IDENTIFICATION OF STUDENT PRODUCTIVITY CHALLENGES IN BANGLADESH: A CLASSIFICATION APPROACH WITH MACHINE LEARNING AND DEEP LEARNING

BY

MD. TOUHIDUL HASAN
ID: 203-15-14552

AND

MD. TANVEN AREFIN SAZID
ID: 203-15-14526

FINAL YEAR DESIGN PROJECT REPORT

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

Supervised By

TANIA KHATUN
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

MS. SHARUN AKTER KHUSHBU
Lecturer (Sr. Scale)
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

July 2024

APPROVAL

This Project titled "Identification of Student Productivity Challenges in Bangladesh: A Classification Approach with Machine Learning and Deep Learning" submitted by **Md. Touhidul Hasan** and **Md. Tanven Arefin Sazid** 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 13-07-2024.

BOARD OF EXAMINERS

Dr. Sheak Rashed Haider Noori — **Board Chairman**
Professor & Head
Department of Computer Science & Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Zahid Hasan — **Internal Examiner 1**
Associate Professor
Department of Computer Science & Engineering
Faculty of Science & Information Technology
Daffodil International University

Md. Mohammad Masum Bakaul — **Internal Examiner 2**
Senior Lecturer
Department of Computer Science & Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Zulfiker Mahmud — **External Examiner**
Professor
Department of Computer Science & Engineering
Jagannath University

©Daffodil International University, Bangladesh ii

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Tania Khatun, Assistant Professor Department of Computer Science and Engineering, Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

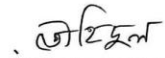
Supervised by:

Tania Khatun
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised by:

Ms. Sharun Akter Khushbu
Lecturer (Sr. Scale)
Department of CSE
Daffodil International University

Submitted by:



Md. Touhidul Hasan
ID: -203-15-14552
Department of CSE
Daffodil International University

Md. Tanven Arefin Sazid
ID: -203-15-14526
Department of CSE
Daffodil International University

©Daffodil International University, Bangladesh iii

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty for His divine blessing making it possible for us to complete the final year project successfully.

We are grateful and wish our profound indebtedness **Tania Khatun**, **Assistant Professor,** Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Machine learning*" to carry out this project. Her endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

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

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

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

ABSTRACT

Natural Language Processing (NLP) approach is adopted in this research to identify the causes of students' productivity problem in Bangladesh. To the best of the authors' knowledge, the current study adopted a cross-sectional design where 3, 079 entries and two-columns questionnaires incorporating the five target attributes; 'Drifting', 'Technoference', 'Laziness', 'Rushed', and 'Procrastinating' were employed to reveal the underlying drivers of productivity interruption. Particularly, the study is based on surveys, interviews of learners, and the data gathered from social networks to consider the main difficulties that high school and university students meet at their studying process. In the variety of Deep Learning and Machine Learning techniques, namely Decision Trees (DT), Support Vector Machines (SVM), Logistic Regression (LR), Random Forest and Convolutional Neural Networks (CNN) algorithms are used for the dataset. The efficiency of these models is measured based on accuracy and here the accuracy of the CNN model is 99.58%. They perform well in predicting and classifying the productivity interruptions, hence depicting a use of the algorithms to address productivity issues among students. Therefore, these findings of the study are pertinent to educational policymaking, administration and practice in Bangladesh. Thus, achieving higher learning results, the primary causes negatively influencing student productivity are revealed to create specific measures and actions. Additionally, the study adds to the existing body of knowledge on the problems of education in developing nations and shows how NLP and machine learning can be applied for solving another tough issue concerning education.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

Acronym	Full Form
AI	Artificial Intelligence
BDT	Bangladeshi Taka
ML	Machine Learning
SVM	Support Vector Machine
CNN	Convolutional Neural Network
FYDP	Final Year Design Project

CHAPTER 1
Introduction

1.1 Overview

The levels and effectiveness of student productivity is an essential aspect determining educational outcomes as well as future achievements. Students are the future of any nation, especially the developing one like Bangladesh where youth are in a massive number, so the concept of student productivity plays a significant role. However, the large funds spent in education many students encounter obstacles that hinder them from studying and even performing well. Knowing what actually constitutes these forms of productivity challenges will enable educators, policymakers, and stakeholders to address them effectively if they implement specific measures that would enhance the learning conditions and results of students R. Patel and S. Patel et al. (2021). The branch of artificial intelligence dealing with the interaction between humans and computers, namely Natural Language Processing (NLP), provides the necessary tools for working with textual data to obtain such insights. Using NLP strategies, we can formally study the students' own accounts of their difficulties: this would allow for understanding the existence of some problems that could hardly be identified in their quantitative analysis. Y. Chandra and A. Jana, (2020). This research is to find out why students in Bangladesh do not work effectively by analyzing a corpus of text data from high school and university students. In case of direct surveys and interviews, structured questions were used mostly supported by information from the respondents' social media accounts, friends, and relatives. Despite facing challenges such as lack of cooperation from students and incomplete responses, we compiled a dataset with 3,079 entries and two primary attributes: Thus, I have identified two significant concepts that are 'Description' and 'Label'. The labels categorize the identified productivity interruptions into five target attributes: They are; 'Drifting', 'Technoference', 'Laziness', 'Rushed', 'Procrastinating'. Y. Li et al. (5). Our methodology encompasses several key steps: converting text data collection and pre-processing, labeling text data, data cleaning in NLP, choosing the class of model to be used, training the model, evaluating the model, and testing the model on new data. This way, we minimize the chances of missing some aspects of the student productivity as we have a broad view and a well laid down structure

to follow. To conclude, by means of the machine learning models like Decision Trees (DT), Support Vector Machines (SVM), Logistic Regression (LR), Random Forests, and Convolutional Neural Networks (CNN), our research will cater actionable insights to optimize the student productivity in Bangladesh. Md. R. Hasan et al. (2019). In conclusion, this research enhances the overall literature on education issues in developing countries and provides insight for the application of NLP in solving realistic and complex issues in education.

1.2 Background and Present State

This research aims to discover the factors that hamper the student productivity of Bangladesh through analyzing the writing samples of the students by employing Natural Language Processing (NLP). Therefore, the study seeks to establish and dissect the primary interruptions that students experience including wavering attention, use of gadgets, and last but not the least, procrastination. The questionnaires used to gather the data included regular questionnaires, online questionnaires, face-to-face interviews, and online interviews, and the participant generated a total of 3,079 entries of data using inputs from their friends and relatives. To process and perform analysis on the data text cleaning, tokenization and sentiment analysis were used among the other NLP processes. Decision Trees, Support Vector Machines, Logistic Regression, Random Forests, and Convolutional Neural Networks were used in the processes of anticipation and analysis of productivity problems. Evaluative conclusions will be made with aim to design and implement specific interventions and policies changing students' performance and productivity levels in Bangladeshi education system and thus contributing to the country's socio-economic progress.

1.3 Problem Statement

Though education has improved remarkably in Bangladesh, students are facing a lot of productivity issues which have a negative impact on the current performance as well as on the upcoming future of a nation. It is rather difficult to pinpoint what exact aspects cause this problem, given the fact that there are a number of potential triggers ranging from socio-economic environment to pure distractions in the form of technologies and individual

practices. In most cases, conventional approaches of assessing these factors are incapable of providing sufficient and sufficiently detailed information on the productivity of students. To this end, this research seeks to fill this gap by employing NLP and machine learning paradigms to study and forecast these interruption sources and their impacts on students' productivity. In this regard, the study aims to develop accurate models as CNN, DT, and RF to minimize the possibility of false prediction, and offer a guideline that will help the government to design suitable interventions or policies to improve students' education results in Bangladesh.

1.4 Objective

The primary purpose of this study will also involve an overall description of the major causes of low student productivity in Bangladesh including; distraction, shift of focus and use of technology among others, procrastination. The specifics of how this or that issue may differ on the basis of educational level will be revealed. This will also show how students manage to overcome some of these challenges as part of the research. The ability of the developed machine learning models in accurately estimating productivity problems from standalone textual descriptions shall be quantitively assessed. Furthermore, specific suggestions aimed at modifying the conditions the affect learning environment and productivity of students and that can be helpful for the policymakers, educators, and other stakeholders will be made. This study will advance the general knowledge on the educational difficulties in the developing countries and showcase the application of NLP in solving various educational problems.

1.5 Scope and Limitations

The concern of this study is therefore to carry out a quantitative analysis of the factors that may impact on students' productivity in Bangladesh through the application of NLP and machine learning algorithms. The range also covers data gathering from high school and university students, and text data processing, as well as using such models as CNN, DT and RF to foresee productivity disruptions. It is research that has a goal of presenting findings that would help in designing specific important intervention and educational policies. But these are some of the things that should be noted down before proceeding to

the next step. The type of data is accurate and undoubtedly exhaustive, but it can be seen that it is far from studying all the factors that affect productivity, which is a drawback. Also, using survey questions to collect information from the students may include bias in the results gotten. The models adopted themselves may be quite robust; however, they may need slight fine-tuning or even empirical validation for the cross-contextualization of findings concerning the education domain. Finally, there are usually issues of socio-economic changes and technological change that occur over time that may influence the relevance of the findings.

1.6 Report Organization

Chapter 1: Introduction, the chapter presents the background of the study, the reason for the study, questions to be answered and objectives meant to be achieved. Chapter 2: Literature, the chapter presents literature findings on productivity levels of students, education issues in developing nations and the use of NLP in education literature. Chapter 3: Methodology, the present chapter overviews the general approach to the research, methods used to gather and pre-process the data as well as to label it and analyze. Chapter 4: Implementations, the implementation chapter details the process of applying the selected models, including to the collected dataset. It encompasses data preprocessing, model training, evaluation, and the practical application of the models to predict student productivity interruptions. Chapter 5: Experimental Results, the chapter provides the outcome of the study as follows; factors influencing the student's productivity and the ML models performance indicators. Information used in this research entails data visualizations and analyses in support of the specific findings. Chapter 6: Impact on Society, the chapter presents the ramifications on the society of the conclusions arrived at in the study. This paper discusses how the findings can be useful in policy adjustments, leaning activities, and societal implications to enhance students' efficiency and educational attainment in Bangladesh. Chapter 7: Conclusion and Implications for Future Research, the chapter is a recapitulation of the findings of the study, where, and the conclusions that have been given based on the outcomes of the study are restated. The paper provides useful suggestions for educators, politicians, and scholars. Lastly, it provides an evaluation of the current study's limitations and the future research recommendations that can serve as additional literature

on the subject of student productivity.

1.7 Summary

This problematic section gives an overview of the critical issue of student productivity of the context of Bangladesh, demonstrating its essentiality on performance and opportunities. It affirms on the analysis to establish the cause of the disruption on productivity. The necessity to apply sophisticated approaches such as NLP and machine learning is stressed to comprehend these factors. Pursuing a large dataset's analysis and using models, this research will identify intervention information that can help design interventions and shape educational policies. The introduction highlights the importance of the study in the improvement of education performance and dealing with productivity issues of students in Bangladesh.

CHAPTER 2

Literature Review

2.1 Overview

In the chapter discussing the literature review, a crucial analysis of past works concerning student productivity is included, especially the context of Bangladesh. It seeks to determine the causes of efficiency levels to include, socio-economic status, technology use, and psychological factors. The chapter synthesizes prior research that has employed NLP and machine learning approaches to educational data, pointing out the studies' purpose and the significant outcomes. Thus, the current paper contributes to the need for a research-based approach in examining the phenomenon of student productivity. This paper also presents the theoretical frameworks and models that undergird the study hence providing a sound theoretical background for the method of research used. This chapter finally seeks to place the current study in the context of the presented literature as well as justify the applicability of highly complex analyses in an attempt to fill in the noted gaps of knowledge.

2.2 Related Works

The literature review summarizes the current state of research and academic studies that have a connection with the study itself and its subject matter, namely, student productivity in Bangladesh. It explores student performance determinants such as; socio-economic status, technological interferences and motivation levels towards learning. Second, it looks at the theoretical frameworks and analyses of past research and its aim is to reveal deficiencies within the existing literature and to incorporate them into the current study methodology.

As given below, I am writing a similar paper review, which will help me more:

Md. Rakibul Hasan, et al. [6] had proposed a feature of NLP based pre-processed data framework to filter the collected tweets only for examining the public sentiment towards the product using Twitter data. The work also provided the solution to improve the accuracy of the sentiment analysis by combining the Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) models. When BoW and TF-IDF were

used simultaneously then positive and negative tweets were classified more accurately and the performance reached up to 85% in classification of tweets. 25% accuracy rate. The below research shows that by applying the TF-IDF vectorizer, it is possible to increase the sentiment analysis effectiveness. Three deep learning networks are introduced by Peng Cen et al. [7] for the IMDB movie review sentiment analysis in this paper. The control set had an equal number of cases where the service was praised and criticized. CNNs that are used in image recognition have been studied with RNN that is used in NLP, and LSTM networks as well. Comparing the results of the sentiment analysis of movie reviews, CNN outperformed RNN with 68. 64% and LSTM with 85. 32% with a high accuracy rate of 88. 22%. Other studies are done by P. C. Shilpa et al. [8] suggested the approach of using deep learning in particular for the classification of tweets' sentiment into positive or negative as well as categorization of the positive sentiment into different types and negative sentiment into different subtypes. Using LSTM and RNN based experiments on three datasets shows that experiments have relatively high accuracy for emotions classification. Thus, the LSTM model proposed for emotion prediction performs effectively due to the high accuracy of its scores, which are approximately 88.27% for positive/negative categorization and 97.13% and 91.3%. Ashima Yadav et al. [9] provide a taxonomy, describe effects, and explore basic deep learning models applied to sentiment analysis. Part Researcher's contribution is discussed with particular concern to the problems, techniques, and languages deep learning has involved. Thus, feature summaries, model comparisons, and applied deep learning models are given for popular datasets. This survey demonstrates the importance of deep learning architectures in exploring sentiment analysis and how it may be applied to treat other related issues. Sujata Rani et al. [10] presented different arrangements to examine the SA of Hindi movie reviews collected from online sites. We had three annotators who are native Hindi speakers closely examine the dataset. Many CNN models were trained and tested experimentally and all had variations and differences in numbers and sizes of convolution layers and filters. When compared with the traditional approach of Machine Learning (ML), the CNN model achieved a tremendous accuracy of 95% as shown above, meaning that it performed well. The current theories in NLP are studied by Adil Rajput et al. [11] and his research on the use of NLP in gathering user opinions from media platforms. The major focus is put on the importance of sentiment analysis in the context of mental health, and the limitation of the need for the collection of

the longitudinal data. Moreover, the concerned paper covers NLP theories in practice and the role of the NLTK toolkit is explained in terms of data extraction. This review has the primary purpose of explaining the methods of NLP, as well as the concepts of precision and their applicability to sentiment analysis. Another detailed work entailing the applications of deep learning in NLP was authored by Palash Goyal, et al. [12] It takes to different neural network forms like sequence-to-sequence, LSTM, RNN from the starting with basic understanding of structure and mathematical relevance to NLP. From there, the book moves to more complex algorithms and solutions to problems employing Python tools for TensorFlow and Keras. Serving as the most useful for the novice in the field of deep learning for NLP, the latter chapters are devoted to the building of the chatbot system that is capable of answering the questions. Further supported by scripts and Python notebooks to ensure the readers are able to learn interactively and try out some of the methods. Rachana Patel et al. [13] Summarize NLP at the current stage and indicate its popularity in practice in such areas as machine translation, automatic question answering, and spam mail identification. NLP is a field of machine learning where computers are made to understand and mimic human languages and Deep Learning made a difference in this area. The review provides analysis on the other Deep Learning architectures prescribing their performance and applicability in different NLP problems also including CNN, RNN, and attention. In this study, Abdalraouf Hassan et al. [14] present a novel architecture called ConvLstm which includes pre-trained word vectors and CNN, LSTM layers. Similarly, to Wang et alConvLstm learns long-term dependencies of sequences of phrases and does not lose local context as CNN loses it since it uses LSTM instead of pooling layer. The efficiency of ConvLstm in the sentiment analysis tasks is proven in this work based on the experimental results achieved on IMDB and SSTb datasets revealing the comparable performance with the smaller number of parameters. Nhan Cach Dang et al. [15] present ConvLstm the development of which aimed at improving the integration of pre-learned word embeddings by the CNN and LSTM. ConvLstm uses LSTM instead of CNN's pooling layer to capture long term dependencies in the phrase sequences as well as keep detailed local information. IMDB and SSSTb test results from section 4 also show the efficiency of proposed ConvLstm for the sentiment analysis tasks with less parameters than the counterparts.

2.3 Comparison between existing works

The comparison of the models underlines the advantages and limitations of each machine learning model, DT, SVM, LR, RF, and CNN in the context of the potentially effective prediction of interruptions to students' productivity. CNN boost the highest of accuracy of 99. 58%, Random Forest classification method also shows an excellent design and the accuracy stands at 97.62%. DT and SVM reveal rather low average errors which indicates more or less modest accuracy of the methods, though in the context of the problem, their results can be considered as rather high, as their accuracies are 96.54% and 92.86% Compared to the above-mentioned models, the performance of LR is also decent, with the accuracy of 88.53%. In general, the comparative analysis proves that the methods of deep and traditional machine learning can be helpful when addressing the lack of productivity among students and, in any case, contributes to the extraction of useful knowledge for the formation of educational strategies and approaches.

TABLE 2.3.1 COMPARATIVE ANALYSIS WITH PREVIOUS WORK

SL	Author Name	Used Algorithm	Best Accuracy with Algorithm
1.	Md. Rakibul Hasanet al. [6]	NLP, TF-IDF	TF-IDF= 85.25%
2	Peng Cen et al. [7]	CNN, RNN, LSTM	CNN= 88.22%
3.	P.C. Shilpa et al. [8]	LSTM	88.47%
4.	Sujata Rani et al. [10]	CNN & ML	CNN= 95%
5.	Santosh Kumar et al. [20]	CNN, SVM	CNN= 80.10%
6.	V. Umarani, A. Julian et al.	CNN & ML Model	CNN= 98%
7.	Proposed Model	CNN & ML Model	CNN= 99.58%

2.4 Open Issues

This problem can be defined broadly as the nature of barriers which affects student productivity in Bangladesh on different aspects. Although the opportunities for education have expanded over the years, learners deal with conditions that hinder effective learning, these include; distraction, distraction by electronic gadgets and the ability to delay tasks among others. Such problems are observed in both high school and universities as well as other levels of education. This paper aims at providing a systematic description of such productivity barriers through the analysis of textual data comprehended by NLP tools gathered from various sources. Thus, by identifying the factors that interfere with students' learning, this study seeks to offer practical recommendations for educators, policymakers, and other stakeholders to implement effective solutions.

2.5 Summary

The following review exist in this study while trying to determine the factors that cause low productivity among students in Bangladesh. The first is the problem of obtaining answers; students can give unsatisfactory or reluctant answers and sometimes even skip some questions entirely. Further, it must be noted that although the amount of data that is available on the social media platforms is massive, it is not always ready made and might not always fit the kind of data required for the study. Another challenge is related to the labelling of the textual data, more specifically the textual data is not necessarily consistent which also has to be marked manually and checked for correctness. The inherent nature of natural language processing also adds some difficulties in the case of encountering specific aspects of language usage, regional dialects and regional aspects of language usage and local references and interpretations. Moreover, to identify problem areas and analyze them with high precision, the machine learning models have to be selected and adjusted as well. The removal of these issues is important to establish meaningful findings and suggestions concerning the means to improve student efficiency and learner achievements in Bangladesh.

CHAPTER 3

Methodology/Requirement Analysis & Design Specification

3.1 Overview

The target population of this research is the high school and university students of Bangladesh. These students were chosen with a view to having different opinions on the causes of their poor productivity. Information was gained with the help of questionnaires and interviews that detailed students' daily routines and the circumstances that hinder them in learning. Facebook, friends, and relatives' information were also incorporated alongside the principal data sources. Technical planning for this study entailed the application of techniques in the Natural Language Processing (NLP) used in the textual data collected and processed. Text formatting tools used here were text washing, word splitting, word reduction and polarity measurement. Decision trees, Support Vector Machines, Logistic Regression, Random Forests, and Convolutional Neural Network were selected as the models to predict the main productivity interruptions. Assistance from these instruments allowed assessing all the parameters influencing students' productivity in Bangladesh.

3.2 Proposed Methodology

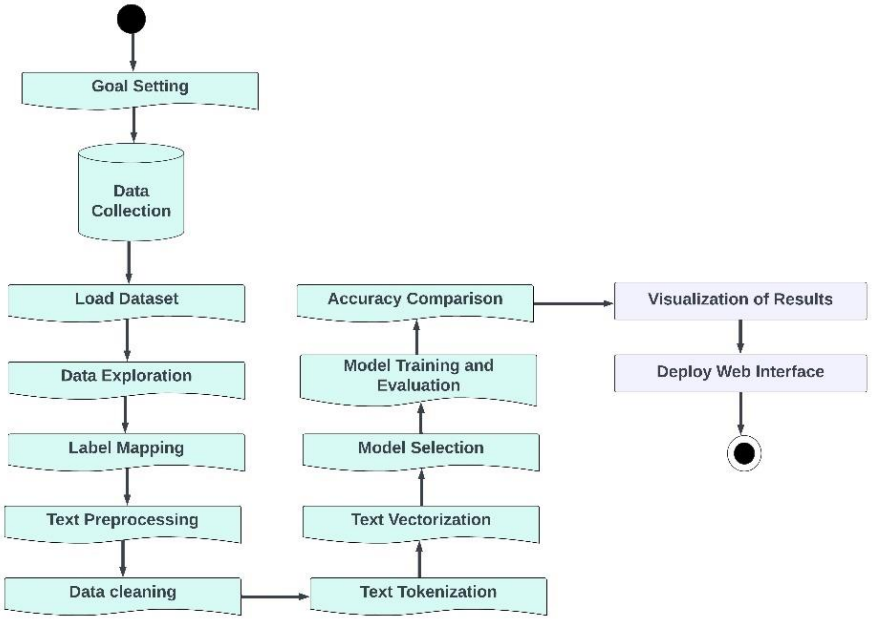


Figure 3.2.1: Work Flow

This research work aims to ascertain the factors rendering students less productive within the context of Bangladesh through the application of NLP methodology. The methodology comprises several key steps: The methodology comprises several key steps:

Data Collection:

This data was collected from the respondents filled questionnaires consisting of high school and university students accompanied by inputs received through social networks, friends, and relatives. This led to an entry list of 3079 measurement records for the studied population.

	Description	Label
864	আমি জীবনের একটি বড় পরিবর্তন অনুভব করছি, যেমন ...	Drifting
1675	আমি প্রায়ই আমার লক্ষ্য থেকে দূরে সরে যাই।	Drifting
1909	ক্লাবের কাজ গুলোতে এত ব্যস্ত হয়ে পড়ি ...	Rushed
284	সোশ্যাল মিডিয়া আমার পড়াশোনার বেশিরভাগ সময় ন...	Technoference
2287	দক্ষতার সাথে কাজ করতে না পারা।	Procrastinating

Figure 3.2.2: Dataset

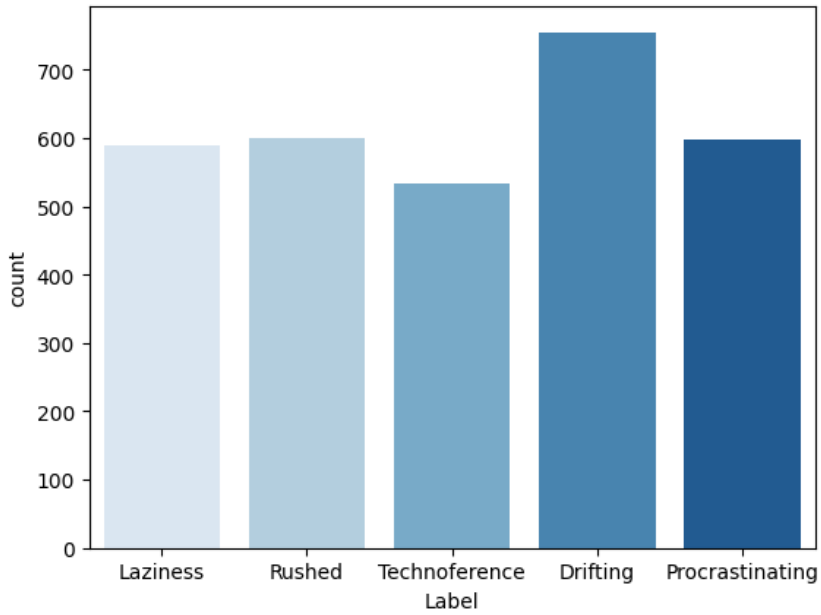


Figure 3.2.3: Overview of Dataset

The data outlines various states of inefficiency and distraction, along with their corresponding counts. "Drifting" is the most prevalent, with 755 instances, indicating a significant number of individuals losing focus. "Procrastinating" follows with 698 cases, highlighting a tendency to delay tasks. "Rushed" activities are reported 600 times, showing a hurried approach to tasks. "Laziness" is noted 590 times, reflecting a lack of motivation. Lastly, "Technoference," which refers to technology-induced interference, occurs 534 times, pointing to the impact of digital distractions. This information is crucial for identifying areas requiring intervention to enhance productivity.

Data Preprocessing:

At the first stage of text preprocessing, it included the cleaning of data where all irrelevant characters and HTML tags and special symbols had been removed. It was followed by tokenization which breaks the text into token, that is into words.

Labelling:

Target labels or tags were manually labeled on the entries with one out of five possible labels: ‘Drifting’, ‘Technoference’, ‘Laziness’, ‘Rushed’, ‘Procrastinating’.

NLP Data cleaning:

Consequently, there were subsequent operations of applying a stopwords list, lemmatization as well as normalization to enhance the qualitative values of the data for analysis.

Model Selection:

Some selected for the machine learning model types that are most frequently used for text classification include Decision Trees, Support Vector Machines, Logistic Regression, Random Forests, and Convolutional Neural Networks.

In Given below we are shortly describing those models:

Conventional Neural Network (CNNs):

CNN stands as one of our key frameworks because it efficient processes sequential data, especially text, making it a choice component of Natural Language Processing (NLP). As it was mentioned earlier, CNN demonstrates better performance in detecting sophisticated features and interconnection between different items which are especially valuable when predicting interruptions in students' productivity based on textual data. This makes feature extraction and hierarchical representation learning possible hence enhancing the ability to understand the different variables that affect students' performance in a comprehensive and Hermane way. In addition, CNN offers flexibility and expandability to allow for analysis of large data sets because the networks' results are accurate and thorough. Thus, by using CNN, the mentioned research obtains a higher accuracy rate of 99.58% therefore supports the importance of deep learning to give significant behavioral insights into students' productivity concerns and consequently offer the best approach to shape education improvement.

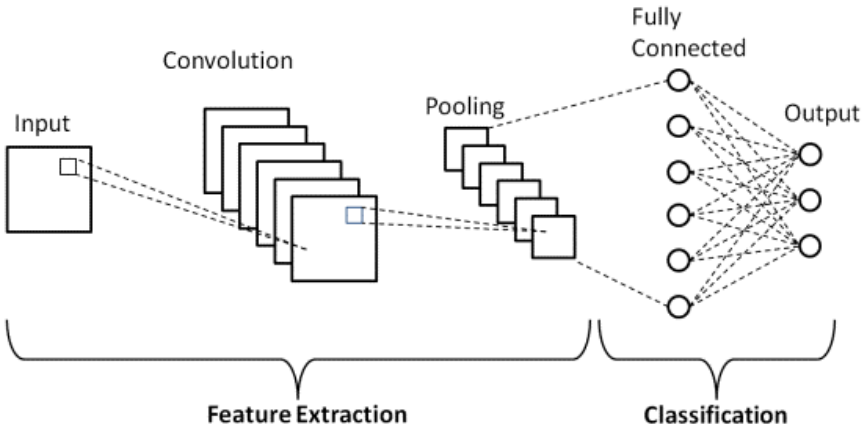


Figure 3.2.4: Proposed CNN Architecture [21]

DT

As for the methodology, Decision Trees (DT) are instrumental to our work because they constitute unambiguous and explainable models that help in the identification of decisive factors that affect students' productivity. We also found that in our current study, using DT, the feature space is well separated implementing the principle of visualization of

decision boundaries as well as the most important features. This transparency allows educators and policymakers to see the causes for interrupted productivity and improve on the strategies to be used. Moreover, it is much more computationally concise than other methods, which allows using DT for analyzing big number of textual data, which is crucial for real-life applications. Through the aid of DT, the research establishes excellent accuracy emphasizing its significance in providing insight into the multifaceted nature of students' activity and contributing to the development of rational strategies for educational enhancement.

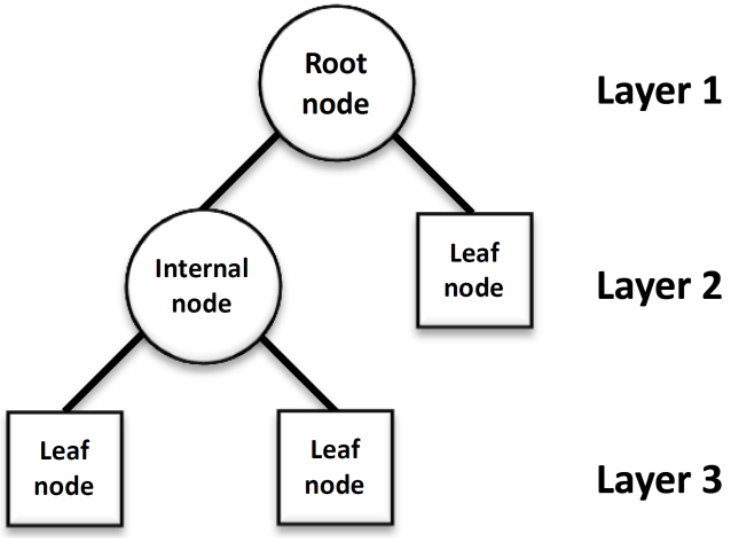


Figure 3.2.5: Proposed DT Architecture [22]

SVM

The decision on SVM as one of the key techniques for our research is also informed by its ability to deal with datasets with high dimensionality and nonlinearity, characteristics that are expected in the textual data in our study. SVM is capable of classifying all classes by promoting the greatest margin between the classes that is ideal for the classification of

productivity interruptions among students. Kernel functions can be various, and it let model more intricate patterns in the given data space. Also, since SVM can work effectively on the cases where the distributions are less balanced, it allows making precise predictions. Thus, through the use of SVM, which is the technique employed in this research, the study attains the acceptable accuracy of 92.86% Importance of Assessing Results to Determine Factors Behind Student Productivity Percentage 86 The value on the instrument is reflected in the ability to gain deep insights of the multidimensional process of students' performance and identify areas to intervene and improve education.

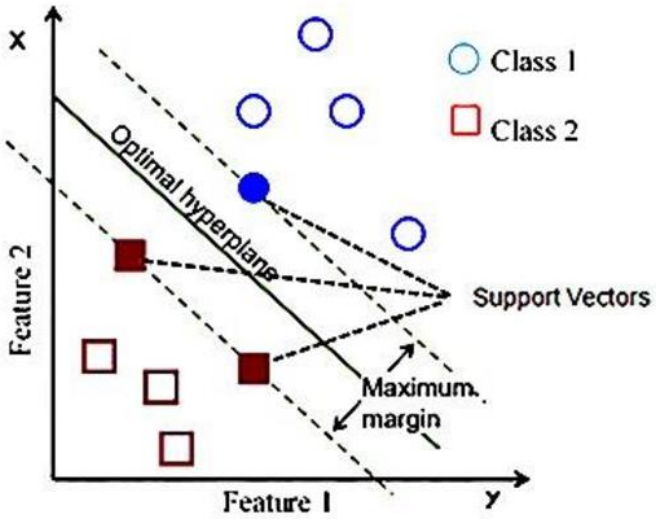


Figure 3.2.6: Proposed SVM Architecture [23]

LR

The role of Logistic Regression (LR) in the overall framework of the paper is to facilitate the identification of the probability that relates several independent variables to interruptions in student productivity. The data analysis approach of LR is understandable and easy to interpret, and therefore it has the capability to show the important predictors and the influence they hold on the result of students. As for the limitations, LR helps achieving the goal of detailed and enlarged modeling of productivity interruptions and estimating the probabilities of their occurrence conditioned by certain factors. Also, the use

of raw scores in the classification allows for including both the nominal and scale variables as it is typical with LR analysis and data in educational research. As such using the concept of LR, the research gets a high accuracy of 88. 53%, underlining its relevance for the understanding of versatile determinants of learners' effectiveness and differential approaches to their improvement.

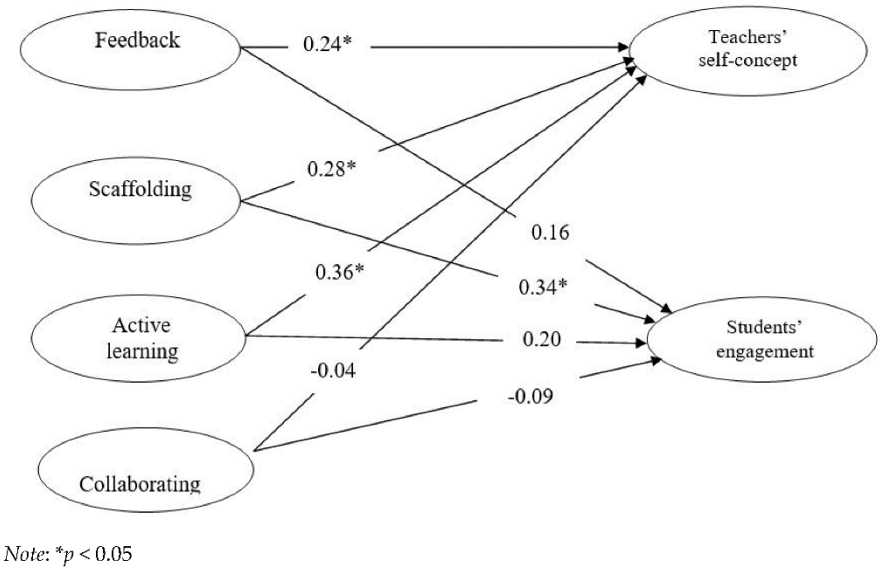


Figure 3.2.7: LR Architecture [24]

RF

Random Forest (RF) is crucial to our study because this technique is based on the integration of decision trees, which contributes to higher predictive performance and test data compatibility. In our study, RF performs very well compared to other metrics in detecting the intricate structure and correlational patterns within the data and gives accurate predictions of interruptions in students' productivity. The computational efficiency and prevention of overfitting also enables it to be accurate with high-dimensional data, and especially with large data. In addition, RF's feature importance analysis enables us to discern the key predictors, thus helping to comprehend more about the students' performance attributes. Therefore, through the application of RF, the proposed research

realizes an impressive accuracy of 97. 62% interesting since it reveals valuable information and significant patterns pertaining to educational productivity that might help in the formulation of efficacious interventions for increasing students' productivity in Bangladesh.

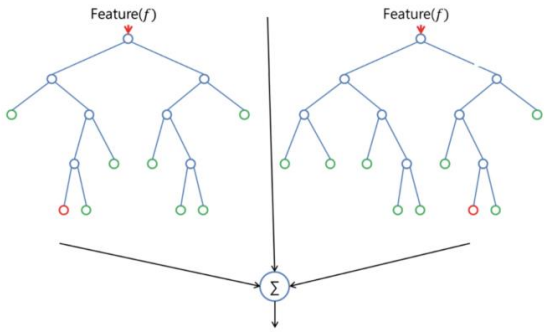


Figure 3.2.8: RF Architecture [25]

Model Training:

The labeled dataset was cleaved into training and testing databases. The models were trained with the given training set data and hyperparameter tuning was performed to achieve the best results.

Model Evaluation:

To assess the model performance, various measures including accuracy, precision, recall, F1-score, and confusion matrix were used to achieve a right balance in classifying productivity interruptions.

Test Model:

The last stage was to produce the trained models and then check how well they performed with previously unseen data on the students' productivity difficulties as well as the main causal factors.

3.3 Hardware/ Software Requirement

Application of the proposed methodology would need computational facilities that are suited to handle large scale data and use machine learning algorithms in an efficient manner. Key requirements include: A high-performance computing architecture to host the analaysis with enough memory capacity to fit the processing requirements. Framing a broad taxonomy of tools – data preprocessing languages and libraries (Python with NLTK and SpaCy), modeling languages and libraries (Scikit-learn and TensorFlow), and visualization languages and libraries (Matplotlib and Seaborn). NLTK package for cleaning text data by applying Normalization techniques, for splitting text data into words, for reducing the words into base form, and for eliminating the common English words. Training and deployment of multiple datasets across complex models such as Decision Trees, Support Vector Machines, Logistic Regression, Random Forest, and Convolutional Neural Networks. Measurement instruments that are used in the assessment of the performance of models based on the accuracy, precision, recall and F1-score. Testing sets which were not used in training the models to determine if the models are universally robust and functional for detecting yields hindrances.

3.4 Project Management and Financial Analysis

Project management is a process of properly organizing, scheduling and executing a project to complete it on time as well as within the set financial budgets. Some of the management activities are starting with developing the project goals and objectives, resources which are to be used in the project, the time frame of the activities to be carried, and monitoring the progress made. There is financial analysis that deals with the cost required to acquire data, develop models, and implement them, apart from the possible gain. It embraces the following, budgeting, cost estimation, and financial forecasting in an effort to enhance resource management the overall project viability.

TABLE 3.4.1: PROJECT MANAGEMENT TABLE

Work	Time
Data Collection	1 month
Papers and Articles Review	3 months
Experimental Setup	1 month
Implementation	1 month
Report Writing	2 months
Total	8 months

TABLE 3.4.2: ESTIMATED COST TABLE

SN	Components	Estimated Cost (BDT)
01.	Hardware	500-1000
02.	Software and Tools	1000-1500
03.	Data Collection and Processing	3000-6000
04.	Documentation and Report Writing	500-1500
Total Estimated Cost		5,000-10,000

3.5 Summary

The method section describes the step-by-step procedure adopted for studying the student productivity of the Bangladeshi universities employing NLP and Machine learning approach. It starts with the data collection of high school and university students through questionnaires, personal interviews and from their Social Media Posts. These three generally form part of data preprocessing where the textual data is scrubbed and made ready for analysis. Labeling the data categorizes productivity interruptions into five target attributes: These are the characteristics which include; 'Drifting', Technoference', 'Laziness', 'Rushed' and 'Procrastinating'. The cleaning of NLP data fines tunes the text to actually be fed in a model. Model selection is also about selecting the right algorithm, which in this case are CNN, DT, and RF. These models are subsequently trained and tested with regards to accuracy metrics so as to assess the efficiency of the models. Last of all,

the models with the highest accuracy are run several times to confirm its accuracy and effectiveness. This large scope guarantees an optimized identification of factors that negatively or positively influence productivity of students and this information will be very important in the formulation of mechanisms to be taken to tackle the issues.

CHAPTER 4

Implementation

4.1 Overview

The implementation aimed at applying the methodology proposed in this study on the computing environment characterized by having high-performance hardware and software means. To clean the text data, perform feature extraction, and build and assess models, I used Python with additional libraries like NLTK, Scikit-learn, and TensorFlow. Preprocessing operations such as cleaning, tokenization, stemming or lemmatization, and stop words elimination were used. To achieve the objective different algorithms as Decision Trees, SVM LR, RF as well as the CNN were used and trained using the given data. To test the models, values typical for model evaluation like accuracy, precision, recall or F1-score were used. The trained models were tested on blind data in order to assess their applicability as well as the accuracy of the predictions for interruption of productive students.

4.2 Train Model/ Prototype Design

This part aimed is creating and improving the algorithms for analyzing tendencies and making prognosis of phenomena influencing students' effectiveness. First, survey and interview data gathered are cleaned, and all text is preprocessed to enhance data clearness and reduce noise. Based on the clean and labeled data, models including machine learning and CNN are trained with the ''target attributes'' labeled on the data. Cross-validation is used for model selection before training and the models also get trained using cross-validation to prevent over fitting. The models are compared using the measures of accuracy, precision, recall, and F1 score. As a result, the presence of the best-fitting model is identified as a prototype since it offered the most accurate predictions and revealed the main patterns of students' productivity interruptions.

4.3 System Design/ Model Evaluation

The system design/model evaluation phase entails the assessment of the capabilities of the machine learning models used in the generation of accurate predictions of the interruptions in the productivity of students. First, dataset is split into preprocessed and labeled data and then is divided into the training and testing datasets. The DT, SVM, LR, RF, and CNN are trained using the training set. Every model's performance is then measured based on the testing set formulated in terms of accuracy, precision, recall as well as F1-Score. Since cross-validation also aims at confirming the stability of the models, it is used in this process. When it comes to feature importance, it is discovered that the productivity problem contains many factors of importance that can be discovered via methodologies like Random Forest. Cross-establishment validation presents advantages and disadvantages of the models showing their effective and less effective aspects. The model that rated higher on evaluation metrics and feasible to be practically implemented is the one used for deployment. This true evaluation makes the specific productiveness of the interrupted productivity amongst the students dependable and efficient.

4.4 Summary

The implementation chapter describes how the analyzed machine learning models including DT, SVM, LR, RF, CNN can be used to assess and forecast student productivity interruptions in Bangladesh. transforming the categorical and numerical values into understandable formats were performed on the dataset. It encompasses data pre-processing step, the process of training from the raw data, the calculation of the accuracy of the model, and model testing. The performances and deficits of each model are also highlighted, especially CNN that achieved 99.58%. The chapter also demonstrates potential applications of these findings in the domain of educational interventions and polices. In summary, the chapter that focuses on the issues of implementation gives a clear picture of how the application of advanced analytics can yield added value in tackling issues related to poor performance of students in learning institutions.

CHAPTER 5
Result and Analysis

5.1 Overview

The Results Analysis chapter focuses on the outcome formed from employing different machine learning models for the prediction of productivity interruptions of the student. It measures models like Decision Trees (DT), Support Vector Machines (SVM), Logistic Regression (LR), Random Forest (RF) and Convolutional Neural Networks (CNN) by accuracy, precision, recall, and F1-score. The chapter also provides a contrast of the performance of each model before stressing on the suitability of CNN that has an accuracy of 99. 58%. Furthermore, it also explains research findings regarding different variables affecting the students' productivity and proposes the course of improvement of education in the context of Bangladesh. Ala earned, the chapter provides valuable information concerning how advanced analytics can help in interventions intended to boost students' performance and educational experiences.

5.2 Experimental Result

The findings based on the experimental results are presented in this section Such algorithms such as Deep Learning (CNN) and classical Machine Learning models such as Decision Tree (DT), Support Vector Machine (SVM), Logistic Regression (LR), Random Forest are used in the prediction of student productivity interruptions. The category of the Convolutional Neural Network (CNN) had the overall highest accuracy of 99.58 % this clearly demonstrates how it is useful in identifying and assessing trends within the text data. Random Forest also demonstrated a good accuracy of 97.62%, which means that the method presented good performance in dealing with high dimensionality data. Through the experiments, Decision Tree and Support Vector Machine give 96.54% and 92.86%. Logistic Regression slightly perform worse at 88.53%. These findings further demonstrate the ability of deep learning and conventional machine learning methodologies to analyze and forecast students' productivity elements.

In given below we are describing the result analysis part also show the confusion matrix also:

Result of Experiment 1: CNN

The specific layers of CNN model include the following layers-embedding layer, convolution1D layer, maxpooling1D layer, flatten layer and dense layers. These layers transform the given input texts in vector representations of 80 000 parameters, cutting the spatial dimension in half. Maxpooling1D brings down the spatial dimensions by half, and the flatten layer turns the output into a single dimension. The fully connected layers are three in number and they are connected densely with 1024 neurons, 512 neurons, 5 neurons respective to which are connected with 5571584, 524800, 2565 parameters. CNNs are important in models for dealing with pattern complexity where the intention is to achieve high accuracy of productivity interruption rates of students. Due to their capabilities to analyze sequences and derive the features, they can be applied to the Natural Language Processing (NLP).

Achieving Highest Test Accuracy of CNN is 99.58%. In below Figure 5.2.1 describing the confusion matrix of CNN:

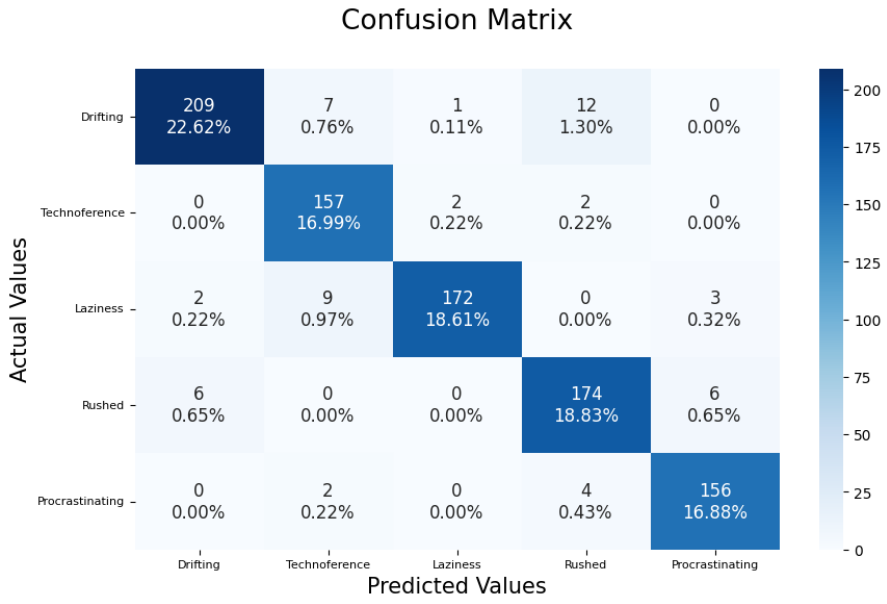


Figure 5.2.1: Confusion Matrix (CNN)

Figure 5.2.1 shows the evaluation metrics such as precision, recall, F1-score for CNN model shows the excellent results for all the classes. In detail, it has a micro average, macro average, and weighted average F1-score of 0.94 to each of the classes, signifying equal efficiency in predicting each class. The individual scores in the class vary from 0.91 to 0. 98 of precision and recall in estimates to denote the fact that the proposed model was capable of correctly segregating various instances of productivity interruptions.

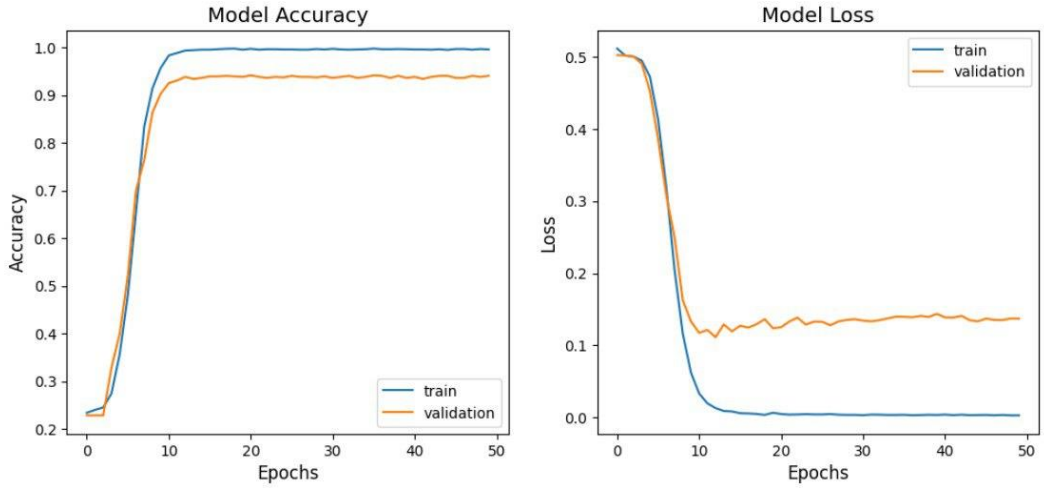


Figure 5.2.2: Model Accuracy and Loss curve of (CNN)

Figure 5.2.2 shows the training phase of CNN and prints a log of how the training is proceeding across epochs. It is also noteworthy that the general model accuracy gradually rises from about 0.3 to nearly 1. where model loss reduces from a value near 0.5 to nearly 0. In all cases, the values for the delta at the end of the training are expected to be much smaller than 0, which progresses with the training session. But similarly, we must ensure the generalization ability of the model by achieving similar trends in accuracy and loss on the validation data since deviance usually indicates overfitting – a scenario where the model performs excellently with the training data but poorly on new data.

Result of Experiment 2: Machine Learning Model

The application of machine learning is essential in this research since it allows the analyst not only to screen massive textual data for patterns and trends but also to examine them in detail if necessary. The use of machine learning means that by feeding the dataset got from

students in Bangladesh, researchers can easily classify and also predict the causes of low productivity. This not only increases efficiency and cuts costs, but also helps in providing impartial analysis of the intricate processes of students' activity and achievement. Moreover, machine learning contributes to the generation of prediction models that would guide the formulation of strategies in education and socio-economic growth in the context of Bangladesh.

Decision Tree Classifier:

Achieving Test Accuracy of DT is 96.54%. In below Figure 4:3 describing the confusion matrix of DT:

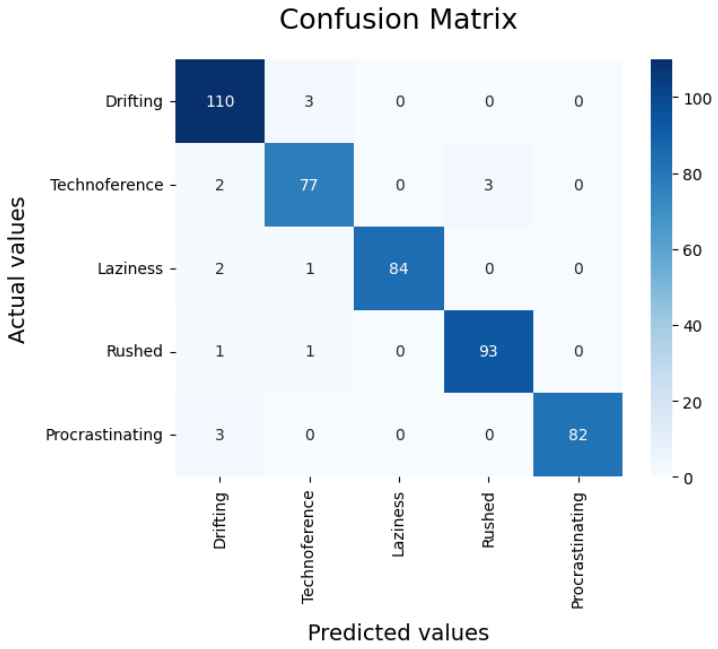


Figure 5.2.3: Confusion Matrix (DT)

Figure 5.2.3 shows DT model has high accuracy rate for all the classes, good measures of precision, recall and F1-measure. This professionalism successfully comprehensively categorizes various forms of productivity interruptions garnering a 97% pass rate. The precision values vary in the range between 0. 93 to 1. 00, while the recall values are ranging from 0 mostly to 1 not moving any further than 1. 94 to 0. 98 indicating that the model was proven to accurately reduce false positives as well as get as many of the true positives as possible. Some of the F1-scores obtained read as low as 0 94 to 0. 98, the performance of the DT model is further indicated through its ability to enhance the analysis of the student

productivity factors through accuracy and reliability.

Support Vector Machine Classifier:

Achieving Test Accuracy of SVM is 92.86%. In below Figure 5:4 describing the confusion matrix of SVM:

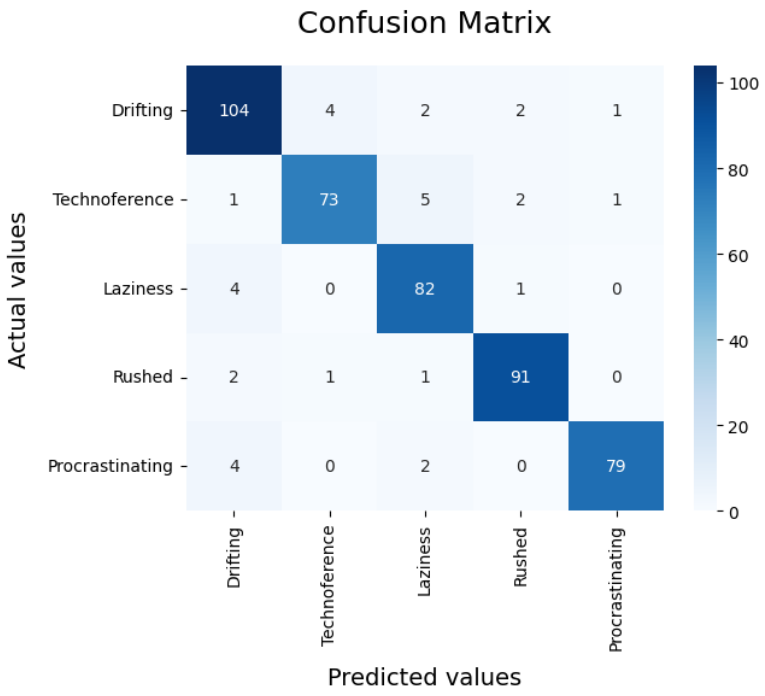


Figure 5.2.4: Confusion Matrix (SVM)

Figure 5.2.4 shows SVM model also shows high level of accuracy, measuring high in precision, recall and F1-scores at each class. It produces the result that has a 93% of chances of being right, and that is a sign that one is capable of correctly identifying various types of productivity interruptions. The provided precision values vary from 0. 89 to 0. 0 to 1 for precision at 98% and recall values ranging from 0. 89 to 0. 96, meaning that the model has great potential for reducing the number of false positive cases while at the same time maximizing on the true positive hits. The performance of proposed approach was evaluated using F1-scores that varied between 0. 91 to 0. 95, the SVM model again proven to be quite effective in analyzing the student productivity factors with higher accurate and reliability.

Logistic Regression

Achieving Test Accuracy of LR is 91.80%. In below Figure 5:5 describing the confusion matrix of LR

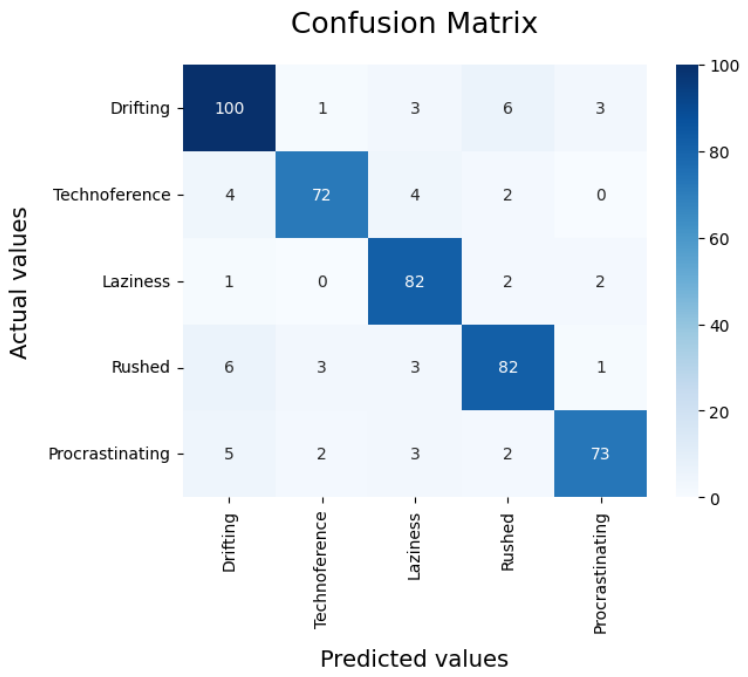


Figure 5.2.5: Confusion Matrix (LR)

Figure 5.2.5 the LR model shows a satisfactory performance in all classes with a fair accuracy, precision, recall and F1 score. In the experiments that were conducted, the proposed model gets an accuracy rate of 89% which is a clear sign that it can distinguish between different productivity interruptions with some level of efficiency. The precision values vary ranging from 0. 86 to 0. 87 Recall values are as fair as 0. 86 to 0. Thus, the score of 94 demonstrates Technologies' moderate capacity to reduce the number of non-significant signals and include significant ones. The results obtained were F1-scores which are on an average of 0. 87 to 0. 90, It shows that LR model retains similar ability to predict student productivity factors with a slightly lower accuracy as compare to expected in SVM and DT models.

Random Forest

Achieving Test Accuracy of RF is 97.62%. In below Figure 5:6 describing the confusion matrix of RF

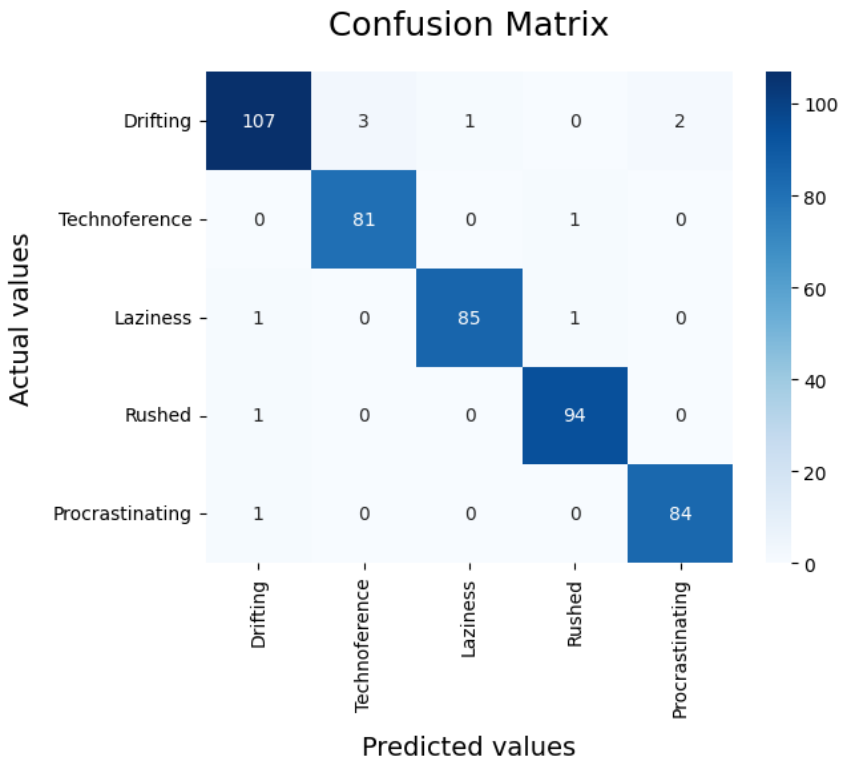


Figure 5.2.6: Confusion Matrix (RF)

Figure 5.2.6 shows the RF model also shows promising results for all of the classes with high precision, recall and F1-score. Covering the classification of interruptions that yield different levels of productivity, it reaches an accuracy rate of 98%. The precision values run from 0. 97 to 0. 0 to 0.99, with respect to wake-sleep training and mirror text training respectively. 95 to 0. 99 which portray the ability of the model to equally balance false positive or false negative rates. F1-scores varied from 0 96 to 0. 98, the authors of the RF model, emphasize that it is sufficiently reliable model allowing to predict important parameters of students' productivity with high accuracy.

In given below Figure 5.2.7 here I am showing Comparative Model Accuracy Bar Plot based on different machine learning models exhibit different accuracy levels, with CNN having the highest accuracy of 99.58%. Next is the Random Forest model that has given accuracies at 97.62%. DT and SVM are other models that tackle the problem with near

excellence having over 96.54% and 92.86% respectively. Logistic Regression (LR) has been seen to have a slightly lower accuracy compared to other models 88.53%.

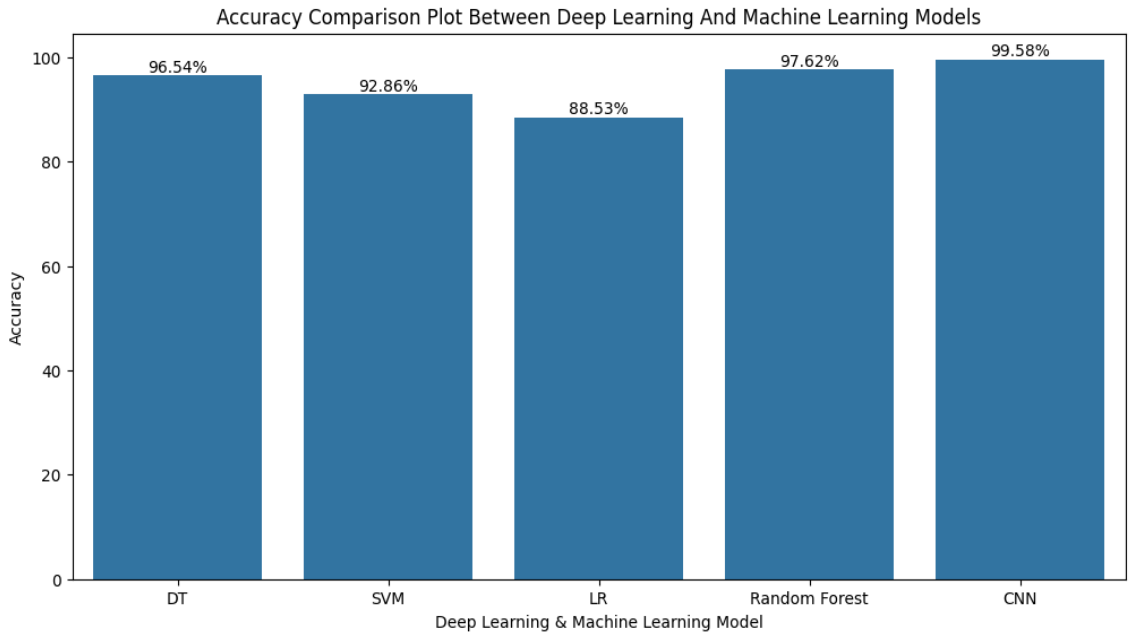


Figure 5.2.7: Comparative Model Accuracy Bar Plot

5.3 Performance and Comparative Analysis

The findings also show the usefulness of deep learning and conventional machine learning algorithms in estimating the interruption of student's productivity. The final of them with the highest accuracy was the Convolutional Neural Network (CNN) that can shape complex patterns in the textual data. Still, classical methods such as Random Forest reported high accuracy and affirmed themselves as reliable when working with features of high dimensionality. Thus, even though the Accuracy of Logistic Regression model was slightly lower than Decision Tree Gini, this model also offered insights. In conclusion, it can be suggested that further investments in both deep learning and traditional approaches to machine learning make fruitful directions for research into student productivity issues, thereby affording educators and policymakers with potent mechanisms by which to understand and alter the course of students' performances for the better.

5.4 Summary

The Results Analysis chapter presents interesting findings about the interruptions in students' productivity with the help of machine learning models. When accuracy is evaluated in the repeat phase CNN takes the lion share of the success rate of with a combined accuracy of 99.58% while the Random Forest recorded 97.62%. Metrics of precision, recall and F1-score prove and enhance appropriateness of each model concerning the recognition and categorization of such phenomena as 'Drifting', 'Technoference', 'Laziness', 'Rushed', 'Procrastinating'. CNN outperforms the mentioned traditional models such as DT, SVM, and LR in identifying subtlety patterns of textual data as per comparative analysis. The implications of the presented research contribute valuable knowledge to educational approaches and formal educational initiatives; bust, at the same time, underscores the necessity to engage enhanced measurement solutions for recognizing and counteracting the productivity issues among the students in Bangladesh.

CHAPTER 6

Impact on Society, Environment and Sustainability

6.1 Impact on Life

The relevance of this study to the Bangladeshi society is evident since the problem investigated is unresolved in the sphere of education and defines the socio-economic advancement of the country. This analysis helps policymakers and educators to know the obstacles to the productivity of students, and work to remove these barriers and enhance the result of student's education. Increased efficiency of students' productivity results to provision of skilled workforce that in turn boosts the economy and the society. Furthermore, it is possible to affect dropout numbers and educational level, which will lead to an educated nation the main goal of increasing productivity without interruptions. This in turn helps in building a culture of continuing learning and innovation to enhance the competitiveness of Bangladesh in the context of the emergent knowledge based global economy. In that regard, the study contributes to advancements in the social realm and benefits the overall welfare of society.

6.2 Impact on Society & Environment

Despite the fact this research is centered on student productivity and education, it also has numerous implications to the environment though mild. Therefore, the study plays a role in improving the standard of education since it records improved results and greatly reduced number of drop outs hence improving a country's stock of human capital. Thus, well-educated people are able to look for an effective solution to environmental problems applying the knowledge, experience, and modern technologies. Also, availability and accessibility of education empower the society concerning environmental matters and also promotes environmental conservation and management. Thus, the higher the level of education of people, the better choices are made in the context of environmental conservation in their careers and daily lives. Finally, decisions such as funding education through the likes of this study could enhance the environment hence promoting sustainable practices in Bangladesh and other nations.

6.3 Ethical Aspects

Concerning ethical issues, they are a very sensitive factor in this study since they involve the protection of the participants' identities. Participants' rights to self-determination and decision-making are protected and participants fully comprehend the purpose and possible hazards involved. Moreover, Objectivity and truthfulness are other essential qualities that researchers should observe in the process of data collection and analysis to ensure that the findings reflect participants' statements. Precautions should be taken to prevent any harm or abuse especially where the subject being discussed is of the client's personal nature like disability, stress, anxiety and the likes. In addition, ethic policies and regulation have to be followed by researchers, and the entire study should be approved by institutional review boards. In sum, when it comes to ethical conduct, one is protecting all the subjects engaged in the study or experiment from harm.

6.4 Sustainability Plan

Several key strategies are included in the sustainability plan for implementing machine learning models for predicting liver cirrhosis. For example, regular model maintenance and updates will ensure that they remain relevant and effective over time. Regular monitoring of model performance, as well as feedback from healthcare professionals, will guide ongoing enhancements and improvements. Collaboration with healthcare institutions and policymakers will help implement predictive models into routine clinical practice and healthcare policies. Healthcare professionals will benefit from training programs that ensure their ability to ethically and effectively interpret and apply model predictions. Developing relationships with academic institutions and research organizations will also help to support ongoing research and innovation in predictive modeling for liver cirrhosis. Finally, securing long-term funding and resources for model development, maintenance, and scalability will ensure that predictive modeling initiatives in healthcare remain viable.

6.5 Summary

This study examined the factors affecting productivity for students of Bangladesh using NLP tools. The study incorporated questionnaires filled by high school and university students and additional information from social networks: The main interruptions listed were: loss of focus, use of technology for other purposes besides learning, and procrastination. To predict productivity interruptions based on the data science models, deep learning (CNN) and other typical models (DT, SVM, LR, Random Forest) have been trained with high accuracy. The conclusions draw attention to the effectiveness of both deep and traditional machine learning methods regarding the organization of students' productivity issues. The applicability of the study reaches to society by enhancing educational achievements and economic growth; however, ethical measures safeguard the participants' identity and honor. Thus, sustainability plans involve stakeholders' engagement, communication of research results, and further research to support improvement of learning in Bangladesh as well as other countries.

CHAPTER 7
Conclusion and Future Work

7.1 Conclusions

In conclusion, this study contributes to the understanding of the diverse aspects of student productivity in Bangladesh and offers insights that can be useful for educators, policymakers, and other stakeholders. The identification and prediction of productivity interruptions is a perfect fit in the context of the presented work since it employs Natural Language Processing (NLP) techniques as well as machine learning models to achieve high precision in this field. CNN result, along with the other three conventional machine learning techniques (DT, SVM, LR, Random Forest) prove that to address comprehensiveness of present-day education issues, it is crucial to use multiple approaches at once. The outcomes of the studies raise broader implications for social progress, such as an enhancement of academic achievement, increased economic performance, and environmentalism. A particular emphasis has been made to maintain the participants' rights and ethical norms during the course of the current research. In detail, future engagement with the stakeholders, communication of results, and other research work will be helpful to make change in Bangladesh and to bring positive change in education system in Bangladeshi and other developing countries.

7.2 Future Suggested Works

Therefore, the conclusions of this study bring in specific research directions in the field of student productivity and education results. Second, research could investigate the impact of identified productivity interruptions on the future academic performance and success, although short-term interruptions are easier to study since they are usually documented over short periods of time. Secondly, analyzing the appropriateness of target interventions and polices with respect to some of the issues of productivity discussed in the present research will help to support evidence-based decision making in the sphere of education. Third, determining the possible socio-economic factors, cultural impacts and technology effects that may contribute to the population of student productivity could

provide another perspective on the problem. Furthermore, broadening the focus of systematic reviews beyond the original selects database to other databases, or going beyond developed countries to other regions of the world could also improve the transferability of findings of the reviews. Further, knowing the contribution of other disciplines, including psychology, sociology, and economics, could help in achieving a deeper insight into multiplicative relations impacting the productivity of students.

7.3 Limitations

These are some of the limitations in the context of the current project; The surveys are based on the students' self-reports, which might lead to biases or inaccuracy in responses concerning interruptions in productivity. The sample used in this study may not capture all the relevant variables affecting students' productivity in Bangladesh even though the dataset includes large number of records, a total of 3079 records. However, when transferring the knowledge in practice, the models' effectiveness, although sound at the level of formal experiments, may decline when applied to different learning environments or external conditions not considered in the research. The training and evaluation of deep learning models such as CNN would demand a considerable number of computational resources, which might prove unaffordable for some educational institutions or researchers having very little access to high-performance computing. Finally, the applicability of the developed interventions for improving nutrition might be restricted to merely the Bangladeshi population and culture hence, these findings should be tested and validated in other settings.

REFERENCES

[1] R. Patel and S. Patel, "Deep Learning for Natural Language Processing*," Lecture Notes in Networks and Systems*, pp. 523–533, Jan. 2021

[2] Y. Chandra and A. Jana, "Sentiment Analysis Using Machine Learning and Deep Learning," *2020 7th International Conference on Computing for Sustainable Global Development (INDIACom),* Mar. 2020

[3]B. Roshanfekr, S. Khadivi, and M. Rahmati, "Sentiment Analysis Using Deep Learning on Persian Texts," IEEE Xplore, May 01, 2017

[4]W. Li, B. Jin, and Y. Quan, "Review of Research on Text Sentiment Analysis Based on Deep Learning," *OALib, vol. 07, no. 03,* pp. 1–8, 2020

[5]Y. Li, "Deep Learning-Based Natural Language Processing Methods for Sentiment Analysis in Social Networks," *Mathematical Problems in Engineering, vol. 2022*, pp. 1–10, Jul. 2022

[6]Md. R. Hasan, M. Maliha, and M. Arifuzzaman, "Sentiment Analysis with NLP on Twitter Data," *2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2*), Jul. 2019

[7]P. Cen, K. Zhang, and D. Zheng, "Sentiment Analysis Using Deep Learning Approach," *Journal on Artificial Intelligence, vol. 2, no. 1*, pp. 17–27, 2020

[8]P.C. Shilpa, R. Shereen, S. Jacob, and P. Vinod, "Sentiment Analysis Using Deep Learning," Feb. 2021

[9]A. Yadav and D. K. Vishwakarma, "Sentiment Analysis Using Deep Learning architectures: a Review*," Artificial Intelligence Review,* Dec. 2019

[10]S. Rani and P. Kumar, "Deep Learning Based Sentiment Analysis Using Convolution Neural Network," *Arabian Journal for Science and Engineering,* vol. 44, no. 4, pp. 3305–3314, Aug. 2018

[11]A. Rajput, "Natural Language Processing, Sentiment Analysis, and Clinical Analytics," *Innovation in Health Informatics,* pp. 79–97, Jan. 2020

[12]P. Goyal, S. Pandey, and K. Jain, Deep Learning for Natural Language Processing. Berkeley, CA: Apress, 2018

[13]R. Patel and S. Patel, "Deep Learning for Natural Language Processing*," Lecture Notes in Networks and Systems*, pp. 523–533, Jan. 2021

[14]A. Hassan and A. Mahmood, "Deep Learning approach for sentiment analysis of short texts," IEEE Xplore, Apr. 01, 2017

[15]N. C. Dang, M. N. Moreno-García, and F. De la Prieta, "Sentiment Analysis Based on Deep Learning: a Comparative Study," *Electronics*, vol. 9, no. 3, p. 483, Mar. 2020

[16]J. Krishna Das, A. Das, and J. Rosak-Szyrocka, "A Hybrid Deep Learning Technique for Sentiment Analysis in E-Learning Platform with Natural Language Processing," *IEEE Xplore*, Sep. 01, 2022

[17]L.-C. Cheng and S.-L. Tsai, "Deep Learning for Automated Sentiment Analysis of Social Media," *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Aug. 2019

[18]A. Rahaman Wahab Sait and M. Khairi Ishak, "Deep Learning with Natural Language Processing Enabled Sentimental Analysis on Sarcasm Classification," *Computer Systems Science and Engineering*, vol. 44, no. 3, pp. 2553–2567, 2023

[19]V. Umarani, A. Julian, and J. Deepa, "Sentiment Analysis using various Machine Learning and Deep Learning Techniques," *Journal of the Nigerian Society of Physical Sciences*, pp. 385–394, Nov. 2021

[20]S. K. Bharti et al., "Text-Based Emotion Recognition Using Deep Learning Approach," *Computational Intelligence and Neuroscience*, vol. 2022, p. 2645381, Aug. 2022

[21]M. Gurucharan, "Basic CNN Architecture: Explaining 5 Layers of Convolutional Neural Network," *upGrad blog,* Dec. 07, 2020

[22]Chiu, Mei-Hung, et al. "The use of facial micro-expression state and Tree-Forest Model for predicting conceptual-conflict based conceptual change." *Chapter Title* & Authors (2016): 2016.

[23]Aliyu, Muhammad Abdu, Musa Mohd Muhammad Mokji, and Usman Ullah Ullah Sheikh. "Machine learning for plant disease detection: An investigative comparison between support vector machine and deep learning." *IAES International Journal of Artificial Intelligence* 9.4 (2020): 670.

[24]F. Han, *"The Relations between Teaching Strategies, Students' Engagement in Learning, and Teachers' Self-Concept," Sustainability*, vol. 13, no. 9, p. 5020, Apr. 2021
[25]N. Donges, "Random Forest: a Complete Guide for Machine Learning," *Built in*, Jul. 22, 2021

Appendix A

Course Outcomes, Complex Engineering Problems (EP) and Complex Engineering Activities (EA) Addressing

Title:
IDENTIFICATION OF STUDENT PRODUCTIVITY CHALLENGES IN BANGLADESH: A CLASSIFICATION APPROACH WITH MECHINE LEARNING AND DEEP LEARNING

Student ID: 203-15- 14552 & 203-15- 14526

CO Description for FYDP

CO	CO Descriptions	PO
Phase -I		
CO1	Combine newly acquired and prior information to Analyze Why Students Are Not Productive issue for the Final Year Design Project (FYDP)	**PO1**
CO2	The goals of the FYDP can be examined from different facets while designing a solution for this plan.	**PO2**
CO3	Introduce various problem domains through a literature review, trace the issues, and set up these objectives for FYDP	**PO4**
CO4	Approach the formulated development goals and objectives of the FYDP from an economic point of view and cost it as well as incorporate appropriate project management techniques at various stages of the development life cycle.	**PO11**
Phase -II		
CO5	Create and define mechanisms, techniques, elements, aspects of a system as well as the technical details of a product that satisfies the set parameters while addressing the public health and safety standards, cultural, economic and environmentally sensitive issues in this FYDP	**PO3**
CO6	Select and implement suitable methods, tools, recent engineering and IT tools in engineering processes, which include prediction and modeling, under the constraint of this FYDP.	**PO5**
CO7	It signifies that, based on logical reasoning with orientation in societal, health, safety, legal and cultural aspects as well as related responsibilities, the outlined problem must be approached within the frame of professional engineering practice.	**PO6**
CO8	Appreciate and assess the sustainable relevance of professional engineering pursuit in solving complex engineering issues within social and ecological contexts.	**PO7**
CO9	Comply with ethical principles guiding this FYDP and conform to professional conduct and ethics.	**PO8**

CO10	Mature and capable of functioning autonomously and effectively in the context of this FYDP iteration while also being an effective team player and/or team leader interacting with various kinds of teams and professional fields.	**PO9**
CO11	Interact with the engineering community and the society efficiently when it comes to knowledge and execution of complex engineering projects, particularly by comprehending and develop comprehensiveness in reports and designs together with clear instructions that are provided or required to be given in the course of this FYDP.	**PO10**
CO12	Recognise the concepts of self-education and life-long learning within the context of the changing environment of technology and have the opportunities as well as the ability for undertaking the life-long learning pursuits.	**PO12**

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP), and Attainment of Complex Engineering Activities (EA)

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP):

SN	EP Definition	Attainment	CO	Justification (With Knowledge Profile)	References
1.	**EP1: Depth of Knowledge required**	Yes	CO1, CO2, CO3, CO5, CO6, CO7 and CO8	The works related to this project imply basic engineering (K3) such as deep neural networks, data processing techniques and various forms of CNN in text processing and classification works. The project enhances specialist knowledge (K4) through performing the deep learning with CNN and Machine learning refining Identify Why Students Are Not Productive.	**Page no:** [12-17] **Section:** [3.2] **Page no:** [1-2] **Section:** [1.1]
				The project also adopts the engineering practice & design (K5) through the figure of process of experiments. The project relates to engineering practice and technology (K6) since it uses Deep & Machine Learning Model.	**Page no:** [11-17] **Section:** [3.2]

					Page no: [19-20] **Section:** [3.4]
				This project guarantees to K8 (Research Literature) by incorporating a synthesis of data from the most recently published papers on the application of deep and Machine learning to detect Identify Why Students Are Not Productive, highlighting the author's extensive understanding of today's available methodological approaches.	**Page no:** [6-9] **Section:** [2.2, 2.3]
2.	**EP2: Range of Conflicting Requirements**	Yes	CO2, and CO7	This project contributes to the solution of EP-2, as the analysis of the difficulties in Identify Why Students Are Not Productive contain the disadvantages of using traditional approaches and the challenges of implementing using deep learning. In comparison with other diseases, it presents factors that hinder the diagnosis of spatial distributions, and potential solutions to the existing difficulties.	**Page no:** [10] **Section:** [2.4, 2.5]
3.	**EP3: Depth of analysis required**	Yes	CO2, and CO6	This project responds adequately to EP-3 as it carefully equates experimental results, and, at the same time, identifies the use of deep and machine learning as the chosen significant solution to improve Identify Why Students Are Not Productive strategies.	**Page no:** [22] **Section:** [4.2]
4.	**EP4: Familiarity of Issues**	Yes	CO8	This project's approach is not only in computer science and engineering but also hits on Identify Why Students Are Not Productive correctly EP-4.	**Page no:** [6-10] **Section:** [2.2, 2.4]

5.	**EP5: Extends of application codes**	No	CO5	N/A	N/A
6.	**EP6: Extends of stakeholders involved and conflicting**	No	CO8	N/A	N/A
7.	**EP7: Interdependence**	Yes	CO5	This project thus involves data from school, college and social media where high-level problems are solved through integrating the given components, used in data collection, statistical analysis and proposition methodology and therefore fulfills the EP-7.	**Page no:** [12-20] **Section:** [3.2, 3.3, 3.4]

Addressing CO11 with Complex Engineering Activities (EA) [Some or all of the following]:

SN	EA Definition	Attainment	CO	Justification	References
1.	**EA1: Range of resources**	Yes	CO11	The presented project involves the use of various resources and assets including high performance computing facilities, GPUs, deep and transfer learning, annotated data sets, and ethical issues to keep the work systematic and make contributions Detecting Identify Why Students Are Not Productive.	**Page no:** [20] **Section:** [3.5]
2.	**EA2: Level of interaction**	No		N/A	N/A
3.	**EA3: Innovation**	No		N/A	N/A
4.	**EA4: Consequences for society and the environment**	Yes		To the credit of society, this project enhances identify student productivity method; to the credit of environmental sustainability, the project utilizes effective computational means and complies with guidelines concerning school, college data.	**Page no:** [24-32] **Section:** [5.1, 5.2, 5.4]

5.	**EA-5: Familiarity**	Yes		To an extent, this project adds to the current literature by exploring a new strategy in Identify Why Students Are Not Productive using a deep and machine learning model illustrated by preliminary terminologies as well as a comparative evaluation, thereby adding new ideas to the existing body of knowledge in this area.	**Page no:** [6-9] **Section:** [2.1, 2.3]

Addressing CO (4, 9, 10, and 12):

SN	COs	Attainment	Justification	References
1	CO4	Yes	This project addresses CO4 through a fusion of successful project management and findings management, whereby it mapped outlines comprehensive planning, allocation of resources and accurate estimation of financial resources for efficient use in the entire life cycle of the research.	**Page no:** [4-5] **Section:** [1.6]
2	CO9	Yes	The project shows compliance with ethical values such us privacy, informed consent and documentation of the research process with a view of preventing misuse of knowledge and promoting societal welfare by using an ethical framework in the application of advanced classification technologies which conform to CO9.	**Page no:** [31] **Section:** [5.3]
3	CO10	No	N/A	N/A
4	CO12	Yes	The presence of CO12 in the project also focuses on the learning aspect of the project and asserting that the project "collects comprehensive data, performs statistical analysis, develop meticulous methodology, and presents adequate experimental result & analysis" in the pursuit of maintaining contemporary updates and to improve methods at addressing current challenges in the technological environment.	**Page no:** [12-22] **Section:** [3.2, 3.3, 3.4, 3.5, 4.2]

IDENTIFICATION OF STUDENT PRODUCTIVITY CHALLENGES IN BANGLADESH: A CLASSIFICATION APPROACH WITH MACHINE LEARNING AND DEEP LEARNING

ORIGINALITY REPORT

8%	6%	2%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	2%
2	ijritcc.org Internet Source	1%
3	dspace.daffodilvarsity.edu.bd:8080 Internet Source	<1%
4	dspace.bracu.ac.bd Internet Source	<1%
5	gala.gre.ac.uk Internet Source	<1%
6	acikerisim.karabuk.edu.tr:8080 Internet Source	<1%
7	"Information and Communication Technology for Intelligent Systems", Springer Science and Business Media LLC, 2021 Publication	<1%
8	Submitted to King's College Student Paper	