Crop yield prediction and Impact of Weather on Crops in Bangladesh

Submitted by

Md. Fazle Rabbee

191-35-2724

Department of Software Engineering

Daffodil International University

Supervised by

Md. Shohel Arman

Assistant Professor

Department of Software Engineering

Daffodil International University

This Thesis report has been submitted in fulfillment of the requirements for the Degree of Bachelor of Science in Software Engineering

Fall-2022

Approval

This thesis titled on "**Crop yield prediction and Impact of Weather on Crops in Bangladesh**", submitted by **Md. Fazle Rabbee (ID: 191-35-2724)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

---------- **Chairman**

Dr. Imran Mahmud
Head and Associate Professor

Department of Software Engineering

Faculty of Science and Information Technology
Daffodil International University

---------- **Internal Examiner 1**

Md. Khaled sohel

Assistant Professor

Department of Software Engineering

Faculty of Science and Information Technology

Daffodil International University

---------- **Internal Examiner 2**

Md. Shohel Arman
Assistant Professor

Department of Software Engineering

Faculty of Science and Information Technology

Daffodil International University

---------- **External Examiner**

Rimaz Khan
Managing Director
Tecognize Solution Limited

THESIS DECLARATION

I announce hereby that I am rendering this study document under Md. Shohel Arman, Assistant Professor, Department of Software Engineering, Daffodil International University. I therefore state that this work or any portion of it was not proposed here therefore for Bachelor's degree or any graduation.

Supervised by

13.12.22

Md. Shohel Arman

Assistant Professor

Department of Software Engineering

Daffodil International University

Submitted by

Fazle Rabbee

Md. Fazle Rabbee

191-35-2724

Department of Software Engineering

Daffodil International University

ACKNOWLEDGEMENT

As a first step, I would like to extend my deepest gratitude and thanks to Almighty God for giving me the opportunity to complete my final year thesis successfully.

It is my privilege and honor to acknowledge and express our profound gratitude to Md Shohel Arman, Assistant Professor, Department of Software Engineering, Daffodil International University, Dhaka. To carry out this project, our supervisor has deep knowledge and a keen interest in the field of "Machine Learning". I have been able to complete this project because of his endless patience, scholarly guidance, constant encouragement, constant and energetic supervision, constructive criticism, valuable advice, and reading many inferior drafts and correcting them at every stage. Daffodil International University's Software Engineering Department and Dr. Imran Mahmud, Head of the Faculty of Software Engineering, deserve my heartfelt gratitude for their kind assistance in completing my thesis.

Last but not least, I would like to acknowledge the support and patience of our parents who have always been there for us.

ABSTRACT

Agriculture is the most important industry that influence the economy in Bangladesh. After freedom, the farming area was Bangladesh's really monetary main impetus. As Bangladesh is an agricultural country, the economy as well as food security of this country mostly depends on production level of different crops over the year. Crop yields are basically reliant on climate. The weather has a significant influence on crop output.AI can blast the farming field by changing the pay situation by developing the ideal harvest. This paper centers around anticipating the yield of the harvest by applying different machine-learning strategies. The expectation made by AI calculations will predict to the government and farmers how much crop yield in next year, by considering factors like temperature, area, rainfall, humidity and so on. Using data on crop yield from the 'Bangladesh Agricultural Research Institute'. The most popular features, based on my data are area, rainfall, humidity, temperature and use of the algorithm is Arima, Long-Short-term Memory (LSTM), and Exponential Smoothing for time series forecasting. Also used classifier model for prediction production based on area, rainfall, humidity and temperature where used linear model and those are Linear Regression, Lasso Regression, Random Forest Regression and Decision Tree. From those algorithm Decision Tree provides maximum accuracy.

CONTENTS

LIST OF FIGURES

CHAPTER 1

1. INTRODUCTION

1.1 BACKGROUND

Bangladesh is an agricultural country. This nation's economy continues to be heavily reliant on the annual output of various crops. However, the creation of yields diminishes because of some unusual negative ways of climate behavior. Environment inconstancy and weaknesses are among the developing worries around the world. The environment of Bangladesh is changing and it is turning out to be more eccentric and consistently. The total populace could reach around 10.6 billion individuals by 2050. From this normal population development, food requests could turn out to be twofold for present-day's utilization. My thesis paper shows the effect of weather conditions on crops for the situation in Bangladesh. The nature of the climate in this classified is named (1) Area (2) Temperature, (3) Windspeed (4) Humidity (5) Rainfall by the meaning of officials. As a matter of fact, the distinction between climate and the environment is a proportion of the time. Contrary to weather, which describes the state of the atmosphere during a brief period of time, climate refers to how the atmosphere "behaves" over similarly long periods of time. I take data from the Bangladesh agriculture institute. The vast majority really take a look at the neighborhood weather conditions gauge to design their days. Furthermore, the adjustment of the environment is really a "hot" subject in the news. a few outrageous climate catastrophes have to some extent or totally harmed local yield creation. A few researchers frame the environment in light of the fact that the normal climate for a chosen locale and period here and there censured 30 years. It's an extremely middle example of climate for a chosen district. At the point when Researchers state environment they were watching midpoints of the climate that happens over a drawn-out sum in an exceptionally unequivocal spot. As a matter of fact, the contrast

between climate and environment is a proportion of the time. Weather conditions state of the environment is over a brief timeframe, and environment is the way the climate "acts" over nearly significant stretches of time. Lately, a few outrageous climate catastrophes have somewhat or totally harmed territorial yield creation. As delayed local records of the obvious impacts of harmful climate fiascoes are still in present, and the worldwide scale results of dry spells, floods, and the intense temperature on crop development and creation are still to be evaluated. Here creators evaluated, to their most extreme information, public cereal assembling misfortunes across the globe shown by detailed intense climate fiascos during 1971-2019. A few specialists likewise show that dry seasons and intense heat poisonously condensed public cereal creation by 9-10%, while their examination could not show an impact from floods and outrageous virus in the public information. Examining the natural cycles, they later tracked down that, creation misfortunes because of dry seasons were related to depreciation in both developed region what's more, yields, while intense intensity principally deteriorated oat yields. Besides, the outcomes feature ~7% bigger creation deterioration from more ongoing dry seasons and 8-11% more in created nations than in creating ones. The principal necessities of the proposed work to contrast the adjustment of climate and the creation level of yields, and get the proper result by utilizing different strategies. AI, which is a part of Computerized reasoning (simulated intelligence) zeroing in on learning is a reasonable methodology that can give a better yield forecast in view of a few highlights. AI (ML) can decide examples and relationships and find information from datasets. The models should be prepared to utilize datasets, where the results are addressed in light of previous experience. The prescient model is fabricated utilizing a few elements, and in that capacity, the boundaries of the models are resolved to utilize verifiable information during the preparation stage. For the testing stage, part of the verifiable information that has not been utilized for preparation is utilized for execution assessment purposes. An ML model can be spellbinding or prescient, contingent upon the research issue and exploration questions. While elucidating models are utilized to

acquire information from the gathered information and make sense of what has occurred, prescient models are utilized to make expectations about what's to come. ML concentrates on comprise of various difficulties when meaning to construct an elite exhibition, prescient model. It is vital to select the right calculations to take care of the front and center issue, and likewise, the calculations and the fundamental stages should be fit for dealing with the volume of information. To get an outline of what has been finished on the utilization of ML in crop yield forecast, we played out a deliberate writing survey'.

1.2 MOTIVATION OF THE RESEARCH

From antiquated days, farming is thought of as the principal cause of supply to fulfill the day to day needs of human lives. The harvests that we develop for food need explicit climatic circumstances to show better execution considering financial yield. A changing environment could significantly affect crops. Remembering the above view, this study researches the effects of environmental change (changes in temperature, rainfall, wind speed, and humidity) on the yield and trimming area of four significant food crops in Bangladesh. The impacts of environmental change on crop creation are global worries, yet they are especially huge for the reasonable agrarian advancement of Bangladesh. This is a nation of variation climatic circumstances all year because of its geographic position and physiographic status. Farming is generally helpless to troublesome atmospheric conditions and environment occasions. Regardless of innovative advancement, climate and environment are as yet key determinants for rural efficiency and manageability. Food security is characterized as admittance to enough and safe food by all individuals consistently for keeping a functioning and solid life. Bangladesh is prevalently an agrarian country with a high populace thickness, where food security is a pivotal issue. Notwithstanding, total homegrown creation and per capita accessibility of food grains have expanded in the country over the course of the last many years.

Environmental change is a possible danger towards accomplishing the previously mentioned objective. It is accordingly important to understand the impacts of environmental change on the creation of significant food crops under manageable ecological circumstances. Assessment of climatology at the public level is generally significant for the cure of rural issues emerging from environmental change. Climatic data not just suggests the most reasonable time for planting and gathering yet in addition goes about as a manual for the determination of the legitimate destinations for a specific yield. In spite of having an extremely low commitment to producing all-out worldwide ozone-harming substance discharges, Bangladesh is presently considered the weak country in the globe because of an unnatural weather change. Environment factors including temperature, precipitation, length of day, moistness adversely affect horticulture. Environmental change is quite possibly the best test confronting humankind, and as AI (ML) specialist, may consider how I can help. Worldwide horticultural creation, specifically, is of expanding worry to the significant global associations responsible for nourishment. Farmers have not enough experience for specific yield, crops and climatic conditions. Climate change has become a major issue affecting the agriculture sector. This work may incredibly help the penniless ranchers who have less information in anticipating the yields for fostering a practical future. In future it might likewise reach out to propose the manure, reasonable rules for cropland and crops for the given info. Furthermore, source of daylight and yield wellbeing are observed at ordinary spans and it is likewise considered for accomplishing a superior harvest yield. Want To create an accurate and efficient model for crop yield estimation based on the weather, Temperature, Humidity, Rainfall, Wind Speed. In other country researcher works with their countries data. But I research from Bangladesh Agricultural Research Institute data set.

1.3 PROBLEM STATEMENT

At the point when the makers of the yields know the exact data on the harvest yield, it limits the misfortune. Machine learning is a quickly developing methodology that is fanning out and helping each area in pursuing feasible choices to make the principal of its applications. Most gadgets these days are worked with by models being broken down before organization. The principal idea is to build the throughput of the farming area with AI models. Another variable that additionally influences the expectation is how much information is being given during the preparation period, as the number of boundaries were higher similarly. The center accentuation would be on accuracy farming, where quality is guaranteed over bothersome ecological elements.

1.4 RESEARCH QUESTIONS

- Q1: Which machine learning algorithms have been used in the literature for crop yield prediction?
- Q2: Which machine learning model detects perfectly from those given multiple features?
- Q3: What are challenges in the field of crop yield prediction using machine learning?

1.5 RESEARCH OBJECTIVE

To increase agricultural crop production is the main goal of crop yield estimation.

It is easy to estimate how much crops can be produced next year.

Early detection and management of problems associated with crop yield indicators can help increases yield and economic growth.

By influencing regional weather patterns, large scale meteorological phenomena can have a significant impact on agricultural production.

1.6 RESEARCH SCOPE

Farmers have not enough experience for specific yield, crops and climatic conditions. Climate change has become a major issue affecting the agriculture sector. In other country researcher works with their countries data. But I will research from Bangladesh Agricultural Research Institute data set. To create an accurate and efficient model for crop yield estimation based on the weather, Sunshine, Humidity, Rainfall, Wind Speed.

CHAPTER 2

2.1. LITERATURE REVIEW

In its earliest, many analyses and models were being done for estimating the negative and positive effect of climate on various types of crops, in Bangladesh and different nations too. Furthermore, they made a lot of algorithms for getting precise result. The worldwide interest in agriculture yields is expected to generally twofold by 2050, driven by expansions in the populace. Obviously, these new developments in worldwide yield produce lingers behind the ordinary requests, leaving us with a critical inquiry: which crops furthermore, which geographic spots subsume the best expectation of satisfying anticipated needs, and where improvements are desperately required?

Aruvansh Nigam, Saksham Garg, Archit Agrawal, Parul Agrawal [1] investigates Indian government dataset and it's been laid out that Arbitrary Timberland AI the calculation gives the best yield expectation precision. Random Forest Regressor gives the highest yield prediction accuracy. Simple Recurrent Neural Network performs better on rainfall prediction LSTM is good for temperature prediction.

It can be improved by extends parameter to predict the more accurate result.

E Setiawati and W A Yusuf [2] is gaining practical experience in the exactness and strength and relationship of irregular woods calculations. Arbitrary timberland calculation makes choice trees on various information tests and afterward predicts the information from every subset and afterward by casting a ballot offers better the response for the framework. Irregular Backwoods utilized the sacking strategy to prepare the information. To support the precision, the arbitrariness infused needs to limit the connection ρ while keeping up with strength.

Md. Toukir Ahmed1 *, Md. Niaz Imtiaz1 , and Nurun Sakiba Mitu1[3] have carried out crop yield expectation by utilizing simply the arbitrary backwoods classifier. Different elements like precipitation, temperature, and season were considered to anticipate the harvest yield. Other AI calculations were not applied to the datasets. With the shortfall of other calculations, correlation and evaluation were missing and subsequently incapable to give the able calculation.

Anakha Venugopal, Aparna S, Jinsu Mani, Rima Mathew, Prof. Vinu Williams[4] has hypothetically portrayed different machine learning strategies that can be applied in different gauging regions. In any case, their work neglects to execute any calculations what's more, consequently can't give a reasonable understanding of the common sense of the proposed work.

Mohsen Shahhosseini', Guiping Hu" and Sotirios V. Archontoulis2 [5] have finished up AI calculations and can foresee an objective/result by utilizing Managed Learning. This paper centers around managed learning procedures for crop yield expectations. To get the predefined yields it requirements to produce a proper capability by a bunch of certain factors which can plan the info variable to the point yield. The paper conveys that the forecasts can be done by Irregular Woods ML calculation which accomplishes the harvest forecast with the best precise worth by thinking about the least number of models

Mohsen Shahhosseini1 , Guiping Hu1*, Isaiah Huber2 & SotiriosV.Archontoulis2 [6] They demonstrated improvements in yield prediction accuracy across all designed ML models when additional inputs from a simulation cropping systemodel.It could concluded that inclusion of additional soil water related variables.

2.2 CONCLUSION

Many different sorts of algorithms have been used. They applied with hopes of getting a good results LSTM, ARIMA, random forest algorithm, and a lot more strategies. Disregard better exactness and ongoing execution. We have applied ARIMA, LSTM, Exponential Smoothing, and ML models for our work on a dataset collected from Bangladesh Agricultural Institute.

CHAPTER 3

3. RESEARCH METHODOLOGY

3.1 RESEARCH METHODOLOGY

I applied ARIMA, LSTM, Exponential Smoothing and ML model for our work on a dataset collected from Bangladesh Agricultural Institute.

3.2 DATA COLLECTION

Data is collected from Bangladesh Agricultural Institute. The agricultural institute supply data for a maximum of 50 years in a row. For crop production prediction and higher crop yields, data sets on climate characteristics, temperature, rainfall, humidity and seed characteristics are used and those are area and production.

Figure 1: Climate characteristics collect from Climate Information System

3.3 DATA PREPROCESSING

I have collected data from Bangladesh Agricultural Institute. That dataset has about 50 years agricultural and weather-related data. It is gathered in raw format, which makes analysis impractical. I can convert data into a comprehensible format by using several strategies, such as substituting missing values and null values. The dataset also contains date of 50 years, area, rainfall, humidity, temperature and production. For time series forecasting I divided dataset and made 5 dataset which was include date and each feature those are area, rainfall, humidity, temperature and production. In Machine learning Model I used all relevant feature to predict crop production. I separated the data into sets for training and validation. Data comprise 80% of the training dataset and 20% of the test dataset.

Year	Total Area	Temperature(avg C)	Wind_speed(m/s)	Humidity(g/kg)	Rainfall(mm)	Production
1981-01-01	10457.60	25.21	2.05	14.34	1260.35	13631.00
1982-01-01	10583.90	25.37	2.05	13.55	922.85	14129.00
1983-01-01	10546.60	25.38	2.39	12.63	606.45	14415.00
1984-01-01	10222.20	25.41	2.34	13.24	1028.32	14622.00
1985-01-01	10397.00	25.69	2.33	12.70	580.08	15041.00
1986-01-01	10607.70	25.57	2.19	13.00	991.41	15407.00
1987-01-01	10321.30	25.90	2.06	13.55	812.11	15414.00
1988-01-01	10222.56	25.95	2.20	13.43	675.00	15544.00
1989-01-01	10411.10	25.60	2.38	11.90	411.33	17710.00
1990-01-01	10430.50	25.58	2.23	13.43	717.19	17785.00
1991-01-01	10243.10	25.25	2.27	13.00	711.91	18255.00
1992-01-01	10177.70	25.09	2.18	11.78	363.87	18341.00
1993-01-01	10073.50	25.08	2.23	12.94	479.88	18041.60
1994-01-01	9921.46	25.48	2.28	12.08	321.68	16833.40
1995-01-01	9941.82	25.75	2.30	12.76	601.17	17686.60
1996-01-01	10177.37	25.65	2.15	12.45	490.43	18881.00

Fig_ 2: Dataset

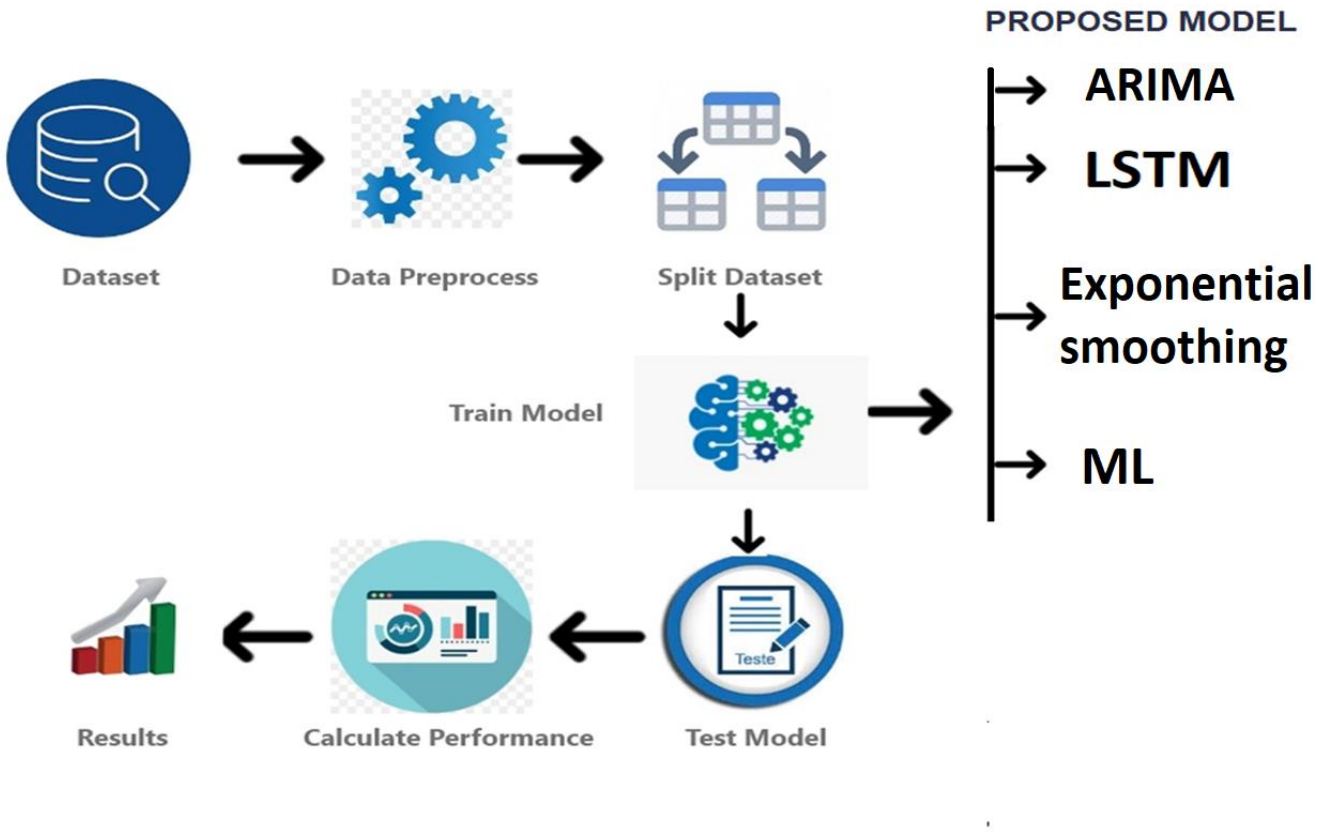


Figure 3: Dataset Labeling

3.4 Time Series

A "time series analysis" is a specific approach to analyzing a set of data points accumulated over an extended period of time. Time series analyzers record the data points over a set period of time at regular intervals rather than randomly or irregularly. However, this type of study entails more than merely collecting data over time.

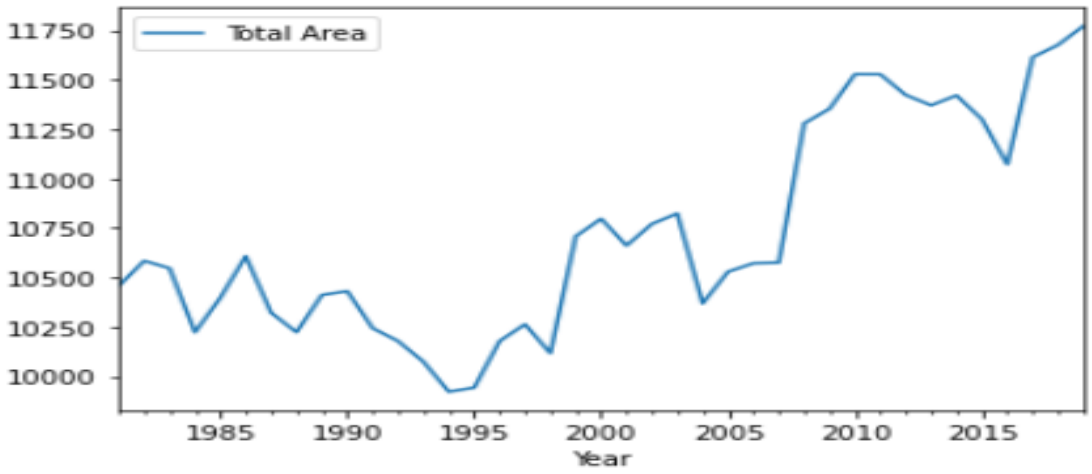


Figure 4: Time Series Plot

3.4.1 Autoregressive Integrated Moving Average (ARIMA)

The Autoregressive Integrated Moving Typical model is curtailed as ARIMA. In time-series information, a sort of model might get different normal fleeting events. ARIMA models are real models that are utilized to break down furthermore, sort out time-series information. In the model, every one of these components is obviously expressed as a line. ARIMA (p, d, q) is a kind of standard documentation in which the lines are supplanted by mathematical characteristics to perceive the ARIMA technique.

$$\Delta Yt = \varphi 0 + \varphi 1 \times \Delta yt{-}1 \ldots + \varphi m \times \Delta yt{-}m + \sigma 0 + \sigma 1 \times \Delta \alpha t{-}1 + \cdots + \sigma k \times \Delta \alpha\, t{-}\mathrm{k}$$

.

3.4.2 Adfuller Tests to check if a series is stationary or not

Augmented Dickey-Fuller test is utilized to gives us different qualities that can help in distinguishing stationarity. The null hypothesis says that a TS is non-Stationary. It includes a Test Measurements and a few basic qualities for some certainty levels. Assuming the Test insights is not exactly the basic qualities, we can dismiss the invalid speculation and say that the series is fixed. THE ADCF test likewise gives us a p-value. Acc to the invalid speculation, lower upsides of p are better.

```
ADF Test Statistics : -0.6338529361929546
p-value : 0.8631477310866358
#Lags Used : 0
Number of Observation Used : 38
```

3.4.3 Differencing

Differencing is a method of transforming a non-stationary time series into a stationary one. This is an important step in preparing data to be used in an ARIMA model.

The reason for differencing is to make the time series stationary. In any case, we ought to be mindful so as to not over-difference the series. An over differenced series might in any case be stationary, which thus will influence the model parameters.

The Formula is $Y_{dt}=Y_t-Y_{t-1}$

3.4.5 Autocorrelation Function (ACF)

Autocorrelation is the relationship between two values in a time series. To put it another way, the time series data are correlated, hence the word. “Lags” are the term for these kinds of connections. When a characteristic is measured on a regular basis, such as daily, monthly, or yearly, time-series data is created. In an ACF plot, each bar represents the size and direction of the connection. Bars that cross the red line are statistically significant.

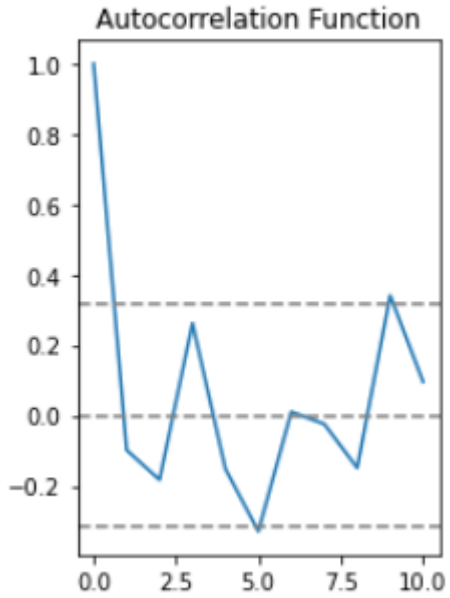


Figure 5: Autocorrelation Function

3.4.6 Partial Autocorrelation Function (PACF)

Similar to the ACF, the partial autocorrelation function merely identifies the correlation between two data sets that the shorter lags between those observations do not account for. Both regression models with time series data and Auto-Regressive Integrated Moving Average (ARIMA) models can be specified using partial autocorrelation graphs.

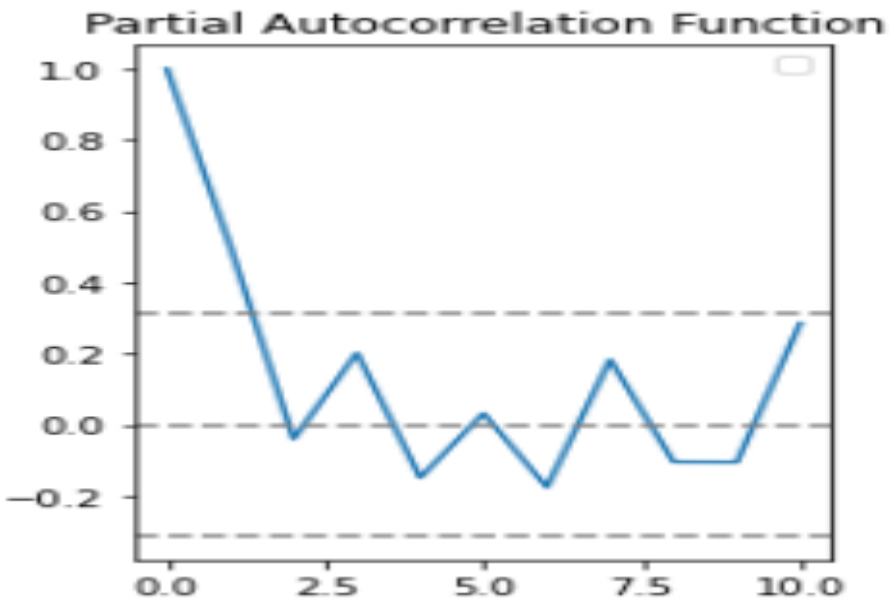


Figure 5: Partial Autocorrelation Function

3.4.7 LSTM

LSTM is a symbol for short-term memory. The memory of recurrent neural networks is increased by a model or design. Recurrent neural networks frequently have "short term memory" in that they continuously use previous data to inform the continuing neural network. In essence, the current task uses the historical data. That suggests that we do not currently have a list of all the historical data that is available for the neural node. Intermittent neural networks benefit from long-term memory thanks to LSTM. It reduces the gradient slope problem, which occurs when a neural network stop learning because the updates to the various loads inside a particular neural network become progressively less substantial. It accomplishes this by employing a series of "gates." These are included.

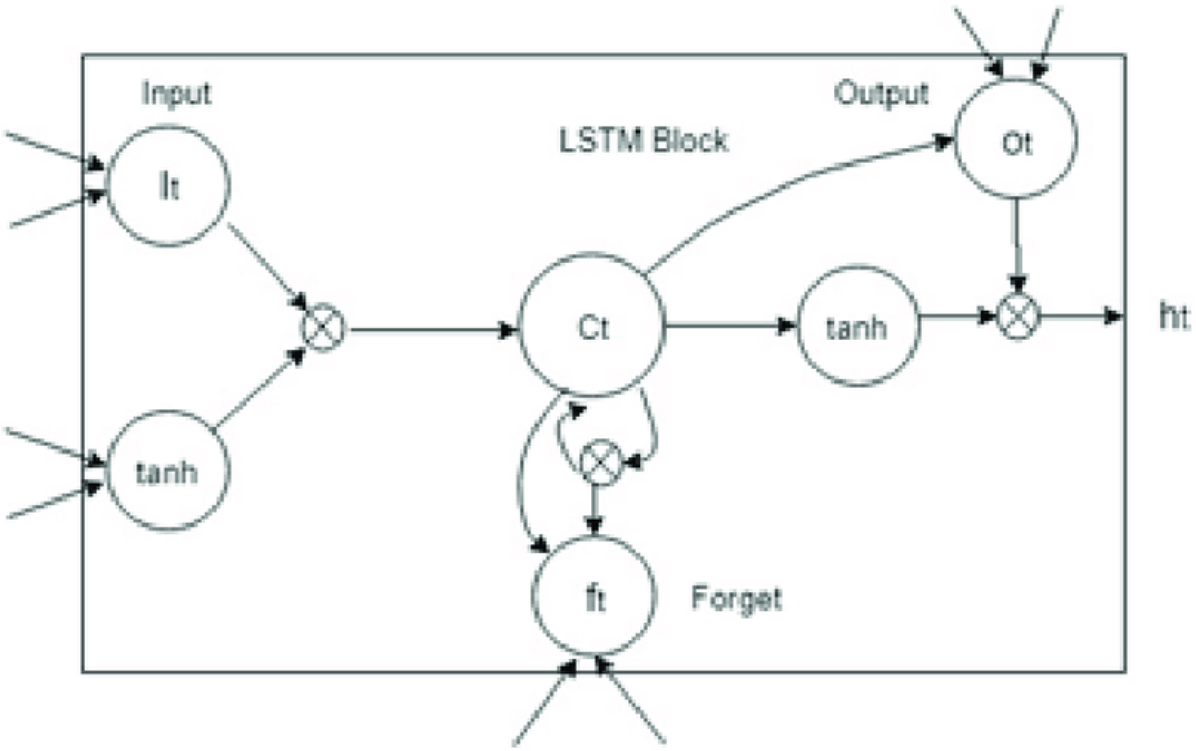


Figure 6: LSTM workflow diagram

LSTM work There are three types of gates inside a unit: Input Gate: Scales contribution to cell (write) Output Gate: Scales result to cell (read) Forget Gate: Scales old cell esteem (reset) Each gate resembles a switch that controls the read/write, consequently integrating the long-term memory capability into the model. Also, it preserves word request, and it doesn't rely upon other etymological highlights to catch the semantics. Expectation is successively in RNN. This will assign a memory to the network. Results from past expectations can work on future forecasts. LSTM gives RNN an additional viewpoint that gives it a fine-grained command over memory.

3.4.8 Decompose

Time series decomposition involves thinking of a series as a combination of level, trend, seasonality, and noise components. Decomposition provides a useful abstract model for thinking about time series generally and for better understanding problems during time series analysis and forecasting.

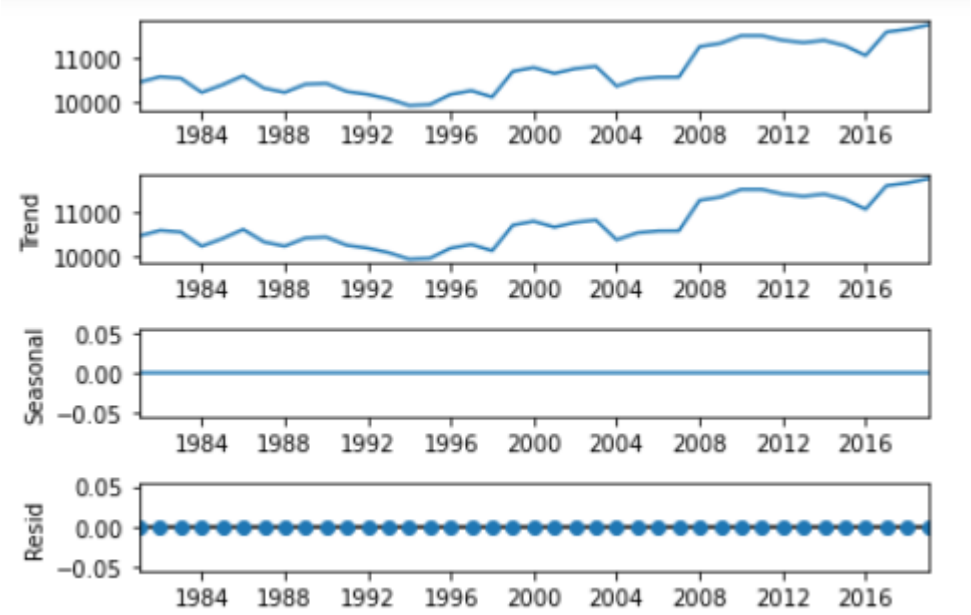


Figure 7: Decompose

3.4.9 Exponential Smoothing

The exponential smoothing function is a general approach for smoothing time series data known as exponential smoothing. Although prior judgments are weighted similarly in the simple moving average, exponential functions are used to assign weights that will drop exponentially with time. It is a usefully informed and easily used strategy for providing some assurance in light of the client's prior assumptions, such as seasonality. Time-series data analysis typically makes use of exponential smoothing.

The crude data succession is frequently addressed by ${x_{t}}$ starting at time $t=0$, and the result of the Exponential smoothing calculation is generally composed as ${s_{t}}$, which might be viewed as a best gauge of what the following worth of x will be. At the point when the grouping of perceptions starts at time $t=0$, the easiest type of Exponential smoothing is given by the equations:$s_{0} = x_{0}$ $s_{t} = α*x_{t} + (1-α)*s_{t-1}$, $t>0$

where $α$ is the smoothing factor, and $0<α<1$.

3.5 ML:

Artificial intelligence (AI) and computer science's machine learning field focuses on using data and algorithms to mimic how people learn, gradually increasing the accuracy of its predictions.

In order to forecast crop production, machine learning models are constructed using a data set that spans from 1971 to 2022. The environment (soil and weather), which includes area, humidity, rainfall, and temperature as input variables, and actual crop production during the research period, which serves as the target variable, make up the data set. Weather and soil information make up the input data.

3.6 ML Model Selection:

Tuning hyperparameters of machine learning models and selecting best models with optimal hyperparameter values is necessary to achieve high prediction accuracies. Cross-validation is commonly used to evaluate the predictive performance of fitted models by dividing the training set to train and validation subsets. Here, we use a random tenfold cross-validation method to tune the hyperparameter of ML models. Grid search is an exhaustive search method that tries all the possible combinations of hyperparameter settings to find the optimal selection. It is both computationally expensive and generally dependent on the initial values

specified by the user. However, Bayesian search addresses both issues and is capable of tuning hyperparameters faster and using a continuous range of values.

3.7 Predictive Model:

Linear regression. Linear regression intends to predict a measurable response using multiple predictors. It assumes the existence of a linear relationship between the predictors and response variable, normality, no multicollinearity, and homoscedasticity.

LASSO regression. LASSO is a regularization method that is equipped with in-built feature selection. It can exclude some variables by setting their coefficient to zero. Specifically, it adds a penalty term to the linear regression loss function, which can shrink coefficients towards zero (L1 regularization)

Random forest. Random forest is built on the concept of bagging, which is another tree-based ensemble model. Bagging tries to reduce prediction variance by averaging predictions made by sampling with replacement. Random forest adds a new feature to bagging, which is randomly choosing a random number of features and constructing a tree with them and repeating this procedure many times and eventually averaging all the predictions made by all trees. Therefore, random forest addresses both bias and variance components of the error and is proved to be powerful

Decision Tree Regression: Decision tree builds regression or classification models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. A decision node (e.g., Outlook) has two or more branches (e.g., Sunny, Overcast and Rainy), each representing values for the attribute tested. Leaf node (e.g., Hours Played) represents a decision on the numerical target. The topmost decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categorical and numerical data

Chapter 4

4. RESULTS AND DISCUSSION

4.1 INTRODUCTION

The models and methods of time series forecasting, machine learning prediction is described. Following the gathering and preprocessing of the data, I outlined the models I used for forecasting and prediction. In that section, I'll go over the outcomes of the model I used.

4.2 RESULT DISCUSSION

After implementing the ARIMA, LSTM, Exponential Smoothing With our dataset and models, we have obtained a future graph prediction and charting. Also implementing machine learning algorithm and those are Linear Regression, Decision Tree Regression, Random Forest Regression and Lasso Regression to predict crop production based on climate and area feature. Preparing exactness and approval precision are thought about for acquiring the diagram of all the model's precision.

4.2.1 Experimental Result and Analysis for Forecasting

1) Area Forecasting:

Arima model: Model are created crop area in BD and what's more, time series calculations are prepared by investigating 50 years data. The ARIMA model forecast area on 50 years fitted data.

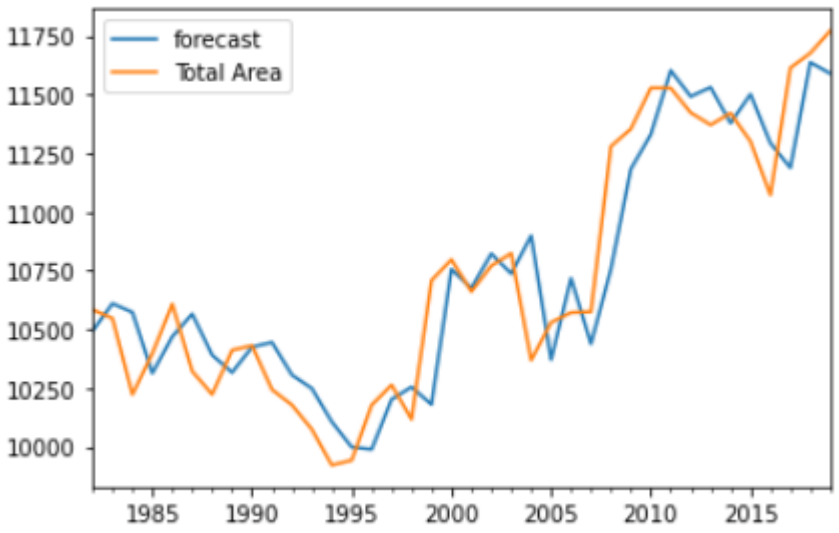


Figure 1: Deviation of anticipated creation with real qualities.

LSTM: This model forecast certain train data from 50 years. Here the model predicts from tarin data before 1994 to 2004.

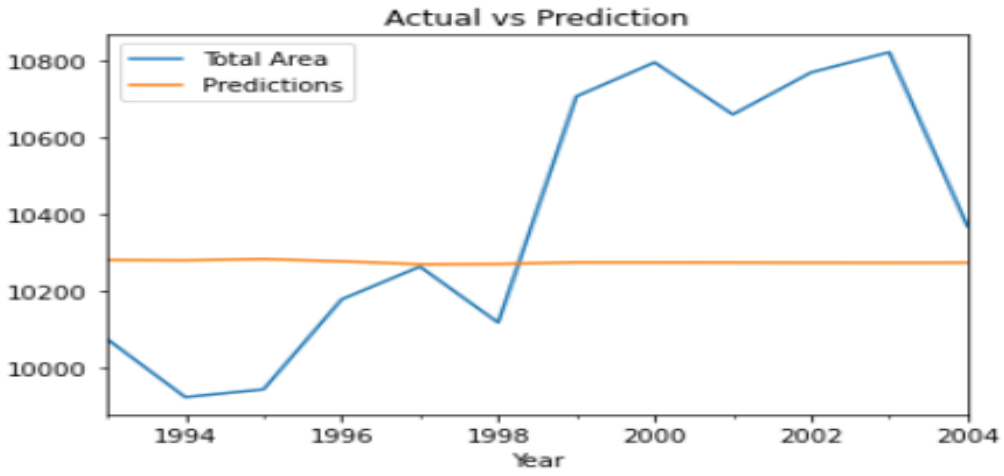


Figure 2: Anticipated creation with real qualities

Exponential Smoothing: This model forecast future in certain period. This model forecasting area upcoming 20 years.

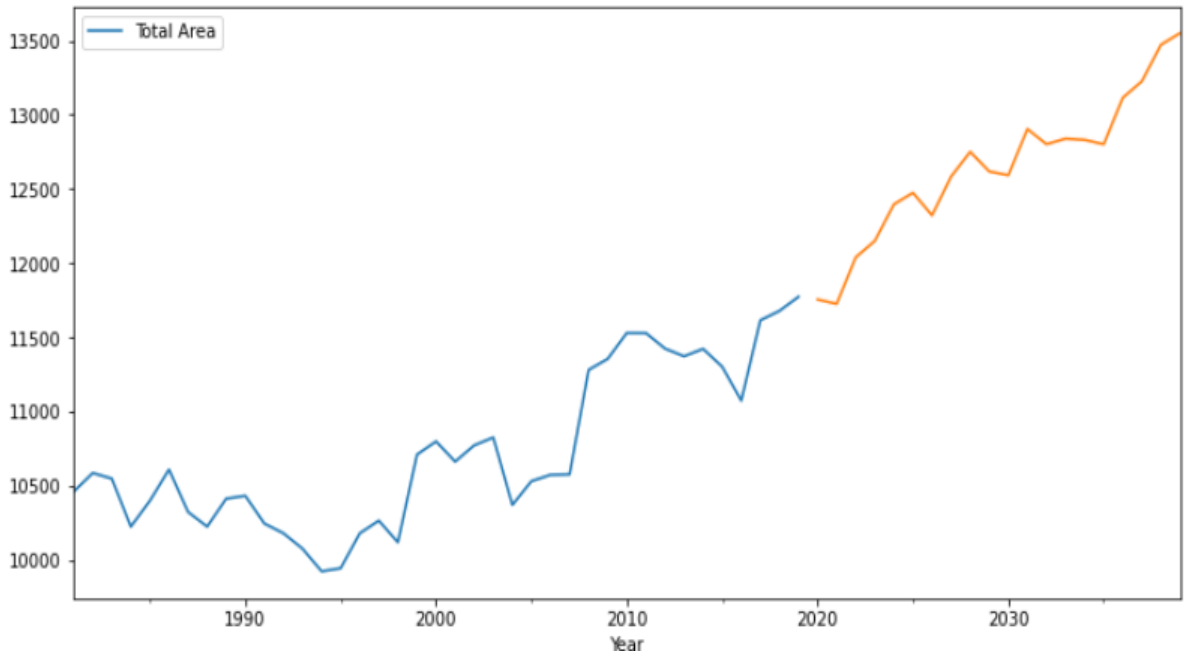


Fig3: Anticipated creation with real qualities

Tab(i). Models Results

Model	Mean Absolute Error
ARIMA	.12
LSTM	.20
Exponential Smoothing	.15

From Fig 1, Fig 2, Fig 3 gives deviation of creation anticipated by ARIMA, LSTM, and Exponential Smoothing individually from the real one. It very well may be seen from Table III and Fig 1 that the forecast given by ARIMA is better nearer to the genuine information as it has a low mean outright mistake than LSTM and Outstanding Smoothing.

2) Temperature Forecasting:

Arima model: Model are created Temperature in BD and what's more, time series calculations are prepared by investigating 50 years data. The ARIMA model forecast area on 50 years fitted data.

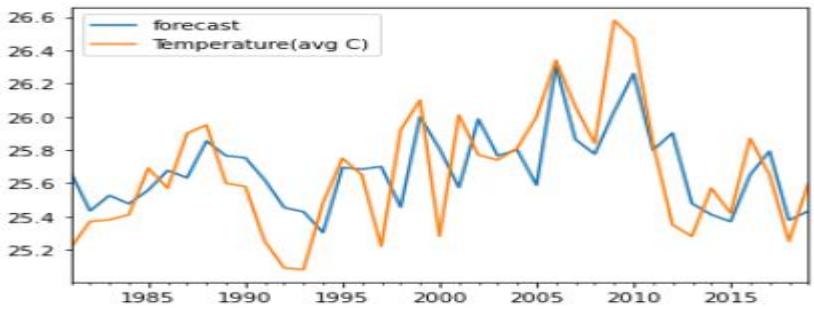


Fig 1: Deviation of anticipated creation with real qualities

LSTM : This model forecast temperature certain train data from 50 years. Here the model predict from tarin data before 1994 to 2004.

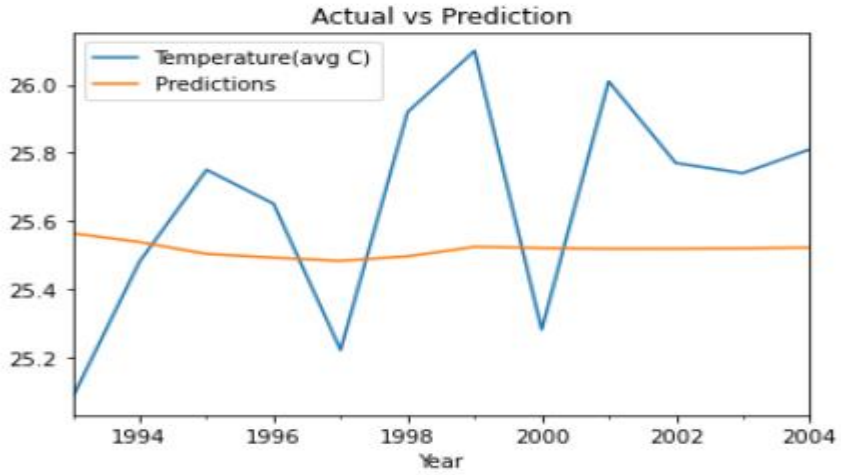


Fig 2: Anticipated creation with real qualities.

Exponential Smoothing: This model forecast future in certain period. This model forecasting area upcoming 20 years.

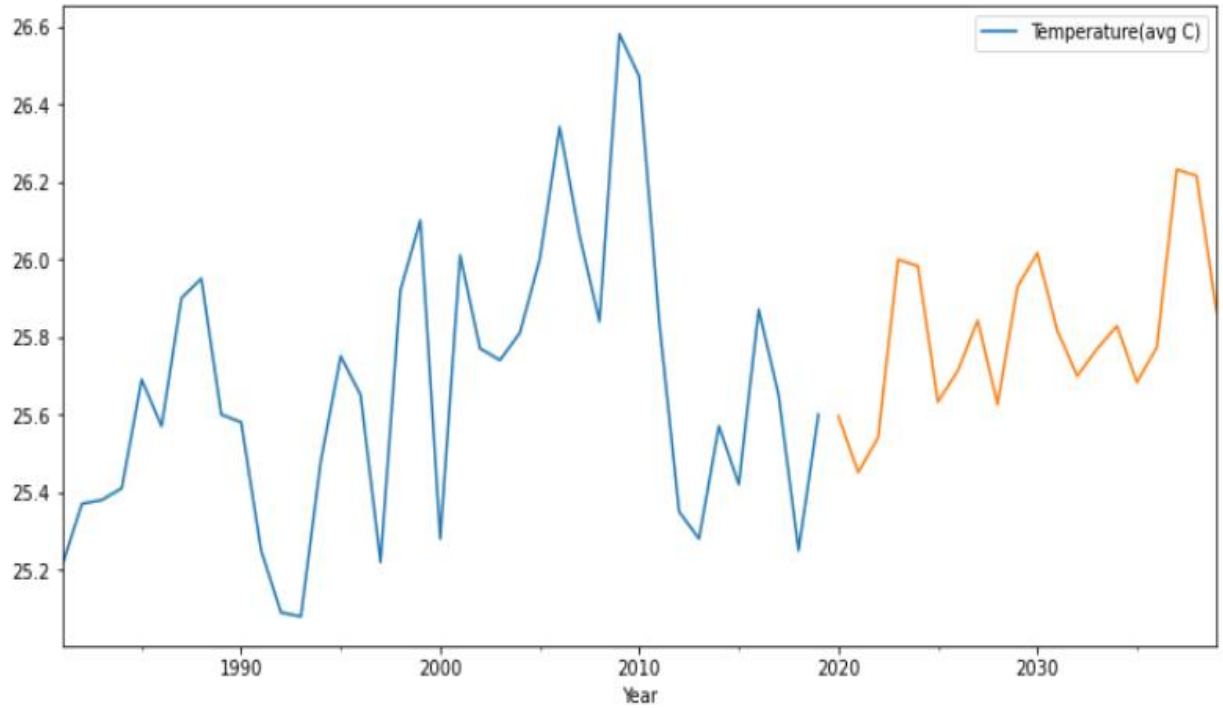


Figure 2: Anticipated creation with real qualities.

Tab(i)_Models result

Model	Mean Absolute Error
ARIMA	.14
LSTM	.25
Exponential Smoothing	.18

From Fig 1, Fig 2, Fig 3 From Fig 1, Fig 2, Fig 3 gives deviation of creation anticipated by ARIMA, LSTM, and Exponential Smoothing individually from the real one. It very well may be seen from Table III and Fig 1 that the forecast given by ARIMA is better nearer to the genuine information as it has a low mean outright mistake than LSTM and Exponential Smoothing.

3) Wind Speed Forecasting:

Arima model: Model are created wind speed in BD and what's more, time series calculations are prepared by investigating 50 years data.. The ARIMA model forecast area on 50 years fitted data.

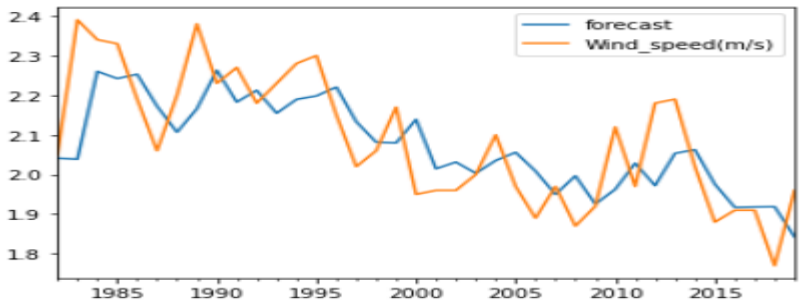


Fig 1: Deviation of anticipated creation with real qualities.

LSTM: This model forecast wind speed certain train data from 50 years. Here the model predicts from tarin data before 1994 to 2004.

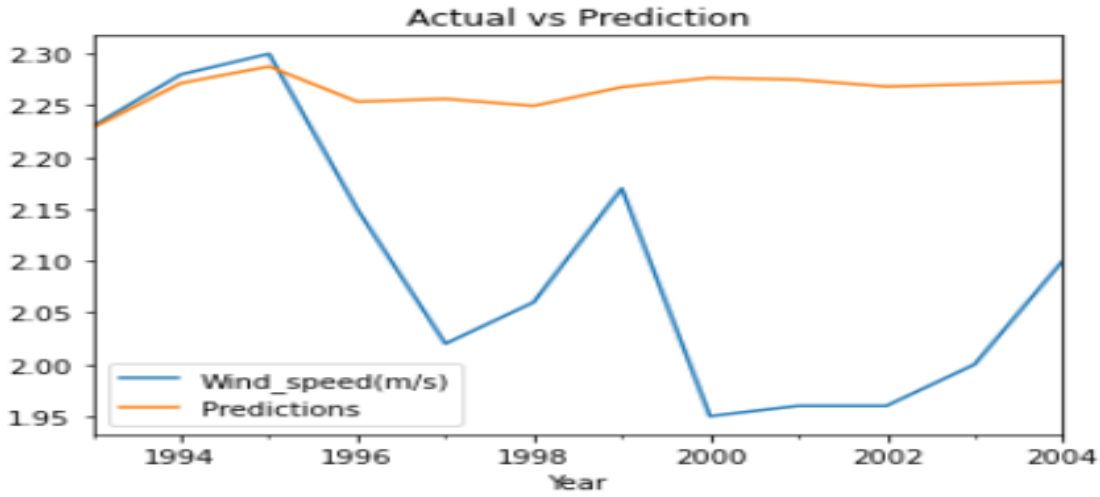


Fig 2: Anticipated creation with real qualities..

Exponential Smoothing: This model forecast future in certain period. This model forecasting area upcoming 20 years.

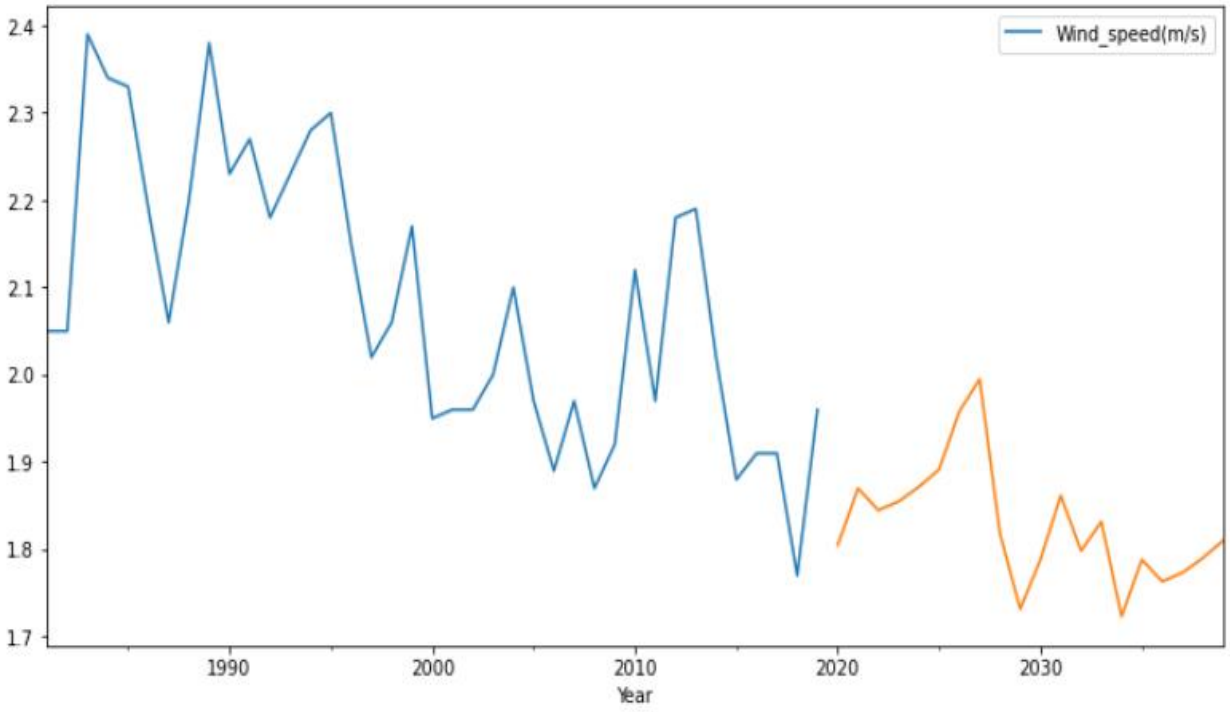


Fig 3: Anticipated creation with real qualities

Table(ii). Models results

Model	Mean Absolute Error
ARIMA	.10
LSTM	.18
Exponential Smoothing	.13

From Fig 1, Fig 2, Fig 3 gives deviation of creation anticipated by ARIMA, LSTM, and Exponential Smoothing individually from the real one. It very well may be seen from Table(ii) and Fig 1 that the forecast given by ARIMA is better nearer to the genuine information as it has a low mean outright mistake than LSTM and Outstanding Smoothing.

4) Humidity Forecasting:

Arima model: Model are created humidity in BD and what's more, time series calculations are prepared by investigating 50 years data. The ARIMA model forecast area on 50 years fitted data.

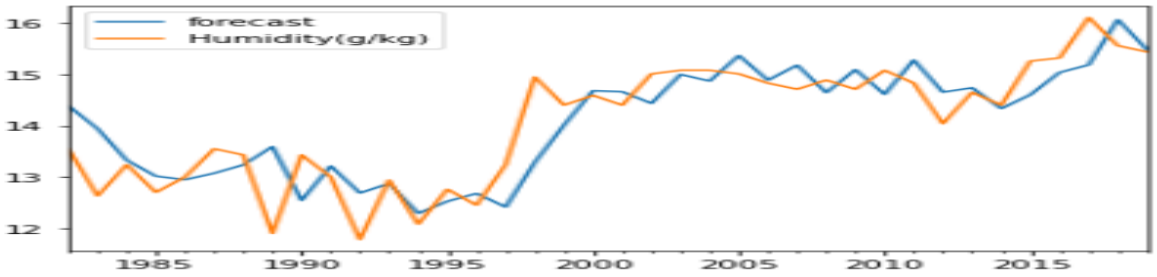


Figure 1: Deviation of anticipated creation with real qualities

LSTM : This model forecast humidity certain train data from 50 years. Here the model predicts from tarin data before 1994 to 2004.

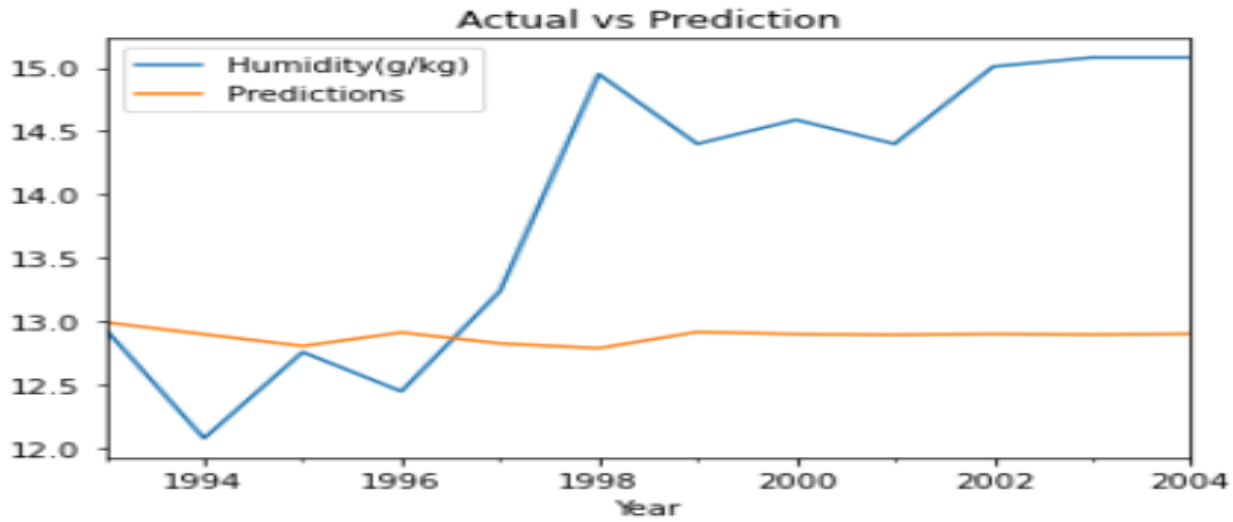


Fig_2: Anticipated creation with real qualities.

Exponential Smoothing: This model forecast future in certain period**. This model forecasting area upcoming 20 years.**

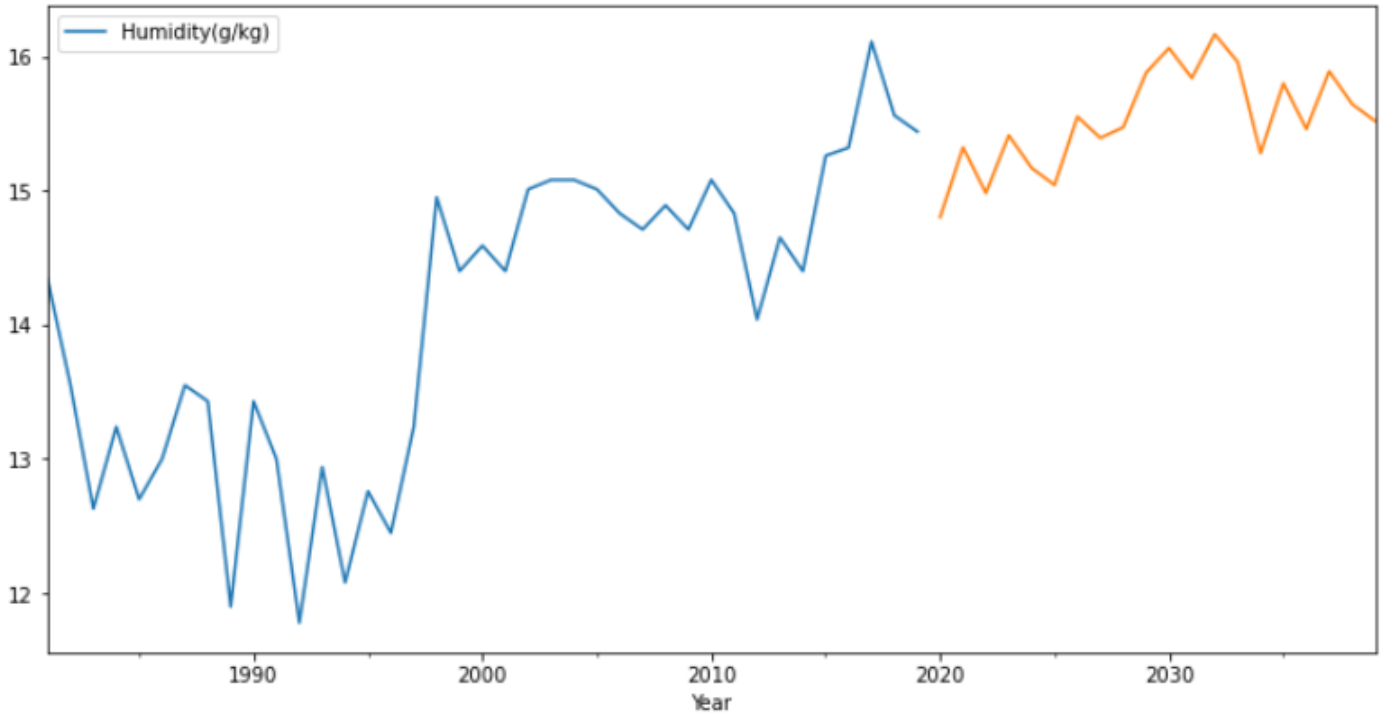


Fig 2: Anticipated creation with real qualities

Table(iii) _Models Results

Model	Mean Absolute Error
ARIMA	.09
LSTM	.22
Exponential Smoothing	.18

From Fig 1, Fig 2, Fig 3 gives deviation of creation anticipated by ARIMA, LSTM, and Exponential Smoothing individually from the real one. It very well may be seen from Table(iii) and Fig 1 that the forecast given by ARIMA is better nearer to the genuine information as it has a low mean outright mistake than LSTM and Outstanding Smoothing.

.5) Rainfall Forecasting:

Arima model: Model are created rainfall in BD and what's more, time series calculations are prepared by investigating 50 years data. The ARIMA model forecast area on 50 years fitted data.

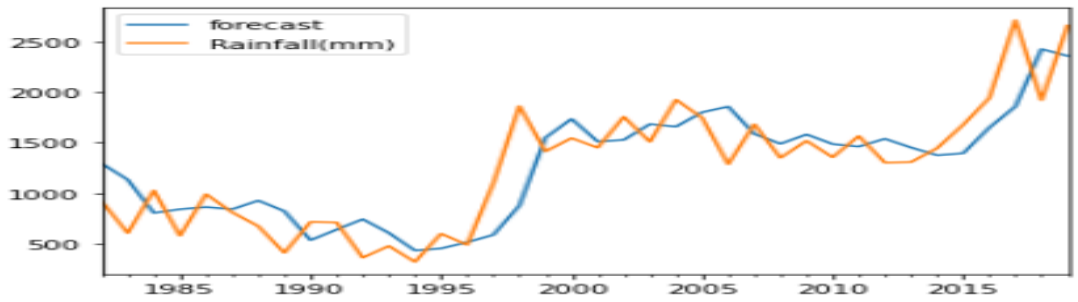


Fig 1_ Deviation of anticipated creation with real qualities

LSTM: This model forecast rainfall certain train data from 50 years. Here the model predicts from tarin data before 1994 to 2004.

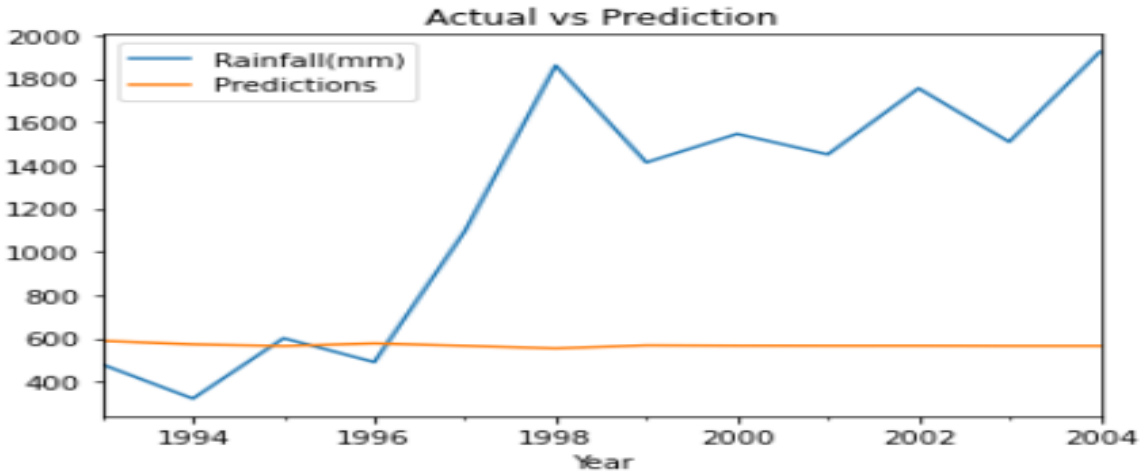


Fig 2: Anticipated creation with real qualities.

Exponential Smoothing: This model forecast future in certain period**. This model forecasting area upcoming 20 years.**

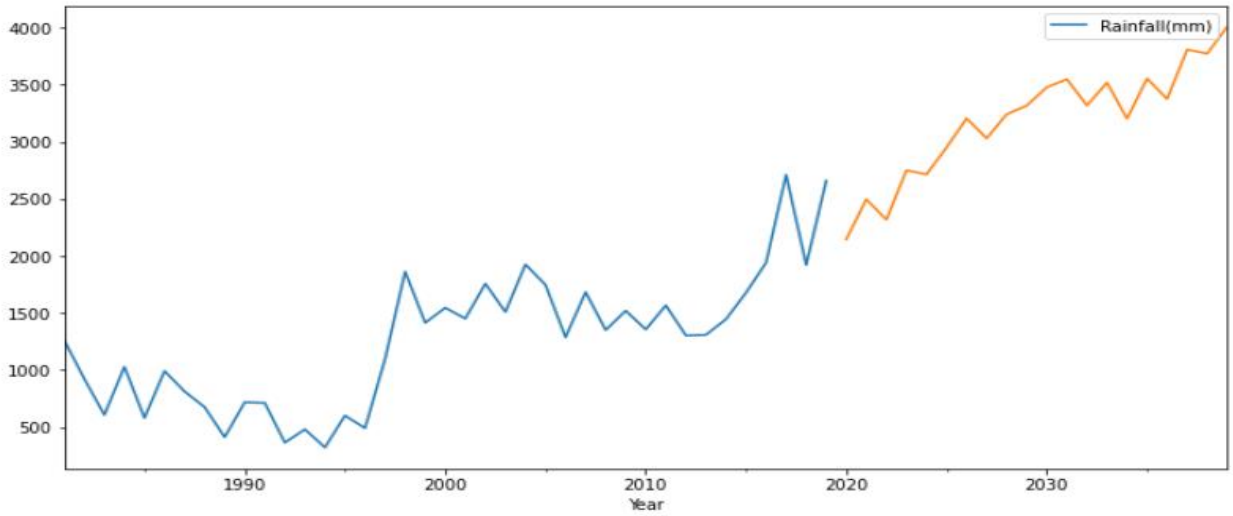


Fig 3: Anticipated creation with real qualities

Table(iv)_Models Results_

Model	Mean Absolute Error
ARIMA	.10
LSTM	.20
Exponential Smoothing	.12

From Fig 1, Fig 2, Fig 3 gives deviation of creation anticipated by ARIMA, LSTM, and Exponential Smoothing individually from the real one. It very well may be seen from Table iv and Fig 1 that the forecast given by ARIMA is better nearer to the genuine information as it has a low mean outright mistake than LSTM and Outstanding Smoothing.

6) **Production Forecasting:**

Arima model: Model are created Temperature in BD and what's more, time series calculations are prepared by investigating 50 years data. The ARIMA model forecast area on 50 years fitted data.

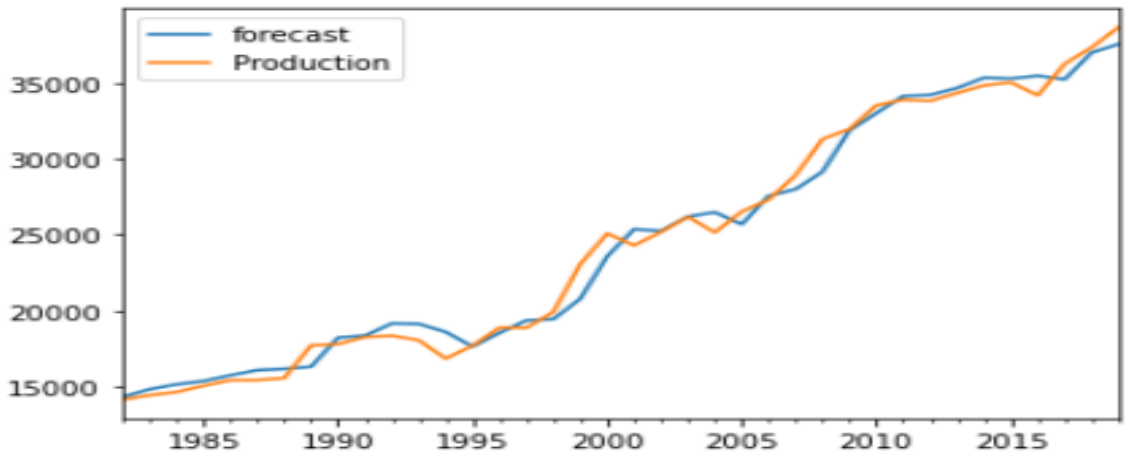


Fig i: Deviation of anticipated creation with real qualities.

LSTM: This model forecast temperature certain train data from 50 years. Here the model predicts from tarin data before 1994 to 2004.

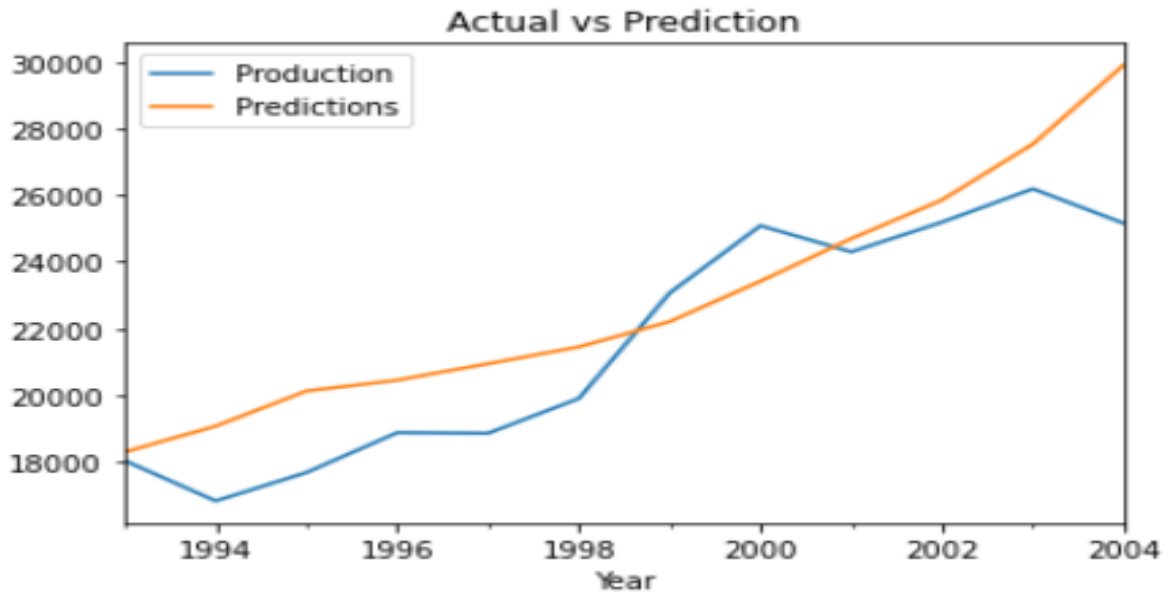


Fig 2: Anticipated creation with real qualities.

Exponential Smoothing: This model forecast future in certain period**. This model forecasting area upcoming 20 years.**

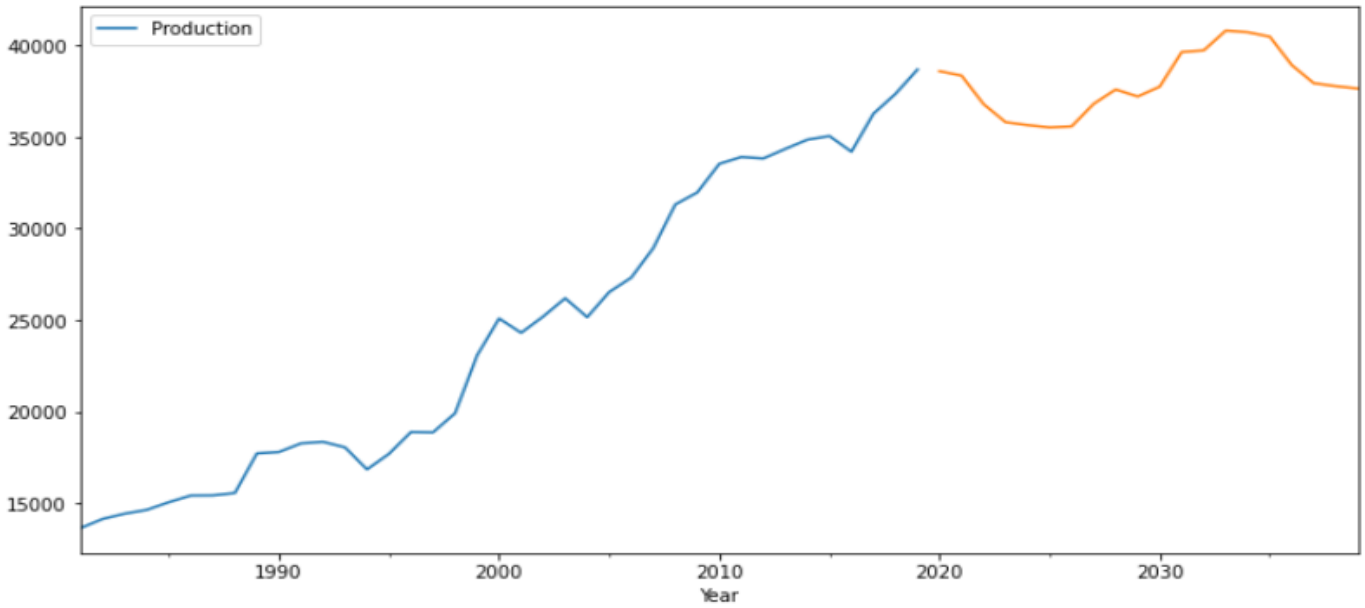


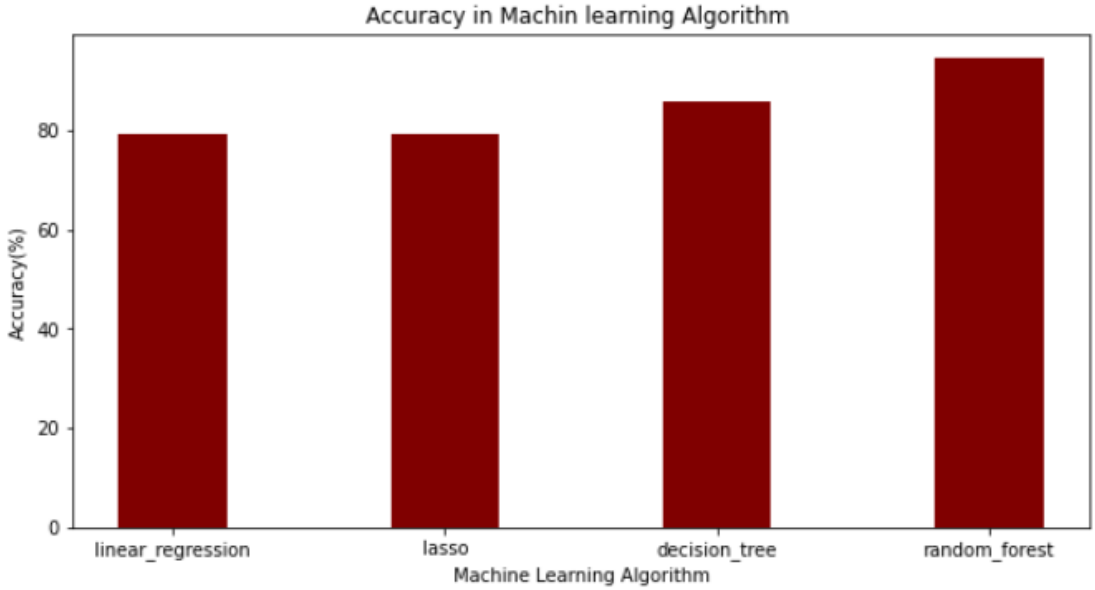
Fig 3: Anticipated creation with real qualities

Tab V. Models Results

Model	Mean Absolute Error
ARIMA	.06
LSTM	.12
Exponential Smoothing	.10

From Fig 1, Fig 2, Fig 3 gives deviation of creation anticipated by ARIMA, LSTM, and Exponential Smoothing individually from the real one. It very well may be seen from Table v and Fig 1 that the forecast given by ARIMA is better nearer to the genuine information as it has a low mean outright mistake than LSTM and Outstanding Smoothing.

Crop Production Prediction for ML Model: Crop Creation in 50 years in Bangladesh was chosen and their creation amount was anticipated based on region, temperature, wind speed, precipitation, and creation. Different kind was prepared on the dataset to feature the exhibition of a model for which the genuine qualities were known.



Model	Accuracy (in Percentage)
Linear Regression	79.40
Lasso Regression	79.39
Decision Tree Regression	84.53
Random Forest	94.63

Figure 2: Correlation of different models.

Accuracy of Models Applied_

It tends to be seen from fig 2 that; Random Forest Classifier gives better results and, in this manner, outflanked the wide range of various methods.

CHAPTER 5

5. CONCLUSION AND LIMITATIONS

5.1 CONCLUSION

The paper introduced the different AI machine leering Algorithms for anticipating the crop yield of the harvest on the premise of mean temperature, wind speed, and area. Tests were led on the Bangladeshi government dataset and it has been laid out that Arbitrary Timberland Regressor gives the best return expectation precision. The consecutive model is Straightforward Repetitive Brain Organization performance measure on precipitation expectation while LSTM is great for temperature forecast. By consolidating precipitation, and temperature alongside different boundaries like season and region, yield expectations for a specific locale can be made. Results uncover that Irregular Woodland is awesome classifier when all boundaries are joined. This will not just assist ranchers in picking the right harvest to fill in the next season yet additionally overcome any barrier between innovation and the agribusiness area. To find the best-performing model, models with more and fewer highlights ought to be tried. Numerous calculations have been utilized in various examinations. The outcomes demonstrate the way that no particular end can be attracted with regard to what the

best model is, yet they obviously show that some AI models are utilized more than others. The most utilized models are the irregular backwoods, brain organizations, straight relapse, and slope-supporting trees. A large portion of the examinations utilized an assortment of AI models to test which model had the best forecast.

5.2 LIMITATIONS

In the work, the output is not much. It may be because of the angle of the weather data which is on basis of whole Bangladesh. Further, I will try to take district wise data in perfect angle get higher prediction, I will try to get the flow of each district data separately and use other models to compare the Time series model.

6. REFERENCES

[1] Nigam, A., Garg, S., Agrawal, A. and Agrawal, P., 2019, November. Crop yield prediction using machine learning algorithms. In *2019 Fifth International Conference on Image Information Processing (ICIIP)* (pp. 125-130). IEEE.

[2] Crane-Droesch, A., 2018. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environmental Research Letters*, *13*(11), p.114003.

[3] Van Klompenburg, T., Kassahun, A. and Catal, C., 2020. Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, *177*, p.105709.

[4] Nigam, A., Garg, S., Agrawal, A. and Agrawal, P., 2019, November. Crop yield prediction using machine learning algorithms. In *2019 Fifth International Conference on Image Information Processing (ICIIP)* (pp. 125-130). IEEE.

[5] Shahhosseini, M., Hu, G. and Archontoulis, S.V., 2020. Forecasting corn yield with machine learning ensembles. *Frontiers in Plant Science*, *11*, p.1120.

[6] Kalimuthu, M., Vaishnavi, P. and Kishore, M., 2020, August. Crop prediction using machine learning. In *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)* (pp. 926-932). IEEE.

[7] Shahhosseini, M., Hu, G., Huber, I. and Archontoulis, S.V., 2021. Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. *Scientific reports*, *11*(1), pp.1-15.

[8] Ahmed, M.T., Imtiaz, M.N. and Mitu, N.S., 2020. Impact of weather on crops in few northern parts of Bangladesh: HCI and machine learning based approach. *Aust. J. Eng. Innov. Technol*, *2*(1), pp.7-15

