Crop Recommendation using Machine Learning

Dr.T. Shekar Reddy[1] , Dr.Animoni Nagaraju[2], Dr.Y. RamaDevi[3]

Dr. T. Shekar Reddy, Assistant Professor, Department of Computer Science, Telangana Mahila Vishwavidyalam(Womens University) , formerly University College For Women(A), OU. Hyderabad, Telangana, India. Email ID: shekarreddy08@gmail.com

Dr. Animoni Nagaraju, Professor, Department of CSE, Malla Reddy College of Engineering for Womens , Malla Reddy Vishwavidyapeeth Deemed to be University, Hyderabad, Telangana, India. Email ID: dranagaraju.cse@gmail.com

Dr.Y.RamaDevi, Professor, Department of AIML ,Chaitanya Bharathi Institute of Technology, Gandipet, Hyderabad, Telangana, India-500075. Email id: yramadevi_cseaiml@cbit.ac.in

I. ABSTRACT

The crop recommendation system using machine learning is an intelligent decision support system that provides recommendations to farmers on the most suitable crop to cultivate based on soil and weather conditions like tempe_rature, humi_dity, rain_fall, nitro_gen, pota_ssium, phos_phorus and PH_value of the soil. This system uses machine learning algorithms like Decision_Tree, Random_Forest, Naive_Bayes, Support_Vector_Machine (SVM), Gaussian_NB, and Logistic_Regression to analyze data on soil properties, climate, and other relevant factors to generate personalized crop recommendations for each farmer.

Keywords: Crop recommendation, tempe_rature, humi_dity, rain_fall, nitro_gen, pota_ssium, phos_phorus, PH_value, Decision_Tree, Random_Forest, Naïve_Bayes, Support_Vector_Machine (SVM), Gaussian_NB, Logistic_Regression, machine learning.

II. INTRODUCTION

Traditional crop recommendation methods often depend on the experience of farmers or agricultural experts, which may not always yield accurate results. In contrast, a machine learning-based crop recommendation system offers a more data-centric approach, enabling farmers to make well-informed decisions grounded in objective data analysis.

This system involves gathering information on various factors like soil characteristics, climate conditions, and crop yields. The data is then processed using machine learning algorithms to develop models that predict the most suitable crops for a specific region. The system generates a list of recommended crops along with their projected yields, guiding farmers in selecting crops that maximize profitability.

By enhancing crop yields, lowering input costs, and supporting sustainable farming practices, this technology has the potential to transform agriculture. As the global population continues to expand, such advancements will become increasingly vital for ensuring food security in the future.

III. LITERATURE SURVEY

Paper no.	Title	Authors	Year	Description	Algorithm/Technique	Key Findings
1.	Machine Learning Facilitated Rice Prediction in Bangladesh	Mohammad Motiur Rahman, Naheena Haq, Rashedur M Rahman	2015	SOM grouped variables; chi-square tested dependencies. Weather conditions (temperature, precipitation) drive crop yield.	Self-Organizing Map, Chi-square Test	Weather conditions significantly impact crop yield.
2.	SVM-Based Classification Scheme of Maize Crop	Suhas S Athani, CH Tejeshwar	2017	Automated weed-crop classification using SVM on maize images, achieving 82% accuracy.	Support Vector Machine (SVM)	SVM effectively classifies weeds in crops.
3	Crop Recommendation System using ML	Dr. D. Kavitha, R. Triveni, S. Harsha Vardhan, N. Akhil, V. Kumar ParasuRam	2023	Uses ML to predict suitable crops based on soil and climate data, providing crop recommendations with estimated yields.	Machine Learning Algorithms	Data-driven crop recommendations enhance farmer decision-making.
4	Review of ML Techniques for Crop Recommendation	Anoushka Pandey, Liza Kansal, Sanidhya Madhav Shukla, Parneeta, Dr. Dhaliwal	N/A	Reviews algorithms (Naive Bayes, SVM, KNN, Random Forest); Naive Bayes performed best in precision, recall, accuracy.	Naive Bayes, Decision Trees, SVM, KNN, Random Forest	Naive Bayes outperforms other models for crop recommendation.
5	Hybrid Approach to Crop Selection Using ML and Climate Data	Singh, K.; Patel, M.	N/A	Combines traditional agronomy with ML to optimize crop selection using climate and soil data, improving yield predictability.	Hybrid Model (Agronomy + Machine Learning)	Hybrid approach enhances crop selection in varying climates.

IV. METHODOLOGY

The following steps are involved while implementing the project to achieve the results.

3.1 Data_Collection:

The first step is data collection which is most important to implementing the project. Data collection is the process of gathering information.

3.2 Data Pre-Processing:

In this step, the collected data will be cleaned and filtered. The real-time data that was collected won't be in our desired format; there will be some noisy data, missing values, redundant values, etc. If we want accurate results the data should be accurate. To achieve accurate data, this step will be used to remove redundant values and to fill the missing data with appropriate values.

3.3 Model_Building:

In this step, we will build and train the recommendation system using the preprocessed data and the ML algorithms. To train the model the dataset needs to be divided into two parts, i.e., training_data_set and testing_data_set. By using the ML algorithms on training_data_set we will train and build the model.

3.4 Crop Recommendation:

In this project, we used five ML algorithms and the accuracies of the five algorithms will be compared, and the one which has the highest accuracy will be used to recommend the suitable crop.

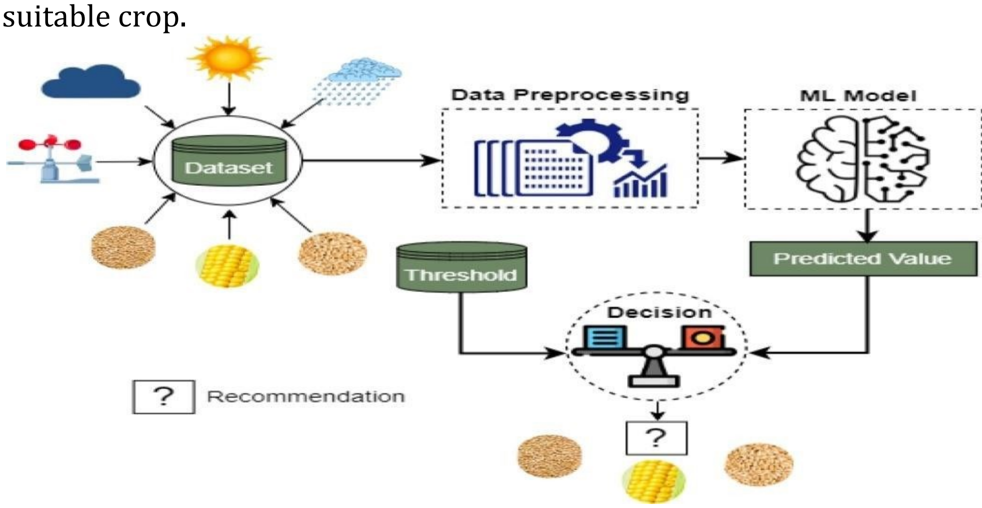


Figure-3.1: System_Architecture

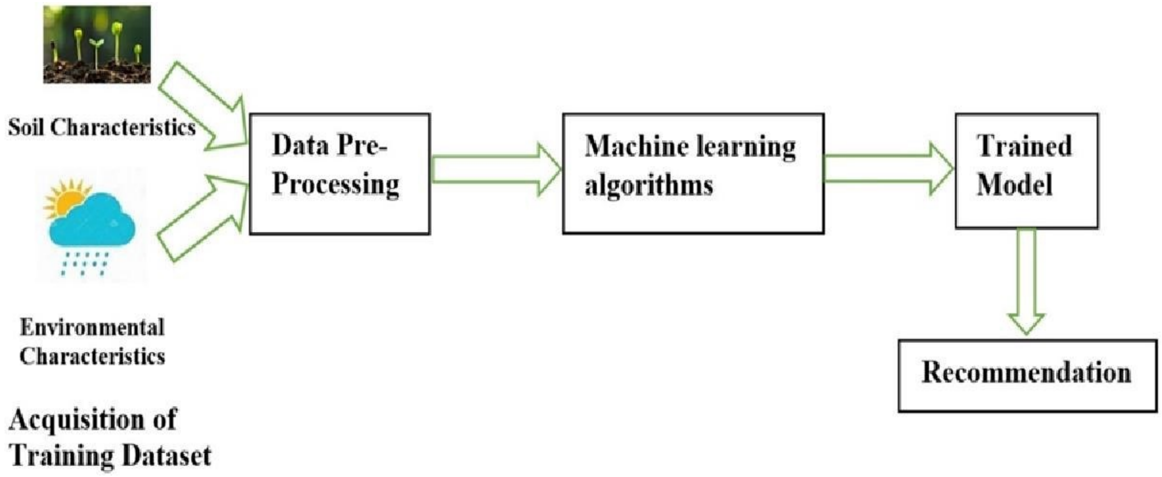


Figure- 3.2: Proposed_System_Model

V. DATASET

The dataset includes parameters such as Nitrogen (N), Phosphorus (P), Potassium (K), soil pH value, humidity, temperature, and rainfall. The data was sourced from the Kaggle website and comprises 2,200 instances based on historical data. It covers 22 different crops, including rice, maize, chickpea, kidney beans, pigeon peas, moth beans, mung beans, black gram, lentil, pomegranate, banana, mango, grapes, watermelon, muskmelon, apple, orange, papaya, coconut, cotton, jute, and coffee.

Attribute Description:
The dataset contains seven attributes, which are described as follows:

- **N:** Nitrogen content in the soil (ratio)
- **P:** Phosphorus content in the soil (ratio)
- **K:** Potassium content in the soil (ratio)
- **Temperature:** Measured in degrees Celsius
- **Humidity:** Relative humidity in percentage (%)
- **pH:** pH value of the soil
- **Rainfall:** Relative rainfall in millimetres (mm)

VI. RESULTS

Six algorithms were used to train and build the model after completing the data preprocessing step. Following this, the accuracies of these machine learning algorithms were compared, and the one with the highest accuracy was selected to be applied to the testing data to recommend the most suitable crops.

S.No.	Algorithm Name	Accuracy
1	Logistic_Regression	0.97
2	Random_Forest_Classifier	1.00
3	Decision_Tree_Classifier	0.99
4	Support _Vector_Machine(SVM)	0.98
5	k-Nearest_Neighbors(KNN)	0.98
6	Gaussian_Naive_Bayes	0.99

Table- **Algorithms Accuracy Values table**

The table presents a comparison of the accuracies of six machine learning algorithms. The **Random Forest Classifier** achieves a perfect accuracy of 1.00. Both the **Support Vector Machine (SVM)** and **k-nearest Neighbors (KNN)** follow closely with an accuracy of 0.98. The **Logistic Regression** model scores 0.97, while the **Decision Tree Classifier** and **Gaussian Naive Bayes** both demonstrate an accuracy of 0.99.

```
In [59]: x1=["LogisticRegression","RandomForest","DecisionTree","SVC","KNN","GaussianNB"]
         y1=[0.97,1.0,0.99,0.98,0.98,0.99]
         colors=["red","aqua","green","grey","blue","orange"]
         plt.figure(figsize=(5,5))
         plt.barh(x1,y1,color=colors,height=0.4)
         plt.xlabel("Accuracy Comparision")
         plt.ylabel("Algorithms")
         plt.title("comparision of Algorithms")
         plt.show()
```

From the **Algorithms Accuracy Values** table, it is evident that the **Random Forest** algorithm achieved the highest accuracy at 1.00, followed by **Gaussian Naive Bayes** with 0.99, and **Support Vector Machine (SVM)** with 0.98. The **Logistic Regression** model provided an accuracy of 0.97, while the **Decision Tree** also reached an accuracy of 0.99. Based on these results, the **Random Forest** algorithm was selected to recommend a suitable crop on the test data. A bar graph was plotted below using the values from the **Algorithms Accuracy Values table Values** table.

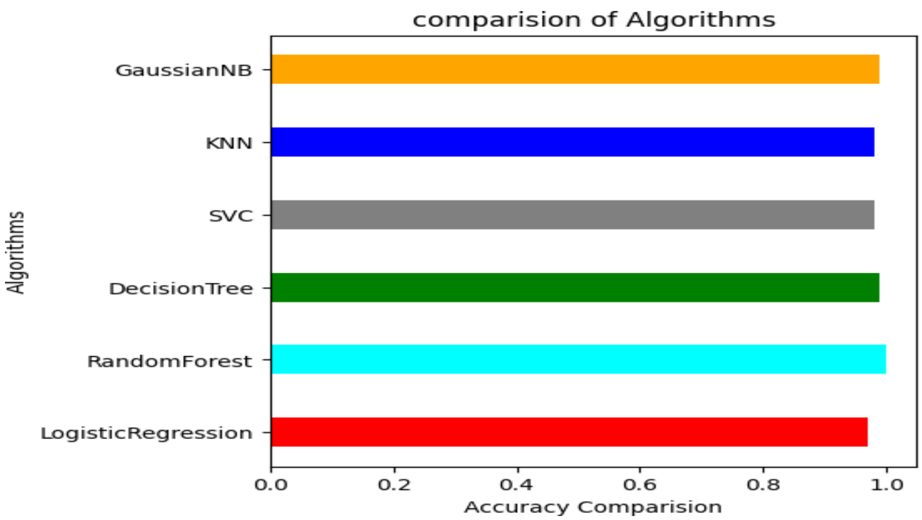


Figure – 5.1: Accuracy comparison of 6 ML algorithms

Random_Forest_Algorithm:

The **Random Forest Classifier** is a versatile and powerful machine learning algorithm, commonly used for both classification and regression tasks. It works by constructing multiple decision trees during the training phase and outputs either the mode of the classes (for classification) or the mean prediction (for regression) from the individual trees.

Due to its ability to handle complex datasets and deliver robust predictions, Random Forest is widely applied across various domains, such as finance, healthcare, and marketing. It is used for tasks like credit scoring, medical diagnosis, and customer segmentation. As a highly effective classifier, Random Forest strikes a balance between accuracy, robustness, and flexibility, making it a popular choice for a wide range of machine-learning applications.

```
# Plot the bar chart
counts_df2.plot(kind='bar', figsize=(7,4))
plt.title('Class Distribution in True and Predicted Labels of Random Forest Classifier')
plt.xlabel('Class')
plt.ylabel('Count')
plt.xticks(rotation=90)
plt.legend(title='Labels', loc='upper left',bbox_to_anchor=(0.2,1.0),title_fontsize='8', fontsize='7')
plt.show()
```

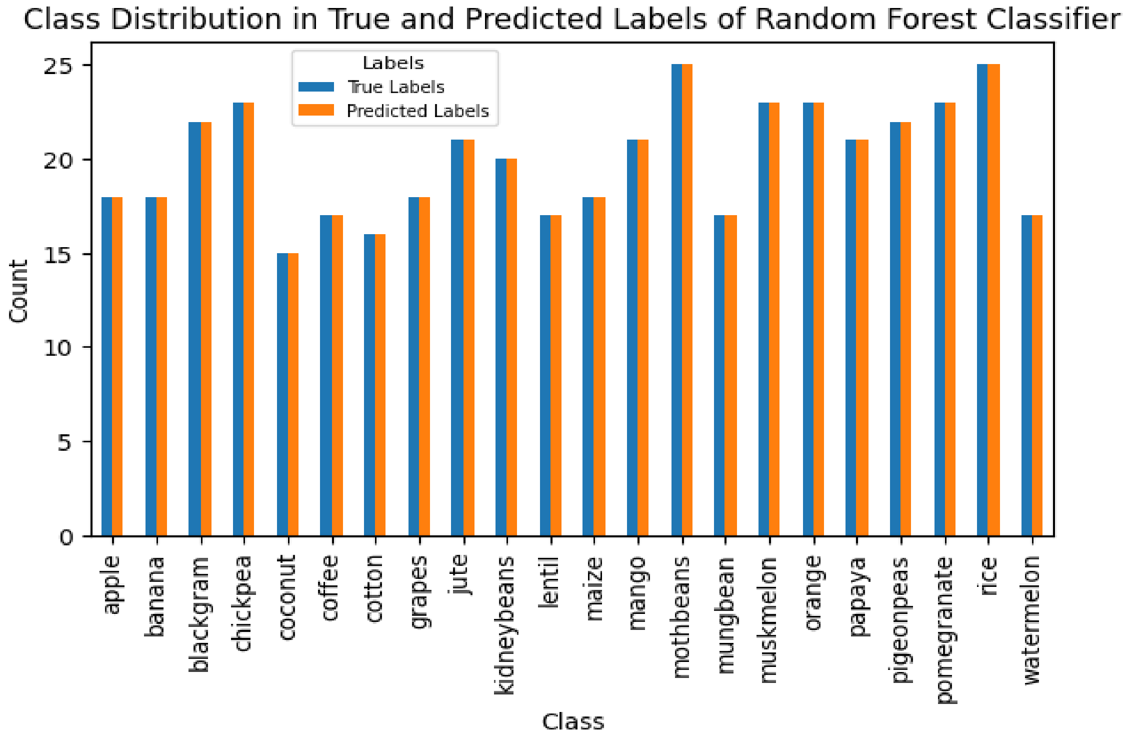


Figure-5.2: Result of Random_Forest algorithm

Description:

The bar chart compares the true and predicted label distributions for the **Random Forest Classifier** across various classes, such as apple, banana, black gram, and others. The blue bars represent the true label counts, while the orange bars display the predicted counts. The x-axis lists the different classes, and the y-axis indicates the corresponding counts. The chart shows that, for most classes, the true and predicted counts are closely aligned, indicating strong model performance. However, a few discrepancies suggest potential areas for improvement.

Crop Recommendation Using Machine Learning:

The machine learning pipeline for crop recommendation begins with the user inputting environmental factors such as humidity, rainfall, pH value, temperature, and others. The collected data is then preprocessed alongside historical crop data. Feature engineering is applied to the data, which is subsequently used to train multiple models, including **Random Forest**, **Support Vector Machine (SVM)**, **k-nearest neighbours (KNN)**, **Gaussian Naive Bayes**, **Logistic Regression**, and **Decision Trees**. After training, the models are evaluated to determine their performance, and the best-performing model is selected. The final model, which is the **Random Forest**, is utilized for prediction, ultimately providing a recommended crop to the user based on the input data.

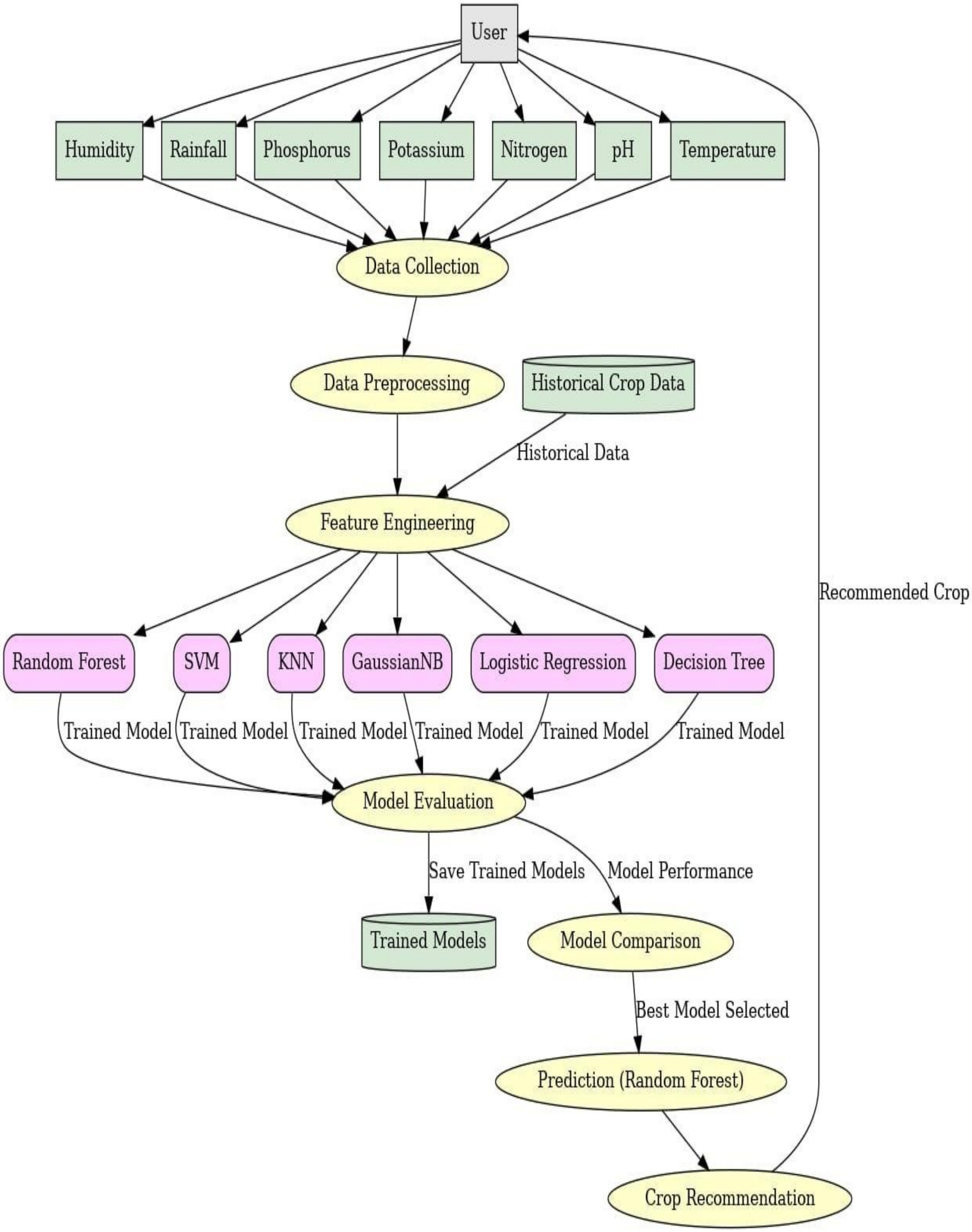
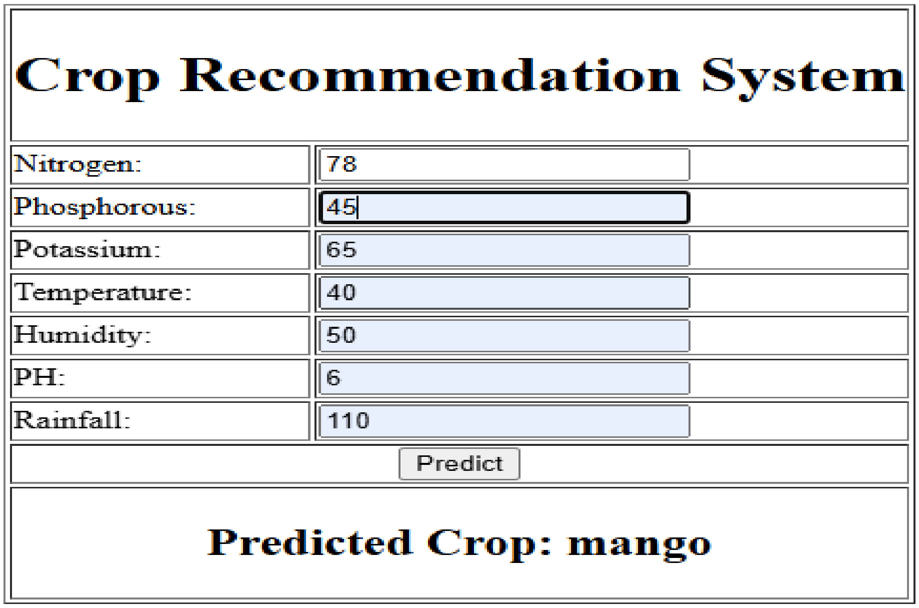


Fig-5.2: Crop Recommendation Using Machine Learning

OUTPUT SCREEN:



Web Application With Predicted Crop

VII. CONCLUSION

The system leverages machine learning algorithms to analyze soil and climate data, delivering accurate results and crop recommendations. This enhances overall efficiency and productivity in agricultural practices, empowering farmers with valuable insights and guidance for successful crop cultivation. Additionally, the system promotes sustainability by encouraging crop diversification and minimizing environmental impact.

The Crop Recommendation System provides a reliable, data-driven approach to enhancing agricultural productivity by accurately suggesting crops tailored to specific environmental conditions. Through rigorous validation methods—including cross-validation, field trials, and user feedback—the system has demonstrated its effectiveness in optimizing yields and resource usage. Its practical application across diverse farming scenarios underscores its value as a tool for sustainable agriculture. With continuous updates and refinements, the system can adapt to evolving agricultural challenges, offering farmers essential guidance for informed decision-making and long-term success.

The **Random Forest Classifier** has proven to be the best algorithm for predicting crop recommendations based on the provided accuracy results. It effectively classifies data and is likely the most robust model in this context. However, **Decision Trees** and **Gaussian Naive**

Bayes are also excellent alternatives, particularly if interpretability or computational simplicity is a priority.

VIII. FUTURE SCOPE

- Currently, the system cannot provide prescriptive analysis for farmers on how to grow a desired crop under existing land conditions. It could benefit from recommendations on necessary changes, such as adjusting pH levels or modifying nitrogen (N), phosphorus (P), and potassium (K) content in the soil.
- The development of a mobile application would facilitate convenient access and usage on smartphones and tablets. Additionally, integrating user feedback and ongoing data collection will help continuously enhance the recommendation system's performance.
- As a result, farmers can select crops that are in demand, ensuring they choose options that maximize their yield.
- This approach ultimately leads to **maximizing their profits!**

IX. REFERENCES

[1] Manpreet Kaur, Heena Gulati, Harish Kundra, "Data Mining in Agriculture on Crop Price Prediction: Techniques and Applications", International Journal of Computer Applications, Volume 99– No.12, August 2014.

[2] J. Meng, "Research on the cost of agricultural products circulation and its control under the new normal economic development," Commercial Times, no. 23, pp. 145147, 2016.

[3] A. Kaloxylos et al., "Farm management systems and the future Internet era," Comput. Electron. Agricult., vol. 89, pp. 130–144, Nov. 2012.

[4] N. N. Li, T. S. Li, Z. S. Yu, Y. Rui, Y. Y. Miao, and Y. S. Li, "Factors influencing farmers' adoption of new technology based on Logistic-ISM model-a case study of potato planting technology in Dingxi City, Gansu Province," Progress in Geography, vol. 33, no. 4, pp. 542-551, 2014.

[5] Y. Wang, "A neural network adaptive control based on rapid learning method and its application," Advances In Modeling and Analysis, Vol. 46(3), pp. 27-34,1994.

[6] .Mahule, Ankit Arun, and A. J. Agrawal. "Hybrid Method for Improving Accuracy of Crop-Type Detection using Machine Learning." International Journal 9, no. 2 (2020).

[7] .Akshatha, K. R., and K. S. Shreedhara. "Implementation of machine learning algorithms for crop recommendation using precision agriculture." International Journal of Research in Engineering, Science and Management (IJRESM) 1, no. 6 (2018): 58-60.

[8] .Dighe, Deepti, Harshada Joshi, Aishwarya Katkar, Sneha Patil, and Shrikant Kokate. "Survey of Crop Recommendation Systems." (2018).

[9] .Jaiswal, Sapna, Tejaswi Kharade, Nikita Kotambe, and Shilpa Shinde. "Collaborative Recommendation System For Agriculture Sector." In ITM Web of Conferences, vol. 32. EDP Sciences, 2020

[10] .Ghadge, Rushika, Juilee Kulkarni, Pooja More, Sachee Nene, and R. L. Priya. "Prediction of crop yield using machine learning." Int. Res. J. Eng. Technol.(IRJET) 5 (2018).

[11] .Kulkarni, Nidhi H., G. N. Srinivasan, B. M. Sagar, and N. K. Cauvery. "Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique." In 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), pp. 114-119. IEEE, 2018.

[12] . Kumar, Y. Jeevan Nagendra, V. Spandana, V. S. Vaishnavi, K. Neha, and V. G. R. R. Devi. "Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector." In 2020 5th International Conference on Communication and Electronics Systems (ICCES), pp. 736-741. IEEE, 2020.

[13] . Medar, Ramesh, Vijay S. Rajpurohit, and Shweta Shweta. "Crop yield prediction using machine learning techniques." In 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), pp. 1-5. IEEE, 2019.

[14] . Jain, Sonal, and Dharavath Ramesh. "Machine Learning convergence for weather based crop selection." In 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), pp. 1-6. IEEE, 2020.

[15]. Gandge, Yogesh. "A study on various data mining techniques for crop yield prediction." In 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), pp. 420-423. IEEE, 2017.

[16] https://www.kaggle.com/datasets/atharvaingle/crop-recommendation- dataset

[17] https://ijisrt.com/assets/upload/files/IJISRT20MAY722.pdf

[18]https://www.javatpoint.com/machine-learning-random-forest-algorithm

[19]https://www.javatpoint.com/machine-learning-naive-bayes-classifier