A Data-Driven Approach to Agricultural Sustainability: Crop Recommendation using Machine Learning

Dr.G.Arutjothi[1], Dr.K.Geetha[2]

[1] *Assistant Professor, Dept. of Computer Applications, Sona College of Arts and Science,Tamil Nadu, India*
[2]*Lecturer, Dept. of Computer Applications, Government Arts College (Autonomous),Tamil Nadu, India*

---***---

Abstract - *This paper aims to develop a crop recommendation system using machine learning techniques with Python, designed to assist farmers in choosing the best crops for their fields. The system analyzes essential soil and environmental parameters such as Nitrogen (N), phosphorus (P), and Potassium (K) content in the soil, temperature, humidity, soil pH level, and rainfall. These parameters are utilized to train various machine learning models, including Logistic Regression, Support Vector Machine(SVM), K-Nearest Neighbor, Decision Tree, Random Forest, Bagging, and Gradient Boosting classifiers. The Random Forest Classifier was identified as the most accurate model through extensive evaluation. The system has been implemented as a user-friendly website to support agricultural officers in providing farmers with real-time data analysis for informed decision-making. This system strives to enhance farming efficiency, optimize resource usage, and improve crop yields by offering tailored crop recommendations. This paper represents a significant advancement in smart agriculture, leveraging technology to support sustainable farming practices and achieve better outcomes.*

Keywords: **Machine Learning, Recommendation System, Random Forest Classifier, Accuracy, Environmental Factors.**

1.INTRODUCTION

The Crop Recommendation System is an innovative initiative aimed at revolutionizing traditional farming methods by integrating advanced machine learning technologies. Its primary objective is to assist farmers in making informed, data-driven decisions regarding crop selection, which is essential for maximizing agricultural productivity and promoting sustainable practices. By utilizing machine learning, the system evaluates intricate datasets that include key soil and environmental factors, providing customized crop recommendations tailored to specific farm conditions. These parameters include soil nutrients such as Nitrogen (N), phosphorus (P), and Potassium (K), along with temperature, humidity, soil pH levels, and rainfall. The project employs Python, a highly adaptable programming language, to build and execute the machine-learning algorithms necessary for crop recommendation. Leveraging Python's powerful libraries, including scikit-learn, pandas, and numpy, the system efficiently handles data analysis and model development. Several machine learning models, such as Logistic Regression, Support Vector Classifier (SVC), K-Neighbors Classifier, Decision Tree Classifier, Extra Tree Classifier, Random Forest Classifier, Bagging Classifier, and Gradient Boosting Classifier, were trained and assessed based on performance metrics. Among these, the Random Forest Classifier stood out for its accuracy and robustness, making it the ideal choice for implementation.

To enhance accessibility, the system is presented as an intuitive web application developed using Visual Studio. The platform allows agricultural professionals and farmers to input real-time data and receive immediate crop recommendations. Features such as user-friendly data input forms, interactive visualizations, and detailed crop analysis reports ensure that users can easily navigate and utilize the system without requiring technical expertise.

The Crop Recommendation System project offers recommendations for various crops, catering to different regions and climatic conditions. Utilizing machine learning and a user-friendly website, it provides data-driven insights for farmers, enabling informed decision-making and improved crop yields. This project demonstrates the potential of technology in transforming agriculture, supporting sustainable farming practices, and contributing to food security and economic stability.This study represents a crucial advancement in promoting sustainable agriculture, enabling well-informed decision-making and contributing to enhanced crop yields and environmental conservation within the farming industry.

2. LITERATURE SURVEY

Grasslands, the world's largest terrestrial ecosystem, are a vital feed source for livestock. Addressing the growing global demand for meat and dairy products sustainably poses significant challenges. Technological advancements such as the Global Positioning System (GPS) and ground-based sensors show promise for grassland and herd management. Additionally, the increasing availability of spaceborne remote sensing data highlights the need to refine methods for exploiting such imagery. Biophysical parameter retrieval for grasslands has progressed from classical regression analysis to more sophisticated modeling techniques, though high-quality calibration and validation data remain critical. The development of hyperspectral satellite instruments and analytical methods could improve habitat discrimination and end-user operations, facilitating sustainable resource management [1].

Agriculture, a cornerstone of human livelihoods, must adapt to challenges such as environmental sustainability, climate change, and population growth while continuing to provide essential resources like food, fibers, and raw materials. Remote sensing offers a powerful tool for monitoring crop status at varying scales, supporting applications such as crop breeding, land use analysis, yield forecasting, and ecosystem services evaluation. Recent advancements in techniques and research enhance the applicability of remote sensing for agricultural applications, strengthening its role in providing efficient, long-term solutions [11].

India, ranking second globally in agriculture, faces challenges related to irrigation, fertilization, and crop rotation. To address these, the ACRIS system integrates Internet of Things (IoT) and machine learning technologies to improve crop yield. It features three key modules: the ARV Recommendation System, which identifies agricultural factors, and the AIM System, which predicts irrigation needs based on soil moisture, temperature, and humidity. This system supports precision farming by optimizing water-resource utilization and enabling effective monitoring of large-scale crop fields [7].

Food shortages driven by climate crises, wars, and the COVID-19 pandemic emphasize the need for efficient farmland management. Machine learning is increasingly applied to soil analysis, incorporating factors such as climate, soil chemistry, nutrients, and geolocations. Advanced systems combining soil analysis with data-driven models, satellite data, and remote sensors enable smart farming practices, promoting sustainability and higher crop yields [5]. Feder et al. propose a machine learning-based crop recommendation system to enhance sustainable agricultural practices. By analyzing parameters such as nitrogen, phosphorus, potassium, and humidity, algorithms like K-Nearest Neighbor (KNN), Decision Tree, Random Forest, and Support Vector Machine (SVM) predict suitable crops. This system aids farmers, researchers, and policymakers in decision-making, promoting increased crop productivity [4]. The rising global population underscores the importance of healthy crop production, which directly impacts economic stability. IoT, cloud computing, and machine learning have revolutionized various agricultural stages, from crop selection to harvest. The IoTSNA-CR model, an IoT-enabled soil nutrient classification and crop recommendation system, aims to reduce fertilizer use while maximizing productivity. This system collects data via IoT sensors, stores it in cloud memory, and applies machine learning techniques for analysis. By integrating hybrid algorithms, the model achieves high accuracy (97.3%) and supports efficient soil management, reducing costs while boosting productivity [8]. Despite their potential to enhance agricultural productivity, technological innovations face barriers to adoption in less developed countries. Challenges such as limited access to credit, information, and infrastructure, coupled with aversion to risk and small farm sizes, hinder widespread implementation. Development initiatives focusing on credit provision, information dissemination, and infrastructure investment have seen only partial success. Uniform adoption of these innovations remains a significant challenge [3].

Machine learning is increasingly applied to solve agricultural problems, particularly in crop recommendation and yield prediction. Reddy et al. propose a three-tiered framework comprising data preprocessing, classification, and performance evaluation. Techniques like correlation analysis, ensembling, and majority voting are utilized, with the Naïve Bayes classifier achieving a high accuracy of 99.54%, outperforming ensemble methods. The study highlights future research directions and challenges in implementing machine learning for agriculture [6]. Similarly, Storm et al. analyze machine learning approaches from an economist's perspective, discussing limitations like inflexible functional forms and complex simulation models, and suggesting solutions to address these issues [10]. Sheykhmousa et al. conduct a meta-analysis of 251 journal papers to review the performance of Random Forest (RF) and Support Vector Machines (SVM) in remote sensing image classification. The study evaluates factors such as spatial resolution, feature engineering, and data types, providing recommendations and identifying challenges to guide future research efforts [9].

Ftom this literature servay we have find the agricultural sector faces challenges such as crop selection based on soil and climatic conditions, climate change variability, and lack of a systematic, data-driven approach to crop recommendation. Traditional methods often rely on intuition and historical practices, leading to suboptimal yields and financial losses. Climate change also complicates crop selection, with changes in temperature, rainfall, and humidity affecting crop growth and yield. The lack of comprehensive data and tools for farmers also contributes to inefficient resource use, increasing costs and affecting soil health and biodiversity. A machine learning-based crop recommendation system can help farmers make more informed decisions, leading to better yields, efficient resource use, and sustainable agricultural practices. This project aims to address these issues by developing a machine learning-based system.

3. METHODOLOGY

3.1 Proposed System

The proposed Crop Recommendation System aims to address the limitations of the existing system by leveraging advanced technologies, particularly machine learning, to provide data-driven crop recommendations. The proposed system utilizes extensive data collection from various sources, including soil nutrient sensors, environmental sensors, and weather stations. This data is used to create a comprehensive profile of the soil and environmental conditions. Machine learning models are trained on this data to identify patterns and make accurate predictions about the most suitable crops for a given piece of land. This data-driven approach ensures that recommendations are based on empirical evidence rather than anecdotal knowledge. Several machine learning models, including Logistic Regression, Support Vector Classifier (SVC), K-Neighbors Classifier, Decision Tree Classifier, Gradient Boosting Classifier, are evaluated to determine the most accurate model. The Random Forest Classifier is chosen for its high accuracy and robustness in handling various data types. This model uses ensemble learning to improve predictive performance by combining multiple decision trees. The proposed architecture is given in figure 1.

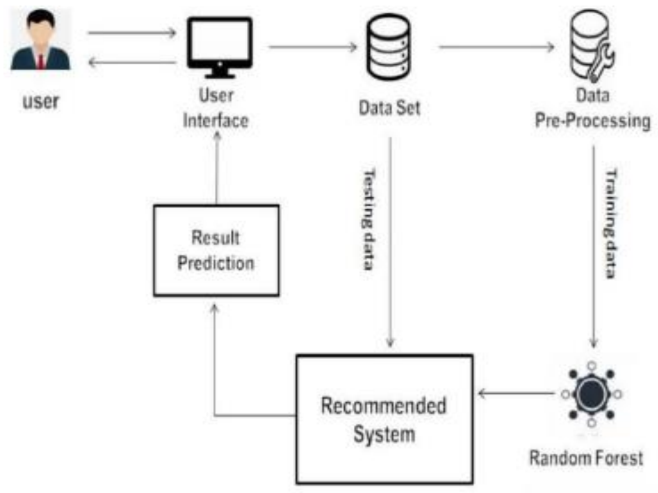

Figure 1: The Proposed Architecture

The proposed system is designed to be adaptive and responsive to changing environmental conditions. It continuously updates its recommendations based on real-time data from sensors and weather stations. This adaptability makes the system more resilient to the effects of climate change, helping farmers select crops that are more likely to thrive under variable conditions.

3.2 Data Collection and Preprocessing

Crop Recommendation System focuses on capturing essential soil and environmental parameters from the user. These parameters include Nitrogen (N) ratio, Phosphorous (P) ratio, Potassium (K) ratio, temperature, humidity, soil pH, and rainfall are collected from sensors and agricultural datasets. This work collected sample dataset by Kaggle [12]. This dataset contain 2200 instances and 7 attributes from this dataset 80% data is for traning and 20% data is for testing. The data is cleaned and preprocessed using normalization and transformation techniques. The attributes are mentiones in the below table.

Table1: Soil Dataset attributes

Attribute Name	Description
nitrogen	Ratio of Nitrogen content (%)
phosphorous	Ratio of Phosphorous content (%)
potassium	Ratio of Potassium content (%)
temperature	Temperature in degrees Celsius
humidity	Relative humidity (%)
pH	Soil pH value
Rainfall	Rainfall in mm

4. RESULTS AND DISCUSSION

This section presents the findings and model accuracy comparisons.The preprocessed data fed into various machine learning models were trained and evaluated, including Logistic Regression, SVC, K-Neighbors, Decision Tree, Extra Tree, Random Forest, Bagging, and Gradient Boosting Classifiers. The Random Forest Classifier was selected for its superior accuracy. The Python language is used to implement the models. Table2 contains the accuracies of various models.

Table2: Results

Table: Model Accuracies Model	Accuracy
Logistic Regression	0.9636
SVC	0.9682
K-Neighbors Classifie	0.9591
Decision Tree Classifier	0.9886
Extra Tree Classifier	0.8977
Random Forest Classifier	0.9932
Bagging Classifier	0.9886
Gradient Boosting Classifie	0.9818

The below Bar chart to show these accuracies visually.

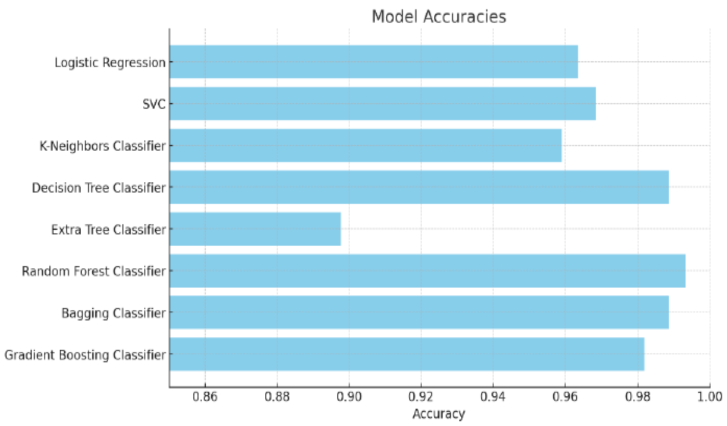

Figure 2: Comparision of Classifier Accuracy

The table2 and bar chart(figure2) demonstrate that Random Forest achieved the highest accuracy among all tested models, making it the best choice for crop recommendation. Due to its ability to provide precise predictions by combining multiple decision trees, Random Forest was selected to ensure reliable crop recommendations.

5. CONCLUSION

The Crop Recommendation System effectively integrates advanced machine learning models with a user-friendly interface to assist farmers in selecting the most suitable crops. This project highlights the potential of data-driven approaches to enhance agricultural productivity and sustainability. The Random Forest Classifier obtained 99.3% of accuracy, which is higher than the other machine learning models. So that this is the suitable model for crop recommandation systems. In conclusion, this research not only demonstrates the effective use of machine learning in agriculture but also highlights the importance of user-centric design and real-time data analysis. The Crop Recommendation System is a valuable tool that supports farmers in optimizing their agricultural practices, leading to better yields, efficient resource use, and sustainable farming. In future, this research will focuses on to implementing Artificial Intelligence(AI) based crop recommandation system.

REFERENCES

[1] Ali, I., Cawkwell, F., Dwyer, E., Barrett, B., & Green, S. (2016). Satellite remote sensing of grasslands: from observation to management. *Journal of Plant Ecology*, *9*(6), 649–671. https://doi.org/10.1093/jpe/rtw005

[2] Breiman, L. (2001). Random forests. *Machine Learning*.

[3] Feder, G., Just, R. E., & Zilberman, D. (1985). Adoption of Agricultural Innovations in Developing Countries: A Survey. *Economic Development and Cultural Change*, *33*(2), 255–298. https://doi.org/10.1086/451461

[4] Gupta, V., & Kumar, M. (2023). Crop Recommendation System Using Random Forest Algorithm. *IEEE Xplore*.

[5] Huang, Y., Srivastava, R., Ngo, C., Gao, J., Wu, J., & Chiao, S. (2023). Data-Driven Soil Analysis and Evaluation for Smart Farming Using Machine Learning Approaches. *Agriculture*, *13*(9), 1777. https://doi.org/10.3390/agriculture13091777

[6] Reddy, L., & Nair, P. (2022). Smart Crop Recommender System. *IEEE Xplore*.

[7] Saleh, A. R. B. M., Venkatasubramanian, S., Paul, N. R. R., Maulana, F. I., Effendy, F., & Sharma, D. K. (2022). Real-time Monitoring System in IoT for Achieving Sustainability in the Agricultural Field. *2022 International Conference on Edge Computing and Applications (ICECAA)*. https://doi.org/10.1109/icecaa55415.2022.9936103

[8] Senapaty, M. K., Ray, A., & Padhy, N. (2023). IoT-Enabled Soil Nutrient Analysis and Crop Recommendation Model for Precision Agriculture. *Computers*, *12*(3), 61. https://doi.org/10.3390/computers12030061

[9] Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Homayouni, S. (2020). Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 6308–6325. https://doi.org/10.1109/jstars.2020.3026724

[10] Storm, H., Baylis, K., & Heckelei, T. (2019). Machine learning in agricultural and applied economics. *European Review of Agricultural Economics*, *47*(3), 849–892. https://doi.org/10.1093/erae/jbz033

[11] Weiss, M., Jacob, F., & Duveiller, G. (2019). Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, *236*, 111402. https://doi.org/10.1016/j.rse.2019.111402

[12] Kaggle: Python and machine learning datasets - https://www.kaggle.com/learn/overview

BIOGRAPHIES

Dr. G. Arutjothi, M.Sc., Ph.D., is an Assistant Professor in the Department of Computer Applications at Sona College of Arts and Science, Salem, Tamil Nadu. She earned her Ph.D. in Computer Science from Periyar University, Salem, in 2024. With expertise in Data Mining, Big Data Analytics, Cloud Computing, and Artificial Intelligence, she has published more than 12 research papers in international journals and conferences. Dr. Arutjothi has over five years of teaching experience and eight years of research experience, with a strong background in programming languages such as Python, C, C++, and Java. She has actively participated in faculty development programs, organized international conferences, and contributed as a reviewer for reputed journals. Her commitment to advancing computer science education and research is evident through her involvement in academic and professional activities.

Dr. K.Geetha completed her MCA in Periyar University and M.Phil in PRIST University, Thanjavur and received Doctorate from Periyar University in 2025. She is doing her service as a ernment Arts College (Autonomous), Salem-7 for the past 12 years along with 10 years of Research experience. She published more 10 papers through conferences and Journels. Which includes IEEE and SCOPUS that are available in online. Her research contribution is for Data Deduplication, Big data Analytics, Social Network analysis, AI and ML.