SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V11I8P106

Volume 11 Issue 8, 59-71, August 2024

Original Article

An Innovative Approach Involves Machine Learning Algorithms to Forecast Future Farmer Revenue

Saikat Banerjee[1], Abhoy Chand Mondal[2]

[1]*Department of Computer Applications, Vivekananda Mahavidyalaya, West Bengal, India.*
[2]*Department of Computer Science, University of Burdwan, West Bengal, India.*

[1]*Corresponding Author : Saikat.banerjee56@gmail.com*

Received: 03 June 2024 Revised: 06 July 2024 Accepted: 03 August 2024 Published: 31 August 2024

Abstract - *Agriculture is crucial for meeting essential human needs and creating job opportunities worldwide. Agriculture provides many jobs and is essential to the economy in emerging countries like India. At this very moment, the features of weather patterns are experiencing unexpected changes due to increased pollution, climate change, and urbanization. The weather has a significant effect on how crops develop and grow. Precipitation, temperature swings, atmospheric moisture content, wind speed, and direction are some of the most essential local climatic factors that determine the success or failure of agricultural cultivation. Machine learning is a cutting-edge innovation that can solve real-world problems. Machine learning is a technique that allows computers to mimic human intelligence by learning from experience and analyzing various data sets. At this time, machine learning techniques are being used a lot in the agricultural sector. Using climatic data to predict crop yields, farmers may increase agricultural production, grow various crops, and use machine learning algorithms better. This research focuses on crop forecasting using weather data, including air temperature, humidity, rainfall, and sunshine hours. The suggested model predicts the maximum income in crops using change system agriculture. This paper proposes an intelligent agricultural system that uses machine learning to help farmers increase their income.*

Keywords - *Crop, Machine Learning (ML), Economy, Minimum Support Price (MSP), Naïve Bayes (NB).*

1. Introduction

In India, agriculture and its related sectors constitute the primary source of livelihood, with nearly 70% of rural households relying on them for their financial sustenance. India ranks among the top 15 leading nations in exporting agricultural commodities, indicating significant potential for augmenting employment prospects [18, 19].

West Bengal is characterized by the prevalence of agriculture, which not only shapes the physical environment but also plays a significant role in the economy. Among all the Indian states, it has one of the most significant proportions of agricultural land. Rice, which necessitates substantial irrigation, is the predominant agricultural product in almost every region. Although West Bengal is comparatively tiny, it plays a substantial role in India's rice production, contributing a large percentage to the yield. Sugarcane and oilseeds are also significant agricultural commodities [22].

Jute is predominantly found in the regions located south of the Ganges River. Mangoes, jackfruit, and bananas are widely grown in the southern and central areas of the State. Winter crops such as wheat and potatoes are cultivated extensively in the southern regions. The regions north of Darjiling and Jalpaiguri have gained a reputation for their long-standing production of tea of exceptional quality. In addition to tea, the Darjiling region cultivates oranges, apples, pineapples, ginger, and cardamom [23, 24].

The agricultural sector is vital to the economy of West Bengal. Even though it just makes up 2.7% of India's landmass, almost 8% of its population lives there. In West Bengal, cropping intensity is the greatest in India. Farmers use their cultivable area 1.85 times in a typical agricultural year, more than 2.5 times the national average. Of the 71.23 lakh farming families in the country, 96 percent are considered small or marginal farmers. .

Approximately 0.77 hectares is the average size of a land holding. However, the state is blessed with many natural resources and unique agroclimatic conditions that allow for the cultivation of an extensive range of crops. There are 52.05 lakh hectares of the planted area, which accounts for 68 percent of the landmass and 92 percent of the arable land [25, 26].

The State is prone to natural disasters like floods, cyclones, hailstorms, and others because of its vicinity to the

Bay of Bengal and its position in the humid tropics. Worldwide, around 70% of freshwater withdrawals are consumed by agriculture. The bulk of these withdrawals are specifically designated for use in irrigation. Farmers may still tap groundwater, a critical freshwater supply of 40% of irrigation during the dry season, despite the lack of surface water [27]. About 60% of India's agricultural water requirements and more than 85% of its potable water supply come from groundwater, making it the world's largest resource consumer [11, 15, 16]. In several central rural locations across the country, groundwater reserves are currently decreasing at a rapid pace. While agricultural intensification has increased food production, it has also led to rapid water extraction and the concomitant depletion of aquifers [28].

By 2025, projected areas with low future groundwater supplies should see a total reduction in cropping intensity of over 68%, while the rest of the country should see a reduction of about 20% [27]. The problematic consequences of climate change are making matters worse; this phenomenon, which affects rainfall patterns, is particularly difficult for increasing agricultural productivity [18, 19]. However, there is hope for the potential of agricultural diversification to alleviate the strain on groundwater resources. The State is India's primary contributor to rice production, accounting for 15 percent of the national total. Rice is a staple meal in most regions of India. West Bengal is the second-largest grower of potatoes, contributing 25 percent to the national total. Kharif rice has increased in output. This high-yielding type necessitates considerable utilization of irrigation, fertilizers, and insecticides [20]. It has substantially reduced the groundwater table in West Bengal during the winter and pre-monsoon seasons, with levels dropping by three to 12 meters or even more.

Over the past decade, the farm area dedicated to cultivating Boro rice has significantly increased. This trend has led to a five to ten percent expansion annually, highlighting the scale of the issue. This expansion has primarily relied on groundwater extraction from unconfined aquifers using privately owned shallow tube wells. Several tube well systems experience water depletion or intermittent water flow during summer, particularly before the monsoon. The consequences of this depletion are severe, with the agricultural sector facing increased uncertainty and risk. [22, 23]. While the State has plenty of potatoes, vegetables, and rice, pulses, oilseeds, and maize are in short supply. Problems with chemical fertilizer imbalances, inadequate farm mechanization, unorganized marketing mechanisms, and a lack of suitable improved seed types slow agricultural expansion and worsen soil health. Nevertheless, it is imperative to promptly address this disparity to ensure that low-cost technology is accessible to small and medium-sized farmers, enabling them to reap its benefits. ML is a subfield of AI that uses statistical approaches to train computers to improve their performance over time. It liberates computers or robots controlled by computers from being rigidly programmed to perform a specific action and instead enables them to respond and make decisions depending on facts. Modern ML applications mainly serve two functions: to classify data according to pre-existing ML models and to foretell potential future outcomes using these same models.

In ML, extracting features is a manual procedure. Supervised, unsupervised, and reinforcement learning are the three primary schools of thought in ML. Supervised learning involves training algorithms to categorize input or predict outcomes using labeled datasets correctly [26]. Detecting spam, recognizing handwriting or voice, classifying documents, and using biometric information are all examples of applications in the real world. For example, Random Forest (RF), Decision Tree, Logistic Regression, Support Vector Machine, and many more are among the most widely used classification methods [2, 3]. It can understand the relationship between our data features and some of the continuous-valued responses through regression, which allows us to make predictions. Market trends, weather forecasts, and other continuous factors can be anticipated using this strategy [26, 29].

Popular regression methods include decision trees, Lassa regression, multivariate regression, basic linear regression, and many more. In unsupervised learning, there is no connection between the input data and any corresponding output variable [2]. The primary objective is to build a model that faithfully depicts the data's distribution or underlying structure. The algorithm's job is to classify the unlabelled datasets based on their shared or unique characteristics [28]. Clustering and association learning are the two main categories of unsupervised learning. In clustering, groups of related data are called "clusters," whereas sets of dissimilar data are called "uniform clusters."

Using neural networks for feature extraction gave rise to the Deep Learning (DL) subfield of ML. On the other hand, deep study is just a learning method that uses how people's minds work. "DL" computers strive to mimic human reasoning by continuously evaluating data according to a predefined logical framework [29]. Neural networks, which are multi-layered structures composed of computational components, are utilized by DL to achieve this purpose. The neural network's architecture was loosely based on how the human brain is structured [30].

Like regular neural networks, deep neural networks use the same overarching principle. Grouping, classifying, and regression analysis are just a few of the many activities made possible by neural networks [2, 9]. On the other hand, it can train the community to use a classified dataset and sort the samples into different categories inside the category case. DL models can address issues that device mastery models cannot handle. All of the latest advancements in AI have been driven by DL [29, 30]. According to the literature, implementing a trust model based on blockchain technology can enhance transparency in allocating organic goods and augment the tamper-proof capabilities of organic farming practices [28]. Prior studies have suggested remedies for the oil from olives supply chain predicament by utilizing the Internet of Things (IoT) and multi-sensor monitoring tools within blockchain intelligent contracts in the realm of blockchain technology for the agri-supply chain, one potential avenue involves overseeing and validating companies' adherence to regulatory and market standards in agri-food traceability by governmental entities. [1]. Weather, soil, solar hour, rainfall, and other parameters play a pivotal role in the success of Indian agriculture, making them crucial elements to understand and manage.

Aerial photography, sensors, and advanced local weather forecasts are all integral components of the innovative field of system farming, which leverages robotics, big data, and advanced analytical capabilities to optimize and manage crop production. System farming is like taking medicine to treat a specific illness. Every aspect of the solutions is carefully customized, from the crops that work best on a given plot to the areas where pesticides are applied. Because each plot's unique requirements are met through precision agriculture and contemporary agricultural techniques, production costs and waste are decreased.

Modern farming techniques rely on technological tools and software for crop management. Sensor-equipped devices in the fields allow for thorough data collection on soil testing, plot measurement, weather pattern analysis, and crop analysis. After the data is adjusted to conclude, a comprehensive and exact set of procedures can be implemented. This research aims to develop an intelligent agriculture system to enhance farmers' profitability. In this study, we initially gathered sensor data from three districts in West Bengal. Subsequently, we employed an ML algorithm to predict the most appropriate crop based on the collected data. Next, we determined the crop's production and compared it to conventional production. We determined the overall revenue and the revenue explicitly generated by the proposed intelligent system. The experimental results demonstrate that implementing the proposed intelligent system leads to an increase in profit.

2. Literature Survey

This study aims to analyze the latest advancements in intelligent applications that use the IoT and Artificial Intelligence (AI) technology in smart agriculture. Additionally, it emphasizes the unresolved issues and challenges related to intelligent applications in smart agriculture, along with the current trends and prospects in this area [1]. Moreover, this research objective is to show the possible integration of IoT and AI technologies in the smart agricultural sector. The results of this analysis will furnish a functional understanding of future research and development of IoT and AI technologies desired to improve the quality norms and profitability of the agriculture business. This paper emphasizes open issues, challenges, trends, and options in the industry as it investigates current improvements in smart applications using IoT and AI for smart agriculture. The effects will benefit the following IoT and AI projects aspiring to raise the agricultural sector's profitability and grade benchmarks through IoT solutions.

Research mostly focuses on agricultural production forecasting using the RF algorithm. The RF in ML is famously resilient when confronted with large datasets and complex interdependencies. The goal of the research was to demonstrate that this technique could accurately forecast crop yield within the agricultural domain. The research would have focused on the outcomes of their experimentation, as well as metrics like recall, accuracy, precision, and possibly invariant F1 score or Root Mean Squared Error (RMSE) [2]. This would have been determined by whether the circumstance was considered a regression or classification problem. The findings provide evidence that the RF model is superior to baseline models or conventional methods when it comes to predicting the amount of crops that will be harvested.

The work employs soil data RF classification to predict crops. The writers look at how ML methods might be applied in agriculture to raise crop prediction accuracy. They underline that crop forecast depends on agricultural planning which depends on soil data. The authors offer the RF algorithm, a strong ML technique, to generate correct forecasts from complex datasets [3]. The method of soil data collecting describes several elements, including pH, nutrient content, moisture levels, etc.). The authors also examine data pretreatment techniques to guarantee that the data has acceptable quality and is compatible with the classifier. This research shows their empirical developments, demonstrating how the RF classifier projects agricultural crop type from soil properties. The authors indicate which soil factors influence agricultural crop performance. Given the tremendous data, the researchers operated a dataset that enclosed several essential features: crop type, soil properties (pH, fertility), and climatic characteristics (temperature, humidity, precipitation). These attributes are relatively vital since they exploit the productivity and growth of the agricultural crop.

This research paper uses the C5.0 approach as an Advanced Decision Tree (ADT) model to analyze agricultural data performed on cloud computing infrastructure. Using cloud computing for agricultural data processing addressed in the paper yields benefits in terms of scalability, cost-effectiveness, and accessibility [4]. Along with the experimental design, the agricultural dataset they detail includes soil qualities, weather, crop varieties, and yield statistics. Everything is set up to test and teach the C5.0 model. They examine its efficiency using F1-score, recall, accuracy, and precision. Strengthening the results can be accomplished by the utilization of cross-valuation techniques. In addition to addressing which parameters, such as the pH and temperature of the soil, have a significant impact on the accuracy of the categorization, the authors may provide particular numerical figures. The authors proceed to jointly consider the discoveries constructed in this research and how they strengthen the impact on farmers' preferences. They examine how farmers might use the C5.0 model to maximize agricultural crop yields, control disease outbreaks, and distribute resources.

This study explores the integration of crop recommendation systems with Explainable Artificial Intelligence (XAI) methodologies to improve agricultural decision-making. The issue of AI-driven suggestions needing more transparency and ease of interpretation is addressed. It is important in complex fields like agriculture, where decisions can have significant economic and environmental impacts [5]. The researchers are concerned with the XAI techniques employed in their research, such as SHAP values, LIME, and characteristic stature analysis.

Different approaches are utilized to explain the generation of suggestions to enhance the reliability and usefulness of the AI model for farmers and decision-makers. This research paper could compare their XAI-enhanced and other traditional methods to agricultural crop recommendations, like statistical measures or expert techniques, and measure user fulfillment with the effects. The authors articulate how the XAI-enhanced approach is beneficial in controlling water and fertilizer, optimizing agricultural crop production, and adjusting to weather variability. This study can demonstrate how the method facilitates eco-friendly farming approaches [5].

Soil assessment and categorization are two areas where the K-Nearest Neighbor (KNN) algorithm shines, and the authors provide it here. They draw attention to its dependence on local neighborhood information for prediction and its non-parametric nature. Various soil properties, including pH, nutrient content, texture, wetness, and geographic or meteorological data, are likely to be discussed in the article as part of the soil dataset utilized for the experiments. Predicting soil kinds or characteristics relies heavily on these parameters [6].

Performance indicators, including accuracy, precision, recall, and F1-score, are the main emphasis of the experiments presented in the paper. These metrics assess the accuracy of the KNN model in predicting soil attributes or sorting soil samples into their correct categories. The issues analyzed by the authors that affect model performance are the choice of distance measure, influence of K value selection, and significance of various soil properties in classification accuracy. They may also propose new lines of inquiry, including how to fine-tune KNN parameters or include more data sources to get better classification results.

The writers present the KNN method and discuss how it might be used to analyze data in agriculture. The key to KNN's success is its ability to classify new data points using the feature space's majority class as a guide. The researchers describe the agricultural dataset that they employed in their tests. Soil characteristics (pH, nutrient levels), weather information (temperature, humidity), crop kinds, and socioeconomic variables impacting agricultural results are likely components of this collection. It necessitates going into the approach to parallelization (whether it be data or task parallelism) and the particular programming frameworks and tools used (such as CUDA for GPU acceleration). The authors examine the scalability of the parallel KNN method in terms of computational resources (e.g., number of processors, GPU cores) and dataset sizes. The algorithm's capacity to handle large-scale agricultural datasets effectively depends on its scalability. In the article, the authors evaluate parallel KNN against conventional sequential KNN using a variety of performance measures, including accuracy and execution time. Parallelization improves algorithm performance without sacrificing accuracy, as this comparison shows [7].

The writers discuss decision tree algorithms and their potential uses in farming. They go over how decision trees make it easier to comprehend the reasons behind classification judgments, which is vital for those involved in agriculture. Decision tree algorithms like ID3, C4.5, CART, and maybe even more recent versions like RFs and Gradient Boosted Trees are probably covered in the article. Each method's merits, shortcomings, and applicability to various agricultural classification tasks might be debated. This information includes soil parameters (pH, moisture, nutrient content), climatic data (temperature, rainfall), crop types, and perhaps socio-economic elements influencing agricultural outputs. Authors may evaluate decision tree algorithms compared to popular agricultural ML methods like neural networks, KNNs, or Support Vector Machines (SVM). Researchers can learn more about the relative merits of decision trees for agricultural categorization jobs by comparing them here [8].

This study presents the issue of agricultural data categorization and stresses the importance of reliable classification methods to enhance agricultural decision-making. Classification and regression are two areas where SVM shine. Its primary goal is to find the best possible hyperplane to maximize the margin between data points as it divides them into classes. Performance metrics for SVM, including F1-score, recall, accuracy, and precision. When there is much space between the classes in the feature space, SVM usually does an excellent job of classifying agricultural data. Agricultural applications of SVM, such as crop disease classification, yield prediction, and soil type identification, may be included in the study [9].

This study discusses the urgent need for reliable crop production forecasting to improve farming methods and guarantee a steady food supply. Because so many variables affect agricultural output, traditional approaches typically fail to provide accurate results. Combining SVM with decision tree concepts is the essence of this method. Decision trees help pick out essential qualities and features when predicting agricultural yields. A prediction model is constructed using SVM using these attributes. The specifics of the data collection and preparation processes for crops, including characteristics like weather, soil, and crop management techniques. Measures used to assess the model's prediction capabilities, including R-squared, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). Researchers will discuss how multi-attribute weighting helps improve prediction accuracy, how healthy decision tree concepts and SVM can be integrated, and ideas for where the field could go regarding research or how the developed approach could inform agricultural decisions [10].

From what we can see in the above literature study, several authors have taken diverse stances while discussing this topic. Most of this field's work has used ML algorithms to aid farmers in crop prediction and increase output generally. Many studies failed to consider a practical issue while evaluating and classifying performance. According to the research outcome, farmers could make better decisions on elements that impact crop development by analyzing diverse agricultural data, particularly current information from IoT devices. This study's findings show that ML algorithms perform well in classification. Thus, the results can be used as a sign for farmers to make timely and informed decisions about crop prediction. Thanks to the integration of various technologies, this improves productivity and, in turn, the economy as a whole.

3. Materials and Methods

3.1. Smart Farming

Implementing innovative farming techniques enables plant growth analysis and the real-time adjustment of system parameters to optimize crop development and enhance agricultural operations. IoT systems, which use sensors to gather application-specific data and perform intelligent processing, efficiently bridge the divide between the virtual and physical domains. The present study outlines the development and implementation of an innovative system for agriculture utilizing an intelligent platform that incorporates ML techniques to facilitate predictive capabilities. The system under consideration is based on wireless sensor network technology. Its successful implementation necessitates the completion of three primary stages:

1. The data collection phase is executed by deploying sensors in an agricultural field. This data includes soil moisture, temperature, NPK, and pH, which are crucial for crop health.
2. The collected data is subjected to cleaning and storage processes to ensure its accuracy and accessibility.
3. Predictive processing is carried out using specific AI techniques.

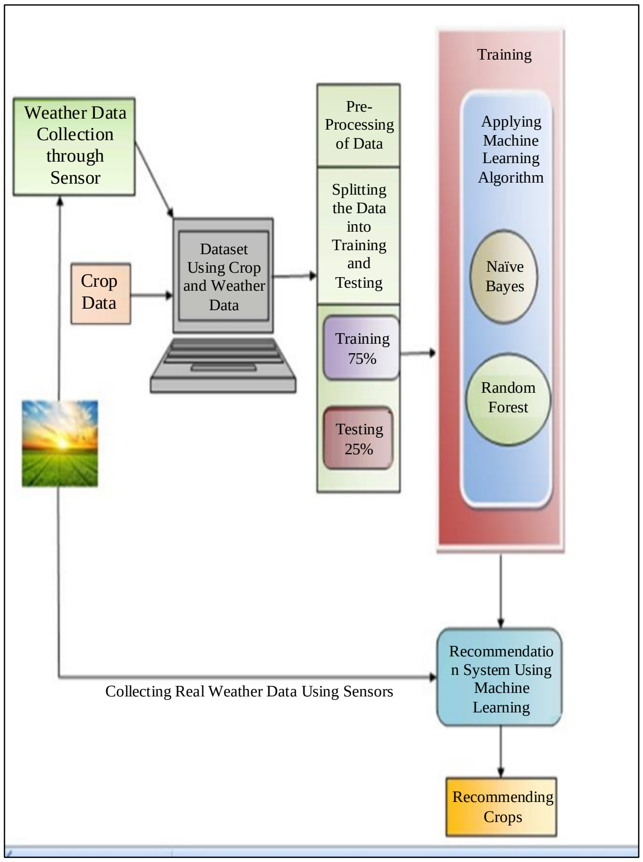

Fig. 1 Workflow of data collection

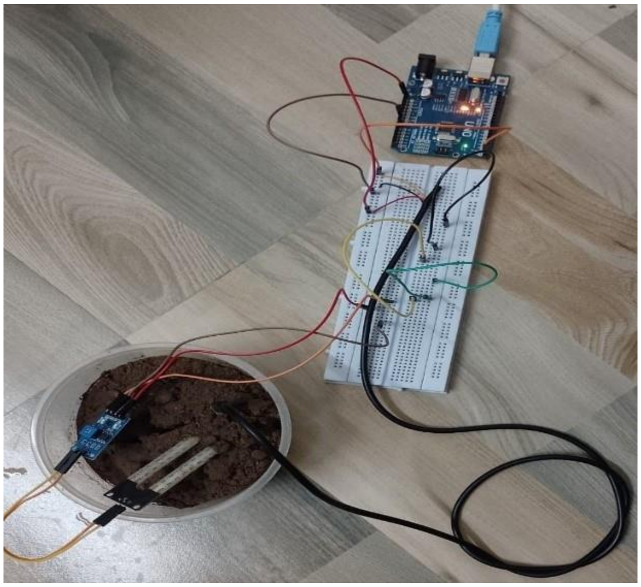

Fig. 2 Soil data collection using sensor devices

Initially, jumper wires connect the moisture devices to their respective modules. Subsequently, the module establishes a connection with the Arduino port through the use of Vnn, Gnd, and A0. The Arduino, a crucial component in our intelligent farming system, is connected to all other devices through jumper wires. Once all the devices are connected, the laptop's voltage is supplied to the Arduino board. Subsequently, the Arduino IDE is installed on our computer to obtain the desired numerical output. Upon installation, the user designs the port and Arduino model, specifically the Arduino Uno. In the Arduino Integrated Development Environment, programmers compose code tailored to particular devices. Subsequently, the data is gathered and consolidated at designated locations such as Hooghly, Howrah, and Burdwan. Once the devices are inserted into the soil, we execute our code and observe the output on the serial monitor. The output signifies the information transmitted from the serial monitor to Excel. To perform this task, Datastreamer must be installed in Microsoft Excel.

Subsequently, the code is executed, and the serial monitor is deactivated. The user then navigates to the Excel interface and selects the datastreamer option. Upon doing so, the user initiates the recording process by selecting the "start recording" button, which results in the data being stored in CSV Excel format. This procedure gathers data from various locations, including Hooghly, Howrah, and Burdwan. Raw food suppliers participate in the sharing of information on the origin and path of food items throughout the whole supply chain. Food manufacturers provide information about the ingredients of the food product and the methods used in its production. Food inspectors and certification agencies verify the authenticity of paperwork regarding the origin and quality of the product. Transportation providers input information about the whereabouts and environmental conditions of food products during transportation. Food product distributors track food supply chain activities to ensure ethical sourcing procedures and the quality and safety of food goods. Consumers can get information on the source of food goods to verify their authenticity.

The experimental configuration for this investigation included an Arduino Uno Board, a DHT11 Temperature and Humidity Sensor, a Rainfall Detection Sensor module, a Buzzer, and multiple jumper wires. The device is outfitted with three pins: GND, VCC, and DATA. The sensor and Arduino were connected using jumper wires. The GND pin of the sensor was linked to the GND pin of the Arduino, the VCC pin of the sensor was linked to the 5V pin of the Arduino, and the DATA pin of the sensor was linked to the digital pin 2 of the Arduino. The Arduino board was connected to a laptop to receive electrical power.

The Arduino board was related to the computer using a USB wire. Afterwards, the Arduino IDE software was used to run the sensor code. When utilizing an IDE, selecting the suitable board type and port is imperative. After the selection procedure is finished, it is essential to download the DHT11 sensor library. Afterwards, we can proceed with the creation and execution of our code. After running the program, the results can be viewed in the serial monitor by choosing the upload function. The data exhibited on the serial monitor experiences intermittent fluctuations occurring within a short span of a few seconds. Temperature in this context can be measured using either the Celsius or Fahrenheit scales.

The Data Streamer utility facilitates acquiring and retaining data from an Integrated Development Environment (IDE) in the .csv file format. The data recording process was halted by choosing the "stop recording" option on the data streamer interface. The rain sensor module is a simple device used to detect rainfall. The module is organized into two separate portions. The system consists of two main components: the detector board and the connector board. The connector board is outfitted with four pins, specifically GND, VCC, AO, and DO. The sensor and Arduino were connected using jumper wires.

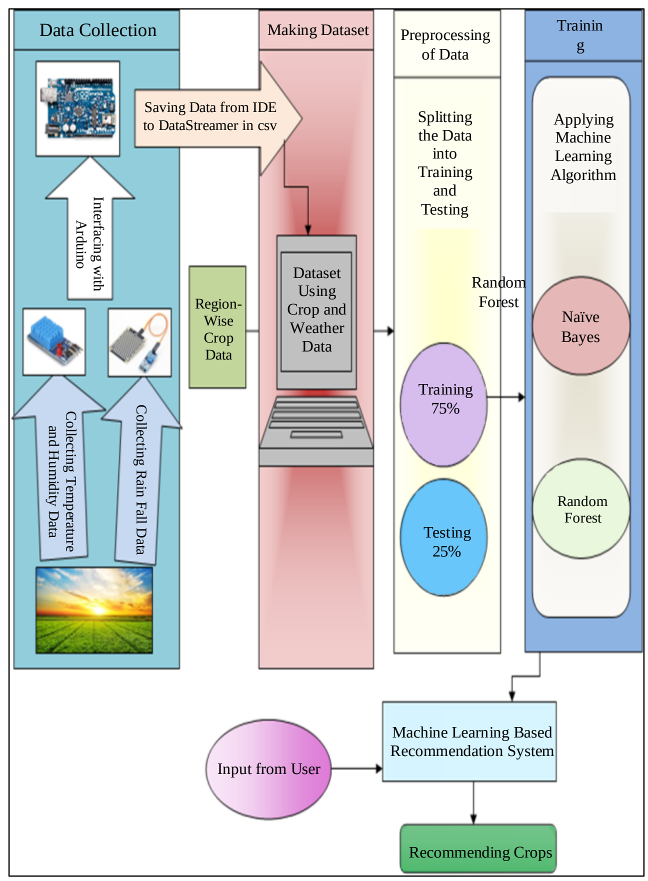

Fig. 3 Workflow of the proposed model

The AO interface is utilized to link the board to the analog pin of the Arduino, while the DO interface is used to connect the board to the digital pin of the Arduino. Jumper wires link the detector board to the sensor's connector board. Similarly, we established a connection between the GND pin of the connector board and the GND pin of the Arduino.

Table 1. Dataset

State_Name	District_ Name	Season	Crop_Name	Label	Temperature (°c)	Rainfall (mm)	Humidity (%)	Sun_Hours (hours)
West Bengal	Bardhaman	Kharif	Jute	1	27.1	1197	80.2	8.3
West Bengal	Hooghly	Kharif	Jute	1	26.9	1390	86	8.1
West Bengal	Howrah	Autumn	Rice	2	27	395	85	8
West Bengal	Bardhaman	Autumn	Rice	2	22.3	326	80	8.2
West Bengal	Hooghly	Autumn	Rice	2	26.4	357	65.9	9.8
West Bengal	Bardhaman	Summer	Moong(Green gram)	3	32.4	211	57.9	9.6
West Bengal	Hooghly	Summer	Moong(Green gram)	3	30.2	253	66.4	9.6
West Bengal	Bardhaman	Rabi	Potato	4	24.74	187	69.2	9.21
West Bengal	Hooghly	Rabi	Potato	4	23.39	203	69.8	9.2
West Bengal	Howrah	Rabi	Potato	4	22.4	222	68.3	9.3
West Bengal	Hooghly	Winter	Boro Rice	5	20.1	46	63	9.5
West Bengal	Howrah	Winter	Boro Rice	5	18.1	60.3	63.9	9.2
West Bengal	Bardhaman	Winter	Boro Rice	5	19.3	43	60	8.8
West Bengal	Hooghly	Annual	Sugarcane	6	28.25	1701	79.8	8.9
West Bengal	Howrah	Annual	Sugarcane	6	29.11	1688	76.4	8.8
West Bengal	Bardhaman	Rabi	Wheat	7	18.99	181	66	9.11
West Bengal	Hooghly	Rabi	Wheat	7	20.77	184	65.7	9.3

Additionally, we linked the AO pin of the connector board to the A0 pin of the Arduino. Subsequently, using jumper wires, we made a physical link between the buzzer and the D3 pin of the Arduino, along with the GND pin of the Arduino. A laptop energized the Arduino board to provide electrical power.

Figure 2 shows that a USB cable connected the Arduino board and the computer. When utilizing an IDE, selecting the suitable board type and port is imperative. Afterwards, we can proceed with the creation and implementation of our code.

After executing the code successfully, we can view the result on the serial monitor by choosing the upload option. The data exhibited on the serial monitor experiences periodic fluctuations during a brief period. It is crucial to utilize Excel Data Streamer instead of the IDE serial monitor for future use.

The Data Streamer utility facilitates acquiring and retaining data from an IDE in the .csv file format. The data recording process was halted by choosing the "stop recording" option on the data streamer interface.

Once we finished collecting data using weather sensors, we continued to save the obtained data in CSV format. Subsequently, we will collect agricultural data that is tailored to the particular regions we have chosen. Using the given information, we have created a thorough dataset that includes the following variables: State_name, District_name, Season, Crop_name, Label, Temperature (°C), Humidity (%), Rainfall (mm), and Sun Hours (hours) [11-17]. Data has been collected from three districts in West Bengal, specifically Bardhaman, Hooghly, and Howrah.

The study have classified our categorization into six seasons: Rabi, summer, autumn, kharif, yearly, and winter [11, 12]. We planted a wide variety of crops, such as Jute, Rice, Moong (Green gram), Potato, Boro rice, Sugarcane, Wheat, Sesame, Chilies, Peas, Beans, Soybean, Maize, Masoor, Rapeseed, and Mustard.

3.2. RF

In order to circumvent these problems, we suggested a crop recommendation system that relies heavily on weather data. Crop data acquired by sensors on a regional scale forms the basis of this recommendation system. It considers climate parameters, including temperature, rainfall, humidity, and solar hours, as inputs to make crop recommendations. In order to determine which ML method would work best with our model, we put them all to the test. At last, we developed a recommendation algorithm that considers weather conditions and delivers more precise results. Farmers can boost their agricultural yields and incomes by adopting this technique. Illnesses can be lessened by using this model. In the grand scheme of things, it would be great for the growth of the agricultural industry. Figure 4 illustrates the functionality of the RF algorithm.

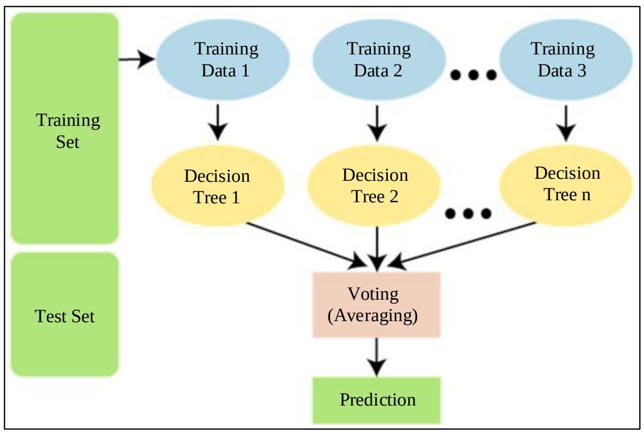

Fig. 4 RF

The RF algorithm functions in two distinct stages. In the initial phase, an RF is created by combining N decision trees. In the second phase, forecasts are generated for each tree created in the initial phase. The following steps and diagram can elucidate the working method:

- Step 1: Select K data points from the training set arbitrarily.
- Step 2: Create decision trees for the selected subsets of data points.
- Step 3: Choose the desired value N to determine the quantity of decision trees to be built.
- Step 4: Recur Steps 1 and 2.
- Step 5: To classify additional data points, ascertain the forecasts of each decision tree and assign the new data points to the category that garners the highest number of votes. This is the culmination of your efforts, where the model is put to use.

3.3. NB Classifier

NB classifiers employ probability to forecast whether an input will be classified into a specific category. The NB algorithm family comprises various classifiers derived from a probability theorem. These classifiers can calculate the likelihood of an input being categorized into one or several categories. When dealing with several categories, the algorithm assesses the likelihood that a data point belongs to each categorization. The system determines the category with the highest likelihood of matching the given text by comparing the probabilities of a match in each category. Numerous firms commonly employ this method to allocate tags to text segments such as email subject lines, customer comments, and articles. Implementation steps:

- Step1: Data acquisition and preparation
- Step 2: The preliminary processing process involves fitting the NB model to the training set.
- Step 3: The model is then used to predict the test results. The accuracy of the predictions is then evaluated by creating a confusion matrix.
- Step 4: Finally, the test set results are visualized.

3.4. Decision Tree

The decision tree algorithm analyzes the dataset to predict and establish its classification. The process commences at the tree's root node, where it evaluates the value of the root attribute relative to the attribute of the record in the dataset. Once the comparison is made, the computer navigates to the next node along the branch. The decision tree algorithm operates iteratively, continuously comparing attribute values as it progresses through each node. This procedure resumes until it reaches the terminal node of the tree, marking the end of the classification process. The method provided below offers a more comprehensive explanation of this iterative procedure.

- Step 1: In building a tree, create the root node, S, which contains the whole dataset.
- Step 2: Use the Attribute Selection Measure (ASM) to find the best attribute in the dataset.
- Step 3: Divide set S into subsets that contain possible values for ideal attributes.
- Step 4: Decide on the tree node with the best quality.
- Step 5: Iteratively generate new decision trees using the dataset subsets obtained in Step 3. This procedure must be carried out until additional categorization becomes impossible.

3.5. KNN

The functioning of K-NN can be elucidated based on the following algorithm:

- Step 1: Decide on a value for K, which is the number of neighbors to be chosen.
- Step 2: Find the KNNs' Euclidean distances.
- Step 3: Evaluate the projected distance to find the K closest neighbors.

- Step 4: Count how often each data item appears among the k closest neighbors.
- Step 5: Sort the updated data by how many neighbors each category has.
- Step 6: The model is ready.

3.6. SVM

One of the most well-known supervised learning methods, SVM, is employed for classification and regression issues. Nevertheless, its main application is in ML classification tasks.

Let us break down the mathematical explanations into manageable steps, which should be sufficient for you to understand how SVM classifiers work:

- Step 1: The SVM algorithm makes class predictions. There is a class labeled one and another with the number -1.
- Step 2: All ML algorithms reduce business problems to equations with unknown variables. For the SVM classifier, the goal is to maximize the margin of error by adjusting a loss function called the hinge loss function, as optimization issues often seek to maximize or minimize something while searching for and adjusting for unknowns.
- Step 3: We introduce the hinge loss function, a cost function that assigns a cost of zero when no class is wrongly predicted for the sake of simplicity. We then add a regularization parameter to formalize these concepts, enhancing our understanding of SVM classifiers.
- Step 4: optimizing weights for SVM, involves applying partial derivatives and other advanced calculus principles to determine the gradients, as is typical for optimization tasks.
- Step 5: Gradients: When the classification is accurate, the regularization parameter is utilized to update the gradients alone, but when misclassification occurs, the loss function is also applied.
- Step 6: Change the gradients when misclassification is absent or present. When the classification is correct, merely update the gradients with the regularization value; when misclassification occurs, step 6 additionally involves utilizing the loss function.

Table 2. Season wise agricultural crop

Area	Crop Cultivated before Monsoon Season	Crop Cultivated in Monsoon Season	Crop Cultivated in the Winter Season
Hooghly, Howrah, Bardhaman	Moong(Green gram)	Jute	Potato
	Chilies	Maize	Masoor
	Watermelon	Soybean	Rapeseed and mustard
	Sesamum indicum	Aus Rice	Peas and beans
		Sugarcane	Boro Rice
			Wheat

4. Experiment

The benchmarks were performed on a system with an Intel Core i7-14th-generation processor, 16 GB of RAM at 2667 MHz, an Nvidia 1650 graphics card, and 1 TB of Solid-State Drive (SSD) space. Anaconda Navigator, a desktop GUI that includes most libraries commonly used for scientific Python work, was utilized for the experiment. We document our code using Google Colab, an offering from Google Research. Colab is an excellent tool for data analysis, teaching, and ML since it lets everyone write and run arbitrary Python code in the browser.

The free cloud-based Jupyter Notebook environment is called Colab. Most importantly, our team members can update the notebooks we produce simultaneously, and no setup is required. We can load numerous popular machine-learning libraries into our notebook using Colab. We conducted experiments to forecast the photos based on their categorization utilizing TensorFlow, keras, matplotlib, OpenCV library, etc.

5. Result

This section explains the results of the research and, at the same time, This analysis's findings point to district-specific and selective strategies for increasing the cultivation of cereals other than rice and potato as West Bengal's best course of action moving forward with the dramatic rise in the irrigation-intensive rice and potato crops. In order to measure water loss during crop production, it is necessary to estimate the evapotranspiration rates. One factor that could affect agricultural output is evapotranspiration, a component of the soil-water balance.

The water required to cultivate crops approximates using evapotranspiration rates. In cases where evapotranspiration rates are low, farmers may change crops to better use existing water resources. Maize has the most significant yield potential compared to other cereals because of its high nutritional density (micro and macronutrients). In most circumstances, maize has a better nutritional productivity than Boro rice. Maize has several uses, including as a biofuel, in the food processing industry, and as a component of chicken feed. The cultivation of maize is seeing rapid growth, with increasing yields in West Bengal. Due to rice's excessive water use, farmers want to grow maize instead. Growing maize provides better returns for farmers and uses only one-fifth as much water as growing rice. Water usage, land use, and nutrient generation are not the only factors that matter when deciding what crops to grow. Factors such as labor needs, storage, transportation expenses, assured markets and prices, seed quality, area expansion, capacity building, and cultivation costs are all in the picture. More comprehensive studies on these topics are required to accomplish the goals of food security and resource preservation. ML algorithms make predictions about manufacturing. The model was developed using sensor data collected from various areas in three districts of West Bengal to forecast agricultural crop production. The production predicts crop income based on the crops' market price or MSP. Figure 5 illustrates a comparative graph of the MSP and market price over 2023-2024 [21-24]. Figure 6, on the other hand, visually represents the total annual income across several crops in the fiscal year 2023-24. The appropriate crop is chosen for each season based on the projected income. Other options are available, but the model is chosen based on its ability to generate the highest income for a specific crop. Figures 9-11 displays the annual income of the agricultural crop in 2023-24, categorized by the several seasons: (a) pre-monsoon, (b) monsoon, and (c) winter.

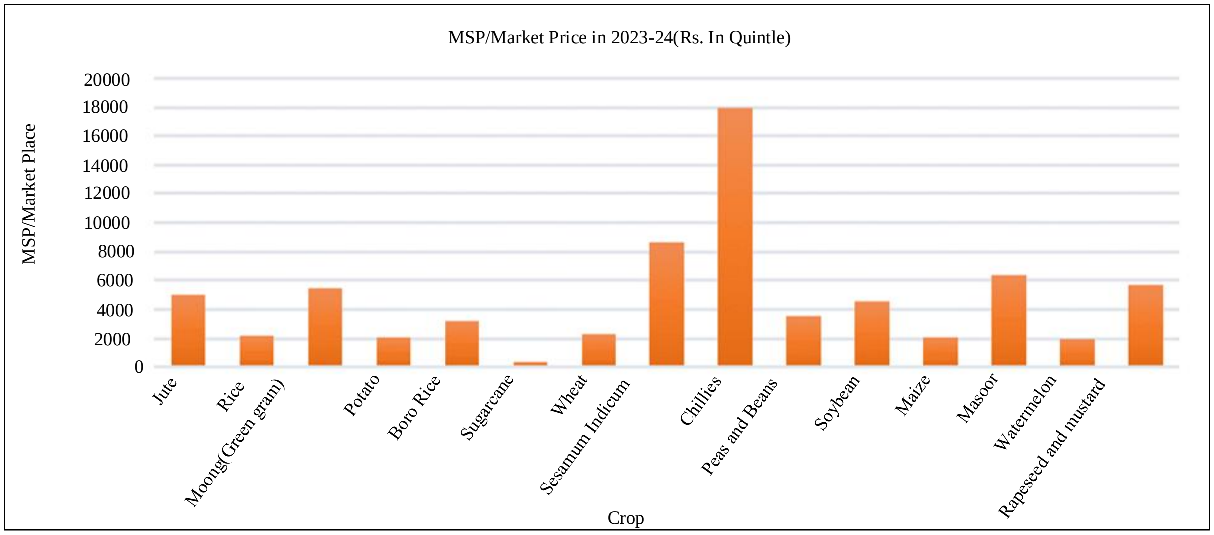

Fig. 5 MSP/ Market price of the agricultural crop in 2023-24

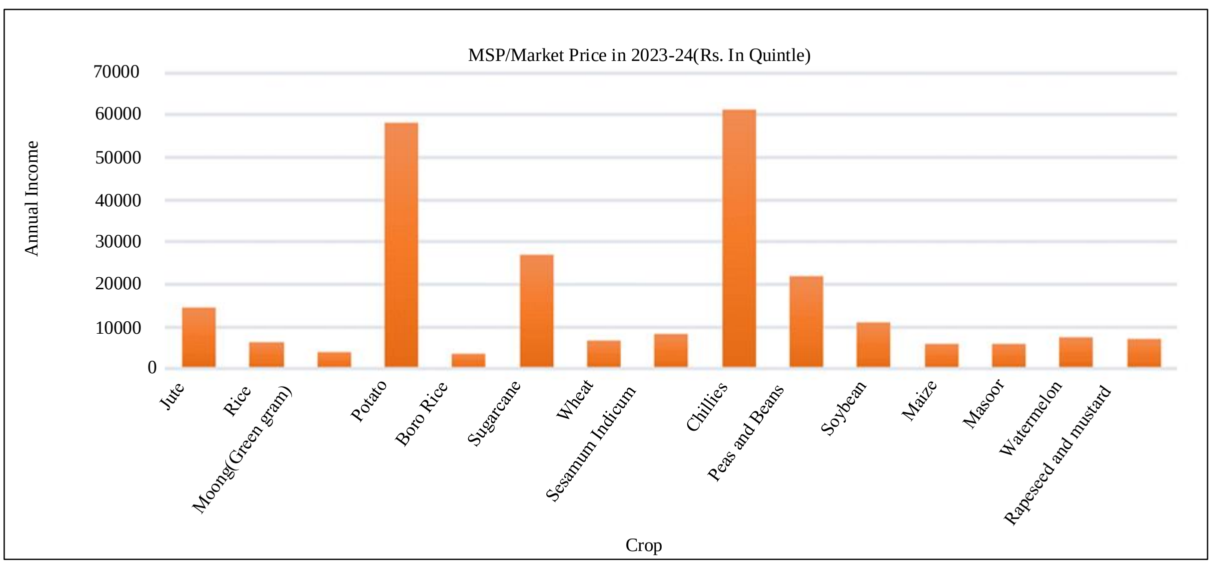

Fig. 6 Total annual income of the agricultural crop in 2023-24

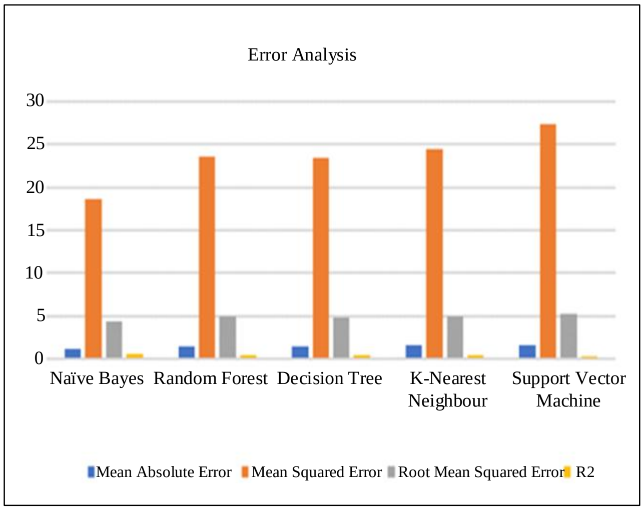

Fig. 7 Error analysis

Table 3. Comparative analysis

ML Methods	Accuracy
Proposed Naïve Bayes	98.97
Proposed Random Forest	98.81
Proposed Decision Tree	98.75
Proposed K-Nearest Neighbour	98.72
Proposed Support Vector Machine	98.53
Random Forest [31]	
Crop selection [31]	97.235
Resource dependency [31]	96.437
Appropriate sowing time [31]	97.647

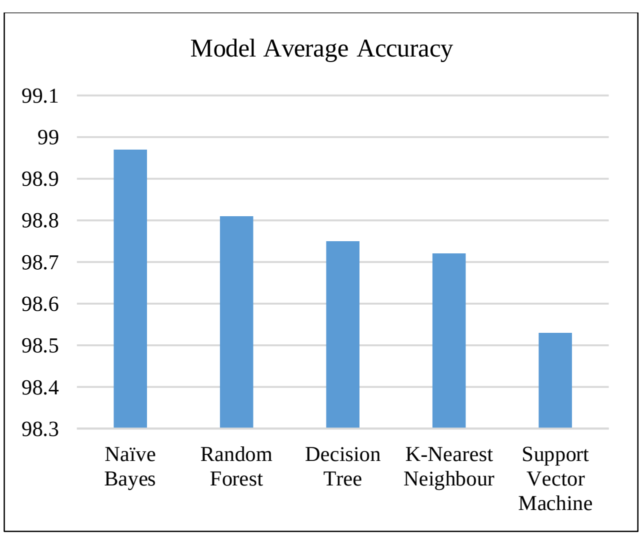

Fig. 8 Model accuracy

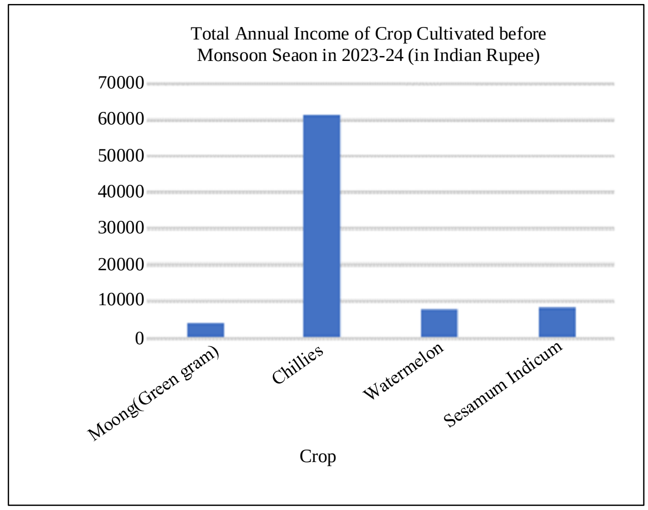

Fig. 9 Season-wise annual income of the agricultural crop in 2023-24 before monsoon

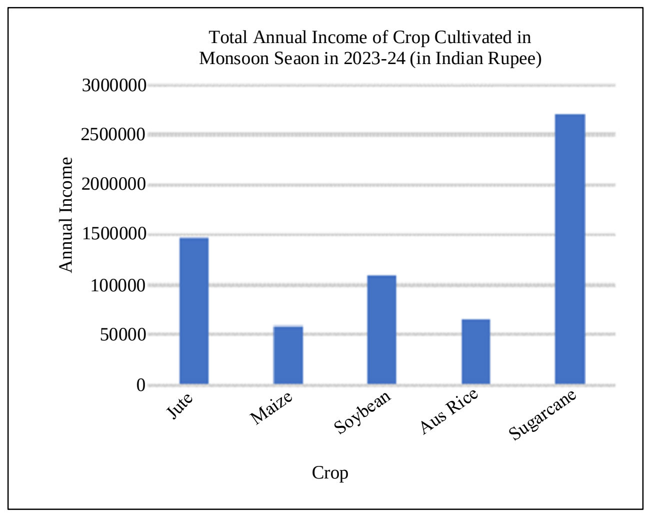

Fig. 10 Season-wise annual income of the agricultural crop in 2023-24 monsoon season

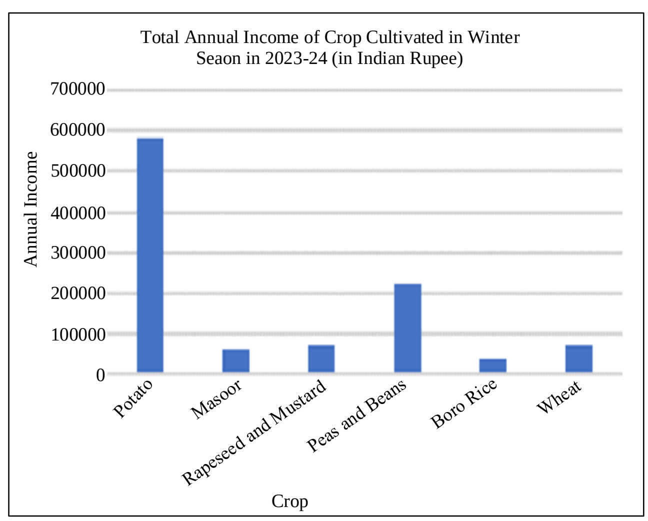

Fig. 11 Season-wise annual income of the agricultural crop in the 2023-24 winter season

According to our study, growing chilies, sugarcane, and potatoes will be profitable before and throughout the monsoon season and winter. Upon analyzing the graphs shown in Figure 7, we observe that the error rate of the Naïve Bayes model is relatively low. At the same time, its accuracy is significantly higher than that of the other four models shown in Figure 8. Therefore, it concluded that the NB model is the most suitable for our investigation. This study [31] proposes the ideal crop for the terrain, its reliance on resources, and the best time for planting. The RF Classifier demonstrated an accuracy of 97.235% for crop selection, 96.437% for predicting resource dependency, and 97.647% for determining the optimal crop sow time [31]. Comparing our proposed model with the RF method [31], as shown in Table 3, it is evident that our model outperforms the RF method.

6. Conclusion

India is renowned for its agricultural prowess and is among the top three global producers of numerous crops. The Indian farmer occupies a central position in the agricultural sector. Nevertheless, most Indian farmers continue to occupy the lowest rung of the social hierarchy. Furthermore, farmers need help determining the most suitable and lucrative crop for their soil despite the limited technological options available, owing to the diverse range of soil types in different geographical areas. Currently, our farmers need to be more efficient in utilizing technology and analysis, which increases the risk of selecting the wrong crop for cultivation, decreasing their income. In order to mitigate such losses, we have devised a system that is user-friendly for farmers.

This system can predict the most suitable crop for a specific piece of land. In order to enhance their income, farmers are taking measures. Future research may necessitate applying the algorithms to more stock markets and examining the influence of different hyperparameters on the acquired outcomes. We aim to explore the feasibility of integrating unstructured textual data into our model.

The analysis may cover a range of criteria, such as investor mood gathered from social networks, policy-related news, and observations made by market analysts. In the future, we will develop a hybrid model exhibiting exceptional performance and accurately estimating crop yields. The crop expenditure needs to be included in this work. In the future, we shall compare the profit by including the crop expenditure and conducting a detailed analysis of the profit.

Acknowledgments

Department of Computer Science at the University of Burdwan, located in Burdwan, West Bengal, India, provided technical support for this research article. Our sincere gratitude goes out to Dr. Sunil Karforma, dean of science at the University of Burdwan, for his assistance in guiding us.

References

[1] Subhrajit Mandal et al., "Adaption of Smart Applications in Agriculture to Enhance Production," *Smart Agricultural Technology*, vol. 7, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[2] Rajesh Yamparla et al., "Crop Yield Prediction Using Random Forest Algorithm," *2022 7th International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore, India, pp. 1538-1543, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] Kranti G. Sapkal, and Avinash B. Kadam, "Random Forest Classifier for Crop Prediction Based on Soil Data," *Journal of Advanced Zoology*, vol. 45, pp. 113-117, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[4] S. Rajeswari, and K. Suthendran, "C5.0: Advanced Decision Tree (ADT) Classification Model for Agricultural Data Analysis on Cloud," *Computers and Electronics in Agriculture*, vol. 156, pp. 530-539, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[5] Mahmoud Y. Shams, Samah A. Gamel, and Fatma M. Talaat, "Enhancing Crop Recommendation Systems with Explainable Artificial Intelligence: A Study on Agricultural Decision-Making" *Neural Computing and Applications*, vol. 36, pp. 5695-5714, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[6] V. Vivek, and Rajat, "K-Nearest Neighbor (KNN), Soil Evaluation, Classifier and Accuracy," *2022 5th International Conference on Contemporary Computing and Informatics (IC3I)*, Uttar Pradesh, India, pp. 698-705, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[7] Vimala Muninarayanappa, and Rajeev Ranjan, "Agriculture Data Analysis Using Parallel k-Nearest Neighbour Classification Algorithm," *International Journal of Reconfigurable and Embedded Systems*, vol. 13, no. 2, pp. 332-340, 2024. [CrossRef] [Publisher Link]

[8] Sudha Bishnoi, and B.K. Hooda, "Decision Tree Algorithms and their Applicability in Agriculture for Classification," *Journal of Experimental Agriculture International*, vol. 44, no. 7, pp. 20-27, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[9] Lei Shi et al., "The Research of Support Vector Machine in Agricultural Data Classification," *Computer and Computing Technologies in Agriculture V*, pp. 265-269, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[10] M. Rajakumaran et al., "Crop Yield Prediction Using Multi-Attribute Weighted Tree-Based Support Vector Machine," *Measurement: Sensors*, vol. 31, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[11] National portal of India, Agriculture. [Online]. Available: https://www.india.gov.in/topics/agriculture

[12] West Bengal State Portal. [Online]. Available: https://wb.gov.in

[13] West Bengal State Agricultural Marketing Board [WBSAMB]. [Online]. Available: http://www.wbagrimarketingboard.gov.in

[14] Government of India, Ministry of Statistics & Programme Implementation, 2024. [Online]. Available: https://mospi.gov.in

[15] Purba Barddhaman, Agriculture, 2024. [Online]. Available: https://purbabardhaman.nic.in/agriculture

[16] District of Hooghly, 2020. [Online]. Available: https://hooghly.nic.in/

[17] Howrah, 2024. [Online]. Available: https://howrah.gov.in/

[18] Press Information Bureau, India, 2024. [Online]. Available: https://pib.gov.in/ & Department of Agriculture & Farmers Welfare, Govt. of India, 2023. [Online]. Available: https://agriwelfare.gov.in/

[19] Indian Agricultural Statistics Research Institute, 2024. [Online]. Available: https://iasri.icar.gov.in/ & Kisaan Helpline Assists Portal, 2024. [Online]. Available: https://www.kisaanhelpline.com

[20] Directorate of Economics & Statistics, Ministry of Agriculture and Farmers Welfare, Government of India, 2024. [Online]. Available: https://eands.da.gov.in/

[21] JioMart, 2024. [Online]. Available: https://www.jiomart.com/c/groceries/fruits-vegetables/fresh-vegetables

[22] Indiamart, 2024. [Online]. Available: https://dir.indiamart.com/impcat/freshvegetables.html

[23] Government of India, Ministry of Agriculture & Farmers Welfare, Department of Agriculture, Cooperation & Farmers Welfare, Directorate of Economics & Statistics, Pocket Book of Agricultural Statistics, 2017. from https://agriwelfare.gov.in/en/Agricultural_Statistics_at_a_Glance

[24] Gurdeep Singh Malhi, Manpreet Kaur, and Prashant Kaushik, "Impact of Climate Change on Agriculture and Its Mitigation Strategies: A Review," *Sustainability*, vol. 13, no. 3, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[25] Sanjeev Kumar, and Sanjay Kumar Upadhyay, "Impact of Climate Change on Agricultural Productivity and Food Security in India: A State Level Analysis," *Indian Journal of Agricultural Research*, vol. 53, no. 2, pp. 133-142, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[26] Sanjay Chaudhary, and P.K. Suri, "Framework for Agricultural e-Trading Platform Adoption Using Neural Networks," *International Journal of Information Technology*, vol. 13, pp. 501-510, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[27] Rajeev Kumar Chaurasia, and Umesh Chandra Jaiswal, 'Spatio-Temporal Based Video Anomaly Detection Using Deep Neural Networks," *International Journal of Information Technology*, vol. 15, pp. 1569-1581, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[28] R. Vikram et al., "Crop Price Prediction Using Machine Learning Naive Bayes Algorithms," *Ubiquitous Intelligent Systems*, pp. 27-34, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[29] Malti Bansal, Apoorva Goyal, and Apoorva Choudhary, "A Comparative Analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory Algorithms in Machine Learning," *Decision Analytics Journal*, vol. 3, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[30] Amudha S. et al., "Deep Learning for Plant Disease Detection and Crop Yield Prediction Based on NPP-WPF Analysis in Smart Agriculture," *2023 7th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Kirtipur, Nepal, pp. 502-510, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[31] Sita Rani et al., "Machine Learning-Based Optimal Crop Selection System in Smart Agriculture," *Scientific Reports*, vol. 13, 2023. [CrossRef] [Google Scholar] [Publisher Link]