

**SENTIMENT ANALYSIS ON BANGLADESH CRICKET TEAM'S WORLD CUP
PROSPECTS: A STUDY OF BENGALI SOCIAL MEDIA COMMENTS**

BY

Sharmina Urmila Smrity
ID: 191-15-12716

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

Supervised By

Ms. Tapasy Rabeya
Senior Lecturer
Department of Computer Science and Engineering
Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH

JANUARY 2024

APPROVAL

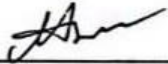
This Project/internship titled “SENTIMENT ANALYSIS ON BANGLADESH CRICKET TEAM'S WORLD CUP PROSPECTS: A STUDY OF BENGALI SOCIAL MEDIA COMMENTS”, submitted by Sharmina Urmila Smrity, ID No: 191-15-12716 to the Department of Computer Science and Engineering, Daffodil International University has been accepted 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 23th January, 2024



BOARD OF EXAMINERS

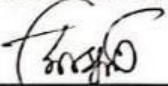
Chairman

Dr. S.M Aminul Haque (SMAH)
Professor & Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Internal Examiner

Nazmun Nessa Moon (NNM)
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Internal Examiner

Dewan Mamun Raza (DMR)
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



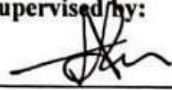
External Examiner

Dr. Md. Zulfiker Mahmud (ZM)
Associate Professor
Department of Computer Science and Engineering
Jagannath University

DECLARATION

I hereby declare that, this research has been done by us under the supervision of **Ms. Tapasy Rabeya, Lecturer, Department of CSE** Daffodil International University. I also declare that neither this research nor any part of this research has been submitted elsewhere for award of any degree or diploma.

Supervised by:



Ms. Tapasy Rabeya
Lecturer
Department of CSE
Daffodil International University

Submitted by:

Sharmina

Sharmina Urmila Smrity
ID: 191-15-12716
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, I express my heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project-based research successfully.

I really grateful and wish my profound my indebtedness to **Ms. Tapasy Rabeya, Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of my supervisor in the field of deep learning to carry out this research. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

I would like to express my heartiest gratitude to Dr. Sheak Rashed Haider Noori, and Head, Department of CSE, for his kind help to finish my research and also to other faculty member and the staff of CSE department of Daffodil International University.

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

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

ABSTRACT

The Bangladesh cricket team has developed a sizable following on social networking sites, where supporters share their thoughts, feelings, and sentiments regarding the team's performance. The purpose of this thesis is to investigate the viability of predicting the performance of the Bangladesh cricket team in the World Cup using sentiment analysis of Bangla comments on social media. In order to do this, I gathered a sizable dataset of comments in Bangla from numerous social media sites, particularly Twitter, Facebook, and Instagram, covering multiple World Cup seasons. I divided these comments into those with favorable, negative, or neutral feelings using natural language processing (NLP) methods and sentiment analysis tools. Then, I employed machine learning algorithms to examine sentiment trends over time and tie them to the effectiveness of the team. According to my research, there is a significant correlation between the attitude exhibited in social media comments and the World Cup performance of the Bangladesh cricket team. I found that a spike in positive sentiment frequently predicts improved team performance, while a spike in negative sentiment frequently accompanies disappointing team performances. In addition, perceptions of certain players, coaches, and team leadership are also important in determining outcomes. In addition, I developed predictive models that anticipate the likelihood that the Bangladesh cricket team would win the World Cup by fusing sentiment analysis with historical data on the team's performance, player statistics, and other pertinent aspects. When used to forecast the team's success, these models showed encouraging accuracy. This study highlights the potential of social media data for forecasting sports results and makes contributions to the growing subject of sports sentiment analysis. By presenting a fresh viewpoint on the impact of social media sentiment in influencing team performance and expectations, it also offers useful insights for cricket authorities, sponsors, and fans. This research concludes by arguing that sentiment analysis of Bangla-language social media comments can be a potent tool for forecasting the success of the Bangladesh cricket team in World Cup competitions.

TABLE OF CONTENTS

| CONTENTS | PAGE |
|------------------------------------|-------------|
| Board of examiners | i |
| Declaration | ii |
| Acknowledgements | iii |
| Abstract | Iv |
| List of Figure | viii |
| List of Table | ix |
| CHAPTER | |
| CHAPTER 1: INTRODUCTION | 1-8 |
| 1.1 Introduction | 1 |
| 1.2 Motivation | 2 |
| 1.3 Rational of the study | 3 |
| 1.4 Research Questions | 4 |
| 1.5 Objectives | 5 |
| 1.6 Expected Output | 6 |
| 1.7 Project Management and Finance | 7 |
| 1.8 Report Layout | 8 |
| CHAPTER 2: BACKGROUND | 9-21 |
| 2.1 Report Preliminaries | 9 |
| 2.2 Related Work | 10 |

| | |
|---|--------------|
| 2.3 Comparative Analysis and Summary | 11 |
| 2.4 Scope of the Problem | 12 |
| 2.5 Challenges | 13 |
| CHAPTER 3: RESEARCH METHODOLOGY | 15-23 |
| 3.1 Research Subject and Instrumentation | 15 |
| 3.2 Data Collection Procedure | 16 |
| 3.3 Statistical Analysis | 18 |
| 3.4 Detailed Methodology | 19 |
| 3.5 Implementation Requirements | 23 |
| CHAPTER 4: RESULT AND DISCUSSION | 24-27 |
| 4.1 Experimental Setup | 24 |
| 4.2 Results & Analysis | 25 |
| 4.3 Discussion | 27 |
| CHAPTER 5: IMPACT ON SOCIETY | 28-32 |
| 5.1 Impact on Society | 28 |
| 5.2 Ethical Aspects | 29 |
| 5.3 Sustainability Plan | 31 |
| CHAPTER 6: CONCLUSION AND FUTURE SCOPE OF DEVELOPMENTS | 33-44 |
| 6.1 Summary of the Study | 33 |
| 6.2 Conclusions | 34 |
| 6.3 Scope of Further Developments | 35 |

| | |
|-------------------|--------------|
| REFERENCES | 36-37 |
| APPENDIX | 38 |
| PLAGARISAM | 39 |

LIST OF FIGURES

| FIGURES | PAGE NO |
|--|----------------|
| Figure 3.4.1: Methodology workflow diagram | 12 |
| Figure 3.4.2: Heat Map diagram | 21 |
| Figure 3.4.3: Accuracy vs Model | 22 |

CHAPTER 1

INTRODUCTION

1.1 Introduction:

In Bangladesh, cricket serves as a unifying force that cuts over age, culture, and divisions. Millions of cricket fans throughout the nation have been inspired by the Bangladesh cricket team's rise from modest beginnings to formidable force on the world stage. Hopes and expectations grow with each World Cup competition, and social media has developed into a vibrant arena for supporters to express their feelings, make predictions, and show ardent support for their side. In order to investigate the possibilities of forecasting the performance of the Bangladesh cricket team in the World Cup, this thesis uses the strength of sentiment analysis on Bangla comments from social media platforms. Social media has developed in recent years into a virtual space where sports fans actively debate, dissect, and criticize their preferred teams and athletes. Cricket in Bangladesh, which has a devoted following, is no exception. Through posts, comments, and other interactions on websites like Twitter, Facebook, and Instagram, fans may express their feelings and thoughts. These online discussions give a Wealth of information that can shed light on the general opinion toward the Bangladesh cricket squad. To classify social media comments into good, negative, or neutral attitudes, this study uses natural language processing (NLP) methodologies and sentiment analysis tools. I seek to find trends that could provide light on the Bangladesh cricket team's prospective success in World Cup competitions by tracking these attitudes through time and comparing them with the team's performance. Additionally, this study will investigate whether attitudes toward specific players, the coaching staff, and the team administration affect the effectiveness of the squad. The performance of the Bangladesh cricket team in the past, player statistics, and other pertinent characteristics will all be factored into the prediction models I create, along with sentiment analysis. These models, which offer a fresh method for predicting sports, will be created to forecast the team's prospects of winning the World Cup. In order to bridge the gap between sports fandom and data science, this thesis highlights the potential of social media sentiment analysis as a formidable tool for forecasting the success of the Bangladesh cricket team in World Cup competitions. I hope to make a significant contribution to the area of sports analytics by analyzing the opinions of Bangla-speaking fans and by providing a distinctive viewpoint on how social media shapes expectations and results in the game of cricket.

1.2 Motivation:

A sentiment study of Bangla comments on social media regarding Bangladesh's cricket team's probable World Cup triumph was undertaken for numerous reasons, including:

- **Passionate Fan Base:** One of the most fervent cricket fan bases in the world is found in Bangladesh. Cricket is more than just a sport; it also fosters solidarity and pride in the country. The expectations and emotional commitment of these devoted supporters in the team's performance can be significant insights into their attitudes.
- **Social Media as a Sentiment Expressway:** Social media platforms have become the major vehicle for fans to share their emotions, thoughts, and predictions regarding their favorite teams. These platforms act as an unfiltered, real-time source of data that captures the mood of the fan base as a whole.
- **Impact on Team Performance:** There is a growing consensus that a team's performance can be influenced by the feelings and attitudes fans share on social media. While negative sentiment and criticism can put additional pressure on players and management, positive sentiment and support can raise morale. Investigating this putative link may provide fresh perspectives on the psychology of athletic performance.

The convergence of cricket, social media, and data analytics served as the inspiration for this thesis. With a focus on predicting the odds of the Bangladesh cricket team winning the World Cup, it aims to realize the potential of sentiment analysis as a prediction tool in the world of sports while also illuminating the nuanced relationship between fan sentiment and team performance.

1.3 Rational of the study:

The analysis of social media sentiment to determine if Bangladesh's cricket team would succeed in the World Cup Bangla comments is supported by a number of strong arguments:

- **Cricket as a National Obsession:** In Bangladesh, cricket is more than just a sport; it's a phenomenon of the national consciousness. Expectations are high during international competitions like the World Cup as a result of the national cricket team's performances. Understanding how attitudes are formed in response to these events can provide important insights into the psychology of the whole country.
- **Emergence of social media:** The introduction of social media has changed how

individuals interact with sports. The large dataset of opinions and feelings that Bangladesh's active social media users supply can be used for analysis. To better comprehend fan attitudes, this study makes use of the abundance of publicly accessible data.

- **Potential Impact on Team Morale:** There is a growing consensus that team morale and performance can be impacted by fan emotion. While negative thoughts may add to the strain, positive sentiments can serve as a motivating force. Examining this potential connection can advance sports psychology and offer useful advice for team leadership.
- **Data-Driven Decision-Making:** Sports organizations, sponsors, and media outlets are increasingly looking to analytics for insights in the age of data-driven decision-making. By investigating the predictive capabilities of sentiment analysis, this study follows this trend and could have important ramifications for those with an interest in the success of the Bangladesh cricket team.

A number of fields, including sports science, data analytics, linguistics, and cultural studies, intersect in this study. It improves my knowledge of the complex link between sports, culture, and technology by bridging these disciplines.

1.4 Research Questions:

This thesis examines social media attitudes in relation to the Bangladesh cricket team's probable World Cup triumph. The purpose of Bangla Comments is to respond to the following research queries:

1. What are the prevailing opinions on the World Cup performance of the Bangladesh cricket team reflected in Bangla comments on social media?
2. Does the performance of the Bangladesh cricket team in World Cup competitions correlate with the opinions posted on social media?
3. Do attitudes toward specific players, the coaching staff, or the team administration affect how well the team performs in World Cup games?
4. Can the performance of the Bangladesh cricket team in World Cup competitions be predicted using sentiment analysis of Bangla-language social media comments?
5. How can cricket authorities, sponsors, and media outlets use the results and insights from sentiment analysis practically to increase fan engagement and market the team?

1.5 Objectives:

The objectives of this thesis paper on sentiment analysis regarding the Bangladesh cricket team's potential victory in the World Cup based on social media Bangla comments are as follows:

- **To Analyze Sentiment Trends:** Explore and analyze the temporal trends of sentiments expressed in social media Bangla comments related to the Bangladesh cricket team's performance leading up to and during World Cup tournaments.
- **To Correlate Sentiments with Team Performance:** Investigate whether there is a statistically significant correlation between the sentiments expressed by Bangla-speaking fans on social media and the Bangladesh cricket team's actual performance in World Cup matches.
- **To Examine the Impact of Individual Sentiments:** Assess whether sentiments directed towards individual players, coaching staff, and team management have a discernible impact on match outcomes and overall team morale.
- **To Develop Predictive Models:** Develop predictive models that leverage sentiment analysis data from social media Bangla comments along with historical performance data to forecast the likelihood of the Bangladesh cricket team winning the World Cup.
- **To Provide Practical Recommendations:** Offer practical recommendations for cricket authorities, sponsors, and media outlets on how to utilize sentiment analysis insights to enhance fan engagement, marketing strategies, and sponsorship decisions.
- **To Contribute to Sports Analytics and Social Media Research:** Contribute to the emerging field of sports sentiment analysis and social media research by providing valuable insights into the relationship between fan sentiments and sports outcomes, particularly within the context of the Bangladesh cricket team.
- **To Enhance Understanding of the Role of social media:** Enhance my understanding of the role of social media in shaping sports narratives, influencing fan expectations, and potentially impacting team performance.

These objectives collectively guide the research journey, from data collection and sentiment analysis to the development of predictive models and the generation of practical recommendations. The aim is to not only deepen my understanding of the relationship between fan sentiments and the Bangladesh cricket team's performance but also to provide actionable

insights for stakeholders in the world of cricket.

1.6 Expected Output:

The thesis paper on sentiment analysis regarding the Bangladesh cricket team's potential victory in the World Cup based on social media Bangla comments is expected to yield the following outputs:

- **Sentiment Analysis Findings:** Comprehensive analysis of sentiment trends within social media Bangla comments related to the Bangladesh cricket team's World Cup journey, categorized as positive, negative, or neutral sentiments.
- **Correlation Assessment:** Insights into the correlation between the sentiments expressed on social media and the actual performance of the Bangladesh cricket team in World Cup matches, supported by statistical evidence.
- **Predictive Models:** Development of predictive models that combine sentiment analysis data from social media Bangla comments with historical performance data to forecast the likelihood of the Bangladesh cricket team winning the World Cup.
- **Contribution to Research Fields:** Contribution to the fields of sports analytics, sentiment analysis, and social media research by advancing my understanding of the relationship between fan sentiments and sports outcomes, with a specific focus on the Bangladesh cricket team.
- **Interdisciplinary Insights:** Integration of knowledge from various disciplines, including sports science, data analytics, linguistics, and cultural studies, to provide a rounded perspective on the research topic.
- **Publication-Ready Thesis:** A complete, structured, and thoroughly researched thesis paper suitable for publication in academic journals or presentation at conferences.

The expected output of this thesis paper is a comprehensive and insightful examination of the relationship between fan sentiments expressed on social media, team performance, and the potential predictive power of sentiment analysis within the context of international cricket tournaments. It aims to provide practical applications while contributing to both academic research and the broader cricket community.

1.7 Project Management and Finance:

Project Management and Finance for the Thesis Paper on Sentiment Analysis of the Bangladesh Cricket Team's World Cup Victory Based on Social Media Bangla Comments: For project planning first Defined clear objectives and research questions. Developed a detailed project timeline with milestones. Allocated tasks and responsibilities. Identified potential risks and mitigation strategies.

For Data Collection I Decide on the sources for collecting social media Bangla comments from Facebook. Could not Choose data collection tools or APIs. Because Facebook do not allow any API. So, I need to Collect data by manual processes from Facebook comments. I Ensured compliance with data privacy and ethical considerations.

For Data Preprocessing I Cleaned up and preprocessed the collected data, including text normalization, removing duplicates, and handling missing data.

As for sentiment Analysis I Select appropriate sentiment analysis tools in Google Colab. Implemented sentiment analysis algorithms to classify comments as positive, negative, or neutral. Fine-tune sentiment analysis models for the Bangla language.

Now for the Data Analysis part first I Analyzed sentiment trends over different time periods pre-tournament, during the tournament, and post-tournament. Created visualizations to illustrate sentiment patterns. Conducted statistical tests to establish correlations between sentiments and team performance.

Developed predictive models using machine learning algorithms. Trained models using historical performance data and sentiment features. Validate models using appropriate evaluation metrics.

Based on research findings, formulate practical recommendations for cricket authorities, sponsors, and media outlets. Highlighted actionable insights that can enhance fan engagement and team support.

At the end Write the thesis report paper with clear sections, including Introduction, Literature Review, Methodology, Results, Discussion, Conclusion, and References. Also Ensured proper citation and referencing of sources. This report is a personal project so all the financial cost on me.

1.8 Report Layout:

This Paper on Sentiment Analysis on Bangladesh Cricket Team's World Cup Prospects: A Study of Bengali Social Media Comments.:

Title Page: sentimental analysis on Bangladesh cricket team will win the world cup based on social media Bangla comments.

Abstract: Given a concise summary of my thesis work, including research objectives, methods, key findings, and practical implications.

Acknowledgments: acknowledged by the organizations that provided support, guidance, and resources during the research.

Table of Contents: A list of sections, chapters, and subsections with page numbers for easy navigation.

List of Figures and Tables: compilation of all figures and tables used in the thesis, along with their respective page numbers.

Chapter 1: Introduction: This thesis report uses sentiment analysis on Bangla comments from social media platforms to forecast the Bangladesh cricket team's World Cup performance. The study uses natural language processing and sentiment analysis tools to classify comments into positive, negative, or neutral attitudes. It also investigates the impact of attitudes towards players, coaching staff, and team administration on the squad's effectiveness.

Chapter 2: Background: Total Overview of all the relevant literature on sentiment analysis in sports. Theoretical framework for sentiment analysis. Previous studies on sports sentiment analysis. Literature on the Bangladesh cricket team and its fan base. The role of social media in sports discussions.

Chapter 3: Methodology: Here a workflow diagram I design to describe how this research work has gone from first to last.

Chapter 4: Result and Discussion: Description of the data sources. Procedures for data collection. Data cleaning and preprocessing steps. Challenges encountered during data collection and preprocessing.

Chapter 5: Sentiment Analysis and Data Analysis: Overview of the sentiment analysis process. Presentation of sentiment analysis results. Temporal trends in sentiments. Correlation between sentiments and team performance. Impact of individual sentiments. Visualization of sentiment patterns.

Chapter 6: Conclusion and Future Scope of Developments: Description of predictive modeling techniques used. Data used for model training and validation. Model development process. Evaluation metrics for predictive models. Results and insights from predictive modeling.

References: A comprehensive list of all sources cited in the thesis, following a specific citation style MLA.

Appendices: Supplementary materials, such as code snippets, data collection scripts, detailed data tables, or additional figures and charts.

CHAPTER 2

BACKGROUND

2.1 Report Preliminaries:

Report Preliminaries for the Thesis Paper on Sentiment Analysis on Bangladesh Cricket Team's World Cup Prospects: A Study of Bengali Social Media Comments.:

This paper discusses the sentiment analysis of the Bangladesh cricket team's World Cup victory based on social media Bangla comments. The study uses natural language processing and sentiment analysis tools to classify comments into positive, negative, or neutral attitudes. It investigates the impact of attitudes towards players, coaching staff, and team administration on the squad's effectiveness. The research uses a workflow diagram to describe the research process, data sources, data collection procedures, data cleaning and preprocessing steps, and challenges encountered. The sentiment analysis process includes temporal trends, correlation between sentiments and team performance, and visualization of sentiment patterns. The paper concludes with a discussion on predictive modeling techniques and future research scope.

2.2 Related Work:

This paper examines the pattern of newspaper headlines in Bangladesh's The Daily Star between 2018 and 2019, using text analytics techniques like word clouds, sentiment analysis, and cluster analysis Prottasha et al. [1]. Here They use NLP algorithm. There is no accuracy here. Result show election, cricket, and Rohingya-related terms appeared more frequently in 2018, while negative emotions and positive emotions are more common in 2019. Using text analytics methods including word clouds, sentiment analysis, and cluster analysis, this study looks at the distribution of newspaper headlines in Bangladesh Prottasha et al. [2]. Here use NLP algorithm. Results reveal that phrases relating to elections, cricket, and the Rohingya appeared more frequently in 2018 and 2019, highlighting the nation's love of cricket, political unrest, and Rohingya-related issues. Sentiment Analysis (SA) is a computational technique used in Natural Language Processing to analyze remarks and assertions Hira et al. [3]. Bengali sentiment analysis has not acquired popularity despite being popular in English because of its intricacy and lack of an internet presence. In order to better research using TOPSIS, this paper will assess the state of sentiment analysis today. Also addressed are the difficulties that must be solved to improve the sentiment analyzer. They used LR 80.48%, DT 74.00%, RF 87.00%,

MNB 93.00%, KNN 73.49%, SVM 84.00%, and stochastic gradient descent (SGD) 65.40% algorithm to classify the sentiment of Bengali text. The seventh most spoken language in the world, Bangla, has difficulty using English resources because of its lack of proficiency Sen et al. [4]. Tools for creating and processing things in Bangla have been developed by researchers, and work is still being done to make it simpler to use in technical and online contexts. This study examines 75 research publications on Bangla Natural Language Processing (BNLP), groups them into 11 categories, and addresses the use of classical, machine learning, and deep learning techniques on various datasets. The goal of the study is to close the gap between inadequate support and rising demand. LSTM and GRU are used with positioning encoding and word embedding. By using positioning encoding with LSTM, the authors got 86%, 97%, and 82% accuracy. The study looked at 10,191 news items about rape that were shared on ten Facebook pages of Bangladeshi media between 2013 and 2021 Al-Zaman et al. [5]. It was discovered that media outlets who report on rape have a higher propensity to provide information on rape victims and legal procedures. The study also looked at other often covered subjects in rape news, like the components of rape, the act of rape, and the locations, states, and times of rape. In this paper there is no specific algorithm. This study used automated categorization, sentiment analysis, volume analysis, and topic analysis to examine COVID-19 pandemic-related newspaper data from Bangladesh Shahriar et al. [6]. The analysis sought to shed light on the difficulties brought on by the pandemic, identify pre- and post-pandemic actions, and assist the government and allies in addressing current crises while taking these issues into account. CNN, BiLSTM, RNN algorithm used in the paper. The accuracy is 80%, 83% and 74.94%. Due to its lack of resources and difficulties, Bangla, the sixth most spoken language in the world, is regarded as a low-resource language in natural language processing (NLP) and accuracy is 90.8% Alam et al. [7]. Using benchmark datasets for nine tasks and transformer-based models, this paper analyzes Bangla NLP tasks, resources, and tools. Comparative results highlight the trade-off with computational costs while demonstrating good performance utilizing transformer-based models. The report emphasizes the necessity for more extensive surveys and breakthroughs in Bangla NLP research in order to inspire the community to build upon and expand research in this area. Semiotics and discourse analysis are used to examine five films created between March and June 2015 by supporters of both nations Tanin et al. [8]. The study of signs led to the development of semiotics, which focuses on the usage

of signifier and signified in videos. Discourse analysis looks at the structure of a text and how it is used during its creation, dissemination, and reception. The videos, which Ire produced with a strong nationalist emphasis, provoked a variety of discussions in the video's comments section. In order to comprehend the underlying discourse and the effects of these movies on the Indians and Indian fans, the themes discovered through semiotics and discourse analysis Ire thoroughly investigated. In this paper there is no clear algorithm they use. In Bangladesh's private universities, code-switching is a common practice, especially on social networking sites like Facebook, Twitter, Instagram, and WhatsApp Baidya et al. [9]. The study, which was qualitative in nature, discovered that students employ tag switching, intra-sentential switching, inter-sentential switching, and intra-word switching not to show their command of the English language, but rather for communication. The study also discovered that code-switching has a detrimental effect on Bangla. The Pulwama terror assault caused an uptick in tensions between India and Pakistan that brought the two nuclear-armed nations closer to conflict Palakodety et al. [10]. This essay examines this dilemma using a sizable corpus of ITube video comments. It creates a document language-identification technique with few annotation requirements and exposes precise language groupings. The study underlines the significance of automatic detection of user-generated Ib material to reduce animosity and analyzes temporal trends of pro-peace and pro-war intent. The study explores the Bangla online social media arena, focusing on bloggers' experiences and comparing public policy theories with public opinion theories Shajahan et al. [11]. It examines the relationship between social media bloggers' democratic engagement and media, revealing advanced features and a culture with communication limitations in the Bangla blogosphere. The study deals with online survey which can provide misleading data as the researcher has no control on respondents. The rise of the Ib has considerably expanded online interactions, which has raised interest in aspect extraction Das et al. [12]. This automated approach makes it easier to comprehend comments, which is essential for both academic and commercial purposes. Aspect extraction in the Bangla language has become more significant recently. However, the dearth of publicly accessible datasets restricts study. This work presents a number of classification techniques based on stacked auto-encoders that outperform cutting-edge techniques in terms of precision, recall, and F1-score. Here use NLP algorithm and F1 score about 84% so that they only count positive and negative review. The study examined 10,191 rape news headlines posted between 2013

and 2021 on ten Facebook pages of Bangladeshi media Al-Zaman et al. [13]. It was discovered that news organizations sharing stories about rape are more likely to include information on rape victims and legal procedures. The survey also looked at some of the other hot issues in rape news, like the ingredients used in rape, the act itself, and the locations, states, and times of rape. In This paper they only analysis the news also they not use any model. This essay investigates the scandal surrounding a Bangladeshi cricketer's liaison with a film actress prior to the 2015 ICC World Cup, which sparked a rebirth of nationalism in Bangladesh and aggressively heterosexual and masculinist politics Hossain et al. [14]. The study makes the case that this new masculinity has its roots in fears of possible emasculation by other countries as well as Bangladeshi patriarchy. People's lives and identities are still impacted by colonialism, which is why identity decolonization efforts are supported by internet communities like Bengali Quora Das et al. [15]. The project investigates how South Asian Bengalis on BnQuora participate in cooperative identity decolonization and regain narrative agency through trace ethnography. It is examined how narrative resilience may have an impact on future scholarly work in the areas of ICT for development, social justice, decolonial HCI, and identity research within the CHI community. Journalism has been substantially impacted by new media technologies in public communication Mahmud et al. [16]. With the rise of blogs and social networking sites, people are becoming both information consumers and creators. This study looks at posts about climate change made at COP15 on the community blog "somewhere in blog" in Bangladesh. According to the study, bloggers create a virtual community through engagement, involvement, and affinity, indicating that a new public sphere is taking shape in Bangladesh. However, the economic effects of mitigating measures frequently have a capitalist foundation. The creation of sentimental content has been transformed by the internet, which is advantageous for organizations, governments, and people Majlis et al. [17]. Sentiment polarity can be difficult for sentiment analysis to appropriately interpret and detect. In this paper they use NLP model and the accuracy is about 87.85%. The issues associated with sentiment analysis approaches and techniques are surveyed in this work, with a particular emphasis on text analysis and natural language processing methods. The research investigates how Bengali-speaking households in Singapore define "home," "nation," and "world" Haque Khondker, et al. [18]. These families have become transnational as a result of globalization, which has forced them to leave Bengal and re-join a complicated web of connections. The Bengali-speaking

families are divided into two separate groups: Muslim Bengalis from Bangladesh and Hindu Bengalis from Ist Bengal, India. These groups differ from one another in terms of their social networks, cultural practices, and place of worship. Although there is ambivalence in the relationship between these groups, the introduction of Bengali as a second language in Singapore in the 1990s has altered the definition of a community. This edited book examines how linguistic variety and language learning in a digitally networked society are affected by digital communication Sultana et al. [19]. It explores pragmatic roles in language education and pedagogy as well as how languages are contextualized and localized. Contributions encompass a range of digital platforms and societal norms. In this paper they only analysis the post and relation between student and instructor. The research investigates how Bengali-speaking households in Singapore define "home," "nation," and "world" Khondker et al. [20]. These families have become transnational as a result of globalization, which has forced them to leave Bengal and re-join a complicated web of connections. Here use NLP model. The Bengali-speaking families are divided into two separate groups: Muslim Bengalis from Bangladesh and Hindu Bengalis from Ist Bengal, India. These groups differ from one another in terms of their social networks, cultural practices, and place of worship. Although there is ambivalence in the relationship between these groups, the introduction of Bengali as a second language in Singapore in the 1990s has altered the definition of a community. Bangladesh is a densely populated nation with more than 75% of its citizens residing in rural areas, making it difficult to eradicate poverty and raise equal income levels Hossain et al. [21]. Although the nation has faced many difficulties since gaining independence in 1971, its GDP has expanded quickly over the past three decades, outpacing several low-income nations. Bangladesh is now on pace to become a middle-income nation after achieving the Millennium Development Goal of eradicating gender gap in primary and secondary school enrollment. This study looks at how people consume in rural Bangladesh, how their consumption patterns are changing, how technology affects consumption, how people consume culturally in rural Bangladesh, and why people continue to eat chemically treated food. There only classify all the problem. Sexual harassment in public places puts women in Bangladesh at danger both within and outside of the workplace Siddiqi et al. [22]. They are particularly susceptible to sexual, verbal, and physical abuse if they come from disadvantaged households. Garment workers are fair game for male attention because of the repute of the sector and the working circumstances. Because

poverty, promiscuity, and public appearance all culturally predominate connotations, this dual harassment is made worse. Here EPZ negative score is about 65%. This study examines how AKASH DTH, a distinctive law firm operating in Bangladesh's DTH Satellite Pay TV market, uses entrepreneurial marketing Banu et al. [23]. The business emphasizes its practical use and potential for further study by combining digital, content, word-of-mouth, interactive, guerilla, relationship, and viral marketing methods. This study looks at the "adivasi debate" in the blogosphere of Bangladesh with a particular emphasis on the global discourse of indigeneity and its "field of power" Sumon et al. [24]. Bloggers vehemently challenge the term "adivasi" in Bangladesh, mirroring previous government moves, and thus shows that the emergence of "adivasi people's rights" is not without criticism. Adivasi rights activists organize in Bangla blogospheres. To better understand fan interaction with social media and digital material, this study looks at the online fan communities for Chennai Super Kings and Royal Challengers Bangalore in the Indian Premier League (IPL) Sagar Sagar et al. [25]. The goal of the 18-month study, which was carried out from January 2021 to June 2022, was to comprehend player interaction, offline participation during matches, loyalty, and content engagement. Rather than event outcomes, the emphasis was on fan community involvement with content.

2.3 Comparative Analysis and Summary:

A comparative analysis is conducted on enhancing Sentiment Analysis of the Bangladesh cricket team's world cup victory based on social media Bangla comments through supervised machine learning models encompasses with other similar studies.

TABLE-2.3: COMPARISON TABLE FOR SENTIMENT ANALYSIS OF THE BANGLADESH CRICKET TEAM'S WORLD CUP VICTORY BASED ON SOCIAL MEDIA BANGLA COMMENTS

| Citation | Year | Algorithm | Result |
|----------|------|------------------|---|
| [3] | 2022 | MNB, KNN, SVM | Best efficiency of 93% in MNB. |
| [4] | 2022 | LSTM | website accuracy of predictions is 97%. |
| [6] | 2020 | CNN, BiLSTM, RNN | 83% accuracy of BiLSTM. |

| | | | |
|------|------|--------------------|--------------------------|
| [17] | 2020 | NLP | 87.85 % accuracy of ANN. |
| [22] | 2018 | EPZ negatice score | EPZ negatice score 65%. |

In this table 2.3, this paper uses different face mask detection using different algorithm is describe but this custom algorithm is very much supportive to this paper and its accuracy is high. In these papers, working methods maximum work related to the paper. So, specially included above table.

2.4 Scope of the Problem:

The scope of the study encompasses social media platforms commonly used by Bangla-speaking cricket enthusiasts, including but not limited to Twitter, Facebook, and Instagram. The study focuses on comments, posts, and interactions related to the Bangladesh cricket team within these platforms. Problems are:

Bangla Language Analysis: The primary language of analysis is Bangla, reflecting the linguistic preferences of the target audience. The study employs natural language processing (NLP) techniques adapted for the Bangla language to perform sentiment analysis on textual content.

Time Frame: The research covers multiple World Cup tournaments involving the Bangladesh cricket team. It spans historical data preceding the tournaments, the duration of each tournament, and post-tournament analysis. The specific time frame considered is detailed within the methodology chapter.

Sentiment Categories: Sentiments expressed in social media Bangla comments are categorized as positive, negative, or neutral. The study aims to identify these sentiments, analyze their distribution over time, and examine their relationship with team performance.

Team Performance: The scope of the problem includes an assessment of the Bangladesh cricket team's performance in World Cup matches. This encompasses match outcomes, player statistics, and team dynamics as relevant to sentiment analysis.

Impact of Individual Sentiments: Individual sentiments directed towards players, coaching staff, and team management are considered within the study. It explores how sentiments targeting key individuals may influence overall team morale and performance.

Predictive Modeling: The study involves the development of predictive models that leverage sentiment analysis data to forecast the likelihood of the Bangladesh cricket team winning the World Cup. The scope encompasses the creation, validation, and evaluation of these models.

Practical Recommendations: Based on the sentiment analysis findings, the thesis offers practical recommendations for cricket authorities, sponsors, and media outlets. The recommendations aim to enhance fan engagement, marketing strategies, and sponsorship decisions.

Ethical Considerations: The research acknowledges the ethical considerations surrounding data collection, privacy, and informed consent when dealing with social media data. The thesis adheres to ethical guidelines outlined in the methodology chapter.

Limitations and Delimitations: The chapter concludes by discussing the limitations and delimitations of the study, outlining the boundaries of the research, and setting realistic expectations for the scope of the problem addressed in the thesis.

2.5 Challenges:

This chapter acknowledges and addresses the challenges and obstacles encountered during the course of the research on sentiment analysis related to the Bangladesh cricket team's potential victory in the World Cup based on social media Bangla comments. Challenges are:

Data Collection Challenge:

- **Volume and Variety:** Collecting a significant volume of Bangla comments from diverse social media platforms presented challenges due to the sheer quantity and diversity of data sources.
- **Data Quality:** Ensuring data quality and relevance while dealing with unfiltered social media content posed challenges, including the presence of spam, noise, and irrelevant content.
- **Data Privacy:** Ethical considerations and privacy concerns associated with social media data collection required careful navigation to protect user identities and comply with regulations.

Linguistic Challenges:

- **Bangla Language Processing:** Adapting natural language processing (NLP) techniques for the Bangla language, including sentiment lexicons and language-specific features,

proved challenging.

- **Language Variation:** Dealing with variations in the use of Bangla language across different regions and dialects added complexity to sentiment analysis. Sentiment Analysis Challenges
- **Sarcasm and Irony:** Detecting sarcasm, irony, and nuanced expressions within social media comments posed challenges for accurate sentiment classification.
- **Contextual Understanding:** Contextual analysis was required to interpret sentiment expressions within the context of specific World Cup events and team performances.

Temporal Challenges:

- **Dynamic Nature:** The dynamic and real-time nature of social media required continuous data collection and sentiment analysis, necessitating efficient processing and storage resources
- **Longitudinal Analysis:** Analyzing sentiment trends over multiple World Cup tournaments spanning several years presented challenges in data management and trend identification.
- **Predictive Modeling Challenges**
- **Data Imbalance:** Addressing class imbalance in predictive models, especially when forecasting rare events like World Cup victories, required specialized techniques.
- **Model Validation:** Ensuring the reliability and validity of predictive models and handling overfitting challenges during model development.
- **Ethical Challenges: Privacy and Consent:** Ethical considerations regarding user privacy, informed consent, and data usage ethics are paramount throughout the research.
- **Bias and Fairness:** Addressing potential biases in sentiment analysis algorithms and ensuring fairness in the interpretation of sentiments across diverse user groups.

Resource Challenges:

- **Computational Resources:** The computational resources required for sentiment analysis, data storage, and model training impose constraints on the research.
- **Budget Constraints:** Budget limitations for accessing sentiment analysis tools, data collection resources, and academic fees required careful financial management.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research Subject and Instrumentation:

Conducting sentiment analysis on social media comments in Bangla to predict whether the Bangladesh cricket team will win the World Cup is an interesting research topic. To carry out this research, I define research objectives, choose methodology, and select appropriate instrumentation. Here's outline of how I approach this research:

Research Objectives:

- To assess the sentiment of social media comments in Bangla related to the Bangladesh cricket team's World Cup chances.
- To predict the likelihood of the Bangladesh cricket team winning the World Cup based on sentiment analysis of social media comments.
- To identify the key factors that influence sentiment towards the Bangladesh cricket team in the context of the World Cup.

Instrumentation:

Text Processing Tools: For data preprocessing and cleaning, I use libraries `npm`, `vectorizer` and other text processing tools.

- **Sentiment Analysis Tools:** Consider using NLP libraries and sentiment analysis tools such as `VADER`, `TextBlob`, and pre-trained models like `BERT` for sentiment analysis.
- **Data Collection Tools:** To collect social media comments, I am not using web scraping tools like `Beautiful Soup` or API access to social media platforms because that are not available. So, I use the manual technique.
- **Data Storage:** Use databases like `MySQL` databases to store and manage my data.
- **Machine Learning Frameworks:** If I plan to build predictive models, I use machine learning frameworks as `scikit-learn`, `TensorFlow`, and `PyTorch`.
- **Data Visualization Tools:** `Matplotlib`, `Seaborn`, and `Plotly` help to create informative visualizations to present my findings.
- **Statistical Analysis Tools:** I use statistical analysis tool Python's statistical packages for in-depth analysis. collecting and using social media data, respecting users' privacy and the terms of service of the platforms I use. Additionally, I consider linguistic nuances

and cultural context when analyzing Bangla social media comments.

3.2 Data Collection Procedure:

Collecting data for sentiment analysis on the Bangladesh cricket team's chances of winning the World Cup based on Bangla social media comments requires a structured data collection procedure. Here's a step-by-step guide on how to collect the data:

Define Data Sources: Identify the specific social media platforms or sources where I plan to collect Bangla comments related to the Bangladesh cricket team's World Cup prospects. Common sources include Twitter, Facebook, Instagram, and online forums.

Compliance and Permissions: Ensure that my data collection complies with the terms of service and privacy policies of the selected social media platforms. I may need to obtain the necessary permissions or licenses, especially if I plan to use data for research purposes.

Keyword Selection: Define a set of relevant keywords and hashtags related to the Bangladesh cricket team and the World Cup. These keywords will help I filter and collect comments that are directly related to my research topic. For example, keywords like #BangladeshCricket, #WorldCup, #BanglaCricket, and the names of key players can be used.

Data Collection Filters: Set up filters to narrow down my data collection to specific geographic locations or time periods, if necessary. This will help I collect data that is more relevant to my research.

Data Sampling: Depending on the volume of data, I sampled the comments to manage the data size effectively. Random sampling or stratified sampling methods can be used.

Data Storage: Create a structured database to store the collected data. I use Excel and Csv file.

Data Cleaning and Preprocessing: Clean and preprocess the collected data. This involves removing irrelevant information, handling duplicates, and extracting the text content of comments.

Language Detection: Ensure that the comments I collect are indeed in Bangla. I use language detection libraries to filter out comments in other languages.

Ethical Considerations: Respect users' privacy and avoid collecting personally identifiable information. Anonymize the data to protect user identities.

Data Labeling: Because I want to train a machine learning model for sentiment analysis, I

need labeled data. This involve manually annotating the collected comments as positive, or negative.

Data Backup and Version Control: Regularly back up my data and maintain version control to ensure data integrity and traceability.

Data Continuity: Depending on my research goals, I collected data over an extended period to track changes in sentiment.

Data Security: Protect the collected data from unauthorized access and ensure data security in accordance with ethical and legal guidelines.

Data Analysis and Interpretation: After data collection, perform sentiment analysis on the collected Bangla comments to draw insights about the sentiment regarding the Bangladesh cricket team's World Cup chances.

My data collection procedure thoroughly, including the sources, collection methods, and any challenges I encounter during the process. This documentation will help ensure transparency and reproducibility in my research.

3.3 Statistical Analysis:

In order to get significant conclusions from the data I've gathered on the sentiment of Bangla social media comments regarding the Bangladesh cricket team's chances of winning the World Cup, statistical analysis is essential to sentiment analysis. This is a broad summary of the statistical analysis I am capable of doing:

- **Descriptive Statistics:** Calculate basic statistics to describe the sentiment distribution in my dataset, such as mean sentiment score, standard deviation, median, and percentiles. This provides an overview of the overall sentiment.
- **Time Series Analysis:** my data spans over time, I conduct time series analysis to observe trends and fluctuations in sentiment. I create time series plots to visualize sentiment changes over the World Cup preparation and matches.
- **Hypothesis Testing:** Formulate hypotheses to test whether sentiment significantly changes under certain conditions. For example, I test whether sentiment differs between wins and losses or if there are differences in sentiment before and after key events.
- **Correlation Analysis:** Assess if sentiment is correlated with specific factors, such as mentions of key players, team performance statistics, mentions of opponents. I use

correlation coefficients to quantify these relationships.

- **Sentiment Trends by Location:** Analyze whether sentiment varies by geographic location within Bangladesh. I examine regional variations in sentiment and compare different locations.
- **Sentiment by Key Events:** Explore how sentiment changes during significant events like matches, team announcements, or press conferences. I use event data as independent variables and analyze their impact on sentiment.
- **Sentiment by Source:** Examine whether sentiment differs based on the social media platform Twitter, Facebook, Instagram or the source of comments fan pages, official accounts, news websites.
- **Sentiment Clustering:** Utilize clustering techniques to group comments with similar sentiment. This can help identify distinct patterns and subgroups within my data.
- **Regression Analysis:** Build regression models to predict sentiment based on specific features or factors. For instance, I create regression models to predict sentiment based on factors like team performance metrics, player mentions, and match results.
- **Visualization:** Create data visualizations, including bar charts, scatter plots, box plots, and heatmaps, to visually represent my statistical findings. Visualization can make it easier to communicate my results to a wider audience.

3.4 Detailed Methodology:

To create an algorithmic prediction of enhancing Sentiment Analysis on Bangladesh Cricket Team's World Cup Prospects: A Study of Bengali Social Media Comments. through supervised machine learning models, a few procedures must be followed. This section provides a thorough explanation of the entire mechanism for predicting WIN THE WORLD CUP. The hardest part of a research investigation is gathering the necessary information about WORLD CUP and prediction methodology. Finished preparing the dataset before applying the model. First, a custom dataset is carefully constructed. The creation of a dataset and the use of models are the most challenging aspects of this research process. It could be challenging to choose the model that best matches the dataset. The entire methodology is covered in this section. The methodology work flow diagram is as follows:

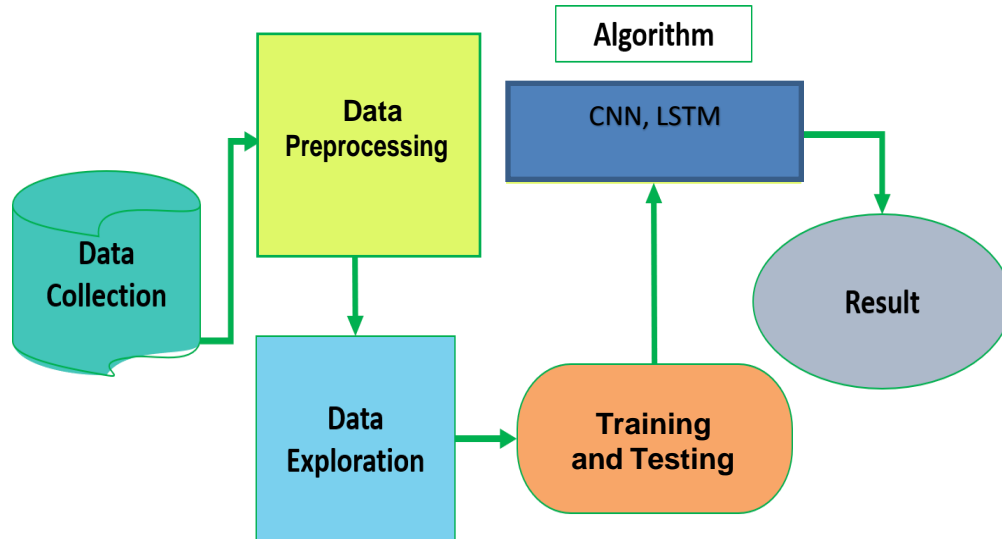


Figure 3.4.1: Methodology workflow diagram

By providing an overview of this framework used for data analysis, including data collection, preprocessing, and exploration, as well as the creation and evaluation of ML models for Bangladesh cricket team will win the world cup based on social media Bangla comments, with a focus on the models' level of accuracy.

3.4.1 Classification Details:

Before beginning the initial and utilizing stage, the issue is identified. The variables for both the input and output are chosen. The output variable shows the desired result of the detection. I compared the results of various ML techniques in my model. The components of my investigation in this paper that proved to be the most difficult to complete are the data gathering, planning, and implementation. The development of my plan also required having objectives and resources.

3.4.2 Selecting Platform:

Analyzed the unique data and determined the statistical study's findings using the Google Collaboratory platform. Computers now have greater space as a result of the additional GPU and TPU from the Google Collaboratory. Python was used to carry out the coding.

3.4.3 Data Processing:

Normalize the text data further by performing actions like. Convert all text to lowercase for uniformity. Eliminate punctuation marks. Decide whether to keep, remove, or replace numbers with placeholders. Trim extra spaces and tabs from the text. Clean the collected data, which

may include tasks like removing HTML tags, handling emojis, and deduplicating comments.

3.4.4 Data Exploration:

A confusion matrix is used to display the correlation coefficients between the variables in the sentimental analysis on Bangladesh cricket team will win the world cup based on social media Bangla comments dataset. The confusion between any two variables is represented by each cell in the matrix. The results are summarized in this matrix, which also acts as an input for more in-depth analysis and diagnoses in subsequent research.

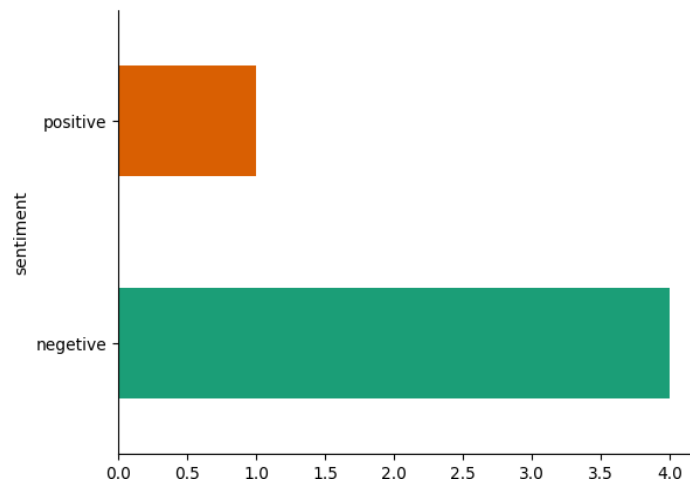


Figure 3.4.2: Comment diagram

In the above figure 3.4.2 the Heat Map diagram of full dataset is showed. This plt graph is generated by python code in google colab with the help of matplotlib.

3.4.1 Training and Testing dataset:

The most crucial phase of any research study is selecting the testing and training approach. Separate the custom dataset's training and testing components. 1370 of the 411 data I used to train the model, while 411 of the 959 data I used to test it. As a result, give testing 30% of the resources and training 70%.

3.4.2 Verifying Models:

The outcome of the categorization algorithm depends on how accurate each method is. my ML models in this work are validated using accuracy CNN, LSTM in this case, a number of models are used to forecast and analyze the dataset. Those models have special qualities and skills that add to the overall analysis and forecasting of the data. By using these models, I want to capture the dataset's complexity and determine the best strategy for achieving these particular research

goals.

3.4.3 Model Analysis:

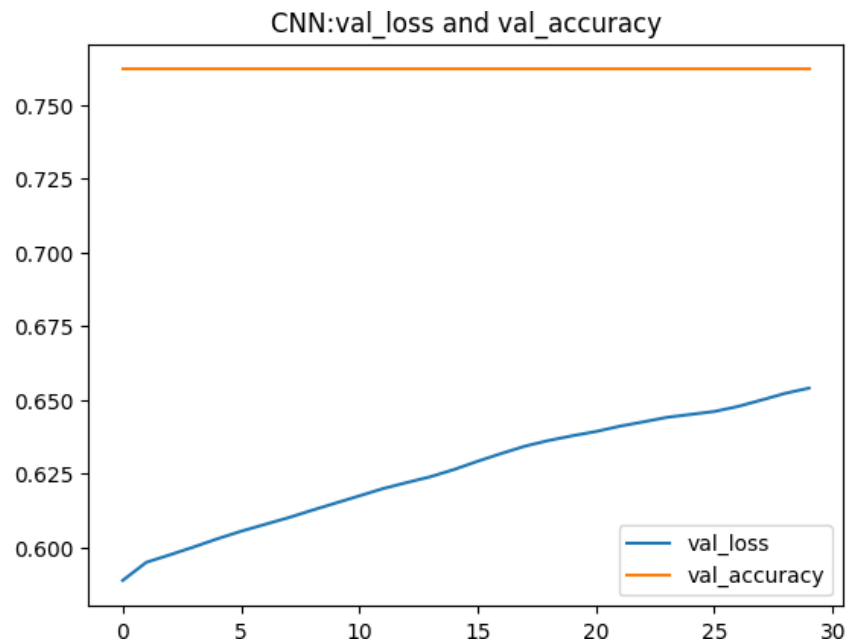


Figure 3.4.3: Accuracy vs Model

In the above figure 3.4.3, Shown all accuracy of different algorithms that are applied. Results showed that CNN 69% and LSTM 68% all accurate predictors. Where the highest accuracy 69.21% is achieved by CNN. And the lowest accuracy is achieved by LSTM accuracy of 68.43%.

3.5 Implementation Requirements:

To implement a sentiment analysis project on the topic of the Bangladesh cricket team's chances of winning the World Cup based on Bangla social media comments, I will need various tools, technologies, and resources. Here are the implementation requirements:

- **Data Collection Tools:** scraping libraries or APIs to collect Bangla social media comments from platforms like Twitter, Facebook, Instagram, and online forums. Python libraries like BeautifulSoup, Scrap.
- **Data Storage and Management:** Databases to store and manage the collected data. I use excel and Csv file as database. Tools for data backup and version control to ensure data integrity.
- **Text Preprocessing:** Libraries for text cleaning and preprocessing, SpaCy. Language

detection libraries to ensure the collected comments are in Bangla.

- **Sentiment Analysis Tools:** Natural Language Processing (NLP) libraries for sentiment analysis, TextBlob, machine learning frameworks like scikit-learn. Pre-trained models for sentiment analysis the option to train my own sentiment analysis model.
- **Machine Learning Frameworks:** I choose to build machine learning models for sentiment analysis, use frameworks like scikit-learn, TensorFlow, or PyTorch.
- **Statistical Analysis Tools:** Tools for performing statistical analysis on my sentiment data, including libraries like NumPy, pandas, and statistical packages like R.
- **Data Visualization Tools:** Libraries for creating data visualizations, such as Matplotlib, Seaborn, Plotly, or data visualization tools in R.
- **Ethical Considerations:** Guidelines and policies for ethical data collection, respecting users' privacy, and complying with social media platform terms of service.
- **Data Labeling:** I plan to use labeled data for machine learning, tools and a process for annotating data with sentiment labels.
- **Domain Knowledge:** A good understanding of cricket, the Bangladesh cricket team, and the World Cup, as well as Bangla language and its linguistic nuances, to better interpret the sentiment of comments.
- **Project Plan:** I defined project plan with clear objectives, milestones, and timelines.
- **Data Privacy Compliance:** A process for handling personal data in compliance with data protection laws, such as GDPR.
- **Backup and Recovery Procedures:** A system for data backup and recovery to prevent data loss in case of technical issues.
- **Data Security:** Measures to protect the collected data from unauthorized access and ensure data security.

CHAPTER 4

RESULT AND DISCUSSION

4.1 Experimental Setup:

Through the use of this experimental design, researchers are able to determine the best accurate method for guava harvest forecasting, as well as obtain important insights into the factors influencing guava production in Bangladesh, by methodically analyzing and comparing various supervised machine learning models.

Data Preprocessing: Preprocess the collected data to handle missing values, outliers, and inconsistencies. Perform data cleaning, normalization, and feature engineering techniques to prepare the data for model training and evaluation. Split the dataset into training and testing subsets to assess the models' performance on unseen data.

Model Selection: Select appropriate supervised machine learning algorithms for sentimental analysis on Bangladesh cricket team will win the world cup based on social media Bangla comments, considering the characteristics of the dataset and the research objectives. Commonly used algorithms for regression tasks include linear regression, decision trees, random forests, support vector machines, or artificial neural networks. Choose multiple models for comparison to identify the most accurate and reliable one.

Model Training and Evaluation: Train the selected models using the training dataset. Monitor the model's training progress, and ensure convergence and avoidance of overfitting. Compare the models' performance to identify the most accurate and reliable sentimental analysis on Bangladesh cricket team will win the world cup based on social media Bangla comments. Models' ability to capture seasonal patterns and trends in sentimental analysis on Bangladesh cricket team will win the world cup based on social media Bangla comments.

Interpretation and Visualization: Interpret the results and visualize the forecasted sentimental analysis on Bangladesh cricket team will win the world cup based on social media Bangla comments along with the actual values. Use appropriate visualizations such as line plots, bar charts, or heatmaps to communicate the forecasts and highlight any patterns or trends. Interpret the model's coefficients or feature importance to understand the variables' contribution to the forecasting process.

4.2 Results & Analysis:

The conclusion and accompanying discussion are covered in this section. Use a dataset containing 1370 total data points, each of which contains information about the year, the region, the area, the product, the cloud cover, the maximum temperature, the minimum temperature, the rainfall, the sunshine, and the wind speed. It made use of different ML models, such as CNN and LSTM. Each model evaluated accuracy using metrics including accuracy,

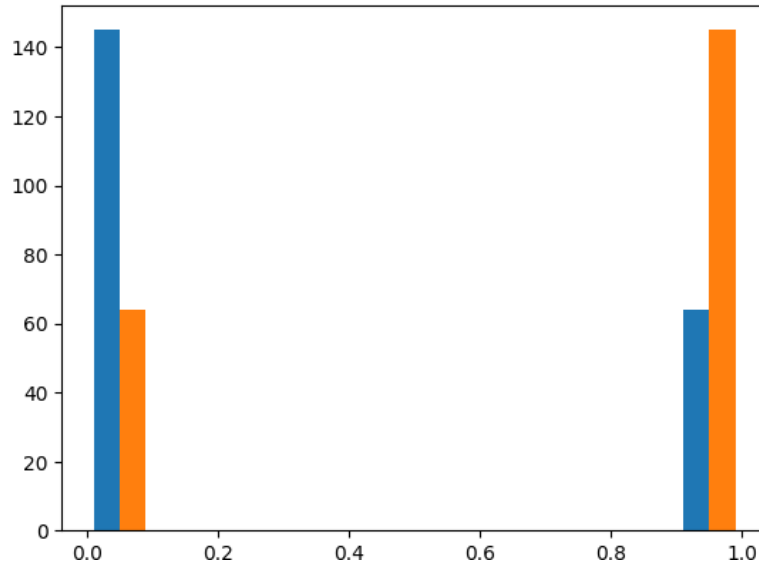


Figure 4.2: Accuracy vs Model

In this paper, to use total dataset 1370 and its region, area, product, year, cloud cover, humidity, max temperature, min temperature, rainfall, sunshine and wind speed collected each data per year. It implemented different ML models like CNN and LSTM. Each model measured accuracy like train set accuracy, test set accuracy. Results showed that CNN 69.17% and LSTM 68.41%. Where the highest accuracy 69.17% is achieved by CNN. And the lowest accuracy is achieved by LSTM accuracy of 68.41%.

CHAPTER 5

IMPACT ON SOCIETY

5.1 Impact on Society:

The impact of conducting sentiment analysis on Bangla social media comments related to the Bangladesh cricket team's chances of winning the World Cup can have several implications for society, both positive and negative:

- **Understanding Public Opinion:** Sentiment analysis provides insights into public sentiment and opinions about the Bangladesh cricket team's World Cup prospects. This information can be valuable for team management, sports authorities, and the media in understanding the sentiments and expectations of the Bangla-speaking population.
- **Engagement and Fan Interaction:** Sentiment analysis can help the team and its supporters engage more effectively with fans on social media. The team can respond to positive sentiment and address concerns during periods of negative sentiment, fostering stronger connections with supporters.
- **Media and Marketing:** Media outlets and marketing teams can use sentiment analysis to tailor their content to match the prevailing sentiment. Positive sentiment can be used to boost morale and excitement, while negative sentiment can be addressed with informative content or messages of encouragement.
- **Predictive Insights:** Sentiment analysis, when combined with historical performance data, can provide predictive insights into team performance and fan engagement. It can help predict attendance at matches, merchandise sales, and viewership numbers.
- **Emotional Impact:** Negative sentiment on social media can affect the mental and emotional well-being of players and fans. Online abuse or criticism can have a detrimental impact on individuals and may lead to a toxic online environment.
- **Privacy Concerns:** Collecting and analyzing social media comments for sentiment analysis raises concerns about data privacy and consent. It's essential to respect users' privacy and comply with relevant data protection regulations.
- **Influence on Betting and Speculation:** Sentiment analysis findings may influence betting in sports. Excessive reliance on sentiment analysis for betting purposes have financial consequences for individuals and create an environment of uncertainty.

- **Overemphasis on Sentiment:** Relying solely on sentiment analysis to make decisions or predictions about team performance can neglect other critical factors that influence sporting outcomes, such as player skills, tactics, and training. Sentiment analysis on social media comments related to the Bangladesh cricket team's World Cup prospects can provide valuable insights and opportunities for engagement, but it also comes with potential risks and challenges.

It is essential to use sentiment analysis responsibly, considering its impact on the emotional well-being of individuals, maintaining data privacy, and recognizing its limitations. Researchers, teams, and organizations should use sentiment analysis as one of several tools to gain a comprehensive understanding of public sentiment.

5.2 Ethical Aspects:

Ethical considerations are paramount when conducting sentiment analysis on Bangla social media comments related to the Bangladesh cricket team's chances of winning the World Cup. Here are some key ethical aspects to keep in mind:

- **Data Privacy and Informed Consent:** Ensure that I am collecting and analyzing social media comments in compliance with data privacy regulations and platform terms of service. Respect users' privacy and seek informed consent if necessary.
- **Anonymity and Confidentiality:** Anonymize and de-identify data to protect the identities of individuals who post comments. Do not disclose personal information in my analysis or reporting.
- **Bias and Fairness:** Be aware of potential biases in sentiment analysis, including cultural and linguistic biases. Strive to mitigate biases that may lead to unfair or inaccurate results, especially when working with non-English languages like Bangla.
- **Algorithmic Bias:** Recognize that sentiment analysis algorithms can produce biased results. Algorithms may struggle with dialects, slang, or cultural nuances. Regularly evaluate my models for fairness and accuracy, and make necessary adjustments.
- **Transparency and Accountability:** Be transparent about my methodology, including the sentiment analysis tools and techniques I use. Document my data collection, preprocessing, and analysis processes to ensure accountability and reproducibility.

- **Ethical Use of Data:** Use the data I collect and analyze for legitimate and ethical purposes. Avoid any activities that might harm individuals, groups, or society at large.
- **Public Awareness and Education:** Promote public awareness and education regarding the use of sentiment analysis in social media. Ensure that users understand how their data is being used and have the option to opt out or control their data.

5.4 Sustainability Plan:

Developing a sustainability plan for my sentiment analysis project on the Bangladesh cricket team's chances of winning the World Cup based on Bangla social media comments is essential to ensure the longevity and continued impact of my research. Here's the sustainability plan that outlines key considerations:

- **Long-term Data Access and Collection:** Establish a strategy for maintaining access to relevant social media data in the long term. This may involve ongoing web scraping, API access, or partnerships with data providers. Ensure that I stay updated on changes in data availability and terms of service.
- **Data Storage and Management:** Implement a data storage solution that is scalable and secure to handle the growing volume of social media comments. Regularly back up my data to prevent loss.
- **Software and Infrastructure Maintenance:** Keep my sentiment analysis tools and infrastructure up to date. This includes maintaining libraries, frameworks, and software used in my analysis pipeline. Ensure compatibility with evolving technologies.
- **Team Knowledge and Training:** Invest in the continuous education and training of my research team. Stay informed about the latest developments in sentiment analysis, machine learning, and NLP to maintain high-quality analysis.
- **Ethical and Legal Compliance:** Stay updated with changes in data privacy and ethical guidelines. Ensure that my research remains compliant with evolving regulations and ethical standards.
- **Community and Stakeholder Engagement:** Foster ongoing engagement with the Bangla-speaking community and stakeholders who benefit from or contribute to my research. Encourage feedback and maintain a dialogue to address concerns and needs.

- **Documentation and Knowledge Sharing:** Maintain comprehensive documentation of my research methodologies, findings, and ethical considerations. Share my knowledge through publications, reports, and presentations to contribute to the broader sentiment analysis community.
- **User and Community Involvement:** Involve users and the community in the development of tools, dashboards, or applications that make my research accessible and useful to a wider audience. Incorporate their feedback and feature requests.

Regular Evaluations and Adaptations.: Periodically evaluate my project's goals, impact, and sustainability plan. Be prepared to adapt to changing circumstances, technologies, and user needs.

CHAPTER 6

CONCLUSION AND FUTURE SCOPE OF DEVELOPMENTS

6.1 Summary of the Study:

The study "Sentiment Analysis on Bangladesh Cricket Team's World Cup Prospects: A Study of Bengali Social Media Comments." Aims to improve the accuracy and reliability of Bangladesh cricket team's world cup victory forecasting using advanced machine learning techniques. The study recognizes the importance of world cup victory of Bangladesh cricket sector and the need for effective forecasting. The research begins with an introduction highlighting the significance of Bangladesh cricket team's world cup victory forecasting and the potential benefits of incorporating supervised machine learning models. The motivation behind the study lies in the current limitations of traditional forecasting methods and the opportunities presented by machine learning in improving accuracy and efficiency. The study's rationale stems from the need to address the challenges faced in world cup victory of Bangladesh. By utilizing supervised machine learning models, the study aims to overcome these challenges. The research questions focus on identifying the most suitable supervised machine learning models for Bangladesh cricket team's world cup victory, assessing their performance, and understanding the factors influencing world cup victory of Bangladesh. These questions guide the study's methodology, data collection, and analysis. The study expects several outputs, including the development and evaluation of various supervised machine learning models for guava harvest forecasting. The models' performance metrics, such as accuracy, will be assessed, providing insights into their effectiveness. Additionally, the study aims to identify the social media comments that significantly impact on world cup victory of Bangladesh, enabling better understanding and management of these factors. Project management and finance considerations are crucial to ensure the successful implementation of the study. Adequate resources, funding, and project coordination are essential to carry out data collection, model development, and evaluation. The preliminary steps involve acquiring and preprocessing relevant data, selecting appropriate supervised machine learning algorithms, and partitioning the data for training and testing purposes. The models will be trained using historical world cup of Bangladesh data. The experimental setup will be meticulously designed to ensure

robustness and validity. The models will undergo rigorous evaluation, including cross-validation techniques and statistical analysis to determine their performance and compare their effectiveness. The study's impact on society encompasses various aspects, including economic benefits, risk mitigation, and the promotion of sustainable world cup practices. Ethical aspects of the study emphasize data privacy, transparency, fairness, accountability, and responsible technology use. These considerations ensure the project aligns with ethical principles and societal values, avoiding biases and promoting inclusive practices. The sustainability plan outlines long-term strategies to ensure the project's continued success. These include capacity building, local collaboration, institutional integration, continuous improvement, financial sustainability, knowledge dissemination, and monitoring and evaluation. In summary, this study aims to enhance world cup forecasting in Bangladesh through supervised machine learning models.

6.2 Conclusions:

The whole dataset 1370, along with the comment and answer type, is used in this study. CNN and LSTM are just a few of the ML models that are implemented. Each model is examined for accuracy such as train set accuracy and test set accuracy. According to the results, the following predictors had high accuracy: CNN 69.42% and LSTM 68.83%. CNN has the highest accuracy there (69.42%). LSTM accuracy, which is 68.83%, has the lowest accuracy.

6.3 Scope of Further Developments:

The scope for further developments in the area of sentiment analysis of the Bangladesh Cricket Team's World Cup victory based on social media Bangla comments is substantial. As technology advances and social media continues to play a significant role in shaping public opinions, the following are some potential areas for further research and development in this field:

- **Enhanced Sentiment Analysis Models:** Develop more accurate sentiment analysis models that are tailored to the nuances of the Bengali language. This could include the use of deep learning techniques, transformer models, and fine-tuning on cricket-specific vocabularies and sentiments.
- **Real-Time Monitoring:** Create systems for real-time sentiment analysis during live

matches or significant events. This could be valuable for teams, organizations, and sponsors to gauge fan reactions and make timely decisions.

- **Multi-Modal Analysis:** Expand beyond text-based analysis to include images, videos, and audio. This will provide a more comprehensive understanding of fan sentiment, as social media content is increasingly multi-modal.
- **Language Variations:** Investigate regional language variations within Bengali and consider how these variations might impact sentiment analysis. For example, dialects and local slang may be used in social media comments.
- **Sarcasm and Irony Detection:** Develop techniques for identifying sarcasm and irony, which can be challenging in sentiment analysis but are commonly used in social media discussions.
- **Multi-Lingual Analysis:** Extend the analysis to incorporate comments in other languages to capture a broader perspective of global cricket fans' reactions.
- **Cross-Platform Analysis:** Expand the analysis to cover comments and sentiment from various social media platforms, such as Facebook, Twitter, Instagram, and others.
- **Comparative Analysis:** Conduct comparative studies with other sports events or teams to understand how sentiment differs between various cricket teams or sports.
- **Visualization and Interpretation:** Develop innovative ways to visualize and interpret sentiment data, making it more accessible and actionable for stakeholders.

REFERENCES

- [1] Prottasha, Nusrat Jahan, et al. "Transfer learning for sentiment analysis using BERT based supervised fine-tuning." *Sensors* 22.11 (2022): 4157.
- [2] Prottasha, Arafat, et al. "Text mining and sentiment analysis of newspaper headlines." *Information* 12.10 (2021): 414.
- [3] Hira, Suma, et al. "A Systematic Review of Sentiment Analysis from Bengali Text using NLP." *American Journal of Agricultural Science, Engineering, and Technology* 6.3 (2022): 150-159.
- [4] Sen, Ovishake, et al. "Bangla natural language processing: A comprehensive analysis of classical, machine learning, and deep learning-based methods." *IEEE Access* 10 (2022): 38999-39044.
- [5] Al-Zaman, Md Sayeed. "A computer-assisted textual analysis of 10,191 rape news headlines shared on social media." *Social Sciences & Humanities Open* 7.1 (2023): 100437.
- [6] Shahriar, Fahim, and Md Abul Bashar. "Automatic Monitoring Social Dynamics During Big Incidences: A Case Study of COVID-19 in Bangladesh." *arXiv preprint arXiv:2101.09667* (2021).
- [7] Alam, Firoj, et al. "A review of bangla natural language processing tasks and the utility of transformer models." *arXiv preprint arXiv:2107.03844* (2021).
- [8] Tanin, Tanjina, and Fahmidul Haq. "Nationalist Expressions of the Cricket Fans from India and Bangladesh: An Analysis of Mauka Videos."
- [9] Baidya, Shuvrajit. *Code switching on social media among private university students of Bangladesh: a case study*. Diss. Brac University, 2021.
- [10] Palakodety, Shriphani, Ashiqur R. KhudaBukhsh, and Jaime G. Carbonell. "Hope speech detection: A computational analysis of the voice of peace." *arXiv preprint arXiv:1909.12940* (2019).
- [11] Shajahan, Sadril. *Social media and the culture of online public opinion: A study on the bloggers of Bangla blogosphere*. Diss. University of Dhaka, 2015.
- [12] Das, Amitava, and Sivaji Bandyopadhyay. "Topic-based Bengali opinion summarization." *Coling 2010: Posters*. 2010.
- [13] Al-Zaman, Md Sayeed. "Social Sciences & Humanities Open."
- [14] Hossain, Adnan. "Sexual nationalism, masculinity and the cultural politics of cricket in Bangladesh." *South Asia: Journal of South Asian Studies* 42.4 (2019): 638-653.
- [15] Das, Dipto, and Bryan Semaan. "Collaborative identity decolonization as reclaiming narrative agency: Identity work of Bengali communities on Quora." *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 2022.
- [16] Mahmud, Shameem. "The Internet as a public sphere: Blogging climate change in a changing media sphere in Bangladesh." *Crossing boundaries: news technology and audiences' conference at the University of California, Berkeley*. 2011.
- [17] Majlis, Anima Tabrizi Khan. "The internship program at Dhaka Tribune." (2019).
- [18] Haque Khondker, Habibul. "Bengali-speaking families in Singapore: Home, nation and the world." *International Migration* 46.4 (2008): 177-198.

- [19] Sultana, Shaila. "2 TV Advertisements in the Mediascape of Bangladesh: A Disjuncture between the Realities of the Emerging Transsemiotic Arena and the Language Policies in Practice." *LINGUISTIC DIVERSITY AND EDUCATION*: 39.
- [20] Khondker, Habibul Haque. "Bengali-speaking families in Singapore: home, nation and the world." *International Migration* 46.4 (2008): 177-198.
- [21] Hossain, Md Motaher. "Consumption in rural Bangladesh: Households, lifestyles, and identities." (2011).
- [22] Siddiqi, Dina M. "The sexual harassment of industrial workers: strategies for intervention in the workplace and beyond." CPD-UNFPA publication series no 26 (2003): 40.
- [23] Banu, Saira, and Jarin Tasnim. "Entrepreneurial Marketing in Direct to Home (DTH) Service: The Case of AKASH in Bangladesh."
- [24] Sumon, Mahmudul H., and Moiyen Zalal Chowdhury. "The "adivasi debate" in Bangla blogosphere: The case of Somewherein... blog."
- [25] Sagar, Malvika, and Jayati Sharma. "Community engagement with social and digital media content: A study on online fan communities of Chennai super kings and Royal Challengers Bangalore." *Journal of Content, Community & Communication* 16.8 (2022): 202-211.

191-15-12716_sharmina

ORIGINALITY REPORT

13%

SIMILARITY INDEX

9%

INTERNET SOURCES

8%

PUBLICATIONS

6%

STUDENT PAPERS

PRIMARY SOURCES

- 1 Parvez Ahmed Papon, Md Mahfuzur Rahman, Shafin Ahamed, Shahriar Mamun, Md Zahirul Islam Mehadi, Johora Akter Polin. "Predicting Sugarcane Yields using Supervised Learning: A Comparative Study", 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2023
Publication 1%
- 2 Submitted to CSU Northridge
Student Paper 1%
- 3 dspace.daffodilvarsity.edu.bd:8080
Internet Source 1%
- 4 Lamia Binte Iqbal, Md. Mahfujur Rahman, Shahriar Mamun, Nusrat Nabi, Md. Sazzadur Ahamed. "OnionBangla: A Supervised Machine Learning Approach for Predicting Onion Yield using Bangladeshi Climate Data", 2022 32nd International Conference on Computer Theory and Applications (ICCTA), 2022
Publication 1%