# ABSTRACT

The goal of this research is to use machine learning techniques to forecast yield in different districts of India on a yearly basis. The agricultural sector is the prime occupation of India. Researchers are developing various scientific technologies in the agriculture field for better yield. In this project, we try to form an ensemble model using various machine learning algorithms for better rice production. Crop production prediction utilizing AI Strategies aims to deliver improved outcomes, but the ensemble model provides better predictive results compared to the individual algorithm. We tried to use a combination of symmetric machine learning algorithms to form an ensemble model for better prediction. Here symmetric algorithms such as random forest, Gradient Boosting, and Logistic Regression are individually used for the prediction of the yield of rice. While combining all the aforesaid algorithms to form an ensemble model of a better result. Crop recommendation depends upon soil characteristics, soil nutrient contents, topography, and other uncontrollable parameters such as weather conditions, rainfall, humidity, environment temperature. Modern technologies need to be integrated into farming practices to enhance crop productivity. The Machine Learning technique is an essential approach for building an appropriate model for crop recommendation.

## TABLE OF CONTENTS


| Chapter no. | Title | Pg no. |
|---|---|---|
| 1 | INTRODUCTION | 1 |
| | 1.1 OUTLINE OF THE PROJECT | 1 |
| | 1.2 MACHINE LEARNING | 1 |
| 2 | LITERATURE SURVEY | 3 |
| 3 | ML ALGORITHM | 6 |
| | 3.1 ALGORITHM USED | 6 |
| | 3.1.1 Random Forest | 6 |
| | 3.1.2 Gradient Boosting | 6 |
| | 3.1.3 Logistic Regression | 7 |
| 4 | DATASET | 8 |
| | 4.1 ATTRIBUTE INFORMATION | 8 |
| | 4.1.1 Summarise data | 8 |
| 5 | SOFTWARE REQUIREMENTS | 14 |
| | 5.1 SOFTWARE USED | 14 |
| | 5.1.1 Anaconda Navigator | 14 |
| | 5.1.2 Python 3.7 (64-bit) | 14 |
| | 5.2 LIBRARIES USED | 15 |
| | 5.2.1. NumPy | 15 |
| | 5.2.2. Sklenar | 15 |
| | 5.2.3. Pandas | 16 |
| | 5.2.4. Matplotlib | 16 |
| | 5.2.5. Seaborn | 17 |
| 6 | METHODOLOGY | 18 |
| | 6.1 WORKFLOW DIAGRAM | 18 |

| | | |
|---|---|---|
| | 6.1.1. Random Forest | 18 |
| | 6.1.2 Logistic Regression | 20 |
| | 6.1.3. Gradient Boosting | 21 |
| | 6.1.4. Ensemble Model | 23 |
| 7 | APPLICATION WORKING | 25 |
| | 7.1 Machine Learning | 26 |
| | 7.2. Web Application | 26 |
| 8 | RESULTS AND DISCUSSION | 32 |
| 9 | CONCLUSION | 35 |
| | REFERENCES | 36 |
| | APPENDIX | 37 |
| | A.SOURCE CODE | 37 |
| | B.SCREENSHOTS | 39 |
| | C.PLAGIARISM REPORT | 42 |
| | D. CERTIFICATES | 43 |
| | E. JOURNAL PAPER | 45 |
| | F. COPYRIGHT FORM | 51 |

# LIST OF FIGURES


| Figure no. | Title | Pg no. |
|---|---|---|
| 4.1 | Count of data points in the data variable Rainfall, N, P, K, Temperature, Humidity, and pH | 9 |
| 4.2 | Count of data points in the data variable N, P, K, Temperature, Humidity, and pH compared with each crop | 10 |
| 4.3 | Checking how (temperature, rainfall), (temperature, humidity), and (temperature, pH) affect the individual crop | 12 |
| 4.4 | Correlation matrix for the dataset | 13 |
| 6.1 | System flow diagram of the proposed method | 18 |
| 6.2 | Defining the random forest classifier | 19 |
| 6.3 | Tuning and fitting the test & train data for best estimators | 19 |
| 6.4 | predictions of random forest model for test and train data | 20 |
| 6.5 | Defining the Logistic Regression classifier | 21 |
| 6.6 | predictions of Logistic Regression for test and train data | 21 |
| 6.7 | Defining the Gradient Boosting classifier. | 22 |
| 6.8 | predictions of Gradient Boosting for test and train data | 22 |
| 6.9 | Defining the Ensemble Model classifier & voting classifier | 23 |

| | | |
|---|---|---|
| 6.10 | prediction of ensemble model test and train data | 24 |
| 7.1 | Classification report of proposed ensemble model | 25 |
| 7.2 | Application Screenshots | 29 |
| 7.3 | Application Output Screenshots | 30 |
| 8.1 | Classification report of proposed ensemble model | 34 |

# LIST OF TABLES


| Table no. | Title | Pg no. |
|---|---|---|
| 6.1 | Matrices report of proposed Random Forest | 20 |
| 6.2 | Matrices report of proposed Logistic Regression | 21 |
| 6.3 | Matrices report of proposed Gradient Boosting | 23 |
| 6.4 | Matrices report of proposed ensemble model | 24 |
| 8.1 | Accuracy of the individual machine learning algorithm and ensemble model | 32 |

# CHAPTER 1

# INTRODUCTION

## 1.1 OUTLINE OF THE PROJECT

Agriculture is the prime occupation in India. Farmers from rural areas of India are mostly dependent on agriculture for their survival. Traditional farming practices pose challenges in terms of predicting abnormal seasonal changes and foresight before flood warning signals, which results in bad crop yields. The consequence of it affects the livelihood of farmers but also badly influences the country's economic status.

Crop recommendation depends upon soil characteristics, soil nutrient contents, topography, and other uncontrollable parameters such as weather conditions, rainfall, humidity, environment temperature. Modern technologies need to be integrated into farming practices to enhance crop productivity. The Machine Learning technique is an essential approach for building an appropriate model for crop recommendation.

Farmers can't accurately predict future possibilities regarding sudden change in climatic conditions, abnormal rise in atmospheric temperature, future flood occurrences. Artificial Intelligence and Machine Learning techniques help farmers in recommending suitable crops, fertilizer recommendations, predicting the average production of the crop based on crop type. The goal of this research is to use machine learning techniques to forecast yield in different districts of India on a yearly basis.

## 1.2 MACHINE LEARNING

Machine learning could be a subfield of computer science (AI). The Goal of machine learning typically is to know the structure of information and match that data into models which will be understood and used by folks. Although machine learning could be a field inside technology, it differs from ancient process

approaches. In ancient computing, algorithms are sets of expressly programmed directions employed by computers to calculate or downside solve. Machine learning algorithms instead give computers to coach on knowledge inputs and use applied math analysis so as to output values that fall inside a particular variable. Thanks to this, machine learning facilitates computers in building models from sample knowledge so as to modify decision-making processes supported knowledge inputs.

# CHAPTER 2

# LITERATURE REVIEW

● [1] (Doshi, Nadkarni, Agarwal & Shah) proposed a recommender system to make a decision about the type of crop to be grown depending upon various factors. They found the predicted result for the total amount of rainfall and combined together with the training model to recommend the type of crop to be harvested.

● The study of [2] (Mythili & Rangaraj, 2021) reveals a new deep learning-based approach for predicting crop yields. The proposed algorithm PSO-MDNN recommended a suitable crop recommender model, to recommend a suitable crop effectively. Its accuracy outperformed different machine learning techniques. The proposed model predicted yields in untested environments and recommended crop cultivation. The weight matrices with L2 regularization and PSO are used to modify the MDNN hyperparameters. The prediction accuracy was enhanced by using a network structure with optimum weights using PSO.

● [3] (Rajak, Pawar & pendke, 2017) Discuss the ensemble model in conjunction with the majority voting approach. SVM using ANN as a learner is combined with Random Tree and NB-classifier to propose a crop with higher efficiency and accuracy.

● [4] (Kuanr, Rath & Mohanty, 2018) depending on the location and agricultural preferences, proposes seed, herbicides, and equipment The agricultural yield in the Kharif season is predicted using fuzzy logic and cosine similarity. Farmers can predict crop production based on previous yield data and expected weather conditions.

● The paper [5] (Kumar, Singh, Kumar & Singh, 2015) Crop selection method (CSM) proposed and advises crop selection to enhance the net production rate.

They recommend the sort of crop to be yielded based on parameters such as rain, temperature, and humidity, as well as the type of soil.

● Another literature [6] (Pudumalar, Ramanujam, Rajashree, & et al. 2017) Precision agriculture, which makes use of soil properties, soil types, and crop production, is addressed. Based on the above characteristics, it recommends the best crop to farmers. The authors suggested an ensemble model that achieved good accuracy by combining Naive Bayes, CHAID, KNN, and Random tree.

● [7] (Lakshmi, Priya, Sahana & Manjunath, 2018) the goal was to create a crop growth monitoring recommendation system. The unique qualities utilized were soil composition, color, drainage, depth, ph. value, water holding capacity, erosion, permeability, and weather. Map-reduce with KNN suggests quickly and accurately.

● [8] (Akshatha & Shreedhara, 2018) suggested an ensemble model for precision agriculture that uses different machine learning algorithms to select the crop that is most suited to the farmer based on the information (factor influencing crop production) that he provides

● Across the US Corn Belt, homogeneous ensembles produce the most accurate yield projections. This model could accurately estimate 2019 corn grain yields with a root mean square error of 866 kg/ha, or 8.5 percent relative root mean square and could explain almost 77% of the spatiotemporal variation in corn grain yields.[9](Shahhosseini M et.al 2021)

●The authors [10](Rashid, M., Bari, B. S., Yusup, Y., Kamaruddin, M. A., & Khan, N.) proposed a system based on a critical evaluation of existing related studies, a prospective architecture for machine learning-based palm oil yield prediction. This system will live up to its promise by tackling new research issues in agricultural yield prediction analysis.

●[11](Thomasvan Klompenburga, and AyalewKassahuna) The most used features are temperature, rainfall, and soil type, and the most applied algorithm is Artificial Neural Networks in these models. After this observation based on the analysis of machine learning-based 50 papers, we performed an additional search in electronic databases to identify deep learning-based studies.

●[12] (Joshua, V.; Priya Harson, S.M.; Kannadasan, R.) did a statistical analysis of possible MLR approaches and machine learning algorithms such as SVM, GRNN, RBFNN, and BPNN are examined for evaluation in order to achieve improved accuracy in agricultural yield prediction. The accuracy level of the various algorithms is scrutinized using model performance indicate