Journal of Computer Allied Intelligence
ISSN: 2584-2676
Vol.02(02), April 2024, pp.1-12

Research Article

Making Crop Recommendations using Machine Learning Techniques

P N S S V Charishma[1,*], S. Rupa Lakshmi[2] and M Vijaya Durga[3]

[1,2,3]Department of CSE, VSM College of Engineering, Ramachandrapuram, Andhra Pradesh, 533255, India.

[*]Corresponding Author: P N S S V Charishma. Email: charishma5034@gmail.com

Received: 25/03/2024; Accepted: 20/04/2024.

DOI: https://doi.org/10.69996/jcai.2024006

Abstract: The agricultural crop recommendation system relies to multiple contain characteristics. This study presents a hybrid model for suggesting suitable crops for the southern states of India. The model takes into account multiple factors including Season, temperature, aquifer level, downpours, soil type, fertilisers, and pesticides. A model for recommenders is constructed functioning as a combination of the two utilising machine learning's classifier method. After you input your crop specifications, the algorithm will give you advice on what to grow. Agricultural crop recommendation systems that utilise technology assist farmers in enhancing agricultural productivity by suggesting appropriate crops based on geographical and meteorological elements. A hybrid recommender model might be developed has been proven to be efficient in suggesting an appropriate crop. In order to better manage agricultural output and inform farmers of shifts in crop market prices, it is highly practical to update crop yield production values. The objective of this work is to apply the crop selection approach in order to address various agricultural challenges and issues faced by farmers. Maximising the crop yield rate enhances the Indian economy. Crop quality is assessed by a rating algorithm. This method also enables the identification of the rates of low- and high-quality crops. The use of an ensemble of classifiers allows for better predictive decision-making by deploying a large number of classifiers. Additionally, a ranking procedure is utilised to make decisions and choose the outcomes of the classifiers. This approach is utilised to forecast the expenditure associated with the crop that is produced for subsequent purposes.

Keywords: - Machine Learning, Crop quality, SVM, NB, KNN

1 Introduction

In the realm of advancing technologies, the effective dissemination of knowledge will greatly assist agriculturists in recognising and harnessing their full potential [1-5]. Information sharing refers to the exchange of important and timely information among agriculturists, whether through formal or informal means. Information sharing willingness pertains to the receptive mindset exhibited by agriculturists. This open approach determines the degree and breadth of information exchange [6-8].

We utilise web technologies such as HTML and CSS to construct the web application. We collect data from various sources to create a dataset, which is then used to forecast crop prices. The results are subsequently tested using non-linear methods. Plants are ranked in accordance with predetermined priorities [9-11]. Put information into our app and share it with farmers whose details are stored in our MySQL database [12-15]. We have software that notifies the farmers through text message of any changes. So, farmers don't have to waste time traveling to nearby cities to get the most recent data [16]. For the next two months, we will forecast the price of agricultural commodities using machine learning techniques [17]. Utilizing the SVM, NB, and KNN algorithms, we will endeavor to foretell the monetary outlay associated with agricultural production. Furthermore, a ranking technique is used to determine choices and choose the results of the classifiers.

1.1 Objective

- Aggregation of data sets from many sources.
- The process of data parsing and purification is utilised to transform raw data into processed data.
- The acquired data undergoes machine learning analysis and runtime analysis to create an efficient method for updating crop values.
- The use of a combination of classifiers improves the robustness and efficiency of the model.
- We can make the best possible evaluations by incorporating a ranking methodology into the project.
- Make data collection and user registration easier by developing a web app.
- Having a wider variety of crops that can be farmed throughout the season is the main objective. Maximizing agricultural output while reducing farmers' struggles with crop selection is the goal of the proposed strategy. In order to predict harvest yields, the model takes into account a wide range of parameters, such as precipitation, temperature, area, season, and soil type.

2 Literature Survey

In order to address nutrient deficiencies, fertilisers are applied to the soil to maintain optimal nutrition levels. Indian agriculturists commonly face the challenge of manually determining and applying the appropriate quantity of fertilisers. Over application or under application of fertilisers can have detrimental effects on plant health and decrease crop productivity. This study by Jignasha M. Jethva et al., provides an overview of different data mining frameworks utilised for analysing soil datasets in order to make fertiliser recommendations.

Agriculture is of utmost importance, particularly in emerging nations such as India. The utilization of information technology in agriculture has the potential to revolutionize decision-making processes and enhance farmers' productivity. This document by M.C.S.Geetha consolidates the contributions of multiple authors into a single source, making it a valuable resource for professionals seeking information on the present state of data mining systems and applications in the context of the farming industry.

In this study by S. Srija et al**.,** the concept for developing the Android application Agro Nutri is presented. It is useful for conveying the amount of fertilizer to be applied based on the harvest. The idea is to use the crop's blanked proposal to figure out how much NPK compost to use. The farmers are able to choose the input they want and receive it through this program. By integrating GPRS, Agro Nutri aims to provide location-based nutrient recommendations.

The field of agriculture has not benefited from the technological developments and adaptation. The universal practices should be followed by Indian farmers. A natural idea, machine learning can be used in every field on any input and output. Through conventional software engineering and measurement computations, its competence has been largely proven. In particular, sensor-based frameworks utilized in precision agriculture have benefited from the increased accuracy brought about by machine learning algorithms. In this research, Karandeep Kaur evaluate the various machine learning applications in agriculture. It also sheds light on the challenges encountered by Indian farmers and offers suggestions for resolving these issues.

Humanity will need more food from less resources of land and water during the next decades. The impact on food production of four choice advancement scenarios from the Special Report on Emission Scenarios and the Millennium Ecosystem Assessment is assessed in this study by Uma A Schneider et al.,. The effects of population growth and specialised change on the land and water supply as well as the changes in demand for forestry and agriculture from population growth and economic growth are partially and jointly taken into account. The effects of income on requests for nourishment are recorded using dynamic flexibilities. From 2010 to 2030, there will be a global increase of up to 14% in arable land.Limits on deforestation have a small impact on world food production and sales, but a large impact on the cost of water and land. While changes in projected income have the greatest partial impact on levels of per capita food consumption, an increase in the population drives up total food production. Alterations in the intensities of land management can amplify or dampen the impact of technological change.

Data generated by the science of rural frameworks aids analysts in making informed decisions regarding farming and the solution of complex problems. Over the course of this field's lengthy history, researchers have examined them across a wide variety of frameworks and scales. This essential tool in the field of agricultural frameworks science has been the product of the ideas and work of scholars from many different fields for sixty years. As agricultural researchers contemplate the "people to come" models, data, and educational resources anticipated to address society's inexorably bewildering framework concerns, it is crucial to revisit this past and its exercises to guarantee that we do not re-invent the wheel and instead attempt to consider all facets of associated problems. James W. Jones et al**.,** hope that by providing this historical context, we can better manage the structure and development of the tools and methods that will make up the next generation of agricultural frameworks. Among other things, the development of farming framework demonstrating has been aided by improvements in process-based bio-physical models of domesticated animals and yields, factual models based on verifiable perceptions, financial streamlining, and reproduction models at family unit and local to worldwide scales, among other events and generally innovative advancements in different fields. The features of rural framework models have evolved substantially over the years, depending on the frameworks used, the size of the models, and the diverse range of applications that inspired their creation and application by specialists from various fields. Recent trends in more widespread collaborative effort across institutions, orders, and private sectors have prepared the ground for the next generation of models, databases, learning objects, and choice emotionally supportive networks, which will necessitate significant advancements in rural frameworks science.

The improvement of the agro-normal and financial associated circumstances in the farming sector is mostly dependent on the system models. There are a number of benefits to conducting field and farm inspections, including gathering information and determining which organizational procedures are most effective. With the right foundation, board, environment, and financial data in place, it can find the organization's managers and help them transcend reality. Decision Support Systems (DSSs) help to create the officials and provide the information for the board. To handle the data, these systems do not apply the traditional strategies. Thus, decide on the matter using the clever system ideas. It is expected to be an essential step in understanding agronomic findings, and their application as truly solid decision-making systems for farmers is growing.

This work by o.Elijah et al., presents an IoT and DA agricultural blueprint. It has been thoroughly covered a few zones connected to the relationship of Iot in agriculture. The study of composition shows that there are groups of work being generated on IoT development that can be used to improve plant and animal productivity and operational efficiency. This research has discovered and evaluated the advantages of Iot and DA as well as outstanding issues. A few advantages are expected from IoT for the agriculture sector. Still, there are some matters that need to be managed to make it reasonable for small and medium-scale farmers. The main problems are price and security. Usually, as rivalry grows in the farming area

By banding together in cooperatives, all of the average production bases and farmers will be able to attain economies of scale. Plus, the plans with downstream companies (shippers or retailers) will put producers in a much better position. Secondly, wholesalers mainly serve lower-level distributors and merchants rather than small-scale consumers in the area. There has been a marked improvement in the quality of both resource utilization and logistics as a result of the redesign of transportation modes, which in turn promotes the movement of new chain joint logistics and pays tribute to the new agricultural items' serious endeavor. Revamp everything that is perceived as related to farming. By that point, Xiurong Sun should have gotten the mind-blowing control to streamline the process and thought about the analysis of the control of ordinary items' enormous value.

Support vector machines (SVMs) allow fuzzy inference systems (FISs) to incorporate self-learning. Utilizing a support vector machine identifier (SVMI) and a fast modified varying metric method (MDFP), an SVM-FIS self-learning controller for a three-phase induction machine adjustable speed system has been created. The suggested controller has the learning capacity and generalization performance of SVM in addition to the robustness, independence from the plant model, and adaptive self-learning capabilities of FIS. Zongkai Shao have given comprehensive explanations of the methods used to create the SVM-FIS, MDFP, and SVMI algorithms. Outstanding static and dynamic performances, as well as robust anti-interference capability, show that the suggested control strategy is realistic, accurate, and efficient according to the simulation results.

In this analysis, Mohammad Motiur Rahman used self-organizing maps (SOM) to classify the information factors' relationships with one another. The level of dependence between the associated variable qualities was then determined using the chi-square test. The most important factors affecting product development, yield, and wine quality were daily adverse weather conditions, including the highest and lowest temperature, precipitation, humidity, and wind speed.

3 Existed System

In general, structural price forecasting places substantial computational and data requirements that are significantly greater than what is typically accessible in developing nations. As a result, for statistical forecasting purposes, researchers frequently depend on concise depictions of price processes. Time series modeling is a critical component of contemporary modest price forecasting methods. A model that describes the fundamental relationship is constructed through the analysis of historical observations of the same variable. Advancement and enhancement about chronology predicting tools has received considerable attention over the last few decades. For accurate and timely price forecasting, time series modeling necessitates less onerous data entry. Consequently, a hybrid or ensemble classification model is required to achieve improved classification.

3.1 Disadvantages of Existing System

- Low efficiency exists.
- The current system for recommending crop yield is either hardware-based, which incurs significant maintenance expenses, or it is not readily available.
- Although numerous solutions have been recently suggested, the development of a user-friendly application for crop recommendation still faces unresolved obstacles.
- A greater quantity of repetitive tasks.

3.2 Proposed System

The suggested system updates the agricultural production rate change using data analysis technologies. This paper's overarching goal is to propose using crop selection as a tool to aid farmers and the agricultural sector in addressing various challenges.This contributes to the growth of the Indian economy through the optimisation of agricultural yield rates. An assortment of land conditions exist. The assessment of crop quality is therefore conducted through a ranking procedure. This approach also allows for the inference of the proportion of low-quality and high-quality crops. An ensemble of classifiers enhances prediction accuracy by harnessing the skills of many classifiers. Moreover, a classification procedure is implemented to determine the outcomes of the classifiers during the decision-making process. The aforementioned system is employed to forecast the future cost of fertilisers. Ensembles of classifiers, including the decision tree and random forest classifiers, are utilised in this endeavour. Furthermore, this endeavour employs the Ranking technique.

3.2.1 Advantages of Proposed System:

- Beneficial for individuals situated at a considerable distance from urban centers.
- Enhanced time management. Almost no duplication of effort.

4 Experimental or Alphabetical Materials and Methods

4.1 Random Forest Algorithm

Among machine learning's supervised learning strategies, the Random Forest algorithm stands out. Classification and regression problems are both within the purview of this approach to machine learning. In order to solve a complicated problem and enhance the model's performance, the idea relies on ensemble learning, which combines multiple classifiers. The

Random Forest classifier gets its name from the fact that it uses a large number of decision trees trained on separate dataset subsets. The classifier takes an average of these trees' predictions in an effort to improve its own prediction accuracy. To arrive at its predictions, the random forest algorithm takes into account the results from multiple decision trees and uses the majority vote to determine the outcome. Having more trees in the forest improves accuracy and reduces the likelihood of overfitting. The Random Forest algorithm operates in two stages: first, it constructs a random forest by combining N decision trees, and then it predicts the outcome using each tree formed in the first step.

The following phases might be used to elucidate the operational procedure:

- Take K random samples from the training dataset.
- Connect the chosen data points to decision trees.
- To specify how many decision trees to construct, enter N.
- Carry out the first two steps.
- Decide which group got the most votes, and then find out what each decision tree predicted about the new data points.

4.2 Decision Tree

Decision trees are a kind of supervised learning algorithm mostly used for solving classification issues, however they may also be applied to regression problems. The classifier is structured like a tree, with the dataset's properties at the root, decision rules at the branches, and potential outcomes at the leaf nodes.

The Decision Node and the Leaf Node are the two main kinds of nodes in a decision tree.

Decision nodes are utilized to make a variety of decisions and contain multiple branches, in contrast to leaf nodes, which only show the results of choices and do not have any further branches.

The evaluations or assessments are based on the characteristics of the provided dataset. It is a graphical tool that shows all the options for fixing a problem or choosing an option depending on some criteria.

The "decision tree" moniker comes from the fact that, like a tree, it has a central node from which branches extend outward.

When making trees, the Classification and Regression Tree algorithm (CART) is what you want to use.

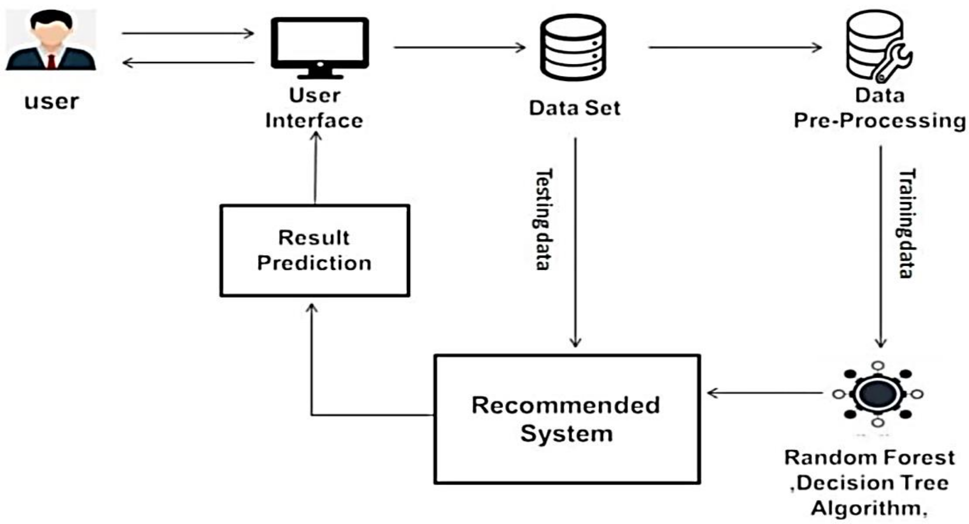
A decision tree functions by presenting a query and then splitting the tree into smaller branches based on the answer provided (either Yes or No).

Here is an algorithm that may enhance your comprehension of the whole process:

- Begin the tree at the root node, as instructed by S, with the whole dataset.
- Utilise the Attribute Selection Measure (ASM) to ascertain the most optimal attribute in the dataset.
- Generate subcategories inside the set S that include probable values for the most exceptional attributes.
- Construct the decision tree node with the most optimal attribute at its center.
- Utilise the dataset subsets created in step 3 to iteratively construct new decision trees.

This procedure should be iterated until the nodes reach a state where they cannot be further categorised. At this stage, the last node is denoted as a leaf node.

4.3



System Architecture

Figure 1: System Architecture

Figure 1 shows System Architecture.

4.3.1 Hardware Requirements

- A computer system equipped with a Pentium Dual Core processor.
- The hard drive has a capacity of 40 gigabytes.
- Screen size: 15 inches,
- Display technology: LED
- Random Access Memory (RAM): 4 gigabytes

4.3.2 Software Requirement

- Windows 7/10 serves as the operating system
- Python is a programming language.

4.4 Modules

- Data Pre-processing
- Admin Login
- Metadata
- Crop Prediction Module
- Crop Recommendation Module

5 Testing

The point of checking a system is to find mistakes. Finding every possible flaw in a work result is what testing is all about. You can test parts, subassemblies, assemblies, and finished goods with it. Software testing makes sure that the system works the way users want it to and doesn't fail in a way that isn't accepted. There are different tests. Each type of test meets a different set of checking needs.

5.1 Types of tests

5.1.1 Unit testing

Unit testing requires creating test cases to verify software logic and input-output validation. Validate all internal code flow and decision branches. Individual application software units are tested.It follows unit completion before integration. This intrusive structural testing requires building knowledge. Unit tests check a business process, a program, or the setup of a system at the component level. Unit tests make sure that each step in a business process follows written instructions and has clear inputs and outcomes that are expected.

Integration testing is used to verify that software components work together as a single application. Event-driven testing focuses on screen or field outcomes. Integration tests show that the components were satisfied, but the combination was right and consistent. Integration testing focuses on component combination issues.

Functional tests make sure that certain features work as expected based on technical and business needs, as well as system manuals and user guides.

5.1.2 System Test

Through the use of system tests, we can ensure that the integrated software system meets all of our requirements. It puts a configuration through its paces to guarantee accurate outcomes. System tests are illustrated by the configuration-oriented system integration test.Using process flows and descriptions, system testing centers on pre-defined integration points and links in the process.

5.1.3 White Box Testing

This kind of testing requires a software tester who has a certain level of understanding and familiarity with the internal mechanisms, organisation, and programming language of the product being tested. It serves a certain function. It is used to test areas that cannot be reached from the perspective of the black box.

5.1.4 Black Box Testing

Testing software in "black box" mode means the tester has no idea what the code, structure, or language used to test it actually is. Like other types of tests, black box tests necessitate a well-defined source document, such as a requirements or design document, from which they are derived. The test program is used like a black box during these tests. You can't "see" into it. The test doesn't think about how the software works; it just gives data and gets results.

6 Results and Discussion

The data processing technology is used in this project to let people know about changes in the rate. In addition, a ranking method is used to make the choice of which classifier results to use. This method is used to guess how much the crop yield will cost.

This study takes into account a total of eighty-eight weather points over a six-year span, from February 2020 to December 2023. A weather point is a physical location that displays current and historical weather conditions for a certain area and time. The weather point in this work is a combination of sun hours, humidity, wind speed and direction, temperature, and wind direction. We trained the model with 60% of the samples and used the remaining 40% to make predictions using the data we had. Maximum temperature, minimum temperature, and amount of precipitation are the three standards used to forecast the most suitable crops for each season. Figures 2, 3, 4 and 5 show the results for these criteria' predictions, correspondingly.

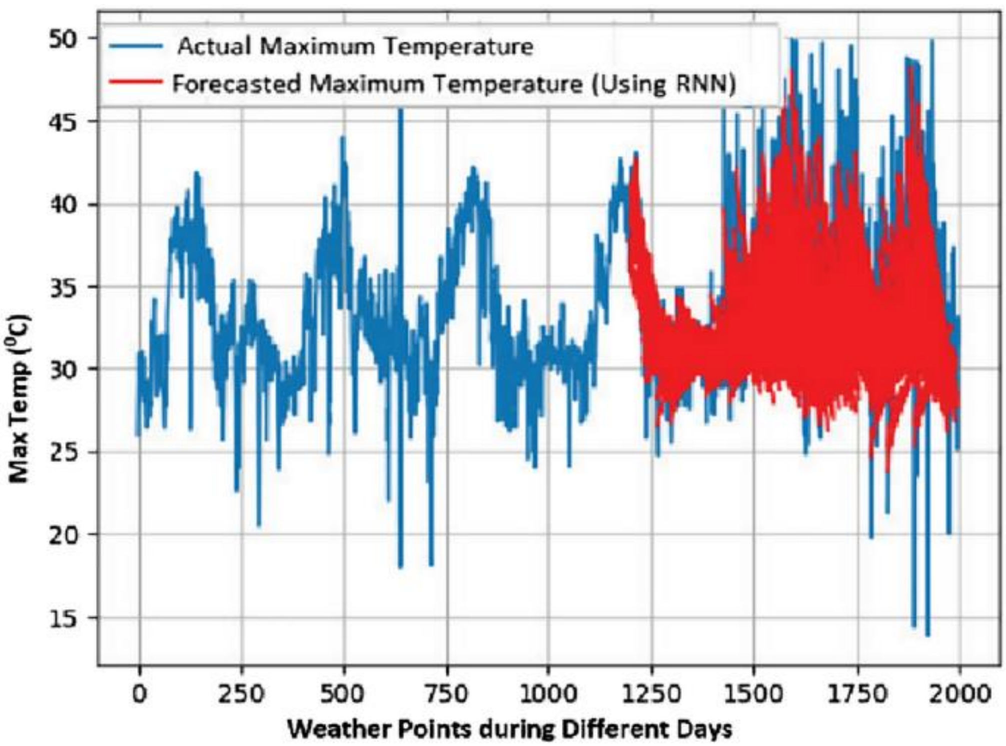


Figure 2: Maximum Temperature Prediction

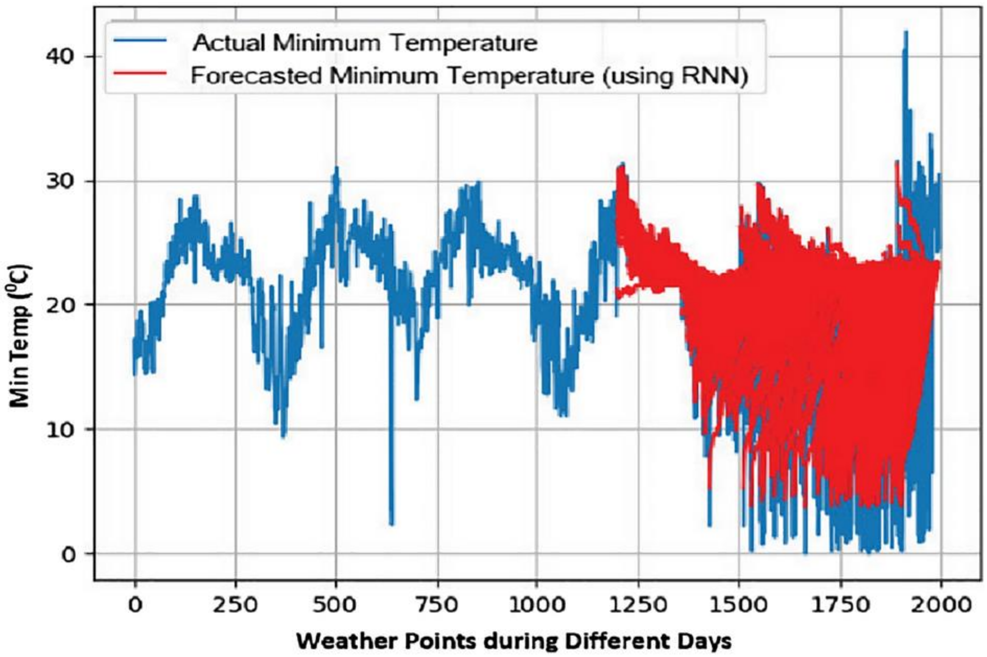


Figure 3: Minimum Temperature Prediction

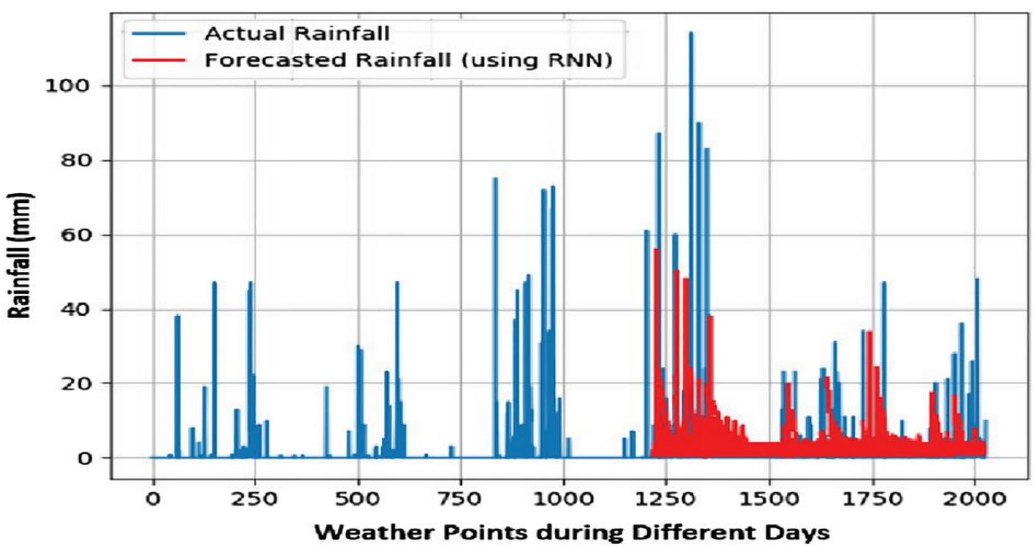


Figure 4: Rainfall prediction

After the weather forecast is finished, the following step is to choose an appropriate crop based on the soil parameters and expected weather. The Random Forest Classification algorithm is used to implement the crop selection process in this work. Table 1 displays the results of the Random Forest classifier's evaluation of the crop's suitability for the Rabi season, taking into account weather and soil parameters.

Table 1: Results of proposed classifier

Selected appropriate crop	Dependency	Appropriate sowing time
Red gram	Rainfall	October-1st Week
Rice	Irrigation	December-3rd Week
Maze	Irrigation	November-3rd week

The results of the Random Forest Classifier's accuracy tests are shown in Table 2. When comparing the used classifier's performance to that of Decision Tree, the Random Forest Classifier comes out on top. Using the specified data size, a random forest classifier takes 5.34 seconds to construct the model.

Table 2: Accuracy Comparison

ML algorithm	Accuracy (%)		
	Crop selection	Dependency (Irrigation/Rainfall)	Appropriate sowing time
Random forest classifier	95.63	95.26	96.52
Decision tree	93.57	92.16	94.56

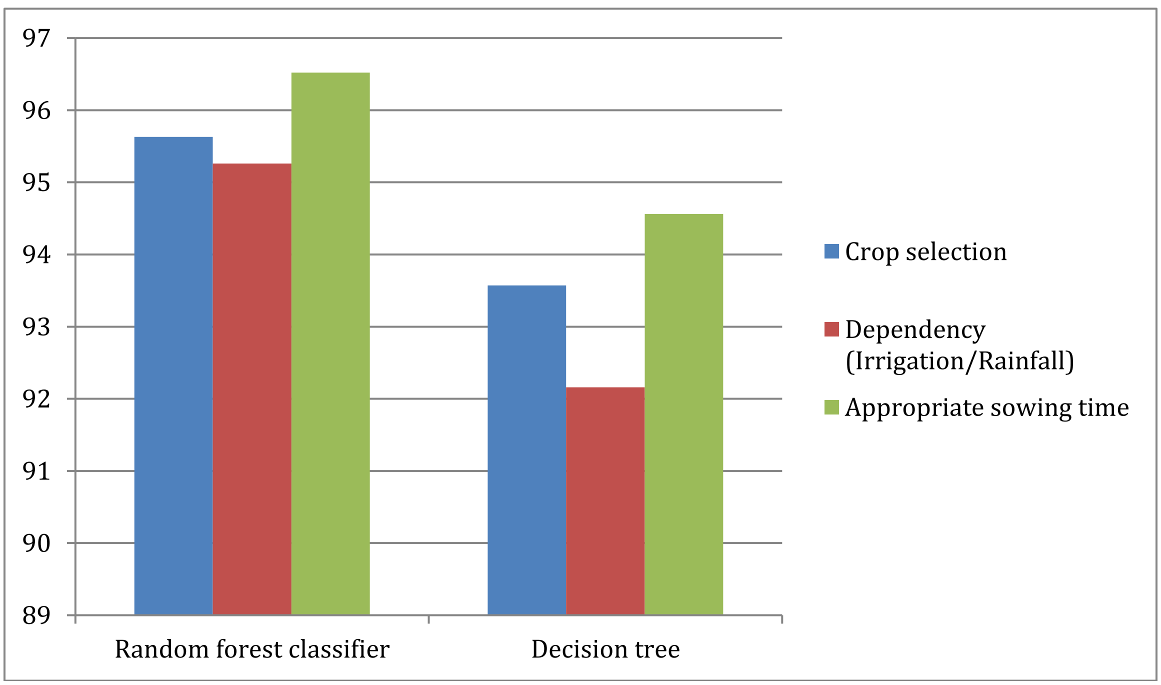


Figure 5: Accuracy Comparison

7 Conclusion

How much and what kinds of information are shared depend on how open people are. The production of crops can be improved with the help of big data research technology. This project suggests a new, smart system for predicting the prices of crops. To make predictions, the main idea is to use a group of algorithms. You can make better guesses when you use an ensemble of classifiers. This is because you use more than one classifier. In addition, a ranking method is used to make the choice of which classifier results to use. This method is used to guess how much the crop will cost in the future.

By suggesting the right crop based on a number of factors, the answer will help farmers increase crop yields, protect cultivated fields from soil erosion, and use less fertilizer in crop production. From a geographical, environmental, and economic point of view, this would give a full forecast.

Acknowledgement: Not Applicable.

Funding Statement: The author(s) received no specific funding for this study.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. N. Abdulla, M. Demirci and S. Ozdemir, "Design and evaluation of adaptive deep learning models for weather forecasting," *Eng. Appl. Artif. Intell*, vol. 116, no. 105440, 2022.
2. A. Haridasan, J. Thomas and E. D. Raj, "Deep learning system for paddy plant disease detection and classification," *Environ. Monit. Assess*, vol. 195, no. 120, 2023.
3. G. Sambasivam and G. D. Opiyo, "A predictive machine learning application in agriculture: Cassava disease detection and classification with imbalanced dataset using convolutional neural networks," *Egypt. Inform J*, vol. 22, pp. 27–34, 2021.
4. R. Manikandan, G. Ranganathan and V. Bindhu, "Deep learning based IoT module for smart farming in different environmental conditions," *Wirel. Pers. Commun*, vol. 128, pp.1715–1732, 2023.
5. A. Gumuscu, M. E. Tenekeci and A. V. Bilgili, "Estimation of wheat planting date using machine learning algorithms based on available climate data," *Sustain. Comput. Inform. Syst*, vol. 28, no. 100308, 2020.
6. A. Chlingaryan, S. Sukkarieh and B. Whelan, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review," *Comput. Electron. Agric*. vol. 151, pp. 61–69, 2018.
7. Jignasha, M. Jethva, Nikhil Gondaliya and Vinita Shah, "A Review on Data Mining Techniques for Fertilizer Recommendation International," *Journal of Scientific Research in Computer Science Engineering and Information Technology*, vol. 2, no. 6, 2017.
8. M. C. S. Geetha, "A Survey on Data Mining Techniques in Agriculture," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 3, pp. 887-892, 2015.
9. S. Srija, R. Geetha Chanda, S. Lavanya and Dr. M. Kalpana, "AgroNutri Android Application," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 5, no. 5, 2016.
10. Karandeep Kaur, "Machine Learning: Applications in Indian Agriculture," *International Journal of Advanced Research in Science, Communication and Technology*, vol. 3, no.1, 2023.
11. A. Uwe, Schneider, Petr Havlik, Erwin Schmid, Hugo Valin, Aline Mosnier et al., "Impacts of population growth, economic development, and technical change on global food production and consumption," *Agricultural Systems*, vol. 104, no. 2, pp. 204-215, 2011.
12. W. James Jones, M. John Antle, O. Bruno Basso, J. Kenneth Boote, T. Richard Conant et al., "Brief history of agricultural systems modeling," *Agricultural Systems*, vol. 155, pp. 240-254, 2017.

13. S. Rajeswari, K. Suthendran and K. Rajakumar, "A smart agricultural model by integrating IoT, mobile and cloud-based big data analytics," *2017 International Conference on Intelligent Computing and Control (I2C2)*, Coimbatore, India, pp. 1-5, 2017.
14. O. Elijah, T. A. Rahman, I. Orikumhi, C. Y. Leow and M. N. Hindia, "An Overview of Internet of Things (IoT) and Data Analytics in Agriculture: Benefits and Challenges," *in IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3758-3773, 2018.
15. X. Sun, J. Zhang, C. Wang and T. Zhang, "Circulation mode selection based on cost analysis," *2017 4th International Conference on Industrial Economics System and Industrial Security Engineering (IEIS)*, Kyoto, Japan, pp. 1-4, 2017.
16. Zongkai Shao, "Support vector machine-based fuzzy self-learning control for induction machines," *2010 International Conference on Computer and Communication Technologies in Agriculture Engineering*, Chengdu, China, pp. 12-16, 2010.
17. M. M. Rahman, N. Haq and R. M. Rahman, "Machine Learning Facilitated Rice Prediction in Bangladesh," *2014 Annual Global Online Conference on Information and Computer Technology*, Louisville, KY, USA, pp. 1-4, 2014.