Multi head text mining applying deep learning and Bangla NLP using COVID-19 feedback from students

BY

Khadijatul Kobra
ID: 191-15-12319

AND

Samrina Sarkar Sammi
ID: 191-15-12532

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

Supervised By

Md. Sanzidul Islam
Lecturer
Department of CSE
Daffodil International University

Co-Supervised By

Sharun Akter Khushbu
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

FEBRUARY 2023

APPROVAL

This Project titled **"Multi head text mining applying deep learning and Bangla NLP using COVID-19 feedback from students"** submitted by **Khadijatul Kobra, ID no: 191-15-12319** and **Samrina Sarkar Sammi, ID no: 191-15-12532** 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 2nd February, 2023.

BOARD OF EXAMINERS

Dr. Touhid Bhuiyan — **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Sheak Rashed Haider Noori — **Internal Examiner**
Professor and Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Md. Sazzadur Ahamed — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Sazzadur Rahman — **External Examiner**
Associate Professor
Institute of Information Technology
Jahangirnagar University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Md. Sanzidul Islam, Lecturer, Department of CSE** Daffodil International University. We 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:

For Khushbu

Md. Sanzidul Islam
Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

Khushbu

Sharun Akter Khushbu
Lecturer
Department of CSE
Daffodil International University

Submitted by:

Khadija

Khadijatul Kobra
ID: -191-15-12319
Department of CSE
Daffodil International University

Sammi

Samrina Sarkar Sammi
ID: -191-15-12532
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Md. Sanzidul Islam**, **Lecturer**, Department of CSE Daffodil International University, Dhaka. To carry out this project, we needed our supervisor's extensive knowledge and keen interest in machine and deep learning in bangla NLP. His never-ending tolerance and academic advice, continuous encouragement , constant and energetic supervision, constructive criticism , valuable advice ,reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Professor Dr Touhid Bhuiyan,** Head**,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We'd like to thank all the learners of Daffodil International University who participated in this discussion while also attending class.

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

ABSTRACT

The COVID-19 epidemic bound administration all around the globe to lock down their frontier, their companies, their schools, and citizens from leaving their houses unless absolutely required. The mental health of both individuals and society as a whole can be seriously harmed by being so imprisoned. This negatively affects students' social and emotional well-being. Having schools, colleges, and universities closed had a significant negative influence on students' academic lives. We physically examined Bangla text that was physically obtained from students using a google form. It was all written in Bangla. With two label classes—positive and negative—through text categorization, we looked at how covid affected social relationships, mental health, and academic success. We tested several ML algorithms like logistic regression, decision trees, random forests, multi-naive Bayes, KNN, SGD, linear SVM, and RBF SVM, and discovered that SGD had the highest accuracy for academic impact. We get the greatest accuracy for the influence on social life column using KNN, and the best score for the impact on mental health with both multi-naive Bayes and SGD. Additionally, we used the CNN, LSTM, CNN-LSTM, BiLSTM, and CNN-BiLSTM as deep learning models in the aforementioned three columns, and we achieved an academic impact accuracy of 80%, mental health impact accuracy of 98.75% and Social life impact accuracy 83.75% using LSTM. The accuracy is 92.50%, 85%, and 92.50% using BiLSTM for academic, mental, and social impact columns. while its accuracy in terms of using CNN, CNN-LSTM, and CNN-BiLSTM is 82.50%, 70%, 92.50%; 85%, 90%, 92.50%; and 85%, 85%, 90% for academic, mental, and social impact columns, respectively.

Keyword: COVID-19, sentiment classification, Deep Learning, , SGD, CNN LSTM, hybrid deep learning

TABLE OF CONTENTS

CONTENTS	PAGE
Board of examiners	ii
Declaration	iii
Acknowledgements	iv
Abstract	v
CHAPTER	
CHAPTER 1: INTRODUCTION	**1-5**
1.1 Introduction	1
1.2 Motivation	2
1.3 Rational of the Study	3
1.4 Research Questions	4
1.5 Expected Outcome	4
1.6 Project Management and Finance	5
1.7 Report Layout	5
CHAPTER 2: BACKGROUND STUDY	**6-10**
2.1 Preliminaries	6
2.2 Related Works	6
2.3 Comparative Analysis and Summary	8
2.4 Scope of the Problem	10
2.5 Challenges	10

CHAPTER 3: RESEARCH METHODOLOGY **11-27**

3.1 Introduction 11

3.2 Subject of Research and Equipment 12

3.3 Procedure for Collecting Data 12

3.4 Data Preprocessing 13

3.4.1 Removal Punctuation 14

3.4.2 Removal of small length of text 14

3.4.3 Removal Stop word 14

3.4.4 Stemming 15

3.4.5 Data Vectorization 15

3.4.6 Tokenization 17

3.4.7 Pad Sequence 17

3.4.8 Label encoding 17

3.4.9 Data splitting 17

3.5 Statistical Analysis 17

3.6 Implementation Requirements 20

3.6.1 Supervised Machine Learning 20

3.6.1.a Logistic Regression Classifier 21

3.6.1.b Decision Tree Classifier 21

3.6.1.c Random Forest Classifier 21

3.6.1.d Multinomial Naive Bayes Classifier 22

3.6.1.e k-Nearest Neighbors Classifier 23

3.6.1.f Linear SVM Classifier 23

3.6.1.g RBF SVM Classifier 24

3.6.1.h Stochastic Gradient Descent 24

3.6.2 Deep Learning 24

3.6.2.a Sequential Model 25

3.6.2.b LSTM 26

3.6.2.c BiLSTM 26

3.6.2.d CNN 26

3.6.2.e CNN LSTM 27

3.6.2.f CNN BiLSTM 27

3.6.3 Confusion Matrix 27

3.6.4 Evaluation of model 27

CHAPTER 4: EXPERIMENT RESULTS AND DISCUSSION 28-58

4.1 Experimental Setup 28

4.2 Experimental Results & Analysis 28

4.2.1 Analysis Experiment result of Supervised Machine Learning models 30

4.2.1.1 Academic Life Impact Visualization 30

4.2.1.2 Mental Health Impact 33

4.2.1.3 Social Life Impact 36

4.2.2 Analysis Experiment result of Deep Learning models 39

4.2.2.1 LSTM 39

4.2.2.2 Bi-LSTM 43

4.2.2.3 CNN 46

4.2.2.4 CNN LSTM 50

4.2.2.5 CNN Bi-LSTM 54

4.3 Discussion 58

CHAPTER 5: Impact on Society, Environment and Sustainability **59-60**

5.1 Impact on Society 59

5.2 Impact on Environment 59

5.3 Ethical Aspect 60

5.4 Sustainability Plan 60

CHAPTER 6: Summary, Conclusion, Recommendation and Implication for Future Research **61-62**

6.1 Synopsis of the Study 61

6.2 Conclusion 62

6.3 Future Work 62

REFERENCES **63-65**

LIST OF FIGURES

FIGURES	PAGE NO
Figure 3.1: Working Flowchart	11
Figure 3.4.1: After data purification and punctuation removal	14
Figure 3.4.4: An illustration of a stemming word	15
Figure 3.5: Unigram CountVectorization	16
Figure 3.5.1: Label Distribution Bar Chart	18
Figure 3.5.2: Statistics inside of a phrase with a class name	18
Figure 3.5.3: Bar Chart of COVID Students' Attitudes Toward Online Learning	19
Figure 3.5.4: Bar Chart of Academic pressure on Students During Covid	19
Figure 3.5.5: Bar Chart of During Covid Anxiety level of students	19
Figure 3.6.1 Process diagram for supervised machine learning	20
Figure 3.6.1.1: Random Forest technique of operation	22
Figure 3.6.1.2: Multi. Naive Bayes' technique of operation	23
Figure 3.6.2: Deep learning process diagram	25
Figure 3.6.2.1: sequential model's data movement	25
Figure 4.2.1.1.1: Bar chart of Vectorization feature for impact on Academic life (a) Bar chart of Unigram feature (b) Bar chart of Bigram feature (c) Bar chart of Trigram feature	31-32
Figure 4.2.1.2.1: Bar chart of Vectorization feature for impact on Mental health (a) Bar chart of Unigram feature (b) Bar chart of Bigram feature (c) Bar chart of Trigram feature	34-35
Figure 4.2.1.3.1: Bar chart of Vectorization feature for impact on Social Life (a) Bar chart of Unigram feature (b) Bar chart of Bigram feature (c) Bar chart of Trigram feature	37-38
Figure 4.2.1: An illustration of the machine learning model's confusion matrix for (a) the academic column, (b) the mental column, and (c) the social column.	39
Figure 4.2.2.1.1: LSTM sequential Model for (a) Academic column (b) Mental column (c) Social column	39

Figure 4.2.2.1.2: Illustration of LSTM Loss curve for (a) Academic (b) Mental (c) Social column 40

Figure 4.2.2.1.3: Visiualization of the LSTM Traning and validation for (a) Academic column (b) Mental column (c)Social column 40

Figure 4.2.2.1.4: Illustration Confusion matrix of LSTM model for (a) Academic column (b) Mental column (c)Social column 42

Figure 4.2.2.2.1: Bi-LSTM sequential Model for Academic, Mental and Social column 43

Figure 4.2.2.2.2: Bi-LSTM Loss curve for the following columns: (a) Academic (b) Mental (c) Social 43

Figure 4.2.2.2.3: Bi-LSTM training and validation curve for the following columns: (a) Academic (b) Mental (c) Social 44

Figure 4.2.2.2.4:Demonstration of the Bi-LSTM model's confusion matrix for (a) Academic column (b) Mental column (c)Social column 45

Figure 4.2.2.3.1: CNN sequential Model for Academic, Mental column and Social column 46

Figure 4.2.2.3.2: An overview of the CNN Loss Curve for the (a) Academic, (b) Mental, and (c) Social Columns 47

Figure 4.2.2.3.3: CNN training and validation curves for (a) academic column (b) mental column (c) social column 47

Figure 4.2.2.3.4: Illustration Confusion matrix of CNN model for (a) Academic column (b) Mental column (c)Social column 49

Figure 4.2.2.4.1: CNN LSTM sequential Model for Academic, Mental column and Social column 50

Figure 4.2.2.4.2: Illustration of CNN-LSTM Loss curve for following columns of (a) Academic (b) Mental (c)Social 51

Figure 4.2.2.4.3: CNN-LSTM Training and validation curve for (a) Academic column (b) Mental column (c)Social column 51

Figure 4.2.2.4.4: Illustration Confusion matrix of CNN-LSTM model for (a) Academic column (b) Mental column (c)Social column 53

Figure 4.2.2.5.1: CNN BiLSTM sequential Model for Academic, Mental column and Social column 54

Figure 4.2.2.5.2: Illustration of CNN BiLSTM Loss curve for (a) Academic column (b) Mental column (c)Social column 55

Figure 4.2.2.5.3: Visiualization of the CNN BiLSTM Traning and validation curve for (a) Academic column (b) Mental column (c)Social column 55

Figure 4.2.2.5.4: Illustration Confusion matrix of CNN BiLSTM model for (a) Academic column (b) Mental column (c)Social column 57

LIST OF TABLES

TABLES	PAGE NO
Table 2.3: Comparative analysis of previous work	9
Table 3.3: Dataset Preview	13
Table 3.5: Example of n-gram distribution	16
Table 4.2.1.1.A : Unigram Feature Performance Table for Impact on Academic Life	30
Table 4.2.1.1.B: Bigram feature Performance Table for impact on Academic life	30
Table 4.2.1.1.C: Trigram feature Performance Table for impact on Academic life	31
Table 4.2.1.1.D: Table of machine learning model accuracy, precision, recall, and F1 scores	32
Table 4.2.1.2.A: Unigram feature Performance Table for impact on mental health	33
Table 4.2.1.2.B: Bigram feature Performance Table for impact on mental health	33
Table 4.2.1.2.C: Trigram feature Performance Table for impact on Mental health	34
Table 4.2.1.2.D: Table for accuracy, precision, recall, and F1 scores for machine learning models	35
Table 4.2.1.3.A: Performance of the Unigram feature as it relates to social life	36
Table 4.2.1.3.B: Bigram feature performance Table for impact on social life	36
Table 4.2.1.3.C: Trigram's Performance Table for Social Life Impact	37
Table 4.2.1.3.D: Table for accuracy, precision, recall, and F1 scores for machine learning models	38
Table 4.2.2.1.A : LSTM Classification report table for (a) Academic column (b) Mental column (c)Social column	40-41

Table 4.2.2.1.B: LSTM Model evaluation 42

Table 4.2.2.2.A: BiLSTM Classification report table for (a) Academic column (b) Mental column (c)Social column 44-45

Table 4.2.2.2.B: Model Performance for BiLSTM 46

Table 4.2.2.3.A: CNN Classification report table for columns (a) Academic (b) Mental (c) Social 47-48

Table 4.2.2.3.B: Model Performance for CNN 49

Table 4.2.2.4.A : CNN-LSTM Classification report table for (a) Academic column (b) Mental column (c)Social column 51-52

Table 4.2.2.4.B: Model Performance for CNN LSTM 53

Table 4.2.2.5.A: CNN BiLSTM Classification report table for (a) Academic column (b) Mental column (c)Social column 55-56

Table 4.2.2.5.B: Model Performance for CNN BiLSTM 57

CHAPTER 1

Introduction

1.1 Introduction

For longer than the last two years, COVID has been a major source of concern for the global population. This fatal condition has affected millions of people. It was first discovered in Wuhan, China, in 2019. On the other side, the WHO labeled it a epidemic in March 2020, claiming that it has rapidly spread over the globe. [27] We must have faced various challenges as a result of COVID-19, including health concerns and other little and major ones. These issues must have caused us some suffering or loss, but with time, everything faded away. When we look back after a while, we see that we gained a lot from going through that challenging time in our lives. We are aware that the corona wreaked ruin on the lives of practically all people living today. It affected our lives greatly and took so much from us. Students all throughout the world are significantly affected by it. Particularly on students' social lives, academic lives, and mental health. Spending a lot of time in isolation alters our thinking, speech, and behavior, which can be harmful to our physical, mental health, and social lives. These alterations cause unreasonable mood swings, inattentiveness, behavioral changes, and a host of other issues. [1]

One of the pandemic's most significant consequences has been in the field of education, which has resulted in the worst disruption in the history of the global educational system. As a result of COVID-19 lockdowns, schools and institutions all around the world closed their doors, keeping both students and staff in the dark about what lies ahead. During this time, students were identified via the online education system. For the students, this was a novel experience. There were numerous resources accessible for different areas; there was no need to go to college or university; one could receive instruction whenever they wanted; and there were other good effects. However, it also has a number of drawbacks, including the lack of online learning system training for teachers, the absence of student gadgets, internet connection issues, the financial crisis, etc. Academic pressure and social problems have a variety of effects on mental health. Due to the emotions that students go through, such as anxiety over their grades, stress from their families, and grief over losing friends, the usefulness of remote learning systems may be diminished. [2] Spending a lot of time on electronics has a bad effect on both physical and mental health, such as headaches, eye difficulties, and temper problems. Every human being values their family. During the lockdown, students spend time with their loved ones. It has a very favorable effect. Students who had poor familial and friendship ties were mentally disturbed in the setting. In that palpable scenario, several students offered assistance to everyone in their vicinity. COVID19 has a significant influence on student life overall. Bangla NLP is now improving daily. The processing of Bangla texts in relation to the effects of COVID on diverse student populations is the main focus of our study.

1.2 Motivation

A highly important area of Bangla NLP to deal with today is sentiment analysis. Our native tongue is Bangla, and we communicate with people using this language to express our feelings. Due to the fact that many applications already integrate sentiment analysis for other languages but not for Bangla, many people update their statuses on social media or text one another throughout this pandemic. It will be quite simple to predict the emotional condition of the person posting a status update or message if there is a sentiment analyzer for Bangla. This concept may be used in various apps, but language apps in particular. Our emotional condition has been greatly impacted by the most recent pandemic COVID-19. A precise dataset is needed to be able to achieve the finest outcomes. To capture the feelings of students at the moment in text form in Bangla, we constructed a dataset. To effectively anticipate emotions, we trained our algorithm. It will be possible to identify the emotional state represented in Bangla through the text connected to the pandemic impact and use this information in a variety of applications.

1.3 Rationale of the Study

The heritage of the Bengali language is quite extensive. Bengali is now the primary language of millions of people worldwide. However, compared to other languages of this present era, Bengali's technology and methods are not as advanced. As a result, that language's architecture has to be enhanced. Many text-based issues may be resolved using NLP techniques and tactics. Bengali emotion analysis is a major NLP challenge right now. There are several text sequences inside it. It enables individuals to rapidly pick up the language with fluency and understand how important it is to identify errors in lengthy emails and consider other people's feelings. The majority of the most relevant and crucial NLP approaches have previously been established for several other languages, such as English and French. However, some inadequate translations of the Bengali materials have been produced. Additionally, the research on Bengali NLP has to be expanded. Bengali text faces the greatest challenge in pretreatment. The machine cannot understand any Bengali letters or symbols. For this problem, the Unicode for these glyphs or characters must be utilized. Because of this, Bengali's approaches don't really work as well as those in other languages. It is important to use sophisticated reasoning and study to find solutions to these types of issues. We thus attempted to show in this work how Bengali text is handled in various preprocessing stages and how these sentences may be fed to various ML and DL models. We must collect public thoughts on a particular subject that has never been used before in order to differentiate our work distinctively. We decided to test students on how COVID-19 has affected their academic, mental, and social lives. It will make it easier to accurately examine various kinds of Bengali text. We employed certain ML techniques and DL models for the investigation. Through this investigation, we can also determine the method or model that can really properly classify text.

1.4 Research Questions

- What is Bengali Sentiment classification?
- What is Multi head text mining?
- How is NLP used to process Bengali text?
- What advantages can Bengali text sentiment categorization provide?
- How does the Machine Learning algorithm work with Bengali text?
- How does the deep learning model work with Bengali text?
- What ML algorithm and DL model work better on text data to recognize the sentiment?

1.5 Expected Output

We're going to compose a study paper on this connected topic using the Bengali language as this is research work and our ultimate objective is to go for publication once we have high accuracy. Many people are researching many niche subjects in an effort to provide Bengali-language-appropriate outcomes. However, finding a clean and excellent data collection for the Bangla language is challenging. There are many opportunities for study but we are unable to do the research since it is tough to get data sets for individuals working with the Bengali language and they are not interested in sharing their information source with everyone. We must construct our own dataset in order to carry out such research projects, but some other researchers may create data sets of a similar nature for their work since they lack access to any publicly available resources for data collection. A machine must learn about Bengali in order to understand Bengali language patterns because we're trying to accomplish sentiment categorization in Bengali. Furthermore, we are employing multiple columns of an Excel sheet that each include various full-length texts for our research, which is also known as multi-head text mining. The algorithm can infer the sentiment from every Bengali phrase once it has learned the language's patterns. With the right data collection, a machine may be taught quickly and effectively, resulting in improved outcomes when we wish to execute all of our tasks. Our machine will be trained using the appropriate data set, and after being trained using several ML and DL models, it should be able to accurately extract sentiment or emotion from any type of Bengali utterance. Additionally, it will demonstrate which ML method and DL model perform better with Bengali sentences.

1.6 Project Management and Finance

Managing a project's financial components, such as its expenditure, sales, and profitability, is known as project finance in the field of project management and finance. No one has made a financial commitment to our efforts. Our project was run by us alone. Our time and effort are the largest investments we have made in this job. By physically visiting the institutes, we gathered information from students in high school, college, and university. In Google Colab, raw Bengali data is pre-processed and trained using a variety of ML, DL, and mixed DL models that make use of freely available library functions.

1.7 Report Layout

- The project introduction, motivation, study's rationale, research questions, project management, finance, and anticipated results are all covered in the first chapter.
- The "Background" section of chapter two will cover the introduction, linked studies, research recap, extent of the issue, and obstacles.
- The third segment focuses on research methodology.
- The discussion of the experimental results will be in Chapter four.
- The effects of our work on sustainability, society, and the environment will be discussed in chapter five.
- The synthesis, completion, and recommendations for further research will all be in Chapter five.

CHAPTER 2

BACKGROUND STUDY

2.1 Preliminaries

An automated procedure can be developed using machine learning techniques. Data gathering and goal setting are the initial steps in creating an ML model, that involves properly characterizing the problem. The genre model may be accurately classified using machine learning. Numerous researchers have attempted to categorize people's health and live in covid situations utilizing novel methodologies in the past that used ML and DL techniques. We systematically displayed a few of these works in this section.

2.2 Related Works

Amar Prashad Chaudhary et. al.[3] described a study among college students in India that examined signs of anxiety, sadness, and terror related to the COVID-19 epidemic. Students were given sociodemographic surveys and psychometric tests to assess the psychological and behavioral effects of COVID-19.

University students' subjective experiences, mental health, personalities, and behaviors during the COVID-19 epidemic were detailed by Herbert, C., et al. [4]. Data analysis methods included correlational analysis, machine learning technologies, language study of one's own perception, personality, and moods during the pandemic, as well as descriptive analysis for prevalence estimation of mental health factors. The regression models and decision tree models used data analysis based on GBR and SVR.

Wang, C. et. al. [6] described, in the new semester of online learning during COVID-19 in China, undergraduate students' anxiety prevalence and intensity. They used a machine learning model based on the XGBoost model and applied the XGBClassifier function of the XGBoost module in the sklearn library. The accuracy percentage was around 80%.

Edeh Michael Onyema, et al.[9] described the impact of the COVID pandemic on education. With 200 data the results show that COVID-19 has negative impacts on education, including disruptions to instruction, a lack of access to facilities for education and research, job losses, and rising student debt. Poor network, power, accessibility, and availability concerns, as well as inadequate digital abilities, hampered online learning. Some negative implications of closing schools due to the coronavirus. Learning disruptions, poor nutrition, increased pressure on institutions that are still operating, and social isolation. The report emphasizes how important it is for instructors, students, and educational institutions to adopt technology and develop their digital literacy.

Matthew H. E. M. Browning et. al. [11] investigate the psychological impacts of COVID-19 and associated risks on college students at seven universities across the United States. Researchers preconfigure to describe the expected psychological effects of the epidemic on schoolchildren. In COVID-19-affected individuals, they examine possible sociodemographic, lifestyle, and awareness variables. 59% of students surveyed reported being negatively affected by the pandemic in some way.

Jana Shafi et al. [12] implemented a machine learning approach regarding factors associated with a different mental health outcome in a pandemic situation. It was discovered that logistic regression has a 67.14% accuracy rate and a 95% sensitivity rate for detecting individual mental stress. Approximately 80% of people were found to be stressed, compared to 140% who were not.

Machuca et al[20] proposed a sentiment analysis of English tweets during the pandemic COVID-19 in 2020. They utilized a dataset with over 16,00,000 tweets that were categorized as either positive or negative. 50,000 top English tweets during each month with the hashtag #coronavirus were downloaded. They achieved a classification accuracy of 78.5% using the logistic regression algorithm.

Rustam et al [21] performed Covid-19 tweet sentiment analysis using a supervised machine learning approach and collected 7528 tweets from the IEEE data port. Researchers have observed that tree-based models performed better for BoW than TF-IDF. With an accuracy score of 0.577 on the chosen dataset, their LSTM model performs poorly. Additionally, employing the BiLSTM and CNN-LSTM, they obtained accuracy scores of 0.579 and 0.61, respectively.

Kunal Chaturbedi et. al. [24] investigated and analyzed the potential consequences of the COVID-19 pandemic on the life of students. According to a study, 38.3% of students felt negative about taking classes online. Mainly, 51.4% of respondents reported that they did not utilize their time during the period of lockdown.

BA Bhat et.al [25] studied to understand the psychological impact, anxiety, depression, and stress that disturbed the social life of people during the initial stage of the COVID-19 outbreak. The spread of COVID-19 infection can result in many new problems, such as psychological problems (67.5% of respondents) and social problems (53.5%).

Shatla et.al [26] sought a study to assess how COVID-19 affected Saudi Arabian medical students' studies, social lives, and mental health. Of the 1253 data collected, 52% reported satisfaction with the utilization of their time, while 36.3% never engage in physical activity, and 42.4% reported increased body weight. Their study had certain drawbacks because it was restricted to medical students in various Saudi Arabian provinces.

2.3 Comparative Analysis and Summary

In this experiment, we have aimed to describe how the COVID epidemic affects people's academic, social, and mental health and also other instances. We have collected data from students. We must filter or otherwise prepare the data after collection. Many library features are available for both English and other languages. We confront several difficulties since there aren't many prebuilt libraries in NLP. We have to alter several libraries in accordance with our study because we are utilizing our data set with Bengali as the language. We did our best to clean up our data. In order to extract the sentiment from any Bengali remarks, we made an effort to include machine learning and deep learning methods in our work.

Comparison work described below in table 2.3

Table 2.3: Comparative analysis of previous work

Reference	Data Type	Data Quantity	Performance Evaluation
[23] Using the BERT model to analyze the sentiment of tweets	English text	596,784 tweet data	BERT model: 94%
[28]Sentiment analysis algorithm study of user sentiment	English Text	25,000 tweet data	VADER: 83% BERT:92% LR:88%
[29]Emotion Detection Based on COVID-19 Social Media Response from Bangladeshi People.	Translated to Bangla language	10,581 social media data	Random Forest: 78.94%, SVM: 75.72%, LSTM: 84.92%, CNN: 81.91%.
[30] Advocates the use of an RNN language model based on LSTM	16000 data+ 61448 data+50000 data+39000 data	travel comments(trip)+travel comments(JD.COM) +English movie reviews+comments from JD.COM	(utilizing travel comments(JD.COM) LSTM:95.62%(Positive), 93.07%(Negative), 86.58%(Neutral) RNN:94.83%(P), 91.7%(Neg), 85.73%(Neu)
[31]Analyzed depression using an LSTM Deep Recurrent Network.	Bangla tweets	1968 tweet data	LSTM model -size 128 of value 78.9%, -Batch size 25 with 10 epochs achieved 81.1%, and for 5 layers with 20 epochs 86.3%.

2.4 Scope of the Problem:

Data collection is the biggest issue we are facing in our research. We faced many problems in collecting data from students. Also, they gave yes-or-no responses, which did not bring any meaning to our work; that's the reason we eliminated those responses. They occasionally responded in English, and as we work on Bangla text classification, we rewrite those comments in Bangla. All those have taken lots of time.

2.5 Challenges:

Sentiment analyses are used to explore human feelings found in documents. For the circumstances students encountered during COVID and sentiment analysis, we will illustrate the supplication of both ML and deep learning. In this problem, text classification techniques are also used. The experiment was carried out on the dataset of Bengali statements, showing how the chosen strategy may be improved and evaluated. Another difficult task is forecasting the feeling of the Bengali text using NLP. To extract the punctuation from the Bengali text, Unicode is employed. There is no library for stop words; we collected Bangla stop words online. A large vocabulary is another challenge for this approach. Comprehensive data collection provides a large vocabulary and precise forecasting of the sensation.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

In this part of the report, we will go through the entire technique of our research work. Here, we go over our data collection approach, pre-processing, algorithms employed, and workflow for our work. The primary goal of our findings is to examine how the COVID epidemic affects people's academic, social, and mental health. We apply machine learning approach to optimize the accuracy of the supervised learning classifier in our dataset. Deep learning techniques were also employed to assess and resolve text-related issues. LSTM and BiLSTM are effective deep-learning methods in this case. We also employed CNN, CNN-LSTM, and CNN-BiLSTM. Each phase of the process is briefly covered in this section.

A work procedure flowchart for our research work shown in figure 3.1

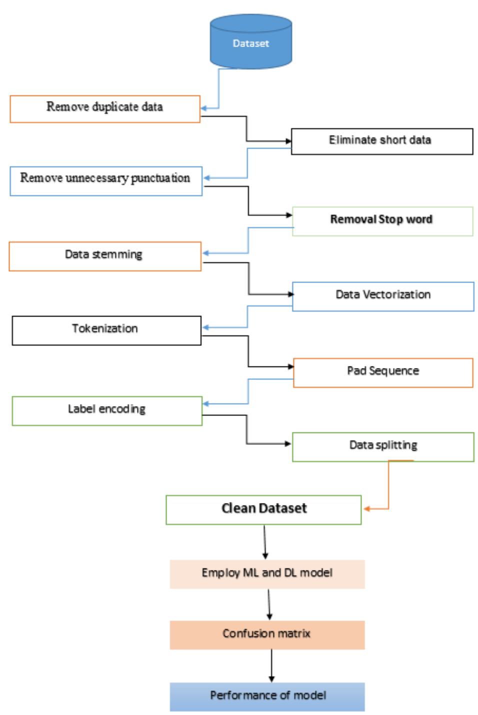


Figure 3.1: Working Flowchart

3.2 Subject of Research and Equipment

Our research focuses on how the COVID epidemic affects students' academic, mental and social lives with the help of Bangla NLP (Natural Language Processing). For this, we use some development tools.

Here is a list of the instruments needed for this model.

- Windows 10
- Pyhton
- Colab
- Bangla NLP
- Pandas
- NumPy
- Matplotlib
- Sklearn library
- Tfidf Vectorization
- Word Cloud
- Tensorflow
- Keras
- Label Encoder
- Spilting function
- Classification Report
- Confusion matrix

3.3 Procedure for Collecting Data

We gathered information physically from students using questionnaires through a Google Form. These students belong to various levels of education, including high school, college, university, and masters. Students respond to the questionnaire by indicating how positively or negatively COVID influences their lives. Out of a total of 400, there were 200 positive and 200 negative data. We deploy the same quantity of information for both positive and negative responses so that training and testing the data would be made easier.

A summary of our dataset is provided in table 3.3 below

Table 3.3: Dataset Preview

Impact on academic life	Impact on mental health	Impact on social life	Label
অনলাইন থেকে আমার কাজের দক্ষতা বৃদ্ধির জন্য আমি অনেক কিছু শিখেছি	পরিবারের সবার সাথে অনেক সময় কাটানো হয়েছে তাই আমার মন সবসময় প্রফুল্ল থাকতো	পরিবার আত্মীয়-স্বজন এবং বন্ধু-বান্ধবের সাথে আগের চেয়ে বেশি যোগাযোগ হতো	Positive
অনলাইন শিক্ষাব্যবস্থা আমার কাছে খুবই চাপ যুক্ত মনে হয়েছে, যার ফলে আমি পড়াই তেমন মনোযোগী হতে পারিনি	অনলাইন শিক্ষা ব্যবস্থার ক্ষেত্রে আমি খুবই মানসিক চাপ অনুভব করতাম	অনলাইন শিক্ষা ব্যবস্থার ক্ষেত্রে আমি খুবই মানসিক চাপ অনুভব করতাম	Negative
আমাদের শিক্ষাব্যবস্থা অনলাইন হওয়ার ক্ষেত্রে পড়ার প্রতি মনোযোগ অনেক কম হয়ে গিয়েছিল	লকডাউন এর কারণে দিন দিন ডিপ্রেশনের কারণ বেড়েই চলছিল	করোনার কারণে সামাজিক দূরত্ব অনেক বেড়ে যায়	Negative
আমি পড়াশোনা অনলাইন হওয়ার জন্য অনেক কিছু শিখতে পেরেছি	লকডাউনে আপনজনদের সাথে থাকার কারণে আমি সব সময় হাসি খুশি থাকতাম এবং খুব ভালো কেটেছে দিন	বাবা মা ভাই বোনের সাথে অনেক ভালো সময় কেটেছে এবং আমাদের মধ্যে সম্পর্ক আরও অনেক ভালো হয়েছে	Positive

3.4 Data Preprocessing

The initial step after collecting data is to clean the dataset. Data preprocessing is the act of converting unstructured and raw data into usable sets of information in order for data mining to work as expected. Procedures such as cleaning null data, eliminating punctuation, label encoding, and tokenization must be performed in order to create clean data. When label encoding is used, the text is first labeled, then an instantaneous label is applied, and the result is processed using a large dataset.

3.4.1 Removal Punctuation

As Bengali has so few unique punctuation marks of its own, we must eliminate them from our data in order to make it more readable. The elimination of punctuation is a crucial stage in the compilation of data.

Figure 3.4.1 shows dataset preview after cleaning

```
Original:
 অনলাইন ক্লাসের পড়া বুঝতে অনেক অসুবিধা হতো, আমি পড়াশোনা ভালো করতে পারিনি
Cleaned:
 অনলাইন ক্লাসের পড়া বুঝতে অনেক অসুবিধা হতো  আমি পড়াশোনা ভালো করতে পারিনি
 Sentiment:--  Negative

Original:
 আর্থিক সমস্যার কারণে আমি স্মার্টফোন ল্যাপটপ কিনতে পারিনি ,তাই প্রথমদিকে অনলাইনে পড়াশুনায় অনেক অসুবিধায় পড়েছিলাম
Cleaned:
 আর্থিক সমস্যার কারণে আমি স্মার্টফোন ল্যাপটপ কিনতে পারিনি  তাই প্রথমদিকে অনলাইনে পড়াশুনায় অনেক অসুবিধায় পড়েছিলাম
 Sentiment:--  Negative

Original:
 সব সময় ডিভাইস নিয়ে ব্যস্ত থাকতাম, পড়াশোনা কিছুই হয়নি, ফলাফল খারাপ হতো
Cleaned:
 সব সময় ডিভাইস নিয়ে ব্যস্ত থাকতাম  পড়াশোনা কিছুই হয়নি  ফলাফল খারাপ হতো
 Sentiment:--  Negative
```

Figure 3.4.1 After data purification and punctuation removal

3.4.2 Removal of small length of text

A complete sentence makes more sense than a small amount of text, because small text doesn’t add any meaning to the sentence. As a result, texts with lengths less than three were removed to clean our dataset.

3.4.3 Removal Stop word

One common NLP operation is to remove stop words from the text. Stop words are mostly used to eliminate superfluous words from sentences. As we mentioned before that bangla NLP is not rich yet, so there are no pre-built libraries to stop word removal. Therefore, it is up to us to eliminate stop words. Prior to anything else, we're compiling every Bengali stop word we can find on the internet. There are a total of 732 stop words, which we then added to a register for future usage.

3.4.4 Stemming

Stemming is the method of stripping the last few letters from a word, which frequently results in inaccurate spelling and meanings. We use stemming to break words down to their most fundamental form. It could or might not be a word that is understood in the language. It is analogous to trimming a tree's branches down to their stems.

For example, the stem of the words 'আমি' 'আমাদের' is 'আমা'

As shown in figure 3.4.4 below, we also created a word cloud using stemming words for additional visualization.

Figure 3.4.4: An illustration of a stemming word

3.4.5 Data Vectorization

To count how many times a word appears in a text or document, we used CountVectorizer. CountVectorization is accomplished using the Python sklearn library.

- Unigram: In order to generate unigrams or 1-grams, we input the value of n=1 to the n-grams function. We also determined the word frequency of the words.

- Bigram: In order to generate bigrams or 2-grams, we input the value of n=2 to the n-grams function. We also determined the word frequency of the words.
- Trigram: In order to generate trigrams or 3-grams, we input the value of n=3 to the n-grams function. We also determined the word frequency of the words.

The provided example in table 3.5 shows incorporating count vectorization into a sentence.

Table 3.5: Example of n-gram distribution

Sentence	Unigram	Bigram	Trigram
লকডাউন এর সময় অনেক নতুন নতুন বন্ধুবান্ধব তৈরি হয়েছে এবং অনেক পুরনো বন্ধুদের হারিয়েছি এবং পরিবারের থেকে অনেক দূরে চলে গেছি মানসিকভাবে	('অন', 3), ('নত', 2), ('বন', 2), ('এব', 2), ('লকড', 1), ('উন', 1), ('এর', 1)	('লকড উন', 1), ('উন এর', 1), ('এর সময', 1), ('সময অন', 1), ('অন নত', 1), ('নত নত', 1), ('নত বন', 1)	('লকড উন এর', 1), ('উন এর সময', 1), ('এর সময অন', 1), ('সময অন নত', 1), ('অন নত নত', 1), ('নত নত বন', 1), ('নত বন ধব', 1)

We also determined the overall word frequency in the three input columns using count vectorization.

Example for a unigram academic input column in figure 3.5

```
unigram_words = get_top_n_words(df['Cleaned_Sentence'],20)
unigram_words
```

```
[('আম', 284),
 ('পড', 231),
 ('ইন', 158),
 ('অন', 145),
 ('কর', 143),
 ('অনল', 140),
 ('মন', 97),
 ('ফল', 96),
 ('রণ', 90),
 ('এব', 82),
 ('সময', 81),
 ('হয', 76),
 ('করত', 67),
 ('হত', 55),
 ('জন', 50),
 ('পর', 48),
 ('নত', 44),
 ('খত', 43),
 ('দক', 38),
 ('ষত', 38)]
```

Figure 3.5: Unigram CountVectorization

3.4.6 Tokenization

A text stream is tokenized when it is divided up into sentences, phrases, symbols, or other relevant components, called tokens. It is the technique of separating distinct pieces from a text corpus.

3.4.7 Pad Sequence

The keras pad_sequences tool was used to preprocess the sequential data. Sample sequences are transformed into 2D numpy arrays using keras pad sequences. To make sure that every sequence in a list is the same length, use pad sequences.

3.4.8 Label encoding

Machine only understand the binary values 0 and 1 instead of any text. So, converting the label text into a machine-readable value for data encoding. This is the most important preprocessing step for machine learning and deep learning models before they can be processed further.

3.4.9 Data spliting

The Split function was used to divide the dataset into two parts: training and testing. We train the model with the training set of data and then use the test set to assess the model's effectiveness. We used 80% of the data for training and 20% for testing in supervised machine learning models. For deep learning models, we used 70% training, 20% validation, and 10% testing data.

3.5 Statistical Analysis

1. We used dataset which was collected from students' responses about their academic, mental, and social lives and divided those responses into two groups positive and negative.

Distributing label bar chat visualize in figure 3.5 below.

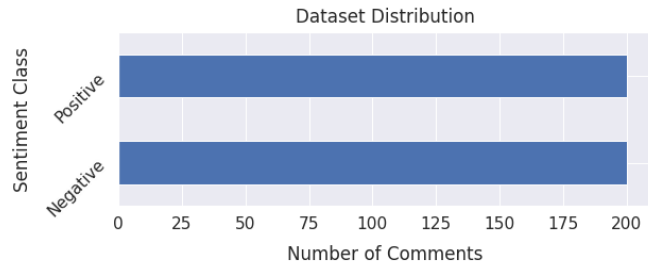


Figure 3.5.1 : Label Distribution Bar Chart

2. The following statistics are based on total sentences, total words, and unique works in good and negative comments:

In figure 3.5.2 visualize the statistics inside of a sentence

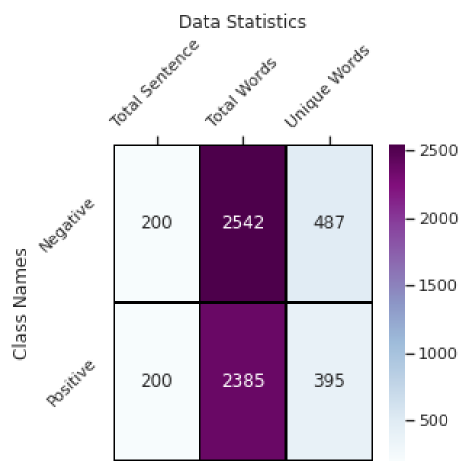


Figure 3.5.2: Statistics inside of a phrase with a class name

3. Figure 3.5.3 visualize a bar Chart of COVID Students' Attitudes toward Online Learning

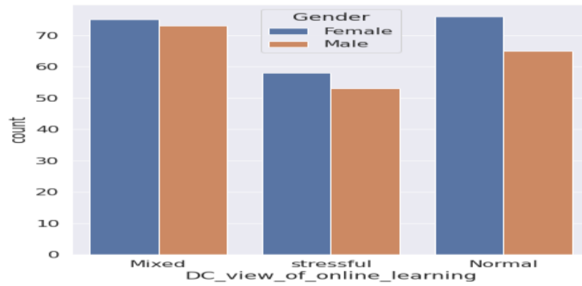


Figure 3.5.3: Bar Chart of COVID Students' Attitudes Toward Online Learning

4. Academic pressure on students is depicted in a bar chart in Figure 3.5.4 below during the COVID.

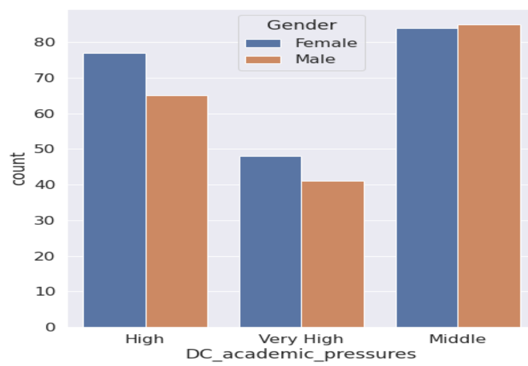


Figure 3.5.4: Bar Chart of Academic Pressure on Students During Covid

5. The Bar Chart of Students' COVID Anxiety Level is shown in Figure 3.5.5 below.

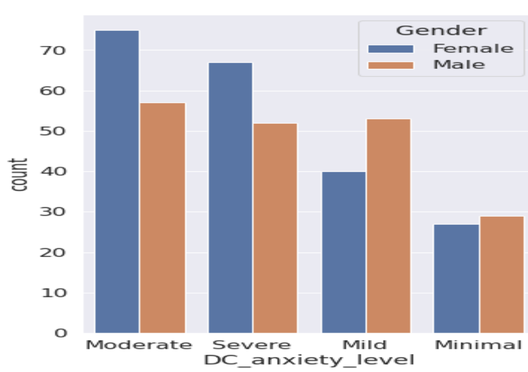


Figure 3.5.5: Bar Chart of During Covid Anxiety level of students

3.6 Implementation Requirements

We have used some supervised ML and DL algorithms.

3.6.1 Supervised Machine Learning

The categorization issue is a prevalent way to formulate the supervised learning problem. Figure 3.8.1 demonstrates how supervised ML is used in a real-world application. The major agenda of our research is to classify ML algorithms while also identifying the most effective method with maximum accuracy and precision.

Process diagram for supervised machine learning shows figure 3.6.1 below

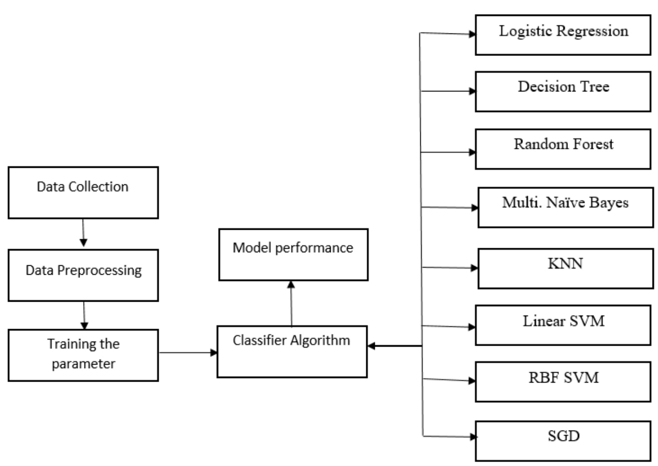


Figure 3.6.1: Process diagram for supervised machine learning

Some Supervised Machine Learning algorithms are:

3.6.1.a Logistic Regression Classifier

The best regression approach to use when the dependent variable is dichotomous is logistic regression. For example, given the student response, will it be positive or negative?

A logistic regression function:

$$\mathrm{Sig}(x) = \frac{1}{1+e^{-x}} \qquad [32]$$

- E stands for log base.
- X represents the numerical value that must be changed.

3.6.1.b Decision Tree Classifier

Decision Tree, a supervised ML technique, uses a set of principles to make decisions, much like humans do. It is the most powerful and widely used tool for categorization and prediction. It also referred CART algorithm: Classification and Regression Tree. Reason being that Decision Trees are capable of doing both classification and regression tasks. A decision tree arranges a succession of roots in a tree form. In essence, a decision tree uses the decision node to partition the dataset recursively until only leaf nodes are left. By maximizing entropy gain, it also determines the optimal split.

3.6.1.c Random Forest Classifier

RF is a collection of several random decision trees that is heavily reliant on training set. For each tree, we'll choose a subset of characteristics at random and utilize just those for training. Random feature selection increases variation by reducing the correlation between the trees. As a result of some of the trees being trained on less significant attributes, they will make poor forecasts, but there will also be some trees that make poor predictions in the opposite way, balancing things out and predicting the proper outcome.

In figure 3.6.1.1 shows random forest technique of operation

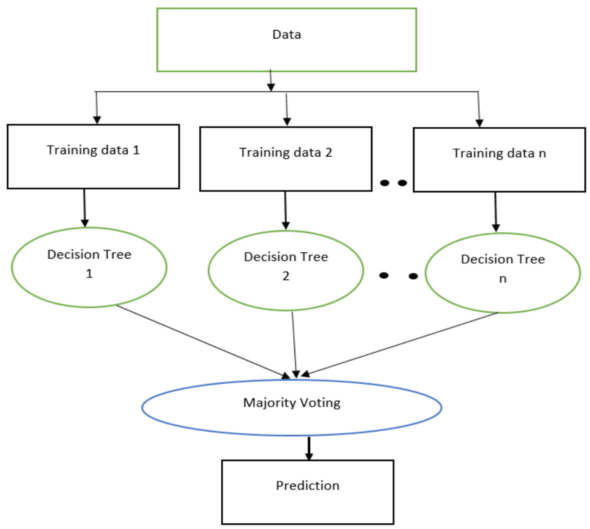


Figure 3.6.1.1: Random Forest technique of operation

3.6.1.d Multinomial Naive Bayes Classifier

The probabilistic learning technique known as the Multinomial Naive Bayes algorithm is mostly utilized in NLP. Since Bangla NLP is the foundation of our research work, we applied it. Using the Bayes principle, the computer provides an educated prediction about the tag of a text, such as an email or news article. In order to provide the tag with the highest likelihood, it calculates the likelihood of each tag for a certain sample.

In figure 3.6.1.2 shows multinomial naïve bayes technique of operation

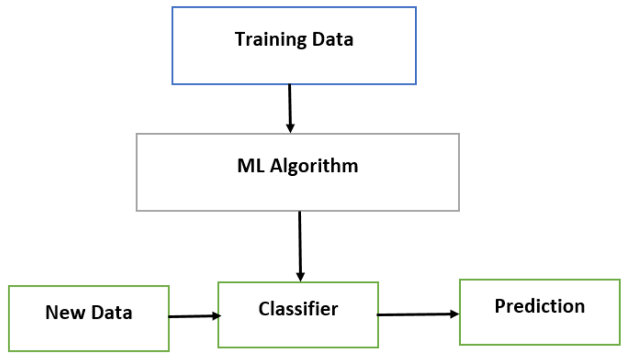


Figure 3.6.1.2: Multi. Naive Bayes' technique of operation

3.6.1.e k-Nearest Neighbors Classifier

In supervised machine learning, KNN is a very straightforward method of classifying data. The point is classed for the classification issue by a vote of its neighbors, and then the point is given the class that is most popular among its k closest neighbors. The k parameter of the k-NN algorithm determines how many neighbors will be looked at in order to categorize a certain query point. When k is low, the bias is often low but the variance is high, and when k is high, the bias is typically high but the variation is low.

3.6.1.f Linear SVM Classifier

Linear SVM are the most basic and, perhaps, the most elegant classification algorithms. To categorize, each item is represented as a point in an n-dimensional space, and the coordinates of the points are commonly referred to as features. By drawing a hyperplane with all the points from one category on one side and all the points from the other category on the other, it conducts the classification test.

3.6.1.g RBF SVM Classifier

One of the fastest, most practical, and well-liked kernels in the Support Vector Machine family of classifiers is the RBF kernel. This kernel performs classification utilizing the fundamentals of Linear SVM by first mapping the data into a high-dimensional space using the dot products and squares of all the features in the dataset. The so-called radial basis function, which may be expressed as:

$$K(X_1, X_2) = \exp\left(-\frac{||X_1 - X_2||^2}{2\sigma^2}\right) \qquad [33]$$

The Squared Euclidean Distance in this case is denoted as ||X1 - X2||^2, and σ is a free parameter that may be utilized to fine-tune the equation.

3.6.1.h Stochastic Gradient Descent

Machine learning applications employ stochastic gradient descent to find the model parameters that best match the predicted and observed results when dealing with large amounts of data. It is a crude yet effective method. SGD would handle the data by selecting one sample at random for each step and only using that sample to compute the derivatives. However, it is more common to select a small subset of data for each step.

3.6.2 Deep Learning

With the use of complex algorithms and vast volumes of data, deep learning is a type of machine learning that trains models. The open-source library TensorFlow was created by Google particularly for deep learning applications.

Process diagram for deep learning shows figure 3.6.2 below

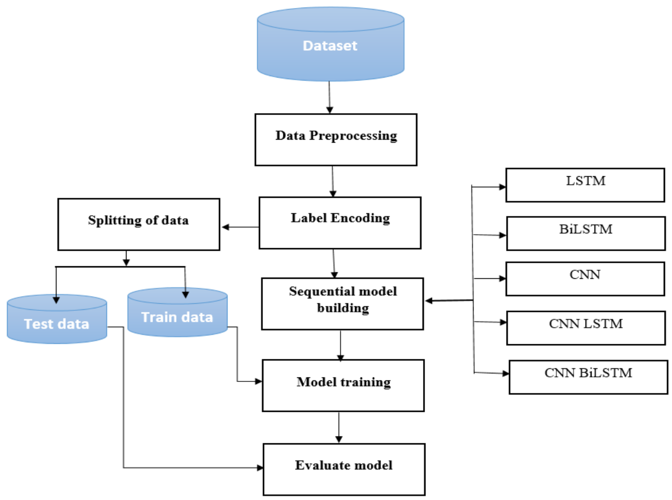


Figure 3.6.2 Deep learning process diagram

3.6.2.a Sequential Model

The sequential approach enables us to define a neural network precisely sequentially, moving from input to output while traversing a sequence of sequential neural layers. Figure 3.8.2.1 shows sequential model's data movement.

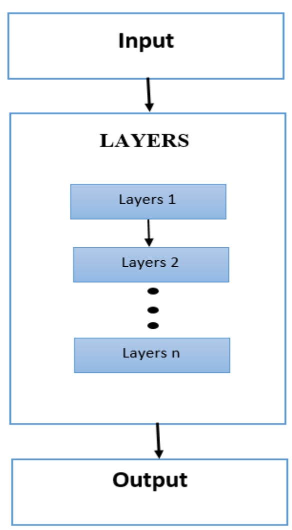


Figure 3.6.2.1 Sequential model's data movement

Details Explanation of the deep learning models used in our research:

3.6.2.b LSTM

Long-Short Term Memory (LSTM) allows a neural network to remember what it needs to remember in order to maintain context, but also to forget what is no longer relevant. By adding something called an internal state to the RNN node, LSTM offers a solution to the long-term reliance issue in RNN. The LSTM cell is made up of three parts: an input gate, a forget gate and an output gate.

The forget gate specifies the types of internal state information that may be erased when they are no longer pertinent to the current context.

The input gate specifies the new data that should be added to or updated in the working storage state data.

The output gate specifies which information, out of all that is kept in that storage state, should be output.

3.6.2.c BiLSTM

Rather than merely encoding the sequence in the forward direction, we also encode it in the reverse way and concatenate the results from both the forward and reverse LSTMs at each time step. Each word's encoded representation now knows the words that come before and after it. That BiLSTM can only be used for sequence encoding.

3.6.2.d CNN

Although Convolution Neural Network (CNN) is typically utilized in computer vision, they have lately been utilized in several NLP tasks, like machine translation, sentiment classification, answer selection etc.To conduct text categorization on our dataset, we will be training a convolutional neural network in this section. To train our data on the CNN architecture, we use Keras. It also assesses the accuracy discovered on the validation set. The following layers are often found in a convolutional neural network: convolutional layers, pooling layers, dense layers.

In addition, we use CNN LSTM and CNN BiLSTM hybrid models in our research.

3.6.2.e CNN LSTM

The definition of CNN with LSTM is to add CNN layers on the front end, LSTM layers, and a Dense layer on the output. CNN layers are used in the CNN LSTM architecture to pull features from input data, and LSTMs are used to assist sequence prediction. In Keras, we may provide a CNN LSTM model that will be jointly trained.

3.6.2.f CNN BiLSTM

Bidirectional LSTM and CNN are combined to create CNN BiLSTM. Through the use of the CNN component, character-level features are induced. In order to extract a new feature vector for each word, the model uses a max pooling layer and a convolution, while word-level characteristics are exploited by BiLSTM.

3.6.3 Confusion Matrix

An analysis of a classifier's estimates' accuracy is called a confusion matrix. It is used to assess the potency of a categorization model. It may be used to determine performance metrics like accuracy, precision, recall, and F1-score to evaluate how well a classification model is doing.

3.6.4 Evaluation of model

After preprocessing the data and building the sequential model, we must now divide the data set into the train set and the test set. Following data analysis, the machine will learn from the train set. To obtain the precise result, the test set will apply the information acquired from the train set. As a result, the model will be used to train and evaluate the pre-processed data.

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental Setup

A crucial area of artificial intelligence is natural language processing. One component of NLP is sentiment analysis. Sentiment analysis in bangla texts is a challenging task. These are used to train the model once the data have been collected and cleaned or preprocessed. The algorithm can then forecast the sentiment based on training when put to the test. Machines are rarely 100% correct, thus they provide the highest likelihood based on their training. Tensorflow, Sklearn, Numpy, Pandas, Matplotlib, Seaborn, and other tools were employed in our work. Logistic regression, decision tree classifier, random forest, multi-naive bayes, KNN, linear SVM, RBF SVM and SGD are a few examples of machine learning models that are employed. Deep learning approach using LSTM, BiLSTM, CNN, CNN LSTM, CNN BiLSTM. The batch size had to be 64, and the number of epochs had to be 10. We employed an effective GPU surface to build our model in Google Cobal and shorten the training period.

4.2 Experimental Results & Analysis

We are aware that if a task is completed manually by a human, it will result in 100% output, however machines cannot. In a similar vein, we train our model in our work after preprocessing the data we obtained and making certain adjustments to obtain a more accurate outcome. As we previously stated, the model's output is occasionally extremely accurate when we train it. It produced the most optimal result. Based on the provided statement, it can determine the appropriate emotion. To determine how well the model can predict the sentiment of Bangla text, we employed several of ML models and the deep learning models. In order to lessen the loss of data, we employed 10 epochs. Additionally, Tansorflow and Matplotlib were used to display various data accuracy in a graphical manner. The model is adequate to provide decent accuracy. We divide our sentiment analysis into two classes: positive class and negative class. For each of these classes, 200 data are present. Based on the Bangla text regarding the COVID-19 influence that was obtained from the student through a survey, our model is capable of predicting both positive and negative attitudes. It will identify the text's mood and provide the appropriate sentiment class.

Classifier Algorithm Name	**Abbreviation**
Logistic Regression	LR
Decision Tree	DT
Random Forest	RF
Multinomial Naive Bayes	Multi.NB
K-Nearnest Neaighoure	KNN
Linear Support Vector Machine	Linear SVM
Radial Basis Function SVM	RBF SVM
Stochastic Gradient Descent	SGD

.

4.2.1 Analysis Experiment result of Supervised Machine Learning models

4.2.1.1 Academic Life Impact Visualization:

- Unigram Feature Performance Table for Impact on Academic Life shown in table 4.2.1.1.A

Table 4.2.1.1.A : Unigram Feature Performance Table for Impact on Academic Life

Classifier Name	Accuracy	Precision	Recall	F1 Score
LR	91.25	96.88	83.78	89.86
DT	82.5	82.86	78.38	80.56
RF	90.0	91.43	86.49	88.89
Multi. NB	87.5	88.57	83.78	86.11
KNN	91.25	87.5	94.59	90.91
Linear SVM	85.0	93.10	72.97	81.82
RBF SVM	88.75	93.75	81.08	86.96
SGD	92.5	91.89	91.89	91.89

- Bigram feature Performance Table for impact on Academic life shows in table 4.2.1.1.B

Table 4.2.1.1.B: Bigram feature Performance Table for impact on Academic life

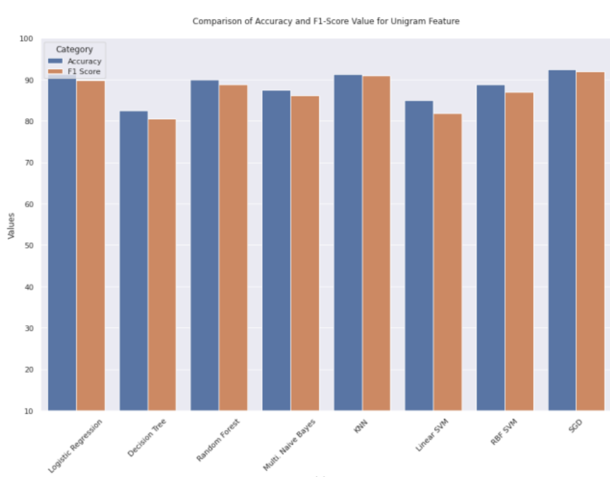
Classifier Name	Accuracy	Precision	Recall	F1 Score
LR	91.25	96.88	83.78	89.86
DT	87.5	86.49	86.49	86.49
RF	87.5	90.91	81.08	85.71
Multi. NB	90.0	91.43	86.49	88.89
KNN	81.25	78.95	81.08	80.0
Linear SVM	86.25	100.0	70.27	82.54
RBF SVM	88.75	100.0	75.68	86.15
SGD	92.50	91.89	91.89	91.89

-Trigram feature Performance Table for impact on Academic life shows in table 4.2.1.1.C

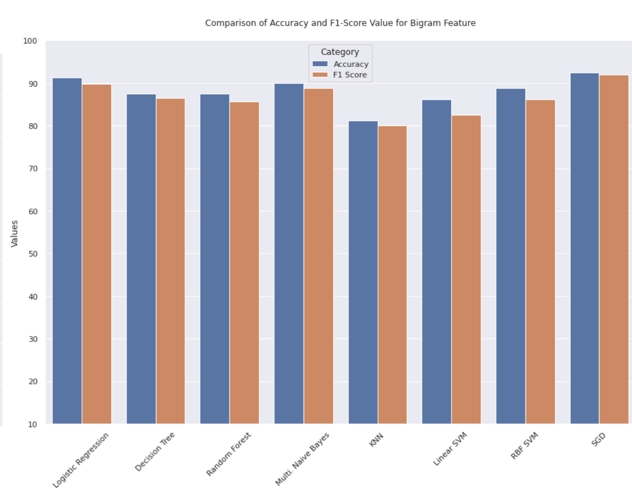
Table 4.2.1.1.C: Trigram feature Performance Table for impact on Academic life

Classifier Name	Accuracy	Precision	Recall	F1 Score
LR	92.50	96.97	86.49	91.43
DT	82.50	89.66	70.27	78.79
RF	88.75	93.75	81.08	86.96
Multi. NB	88.75	88.89	86.49	87.67
KNN	83.75	81.58	83.78	82.67
Linear SVM	81.25	71.15	100.0	83.15
RBF SVM	88.75	100.0	75.68	86.15
SGD	95.0	97.14	91.89	94.44

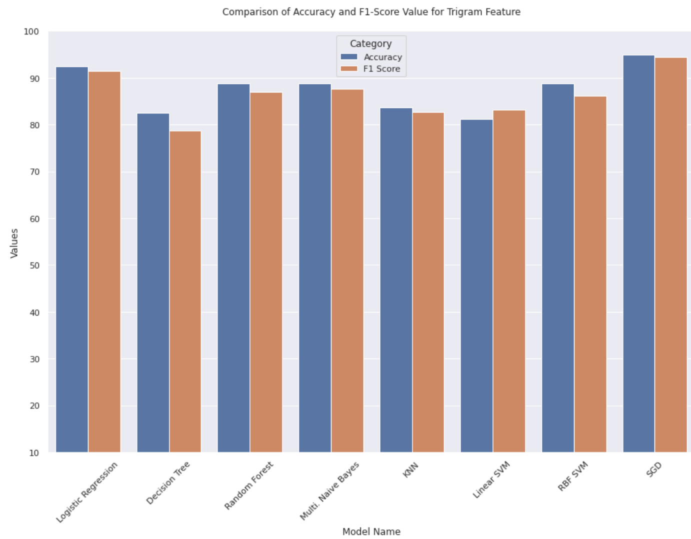
Visualization of bar chart of Unigram, Bigram and Trigram feature Accuracy and f1 score for different ML model for impact on Academic life Figure 4.2.1.1.1



(a)



(b)



(c)

Figure 4.2.1.1.1 : Bar chart of Vectorization feature for impact on Academic life (a) Bar chart of Unigram feature (b) Bar chart of Bigram feature (c) Bar chart of Trigram feature

Accuracy, recall, precision, and F1 score table for ML models for academic life in 4.2.1.1.D

Table 4.2.1.1.D: Table of machine learning model accuracy, precision, recall, and F1 scores

Classifier Name	Accuracy	Precision	Recall	F1 Score
LR	92.50	96.97	86.49	91.43
DT	82.50	89.66	70.27	78.79
RF	88.75	93.75	81.08	86.96
Multi. NB	88.75	88.89	86.49	87.67
KNN	83.75	81.58	83.78	82.67
Linear SVM	81.25	71.15	100.0	83.15
RBF SVM	88.75	100.0	75.68	86.15
SGD	95.0	97.14	91.89	94.44

4.2.1.2 Mental Health Impact:

Unigram Feature Performance Table for Impact on Mental health shows in table 4.2.1.2.A

Table 4.2.1.2.A: Unigram feature Performance Table for impact on mental health

Classifier Name	Accuracy	Precision	Recall	F1 Score
LR	90.0	89.19	89.19	89.19
DT	85.0	90.32	75.68	82.35
RF	91.25	94.12	86.49	90.14
Multi. NB	86.25	86.11	83.78	84.93
KNN	93.75	92.11	94.59	93.33
Linear SVM	90.0	93.94	83.78	88.57
RBF SVM	90.0	93.94	83.78	88.57
SGD	93.75	97.06	89.19	92.96

Bigram Feature Performance Table for Impact on Academic Life shows in table 4.2.1.2.B

Table 4.2.1.2.B: Bigram feature Performance Table for impact on mental health

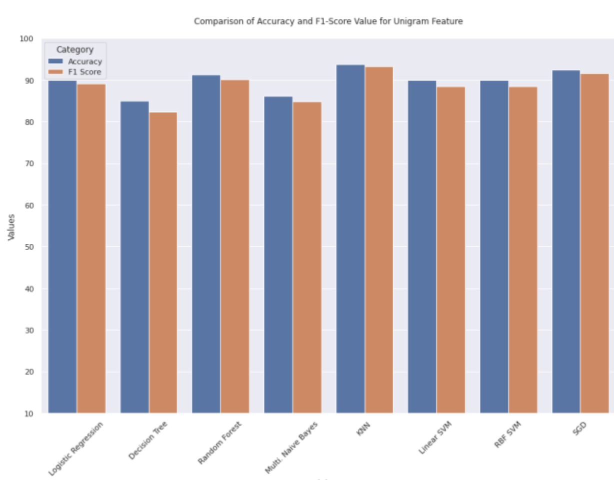
Classifier Name	Accuracy	Precision	Recall	F1 Score
LR	90.0	89.19	89.19	89.19
DT	80.0	83.87	70.27	76.47
RF	86.25	93.33	75.68	83.58
Multi. NB	93.75	97.06	89.19	92.96
KNN	93.75	97.06	89.19	92.96
Linear SVM	87.50	90.91	81.08	85.71
RBF SVM	87.50	90.91	81.08	85.71
SGD	92.50	96.97	86.49	91.43

Trigram Feature Performance Table for Impact on Mental health shows in table 4.2.1.2.C

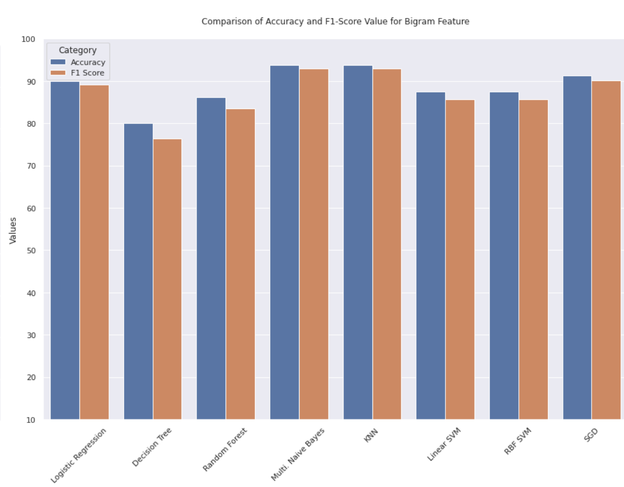
Table 4.2.1.2.C: Trigram feature Performance Table for impact on Mental health

Classifier Name	Accuracy	Precision	Recall	F1 Score
LR	90.0	89.19	89.19	89.19
DT	86.25	93.33	75.68	83.58
RF	85.0	90.32	75.68	82.35
Multi. NB	92.50	96.97	86.49	91.43
KNN	93.75	97.06	89.19	92.96
Linear SVM	88.75	83.33	94.59	88.61
RBF SVM	90.0	93.94	83.78	88.57
SGD	91.25	94.12	86.49	90.14

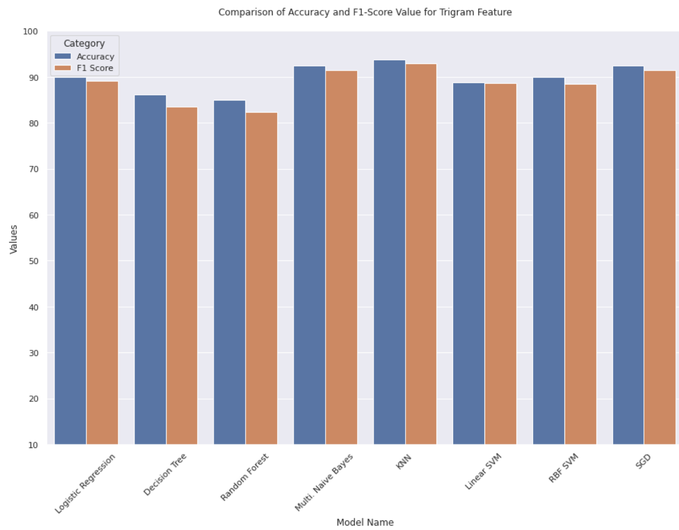
Figure 4.2.1.2.1 shows a representation of the Unigram, Bigram, and Trigram feature bar chart accuracy and f1 score for several ML models for an influence on mental health.



(a)



(b)



(c)

Figure 4.2.1.2.1: Bar chart of Vectorization feature for impact on Mental health (a) Bar chart of Unigram feature (b) Bar chart of Bigram feature (c) Bar chart of Trigram feature

Accuracy, precision, recall and F1 score table for machine learning models for mental health in 4.2.1.2.D below

Table 4.2.1.2.D: Table for accuracy, precision, recall, and F1 scores for machine learning models

Classifier Name	**Accuracy**	**Precision**	**Recall**	**F1 Score**
LR	90.0	89.19	89.19	89.19
DT	86.25	93.33	75.68	83.58
RF	85.0	90.32	75.68	82.35
Multi. NB	92.50	96.97	86.49	91.43
KNN	93.75	97.06	89.19	92.96
Linear SVM	88.75	83.33	94.59	88.61
RBF SVM	90.0	93.94	83.78	88.57
SGD	91.25	94.12	86.49	90.14

4.2.1.3 Social Life Impact:

Unigram Feature Performance Table for Impact on Social life shows in table 4.2.1.3.A

Table 4.2.1.3.A: Performance of the Unigram feature as it relates to social life

Classifier Name	Accuracy	Precision	Recall	F1 Score
LR	90.0	87.18	91.89	89.47
DT	83.75	78.57	89.19	83.54
RF	86.25	86.11	83.78	84.93
Multi. NB	95.0	97.14	91.89	94.44
KNN	91.25	89.47	91.89	90.67
Linear SVM	90.0	87.18	91.89	89.47
RBF SVM	91.25	89.47	91.89	90.67
SGD	93.75	100.0	86.49	92.75

Bigram Feature Performance Table for Impact on social life shows in table 4.2.1.3.B

Table 4.2.1.3.B: Bigram feature performance Table for impact on social life

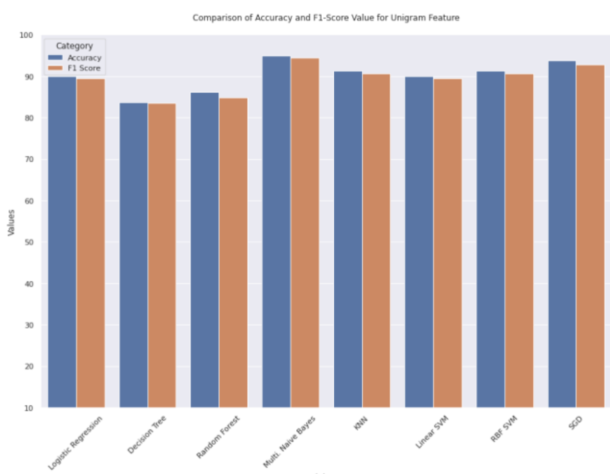
Classifier Name	Accuracy	Precision	Recall	F1 Score
LR	91.25	87.50	94.59	90.91
DT	81.25	73.91	91.89	81.93
RF	90.0	85.37	94.59	89.74
Multi. NB	93.75	90.0	97.30	93.51
KNN	90.0	87.18	91.89	89.47
Linear SVM	87.50	81.40	94.59	87.50
RBF SVM	91.25	87.50	94.59	90.91
SGD	95.00	92.31	97.30	94.74

Trigram Feature Performance Table for Impact on social life shows in table 4.2.1.3.C

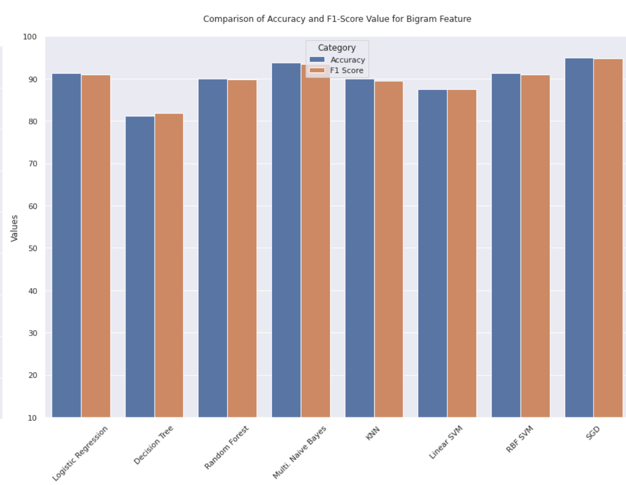
Table 4.2.1.3.C: Trigram's Performance Table for Social Life Impact

Classifier Name	Accuracy	Precision	Recall	F1 Score
LR	92.50	87.80	97.30	92.31
DT	78.75	69.23	97.30	80.90
RF	90.0	85.37	94.59	89.74
Multi. NB	95.0	92.31	97.30	94.74
KNN	90.0	85.37	94.59	89.74
Linear SVM	81.25	72.0	97.30	82.76
RBF SVM	88.75	81.82	97.30	88.89
SGD	95.0	92.31	97.30	94.74

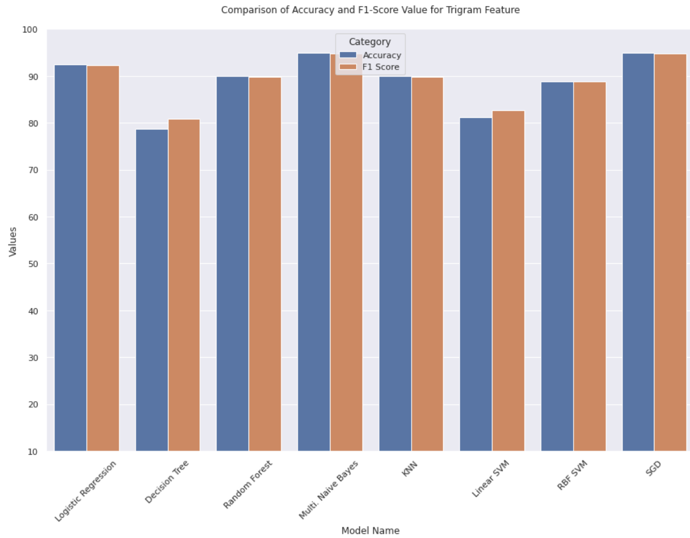
Bar graph visualization of Unigram, Bigram, and Trigram feature accuracy and f1 score for various ML models for impact on Social Life Figure 4.2.1.3.1



(a)



(b)



(c)

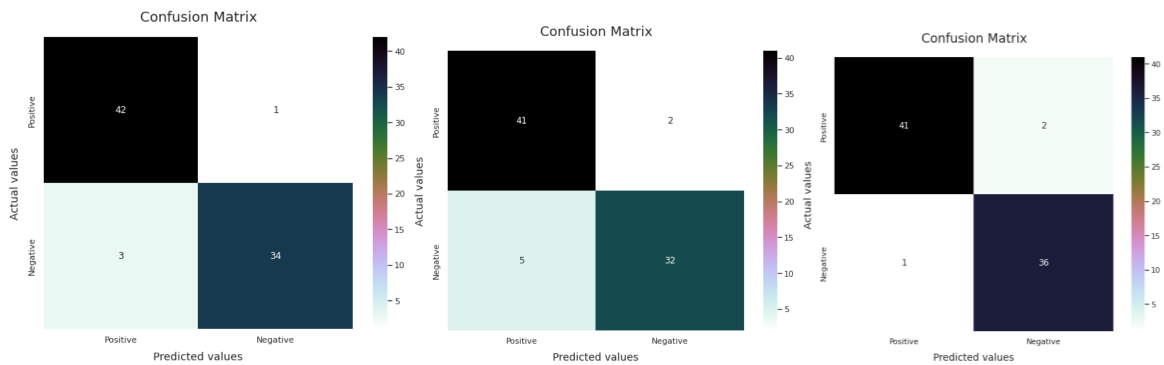
Figure 4.2.1.3.1: Bar chart of Vectorization feature for impact on Social Life (a) Bar chart of Unigram feature (b) Bar chart of Bigram feature (c) Bar chart of Trigram feature

Table of F1 scores, recall, accuracy, and precision for machine learning models for Social life in 4.2.1.3.D

Table 4.2.1.3.D: Table for accuracy, precision, recall, and F1 scores for machine learning models

Classifier Name	Accuracy	Precision	Recall	F1 Score
LR	92.50	87.80	97.30	92.31
DT	78.75	69.23	97.30	80.90
RF	90.0	85.37	94.59	89.74
Multi. NB	95.0	92.31	97.30	94.74
KNN	90.0	85.37	94.59	89.74
Linear SVM	81.25	72.00	97.30	82.76
RBF SVM	88.75	81.82	97.30	88.89
SGD	95.0	92.31	97.30	94.74

The confusion matrix of machine learning techniques is shown in a graph in Figure 4.2.1.



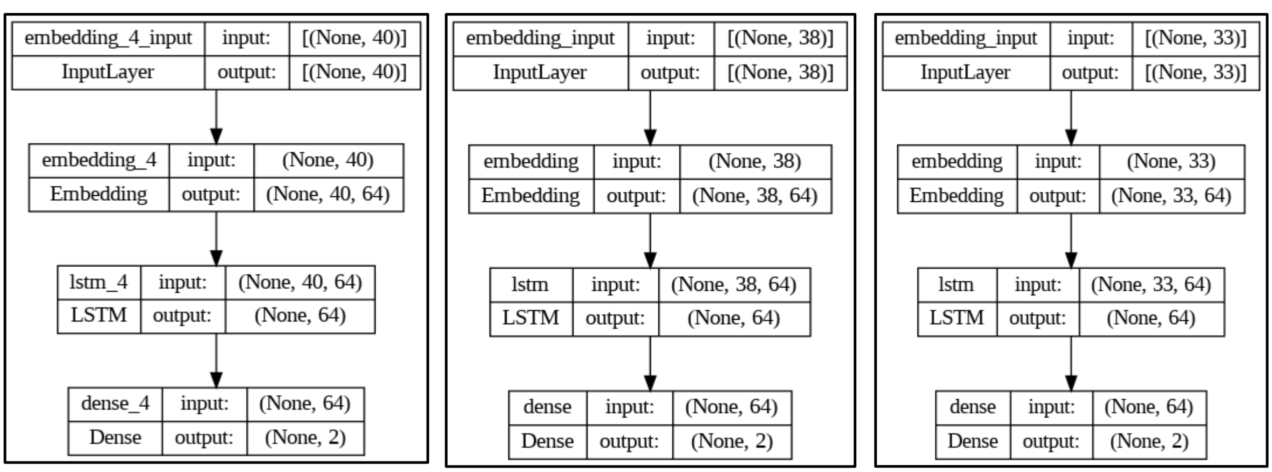
(a) (b) (c)

Figure 4.2.1: An illustration of the machine learning model's confusion matrix for (a) the academic column, (b) the mental column, and (c) the social column.

4.2.2 Analysis Experiment result of Deep Learning models

4.2.2.1 LSTM

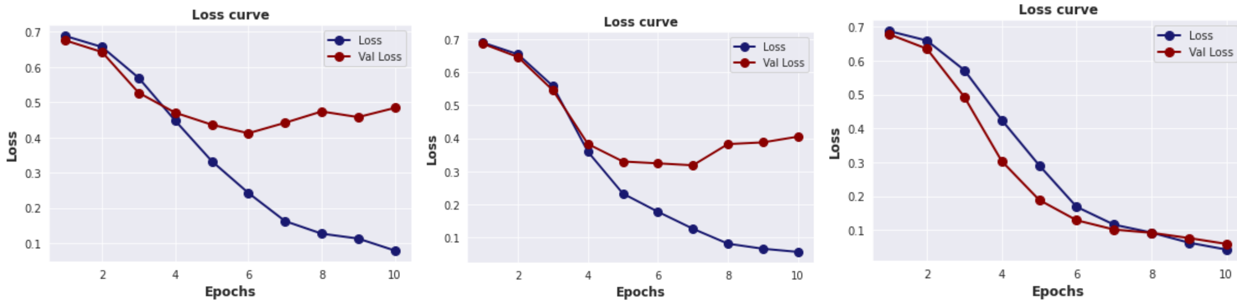
The sequential model chart for each of the three input columns is displayed in figure 4.2.2.1.1



(a) **(b)** **(c)**

Figure 4.2.2.1.1: LSTM sequential Model for (a) Academic column (b) Mental column (c) Social column

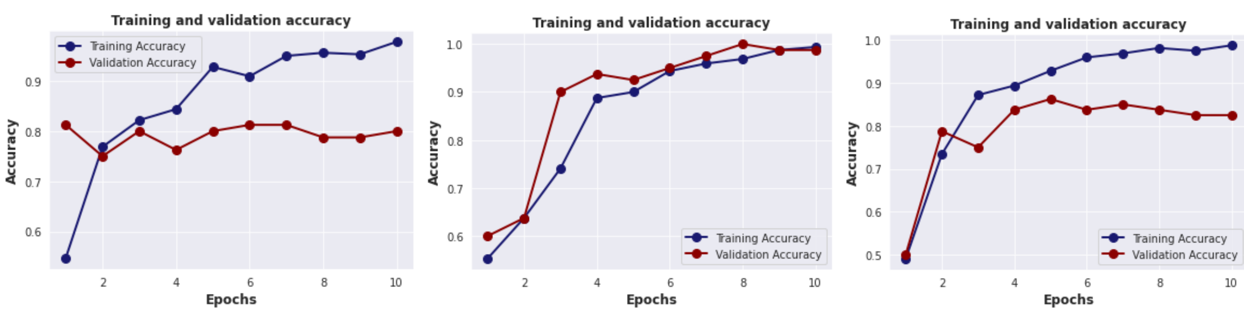
- Visualize the epochs and validation accuracy plot in figure 4.2.2.1.2



(a) **(b)** **(c)**

Figure 4.2.2.1.2: Illustration of LSTM Loss curve for (a) Academic (b) Mental (c)Social column

- Visualize the training and validation accuracy plot in figure 4.2.2.1.3



(a) (b) (c)

Figure 4.2.2.1.3: Visiualization of the LSTM Traning and validation curve for (a) Academic column (b) Mental column (c)Social column

- Precision, recall, f1 score and support value corresponding to label columns, macro and weighted avg are shown in table 4.2.2.1.A

Table 4.2.2.1.A : LSTM Classification report table for (a) Academic column (b) Mental column (c)Social column

Classification model	Precision	Recall	F1 score	Support
Positive	76.32	80.56	78.38	36.0
Negative	83.33	79.55	81.40	44.0
macro avg	79.82	80.05	79.89	80.0
weighted avg	80.18	80.0	80.04	80.0

(a)

Classification model	Precision	Recall	F1 score	Support
Positive	100.0	97.22	98.59	36.0
Negative	97.78	100.0	98.88	44.0
macro avg	98.89	98.61	98.73	80.0
weighted avg	98.78	98.75	98.75	80.0

(b)

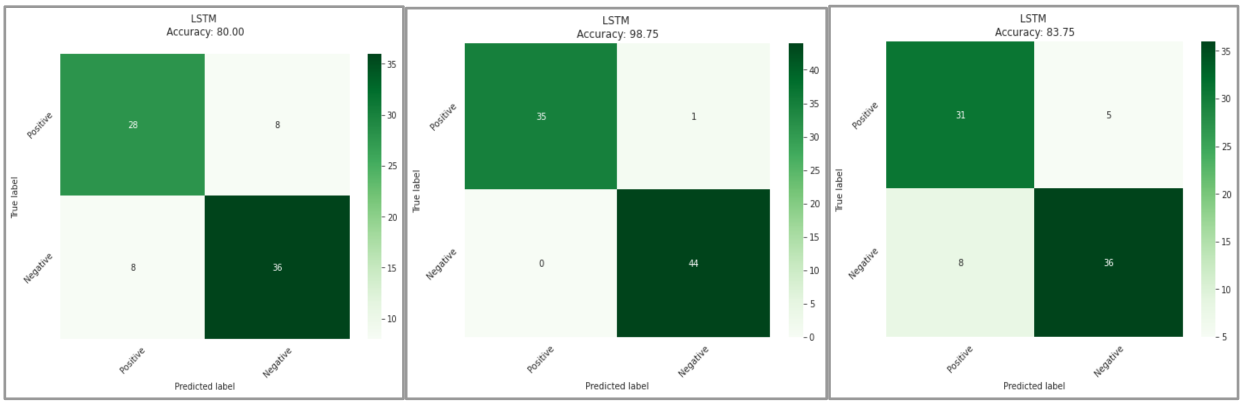
Classification model	Precision	Recall	F1 score	Support
Positive	79.49	86.11	82.67	36.0
Negative	87.80	81.82	84.71	44.0
macro avg	83.65	83.96	83.69	80.0
weighted avg	84.06	83.75	83.79	80.0

(c)

- The accuracy for three input columns for LSTM model:

Column Name	Accuracy
Academic	80
Mental	98.75
Social	83.75

- The confusion matrix for the academic, mental, and social input columns is shown in figure 4.2.2.1.4.



(a) (b) (c)

Figure 4.2.2.1.4: Illustration Confusion matrix of LSTM model for (a) Academic column (b) Mental column (c)Social column

Program Evaluation for LSTM model shows in tabel 4.2.2.1.B:

Table 4.2.2.1.B: LSTM Model evaluation

Column name	**Test Sentence**	**Sentiment**
Academic	ভালোভাবে পড়াশোনায় মনোযোগ দিতে পারেনি	Negative
Mental	মানসিকভাবে আমি সুস্থ ছিলাম সবসময় আমি বিভিন্ন কিছু নিয়ে ব্যস্ত থাকতাম	Positive
Social	আশেপাশের কোন মানুষ অসুস্থ হলে তাকে যথাসাধ্য সাহায্য করার চেষ্টা করেছে	Positive

4.2.2.2 Bi-LSTM

BiLSTM sequential model chart for the three input columns is displayed in figure 4.2.2.2.1.

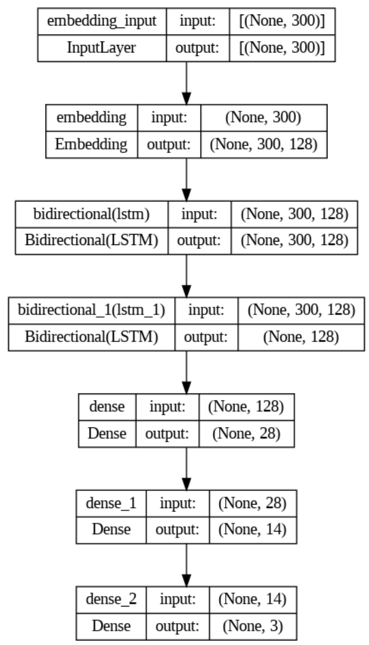
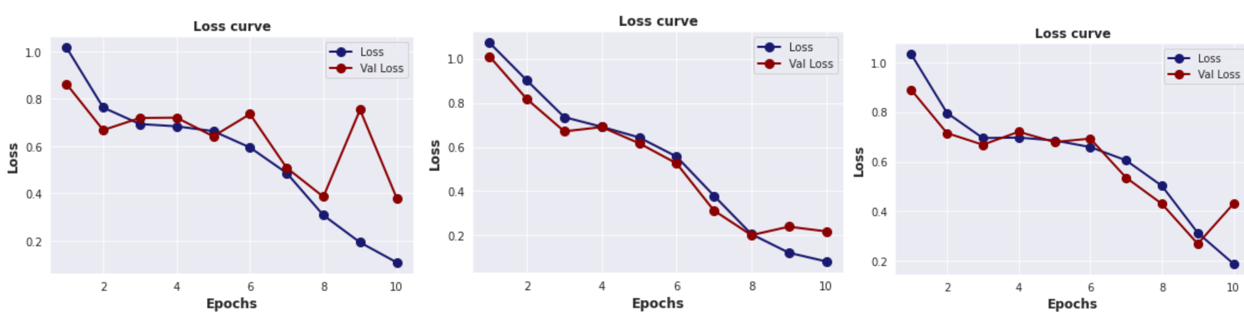


Figure 4.2.2.2.1: Bi-LSTM sequential Model for Academic, Mental and Social column

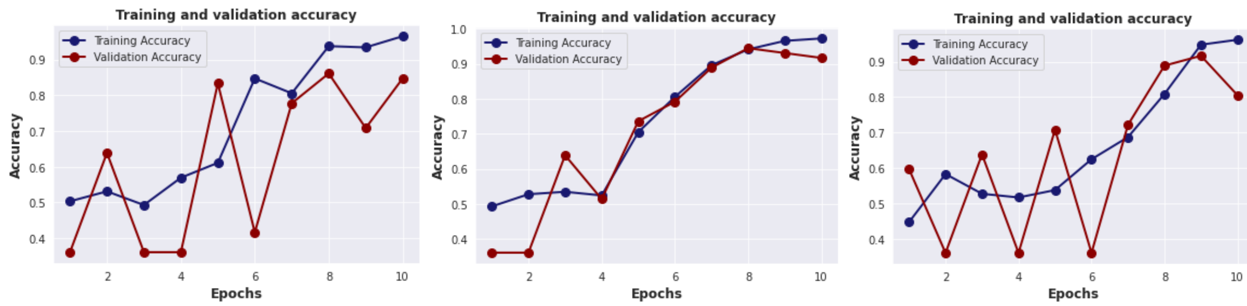
- The plot of epochs and validation accuracy is shown in Figure 4.2.2.2.2.



(a) **(b)** **(c)**

Figure 4.2.2.2.2: Bi-LSTM Loss curve for the following columns: (a) Academic (b) Mental (c) Social

- The accuracy plots for training and validation are displayed in Figure 4.2.2.2.3.



(a) (b) (c)

Figure 4.2.2.2.3: Bi-LSTM training and validation curve for the following columns: (a) Academic (b) Mental (c) Social

- Table 4.2.2.2.A displays precision, recall, f1 score, and support value for label columns, macro, and weighted avg.

Table 4.2.2.2.A: BiLSTM Classification report table for (a) Academic column (b) Mental column (c)Social column

Classification model	Precision	Recall	F1 score	Support
Positive	95.24	90.91	93.02	22.0
Negative	89.47	94.44	91.89	18.0
Macro avg	92.36	92.68	92.46	40.0
Weighted avg	92.64	92.50	92.51	40.0

(a)

Classification model	Precision	Recall	F1 score	Support
Positive	83.33	90.91	86.96	22.0
Negative	87.50	77.78	82.35	18.0
Macro avg	85.42	84.34	84.65	40.0
Weighted avg	85.21	85.0	84.88	40.0

(b)

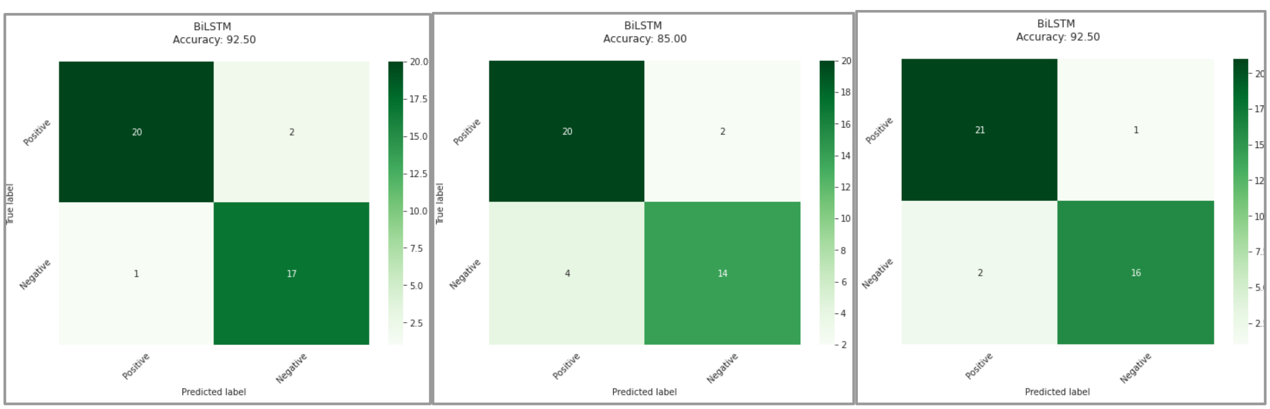
Classification model	Precision	Recall	F1 score	Support
Positive	91.30	95.45	93.33	22.0
Negative	94.12	88.89	91.43	18.0
Macro avg	92.71	92.17	92.38	40.0
Weighted avg	92.57	92.50	92.48	40.0

(c)

- For three input columns, the accuracy for BiLSTM model is as follows:

Column Name	**Accuracy**
Academic	92.50
Mental	85.00
Social	92.50

- The confusion matrix for the academic, mental, and social input columns is shown in figure 4.2.2.2.4.



(a) (b) (c)

Figure 4.2.2.2.4:Demonstration of the Bi-LSTM model's confusion matrix for (a) Academic column

For program evaluation of the Bi-LSTM model, see table 4.2.2.2.B.

Table 4.2.2.2.B: Model Performance for BiLSTM

Column name	Test Sentence	Sentimen
Academic	পড়াশোনায় মনোযোগ দিতে পারেনি	Negative
Mental	সবসময় আমি বিভিন্ন কিছু নিয়ে ব্যস্ত থাকতাম	Positive
Social	আশেপাশের কোন মানুষ যথাসাধ্য সাহায্য করার চেষ্টা করেছে	Positive

4.2.2.3 CNN:

The CNN sequential model chart for the three input columns is shown in Figure 4.2.2.3.1.

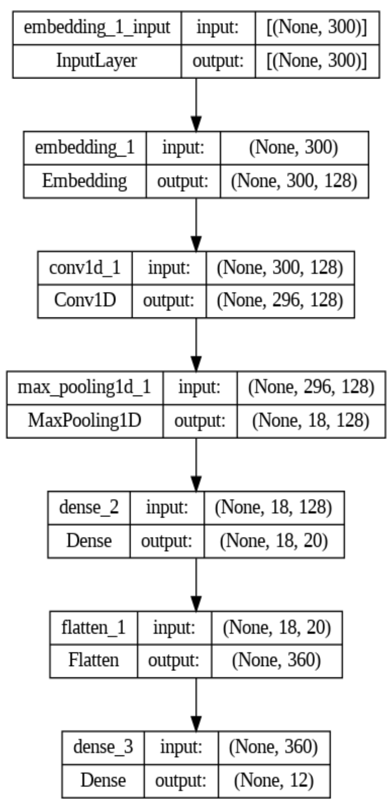
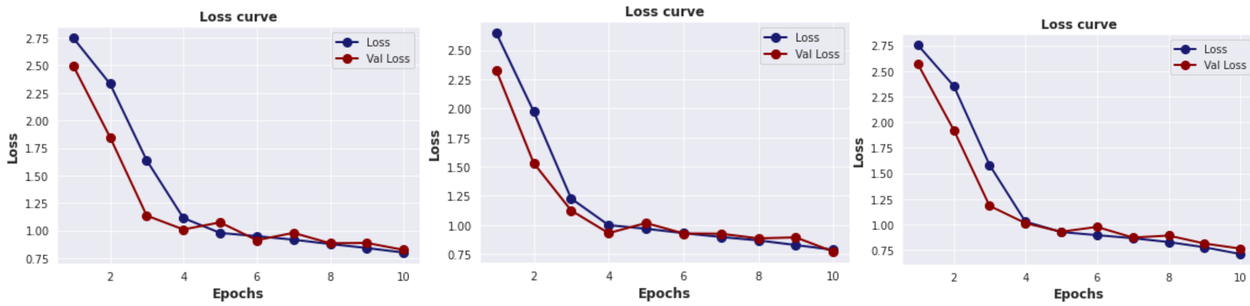


Figure 4.2.2.3.1**:** CNN sequential Model for Academic, Mental column and Social column

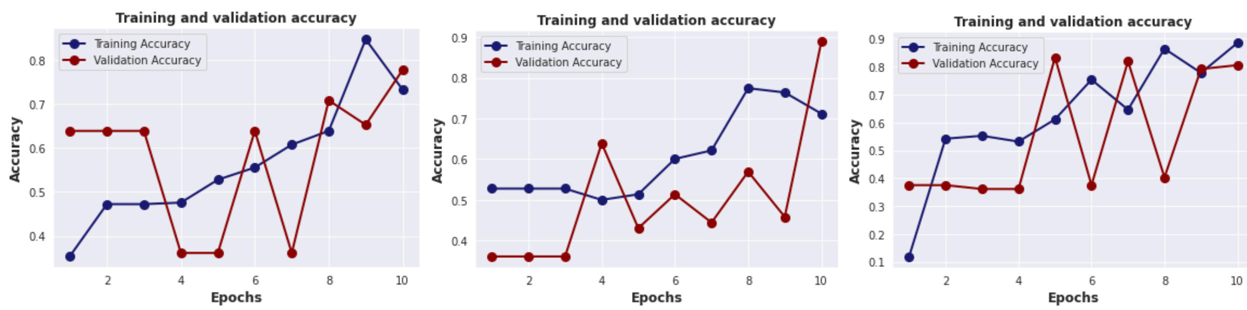
- Epochs and validation accuracy plots are depicted in figure 4.2.2.3.2 below



(a) **(b)** **(c)**

Figure 4.2.2.3.2: An overview of the CNN Loss Curve for the (a) Academic, (b) Mental, and (c) Social Columns

- The training and validation accuracy plots are shown in Figure 4.2.2.3.3.



(a) (b) (c)

Figure 4.2.2.3.3: CNN training and validation curves for (a) academic column (b) mental column (c) social column

- For the label columns, macro, and weighted avg, shows the precision, recall, f1 score, and support value table in 4.2.2.3.A

Table 4.2.2.3.A: CNN Classification report table for columns (a) Academic (b) Mental (c) Social

Classification model	Precision	Recall	F1 score	Support
Positive	80.0	90.91	85.11	22.0
Negative	86.67	72.22	78.79	18.0
Macro avg	83.33	81.57	81.95	40.0
Weighted avg	83.0	82.50	82.26	40.0

(a)

Classification model	Precision	Recall	F1 score	Support
Positive	100.0	77.27	87.18	22.0
Negative	78.26	100.0	87.80	18.0
Macro avg	89.13	88.64	87.49	40.0
Weighted avg	90.22	87.50	87.46	40.0

(b)

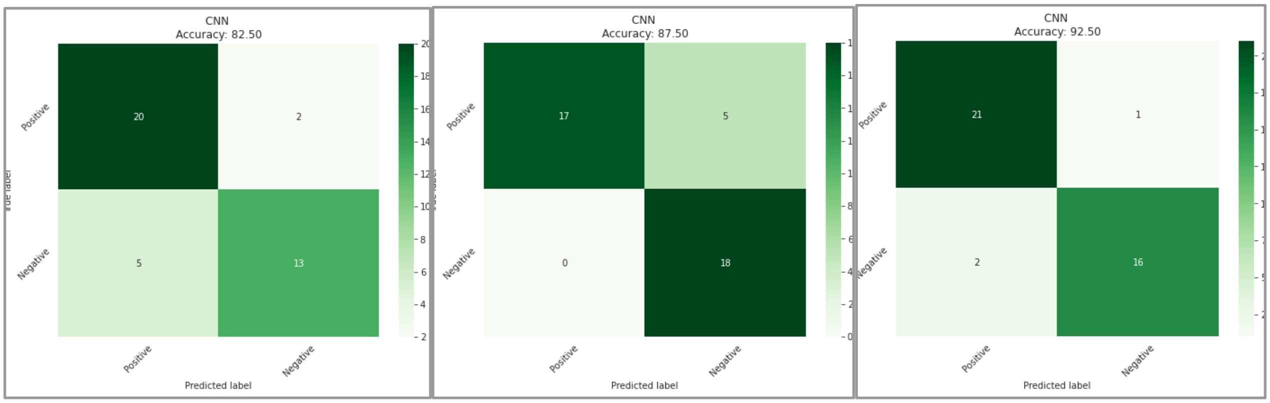
Classification model	Precision	Recall	F1 score	Support
Positive	91.30	95.45	93.33	22.0
Negative	94.12	88.89	91.43	18.0
Macro avg	92.71	92.17	92.38	40.0
Weighted avg	92.57	92.50	92.48	40.0

(c)

- The accuracy for three input columns for CNN model:

Column Name	Accuracy
Academic	82.50
Mental	87.5
Social	92.50

-The confusion matrix for the academic, mental, and social input columns is shown in figure 4.2.2.3.4.



(a) (b) (c)

Figure 4.2.2.3.4: Illustration Confusion matrix of CNN model for (a) Academic column (b) Mental column (c)Social column

CNN model program evaluation shows in table 4.2.2.3.B :

Table 4.2.2.3.B: Model Performance for CNN

Column name	**Test Sentence**	**Sentimen**
Academic	আমার কাজের দক্ষতা বৃদ্ধির জন্য আমি অনেক কিছু শিখেছি	Positive
Mental	দিন দিন ডিপ্রেশনের কারণ বেড়েই চলছিল	Negative
Social	সামাজিক দূরত্ব অনেক বেড়ে যায়	Negative

4.2.2.4 CNN LSTM:

The CNN LSTM sequential model chart for the three input columns is displayed in figure 4.2.2.4.1.

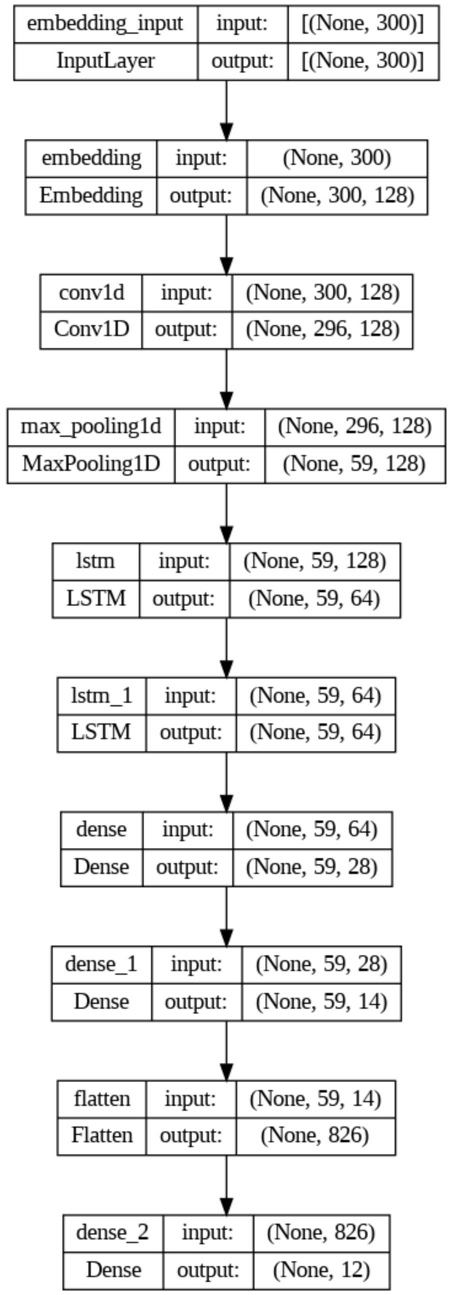
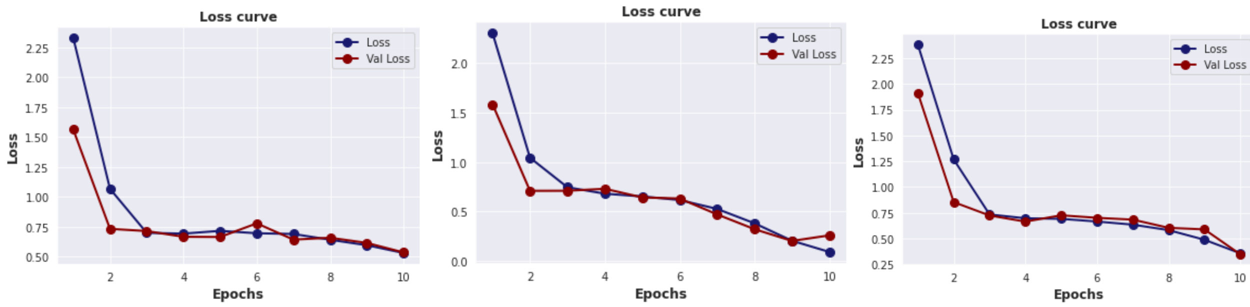


Figure 4.2.2.4.1: CNN LSTM sequential Model for Academic, Mental column and Social column

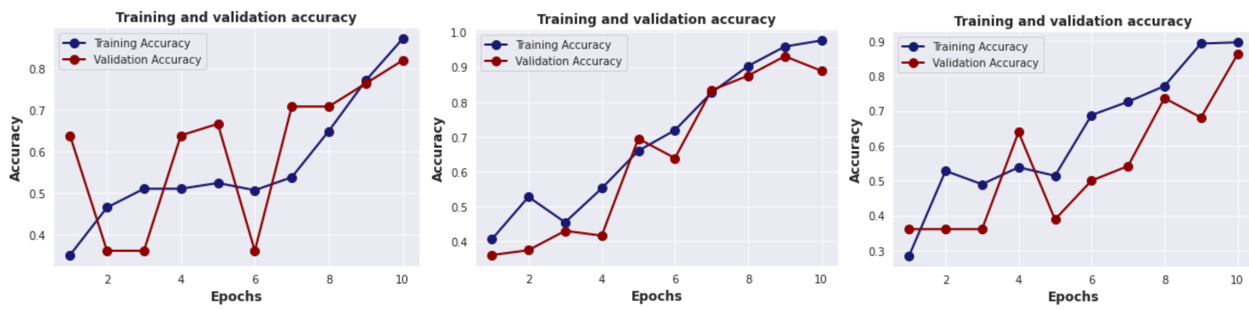
- The plot of epochs and validation accuracy is shown in Figure 4.2.2.4.2.



(a) **(b)** **(c)**

Figure 4.2.2.4.2: Illustration of CNN-LSTM Loss curve for following columns of (a) Academic (b) Mental (c)Social

The accuracy plots for training and validation are displayed in Figure 4.2.2.4.3.



(a) (b) (c)

Figure 4.2.2.4.3: CNN-LSTM Traning and validation curve for (a) Academic column (b) Mental column (c)Social column

- For the label columns, macro, and weighted avg, shows the precison, recall, f1 score, and support value table in 4.2.2.4.A

Table 4.2.2.4.A : CNN-LSTM Classification report table for (a) Academic column (b) Mental column (c)Social column

Classification model	Precision	Recall	F1 score	Support
Positive	90.0	81.82	85.71	22.0
Negative	80.0	88.89	84.21	18.0
Macro avg	85.0	85.35	84.96	40
Weighted avg	85.5	85.0	85.04	40.0

(a)

Classification model	Precision	Recall	F1 score	Support
Positive	87.50	95.45	91.30	22.0
Negative	93.75	83.33	88.24	18.0
Macro avg	90.62	89.39	89.77	40.0
Weighted avg	90.31	90.0	89.92	40.0

(b)

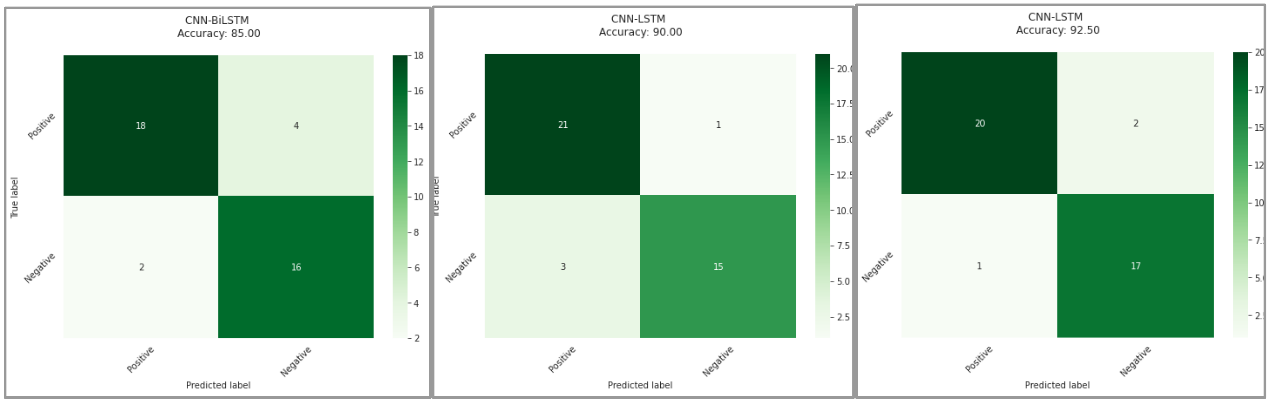
Classification model	Precision	Recall	F1 score	Support
Positive	95.24	90.91	93.02	22.0
Negative	89.47	94.44	91.89	18.0
Macro avg	92.36	92.68	92.46	40.0
Weighted avg	92.64	92.50	92.51	40.0

(c)

- The accuracy for three input columns for CNN LSTM:

Column Name	Accuracy
Academic	85.0
Mental	90.00
Social	92.50

- The confusion matrix for the academic, mental, and social input columns is shown in figure 4.2.2.4.4.



(a) (b) (c)

Figure 4.2.2.4.4: Illustration Confusion matrix of CNN-LSTM model for (a) Academic column (b) Mental column (c)Social column

The CNN-LSTM model is evaluated in the following table 4.2.2.4.B:

Table 4.2.2.4.B: Model Performance for CNN LSTM

Column name	Test Sentence	Sentimen
Academic	অনলাইন ক্লাসের পড়া বুঝতে অনেক অসুবিধা হতো	Negative
Mental	অনলাইন শিক্ষা ব্যবস্থার ক্ষেত্রে আমি খুবই মানসিক চাপ অনুভব করতাম	Negaitive
Social	বন্ধু-বান্ধবের সাথে বেশি যোগাযোগ হতো	Positive

4.2.2.5 CNN Bi-LSTM:

The sequential model chart for three input columns is displayed in figure 4.2.2.5.1

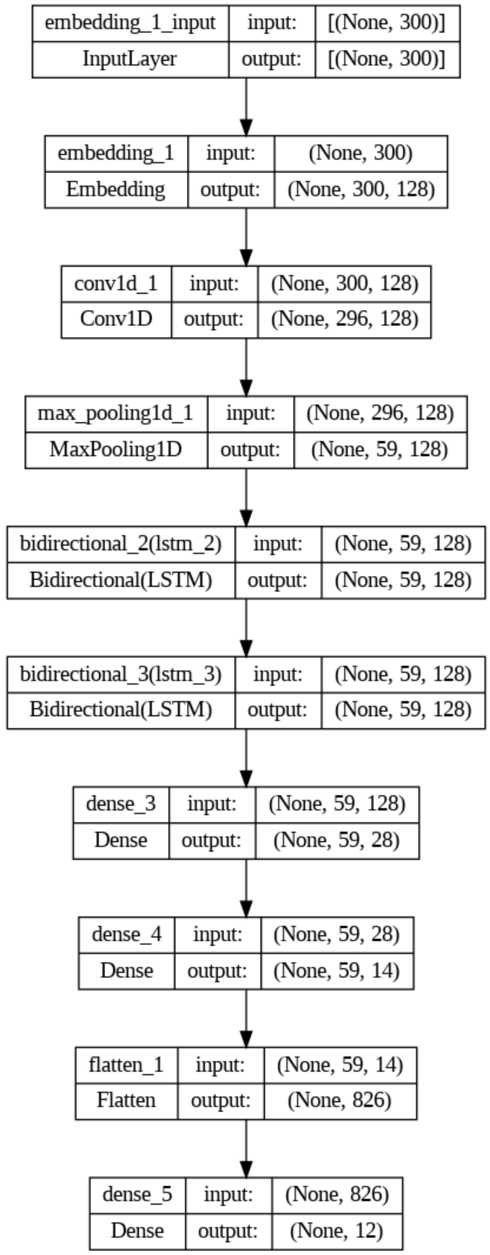
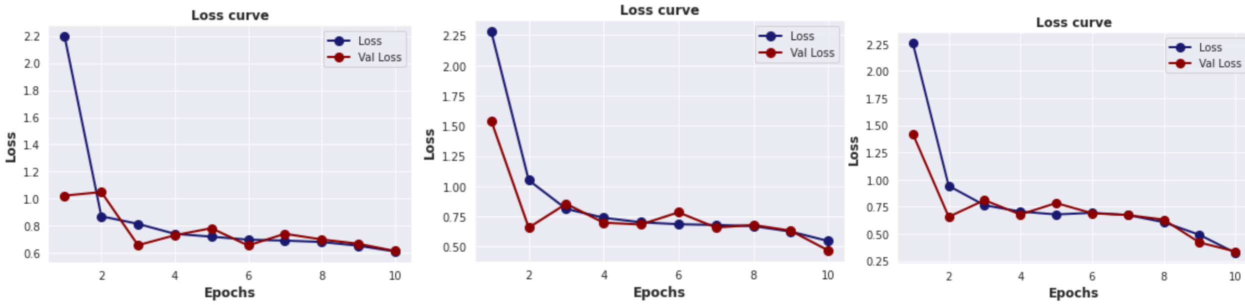


Figure 4.2.2.5.1: CNN BiLSTM sequential Model for Academic, Mental column and Social column

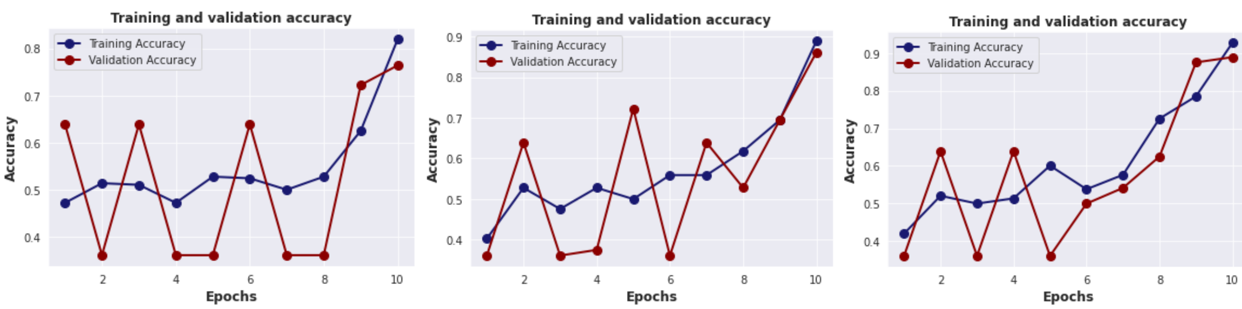
- Visualize the epochs and validation accuracy plot in figure 4.2.2.5.2



(a) **(b)** **(c)**

Figure 4.2.2.5.2: Illustration of CNN BiLSTM Loss curve for (a) Academic column (b) Mental column (c)Social column

- Visualize the training and validation accuracy plot in figure 4.2.2.5.3



(a) (b) (c)

Figure 4.2.2.5.3: Visiualization of the CNN BiLSTM Traning and validation curve for (a) Academic column (b) Mental column (c)Social column

- Precision, recall, f1 score and support value corresponding to label columns, macro and weighted avg are shown in table 4.2.2.5.A

Table 4.2.2.5.A: CNN BiLSTM Classification report table for (a) Academic column (b) Mental column (c)Social column

Classification model	Precision	Recall	F1 score	Support
Positive	83.33	90.91	86.96	22.0
Negative	87.50	77.78	82.35	18.0
Macro avg	85.42	84.34	84.65	40
Weighted avg	85.21	85.0	84.88	40.0

(a)

Classification model	Precision	Recall	F1 score	Support
Positive	86.36	86.36	86.36	22.0
Negative	83.33	83.33	83.33	18.0
Macro avg	84.85	84.85	84.85	40.0
Weighted avg	85.0	85.0	85.0	40.0

(b)

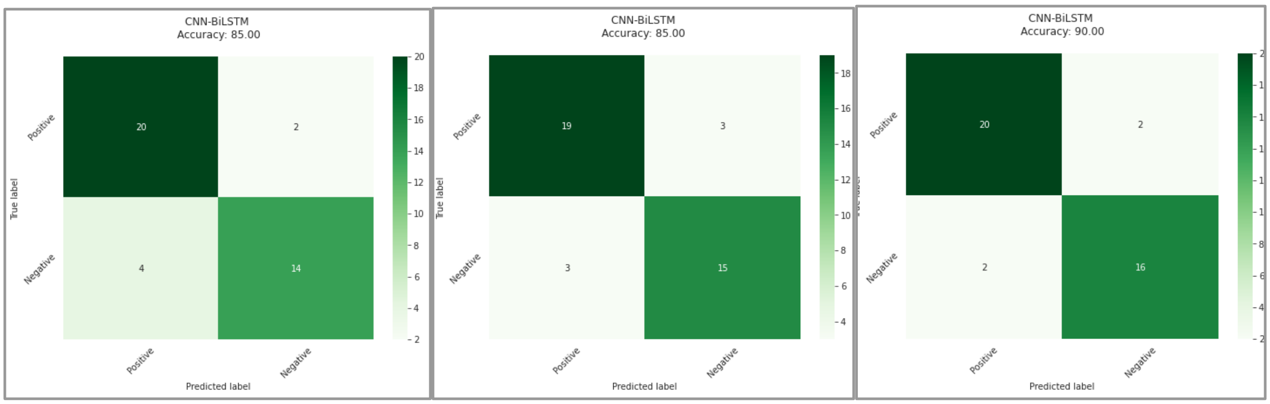
Classification model	Precision	Recall	F1 score	Support
Positive	90.91	90.91	90.91	22.0
Negative	88.89	88.89	88.89	18.0
Macro avg	89.90	89.90	89.90	40.0
Weighted avg	90.0	90.0	90.0	40.0

(c)

The accuracy for three input columns for CNN BiLSTM:

Column Name	Accuracy
Academic	85.0
Mental	85.0
Social	90.0

The confusion matrix for the academic, mental, and social input columns is shown in figure 4.2.2.5.4.



(a) (b) (c)

Figure 4.2.2.5.4: Illustration Confusion matrix of CNN BiLSTM model for (a) Academic column (b) Mental column (c)Social column

The CNN BiLSTM model program evaluation is shown below in table 4.2.2.5.B.

Table 4.2.2.5.B: Model Performance for CNN BiLSTM

Column name	**Test Sentence**	**Sentimen**
Academic	অনলাইন হওয়ার জন্য অনেক কিছু শিখতে পেরেছি	Positive
Mental	রাগ এবং জিদের মাত্রা অনেক বেড়ে গিয়েছিল	Negative
Social	পরিবারের সবার সঙ্গে সম্পর্ক অনেক ভালো হয়ে গিয়েছিল	Positive

4.3 Discussion

In this research work, we have created our own dataset for three input columns from student responses. For text classification, we used COVID related text data from student's academic, mental and social life. Labeling the dataset into 2 groups: positive and negative To add words one after the other, the Bangla sentence employs extensive punctuation, these marks should be eliminated to train the model. As there is no pre-built library for stop words in Bangla NLP, we created an excel file with Bangla words collected online. After the data cleaning process, we split the dataset into train and test sets. Nearly 80–90% of the data we use is for training, and 10–20% is for testing. We employed both supervised machine learning and deep learning models to analyze the data. Supervised machine learning was utilized for the categorical dataset. We used a diverse number of ML algorithms including Logistic regression, Decision tree, Random forest, KNN, Multinomial naive Bayes, Linear SVM, RBF SVM, and SGD into three input columns separately. In all of these algorithms, SGD, RBF SVM, and multinomial naive Bayes performed admirably for all column data. To predict the actual value, we used SVM classification analysis and a confusion matrix to outline the execution. Then, a range of deep learning models was utilized, including CNN, LSTM, CNN LSTM, BiLSTM, and CNN BiLSTM. Deep-learning sequential models are used to build these models and train the data. After training the dataset we employed it to test the value. In comparison to other models, CNN and CNN LSTM fared well.

Then the algorithms and deep learning model classify any new data that we input before providing us with a prediction.

Chapter 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society

The detection or prevention of damage, such as by classifying hate speech or recognizing depressed symptoms, is a major focus of current research on NLP that promotes social good. NLP analysis, however, also has the ability to be used in positive purposing, such as to improve user wellbeing or promote constructive communication. The social effect aspects of NLP are likewise derived from the interrelationships among linguistics, culture, and the individual. We look at the autonomous inference of user traits, a well-known and intriguing NLP problem with a promising solution. Thus, as part of our investigation, we gathered information concerning the impact of COVID on youngsters under the age of 10. We gathered information on how social, mental, and academic life affected the style of Bengali text. With various ML techniques, DL models, and mixed DL models, we trained a dataset. It will assist us in determining which model is more reliable for various text types and can also be used to detect both negative and positive feelings.

5.2 Impact on Environment

NLP has been employed recently for various beneficial applications, including recommenders and scam or fraud identification. In general, automatic categorization works by applying ML techniques to categorize documents into certain groups. In general, organizing and using the massive amounts of information that are accessible in disorganized text format is one of the most important data mining methodologies. Sentiment analysis is a crucial text categorization application. Text mining includes the discipline of sentiment analysis. Algorithms are used in text mining to separate essential information from unorganized materials. In a perfect world, this kind of algorithm would mimic how a person thinks when reading. The method of reading and comprehending human-written information using computers for actionable information is known as text analysis. Text-mining systems can categorize, organize, and retrieve data on their own from texts to find attitudes, correlations, trends, and other necessary details. As a result, because the proposed study focuses on text classifiers, it may be incorporated as a feature in any practical application and applied to AI systems to make them more valuable for the surroundings. Machines will be capable to determine what feeling a text is displaying.

5.3 Ethical Aspects

A person's style of writing can be a distinguishing characteristic of their personality. Online communication has evolved into a new medium for writing, expressing oneself, or recording events ever since social networking first appeared. Due to this, there is also a quantity of unstructured information that must be understood. This is a situation where text categorization may work to your benefit. Unorganized data is categorized into several groups through the method of text classification. Therefore, to train the model for our task, we employed some unorganized sentences. We think that no one's privacy has been violated and that we have not used anyone's sensitive information because we obtained data from students with their consent. For our code implementation, we used Google Colab, which is available to everyone. Thus, we haven't harmed society in this situation.

5.4 Sustainability Plan

The economy, social conscience, and the ecosystem are the three primary pillars of durability. Accordingly, our multi-head text categorization method that combines ML, DL, and combination DL is extremely sustainable. Because it may be included in social networking sites, online apps, language apps, etc. as Bengali NLP needs to be more robust, our effort is beneficial for the environment. Bangla NLP isn't as advanced as other languages like French and English in terms of development. Economically speaking, our work is extremely cost-effective because all crucial library functions may be imported for free and will be useful when used. Regression and the addition of new classes are two potential developments we have planned for this project.

CHAPTER 6

Summary, Conclusion, Recommendation and Implication for Future Research

6.1 Synopsis of the Study

Bengali NLP is used throughout our research study. For the Bengali Text Sentiment Analysis, we applied ML and DL models in this work.

A detailed explanation of the task is provided below, in stages.

- Collecting data from students through questionnaire
- Labeling the data into 2 groups (positive and negative)
- preparation of data
- StopWord Removal
- Stemmer
- TF-IDF for Data Vectorization
- Tokenization
- Label Encoding
- Dataset Splitting
- Apply ML classifier models to find the best accuracy, precision, recall
    - Logistic Regression
    - Decision tree
    - KNN
    - Multinomial Naive Bayes
    - Linear SVM
    - RBF SVM
    - Random Forest
    - SGD
- Confusion matrix
- Pad Sequence
- Build Sequenquial model
- Use Deep Learning models
    - BiLSTM

- CNN
- LSTM
- CNN BiLSTM
- CNN LSTM

- Model Training
- Illustration Loss and Accuracy curve
- Confusion matrix
- Evaluate the outcome and the machine's response.

6.2 Conclusion

We reported our key conclusions from the text data in this study. In three text fields, we have used machine learning and deep learning approaches. The data was cleaned and separated into two sets: training and testing. We first trained the model, then tested it to see how accurate it was. We applied machine learning techniques and got the highest output for the SGD algorithm at 95% on academic impact texts. Moreover, the greatest accuracy on mental impact was achieved with KNN, which is 93.75%. Multi. NB and SGD both provide the greatest accuracy on social life effect texts, which is 95%. In deep learning approaches, BiLSTM got the highest accuracy for the academic impact column which is 92.5%. The LSTM model achieved the highest accuracy of 98.75% in the mental impact column. BiLSTM, CNN, and CNN LSTM provide the highest accuracy of 92.5% in the social impact column.

6.3 Future Work

The field of natural language processing is vast, yet resources for Bangla NLP are poor. We are looking for various areas of research to expand NLP research in future. To improve the efficiency of emotion categorization, we will try to enhance our dataset and use more deep learning methods and machine learning algorithms. We particularly paid attention to the behavioral analysis of the pupils and how the COVID epidemic affected them. From the perspectives of teachers and parents, we shall attempt to assess sentiment. Additionally, we'll attempt to employ various regression techniques on our dataset and compare all of these models and methods

Reference:

[1] Salehin, I., Dip, S.T., Talha, I.M., Rayhan, I. and Nammi, K.F., 2021. Impact on human mental behavior after pass through a long time home quarantine using machine learning. *International Journal of Education and Management Engineering (IJEME)*, *11*(1), pp.41-50.

[2] Akour, I., Alshurideh, M., Al Kurdi, B., Al Ali, A. and Salloum, S., 2021. Using Machine Learning Algorithms to Predict People's Intention to Use Mobile Learning Platforms During the COVID-19 Pandemic: Machine Learning Approach. *JMIR Medical Education*, *7*(1), pp.e24032-e24032.

[3] Chaudhary, A.P., Sonar, N.S., Jamuna, T.R., Banerjee, M. and Yadav, S., 2021. Impact of the COVID-19 pandemic on the mental health of college students in India: cross-sectional web-based study. *JMIRx med*, *2*(3), p.e28158.

[4] Herbert, C., El Bolock, A. and Abdennadher, S., 2021. How do you feel during the COVID-19 pandemic? A survey using psychological and linguistic self-report measures, and machine learning to investigate mental health, subjective experience, personality, and behaviour during the COVID-19 pandemic among university students. *BMC psychology*, *9*(1), pp.1-23.

[5] Naughton, F., Ward, E., Khondoker, M., Belderson, P., Marie Minihane, A., Dainty, J., Hanson, S., Holland, R., Brown, T. and Notley, C., 2021. Health behaviour change during the UK COVID-19 lockdown: Findings from the first wave of the C-19 health behaviour and well-being daily tracker study. *British journal of health psychology*, *26*(2), pp.624-643.

[6] Wang, C., Zhao, H. and Zhang, H., 2020. Chinese college students have higher anxiety in new semester of online learning during COVID-19: a machine learning approach. *Frontiers in psychology*, *11*, p.587413.

[7] Sharma, M., Batra, K., Davis, R.E. and Wilkerson, A.H., 2021, January. Explaining handwashing behavior in a sample of college students during COVID-19 pandemic using the multi-theory model (MTM) of health behavior change: a single institutional cross-sectional survey. In *Healthcare* (Vol. 9, No. 1, p. 55). MDPI.

[8] Boca, G.D., 2021. Factors influencing students' behavior and attitude towards online education during COVID-19. *Sustainability*, *13*(13), p.7469.

[9] Onyema, E.M., Eucheria, N.C., Obafemi, F.A., Sen, S., Atonye, F.G., Sharma, A. and Alsayed, A.O., 2020. Impact of Coronavirus pandemic on education. *Journal of education and practice*, *11*(13), pp.108-121.

[10] Gupta, P., Kumar, S., Suman, R.R. and Kumar, V., 2020. Sentiment analysis of lockdown in india during covid-19: A case study on twitter. *IEEE Transactions on Computational Social Systems*, *8*(4), pp.992-1002.

[11] Browning, M.H., Larson, L.R., Sharaievska, I., Rigolon, A., McAnirlin, O., Mullenbach, L., Cloutier, S., Vu, T.M., Thomsen, J., Reigner, N. and Metcalf, E.C., 2021. Psychological impacts from COVID-19 among university students: Risk factors across seven states in the United States. *PloS one*, *16*(1), p.e0245327.

[12] Shafi, J. and Rahman, M., 2021. Application of machine learning models for detecting mental stress during COVID-19 pandemic. *Application of Machine Learning Models for Detecting Mental Stress during COVID-19 Pandemic (February 3, 2021)*.

[13] Kuo, C.P. and Fu, J.S., 2021. Evaluating the impact of mobility on COVID-19 pandemic with machine learning hybrid predictions. *Science of The Total Environment*, *758*, p.144151.

[14] Mohan, S., Abugabah, A., Kumar Singh, S., Kashif Bashir, A. and Sanzogni, L., 2022. An approach to forecast impact of Covid-19 using supervised machine learning model. *Software: Practice and Experience*, *52*(4), pp.824-840.

[15] Faisal, F., Nishat, M.M., Mahbub, M.A., Shawon, M.M.I. and Alvi, M.M.U.H., 2021, August. Covid-19 and its impact on school closures: a predictive analysis using machine learning algorithms. In *2021 International Conference on Science & Contemporary Technologies (ICSCT)* (pp. 1-6). IEEE.

[16] Atlam, E.S., Ewis, A., Abd El-Raouf, M.M., Ghoneim, O. and Gad, I., 2022. A new approach in identifying the psychological impact of COVID-19 on university student's academic performance. *Alexandria Engineering Journal*, *61*(7), pp.5223-5233.

[17] Ho, I.M.K., Cheong, K.Y. and Weldon, A., 2021. Predicting student satisfaction of emergency remote learning in higher education during COVID-19 using machine learning techniques. *Plos one*, *16*(4), p.e0249423.

[18] Shim, J.G., Ryu, K.H., Lee, S.H., Cho, E.A., Lee, Y.J. and Ahn, J.H., 2021. Text mining approaches to analyze public sentiment changes regarding COVID-19 vaccines on social media in Korea. *International Journal of Environmental Research and Public Health*, *18*(12), p.6549.

[19] Mansoor, M., Gurumurthy, K. and Prasad, V.R., 2020. Global sentiment analysis of COVID-19 tweets over time. *arXiv preprint arXiv:2010.14234*.

[20] Machuca, C.R., Gallardo, C. and Toasa, R.M., 2021, February. Twitter sentiment analysis on coronavirus: Machine learning approach. In *Journal of Physics: Conference Series* (Vol. 1828, No. 1, p. 012104). IOP Publishing.

[21] Rustam, F., Khalid, M., Aslam, W., Rupapara, V., Mehmood, A. and Choi, G.S., 2021. A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. *Plos one*, *16*(2), p.e0245909.

[22] Tan, H., Peng, S.L., Zhu, C.P., You, Z., Miao, M.C. and Kuai, S.G., 2021. Long-term effects of the COVID-19 pandemic on public sentiments in mainland China: sentiment analysis of social media posts. *Journal of Medical Internet Research*, *23*(8), p.e29150.

[23] Singh, M., Jakhar, A.K. and Pandey, S., 2021. Sentiment analysis on the impact of coronavirus in social life using the BERT model. *Social Network Analysis and Mining*, *11*(1), pp.1-11.

[24] Chaturvedi, K., Vishwakarma, D.K. and Singh, N., 2021. COVID-19 and its impact on education, social life and mental health of students: A survey. *Children and youth services review*, *121*, p.105866.

[25] Bhat, B.A., Khan, S., Manzoor, S., Niyaz, A., Tak, H.J., Anees, S.U.M., Gull, S. and Ahmad, I., 2020. A study on impact of COVID-19 lockdown on psychological health, economy and social life of people in Kashmir. *International Journal of Science and Healthcare Research*, *5*(2), pp.36-46.

[26] Shatla, M.M., Alotaibi, N.T., Munshi, S.A., Yaseen, E.E., Qanadeely, E.M., Abdulaziz, W., Alshaddadi, M.A.G. and Ali, I., 2021. Covid-19 and its impact on education, social life, and mental health among medical students in Saudi Arabia. *Medical Science*, *25*(116), pp.2708-2717.

[27]https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020

[28] Nair, A.J., Veena, G. and Vinayak, A., 2021, April. Comparative study of twitter sentiment on covid-19 tweets. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 1773-1778). IEEE.

[29] Rahib, M.R.H.K., Tamim, A.H., Tahmeed, M.Z. and Hossain, M.J., 2022. Emotion Detection Based on Bangladeshi People’s Social Media Response on COVID-19. *SN Computer Science*, *3*(2), p.180.

[30] Li, D. and Qian, J., 2016, October. Text sentiment analysis based on long short-term memory. In *2016 First IEEE International Conference on Computer Communication and the Internet (ICCCI)* (pp. 471-475). IEEE.

[31] Uddin, A.H., Bapery, D. and Arif, A.S.M., 2019, July. Depression analysis from social media data in Bangla language using long short term memory (LSTM) recurrent neural network technique. In *2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2)* (pp. 1-4). IEEE.

[32]https://learn.theprogrammingfoundation.org/getting_started/intro_data_science/module4/?gclid=CjwKCAiA2rOeBhAsEiwA2Pl7Q8rQ3i_7E4kC8rn-WU5jJHBHrXkOB6nR4D-cpE1SGgecIaUigXf9cBoC-x4QAvD_BwE

[33]https://www.pycodemates.com/2022/10/the-rbf-kernel-in-svm-complete-guide.html

Report

ORIGINALITY REPORT

18%	14%	11%	12%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.fedoa.unina.it Internet Source	1%
2	Submitted to University of Hertfordshire Student Paper	1%
3	Submitted to University of Wales Institute, Cardiff Student Paper	1%
4	Submitted to Cardiff University Student Paper	1%
5	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
6	link.springer.com Internet Source	1%
7	Submitted to University of Bristol Student Paper	<1%
8	"Sentimental Analysis and Deep Learning", Springer Science and Business Media LLC, 2022 Publication	<1%