AgriRec: Decision Tree-Based Agricultural Crop Recommendation with Web Platform Integration

Basant Rauniyar
UG, CE-SE
FET, Jain
(Deemed-to-be-University)
Bangalore-562112
basantrauniyar15@gmail.com

Prem Rauniyar
UG, CE-SE
FET, Jain
(Deemed-to-be-University)
Bangalore-562112
premrauniyar2019@gmail.com

Man Bahadur Thapa
UG, CE-SE
FET, Jain
(Deemed-to-be-University)
Bangalore-562112
thaparitik45@gmail.com

Karthikeyan S.
Assistant Professor
Department of CE-SE
Bangalore-562112
link2karthikcse@gmail.com

***Abstract*—In the pursuit of sustainable and enhanced agriculture, with growing food demands and environmental challenges, precision agriculture has become a pivotal innovation to optimize crop production. This research presents AgriRec, an AI-powered web platform designed to deliver precise crop cultivation using machine learning. The following study is based on the Crop Recommendation dataset, which comprises 2,200 records with eight agronomic factors consisting of the soil nutrients and environmental factors such as Nitrogen (N), Phosphorus (P), and Potassium (K) along with pH of soil, temperature, humidity and rainfall. The dataset also has a label column, providing the different crop types. This study uses data preprocessing techniques like filling missing values and removing duplicates to ensure consistency. Different machine learning models including Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree and Logistic Regression, are utilized and evaluated using performance metrics such as accuracy, recall, precision and F1-Score. Among these four, Decision Tree model achieved the highest accuracy of 98.18%. Thus, the optimal model, Decision Tree is deployed in a web-interface using HTML, CSS and JavaScript facilitating user engagement to get real-time inputs and predictions. The following study showcases the considerable capability of machine learning in revolutionizing traditional farming practice, offering farmers with a practical tool to improve their crop yield and decision-making, promoting sustainable farming methods to mitigate food needs.**

Keywords—AgriRec, LabelEncoder, AI-Driven Crop Recommendation, Web-based Decision Support, Agronomic Data Analysis, Precision Agriculture, Data-Enhanced Crop Yield, Sustainable Farming, Optimal Crop Selection

I. INTRODUCTION

The agriculture now also faces a lot of problems and challenges in meeting the demands of the food production although everything here is being advanced the agriculture now is not being advanced [1]. Farmers are using the traditional ways of farming and face a lot of problems like climate change and land requirements. Precise agriculture has emerged a game changing strategy, levering cutting-edge technology to maximize and enhance the crop yield and optimize the resources.

The inclusion of machine learning into agricultural practices provides a vast opportunity to evaluate and examine the large datasets and provide valuable trends and patterns, which can allow the farmers to make well-planned decisions for the specific soil and environmental characteristics [2].

AgriRec, an AI-powered web platform created to offer the best crop suggestions depending on certain agronomic circumstances, is presented in this research. The objective of this paper is to create a robust and reliable model that will help the farmers choose the best crops for the field and the crops which can be grown in that field using target datasets with 2,200 entities and eight essential attributes of the agronomic factors of the soil and environmental characteristics of a specific location.

Choosing the optimal crops for the soil is very difficult for farmers and requires an expert [3]. To overcome this situation, the plan is to close the gap by developing an intuitive platform using machine learning algorithms that will evaluate important soil parameters and offer accurate suggestions to the farmers [4]. By this, the accuracy of the crops also increases, eventually raising agricultural output.

This study makes the model performance more accurate by using data preparation, analytical analysis, and the use of different ML algorithms, such as Support vector machine (SVM), decision tree, K-nearest neighbors (KNN), and logistic regression. This will help farmers enhance crop cultivation techniques; the finding is used to expand the knowledge in agriculture and know the exact crops for the cultivation and increasing the productivity.

II. LITERATURE REVIEW

The use of machine learning has become more popular nowadays in the field of precision agriculture. Many study and research has also shown that these ML techniques can have a greater impact on improving the current agricultural methods and bring solutions to the problems created majorly by the resource constraints and ongoing climate change.

A. *Machine Learning in Agriculture*

The advancement of precision agriculture heavily relies on the use of machine learning. Much research has been done on the use of ML in variety of field including the irrigation management, pest control, and precision farming [5][6]. The effective and accurate crop recommendations for higher yield can be made possible with the help of supervised learning approaches, which can predict based on past data [7]. The higher accuracy and scalability of model is necessary to estimate the agricultural yields using ML techniques [8].

B. *Crop Recommendation Systems*

The importance of the precision agriculture with the help of crop recommendation systems are essential for farmers to choose the best crops for them based on their local soil and environmental conditions. These recommendation systems used to depend traditionally on human analysis and expert knowledge, but the recent advancements in data science have made it easier, accurate and effective [9]. The reliability of ML techniques, including Decision Trees, KNN, SVM, etc. has helped a lot in the journey of precision agriculture [10]. The Decision Tree model can deal with both data including complete and incomplete data to predict and classify agricultural data [11]. Similarly, the use of hybrid model performed better with less agricultural data, thereby making the hybrid model an optimal and favourable choice for a practical use, compared to the individual models as less and smaller datasets on agriculture are available [12].

C. *Data-Driven Decision Support Systems*

The precision agriculture advances with the adoption of web-based decision support systems, allowing the farmers to get quick advice for farming and forecasting tools [13]. The use of ML can be a powerful ally to the farmers in rural areas to make well-informed decisions to predict crop prediction based on the soil and weather patterns [14]. Likewise, the support advisory platform as suggested by S. Babu based on the soil conditions and dynamic weather patterns and trends also proved to be a successful recommendation tool for small and marginal farmers [15].

D. *Challenges and opportunities in Precision Agriculture*

Although the use of the ML-driven crop recommendation systems has made a significant progress, several challenges remain. The major challenge in the performance and reliability of the model is the lack of high-quality and region-specific data [16]. The scope of Big Data applications in Smart Farming goes beyond primary production; it is influencing the entire food supply chain. Big data are being used to provide predictive insights in farming operations, drive real-time operational decisions, and redesign business processes for game-changing business models [17]. The other challenge is the cost of running the complex machine learning algorithms for the massive farms in the real-time, making it a highly expensive computational system [18]. Although facing the obstacles, precision agriculture offers a lot of creative and innovative solutions and options to overcome these challenges. With the inclusion of real-time weather data, IoT devices and sensors with the satellite data, the crop recommendation systems can be improved significantly [19].

As the machine learning algorithms gets more advanced and sophisticated, they can optimize and predict the long-term agricultural trends and patterns while optimizing the resource allocation with best crop selection.

III. PROPOSED METHODOLOGY

AgriRec aims to create a platform powered by AI and Machine learning techniques that can effectively and efficiently predict the optimal crop suited for the location based on its soil and environmental characteristics. The Fig. 1 illustrates the system architecture and working of the AgriRec system to predict the optimal crop recommendation based on agronomic factors.

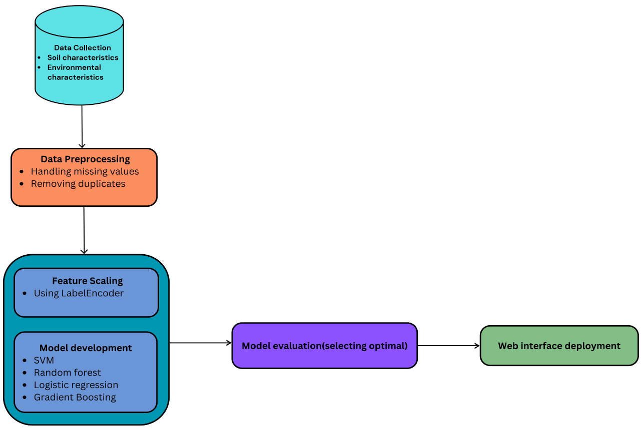

Fig. 1: System architecture of AgriRec

A. *Data Collection*

The AgriRec system utilizes the "Crop Recommendation Dataset" [20], which consists of 2200 entries with eight attributes including soil factors like pH of the soil, Nitrogen (N), Potassium (K) and Phosphorus (P); and environmental factors like temperature, humidity and rainfall. The label on these entries indicates the type of crop. Each crop class consists of 100 rows making up the equal proportion in the dataset. The balanced distribution of the crop class in the dataset allows the model to predict accurately.

B. *Data Preprocessing*

To ensure the quality and usability of the dataset, following preprocessing methods are utilized:

- Handling Null Values: The absence of the values in the dataset can negatively impact the quality of the dataset. Thus, to counter the missing or null values, the data with will null values are firstly identified and filled using mean imputation for numerical features.
- Removing Duplicates: To maintain and manage the integrity and reliability of the dataset, the duplicate entries are erased. It prevents the replicated and duplicated information of the dataset to disrupt the performance of the model.
- Label Encoding: The label column in the dataset represents the various crops. Each crop is given a unique numerical value. The numerical encoding is done with the help of Label Encoding using the LabelEncoder.

- Normalizing the Data: The numerical features such as Nitrogen (N), pH of soil, Phosphorus (P), temperature, Potassium (K), etc. are normalised making sure that the improper proportion of the feature doesn't rule down the quality of the model due to scale variations.

C. Data Splitting

The dataset is divided into two sets: training and testing sets using ratio of 80:20 split. The ML models are trained using the training data (80%) and testing data (20%) is used to measure the performance of the model.

D. Model Development

The study utilizes four ML algorithms that were applied to the "Crop Recommendation Dataset" to determine the optimal crop based on the soil and environmental conditions. The performance of each model was evaluated based on performance metrics like accuracy, precision, etc.

- Logistic Regression:

The multiclass categorization is used in Logistic Regression model to predict the optimal crop recommendation based on the agronomic conditions. The probability of each crop class is determined with the help of Sigmoid Activation Function.

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

$where,$
$h_\theta(x)$ = The probability of each crop class
θ = Model parameters (weight)
x = Parameters for input like pH, rainfall, Nitrogen, etc.

Logistic Regression is a straightforward approach. It is effective in predicting estimates across multiple categories. Thus, it is a reliable model to predict the optimal crop based on the agronomic factors.

- Decision Tree:

Decision Tree model also uses multiclass categorization to predict the optimal crop based on the agronomic characteristics of the dataset like Logistic Regression as it can work on both numerical and symbolic data. Using Gini impurity criteria, the purity of the class is maximized to ensure that the dataset is split properly.

$$Gini(D) = 1 - \sum_{i=1}^{m} p_i^2$$

$where,$
D = Dataset.
P_i= Samples in class III of dataset.

Decision Tree can capture the complex and non-linear relationships between the categories of crops and the agronomic features. The ability to handle complex correlations allows it to be a reliable model for optimal crop recommendation system.

- K-Nearest Neighbors (KNN):

KNN model utilizes the proximity or closeness of the data points with the comparable agronomic factors in the dataset to categorize and predict the optimal crop selection suited for the location. It is a non-parametric model where samples are categorized by providing them with label like their neighbors.

$$\hat{y} = majority\ vote\ of\ K\ nearest\ neighbors$$

$where,$
ŷ = Predicted crop class type.
K = The number of neighbors considered while performing a majority vote.

KNN is a effective and efficient model to identify the local patterns and trends in data to look after the variability of agronomic data in the dataset improving the reliability and robustness of the recommendation system.

- Support Vector Machine (SVM):

With the use of Radial Basis Function (RBF) kernel, the Support Vector Machine (SVM) creates a decision boundary to optimize the margin between the crop classes ensuring the model is reliable and robust to predict the optimal crop suited based on the agronomic characteristics.

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b$$

$where,$
α_i = The Lagrange multipliers.
y_i = Label for each crop class
$K(x_i, x)$ = Kernel function (non-linear relationships are captured using the RBF Kernel)
b = The Bias term.

The use of RBF kernel allows the SVM model to handle complex, non-linear patterns within the dataset, making the model ideal for identifying the minute influences of agronomic features on crop suitability. The SVM model with the RBF kernel makes it the optimal ML model for best crop recommendations.

IV. RESULTS

A. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) was conducted to identify and understand the patterns and relationships between the agronomic characteristics. It majorly included Correlation Analysis and Distribution Analysis.

- Correlation Analysis:

The following Fig. 2 is the correlation matrix of the attributes in the dataset. The following graph suggest the strong correlation between Potassium (K) and Phosphorus (P), indicating that the higher potassium levels usually correspond to the higher phosphorus levels. The relationship between them suggests a potential synergistic effect that might influence the crop growth and yield in the nutrient-rich environment and soil. With the study of the correlation matrix, we can better understand the relationship and correlation of the different agronomic factors of the dataset which enables to identify the agronomic trends and patterns leading to a better understanding of the different soil and environmental factors role in determining the optimal crop selection for a specific location and develop a robust model which can look after all these criteria.

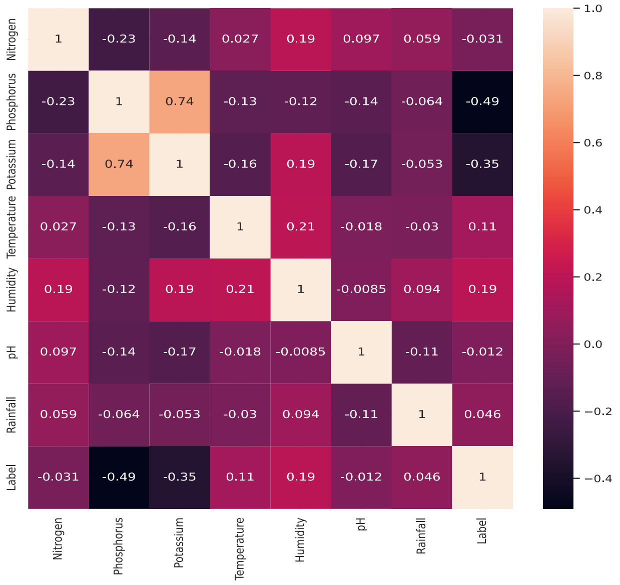

Fig. 2: Correlation matrix

- Distribution Analysis:

In this analysis, every feature of the dataset is analysed and examined to detect any outliers and skewness, which can have negative impact on the performance of the model.

a. Soil Nutrient Analysis (N, P, K): The Fig. 3 illustrates the distribution of the soil compositions, namely Nitrogen (N), Potassium (K) and Phosphorus (P) within the dataset. The graph demonstrates that the distribution of these agronomic factors is skewed towards the lower values showcasing the variability of the data points. This skewness suggests that most samples contain moderate to low nutrient concentrations, which can have impact on the crop growth and yield. Visualizing these distributions allows to identify the soil compositions and adjust the recommendations for the nutrient-deficient soils, enabling for more tailored fertilization techniques and make the optimal crop selection depending on the soil composition.

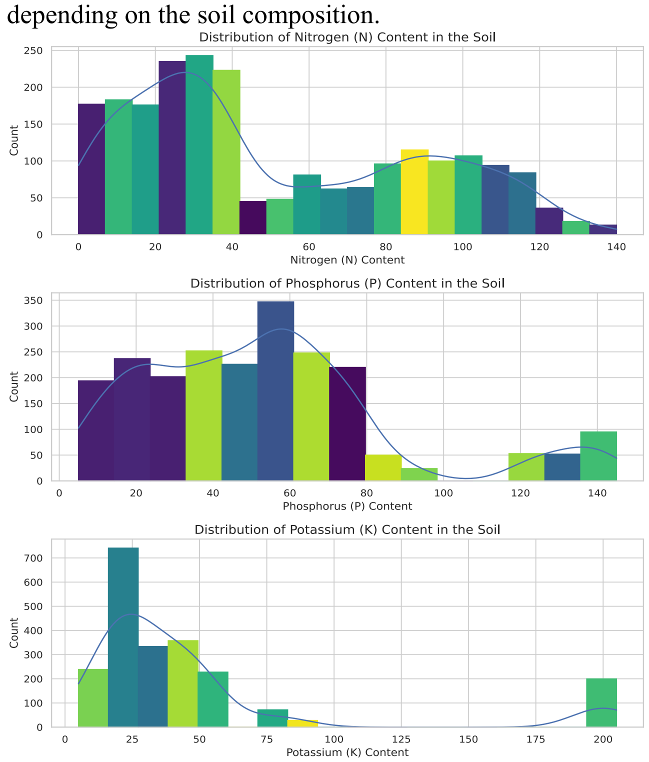

Fig. 3: Soil Nutrient Analysis (N, P, K)

b. Environmental Factors Analysis (Temperature, Humidity and Rainfall): The study on the distribution of the environmental factors like temperature, humidity and rainfall provides insights on the regional environmental conditions which can heavily impact the crop growth and yield. The Fig. 4 suggests that the temperature and humidity display the ideal distribution patterns optimal for crop growth, while rainfall is seen to be showing the seasonal peaks. These factors are key parameters crucial for predicting the crop viability, as different crops need specific environment to grow. By examining these patterns, it is ensured that the recommendations made align with the local climatic conditions, allowing the farmers to make well-informed decisions.

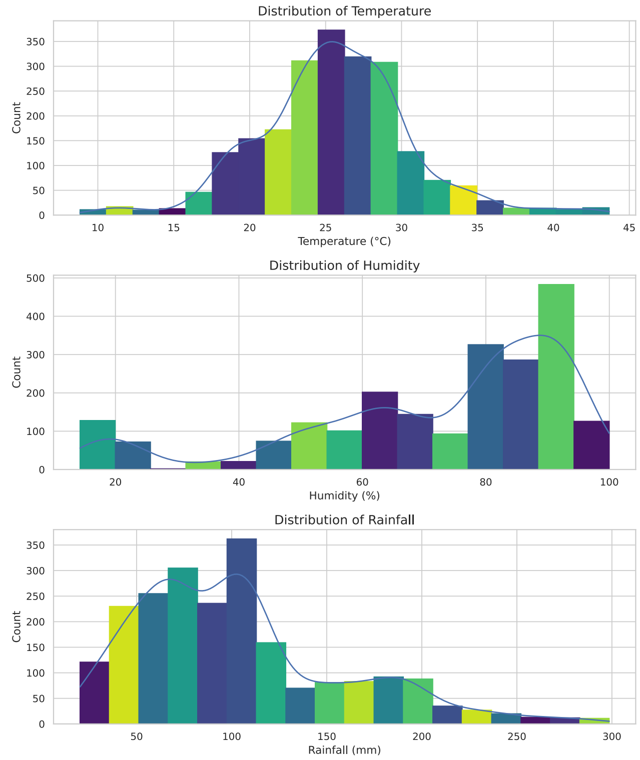

Fig. 4: Environmental Factor Analysis (Temperature, Humidity, Rainfall)

B. *Model Evaluation*

The development of the ML models for the optimal crop recommendation provided the significant insights and the capabilities of them. The Decision Tree model achieved the highest accuracy of astonishing 98.18%. Furthermore, these models were evaluated on following basis:

- Accuracy:

The models utilized were compared based on their accuracy ratings. The Fig. 5 illustrates that the Decision Tree achieved the accuracy of 98.18% outperforming other three models showing it's robustness and reliability in accessing the patterns in agronomic data. KNN achieved the accuracy score of 97.05%, followed by SVM with 96.14% and Logistic Regression with 95.23%. Thus, with the comparison of the accuracy of these four models, Decision Tree can be considered an optimal choice for crop recommendation system because of its ability to handle multiple variability in the dataset with the complex and non-linear relationships among the agronomic characteristics.

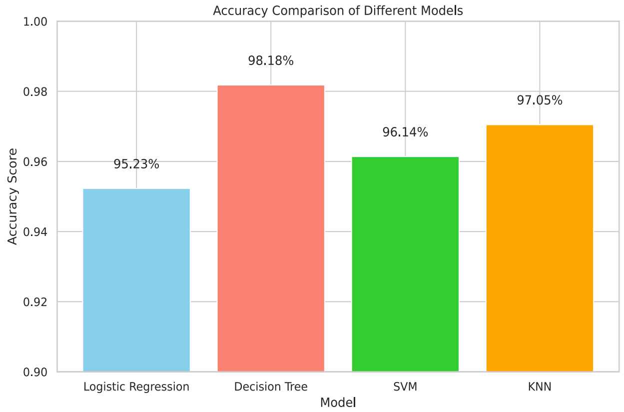

Fig. 5: Accuracy Comparison

- Performance Metrics:

The following Fig. 6 illustrates the brief performance comparison of the four models in metrics such as Accuracy, Recall, Precision and F1-Score. In the graph below, we can clearly see that the Decision Tree model outclassed all other three models in term of all these metrics, thereby stating its resilience and robustness in predicting the optimal crop types. Similarly, KNN also demonstrated excellent precision but has lower scores in other metrics showcasing it's limitations. SVM and Logistic Regression also shared similar patterns with even lower scores. The overall performance evaluation of these models shows the effectiveness and efficiency of the Decision Tree model in providing with reliable and robust crop recommendation system under diverse and variable agronomic characteristics.

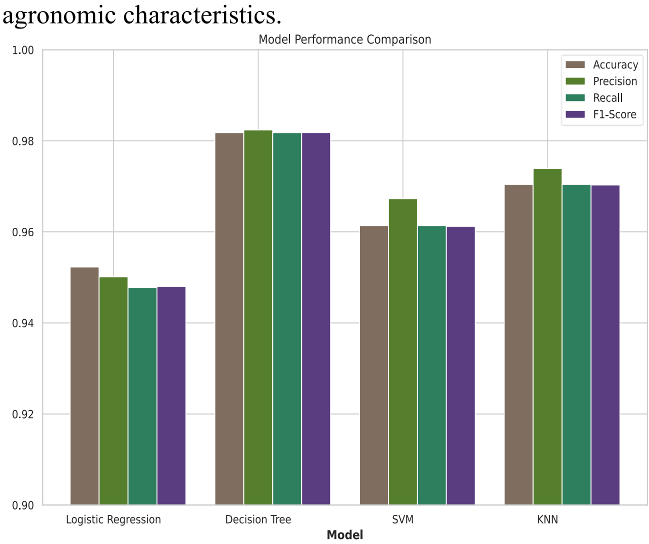

Fig. 6: Performance Metrics Comparison

- Confusion Matrix

To further access and evaluate the outperforming Decision Tree model, a confusion matrix was used. It provides a greater and deeper look into the accuracy of the Decision Tree model in classifying each type of crop category. The Fig. 7 illustrates the strong confusion matrix of the Decision Tree model further boosting its reliability and robustness. The confusion matrix consists of cells representing true positives, false positives, false negatives and true negatives. The misclassifications were almost negligible, and most predictions fell within the correct boundaries. The excellent accuracy and high true positive rates across all crop classes in dataset of the Decision Tree model as visualized by its confusion matrix demonstrates its applicability practical application of crop recommendation system.

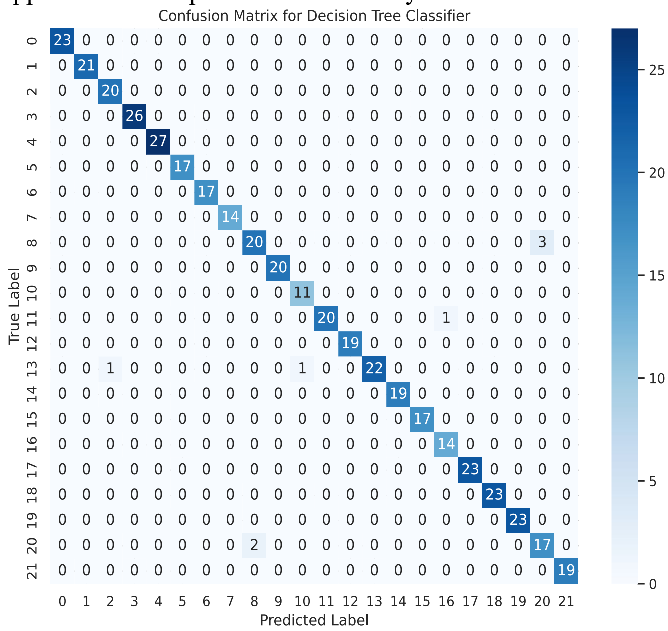

Fig. 7: Confusion Matrix of Decision Tree Classifier

V. WEB APPLICATION DEVELOPMENT

The web application is developed to seamlessly integrate the crop recommendation system. The implementation of this system can help farmers to optimize their crop production, reduce costs, and increase efficiency in agricultural activities [21]. A user-friendly interface is created, enabling farmers and those responsible for the agriculture industry to receive crop predictions according to their lands and their data. The website recommends the best crops for the given inputs. The following technologies are used in the platform:

A. *Frontend*

The web interface design is created using HTML, CSS, and JavaScript, which helps the user with responsive websites and good user experience. The website has the following algorithm inputs, such as nitrogen (N), phosphorus (P), potassium (K), pH of the soil, temperature, humidity, rainfall, etc. It provides an easily accessible website for the farmers to get well-informed decisions to have optimal crop selection with higher yield.

B. *Backend*

The backend of the application uses Python, where the flask is used to build the server that manages the machine learning model and its predictions with the help of the ML model packed into a pickle file. When the user submits the inputs in frontend, the backend comes into play and processes the data and passes through the trained Decision tree model which predicts the suitable crop. The server-side logic is seen by the flask, processing the inputs with the ML model.

C. *Model Integration*

The trained model, the Decision tree, is saved using Pickle, and is integrated into the Flask server. The model developed processed all the inputs and generated the prediction on the optimal crop. This helps to recommend the crop in real-time user inputs.

D. User Feedback

When the user submits the given inputs, the web application provides tailored feedback, including the best crops for the given inputs.

VI. CONCLUSION

This study successfully developed an AI-powered web platform that recommends the crop suitable for the land. By using Machine learning techniques and calculations from the dataset, the crops are predicted accurately and the factors that are included in the dataset are soil nutrients, pH, temperature, humidity, and rainfall. The accuracy of our model i.e. Decision Tree Model is about 98.18%.

The web application that is created provides a seamless interface that will help farmers with easy access, and it enables them to input their data and get the crop recommendation easily. The tool not only helps farmers in decision-making but also helps them grow sustainable farming by optimizing the resources and enhancing agriculture practices. The model's integration into a web platform ensures the best solution for the farmers, enabling the farmers to access diverse regions and environments.

Future work for the model is to focus on the platform's capabilities, adding new environmental factors. The focus is to expand the model's scope to cover all the variety of crops. Additionally, more advanced machine learning techniques should be integrated to improve prediction accuracy and robustness. The main aim is to make AgriRec contribute to the global effort to achieve good crops and sustainable agriculture by providing farmers with the tools that can be easily accessible and help farmers grow more crops and get more profitable.

REFERENCES

[1]. R. Kumar *et al.*, "Crop Selection Method to maximize crop yield rate using machine learning technique," 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), Avadi, India, 2015, pp. 138-145, doi: 10.1109/ICSTM.2015.7225403.

[2]. K. P. Sasi *et al.*, "A Machine Learning-Enabled System for Crop Recommendation," *Engineering Proceedings*, vol. 67, no. 1, p. 51, 2024, doi: 10.3390/engproc2024067051.

[3]. Z. Doshi *et al.*, "AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-6, doi: 10.1109/ICCUBEA.2018.8697349.

[4]. S. A. Z. Rahman *et al.*, "Soil Classification Using Machine Learning Methods and Crop Suggestion Based on Soil Series," 2018 21st International Conference of Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 2018, pp. 1-4, doi: 10.1109/ICCITECHN.2018.8631943.

[5]. H. Hammouch *et al.*, "A Systematic Review and Meta-Analysis of Intelligent Irrigation Systems," in *IEEE Access*, vol. 12, pp. 128285-128304, 2024, doi: 10.1109/ACCESS.2024.3421322.

[6]. S. M and S. Manju Priya, "Empowering Farmers with AI: Risk Mitigation Through Machine Learning in Smart Irrigation Systems," *2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT)*, Kollam, India, 2024, pp. 197-201, doi: 10.1109/ICCPCT61902.2024.10673341.

[7]. G. Chauhan and A. Chaudhary, "Crop Recommendation System using Machine Learning Algorithms," *2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART)*, MORADABAD, India, 2021, pp. 109-112, doi: 10.1109/SMART52563.2021.9676210.

[8]. N. H. Kulkarni *et al.*, "Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique," 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), Bengaluru, India, 2018, pp. 114-119, doi: 10.1109/CSITSS.2018.8768790.

[9]. S. Pudumalar *et al.*, "Crop recommendation system for precision agriculture," 2016 Eighth International Conference on Advanced Computing (ICoAC), Chennai, India, 2017, pp. 32-36, doi: 10.1109/ICoAC.2017.7951740.

[10]. A. Savla *et al.*, "Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture," 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, India, 2015, pp. 1-7, doi: 10.1109/ICIIECS.2015.7193120.

[11]. J. Wu *et al.*, "The Development and Application of Decision Tree for Agriculture Data," 2009 Second International Symposium on Intelligent Information Technology and Security Informatics, Moscow, Russia, 2009, pp. 16-20, doi: 10.1109/IITSI.2009.10.

[12]. P. Pawar *et al.*, "An Automated Combined System for Crop Prediction and Yield Prediction Using Deep Hybrid Learning Technique," 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Pune, India, 2023, pp. 1-5, doi: 10.1109/ICCUBEA58933.2023.10392068.

[13]. B. Varshitha *et al.*, "Farmeasy - A Web Portal for Farmers," 2023 7th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), Bangalore, India, 2023, pp. 1-5, doi: 10.1109/CSITSS60515.2023.10334226.

[14]. S. S. R. Jasti *et al.*, "Crop Intelligent: Weather based Crop Selection using Machine Learning," 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2023, pp. 1594-1600, doi: 10.1109/ICSCDS56580.2023.10104898.

[15]. S. Babu, "A software model for precision agriculture for small and marginal farmers," 2013 IEEE Global Humanitarian Technology Conference: South Asia Satellite (GHTC-SAS), Trivandrum, India, 2013, pp. 352-355, doi: 10.1109/GHTC-SAS.2013.6629944.

[16]. S. Veenadhari *et al.*, "Machine learning approach for forecasting crop yield based on climatic parameters," 2014 International Conference on Computer Communication and Informatics, Coimbatore, India, 2014, pp. 1-5, doi: 10.1109/ICCCI.2014.6921718.

[17]. S. Wolfert *et al.*, "Big Data in Smart Farming – A Review," *Agricultural Systems*, vol. 153, pp. 69-80, 2017, doi: 10.1016/j.agsy.2017.01.023.

[18]. J. W. Jones *et al.*, "Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science," *Agricultural Systems*, vol. 155, pp. 269-288, 2017, doi: 10.1016/j.agsy.2016.09.021.

[19]. H. Andrianto *et al.*, "Future Research on Smart Farming Platforms," 2022 International Conference on Information Technology Systems and Innovation (ICITSI), Bandung, Indonesia, 2022, pp. 358-362, doi: 10.1109/ICITSI56531.2022.9970957.

[20]. S. Mali, "Crop Recommendation Dataset," *figshare*, Jul. 16, 2024. [Online]. Available: https://doi.org/10.6084/m9.figshare.26308696 [Accessed: Jan. 17, 2025].

[21]. D. Balakrishnan *et al.*, "Agricultural Crop Recommendation System," 2023 3rd International Conference on Intelligent Technologies (CONIT), Hubli, India, 2023, pp. 1-5, doi: 10.1109/CONIT59222.2023.10205756.