2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal

MACHINE LEARNING-BASED CROP RECOMMENDATION SYSTEM FOR ENHANCED YIELD PRODUCTION

M. Madhavi Latha[1], Bhumagouni Pranay[2], Gattu Rishitha[2], Shaik Shouqib Ali[2], Md. Amer[2], Ranganamaina Naresh[2]

[1]Assistant Professor,[2]UG Student,[1,2]Department of Computer Science and Engineering (AI & ML)
[1,2]Malla Reddy Engineering College and Management Science, Kistapur, Medchal-50140l, Hyderabad, Telangana, India

ABSTRACT

Agriculture plays a critical role in ensuring that there is enough food to feed our ever-growing global population. With the challenges posed by climate change and its uncertainties, it has become increasingly important to adopt sustainable practices in agriculture. One crucial aspect of this is optimizing how we use our land, water resources, and fertilizers while also increasing crop yields. Traditionally, farmers relied on their own experience, local knowledge, and trial-and-error methods to decide which crops to grow in their specific regions and soil types. Unfortunately, this subjective approach often led to uncertainties, reduced productivity, and economic losses. Inappropriate crop choices were also common, resulting in suboptimal yields and wasted resources. To address these issues, we are working on developing a Machine Learning-based Crop Recommendation System. This system holds immense promise in transforming agriculture by offering farmers data-driven insights to make informed decisions about their crop selection and management practices. The system uses a wealth of data, including historical crop performance, soil characteristics, and climate patterns, to empower farmers with intelligent recommendations. By analysing this vast array of information, the system guides farmers in making data-driven choices, reducing the risk of crop failure, and increasing profitability. Furthermore, it promotes the adoption of sustainable farming methods, thereby minimizing environmental impacts and preserving precious natural resources for future generations. Implementing this recommendation system can have far-reaching benefits. It can lead to economic advantages for farmers, ensuring food security, and encouraging the adoption of sustainable agricultural practices to protect our environment. By providing valuable insights, the system empowers farmers to make informed decisions that ultimately lead to improved crop yields, reduced resource wastage, and increased profitability. With the potential to revolutionize crop selection, this novel approach paves the way toward a more sustainable and food-secure future for all of us.

Keywords: Smart agriculture, Recommendation systems, Crop management, Machine learning.

1. INTRODUCTION

Agriculture is the cornerstone of global food security and sustenance, making it an industry of paramount importance. However, the efficiency and productivity of agricultural practices are often constrained by factors such as soil quality, climate conditions, and crop choice. In this context, the development of a "Machine Learning-Based Crop Recommendation System for Enhanced Yield Production" emerges as a pivotal endeavour that holds the potential to revolutionize the way crops are cultivated and harvested. At its core, this research aims to bridge the gap between traditional agricultural practices and cutting-edge technology [1]. It leverages the power of machine learning, a subset of artificial intelligence, to optimize crop selection decisions made by farmers. This innovation is of immense significance given the growing challenges faced by the agriculture sector, including the need to feed an expanding global population, mitigate climate change impacts, and reduce resource wastage.

The primary motivation behind this research lies in addressing several pressing issues. Firstly, the unpredictable effects of climate change have significantly altered weather patterns and precipitation levels, making it increasingly difficult for farmers to make informed crop choices based on historical data alone [2]. Additionally, soil quality can vary dramatically even within a single farm, necessitating a more nuanced approach to crop selection. Traditional methods, which often rely on anecdotal

2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal

knowledge and experience, are no longer sufficient to ensure optimal yield and resource efficiency. Machine learning offers a novel and data-driven solution to this dilemma. By analyzing a myriad of factors, including historical weather data, soil composition, and past crop performance, machine learning algorithms can provide personalized and accurate crop recommendations. This not only optimizes yield but also promotes sustainable farming practices by reducing the unnecessary use of water, pesticides, and fertilizers [3].

Furthermore, the research recognizes the global imperative of sustainable agriculture. In a world grappling with issues of food security and environmental conservation, the ability to recommend crops that are not only high-yielding but also environmentally responsible is crucial. Machine learning models can factor in environmental impact considerations, helping farmers make choices that align with conservation and resource preservation goals [4].

In this introductory overview, we will delve into the key components and objectives of this research. Firstly, we will explore the challenges faced by farmers and the limitations of conventional crop selection methods. Next, we will introduce the concept of machine learning-based crop recommendation systems, highlighting their potential to revolutionize agriculture. We will also discuss the role of data sources, including historical data, satellite imagery, and soil assessments, in training these models. Additionally, ethical considerations, such as data privacy and equitable access to technology, will be addressed [5]. The research on a "Machine Learning-Based Crop Recommendation System for Enhanced Yield Production" seeks to empower farmers with data-driven decision-making tools that can optimize crop selection, enhance yield, and promote sustainable agricultural practices. By harnessing the capabilities of machine learning, this research endeavours to usher in a new era of precision agriculture that is not only more productive but also environmentally responsible and equitable [6].

The research motivation for a "Machine Learning-Based Crop Recommendation System for Enhanced Yield Production" is deeply rooted in the critical challenges facing agriculture today. As the global population continues to grow, the pressure on the agricultural sector to produce more food intensifies. This demand necessitates not only increased agricultural productivity but also a more efficient approach to crop selection [7]. Traditional farming methods often rely on historical practices and lack the adaptability required to optimize crop choices in the face of evolving climate conditions and soil quality. This research is motivated by the urgent need to bridge this gap between traditional farming practices and cutting-edge technology, as it holds the potential to significantly enhance global food security.

Climate change further compounds the challenges in agriculture, introducing unpredictability into weather patterns and exacerbating the uncertainty surrounding crop selection. Farmers require adaptive strategies to cope with these changing conditions, and machine learning-based crop recommendation systems provide a promising solution by analyzing historical weather data and climate projections to offer data-driven guidance. Additionally, resource inefficiency in agriculture, including the excessive use of water, fertilizers, and pesticides, poses significant environmental and economic concerns [8]. By recommending crops that align with local conditions and resource availability, this research aims to optimize resource utilization, reduce waste, and promote sustainable farming practices, thereby mitigating environmental degradation [9].

Sustainability is a global imperative, and agriculture plays a pivotal role in this pursuit. The research is motivated by the goal of fostering sustainable farming practices that address environmental impact issues such as soil erosion, water pollution, and habitat destruction. Machine learning-based recommendations not only prioritize high-yield crops but also consider their environmental footprint, contributing to responsible land management and conservation efforts. Moreover, this research is driven by the desire to empower smallholder farmers, who often lack access to advanced agricultural technology and expertise. By providing accessible and user-friendly crop recommendation systems, it seeks to empower these farmers, improve their livelihoods, and reduce poverty [10].

2. LITERATURE SURVEY

Prakash, et al. [11] proposed a Machine learning-based crop suggestion system. an intelligent model in crop suggestion predicts the right kind of crop that suits the soil and other environmental factors by using a machine learning algorithm. Temperature, wetness of the soil, environment dampness and pH are the essential information for crop suggestion in the agriculture field. Many existing works of IoT in the field of agriculture are related to yield prediction and few related to crop suggestion use fewer parameters and learning algorithms. Indira, et al. [12] proposed A Machine Learning based New Recommendation System to the Farmer. An algorithm called MobileNet uses an image of a leaf to identify the disease present in a plant. The XGBoost model predicts a suitable crop based on the local soil nutrients and rainfall. Random Forest [RF] model was used to propose fertilizer and develop ideas for improving soil fertility depending on nutrients present in the soil. When compared to other approaches, the proposed model delivers a high level of accuracy. Jeevaganesh, et al. [13] proposed a machine learning algorithm: AdaBoost to predict the yield of crops based on the parameters like state, district, area, seasons, rainfall, temperature, and area. To enhance the yield, this research study also suggests a fertilizer based on soil conditions like NPK values, soil type, soil PH, humidity, and moisture. Fertilizer recommendation is primarily done by using the Random Forest [RF] algorithm.

Wen, et al. [14] proposed Optimizing machine learning-based site-specific nitrogen application recommendations for canola production. This work employed four machine learning models, namely, support vector machine (SVM), gradient boosting (GB), random forest (RF), and ridge regression (RR) to predict the site-specific EONR values of canola crops, based on a 22-site-year field study across eastern Canada. SSL, et al. [15] proposed An Intelligent Crop Recommendation System using Deep Learning. The purpose of this work is to examine the approaches employed in extracting the water bodies utilizing the mode of satellite remote sensing. The goal of the proposed work is to collect the data of temperature and humidity and utilize the algorithm of clustering with the method of k-Nearest Neighbor to find out the patterns which are all hidden in them with a help of huge amount of dataset.

Chowdhary, et al. [16] proposed An Ensemble Model to Predict Crops using Machine Learning Algorithms. The authors tried to form an ensemble model using various machine-learning algorithms for better rice production. Crop production prediction utilizing AI Strategies aims to deliver improved outcomes, but the ensemble model provides better predictive results compared to the individual algorithm. we tried to use a combination of symmetric machine-learning algorithms to form an ensemble model for better prediction. Suresh, et al. [17] proposed a solution based on IoT and deep learning for improving paddy output. The system collects data through sensors and transfers it to the cloud in order to diagnose plant stress caused by soil fertility, environmental imbalance, and crop diseases (AWS EC2 Server). The proposed system is implemented in three stages. The stages of production are divided into three categories: the start stage (planting to panicle initiation), the middle stage (panicle initiation to flowering), and the end stage (flowering to maturity).

Raviraja, et al. [18] proposed machine learning-based mobile applications for autonomous fertilizer suggestion. In this work, the smartphone app will be used to construct an autonomous fertilizer suggestion system for farmers. The dataset was gathered from the website Kaggle. The encoding technique is used to make the data entries numeric. Three Machine Learning (ML) models K-Nearest Neighbour (KNN), Random Forest (RF), and Decision Tree (DT) are employed to predict fertilizer based on environmental and soil characteristics. Mamatha, et al. [19] proposed Machine learning-based crop growth management in a greenhouse environment using hydroponics farming techniques. In this proposed system, in hydroponics for germination organic coconut coir medium is used rather than rock wool, because rock wool is not bio-degradable and is composed of volcanic materials. In the proposed research of hydroponics, the system is automated on a wide scale covering the entire green house with different crops produced in different climatic conditions. Oikonomidis, et al. [20] proposed Hybrid deep learning-based models for crop yield prediction. The authors had developed deep learning-based models to evaluate how the underlying algorithms perform with respect to different performance criteria. The algorithms evaluated in our study are the XGBoost machine learning (ML) algorithm, Convolutional Neural Networks (CNN)-Deep Neural Networks (DNN), CNN-XGBoost, CNN-Recurrent Neural Networks (RNN), and CNN-Long Short-Term Memory (LSTM).

2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal

3. PROPOSED METHODOLOGY

3.1 Overview

The system leverages data analysis, visualization, machine learning models, and evaluation techniques to provide farmers with informed crop recommendations. It ensures that data is appropriately prepared and understood, models are trained and evaluated, and predictions are made using the learned patterns from historical data. The goal is to optimize crop yield based on environmental conditions and nutrient levels. Figure 1 shows the proposed system model. The detailed operation illustrated as follows:

step 1: Data Import and Preparation:

- Data is a critical component in machine learning-based systems. In this context, data includes information about different crops, such as their nutrient requirements (N, P, K), temperature preferences, humidity tolerance, pH levels, and rainfall needs.
- Data is typically stored in structured formats like CSV files. The system uses libraries like Pandas and NumPy for data manipulation. Pandas allows for efficient data storage, retrieval, and manipulation, while NumPy provides numerical operations for data analysis.

step 2: Exploratory Data Analysis (EDA):

- EDA is the process of summarizing and visualizing key characteristics of a dataset. It helps understand the data's distribution, potential outliers, and relationships among variables.
- The **df.tail()** function displays the last few rows of the dataset, providing a glimpse of the data.
- **df.size** and **df.shape** reveal the overall size and dimensions of the dataset, respectively.
- **df.columns** lists the column names, which are essential for identifying features and the target variable.

step 3: Data Visualization:

- Data visualization techniques, like count plots and heatmaps, are employed to gain insights from data.
- Count plots help visualize the distribution of different crops in the dataset, which is vital for understanding class imbalances and model biases.
- A heatmap is used to visualize the correlation between features. High positive or negative correlations between features can impact the model's performance. For instance, if temperature and humidity are highly correlated, the model may struggle to distinguish their individual effects on crop growth.

step 4: Feature Selection and Target:

- Features are the variables used as input to the machine learning model. In this case, features include factors like nutrient levels (N, P, K), environmental conditions (temperature, humidity, pH), and rainfall.
- The target variable ('label') represents the outcome or prediction the model is designed to make, which is the type of crop to recommend.

step 5: Model Training and Evaluation:

- Model training involves using historical data (features) and their corresponding outcomes (target) to teach the machine learning model how to make predictions.
- Two models are used here: Logistic Regression and Random Forest.

- Evaluation metrics like accuracy scores, precision, recall, F1-score, and classification reports are used to assess model performance. Accuracy measures how often the model makes correct predictions, while precision and recall provide insights into the model's ability to avoid false positives and negatives.

step 6: Confusion Matrix:

- A confusion matrix is a table used to describe the performance of a classification model. It shows the number of true positives, true negatives, false positives, and false negatives.
- True positives are instances where the model correctly predicts a positive class, while true negatives are instances where the model correctly predicts a negative class. False positives and false negatives represent prediction errors.

step 7: Accuracy Comparison Plot:

- The bar plot comparing the accuracies of different models helps users understand which model is more effective in making crop recommendations. Accuracy is a widely used metric for comparing models, but it may not tell the whole story, especially if class imbalances exist.

step 8: Making Predictions:

- Once the model is trained and evaluated, it can be used to make predictions for new, unseen data points. Users can input environmental conditions, and the model will predict the most suitable crop based on the patterns it learned during training.

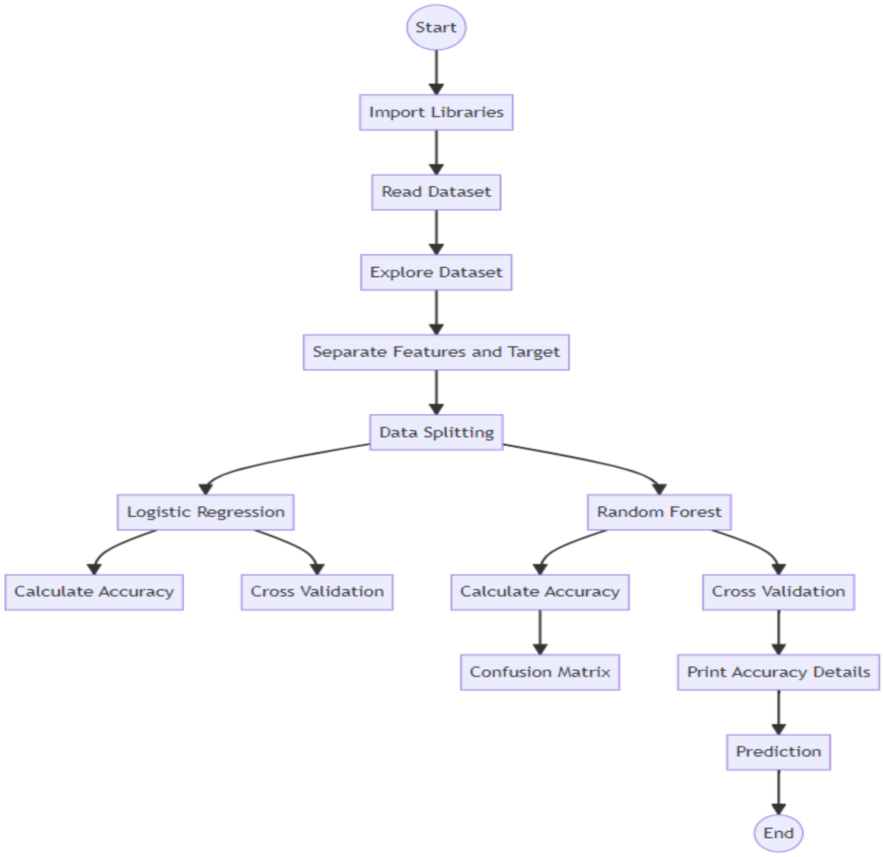

Figure 1. Proposed System model.

3.2 Data Preprocessing

- Each decision tree created is independent of the other thus it shows the property of parallelization.
- It is highly stable as the average answers given by a large number of trees are taken.
- It maintains diversity as all the attributes are not considered while making each decision tree though it is not true in all cases.
- It is immune to the curse of dimensionality. Since each tree does not consider all the attributes, feature space is reduced.

4. RESULTS

The dataset contains agricultural-related information with various features related to the cultivation of different crops. Here are columns in the dataset:

- N: This column appears to represent the nitrogen content in the soil, which is an essential nutrient for plant growth.
- P: This column likely represents the phosphorus content in the soil, another crucial nutrient for plant development.
- K: This column could represent the potassium content in the soil, which is important for various plant processes.
- Temperature: This column represents the temperature of the environment in which the crops are grown. Temperature significantly affects plant growth and development.
- Humidity: This column likely represents the humidity level of the environment. Humidity can influence plant transpiration and overall moisture conditions.
- pH: This column represents the pH level of the soil. Soil pH affects nutrient availability to plants.
- Rainfall: This column could represent the amount of rainfall in the region where the crops are cultivated. Rainfall is a critical factor in determining irrigation needs and water availability.
- Label: This column appears to be the target variable or the label that indicates the type of crop being grown. It seems to be a categorical variable representing different types of crops, such as "rice" and "coffee."

The dataset is organized with rows corresponding to different instances or observations, each with values for the different features (N, P, K, Temperature, Humidity, pH, Rainfall) and a label indicating the type of crop. This kind of dataset is used for various machine learning tasks, such as classification (predicting the crop type) or regression (predicting some numerical outcome based on the features.

2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal

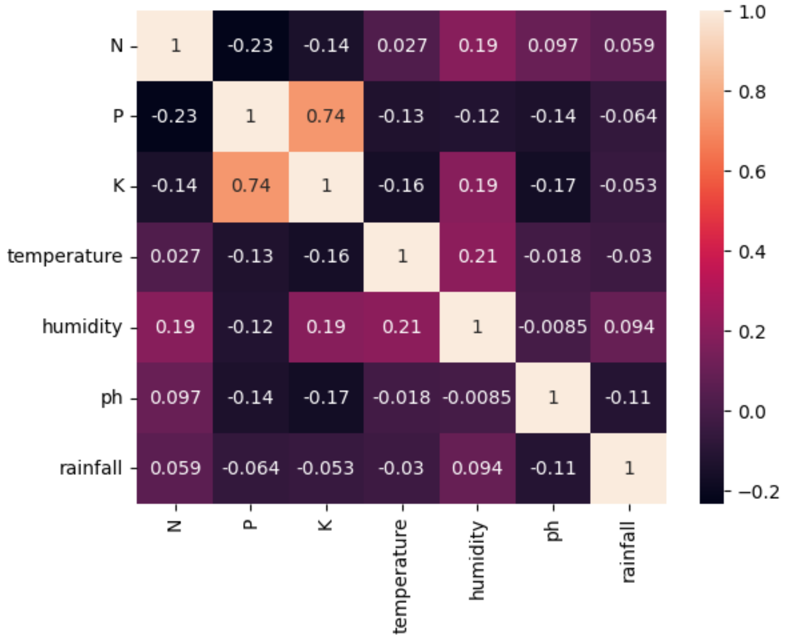

Figure 2: correlation between numerical columns in the Data Frame

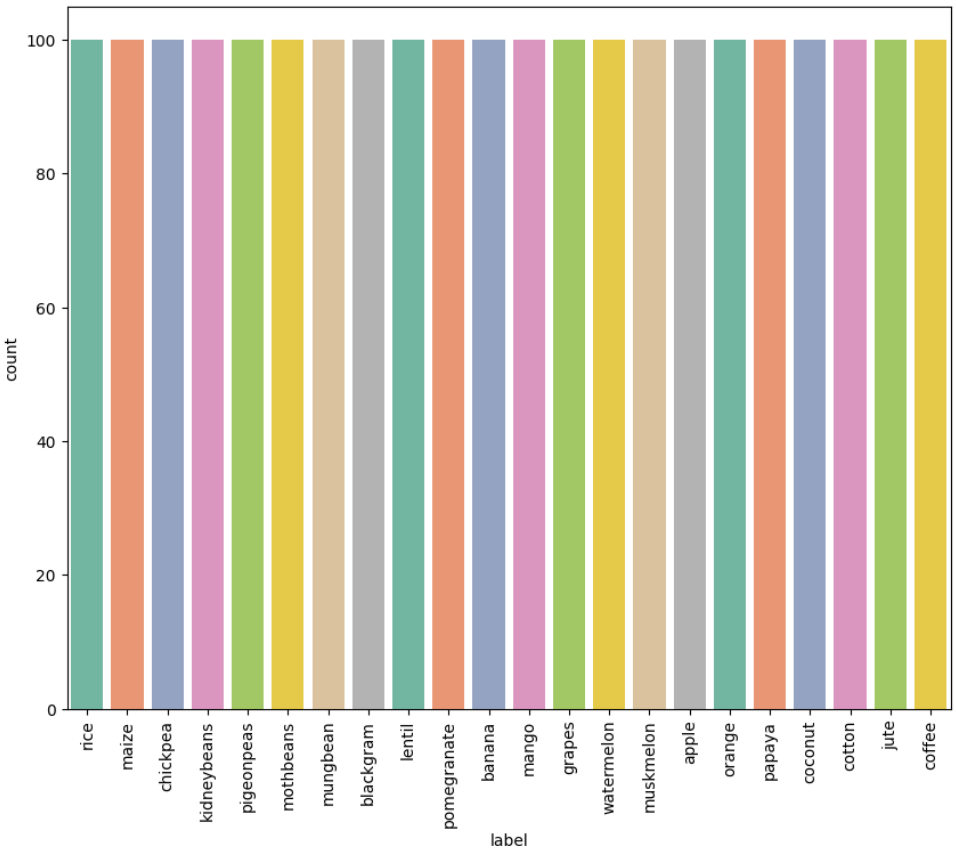

Figure 3: showcase the distribution of categories in the "label" column of the Data Frame

2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal

```
classification_report:

               precision    recall  f1-score   support

       apple       1.00      1.00      1.00        13
      banana       0.81      1.00      0.89        17
   blackgram       0.67      0.75      0.71        16
    chickpea       1.00      1.00      1.00        21
     coconut       1.00      0.95      0.98        21
      coffee       1.00      1.00      1.00        22
      cotton       0.74      1.00      0.85        20
      grapes       1.00      1.00      1.00        18
        jute       0.89      0.57      0.70        28
 kidneybeans       0.93      1.00      0.97        14
      lentil       0.81      0.91      0.86        23
       maize       1.00      0.43      0.60        21
       mango       1.00      1.00      1.00        26
   mothbeans       0.76      0.68      0.72        19
    mungbean       1.00      1.00      1.00        24
   muskmelon       1.00      1.00      1.00        23
      orange       1.00      1.00      1.00        29
      papaya       0.78      0.74      0.76        19
  pigeonpeas       1.00      0.83      0.91        18
 pomegranate       1.00      0.94      0.97        17
        rice       0.48      0.88      0.62        16
  watermelon       1.00      1.00      1.00        15

    accuracy                           0.89       440
   macro avg       0.90      0.89      0.89       440
weighted avg       0.91      0.89      0.89       440
```

Figure 4: Classification report of Logistic Regression

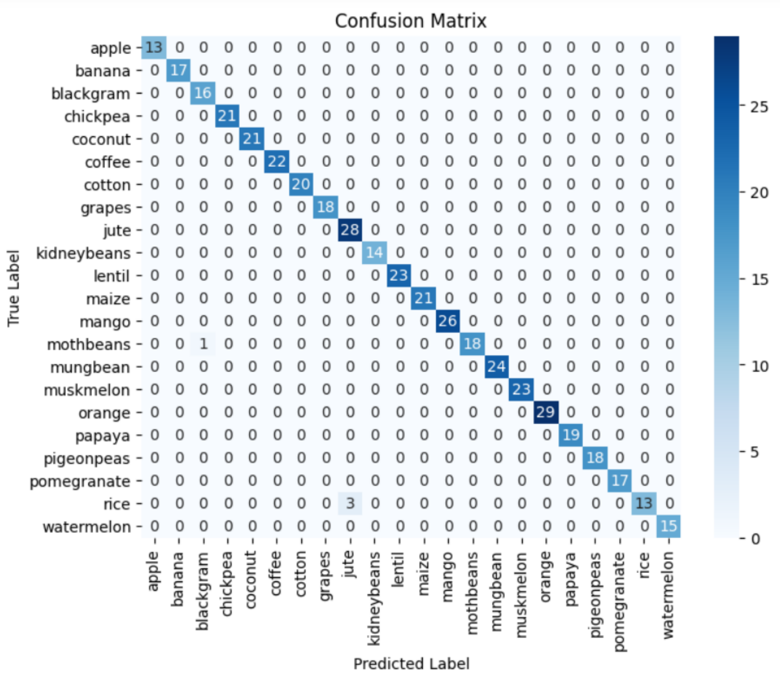

Figure 5: Confusion matrix of Random Forest classifier.

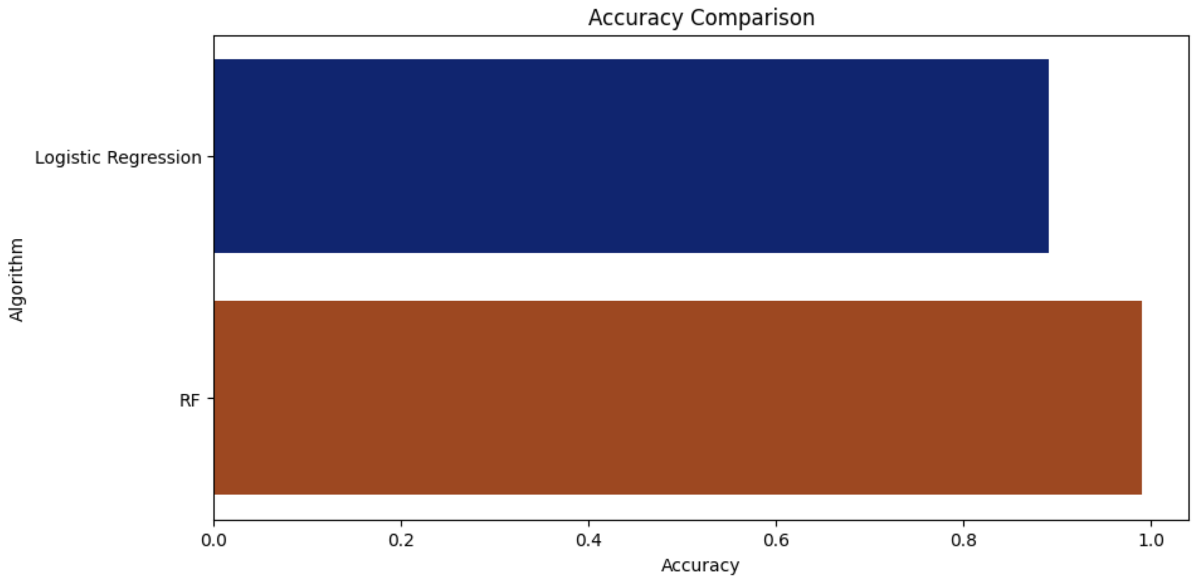

Figure 6: Accuracy Comparison Graph for ml models

Table 1: Overall performance comparison of ML models.

Model name	Accuracy (%)	Precision (%)	Recall (%)	F1-score
Random Forest	99	99	99	99
Logistic Regression	89	90	89	89

5. Conclusion

In conclusion, the "Machine Learning-based Crop Recommendation System for Enhanced Yield Production" represents a comprehensive approach to empower farmers with data-driven insights for making informed decisions about crop selection. This system operates on fundamental principles of data science and machine learning. Beginning with data import and preparation, the system recognizes the critical role of data in its operations. It leverages libraries like Pandas and NumPy for efficient data handling, ensuring that historical information about various crops' growth requirements is readily accessible. Exploratory Data Analysis (EDA) forms the foundation for understanding the dataset's characteristics. By utilizing functions like **df.tail()**, **df.size**, and **df.shape**, it gains insights into the data's structure and size. Furthermore, examining the dataset's columns with **df.columns** sets the stage for feature selection and target variable identification. Data visualization techniques, including count plots and heatmaps, enable a visual understanding of crop distribution and feature correlations. These visuals provide invaluable context, aiding in better-informed decisions during subsequent stages of model development. Feature selection involves extracting relevant factors such as nutrient levels (N, P, K), environmental conditions (temperature, humidity, pH), and rainfall. These features become the basis for training machine learning models, and the target variable ('label') defines the model's prediction goal. The training and evaluation of models, specifically Logistic Regression and Random Forest, embody the heart of this system. Model performance is rigorously assessed using metrics like accuracy, precision, recall, and F1-score. These evaluations ensure that the models can make accurate crop recommendations. A critical aspect is the confusion matrix, which provides a detailed breakdown of prediction results, highlighting true positives, true negatives, false positives, and false negatives. This aids in understanding model strengths and weaknesses. The system also facilitates model selection by comparing accuracies through a bar plot, enabling users to identify the most suitable model for their specific needs. Finally, the ability to make predictions for new data points based on environmental conditions allows farmers to apply the system's insights directly to their farming practices, potentially leading to optimized crop yield production.

REFERENCES

[1] Kilaru, Ruthvik, and Kommisetti Murthy Raju. "Prediction of maize leaf disease detection to improve crop yield using machine learning based models." 2021 4th International Conference on Recent Trends in Computer Science and Technology (ICRTCST). IEEE, 2022.

[2] Kheir, Ahmed MS, et al. "Machine learning-based cloud computing improved wheat yield simulation in arid regions." Computers and Electronics in Agriculture 203 (2022): 107457.

[3] Pham, Hoa Thi, Joseph Awange, and Michael Kuhn. "Evaluation of three feature dimension reduction techniques for machine learning-based crop yield prediction models." Sensors 22.17 (2022): 6609.

[4] Khan, Rijwan, et al. "Machine learning-based agriculture." Application of Machine Learning in Agriculture. Academic Press, 2022. 3-27.

[5] Manjula, E., and S. Djodiltachoumy. "Efficient prediction of recommended crop variety through soil nutrients using deep learning algorithm." Journal of Postharvest Technology 10.2 (2022): 66-80.

[6] Gopi, P. S. S., and M. Karthikeyan. "Red fox optimization with ensemble recurrent neural network for crop recommendation and yield prediction model." Multimedia Tools and Applications (2023): 1-21.

[7] Nihar, Ashmitha, N. R. Patel, and Abhishek Danodia. "Machine-Learning-Based Regional Yield Forecasting for Sugarcane Crop in Uttar Pradesh, India." Journal of the Indian Society of Remote Sensing 50.8 (2022): 1519-1530.

[8] Li, Kai-Yun, et al. "Toward automated machine learning-based hyperspectral image analysis in crop yield and biomass estimation." Remote Sensing 14.5 (2022): 1114.

[9] Sagana, C., et al. "Machine Learning-Based Crop Recommendations for Precision Farming to Maximize Crop Yields." 2023 International Conference on Computer Communication and Informatics (ICCCI). IEEE, 2023.

[10] Ennaji, Oumnia, Leonardus Vergutz, and Achraf El Allali. "Machine learning in nutrient management: A review." Artificial Intelligence in Agriculture (2023).

[11] Prakash, R. Varun, M. Mohamed Abrith, and S. Pandiyarajan. "Machine learning based crop suggestion system." 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, 2022.

[12] Indira, D. N. V. S. L. S., et al. "KRISHI RAKSHAN-A Machine Learning based New Recommendation System to the Farmer." 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, 2022.

[13] Jeevaganesh, R., D. Harish, and B. Priya. "A Machine Learning-based Approach for Crop Yield Prediction and Fertilizer Recommendation." 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI). IEEE, 2022.

[14] Wen, Guoqi, et al. "Optimizing machine learning-based site-specific nitrogen application recommendations for canola production." Field Crops Research 288 (2022): 108707.

[15] SSL, Durai Arumugam, R. Praveenkumar, and V. Balaji. "An Intelligent Crop Recommendation System using Deep Learning." International Journal of Intelligent Systems and Applications in Engineering 11.10s (2023): 423-428.

[16] Chowdhary, Muppala Sunny, et al. "InteliCrop: An Ensemble Model to Predict Crop using Machine Learning Algorithms." 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI). IEEE, 2022.

[17] Suresh, Chalumuru, et al. "Paddy Crop Monitoring System Using IoT and Deep Learning." Innovative Data Communication Technologies and Application: Proceedings of ICIDCA 2021. Singapore: Springer Nature Singapore, 2022. 791-807.

[18] Raviraja, S., et al. "Machine learning based mobile applications for autonomous fertilizer suggestion." 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA). IEEE, 2022.

[19] Mamatha, V., and J. C. Kavitha. "Machine learning based crop growth management in greenhouse environment using hydroponics farming techniques." Measurement: Sensors 25 (2023): 100665.

[20] Oikonomidis, Alexandros, Cagatay Catal, and Ayalew Kassahun. "Hybrid deep learning-based models for crop yield prediction." Applied artificial intelligence 36.1 (2022): 2031822.