# Transforming Indian Agriculture: A Machine Learning Approach for Informed Decision-Making and Sustainable Crop Recommendations

Smitta Ranjan Dutta
Assistant Professor, Ajay Binay Institute of Technology, Master in Computer Application
E-mail: smittaranjandutta@gmail.com

Sanata Kumar Swain
Assistant Professor, Ajay Binay Institute of Technology, Master in Computer Application
E-mail: sanu.sanata@gmail.com

Dr. Amaresh Sahu
Associate Professor, Ajay Binay Institute of Technology, Master in Computer Application
E-mail: amaresh_sahu@yahoo.com

Bijoy Laxmi Behera
Student, Ajay Binay Degree College, Information Technology and Management
E-mail: bijoy.2017pp@gmail.com

Corresponding author: Dr. Amaresh sahu, E-mail: amaresh_sahu@yahoo.com

## Abstract

In India's economic landscape, agriculture holds a pivotal position. However, the evolving structure of Indian agriculture presents a pressing challenge. The critical solution to overcoming this crisis lies in transforming agriculture into a profitable sector, incentivizing farmers to continue their cultivation practices. This research article is dedicated to facilitating this transformation by introducing machine learning as a tool for informed decision-making among farmers. The focus of this study is to harness machine learning algorithms to enhance agricultural productivity and sustainability through data-driven crop recommendations. By dissecting the fundamental components of machine learning-based crop suggestions, we delve into the advantages demonstrated through real-world case examples. These instances serve as tangible proof of how data-driven insights can optimize crop yields and resource utilization. This research aims to provide a comprehensive understanding of the complexities involved, paving the way for potential future solutions and paths. The ultimate objective is to illustrate how machine learning is reshaping precision agriculture, ushering in a new era of environmentally conscious farming practices. Through this exploration, we aspire to showcase not only the revolutionary impact but also the potential for a harmonious coexistence between technological advancements and sustainable farming methods.

**Keywords:** Machine learning**,** Logistic Regression, Decision Tree, Random Forest, Ada Boost, Bagging, SVM (RBF & Linear Kernel), K-Nearest Neighbours, Naive- Bayes (multinomial & Gaussian), Multilayer Perceptron, XGBoost, LightGBM, Catboost and SGD Classifier.

## Introduction

The emerging field of crop suggestion in agriculture is capturing public attention, as farmers often lack awareness of the most suitable crops for their farms, leading to productivity hindrances and confusion. To address this, we have expanded existing databases on crop production by creating a tailored dataset for India. This dataset incorporates crucial factors such as rainfall, humidity, temperature, pH, and season, offering a comprehensive understanding of environmental and geographic influences on crop patterns [1]. Utilizing this dataset, we aim to develop a machine-learning model that can assist in determining the optimal crops for specific regions. The implementation of machine learning in agriculture holds the potential to revolutionize the industry by providing early guidance on raw material and resource requirements. This proactive approach addresses issues like nutrient shortfalls resulting from the cultivation of inappropriate crops, thereby enhancing overall output efficiency.

Recognizing India's lag in adopting modern agricultural solutions, especially given its significance as the primary source of income for a majority of the population, there is a pressing need for scientific advancements in the sector. The primary objective of our model is to furnish farmers with guidance based on variables such as soil composition, temperature, humidity, rainfall, and geographical impact.

In our research, we present an overview of the data collection and analysis procedures used to evaluate the effectiveness of machine learning algorithms in predicting crop yields for various crops. The dataset encompasses information from 22 different crops between 2000 and 2014. By employing machine learning methods, we train and test the model to assess its accuracy. The methodology involves data preparation using Pandas and NumPy, data visualization with Matplotlib and Seaborn, and machine learning model development using Scikit-learn, XGBoost, and CatBoost. The integration of gradient boosting algorithms like XGBoost and CatBoost proves essential in creating precise prediction models that account for intricate interactions between variables.

## Literature review

In recent years, numerous studies have delved into the convergence of agriculture and advanced machine learning, shedding light on the potential applications of contemporary technology in the agricultural sector.

Kevin Tom Thomas et al. specifically focused on crop prediction using KNN with cross-validation, giving prominence to soil factors. Mahendra N. and colleagues opted for the Decision Tree method to predict crops, taking into account both soil and weather factors. Some researchers, including Kevin Tom Thomas, utilized the Support Vector Machine (SVM) algorithm to determine rainfall, a critical weather parameter.

Mansi Shinde et al. cantered their study on NPK (Nitrogen, Phosphorous, and Potassium) levels in the soil for crop recommendations, employing the Random Forest algorithm. Sonal Jain et al. underscored the influence of weather and soil characteristics on crop selection, noting the absence of crucial soil features like NPK levels. A. Suresh et al. applied modified KNN and K-Means to predict crop yields in Tamil Nadu, concentrating on main crops.

For detecting soil type for recommendation of crops, S. Pudumalar and co-authors employed an ensemble strategy that combined Naive Bayes, K-Nearest Neighbour, and Random Forest algorithms. R. Kumar and his research team conducted an investigation where they took into account factors such as governmental policies, market prices, and production rates during the meticulous process of crop selection. The introduction of the Crop Selection Method (CSM) was proposed as a means to maximize the crop yield rate.

The integration of the Internet of Things (IoT) into crop recommendation takes center stage in the investigation led by Angu Raj and his research team. They employed sensors to collect soil characteristics, including temperature, humidity, soil moisture, and ph. In a similar vein, Lakshmi N. and her research team proposed a comprehensive crop recommendation system by leveraging advanced big data methodologies. Their approach involved considering a range of factors, including drainage, texture, color, depth, soil erosion, pH, permeability, and water retention.

Vivek, M.V.R., and his colleagues, undertook a comprehensive investigation into the application of diverse machine learning algorithms for crop recommendation. Their research was grounded in the analysis of meteorological data, crop data, and soil data, incorporating a variety of techniques such as Naive Bayes, multi-layer perceptron, JRIP, Jf48, and SVM.

## Data and variables

Using the data gathered from kaggle.com, we were able to create a predictive model that will suggest the best crops to grow on a certain farm depending on a variety of factors. The size of the dataset is (2200, 8). Nitrogen, phosphorous, and potassium are the components of soil. The following list of data fields:

Ratio of Nitrogen content in soil is represented as N.
Ratio of Phosphorous content in soil is represented as P.
Ratio of Potassium content in soil is represented as K.
Temperature in degree Celsius is represented as temperature.
Relative humidity in % is represented as humidity.
Ph value of the soil is represented as ph.
Rainfall in mm is represented as rainfall.
crop suitable is represented by label

The snapshot of the data set is depicted in Figure 1.

**Figure 1.**

| | N | P | K | temperature | humidity | ph | rainfall | label |
|---|---|---|---|---|---|---|---|---|
| 0 | 90 | 42 | 43 | 20.879744 | 82.002744 | 6.502985 | 202.935536 | rice |
| 1 | 85 | 58 | 41 | 21.770462 | 80.319644 | 7.038096 | 226.655537 | rice |
| 2 | 60 | 55 | 44 | 23.004459 | 82.320763 | 7.840207 | 263.964248 | rice |
| 3 | 74 | 35 | 40 | 26.491096 | 80.158363 | 6.980401 | 242.864034 | rice |
| 4 | 78 | 42 | 42 | 20.130175 | 81.604873 | 7.628473 | 262.717340 | rice |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 2195 | 107 | 34 | 32 | 26.774637 | 66.413269 | 6.780064 | 177.774507 | coffee |
| 2196 | 99 | 15 | 27 | 27.417112 | 56.636362 | 6.086922 | 127.924610 | coffee |
| 2197 | 118 | 33 | 30 | 24.131797 | 67.225123 | 6.362608 | 173.322839 | coffee |
| 2198 | 117 | 32 | 34 | 26.272418 | 52.127394 | 6.758793 | 127.175293 | coffee |
| 2199 | 104 | 18 | 30 | 23.603016 | 60.396475 | 6.779833 | 140.937041 | coffee |

## Methodology and model specifications

The development of this research code involved the utilization of diverse technologies, including but not limited to NumPy, Pandas, and Matplotlib. Pytorch, along with several other Python libraries, was also instrumental in the coding process. The models for this project were prepared through the utilization of Python files, .pynb files, and .pkl files. A range of machine learning algorithms was employed on the provided dataset, leading to the training and evaluation of numerous models. The outcomes of these models enable users to predict the optimal crop for cultivation based on specific conditions through an intranet-based system. This system facilitates the addition, viewing, and updating of crops suitable for cultivation in a particular area under specific environmental conditions, thereby enhancing agricultural management practices.

## Empirical results

Figure 2. represents confusion matrices for various models: Logistic Regression, Decision Tree, Random Forest, Ada Boost, Bagging, SVM (RBF & Linear Kernel) , K-Nearest Neighbours, Naive-Bayes (multinomial & Gaussian), Multilayer Perceptron, XGBoost, LightGBM, Catboost and SGD Classifier.

**Figure 2.**

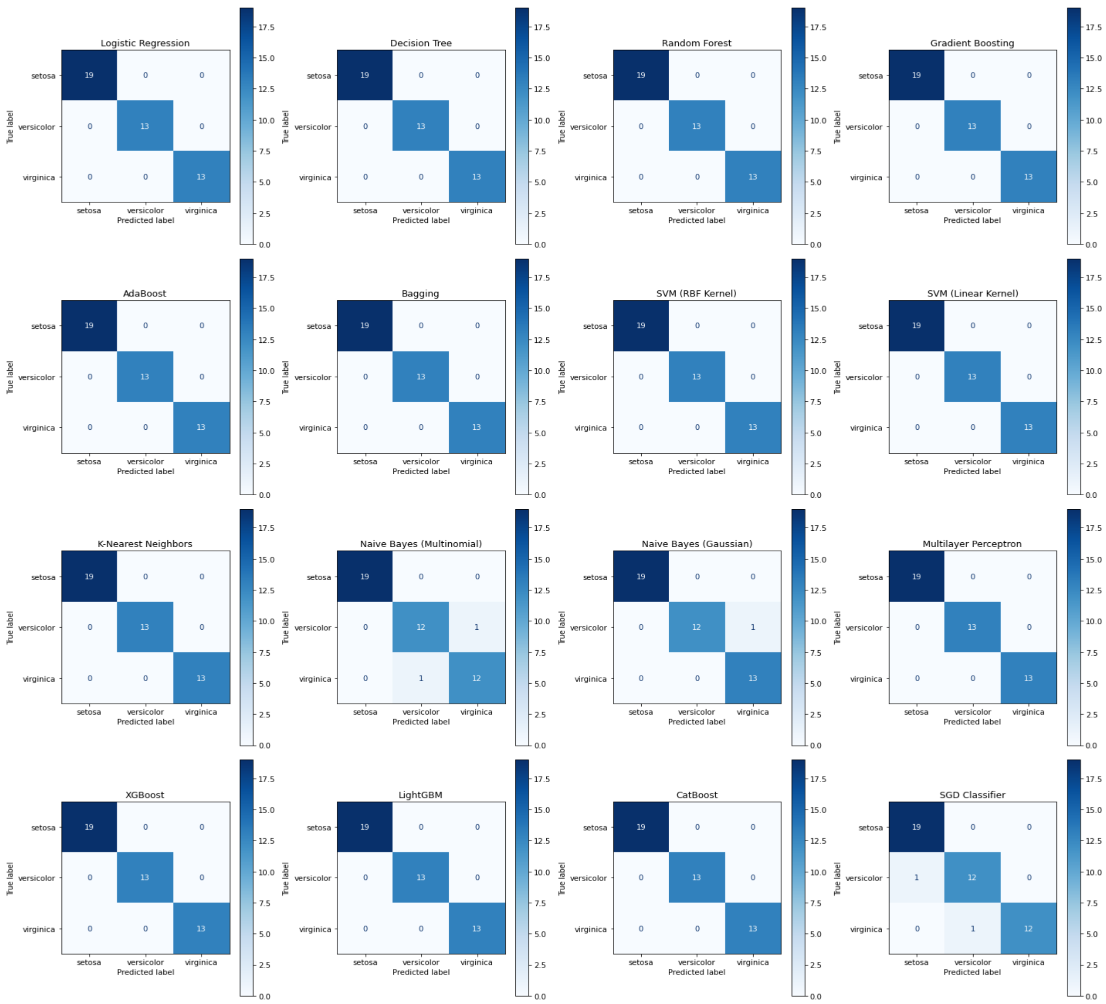
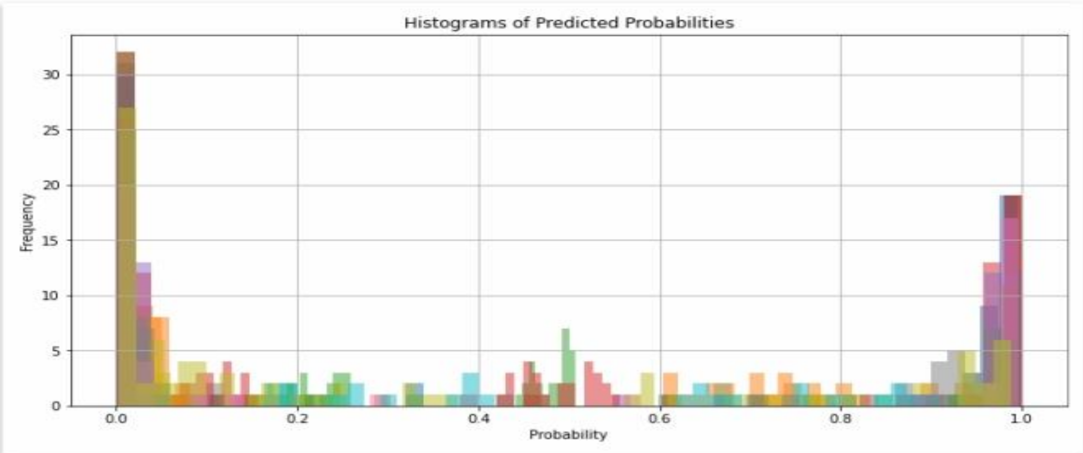


Figure 3. shows histogram of predicted probabilities of various ML algorithms on parameters

**Figure 3.**




On the selected dataset, all the algorithms are evaluated and trained, and various models are produced by running predictions through this model. The algorithm that must be selected must be compatible with the prepared datasets. As a result, a variety of models are generated after all the algorithms have been trained and tested on the dataset. In order to develop models and make accurate predictions, ML Algorithms must be coded once all of those algorithms have been examined. Naive Bayes, XGBoost delivers the highest and equal accuracy, or 99.55 percent, out of the all algorithms used. The accuracy percentage obtained with Random Forest and SVM is 99.31. Regression using logistics and k-NN yields accuracy scores of 98.86 and 98.63, respectively. Decision trees only provided an accuracy of 85.91 percent in the case, which was their worst performance.

## Conclusion

We utilized a sample dataset from Kaggle that contained data culled from a sizable agricultural demographic. Farmers frequently take a hit-and-miss technique that wastes resources and land or even makes crops grow uncontrollably. We are trying to remove all these heavy hurdles by providing them with access to an accurate and compelling model generated by machine learning using a random forest classifier to decide the correct crop to be produced in their fields. This will enable them to raise their agricultural production's quality and volume. By doing this, they will be able to maintain the nutrients and quality of the soil. Before harvesting a particular crop, farmers could run into problems or sickly crops. They are then allowed to upload the crop and soil reports. Subsequently, addressing challenges involves the exploration of viable AI-driven approaches to present practical remedies. Furthermore, the implementation of AI models opens avenues for the provision of APIs and Virtual agents, facilitating IOT solutions. This technological integration empowers farmers to engage seamlessly with suppliers of essential resources such as seeds and fertilizers, tailored to the specific needs of each crop.

## References


Naresh G. Gadge and Ruchira C. Mahore. A Novel Application of Machine Learning for Crop Analysis and Prediction, 2022 ICWITE 2022: IEEE International Conference for Women in Innovation, Technology, and Entrepreneurship

http://www.itm-conferences.org

Kevin Tom Thomas, Varsha S, Merin Mary Saji, Lisha Varghese and Er. Jinu Thoma, “Crop Prediction Using Machine Learning”, International Journal of Future Generation Communication and Networking, Vol. 13, Issue: 3, pp. 1896-1901, 2020

Mahendra N , Dhanush Vishwakarma, Nischitha K , Ashwini, and Manjuraju M. R, “Crop Prediction using Machine Learning Approaches”, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 08 (August 2020).

Mansi Shinde, Kimaya Ekbote, Sonali Ghorpade, Sanket Pawar and Shubhada Mone, "Crop Recommendation and Fertilizer Purchase System", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (2), 665- 667,2016.

S. Jain and D. Ramesh, "Machine Learning convergence for weather-based crop selection," IEEE International Students' Conference on Electrical, Electronics, and Computer Science (SCEECS), pp. 1-6, doi: 10.1109/SCEECS48394.2020.75, 2020.

Suresh, P. Ganesh Kumar and M. Ramalatha, "Prediction of major crop yields of Tamilnadu using K-means and Modified KNN," 3rd International Conference on Communication and Electronics Systems (ICCES), pp. 88-93, doi: 10.1109/CESYS.2018.8723956, 2018.

S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika and J. Nisha, "Crop recommendation system for precision agriculture," Eighth International Conference on Advanced Computing (ICoAC), pp. 32-36, doi: 10.1109/ICoAC.2017.7951740, 2017.

R. Kumar, M. P. Singh, P. Kumar and J. P. Singh, "Crop Selection Method to maximize crop yield rate using machine learning technique," International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), pp. 138- 145, doi: 10.1109/ICSTM.2015.7225403, 2015.

S. M. PANDE, P. K. RAMESH, A. ANMOL, B. R. AISHWARYA, K. ROHILLA, and K. SHAURYA, "Crop Recommender System Using Machine Learning Approach," 5th International Conference on Computing Methodologies and Communication (ICCMC), , pp. 1066-1071, doi: 10.1109/ICCMC51019.2021.9418351, 2021

Raj, Angu & Balashanmugam, Dr.Thiyaneswaran & Jayanthi, J & Yoganathan, N & Srinivasan, PCrop. . Recommendation on Analyzing Soil Using Machine Learning. Turkish Journal of Computer and Mathematics Education (TURCOMAT). 12. 1784- 1791, 2021.

Lakshmi N, Priya M, Mrs. Sahana Shetty and Mr. Manjunath C.R ,"Crop Recommendation System for Precision Agriculture ", International Journal for Research in Applied Science & Engineering Technology (IJRASET), ISSN: 2321-9653; IC Value: 45.98; SJImpact Factor: 6.887, Volume 6 Issue V, May 2018

M.V.R. Vivek, D.V.V.S.S. Sri Harsha and P. Sardar Maran, “A Surveyon Crop Recommendation Using Machine Learning”, International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277- 3878, Volume-7, Issue-5C, February 2019.

https://www.kaggle.com/datasets/siddharthss/crop-recommendation-dataset.