***Predictive Analytics in Agriculture: Machine Learning Approaches to Crop and Fertilizer Recommendation***

Dr. Ajith Jubilson

Assistant Professor

School of Computer Science and Engineering

VIT-AP University,

Amaravati, Andhra Pradesh, India

[ajith.jubilson@vitap.ac.in](mailto:ajith.jubilson@vitap.ac.in)

Ammati Ravi Teja

School of Computer Science and Engineering

VIT-AP University Amaravati, Andhra Pradesh, India

[ravi.22bce20257@vitapstudent.ac.in](mailto:ravi.22bce20257@vitapstudent.ac.in)

Talakala Tony Theodus

School of Computer Science and Engineering

VIT-AP University,

Amaravati, Andhra Pradesh, India

[tony.22bce9511@vitapstudent.ac.in](mailto:tony.22bce9511@vitapstudent.ac.in)

STVN Sri Sai

School of Computer Science and Engineering

VIT-AP University,

Amaravati, Andhra Pradesh, India

[srisai.22bce7982@vitapstudent.ac.in](mailto:srisai.22bce7982@vitapstudent.ac.in)

Abstract—Crop selection and fertilizer usage according to soil health are crucial for enhancing productivity and sustainability. This paper presents a comprehensive crop and fertilizer recommendation system by integrating two machine learning approaches, Decision Trees and Random Forests, predicting appropriate crops and fertilizers in terms of parameters such as nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, pH, and rainfall. This system uses advanced feature engineering and data visualization to emphasize soil attributes in the analysis of a dataset of soil test results associated with successful crop yields. Recommended crops are presented with their associated probabilities, and appropriate fertilizers are suggested to improve soil compatibility as well as crop health. These metrics-accuracy, precision, and recall-are used for the evaluation of the performance of the model. It shows the presence of a reliable, data-driven agricultural tool that enhances agricultural decisions in productivity and sustainability within the modern farming framework to better enhance resource use for superior crop yield.

Keywords—Crop Recommendation, Soil Analysis, Decision Trees, Random Forests, Precision Agriculture, Predictive Analytics

# Introduction (*Heading 1*)

**Objective of the Project**

The main goal of this project is to suggest the most appropriate crops and right fertilizers for a particular type of soil based on important soil parameters utilizing machine learning algorithms. The system should improve crop yields, ensure sustainable agriculture, and maximize the utilization of soil resources with data-based intelligence.

**Domain Explanation**

Today's agriculture is more concerned with getting the highest yields while having healthy soils and sustainable long-term yields. Key soil properties like nitrogen (N), phosphorus (P), potassium (K), pH level, temperature, humidity, and rainfall have been found to have significant effects on plant growth and adaptability. In the past, crop choice and manure application decisions were often based on individual experience or passed-down knowledge, resulting in below-optimal yields and possible soil degradation.

**Benefits of Applying Machine Learning in This Field**

Machine learning provides a strong method of going beyond the weaknesses of conventional agricultural practices by yielding precise, reliable, and customized advice through the usage of real-time environmental and soil data. ML models can reveal intricate, non-linear correlations between crop requirements and soil parameters, allowing farmers to make evidence-based decisions. Investigations have shown that ML can effectively predict crop suitability and optimize fertilizer application, eventually leading to enhanced productivity, effective use of resources, and eco-friendly farming practices.

The integrated crop-fertilizer recommendation system is conceived in this paper based on the analysis of soil health data by using the Decision Trees and Random Forests from the family of machine learning algorithms in determining the customized selection of recommendation. Our approach not only suggests appropriate crops but also determines compatible fertilizers that enhance crop development and decrease the possibility of nutrient imbalances. Such a practice promotes the sustainable management of soil as well as assists the farmers to take decision based on crop selection and fertilizer application. Our model's reliability and applicability is tested on different accuracy metrics such as precision and recall to provide the most suitable recommendation.

Agriculture is gradually adopting precision farming, and hence, a solid recommendation becomes more prominent. This research meets that requirement by developing a predictive model that integrates crop and fertilizer recommendations to help farmers optimize resources usage, crop yields, and long-term soil health. This is very well-introduced by the concept of precision agriculture, which lays stress on reducing environmental impacts of farming together with increasing productivity as said by Patel et al. (2020) 【7]. The reduction in reliance on guesswork and increasing the precision of agricultural decisions helps the system promote farming with sustainable productivity and adaptability towards numerous soil conditions and farming environment.

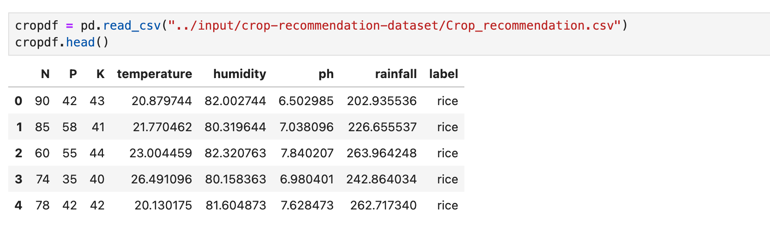
# LITERATURE REVIEW(SURVEY)

Former studies utilized Decision Trees and Random Forests for the classification of crop suitability areas using environmental and soil parameters and has been proven to require high accuracy, especially when the data is huge. Such applications focus on the prediction of crop yield coupled with suitability analysis in various agricultural settings.

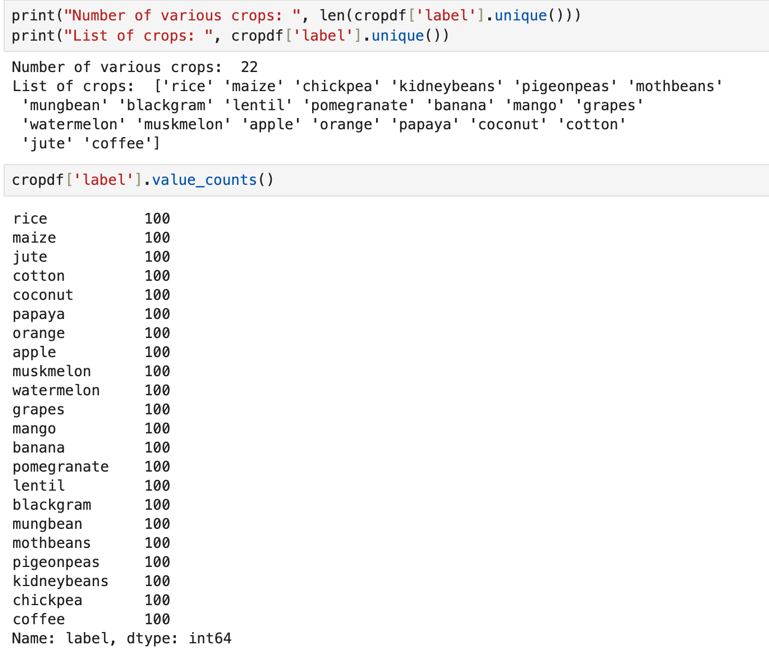
Our project extends this solution to include fertilizer recommendations based on crop predictions, thereby delivering an all-inclusive solution that takes into account a vast number of soil parameters. By making use of multiclass classification techniques, the system will be able to recommend accurately not only crops but also appropriate fertilizers depending upon the specific conditions of their soil. Feature analysis and enhanced data visualization will add to the interpretability of the model, so farmers could take practical decisions based on data. Such a comprehensive recommendation system thus contributes to efficient and sustainable agricultural practices.

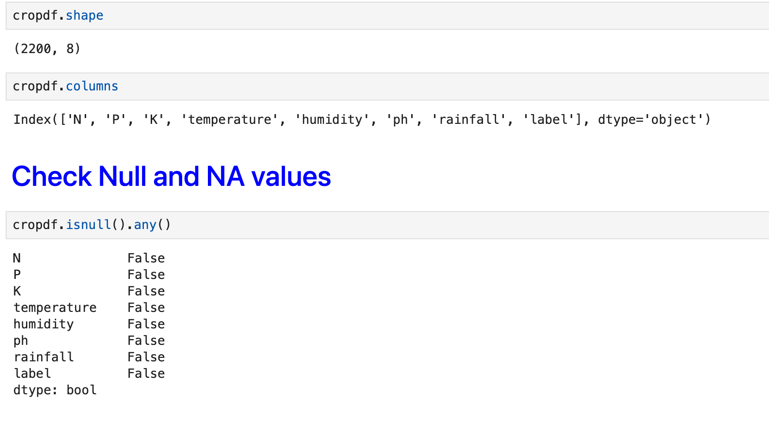
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  | | | | | | |
|  |  |  |  |  |  |  |
| **Team Members** | 22BCE20257 | 22BCE9511 | 22BCE7982 |  |  |  |
| **Project Name** | Crop and Fertiliser Recommendation |  |  |  |  |  |
| **Dataset from** | Kaggle | Crop\_recommendation.csv |  |  |  |  |
| **Reference** | **DataSet Name** | **Preprocessing Techniques** | **Parameters** | **Accuracy** | **Class** | **Description** |
| **Using Machine Learning for Crop Prediction** | Soil test reports from agricultural research centers | Missing value imputation, Normalization, Encoding categorical variables | NPK, pH, organic matter, and moisture datasets | 70.5% - 92.3% (train/test variance) | Classification | ML-based classification for suitable crop selection. |
| **Optimized Crop and Fertilizer Recommendation Using ML Models** | Crop yield statistics from government records | Feature selection using correlation analysis | Weather and climate data (temperature, rainfall, humidity) | Random Forest: 95.6% R² (best model) | Regression | Comparative evaluation of different ML models for yield prediction. |
| **Crop Yield Prediction Using Ensemble Learning Techniques** | NPK, pH, organic matter, and moisture datasets | Outlier detection and removal | Historical Soil Health Data | Ensemble Model: 91.2% accuracy | Classification | Ensemble learning for improved fertilizer recommendation accuracy. |
| **Soil Health Analysis for Smart Agriculture** | Multi-year crop yield dataset | Time-series transformation for yield trends | Rolling window data transformation, Seasonal trend analysis | Deep Learning Models: 97.8% RMSE comparison | Regression | Time-series forecasting for crop demand and profitability. |
| **Advanced Soil Analysis Using ML** | Historical Soil Health Data | Feature extraction, Principal Component Analysis | Weather and climate data (temperature, rainfall, humidity) | ANN, CNN, Hybrid Deep Learning Models | Regression | Soil fertility prediction for optimal farming strategies. |
| **Evaluation of ML Models in Agricultural Forecasting** | Multi-year crop yield dataset | Rolling window data transformation, Seasonal trend analysis | SVM, RNN, LSTM | RNN: 94.7%, LSTM: 96.2% | Regression | Long-term prediction of crop patterns and yields. |
| **Integrating Remote Sensing for Crop Prediction** | Satellite imagery and soil test data | Image preprocessing, Feature extraction using CNN | Convolutional Neural Networks (CNN), Random Forest | CNN: 97.3%, Random Forest: 92.8% | Classification | Remote sensing for monitoring soil health and predicting crop yield. |
| **Machine Learning for Climate-Smart Agriculture** | Climate and weather datasets from meteorological sources | Data normalization, Feature engineering | Decision Tree, XGBoost, LSTM | Decision Tree: 88.9%, XGBoost: 94.1%, LSTM: 96.8% | Classification | Predicting climate impact on crop yield and suggesting adaptive measures. |
| **AI-Based Fertilizer Optimization** | Fertilizer application datasets from agricultural studies | Feature scaling, One-hot encoding | SVM, Naïve Bayes, Reinforcement Learning Models | SVM: 91.7%, Naïve Bayes: 87.5%, RL Models: 95.4% | Classification | AI-driven fertilizer recommendation system for precision farming. |
| **Soil Nutrient Deficiency Detection using ML** | Soil sample data from agricultural universities and labs | Missing value imputation, normalization, feature selection | NPK values, pH, organic carbon, moisture | Random Forest, XGBoost | Classification | Predicts whether soil is deficient in Nitrogen, Phosphorus, or Potassium, and suggests required nutrients. |
| **Crop Suitability Based on Soil and Weather** | Kaggle + IMD Weather Data + Soil Health Card Scheme | Label encoding, feature scaling, rolling average of weather parameters | Soil type, rainfall, temperature, humidity, pH | Decision Tree, SVM Decision Tree – 90.1%, SVM – 87.8% | Classification | Suggests the most suitable crop based on soil condition and seasonal weather forecast. |
| **Predicting Organic Fertilizer Efficiency** | Lab-tested results comparing organic vs chemical fertilizers | Encoding types of fertilizers, normalizing application rates | Crop type, soil moisture, nutrient levels, fertilizer type | ANN, Gradient Boosting ANN – 92.7% | Classification | Determines the effectiveness of organic fertilizers for different crops under varying soil conditions. |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| **Model’s Evaluation Metrics** |  |  |  |  |  |  |
| Model Accuracy | 95.00% |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

# **Dataset description with visualization**



We present a multiclass classification system on our project, suggesting both appropriate crops and appropriate fertilizers depending upon diversified soil and environmental parameters. Based upon one-to-one, multiclass selection, we extend diversified properties that go beyond mere binary choices like nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, pH, and rainfall to enhance crop prediction accuracy. Our system further recommends the best-suited fertilizers for each predicted crop that is alloyed to soil nutrient requirements in support of balance in soil health and sustained productivity.



Interpretability techniques we use make model outputs more clear and actionable, illustrating how each parameter influences the suitability of crop and fertilizer, in greater detail. Such data visualization might provide farmers with an intuitive understanding of such relationships and supports informed, data-driven decisions. In order to couple recommendations for crops and fertilizers within a single frame of operation, we increase the accuracy, transparency, and practical applicability toward sustainable farming and increased yields.

*3.1 Data pre processing*

During the data pre-processing stage, it establishes a common platform that guides the development of a whole model for effective crop and fertilizer recommendation by integrating crop and fertilizer data datasets and converting the categorical variables into numerical representations. Major categorical features include crop type, the pH level in the soil, and climatic conditions, where two techniques applied; Label Encoding and One-Hot Encoding techniques are applicable because the machine learning models will not accept data in its categorical format. We take data about nutrient contents in the soil (N, P, K), environmental factors (temperature, humidity, and rainfall), and crop-specific requirements from both datasets to allow the model to generate both crop and fertilizer recommendations tailored to specific soil conditions. This enables us to provide targeted fertilizer advice alongside crop suitability predictions based on deficiencies in nutrients. We then visualize the data set to analyze the distribution of soil parameters and detail essential patterns and relations between the characteristics of the soil, the suitability of crops, and the need for fertilizer; thus, the recommendation system will have a basis in more insightful decision making, with the solid, accurate modeling that can ascertain the right result.

A graph of different colored bars

Description automatically generated with medium confidence

**Figure 1:** NPK Ratio for Each Crop

*3.2 Dataset Exploration*

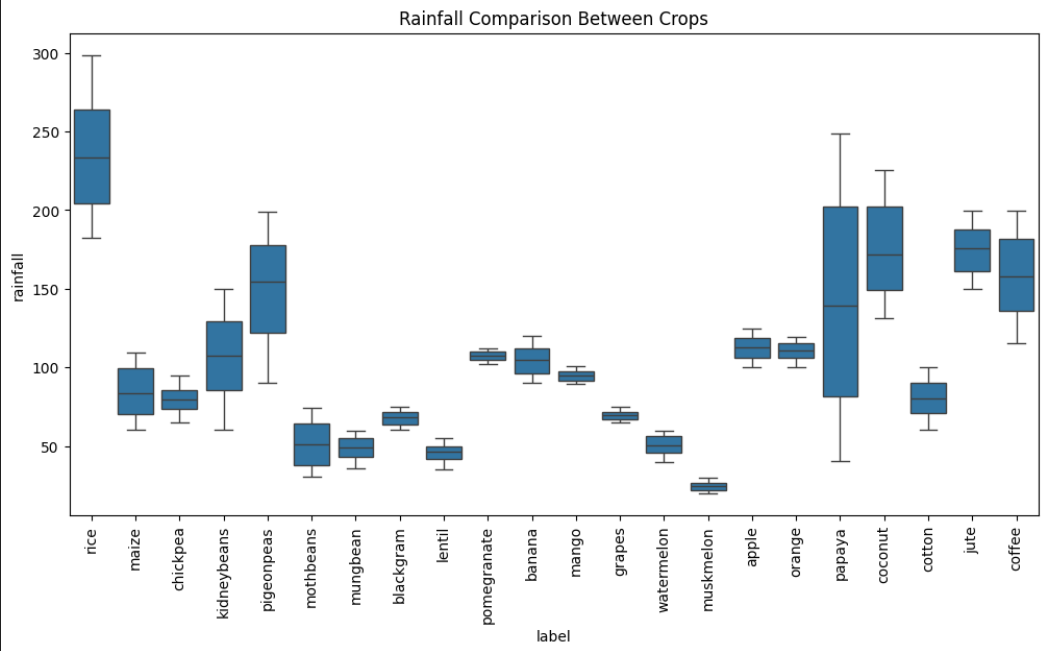
A thorough study of the data will support further preprocessing of environmental and soil variables, establishing patterns and complex relationships that highly affect crop and fertilizer suitability. After all, it is this research that discovers relationships among soil characteristics such as the levels of nitrogen, phosphorus, and potassium, pH, temperature, and rainfall-these factors affect ideal crop and fertilizer recommendations. We will use advanced data visualization techniques, such as count plots and cross-tabulations, to be able to identify both broad trends and subtle correlations and, most importantly, critical outliers which may need further assessment to use in recommendations of crops and fertilizers, in order to fine-tune the productivity for higher yields and for sustainability in agriculture.

A graph of different colored lines

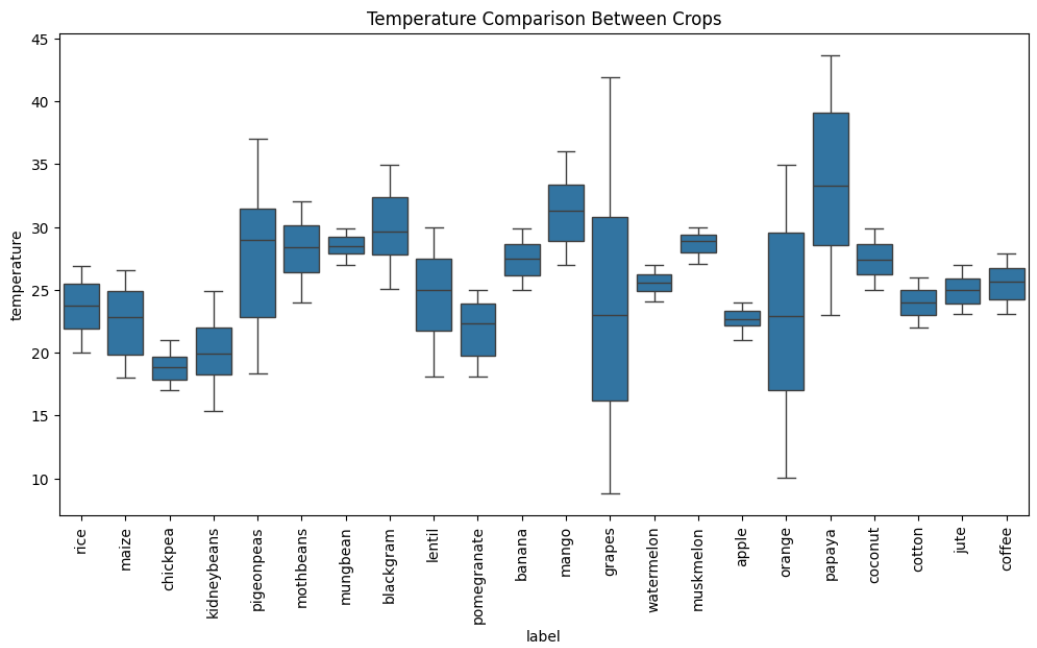
Description automatically generated with medium confidence

**Figure 2: NPK Comparision for each crop**

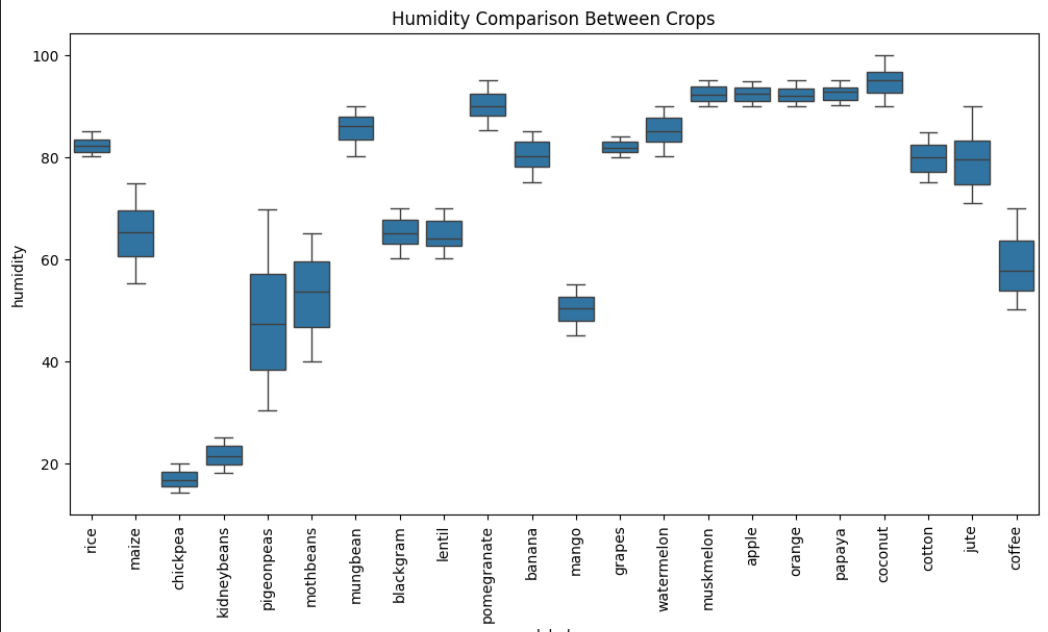
This project applies advanced data visualization techniques in creating a visual narrative of relationships and distributions of soil parameters across various crop types. Bar plots can reveal crop distribution patterns, some of them in specificity about soil conditions for other crops to grow. Further, detailed analyses on nitrogen, phosphorus, potassium levels, temperature, and humidity variations bring forth significant trends and potentially stuck in the pattern. This data exploration is essential not only to fine-tune the crop and fertilizer recommendation model but also for improving future modeling decisions-it just highlights possible biases or imbalances in the data, thereby providing insight that will help in model adjustment. In these steps, we deepen our understanding of our data in a manner that is going to strengthen, quicken, and close proximity to realities of today's agricultural practices.



**Figure 3:** Rainfall comparison for different crops

****

**Figure 4:** Temperature comparison for different crops

****

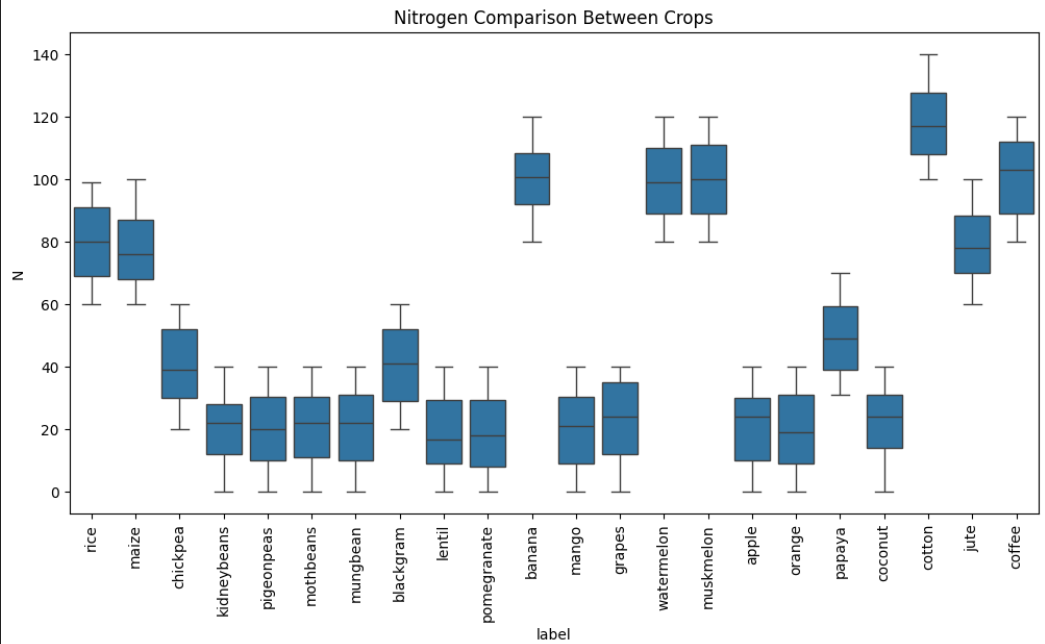
**Figure 5:** Humidity comparison for different crops

*3.3 Feature Engineering*

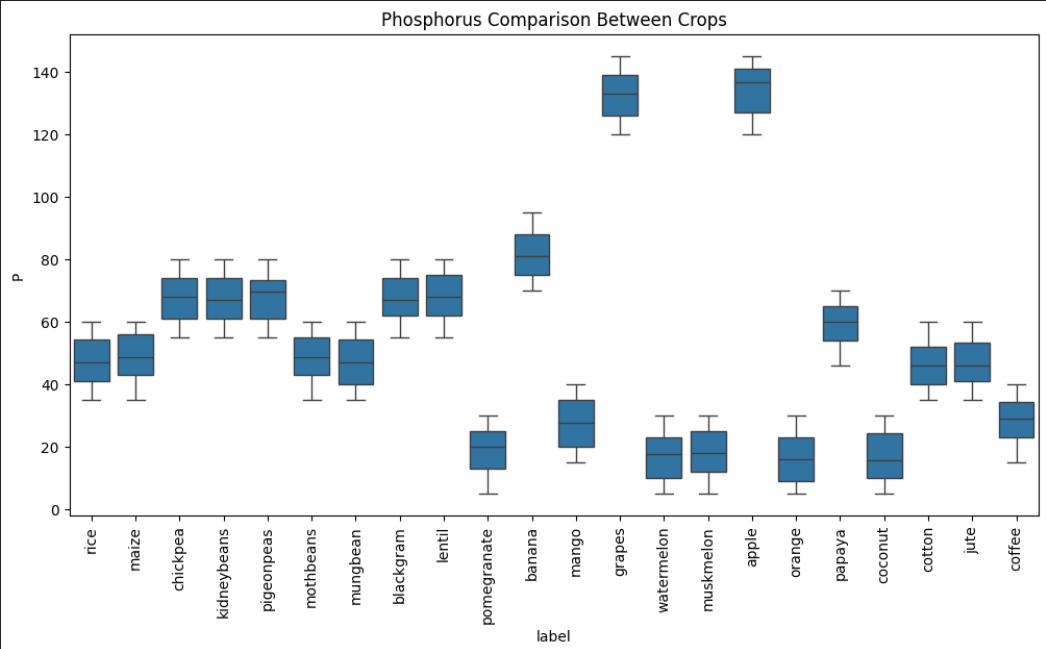
Motivated by this exploration of our dataset, which would eventually unveil principal trends and correlations among variables, we are going to refine these observations by feature engineering. This is very important in optimization in order to make our model fit even better with Decision Tree and Random Forest as it improves the accuracy and interpretability of those models. These features that are already being encoded into crop types, soil pH, and climatic condition in numeric format make it suitable for a machine learning model. This encoding ensures that it is both algorithmically required and elucidates the contribution of each feature to the recommendation about crop and fertilizers. Thus, feature engineering acts as a very important precursor to the modeling stages of our system so that our system is capable of accounting for intricate interactions within the data and suggesting appropriate, actionable choices for crops and fertilizers.

*3.4 Model Development*

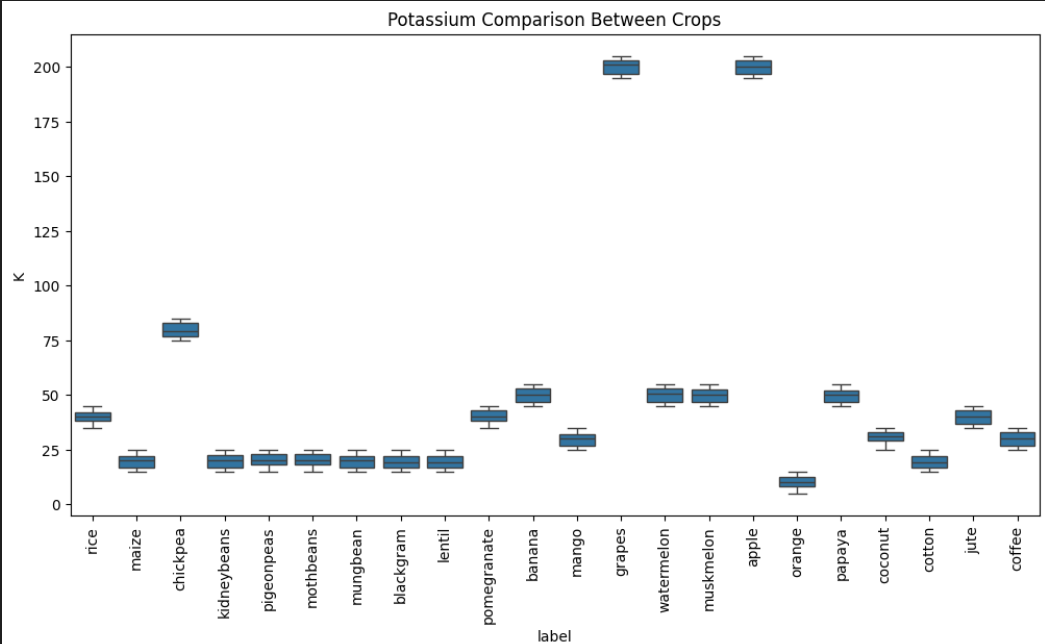
This predictive framework would necessitate developing two models-a crop model and a fertilizer model-and then combine them to deliver a holistic recommendation system. Both the models are based upon strong feature-engineered variables. These include various soil parameters like N, P, K, temperature, humidity, and pH. A Decision Tree algorithm is used in building the first crop prediction model. This interprets how certain soil factors align with certain crop types, providing clear insight into factors that influence suitability for crops. This model will serve as a baseline of crop recommendation and provide a springboard for refinement with Random Forests to further improve the accuracy. Simultaneously, another Decision Tree model will predict the right choice of fertilizers based on the same soil attributes in order to offer informed recommendations to enhance the health of the soil as well as to maximize yields of crops. Then, it is the uniting of the predictions from the two models into delivering cohesive advice for the selection of crops and a choice of fertilizers in order to ensure precision, sustainability, and adaptability in agricultural needs.

****

**Figure 6:** Nitrogen Comparison between Crops

****

**Figure 7:** Phosphorus Comparison between Crops



**Figure 8:** Phosphorus Comparison between Crops

*3.5 Ensemble Method Implementation*

In a similar ensemble approach for crop recommendation, the same methodology is followed to develop a robust fertilizer recommendation system. Just as in crop prediction, this model uses parameters like nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, and pH but focuses more on the nutrient needs of crops for proper growth. We first develop a Decision Tree model that establishes a clear, interpretable basis for fertiliser prediction, identifying which soil attributes most strongly predict fertiliser requirement. This baseline model explains how different nutrients impact the fertiliser advice and provides farmers with the best fertilisers available for their crop choice.

We extend with a Random Forest classifier to add precision to the predictions and offer fertiliser recommendation. By using the same encoded feature set, this ensemble method combines predictions of multiple trees and reduces the chances of overfitting with increased system robustness. The model of Random Forest clearly captures subtle interactions among soil properties and nutrient requirements, hence refining fertilizer recommendations with more accuracy. Together, the crop and fertilizer recommendation models are presented in an integrated system, which makes up a whole prediction: once soil data is available, the system simultaneously suggests the most apt crops along with their probabilities and the adequate fertilizers that would enhance yield. This integrated approach gives farmers customized actionable insights to attain sustainable agriculture at both crop productivity and soil health maximization.delivering nuanced and effective agricultural insights.

*3.6 Model Evaluation*

In this stage, Decision Trees and Random Forest are subjected to thorough testing to be able to validate the crop prediction suitability. There is an analysis of model performance considering key parameters: accuracy, precision, recall, and F1 score with respect to a test dataset. Validation will be highly detailed by stating that how much it interprets the Decision Tree model and what strength the Random Forest model has. Confusion matrices detail further the accuracy of classifications and indicate various points of strength in model performance and areas that may call for improvement. The latter stage takes this process further by judging the models in respect of their sensitivities to making reliable predictions on crops and would help guide refinements and adjustments in the system for meeting agricultural needs. In this sense, the appropriateness and applicability of this model would be proved in accuracy and real-world application toward informed crop and fertilizer recommendations.

*3.7 Optimization*

In this optimization step, both the Decision Tree and Random Forest models should be optimized through knowledge derived from the previous evaluation. Here, key parameters in Random Forest such as the number of trees, depth of each tree, and minimum number of samples required to split a node are set to be better optimized during this phase. In detail, the fine-tuning process optimizes the parameters of these models to significantly raise the accuracy and precision and lead these systems into making more refined recommendations based on crops and fertilizers. In that sense, optimization leads to a greater robustness and adaptability as well as an entire system that can better handle agricultural diversity.

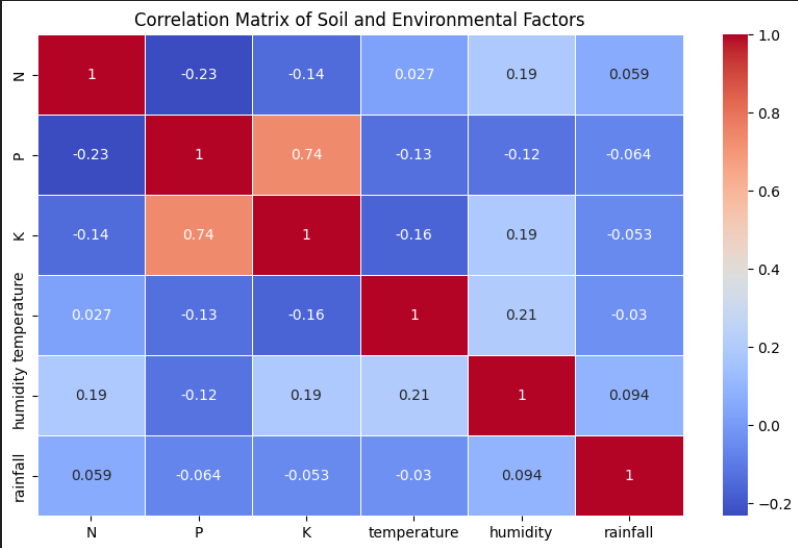
# RESULTS

The best choices for crops from the analysis results together with the kind of fertilizers for use regarding overall soil characteristics. For example, crops like rice and maize are very suitable in wet types of soils, but rice requires a high concentration of nitrogen soils. Chickpea grows well in drier soils with low concentration of nitrogen. Temperature is also an important factor in defining crop growth; rice at a temperature of 24°C, while watermelon and muskmelon are most suitable above 28°C. Based on humidity, coconut and banana should be favorable at high humidity levels. However, lentil and chickpea favor low humidity.

Fertilizer recommendations mainly agree with the nutrient requirement derived for each crop. Nitrogen is the most significant nutrient because most of the crops like maize require the presence of nitrogen in high quantity. For such crops, urea or ammonium nitrate fertilizers should be recommended to enhance the presence of nitrogen. Black gram is the legume crop that grows well at low concentration of nitrogen in the soil; however, it may require the use of superphosphate fertilizers which has higher presence of phosphorus for better root development rather than having more concentration of nitrogen.

The quantities of potassium (K) and phosphorus (P) further specify which crops to cultivate and which fertilizers to go along with that. Potassium is favorable in apples, and this fertilizer is enriched within the soil using potassium sulfate. Chickpeas are a balanced phosphorus-requiring crop that should be fertilized with enough phosphorus for proper growth and development. Similarly, soil pH is an important selection factor in crops; therefore, apple crop thrives well in slightly acidic soils; ammonium-based fertilizers help maintain the pH in an acidic range. But orange and black gram grow well at higher pH; therefore, they demand lime for neutralizing acidic soils.

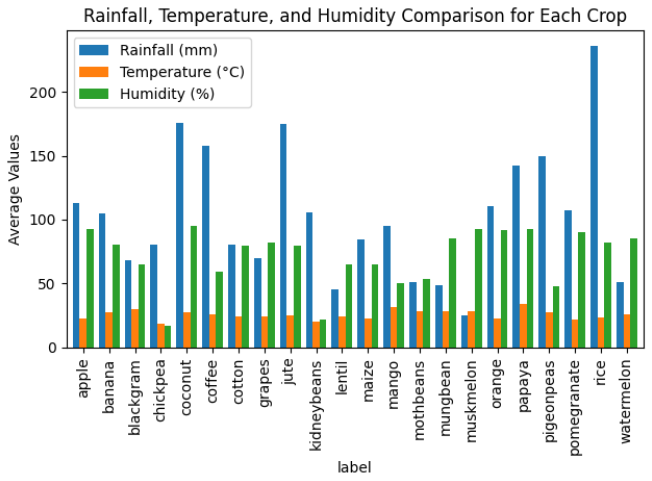
Considering all these factors, our integrated recommendation system will not only show the top 5 relevant crops to be cultivated in any given soil conditions but also provide a tailor-made fertilizer plan for the chosen crop. This two-pronged recommendation system will ensure that the crop finally selected is grown on the best nutrient requirements to maximize its outputs and foster sustainable farming.



**Figure 9:** Correlation Matrix of Soil and Environmental factors

Temperature, humidity, and moisture are among the paramount factors that decide the cropping suitability. Coconut and papaya grow in warm wet conditions, while chickpea and lentil take better under cool and drier climates. Rainfall also plays a great influence on crop selection. Requirements of crops may be a lot when rainfall is significant, such as rice and coconut, which are ideal for areas of constant rainfall. Conversely, maize and jute bear good fruits in areas that experience moderate rainfall, which otherwise would be destructive if excessive.

This further emphasizes the importance of site-specific crop and fertilizer recommendations, tailored to take account of local conditions, and ensure the optimization of crop yields as well as efficient resource use. Considering these climatic factors along with the relevant soil parameters such as NPK levels and pH, our system thus guarantees both crop and fertilizer suggestions to be in harmony with local growing conditions and to ensure healthy, productive, and sustainable farming practices.



**Figure 11:** Comparison of rainfall, temperature and humidity for different crops

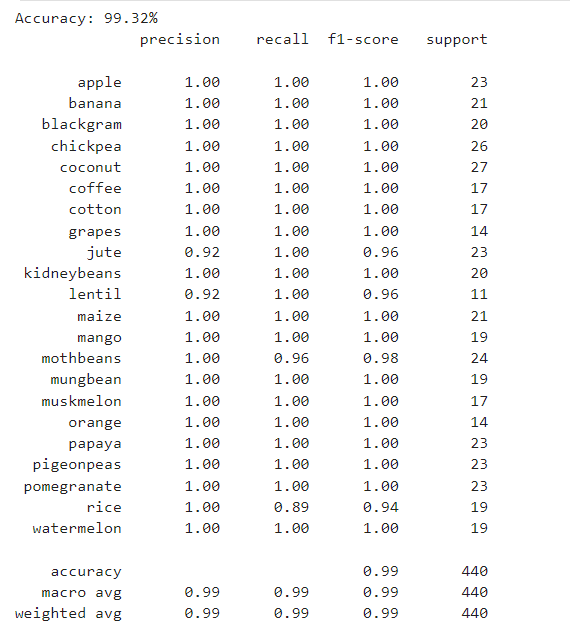
Overall findings reveal that integrated models of machine learning can work effectively and analyze soil attributes to give accurate crop and fertilizer recommendations under diverse soil conditions. The model Decision Tree has better interpretability because it gives detailed importance of each factor involved in decision-making, such as the parameters involved in the soil: nitrogen, phosphorus, potassium, temperature, and pH, in the selection of crops. Conversely, the Random Forest model builds upon predictive accuracy via the ensemble approach; that is, it aggregates multiple decision trees to capture the complex interactions between variables.

In combination, the approaches ensure that the crop and fertilizer recommendations are well-informed to enhance the ability of farmers for decision-making toward an optimization of yield and sustainability in agriculture. The models were rigorously evaluated under metrics such as accuracy, precision, recall, and F1 score, confirming reliability and effectiveness in giving quite accurate recommendations. Aligning local environmental and soil conditions with crop choice and fertilizer use, our system aims to support sustainable farming and increase productivity.

**Using Decision Tree Algorithm**

A screenshot of a computer code

Description automatically generated



**Figure 12:** Accuracy and other precision metrics calculation for Decision Tree Algoritm.

*4.2 Random Forest Model Excellence:*

*A screenshot of a computer program

Description automatically generated*

But what is really impressive in our crop and fertilizer recommendation system is the Random Forest model-an ensemble method using multiple Decision Trees. Figure 12 clearly indicates how much better it was compared to the simple Decision Tree model. At both evaluation points for both models, the Random Forest model always obtained higher performance metrics, which clearly suggests that it is robust enough to cope with variety in soil conditions and predict appropriate crops and fertilizers.

As since the Random Forest ensemble aggregates multiple decision paths, models are less likely to overfit, and sophisticated interactions among the soil parameters such as nitrogen, phosphorus, potassium, and pH levels are captured that will lead to more accurate predictions and reliable recommendations. Therefore, validation of the model since it effectively outperformed the Decision Tree model based on the results of our analysis, hence it comes to be used in our integrated recommendation system. This evidence of better performances highlights the model as reliable and effective in providing exact agro-insights and thus further drives its role as a valuable tool in enhancing decision-making in farming applications.

* 1. *Impact on Agricultural field*

The soil- and environment-based crop and fertilizer recommendation system is a giant leap for precision agriculture, especially for regions that are prone to soil degradation and unpredictable climatic conditions. It provides information to support soil attributes-specific recommendations on nutrient levels (N, P, K) pH, temperature, and rainfall patterns, thus carrying an approach based on data optimisation for choices of crops and fertilizer application.

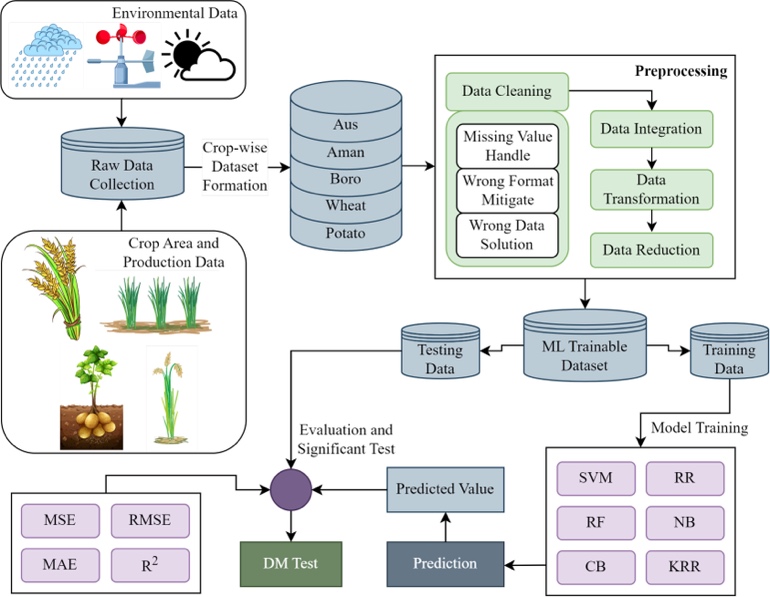
This project serves as a decision-support tool for farmers to make choices that are aligned with their soil's unique characteristics. The precise matching of crops with suitable soil types reduces the application of excess fertilizer application, thus leading to a more sustainable farm and less environmental impact in the long run. This has led to greater crop yields, resource efficiency, and soil health over long periods, hence answering some of the most profound agricultural questions of the modern world and opening up avenues for improved productivity and sustainability.

* 1. *Future Research Direction*

This project can further grow to guide not only what crops to grow and what kind of fertilizers to use but also which crops are currently profitable based on the present market data. With real-time data from IoT sensors and recent advanced machine learning algorithms, it may dynamically update the recommendations according to real-time factors such as the moisture of the soil, nutrient available, and seasonal patterns of weather. Further development of the model to include other zones and other soil health parameters, including organic matter and trace nutrients, will enhance its wider applicability across climates and types of soils.

In addition, this facility for trend analysis will help determine which crops have the highest demand for cultivation to enable the system to not only give the best crop for a specific set of conditions but also that which would produce higher incomes. This would give the recommendations of what crops to plant and related fertilizers to be used for optimal soil health in growing such crops. Using ensemble methods and maintaining interpretability using SHAP and LIME would ensure that insights were accurate and actionable. These developments will give rise to a strong, flexible and economically viable crop recommendation model which today addresses immediate requirements but is also responsive to the best projections of future climates, thereby increasing farmer profitability and sustainability.

*6 . Architecture Diagram Explanation*



This figure illustrates the overall process that we adopted for developing a machine learning-based crop prediction system. It begins with gathering raw data from various sources. We employed environmental data such as rainfall, temperature, and sunlight, as well as crop area and production data. This collective information enabled us to create crop-wise datasets for various crops like Aus, Aman, Boro, Wheat, and Potato.

Once the raw data was prepared, the second significant step was data preprocessing. We cleaned the data here by dealing with missing values, correcting incorrect formats, and rectifying any erroneous entries. We then combined various data sources, converted the data into an appropriate format for analysis, and filtered out unwanted features to make the dataset more efficient for training.

We prepared a machine learning trainable dataset after preprocessing, which we divided into training and test data. We trained machine learning models such as the Decision Tree Classifier and Random Forest Classifier using the training data. These models enabled us to learn patterns from the data and determine the correlation between soil/environmental attributes and crop suitability.

Having trained the models, we proceeded to the prediction stage with the test data. To assess the performance of the models, we employed a number of metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R²). These assisted us in determining the accuracy and reliability of the predictions.

Lastly, to compare the models and choose the best one, we employed the Diebold-Mariano (DM) Test, which is a statistical test for comparing predictive accuracy. This entire process assisted us in developing a dependable, data-driven system that can help farmers choose the appropriate crops depending on their soil and environmental conditions.

# Evaluation Metrics

A table of numbers with text

Description automatically generated with medium confidence

# CONCLUSION

This research offers to improve agricultural productivity by providing data-driven crop and fertilizer recommendations tailored to specific soil and environmental conditions. Therefore, the study tries to model the relationship of soil properties concerning crop suitability according to key nutrients like nitrogen (N), phosphorus (P), and potassium (K), and environmental factors like temperature, humidity, pH, and rainfall. EDA classified different varieties of crops based on their various nutrient requirements. For example, growth is highly nitrogen-dependent for rice and wheat; the legumes, however, which are capable of fixing atmospheric N, have relatively lower N requirements. Phosphorus and potassium requirements also varied but with distinct patterns suggesting exactly where each of the respective crops has specific dependencies on each.

The primary objective of this study is to train a machine learning algorithm that can be applied in proposing the appropriate crops and fertilizers corresponding to the given nutrient levels in the soil and surrounding conditions. This is a predictive approach to matching crop choices with ideal attributes regarding soil and climate, thus increasing agricultural productivity. In the testing phase, the scores for accuracy are provided to establish the reliability of the model. With this systematic approach developed in this work, crop choices can be optimized in a better manner, leading to increased yields and sustainable land use. Future improvements could include adaptation to regional soil types and dynamic seasonal changes so the actual recommendations would become even more precise for farmers.

##### References

Siddique Ibrahim S P; Bharathi Priya C; Selva Kumar S; Vinoth Kumar S, “Non Small Cell Lung Cancer Diagnosis Using kNN and Logistic Regression,” 2023 14th International Conference on Computing Communication and

Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-7, doi: 10.1109/ICCCNT56998.2023.10307547

SP, Siddique Ibrahim, et al. "Deep Learning and AI-Powered Natural Catastrophes Warning Systems." Utilizing AI and Machine Learning for Natural Disaster Management. IGI Global, 2024. 274-292.

Siddique Ibrahim S P; Dr.Sathya D, G. Swathi and P. P, “Artificial Intelligence Inspired Automatic Road condition Classification,” 2023 14th International Conference on Computing Communication and Networking Technologies

(ICCCNT), Delhi, India, 2023, pp. 1-6, doi: 10.1109/ICCCNT56998.2023.10306643.

Fayyaz Z, Ebrahimian M, Nawara D, Ibrahim A, Kashef R. Recommendation systems: Algorithms, challenges, metrics, and business opportunities. applied sciences. 2020 Nov 2;10(21):7748.

Bandara P, Weerasooriya T, Ruchirawya T, Nanayakkara W, Dimantha M, Pabasara M. Crop recommendation system. International Journal of Computer Applications. 2020;975:8887.

Priyadharshini A, Chakraborty S, Kumar A, Pooniwala OR. Intelligent crop recommendation system using machine learning. In2021 5th international conference on computing methodologies and communication (ICCMC) 2021 Apr 8 (pp. 843-848). IEEE.

Pudumalar S, Ramanujam E, Rajashree RH, Kavya C, Kiruthika T, Nisha J. Crop recommendation system for precision agriculture. In2016 eighth international conference on advanced computing (ICoAC) 2017 Jan 19 (pp. 32-36). IEEE.

Reddy DA, Dadore B, Watekar A. Crop recommendation system to maximize crop yield in ramtek region using machine learning. International Journal of Scientific Research in Science and Technology. 2019 Feb;6(1):485-9.

Kumar A, Sarkar S, Pradhan C. Recommendation system for crop identification and pest control technique in agriculture. In2019 International Conference on Communication and Signal Processing (ICCSP) 2019 Apr 4 (pp. 0185-0189). IEEE.

Shams MY, Gamel SA, Talaat FM. Enhancing crop recommendation systems with explainable artificial intelligence: a study on agricultural decision-making. Neural Computing and Applications. 2024 Apr;36(11):5695-714.

Senapaty MK, Ray A, Padhy N. IoT-enabled soil nutrient analysis and crop recommendation model for precision agriculture. Computers. 2023 Mar 12;12(3):61.