International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Smart Crop Recommendation Using Machine Learning for Precision Agriculture

[1]Ramathirtham Vaishnavi ,[2]Tejavath Sravya,[3]Shaik Vaseem , [4]Dr. K. Madan Mohan
[1,2,3]UG Scholars, [4]Associate Professor
[4]drmohan.csegnit@gniindia.org
[1,2,3,4]Department of Computer Science and Engineering
[1,2,3,4]Guru Nanak Institute of Technology, Hyderabad, Telangana, India.

Abstract - Agriculture and its allied sectors are undoubtedly the largest providers of livelihoods in rural India. The agriculture sector is also a significant contributor factor to the country's Gross Domestic Product (GDP). Blessing to the country is the overwhelming size of the agricultural sector. However, regrettable is the yield per hectare of crops in comparison to international standards. This is one of the possible causes for a higher suicide rate among marginal farmers in India. This paper proposes a viable and user-friendly yield prediction system for the farmers. The proposed system provides connectivity to farmers via a mobile application. GPS helps to identify the user location. The user provides the area & soil type as input. Machine learning algorithms allow choosing the most profitable crop list or predicting the crop yield for a user selected crop. To predict the crop yield, selected Machine Learning algorithms such as Support Vector Machine (SVM), Artificial Neural Network (ANN), Random Forest (RF), Multivariate Linear Regression (MLR), and K-Nearest Neighbour (KNN) are used. Among them, the Random Forest showed the best results with 95% accuracy. Additionally, the system also suggests the best time to use the fertilizers to boost up the yield.

Key Words: Machine Learning, Random Forest, Support Vector Machine (SVM), Artificial Neural Network (ANN), Multivariate Linear Regression (MLR), K-Nearest Neighbour (KNN)

I. INTRODUCTION

An Agriculture has an extensive history in India. Recently, India is ranked second in the farm output worldwide [15]. Agriculture-related industries such as forestry and fisheries contributed for 16.6% of 2009 GDP and around 50% of the total workforce. Agriculture's monetary contribution to India's GDP is decreasing [1]. The crop yield is the significant factor contributing in agricultural monetary. The crop yield depends on multiple factors such as climatic, geographic, organic, and financial elements [6]. It is difficult for farmers to decide when and which crops to plant because of fluctuating market prices [7]. Citing to Wikipedia figures India's suicide rate ranges from 1.4-1.8% per 100,000 populations, over the last10 years [15]. Farmers are unaware of which crop to grow, and what is the right time and place to start due to uncertainty in climatic conditions. The usage of various fertilizers is also uncertain due to changes in seasonal climatic conditions and basic assets such as soil, water, and air. In this scenario, the crop yield rate is steadily declining [2]. The solution to the problem is to provide a smart user-friendly recommender system to the farmers.

The crop yield prediction is a significant problem in the agriculture sector [3]. Every farmer tries to know crop yield and whether it meets their expectations [4], thereby evaluating the previous experience of the farmer on the specific crop predict the yield [3]. Agriculture yields rely primarily on weather conditions, pests, and preparation of harvesting operations. Accurate information on crop history is critical for making decisions on agriculture risk management [5].

II. EXISTING SYSTEM:

The system predicts crop yields using inputs like state, district, season, and area. It combines machine learning techniques like Kernel Ridge, Lasso, ENet, and stacking regression, with Random Forest achieving 95% accuracy. It includes a Predictor System to estimate yields based on soil type and area, and a Recommender System to suggest optimal crops.

Data preprocessing involves filtering NA values and using robust scaling. The stacked regression approach improves predictions by combining multiple models, feeding them into a meta-model for enhanced accuracy.

DISADVANTAGES:

i Overfitting Risk

ii Data Dependency

iii Limited Adaptability

III. PROPOSED SYSTEM:

The system predicts crop yields in India using simple parameters like State, District, Season, and Area. It employs advanced regression techniques such as Kernel Ridge, Lasso, and ENet, and enhances prediction accuracy using Stacking Regression. The system includes an inbuilt Predictor to estimate yield and a Recommender to help farmers select the best crops. Various machine learning algorithms, including Random Forest, ANN, SVM, MLR, and KNN, were tested on datasets from Maharashtra and Karnataka. Among these, the Random Forest algorithm achieved the best accuracy of 95%.

ADVANTAGES:

i Works well with large datasets

ii High accuracy with ensemble learning approach

iii Can be used for both classification and regression

IV. LITERATURE SURVEY:

A) TITLE: BIoT: Blockchain-based IoT for Agriculture

AUTHORS: Uma Maheswari S, Sreeram S, Kritika N, Prasanth DJ

YEAR:2019

Summary: This paper proposes using blockchain and IoT in agriculture to improve trust and transparency. Blockchain allows direct (peer-to-peer) transactions without intermediaries. Smart contracts on Ethereum automate buying and selling of crops and land. It helps farmers get real-time data on seed quality, climate, payments, soil, and prices—everything on one secure platform.

B) TITLE: Analysis of growth and instability in the area, production, yield, and price of rice in India

AUTHOR: Jain A.

YEAR: 2018

Summary: This study looks at 41 years of data (1970–2012) on rice area, production, yield, and price in India. While growth was positive, it became less stable in recent years, especially from 2000–2012. Reasons include less irrigation, fewer seeds/fertilizers used, and other input shortages. Price instability increased after economic reforms, especially in wholesale prices.

C) TITLE: A model for prediction of crop yield.

AUTHOR: Manjula E, Djodiltachoumy S

YEAR: 2017.

Summary: In this chapter, we give an overview of the main Data Mining techniques used in the context of Recommender Systems. We first describe common preprocessing methods such as sampling or dimensionality reduction. Next, we review the most important classification techniques, including Bayesian Networks and Support Vector Machines. We describe the k-means clustering algorithm and discuss several alternatives. We also present association rules and related algorithms for an efficient training process. In addition to introducing these techniques, we survey their uses in Recommender Systems and present cases where they have been successfully applied.

V. METHODOLOGIES

MODULES NAME:

1. Data Collection
2. Dataset
3. Data Preparation
4. Model Selection
5. Analyse and Prediction
6. Accuracy on test set
7. Saving the Trained Model

VI. MODULES EXPLANATION

1. Data Collection:

The first step in building a machine learning model is data collection. In this project, data is collected through techniques like web scraping and manual entry. The dataset used, named *crop_yield.csv*, contains 94,375 records. Quality and quantity of data significantly affect the model's performance.

2. Dataset

The dataset includes six columns: State Name, Season, Crop, Area, Production, and Soil Types. These features cover critical aspects of agriculture across India's 29 states. Seasons are categorized as Winter, Summer, Monsoon, Post-monsoon, and Whole Year, capturing varied cropping patterns.

3. Data Preparation

Data preparation involves cleaning and transforming the dataset. Unnecessary columns are removed, and missing values are dropped to ensure data quality. Only relevant features are retained for training the model, such as State Name, Season, Soil Types, and Area.

4. Model Selection

The data is split into training and testing sets in an 80:20 ratio using the train_test_split function from sklearn. For classification tasks, the **Random Forest Classifier** is used, and for regression tasks like predicting crop yield, the **Random Forest Regressor** is implemented. These models build multiple decision trees and aggregate their results for improved accuracy and robustness.

5. Analyze and Prediction

The model analyzes the relationship between features such as soil type, area, and crop to make accurate predictions. It identifies patterns and uses them to recommend suitable crops or predict yield. Feature importance is evaluated to determine which attributes contribute most to the prediction.

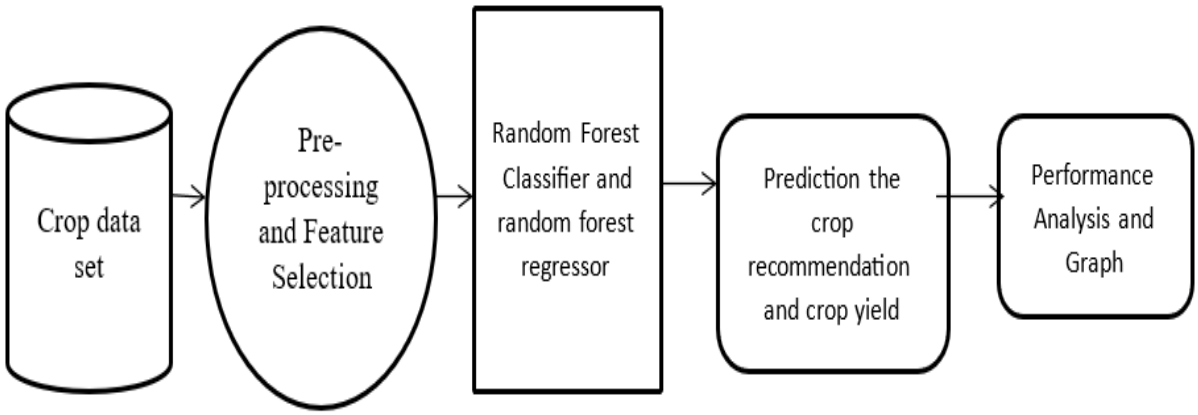
6. Accuracy on test set

The model achieves an accuracy of **87%** for crop recommendation and **96%** for crop yield prediction. This indicates strong performance and reliability in real-world agricultural applications.

7. Saving the Trained Model

Once the model is trained and validated, it is saved using the **pickle** library in a .pkl file. This allows the model to be reused later without needing to retrain it, making it production-ready

VII. SYSTEM ARCHITECTURE



VII. CONCLUSION

This paper highlighted the limitations of current systems and their practical usage on yield prediction. Then walkthrough a viable yield prediction system to the farmers, a proposed system provides connectivity to farmers. The inbuilt predictor system helps the farmers to predict the yield of a given crop. The inbuilt recommender system allows a user exploration of the possible crops and their yield to take more educated decisions. For yield to accuracy, machine learning algorithm were implemented and tested on the given datasets. The proposed model also explored the timing of applying fertilizers and recommends appropriate duration.

VIII. REFERENCES

[1]. Uma Maheswari S, Sreeram S, Kritika N, Prasanth DJ, "BIoT: Blockchain-based IoT for Agriculture", 11th International Conference on Advanced Computing (ICoAC), 2019 Dec 18 (pp. 324-327). IEEE.

[2] Jain A. "Analysis of growth and instability in the area, production, yield, and price of rice in India", Journal of Social Change and Development,2018;2:46-66.

[3] Manjula E, Djodiltachoumy S, "A model for prediction of crop yield "International Journal of Computational Intelligence and Informatics,2017 Mar;6(4):2349-6363.

[4] Sagar BM, Cauvery NK., "Agriculture Data Analytics in Crop Yield Estimation: A Critical Review", Indonesian Journal of Electrical Engineering and Computer Science, 2018 Dec;12(3):1087-93.

[5] Wolfert S, Ge L, Verdouw C, Bogaardt MJ, "Big data in smart farming–a review. Agricultural Systems", 2017 May 1;153:69-80.

[6] Jones JW, Antle JM, Basso B, Boote KJ, Conant RT, Foster I, GodfrayHC, Herrero M, Howitt RE, Janssen S, Keating BA, "Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science, Agricultural systems", 2017 Jul1;155:269-88.

[7] Johnson LK, Bloom JD, Dunning RD, Gunter CC, Boyette MD, Creamer NG, "Farmer harvest decisions and vegetable loss in primary production. Agricultural Systems", 2019 Nov 1;176:102672.