis, considering the diverse characteristics within the text. Given the multilingual nature of social media data, sentiment analysis becomes intricate. Addressing this, the creation of a bilingual sentiment analysis method emerges as a crucial undertaking. Noteworthy is the current drawback of treating each opinion uniformly, warranting the exploration of methods to identify subjects and ascertain subject-level sentiments within messages.

CHAPTER 5

IMPACTS ON ENVIRONMENT, SOCIETY, SUSTAINABLE DEVELOPMENT

5.1 Impact on Environment

The primary environmental consequence of this study is from the high processing power needed for machine learning models, such as AutoML, which might result in higher energy use and carbon footprints. Better sentiment analysis tools, for example, may be able to offset this by encouraging more effective information processing and long-lasting decision-making processes.

5.2 Impact on Society

What makes the research valuable to society is its potential to enhance decision-making processes across a range of domains. Through the identification and evaluation of effective machine learning models, the project provides advertisers, policymakers, and businesses with valuable information on sentiment analysis. Improved sentiment analysis technology can lead to more public opinion understanding, better customer engagement, and more informed strategies. Ultimately, the application of sentiment analysis to quantify user attitudes and preferences will advance domains where society benefits from more accurate, data-driven decision-making.

5.3 Ethical Aspects

The proper use of sentiment analysis techniques is the primary ethical problem in this work. It is essential to respect individuals' privacy and obtain their agreement before using their data in datasets. Minimizing biases, addressing any unexpected repercussions, and maintaining transparency while developing models are crucial ethical imperatives, especially in sensitive domains such as the social sciences. Moreover, respecting moral standards in the quickly evolving domain of sentiment deciphering hinges on the proper model implementation, performance limitations and related social consequences.

5.4 Sustainability Plan

A key element of the research's sustainability plan is to promote relevance and long-term effects. To stay up to speed with research, keep an eye on breakthroughs in machine learning and sentiment analysis. Public availability of the datasets and coding will promote repeatability and future study.Collaborating with corporate partners will facilitate real-world implementations and provide insights for ongoing model enhancement. Regular updates and new research that addresses fresh challenges will preserve the research's value. In the ever-evolving area of sentiment analysis, engaging with the academic community through conferences and publications will foster a vibrant exchange of ideas and extend the impact and longevity of research.

CHAPTER 6

SUMMARY OF THE STUDY & FUTURE SCOPE

6.1 Summary of the Study

The comparison analysis assists practitioners in choosing the finest sentiment analysis tools by offering helpful insights about the advantages and disadvantages of each model. The findings show how sentiment analysis is developing and the necessity for advanced techniques and constant technological advancement. This paper addresses concerns and offers a thorough analysis of machine learning techniques, which adds to the scholarly discourse. It also provides helpful guidance for sectors whose decision-making relies on sentiment analysis. The related works use a variety of machine learning algorithms to demonstrate how different algorithms work and evaluate their performances . This study which made use of SVM and BERT demonstrated good accuracy.

6.2 Conclusions

This study presents an in-depth analysis of several sentiment analysis machine learning models from related studies. The necessity of selecting a model that meets the particular requirements of the application was made evident by significant problems and challenges.

The comparative study provides useful information about the benefits and drawbacks of each model, assisting practitioners in selecting the best sentiment analysis tools. The results demonstrate the evolution of sentiment analysis and the need for cutting-edge methods and ongoing technical development. This study contributes to the scholarly conversation by addressing issues and providing a complete overview of machine learning approaches. It also offers beneficial advice to industries where sentiment analysis is used to make decisions.

6.3 Further Study & future work

This paper examines the benefits and drawbacks of machine learning models, emphasizing the need of comprehending ensemble methods and hyperparameter modifications. It highlights the necessity of improving sentiment recognition techniques, with an emphasis on model predictions' interpretability and scalability to bigger datasets. The study also emphasizes how important it is to comprehend language nuances and how they affect model performance. Online environments are dynamic, thus they require constant improvement and modification. Through the identification of gaps and the provision of innovative possibilities for sentiment analysis techniques, the findings offer new paths for future study. Paths like hyper models, hyper tuning etc. are some of the near future for this study.The paper lays the foundations for future research into uncharted areas of machine learning applications in sentiment analysis and advanced sentiment analysis methodology.

REFERENCE

[1] Mahmud Isnan, Gregorius Natanael Elwirehardja, Bens Pardamean,Sentiment Analysis for TikTok Review Using VADER Sentiment and SVM Model,Procedia Computer Science,Volume 227, 2023

[2] Handan Cam, Alper Veli Cam, Ugur Demirel, Sana Ahmed,
Sentiment analysis of financial Twitter posts on Twitter with the machine learning classifiers, Heliyon, Volume 10, Issue 1, 2023

[3] Orestes Appel, Francisco Chiclana, Jenny Carter, Hamido Fujita,
A hybrid approach to the sentiment analysis problem at the sentence level,
Knowledge-Based Systems,Volume 108,2016,

[4] Sharma, A., & Dey, S. (2012, October). A comparative study of feature selection and machine learning techniques for sentiment analysis. In *Proceedings of the 2012 ACM research in applied computation symposium* (pp. 1-7).

[5]Hassan, A. U., Hussain, J., Hussain, M., Sadiq, M., & Lee, S. (2017, October). Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression.

[6] Bhagat, Abhishek and Sharma, Akash and Chettri, Sarat, Machine Learning Based Sentiment Analysis for Text Messages (June 20, 2020). International Journal of Computing and Technology, June 2020,

[7]Agarwal, B., Mittal, N., Agarwal, B., & Mittal, N. (2016). Machine learning approach for sentiment analysis. *Prominent feature extraction for sentiment analysis*

[8]Singh, Rameshwer, Rajeshwar Singh, and Ajay Bhatia. "Sentiment analysis using Machine Learning techniques to predict outbreaks and epidemics." *Int. J. Adv. Sci. Res* 3.2 (2018)

[9]Neethu, M. S., and R. Rajasree. "Sentiment analysis on twitter using machine learning techniques." *2013 fourth international conference on computing, communications and networking technologies (ICCCNT)*. IEEE, 2013.

[10]A. J. Nair, V. G and A. Vinayak, "Comparative study of Twitter Sentiment On COVID - 19 Tweets," *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India, 2021

[11] Q. Dong, T. Sun, Y. Xu, X. Xu, M. Zhong and K. Yan, "Network Public Opinion Sentiment Analysis based on Bert Model," *2022 IEEE 10th International Conference on Information, Communication and Networks (ICICN)*, Zhangye, China, 2022

[12] A. Topbaş, A. Jamil, A. A. Hameed, S. M. Ali, S. Bazai and S. A. Shah, "Sentiment Analysis for COVID-19 Tweets Using Recurrent Neural Network (RNN) and Bidirectional Encoder Representations (BERT) Models," *2021 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube)*, Quetta, Pakistan, 2021

[13] Li, M., Li, W., Wang, F. et al. Applying BERT to analyze investor sentiment in the stock market. Neural Comput & Applic 33, 4663–4676 (2021)

[14] Karabila, I., Darraz, N., EL-Ansari, A. et al. BERT-enhanced sentiment analysis for personalized e-commerce recommendations. Multimed Tools Appl 83, 56463–56488 (2024)

[15] N. A. Osman, S. A. M. Noah, and M. Darwich, "Contextual Sentiment Based Recommender System to Provide Recommendation in the Electronic Products Domain," International Journal of Machine Learning and Computing vol. 9, no. 4, pp. 425-431, 2019.

[16] D. Jiang, X. Luo, J. Xuan and Z. Xu, "Sentiment Computing for the News Event Based on the Social Media Big Data," in IEEE Access, vol. 5, pp. 2373-2382, 2017

[17] Vinodhini, G., & Chandrasekaran, R. (2017). A sampling based sentiment mining approach for e-commerce applications. In Information Processing & Management (Vol. 53, Issue 1, pp. 223–236). Elsevier BV.

[18] B. Roshanfekr, S. Khadivi and M. Rahmati, "Sentiment analysis using deep learning on Persian texts," 2017 Iranian Conference on Electrical Engineering (ICEE), Tehran, Iran, 2017

[19] Chauhan, A., & Mohana, R. (2023). Improving Bert Model Accuracy for Uni-modal Aspect-Based Sentiment Analysis Task. In Scalable Computing: Practice and Experience (Vol. 24, Issue 3, pp. 277–286). Scalable Computing: Practice and Experience

[20] Sallam, R. M., Hussein, M., & Mousa, H. M. (2022). Improving collaborative filtering using lexicon-based sentiment analysis. In International Journal of Electrical and Computer Engineering (IJECE) (Vol. 12, Issue 2, p. 1744)

Sentiment

ORIGINALITY REPORT

21%	17%	10%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	8%
2	Submitted to Daffodil International University Student Paper	2%
3	Qian Dong, Tingting Sun, Yan Xu, Xuguang Xu, Mei Zhong, Kai Yan. "Network Public Opinion Sentiment Analysis based on Bert Model", 2022 IEEE 10th International Conference on Information, Communication and Networks (ICICN), 2022 Publication	1%
4	www.researchgate.net Internet Source	1%
5	www.scpe.org Internet Source	1%
6	centaur.reading.ac.uk Internet Source	1%
7	Behnam Roshanfekr, Shahram Khadivi, Mohammad Rahmati. "Sentiment analysis using deep learning on Persian texts", 2017	<1%