**Enhancing Crop Yield Prediction using Advanced Data Analysis Techniques & Hybrid Model Evaluation**

**Submitted for**

**DATA VISULIZATION AND DASHBOARD**

Submitted by:

**(E23MCAG0048) Arjun Kumar Sahu**

**(E23MCAG0060) Ayush Singh**

**(E23MCAG0082) Harsh Deep**

Submitted to-

**Dr. Nitin Arvind Shelke**

**July-Dec 2023**

**SCHOOL OF COMPUTER SCIENCE AND ENGINEERING**

A close-up of a logo

Description automatically generated

**INDEX**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | | Sr No | Content | Page No |  | | 1 | ABSTRACT | 5 |  | | 2 | INTRODUCTION | 5 |  | | 3 | RELATED WORKSS | 7 |  | | 4 | SOFFTWARE USES | 8 |  | | 5 | METHODOLOGY | 8 |  | | 6 | EXPERIMENTAL RESULTS | 8 |  | |  |  |  |  | |  |  |  |  | |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Title** | **Page No.** |
| **1** | Box Plot Of N | 10 |
| **2** | Box Plot Of P | **10** |
| **3** | Box plot of K | **11** |
| **4** | **Box Plot Of temperature** | **11** |
| **5** | **Box Plot Of humidity** | **12** |
| **6** | **Ph** | **12** |
| **7** | **Rainfall** | **13** |
| **8** | **Nitrogen content is higher in cotton, lowest in lentil** | **14** |
| **9** | **Higher phosphorus in apple and lower in watermelon** | **14** |
| **10** | **Higher potassium in grapes and lower in orange** | **15** |
| **11** | **Higher nitrogen in cotton and lowest in coconut** | **15** |
| **12** | **Crop require more potassium and nitrogen, highest potassium in apple and grapes and lowest in orange** | **16** |
| **13** | **Crop required more nitrogen and phosphorus and potassium, highest phosphorus in apple and pomegranate, rice and coconut** | **17** |
| **14** | **Correlation between features** | **19** |
| **15** | **Confusion matrix** | **21** |
| **16** | **Confusion matrix** | **22** |
| **17** | **Confusion matrix** | **23** |
| **18** | **Confusion matrix** | **25** |
| **19** | **Confusion matrix** | **25** |
| **20** | **Final reasult** |  |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **Table No.** | **Title** | **Page No.** |
| **1** | **Table 1** | 19 |
| **2** | **Table 2** | **20** |
| **3** | **Table 3** | **21** |
| **4** | **Table 4** | **22** |
| **5** | **Table 5** | **23** |
|  |  | **24** |
|  |  | **26** |

# ABSTRACT

Traditional farming practices are rapidly being replaced by data-driven approaches in the field of contemporary agriculture. The main goal of this project is to develop and implement a crop recommendation system using the extensive dataset provided by Atharva Ingle on Kaggle. The technology uses machine learning and data analysis to enable farmers to choose the best crops for their unique geographical and environmental conditions by providing educated choices. The main component of the project, the Atharva Ingle dataset, includes a wide range of important data, such as soil characteristics, climate change and past crop yields. Crop recommendation systems learn to recognize the complex patterns and connections that govern effective crop development by consuming this dataset. Predictive models are developed using machine learning techniques to predict crop performance by combining multiple input parameters. Real-time integration of soil and climate data, user-friendly interfaces that allow farmers to enter the characteristics of their farms, and a robust recommendation engine that can propose crops are some of the main aspects of a crop recommendation system. These suggestions are supported by historical data analysis and relationships between different crops and environmental factors. Furthermore, the approach considers economic feasibility, providing a holistic framework to decision makers.

# INTRODUCTION

A turning point is approaching in the history of agriculture, the foundation of human civilization. Global population growth, changing climate patterns and the need for sustainable practices have combined to necessitate creative solutions. The convergence of data science and agriculture offers a game-changing possibility for building systems that maximize productivity, resource allocation and crop selection. The development and implementation of a crop recommendation system, a technological marvel aimed at helping farmers make informed decisions about the crops they produce, is the focus of this project. With the use of Atharva Ingle dataset, a large collection of agricultural factors and sophisticated machine learning algorithms, this recommendation system can provide customized crop recommendations based on the current environmental conditions. When choosing crops, farmers in traditional rural settings often rely on generational knowledge and experience. But a more intensive and data-driven strategy is necessary due to the complex interplay between crop needs, meteorological conditions and soil properties. This research acknowledges the intrinsic potential of data analytics to unravel these complex relationships and gain accurate insights, improving agricultural sustainability and production. The Atharva Ingle dataset obtained from Kaggle provides an extensive number of data points including agricultural yield, temperature, humidity, precipitation, and soil properties. The project aims to feed this dataset into a crop recommendation system to uncover hidden patterns and correlations that are beyond the scope of traditional human analysis. Using this information, machine learning algorithms can be trained to understand the nuances of effective crop production under different conditions. The main goal of this project is to provide easy-to-use equipment to the farmers. Farmers can enter unique field information into the crop recommendation system interface to receive customized suggestions. This method uses predictive algorithms and historical data to generate recommendations that consider both crop suitability and economic feasibility. In the context of agriculture and crop production, three important nutrients nitrogen, phosphorus and potassium are together called NPK. Proper growth, development and general productivity of plants depend on these nutrients. When it comes to choosing the best crops, nitrogen, phosphorus, and potassium use will likely be important in your crop recommendation project. Here's why they're important:

**Nitrogen(N)**

Nitrogen is a fundamental nutrient that plants require for various processes including photosynthesis, protein synthesis and overall growth. It is a primary component of chlorophyll, the pigment responsible for capturing light energy during photosynthesis. Adequate nitrogen levels promote vigorous leaf growth, resulting in healthy plants with vibrant green leaves. Nitrogen deficiency can result in stunted growth, yellowing of leaves and reduced crop yield. However, excessive nitrogen application can lead to unbalanced growth and environmental issues such as water pollution.

Phosphorus(P)

Phosphorus is essential to aid in processes such as energy transfer within plants, root development, flower formation and seed production. It is an important component of ATP (adenosine triphosphate), which serves as the primary energy currency in cells. Phosphorus deficiency can result in poor root growth, delayed flowering, and reduced fruit set. Incorporating adequate phosphorus into the soil helps establish healthy plants and increases overall crop quality. It is important to balance phosphorus application, as excess phosphorus can have negative effects on water bodies and aquatic ecosystems.

Potassium(K)

Potassium is important for various physiological functions in plants, including water uptake, enzyme activation, and overall stress resistance. It helps regulate water balance within cells, boosts disease resistance, and improves the overall resiliency of plants to environmental stresses. Adequate potassium levels contribute to stronger stems, better root growth and increased crop yield. Potassium deficiency can lead to weakened plant structure, reduced disease resistance and reduced tolerance to extreme conditions.

# RELATED WORKS

The paper "Ensemble machine learning-based recommendation system for effective prediction of suitable agricultural crop cultivation" compares various machine learning models for predicting crop production. The proposed KRR model outperforms other models in terms of error metrics such as MSE, MAE, RMSE, and R2. The study emphasises the importance of increasing wheat production in South Asia and discusses the challenges and potential solutions for achieving this. The results suggest that the KRR model is more effective in predicting crop production compared to other models considered in the study.[1]

The paper "Multimodal Machine Learning Based Crop Recommendation and Yield Prediction Model" proposes a model for crop recommendation and yield prediction. It utilizes a multimodal machine learning approach that integrates various data sources such as soil, weather, and crop information. The model employs a novel optimization algorithm called KELM, which aims to balance exploitation and exploration in the search for the best solutions. The algorithm uses concentration vectors and an exponential term to achieve this balance. The paper provides detailed equations and explanations of the algorithm's components and parameters, aiming to improve crop yield prediction and recommendation accuracy.[2]

The paper discusses the use of different machine learning algorithms for predicting crop yield and the validation of prediction accuracy using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Squared Error (MSE). It explains that the selection of a particular algorithm is based on the nature of the application and the accuracy of the prediction algorithm. The paper also provides the computations for MSE, RMSE, and MAE, and how these metrics indicate the accuracy of the classifier. The proposed approach aims to provide insight into how different research works have employed various machine learning algorithms for crop yield prediction .[3]

The paper "Machine learning approach for forecasting crop yield based on climatic parameters" by Dr. CD Singh (2019) presents a machine learning-based method for predicting crop yields using climatic data. The study explores various machine learning techniques, such as regression, decision trees, and neural networks, to identify the most suitable model for predicting crop yields. The research highlights the importance of accurate crop yield predictions in improving agricultural productivity and sustainability.[4]

The paper discusses the use of entropy, Gini index, and information gain in the context of crop prediction using machine learning. It explains that entropy and Gini index are metrics for measuring the impurity of nodes in a decision tree, and information gain is a measurement of how much the reduction in entropy is achieved by a particular attribute. The paper also covers the Random Forest method, which consists of multiple decision tree classifiers to enhance the model's performance, and explains how ensemble learning is used to address overfitting. The paper provides formulas and criteria for calculating entropy, Gini index, and information gain, and discusses their role in building decision trees for crop prediction.[5]

The paper proposes a farmer-friendly system that predicts the best suitable crop for a particular land by considering parameters such as annual rainfall, temperature, humidity, and soil pH. The system displays the recommended crop, required seeds/acre, market price, and approximate yield of the recommended crop. The system takes NPK values in the input section to display the required NPK for the recommended crop. The system was tested for various datasets collected from different farmers for their land conditions. The proposed system helps farmers to take the right decision in selecting the crop for cultivation, which will reduce their losses and increase their income.[6]

The paper titled "Crop Selection Method to maximise crop yield rate using machine learning technique" by R. Kumar et al. proposes a method to maximise crop yield rate using machine learning. The authors suggest using a decision tree algorithm to select the best crop for a given area based on factors such as soil type, climate, and water availability. The proposed method was tested on a dataset of crops grown in India and showed promising results in terms of maximising crop yield rate.[7]

The paper "AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms" presents a system for recommending crops to farmers based on various factors such as soil type, climate, and historical crop data. The system uses machine learning algorithms to analyze the data and provide recommendations to the farmers. It also includes features such as rainfall prediction and crop suitability prediction. The system aims to help farmers maximize their crop yield rates and improve the agricultural economy. [8]

# SOFTWARE USED

Jupyter Notebook, Anaconda Navigator, Power BI

# METHODOLOGY

The above code uses an approach that involves several steps for crop production analysis and prediction based on various agricultural parameters. Data exploration and visualization using tools like Pandas, NumPy, Matplotlib, Seaborn, and Plotly form the first step. To learn more about the dataset, descriptive statistics, correlation analysis, and visualization methods such as box and bar graphs are used. To improve the data quality, outliers in the precipitation column are found and removed. For agricultural production forecasting, several machine learning models are used, such as decision trees, random forests, logistic regression, support vector machines (SVM), and an ensemble model that combines logistic regression and decision trees. Confusion matrix is used to calculate and display model performance indicators including precision, accuracy, recall, and F1-score. A bar plot representing a comparative study of the accuracy ratings of the models represents the findings of the research. Hybrid ensemble models, which combine decision trees with logistic regression, show promise as a method of accurately predicting crop production based on certain agricultural variables. Thus, this study provides a holistic framework for using machine learning in agriculture, providing decision-making and crop management insights to farmers.

# EXPERIMENTAL RESULTS

## Boxplot for Outlier Detection and Training Data Analysis

In the pursuit of understanding and enhancing the performance of machine learning models, data visualization plays a crucial role. Boxplots, also known as box-and-whisker plots, stand as powerful tools for analyzing data distribution, identifying outliers, and gaining insights into the central tendencies of the dataset. In this project, boxplots were employed to assess and address outliers within the dataset, particularly during the phases of checking and training.

## Boxplot Components

A box plot consists of several components that collectively provide a comprehensive view of the data distribution:

Box: The box in the centre of the plot represents the interquartile range (IQR), which encompasses the middle 50% of the data. The box is divided into the lower quartile (Q1) and the upper quartile (Q3), with the median marked by a line inside the box. Whiskers: The whiskers extend from the box to the minimum and maximum values within a specified range, often 1.5 times the IQR. Data points beyond the whiskers are considered potential outliers. Outliers: Outliers are data points that fall significantly outside the whiskers' range. They are represented as individual points on the plot and could be indications of anomalies or errors in the data.

## Outlier Detection

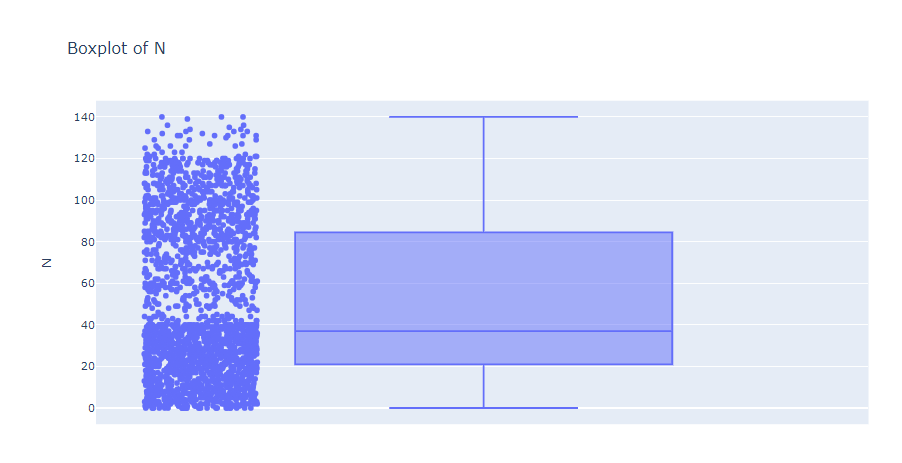
In the context of this project, boxplots were employed as a means of outlier detection. Outliers can significantly impact the performance of machine learning algorithms, leading to skewed models and erroneous predictions. By visualizing the distribution of data through boxplots, it becomes easier to identify these anomalies and assess their impact on the training process.

## Training Data Analysis

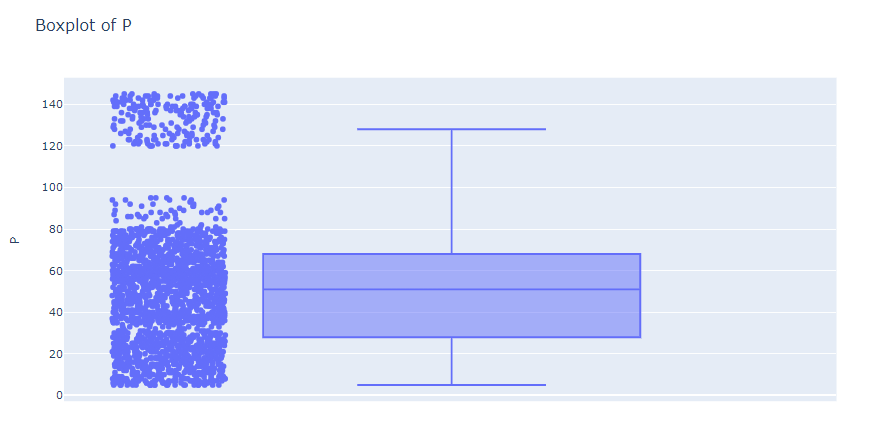
Before training machine learning models, it is essential to gain a holistic understanding of the dataset's characteristics. Boxplots aid in this analysis by revealing the distribution of features and their potential variations. By visualizing the spread of data within each feature through boxplots, one can determine whether the dataset exhibits skewness, the presence of outliers, and the overall distribution pattern.

## Interpreting results

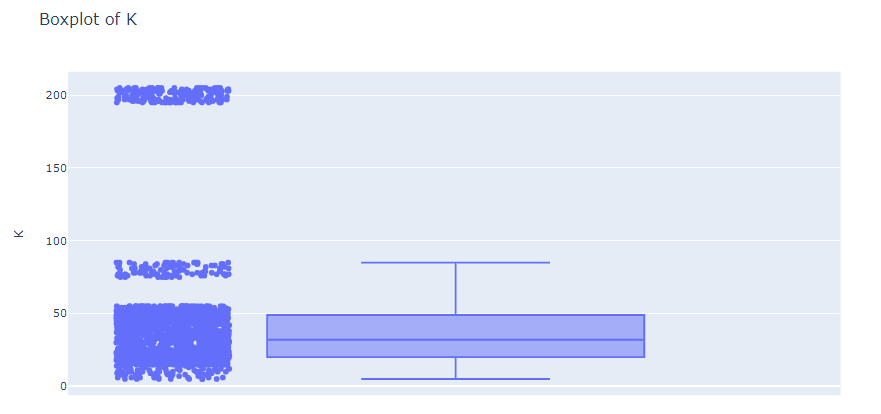
During the checking and training phases, the boxplots help in making informed decisions. If outliers are identified, it prompts a closer inspection to determine whether they are genuine data points or errors that need to be addressed. Outliers might necessitate data cleaning, transformation, or more advanced handling techniques. Additionally, analyzing boxplots for different features provides insights into the range of values and the relative magnitude of variations, which can guide feature engineering and model selection.



Figure



Figure



Figure

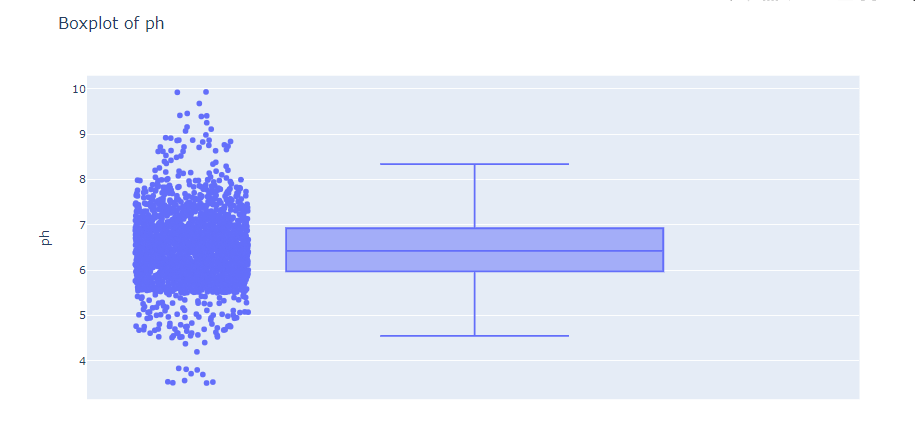
A blue and purple rectangular object

Description automatically generated with medium confidence

Figure



Figure



Figure



Figure

Outliers play a crucial role in data analysis for several reasons. Primarily, they are instrumental in identifying fraudulent or anomalous instances within a dataset. Outliers are values that significantly deviate from the mean of the dataset. Detecting these outliers is commonly achieved through methods such as boxplots, scatter plots, histograms, and the calculation of statistical measures like the Interquartile Range (IQR), Z-Score, and Standard Deviation. The IQR, representing the range between the first quartile (Q1) and the third quartile (Q3), serves as a robust metric for identifying outliers. Another method, the Z-Score, involves standardizing values based on the mean and standard deviation. Both IQR and Z-Score methodologies aid in pinpointing extreme values that may indicate irregularities or errors. Various techniques are employed to remove outliers from datasets. The IQR method involves eliminating data points falling outside a specified range, while the Z-Score method removes values beyond a certain standard deviation threshold. Additionally, outliers can be addressed through techniques such as the standard deviation method, scatter plot analysis, histogram evaluation, percentile assessment, and logarithmic or square root transformations. In summary, the identification and removal of outliers are critical steps in data preprocessing and analysis, contributing to the accuracy and reliability of statistical models and ensuring meaningful insights are derived from the dataset.

The bar plot in Fig -8 shows that the nitrogen content is highest in cotton and lowest in lentil.

A graph of different colored bars

Description automatically generated

Figure

The bar plot in Fig-9 shows that the phosphorus content is highest in apple and lowest in watermelon.

A graph of different colored bars

Description automatically generated

Figure

The bar plot in Fig-10 shows that the potassium content is highest in grapes and lowest in orange.

A graph of different colored bars

Description automatically generated

Figure

Fig-11 shows that the nitrogen content is highest in cotton and lowest in coconut.

A graph of different fruits

Description automatically generated with medium confidence

Figure

This shows that the crop which requires more nitrogen also requires more potassium and potassium content is highest in apple and grapes and lowest in orange.

A graph with text on it

Description automatically generated

Figure

This shows that the crops which require more nitrogen also require more phosphorous and potassium and phosphorus is highest in apple and pomegranate, rice and coconut.

A graph of a bar chart

Description automatically generated with medium confidence

Figure

## CORRELATION

Correlation analysis is a fundamental statistical technique used to measure the strength and direction of relationships between different variables. In the context of the crop recommendation project, correlation analysis involves exploring how various factors, such as soil properties, weather conditions, and nutrient levels, are interconnected and potentially influence crop growth and yield. The strength of correlation is quantified by correlation coefficients, which range from -1 to 1. A coefficient of 1 signifies a perfect positive correlation, while -1 indicates a perfect negative correlation. A value close to 0 suggests a weak or no linear correlation.

The direction of correlation, indicated by the sign of the coefficient, helps interpret relationships. Positive correlation implies that as one variable increases, the other tends to increase as well, while negative correlation suggests an inverse relationship. Common correlation coefficients include the Pearson coefficient for linear relationships, Spearman rank correlation for non-linear relationships, and Kendall Tau rank correlation. These coefficients provide numerical values that convey both the strength and direction of the relationship between variables.

Correlation matrices offer a comprehensive view of how each variable correlates with others. In the crop recommendation project, correlation analysis can unveil insights such as the interplay between different soil properties, the correlation of specific weather conditions with crop yields, and the presence of multicollinearity (high correlation) between features. Identifying these patterns is crucial for making informed decisions about feature selection, model training, and understanding the factors that significantly impact crop performance. By leveraging correlation analysis, the project aims to enhance predictive modelling and facilitate more effective crop management strategies.

A screenshot of a graph

Description automatically generated

Figure

1. MODELS USED IN PROJECT

In the agricultural innovation landscape, various machine learning algorithms serve as invaluable tools for unravelling intricate patterns within agricultural data. This project encompasses a diverse array of these algorithms, including Logistic Regression, Random Forest, Decision Tree, Hybrid Models, and Support Vector Machine (SVM), with the collective goal of constructing a robust Crop Recommendation System that empowers farmers with informed and optimal crop suggestions.

Logistic Regression:

Despite its apparent simplicity, Logistic Regression proves to be a foundational algorithm for classification tasks. Tailored for scenarios where the relationship between input variables and outcomes is linear, Logistic Regression in the context of crop recommendation aids in making binary decisions. For instance, it can assess the suitability of a specific crop based on prevailing conditions.

Random Forest:

Renowned for its robustness in handling complexity and mitigating overfitting, Random Forest stands as an ensemble algorithm comprising multiple decision trees. Each tree, trained on different data subsets, contributes to more accurate and stable predictions. In the Crop Recommendation System, Random Forest excels in capturing nonlinear relationships between factors like soil type, weather conditions, and historical data, facilitating precise crop suggestions.

Decision Tree:

Offering transparency in decision-making based on input features, Decision Trees recursively partition data into subsets by selecting optimal features at each node. Their interpretability makes Decision Trees an effective baseline model for the project. By employing Decision Trees, the project aims to visually represent the crop selection decision process while considering various factors.

Hybrid Models (Logistic Regression and Decision Tree):

Hybrid Models leverage the strengths of different algorithms to create synergistic solutions. In this project, the combination of Logistic Regression and Decision Tree Classifier aims to concurrently harness linear and nonlinear relationships. This nuanced approach to crop recommendation considers the simplicity of Logistic Regression and the complexity of Decision Trees.

Support Vector Machine (SVM):

Known for their efficacy in classifying data through optimal hyperplane identification, Support Vector Machines excel in handling datasets with intricate decision boundaries. In crop recommendation, SVMs shine in mapping input variables to higher-dimensional spaces, capturing subtle nuances in the data for accurate crop predictions.

In summary, the amalgamation of these machine learning algorithms provides a multi-faceted perspective on crop recommendation. By synergistically leveraging the strengths of Logistic Regression, Random Forest, Decision Tree, Hybrid Models, and Support Vector Machine, the Crop Recommendation System aims to deliver comprehensive, accurate, and actionable insights, revolutionizing modern agriculture through the power of data science.

1. RESULTS
2. Decision TreeClassifier

Accuracy: decision tree model accuracy score: 0.9905

|  |
| --- |
| precision recall f1-score support |
|  |
| apple 1.00 1.00 1.00 29 |
| banana 1.00 1.00 1.00 37 |
| blackgram 1.00 0.97 0.99 34 |
| chickpea 1.00 1.00 1.00 30 |
| coconut 1.00 1.00 1.00 23 |
| coffee 1.00 1.00 1.00 33 |
| cotton 0.96 1.00 0.98 23 |
| grapes 1.00 1.00 1.00 30 |
| jute 0.94 0.94 0.94 31 |
| kidneybeans 1.00 1.00 1.00 33 |
| lentil 1.00 1.00 1.00 29 |
| maize 0.96 0.96 0.96 28 |
| mango 1.00 1.00 1.00 23 |
| mothbeans 1.00 1.00 1.00 38 |
| mungbean 1.00 1.00 1.00 30 |
| muskmelon 1.00 1.00 1.00 32 |
| orange 1.00 1.00 1.00 30 |
| papaya 1.00 1.00 1.00 23 |
| pigeonpeas 1.00 1.00 1.00 25 |
| pomegranate 1.00 1.00 1.00 29 |
| rice 0.86 0.86 0.86 14 |
| watermelon 1.00 1.00 1.00 27 |
|  |
| accuracy 0.99 631 |
| macro avg 0.99 0.99 0.99 631 |
| weighted avg 0.99 0.99 0.99 631 |

Confusion Matrix:

A blue grid with yellow and orange numbers

Description automatically generated

Figure

1. Logistic Regression

Accuracy: Logistic Regression Model accuracy score: 0.9445

|  |
| --- |
| precision recall f1-score support |
|  |
| apple 1.00 1.00 1.00 29 |
| banana 1.00 1.00 1.00 37 |
| blackgram 0.87 0.79 0.83 34 |
| chickpea 1.00 1.00 1.00 30 |
| coconut 0.92 1.00 0.96 23 |
| coffee 1.00 1.00 1.00 33 |
| cotton 0.78 0.91 0.84 23 |
| grapes 1.00 1.00 1.00 30 |
| jute 0.81 0.94 0.87 31 |
| kidneybeans 1.00 1.00 1.00 33 |
| lentil 0.88 1.00 0.94 29 |
| maize 0.81 0.79 0.80 28 |
| mango 1.00 1.00 1.00 23 |
| mothbeans 0.91 0.76 0.83 38 |
| mungbean 0.97 1.00 0.98 30 |
| muskmelon 1.00 1.00 1.00 32 |
| orange 1.00 1.00 1.00 30 |
| papaya 0.95 0.91 0.93 23 |
| pigeonpeas 0.96 1.00 0.98 25 |
| pomegranate 1.00 1.00 1.00 29 |
| rice 0.88 0.50 0.64 14 |
| watermelon 1.00 1.00 1.00 27 |
|  |
| accuracy 0.94 631 |
| macro avg 0.94 0.94 0.94 631 |
| weighted avg 0.95 0.94 0.94 631 |
|  |

A blue and yellow graph with numbers

Description automatically generated

Figure

1. Random Forest Classifier:

Random Forest Model accuracy score: 0.9857

|  |
| --- |
| precision recall f1-score support |
|  |
| apple 1.00 1.00 1.00 29 |
| banana 1.00 1.00 1.00 37 |
| blackgram 1.00 0.97 0.99 34 |
| chickpea 1.00 1.00 1.00 30 |
| coconut 1.00 1.00 1.00 23 |
| coffee 1.00 1.00 1.00 33 |
| cotton 0.96 1.00 0.98 23 |
| grapes 1.00 1.00 1.00 30 |
| jute 0.82 1.00 0.90 31 |
| kidneybeans 1.00 1.00 1.00 33 |
| lentil 1.00 1.00 1.00 29 |
| maize 0.96 0.96 0.96 28 |
| mango 1.00 1.00 1.00 23 |
| mothbeans 1.00 1.00 1.00 38 |
| mungbean 1.00 1.00 1.00 30 |
| muskmelon 1.00 1.00 1.00 32 |
| orange 1.00 1.00 1.00 30 |
| papaya 1.00 1.00 1.00 23 |
| pigeonpeas 1.00 1.00 1.00 25 |
| pomegranate 1.00 1.00 1.00 29 |
| rice 1.00 0.50 0.67 14 |
| watermelon 1.00 1.00 1.00 27 |
|  |
| accuracy 0.99 631 |
| macro avg 0.99 0.97 0.98 631 |
| weighted avg 0.99 0.99 0.98 631 |

A graph of numbers and a number

Description automatically generated with medium confidence

Figure

1. Support Vector Machine

SVM Model accuracy score: 0.9842

|  |
| --- |
| precision recall f1-score support |
|  |
| apple 1.00 1.00 1.00 29 |
| banana 1.00 1.00 1.00 37 |
| blackgram 1.00 1.00 1.00 34 |
| chickpea 1.00 1.00 1.00 30 |
| coconut 1.00 1.00 1.00 23 |
| coffee 1.00 0.97 0.98 33 |
| cotton 1.00 1.00 1.00 23 |
| grapes 1.00 1.00 1.00 30 |
| jute 0.81 0.94 0.87 31 |
| kidneybeans 1.00 1.00 1.00 33 |
| lentil 0.97 1.00 0.98 29 |
| maize 1.00 1.00 1.00 28 |
| mango 1.00 1.00 1.00 23 |
| mothbeans 1.00 0.97 0.99 38 |
| mungbean 1.00 1.00 1.00 30 |
| muskmelon 1.00 1.00 1.00 32 |
| orange 1.00 1.00 1.00 30 |
| papaya 0.96 1.00 0.98 23 |
| pigeonpeas 1.00 1.00 1.00 25 |
| pomegranate 1.00 1.00 1.00 29 |
| rice 0.89 0.57 0.70 14 |
| watermelon 1.00 1.00 1.00 27 |
|  |
| accuracy 0.98 631 |
| macro avg 0.98 0.98 0.98 631 |
| weighted avg 0.98 0.98 0.98 631 |

A graph of numbers and symbols

Description automatically generated with medium confidence

Figure

1. Hybrid Model (Logistic Regression + Decision Tree Classifier):

Accuracy score of ensemble model is: 0.9587955625990491

|  |
| --- |
| precision recall f1-score support |
|  |
| apple 1.00 1.00 1.00 29 |
| banana 1.00 1.00 1.00 37 |
| blackgram 0.86 0.94 0.90 34 |
| chickpea 1.00 1.00 1.00 30 |
| coconut 0.92 1.00 0.96 23 |
| coffee 1.00 1.00 1.00 33 |
| cotton 0.79 1.00 0.88 23 |
| grapes 1.00 1.00 1.00 30 |
| jute 0.81 0.97 0.88 31 |
| kidneybeans 1.00 1.00 1.00 33 |
| lentil 0.91 1.00 0.95 29 |
| maize 0.92 0.79 0.85 28 |
| mango 1.00 1.00 1.00 23 |
| mothbeans 1.00 0.79 0.88 38 |
| mungbean 1.00 1.00 1.00 30 |
| muskmelon 1.00 1.00 1.00 32 |
| orange 1.00 1.00 1.00 30 |
| papaya 1.00 0.91 0.95 23 |
| pigeonpeas 1.00 1.00 1.00 25 |
| pomegranate 1.00 1.00 1.00 29 |
| rice 0.88 0.50 0.64 14 |
| watermelon 1.00 1.00 1.00 27 |
|  |
| accuracy 0.96 631 |
| macro avg 0.96 0.95 0.95 631 |
| weighted avg 0.96 0.96 0.96 631 |

Confusion Matrix:

A blue grid with yellow and orange numbers

Description automatically generated

Figure

1. Accuracy of all models :

A chart of different colored rectangular shapes

Description automatically generated

Figure

1. Final Result :

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | N | P | K | temperature | humidity | ph | rainfall |
| 1203 | 36 | 125 | 196 | 37.465668 | 80.659687 | 6.155261 | 66.838723 |

1203 grapes

Name: label, dtype: object