Journal of Current Research in Engineering and Science
JCRES
Bi-Annual Online Journal (ISSN: 2581 - 611X)

VOLUME:8, ISSUE:1, FEBRUARY 2025, PAPER NUMBER:48

SOIL TESTING AND CROP PREDICTION WITH YIELD PREDICTION USING MACHINE LEARNING

Swastid Shiwam Tripathi
Dept. of Computing and Technology
SRM Inst. Of Science and Technology
Kattankulathur, India
swastidt21@gmail.com

P Harsh Sharma
Dept. of Computing and Technology
SRM Inst. Of Science and Technology
Kattankulathur, India
harshpsharma98842@gmail.com

Dr. Manoranjitham T
Dept. of Computing and Technology SRM Inst. Of Science and Technology
Kattankulathur, India
manorant@srmist.edu.in

***Abstract*— The results of a thorough investigation of the use of machine learning algorithms for yield estimation, nutrient advice, and crop suitability prediction in agricultural settings are presented in this research. Three different machine learning models—Linear Regression, Random Forest Regression, and Decision Tree Regression—were assessed using a broad dataset that included variables like soil type, climate, past yield data, and nutrient levels. To evaluate each model's performance in yield estimation, crop suitability prediction, and nutrient recommendation, it underwent rigorous training and testing. The efficacy of the models was assessed using performance indicators such as mean squared error (MSE), mean absolute error (MAE), recall, accuracy, and precision.**

Agricultural decision-making procedures were able to identify the best appropriate method by gaining insights into the strengths and limitations of each model through rigorous experimentation and analysis. The study's conclusions promote precision agriculture by providing useful tools for boosting yields, encouraging sustainable agricultural practices, and optimizing crop management techniques

Keywords*— *Machine Learning, Agriculture, Yield, Regression, Crop Recommendation

I. INTRODUCTION

In the pursuit of maintaining rural livelihoods and ensuring global food security, agriculture plays a pivotal role. As the world's population continues to grow and environmental challenges escalate, there is an increasing demand for innovative solutions to enhance agricultural productivity, optimize resource utilization, and mitigate hazards. This presents an unprecedented opportunity to revolutionize traditional farming practices and usher in a new era of precision agriculture leveraging machine learning technology.

Precision agriculture, with its focus on managing agricultural inputs based on site-specific variability, holds immense potential to enhance resource efficiency, minimize environmental impact, and maximize crop yields. Central to precision agriculture is the ability to leverage vast amounts of data to make informed decisions tailored to the unique needs of individual fields and crops. Machine learning, with its capacity to derive insights from diverse and extensive datasets, emerges as a powerful tool for driving advancements in precision farming.

Key challenges in agricultural decision-making, such as predicting crop suitability, assessing yield potential, and recommending nutrient management strategies, are thoroughly explored through the application of machine learning algorithms in this research. The objective is to develop prediction models that empower farmers with actionable insights drawn from various data sources, including soil characteristics, climate conditions, historical yield data, and nutrient levels.

The research evaluates the performance, interpretability, and computational efficiency of machine learning algorithms—specifically Linear Regression, Random Forest, and Decision Tree models—in addressing agricultural prediction tasks. Each model offers distinct advantages and trade-offs, and this study aims to elucidate their relative merits within the agricultural context through rigorous testing and evaluation.

Ultimately, the overarching goal of the research project is to advance precision agriculture by equipping farmers and agricultural stakeholders with valuable information and tools to make data-driven decisions. Through the development of a mobile application interface, the project aims to provide users with intuitive access to actionable insights, thereby enabling them to optimize crop management strategies, enhance productivity, and foster sustainability. By harnessing machine learning and data-driven approaches, the project endeavors to catalyze agricultural innovation and bolster resilience in the face of evolving environmental and socioeconomic challenges.

II. LITERATURE SURVEY

The authors provide a study on soil analysis and crop fertility prediction using machine learning approaches in this publication. The research makes use of three different datasets: soil, crop, and yield. The soil dataset has 15 variables, and machine learning algorithms are used to determine whether or not the soil is fertile. The crop dataset contains parameters like temperature, humidity, pH, and rainfall, and algorithms are used to predict the sort of crop that will be cultivated. The yield dataset includes nitrogen, phosphorus, potassium, organic care, pH, and temperature attributes, and models are used to forecast crop output. For classification and clustering problems, many methods such as Naive Bayes, JRip, and 48 are used. The results suggest that the proposed models perform well.[1]

The paper outlines a proposed system for recognising soil series and identifying possible crops based on geographical features using machine learning algorithms. The process includes a training and testing step that uses soil and crop datasets. To determine the soil class, various machine learning approaches such as weighted K-NN, Gaussian Kernel based SVM, and Bagged Tree are utilized. The document also examines the development of soil and crop databases. The suggested methodology recognises soil series and suggests crops based on geographical features using machine learning algorithms.

The process includes a training and testing step that uses soil and crop datasets. To accomplish this, various machine learning approaches such as weighted K-NN, Gaussian Kernel based SVM, and Bagged Tree are employed.[2]

The study examines various categorization machine learning techniques, such as neural networks, decision trees, naive bayes, and support vector machines (SVM). The goal is to find the most accurate algorithm for identifying soil data. The report also describes the comparative approach and presents results in terms of accuracy and class recall. When compared to other algorithms, SVM has the highest accuracy. With the Important points being : The study compares classification machine learning techniques such as neural networks, decision trees, naive bayes, and SVM;The goal is to find the most accurate algorithm for identifying soil data;When compared to the other algorithms, SVM has the highest accuracy;The study presents accuracy and class recall results for each algorithm.[3]

The purpose of this work is to discuss the use of machine learning algorithms for soil categorization and crop detection. To improve the edges and contrast in soil photographs, decision trees, artificial neural networks (ANN), and support vector machines (SVM) are used. For subsurface soil categorization, the Cone Penetration Testing (CPT) approach is employed, and multiple algorithms are used depending on soil type. picture acquisition, picture pre-processing, feature extraction, and segmentation are all part of the proposed system. Using machine vision systems and pattern recognition techniques, the project intends to give a realistic solution for quick identification of soil types and crop detection.[4]

A big data analytics approach for predicting crop output and regulating fertilizer in agriculture is presented in this research article. The model receives and preprocesses data from several sources, employs a MapReduce technique based on nearest neighbors for data clustering, and employs a convolutional neural network for data classification. Using deep reinforcement learning, it makes recommendations to farmers based on optimal circumstances and fertilizer consumption given by an agronomist. The model's performance is assessed using multiple indicators and compared to existing approaches, confirming its superiority.[5]

The study paper focuses on applying machine learning approaches to estimate soil moisture to improve agricultural practices. It makes use of three datasets from distinct places and predicts using multiple linear regression (MLR), support vector regression (SVR), and recurrent neural network (RNN). MLR surpasses SVR and RNN in predicting soil moisture for 1, 2, and 7 days ahead, according to the data. The report also advises that more metrics, such as soil temperature and pH, be included in future soil health monitoring.[6]

The study paper in question examines soil data analysis utilizing data mining approaches, with a specific emphasis on soil fertility and attribute prediction. It introduces an automated system that uses a rule-based approach to classify soil samples based on fertility. The paper also uses the WEKA data mining tool to compare three classification algorithms - Naive Bayes, J48 (C4.5), and JRip - and concludes that J48 (C4.5) is the most successful classifier for soil samples. Furthermore, the paper uses regression algorithms to estimate untested numerical properties of soil samples, such as phosphorus concentration, and discovers that Linear Regression delivers the best accurate and efficient prediction. This prediction method is offered for soil testing facilities as a potential time and cost-saving measure.[7]

The paper is a research work that discusses a cluster-based approach to crop yield prediction utilizing the fuzzy c-means algorithm and support vector machine. It examines existing clustering algorithms and employs a real-time crop production dataset from India. Preprocessing, categorization, and clustering are proposed as steps in the paper. The suggested method's efficiency and accuracy are examined, and it is determined to be superior to both k-means and fuzzy c-means algorithms. The research finds that by comparing multiple variables on a large dataset, the suggested technique may effectively forecast crop yield.[8]

The research investigates the use of data mining methods in classifying agricultural land soils, with the goal of evaluating several classification strategies using a large soil profile dataset from Kanchipuram, India's Soil Science and Agricultural Department. To categorize soil data, which is initially prepared as an Excel spreadsheet and then transformed into a CSV file for analysis, several classifiers such as Naive Bayes, k-means, and SVMs are used. The results show that data mining approaches are successful in soil profile analysis, with Naive Bayes displaying noteworthy accuracy. The study emphasizes the importance of weather and soil features in agricultural decision-making by emphasizing the potential of data mining approaches in boosting soil management across agriculture, horticulture, and land use planning. These findings contribute to our understanding of soil classification and provide a foundation for future research in this area.[9]

III. Proposed System

With the use of machine learning algorithms, the suggested system seeks to transform agricultural decision-making by predicting crop compatibility, estimating production, and providing advice on nutrient management techniques. This is a detailed description of the system:

1. Predicting Crop Suitability: The system makes predictions about which crops would perform best in a given area or on a particular agricultural plot based on past data on soil type, climate, and crop performance.

- Using the dataset, machine learning algorithms—like decision trees or random forests—are trained to categorise crops according to how well-suited they are for various environmental circumstances.

- Through the analysis of variables like temperature, precipitation, soil pH, and nutrient levels, the system can advise farmers on which crops have the best chance of thriving in their particular area.

2. Estimating Crop Yield : - The system uses regression models, like random forest or linear regression, to predict crop yields depending on a number of variables, such as management techniques, soil properties, and weather patterns.

- Using historical yield data and environmental factors as a guide, the algorithm forecasts crop yields for a specific growing season. Farmers are able to prepare for post-harvest tasks like marketing and storage, predict harvest results, and allocate resources as efficiently as possible thanks to this knowledge.

3. User Interface and Accessibility: With a mobile application for convenient access from any device, the system is made to be user-friendly and accessible for farmers and agricultural stakeholders.

Users can enter pertinent data into the interface, such as crop preferences, soil type, and climate conditions, to receive predictions and recommendations that are specific to them.

- Users may understand the data and decide on crop management strategies with the use of dashboards and visualization tools.

- In order to maintain prediction relevance and accuracy, machine learning models are routinely retrained using the most recent data.

All things considered, the suggested method provides a thorough approach to raising agricultural resilience, sustainability, and production. Farmers may make well-informed decisions that maximize crop production, reduce risks, and ultimately support global food security and sustainability initiatives by utilizing machine learning and data-driven insights.

IV. METHODOLOGY

1. Data Collection and Preprocessing

- Data Source: Describe the source of the dataset used in the research, including its origin, format, and any preprocessing steps taken.
- Data Cleaning: Detail any data cleaning steps, such as handling missing values, removing duplicates, and ensuring data consistency.
- Feature Engineering: Explain any feature engineering techniques applied, such as encoding categorical variables or creating new features from existing ones.

2. Exploratory Data Analysis (EDA)

- Descriptive Statistics: Present descriptive statistics of the dataset, including measures of central tendency, dispersion, and distributions.
- Data Visualization: Showcase visualizations like histograms, box plots, and scatter plots to explore relationships between variables and uncover patterns in the data.

3. Feature Mapping

- District, Season, and Crop Mapping: Explain the process of mapping categorical variables like district names, seasons, and crop types to numerical values for modeling purposes.
- Mapping Dictionary: Describe the creation and usage of mapping dictionaries to convert string values to numeric representations.

4. Model Training

- Data Splitting: Discuss the method used to split the dataset into training and testing sets, including the proportion of data allocated to each.
- Model Selection: Explain the rationale behind choosing specific machine learning models for the task, such as Linear Regression, Random Forest Regressor, and Decision Tree Regressor.
- Model Training: Detail the training process for each selected model, including parameter tuning and cross-validation if applicable.

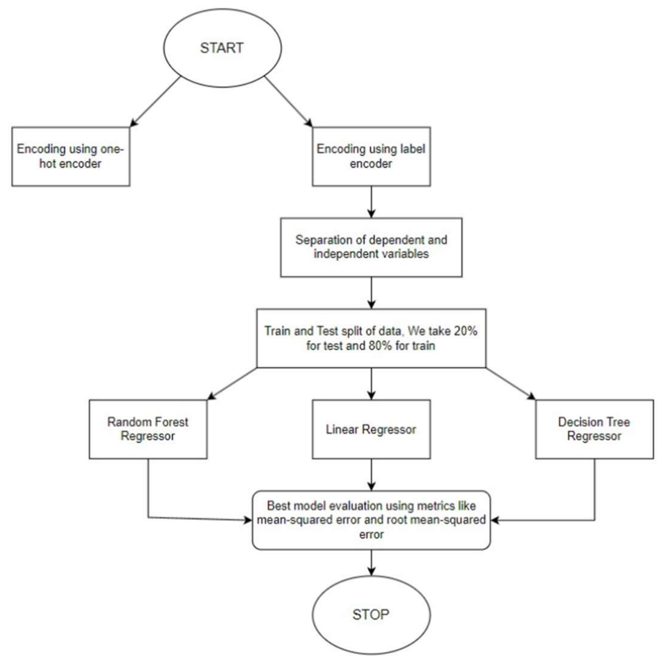
5. Model Evaluation

- Performance Metrics: Define the evaluation metrics used to assess the performance of the trained models, such as R-squared (R2) score and Mean Squared Error (MSE).
- Evaluation Results: Present the performance results of each model on both the training and testing datasets, highlighting any overfitting or underfitting issues.

6. Model Deployment

- Model Persistence: Discuss the saving of trained models to disk using pickle files for future deployment and inference.
- Deployment Considerations: Address considerations for deploying machine learning models in real-world applications, such as scalability, latency, and model versioning.

V. Architecture Diagram



VI. Results and conclusions

For the yield prediction model, we conducted comprehensive testing using three machine learning algorithms: Linear Regression, Random Forest, and Decision Tree. Through rigorous evaluation, we observed variations in prediction performance, interpretability, and computing efficiency across the models. Our results indicated that Random Forest exhibited the highest prediction accuracy, followed closely by Decision Tree, while Linear Regression showed slightly lower performance. Additionally, we constructed a comparative table showcasing the predictive capabilities of each model across various agricultural scenarios, providing valuable insights for farmers and stakeholders.

In the soil classification and crop recommendation model, we employed a diverse set of machine learning algorithms, including Support Vector Machine (SVM), Random Forest, Naive Bayes, Linear Regression, Multilayer Perceptron (MLP), and Artificial Neural Networks (ANN). Our experimentation revealed that the ANN-based method, leveraging deep learning architecture with multiple interconnected layers, outperformed other methodologies in terms of accuracy. Notably, the ANN model demonstrated superior performance in soil fertility assessment, crop recommendation, and yield prediction based on soil features.

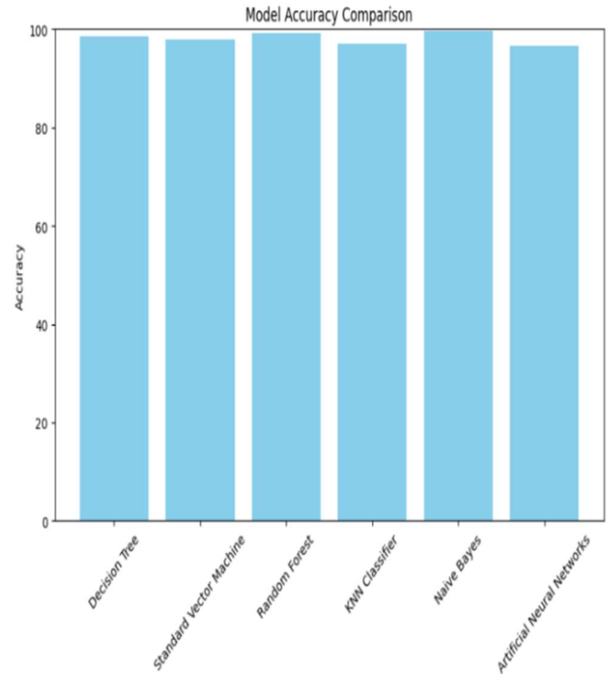
This research underscores the pivotal role of machine learning in advancing precision agriculture and optimizing agricultural practices. By harnessing extensive datasets and leveraging machine learning algorithms, we have demonstrated the efficacy of predictive models in enhancing crop management strategies, maximizing resource efficiency, and promoting sustainability in agriculture.

The comparative analysis of machine learning algorithms in both yield prediction and soil classification tasks highlights the importance of selecting appropriate models tailored to specific agricultural contexts. While Random Forest emerged as the top performer in yield prediction, the ANN-based approach showcased unparalleled accuracy in soil classification and crop recommendation, surpassing traditional methodologies.

Moreover, the development of a mobile application interface facilitates seamless access to actionable insights derived from machine learning models, empowering farmers and stakeholders to make informed decisions regarding crop selection, cultivation, and yield optimization based on soil characteristics.

Through the integration of machine learning techniques, this research endeavors to drive agricultural innovation, foster resilience against environmental challenges, and contribute to global food security. By bridging the gap between agricultural science and machine learning, we aim to pave the way for a sustainable and efficient agricultural ecosystem in the face of evolving socio-economic dynamics. Additionally, to visually represent the comparative performance of the models, we have included a table for yield prediction and a bar graph for soil classification, providing a comprehensive overview of their respective strengths and limitations.

	R2 score (Train Set)	R2 score (Test Set)	Mean Square error (Train Set)	Mean Square Error (Test Set)	Root mean Square error (Train Set)	Root mean Square Error (Test set)
Random forest regressor	0.988314	0.947761	36312350299 31.65	112399271767 79.188	1905579.97	3352600.0621576067
Decision tree regressor	1.0	0.936574	0.0025757216 06	13647052342515.98	0.050751567	3694191.703541653
Linear regressor	0.002758	0.00123579335 8	30988565611 0614.9	214901506930627.2	17603569.41	14659519.328089416



VII. Future Scope

The agricultural decision support system has a great deal of room to grow and improve.

1.Incorporation of Up-to-date Data Sources: Predictions and recommendations made by the system can be made more accurately and promptly by incorporating real-time data sources including satellite imaging, soil moisture sensors, weather forecasts, and Internet of Things devices. Farmers are better able to adjust to changing environmental conditions and make more educated decisions when they have access to current information.

2. Expansion of Crop and Region Coverage: By expanding the system's reach to encompass a larger spectrum of crops and locations, a greater number of farmers and agricultural stakeholders can benefit from its increased value. Including more crop varieties, growth environments, and geographic areas can enhance the system's worldwide adoption while accommodating a variety of agricultural landscapes.

3. Sophisticated Methods of Machine Learning: The predictive power and resilience of the system can be further increased by investigating cutting-edge machine learning techniques like deep learning, ensemble learning, and reinforcement learning. These methods can improve the accuracy of crop suitability forecasts, production estimates, and suggestions for nutrient management by revealing intricate patterns in agricultural data.

4. Crop Pest and Disease Management Integration: By incorporating modules for crop pest and disease management, farmers can receive complete assistance in reducing risks and maintaining crop health. The technology can assist farmers in preventing yield losses and minimizing the use of chemical pesticides by analyzing pest and disease outbreaks, finding early warning indications, and advising relevant actions.

5. Mobile Application Development: For farmers who would rather access information while on the go, creating a mobile application version of the system can improve accessibility and convenience. Features like customized alerts, offline data

access, and smooth data input and collection integration with mobile devices are all possible with a mobile app.

6. Community Engagement and Knowledge Sharing: You may encourage a culture of learning and innovation in agriculture by setting up venues for farmers, researchers, and agricultural specialists to collaborate, share knowledge, and engage with the community. Farmers can be empowered to adopt sustainable agricultural practices and adjust to changing conditions by the system by promoting the exchange of best practices, success stories, and local insights.

7. Integration with Supply Chain and Market intelligence: Farmers may increase profitability and optimize their marketing plans by integrating modules for supply chain management and market intelligence. The technology can help farmers make well-informed decisions about crop selection, harvest timing, and marketing channels by offering insights about customer preferences, market trends, and price fluctuations.

8. Backing for Agriculture That Is Climate-Smart: The system can be brought into line with more general sustainability objectives by including climate-smart agriculture concepts like carbon sequestration, resource efficiency, and climate resilience. Through the promotion of techniques like water-efficient irrigation, agroforestry, and conservation agriculture, the system can aid in the mitigation and adaptation of climate change in the agricultural sector.

References

[1].Jagdeep,Shalu,Viyayalakshmi(2021). Soil Analysis and Crop Fertility Prediction using Machine Learning.IJIRAE::International Journal of Innovative Research in Advanced Engineering,Vol: VIII,41-49

[2].SK AL Zaminur Rahman, Kaushik Chandra Mitra and S.M Mohidul Islam, "Soil Classification using Machine Learning Methods and Crop Suggestion Based on Soil Series", 2018 IEEE, 21st International Conference of Computer and Information Technology (ICCIT), pp.978-1-5386-9242-4/ 18.

[3] Pramudyana Agus Harlianto, Teguh Bharata Adji and Noor Akhmad Setiawan, "Comparison of Machine Learning Algorithm for Soil Type Classification", 2017 3rd International Conference on Science and Technology-Computer(ICST).

[4]. Ashw ini Rao, Janhavi U, Abhishek Gowda NS, Manjunatha and Mr s.Rafega Beham A, "Machine Learning in Soil Classification and Crop Detection", IJSRD International Journal for Scientific Research and Development, Vol-4,Issue 01, 2016, ISSN: 2321-0613

[5].Sini Anna Alex and Anita Kanavalli, "Intelligent Computational Techniques for Crops Yield Prediction and Fertilizer Management over Big Data Environment", International Journal of Innovative Technology and Exploring,Engineering (IJITEE), Vol-8, Issue-12, October 2019, ISSN: 2278-3075.

[6].Shikha Prakash, Animesh Shar ma and Sitanshu Shekhar Sahu, "Soil Moisture prediction using Machine Learning",2018 IEEE, 2nd International Conference on Inventive Communication and Computational Technologies(ICICCT).

[7]. Jay Gholap, Anurag Ingole, Jayesh Gohil, Shailesh Gargade and Vahida Attar, "Soil Data Analysis Using Classification Techniques and Soil Attribute Prediction", arXiv preprint arXiv:1206.1557.

[8].Dr.V.Geetha,M.Karthika, "A Cluster Based Approach for Crop Yield PredictionUsing Fuzzy C-Means Algorithm",ISSN NO: 1301-2746.

[9].Ramesh and K. Ramar, "Classification of Agricultural Land Soils: A Data Mining Approach",Agricultural Journal 6 (3): 82-86, 2011,ISSN: 1816-9155,O Medwell Journals, 2011.