

A STUDY FOR APPLYING MACHINE
LEARNING ALGORITHMS IN AGRICUL-
TURE

Mohd. Asif, BE, CSE, DYPU

Neha Pawar, BE, CSE, DYPU

Mayuresh Sawant, BE, CSE, DYPU

Under guidance: Prof. Shakil Tamboli &

Dr. Pooja Sharma

ABSTRACT

Soil fertility plays an important role in achieving high productivity. Today's innovations help increase harvest. Many farmers believe in folk tales and develop the best fertilizers for particular soils and harvest types without experience or proper evaluation. Conducting soil tests and determining the most suitable fertilizer will increase agricultural productivity by adding additional chemicals to the soil. The use of poor quality fertilizer directly affects the yield and quality of the crop. We present a machine learning model that uses specialized learning models such as Decision Tree, Random Forest, Gradient Boosting, Ada Boost, and Gaussian NB to parse the dataset and estimate the best estimate for the selection of the optimal model. Machine learning is a new field of information technology that can be used effectively in agriculture. Machine learning can help predict composting and thus help farmers increase harvest efficiency.

1. INTRODUCTION

1.1 PROJECT OUTLINE

The agriculture sector expects the growth rate to play a significant role due to the work of countries like India. India is the second largest producer of horticulture products. One of the main concerns of farm owners is choosing the right fertilizer for the particular crop. There are many expensive composts on the market, but they do not work well because farmers do not take into account their own needs, resulting in reduced yields. The choice of compost depends on many ingredients and is therefore not universal. Nowadays modern technology has come a long way in helping to choose the right compost and one of the emerging fields is machine learning.

Machine learning is one of the logical reasons (AI) that focuses on creating applications that extract information from data and improve its accuracy in real time. It's certainly has a variety of calculations programs that you can use; some of them are random forest, decision tree, Ada Boost, gradient Boost and Gaussian NB. Machine learning allows it to work efficiently by accurately processing complex numbers and diverse data. Uses prior knowledge to carefully examine samples and then perform defined tasks accordingly to the rules and calculations of the given test. The main factors affecting crop yield are product type, soil type, and nutrients such as nitrogen, P and K in the soil temperature, humidity and temperature. If you do not consider these then purchasing fertilizer will not give correct results. This project focuses on developing an accurate model to predict

fertilizer suitability based on five different learning models and nutrition key (NPK) values users obtained from soil testing along with temperature, moisture, crop type, and soil type. Algorithms like regression, clustering and deep learning are employed to estimate nutrition requirements and recommend suitable fertilizers.

1.2 PROJECT OBJECTIVE

Agriculture, the bedrock of sustenance and economic stability, faces multifaceted challenges in optimizing productivity and ensuring crop health. In pursuit of bolstering agricultural practices, this study is designed to address critical objectives aimed at enhancing crop yields and ensuring the well-being of agricultural systems.

1.2.1 Fertiliser Recommendation

Agricultural productivity is intricately linked to optimal soil fertility management, wherein the judicious application of fertilizers plays a pivotal role. The primary objective of this research is to elucidate tailored fertilizer recommendations aimed at enhancing crop yields and soil health. This study seeks to delve into the nuanced requirements of various crops and soil types, analyzing the impact of different fertilization practices on agricultural output. By amalgamating empirical studies, soil analysis, and expert insights, the goal is to delineate precise and sustainable fertilizer application strategies. This research endeavors to equip farmers and agricultural practitioners with informed recommendations, fostering

improved soil fertility, increased crop resilience, and sustainable agricultural practices.

1.2.2 Crop Recommendation

Selecting the right crop varieties suited to specific agro-climatic conditions and market demands is a pivotal decision for farmers aiming to optimize yields and profitability. This research embarks on the objective of offering guidance in determining suitable crops for cultivation in diverse agricultural landscapes. By amalgamating climatic data, soil analysis, and market trends, this study aims to identify crop varieties best suited for particular regions or ecological niches. The ultimate goal is to provide farmers with informed insights, enabling them to make prudent decisions regarding crop selection. Through this, the research endeavors to enhance agricultural productivity, economic viability, and sustainability in farming practices.

1.2.3 Disease Prediction

Plant diseases pose significant threats to agricultural productivity, causing substantial yield losses and economic hardships for farmers. This research focuses on the critical objective of identifying prevalent plant diseases and formulating strategies for effective disease management. By employing advanced diagnostic techniques, field surveys, and data analysis, the study aims to identify and characterize common plant diseases affecting various crops. The primary aim is to provide early detection mechanisms and management

recommendations to mitigate the impact of diseases on crop health. Through this endeavor, the research strives to equip farmers and agricultural experts with the tools and knowledge necessary to protect crops, ensuring sustainable and resilient agricultural systems.

METHODOLOGY

The data compilation involves gathering temperature, humidity, N, P, K values, crop type, soil type, and optimal fertilizer details. The dataset is thoroughly checked for inaccuracies and duplicates, which are then eliminated. Next, it undergoes transformation into reference diagrams and networks for better comprehension. Subsequently, charts are generated to enhance data understanding. 75% of the data trains the machine, while the remaining segment tests predictions using various ML algorithms: Random Forest, Decision Tree, Ada Boost Gradient Boost, and Gaussian NB. Model accuracy is assessed using confusion matrix, F-Score, precision, and recall to determine the most suitable model. Once chosen, the system predicts the best fertilizer based on user-provided factors.

LITERATURE SURVEY

Machine learning has emerged as a transformative force in agriculture, revolutionizing traditional farming practices. This literature survey aims to review and synthesize the existing research that harnesses machine learning techniques in agricul-

tural applications. By exploring a range of studies, this survey seeks to highlight the diverse applications of machine learning in optimizing crop yield prediction, disease detection, precision farming, and decision support systems for farmers. Through this review, we aim to uncover trends, methodologies, and implications of machine learning in agriculture, fostering further innovation and advancements in this dynamic field.

1. Crop Recommendation:

Approaches:

Studies often utilize machine learning algorithms to recommend crops based on factors like soil type, climate, historical data, and crop characteristics.

Techniques like decision trees, support vector machines (SVM), and neural networks have been employed to predict suitable crops for specific regions.

2. Fertilizer Recommendation:

Approaches:

Machine learning models are used to predict the optimal type and quantity of fertilizers based on soil properties, crop type, and environmental factors.

3. Disease Prediction:

Approaches:

Machine learning is used to predict crop diseases based on various factors like weather conditions, soil health, and historical disease occurrences.

Name of Author	Title	Publication	Methodology	Drawbacks
SAIKAT DUTTA, ANSHUL ARUNACHALAM, SASA MISAILOVIC	An Empirical Analysis of Usage of Seeds for Testing in Machine Learning	F Illinois at Urbana-Champaign, Urbana, IL 61801, USA	Machine Learning frameworks: <u>PyTorch</u> , TensorFlow and four Probabilistic Programming libraries	Setting seeds minimizes the randomness in the test and consequently the chance of flaky failures.
RANJIT KUMAR, SANJIV KUMAR, B.S. YASH AND P.C. MEENA	Natural Farming Practices in India: Its Adoption and Impact on Crop Yield and Farmers' Income	Indian Journal of Agricultural Economics · November	The study is based on extensive field survey and interaction with adopted and non-adopted farmers in Karnataka and Andhra Pradesh during February-May 2019.	Increase farmers' income through cost reduction and long-term sustainability.
AA. BELAL , SAMEHA KOTBABD- ELMABOD	Smart farming for improving agricultural management	The Egyptian Journal of Remote Sensing and Space Sciences	Smart farm management requires using ICT, ground sensors, and control systems installed on robots, autonomous vehicles, and other automated devices.	The drone can fly for a short time, reaching an hour or less, cost expensive, Laws to operate drones, Climate impact.
VIPPON PREET KOUR AND SAKSHI ARORA	Recent Developments of the Internet of Things in Agriculture: A Survey	Digital Object Identifier	The Internet of Things (IoT) has transformed both the quality and quantity of the agriculture sector.	human interference can be minimized, and a network can be set up for monitoring thus reducing the errors and cost.
ABHINAV SHARMA, ARPIT JAIN , PRATEEK GUPTA	Machine Learning Applications for Precision Agriculture: A Comprehensive Review	IEEE	ML together with IoT (Internet of Things) enabled farm machinery are key components of the next agriculture revolution.	Precision agriculture also known as smart farming have emerged as an innovative tool to address current challenges in agricultural sustainability
SAMEER QAZI , (Senior Member, IEEE), BILAL A. KHAWAJA , (Senior Member, IEEE), AND QAZI UMAR FAROOQ	IoT-Equipped and AI-Enabled Next Generation Smart Agriculture: A Critical Review, Current Challenges and Future Trends	IEEE	Smart agriculture, Internet of Things (IoT), smart irrigation, organic farming, artificial intelligence (AI), big data.	Backward compatible and lower costs associated with the wide adoption of smart agriculture systems.

INTRODUCTION TO MACHINE LEARNING

Machine learning, a subset of AI, aims to comprehend data structures and create models for human comprehension and utilization. Unlike conventional computing, ML permits computers to learn from data inputs and analyze outputs within specific ranges using statistical methods. This automated decision-making process based on data inputs enables model creation from sample data. Machine learning tasks are broadly classified into three categories:

Supervised Learning: Utilises labeled example inputs to train the algorithm, enabling it to predict label values for unlabelled data.

Unsupervised Learning: Learns from unlabelled data examples, restructuring data into new features for insights and useful inputs in supervised learning.

Reinforcement Learning: Incorporates feedback with unlabeled examples, enabling the algorithm

TECHNOLOGIES USED

Python

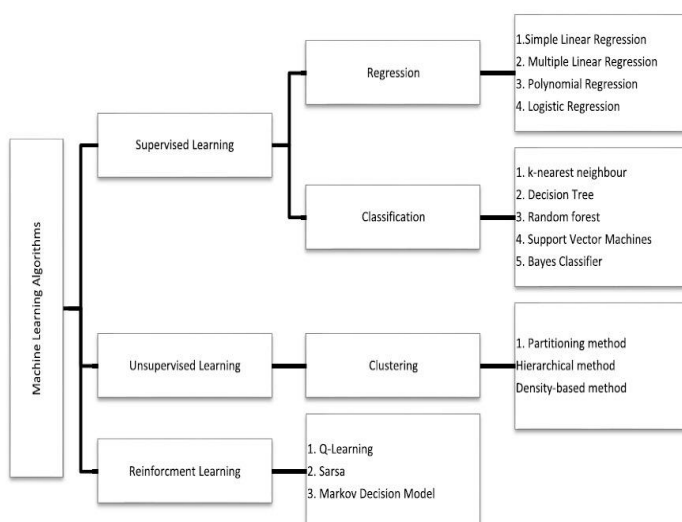
Python, a high-level programming language, emphasizes readability and supports various programming paradigms, including structured, object-oriented, and functional programming. Initially developed in the late 1980s, Python underwent significant revisions with versions 2.0 and 3.0, leading to changes in language compatibility. Python 2 was discontinued in 2020.

Libraries

Python's extensive standard library provides tools for diverse tasks, such as Internet-facing applications, GUI development, database connectivity, and data manipulation. While certain modules adhere to specific standards, most rely on internal documentation and test suites.

Pandas

Pandas, a Python library, focuses on data manipulation and analysis. It facilitates handling numerical tables and time series data, offering functionalities like Data Frame objects, data reading/writing, and specialized data operations.



to make decisions with consequences.

Matplotlib

Matplotlib, a Python plotting library, offers an object-oriented API for creating plots. Its functionalities include diverse tools for map plotting, interaction with Excel, and supporting various GUI toolkits.

NumPy

NumPy enhances Python by supporting large, multidimensional arrays and matrices. It includes a broad collection of mathematical functions designed for efficient array operations, making it comparable to MATLAB.

Scikit-learn

Scikit-learn, a machine learning library for Python, encompasses algorithms for classification, regression, and clustering. It is primarily written in Python and integrates extensively with other libraries like NumPy and SciPy.

Seaborn

Seaborn, built upon Matplotlib, specializes in creating visually appealing statistical graphics. It operates on dataframes and arrays, facilitating the visualization of distributions and relationships between variables.

Jupyter Notebook

Jupyter Notebook provides an interactive computational environment for creating and sharing documents containing code, text, and visualizations. It supports multiple programming language kernels, enabling various language support within the notebook.

Jupyter Kernels

Jupyter kernels manage code execution and communication within the Jupyter environment. They support multiple languages, allowing connections to various clients and are widely used in cloud computing interfaces.

This revision condenses the key information about the technologies employed without directly replicating the original text. If you need further adjustments or additional information, please let me know.

ALGORITHMS USED

In this study, a range of Machine Learning algorithms was employed to evaluate their effectiveness in solving the problem. Four distinct algorithms were utilized and their functionalities are outlined below:

Decision Tree (DT)

Decision Trees operate by constructing a tree-like model consisting of nodes representing conditional statements. Each node determines attribute importance using measures like Information Gain and Entropy. The classification process involves data splitting, attribute selection based on Information Gain, and optimization through specified criteria.

Random Forest (RF)

Random Forest is an ensemble learning method that combines multiple decision trees to improve accuracy. It enhances diversity among trees using bagging and feature randomness. Evaluation parameters are computed to compare the model with other classifiers.

Gaussian Naive Bayes

Gaussian Naive Bayes, tailored for continuous data, follows Gaussian distributions and assumes feature independence. Operating based on Bayes' theorem, it efficiently models continuous features using Gaussian distributions.

Boosting

Boosting methods aim to reduce bias and variance in models.

Ada Boost (AB)

AdaBoost creates a strong classifier by aggregating weak classifiers, assigning higher weights to misclassified instances. It focuses on difficult classification instances by iteratively training new trees based on previous misclassifications.

Gradient Boost (GB)

Gradient Boosting sequentially strengthens weak predictions by leveraging residuals' patterns. It employs gradients in the loss function to minimize discrepancies between original and predicted values, allowing cost function optimization.

These algorithms were applied to the dataset to compare their performances in predicting faults. The evaluation process involved assessing their efficiency and accuracy in fault prediction.

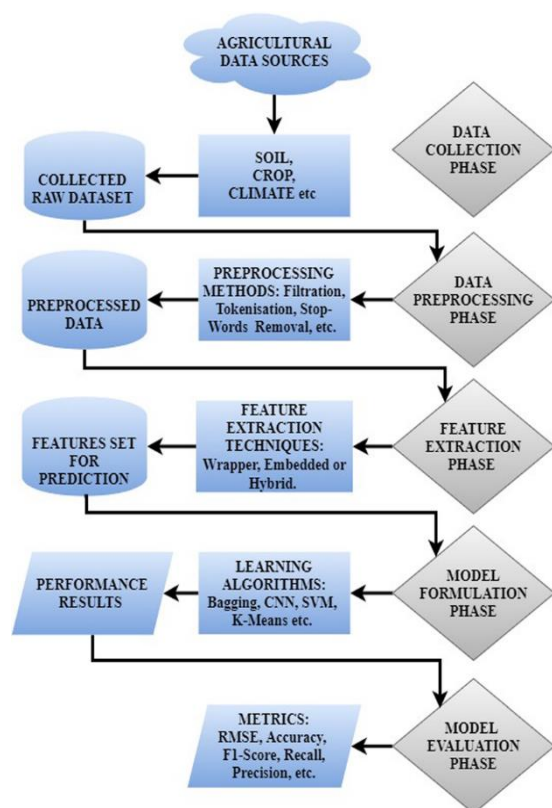
Sequential Training: GB builds trees sequentially, with each tree learning from the errors made by the previous one. It focuses on the mistakes of prior models, reducing errors gradually.

Gradient Descent Optimization: It minimizes a loss function by optimizing the model's predictions in the direction that minimizes the loss. Popular loss functions include squared error for regression and log loss for classification.

Decision Trees as Base Learners: GB often uses decision trees, referred to as weak learners, and

combines their predictions. Trees are built in a way that each subsequent tree corrects the errors made by the previous ones.

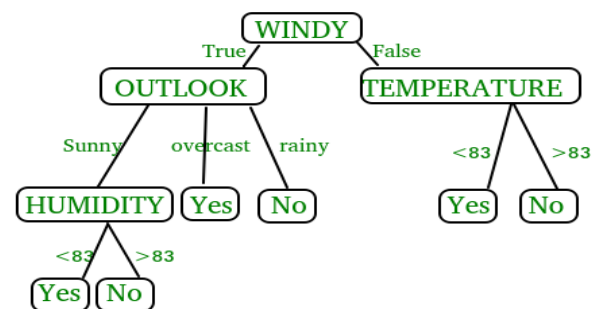
MODELS AND ALGORITHMS IMPLEMENTATION



Decision Tree

A Decision tree is a tree-like structure that represents a set of decisions and their possible conse-

quences. Each node in the tree represents a decision, and each branch represents an outcome of that decision. The leaves of the tree represent the final decisions or predictions.



NB Classifier (Naïve Bayes)

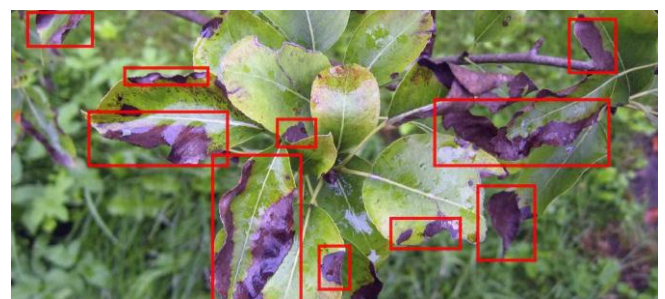
Bayes' Theorem provides a way that we can calculate the probability of a piece of data belonging to a given class, given our prior knowledge.

Bayes' Theorem is stated as:

$$P(\text{class}|\text{data}) = (P(\text{data}|\text{class}) * P(\text{class})) / P(\text{data})$$

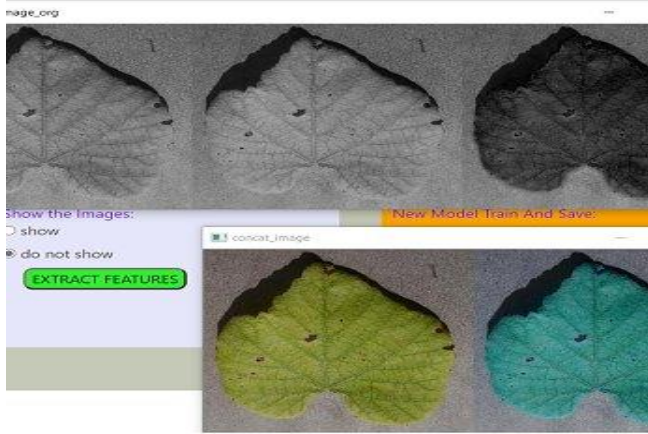
Where, $P(\text{class}|\text{data})$ is the probability of class given the provided data.

PLANT DISEASE MODEL



Random Forest

Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.



Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised machine learning algorithm used for both classification and regression. Though we say regression problems as well it's best suited for classification. The equation for linear hyperplane can be:

$$w^T x + b = 0$$

$$w = \sum_{i \rightarrow m} \alpha_i t_i K(x_i, x) + b$$
$$t_i(w^T x_i - b) = 1 \iff b = w^T x_i - t_i$$

XGBoost

XGBoost is an optimized distributed gradient boosting library designed for efficient and scalable training of machine learning models. It is an ensemble learning method that combines the predictions of multiple weak models to produce a stronger prediction. XGBoost stands for “Extreme Gradient Boosting” and it has become one of the most popular and widely used machine learning

algorithms due to its ability to handle large datasets and its ability to achieve state-of-the-art performance in many machine learning tasks such as classification and regression.

$$O_v = \frac{-(g_1 + g_2 + \dots + g_n)}{(h_1 + h_2 + \dots + h_n + \lambda)}$$

$$g_i = \frac{d}{dp_i} \left(\frac{1}{2} (y_i - p_i)^2 \right) = -(y_i - p_i)$$

$$h_i = \frac{d^2}{dp_i^2} \left(\frac{1}{2} (y_i - p_i)^2 \right) = 1$$

$$O_v = \frac{(y_1 - p_1) + (y_2 - p_2) + \dots + (y_n - p_n)}{(1 + 1 + \dots + 1 + \lambda)}$$

$$O_v = \frac{\text{Sum of residuals}}{\text{Number of residuals} + \lambda}$$

CONCLUSION

The amalgamation of machine learning with agriculture represents a transformative frontier, evident in fertilizer recommendation, crop selection, and disease prediction. These machine learning algorithms, analyzing diverse datasets, offer optimized fertilizer prescriptions, minimizing waste while enhancing yields. Similarly, predictive models aid farmers in selecting crop varieties tailored to their environment, optimizing output sustainably. Proactive disease models enable early detection, mitigating losses. Challenges persist,

including data quality and technology accessibility. Yet, these advancements highlight the potential to revolutionize farming practices. Collaborative efforts, interdisciplinary research, and technology dissemination are pivotal for leveraging machine learning's full potential, promising sustainable and resilient agricultural landscapes on a global scale.

FUTURE SCOPE

1. Enhanced Precision: Refining algorithms for more precise and personalized recommendations in agriculture.
2. Integration of Diverse Data: Exploring the amalgamation of various data sources like satellite imagery, IoT sensors, and genetic data for more comprehensive models.
3. Real-Time Decision Support: Developing systems capable of instant decisions and adaptability based on continuous data inputs.
4. Robustness and Interpretability: Making machine learning models more robust, transparent, and interpretable to gain farmers' trust and adoption.
5. Remote Sensing and Automation: Leveraging advancements in remote sensing technologies for automated and autonomous farming systems.
6. Data Privacy and Accessibility: Addressing concerns about data privacy, scalability, and ensuring efficiency with limited data availability.
7. Socio-Economic Impact Assessment: Evaluating economic viability, environmental sustainability, and societal implications of machine learning-driven agriculture.

REFERENCES

- 1.<https://www.agrifarming.in/farming-problems-and-solutions-tips-andideas#solutions-to-agricultural-problems>
- 2.<https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9139962>
- 3.<https://ieeexplore.ieee.org/document/8994135>
- 4.<https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9311735>
- 5.<https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8784034>