PADDY YIELD ESTIMATION BY DEEP LEARNING APPROACH

BY

ASHRAFUL ISLAM

ID: 182-15-11445

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

Supervised By

Mr. Mahimul Islam Nadim

Lecturer

Department of CSE

Daffodil International University

Co-Supervised By

Ms. Tapasy Rabeya

Senior Lecturer

Department of CSE

Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

26 JANUARY 2024

APPROVAL

This Project titled **"Paddy Yield Estimation by Deep Learning Approach"**, submitted by Ashraful Islam, ID No 182-15-11445 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 26 January 2024

BOARD OF EXAMINERS

Md Taimur Ahad

Chairman

Dr. Md. Taimur Ahad (MTA)
Associate Professor & Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Abdus Sattar(AS)
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Tapasy Rabeya(TRA)
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

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

DECLARATION

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

Mr. Mahimul Islam Nadim
Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

Ms. Tapasy Rabeya
Sr. Lecturer
Department of CSE
Daffodil International University

Submitted by:

Ashraful Islam

Ashraful Islam

ID: **182-15-11445**

Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

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

I really grateful and wish our profound our indebtedness to Mr. Mahimul Islam Nadim**,** Lecturer Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "paddy yield prediction" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

I would like to express our heartiest gratitude to **Dr Sheak Rashed Haider Noori,** 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.

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

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

ABSTRACT

Three crucial crops in Bangladesh are farmed concurrently based on the country's climate and seasons. The individuals mentioned are Aush, Aman, and Boro. Various sorts of damage occur in Bangladesh at different times as a result of natural catastrophes, in accordance with the country's climate. Examples include storms, torrential downpours, floods, and river overflows. They inflict harm. Rice is the primary agricultural product of Bangladesh. The rice yield is significantly impacted by these natural disasters. Consequently, a multitude of different natural disasters transpire. These encompass agricultural yield decline, insufficiency in food supply, and potentially even widespread starvation. In order to address these issues, I have devised a model that can accurately predict the rice yield for the current season by examining historical data. For this particular situation, I have employed D planning. Among the three models I have tested, the LSTM model demonstrated superior performance. The third model demonstrated an R2 square A score of 0.77% with less of loss at 0.22%. Here use of 1560 data. The model has achieved unprecedented advancements.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Introduction

As a result of Bangladesh's substantial reliance on the agricultural sector, agriculture is highly important to the country's economy. For the purpose of ensuring that future generations would have access to an adequate quantity of food, it is of the utmost importance to build an agricultural system that is not only free from dangers but also environmentally beneficial, sustainable, and economical. Rice, jute, fish, together with fruits and vegetables, are the primary products that come from agricultural production. There has been an increase in the amount of wheat harvested in the most recent harvests. A rise in maize yields has occurred as a direct result of the demand for chickens. Not only are textiles, leather, leather goods, ceramics, and ready-to-wear apparel produced locally, but also leather goods and leather products. In Bangladesh, there are many different kinds of rice paddies that are cultivated. Aus, Aman, and Boro are the three sub types of rice farming that are generally used. In a manner comparable to the cultivation of rice and wheat, Bangladesh is responsible for the cultivation of a significant quantity of potatoes. Bangladesh is largely the world's largest consumer of vegetables. Because of the high yield it produces and the high nutritional value it possesses, it is considered to be the fourth most important crop in the country. It is essential to conduct a thorough analysis of agricultural parameters in order to get the highest possible crop yield. According to the data provided by the bureau of statistics (BBS), Boro was responsible for 54% of the country's rice production for the year 2021-2022. Aman supplied 38%, while Australia produced 8%. This year, the Department of Agriculture and Environment forecasts a yield of 34,000 tonnes of Australian exports from 14,000 acres and a production of 2,000,000 tonnes of Boro from 47.54 million hectares. It is the primary source of sustenance for the bulk of Bangladesh's population, which is estimated to be 149 million people. In 2009, the average amount of milled rice consumed by each individual was 173.3 kilograms. A number of efforts are now being put into action by the government in order to boost rice production and reduce the amount of rice that is imported. In order to keep prices at a low level, rice farmers are given subsidies for their inputs. A total of 712 million dollars was allocated for subsidies in the year 2010. Through the use of a government input distribution card, small and marginal farmers were provided with financial subsidies for energy, irrigation fuel, fertilizer, and other forms of government support. Various places in Bangladesh, including the highlands, irrigated lowlands, rain-dependent lowlands, stagnant medium-deep water, salty parts, and tidal non-saline regions, are all used for the production of rice. Prior to the beginning of the monsoon season, Bangladesh receives up to 400 millimeters of precipitation between the months of March and May. This precipitation happens before the monsoon season begins. The rain that has fallen has made it possible for farmers to plant crops that are resistant to the effects of drought. For the purpose of this investigation, I take advantage of deep learning strategies to forecast crop yields.

1.2 Motivation

Rice is a staple in our diet. Bengali identity is demonstrated by the love of fish and rice. Rice is a staple food in our household. I am able to fulfill our requirements if I precisely forecast the output of rice. Potatoes are classified as a vegetable in Bangladesh, although in other countries they are regarded a staple crop and the primary source of carbohydrates they contain. As reported by the Bangladesh Bureau of Statistics in 2018, this plant is responsible for producing over 63 percent of the veggies that are consumed in Bangladesh. Because of the vast quantity of potatoes that are processed in Europe, a variety of products are produced, including starch, beer, potato-based meals, flour, and dextrose. By utilizing cutting-edge technologies, I can forecast the yields of food crops. To resolve the issue, I contemplated employing machine learning.

1.3 problem definition

It is important to note that the notion of "Machine Learning" is of great significance in the context of the modern world of advanced information and communication technology (ICT). The implementation of machine learning will make a substantial contribution to the expansion of our agricultural manufacturing industry. It is vital to have a comprehensive understanding of the issues that are currently being faced in this business as well as the needs that are related to them in order to devise an effective solution. In addition to the use of machine learning in agriculture, it is vital to have a thorough awareness of government regulation, the standards of the technology industry, and the numerous educational possibilities that are available.

1.4 Research Question

- When evaluating crop output, it is important to consider many variables of the climate.
- Which variety of rice would yield the most quantity of grains?
- Which algorithms would be employed?
- Can the productivity of climate change-resilient crops be enhanced?
- Can one implement multiple strategies to enhance fruit yield?
- Which hybrid strategies can be employed to effectively discriminate yield rates?
- Is this strategy universally applicable and correctly implementable at the field level by all individuals?
- Who will be the project's intended users?
- Is the necessary data accessible for this project?
- Is it possible for farmers to adopt this strategy automatically?

1.5 Research Methodology

In the methodology section of our research study, I provided a full explanation of how I obtained data, how I prepossessed that data, how I categorized the data that emerged from the data collection, how I choose algorithms, how I constructed those algorithms, and how I evaluated those algorithms. My methods for carrying out our research were discussed in this section, which was devoted to those methods. At the end of this part, which will bring the discussion to a close, a definition of the output of the model that was offered will be presented.

1.6 Research Objective

In the methodology section of our research study, I provided a full explanation of how I obtained data, how I prepossessed that data, how I categorized the data that emerged from the data collection, how I choose algorithms, how I constructed those algorithms, and how I evaluated those algorithms. My methods for carrying out our research were discussed in this section, which was devoted to those methods. At the end of this part, which will bring the discussion to a close, a definition of the output of the model that was offered will be presented.

1.7 Report layout

Chapter 1 In preliminary investigation crucial part was this first component. In addition, this chapter discusses why I decided to do such research. The most important component of this chapter is the problem description. The research issue, very challenging is product review, are discussed in this part.

Chapter 2 As it's a input analysis that gives a quick overview of the work done in this area. The work using machine learning that is linked to this is explained here.

Chapter 3 Here it's a brief overview of a technique or procedure is provided. What was the outcome of the analysis in this segment?

Chapter 4 It's in the evaluation of the outcomes. It comprises the graphical analysis' findings.

Chapter 5 Here it's the final section of the research. Here discussion about the model's output. This portion also demonstrates the precision of the relationship. This part also includes the concept and performance's online implementation. The chapter ends with a discussion of the work's limitations. The study's potential was also encoded.

CHAPTER 2

BACKGROUND STUDY

2.1 Terminology

There is currently no technology available that can accurately forecast crop production in our specific region, and I have not yet discovered a reliable method. The problem at hand is being analyzed within the framework of the decline in agricultural productivity in Bangladesh and the application of machine learning techniques. Deep learning is an area of artificial intelligence that enables computers to learn and enhance their performance without human intervention. Deep learning enables autonomous computers to extract and derive conclusions from data. Defining "learning" in the context of deep learning algorithms can be challenging due to the various methods of extracting knowledge from data, which depend on their formation. Defining "learning" becomes challenging due to this. In order to acquire new knowledge, a substantial amount of consistent data is required. The algorithm's performance increases proportionally with the number of samples it learns from. Every input-output pair is encompassed inside a problem domain, which can be represented as a line, cluster, or another statistical description. I employed these techniques to achieve the most favorable outcomes.

2.2 Related work

There is a growing prevalence of utilizing deep learning to address challenges in predicting. Applying deep learning to address yield prediction has been extensively contemplated. The utilization of deep learning has greatly facilitated the streamlining of this technique.

Ganguly et al. [3] present a deep learning-based rice yield prediction model utilizing satellite imagery. Unreliable crop yield estimation can cause crop failures, food shortages, and market instability. They forecast rice yield using remote sensing, meteorology, and machine learning. CNN extract satellite image features, and LSTM networks model temporal dependencies. The authors test their model on Odisha rice yield data from 2000 to 2016. Their model significantly outperforms traditional regression models. Their model has an MAE of 0.73 tons per hectare, 26% better than the best regression model. Their model accurately predicts rice yield and provides actionable insights for farmers, policymakers, and the food industry.

Ruggiero et al. [4] offer a precision agriculture yield prediction deep learning model. Yield unpredictability can cause crop losses and lower farmer profits. Their model predicts crop yield using remote sensing, field data, and machine learning. A CNN extracts characteristics from remotely sensed imagery, and an LSTM network models temporal dependency. The authors tested their model on Italian maize yield data. Their model significantly outperforms traditional regression models. Their model has an MAE of 3.3%, 19.7% better than the best regression model. Their model accurately predicts crop yield and provides information for precision agriculture practices like variable-rate fertilization and irrigation.

Zhongqi et al. [5] deep learning precision agriculture model predicts crop yield. Crop yield changes might reduce farmer income. Remote sensing, meteorology, and machine learning drive their agricultural output. CNN analyze remote sensing data and LSTM networks simulate temporal dependencies. Iowa soybean yield data tests the authors' model. They outperform regression models. Their MAE is 1.77 bushels per acre—25% better than the best regression model. Their model accurately estimates crop production and aids precision agriculture practices like variable-rate fertilization and irrigation.

Sung-Ju Jang et al. [6] found that semiconductor companies' competitiveness depends on manufacturing efficiency. Optimizing wafer maps before production is one of the best ways to boost efficiency. Various criteria, including as the number of shots, total number of dies, lithographic performance rates, minimum film objective (MFO), and cost, can be employed to assess the productivity of a wafer map. This research employs machine learning to ascertain agricultural yield. Their method is based on the correlation between the position of the wafer and the fluctuations in yield at the die-level, which is independent of the testing process. Enhancing yield forecasts through the use of spatial modeling. The return model and method may develop wafer maps with productivity benefits of 8.59 percent, according to the trial results.

Niketa Gandhi et al. [7] review India's rice-growing regions' machine-learning approaches. India produces a lot of rice, wheat, and legumes. Rice-growing areas need good weather to thrive. This paper reviews the WEKA approach's SMO classifier application to 27 Maharashtra districts. Publicly available Indian government statistics were utilized to predict rice crop output. SMO fared poorly on the same dataset.

Rakesh Kumar et al. [8] emphasized that in agriculturally dependent societies, good agricultural planning is crucial to economic growth and food security. Production rates, market tendencies, and government policies are all examples of inputs. The forecast of crop output rates, the prediction of weather, the classification of soil, and the categorization of crops are only some of the areas that have benefited from research applying statistical or mathematical methods that are pertinent to agricultural planning. It might be challenging to pick what to grow on a plot of land if more than one crop is viable at different times of the year. In order to maximize national economic growth, this research shows that the CSM can address the yield selection problem and boost seasonal plants per production. The potential for the proposed method to boost agricultural net yields has also been discussed.

Anshal Savla el. at. [9] research, precision agriculture requires the utilization of agricultural technology that is at the cutting edge of its industry. In this specific piece of research, they discussed a number of different categorization methods that can be used for data mining. After then, Through the application of these algorithms to data collected over time, predictions about soybean crop yields are generated. Furthermore, among the several categorization approaches, a comparison of algorithms is performed to determine which algorithm is most appropriate for yield estimation. This is done so that the most accurate yield estimate may be obtained.

Yogesh Gandge et al. [10] The economy is dominated by farming. Like other countries, India's agricultural output can suffer from extreme weather disasters like flooding and drought. Crop projections can only be made by carefully analyzing soil quality, pH, EC, N, P, and K. Crop prediction requires several datasets, making it a popular data mining choice. Data mining helps them gain insights from vast databases. I examine harvest yield prediction data mining methodologies in this research. The accuracy of feature extraction and classifier application determines a crop yield estimation system's success. This paper summarizes the results of many agricultural output forecast algorithms employed by different authors, including their accuracy and suggestions.

Farmers today frequently utilize the yield prediction approach created by Monali Paul et al. [11] to choose the most fruitful crops. It is therefore an interesting challenge to try to predict future harvests. A farmer's past experience and familiarity with a particular plot of land and crop were crucial to his or her future success. Here, I present a data mining strategy for estimating the most probable soil-related variables. Success in harvesting will be indicated by the projected category yield. A classification law is explicitly recognized for the problem of predicting agricultural yields using the Naive Bayes classifier and the K-Nearest Neighbor approaches.

According to research by Mohammad Motiur Rahman et al. [12], these limestone features have a substantial influence on environmental factors as erosion, wind speed, and humidity. Bangladesh, which is located in the foothills of the Himalayas, has a diversified landscape. Long-term human occupancy has resulted in the formation of microregions. Each of these locations has a unique microclimate of its own. To do this, it is essential for a food business owner to carefully consider which parts of a property would provide the most profit. The goal of this research was to predict agricultural output in the future by applying machine learning techniques. The models were subsequently "trained" to associate present meteorological conditions with agricultural yield success. The next step is to test the models' abilities to predict future climatic factors that have not yet been identified.

S. Bhanumathi et al. [13] addressed how to evaluate crop yields, however data mining is a relatively new field of study. Increased harvests are a top priority for farmers. No farmer ever stops thinking about how much he might reap. Learn more about how soil alkalinity is calculated, and how it is affected by variables like pH and location. Third-party approaches, such as APIs for the environment and temperatures, soil type, soil nutritional content, precipitation in the region, and soil conditions (K), are used to determine nutrient percentages like nitrogen (N), phosphorus (P), and potassium (K). To build a model, I will investigate a selection of these metrics and apply several machine-learning techniques to the gathered data. The module employs a model that consistently calculates crop production and gives the typical customer an appropriate fertilizer proportion based on environmental and field-specific parameters, hence increasing agricultural productivity and farmer income.

Ratchaphum Jaikla et al. [14] developed a technique for predicting rice yield. Most authors have attempted precise rice yield forecasts, but traditional methods are laborious and often off. In this paper, I will use SVR, the most widely used image prediction model, to create a method for forecasting rice output. Estimates of soil nitrogen, measurements of mosaic virus weight, and yield projections for rice are used in this article's three-stage prediction process. They evaluate the results of the Crop Yield Model (CSM-Rice simulation model) and compare them to those of commercial implementations such as DSSAT4. The results agree with those predicted by the CSM-Rice simulation model, providing further evidence that their method works. In the same vein, their model's mistake is manageable.

M. M. Hasan et al. [15] show the historical ups and downs of the rice price in Bangladesh. To mitigate the rapidity of the market, they speculated on the future cost of rice. I provide a baseline technique for estimating yields for Aus, Aman, Boro, and potato using regression algorithms, taking into account common yield-impacting characteristics, based on our expertise and the research gap in previous studies.

Madhu et al. [16] present a new model to estimate paddy yield using meteorological and remote sensing data. The authors use deep learning to improve agricultural productivity prediction. Weather, remote sensing, and vegetation indicators are used in the proposed model. Both data sets are used to capture the complicated interactions between environmental conditions and paddy production. The model outperforms traditional paddy yield estimation methods. The deep learning algorithm learns and extracts key patterns and characteristics from meteorological and remote sensing data, improving forecasts. This method helps farmers, policymakers, and researchers make better agricultural decisions.

2.3 Comparison of this study

I have chosen to specialize in computer vision and deep learning due to their compatibility with domains such as agriculture and the natural environment, as evidenced by my review of relevant academic articles and initiatives.

- For the achievement of my objectives, it is most effective when it aligns with the requirements of both parties involved.

- The LSTM algorithm regularly achieves the best level of accuracy when classifying numerical data, consistently obtaining 100% or higher accuracy.
- It can be augmented with various resources and is user-friendly.
- Evaluating quantitative data, such as projected output, in order to determine the optimal strategy.
- By appropriately utilizing LSTM layers and ensuring sufficient training, it is possible to attain commendable accuracy in the outcomes.

2.4 Scope of this problem

One of the most important goals that I have set for myself is to create a system that is capable of fast differentiating between different yield rates. The LSTM has demonstrated an amazing level of performance, as demonstrated by our research. In the not too distant future, I will be able to make use of these technologies in a variety of contexts, particularly in significant agricultural initiatives. My goal is to make the duties at hand as easy to complete as possible, and I will make it a point to make myself easily accessible to everyone. Through the utilization of our technology, prosecutors are able to efficiently carry out their duties in a variety of sectors of the agriculture business in the country, which are responsible for the cultivation of considerable crops.

2.5 Challenges

The primary challenge of this task is the preparation of data for future management. I utilized advanced deep learning techniques to accurately ascertain our dataset and predict future alterations. Another prevailing concern in Bangladesh is the dearth of essential provisions and employment opportunities.

Efficient data acquisition is essential for deep learning. Gathering a substantial amount of data was challenging for this research endeavor. Processing this system on a conventional device or computer proved to be challenging. Due to the absence of complete data, the algorithm requires our data to be in a specific format in order to accurately detect numerical values.

When choosing a model for deep learning, there are multiple alternatives available. Choosing the appropriate model is crucial for them. Choosing reliable data and selecting the appropriate model is a more straightforward task. There are numerous data classification models available. I utilize Artificial Neural Networks (ANN), bidirectional Long Short-Term Memory (bi. LSTM), and Google's TensorFlow framework to create our model.

The system necessitates a meticulously configured apparatus, therefore rendering the processing technique intricate.

2.6 Summary

A discussion of the earlier research on this subject may be found in Chapter 2, which can be accessed by clicking on this link. Additionally, deep learning and regression methods are utilized in conjunction with an investigation into the English language. There is also an investigation of the English language. It has been demonstrated through all of the background study that our regression analysis project is the most accurate.

CHAPTER 3
METHODOLOGY

3.1 Methodology

Deep learning integrates statistical analysis, deep learning techniques, and database management systems. This approach is employed to examine extensive data sets in order to identify trends. I employed the six phases of Data Mining to successfully accomplish the research project. During the data preprocessing stage, I observed three specific tasks: noise removal, mean imputation for null values, and level encoding. Initially, I gathered project data from many websites.

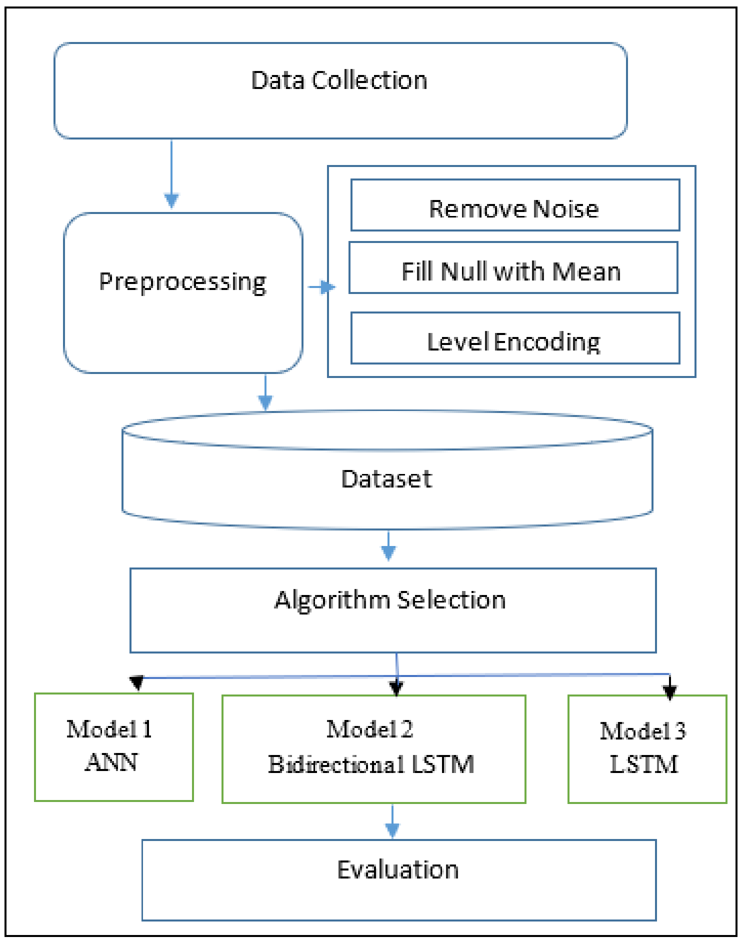

Figure 3.1 Methodology Diagram

The data was preprocessed using the following steps: noise reduction, mean imputation for null values, and level encoding. Upon establishing a project dataset, I opted to utilize the deep learning techniques of ANN, Bidirectional LSTM, and LSTM Regression. I am prepared to proceed. Subsequently, I executed our algorithms, which performed proficiently as evaluators. Presented below is a concise elucidation of each item.

3.2 Data collection

Whenever a scientific project is being undertaken, the collection of data can be a challenging endeavor. I was able to obtain the required information by visiting the Bangladesh Agricultural Research Council's official website, which may be accessed at http://www.barc.gov.bd/. The data is separated into two distinct sections that are well delineated. This section of the document was utilized for the purposes of learning and instruction when it was first implemented. It was through the utilization of an additional component that projections were produced. The initial segment was allocated for the purpose of training and testing, whilst the subsequent portion was designated for forecasting. Both instances occurred. For the preliminary examination, I chose 1560 daily expenditures from the years 2022 and 2023 in eight places in Bangladesh. The cities include Barisal, Chattogram, Dhaka, Khulna, Mymensingh, Rajshahi, Rangpur, and Sylhet. All cities are located inside the borders of Bangladesh. The debate commenced with their participation. For the forecasts section, I employed the mean value of 120 daily prices encompassing all cities during the year 2023.

3.3 Preprocessing

During the preprocessing phase of data mining, our team went through all three of these steps. The following is a summary of the steps:

Remove Noise:

The data collection I prepared includes the variables Area, Production, Yield, and Production as independent variables. However, I have only considered yield as the sole factor. I have reached the determination that it would be most advantageous to cease production.

Simply substitute the mean for the null:

In each category, the mean of all the parameters that were selected. Regarding each district, I am in possession of division-specific data. The missing argument needs to be filled in using a specific method prescribed by the system.

Label encoding:

I have complete control over our 64 national districts. I am aware that computers lack the ability to comprehend strings. Consequently, I converted them into numerical values ranging from 0 to 63. Furthermore, the crop names were represented by numerical codes 1, 2, 3, and 4, corresponding to Aus, Aman, and Boro, respectively. Level Encoding utilizes advanced data analysis techniques to enable robots to comprehend and interpret complex operational data.

Table 3.1 Label Encoding Representation of data

	District	Division	Year	Crop name	Avg WindSpeed	Avg Sunshine	AvgMax Temp	Avg MinTemp	Avg Cloud Coverage	Avg Humidity
0	0	3	0	1	1.35	6.03	31.59	21.64	3.08	79.59
1	1	1	0	1	1.21	5.75	30.59	21.68	3.31	74.92
2	2	0	0	1	1.08	7.02	31.05	21.85	3.48	82.76
3	3	0	0	1	1.08	5.02	31.05	21.85	3.48	82.76
4	4	0	0	1	1.08	5.02	31.05	21.85	3.48	87.76

Feature Engineering:

Figure 3.2 represents the correlation analysis of my dataset. To reduce the dataset dimension, I applied correlation analysis in my dataset and I set a threshold over -0.75 or 0.75. If any two features contain correlation score more than these two values then these two feature selected as highly correlated then I choose any one of them and another is removed.

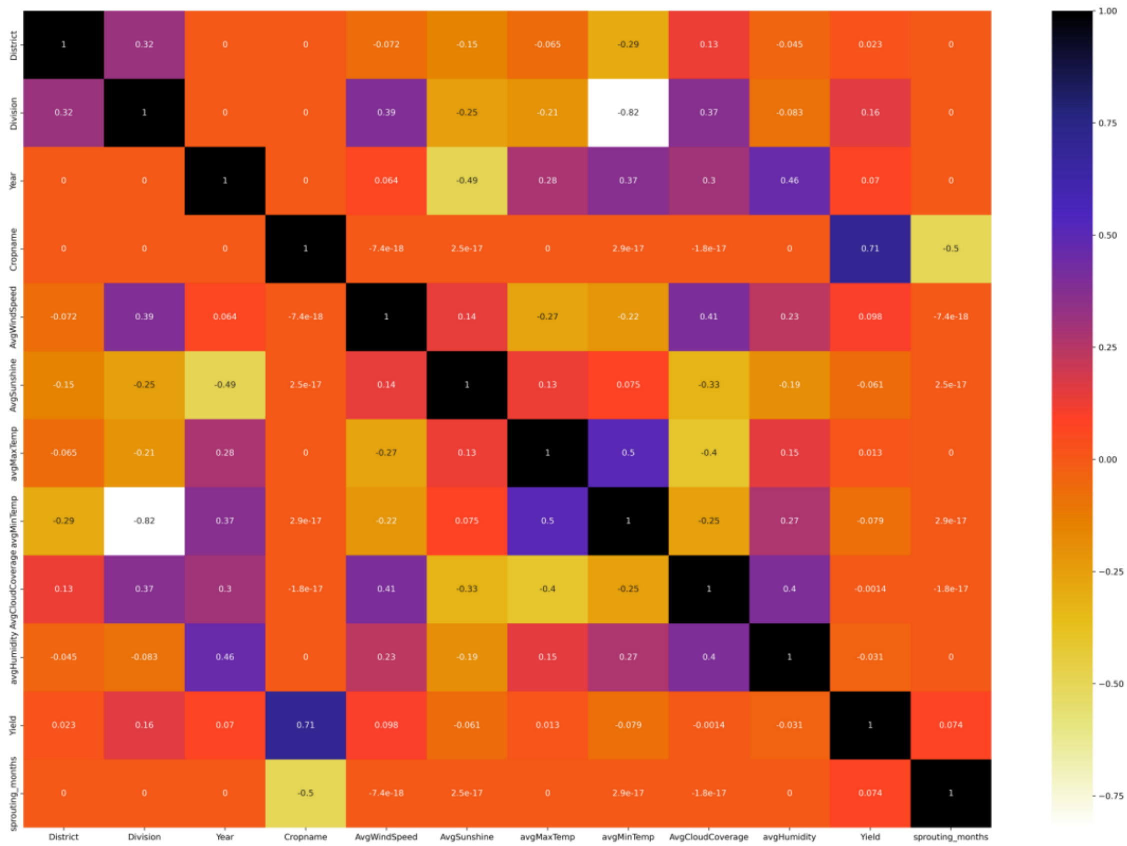

Figure 3.2: Correlation Analysis

3.4 Algorithm Implementation

I attained the maximum attainable precision while upholding a fair rate of data usage following the implementation of the requisite algorithms. The three deep learning algorithms also attained quite commendable outcomes. I selected the LSTM approach based on the expectation that it will provide the most precise prediction of crop yield. In this part, I employ three specific methodologies and tactics to apply deep learning techniques: Artificial Neural Networks (ANNs), Bidirectional Long Short-Term Memory (LSTM), and Long Short-Term Memory (LSTM). To address this particular scenario, I employed three distinct solutions, each corresponding to a different model. The initial model is an Artificial Neural Network (ANN), the second model is a Bidirectional Long Short-Term Memory (LSTM), and the third model is a Long Short-Term Memory (LSTM).

3.5 Implementation Requirement

Programming for computers is the subject of study.

For deep learning and natural language processing, choose a programming language. The Python packages Scikit-Learn and TensorFlow are among the most well-known tools.

Models and Structures for Deep learning:

It is important to make use of frameworks for machine learning that are compatible with computer languages. TensorFlow and Scikit-Learn are two examples of well-known frameworks that offer a diverse selection of implementation techniques respectively.

The feature of FastText that allows it to be integrated or merged into another framework or system.

Make use of the LSTM tool in order to generate word embeddings that are specifically customized for the Bangla language by applying complex approaches.

CHAPTER 4

EXPERIMENT RESULT AND DISCUSSION

4.1 Experiment Result

The Artificial Neural Network (ANN), Bidirectional Long Short-Term Memory (LSTM), and Long Short-Term Memory (LSTM) are the three approaches and strategies that I use in this section to execute deep learning techniques. I utilized three distinct strategies, each of which corresponded to a different model, in order to deal with this particular circumstance. ANN is the first model, Bidirectional LSTM is the second model, and LSTM is the referent of Model 3, which is the third model. With an R-squared value of 77%, model 3 (LSTM) demonstrated the most favorable findings, despite the fact that all three models produced results that were satisfactory. A greater reduction in the rate of loss was achieved by the model 3.

Table 4.1: Accuracy Table

Model	R_2Score	Loss
1	0.75	0.22
2	0.75	0.22
3	0.77	0.20

4.2 Result Analysis

In this part of the presentation, I will use sophisticated methods to filter all of the models by utilizing a pipeline in order to determine the best possible outcome. It is not only the R-squared score that is used to select the final model; other superior measures that my model performs well in are also taken into consideration.

ANN Model 1

The initial model is the Artificial Neural Network (ANN) foundation. This model has demonstrated commendable precision. In this particular scenario, the loss amounts to around 22%, which corresponds to a rate of 75%. I have further analyses that I will implement to assess their effectiveness.

Architecture

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense (Dense)               (None, 128)               1408

 dropout (Dropout)           (None, 128)               0

 dense_1 (Dense)             (None, 64)                8256

 dropout_1 (Dropout)         (None, 64)                0

 dense_2 (Dense)             (None, 32)                2080

 dropout_2 (Dropout)         (None, 32)                0

 dense_3 (Dense)             (None, 1)                 33

=================================================================
Total params: 11777 (46.00 KB)
Trainable params: 11777 (46.00 KB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

Figure 4.2: ANN Model 1 Architecture

Model 1 of the Artificial Neural Network is seen in Figure 1. A total of 128 forms were produced as a result of my use of a dense layer in conjunction with a dropout layer. Both Dense_1 and Dense Dropout_1 provide output forms that are comprised of 64. Both Dense_2 and Dense Dropout_2 produce output forms that are equal to sixty-two. Dense_3, which is the final layer, has an output shape of 1, is the final layer. The last layer is required to have a single output shape, which is of the utmost significance. As a whole, this model contains 11,777 different parameters. This model started out with 128 layers and ended up with just one layer when it was all said and done.

Training Vs validation loss

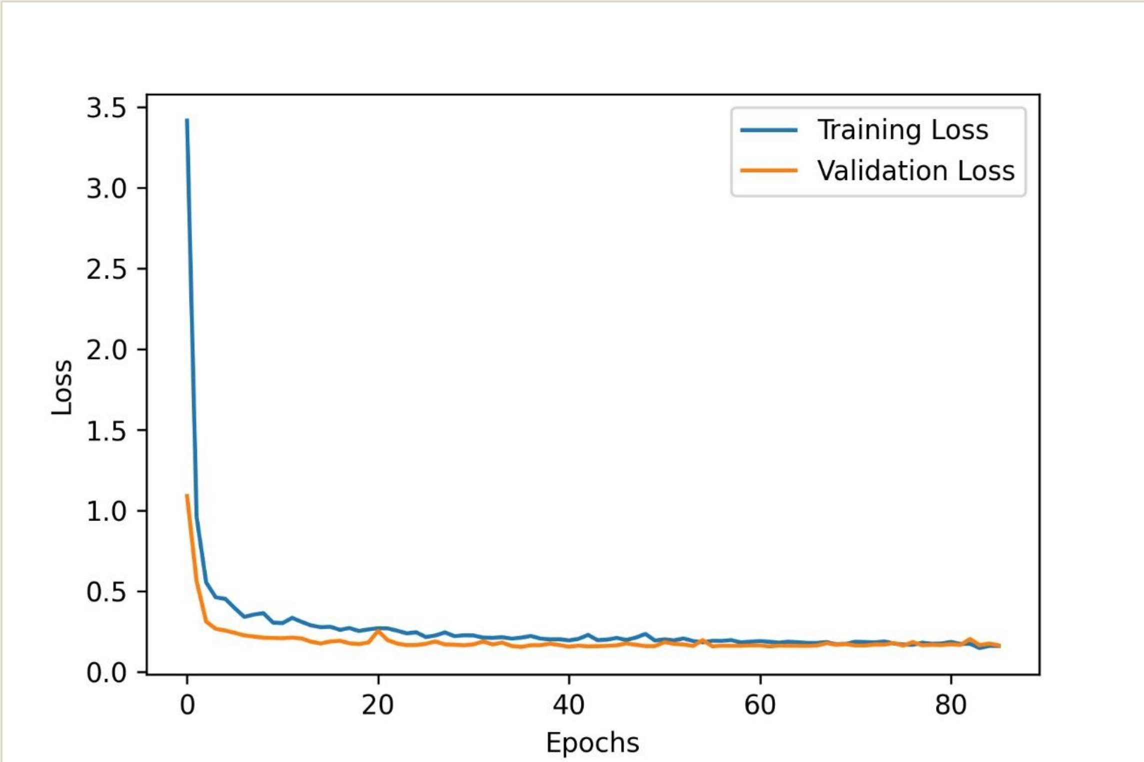

Figure 4.3: Model 1 Training vs validation loss

Figure 2 displays a comparison between the training loss and the validation loss. The blue line corresponds to the training loss, whereas the orange line corresponds to the validation loss. According to this chart, there are few inaccuracies throughout the range of 0 to 60 epochs. No faults are detectable here beyond sixty. There is a lack of evidence indicating overfeeding in this case. Despite the lesser level of data loss owing to validation loss, this model fails to satisfy the primary objective as indicated by its R-squared score of 0.75%. Consequently, I have made the decision to abstain from employing this method for future implementation due to its tendency to cause substantial losses.

Bidirectional LSTM Model 2

The Bidirectional LSTM foundation is the second model that is being discussed. A respectable level of precision has been achieved by this model. Taking into consideration this particular scenario, the loss amounts to around 22%, which is equivalent to a rate of 75%. For the purpose of determining how effective they are, I have other analyses that I will put into action.

Architecture

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 bidirectional (Bidirection  (None, 1, 256)            142336
 al)

 dropout (Dropout)           (None, 1, 256)            0

 bidirectional_1 (Bidirecti  (None, 1, 128)            164352
 onal)

 dropout_1 (Dropout)         (None, 1, 128)            0

 bidirectional_2 (Bidirecti  (None, 64)                41216
 onal)

 dropout_2 (Dropout)         (None, 64)                0

 dense (Dense)               (None, 16)                1040

 dense_1 (Dense)             (None, 1)                 17

=================================================================
Total params: 348961 (1.33 MB)
Trainable params: 348961 (1.33 MB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

Figure 4.5: Bidirectional LSTM Model 2 Architecture

A representation of the architecture of the Bidirectional LSTM model 2 is shown in Figure 1. An output of 256 forms was produced as a result of my utilization of a dropout layer in addition to a bidirectional layer. 128 is the value that is produced by both the bidirectional _1 and Dropout output types. 64 is the value that is produced by both the bidirectional _2 and the Dropout_2 output forms. The final layer is called Dense_1, and it will produce a shape of 32 as its output. The last layer is required to have a single output shape, which is of the utmost significance. In total, there are 348961 parameters assigned to this model. The beginning of this model consisted of 256 layers, and it ended with just one layer altogether.

Training Vs validation loss

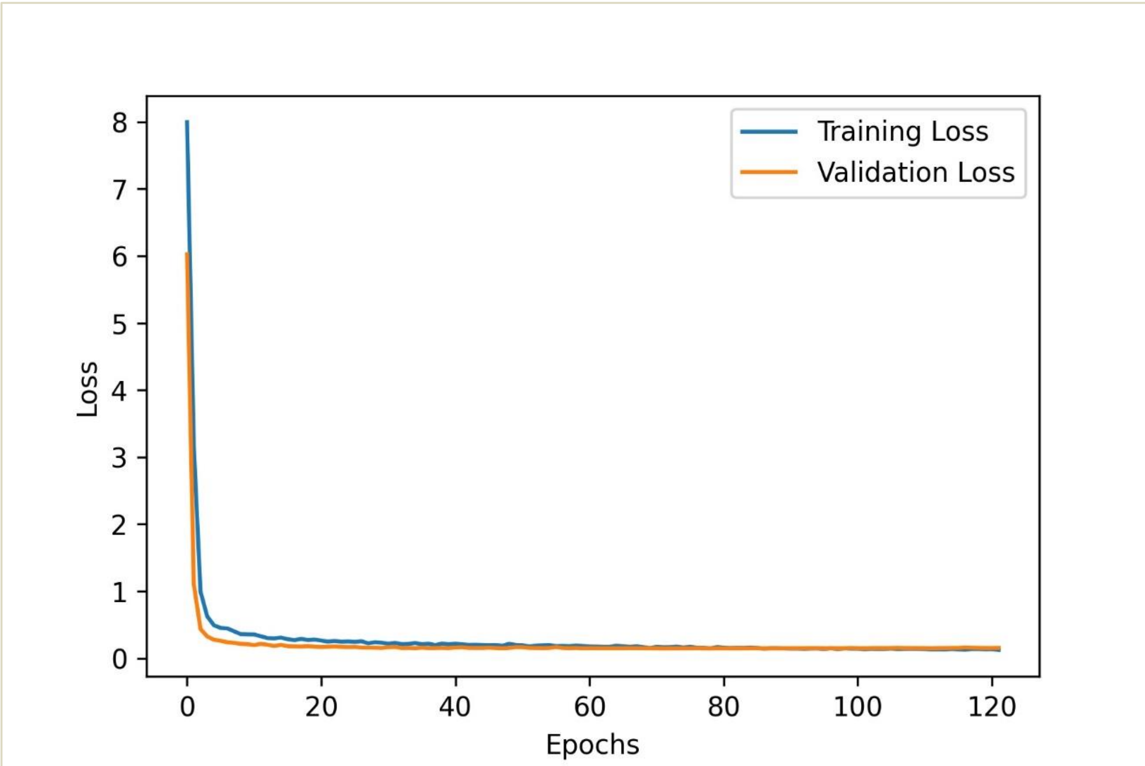

Figure 4.6: Model 2 Training vs validation loss

A comparison of the training loss and the validation loss is shown in Figure 3. It is important to note that the blue line represents the training loss, whereas the orange line represents the validation loss. Within the range of 0 to 40 epochs, this figure observation appears to demonstrate that there is just one inaccuracy total. If you wait until forty, you won't see any faults here. There is no evidence of overfeeding here. With a Squared score of 75%, this model does not meet the primary objective, despite the fact that it has a somewhat lower amount of data loss due to validation loss. Therefore, I have not adopted this approach for the purpose of implementing faults that result in a significant amount of loss.

LSTM Model 3

For the third model that is being discussed, the LSTM foundation is being considered. This model has accomplished a level of precision that is reasonable in its level of accuracy. When this particular scenario is taken into consideration, the loss amounts to around 20%, which is equivalent to a rate of 77%. I have additional analyses that I will put into action in order to determine how effective they are. I say this because I have other analyses. The best results and the lowest loss rate were achieved by this model in comparison to other models.

Architecture

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm (LSTM)                 (None, 1, 128)            71168

 dropout (Dropout)           (None, 1, 128)            0

 lstm_1 (LSTM)               (None, 64)                49408

 dropout_1 (Dropout)         (None, 64)                0

 dense (Dense)               (None, 32)                2080

 dense_1 (Dense)             (None, 1)                 33

=================================================================
Total params: 122689 (479.25 KB)
Trainable params: 122689 (479.25 KB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

Figure 4.7: LSTM Model 3 Architecture

LSTM model 3 is depicted in Figure 4, which shows its structure overall. By utilizing a Long Short-Term Memory (LSTM) layer in conjunction with a dropout layer, I was able to generate an output that was comprised of 128 different forms. There is a 64-bit output size for both the LSTM_1 and the Dense Dropout_1 class. It is important to note that both Dense_2 and Dense Dropout_2 provide output dimensions of 32. The final layer is called Dense_1, and it has a shape of 1 as its output. One of the most important requirements is that the final layer must have a single output shape. This particular model contains a total of 12,2689 parameters in its entirety. After beginning with 128 layers, this model was reduced to just one layer by the end.

Training Vs validation loss

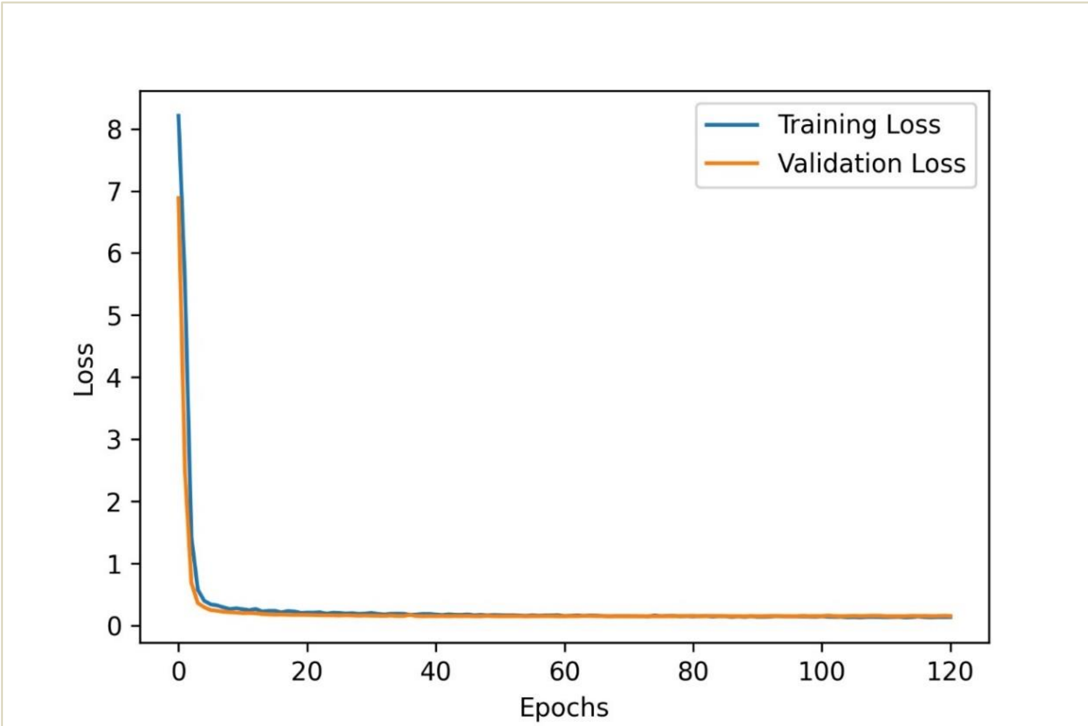

Figure 4.8: Model 3 Training vs validation loss

Figure 5 refers to the training vs validation loss graph. The blue line indicates the training loss and the orange color indicates to validation loss. This figure observation seems to show there are no any single errors at any epochs. No overfeeding is seen here. There is no option to find out the deferent between the two lines. The two lines overlap each other, which means there is no error and how much fits this project dataset. It shows a R_squre score of 77%. That is the best score of other models.

4.3 Discussion

The results of the various analyses shown above indicate that the LSTM model 3 outperforms the others, with the best Rescuer score coming in at 77%. There is a significant amount of loss in every other model, which is discovered to be the validation loss graph. In contrast, the graph of model 3 reveals that it is the one that produces the best results. It can be seen from the graph that there is no loss with only 20% of the total. As I am all aware, the dataset that has the least amount of loss is the one that is deemed to be the best for review. I decided to use this model for my further work. Within the context of this project, the implementation of this LSTM model 3 is the next evaluation task.

4.4 Evaluation

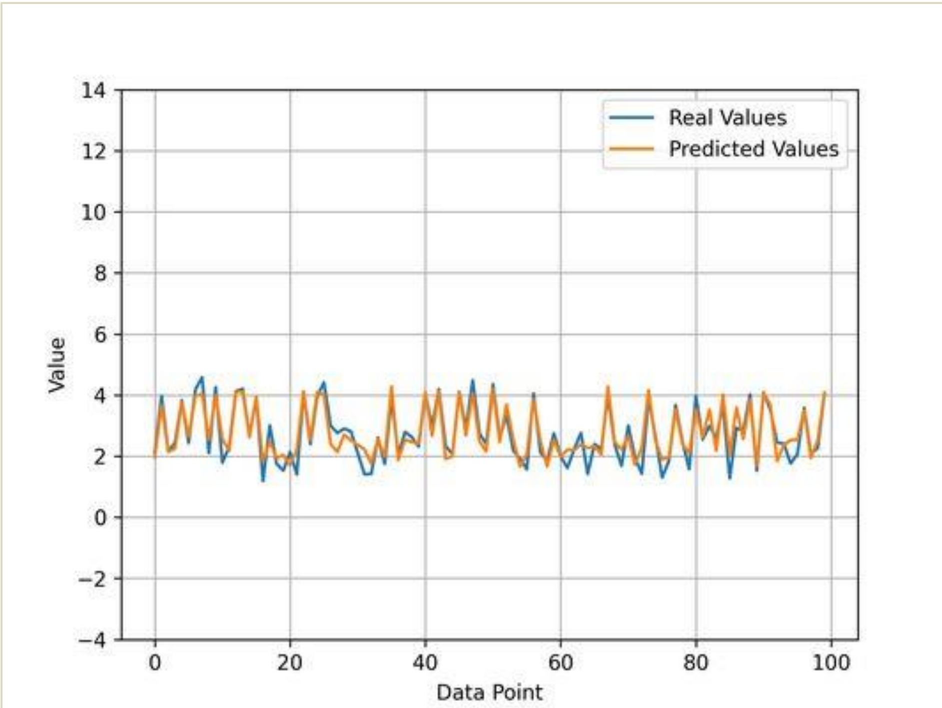

Figure4.9: Evaluation graph for LSTM model 3

This graph illustrates the most important component of the project, which is the degree of accuracy it possesses in comparison to cases that actually occur in the real world. The blue line on the graph represents the actual numbers, while the orange curve represents the expected values. Both sets of values are displayed on the graph. As a result of the formation of a zigzag line at a very short distance, both lines have an upward tendency. At the point where the graph reaches its highest point, the actual value and the expected value are said to coincide. There is not a single substantial inaccuracy that can be found in this graph.

Because of this, it is possible to assert that the Model 3 LSTM offers great predictive outcomes.

CHAPTER 5

IMPACT ON SOCIETY ENVIRONMENT AND SUSTAINABLY

5.1 Impact on Society

These impacts will show in society,

- Helping farmers improve both production and distribution with accurate crop output estimates will boost food security.
- This prevents food scarcity and price fluctuation by ensuring an even supply.
- Helping farmers improve both production and distribution with accurate crop output estimates will boost food security.
- This prevents food scarcity and price fluctuation by ensuring an even supply.
- Deep learning algorithms can detect pests, diseases, and weather hazards.
- Farmers can avert crop failures and financial losses by taking precautions.

5.2 Impact on Environment

- Deep learning models optimize water, fertilizer, and pesticide use for sustainable agriculture.
- This reduces resource use and runoff, making farming more effective and environmentally benign.
- With accurate production projections, precision agriculture may target actions to reduce environmental impact.
- This reduces agricultural impacts on nearby ecosystems, preserving biodiversity.

5.3 Impact on Sustainability

- Deep learning algorithms can use climate data to help farmers adjust to shifting weather.
- For agriculture to survive climate change, adaptation is essential.
- Accurate yield projections improve harvesting and storage plans, lowering post-harvest losses.
- This reduces food waste and optimizes supply chain resource use, promoting sustainability.

- Deep learning models can promote regenerative agriculture by revealing optimal farming practices.
- These practices promote a holistic and resilient agricultural system by improving soil health, biodiversity, and sustainability.

Chapter 6

CONCLUSION, RECOMMENDATION, AND FUTURE IMPLEMENTATION

6.1 Conclusion

When it comes to food production, Bangladesh can mostly rely on its own rice and potato crops. Half of Bangladesh's agricultural GDP and one-sixth of the country's government revenue come from rice. A total of 27,26,000,000 metric tons of rice will be needed by Bangladesh in the year 2022. A total of 10.28 million hectares would remain after rice production was reduced. The present rice production of 2.74 t/ha needs to be increased to 3.74 t/ha. Based on our research, we can predict the potato and rice yields (Aus, Aman, and Boro) using a range of climatic data from different regions. In terms of predicting future rice yields, our results show that LSTM is the best of the four methods. With less data rate, LSTM got an R-squared score of 0.77%. Researchers hope that farmers will use this study's results to improve their weather forecasting and, in turn, lessen the possibility of future financial losses. This will have a more potent impact on our ability to think creatively. A rise in production can fix any problems. Using machine learning to forecast outcomes can boost yield. Machine learning is helping more and more farmers. Our capacity to enhance land productivity is critical to the success of this endeavor. In our suggested model, I included seven popular Machine Learning regression techniques.

6.2 Limitation

In spite of the fact that I exerted all possible effort to ascertain the best possible outcome, I ran into a few obstacles along the route. There was a limitation in the form of a lack of access to data, particularly data pertaining to soil. The majority of online government websites placed restrictions that made it difficult to obtain additional data, which made the task of gathering this information a difficult one. In the event that I was able to acquire additional facts. There are several reasons why having a larger quantity of data would be beneficial.

It can be said that; data consistency is very important limitation for this project.

6.3 Recommendation

- In order for users to fulfill the needs of the model and algorithm, they are required to make modifications to the project's visual and functional elements. Taking into consideration the project, these alterations are required to be implemented.
- Every project is required to have an ethical rationale, which enables a transparent explanation of the decision-making process and a fundamental description of the algorithm that was utilized.
- Due to the fact that this is an essential aspect to take into consideration, persons who are very proactive in the field of harvesting work should be given priority.
- When it comes to successfully completing projects, working together with others and conducting research are both essential.
- Through the application of my concept, I intend to achieve my objective of enhancing the existing agricultural system by having the ability to precisely estimate the rate of yield.
- For the purpose of ensuring that it is accessible to farmers and field workers all over the world, the project must to be built with a global perspective.

6.4 Future work

It is my intention to undertake such expansion stages in the future.

- In this experiment, three distinct types of rice were utilized. Wheat, oil seeds, maize, legumes, and other types of crops are going to be utilized in my future work.
- I have the intention of enhancing the process of collecting project data primarily in the future by gathering information from numerous years.
- Creating an Android application that is user-friendly and accessible to a large number of people.
- In addition to that, I will also put together a web-based system that will offer recommendations.
- In future I will apply more models for get more good result.
- Next this project integrated AI tools for make it more useful.

APPENDIX

Appendix A: Agricultural Website here the source of data.

The challenging performance was deciding the methodological methods for our report. Though it wasn't as expected work, and nothing had been accomplished in this area of analysis previously. None of the source was as supportive as I need. But doing hard as giving effort, I might be able to do it.

Heres a site that is mentionable. (http://www.barc.gov.bd/)

References

1. Ali, "Save Boro first, then Aush," The Daily Star, Apr. 16, 2020. [Online]. Available: https://www.thedailystar.net/frontpage/news/save-boro-first-then-aush-1893349
2. Ricepedia, "Recent developments in the rice sector," Ricepedia, Sep. 2012. [Online]. Available: http://ricepedia.org/bangladesh
3. S. Ganguly, R. R. Nemani, and R. B. Myneni, "A Deep Learning-Based Model for Rice Yield Prediction using Satellite Imagery," Remote Sensing, vol. 11, no. 19, pp. 2270, 2019.
4. F. Ruggiero, D. Ventrella, S. Pascucci, and A. Matese, "A Deep Learning Model for Yield Prediction in Precision Agriculture," Sensors, vol. 20, no. 18, pp. 5255, 2020.
5. Z. Li, Q. Sun, L. Zhang, S. Feng, and J. Liu, "Deep Learning-Based Crop Yield Prediction Using Remote Sensing Data," IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no. 9, pp. 6299-6311, 2020.
6. S. Jang, J. Kim, T. Kim, H. Lee, and S. Ko, "A Wafer Map Yield Prediction Based on Machine Learning for Productivity Enhancement," in IEEE Transactions on Semiconductor Manufacturing, vol. 32, no. 4, pp. 400-407, Nov. 2019.
7. N. Gandhi, L. J. Armstrong, O. Petkar, and A. K. Tripathy, "Rice crop yield prediction in India using support vector machines," in 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), Khon Kaen, Thailand, 2016, pp. 1-5.
8. R. Kumar, M. P. Singh, P. Kumar, and J. P. Singh, "Crop Selection Method to maximize crop yield rate using machine learning technique," in 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), Avadi, India, 2015, pp. 138-145.
9. N. Savla, P. Israni, A. Dhawan, A. Mandholia, H. Bhadada, and S. Bhardwaj, "Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture," in 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, India, 2015, pp. 1-7.
10. Y. Gandge and Sandhya, "A study on various data mining techniques for crop yield prediction," in 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), Mysuru, India, 2017, pp. 420-423.
11. M. Paul, S. K. Vishwakarma, and A. Verma, "Analysis of Soil Behaviour and Prediction of Crop Yield Using Data Mining Approach," in 2015 International Conference on Computational Intelligence and Communication Networks (CICN), Jabalpur, India, 2015, pp. 766-771.
12. M. M. Rahman, N. Haq, and R. M. Rahman, "Machine Learning Facilitated Rice Prediction in Bangladesh," in 2014 Annual Global Online Conference on Information and Computer Technology, Louisville, KY, USA, 2014, pp. 1-4.

13. S. Bhanumathi, M. Vineeth, and N. Rohit, "Crop Yield Prediction and Efficient use of Fertilizers," in 2019 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2019, pp. 0769-0773.
14. R. Jaikla, S. Auephanwiriyakul, and A. Jintrawet, "Rice yield prediction using a Support Vector Regression method," in 2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, Krabi, Thailand, 2008, pp. 29-32.
15. M. M. Hasan, M. T. Zahara, M. M. Sykot, A. U. Nur, M. Saifuzzaman, and R. Hafiz, "Ascertaining the Fluctuation of Rice Price in Bangladesh Using Machine Learning Approach," in 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, pp. 1-5.
16. N. Madhu et al., "A Deep Learning Approach for Estimating Paddy Yield using Weather and Remote Sensing Data," 2021.

PLAGIARISM

Yield Prediction

ORIGINALITY REPORT

12%	11%	2%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	9%
2	Submitted to University of Hertfordshire Student Paper	<1%
3	Zihad Hossain Nayem, Iqbal Jahan, Abdul Aziz Rakib, Solaiman Mia. "Detection and Identification of Rice Pests Using Memory Efficient Convolutional Neural Network", 2023 International Conference on Computer, Electrical & Communication Engineering (ICCECE), 2023 Publication	<1%
4	amslaurea.unibo.it Internet Source	<1%
5	www.tandfonline.com Internet Source	<1%
6	www.mdpi.com Internet Source	<1%
7	Md. Mobinur Rahman, Umakanta Sarker, Md Ahsanul Haque Swapan, Mohammad Sharif	<1%