International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 06 | June - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

CROP RECOMMENDATION USING XGBOOST ALGORITHM

D. Sunil[1], B. Udaysantoshkumar[2], K. Nandini[3], G. Pooja Rao[4], A. Sai Deepak[5]

[1]*Assistant Professor,*
[2-5] *B. Tech Student, LIET*
[1,2,3,4,5] Computer Science & Engineering, Lendi Institute of Engineering and Technology, Vizianagaram

---

Abstract - Precision agriculture, a modern farming approach leveraging soil characteristics, weather data, and crop yield information, minimizes crop failures and aids farmers in strategic decision-making for optimal yields and profits. To address the agrarian crisis, this project focuses on enhancing recommendation systems for informed crop cultivation decisions. Using an augmented dataset from various Indian sources, including soil and weather information, the study employs the XGBoost algorithm to predict the most suitable crops. Unlike the baseline Random Forest model with 87% accuracy, our proposed XGBoost model achieves an impressive 99.32% accuracy, showcasing its superior performance. This project underscores the transformative impact of advanced machine learning techniques in revolutionizing crop recommendation accuracy for farmers, ultimately contributing to sustainable and profitable agriculture.

***Key Words*:** *XGBoost*

1. INTRODUCTION

Agriculture plays an essential part in an economy's life. They are the backbone of our country's economic system. One of the key problems confronting farmers is selecting the right crop for cultivation. Selection of crops is determined by several factors such as temperature, soil composition, market prices etc. Yield forecasting is a very important issue in agriculture. Any farmer is interested in knowing how much yield he is about to expect and what is the crop that is suitable for the land. Analyze the various related attributes like location, pH value from which alkalinity of the soil is determined. Along with it, the percentage of nutrients like Nitrogen (N), Phosphorous (P), and Potassium (K).

2. Existing System

Maharashtra is a leading State in agriculture. Agriculture is the one that plays important role in the economy of India. India is an agricultural country and its economy largely based upon crop production. Hence one must say that agriculture is often the backbone of all businesses in the a-part-of-us country. Basically, paper focuses on predicting the yield of the crop by using a different machine learning algorithm. The application (Smart Farm) developed in this research helps users to predict the crop yield using different climatic parameters. Different methods of predicting crop yield are developed within several years with different outcomes of success, but many of them do not take into consideration the climate. Machine Learning is the best technique which gives a better practical solution to crop yield problem. So, the Random Forest algorithm which we decided to used to train our model to give high accuracy and best prediction., we chose 5 climatic parameters to train the model. Agriculture inputs such as pesticides, fertilizers, chemicals, soil quality, etc. were not used as it depends upon the type of field. The model is trained and designed using 20 decision trees build the random forest algorithm which gives better accuracy of the model. 10-fold cross-validation technique used to improve the accuracy of the model. The predicted accuracy of the model is analyzed 87%.

Challenges:

Limited Feature Set
City-Specific Dataset.
Temperature-Centric Focus.
Temporal Constraints.
Algorithm-Specific Limitations.

3. Proposed Algorithm

XGBoost (Extreme Gradient Boosting) is a powerful ensemble learning algorithm that belongs to the family of boosting algorithms. While it's not typically referred to as a "voting classifier" like the one you described, it shares some similarities in terms of ensemble learning.
Here's how the XGBoost algorithm works:

Gradient Boosting: XGBoost works by building an ensemble of weak learners, typically decision trees, sequentially. Each new tree is trained to correct the errors made by the ensemble of trees built so far. This process is called gradient boosting.

Loss Function Optimization: XGBoost minimizes a specific loss function during training, which measures the difference between the actual target values and the predicted values. It uses gradient descent optimization to iteratively minimize this loss function.

Regularization: XGBoost incorporates both L1 (Lasso) and L2 (Ridge) regularization techniques to control the complexity of the model and prevent overfitting. Regularization helps in improving the generalization capability of the model.

Tree Pruning: XGBoost applies tree pruning techniques to control the growth of individual decision trees within the ensemble. Pruning removes unnecessary branches of a tree, reducing its complexity and improving its predictive power.

Feature Importance: XGBoost provides a mechanism to assess the importance of features in the dataset. By analyzing how frequently each feature is used in the ensemble of trees and how much it contributes to reducing the loss function, XGBoost assigns an importance score to each feature.

Parallel and Distributed Computing: XGBoost is designed to be highly scalable and efficient. It supports parallel and distributed computing, allowing it to handle large datasets and leverage multicore CPUs or distributed computing frameworks like Apache Spark.

4. System Architecture

Data Collection: This involves collecting data from various sources such as weather stations, satellite imagery, soil sensors, and historical crop yield data.

Data Preprocessing: The collected data needs to be cleaned, transformed, and standardized before it can be used for analysis. This includes tasks such as data normalization, outlier removal, missing value imputation, and data aggregation.

Crop Forecasting: The next step is to use machine learning or statistical models to predict the crop yield for the upcoming season. These models typically use historical crop yield data, weather data, soil data, and other relevant factors to make accurate predictions.

Fertilizer Recommendation: Based on the predicted crop yield and soil nutrient levels, the system can recommend the optimal amount and type of fertilizer to be used for each crop.

Visualization and Reporting: The system can generate visualizations such as charts and maps to help farmers and agricultural experts better understand the data and recommendations. Reports can also be generated to summarize the results of the analysis and provide actionable insights for decision-making.

Integration: The system architecture can also include integration with other agricultural tools such as irrigation systems and crop management software to provide a comprehensive solution for farmers.

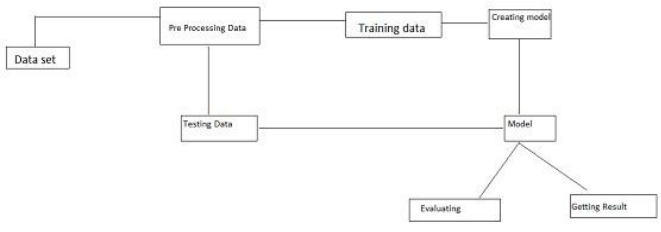

Fig -1: System Architecture

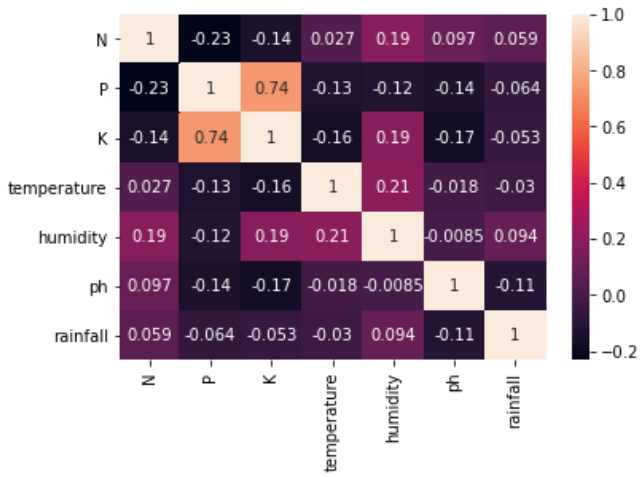

Fig -2

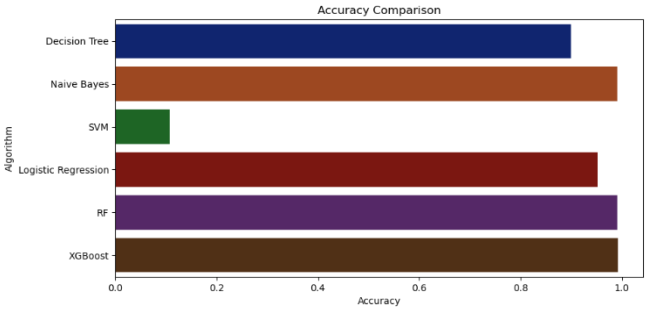

Fig -3: Accuracy Comparision

5. CONCLUSIONS

A farmer was curious about how much output he should expect during the rising season. Previously, this yield estimate was based on the farmer's long-term expertise with specific yields, crops, and meteorological circumstances. Instead of worrying about crop forecasts, farmers go straight for yield forecasting with the current technique.A recommendation model which may be used for future recommendation of crop .

ACKNOWLEDGEMENT

We would like to thank the Department of Computer Science & Engineering Lendi Institute of Engineering and Technology, Vizianagaram for helping us to carry out the work and supporting us throughout the research.

REFERENCES

1. Aruvansh Nigam,Saksham Garg,Archit Agrawa,*Parul Agrawal,Crop Yield Prediction Using Machine Learning Algorithms",2019 Fifth International Conference on Image Information Processing (ICIIP),pp 125-130

2. Rakesh Kumar1, M.P. Singh2, Prabhat Kumar3 and J.P. Singh4,"2015 International Conference on Smart Technologies and Management for Computing,

Communication, Controls, Energy and Materials (ICSTM), Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, T.N., India. 6 - 8 May 2015. pp.138-145.

3. Jeong JH, Resop JP, Mueller ND, FleisherDH, Yun K, Butler EE, et al. (2016)"Random Forests for Global and Regional Crop Yield Predictions" ,PLoSONE 11(6): e0156571. doi:10.1371/journal.ppone.0156571 pp-1-15

4. Narayanan Balakrishnan1 and Dr. Govindarajan Muthukumarasamy2,"Crop Production-Ensemble Machine Learning Model for Prediction",International Journal of Computer Science and Software Engineering (IJCSSE), Volume 5, Issue 7, July 2016 ISSN (Online): 2409-4285 www.IJCSSE.org Page: 148-153

5. Sami Khanala,∗John Fultonb,Andrew Klopfensteinb,Nathan Douridasc,Scott Shearerb,"Integration of high resolution remotely sensed data and machine learning

6. techniques for spatial prediction of soil properties and corn yield",Computers and Electronics in Agriculture Volume 153, October 2018, Pages 213-225

7. Louis Kouadioa,∗Ravinesh C. Deob,∗Vivekananda Byrareddya,Jan F. Adamowskic,Shahbaz Mushtaqa,Van Phuong Nguyend,"Artificial intelligence approach for the prediction of Robusta coffee yield using soil fertility\ properties",Computers and Electronics in Agriculture Volume 155, December 2018, Pages 324-338