

Crop Recommendation

Partha Koley, Somsubhra Gupta

Department of Computer Science & Engineering

Swami Vivekananda University

Partha.koley13@gmail.com

Abstract

Agriculture plays a leading role in the economic growth and development of the countries that make up the country. This project is full of Machine Learning algorithms in the field of agriculture to provide efficient and effective solutions to the problems faced by farmers. Some of the problems with systems that have been implemented are that they focus on a single parameter (moisture or temperature or soil pH etc.) to predict crop growth adequacy. In this project, however, many factors must be considered simultaneously to make the best and most accurate predictions for crops, such as pH, Nitrogen (N), Phosphorus (P), Potassium (K) soil content, temperature, humidity. The proposed project provides solutions for smart farming by monitoring soil and weather characteristics that can help farmers significantly increase their productivity. This project uses machine learning algorithms such as K- Nearest Neighbors, Random Forest and a dataset with approximately 2200 rows of data. The project provided farmers with crops to grow. Therefore, it will help farmers to choose the right crops for their fields.

keywords: Machine Learning, K-nearest neighbor, Random Forest

Introduction

As we all know, agriculture is the foundation of Indian economy. Agriculture is an important occupation in India. Over 60% of the country's land is used for agriculture, feeding 1.3 billion people. In India, agriculture gave birth to civilization. We need land to grow crops. Therefore, land is a

key factor in agriculture. Soil health is essential for good food production. It provides essential nutrients, water, oxygen and root support. Soil is the foundation of the food system and is home to all plants used for food production. In India, there are different varieties of soil. These are alluvial soils (cotton, rice), black soils (sugar cane, sunflower), red soils (maize, ragi), red soils (beans, tea, coffee), etc. Much research has been done to improve agricultural planning. Crop recommendations can be made using machine learning techniques

Generally, farmers choose to plant crops based on what nearby farmers are growing in their fields or crops that have a high market at that time without knowing their soil conditions and what crops are most suitable in their fields according to NPK planting percentage values, rainfall conditions, temperature, etc. Because of them did

Previous Work

Before proceeding with our project, we referred to the following research papers. In referring to each of

these documents, we came across several conclusions discussed below. Each entry below discusses the title of the article, the algorithms used, and the general conclusions drawn from that research article.

[1] Kumar, Y. Jeevan Nagendra, V.Spandana ETAL "A supervised machine learning approach for crop yield prediction in agriculture". In this proposed system, crop forecasting can be done using data such as temperature, humidity, pH, and rainfall. This harvest prediction can be done with random forest and decision tree algorithms. More accurate results can be found by applying random forests.

[2] Anushal Savla, Parul Dhawan ETAL presents "A Survey of -Class Algorithms for Developing Yield Forecasting Accuracy in Precision Agriculture". The data sets are collected over multiple years in soybean crop yield forecasting. The algorithms used for prediction in this article include Support vector machine, Random Forest, Neural networks, REP tree, Bagging and Bayes. The conclusion is that Bagging is the best predictive algorithm among the above algorithms.

[3] Monali Paul, Santosh ETAL in “Soil Behavior Analysis and Crop Yield Prediction Using Data Mining Methods”. in this document. Analyze the ground dataset and predict the class. From the predicted soil class, crop was identified. Naive Bayes and KNN algorithms are used for predictions. Work on the future has already begun.

[4] Reddy, D. Anantha ETAL presents “Crop recommender system to maximize yield of crops in ramtek region using machine learning”. This proposed system deals with parameters such as soil properties, soil type, and crop yield data collection that advises farmers of suitable crops to grow based on that parameter. This proposed system is applicable to different machine learning algorithms such as random forests, K nearest neighbors and naive bias. This proposed system can predict crops in weather conditions, state and district values.

[5] Kulkarni, Nidhi ETAL “Crop productivity improvement by crop recommender system using ensemble techniques”. This recommender system uses to recommend crops based on soil type, average rainfall, and temperature. This proposed system is applicable to various machine learning algorithms such as random forests, naive arrays, and linear SVM. The system will recommend crop types like Kharif and Rabi.

[6] Rajak, RohitKumar ETAL “Crop recommendation system to maximize yield of crops using machine learning techniques”. the proposed system. Using to identify cream based on soil database. This proposed system is applicable to different crops such as peanuts, beans, cotton, vegetables. This proposed system is used in various machine learning algorithms such as.

Research Methodologies

Crop recommendation systems leverage machine learning (ML) to optimize agricultural practices by suggesting the most suitable crops for specific locations, environmental conditions, and soil properties. The primary research methodologies employed in these systems involve data collection, preprocessing, model selection, and evaluation.

3.1 Data Collection & Processing

Data collection and processing are critical components of a crop recommendation system powered by machine learning (ML). The success of such systems hinges on the quality and relevance of the data used to train ML models. **Data Collection** involves gathering diverse datasets, including soil characteristics (pH, moisture, texture), weather conditions (temperature, rainfall, humidity), crop yield records, and local farming practices. Satellite imagery and IoT-based sensor data can also provide real-time insights into environmental factors.

Data Processing ensures the collected data is clean, consistent, and ready for ML algorithms. This stage includes data cleaning (removing errors, handling missing values), normalization, and feature extraction to highlight relevant parameters for crop selection. Advanced processing techniques like dimensionality reduction and data augmentation enhance the dataset's utility, enabling the development of robust, accurate crop recommendation models.

Effective data collection and processing lay the groundwork for reliable predictions, supporting farmers in making informed decisions for sustainable agriculture.

3.2 Data Analysis & Feature Selection

Data analysis and feature selection are critical steps in building an effective crop recommendation system using machine learning. The primary objective is to identify the key factors influencing crop yield and suitability for specific geographic regions, climate conditions, and soil types. This enables more accurate predictions, optimizing agricultural practices.

Data Collection: The system collects a variety of data, such as weather patterns, soil conditions, crop performance, and geographic characteristics. This data may come from sources like satellite imagery, climate databases, and agricultural sensors.

Data Preprocessing: Raw data is cleaned and normalized to handle missing values, outliers, and inconsistencies. Techniques like scaling, encoding, and imputation are used to prepare the dataset for modeling.

Feature Selection: Feature selection is a crucial process to reduce the dimensionality of the dataset and improve model performance. It involves identifying the most relevant features that contribute to crop growth prediction. Methods like correlation analysis, mutual information, and techniques like Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA) are commonly used to select optimal features.

Machine Learning Models: Once the features are selected, various machine learning algorithms like Random Forest, Support Vector Machines (SVM), and Neural Networks are employed to develop predictive models. These models are trained on historical data to predict the most suitable crops based on the selected features.

Evaluation & Optimization: The model is evaluated using performance metrics like accuracy, precision, recall, and F1-score. Hyperparameter tuning and cross-validation help optimize the model, ensuring robustness and generalizability.

By effectively analyzing data and selecting the most significant features, crop recommendation systems can provide farmers with personalized and data-driven crop recommendations, ultimately improving agricultural productivity and sustainability.

3.3 Model Architecture

The crop recommendation system aims to assist farmers in selecting the most suitable crops for their fields based on various environmental, soil, and climatic parameters. Machine learning (ML) plays a crucial role in predicting optimal crop choices by analyzing historical data and external factors that influence crop growth.

Data Collection and Preprocessing: The system begins by gathering data from multiple sources, including weather forecasts, soil properties, geographical data, and past crop yields. This data undergoes preprocessing steps such as normalization, missing value imputation, and data transformation to make it suitable for ML modeling.

Feature Selection: Key features influencing crop growth, such as temperature, rainfall, soil pH, and nutrient levels, are selected. Feature engineering techniques like principal component analysis (PCA) or correlation analysis may be applied to reduce dimensionality and improve model performance.

Model Development: A supervised learning model such as Decision Trees, Random Forest, Support Vector Machines (SVM), or Neural Networks is trained on historical crop data. These models are used to learn the relationships between the input features (soil type, weather conditions, etc.) and the target output (recommended crop types).

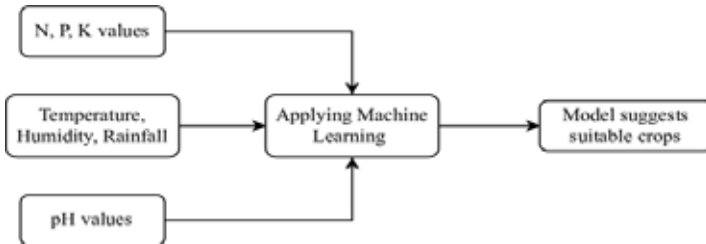
Model Evaluation: The model is evaluated using metrics such as accuracy, precision, recall, and F1-score. Cross-validation and hyperparameter tuning help improve generalization and prevent overfitting.

Recommendation Output: Once the model is trained, it can predict the best crops for a given set of inputs. The system provides tailored recommendations based on real-time data or user input, considering various environmental constraints.

3.4 Integration and Deployment:

The trained model is deployed in a user-friendly interface, such as a mobile app or web portal, where farmers can input their farm parameters and receive crop suggestions. The system can be further enhanced with continual learning mechanisms to adapt to changing climate patterns and crop yield outcomes.

By utilizing machine learning techniques, the crop recommendation system offers efficient and scalable solutions for agricultural planning, optimizing crop selection, and improving yields.



Various Algorithm used in Crop Recommendation

Various algorithms can be employed for crop recommendation systems, each offering unique strengths based on the complexity of the problem and the available data:

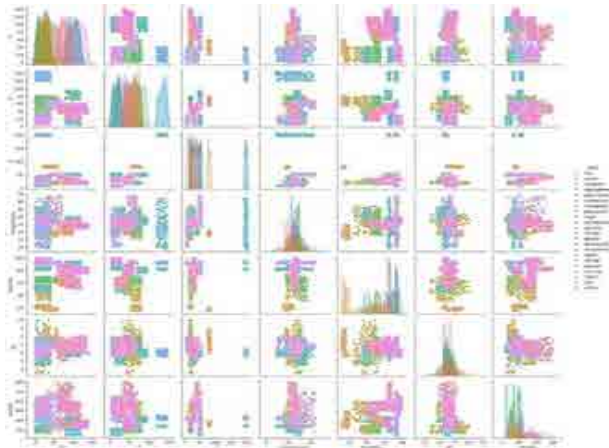
- 1. Random Forest:** An ensemble learning method that combines multiple decision trees to improve accuracy and reduce overfitting. It's effective for handling complex relationships and large datasets, making it suitable for crop recommendation based on environmental and soil factors.
- 2. Support Vector Machine (SVM):** A powerful classifier that works well for binary and multi-class problems. It can efficiently handle high-dimensional data and is useful in crop recommendation when distinguishing between multiple crop types based on features like soil and climate conditions.
- 3.K-Nearest Neighbors (KNN):** A simple, non-parametric algorithm that classifies crops by comparing the features of a new data point to the most similar ones in the dataset. It is easy to implement but can be computationally expensive for large datasets.
- 4. Artificial Neural Networks (ANN):** A deep learning model capable of learning complex patterns in large datasets. ANNs can be effective in crop recommendation systems, especially when incorporating diverse data sources like weather patterns, soil conditions, and historical crop yield data.
- 5. Decision Trees:** These models split data based on feature values to create a tree structure for decision-making. They are easy to interpret and can be used for crop recommendation by understanding how different factors impact crop suitability.

6. Gradient Boosting Machines (GBM): A boosting algorithm that sequentially builds trees, each correcting the errors of the previous one. GBM is known for its high predictive performance and is useful for crop recommendation systems when accuracy is a key priority.

7. Naive Bayes: A probabilistic classifier based on Bayes' theorem, useful for crop recommendation when the features are assumed to be conditionally independent. It can be efficient for large datasets with relatively simple relationships

8. Clustering Algorithms (e.g., K-means): Used for segmenting regions or soil types into clusters, making it easier to recommend the most suitable crop for each cluster based on shared characteristics.

Each algorithm offers unique advantages, and the choice of algorithm depends on the dataset's complexity, the nature of the problem, and the specific requirements of the crop recommendation system.



Confusion Matrix

A confusion matrix is a key tool for evaluating the performance of a classification model, including a crop recommendation system using Machine Learning (ML). It provides a comprehensive breakdown of the model's classification results, showing how well the model distinguishes between

different crop types under various conditions. Here's a general explanation of the confusion matrix in the context of a crop recommendation system: In a typical confusion matrix for a multi-class classification problem (where each class corresponds to a specific crop type), the rows represent the actual crop labels, and the columns represent the predicted crop labels.

Model Classification Report

The Model Classification Report of a Crop Recommendation System using Machine Learning (ML) provides an evaluation of the predictive model's performance, particularly focusing on the classification accuracy of crop types based on environmental and soil data. In a crop recommendation system, the model predicts the most suitable crops for a given set of conditions, such as soil quality, weather patterns, and geographical location. Key metrics typically included in the classification report are:

- 1. Accuracy:** Represents the overall percentage of correct predictions made by the model, indicating its effectiveness in recommending suitable crops.
- 2. Precision:** Measures the proportion of true positive predictions among all predicted positives. It shows how many of the recommended crops were relevant to the user.
- 3. Recall (Sensitivity):** Reflects the proportion of actual positives (correct crop suggestions) that were identified by the model. High recall indicates that the model successfully recommends most of the possible relevant crops.
- 4. F1-Score:** The harmonic mean of precision and recall. It provides a balanced view of the model's performance, especially when there is an imbalance in the class distribution.
- 5. Support:** The number of instances of each class in the dataset, helping to understand the distribution of crop classes across the model.

The classification report serves as a critical tool to assess and fine-tune the ML model, ensuring its optimal performance in real-world crop recommendations. By leveraging algorithms

such as Decision Trees, Random Forests, or Support Vector Machines, the system can be continually improved to enhance the prediction accuracy and provide more reliable recommendations to farmers and agriculturalists.

while Random Forest excels in generalization, the conclusion also emphasizes the importance of continuous refinement through additional techniques such as hyperparameter tuning, model evaluation using precision, recall, and F1-scores, and incorporating feedback loops from users.

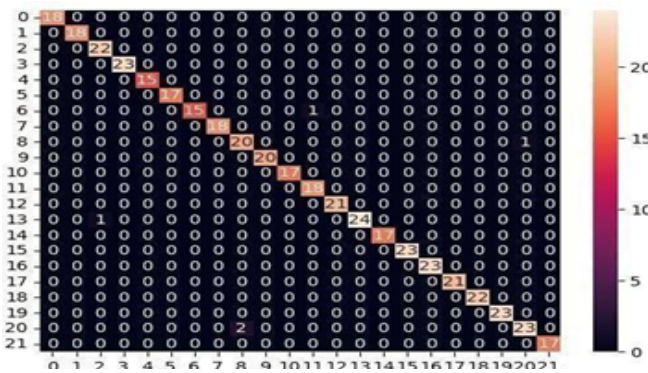


Figure : Heatmap of Confusion Matrix

Modal Efficiency

The efficiency of a crop recommendation system using the Random Forest model lies in its ability to handle complex, non-linear relationships between various agricultural factors (such as soil quality, weather, and crop characteristics) without overfitting. Random Forest, an ensemble learning method, combines multiple decision trees to make predictions, ensuring better generalization and robustness.

The algorithm's ability to rank features by importance is crucial for crop recommendation, as it allows agronomists to identify the most influential factors affecting crop suitability. This interpretability helps in making informed decisions and fine-tuning agricultural practices.

Its efficiency stems from its capacity to handle large datasets, manage feature interactions, and work well with both numerical and categorical data.

In crop recommendation, Random Forest can provide accurate predictions by considering multiple factors simultaneously, reducing the risk of bias in favor of specific crops. Additionally, its built-in feature importance calculation helps identify the most influential factors, improving the interpretability of the model. It is also computationally efficient, especially when combined with parallel processing, making it suitable for real-time applications in agricultural decision-making.

Accuracy in terms of numbers:

Decision tree = 0.90681

naïve bayes = 0.98863

SVM= 0.97727

Random forest = 0.99545

KNN= 0.96818

the crop recommendation system can become a more reliable and adaptable tool for farmers, ultimately leading to better agricultural outcomes.

Analysis of Conclusion

Random Forest for crop recommendation proves effective due to its ability to handle complex data and model non-linear relationships. Its ensemble approach ensures accurate predictions, while feature importance helps identify key factors influencing crop suitability. The model's computational efficiency allows for scalability and real-time application in agriculture. Continuous refinement through tuning and user feedback further enhances its performance, making it a reliable and adaptable solution for dynamic agricultural environments.

Areas for Improvement

Improving a crop recommendation system requires addressing several critical areas to enhance its performance and applicability. One of the primary concerns is data quality and preprocessing.

Inaccurate or incomplete data can negatively impact the model's ability to make correct predictions. Ensuring clean and well-structured data through techniques like data imputation and outlier removal is essential. Additionally, proper feature engineering, such as including relevant agricultural factors like soil quality and weather conditions, can significantly improve the system's predictive accuracy.

Model selection and complexity are also crucial. Overfitting and underfitting must be avoided by fine-tuning the model through techniques such as hyperparameter optimization and cross-validation. Experimenting with different algorithms, including ensemble methods, can help balance performance and robustness. Class imbalance is another challenge, which can be addressed by using techniques like oversampling or weighted loss functions to improve prediction accuracy across all crop types.

To ensure the system remains relevant in diverse agricultural settings, it is important to incorporate domain adaptation techniques. Models trained in one region may not generalize well to others, so strategies to adapt the system to new geographical areas are necessary. Additionally, integrating real-time data from sensors and weather stations can provide dynamic recommendations that reflect up-to-date conditions, enhancing the model's usefulness.

User feedback also plays a critical role in improving the system. By allowing farmers to rate and provide feedback on crop recommendations, the system can continuously evolve and become more personalized over time. Incorporating explainable AI methods will improve transparency and help users trust the recommendations provided. Finally, evaluating the model with metrics beyond accuracy, such as precision, recall, and F1-score, ensures a balanced understanding of its performance and capabilities.

Conclusion

In this article, we propose and effectively implement a crop recommendation system that can be easily used by farmers around the world. The system can help farmers decide which crops to plant based on parameters such as nitrogen,

phosphorus, potassium, pH, humidity, temperature, and rainfall. By using this research we can increase the productivity of the nation

and generate profits. In this way the farmer can plan for a good harvest, increasing the overall profit from his land and the country . proposed system. We will also stimulate the Indian economy and above all help farmers to increase their income. To realize this proposed system, we recommend farmers to plant their land on the best crops so that farmers can benefit from it.

Acknowledgment

We would like to express our heartfelt gratitude to sir Somsubhra Gupta our Swami Vivekananda University and all those previous contributed in this project. We extend our sincere thanks to our advisors and mentors for their invaluable guidance, insights, and support throughout this research. Their expertise and encouragement have been instrumental in shaping the direction and outcomes of this work.

We are also grateful to the agricultural experts and farmers who provided essential data and practical insights, helping us to understand the real-world challenges and requirements of crop recommendation. Their contributions have been crucial in ensuring the practical relevance and applicability of our system.

Additionally, we acknowledge the support from our institution and funding bodies, whose resources and infrastructure facilitated the development and implementation of this project. Finally, we would like to thank our peers and colleagues for their constructive feedback and collaboration, which have enriched our research process. This work would not have been possible without the collective efforts and support of everyone involved.

References

- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.

This seminal paper introduces the Random Forest algorithm, detailing its methodology and advantages in classification tasks.

a. Random Forests | Machine Learning

b. Gosai, H., & Raval, M. (2021). Crop Recommendation System using Machine Learning. International Journal

c. of Scientific Research in Computer Science, Engineering and Information Technology, 7(3), 129-134.

This research paper recommends a crop recommendation system based on machine learning methods that evaluate climatic variables like temperature, humidity, and rainfall.

[Link to the paper](#)

a. Biau, G., & Scornet, E. (2016). A Random Forest Guided Tour. Test, 25(2), 197-227.

This article provides a comprehensive review of the Random Forest algorithm, discussing its theoretical foundations and practical applications.

[Link to the paper](#)

a. Scornet, E., Biau, G., & Vert, J.-P. (2015). Consistency of Random Forests. The Annals of Statistics, 43(4), 1716-1741.

This paper analyzes the consistency of Random Forests, providing insights into their performance in various scenarios.

[a. Link to the paper](#)

b. Livingston, F. (2005). Implementation of Breiman's Random Forest Machine Learning Algorithm.

This document provides tools for exploring Breiman's Random Forest algorithm, focusing on its development, verification, and significance.

[Link to the paper](#)

A Machine Learning Approach to Crop Recommendations

This article presents a machine learning approach tailored for crop recommendations, focusing on the integration of various data inputs to improve agricultural productivity.

[Link to the paper](#)