Applying machine learning approach to predict annual yield of major crops and recommend planting different crops in different seasons in Bangladesh

BY

Arminara Jemi
ID: 191-15-2442
AND

Faysal Miah
ID: 191-15-2492

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

Supervised By

Mr. Amit Chakraborty Chhoton
Senior Lecturer
Department of CSE
Daffodil International University

Co-Supervised By

Mr. Md. Sabab Zulfiker
Senior Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2023

APPROVAL

This Project titled **"Applying machine learning approach to predict annual yield of major crops and recommend planting different crops in different seasons in Bangladesh"**, submitted by Arminara Jemi ID: 191-15-2442, and Faysal Miah ID: 191-15-2492 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfilment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held *on 25 January 2023*

BOARD OF EXAMINERS

Chairman

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

Internal Examiner

Dr. Md. Atiqur Rahman
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Shayla Sharmin 25.1.23
Shayla Sharmin
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

Farid 25-01-23
Dr. Dewan Md Farid
Professor
Department of Computer Science and Engineering
United International University

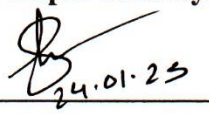
DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Mr.Amit ChakrabortyChhoton,Senior Lecturer, Department of CSE**, Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Mr.Amit ChakrabortyChhoton
Senior Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:



Mr.Md.SababZulfiker
Senior Lecturer
Department of CSE
Daffodil International University

Submitted by:

Arminara Jemi
ID: -19-15-2442
Department of CSE
DaffodilInternationalUniversity

Faysal Miah
ID: -19-15-2492
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

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

We really grateful and wish our profound our indebtedness to **Mr.Amit Chakraborty Chhoton**, **Senior Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "**Machine Learning**" to carry out this project. His endless patience ,scholarly guidance ,continual encouragement , constant and energetic supervision, constructive criticism , valuable advice ,reading many inferior draft and correcting them at all stage have made it possible to complete this project.

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

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

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

ABSTRACT

The Bangladesh economy heavily depends on agriculture. Bangladesh's agricultural sector is crucial for providing jobs, income, and GDP. Considering how dramatically the human population is growing, crop output is the primary factor in determining food security. In this study machine learning is used to predict Annual yield of major crops and recommend planting different crops in different seasons which are mostly cultivated all over Bangladesh. For getting the best accuracy, this study uses Decision tree, Random forest (RF), Support Vector Machine (SVM), Adaboost Classifier (ADB), KNN, Logistic regression (LOR), and the Naive Bayes (NB) algorithm. Algorithms for machine learning are used to analyze four most planted yields in Bangladesh. Those crops include: Rice (Aman, Aus, Boro), Potato, jute and wheat.

Keywords: Decision tree, Random forests, Support Vector Machines, Adaboost Classifier (ADB), KNN, Logistic regression, and the Naive Bayes, Bangladesh economy, Machine learning.

TABLE OF CONTENTS

FIGURES NAME **PAGE**

LIST OF TABLE PAGE

CHAPTER 1
Introduction

1.1Overview

Bangladesh is mostly an agricultural nation. The ground is quite fertile, and the climate is perfect, so many different kinds of crops flourish here. More than 45 percent of the workforce is employed in the agriculture industry, which makes for around 17 percent of the nation's GDP. Our country has more than 70% of its land used for agriculture [11]. For this paper we collect 4 types of crops. There are 3 types of rice (Aman, Aus, Boro), potato, jute and wheat. Seasonal variations interfere with crop planting and harvesting dates, which have a substantial impact on crop yield. Bangladesh has three seasons. Those are summer (March to May), the rainy season (June to October), and winter (November to February) (December to February). While Aman can get by with just the monsoon rain, Aus rice needs irrigation in order to thrive past the first plantation stage. Though it thrives in relatively dry winter and late summer, Boro needs a lot of water for its growth and development. In Bangladesh, the majority of kinds are harvested between February and March. In Bangladesh, the tubers are typically collected manually using a spade or other tools. The potato is the most significant food crop in the world in terms of fresh product volume. The rainy season is when jute is grown. Depending on the species, sowing in Bangladesh often begins at the end of February and lasts until the end of May. Pre-monsoon showers and moist conditions are crucial for cultivation. March and April were harvest months for wheat. Farmers put forth a lot of effort to improve our economic growth, working long hours and under a lot of stress to make our nation a better place to live. However, the farmers' current methods of irrigation and farming are still out of date and falling behind. To find patterns in the data, we train a model with the machine learning models [2]. It assists farmers in cultivating the best crops possible for their farms [1]. Machine learning has completely transformed the agricultural industry. High-performance computing, big data and a component of artificial intelligence known as machine learning has emerged to open up new possibilities for data-intensive science in the multidisciplinary field of agritechnology. In agriculture, for instance, machine learning is not a hidden technique or a magic trick; rather, it is a collection of well-defined models that gather specific data and use particular techniques to create desired the results [12].

The purpose of this paper is to forecast annual yields of major crops and recommend different crops for planting in different seasons in Bangladesh. We can also see the visual pattern of major crops each year.

- A variety of techniques were used to pre-process the dataset., including encoding categorical data and addressing imbalanced data.
- We are using the traditional machine learning algorithms (LoR, SVM, ADB, NB, DT, RF, KNN) with better accuracy.
- In order to assign the optimum settings and increase accuracy, we have incorporated hyperparameter adjustment.
- We utilized ensemble learning classification models to optimize in addition to making sure real-world performance.
- We defined the results of numerous analysis methods in order to compare our efficiency to that of others.
- We used different parameters for accurate output.
- We used supervised classifier for better prediction.
- We used voting classifier. A voting classifier is a type of machine learning predictor that develops a number of training sets or predictor variables and makes predictions based on averaging their results.
- We also find out run time of every different classifiers.

1.2. Motivation

Humans depend on agriculture to meet their basic requirements for things like food, clothes, and shelter. Climate change has had a significant impact on crop performance in previous decades in the field of agriculture. From an agricultural standpoint, there is a lot of manual labor for farmers to do. With the help of Agriculture can combat poorness, improve incomes, and increase food security for 80% of the world's poor people who live in rural areas and mostly work in agriculture. Furthermore, for agricultural production, proper planning of the what, how, why, and where to produce crops or livestock is required. Planning ensures that available resources, labor, and time are used efficiently and effectively. This projection can help farmers decide what plant and when to plant it. Agricultural Officers provide detailed instructions for growing seasonal crops. Farmers can easily obtain seasonal seeds from the government.

This work is an effective strategy for higher-quality farming and higher profits. In the realm of machine learning, using a data clustering method to extract relevant information and provide predictions is an effective strategy. This prediction sparked the development of new crops that can grow in any season. We know that Bangladesh is an agricultural country where most of the people are farmers. Our country provides various types of production of crops in the world. Increasing productivity of crops can help Bangladeshi farmers to spend their livelihood with proper ways. This Paper helps us to gain proper knowledge about seasonal major crops cultivation. Proper knowledge of cultivation can grow our crops productivity.

1.3. Research Objective

- To determine the crop prediction outcome
- Evaluate the related issue
- Benefit of the prediction of normal people

1.4. Research Question

- What exactly seer database?
- What are the main characteristics of this database?
- How does the algorithm work in this study?
- How might you anticipate the early detection of seasonal crops?
- What will be the main purpose of this research?

1.5. Report Layout

- Literature review
- Methodology
- Result
- Conclusion and Future work
- References

CHAPTER 2

Literature Review

Based on machine learning, this study proposes a framework for agricultural yield prediction. SVM, Random Forest, and ID3 are the machine learning methods employed for the inquiry. The SVM approach offers the highest accuracy and lowest error **[1]**. In this paper, machine learning techniques are used to train a model to find out the patterns in data, which is then used to predict crop yield. Machine learning is used in this study to forecast yields of India's four most widely grown crops. These crops include wheat, potatoes, paddy rice, and corn. The decision tree regressor achieved the maximum score of 96% for fitting observed data **[2].** In these case studies, they forecasted yield at the regional level for five commodities, three nations, and France (FR). To confirm the explainability of the features, they examined feature selection frequencies for each crop across many countries and algorithms. All findings were combined a national level, and then they were compared to previous MCYFS forecasts. In this section, they provide the measured RMSE for a variety of studies in practice **[3]**. In this study, in Pakistan's Punjab province wheat and rice yields were estimated using the crop-growing season's highest Enhanced Vegetation Index (EVI) combined with Machine Learning Regression models. The predicted accuracy of five MLR models was compared using a fivefold cross-validation procedure. The regression model based on the Gaussian process outperformed the other models, according to the study's findings. The best-performing wheat model had coefficients of determination (R2), mean absolute error (MAE), and root mean square error (RMSE, t/ha) values of 75%, 28%, and 23%, respectively, and the best rice model had values of 68%, 11%, and 0.91% **[4]**. In order to help farmers anticipate crop yield, this research suggests a workable and user-friendly approach using the machine learning techniques. Random Forest provided the most accurate findings, 95% of the time **[5]**. They present a machine learning-based prediction method for forecasting the annual yield of six different crops at the national level in West Africa, including rice, maize, cassava, seed cotton, yams, and bananas.

The correlation between the predicted outcomes of the decision tree model and the K-Nearest Neighbor model and the anticipated data demonstrates the model's effectiveness.**[6]**.

By utilizing machine learning, it is one of the most cutting-edge technologies in crop prediction, this research aids the novice farmer in a way that directs them for sowing the reasonable crops. In this precision agriculture, the Naive Bayes classifier, a Supervised learning technique, aids in building the most accurate and effective model to predict crop and production **[7].** This research focuses on developing a prediction of the model that could be applied to crop yield forecasting in the future. Random Forest's accuracy is 86.35%, whereas Support Vector Machine's accuracy is 73.75%, making it a good algorithm for classifying soil. SVM algorithm accuracy for yield prediction is 99.47%, while RF accuracy is 97.48%. SVM method is therefore effective for predicting agricultural yield **[8]**. The study focuses on applying neural network regression modeling to estimate various crops' yields. With 82% accuracy and only a small loss, the dependent variable was predicted using Artificial Neural Network and linear regression**[9].**

Two distinct Machine Learning (ML) techniques are suggested in this paper to analyze crop yield. Support Vector Regression and Linear Regression are two techniques that can be used to validate variable parameters in the prediction of continuous variables using 140 data points collected. **[10]**.

CHAPTER 3

Methodology

3.1Data Collection:

The most effective way to gather and measure data from many sources is through data collection.As reported by the Bangladesh Bureau of Statistics, six main crops grown in Bangladesh are Aus, Aman, Boro, Jute, Wheat, and Potato. We have used seven machine learning algorithms. Numerous characteristics of the data include Crop, Year, Area_(In_Acre),Area_(In_Lac_Hectare),Total_Production_(M.ton),Total_Production_(Lac_M.ton), Yield_rate_(M.ton_per_Hectare), Month_range and Season.

Season- Summer, Rainy Season, Winter.

The spin axis of the Earth is inclined with particular regard to its elliptical orbit. This is the source of the seasons. Summer is the time when the earth's axis points toward the sun. Only when Earth's axis is turned away from the sun, winter is postulated.

Year- 1970 to 2015

Here we can use 45 years of crop data. Bangladesh's agricultural and rural sectors are critical to the country's large, dense, and quickly developing nation 's long-term food security and sustainability. In a country with limited natural resources, rural areas are pursued intensively for the crop and allied sectors. This is the most comprehensive source of agricultural series data.

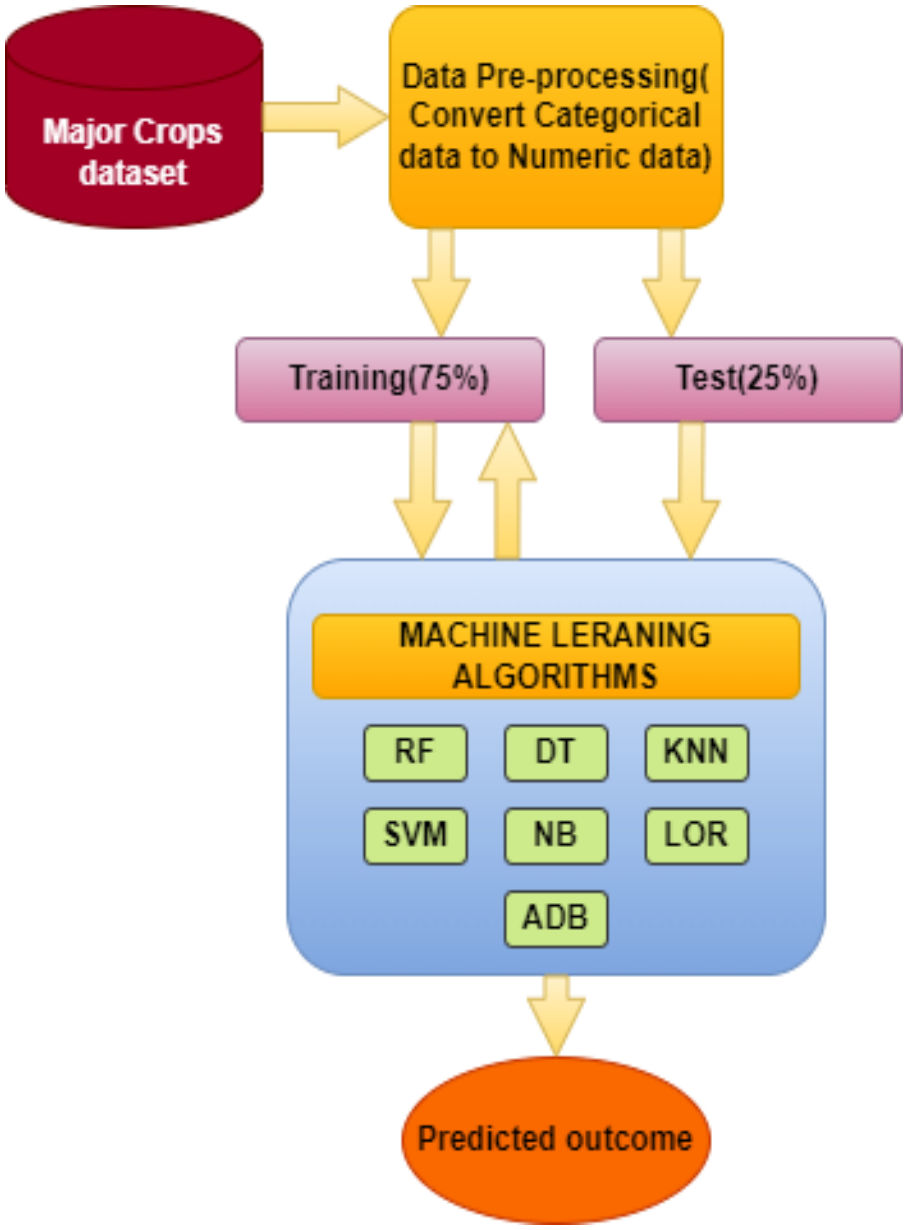


Figure 01. Overall Process Model

Figure 2 shows the correlation between crops and seasons. A resulting graph known as correlation expresses how closely two or more variables are related linearly. We know that different major crops are planted in different seasons in agricultural countries. The relationship between two different features in our dataset is shown in the diagram on the page. The degree of association between the attributes is shown in the right side bar. The stronger the association between two separate features, the value is higher and the darker in the bar.

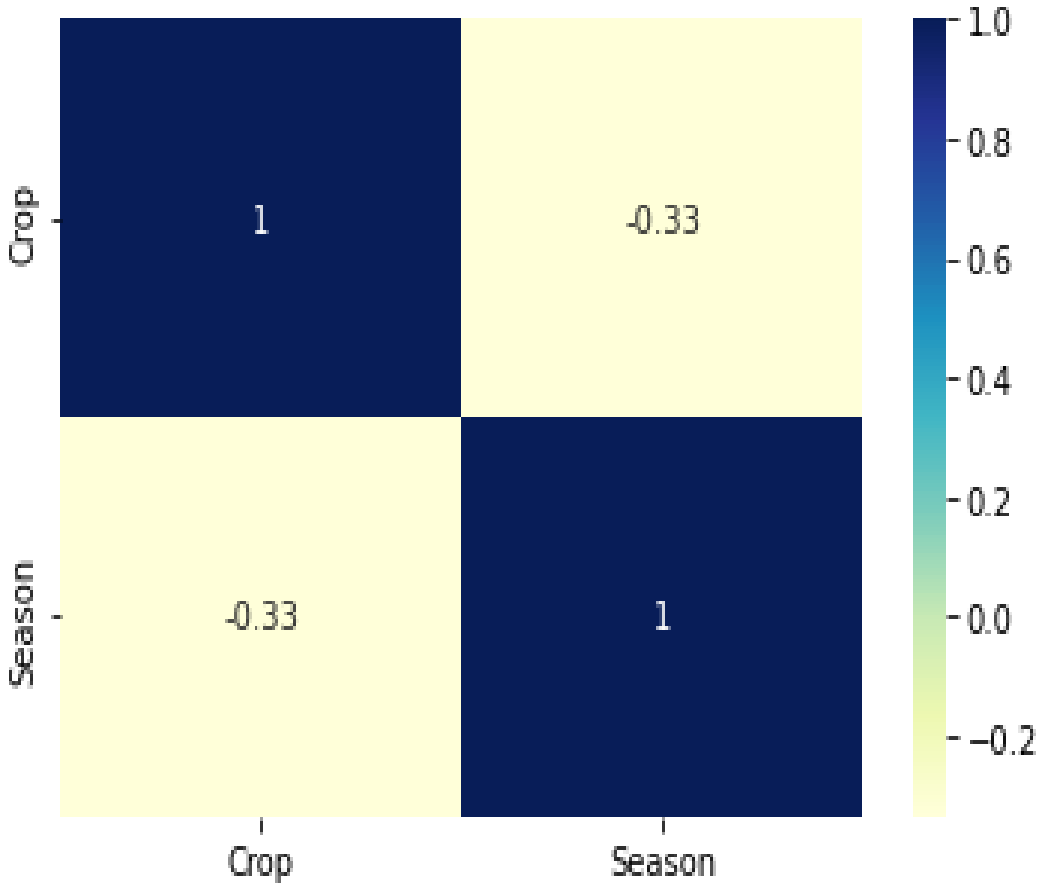


Figure 02. Correlation between season and crop

3.2Data Processing:

After collecting 270 different types of data, create a dataset. Before training and testing the model, the dataset needs to be preprocessed. Therefore, in order to improve accuracy, it is necessary to remove unnecessary attributes and fill up any missing values in datasets with undesirable nan values. Having to follow data cleansing or processing, our dataset will be divided into a training set and a test set using the sklearn package. 25% of the data are utilized for the test, 75% are used for training.

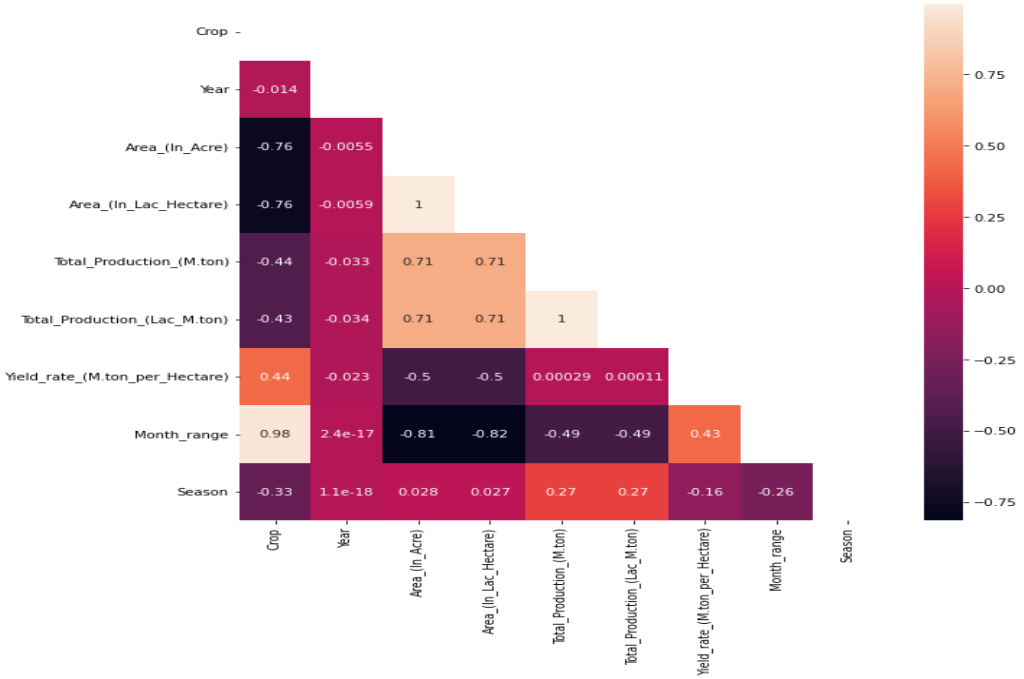


Figure03.Correlation Between Data

3.2.1. Categorical data encoding

The process of converting categorical data to a numerical value is known as encoding. Machins can not understand the categorical data. We are aware that the input and output variables for machine learning models must be numbers, and that if the dataset contains categorical data, it must first be converted to numbers before the model can be fitted. Our Crop dataset includes both the category and numerical data, and all columns contain String types of data. Thus, the dataset has been encoded.

3.2.2. Missing Value Imputation

The process of replacing missing data is known as missing value imputation. With imputed values that have been determined by research, including other dataset data. We know that machine learning algorithms are not support missing values.

Fortunately, there are no missing values in any of our datasets. That's why our useable machine learning algorithms can easily support our dataset.

3.2.3. Scaling Feature

Scaling is a feature in a machine learning model can accelerate the minimization of the cost function and make the gradient descent flow more fluid, which both benefit the optimization process. Absent scaling features, the system might favor characteristics with larger relates to increased. FS (Feature scaling) is a technique for normalizing the variety of distinct feature values. We have used Maximin scaler, which scales all data features between [0, 1] if there are no negative values and [-1, 1] otherwise.

3.2.4. Imbalanced Data Handling

In the domain of supervised machine learning including a number of classes, an unbalanced dataset is especially applicable. In our dataset we can face any problem for imbalanced data. That's why we can use any technique foe smote our data.

A pie plot is a type of visual illustration that uses a round graph to visualize information. It is a fixed composite chart that fits best when there are few elements. Pie charts are frequently used to display examples of information.

Figure 4 shows the majority of seasons when the major crops of our dataset can plant many more. Here we can see a pie plot which can separate 3 types of different color with different percentage. 50.0% declare the winter season. The second position of this pie plot is the rainy season which is 33.3%. And the last slice of this pie plot is 16.7% which is summer.

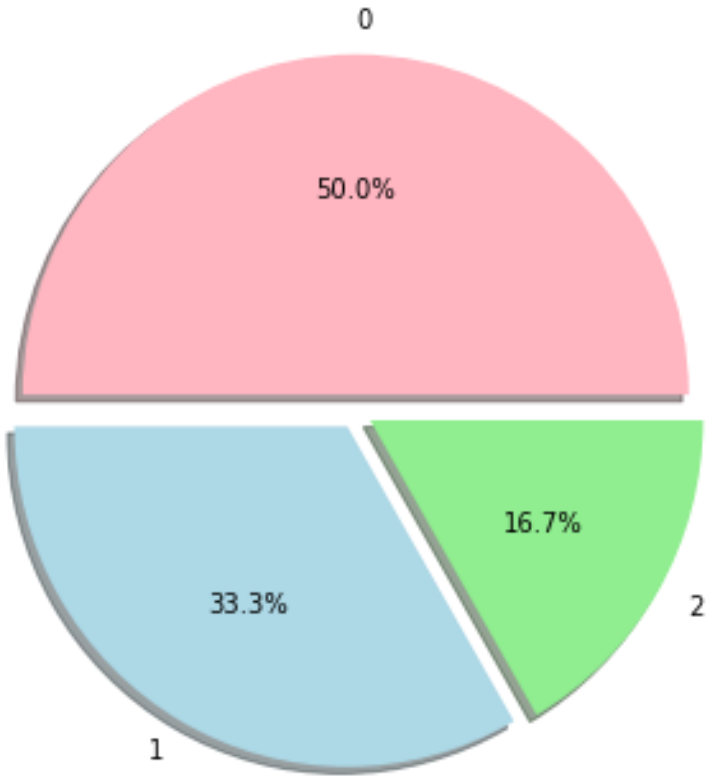


Figure 04. Majority of seasons

3.3Machine Learning Algorithm for Prediction:

Appropriate statistical construction is automated using machine learning, a technique for data analysis. It is a category of artificial intelligence built on the premise that machines can learn from data, spot connections, and make judgments with little help from people. Based on training data, the machine learning predictive algorithms have high efficiency projections of expected outcomes. A classifier in machine learning is a method which intelligently categorizes pieces of data into several types or groups. There seem to be two main models supervised and unsupervised within the classifier category. The classifiers are trained to distinguish between unlabeled and labeled data in the supervised model. As classification and regression algorithms, we used supervised machine learning algorithms in our paper.

Our paper will benefit most from the classification algorithm. Machine learning focuses committed toward developing tools which use information to enhance function of support forecasts. We used 7 types of different classification and regression algorithms for better accuracy. We know that machine learning is the field of prediction. Here is our algorithms:

- Random Forest (RF)
- K- Nearest Neighbours (KNN)
- SVM
- Decision Tree (DT)
- Naive Bayes (NB)
- Adaboost Classifier (ADB)
- Logistic regression (LoR)
- Voting Classifier

3.3.1. Random Forest (RF)

The Random forest is a reliable supervised machine learning technique for regression and classification tasks. It constructs decision trees from various data sets, using the major part of those samples for classification scheme and the score of those samples for regression. One of the most important characteristics of Random Forest Algorithm is its capability to handle data sets with both dependent variables, as in regression, and predictor data, as in classifiers. It produces better results when issues are classified.

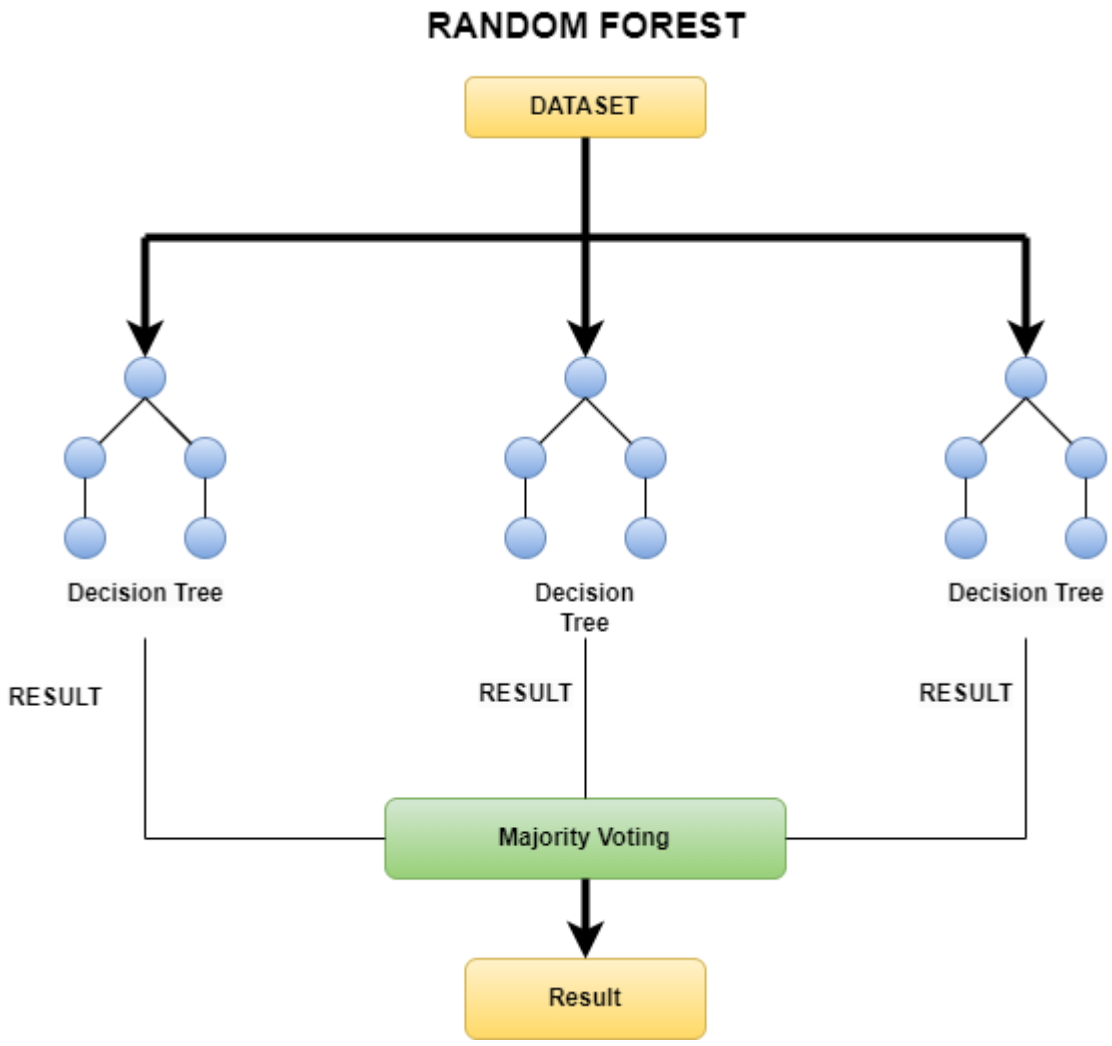


Fig:5. Random Forest Classifier

3.3.2. Decision Tree DT)

A decision tree, a supervised learning technique, can be used to solve the classification and regression problems. It is a tree-structured classifier with internal nodes that represent dataset features, branch offices for selection, and leaf nodes that represent classification results. A decision tree contains nodes such as the Decision Node and the Leaf Node. Decision nodes are utilized to make decisions and have many branches, whereas leaf nodes represent results and have no increased branches.

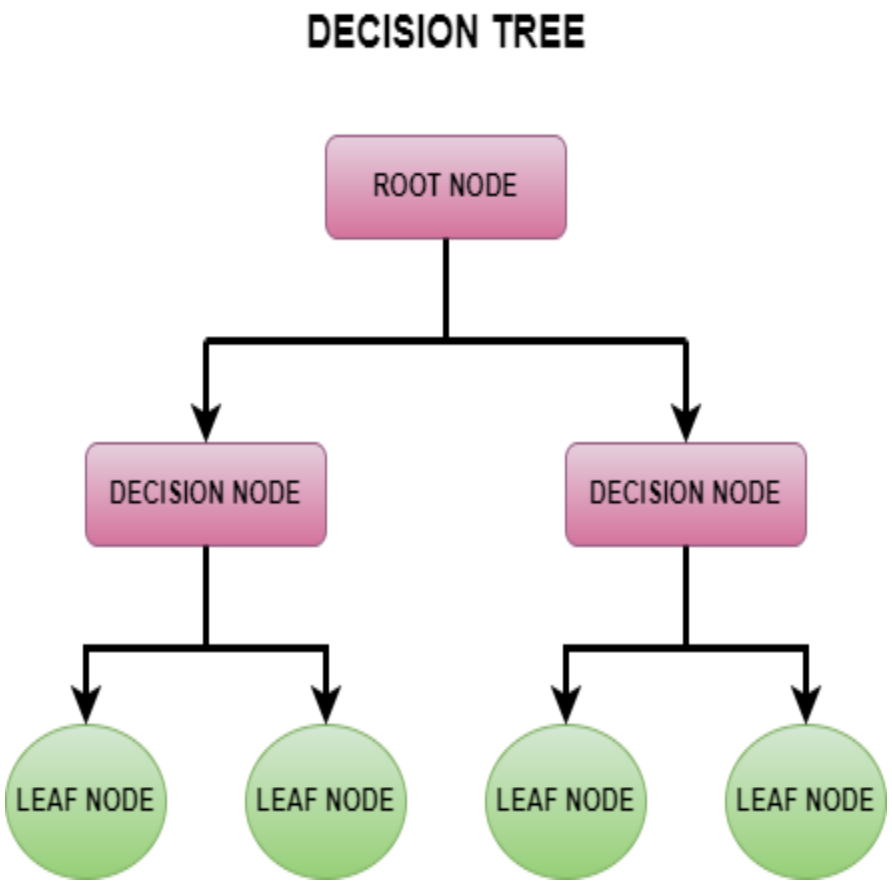


Figure 06. Decision Tree Classifier

3.3.3.K-nearest Neighbors (KNN)

The k-nearest neighbors algorithm, also known as KNN or k-nn, is a supervised learning classifier that employs similarity to generate classifications or predictions about how to identify a single piece of information. Although it can be used to solve classification or regression problems, it is most commonly used to solve classification problems because it is based on the idea that similar points can be found close together.

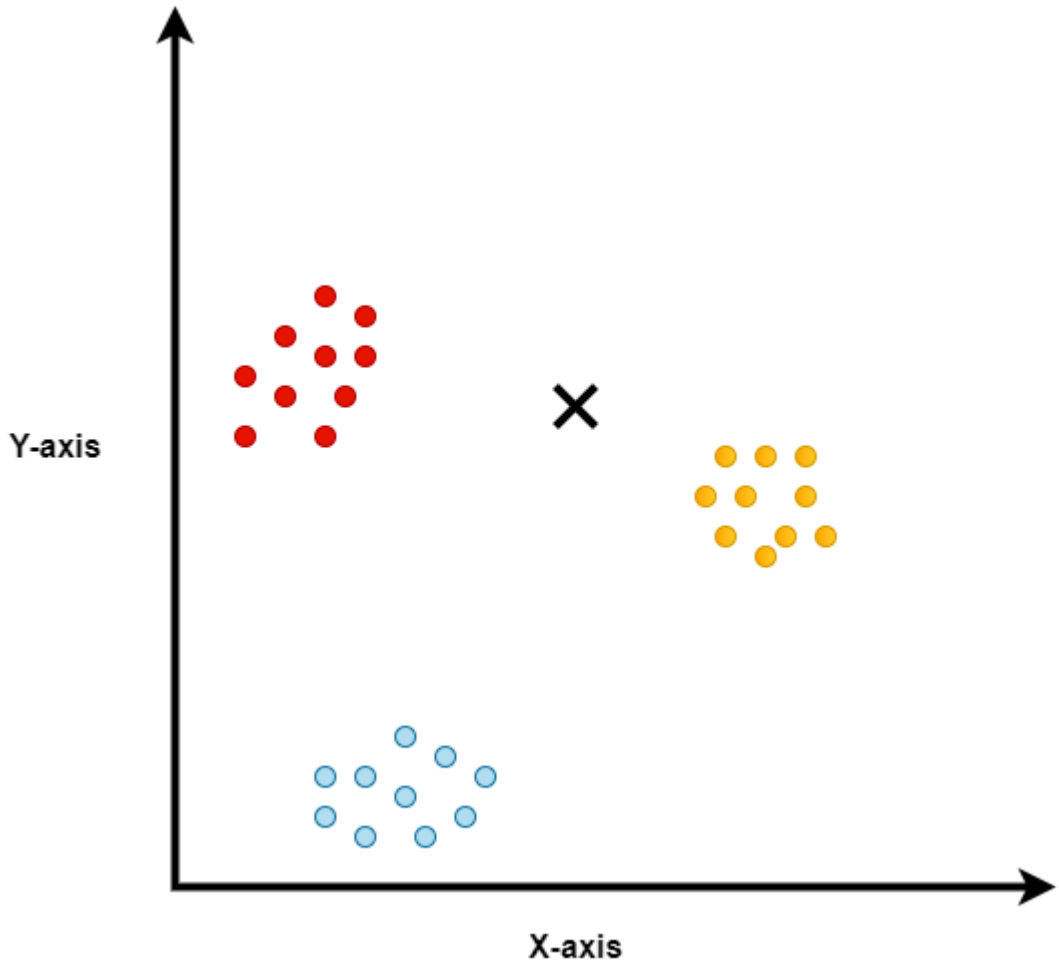


Fig:7. K-nearest Neighbor Classifier

3.3.4.SVM

Using supervised machine learning also called the Support Vector Machine (SVM) method is utilized for classification and regression. Even though we also use the word "regression concerns," classification is the preferred term. The SVM approach seeks to locate a hyperplane that categorizes the data points with clarity in an N-dimensional space.

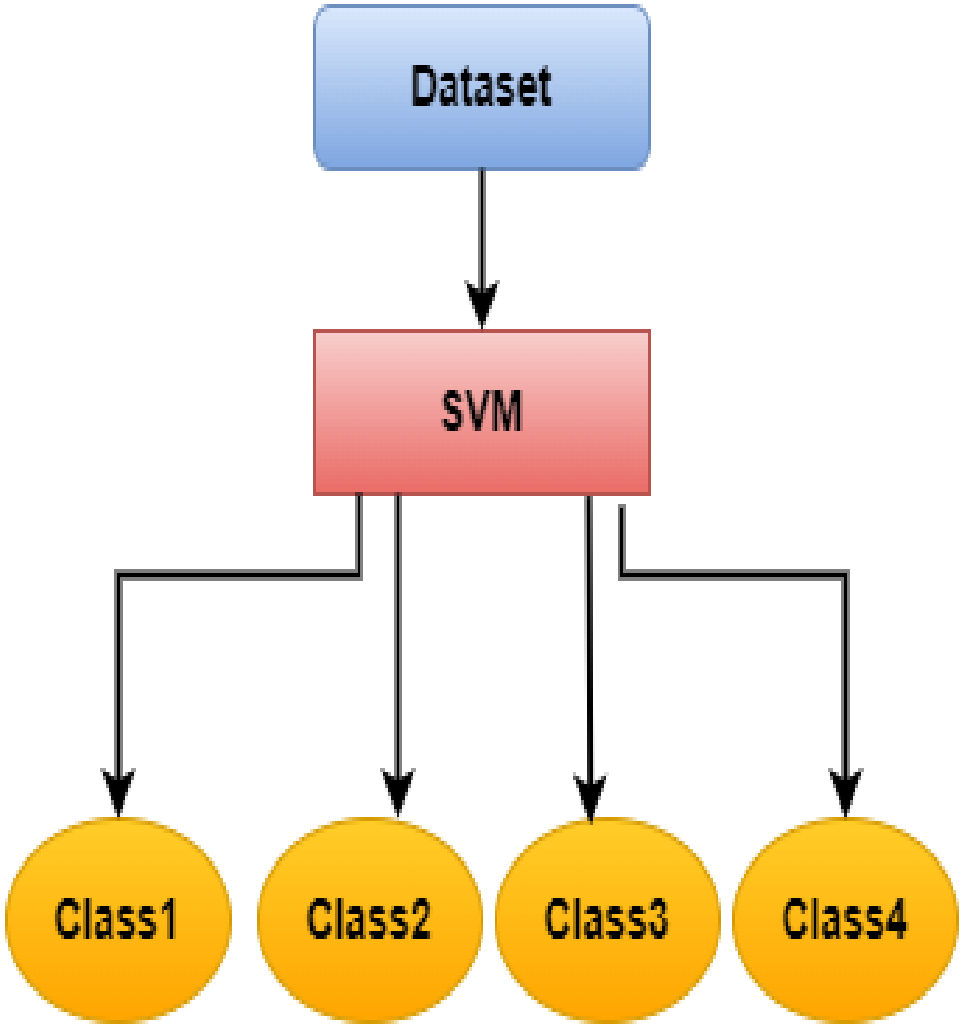


Figure 08.SVM Classifier

3.3.5. Naïve bayes (NB)

The Bayes Theorem and the concept of predictor autonomy are the basis of this classification method. Simple words, a Naive Bayes classification appears to believe that the availability of one functionality in a class has no direct effect on the presence of other functionalities in the class. A fruit is categorized as an apple if it is green, rounded, and has a surface area of estimated 4 inches. Although each of the above characteristics increases the possibility that this fruit is an apple even if they are interdependent or dependent on the existence of other characteristics, that is why it is given the label "Naive." The Naive Bayes model is easy to construct and particularly beneficial for very big data sets.

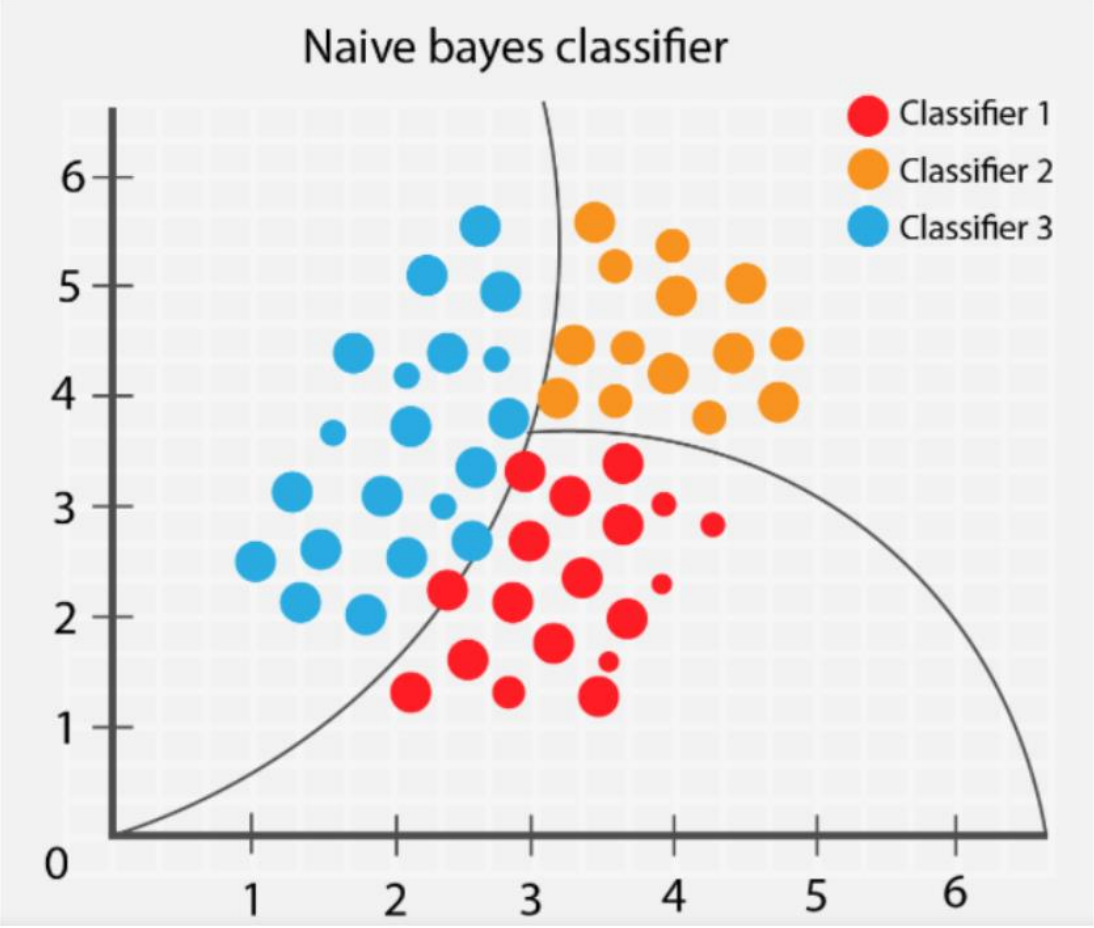


Figure 09. Naive Bayes Classifier

3.3.6. Adaboost Classifier (ADB)

The AdaBoost algorithm, relatively brief for Supervised Algorithms, is a Boosting methodology used in Machine Learning as an Optimization Learning. Because the weights are reassigned to each instance, it is called Adaptive Boosting. Higher weights are assigned to incorrectly classified instances. Boosting is a technique used in supervised learning to reduce both bias and variance. It is based on the principle of learners' sequential growth. Each subsequent learner, with the exception of the first, is developed from previously developed learners. Simply put, weak students are converted into strong students. Except for one exception, the AdaBoost algorithm follows the same principles as boosting.

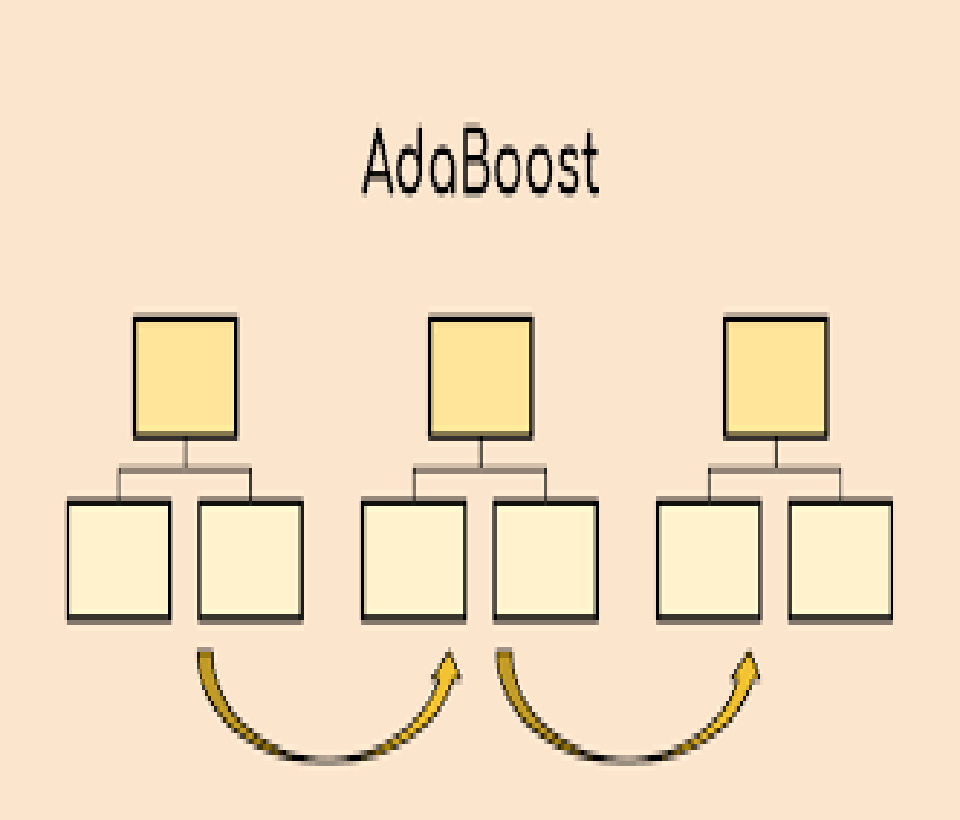


Figure 10. Adaboost Classifier (ADB)

3.3.7. Logistic regression (LoR)

When doing a logistic regression analysis, a dichotomous dependent variable should be considered (binary). Similar to other regression techniques, logistic regression is a prediction analysis. A logistic regression analysis can be used to describe the relationship between a binary dependent variable, and one or more independent word, numerical, period, or percentage predictor variables. The analytical procedure is made simpler and the results are explained in a plain text by the Intellectus Statistics program. Logistic regressions can be challenging to read at times.

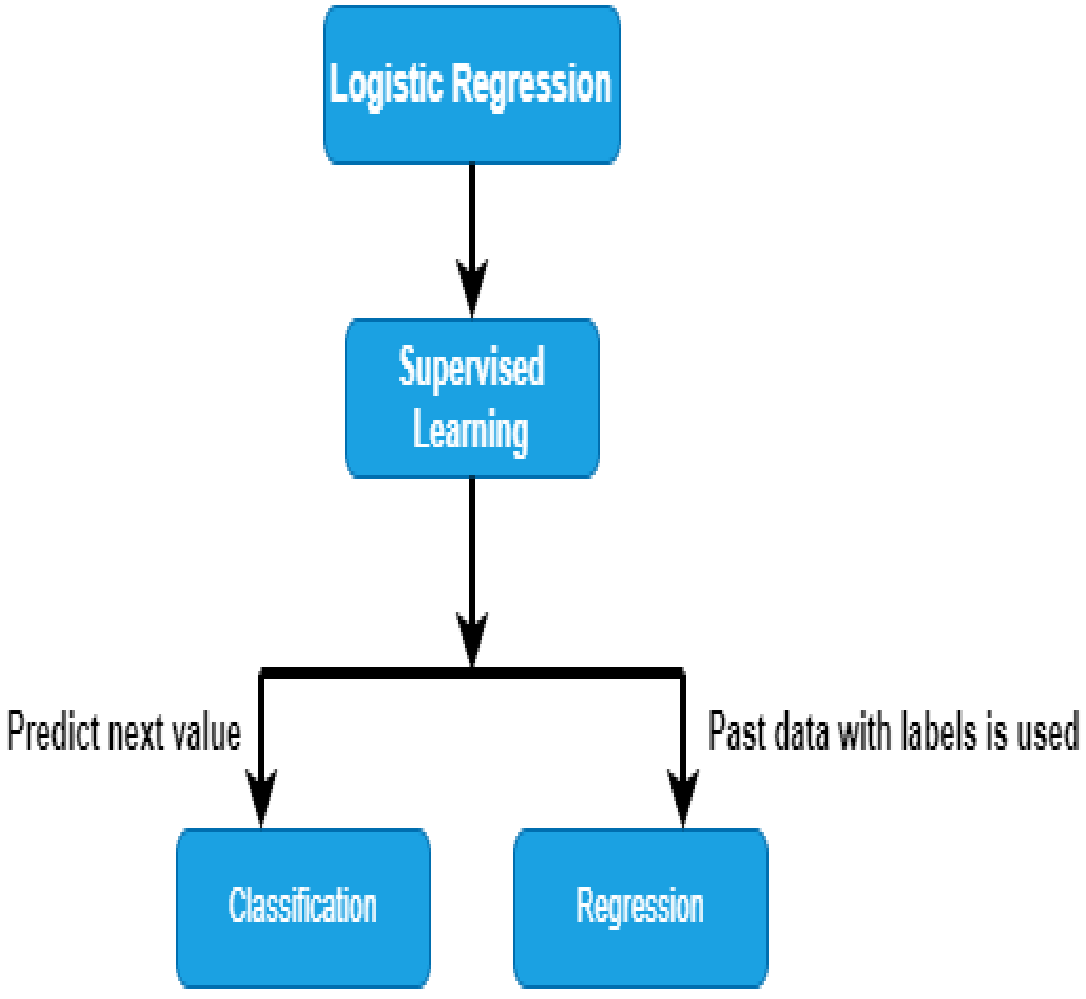


Figure 11. Logistic Regression Classifier

3.3.8. Voting Classifier

The Voting classifier is intended to function as just an estimator that incorporates two classification techniques. Subsequently, it evolves into a robust schema that manages the shortcomings of each classifier on a specific dataset. This classifier works as a grouping of individual models, producing an output (class) prediction based on the selected class's maximum probability. Voting allows the output class to be predicted based on its most substantial majority. We have chosen soft voting over hard voting for our study. Our performance category is the prediction based on that class's mean probability. All seven classification algorithms are used by the soft voting classifier. The Voting classifier has a prediction accuracy of 98%.

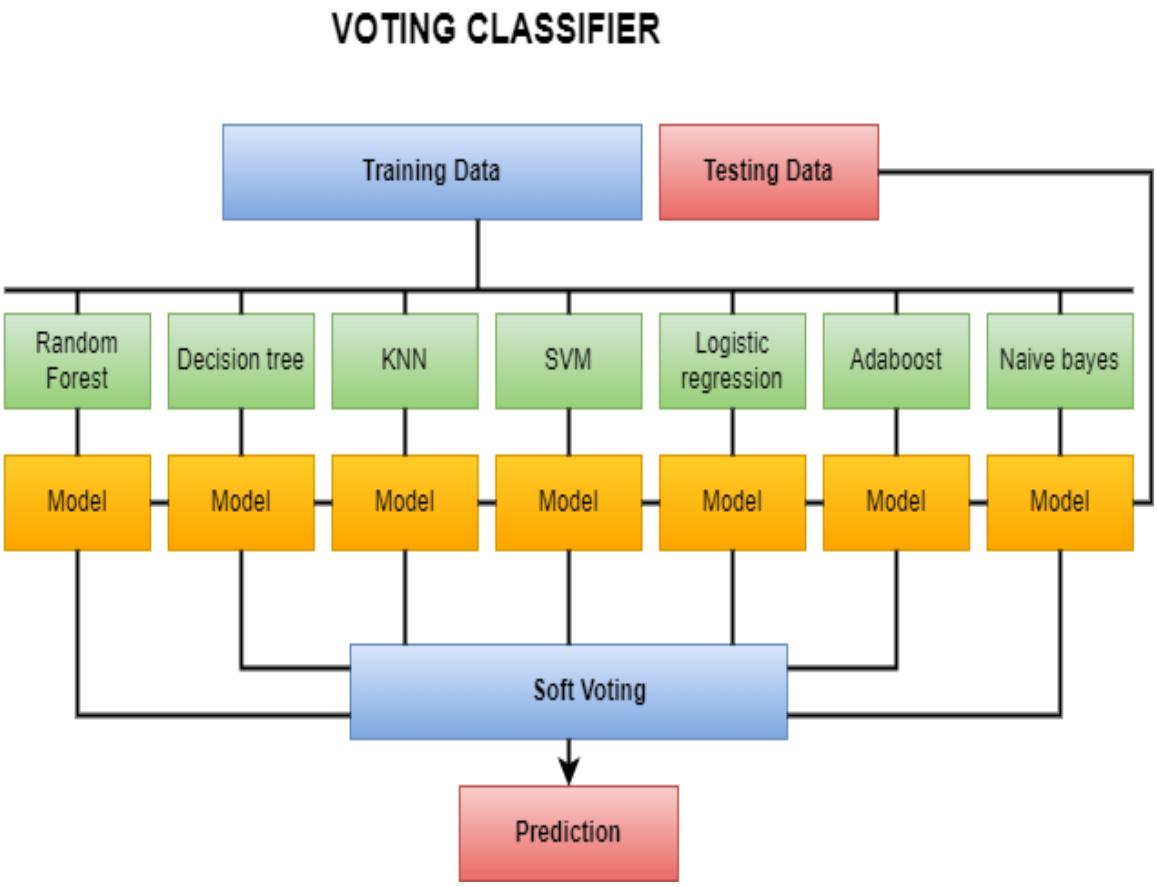


Figure 12. Voting Classifier

Table 01. Details of Datasets

Attributes	**Description**	**Types of values**
Crop	Aman, Aus, Boro, Jute, Potato, Wheat	Nominal
Year	1970-2015	Numerical
Area_(In_Acre)	Integer types of amount	Numerical
Area_(In_Lac_Hectare)	Float types of amount	Numerical
Total_Production_(M.ton)	Integer types of amount	Numerical
Total_Production_(Lac_M.ton)	Float types of amount	Numerical
Yield_rate_(M.ton_per_Hectare)	Float types of amount	Numerical
Season	Months Name	Nominal

3.4. Performance Evaluation Measure

There are a few performance evaluation metrics that we can use to determine the precise performance of our current model. These techniques evaluate overall performance based on unobserved data.

Accuracy: It refers to the percentage of correctly predicted test data. Accuracy is used to perform ease of access of the measures of actual measurements. It was based on just one variable. Accuracy addresses deliberate mistakes.

Accuracy = (TP+TN) / (TP+FP+TN+FN)

Precision: The percentage of accurately predicted positive observation is what it refers to. In actuality, precision identifies the real true portion of all the occasions when they would have anticipated true.

Precision = TP / (TP + FP)

Recall: It measures the percentage of here expected positive observations.

Recall = TP / (TP + FN)

F1 Score: Essentially, it is called the harmonic mean of recall and precision.

F1 Score = (2*P*R) / (P+R)

Figure 13 shows the visualization of confusion matrix. Here confusion matrix plays the role of a classification problem's prediction outcomes are compiled. Count values are used to describe the number of accurate and inaccurate predictions for each class. Right side bar become darker to lighter. Darker part means the higher value. Values are darker to lighter means the confusion matrix values are become stronger to weak.

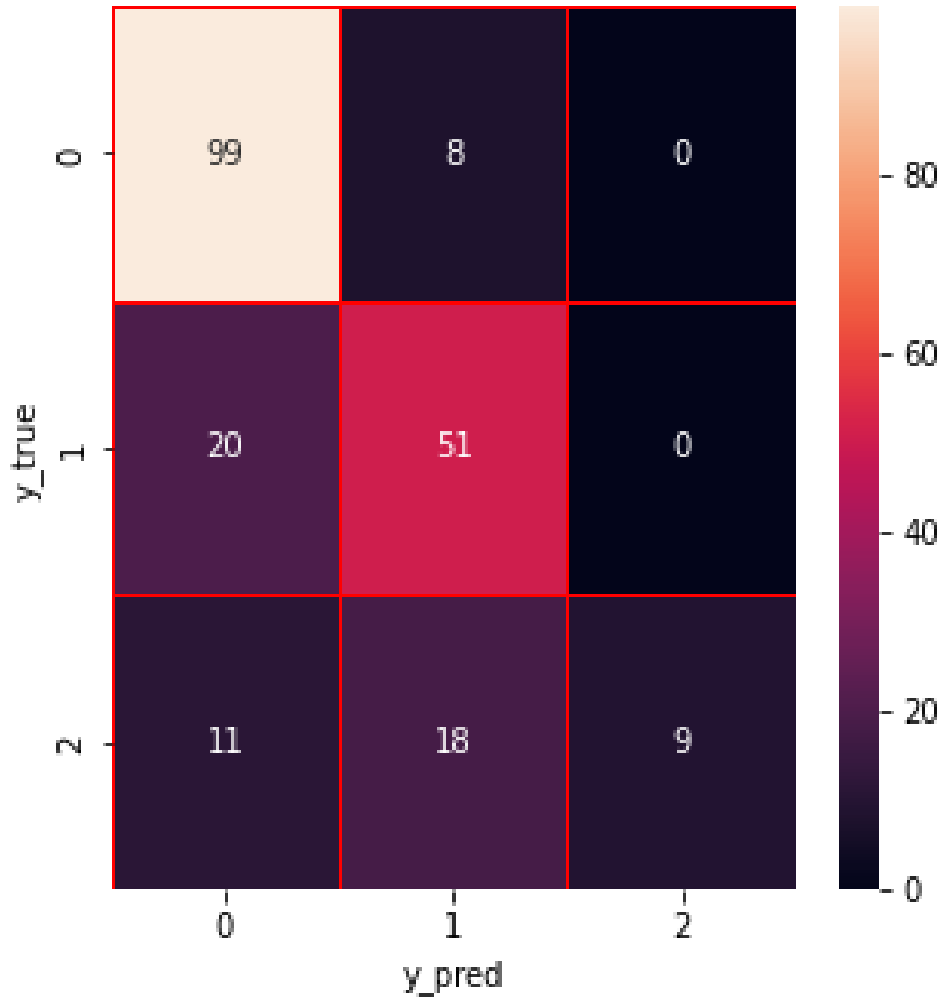


Fig13.Confusion Matrix Visualization

CHAPTER 4

Result

We must conduct a comparative analysis based on the experimental results of the evaluation measures of the machine learning models of major crops dataset in this segment. Once the data has been prepared for use in the models, it is divided into training and testing data. Using 7 types of different machine learning algorithms, K-Nearest Neighbor algorithm provides the best accuracy. And the accuracy is 98%. Second best algorithm is Random Forest (RF), its accuracy is 95%. Logistic Regression provides 91% accuracy. On the other hand SVM provides 73% accuracy. Decision Tree provides the same accuracy of SVM. Using the Adaboost classifier we got 86% accuracy. Naive Bayes provides the lowest accuracy which is 64%.

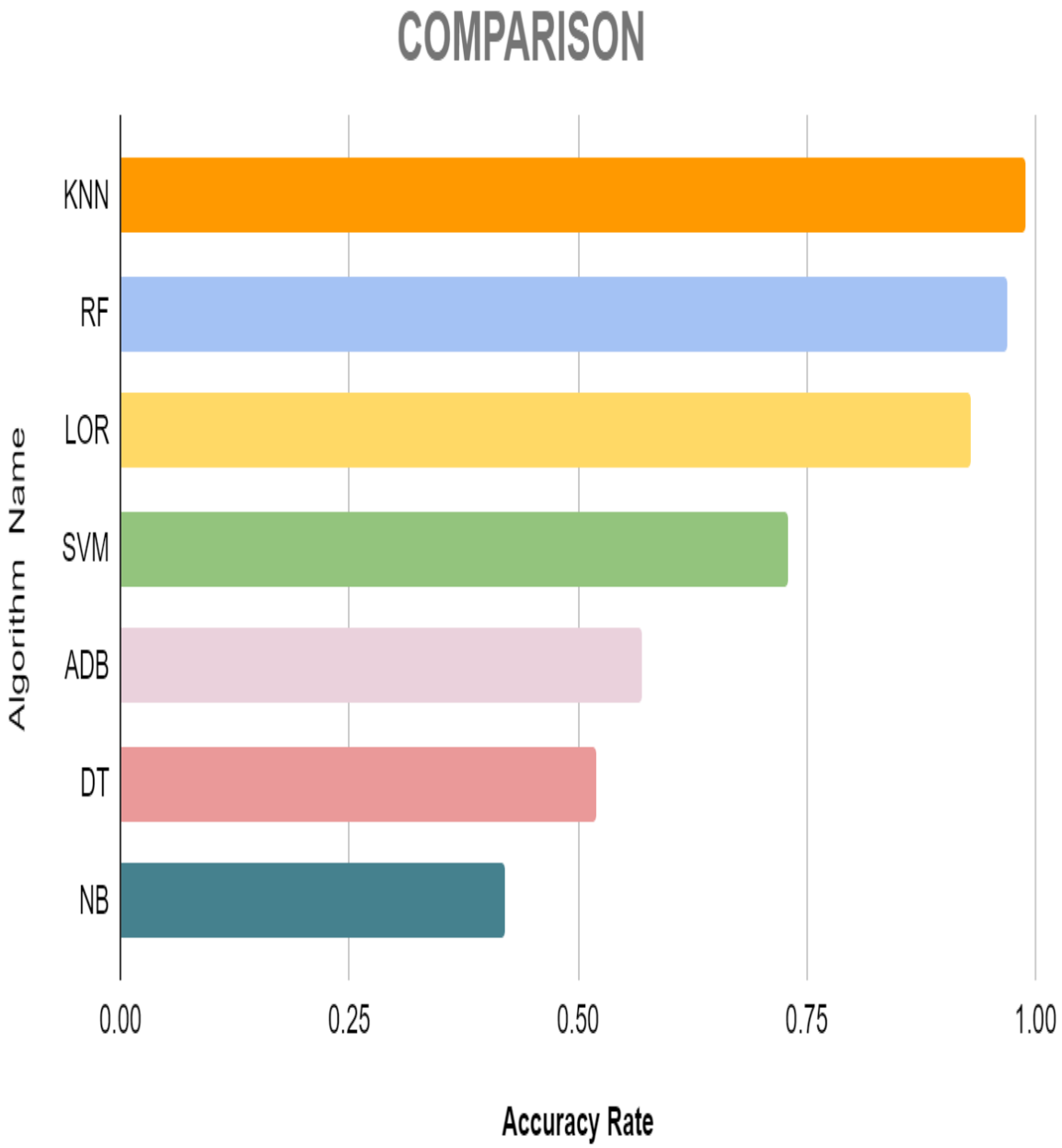


Figure14. Comparison of Algorithms Accuracy Rate

In Figure 4 here we visualize the density between Yield_rate_(M.ton_per_Hectare), Month_range and Seasons. Here we can see three types of different seasons.

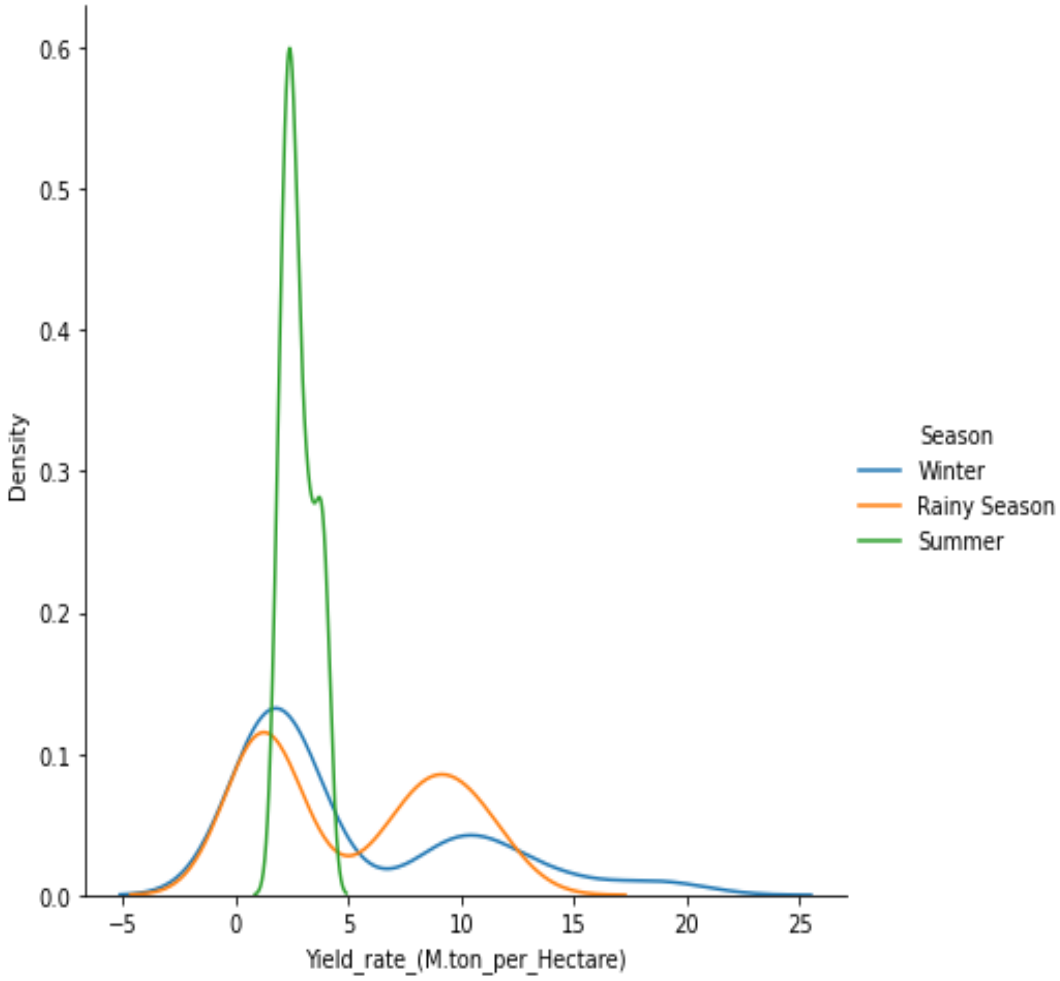


Figure15.Season wise Yield_rate_(M.ton_per_Hectare)

Table 02. Accuracy of Algorithms

Name of Algorithms	Accuracy (%)
Random Forest (RF)	95%
Decision Tree (DT)	73%
K-Nearest Neighbors (KNN)	98%
Adaboost Classifier (ADB)	86%
SVM	73%
Naive Bayes (NB)	64%
Logistic Regression (LoR)	91%

Table 03. Precision, Recall, F1 Score

Algorithm Name	Precision	Recall	F1 Score
Random Forest (RF)	0.9744	0.8889	0.9200
Decision Tree (DT)	0.5203n	0.5833	0.5253
Adaboost Classifier (ADB)	0.9331	0.7778	0.7977
SVM	0.5203	0.5833	0.5253
Naive Bayes (NB)	0.4257	0. 4964	0.4571
Logistic Regression (LoR)	0.9331	0.7778	0.7977
KNN	0.9910	0.9855	0.9880

CHAPTER 5
Conclusion and Future Work

In order for humanity to survive, agriculture is essential. A significant section of the global population depends on agriculture for their livelihood. Additionally, it provides several job options for residents. Many farmers wish to keep practicing outdated farming methods that bring in little money. The long-term growth and prosperity of the economy depends on agriculture and related industries. The three main challenges in agricultural production are the choice of the crop, supporting systems for increased crop output, and decision-making. In this study, the various machine learning algorithms are utilized to forecast Bangladesh's key crops' annual yields and suggest planting various crops at various times of year. We have made predictions using the data set for four major crops, including potato, 3 types of rice, wheat, and jute. The KNN predicts crop productivity with the best degree of accuracy. Future forecasts from the model could be improved by including a few additional pertinent features. It helps farmers grow the most advantageous crop for their agricultural property.

Reference

[1] Sajja, G.S., Jha, S.S., Mhamdi, H., Naved, M., Ray, S. and Phasinam, K., 2021, September. An investigation on crop yield prediction using machine learning. In *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 916-921). IEEE.

[2] Pant, J., Pant, R.P., Singh, M.K., Singh, D.P. and Pant, H., 2021. Analysis of agricultural crop yield prediction using statistical techniques of machine learning. *Materials Today: Proceedings*, *46*, pp.10922-10926.

[3] Paudel, D., Boogaard, H., de Wit, A., Janssen, S., Osinga, S., Pylianidis, C. and Athanasiadis, I.N., 2021. Machine learning for large-scale crop yield forecasting. *Agricultural Systems*, *187*, p.103016.

[4] Zhou, Q. and Ismaeel, A., 2021. Integration of maximum crop response with machine learning regression model to timely estimate crop yield. *Geo-spatial Information Science*, *24*(3), pp.474-483.

[5] PANDE, S.M., RAMESH, P.K., ANMOL, A., Aishwarya, B.R., ROHILLA, K. and SHAURYA, K., 2021, April. Crop recommender system using machine learning approach. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 1066-1071). IEEE.

[6] Cedric, L.S., Adoni, W.Y.H., Aworka, R., Zoueu, J.T., Mutombo, F.K., Krichen, M. and Kimpolo, C.L.M., 2022. Crops Yield Prediction Based on Machine Learning Models: Case of West African Countries. *Smart Agricultural Technology*, p.100049.

[7] 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 (ICCSIT)* (pp. 926-932). IEEE.

[8] Bondre, D.A. and Mahagaonkar, S., 2019. Prediction of crop yield and fertilizer recommendation using machine learning algorithms. *International Journal of Engineering Applied Sciences and Technology*, *4*(5), pp.371-376.

[9] Kale, S.S. and Patil, P.S., 2019, December. A machine learning approach to predict crop yield and success rate. In *2019 IEEE Pune Section International Conference (PuneCon)* (pp. 1-5). IEEE.

[10] Haque, F.F., Abdelgawad, A., Yanambaka, V.P. and Yelamarthi, K., 2020, June. Crop yield analysis using machine learning algorithms. In *2020 IEEE 6th World Forum on Internet of Things (WF-IoT)* (pp. 1-2). IEEE.

[11] {Bangladesh Bureau of Statistics. (2008-2017). Bangladesh Bureau of Statistics. Retrieved from http://www.bbs.gov.bd/.}

[12] Nischitha, K., Dhanush Vishwakarma, M.N. and Ashwini, M.M., 2020. Crop prediction using machine learning approaches. *International Journal of Engineering Research & Technology (IJERT)*, *9*(08).

Project Report

ORIGINALITY REPORT

22%	18%	15%	16%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University of Wales Institute, Cardiff Student Paper	2%
2	scholars.hkbu.edu.hk Internet Source	2%
3	Submitted to University of Sunderland Student Paper	1%
4	Submitted to University of Hertfordshire Student Paper	1%
5	www.ijert.org Internet Source	1%
6	Submitted to The University of the West of Scotland Student Paper	1%
7	Submitted to National College of Ireland Student Paper	1%
8	docplayer.net Internet Source	1%
9	expert.psru.ac.th	

	Internet Source	1%
10	Nusrat Songita Khan, Fahad Molla, Rubayat Shusmita Khan, Enamul Haque Shamim, Shahadat Hossain, Md. Manzurul Hasan. "Chapter 41 Exploration ofOnline Fake News Through Machine Learning andSentiment Analyses", Springer Science and Business Media LLC, 2023 Publication	1%
11	eudl.eu Internet Source	1%
12	Submitted to Coventry University Student Paper	1%
13	dergipark.org.tr Internet Source	1%
14	Submitted to Bournemouth University Student Paper	1%
15	www.webology.org Internet Source	1%
16	Submitted to University of Bedfordshire Student Paper	1%
17	Submitted to Ahsanullah University of Science and Technology Student Paper	<1%

18	Submitted to Far Eastern University Student Paper	<1%
19	Submitted to BRAC University Student Paper	<1%
20	Submitted to Middlesex University Student Paper	<1%
21	Rithesh Pakkala Permanki Guthu, Shamantha Rai Bellipady. "A Formal Statistical Data Modeling for Knowledge Discovery and Prognostic Reasoning of Arecanut Crop using Data Analytics", International Journal of Software Science and Computational Intelligence, 2022 Publication	<1%
22	Tanhim Islam, Tanjir Alam Chisty, Amitabha Chakrabarty. "A Deep Neural Network Approach for Crop Selection and Yield Prediction in Bangladesh", 2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), 2018 Publication	<1%
23	asianssr.org Internet Source	<1%
24	Submitted to University of Reading Student Paper	<1%
25	Submitted to Wawasan Open University	

	Student Paper	<1%
26	ijcrt.org Internet Source	<1%
27	www.ijariit.com Internet Source	<1%
28	www.medrxiv.org Internet Source	<1%
29	www.researchgate.net Internet Source	<1%
30	Submitted to King Fahd University for Petroleum and Minerals Student Paper	<1%
31	www.coursehero.com Internet Source	<1%
32	Kalpana Jain, Naveen Choudhary. "Comparative analysis of machine learning techniques for predicting production capability of crop yield", International Journal of System Assurance Engineering and Management, 2022 Publication	<1%
33	link.springer.com Internet Source	<1%
34	www.itm-conferences.org Internet Source	

<1%

35 Easa Alalwany, Imad Mahgoub. "Classification of Normal and Malicious Traffic Based on an Ensemble of Machine Learning for a Vehicle CAN-Network", Sensors, 2022 <1%
Publication

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		