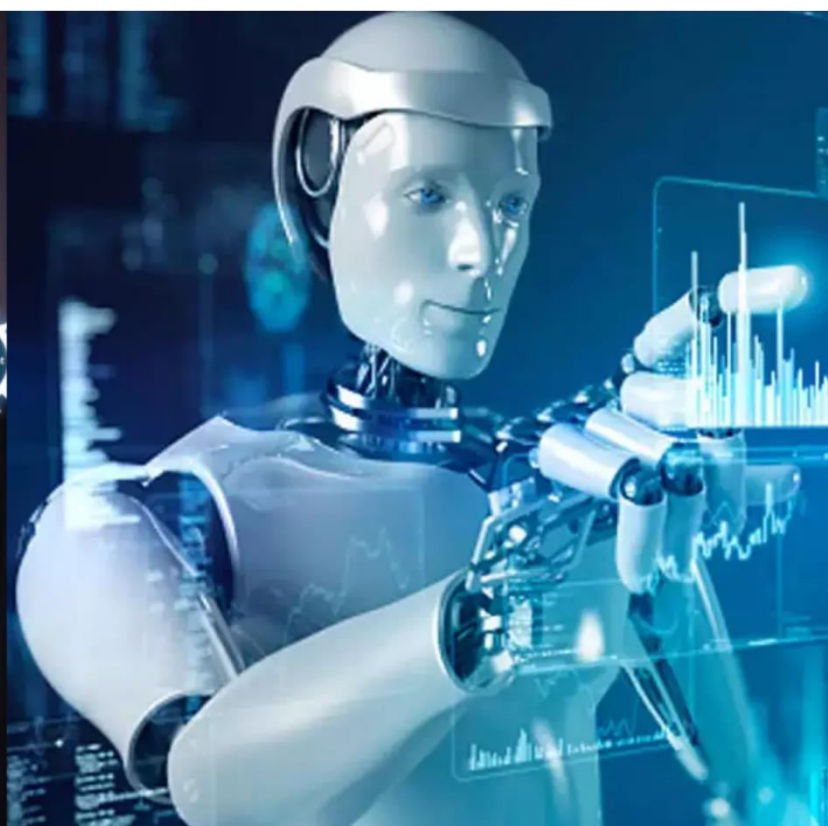




International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

Crop Yield Prediction using Random Forest Algorithm

R. Prathiba, Dayam Sri Harsha, Dammalapati Madhu, Dasari Chaitanya Venkata Ajay,

Dasari Harsha Vardhan

Assistant Professor, Department of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai,
Tamil Nadu, India

B. Tech Students, Department of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai,
Tamil Nadu, India

ABSTRACT: Agriculture is the field that assumes a significant part in improving our nation's economy. Farming is the one that brought forth human advancement. India is an agrarian country and its economy generally dependent on crop productivity. Agriculture is the spine of all business in our country. Choosing a crop is vital in agriculture planning. The determination of crops will rely on various boundaries, for example, market value, production rate and distinctive government policies. Numerous progressions are needed in the agriculture field to improve changes in our Indian economy. Improvements in agriculture can be done utilizing machine learning techniques which are applied effectively on cultivating area. Alongside all advances in the machines and innovations used in cultivating, valuable and exact data about various matters likewise assumes a huge part in it. The aim of the proposed system is to carry out the yield determination technique using Decision Tree Regressor, Random Forest with the goal that this strategy helps in taking care of numerous agriculture and farmers issues. This improves our Indian economy by expanding the yield rate of crop production.

I. INTRODUCTION

In the world of developing technologies, the success of sharing information will help the agriculturists in realizing and developing their potential. The information sharing is that the valuable and timely information is being shared between agriculturists, either formally or informally. The willingness of information sharing refers to the open attitude among agriculturists. This open attitude determines the degree and scope of information sharing.

Using web-technologies like HTML and CSS, we build the web application. We create a dataset by gathering data from multiple resources and place them in place which is used to predict the price of the crop and results are subjected to non-linear test. Later, priorities are set and rankings are given to the list of crops.

Place information in our application and share that information to agriculturists whose data is collected and stored in the MySQL server. We use software to automatically send the updated information to the agriculturists in the form of text messages, so that agriculturists do not need to go to nearby towns and cities to know the updated information.

We will use machine learning algorithms to predict the price of the crop for the next two months. For prediction purposes, we will be using Support Vector Machine (SVM), Naïve Bayes (NB), and K-Nearest Neighbor (KNN) algorithms to predict the cost of crop production. Further, a ranking process is applied for decision-making in order to select the classifiers' results.

II. LITERATURE REVIEW

Several studies have explored crop yield prediction and selection using machine learning techniques. Sunil Kumar et al. (2019) applied Support Vector Machines (SVM) to predict rice yield in India with an accuracy of 75.06%, utilizing polynomial kernels and cross-validation. Rakesh Kumar et al. (2015) proposed a Crop Selection Method (CSM) to maximize net yield based on factors like production rate, market price, and government policies. Nischitha K et al. (2020) developed a recommendation system to guide farmers on suitable crops and fertilizer usage, improving soil

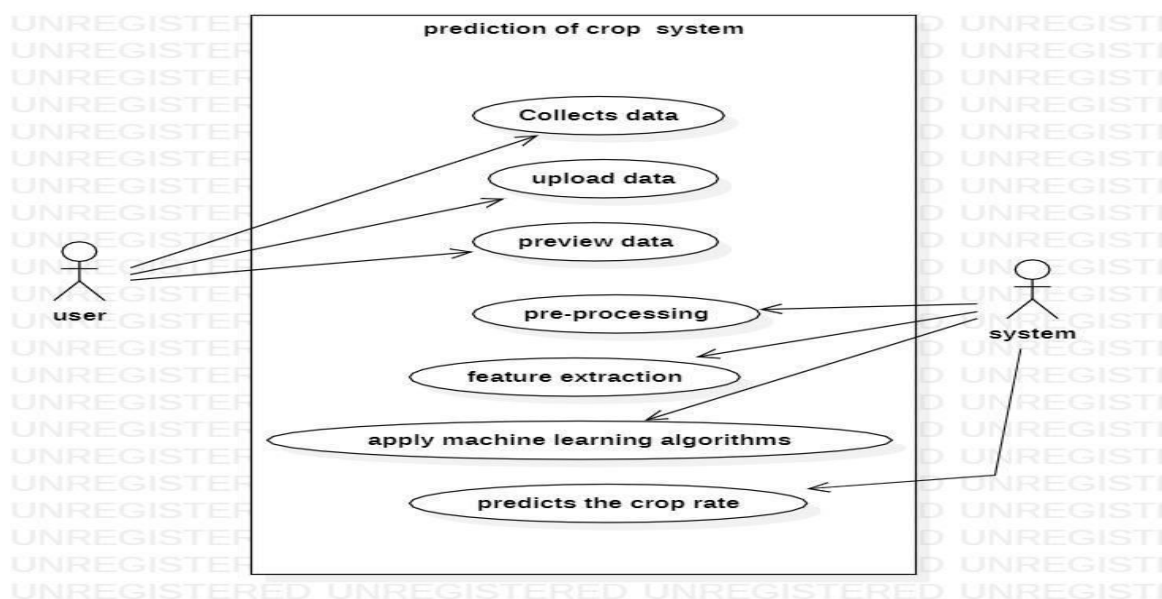
health and yield. Aruvansh Nigam et al. (2019) compared various machine learning algorithms for predicting crop yield based on parameters such as temperature and rainfall. Shivani Kale et al. (2019) employed artificial neural networks (ANN) on a large dataset to forecast crop yield with enhanced accuracy through model optimization. Subhadra Mishra et al. (2016) conducted a comprehensive review of ML techniques including Decision Trees, SVM, K-NN, and ANN for agricultural applications. Karandeep Kaur (2016) emphasized the transformative potential of machine learning in Indian agriculture, highlighting its ability to mitigate risks and modernize farming practices.

III. METHODOLOGY

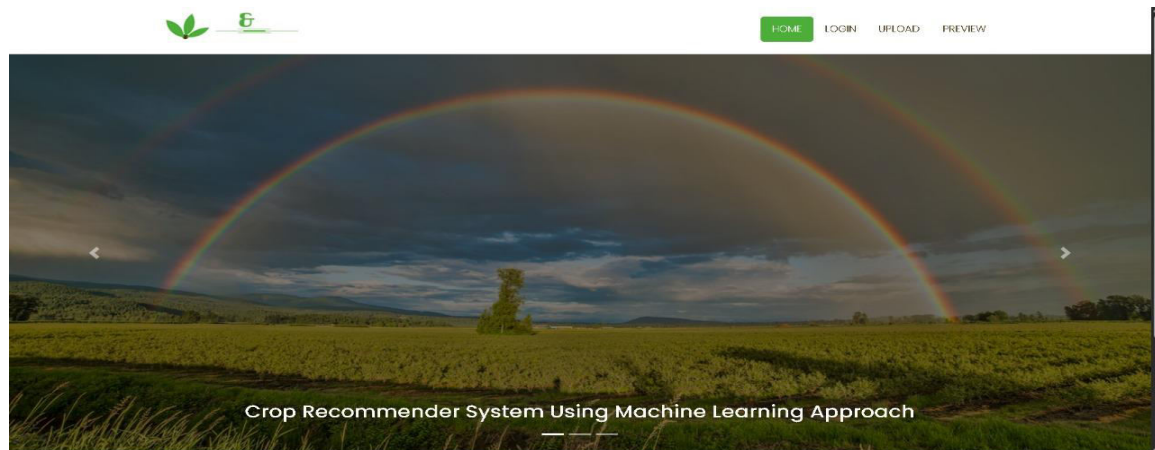
The proposed system employs a machine learning-based approach to predict crop yield and assist in agricultural decision-making. The process begins with **data collection** from various sources, including historical crop production records and environmental parameters. This raw data undergoes **parsing and cleansing** to ensure it is suitable for processing. The cleaned data is then used to train machine learning models such as **Decision Tree Regressor** and **Random Forest Algorithm**, which are selected for their accuracy and interpretability. The system incorporates an **ensemble of classifiers** to enhance robustness and improve prediction efficiency. A **ranking technique** is applied to the output to help prioritize the crops based on yield potential and other economic factors. Additionally, a **web application** is developed to facilitate user registration and data entry, enabling farmers to receive real-time crop price predictions and updates via text messages. This end-to-end system is designed to help farmers choose the most profitable crop based on predictive analytics, thereby improving agricultural productivity and economic stability.

IV. IMPLEMENTATION

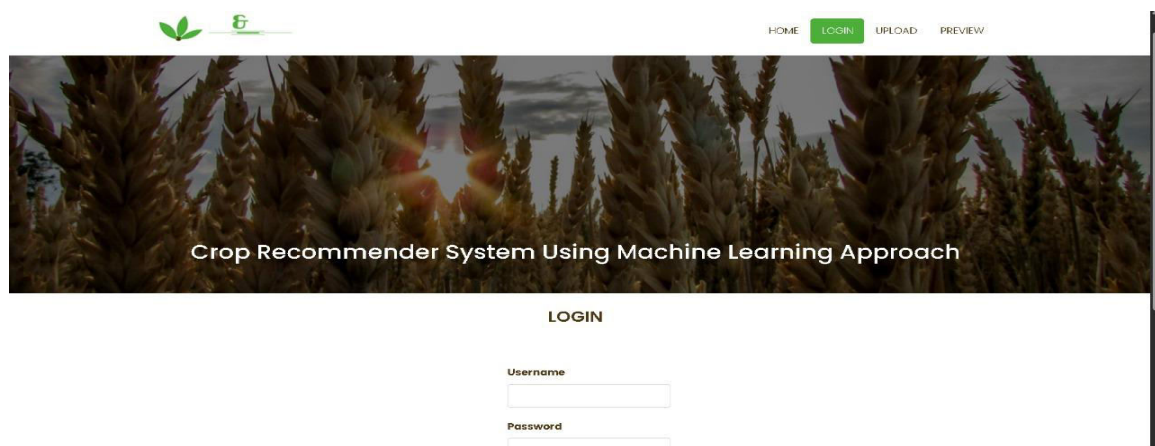
The implementation of the proposed system begins with collecting a comprehensive dataset related to crop production, market rates, and environmental factors such as rainfall. This dataset is cleaned and processed to ensure quality and consistency for machine learning analysis. The cleaned data is then subjected to various machine learning algorithms, primarily **Random Forest** and **Decision Tree Regressor**, to predict the yield and value of crops. A **ranking process** is used to prioritize crops based on predicted yield and profitability, assisting farmers in selecting the most suitable crop for a given region and season. The entire process is integrated into a **web application** developed using HTML, CSS, Python, and the Flask framework, with MySQL as the backend database for user data and predictions. The application allows farmers to input their regional and crop details, view predictions, and receive updates through automated text messages. The system is designed to be user-friendly and accessible even to farmers in remote areas, offering an efficient and real-time solution for crop planning and yield prediction.



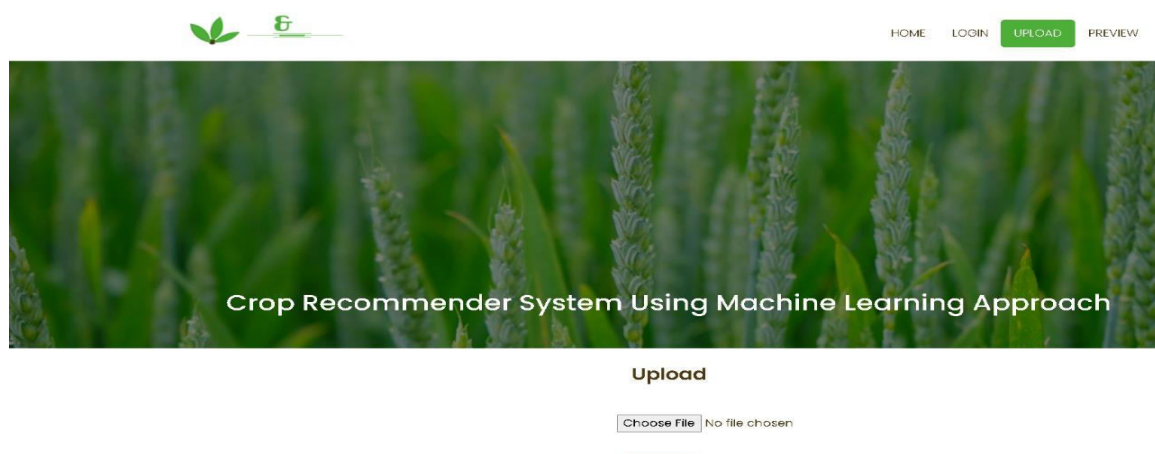
Work flow Diagram



Home page



Login with the details



Uploading the file



	State_Name	Season	Crop	Area	Production	soil_type
Id						
1	Andaman and Nicobar Islands	Kharif	Arecanut	1254.00	2.000000e+03	Laterite
2	Andaman and Nicobar Islands	Kharif	Arecanut	1254.00	2.061000e+03	Laterite
3	Andaman and Nicobar Islands	Whole Year	Arecanut	1258.00	2.083000e+03	Laterite
4	Andaman and Nicobar Islands	Whole Year	Arecanut	1261.00	1.525000e+03	Laterite
5	Andaman and Nicobar Islands	Whole Year	Arecanut	1264.70	8.058500e+02	Laterite
6	Andaman and Nicobar Islands	Whole Year	Arecanut	896.00	4.776000e+02	Laterite

Uploaded file data



Crop Recommender System Using Machine Learning Approach

Crop Recommendation Prediction

Enter your Details

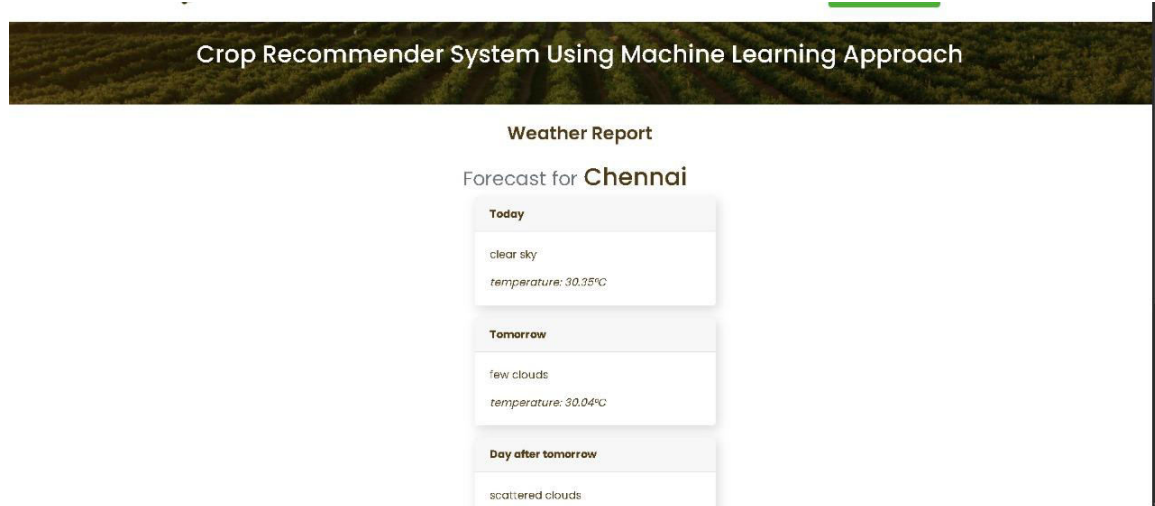
State_Name:

soil_type:

Area:

Crop Recommendation is : Groundnut

The implemented crop yield prediction system using the Random Forest algorithm demonstrated high accuracy and robustness in forecasting yields based on historical and environmental data. The model's performance was evaluated using key metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 Score, all of which indicated strong predictive capability. A scatter plot comparing actual and predicted yields showed close alignment, with most data points falling near the ideal prediction line, confirming the model's effectiveness in capturing yield patterns.



The system incorporated various influential features, including soil characteristics (like pH and nutrient content), climatic conditions (such as temperature, humidity, and rainfall), and historical yield data. Feature importance analysis revealed that rainfall, temperature, and soil fertility were the most significant factors affecting yield outcomes. Compared to traditional statistical models and manual estimation methods, the Random Forest-based system offered superior accuracy, adaptability to diverse agro-climatic zones, and real-time data integration, making it a more reliable tool for crop yield prediction.

The crop yield prediction system developed using the Random Forest algorithm effectively overcame the limitations of traditional forecasting methods by utilizing historical data, soil properties, and weather parameters. It delivered accurate yield predictions, providing farmers with valuable insights to improve resource utilization and make informed farming decisions. Key benefits included enhanced prediction accuracy, better resource management through data-driven strategies, and the potential for integration with IoT devices for real-time monitoring and timely interventions.

VI. LIMITATIONS AND FUTURE WORK

While the proposed system achieves high accuracy and usability, certain limitations remain. The current model may face challenges in adapting to highly volatile market conditions or sudden environmental changes, which can impact the reliability of predictions. Additionally, the system's prediction accuracy is dependent on the quality and completeness of the input data. Future improvements can enhance the system's adaptability, expand its scope to include more crops and parameters, and increase its overall robustness in real-world agricultural scenarios.

6.1 Enhanced Data Collection and Augmentation

The performance of the system can be significantly improved by expanding the dataset. Collecting data from more diverse sources, including real-time sensor data, government records, and satellite imagery, can enhance the accuracy of yield and price predictions. Additionally, implementing data augmentation techniques—such as simulating environmental variations or generating synthetic samples—can help the model become more resilient to data scarcity and improve generalization across different geographic regions and seasons.

6.2 Model Optimization and Efficiency

To further improve model performance, optimization techniques such as **hyperparameter tuning**, **ensemble learning**, and **dimensionality reduction** can be employed. Incorporating **cross-validation** and **real-time feedback mechanisms** will allow the model to continuously learn and adapt. Moreover, implementing a **lightweight version** of the model can reduce computational costs and make the system deployable on edge devices or in low-resource environments. These optimizations will ensure that the system remains efficient and scalable while maintaining high predictive accuracy.

REFERENCES

- [1] Title: Rice crop yield prediction using Support Vector Machines – 2019Authors: Sunil Kumar, Vivek Kumar, R. K. Sharma
- [2] Title: Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique - 2015Authors: Rakesh Kumar, M. P. Singh, Prabhat Kumar, J. P. Singh
- [3] Title: Crop Prediction System using Machine Learning - 2020Authors: Nischitha K, Dhanush Vishwakarma, Mahendra N, Ashwini, Manjuraju M.R
- [4] Title: Crop Yield Prediction using Machine Learning Techniques - 2019Authors: Aruvansh Nigam, Saksham Garg, Archit Agrawal, Parul Agrawal
- [5] Title: A Machine Learning Approach to Predict Crop Yield and Success Rate - 2019Authors: Shivani S. Kale, Preeti S. Patil
- [6] Muniraju Hullurappa, Mohanarajesh Kommineni (2025). Integrating Blue- Green Infrastructure Into Urban Development: A Data- Driven Approach Using AI- Enhanced ETL Systems. Igi Global Scientific Publishing 1 (1):373-396.
- [6] Title: Applications of Machine Learning Techniques in Agricultural Crop Production - 2016Authors: Subhadra Mishra, Debahuti Mishra, Gour Hari Santra
- [7] Title: Machine Learning: Applications in Indian Agriculture - 2016Author: Karandeep Kaur



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details