A Data-Driven Approach to Precision Agriculture: Crop Recommendation Using ML

|  |  |
| --- | --- |
| Nikhil Sharma | Pawan Negi |
| *School of Computer Science and Engineering* | *School of Computer Science and Engineering* |
| *Galgotias University* | *Galgotias University* |

nikhilvasisth0309@gmail.com negipawan787@gmail.com

*Abstract-* ***The global frugality and food force heavily depend on husbandry, yet growers face challenges in choosing the stylish crops to grow due to changeable rainfall, soil conditions, and limited coffers. This study introduces a machine literacy- grounded crop recommendation system to help growers in making informed opinions about crop selection. The system analyses factors similar as soil composition, pH situations, rush, temperature, and humidity to suggest suitable crops for specific areas. We estimated several Machine learning techniques, such as Decision Trees, Random Forests, and Support Vector Machines (SVM), to identify the most effective model. Keywords – husbandry, Recommendation system, Ensemble Model, SVM.***

1. Introduction

The agricultural sector is crucial for economic sustainability and ensuring food security still growers face considerable obstacles when deciding which crops to grow. Factors including climate variability, erratic downfall, dwindling soil fertility, and shifting environmental conditions make this decision- making process complex. Choosing infelicitous crops can affect in reduced yields, financial lapses, and hamstrung use of coffers similar as water and diseases. also, as climate change introduces fresh misgivings, counting solely on experience and original knowledge is getting less effective, numerous growers, particularly those in remote and resource- limited areas, warrant access to ultramodern tools or expert guidance.

The consequences of poor crop selection extend beyond individual growers, affecting indigenous food security and agrarian sustainability. also, growing populations bear increased agrarian output, creating pressure to enhance yields without causing environmental detriment. Accordingly, there's an critical demand for data- driven results to empower growers and enhance their crop selection styles.

This study introduces a machine learning based crop recommendation system that utilizes data on soil parcels, climate conditions, and environmental factors to suggest applicable crops. By assaying information similar as soil type, pH position, temperature, moisture, and rush, the system provides accurate recommendations, enabling growers to make well- informed opinions and alleviate pitfalls.

1. Background

Machine Learning is a part of computer wisdom, wherein, computers can learn from former gests, and give an affair grounded on the former gests. In other words, we can say that the computers have the capability to learn

It can be classified in 3 types-

1. Supervised Machine Literacy
2. Unsupervised Machine Literacy
3. Semi Supervised Machine Literacy.
	1. Supervised Machine Literacy

Supervised machine literacy is similar to a school teacher tutoring a pupil, and a pupil literacy. In supervised learning, the algorithm learns from a training dataset. In supervised learning, there are input variables (A) and an outcome variable (B). The ideal of the algorithm is to learn the function which maps the input variable to the affair variable (1). Exemplifications of Supervised machine literacy are Decision Trees, Random timbers, Naive Baye

* 1. Unsupervised Machine Literacy

Unsupervised Machine Literacy is a subset of machine learning. The primary ideal is to uncover concealed Patterns, relationships, or structures within the input data are identified without explicit supervision. In discrepancy to supervised literacy, which involves training algorithms using labelled input, unsupervised literacy concentrates on revealing the essential data distribution or organizing it into groups. This approach allows the algorithm to identify underpinning structures in the data autonomously, without counting on pre- assigned markers or issues.

* 1. Semi Supervised Machine Literacy

Semi-Supervised Machine Literacy is deposited between supervised and unsupervised literacy. Semi-supervised machine literacy employs a mixture of labelled and unlabelled data is used for training the model. This approach generally involves a limited collection of labelled exemplifications alongside a larger collection of unlabelled information. The primary ideal is to harness the limited labelled data to guide the model in rooting meaningful patterns from the abundant unlabelled data, thereby enhancing its capability to generalize to new cases.

1. Problem Statement

Agriculture is a cornerstone of human sustenance, playing a critical role in global food security and economic stability. However, one of the most persistent challenges in agriculture is the lack of accurate and scientific guidance for crop selection. Farmers often rely on traditional practices or intuition to decide which crops to grow, which may not align with the specific soil and climatic conditions of their region. Current Challenges are**:**

1. **Inadequate Knowledge:** Many farmers lack access to technical expertise or agronomical data, leaving them unable to make informed decisions about crop selection.
2. **Variability in Conditions:** Soil fertility, represented by nutrient levels like Potassium (K), Phosphorus (P), and Nitrogen (N), varies significantly across regions, and these parameters must be optimized for different crops.
3. **Climatic Dependencies:**

Environmental factors such as Rainfall, Humidity, and Temperature play a crucial role in crop growth. Inconsistent or unsuitable climatic conditions further exacerbate the risks of crop failure.

1. **Unsustainable Practices:** Improper crop selection can lead to soil depletion and unsustainable farming practices, affecting long-term agricultural productivity.
2. Literature Survey

The research paper [2] examines the essential components and strategic planning required for creating a software model that supports precision farming. It provides a

comprehensive analysis of the fundamental principles of this agricultural approach. The paper begins by introducing fundamental principles of precision farming before advancing to the development of a supportive framework. It presents a model that applies Precision Agriculture (PA) concepts to small, open farms at the individual farmer and crop level, with the aim of addressing variability to a certain degree. The primary objective of this model is to deliver direct guidance to farmers, irrespective of their land size, even for the smallest plots.

The research paper [3] provides a thoughtful analysis of various sorting techniques and their efficiency in predicting crop yields for precision agriculture. These methods are applied to a dataset spanning multiple years, focusing on predicting soybean crop yields. The research evaluates several machine learning algorithms, including Support Vector Machine, Random Forest, Neural Network, REP Tree, Bagging, and Bayes, for yield prediction.

1. METHODOLOGY

Our System employs a process into different steps such as Data Collection, Data Preprocessing and evaluation , as illustrated in Figure 1.



Figure 1: Block Diagram of the Methodology of the System

5.1. Dataset Collection

The effectiveness of a crop recommendation system utilizing machine learning is largely dependent on the quality and relevance of available data. This dataset contains soil metrics such as N, P, K, humidity, rainfall, Ph and temperature for several types of soil as well as the crops that can be cultivated in those soil as labels. The Figure 2 shows the sample dataset. Dataset consists of 2200 rows and 8 columns. The last column is label that defines crop.



Fig 2. Dataset

* 1. Data Preprocessing

The initial dataset frequently contains missing data, inconsistencies, and noise, which can impact the model's effectiveness. To improve accuracy, data preprocessing methods such as imputing missing values with the mean or mode, normalizing numerical data, and encoding categorical variables were implemented. Feature scaling techniques like Min-Max Scaling were employed to maintain consistency across various features.

* 1. Spilt the Dataset

During this process, the pre-processed data is divided into a training set and a test set, following an 80:20 ratio. The training set comprises 1760 rows x 8 columns, while the test set includes 440 rows x 7 columns.

* 1. Algorithms Used in the System
		1. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a crucial supervised machine learning algorithm applicable to both regression and classification problems. SVM aims to optimize the margin between the nearest data points (support vectors) from distinct classes and the hyperplane, ensuring the identification of the optimal separating hyperplane.

This approach is particularly useful when dealing with non-linearly separable data. SVM performs effectively even with limited datasets and is particularly advantageous in high dimensional contexts.

5.4.2. Random Forest

Random Forest is a machine learning technique that combines multiple decision trees to enhance predictive accuracy. This ensemble method creates numerous trees using random subsets of the training data. Each tree independently generates a prediction, and the final result is determined by combining all tree predictions. The versatility of Random Forest allows it to process both categorical and numerical inputs, making it applicable to various applications. Furthermore, it offers an evaluation of feature importance, helping identify which variables have the most significant impact on the model's predictions. Random Forest is highly scalable and performs well with extensive datasets.

5.4.3. Decision Tree classifier

The Decision Tree Classifier is a widely-used machine learning technique employed for classification and regression purposes. Its operation involves iteratively dividing the dataset into subgroups based on the most effective feature for data separation. This

division continues until each subgroup becomes homogeneous or meets a predetermined stopping condition, such as reaching a maximum depth or minimum sample size per leaf. Within the tree structure, internal nodes represent features, while branches signify decisions based on these features. The tree's leaves indicate the final predicted class or value. One of the main advantages of decision trees is their ease of interpretation and visualization, making them highly understandable models. They don't require feature scaling since they rely on data splitting rather than distance calculations. However, decision trees are susceptible to overfitting, particularly when dealing with deep trees or noisy data. To mitigate this issue, techniques such as pruning (reducing branch complexity) or ensemble methods like Random Forests are employed.

1. Existing System

The existing systems for crop recommendation, ranging from traditional farming practices to modern technological solutions, have notable limitations in addressing the complex and dynamic needs of farmers. Agricultural extension services and soil testing laboratories provide more scientific recommendations but are often inaccessible to farmers in remote areas due to limited infrastructure, high costs, and time-consuming processes. Advanced systems such as IoT-based platforms and machine learning models have shown potential in research and pilot stages, but their high costs, technical complexity, and limited accessibility prevent widespread adoption. The shortcomings in current systems underscore the urgent demand for a robust, expandable, and user-friendly crop recommendation platform powered by machine learning. Such a system would offer farmers personalized, evidence-based guidance to boost yields and encourage sustainable farming methods.

1. Result

The project is expected to produce a reliable and efficient crop recommendation system that precisely determines the most appropriate plant for cultivation using various input factors including K, P, N, precipitation, moisture levels, and atmospheric conditions. It will feature a user friendly interface, such as a web or mobile application, for easy access and real-time recommendations. Additionally, the system's use can encourage sustainable farming practices by recommending crops that align with specific soil and climatic conditions.

The crop recommendation system was evaluated using multiple machine learning algorithms. Among these, the Random Forest Classifier achieved the highest accuracy of 99.31%, showing strong predictive performance and reliability in handling diverse input data. The Decision Tree Classifier also performed well with an accuracy of 98.86% and provided easy-to-understand results, which can be useful in real-world applications where model transparency is important. The Support Vector Classifier (SVC) reached an accuracy of 96.81%, effectively managing complex patterns in the data. Overall, while all three models gave accurate recommendations, Random Forest proved to be the most effective, combining high accuracy with consistent results across different scenarios. The work can be extended further to add functionalities like fertilizer recommendation etc.

1. Conclusion

This Paper introduces a machine learning based system for crop recommendations which aims to aid farmers in choosing the most appropriate crops according to the environmental and soil conditions. The system utilizes critical factors such as pH level, temperature, precipitation, and humidity to offer data-supported suggestions, thus minimizing the uncertainties linked with conventional crop selection methods. By utilizing algorithms like Decision Trees, Random Forest, and Support Vector Machine (SVM), the system ensures precise forecasts, enabling farmers to enhance their yields and make well informed choices.

To sum up, the Crop recommendation system presents a pragmatic approach for modern agriculture by merging technology with age old farming techniques. has the capacity to significantly improve agricultural outcomes, strengthen food security, and encourage sustainable farming practices in the face of environmental challenges.



1. References
2. Tanmay Banavlikar, Aqsa Mahir, Mayuresh Budukh, Soham

Dhodapkar(Dec-2017), ‘CROP

RECOMMENDATION SYSTEM USING

NEURAL NETWORKS’, International Research Journal of Engineering and Technology (IRJET).

1. Satish Babu (2013), “A Software Model for Precision Agriculture for Small and Marginal Farmers”, at the International Centre for Free and Open Source Software (ICFOSS) Trivandrum, India.
2. Anshal Savla, Parul Dhawan, Himtanaya Bhadada, Nivedita Israni,

Alisha Mandholia , Sanya Bhardwaj(2015), ‘Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture', Innovations in Information, Embedded and Communication systems (ICIIECS).

1. Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh (2015), ‘Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique’, International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM).
2. Liying Yang (2011), ‘Classifiers selection for ensemble learning based on accuracy and diversity’ Published by Elsevier Ltd. Selection and/or peer review under responsibility of [CEIS].
3. Rohit Kumar Rajak, Ankit Pawar, Mitalee Pendke, Pooja Shinde, Suresh Rathod, Avinash Devare(2017), ‘Crop Recommendation System to Maximize

Crop Yield using Machine Learning Technique’ , International Research Journal of Engineering and Technology (IRJET).

1. Modi, D., Sutagundar, A. V., Yalavigi, V., & Aravatagimath, A. (2021, October). Crop recommendation using machine learning algorithm. In 2021 5th International Conference on Information Systems and Computer Networks (ISCON) (pp. 1- 5). IEEE.
2. Priyadharshini, A., Chakraborty, S., Kumar, A., & Pooniwala, O. R. (2021, April). Intelligent crop recommendation system using machine learning. In 2021 5th international conference on computing methodologies and communication (ICCMC) (pp. 843- 848). IEEE.
3. Gosai, D., Raval, C., Nayak, R., Jayswal, H., & Patel, A. (2021). Crop recommendation system using machine learning. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 7(3), 558-

569.