e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

COMPARATIVE STUDY FOR CROP RECOMMENDATION USING MACHINE LEARNING ALGORITHMS

Anuja AV*1, Poojanaa K*2, Joshika V*3

*1Assistant Professor, Department Of Software Systems, Sri Krishna Arts And Science College, India.

*2,3Student, Department Of Software Systems, Sri Krishna Arts And Science College, India.

DOI : https://www.doi.org/10.56726/IRJMETS45288

ABSTRACT

A large portion of the Indian population is involved in agriculture as their primary livelihood. Typically, farmers adhere to traditional methods, such as planting the same crop repeatedly, increasing their use of fertilizers, and following established routines. Nevertheless, recent years have witnessed notable advancements in the utilization of machine learning across various sectors and research domains. Given these advancements, our goal is to implement a machine learning- based system within the agricultural industry to support and benefit farmers. Our strategy involves integrating multiple factors to achieve more favourable outcomes. This enhancement is expected to result in improved crop yields and the recognition of patterns that contribute to accurate predictions. Through the utilization of this system, we can effectively determine the most suitablecrops for specific regions, thereby offering valuable insights to farmers and optimizing their agricultural output.

Keywords: Decision Tree (DT), Naïve Bayes (NB), Random Forest (RF), XG BOOST.

I. INTRODUCTION

Agriculture stands as the fundamental source of food supply for countries worldwide, whether they are underdeveloped developing, or even developed. Predictions indicate that the global population will reach approximately 9.7 billion by 2025. Coupled with unpredictable weather patterns, this scenario creates challenges in ensuring sustainable food production. The ongoingshift in climate poses a persistent threat to crops, driving farmers into debt and, unfortunately, even leading some to considersuicide [18]. Such risks could be mitigated by employing mathematical and statistical methods on the data. These methods have thepotential to recommend the most suitablecrops for a farmer's agricultural land, therebyaiding in maximizing profits [12]. In thecontemporary context, agriculture has witnessed substantial growth in India. Precision agriculture, often referred to as "site-specific" farming, emerges as a pivotal approach. Despite the advancements in precision agriculture, certain challenges persist. Its role in crop recommendations relies on diverse parameters [8]. Precision agriculture strives to analyze these parameters in a site-specific manner to uncover underlying issues. However, not allresults generated by precision agriculture can be deemed accurate. In the agricultural realm, precise and accurate recommendations hold significant value, as errors could result in substantial material and capital losses. Numerous research initiatives to develop a more efficient and accurate model for predicting crop outcomes[11]. Machine Learning, encompassing algorithms like supervised, unsupervised, and reinforcement learning, plays a central role in this pursuit. Each algorithm category brings forth its strengths and weaknesses. Supervised learning involves constructing mathematical models from incomplete training data, where portions of input lack labels [8]. This paper's objective revolves around recommending the most suitable cropbased on input parameters such as Nitrogen (N), Phosphorous (P), Potassium (K), soil pH value, humidity, temperature, and rainfall. The paper goes further to predict theaccuracy of future yields for eleven distinct crops, including rice, maize, chickpea, kidney beans, pigeon peas, moth beans, mungbean, black gram, lentil, pomegranate, banana, mango, grapes, apple, orange, papaya, coconut, cotton, jute, and coffee. Advancements in agriculture, challenges persist. To achieve this, a variety of supervised machine learning approaches are employed within the context of India. The dataset incorporates parameters like Nitrogen (N), Phosphorous (P), Potassium (K), soil pH value, humidity, temperature, and rainfall. The proposed system leverages diverse Machine Learning algorithms such as Decision Trees, Naïve Bayes (NB), Support Vector Machine(SVM), Logistic Regression, Random Forest (RF), and XGBoost [12].

PROBLEM STATEMENT

In the context of a country like India, crop production is significantly impacted by a multitude of variables. Key factors such as humidity, temperature, rainfall, and soil type are critical in predicting crop outcomes, and their dynamics vary considerably across regions. Despite this, many Indian farmers continue to rely on traditional methods passed down through generations. While these methods may have been effective in the past when climatic conditions were more stable, the current scenario, marked by global warming and environmental pollution, necessitates a shift towards modern techniques. The present moment demands a comprehensive examination of extensive datasets and the development of a system capable of furnishing valuable insights into crop yield forecasts. This modern approach hinges on the availability of vast,well structured datasets and an algorithm that can effectively harness this information to generate meaningful solutions. By using only one methodology, may not give a satisfied or notable change in the prediction of crops. So we are using panel of algorithms to get themost accurate crops in the different regions bycomparison.

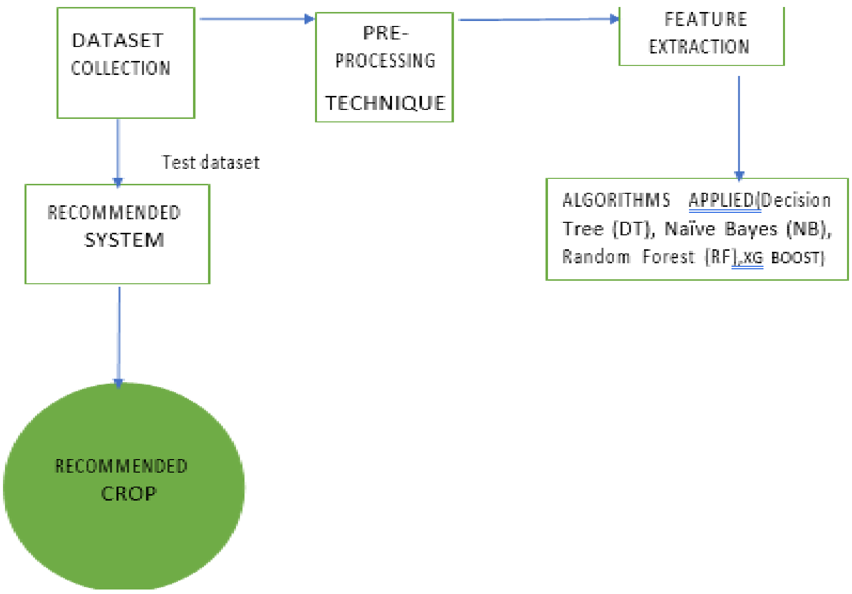
EXISTING SYSTEM

The existing system employs the Support Vector Machine (SVM) algorithm. In SVM, individual data points are represented as points in an n-dimensional space, with each feature value corresponding to a specific coordinate. The classification task involves identifying a hyperplane that effectively separates the two classes. However, the existing SVM implementation has yielded an accuracy rate of just 10.68%, which represents a notably low level of accuracy.

PROPOSED SYSTEM

The proposed system ensures the comparison of various machine learning algorithms, which helps farmers to make better choice bythe comparison of algorithms. In the same region applying various methodologies it turns out various predicted outputs.It makes and creates chance for the farmers to choose according to the accuracy and their own choice by considering or comparing pro's and con's of those algorithms. In the proposed system algorithms such as Decision Tree (DT), Naïve Bayes (NB), Random Forest (RF),XG BOOST are used.

II. METHODOLOGY



Collection of Datasets: Gather relevant data related to crops, agricultural practices, weather conditions, soil types, and other factors that might influence crop growth and yield. This data will be used to train andevaluate your recommendation system.

Pre-processing (Noise Removal): Clean the collected data by removing any inconsistencies, errors, or outliers. This step involves data cleaning, normalization, and handling missing values to ensure the quality of the data you'll be using.

Feature Extraction: Identify and extract meaningful features from the pre-processed data. These features could include weather patterns, soil characteristics, historical crop yields, and more. Properly selected features

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

are crucial for building an effective recommendation system. Applying Various Machine Learning Algorithms: Train and test various machine learning algorithms using your pre-processed and feature-extracted dataset. These algorithms could include decision trees, random forests, support vector machines, neural networks, or any other suitable algorithm for making predictions based on the given data.

Recommendation System: Build a recommendation system that takes in input data related to weather, soil, and other relevant factors. The system will use the trained machine learning models to make predictions about which crops are likely to perform well under those conditions.

Recommended Crop: Once the recommendation system is in place, it will provide suggestions for the best-suited crops based on the input data. The recommendation can be in the form of a ranked list of crops, along with predicted yields or other relevant information. Remember that building an effective recommendation system involves iterative development, testing, and refinement. You'll need to continuously improve your models by incorporating user feedback and updating your datasets to ensure accurate and up-to-date recommendations. Additionally, the success of your system will heavily depend on the quality and relevance of the data you collect and the features you extract.

2.1 COMPARISON OF ALGORITHMS

A. DECISION TREE

Decision tree classifiers operate using a greedy approach, representing a supervised learning algorithm wherein attributes and class labels are organized within a tree structure [15]. The fundamental objective of utilizing Decision Trees is to create a training model that can be applied to predict the class or value of target variables by learning decision rules derived from historical data(training data). Decision trees are comprised of two primary types: decision nodes and leaves. The leaves signify outcomes or final results. Within the tree, each node functions as a test for a specific attribute, and each branch stemming from that node corresponds to a potential answer for the test. This recursive process extends to every sub-tree originating from the new nodes [22].

The fundamental objective of utilizing Decision Trees is to create a training model that can be applied to predict the class or value of target variables by learning decision rules derived from historical data (training data). Decision trees are comprised of two primary types: decision nodes and leaves. The leaves signify outcomes or final results. Within the tree, each node functions as a test for a specific attribute, and each branch stemming from that node corresponds to a potential answer for the test. This recursive process extends to every sub-tree originating from the new nodes [22]. The leaves signify outcomes or final results. Within the tree, each node functions as a test for a specific attribute, and each branch stemming from that node corresponds to a potential answer for the test.This recursive process extends to every sub- tree originating from the new nodes [22].

B. NAIVE BAYES

Naive Bayes serves as a valuable computational tool for addressing binary and multi-class classification challenges. It particularly excels when dealing with binary or categorically structured input values. The Naive

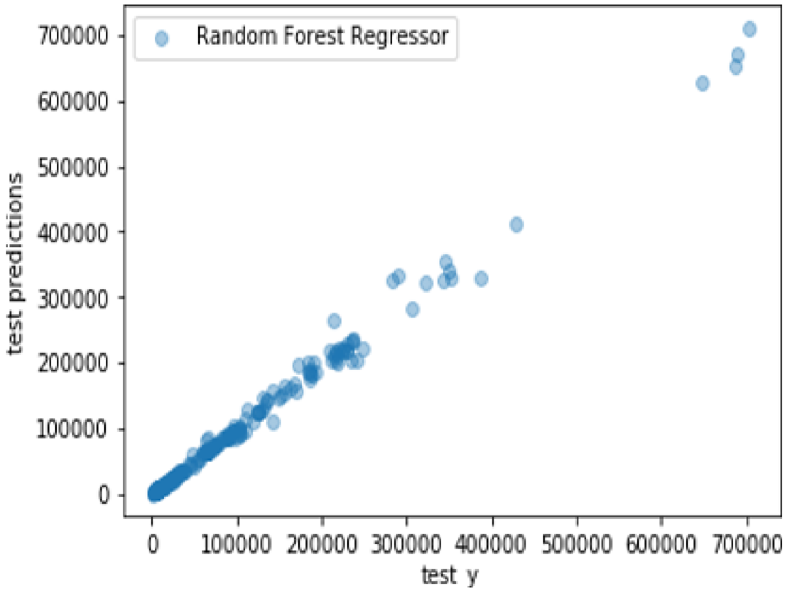
e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

Bayes strategy remains exceptionally straightforward, operating on the premise thatthe presence of a specific attribute within a class is entirely independent of the presence ofany other attribute. This classifier derives its foundation from Bayes' theorem, proving especially advantageous when handling high dimensional datasets. Naive Bayes holdsa wide array of applications, including realtime predictions, the probabilistic prediction of multiple classes for target attributes, spam filtering, and its synergy with collaborative filtering to construct recommendation systems. The initial step involves determining the probability of each attribute within the dataset, often referred to as class probability. Subsequently, conditional probability quantifies the likelihood of each data value given a particular class value [21].We have implemented the NaïveBayes (NB) approach in our model as follows:

- We import NaiveBayesClassifierlibrary from the sklearn.tree class.
- Subsequently, a NaiveBayes Classifierobject is instantiated.
- Finally, the data is fitted to the classifier.

C. RANDOM FOREST

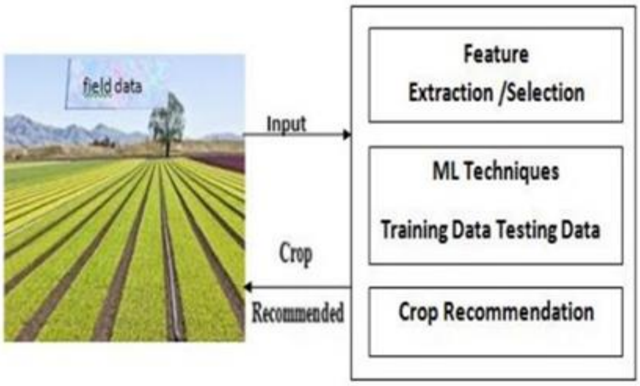
Random Forest is a machine learning algorithm that operates by generating multipledecision trees during the training phase. The outputs are then combined based on the number of classes for classification tasks or used for prediction in regression tasks. The accuracy of prediction is often influenced by the number of trees used. The dataset employed encompasses variables such as rainfall, perception, temperature, and production. These factors are utilized for training purposes, with only two-thirds of thedataset being utilized for this purpose. The remaining portion of the dataset is reserved for experimental purposes. The RandomForest algorithm relies on three key parameters: n_estimators: This parameterdetermines the number of trees to be grown inthe forest, and it has a direct relationship with prediction accuracy.max_ features (m_try): It specifies the number of variables to be considered when making a node split ina tree. min_samples_split (node size): This parameter guides the algorithm on the minimum number of observations required fora split in terminal nodes. We have incorporatedthe Random Forest (RF) algorithminto our model as follows:



D. XGBOOST

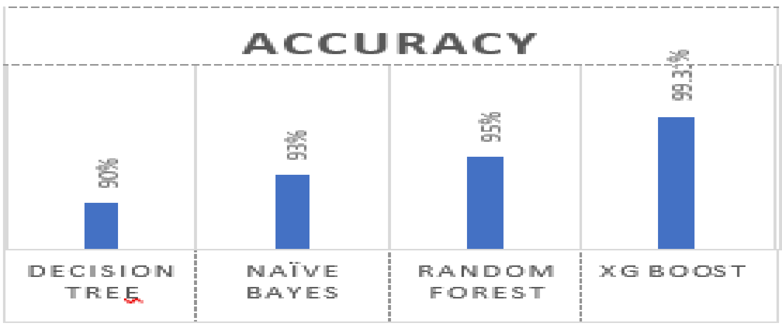
Extreme Gradient Boosting (XGBoost) represents an enhanced and an adaptable iteration of the gradient boosting algorithm, specifically it engineered for efficacy, performance. Renowned for its superior solutions compared to other machine learning algorithms is open-source library and an integral part of the Distributed Machine Learning Community. XG Boost introduces a parallel tree boosting approach (also referred to as GBDT, GBM), which adeptly addresses a multitude of data science challenges with speed and precision. We have integrated XGBoost into our model through the following steps:

IRJMETS
e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com



III. RESULT ANALYSIS

ALGORITHM	ACCURACY
DECISION TREE(DT)	90%
NAÏVE BAYES(NB)	99%
RANDOM FOREST(RF)	99%
XG BOOST	99.31%



IV. CONCLUSION

In this research paper, we have successfully introduced and implemented an intelligent crop recommendation system, designed to be accessible to farmers throughout India. The system serves as a valuable tool for farmers, aiding them in making well-informed decisions regarding crop selection based on factors such as Nitrogen, Phosphorous, Potassium, pH Value, Humidity, Temperature, and Rainfall. By employing this research, we have the potential to enhance the country's agricultural productivity and generate profits through this innovative approach. Consequently, farmers can choose the most suitable crops, thereby increasing their yields and contributing to the overallprofitability of the nation. This study has detailed the recommendation of various cropsin India by utilizing diverse machine learningalgorithms, including Decision Tree, Naïve

V. REFERENCES

[1] S. Agarwal and S. Tarar, A hybrid approach for crop yield prediction using machine learning and deep learning algorithms, in Journal of Physics: Conference Series,2021,no. 1:10.1 088/17426596/1714/ 1/012012.

[2] D. Paudel, H. Boogaard, A. de Wit, â€¦ S. J.- A., and undefined 2021, Machine learning for large- scale crop yield forecasting, Elsevier, Accessed: Jun. 06, 2021.

[3] S. A. Shetty, T. Padmashree, B. M. Sagar, and N. K. Cauvery, Performance Analysis on Machine Learning Algorithms with Deep Learning Model for Crop Yield Prediction, in Springer, 2021, pp. 739750.

[4] D. Elavarasan and P. M. D. R. Vincent, A reinforced random forest model for enhanced crop yield prediction by integrating agrarian parameters, J. Ambient Intell Humaniz. 2021:10.1007/s12652-020-02752-y.

[5] N. R. Prasad, N. R. Patel, and A. Danodia, Crop yield prediction in cotton for regional level using random forest approach, Spat. Inf. Res., vol. 29, no. 2, pp. 195 206, 2021: 10.1007/s41324-020- 00346-6.

[6] Banco Mundial, Indicators, Data, World Development Indicators, 2019

[7] Government of India (GOI), Open Government Data (OGD) Platform India, Website of Open Government Data (OGD) Platform India, 2016. https://data.gov.in/ (accessed Jun. 06, 2021).

[8] S. R. Prathibha, A. Hongal, and M. P. Jyothi, IOT Based Monitoring System in Smart Agriculture, in Proceedings - 2017 International Conference on Recent Advances in Electronics and Communication Technology, ICRAECT 2017, 2017, pp. 8184.

[9] A. Tatapudi and P. Suresh Varma, Prediction of crops based on environmental factors using iot & machine learning algorithms, Int. J. Innov. Technol. Explore. Eng., vol. 9, no. 1, pp. 5395, 5401, 2019.

[10] R. Medar, V. S. Rajpurohit, and S. Shweta, Crop Yield Prediction using Machine Learning Techniques, in 2019 IEEE 5th International Conference for
Convergence in Technology, I2CT 2019,2019

[11] K. Haseeb, I. U. Din, A. Almogren, and N. Islam, energy efficient and secure IoT-based WSN framework: An application to smart agriculture, Sensors (Switzerland), vol. 20, no. 7,2020.

[12] P. Rekha, V. P. Rangan, M. V. Ramesh, and K.V. Nibi, High yield groundnut agronomy: An IoT based precision farming framework, in GHTC 2017 - IEEE Global Humanitarian Technology Conference, Proceedings, 2017, vol. 2017 Janua.