

A Predictive Analytics Framework for Crop Recommendation Using Ensemble Learning Techniques

Prasanth M¹, Nagasundaram S²

¹MCA Student, Department Of Computer Application-PG VISTAS

²Assistance professor, Department Of Computer Application-PG VISTAS

DOI : <https://doi.org/10.51583/IJLTEMAS.2025.140400065>

Abstract: This paper presents the development and implementation of a data-driven Crop Recommendation System designed to assist farmers in making informed decisions regarding optimal crop selection. The system is deployed as a responsive web application featuring a user-friendly interface, allowing easy access and interaction for end-users. The application comprises three main modules: the Home Page, which introduces the system and outlines its significance; the User Guide Page, which provides step-by-step instructions for effective usage; and the Crop Recommendation Page, where users input specific environmental and soil parameters to receive crop suggestions tailored to their conditions. The core of the recommendation engine is a Random Forest Classifier, selected for its high accuracy and robustness in handling agricultural datasets. The model was trained on a comprehensive dataset that includes critical features such as nitrogen, phosphorus, potassium levels, temperature, humidity, pH, and rainfall. By analyzing these inputs, the system delivers personalized crop recommendations that align with the user's agro-climatic conditions, thereby enhancing decision-making for better agricultural outcomes. The model's performance was evaluated using standard classification metrics, confirming its effectiveness in providing reliable crop suggestions. This intelligent system addresses a crucial need in modern agriculture for precision-based recommendations that can lead to increased crop productivity and resource optimization. The solution promotes sustainable farming practices by leveraging machine learning techniques and facilitating easy access through a web-based platform.

Keywords: Crop Recommendation System, Machine Learning, Random Forest Classifier, Precision Agriculture, Web Application, Smart Farming, Agricultural Decision Support, Soil and Climate Parameters, Data-Driven Farming

I. Introduction

Agriculture is a critical sector that sustains the livelihood of a significant portion of the global population. Despite technological advancements, many farmers, especially in developing regions, still rely on traditional methods and intuition for crop selection. This often leads to suboptimal crop yields, inefficient use of resources, and economic losses. One of the major challenges faced by farmers today is the lack of access to scientific and data-driven tools that can guide them in choosing the most suitable crops based on environmental and soil conditions.

With the increasing availability of agricultural datasets and the advancement of machine learning techniques, it has become feasible to build intelligent systems that can assist farmers in making informed decisions. In this context, crop recommendation systems offer a promising solution. By analyzing various soil and climatic parameters, such systems can provide personalized recommendations, thereby enhancing crop productivity and promoting sustainable farming practices.

This paper presents a web-based Crop Recommendation System that leverages a Random Forest Classifier to suggest the most appropriate crops to cultivate. The system takes into account essential parameters such as nitrogen, phosphorus, potassium levels, soil pH, temperature, humidity, and rainfall. The choice of the Random Forest algorithm is motivated by its accuracy, scalability, and ability to handle non-linear relationships within agricultural data.

The proposed system is designed with a focus on usability and accessibility. It features an intuitive web interface with three primary modules: Home, User Guide, and Crop Recommendation. The Home page provides an overview of the application and its benefits, while the User Guide offers instructions to help users navigate the system efficiently. The Crop Recommendation module allows users to input relevant data and obtain crop suggestions instantly.

By integrating data-driven methods into crop selection, the proposed system aims to empower farmers with actionable insights, reduce dependency on guesswork, and contribute to improved agricultural productivity. This work highlights the potential of machine learning applications in agriculture and sets the foundation for future developments in intelligent farming solutions.

Existing System

In the current agricultural landscape, many farmers depend on traditional knowledge, experience-based decision-making, and government-issued crop calendars for crop selection. These conventional methods do not consider the dynamic nature of environmental and soil parameters, leading to inconsistent crop yields and inefficient resource usage. While such approaches may have worked in the past, they are increasingly inadequate in the face of modern agricultural challenges like climate change, soil degradation, and fluctuating rainfall patterns.

Some existing digital solutions and mobile applications attempt to assist farmers in crop planning. However, these systems often suffer from the following limitations:

- **Lack of Personalization:** Recommendations are usually generic and do not account for the farmer's specific soil conditions, climate, or geography.
- **Limited Use of Machine Learning:** Many systems use basic rule-based algorithms or simple data lookups instead of advanced predictive models.
- **Static Datasets:** Most existing systems rely on fixed datasets that are not frequently updated, reducing the reliability of recommendations over time.
- **Poor User Interface:** Some applications lack an intuitive interface, making them difficult for non-technical users, especially in rural areas.
- **No Real-Time Interaction:** Existing platforms typically do not allow users to input real-time environmental data (e.g., current pH, rainfall, or temperature) to get updated recommendations.

These shortcomings emphasize the need for a more robust and intelligent system that can deliver **accurate, personalized, and data-driven crop recommendations**. The proposed system addresses these gaps by leveraging **ensemble learning techniques**, specifically the **Random Forest algorithm**, within a responsive and accessible web-based interface. This ensures that recommendations are not only accurate but also easy to access and interpret, empowering farmers to make better decisions and improve crop productivity.

An existing system called Climate Field View, which is a crop recommendation system based on machine learning. Climate Field View: Climate Field View by The Climate Corporation is an advanced digital farming platform that offers crop recommendation capabilities. It uses machine learning algorithms to analyse field data, including soil moisture, weather conditions, and crop health, to provide personalized recommendations for planting.

II. Proposed System

The proposed crop recommendation system aims to provide farmers with an intelligent and data-driven solution for making informed decisions about crop selection. The system utilizes a combination of machine learning techniques and agricultural data to generate personalized crop recommendations. At its core, the system collects and analyzes various data inputs of soil characteristics. These inputs serve as the foundation for training machine learning models, which are capable of identifying patterns and relationships between different variables. The system employs advanced algorithms, such as Random Forest (RF), Support Vector Machine (SVM), or other suitable models, to process the collected data and generate accurate predictions. These predictions are based on the relationships identified during the training phase, allowing the system to recommend the most suitable crops for specific farming conditions. To ensure user-friendliness and accessibility, the crop recommendation system is designed as a web application with an intuitive interface. The application consists of multiple pages, including a Home page providing an overview of the system's benefits, a User Guide page with instructions on how to use the application, and a Crop Recommendation page where farmers can input their specific crop features and receive personalized recommendations. In the development of the system, a major focus is placed on the implementation and evaluation of the chosen machine learning models. The system utilizes a Random Forest classifier or other suitable algorithms to generate crop recommendations based on the trained model. The trained model is saved as a pickle file, allowing for efficient and convenient future use. By leveraging the power of machine learning and incorporating relevant agricultural data, the proposed crop recommendation system offers a reliable and effective tool for farmers to optimize their crop selection process. It considers environmental and soil factors to provide accurate and personalized recommendations, empowering farmers to make informed decisions and potentially improve their yield and profitability. The limitation in the existing system is improved in proposed system, that the data are collected and analyse simultaneously.

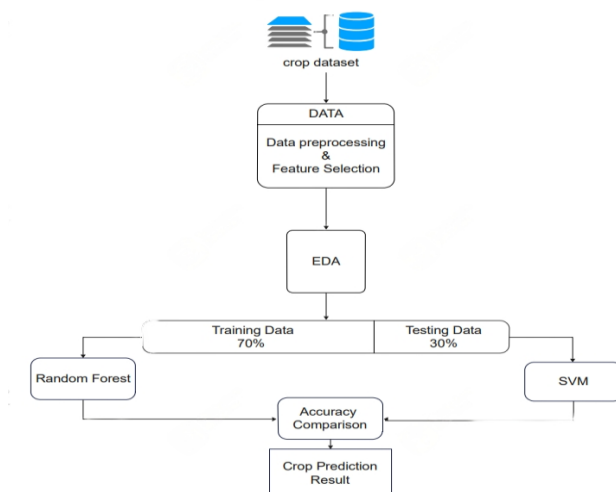


Fig 1. Architecture Diagram

III. METHODOLOGY

- **Data collection:** Collecting data on crop characteristics such as nutrient levels, temperature, humidity, pH level, and rainfall, as well as the corresponding crop yield or label. The dataset used in the code is likely a pre-existing dataset used for this purpose.
- **Data preprocessing:** Preprocessing the collected data to remove duplicates, replace missing values, and handle any outliers or anomalies in the data.
- **Feature selection:** Selecting the most relevant features for the model using techniques such as correlation analysis.
- **Model creation:** Creating a machine learning model using algorithms such as SVM, random forest, or logistic regression, and training it on the preprocessed data.
- **Model evaluation:** Evaluating the accuracy of the model using metrics such as accuracy, precision, recall, and F1-score. Cross-validation can also be used to ensure the model is not overfitting the training data.
- **Deployment:** Once a satisfactory model has been built, it can be deployed as a crop recommendation system. The system could take inputs such as nutrient 12 levels, temperature, humidity, pH level, and rainfall, and provide recommendations on the most suitable crop to grow based on the trained model.
- **Maintenance:** The deployed model would need to be maintained over time, as new data becomes available or changes occur in the environment. The model may need to be retrained periodically to ensure its accuracy and effectiveness in making crop recommendations.

Machine Learning

It is the scientific study of algorithms and statistical models that computer systems use to effectively perform a specific task without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model of sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task. Machine learning algorithms are used in a wide variety of applications, such as email filtering, and computer vision, where it is infeasible to develop an algorithm of specific instructions for performing the task. Machine learning is closely related to computational statistics, which focuses on making predictions using computers. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a field of study within machine learning, and focuses on exploratory data analysis through unsupervised learning.

Random Forest

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

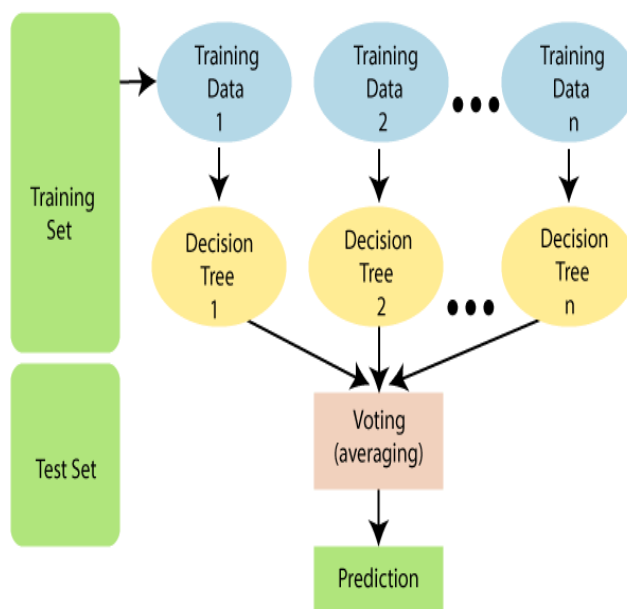


Fig 2. Random Forest

Support Vector Machine

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine.

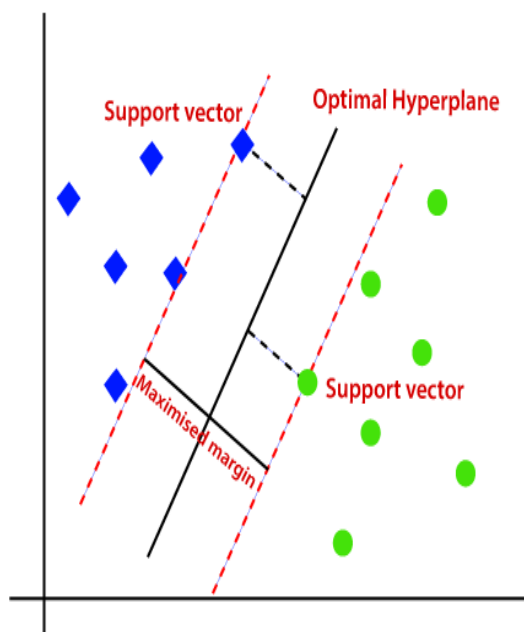


Fig 3. Support Vector Machine

IV. Implementation

Implementation includes all those activities that take place to convert from old system to the new. The new system may be totally new, replacing an existing system or it may be major modification to the system currently put into use. In this project, design is done by using the Python.

Data collection and Pre-processing

Text based classification dataset is collected from the Kaggle website. Dataset is preprocessed to clean the noise data by removing the Null values.

Training Data and Test Data

Machine learning concept includes learning some information or we can say properties from dataset and testing properties of another data can be classified. Mostly we are splitting data in two categories as training set and testing set. Training set (70%) is the set of learning properties from given data. Testing set (30%) is to check properties of the data given for query which may include some different properties. Depending on Training set with different classifier we may get different results with different accuracy. The classifier of type deep learning which is having highest classification accuracy can be considered for further analysis.

Modules

- Data Analysis
- Descriptive Analytics
- Predictive Analytics
- Web App Building

Data Analysis

Data analysis is the process of cleaning, changing, and processing raw data and extracting actionable, relevant information that helps businesses make informed decisions. The procedure helps reduce the risks inherent in decision-making by providing useful insights and statistics, often presented in charts, images, tables, and graphs.

Descriptive Analytics

Descriptive Analytics help describe and understand the features of a specific data set by giving short summaries about the sample and measures of the data. The most recognized types of descriptive statistics are measures of center: the mean, median, and mode, which are used at almost all levels of math and statistics. The mean, or the average, is calculated by adding all the figures within the data set and then dividing by the number of figures within the set.

Predictive Analytics

Random Forest:

Random Forest is a machine learning algorithm that is commonly used for both classification and regression tasks. It is an ensemble learning method that constructs multiple decision trees and combines their predictions to produce the final output. In a Random Forest model for classification, each tree is trained on a random subset of the training data and a random subset of the input features. During prediction, each tree in the forest independently predicts the class of a given input, and the final prediction is made by combining the individual tree predictions through voting or averaging. Random Forest models are particularly useful for handling large datasets with many input features, as they are less prone to overfitting than single decision trees. They are also able to handle missing data and noisy features effectively, making them a popular choice for predictive analytics and data mining tasks.

```
In [46]: from sklearn.ensemble import RandomForestClassifier

RF = RandomForestClassifier(n_estimators=20, random_state=0)
RF.fit(X_train, y_train)

predicted_values = RF.predict(X_test)

x = metrics.accuracy_score(y_test, predicted_values)
acc.append(x)
model.append('RF')
print("RF's Accuracy is: ", x)

print(classification_report(y_test, predicted_values))
```

	precision	recall	f1-score	support
Beans	0.11	0.16	0.13	19
blackgram	1.00	0.92	0.96	13
chickpea	1.00	1.00	1.00	20
coffee	1.00	1.00	1.00	22
cotton	1.00	1.00	1.00	24
cowpeas	1.00	1.00	1.00	21
groundnuts	0.13	0.14	0.13	22
jute	0.82	1.00	0.90	14
kidneybeans	0.00	0.00	0.00	24
lentil	1.00	1.00	1.00	23
maize	0.95	1.00	0.98	20
mothbeans	0.14	0.13	0.13	23
mungbean	1.00	1.00	1.00	17
peas	1.00	1.00	1.00	13

Fig 4. Predictive Analytics – Random Forest

Support Vector Machine:

Support Vector Machine (SVM) is a type of machine learning model that is commonly used for classification and regression tasks. SVM is a supervised learning algorithm that learns to classify data points by finding the hyperplane (decision boundary) that maximizes the margin between the classes. SVM works by transforming the input data into a high-dimensional feature space and finding the optimal hyperplane that separates the data into different classes. The hyperplane is chosen to maximize the margin between the classes, which is defined as the distance between the hyperplane and the closest data points from each class. SVM can be used for both linear and nonlinear classification tasks. In linear SVM, a linear hyperplane is used to separate the classes, while in nonlinear SVM, a nonlinear function is used to transform the data into a higher-dimensional space, where a linear hyperplane can be used to separate the classes. SVM has several advantages over other machine learning models, such as its ability to handle high-dimensional data, its robustness to outliers, and its effectiveness with small datasets. However, SVM can be computationally expensive and requires careful tuning of its parameters to achieve optimal performance.

```
In [118]: from sklearn.svm import SVC

svm = SVC(kernel='rbf', random_state=0)
svm.fit(X_train, y_train)

predicted_values = svm.predict(X_test)

x = metrics.accuracy_score(y_test, predicted_values)
acc.append(x)
model.append('svm')
print("svm's Accuracy is: ", x)

print(classification_report(y_test, predicted_values))
```

	precision	recall	f1-score	support
Beans	0.36	0.74	0.48	19
blackgram	1.00	1.00	1.00	13
chickpea	0.42	1.00	0.59	20
coffee	1.00	1.00	1.00	22
cotton	1.00	1.00	1.00	24
cowpeas	0.00	0.00	0.00	21
groundnuts	0.49	1.00	0.66	22
jute	0.67	1.00	0.80	14
kidneybeans	0.38	0.12	0.19	24
lentil	1.00	1.00	1.00	23
maize	1.00	1.00	1.00	20
mothbeans	0.00	0.00	0.00	23
mungbean	0.45	1.00	0.62	17
peas	0.38	0.77	0.51	13
pigeonpeas	0.00	0.00	0.00	17

Fig 5. Predictive Analytics – Support Vector Machine

Web App Building

A web application can be developed for several uses, which can be used by anyone like it can be used as an individual or as a whole organization for several reasons. Here we are creating the Web app using Streamlit library.

V. Conclusion

Machine learning has been widely used as a powerful tool to solve problems in the agriculture environment, particularly in crop recommendation systems. By analyzing parameters such as n, p, k, temperature, pH, humidity, rainfall, and soil type, machine learning algorithms can predict which crops are most suitable for a given area and optimize resource allocation. Despite this, several challenges remain in fully applying machine learning approaches in this field: (1) Machine learning is usually dependent on large amounts of high-quality data. Obtaining sufficient data with high accuracy in crop prediction and management systems is often difficult owing to the cost or technology limitations. (2) As the conditions in crop prediction and management systems can be extremely complex, the current algorithms may only be applied to specific systems, which hinders the wide application of machine learning approaches. (3) The implementation of machine learning algorithms in practical applications requires researchers to have certain professional background knowledge.

Future Enhancement

To overcome the above-mentioned challenges, the following aspects should be considered in future research and engineering practices: (1)Integration with IoT and sensor technologies: By integrating machine learning algorithms with IoT devices and sensors in the field, it is possible to collect real-time data on environmental conditions such as temperature, humidity, and rainfall, which can then be used to make more accurate crop recommendations.(2)Incorporation of satellite and remote sensing data: The use of satellite and remote sensing data can provide a broader view of the environment and help to identify patterns and trends that may not be visible at ground level. Integrating this data with machine learning algorithms can enhance the accuracy of crop recommendations.(3)Application of deep learning techniques: Deep learning algorithms such as convolutional neural networks and recurrent neural networks have shown promise in image recognition and sequence prediction tasks, respectively. These techniques could be applied to crop recommendation systems to improve the accuracy of crop identification and yield prediction.

Reference

1. "Crop Recommendation System Using Data Mining Technique" by Ashish Kadam and Manisha Sharma. International Journal of Scientific and Research Publications, Volume 7, Issue 11, November 2017.
2. "Crop Recommendation System Using Data Mining Techniques for Precision Agriculture" by M. Mohamed Thasin, S. Anandhi, and K. Santhi. International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-8 Issue-9, July 2019.
3. "A Machine Learning Based Crop Recommendation System for Precision Agriculture" by S. Deepika, S. Aruna, and K. Gowtham Kumar. International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-9 Issue-1, November 2019.
4. "Crop Recommendation System Using Machine Learning Algorithms" by Ravi Kumar K V and P K Giridhar. International Journal of Engineering and Advanced Technology (IJEAT), Volume-9 Issue-5, June 2020.
5. "Crop Recommendation System using Hybrid Machine Learning Technique" by T. Saravanakumar, V. Vanitha, and R. Amalraj. Indian Journal of Science and Technology, Volume-13 Issue-24, June 2020.
6. "Crop Recommendation System using Machine Learning Algorithms for Precision Agriculture" by S. S. Gupta, S. Bhatt, and M. C. Trivedi. International Journal of Computer Sciences and Engineering, Volume-8 Issue-2, February 2020.
7. Mayank Champaneri, ChaitanyaChandvidkar, DarpanChachpara, MansingRathod, "Crop yield prediction using machine learning" International Journal of Science and Research, April 2020.
8. Pavan Patil, VirendraPanpatil, Prof.ShrikantKokate, "Crop Prediction System using Machine Learning Algorithms", International Research Journal of Engineering and Technology, Feb 2020.
9. Ramesh Medar, Shweta, Vijay S. Rajpurohit, "Crop Yield Prediction using Machine Learning Techniques", 5th International Conference for Convergence in Technology, 2019.
10. TruptiBhange, Swati Shekapure, Komal Pawar, HarshadaChoudhari, "Survey Paper on Prediction of Crop yield and Suitable Crop", International Journal of Innovative Research in Science, Engineering and Technology, May 2019.
11. E. Manjula, S. Djodiltachoumy, "A Modal for Prediction of Crop Yield", International Journal of Computational Intelligence and Informatics, March 2017.
12. Nishit Jain, Amit Kumar, SahilGarud, Vishal Pradhan, Prajakta Kulkarni, "Crop Selection Method Based on Various Environmental Factors Using Machine Learning", International Research Journal of Engineering and Technology (IRJET), Feb 2017.
13. Rakesh Kumar, M.P. Singh, Prabhat Kumar, J.P. Singh, "Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique", 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, T.N., India., May 2015.
14. Liu Qingyuan, and Wang Tianchuan, "Market price theory and practice". Dalian: Northeast University of Finance and Economics Press, 1998, pp. 15-20.

15. R. S. Pindyck, and D. L. Rubinfeld, "Econometric Models and Economic Forecasts", The McGraw-Hill Companies, Inc, 1998, pp. 251-260.
16. Yao Xia, Peng Hangen, and Zhu Yan, "ARIMA Time Series Modeling and Applying on Fresh Agricultural Products", System Sciences and Comprehensive Studies in Agriculture, vol. 23, Feb. 2007, pp. 88-94.
17. Nie Rong, Qin Keming, and Zhang Xiaohong, "Stochastic Model and Risk Measuring on the Farm-produce's Price", Mathematics In Practice and Theory, vol. 34, Nov. 2004, pp. 108-112.
18. Zhang Xiaoshuan, "Research of Aquatic Products Price Forecasts Support System", Beijing: China Agricultural University, 2003, pp. 10-15.