Supply Chain Insider
Volume 14, Issue 01, 1-39. 10-10-24
Article Received: 20-08-2024
Accepted: 15-09-2024
Available Online: 10-10-2024
ISSN: 2617-7420 (Print), 2617-7420 (Online)
DOI: 10.5281/zenodo.13865817
supplychaininsider.org

Optimization of rice supply chain through multivariate forecasting using ensemble learning models

Faiyaz Bin Mahmud[1], Md Jawad Bin Rouf[1], Mohammad Mynul Islam Mahin[1]

[1]Department of Mechanical and Production Engineering, Ahsanullah University of Science and Technology, Dhaka-1208, Bangladesh

Correspondence: Md Jawad Bin Rouf, Department of Mechanical and Production Engineering, Ahsanullah University of Science and Technology, Dhaka-1208, Bangladesh

Email: mdjawadbinrouf@gmail.com

Abstract

Accurate forecasting of crop yields and production levels is critical for optimizing agricultural supply chains and enhancing food security. This study evaluates six ensemble learning techniques for simultaneously forecasting the yield rate and production of rice utilizing historical data of several features from various districts in Bangladesh. The models are assessed based on five performance metrics. Among the models, Gradient Boosting (GB) demonstrated superior performance, achieving the lowest mean absolute error (MAE), mean squared error (MSE), median absolute error (MeAE), and the highest R2 score of 0.9943. Category Boosting (CatBoost) and Extreme Gradient Boosting (XGBoost) models also performed strongly, with R^2 values of 0.9917 and 0.9892, respectively. These findings offer significant contributions to enhancing supply chain efficiency by enabling better resource allocation, demand planning, and distribution strategies. Furthermore, the outcomes have extensive implications for policymakers, stakeholders, and food security initiatives, supporting informed decision-making in the face of growing demand and environmental challenges.

Keywords: Supply Chain Management, Ensemble Learning, Agriculture, Forecasting, Inventory Management

1. Introduction

1.1 Background

In this world of technology, it becomes not a fad but a norm to apply scientific approaches and training by machine learning on the solution of existing problems (Sarker, 2021). This has led first-world countries to develop in areas of agriculture, health and education (Arshad et al., 2024). For example, the over dependence on internet in the last two decades has seen definite advantages in organizations and to citizens from actual time production and consumption (Farooq et al., 2020). Nowadays IoT is not only enhancing the efficiency of the user perception and ability of changing working environments but also offering solutions in health, retailing, traffic, security problems, smart homes, cities, and is in the process of doing so in agriculture. The IoT is suitable for constant and close monitoring in agriculture, including services such as farming, cattle, and greenhouses (Maraveas et al., 2022; Tuser et al., 2023). These IoT-based setups employ wireless sensor networks (WSNs) to gather data using the sensing devices and cloud solutions for analyzing and processing physical data collected from distant locations so that more effective decisions can be made by the research scholars and agriculturists (Farooq et al., 2020). On the other hand, the countries in the developmental stage such as Bangladesh fail to integrate these concepts due to their unwillingness to spend a lot of income in new technologies, and their commonly used excuse being high costs (Sultana & Tamanna, 2021; Ahmed et al.,2023). Poor investment in AI technology and IoT solutions can be said to have contributed to the slowing down of this nation's progress in fields like agriculture, healthcare among others (Mhlanga, 2021). Therefore, Bangladesh is depriving itself of the wonderful changes that are brought by technologies such as increase in productivity, better health facilities, and improvement in the lives of people.

Agriculture plays a pivotal role in Bangladesh, both as a key component of the country's economy and as the primary source of livelihood for the majority of its population (Alam et al., 2009; Hamid et al., 2023). Bangladesh's agricultural sector is characterized by its diverse crop production (Ruane et al., 2013). The country's favorable climatic conditions and fertile lands support a variety of agricultural activities, making it a crucial player in ensuring food security and sustaining economic growth (Khan, 2021). The agricultural sector employs about 45% of the workforce and contributes nearly 17% to the Gross Domestic Product (GDP) of Bangladesh (Rahman, 2017). Rice dominates the agricultural landscape, accounting for approximately 75% of the country's total crop production (Gumma et al., 2014; Dey et al., 2023). It is not only the staple food for the Bangladeshi population but also a major export commodity, contributing significantly to the national economy (Bandumula, 2017). Despite its importance, the agricultural sector in Bangladesh faces numerous challenges. These include vulnerability to climate change, frequent natural disasters such as floods and cyclones, and the limited use of modern farming techniques and technologies (Hossain et al., 2011). Additionally, inefficiencies in the supply chain, from production to distribution, often result in significant post-harvest losses and lower profitability for farmers (Ali et al., 2021). Addressing these challenges through technological advancements and optimized supply chain management is crucial for enhancing agricultural productivity and ensuring sustainable economic development in Bangladesh. Agricultural technological advancement is thus a key to success in agricultural growth whereby agriculture technology is aligned with ecological and economic endowments and fine-tuned to the resources obtainable (Abegunde et al., 2019). The purpose of this study is to get a learning agent that could help in the decision-making process for introducing technology and thus enhance the profitability of the agricultural sector (Saiz-Rubio & Rovira-Más, 2020).

1.2 Challenges in Optimizing the Rice Supply Chain

Supply chain management of rice in Bangladesh has been engaging different phases like farming, harvesting, milling and marketing (Gazi, 2020). This chain is vital for ensuring food security and economic stability (Alabi & Ngwenyama, 2022). However, inefficiencies and bottlenecks at different stages can significantly impact overall productivity and profitability. Optimizing the rice supply chain faces several challenges, particularly concerning the Yield Rate (M. Ton) and Production (M. Ton). Variability in weather conditions significantly affects these key metrics, complicating forecasting and planning (Ukhurebor et al., 2022). Inadequate access to modern farming techniques and technologies exacerbates this issue, as farmers often lack the resources or knowledge to stabilize and improve yields (Khatri et al., 2023). Additionally, data availability and quality are critical obstacles, with accurate forecasting hindered by scarce or inconsistent data. Inefficient logistics and distribution networks further complicate the process, as poor infrastructure and transportation issues lead to delays and losses (Kumar et al., 2022). Financial constraints add another layer of complexity, preventing many farmers from investing in technologies and practices that could enhance productivity and profitability (Akpan & Zikos, 2023).

1.3 Significance of the Study

Machine learning is a part of Artifical Intelligence (AI) which can be defined as the training of an algorithm on data and these algorithms are used to make predictions or decisions (Dhall et al., 2019). It is crucial since it is capable of changing several fields by obtaining and forecasting data that was inaccessible in the past (Wang et al., 2022). Ensemble learning is particularly a kind of machine learning approach acquired by combining different models to improve both the accuracy and, thus the stability of the forecast made, then when it is made by a single model (Zounemat-Kermani et al., 2021).

To overcome the challenges, this research proposes the usage of machine learning, including ensemble learning, to enhance the rice supply chain (Sharma et al., 2020). The potential effects comprise refining of the agricultural planning and policies through accurate prediction, which shall enhance efficient utilization of the resources towards the enhancement of agricultural strategies (Weiss et al., 2020). Farmers, distributors, and consumers stand to benefit significantly, as informed decision-making can help maximize profit margins and create a more reliable and efficient supply chain (Mu et al., 2021). Furthermore, this research contributes to the broader field of supply chain optimization by demonstrating the effectiveness of ensemble learning techniques in improving predictive accuracy and operational efficiency (Pasupuleti et al., 2024).

This study plans to address the following research questions (**RQs**):

RQ1: How can ensemble learning techniques effectively forecast rice yield and production in Bangladesh?

RQ2: Which ensemble learning model demonstrates the highest effectiveness for predicting rice yield and production?

RQ3: How can the findings aid in decision-making processes with respect to agricultural planning and supply chain management in Bangladesh?

The study endeavors to accomplish the following research objectives (ROs) by addressing the **RQs:**

RO1: To deploy ensemble learning algorithms for developing accurate forecasts of rice yield and production.

RO2: To assess and contrast the performance of various ensemble learning models to identify the most effective one for predicting rice yield and production.

RO3: To provide actionable insights to policymakers, farmers, and supply chain stakeholders, and assist decision-making processes in regard to effective agricultural planning and supply chain management strategies in Bangladesh.

2. Literature Review

Bangladesh is relatively new to this field of study within agriculture and technology, and the data sets are not very organized which is why it is hard to determine the exact outcome. Optimization of forecasting has become a prominent field of interest within data analysis and predictive modeling. The primary objective is to enhance the precision and dependability of future predictions by refining the models used. This approach involves adjusting the parameters of a forecasting model to minimize the variance between forecasted and observed values, ultimately improving the model's predictive capability (Abolghasemi. 2023). This optimization process is crucial for making more informed decisions in a variety of sectors, such as finance, logistics, and environmental planning (Abolghasemi. 2023).

Optimization techniques have proven to be highly valuable in numerous fields, particularly in forecasting. For example, optimization has boosted supply chain management efficiency in logistics, resulting in lower costs and higher service levels (Juan et al.,2021). For perishable goods, optimizing is vital to maintain product quality, protect consumers, and manage supply chain expenses efficiently (Azadi, 2018). The application of optimization in forecasting has been demonstrated in numerous studies across different domains. For example, in the financial sector, optimization techniques have been used to improve the accuracy of stock price predictions, thereby enabling investors to make more informed decisions (Dey & Chatterjee, 2023) . In the logistics industry, optimization has been applied to enhance the efficiency of supply chain management, leading to reduced costs and improved service levels. In the environmental sector, optimization has been crucial for predicting weather patterns and climate change impacts, aiding in the development of more effective mitigation strategies (Dey & Chatterjee, 2023).

Intelligent methods of forecasting, such as machine learning (ML) and deep learning (DL), leverage advanced algorithms to analyze large volumes of data and uncover complex patterns (Begum, 2023). In ML-based forecasting, algorithms learn from historical data to identify relationships and make predictions. Common ML techniques for forecasting include decision trees, random forests, support vector machines, and gradient boosting. These methods excel at handling non-linear relationships and can adapt to changing patterns in data . Deep learning, a subset of ML, employs neural networks with multiple layers to learn intricate patterns from data (Begum, 2023). Machine learning (ML) and deep learning (DL) have been increasingly applied in the field of forecasting, demonstrating their potential to enhance the accuracy and efficiency of predictions across various domains. These technologies have been utilized in a wide range of applications, from weather forecasting to financial market predictions, and from energy demand forecasting to disease outbreak modeling (Mendonça et al., 2024). In weather forecasting, ML and DL models have been employed to predict weather patterns with remarkable precision, improving the accuracy of weather forecasts and enabling more effective planning and response to extreme weather events. In the financial sector, ML and DL have been instrumental in predicting market trends and movements, improving the accuracy of market predictions and revolutionizing the way financial markets are analyzed and managed. In energy demand forecasting, ML and DL have been used to predict future energy consumption patterns, enabling utility companies to optimize their energy production and distribution networks. In public health, ML and DL have been utilized to model disease outbreaks and predict their potential impact, informing public health policies and enabling more effective responses to disease outbreaks (Mendonça et al., 2024). The applications of ML and DL in forecasting demonstrate their potential to revolutionize various sectors by providing accurate and timely predictions, enhancing the accuracy of predictions, and opening new possibilities for innovation and improvement in various fields (Shomoye, 2024).

2.1 Study Context

In Bangladesh, the integration of machine learning and ensemble learning techniques in agriculture is becoming increasingly important for enhancing predictive accuracy and decision-making processes (Nti et al., 2023). As rice is a staple food and a significant part of the national economy, optimizing its supply chain is crucial for food security and economic stability (Bala et al., 2016; Shikder et al., 2022). However, the rice supply chain in Bangladesh faces several challenges, including weather variability, limited access to modern farming technologies, and inefficient logistics (Jamal et al., 2023). Traditional methods of predicting rice yield and production in Bangladesh often fall short due to their inability to account for the complex factors affecting agricultural output. Ensemble learning techniques, which combine multiple machine learning models, offer a promising solution by improving predictive performance (Ganaie et al., 2022). To obtain better forecasts of rice yield and production with higher accuracy, this study utilizes certain algorithms like Random Forest, AdaBoost, Gradient Boosting, LightGBM, CatBoost, and XGBoost. The research focuses on using historical data, including details on crop yield per hectare, production, temperature averages, production area, year range, and region, to predict the yield rate (M.Ton) and production (M.Ton) of rice across Bangladesh. By leveraging these data, the study aims to provide actionable insights that can enhance agricultural planning and policymaking, leading to better resource allocation and more effective agricultural strategies. Farmers in Bangladesh stand to benefit significantly from this research by gaining the ability to make informed decisions that maximize their profit margins and improve productivity (Mainuddin et al., 2021). Additionally, a more reliable and efficient supply chain will benefit distributors and consumers by stabilizing prices and ensuring a consistent supply of rice.

2.2 Related Works

Machine learning has increasingly become a significant tool in agriculture, particularly for forecasting crop output. Dahikar and Rode (2014) elaborated on the application of Artificial Neural Networks (ANN) in crop yield estimation where one of the type of models being used was feed-forward back propagation neural network. The set parameters that were taken into consideration include pH, Nitrogen, Temperature, and Rainfall to make a perfect prediction. Some of the methods needed one or multiple variables or described complicated techniques, relying on well-structured data primarily from technologically developed nations. Shakoor et al. (2017) for improving the decision-making regarding farmers' crops in Bangladesh utilized supervised machine learning approaches to predict agricultural production outputs. The study availed itself to tackling the problem, which normally emanates from farmers employing past experiences in arriving at decisions for growing crops that are less profitable. Such an informative approach of the research pursued the objective of making agriculture business more efficient and profitable through the provision of intelligent information prediction analysis. The study focused on six major crops: Aus rice, Aman rice, Boro rice, potato, jute and wheat. The researchers used the information gathered from the Yearbook of Agricultural Statistics and Bangladesh Agricultural Research Council Information System to take Decision Tree Learning (ID3) and K-Nearest Neighbors Regression then proposed more economical crop for the concerned sector. Priyadarshi et al., 2019 analyzed that number of studies have shown that algorithms like Support Vector Regression and Long Short-Term Memory networks yielded better results than the other possible models. They found that their approach, LSTM and SVR, led to enhancement on the revenues as well increased efficiency in the reduction of the forecast errors. Although the research applies to all fruits and vegetables, it indicates that easy-to-use systems can be developed to meet the daily demand of the fresh produce, hence the low perishable products and inventory. Condran et al. (2022) provided a comprehensive review on the use of machine learning in agriculture, that also encompasses the review of experimental scenarios for benchmarking of the model performance. Speaking of the research limitations, they pointed out the future research avenue in the area of Internet of Things (IoT) application in agriculture, which looks promising for improving the agricultural practices. Specifically, the paper by Moon et al. (2023) synthesizes the performance of different machine learning approaches from the perspective of wheat and rice production in Bangladesh. Some of the methods used in the study included K-nearest Neighbor, Random Forest, Ridge Regression, Support Vector Regression, Naïve Bayes, and CatBoost. The authors recommended that subsequent studies should increase learnable datasets which contain other data from other sources like characteristics of the environment and farming areas. Similarly, Kurek et al. (2023) developed three models for agricultural yield estimation: These are the hybrid, satellite, and non-satellite ones. That is proved by the fact that the identified SVM-based hybrid model which gives the best result has the value of 5. Hence, the sample model, producing approximately 85% Mean Absolute Percentage Error (MAPE), possessed the lowest error rate and proved to yield the best results compared to the satellite as well as the non-satellite models. The study pointed out the possibility of improvement of yield estimates by agricultural machine learning.

The studies mentioned above were based on well managed databases. Bangladesh experiences numerous problems in this area because of insufficient and ineffective data management methods. Hence, the number of predictors for detailed predictions is kept to a minimal. Therefore, a machine learning based prediction model inclusive of algorithms is deployed in this study with the scarce predictor information coming through different sources and within the purview of Bangladesh's agricultural data spectrum.

3. Methodology

The framework proposed for multivariate energy forecasting in Bangladesh is illustrated in **Figure 1**. The study comprises several steps: (1) data collection; (2) data processing, involving preprocessing to ensure data quality and feature engineering for meaningful insights; (3) partitioning the dataset; (4) utilizing the training set for model training; (5) assessing models on the test set; and (6) selecting the best model based on performance metrics.

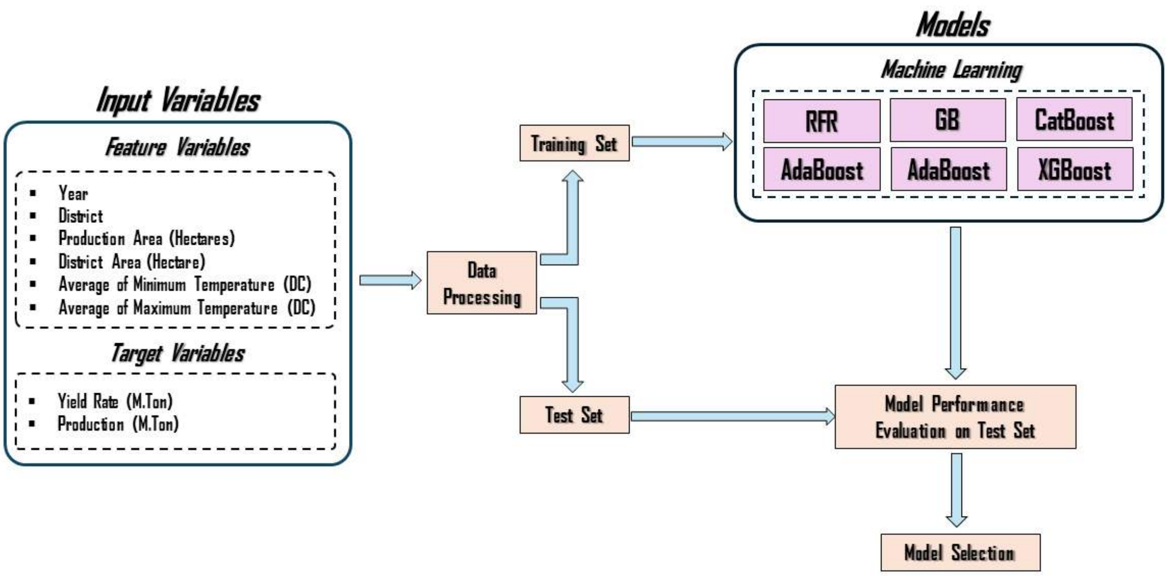

Figure 1: Framework for Multivariate Rice Supply Chain Optimization in Bangladesh

3.1 Data collection and analysis

Historical public data are collected from the Department of Agricultural Extension published by Bangladesh Bureau of Statistics. Yearly district wise data as exhibited in **Table 1** are recorded as a Microsoft Excel file from 2018 to 2023.

Table 1: Accumulated data

Feature No.	Data Collected	Data Type	Source
1	Year	int64	Bangladesh Bureau of Statistics (BBS)
2	District	object	
3	Production Area (Hectares)	float64	
4	Yield Rate (M.Ton) **(Target Variable 1)**	float64	
5	Production (M.Ton) **(Target Variable 2)**	int64	
6	District Area (Hectare)	float64	
7	Average of Minimum Temperature (DC)	float64	
8	Average of Maximum Temperature (DC)	float64	

Figure (2 – 8) depicts the features distribution. The frequency of values that fall within specific ranges for a particular feature is represented by each histogram.

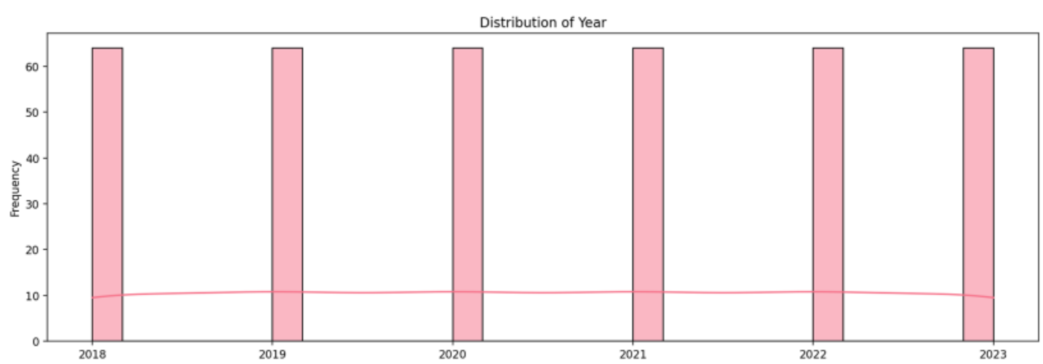

Figure 2: Distribution of Year

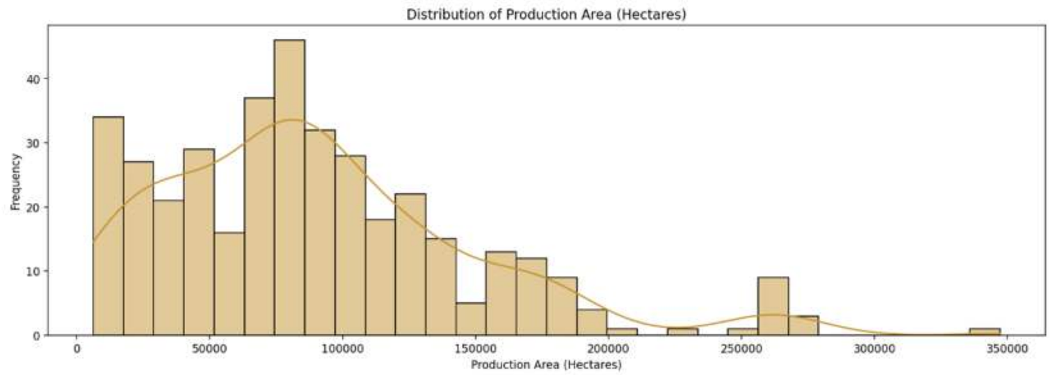

Figure 3: Distribution of Production Area (Hectares)

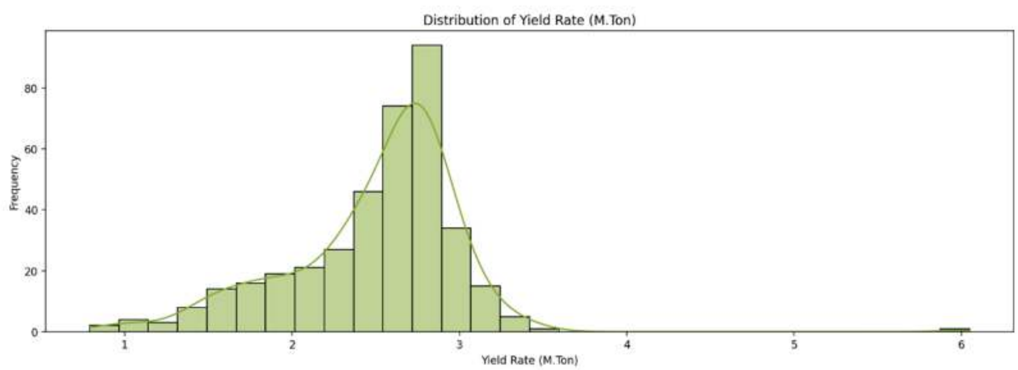

Figure 4: Distribution of Yield Rate (M. Ton)

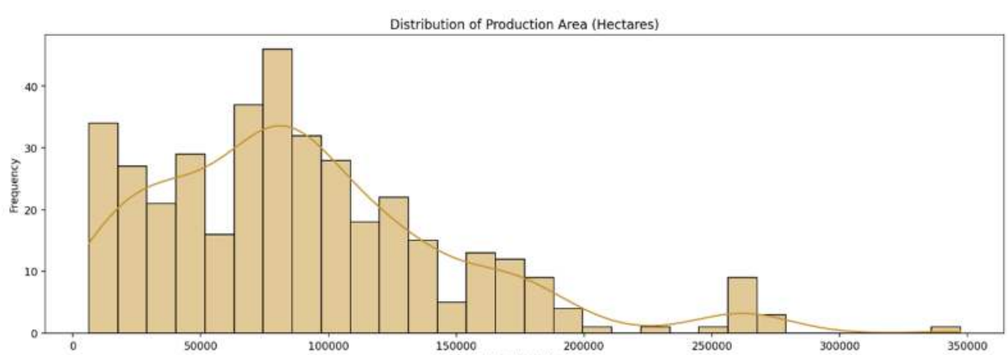

Figure 5: Distribution of Production Area (Hectares)

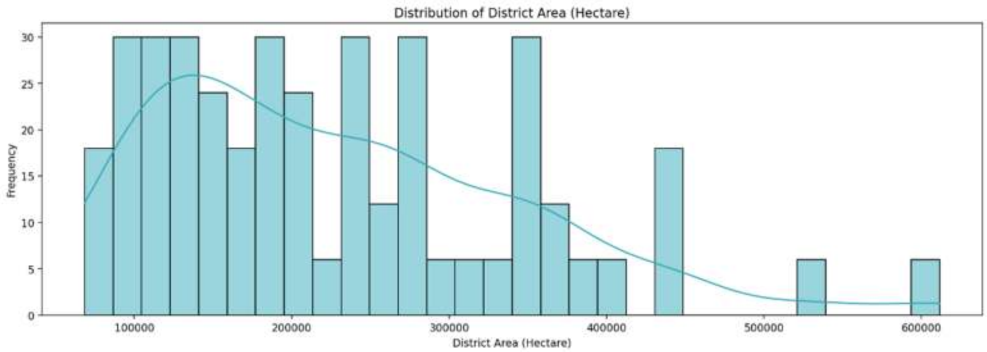

Figure 6: Distribution of District Area (Hectare)

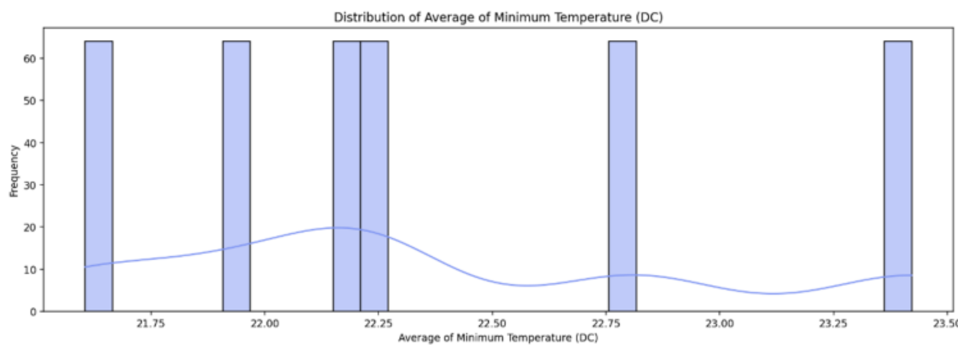

Figure 7: Distribution of Average of Minimum Temperature (DC)

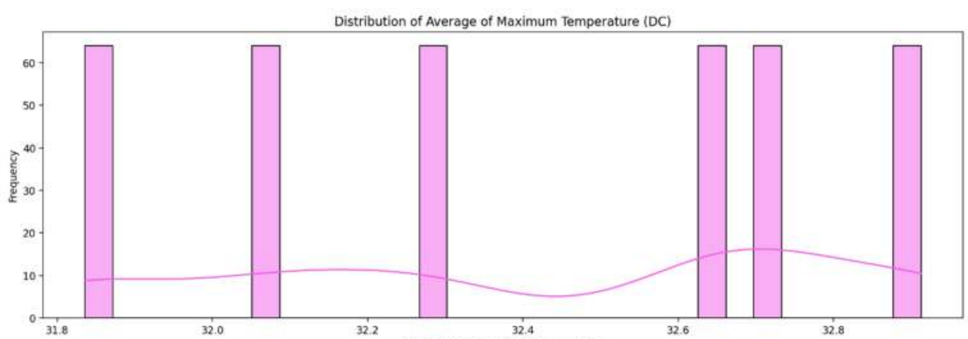

Figure 8: Distribution of Average of Maximum Temperature (DC)

Figure 9 is a graphical representation of the correlation matrix which helps to visualize correlation coefficients between each pair of features.

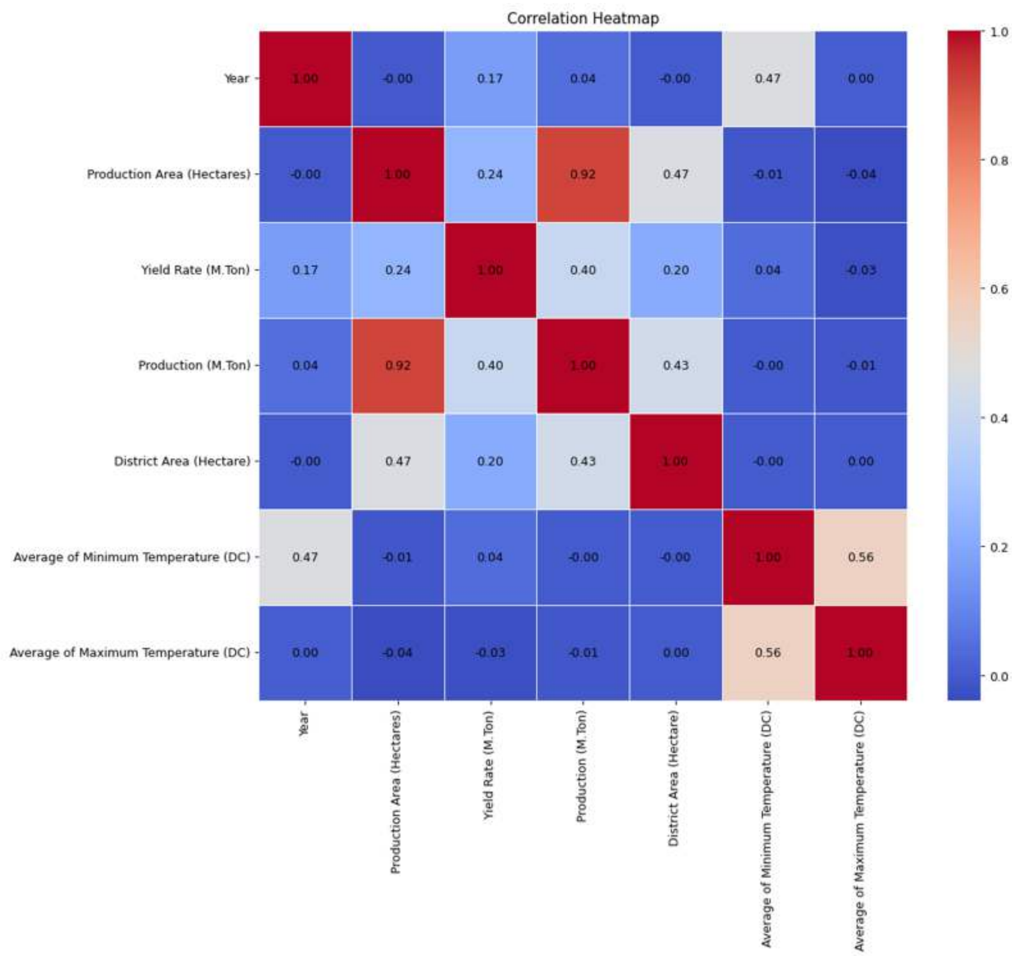

Figure 9: Correlation Heatmap.

Figure (10 – 11) present line plots depicting trends in Yield Rate (M.Ton) and Production (M.Ton) across the years 2018 to 2023. **Figure 10** shows the yield rate (in metric tons) for various years from 2018 to 2023, with a general increasing trend and notable variability each year. **Figure 11** shows the production (in metric tons) for various years from 2018 to 2023, displaying high variability each year with a noticeable downward trend over the years.

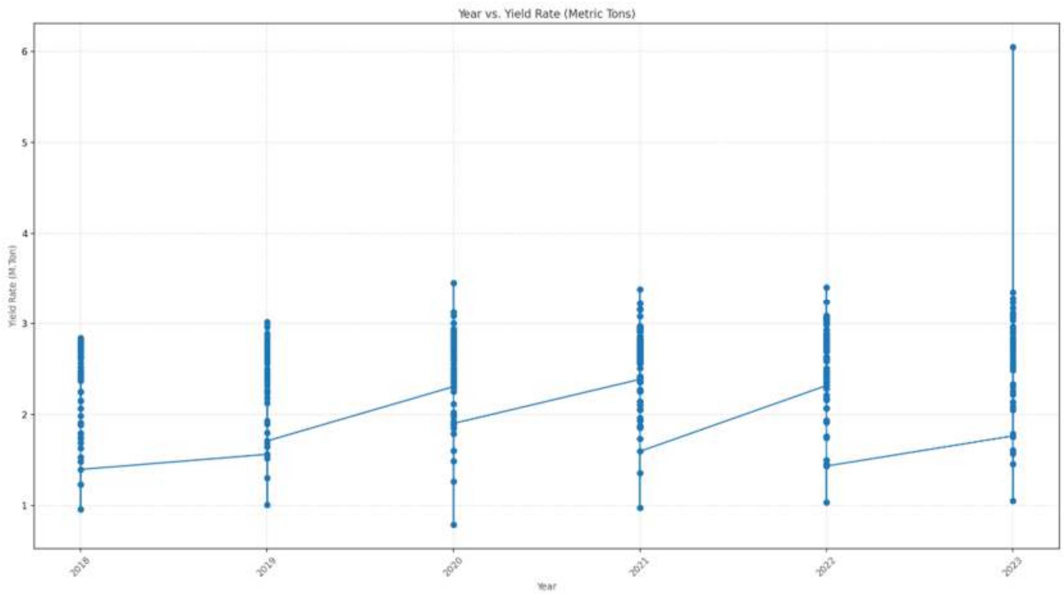

Figure 10: Year vs. Yield Rate (Metric Tons)

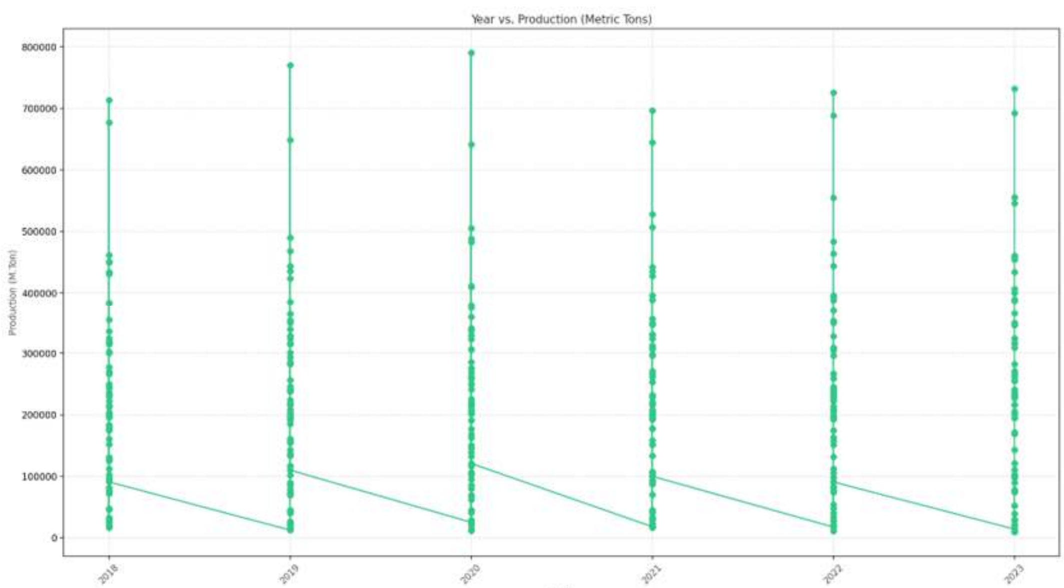

Figure 11: Year vs. Production (Metric Tons)

The scatter plot in **Figure 12** illustrates the relationship between Production area (Hectares) and Production (M.Ton) from 2018 to 2023. This scatter plot shows the relationship between production area (in hectares) and production (in metric tons), with a positive correlation indicated by the upward-sloping trend line and a confidence interval shaded around it.

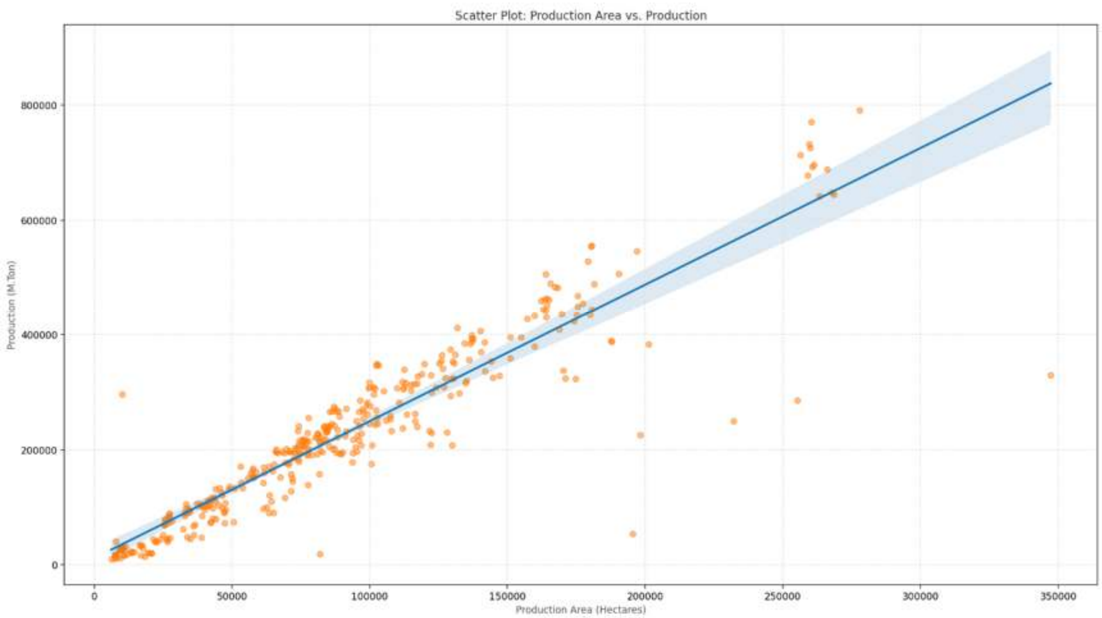

Figure 12: Production Area vs. Production

3.2 Data processing

3.2.1 Managing missing data

Given the time span, missing values are present within the dataset. Ensuring the quality and completeness of the experimental data is essential. Missing values are filled in using last observation carried forward (LOCF) or forward fill method, as shown in **Equation (1)**:

$$x_n = \begin{cases} x_n & \text{when } x_n \text{ is not missing} \\ x_{n-1} & \text{when } x_n \text{ is missing} \end{cases} \quad (1)$$

Let x_n represents the value at time m. When a_m is missing, it is substituted by x_{n-1}, the value from the preceding time step. LOCF propagates the last valid observation forward to fill the missing values along each column (Hadeed et al., 2020).

3.2.2 Data splitting and scaling

It is often common to split data 80:20 to have sets for use in training and testing where 80% of the data is used for training purposes while the rest of 20 % is used in testing. However, to maintain the chronological sequence, the 'shuffle' parameter is set to 'True.' The 'random_state' is also set to 42 to guarantee the reproducible nature of subsequent runs.

The features are normalized using 'StandardScaler' from 'scikit-learn' library, as demonstrated mathematically in **Equation (2)**:

$$X = \frac{M-N}{V} \quad (2)$$

where, transformed value of the feature is represented by X, the original value by M, the mean by N, and the standard deviation by V. Both the training and testing sets of features are scaled to maintain consistency between their scales. Similarly, the target variables are scaled independently to prevent any potential data leakage during model training.

3.3 Machine learning (ML) models

The considered ML models include Random Forest (RF), Gradient Boosting (GB), Adaptive Boosting (AdaBoost), Light Gradient Boosting Machine (LightGBM), Category Boosting (CatBoost) and Extreme Gradient Boosting (XGBoost) to which benchmarking is performed in this research. All the hyperparameters of each model are carefully tuned by pre-processing by Grid Search with Cross-Validation (GridSearchCV) and Randomized Search with Cross

Validation (RandomizedSearchCV). These techniques are applied to select the right hyperparameters for a model while using cross validation to boost the reliability of the process. Concerning the training and validation, the employed scheme is the ‘k-fold cross-validation,’ where the model is trained k times, and the data set is divided into k subsets: one subset is used for validation, while the rest is trained for each iteration of the procedure. An accurate selected statistic is used to evaluate the performance, for example, R^2 or MSE that are calculated on each of the k splits. The best hyperparameters are achieved from the combination that gives an average result from all the folds of the dataset used.

3.3.1 Random Forest (RF)

Random Forest (RF) method is an averaging algorithm aimed at utilizing many randomized decision trees to solve specific tasks in the field of data analysis. Both are techniques belonging to the perturb-and-combine family suitable for trees. (Plaia et al., 2021). In RF each tree is developed by starting with a sample and then making use of substitution from the training data (Bernard et al., 2012). The best split is accomplished by thoroughly searching for the best value of input features or a random subset of size max_features. Two sources of randomness are utilized to diminish the variance of a forest estimator, as individual decision trees frequently overfit because of high variance. Incorporated randomness in forests lead to independent prediction errors, these can be updated through averaging of predictions. RFs connect to different trees, occasionally with a slight increase in bias level, inducing substantial variance decrease and a better model. This model gets an (x) input vector, comprised of the values of various evidential features assessed for a given training region and constructs K number of regression trees and average the outcomes. The RF regression predictor is

$$\hat{f}_{rf}^{K}(x) = \frac{1}{K}\sum_{k=1}^{K} T(x) \qquad \textbf{(3)}$$

after K such trees $\{T(x)\}_1^K$ are developed.

3.3.2 Adaptive Boosting (AdaBoost)

This famous boosting algorithm was presented in 1995 by Freund and Schapire (Schwenk & Bengio, 2000). AdaBoost employs a series of weak learners to predict data, and these are then combined by adding weights to the majority vote to result in the final forecast (Ding et al., 2022). The data modifications at each alleged boosting iteration comprises of applying weights w_1, w_2, …, w_N to each of the training samples. At first, those weights are all set to $w_i = 1/N$, with the goal that the initial step essentially trains a weak learner on the original data. Each time it undergoes the same learning algorithm on reweighted data where weights of examples predicted correctly are reduced and weights of examples predicted incorrectly are boosted. As iterations progress, difficult-to-predict examples acquire more impact, constraining weak learners to focus on missed examples in the sequence. This process guarantees efficient learning and accurate predictions.

3.3.3 Gradient Boosting (GB)

Gradient Boosted Regression Trees (GBRT) regressors are additive techniques that predict $\hat{y}_i$ for a given input x_i in the following form:

$$\hat{y}_i = F_M(x_i) = \sum_{m=1}^{M} h_m(x_i) \quad \textbf{(4)}$$

where the h_m are estimators, with regards to boosting are called *weak learners.* Gradient Tree Boosting uses fixed size decision tree regressors as weak learners. The constant M corresponds to the n_estimators parameter (Yang et al., 2022). A GBRT is built in a greedy fashion similar to other boosting algorithms:

$$F_m(x) = F_{m-1}(x) + h_m(x), \quad \textbf{(5)}$$

where to minimize a sum of losses L_m, the newly added tree h_m is fitted, given the previous ensemble F_{m-1}:

$$h_m = \arg\min_h L_m = \arg\min_h \sum_{i=1}^{n} l(y_i, F_{m-1}(x_i) + h(x_i)),$$

(6)

where $l(y_i, F(x_i))$ is established by the loss parameter. The initial model F_0 is chosen as the constant that minimizes the loss by default: for a least-squares loss, this is the empirical mean of the target values, and can be specified via the init argument. The value of l can be approximated using a first-order Taylor approximation as follows:

$$l(y_i, F_{m-1}(x_i) + h_m(x_i)) \approx l(y_i, F_{m-1}(x_i)) + h_m(x_i)[\frac{\partial l(y_i, F(x_i))}{\partial F(x_i)}]_{F=F_{m-1}}.$$

(7)

The quantity $[\frac{\partial l(y_i, F(x_i))}{\partial F(x_i)}]_{F=F_{m-1}}$ is the loss's derivative in relation to the second parameter, assessed at $F_{m-1}(x)$ and easily computed for $F_{m-1}(x_i)$ in a closed form as the loss is differentiable, and is denoted by g_i. Removal of the constant terms:

$$h_m \approx \arg\min_h \sum_{i=1}^{n} h(x_i) g_i \quad \textbf{(8)}$$

If $h(x_i)$ is fitted to predict a value that is proportional to the negative gradient $-g_i$, this is minimized. T**he estimator h_m is therefore fitted and updated at each iteration to predict the negative gradients of the samples,** resulting in a gradient descent in a functional space.

3.3.4 Light Gradient Boosting Machine (LightGBM)

LightGBM is a GBDT algorithm used in ordering, regression, and classification problems, available features and it corresponds for exclusive feature bundling and gain-based one-side sampling techniques (Hajihosseinlou et al., 2023). Supervised training set is given: $X = \{(x_i, y_i)\}_{i=1}^{n}$, Intention of the method is to obtain an estimation $\hat{f}(x)$ to a specific function $f^*(x)$ that limits the anticipated value of a specific loss function $L(y, f(x))$ as follows:

$$\hat{f} = \arg\min_f E_{y,X} L(y, f(x)) \quad \textbf{(9)}$$

To approximate the final model, LightGBM technique combines several T regression trees $\sum_{t=1}^{T} f_t(X)$, that is:

$$f_T(X) = \sum_{t=1}^{T} f_t(X) \quad \textbf{(10)}$$

Regression trees are be expressed as $w_{q(x)}, q \in \{1,2,\ldots,J\}$, where J signifies the quantity of leaves, q represents the decision rules of the tree and w is a vector that indicates the sample weight of leaf nodes. Subsequently, the method can be trained in an additive form at step t as follows:

$$\Gamma_t = \sum_{i=1}^{n} L(y_i, F_{t-1}(x_i) + f_t(x_i)) \quad \textbf{(11)}$$

Objective function is quickly approximated with Newton's strategy. The formulation can be transformed as follows subsequent to eliminating the constant term in (iii) for straightforwardness:

$$\Gamma_t \cong \sum_{i=1}^{n} \left(g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)\right)$$

(12) where g_i and h_i denote the first- and second-order gradient statistics of the loss function. Let I_j denote the sample set of leaf j, and (iv) could be transformed as follows:

$$\Gamma_t = \sum_{j=1}^{j} \left((\sum_{i \in I_j} g_i) w_j + \frac{1}{2} (\sum_{i \in I_j} h_i + \lambda) w_j^2)\right)$$

(13)

Optimal leaf weight scores of each leaf node w_j^* and the extreme value of Γ_K could be solved for a specific tree structure $q(x)$ as:

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}$$

(14)

$$\Gamma_T^* = -\frac{1}{2} \sum_{j=1}^{J} \frac{\left(\sum_{i \in I_j} g_i\right)^2}{\sum_{i \in I_j} h_i + \lambda}$$

(15) where Γ_T^* can be seen as the scoring function that measures the nature of the tree structure q. Objective function after adding the split is:

$$G = \frac{1}{2} \left(\frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \lambda}\right)$$

(16) where I_L and I_R are the sample sets of the right and left branches, individually.

3.3.5 Category Boosting (CatBoost)

Catboost model uses decision trees as the base learner in regression and each tree reflects output value and the feature space division used in splitting as well as the splitting criteria/ decision rules used for tree splitting (Saber et al., 2021). Criteria of splitting individuals look similar to a pair $p = (q, m)$ having a feature indicator $q = 1, 2, \ldots\ldots, n$, and threshold value $m \in D$When implementing the splitting criteria/decision rule a given set of feature vectors X can be partitioned into two disjoint mathematical subset of X^D and X^C, in order that for every $x = (x^1, x^2, x^3, \ldots \ldots \ldots, x^n) \in X$,

$$x \epsilon \begin{cases} X^C & if x^q \leq m \\ X^D & if x^q > m \end{cases} \quad \textbf{(17)}$$

Subsequent to carrying out splitting criteria/decision rule to e disjoint sets $X_1, X_2, \ldots \ldots \ldots, X_e \in D^n$, $2e$ disjoint sets $X_1^C, X_1^D, X_2^C, X_2^D \ldots \ldots \ldots X_e^C, X_e^D$. For a predetermined collection of sets of $N = \{X_1, X_2, \ldots \ldots \ldots, X_e \in D^n\}$ and the target variable $Y : D^n \rightarrow D$, the splitting criteria/decision rule can be given as $argmin_p\{G(p, Y, N)\}$, where N functions to estimate the optimality of the decision rule/splitting criteria p and the assortment N with respect to the target variable Y. For an negligent decision tree, G can be characterized as:

$$G(p, Y, N) = \frac{1}{\sum_{a=1}^{c} |X_a|} [\sum_{a=1}^{e} |X_a^C| Var\{Y(X_a^C)\} + |X_a^D| Var\{Y(X_a^C)\}]$$

(18)

where $Y\ (X_a)$ is the target variable score as for the sample X_a.

3.3.6 Extreme Gradient Boosting (XGBoost)

XGBoost, which stands for" Extreme Gradient Boosting," is an advanced machine learning algorithm that has gained widespread recognition for its efficiency and performance in handling large datasets (Kiangala & Wang, 2021). Developed by Tianqi Chen and Carlos Guestrin, XGBoost is designed to optimize speed and accuracy through its scalable and distributed computing capabilities (Drahokoupil, 2022). The primary goal of XGBoost is to minimize an objective function that balances model accuracy and complexity. The objective function $Obj(\Theta)$ is a combination of the training loss $L(\Theta)$ and a regularization term $\Omega(\Theta)$:

$$Obj(\Theta) = L(\Theta) + \Omega(\Theta)$$

(19)

Where $L(\Theta)$ measures how well the model fits the training data, and $\Omega(\Theta)$ penalizes the complexity of the model to prevent overfitting. The training loss L can be defined for n training examples as follows:

$$L = \sum_{i=1}^{n} l\,(y_i, \hat{y}_i) \quad \textbf{(20)}$$

For regression problems, the squared loss is typically used:

$$l(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2 \quad \textbf{(21)}$$

Where y_i is the actual value and $\hat{y}_i$ is the predicted value. When a new tree is included in the model, the objective at iteration t is changed as follows:

$$Obj^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t)}\right) + \sum_{j=1}^{t} \Omega\left(f_j\right)$$

(22)

This can be expanded using the predictions from the previous iteration $t-1$:

$$Obj^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) + constant$$

(23)

To simplify the computation, XGBoost employs a second-order Taylor ex- pansion of the loss function, discarding constant terms:

$$Obj^{(t)} \approx \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)\right] + \Omega(f_t)$$

(24)

Where $g_i = \frac{\partial l\left(y_t, \vartheta_1^{(t-1)}\right)}{\partial \vartheta_t^{(t-1)}}$ is the first-order partial derivative, and $h_i = \frac{\partial^2 l\left(y_t, \vartheta_1^{(t-1)}\right)}{\partial \vartheta_t^{(t-1)}}$ is the second-order partial derivative. The regularization term $\Omega(f)$ helps to

control the complexity of the trees and is defined as:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} w_j^2 \quad \textbf{(25)}$$

Where T indicates the number of the leaf nodes, γ stands for the penalty coefficient for the number of T, λ is the regularization parameter for the leaf weights, and w_j is the weight of leaf j. Rewriting the objective function to include the complexity term, we get:

$$Obj^{(t)} = \sum_{i=1}^{n} \left[g_i w_{q(x_i)} + \frac{1}{2} h_i w_{q(x_i)}^2 \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} w_j^2$$

(26)

The gain from splitting a leaf node, which measures the improvement brought by the split, is given by:

$$Gain = \frac{1}{2} \left[\frac{G_L^2}{H_L+\lambda} + \frac{G_R^2}{H_R+\lambda} - \frac{(G_L+G_R)^2}{H_L+H_R+\lambda} \right] - \gamma$$

(27)

Where G_L and G_R are the summed gradients for the left and right nodes, respectively, and H_L and H_R are the summed Hessians (second-order gradients) for the left and right nodes, respectively.

3.4 Evaluation metrics for ML models

Five metrics have been utilized for comparison in order to evaluate the efficacy of the suggested models. MAE, MSE, MAPE, median absolute error (MeAE), and coefficient of determination (R^2) indices are utilized to quantify the differences between the forecasted and actual values. The following are the calculation formulas over n_{sample} samples:

$$\text{MAE}\,(y, \hat{y}) = \frac{1}{n_{\text{sample}}} \sum_{t=0}^{n_{\text{sample}}-1} |y_t - \hat{y}_t|$$

(28)

$$\text{MSE}\ (y, \hat{y}) = \frac{1}{n_{\text{sample}}} \sum_{t=0}^{n_{\text{sample}}-1} (y_t - \hat{y}_t)^2$$

(29)

$$\text{MAPE} = \frac{1}{n_{\text{sample}}} \sum_{t=1}^{n_{\text{sample}}} \left| \frac{y_t - \widehat{y_t}}{y_t} \right| \times 100$$

(30)

$$\text{MeAE}\ (y, \hat{y}) = \text{median}\ (|y_1 - \hat{y}_1|, \dots, |y_n - \hat{y}_n|)$$

(31)

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{t=0}^{n_{\text{sample}}-1} (y_t - \hat{y}_t)^2}{\sum_{t=0}^{n_{\text{sample}}-1} (y_t - y)^2} \quad \textbf{(32)}$$

where, $\bar{y} = \frac{1}{n_{\text{sample}}} \sum_{t=0}^{n_{\text{sample}}-1} y_t$

Here, y_t is the true value and $\hat{y}_t$ is the predicted value corresponding to the t-th sample.

4. Results, comparative analysis, and discussion

Forecast accuracy indices for the suggested ML models are presented in **Table 2**.

Table 2: Indices of ML Models

Model	MAE	MSE	MAPE (%)	MeAE	R^2
RF	0.0588	0.0096	5.133	0.0258	0.9850
AdaBoost	0.1293	0.0289	31.863	0.0957	0.9608
GB	**0.0376**	**0.0048**	6.375	**0.0161**	**0.9943**
LightGBM	0.0886	0.0366	15.692	0.0394	0.9523
CatBoost	0.0444	0.0052	**4.341**	0.0278	0.9917
XGBoost	0.0537	0.0071	4.899	0.0329	0.9892

4.1 MAE score

Figure 13 illustrates the comparison of MAE scores among the models. GB achieves the lowest MAE of 0.0376, demonstrating superior performance in minimizing absolute errors. Catboost follows with an MAE of 0.0444, and XGBoost achieves an MAE of 0.0537. RF records an MAE of 0.0588. LightGBM and AdaBoost exhibit the highest MAE values of 0.0886 and 0.1293 respectively.

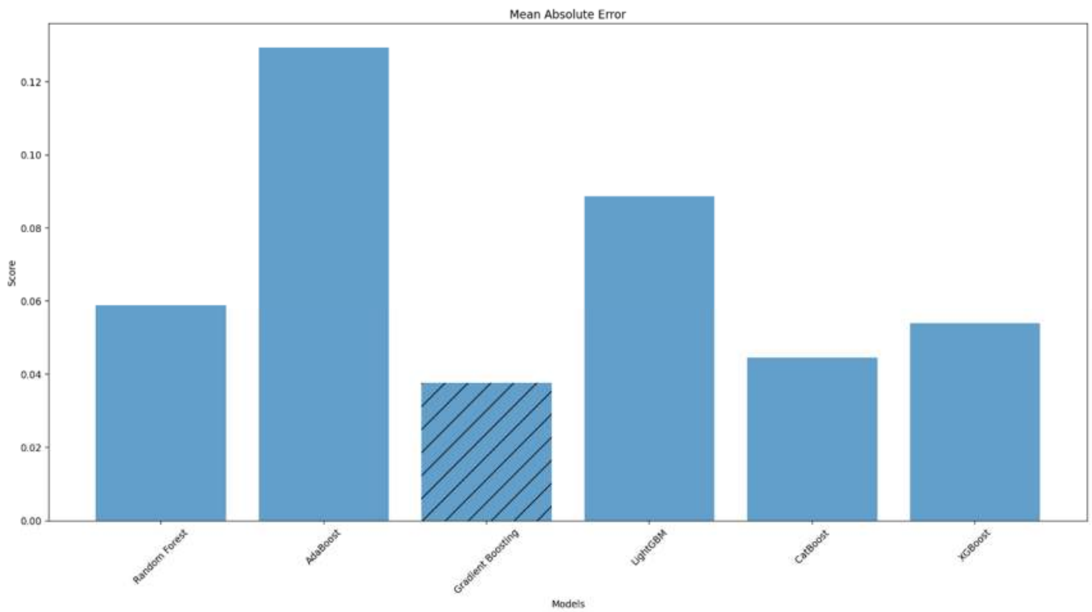

Figure 13: Comparison of Mean Absolute Error (MAE)

4.2 MSE score

Figure 14 illustrates the comparison of MSE scores among the proposed models. GB achieves the lowest MSE of 0.0048, followed by CatBoost with an MSE of 0.0052 and XGBoost with an MSE of 0.0071. RF records an MSE of 0.0096. AdaBoost shows an MSE of 0.0289, indicating higher sensitivity to larger errors. Respectively, while LightGBM has the highest MSE at 0.0366.

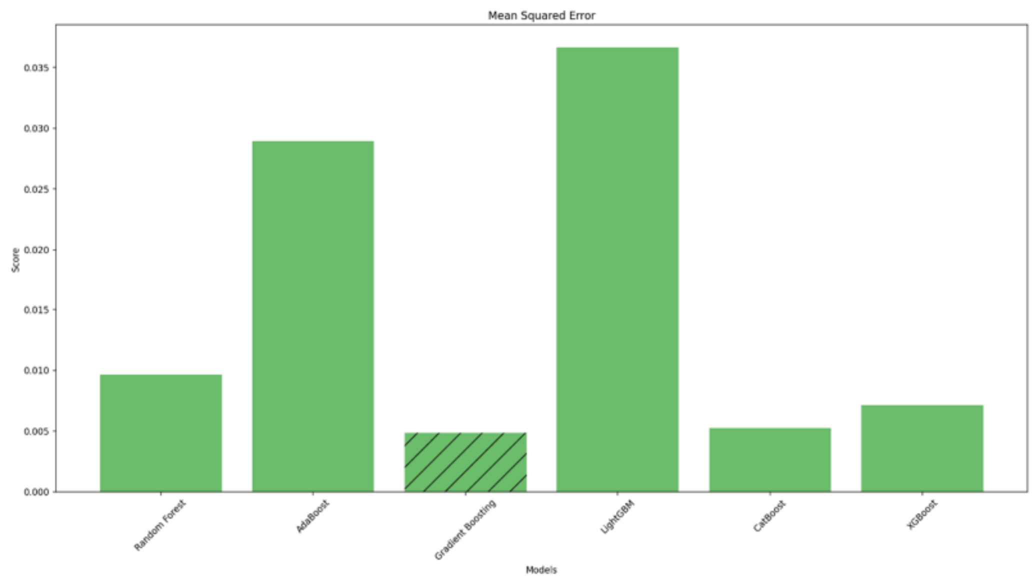

Figure 14: Comparison of Mean Squared Error (MSE)

4.3 MAPE score

Figure 15 illustrates the comparison of MAPE scores among the proposed models. The CatBoost model achieves the lowest MAPE of 43.41%, followed by XGBoost with an MAPE of 48.99% and RF with an MAPE of 51.33%. GB records an MAPE of 63.75%. LightGBM shows an MAPE of 156.92%, indicating higher sensitivity to larger errors. Respectively, while AdaBoost has the highest MAPE at 316.63%.

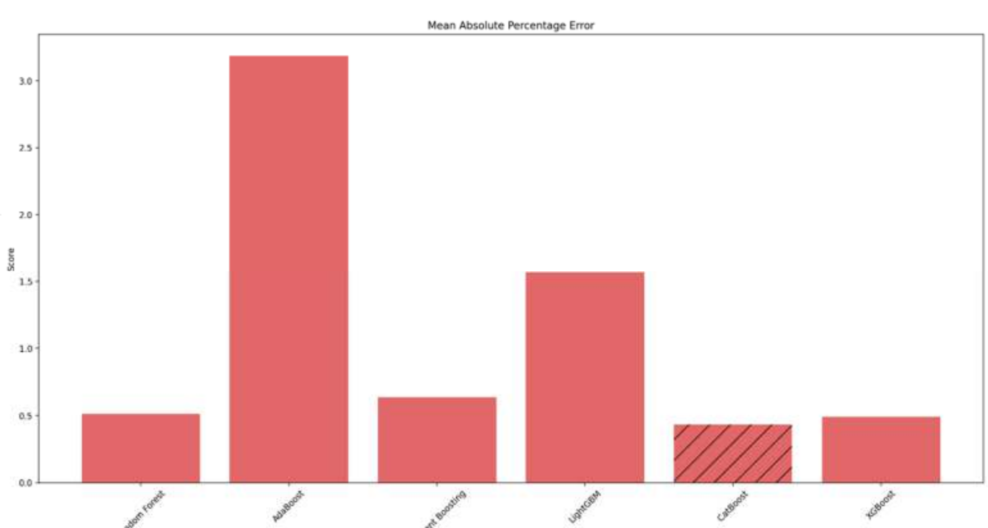

Figure 15: Comparison of Mean Absolute Percentage Error (MAPE)

4.4 MeAE score

Figure 16 illustrates the comparison of MeAE scores among the proposed models. GB achieves the lowest MeAE of 0.0161, followed by RF with an MeAE of 0.0258 and CatBoost with an MeAE of 0.0278. XGBoost records an MeAE of 0.0329. LightGBM shows an MeAE of 0.0394, indicating higher sensitivity to larger errors. Respectively, while AdaBoost has the highest MeAE at 0.0957.

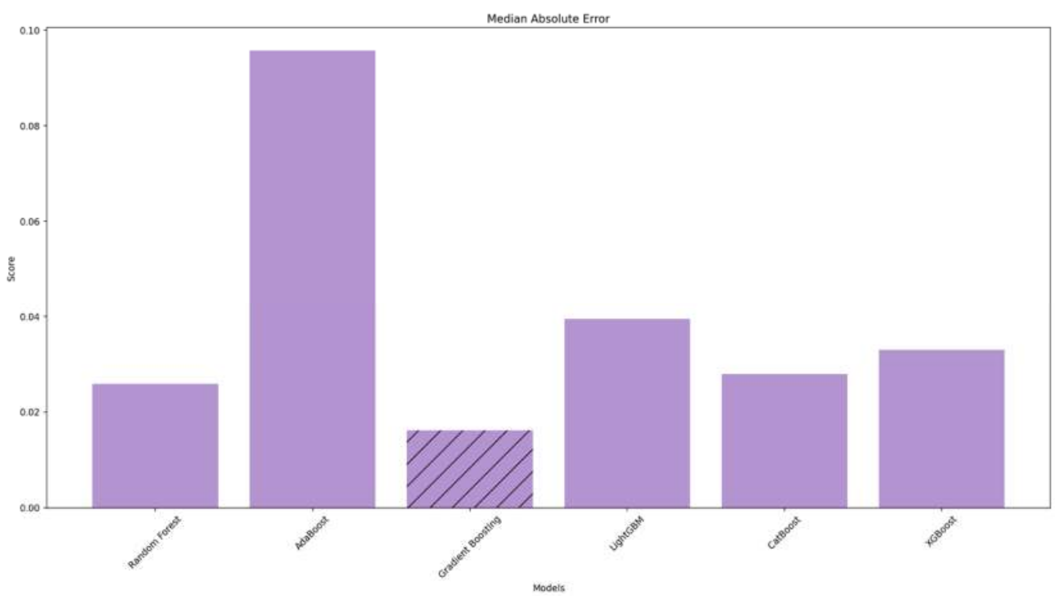

Figure 17: Comparison of Median Absolute Error (MeAE)

4.5 R^2 score

Figure 17 illustrates the comparison of R^2 scores among the proposed models. GB has the highest R^2 score of 0.9943, indicating excellent explanatory power. CatBoost follows with an R^2 of 0.9917, and XGBoost records an R^2 of 0.9892. RF achieves an R^2 of 0.9850. Respectively, AdaBoost shows an R^2 at 0.9523 and LightGBM has the lowest R^2 at 0.9608.

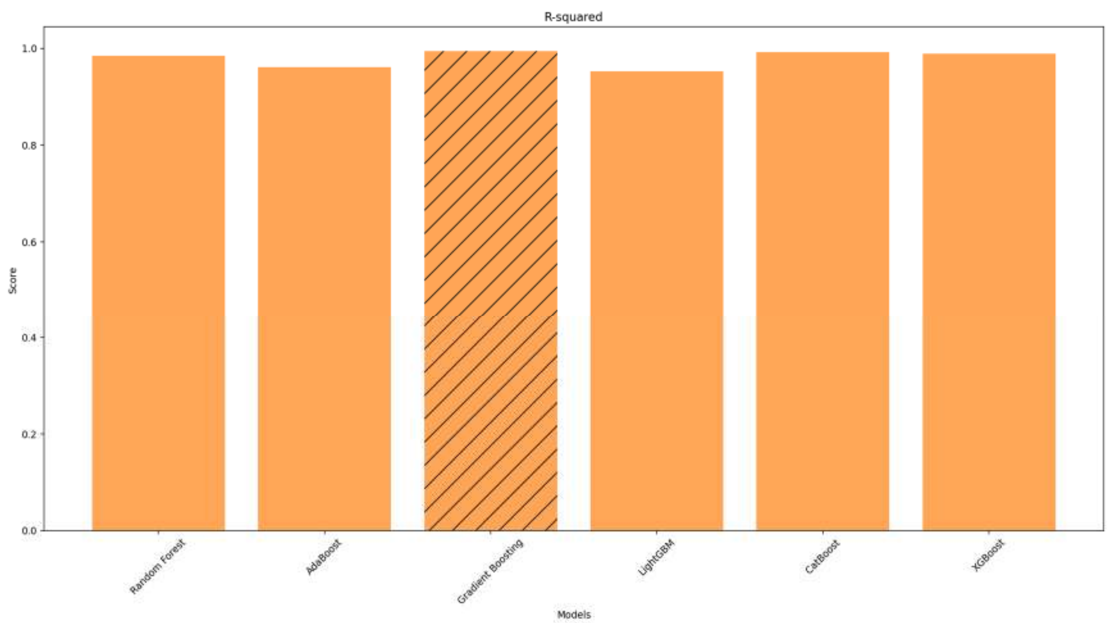

Figure 18: Comparison of R-squared (R^2) Score

Upon evaluating the models across five performance metrics, it is evident that GB outperforms all other models utilized in this study, consistently demonstrating superior performance across all evaluation metrics. CatBoost and XGBoost also exhibit strong performance.

4.6 Significance of the findings

Integrating advanced technology in Bangladesh's agricultural sector, specifically in optimizing the rice supply chain, holds significant importance (Jamal et al., 2023). Therefore, utilizing state-of-art machine learning techniques as well as techniques related to ensemble learning, this study targets to improve the accuracy of rice yield and production that plays an essential role in efficient utilization of resources and overall planning and formulation of agricultural and development policies. Since agriculture, especially rice production, forms the basis of Bangladesh's economic, this proposal can enhance the farmers' profit status, rice price, and supply, which can be good news to both buyers and sellers of rice (Mamun et al., 2021). It also manages different supply chain issues like fluctuating weather patterns, the unavailability of advanced equipment and agricultural technology, and inadequate supply logistics using a more progressive ensemble learning method to build a more efficient supply chain system (Kumar & Agrawal, 2023). In this study, GB is found to perform best in the prediction of the rice yield and production as seen in the assessments of the accuracy of the various models and their performances. By employing ensemble learning, this model is very effective in solving the numerous factors that influence agricultural productivity since it amalgamates the predictions from multiple weak learners known to develop a more robust model. From this model, it is possible to come up with a consolidated rice yield and production forecast over a given period depending on the data on crop yield per hectare, average temperatures, production area, and general information for a specific region. Such forecasts can be put into application in order to produce sound decisions, efficiently utilize resources, and thus enable farmers to adopt suitable agricultural policies, therefore enhancing production and returns. Furthermore, the study relates to the production of other research works in the field of supply chain enhancement and predictive modeling and propels a similar approach in other agricultural segments and other worlds. In addition to these economic impacts, socio-economic impacts are also immense as increased efficiency and productivity in agriculture positively affects food security, poverty diminishing, and welfare improvement of Bangladeshi rural people (Rahman & Anik, 2020). Therefore, this study proves the technology can be a powerful tool in agriculture, and provides the solution to some real-life issues, leading to the development of agriculture in Bangladesh.

5. Conclusion

This study comprehensively investigates the potential of ensemble learning techniques to optimize the rice supply chain in Bangladesh. By leveraging historical data and employing advanced machine learning algorithms, the research demonstrates significant improvements in predictive accuracy and operational efficiency. Specifically, the models—RF, AdaBoost, GB, LightGBM, CatBoost, and XGBoost—have been meticulously analyzed to provide insights into their performance in predicting rice yield and production. The findings indicate that ensemble learning methods, particularly GB, CatBoost, and XGBoost, exhibit superior performance across multiple evaluation metrics. Among these, Gradient Boosting stands out as the best-performing model, achieving an MAE of 0.0376, an MSE of 0.0048, a MAPE of 0.6375, a MeAE of 0.0161, and an R^2 score of 0.9943. This model consistently outperforms others, showcasing its robustness in handling the complexities of agricultural data in Bangladesh. The study furthermore highlights the importance of accurate forecasting in enhancing agricultural planning and policymaking, which in turn can lead to better resource allocation and more effective agricultural strategies. The study also points out the areas of technological imperatives in helping to overcome the problems encountered in the agriculture industry in Bangladesh. If the rice supply chain management is enhanced, the farmers, distributors, and ultimately the consumers can gain tremendously through better decisions and decision-making tools that maximize the profit possible and, also regulate the prices of rice while at the same time ensuring a steady supply. All in all, this research not only enriches the context of supply chain optimization and future forecasting but also establishes a reference case for the use of machine learning methods in developing nations. Speaking about the contribution of modern approaches of predictive models' integration into the agricultural field seen as having the potential to revolutionize conventional farming, promote sustained economic growth and guarantee food security in Bangladesh.

References

1. Abdur Noor Tuser, Md. Rahat Hossain, Pranta Barua, Maliha Khan Puspita, & Ifat Sadek Pushpo. (2023). A Scenario of Adopting AI Technology in Supply Chain: A Study of E-Commerce in Bangladesh. Supply Chain Insider | ISSN: 2617-7420 (Print), 2617-7420 (Online), 10(1).
2. Abegunde, V. O., Sibanda, M., & Obi, A. (2019). The dynamics of climate change adaptation in Sub-Saharan Africa: A review of climate-smart agriculture among small-scale farmers. Climate, 7(11), 132.
3. Abolghasemi, M. (2023). The intersection of machine learning with forecasting and optimisation: theory and applications. In Forecasting with Artificial Intelligence: Theory and Applications (pp. 313-339). Cham: Springer Nature Switzerland.
4. Abolghasemi, M., Tarr, G., & Bergmeir, C. (2022). Machine learning applications in hierarchical time series forecasting: Investigating the impact of promotions. International Journal of Forecasting.
5. Afjal Hossain, M., Imran Reza, M., Rahman, S., & Kayes, I. (2012). Climate change and its impacts on the livelihoods of the vulnerable people in the southwestern coastal zone in Bangladesh. Climate change and the sustainable use of water resources, 237-259.
6. Akpan, A. I., & Zikos, D. (2023). Rural Agriculture and Poverty Trap: Can Climate-Smart Innovations Provide Breakeven Solutions to Smallholder Farmers?. Environments, 10(4), 57.
7. Al Mamun, M. A., Nihad, S. A. I., Sarkar, M. A. R., Aziz, M. A., Qayum, M. A., Ahmed, R., ... & Kabir, M. S. (2021). Growth and trend analysis of area, production and yield of rice: A scenario of rice security in Bangladesh. PloS one, 16(12), e0261128.
8. Alabi, M. O., & Ngwenyama, O. (2023). Food security and disruptions of the global food supply chains during COVID-19: building smarter food supply chains for post COVID-19 era. British Food Journal, 125(1), 167-185.

9. Alam, G. M., Hoque, K. E., Khalifa, M. T. B., Siraj, S. B., & Ghani, M. F. B. A. (2009). The role of agriculture education and training on agriculture economics and national development of Bangladesh. African Journal of Agricultural Research, 4(12), 1334-1350.
10. Ali, A., Xia, C., Mahmood, I., & Faisal, M. (2021). Economic and environmental consequences' of postharvest loss across food supply Chain in the developing countries. Journal of cleaner production, 323, 129146.
11. Alvee Jawad Bin Hamid, M. Abrar Faiyaz Rabbi, & Syeda Tahsina Hasan. (2023). Implementation of The Kanban and 5S Methodology in The Upstream Supply Chain of The Jute Industry: A Case Study in Bangladesh. Supply Chain Insider | ISSN: 2617-7420 (Print), 2617-7420 (Online), 10(1).
12. Ankon Dey, Jawata Afnan, Naymul Hasan Nayem, Arindam Dey, & Fariha Tanha Tonu. (2023). Supply Chain Modelling of Daily Consumed Cereal Grain in Bangladesh. Supply Chain Insider | ISSN: 2617-7420 (Print), 2617-7420 (Online), 11(1).
13. Arshad, M. F., Burrai, G. P., Varcasia, A., Sini, M. F., Ahmed, F., Lai, G., ... & Parpaglia, M. L. P. (2024). The groundbreaking impact of digitalization and artificial intelligence in sheep farming. Research in Veterinary Science, 105197.
14. Azadi, Z. (2018). Stochastic optimization models for perishable products (Doctoral dissertation, Clemson University).
15. Bala, B. K., Bhuiyan, M. G. K., Alam, M. M., Arshad, F. M., Sidique, S. F., & Alias, E. F. (2017). Modelling of supply chain of rice in Bangladesh. International Journal of Systems Science: Operations & Logistics, 4(2), 181-197.
16. Bandumula, N. (2018). Rice production in Asia: Key to global food security. Proceedings of the National Academy of Sciences, India Section B: Biological Sciences, 88, 1323-1328.
17. Begum, T. (2023). Predictive analytics for machine learning and deep learning. In Handbook of Big Data Research Methods (pp. 148-164). Edward Elgar Publishing.
18. Bernard, S., Adam, S., & Heutte, L. (2012). Dynamic random forests. Pattern Recognition Letters, 33(12), 1580-1586.

19. Condran, S., Bewong, M., Islam, M. Z., Maphosa, L., & Zheng, L. (2022). Machine learning in precision agriculture: a survey on trends, applications and evaluations over two decades. IEEE Access, 10, 73786-73803.
20. Dahikar, S. S., & Rode, S. V. (2014). Agricultural crop yield prediction using artificial neural network approach. International journal of innovative research in electrical, electronics, instrumentation and control engineering, 2(1), 683-686.
21. Dey, P., & Chatterjee, K. (2023). Application of Machine Learning for Optimization. Handbook of Research on AI and Machine Learning Applications in Customer Support and Analytics, 113-127.
22. Dhall, D., Kaur, R., & Juneja, M. (2020). Machine learning: a review of the algorithms and its applications. Proceedings of ICRIC 2019: Recent innovations in computing, 47-63.
23. Ding, Y., Zhu, H., Chen, R., & Li, R. (2022). An efficient AdaBoost algorithm with the multiple thresholds classification. Applied sciences, 12(12), 5872.
24. Drahokoupil, J. (2022). Application of the XGBoost algorithm and Bayesian optimization for the Bitcoin price prediction during the COVID-19 period. FFA Working Papers.
25. Farooq, M. S., Riaz, S., Abid, A., Umer, T., & Zikria, Y. B. (2020). Role of IoT technology in agriculture: A systematic literature review. Electronics, 9(2), 319.
26. Foysal Ahmed, Sohaib Khan, & Fatema Akter Akhi. (2023). Integrating the Internet of Everything and Artificial Intelligence in Reverse Logistics to Prepare for Industrial Revolution 5.0. Supply Chain Insider | ISSN: 2617-7420 (Print), 2617-7420 (Online), 10(1).
27. Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., & Suganthan, P. N. (2022). Ensemble deep learning: A review. Engineering Applications of Artificial Intelligence, 115, 105151.
28. Gazi, M. A. I. (2020). Supply chain management for agro products in Bangladesh; logistics support for capturing market by ensuring balanced distribution. International Journal of Management, Accounting and Economics, 7(6), 277-297.

29. Gumma, M. K., Thenkabail, P. S., Maunahan, A., Islam, S., & Nelson, A. (2014). Mapping seasonal rice cropland extent and area in the high cropping intensity environment of Bangladesh using MODIS 500 m data for the year 2010. ISPRS journal of photogrammetry and remote sensing, 91, 98-113.
30. Hadeed, S. J., O'rourke, M. K., Burgess, J. L., Harris, R. B., & Canales, R. A. (2020). Imputation methods for addressing missing data in short-term monitoring of air pollutants. Science of the Total Environment, 730, 139140.
31. Hajihosseinlou, M., Maghsoudi, A., & Ghezelbash, R. (2023). A novel scheme for mapping of MVT-type Pb–Zn prospectivity: LightGBM, a highly efficient gradient boosting decision tree machine learning algorithm. Natural Resources Research, 32(6), 2417-2438.
32. Jamal, M. R., Kristiansen, P., Kabir, M. J., & Lobry de Bruyn, L. (2023). Challenges and adaptations for resilient rice production under changing environments in Bangladesh. Land, 12(6), 1217.
33. Juan, A. A., Rabe, M., Goldsman, D., & Faulín Fajardo, J. (2021). Simulation-optimization in logistics, transportation, and SCM. MDPI.
34. Khan, N. A. (2021). Climate Change and Sustainability of Agriculture in Bangladesh. Climate Change in Bangladesh: A Cross-Disciplinary Framework, 65-84.
35. Khatri, P., Kumar, P., Shakya, K. S., Kirlas, M. C., & Tiwari, K. K. (2023). Understanding the intertwined nature of rising multiple risks in modern agriculture and food system. Environment, Development and Sustainability, 1-44.
36. Kiangala, S. K., & Wang, Z. (2021). An effective adaptive customization framework for small manufacturing plants using extreme gradient boosting-XGBoost and random forest ensemble learning algorithms in an Industry 4.0 environment. Machine Learning with Applications, 4, 100024.
37. Kumar, A., & Agrawal, S. (2023). Challenges and opportunities for agri-fresh food supply chain management in India. Computers and Electronics in Agriculture, 212, 108161.

38. Kumar, P., Singh, R. K., & Shahgholian, A. (2024). Learnings from COVID-19 for managing humanitarian supply chains: systematic literature review and future research directions. Annals of Operations Research, 335(3), 899-935.
39. Kurek, J., Niedbała, G., Wojciechowski, T., Świderski, B., Antoniuk, I., Piekutowska, M., ... & Bobran, K. (2023). Prediction of Potato (Solanum tuberosum L.) Yield Based on Machine Learning Methods. Agriculture, 13(12), 2259.
40. Mainuddin, M., Alam, M. M., Maniruzzaman, M., Kabir, M. J., Mojid, M. A., Hasan, M. M., ... & Islam, M. T. (2021). Yield, profitability, and prospects of irrigated Boro rice cultivation in the North-West region of Bangladesh. PloS one, 16(4), e0250897.
41. Maraveas, C., Piromalis, D., Arvanitis, K. G., Bartzanas, T., & Loukatos, D. (2022). Applications of IoT for optimized greenhouse environment and resources management. Computers and Electronics in Agriculture, 198, 106993.
42. Mendonça, M. O., Netto, S. L., Diniz, P. S., & Theodoridis, S. (2024). machine learning: Review and trends. Signal Processing and Machine Learning Theory, 869-959.
43. Mhlanga, D. (2021). Artificial intelligence in the industry 4.0, and its impact on poverty, innovation, infrastructure development, and the sustainable development goals: Lessons from emerging economies?. Sustainability, 13(11), 5788.
44. Moon, S., Lee, S., Jeon, W., & Park, K. J. (2023). Learning-Enabled Network-Control Co-Design for Energy-Efficient Industrial Internet of Things. IEEE Transactions on Network and Service Management.
45. Mu, J., Li, J., Li, Y., & Liu, C. (2021). The dynamics of brand-driven quality improvement decision-making in multi-small-supplier agri-food supply chain: The case of china. Sustainability, 13(19), 10815.
46. Nti, I. K., Zaman, A., Nyarko-Boateng, O., Adekoya, A. F., & Keyeremeh, F. (2023). A predictive analytics model for crop suitability and productivity with tree-based ensemble learning. Decision Analytics Journal, 8, 100311.
47. Pasupuleti, V., Thuraka, B., Kodete, C. S., & Malisetty, S. (2024). Enhancing supply chain agility and sustainability through machine learning: Optimization techniques for logistics and inventory management. Logistics, 8(3), 73.

48. Plaia, A., Buscemi, S., Fürnkranz, J., & Mencía, E. L. (2022). Comparing boosting and bagging for decision trees of rankings. Journal of Classification, 39(1), 78-99.
49. Rahman, M. T. (2017). Role of agriculture in Bangladesh economy: uncovering the problems and challenges. International Journal of Business and Management Invention, 6(7).
50. Rahman, S., & Anik, A. R. (2020). Productivity and efficiency impact of climate change and agroecology on Bangladesh agriculture. Land Use Policy, 94, 104507.
51. Rasel Shikder, Md. Zumman Siddique, Esrak Fardin Ratul, & Nazifa Tabassum. (2022). A Roadmap for the Implementation of Blockchain Technology throughout the Rice Supply chain in Bangladesh. Supply Chain Insider | ISSN: 2617-7420 (Print), 2617-7420 (Online), 8(1).
52. Ruane, A. C., Major, D. C., Winston, H. Y., Alam, M., Hussain, S. G., Khan, A. S., ... & Rosenzweig, C. (2013). Multi-factor impact analysis of agricultural production in Bangladesh with climate change. Global environmental change, 23(1), 338-350.
53. Saber, M., Boulmaiz, T., Guermoui, M., Abdrabo, K. I., Kantoush, S. A., Sumi, T., ... & Mabrouk, E. (2022). Examining LightGBM and CatBoost models for wadi flash flood susceptibility prediction. Geocarto International, 37(25), 7462-7487.
54. Saiz-Rubio, V., & Rovira-Más, F. (2020). From smart farming towards agriculture 5.0: A review on crop data management. Agronomy, 10(2), 207.
55. Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. SN computer science, 2(3), 160.
56. Schwenk, H., & Bengio, Y. (2000). Boosting neural networks. Neural computation, 12(8), 1869-1887.
57. Shakoor, M. T., Rahman, K., Rayta, S. N., & Chakrabarty, A. (2017, July). Agricultural production output prediction using supervised machine learning techniques. In 2017 1st international conference on next generation computing applications (NextComp) (pp. 182-187). IEEE.
58. Sharma, R., Kamble, S. S., Gunasekaran, A., Kumar, V., & Kumar, A. (2020). A systematic literature review on machine learning applications for sustainable agriculture supply chain performance. Computers & Operations Research, 119, 104926.

59. Shomoye, M. A. (2024). Exploring Emotion Recognition of Students in Virtual Reality Classrooms Through Convolutional Neural Networks and Transfer Learning Techniques.
60. Sultana, N., & Tamanna, M. (2021). Exploring the benefits and challenges of Internet of Things (IoT) during Covid-19: a case study of Bangladesh. Discover Internet of Things, 1(1), 20.
61. Ukhurebor, K. E., Adetunji, C. O., Olugbemi, O. T., Nwankwo, W., Olayinka, A. S., Umezuruike, C., & Hefft, D. I. (2022). Precision agriculture: Weather forecasting for future farming. In Ai, edge and iot-based smart agriculture (pp. 101-121). Academic Press.
62. Wang, J., Bretz, M., Dewan, M. A. A., & Delavar, M. A. (2022). Machine learning in modelling land-use and land cover-change (LULCC): Current status, challenges and prospects. Science of The Total Environment, 822, 153559.
63. Weiss, M., Jacob, F., & Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. Remote sensing of environment, 236, 111402.
64. Yang, H., Liu, X., & Song, K. (2022). A novel gradient boosting regression tree technique optimized by improved sparrow search algorithm for predicting TBM penetration rate. Arabian Journal of Geosciences, 15(6), 461.
65. Zounemat-Kermani, M., Batelaan, O., Fadaee, M., & Hinkelmann, R. (2021). Ensemble machine learning paradigms in hydrology: A review. Journal of Hydrology, 598, 126266.