IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.13476

Crop Recommendation System Using Machine Learning

Prof. A. M. Ghime[1], Akshay Kshirsagar[2], Shahuraj Lohakare[3], Fardeen Quadri[4], Vikas Kare[5]

Assistant Professor, Department of Computer Engineering, TSSM BSCOER Narhe Technical Campus, Pune, India[1]

Student, Department of Computer Engineering, TSSM BSCOER Narhe Technical Campus, Pune, India[2]

Abstract: Agriculture is the foundation of many large economies, like India and Maharashtra. The newcomers in the agricultural sector face the challenge of deciding which crops to cultivate on their farms. This problem needs addressing, and we are taking steps to solve it. To address this, we have developed a system that predicts suitable crops for farmers based on the natural content present in the soil and another parameters like weather, rainfall, humidity and many more. This solution aims to reduce farmers' losses and enhance production. Unlike existing systems, which are not fully functional and cannot effectively guide farmers in selecting crops, ours utilizes classification and regression algorithms for crop prediction. Our system can be used by the farmer's on the web as well as the android phone's as well. Proposed system uses a dataset which contains the samples of the crops with the required nutrition's such as Potassium. Phosphorus, Nitrogen, pH, Humidity, Rainfall and many more features. We are using the K-Nearest Neighbor (KNN) machine learning algorithm which is Supervised Learning algorithm used for the classification and regression. System uses a pickle library of python to create the Machine Learning Model which we actually recommend the crop to be cultivate. Model then takes the input as the ingredient in the soil as a parameter and then KNN finds the best suitable crop for that particular type of soil.

Keywords: Machine Learning, Crop Prediction, K-Nearest Neighbour, Crop Recommendation, Classification, Regression, Machine Learning Model, Supervised Algorithm.

I. INTRODUCTION

Farming is a crucial field for Indian economic situation along with additionally human future. In India, agriculture is an important source of income for many peoples. The use of Information Technology may alter the state of decision making, allowing farmer's to get and produce the best results[1][4]. Data mining processes are basically connected to the agriculture are employed in the decision making process. The process of extracting the most important data form the large amount of data is called the data mining[3]. As agriculture involves the many type of the data, such as soil data, crop data, and weather data, fertilizers data, we now employ a machine learning algorithm approach to for the crop prediction. Machine Learning algorithm can be efficiently used for the crop prediction and recommendation system.

Data Mining techniques and algorithms are used for the purpose of crop recommendations. These techniques can be used for the crop recommendation. And recommendation will be based on the Nitrogen (N), Phosphorous (P), Potassium (K) values [2]. This paper focuses on the building a Machine Learning model which can help to predict the crops based on the soil ingredients. The resigned system will be recommend the best suitable crop for the particular land. Based on the weather parameters and the soil content such as Rainfall, Temperate, Humidity, and pH[6]. They are collected from the Maharashtra government agriculture department website and some datasets are collected form the Kaggle's Data set which is basically famous for the real world datasets.

II. LITERATURE SURVEY

Nowadays, many experts are using automated farming. They're using well-known algorithms like Decision Trees and multiple linear regression to predict soybean productivity based on climate conditions. To make it easier for farmers to understand, they've created different rules from the Decision Tree[5]. They used a paper by Md. Tahmid Shakoor and his colleagues to choose various attributes like soil depth, slope, and drainage. They implemented two machine learning algorithms to classify the data. Their system automatically collects weather and soil data for a given area[3]. It works over large regions and provides forecasts at a resolution compatible with the best input data resolution, usually from soil data. This helps in forecasting crop performance before the crop season starts[4]. This way, farmers can make changes in their strategies, like choosing more resilient crops or changing the crop type, to adapt to extreme weather conditions later in the crop cycle.

This system aims to bring agriculture and technology, especially Android technology, closer together. It helps users buy products online through an Online Shopping Platform, where they can select items from the website[5][6]. Users can easily find different products in various categories, making purchasing easier and helping vendors reach more customers.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024
DOI: 10.17148/IJARCCE.2024.13476

Recommended fertilizers can also be purchased directly from the website. The site offers a selection of fertilizers that are recommended for use. Additionally, fertilizers that are often bought together will be suggested to users. The apriori algorithm, which finds frequent item sets, is used for this purpose. Users will receive notifications about their progress via email or SMS. They will also be notified about the shopping process, order details, and billing information.

III. RELATED WORK

Machine learning (ML) algorithms have been increasingly applied to crop and fertilizer recommendation systems, aiming to optimize agricultural practices and improve crop yields. Crop and fertilizer recommendation systems utilizing machine learning algorithms have collected significant attention due to their potential to optimize agricultural practices and enhance crop yields. In recent years, researchers have explored various approaches to address this challenge, leveraging data-driven methodologies and advancements in machine learning techniques. This section presents an overview of existing literature in this domain.

1. Traditional Methods

Traditional methods for crop and fertilizer recommendation often rely on expert knowledge, soil testing, and manual observation. While these approaches have been used effectively in agriculture for decades, they are limited in scalability and may not fully leverage the vast amount of data available in modern agricultural settings.

2. Data-Driven Approaches

a) Supervised Learning Techniques

Researchers have explored various supervised learning algorithms such as decision trees, support vector machines (SVM), and random forests to predict crop yields and recommend appropriate fertilizers. These approaches typically utilize historical crop data, soil attributes, and climate information to make recommendations (Reference).

b) Deep Learning Models

Deep learning models, particularly Convolutional Neural Networks (CNNs) and recurrent neural networks (RNNs), have shown promise in analyzing large-scale agricultural data. These models can integrate satellite imagery, weather patterns, and soil composition to generate accurate recommendations for crop types and optimal fertilizer usage (Reference).

3. IoT and Sensor Data Integration

Integration of Internet of Things (IoT) devices and sensor data has enabled real-time monitoring of environmental conditions in fields. By combining sensor data with ML algorithms, researchers have developed systems capable of dynamically adjusting fertilizer recommendations based on current soil moisture levels, weather forecasts, and other relevant factors (Reference).

4. Hybrid Approaches

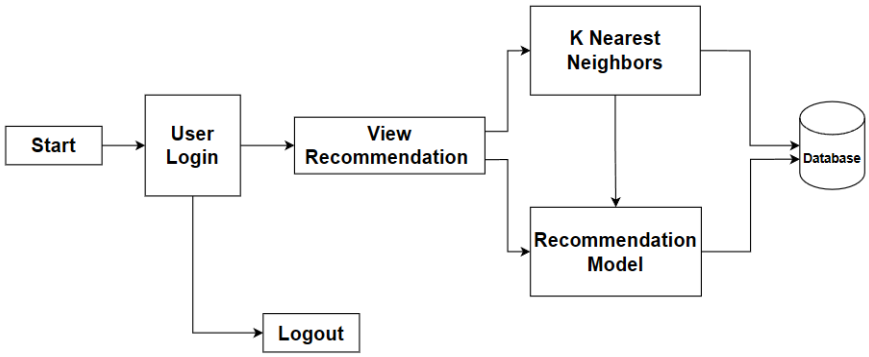
Hybrid approaches, which combine domain knowledge with data-driven ML techniques, have emerged as effective solutions for crop and fertilizer recommendation. These approaches leverage both expert insights and algorithmic predictions to provide tailored recommendations for specific soil types, crop varieties, and environmental conditions (Reference).

Challenges and Future Directions -

Despite the progress made, several challenges remain in the development of crop and fertilizer recommendation systems. These include the need for large, high-quality datasets, the incorporation of multi-modal data sources, and the integration of domain knowledge into ML models. Future research directions may involve the exploration of advanced techniques such as reinforcement learning, ensemble methods, and transfer learning to further enhance the accuracy and scalability of recommendation systems in agriculture (Reference).

IV. DIAGRAMS

A. System Architecture:



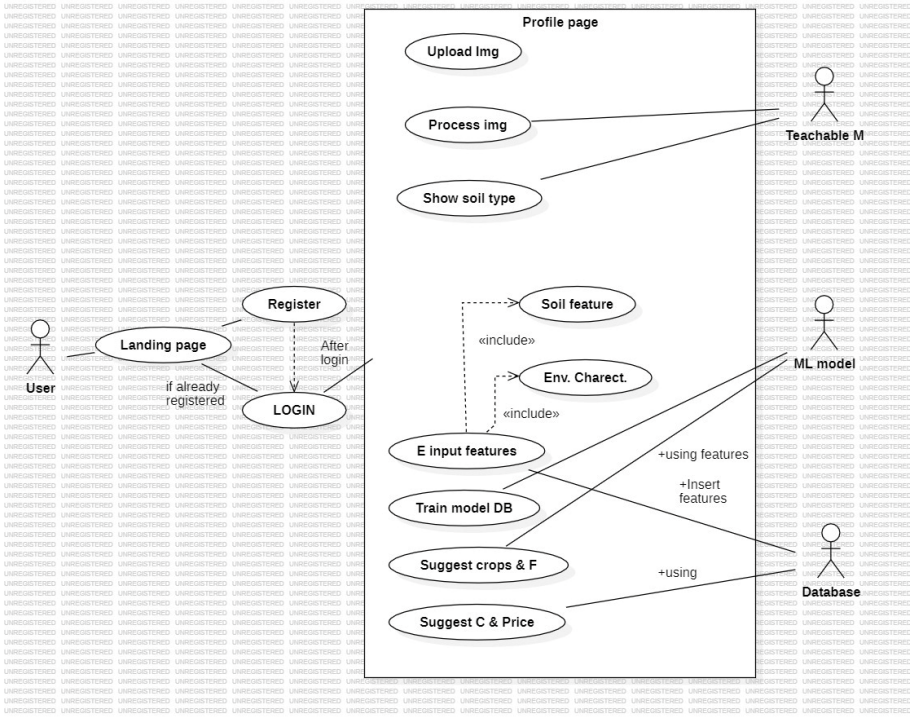
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.13476

B. Use Case:



V. ALGORITHMS

1. K-Nearest Neighbour's Algorithm:

K-Nearest Neighbors Algorithm is one of the simplest Machine Learning Algorithm which is based on the Supervised Learning Technique. KNN is one of the most basic and essential classification algorithm in machine learning. It belongs to the supervised learning domain and finds the intense application in the pattern recognition, data mining and many intrusion detection and many more. This algorithm is non-parametic, means it does not make any underlying assumptions. It can handle the both numerical as well as the categorical data, which makes it as the best choice for the classification as well as regression tasks. KNN works by finding the K nearest neighbors to the given data point or the data set based on the distance matric, such as Euclidean distance. The class or the value of the data point is then determined by the majority vote or the average of the K neighbors. This approach allows the algorithm to adapt the different patterns and make the prediction based on the local data structure of the provided data.

2. Support Vector Machine:

The Support Vector Machine (SVM) is a machine learning tool used for both classification and regression, but it's mainly used for classification. Its main job is to find the best line, plane, or hyperplane that can separate different groups of data points in space. This line or plane tries to have the largest possible gap between the closest points of different groups. The dimension of this line or plane depends on how many features we have. If we have two features, it's just a line. With three features, it's a plane. Visualizing it gets tricky when we have more than three features.

3. Classification:

A classification algorithm is a type of machine learning tool that sorts data into different categories based on patterns it learns from existing examples. For instance, it might determine if an email is spam or not spam, whether an image contains a cat or a dog, or if a patient has a certain disease or not. Unlike regression, which predicts a numerical value, classification predicts a category, like "yes" or "no," "red" or "green." This type of algorithm needs labeled data, meaning it learns from examples where the correct category is already known. Essentially, it maps input data to specific output categories.

4. Regression:

Regression helps predict values based on known data. It looks at how one thing depends on others, like how house prices depend on factors like size and location. The goal is to find the best model that shows this relationship, so we can use it to make predictions. Random forest regression combines many decision trees to make predictions. Each tree is trained on a different part of the data, and then their predictions are averaged to get the final result. This helps improve accuracy.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024
DOI: 10.17148/IJARCCE.2024.13476

Linear regression is simple and widely used. It assumes a straight-line relationship between variables. Polynomial regression handles more complex, nonlinear relationships by adding terms to the linear model. Support Vector Regression (SVR) uses a hyperplane to minimize prediction errors. Decision tree regression creates a tree structure to predict values. Random forest regression combines many trees for better accuracy by averaging their predictions.

VI. METHODOLOGY

We're creating a Machine Learning system focusing on Maharashtra, India, considering how climate varies across districts. We gathered historical data on crops and climate from government websites like data.gov.in and imd.gov.in. For each district, we collected monthly data on factors like precipitation, temperature, and cloud cover, which affect crops. First, we collected data from various sources and prepared datasets for analysis. We split the dataset into two parts: 75% for training the model and 25% for testing it. To predict future events, we're using Supervised and Unsupervised learning algorithms. Supervised learning uses labelled examples from the past to predict new data. Unsupervised learning looks at data without labels to uncover hidden patterns.

We're particularly using a Random Forest Classifier, a powerful algorithm that builds many decision trees during training. It can classify or predict outcomes by combining the results of these trees. The more trees in the 'forest,' the better the prediction. "Random Forest Classifier: Random forest is the most popular and powerful supervised machine learning algorithm capable of performing both classification and regression tasks, that operate by constructing a multitude of decision trees at the time of training and generating outputs of the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. The more trees in a forest the more robust the prediction.
Creating a user-friendly application for crop recommendation still faces challenges, but this solution aims to address them. It focuses on factors like rainfall, temperature, and soil type, which directly affect farming. The goal is to suggest a wider range of crops that can be grown throughout the season. This system aims to help farmers choose crops more easily and increase their yield to reduce suicide rates. It predicts crop yields based on data from the region. By combining agriculture with machine learning, it can improve farming practices and optimize resources.

Data from past years is crucial for forecasting current performance. We gather historical data from reliable sources like data.gov.in, kaggle.com, and indianwaterportal.com, focusing on Maharashtra and Karnataka regions. This data includes details like the state, district, year, season, crop type, cultivation area, and production. We also include information about soil type, temperature, and rainfall specific to each region. After cleaning and pre-processing the data, we replace null values with averages and convert categorical attributes into labels. One hot encoding is used to handle categorical values in the datasets.

CONCLUSION

This paper highlighted the limitations of current systems and their practical usage on crop prediction. Then walks through a viable crop prediction system to the farmers, a proposed system provides connectivity to farmers via a mobile application. The mobile application includes multiple features that users can leverage for the selection of a crop. The built-in predictor helps farmers forecast how much crop they'll get. The inbuilt recommender system allows a user exploration of the possible crops and their yield to take more educated decisions. For yield to accuracy, various machine learning algorithms such as Random Forest, ANN, SVM, MLR, and KNN were implemented and tested on the given datasets from the Maharashtra and Karnataka states. The various algorithms are compared with their accuracy. The results obtained indicate that Random Forest Regression is the best among the set of standard algorithms used on the given datasets with an accuracy of 95%. The proposed model also explored the timing of applying fertilizers and recommends appropriate duration

ACKNOWLEDGMENT

"Thankfulness is not just the best of virtues yet the moms and dad of all others." Consequently, it is fantastic enjoyment for me to recognize the help along with payment of variety of people that assisted me in establishing "Plant Referral System" job. First of all, I want to tape-record my thankfulness and also many thanks to **Prof. A. M. Ghime** for her passionate assistance plus aid in effective conclusion of workshop. I reveal my many thanks to **Dr. G. A. Hinge** (Principal), **Dr. P. A. Kadam** (Head of Computer Department) for their important assistance. I am additionally thankful to various other educators as well as non-teaching team of Computer Engineering Department as well as Library for their co-operation and also aiding us as well as leading us to finish our job properly."

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.13476

REFERENCES

[1]. Sundari V, Anusree M, Swetha U and Divya Lakshmi R "Crop recommendation and yield prediction using machine learning algorithms".
[2]. Mansi Shinde, Kimaya Ekbote, Sonali Ghorpade, Sanket Pawar, Shubhada Mone "Crop Recommendation and Fertilizer Purchase System".
[3]. Dhanush Vishwakarma, Mahendra N, Manjuraju M.R, Ashwini "Crop Prediction Using Machine Learning Approaches".
[4]. Pavan Patil, Virendra Panpatil, Prof. Shrikant Kokate "Crop Prediction System Using Machine Learning Algorithms".
[5]. Shilpa Mangesh Pandey, Dr. Prem Kumar Ramesh, Anmol, B.R. Asishwarya, Karuna Rohilla, Kumar Shaurya "Crop Recommender Systerm Using Machine Learning Approach".
[6]. D. Jaynarayana Reddy, Dr. Rudra Kumar "Crop Yeild Prediction using Machine Learning Algorithms".