iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** IV **Month of publication:** April 2024

DOI: https://doi.org/10.22214/ijraset.2024.61326

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

CRISHI: A Crop Recommendation System

Prof. Dr. Pooja Kherwa[1], Om Srivastava[2]
MSIT

Abstract: ***The Crop Recommendation System proposes a mixed model to suggest crops for Indian states, considering various factors such as soil pH, moisture, NPK levels, crop data, and temperature. This suggestion model is constructed as a sample model incorporating supervised machine learning algorithms. The technology- based agricultural crop identifier model aims to enhance farmers crop yield by identifying appropriate crops based on their parameters. The efficacy of the recommendation model lies in its ability to recommend the right crop for specific conditions, which in turn contributes to addressing agricultural and farming challenges and boosting the Indian economy by maximizing crop production. The system also includes a grading process to identify crop quality and signalling the presence of low and high-quality crops. The utilization of a set of classifiers facilitates improved decision-making through multiple classifiers. The assessment process includes a request for decision-making to determine the outcomes of the classifiers.***

Keywords: Crop Recommendation, Soil PH, Humidity, NPK levels, Crop data, Temperature, Classification, Regression, Geographic and the climatic parameters, Crop Yield, Prediction and Ranking.

I. INTRODUCTION

In a rapidly advancing technological environment, integrating tech and AI techniques in agriculture becomes crucialto help farmers recognize and amplify the significance of their work in contributing to the Indian economy. Timelysharing of accurate information among farmers determines the degree of information exchange and outreach. By using languages like HTML and CSS, we develop a website, gather data from diverse sources and surveys, and use it to predict crop types, titles, and prices subjected to non-linear testing. Prioritization and ranking of crops are thenperformed. By providing this information through our app, farmers eliminate the need to seek updates from urban centers, thus enhancing accessibility to pertinent details. Implementing ml algorithms such as SVM, NB, KNN, Decision Tree, Random Forest, and Gradient Boosting enhances predictive abilities for crop yields and production costs.

A. Crop Recommendation System

The Crop Identifier Technology for crop cultivation empowers farmers to elevate crop yield by suggesting suitablecrops based on geographical and weather conditions. This is a hybrid model that has demonstrated its effectivenessin recommending appropriate crops.

B. Need of a Recommendation System

The Crop Recommendation Technology System for Agriculture supports farmers in maximizing crop yield through tailored recommendations determined by geographical and climatic parameters.

C. Crop Price Prediction Model

Utilizing a Linear Regression model on historical data enables farmers to foresee crop prices with potential enhancement through larger datasets. Segmentation into training and testing datasets, coupled with the implementation of Random Forest and linear regression models, refines crop price forecasting.

D. Need of a Price Prediction Model

Precise crop price estimation aids in informed decision-making before crop cultivation, offering insights to mitigate losses and manage price fluctuations effectively.

II. CHALLENGES

Throughout the Literature Review, various research gaps have surfaced concerning building a crop forecasting model using machine learning through Flask. The process presents challenges like data collection, appropriate model selection, scalability, integration, user interface design, real-time data updates, security concerns, model interpretability, deployment and hosting, and continuous monitoring and maintenance, each posing unique hurdlesin the project's efficacy.

III. OBJECTIVES

To bridge existing research gaps and enhance project performance, the established objectives encompass information acquisition, data cleansing and organization, real-time data analysis for prompt value updates, operational robustness through classifier combination, ranking methodology implementation for informed choices, and website development for user information exchange. The primary aim remains identifying a variety of viable crops, aiding farmers in optimal crop selection and yield optimization through predictive analysis of climate variables, area details, seasonal insights, and soil composition.

IV. RESEARCH METHODOLOGY

Designing a research methodology for a Crop Recommendation System project involves several key to ensure a systematic and reliable approach. Here's a suggested research methodology:

Problem Definition and Scope: Define the scope of your project. Clearly articulate the goals of CropRecommendation System and the specific challenges you aim to address.

A. Data Collection

The first set is the collection of the dataset. The second step is the subsequent processing of the dataset "Crop Recommendation with water data". We have collected this dataset from Kaggle and it has parameters N(nitrogen in the soil), P(phosphorus in the soil), K(potassium in the soil). These are the most important parameters taken into account. Weather conditions such as humidity, temperature, rainfall are also present in the dataset. A new feature has been added to the data that is total water available on the farm which tells the purity of the water to be used in the crops.

B. Processing of Data

After collecting the data, the desired feature are filtered and extracted from the data. Then preprocessing of the datais started that is cleaning of data, converting the numerical in one format, removing null values and reducing the data. Then the dataset is divided into training and testing dataset and as per rule 80% is the training data and the rest20% is the testing data. The training model is fed into the model and neural network and is developed and the testingdata is then used to test the result.

C. Models Used

1) *Random Forest Algorithm:* This algorithm comes under supervised learning. More trees help to achieve higher accuracy and also prevents overfitting problems. Some steps are to be taken for this algorithm. K random data points are selected from the dataset. Associated decision trees are then built. Decide the number N of decision trees you want to build for your training set. Selection and building steps are repeated. Find prediction of each tree for new data points and assign them to the category in which it fitsthe most.

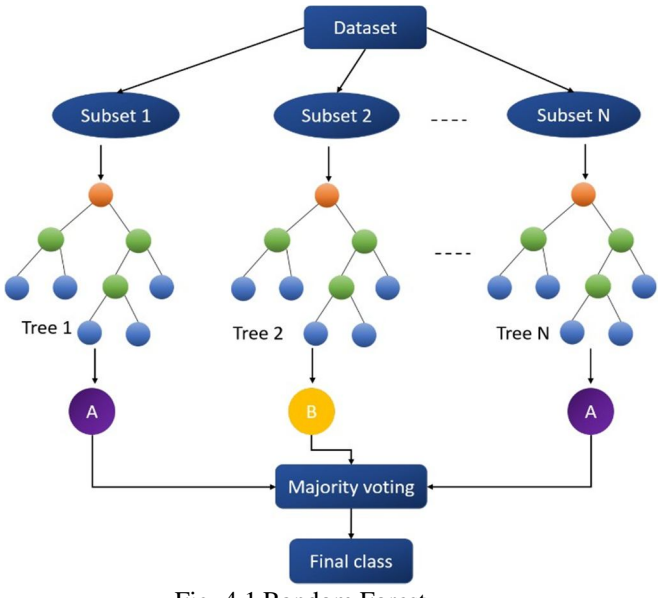


Fig. 4.1 Random Forest

2) Logistic Regression Algorithm: It is a statistical algorithmic method used mainly for binary classification task. Its main goal is to predict the probability that an instance belons to one of the classes. To map a linear combination of input features and their associated weights to a 0 to 1 probability, it utilizes the sigmoid function. The hypothesis function estimates the probability that the dependent variable is 1, given the input features. The model is trained by minimizing a cost function, often the log loss, using techniques like gradient descent. Logistic regression is widely applied in areas such as finance and healthcare for tasks like fraud detection and disease prediction, providing a simple yet effective approach to binary classification problems.

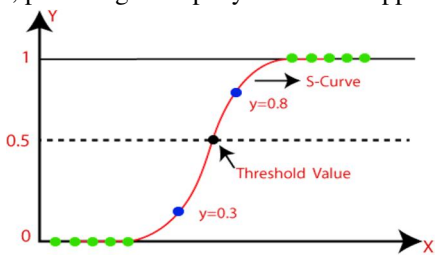


Fig. 4.2 Logistic regression

3) Gaussian Bayes Algorithm: It is a classification algorithm for probabilities given by Bayes. In his theorem he made use of the "naive" assumption of feature independence. It is most used in text classification and spam filtering. Despite its simplistic assumption of feature independence, Naive Bayes often performs surprisingly well in practice and is computationally efficient. It is mostly used in NLP tasks, classification of documents and sentiment analysis because it is very simple, fast and effective in handling high dimensional data with many features.

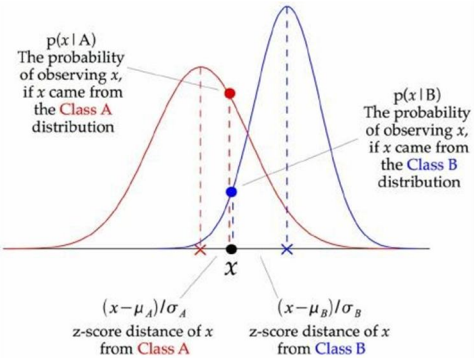


Fig. 4.3 Gaussian Naïve Bayes

4) Support Vector Machine Algorithm: The Support Vector Machines (SVMs) represent a potent group of supervised learning methods, and Support Vector Classification (SVC) specifically serves classification purposes. SVC works by identifying the hyperplane that best separates distinct classes within the feature space while increasing the margin. It is explained as the hyperplane's distance from the nearest data point of any class. The main goal of the algorithm is to strike a balance between fitting the data accurately and generalizing well to new, unseen data. SVC proves effective in high-dimensional spaces and demonstrates versatility, accommodating both linear and non-linear classification tasks using diverse kernel functions. Despite its advantages, SVC may exhibit sensitivity to hyperparameter selection, necessitating meticulous tuning. Nevertheless, SVMs, including SVC, see widespread utilization across various domains like image recognition, text classification, and bioinformatics due to their capacity to handle intricate decision boundaries and offer reliable classification performance.

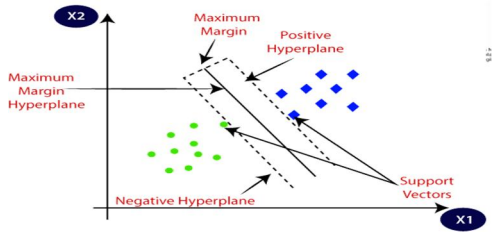


Fig. 4.4 Support Vector Classification

5) K- Nearest Neighbors Algorithm: K-Nearest Neighbors (KNN) stands as a straightforward yet efficient supervised learning tool applied in both classification and regression assignments. KNN's core lies in its intuitive methodology: when predicting a data point's class or value, it examines the K-nearest neighbors in the feature space, determined by a distance metric like the Euclidean distance. The target data point is then assigned the majority class or average value of these neighboring points. KNN operates as a non-parametric model, avoiding assuming of underlying data distribution and adapting well to diverse datasets. However, the algorithm's performance may hinge on the choice of the number of neighbors (K) and the distance metric. KNN notably fits scenarios where irregular decision boundaries challenge representation using simple mathematical formulas. Despite being computationally intensive compared to certain counterparts, KNN's simplicity and effectiveness render it a valuable asset across multiple domains including pattern recognition, image analysis, and recommendation systems.

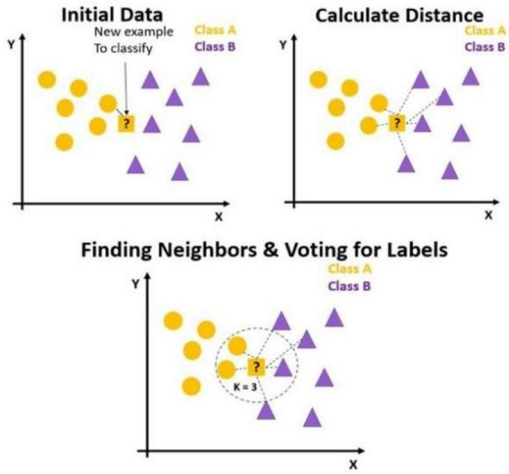


Fig. 4.5 KNN

6) Decision Trees Algorithm: Decision Trees emerge as a prevalent and interpretable machine learning technique serving both classification and regression objectives. Operating akin to flowcharts, a decision tree splits the dataset into smaller units based on the most important features, forming a tree structure where each node represents a decision predicated on a specific feature. The fundamental aim revolves around constructing a tree that efficiently segregates the data, leading to leaves that signify the predicted outcome. Decision Trees entice users with their transparency, easy understandability, and capability to work on both numerical and categorical data without extensive preprocessing requirements. Nonetheless, susceptibility to overfitting, particularly on noisy datasets, poses a challenge often addressed via strategies like pruning. Ensembles such as Random Forests, comprising multiple decision trees, are frequently deployed to uplift predictive accuracy and robustness. Decision Trees discover applications across an array of fields including finance, healthcare, and natural language processing owing to their adaptability and knack for capturing intricate decision-making processes.

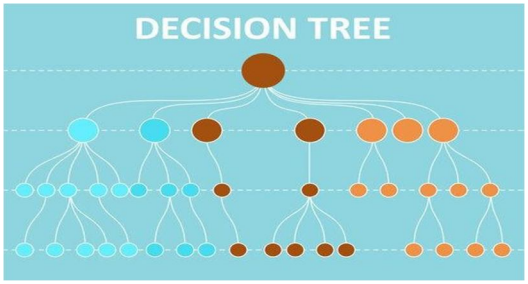


Fig. 4.6 Decision Trees

7) Bagging Classifier Algorithm: A Bagging Classifier, shortened from Bootstrap Aggregating Classifier, stands as an ensemble learning mechanism amalgamating the predictions from numerous base classifiers to enhance overall performance and robustness. The premise behind bagging involves training each base classifier on a distinct bootstrap sample derived from the original training dataset. A bootstrap sample materializes through the random selection of instances from the dataset with replacement, permitting some instances to feature multiple times while excluding others entirely.

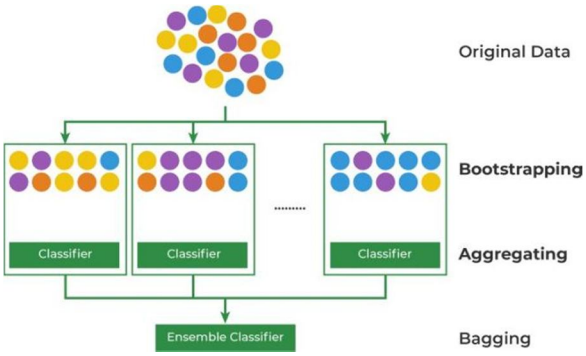


Fig. 4.7 Bagging Classifier

8) AdaBoost Algorithm: AdaBoost, an abbreviation for Adaptive Boosting, garners recognition as a prevalent ensemble learning algorithm crafted to elevate the efficacy of feeble learners and forge a resilient predictive model. Within AdaBoost, a weak learner, typically a simple model like a decision tree with restricted depth, undergoes training on the dataset. The algorithm assigns higher weights to misclassified instances from the weak learner, rendering these instances more impactful in subsequent iterations. This adaptive characteristic empowers AdaBoost to concentrate on data points posing challenges for the prevailing ensemble, progressively refining overall performance. Each iteration sees the training of a fresh weak learner, with its weight in the ultimate ensemble contingent on its accuracy. This iterative process persists for a predetermined number of iterations or until a stipulated accuracy threshold gets attained. The final prediction materializes as a weighted amalgamation of the weak learners' outputs.

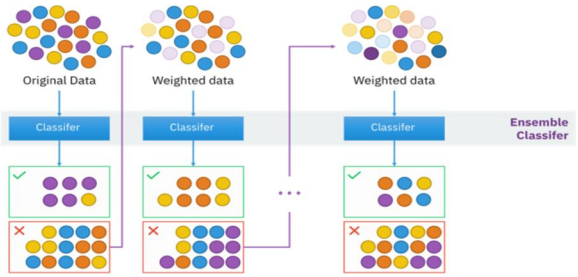


Fig. 4.8 AdaBoost Classifier

9) Gradient Boosting Algorithm: Gradient Boosting emerges as a robust ensemble learning strategy constructing a predictive model incrementally by blending the strengths of inferior learners. Unlike AdaBoost, which focuses on tweaking the misclassified instances' weights, Gradient Boosting crafts each weak learner to rectify the preceding ones' errors systematically. The process unfolds with a rudimentary model, frequently a shallow decision tree, progressing iteratively by adding fresh trees that encompass the residuals (the deviations between the anticipated and actual values) of the current model. Every new tree is honed to minimize the ensemble's error and receives a weight factor based on its contribution toward the collective enhancement. This journey is steered by a loss function quantifying the variance between predicted values and genuine values.

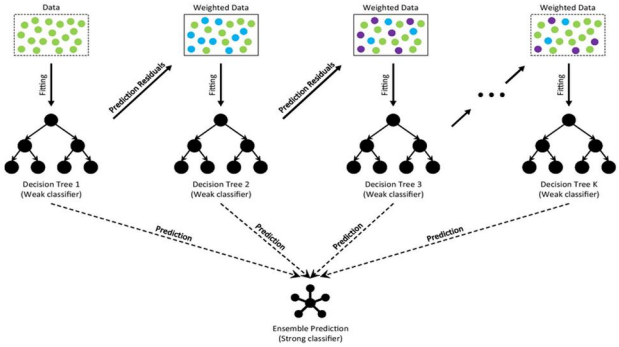


Fig. 4.9 Gradient Boosting

10) Extra Trees Algorithm: The Extra Algorithm Trees is a method in ensemble learning that improves upon the concept of Random Forests by adding an extra layer of randomness in the tree-building process. Like Random Forests, Extra Trees builds multiple decision trees, but with two main modifications to increase diversity and combat overfitting. Firstly, while Random Forests pick a subset of features for each split, Extra Trees takes it up a notch by randomly selecting the feature thresholds at each node. This additional randomness helps in creating more varied trees within the ensemble, reducing sensitivity to noise and outliers. Secondly, Extra Trees usually employs the entire dataset without bootstrapping when training each tree, fostering more diversity among individual trees. This approach proves valuable, especially with small datasets. The ensemble's prediction in Extra Trees utilizes voting for classification tasks or averaging for regression tasks.

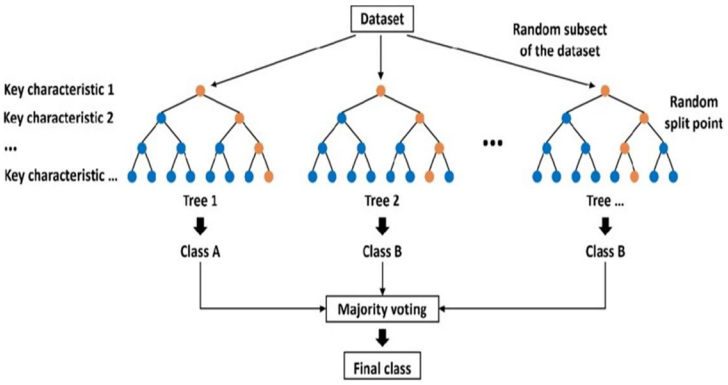


Fig. 4.10 Extra Trees

V. RESULTS AND DISCUSSION

The model was trained with various algorithms, including Random Forest Algorithm, Logistic Forest Algorithm, Gaussian Naive Bayes Algorithm, Support Vector Classification algorithm, K Nearest neighbors Algorithm, Decision tree Algorithm, Bagging Classifier Algorithm, ADABOOST Classifier Algorithm, Gradient Boosting Algorithm, and Extra Trees Algorithm, among which Random Forest Algorithm yielded the best results.

Logistic Regression algorithm	94.45 %
Naïve Bayes algorithm	98.73 %
Support Vector Machine algorithm	95.72 %
K-Nearest Neighbors algorithm	97.14 %
Decision Tree algorithm	98.41 %
Random Forest algorithm	98.73 %
Bagging algorithm	98.73 %
AdaBoost algorithm	18.06 %
Gradient Boosting algorithm	98.41 %
Extra Trees algorithm	89.85 %

Table No. 3.1 – Model Predictions

RANDOM FOREST ALGORITHM: This algorithm comes under supervised learning. More trees help to achieve higher accuracy and also prevents overfitting problems. Some steps are to be taken for this algorithm. K random data points are selected from the dataset. Associated decision trees are then built. Decide the number N of decision trees you want to build for your training set. Selection and building steps are repeated. Find prediction of each tree for new data points and assign them to the category in which it fitsthe most.

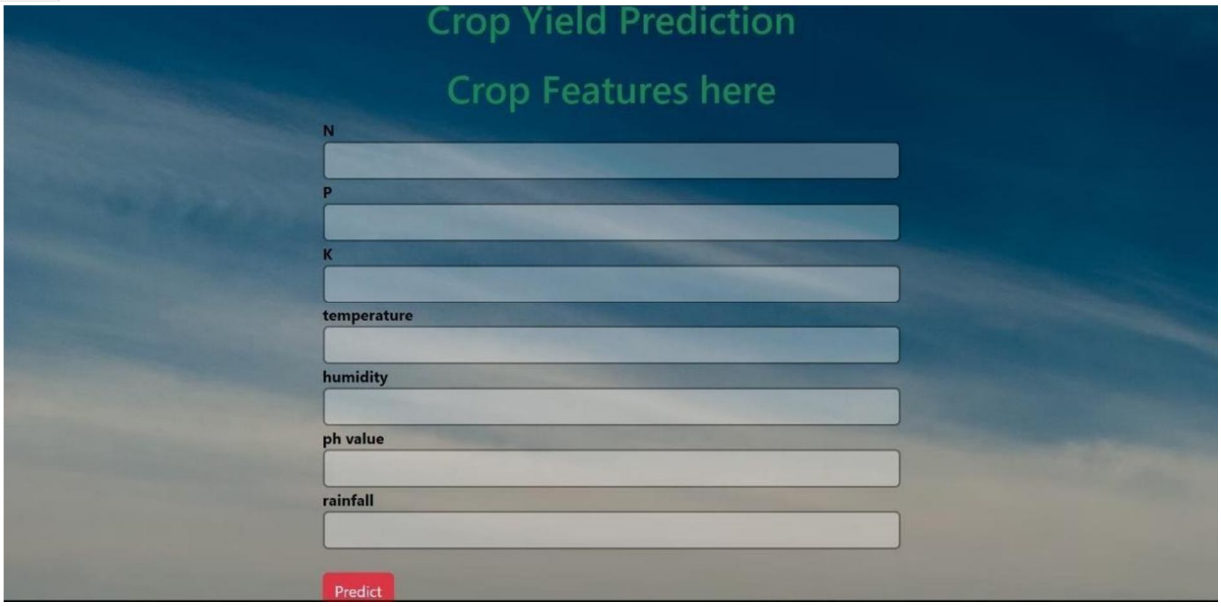


Fig. 5.1 Original Prediction Page

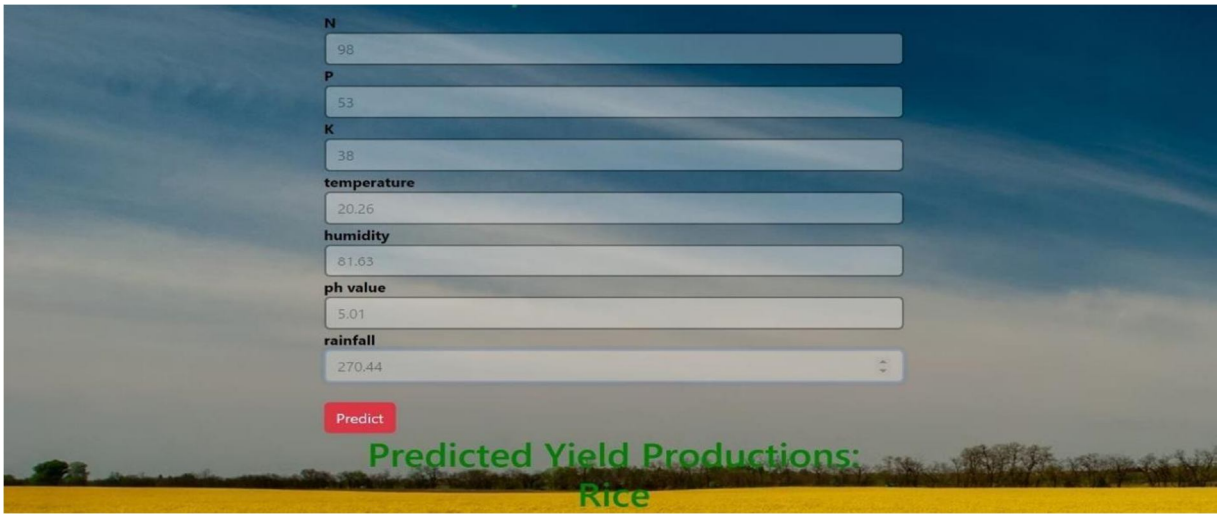


Fig. 5.2 Crop Predicted Page

VI. CONCLUSION

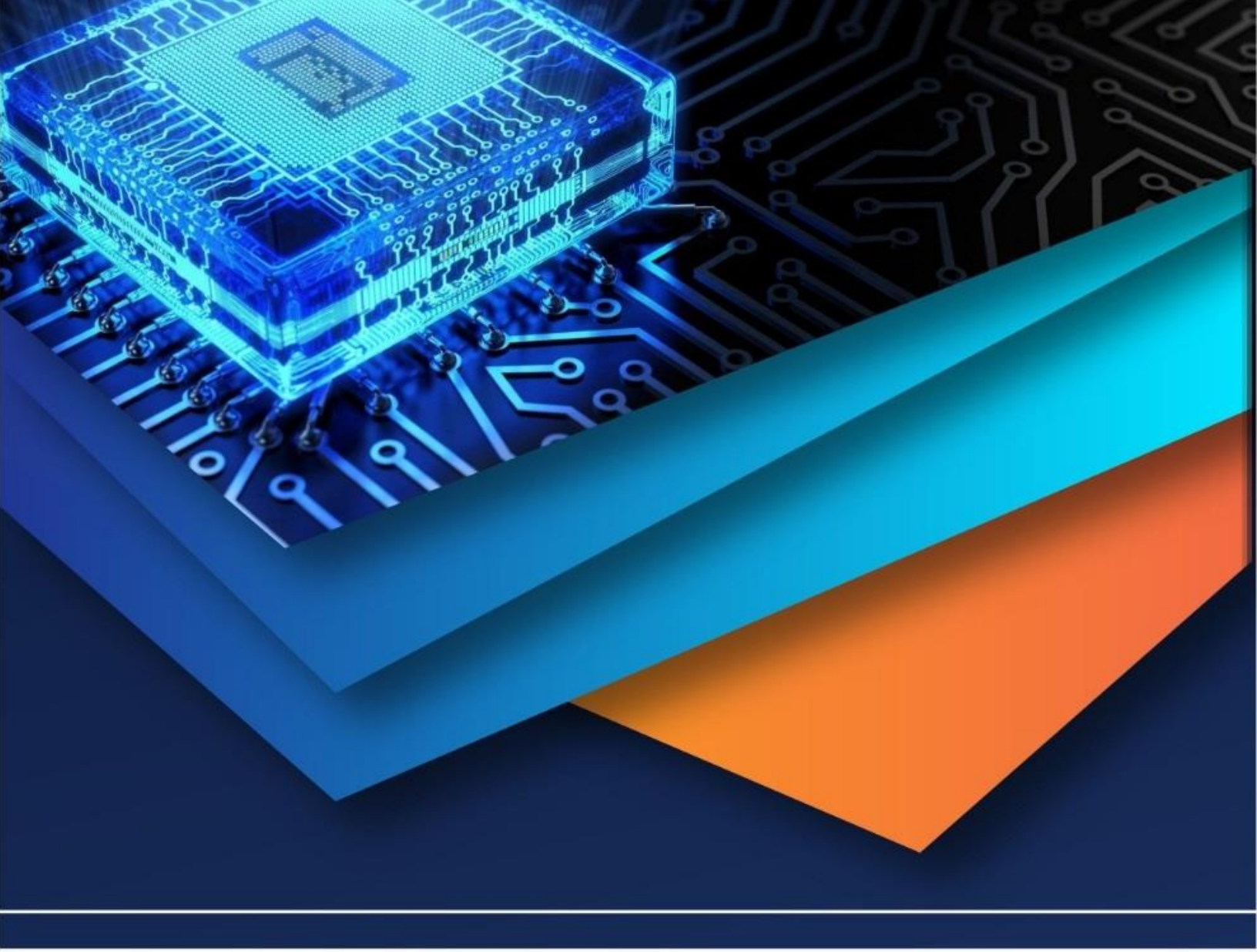
In summary, incorporating a machine learning model into a Flask web application for crop recommendations presents significant potential for modern agriculture. By utilizing machine learning algorithms like decision trees or random forests, farmers can receive tailored crop suggestions based on crucial environmental factors. The Flask web application delivers a user-friendly platform for farmers to input specific parameters and receive accurate recommendations. This fusion of technology streamlines decision-making for farmers and promotes sustainable agricultural practices. By continuously updating and learning from fresh data, the system excels in providing preciseand timely crop recommendations, further paving the way for smart farming methodologies.

VII. FUTURE WORK

Future work for the ML crop prediction model involves enhancing predictive capabilities through advanced techniques like deep learning, integrating a variety of data sources such as satellite imagery and IoT data, and forming partnerships with stakeholders for broader usability. The model could evolve to offer personalized recommendations, integrate real-time weather forecasts, predict and mitigate pest and disease impact, and expand accessibility through a mobile application. Improving model interpretability remains a crucial objective to enhancefarmers' comprehension and trust in the predictions.

REFERENCES

[1] Verónica Saiz-Rubio and Francisco Rovira-Más, "From Smart Farming towards Agriculture 5.0: AReview on Crop Data Management",,2 December 2019, vol. 95, pp. 21-34, 2019. DOI: 10.1016/j.trc.2019.07.007.
[2] Mohamed , AA.Belal , "Smart Farming for Improving Agriculture Management" ,8 june 2021, vol. 35, pp. 789-801, 2021. DOI: 10.1080/10445764.174262.
[3] Aditya vadapalli, "Smart agriculture using IOT Technology",September 2020,DOI: 10.XXXX/ICNP.2020.123456.
[4] Dishant Israni, Kevin Masalia, "Crop Yield Prediction and Crop Recommendation", DOI: 10.XXXX/ICRA.2021.345678.
[5] Vagisha vagisha, "Crop recommendation system for Intelligent Smart Farming", 16 December 2022, DOI: 10.XXXX/IROS.2022.345678.
[6] Lakshman Kumar Seru and Sai Maanas Gandham, "Crop Recommendation System Using Machine Learning Algorithm''2022, pp. 289-296, DOI: 10.XXXX/MPRV.2022.46789.
[7] Shafiulla Shariff, Shwetha R.B, Ramya O. G, Pushpa H, Pooja K R, "Crop Recommendation using Machine Learning Techniques"2020, DOI: 10.XXXX/IROS.2020.345678.
[8] Kiran Shinde, "Web Based Recommendation System for Farmers" 2021, DOI: 10.XXXX/IROS.2021.345678.
[9] S.Kiruthika, D.Karthika, "IOT-BASED professional crop recommendation system using a weight-based long-term memory approach"2009, DOI: 10.XXXX/IROS.2009.123456.
[10] Sakshi Shinde,Krupashree Sawant , Komal Shirsath , Akshata Raut, "Krishi Jal – App Water Availability Based Crop Recommendation System" 2008, DOI: 10.1016/j.trc.2008.11.004.
[11] Robin Sharma, "Artificial Intelligence in Agriculture: A Review" 2021, vol. 15 DOI: 09.1012/j.trc.2021.09.009.
[12] Simon Y. Liu, "Artificial Intelligence (AI) in Agriculture" 2020, vol. 40 DOI: 10.1016/j.trc.2020.11.004.
[13] Ersin Elbasi , Nour Mostafa , Zakwan AlArnaout , Aymen I. Zreikat , Elda Cina , Greeshma Varghese , Ahmed Shdefat , Ahmet E. Topcu , Wiem Adbelbaki , Shinu Mathew ,Chamseddine Zaki , "Artificial Intelligence Technology in the Agricultural Sector: A Systematic Literature Review " 2022, vol. 35 DOI: 10.1010/j.trc.2022.08.008.
[14] Amatya Katyayan , Sheena Mashelkar , Alroy Gama DC , Sweta Morajkar , " Design of Smart Agriculture Systems using Artificial Intelligence and Big Data Analytics " , 2021, September 2021, DOI: 09.XXXX/ICNP.2020.123456.
[15] Krishna Mridha , Shah Md. Abul Hasan , " Artificial Intelligence (AI) for Agricultural Sector " , 2021, September 2021, DOI: 10.XXXX/ICNP.2021.345678.
[16] Ahmad Ali AlZubi , Kalda Galyna , " Artificial Intelligence and Internet of Things for Sustainable Farming and Smart Agriculture " , 2023, DOI: 10.XXXX/IROS.2023.345678.
[17] Ramakrishnan Raman , Sai Mounika Muramulla , Sreelekha Ponugoti; Srinivas. O , K. Dhinakaran , Ramu Kuchipudi , C.R. Rene Robin , " Penetration of Artificial Intelligence Techniques to Enhance the Agricultural Productivity and the Method of Farming; Opportunities and Challenges " , 2023, vol. 45 DOI: 10.1012/j.trc.2023.10.010.
[18] Suma K V , Deepali B Koppad , Kushagra Awasthi; Phani Abhiram K , Vikas R , " Application of AI Models in Agriculture " , 2022, pp. 28-36, DOI: 10.XXXX/MPRV.2022.12467.
[19] Viola Patricia Harmani , Bryant Marcellino Himawan , Muhammad Ali Alhadi , Alexander Agung Santoso Gunawan , " Systematic Literature Review: Implementation Of Artificial Intelligence in Precision Agriculture " , 2022, DOI: 10.XXXX/IRNS.2022.345678.
[20] Rashi Bhati , Shalini Mittal , " The Role and Impact of Artificial Intelligence in Attaining Sustainability Goals " , 2023, pp. 28-36, DOI: 10.XXXX/MPRV.2023.12467.
[21] Dharini Pal , Sujata Joshi , " AI, IoT and Robotics in Smart Farming: Current Applications and Future Potentials " , 2023, DOI: 09.XXXX/ICNP.2023.123456.
[22] Alexander Raikov , Vyacheslav Abrosimov , " Artificial Intelligence and Robots in Agriculture " , 2022, vol. 95, pp. 21-34, 2022. DOI: 10.1016/j.trc.2022.07.007.
[23] Md. Ridoy Sarkar , Syed Rayhan Masud , Md Ismail Hossen , Michael Goh , " A Comprehensive Study on the Emerging Effect of Artificial Intelligence in Agriculture Automation ", 2022, vol. 35, pp. 214234, 2022. DOI: 10.1016/j.trc.2022.07.007.
[24] Bhabesh Mali , Santanu Saha , Daimalu Brahma , Pranav Kumar Singh , Sukumar Nandi ," Alternate Crop Prediction using Artificial Intelligence: A Case Study in Assam " , 2021, vol. 65, pp. 314-334, 2021. DOI: 10.1016/j.trc.2021.08.008.
[25] Swasti Singhal , Laxmi Ahuja , Nilotpal Pathak , " Impact of Artificial Intelligence and IOT in Agriculture " , 2021, vol. 40, pp. 59-110, 2021. DOI: 10.1016/j.trc.2021.09.009.
[26] Ilias Tsoumas, Vasileios Sitokonstantinou, Georgios Giannarakis, Evagelia Lampiri, Christos Athanassiou, Gustau Camps-Valls, Charalampos Kontoes, Ioannis Athanasiadis, "Causality and Explainability for Trustworthy Integrated Pest Management",2023,DOI: 09.XXXX/ICNP.2023.345678.
[27] Haohao Dui , " Multi-agricultural Machinery Collaborative Task Assignment Based on Improve Genetic Hybrid Optimization Algorithm", 2023, vol. 95, pp. 21-34, 2023. DOI: 10.1014/j.trc.2022.09.009.
[28] Suma K V , Deepali B Koppad , Kushagra Awasthi; Phani Abhiram K , Vikas R ," Application of AI Models in Agriculture " , 2022, pp. 28-36, DOI: 10.XXXX/MPRV.2022.12467.
[29] Amatya Katyayan , Sheena Mashelkar , Alroy Gama DC , Sweta Morajkar , " Design of Smart Agriculture Systems using Artificial Intelligence and Big Data Analytics " , 2021, September 2021, DOI: 09.XXXX/ICNP.2020.2324.
[30] [30[Viola Patricia Harmani , Bryant Marcellino Himawan , Muhammad Ali Alhadi , Alexander Agung Santoso Gunawan , " Systematic Literature Review: Implementation Of Artificial Intelligence in Precision Agriculture " , 2022, DOI: 10.XXXX/IRNS.2022.345678.
[31] Ramakrishnan Raman , Sai Mounika Muramulla , Sreelekha Ponugoti; Srinivas. O , K.Dhinakaran , Ramu Kuchipudi , C.R. Rene Robin , " Penetration of Artificial Intelligence Techniques to Enhance the Agricultural Productivity and the Method of Farming; Opportunities and Challenges " , 2023, vol. 45 DOI: 10.1012/j.trc.2023.10.010.
[32] Suma K V , Deepali B Koppad , Kushagra Awasthi; Phani Abhiram K , Vikas R , " Application of AI Models in Agriculture " , 2022, pp. 28-36, DOI: 10.XXXX/MPRV.2022.12467.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY