Agro-Insight: Optimal Crop Recommendation for Soil Properties and Climatic Factors using Machine Learning Algorithms

Siyona Snotra[1], Vaibhav Tomar[2], Neha Singhal[3*]*

[1]PG Scholar, Dept. of Computer Science,
Christ (Deemed to Be University), Bengaluru, India
E-mail: siyona.snotra@mca.christuniversity.in
[2]PG Scholar, Dept. of Computer Science,
Christ (Deemed to Be University), Bengaluru, India
E-mail: vaibhav.tomar@mca.christuniversity.in
[3]Assistant Professor, Dept. of Computer Science,
Christ (Deemed to Be University), Bengaluru, India
E-mail: neha.singhal@christuniversity.in

Abstract

Agriculture is confronted with major challenges in choosing the most appropriate crops to be cultivated, as soil quality, climate fluctuation, and inefficient management of resources tend to result in low yields. To overcome this, we suggest a Machine Learning-based Crop Recommendation System that utilizes algorithms like SVM, Decision Trees, Random Forest, and Naive Bayes where Random-Forest gave us 99.32% accuracy. Our approach analyses key agricultural factors—soil nutrients, temperature, humidity, and rainfall—to provide evidence-based recommendations to farmers. Also, we have a Recommendation History Page, where past predictions and user comments are kept, enabling personalized insights and increasing the accuracy of the system in the long run. This enables better decision-making via the capability to monitor trends and alter decisions. Our system is especially useful for precision agriculture, resource management, and sustainable farming, ultimately empowering farmers with affordable, AI-based solutions to enhance productivity and efficiency. Future development will focus on enhancing usability and scalability to support mass deployment.

Keywords: Agriculture, Crop Recommendation, SVM, Random Forest, Naive Bayes.

1. Introduction

It is one of the principal components of human survival and economic development, especially in India, and accounts for a considerable percentage contribution to the GDP and absorbs a major chunk of the population. However, the agricultural sector is replete with challenges such as climate change, soil degradation, and erratic rainfall patterns. These problems, not only reducing crop productivity, are causing extreme economic distress to farmers, and some are even compelled to extreme steps due to financial instability (N. K. et al., 2020). Such problems need innovative solutions, and one of the promising approaches has been precision agriculture. Precision agriculture focuses on site-specific farming practices and advanced data analysis to optimize the use of resources and enhance crop yield (Patel et al., 2024). Predictive

M. Sharma et al. (eds.), *Proceedings of the Third International Conference on Recent Advances in Computing Sciences (RACS 2025)*, Advances in Intelligent Systems Research 212,
https://doi.org/10.2991/978-94-6239-768-2_3

algorithms such as Decision Trees, Support Vector Machines (SVM), Random Forest, and XG-Boost have shown to effectively present suitable crops and predict yield with an increased level of accuracy (Reddy et al., 2021). Systematic literature reviews highlight the importance of choosing the appropriate algorithms and building scalable data-driven solutions specific to agricultural needs (van Klompenburg et al., 2020).

In 2021, this study claims analysis was done in such parameters to help the farmer tailor insight to give the right decisions regarding crop selection. (Gosai et al., 2021) The study undertakes the thorough evaluation of the performance of a variety of machine learning algorithms on crop prediction performance. This allows an identification of their strengths and weaknesses, as well as how different agricultural scenarios play out under application and their restrictions (Wani et al., 2022).

2. Literature review

An extremely informative survey in the year 2024 related to the integration of Internet of Things (IoT) technology with ML methodologies was conducted to improve the smart farming practices. The present study, therefore, gives a background as to how real-time data collection is possible through measures of environmental conditions such as temperature, soil moisture, and humidity as backbone data to develop correct models. The authors have computed the data with Different ML techniques such as linear regression, decision trees, random forests, and extreme gradient boosting. (Widianto et al., 2024).

A published study in 2023, Garg and Alam, who design the Wrapper PART Grid method, based on crop recommendations, showed real-time data collection. The wrapper technique for feature selection was used along with the classification methodology model. The accuracy results obtained were 99.31% with very high accuracy, results meant to help farmers in decisions to do with crop production in favor of production increases and minimizing crop losses (Garg et al., 2022).

In 2023, a systematic literature review of the proposed crop and market price prediction systems was proposed using the machine learning technique to emphasize how important it is to help farmers in decision-making using data. Crop and Market Price Prediction: Overview of Methods for Decision Tree Algorithms through Artificial Neural Network Prediction of Crop Yield from Temperature and Rain to Soil Types (Sagar et al., 2023).

Kharde et al. carried out comprehensive review work on front-runners of applications of ML and the Internet of Things in agriculture. The authors examine several machine learning techniques, including Support Vector Machines (SVM), Random Forests (RF), Decision Trees (DT), and Artificial Neural Networks (ANN), which are frequently employed in studies related to soil health assessment and crop monitoring. Furthermore, they emphasize the significant benefits these methods offer for decision-making, enhancing productivity, optimizing resource utilization, and promoting sustainability within the agricultural framework in India (Kharde et al., 2022).

In 2021, publication" AgroSys– A Crop Recommendation System" released the new principle that can increase efficiency in agriculture via a machine learning based recommendation system. The authors built a neural network named Multilayer

Perceptron as AgroSysNN. The system used ensemble learning techniques with majority voting and combined the K Nearest Neighbor and Naive Bayes models to boost the predictive accuracy. A great model that obtains 99.7% precision has high opportunities for application to precision agriculture with this new technology. (Kadam et al., 2021).

In 2021, a new approach came in crop recommendation that relied on using Machine Learning and IoT technology in ”Crop Recommendation on Analyzing Soil Using Machine Learning.” Their study suggested the use of precision farming methods in engaging with environmental and climatic variability while using IOT sensors to determine critical soil parameters such as moisture, temperature, and humidity linked to a certain pH. The authors provided a machine learning model that successfully managed to attain a ratio of about prediction accuracy of 96.89 percent in predicting an appropriately suited crop. (Anguraj et al., 2021).

A paper by Parikh et al. has been put up, and in this paper, the identification of a Machine Learning crop determination System was found to aid farmers in propounding the best crops for the land by improving farm production and waste of resources. Applying the available reliable dataset on Kaggle, the authors used and implemented several different algorithms from machine learning designed for determining factors such as soil type, temperature, humidity, and rainfall through the Random Forest Classifier algorithm. (Parikh et al., 2021).

3. Methodology

In the study, a clearly defined machine learning pipeline was implemented to build a stable crop classification model, which was reproducible process, beginning with data acquisition and ending at prediction, while allowing efficient model training and implementation into real-world situations relevant to agriculture.

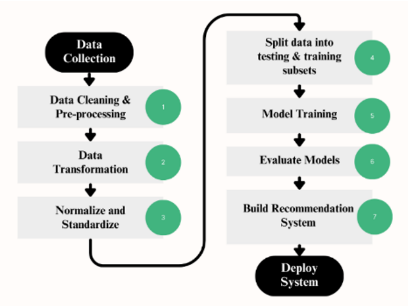


Figure 1: Flowchart 0f the proposed system

Figure 1 involves the following steps:

3.1 Data Collection and Pre-processing

The dataset from this investigation contained several agricultural parameters such as temperature, humidity, pH, rainfall, phosphorous (P), potassium (K), nitrogen (N), and rainfall. The following steps were taken in the pre-processing stage:

Data Cleaning: The cleaning stage included the process of reviewing the entire data set to identify duplicate records, missing values, or inconsistencies in data quality. Removal of outliers was considered to reduce bias during training to ensure that the model maintains high predictive power

Feature Engineering: Categorical variables were converted to numerical values using a mapping dictionary. This was to make things easier for the algorithms in processing the categorical data in this manner.

Exploratory Data Analysis: Correlation analysis was performed to identify relationships between input features and their impact on crop classification. A heat-map visualization was used to highlight significant correlations, ensuring meaningful attributes were retained for model training.

3.2 Normalization and Standardization:

Two scaling methods were applied in order to achieve better model performance:

Min-Max scaling: Rescaled features to a fixed range for easy comparison between different attributes.

Standard scaling: Ensured that the distribution of data was uniform so as to improve the classification process, thus providing better accuracy through better functioning of distance-based models.

3.3 Feature Selection and ML Model Training

Feature-Selection was used to identify the most influential parameters for crop categorization. This, increases the model's interpretability, removes correlations between features, and escapes from noise in the dataset. Important aspects have been correlation matrices and wrappers. The major performance metric was the accuracy of algorithms evaluated using the training dataset. For classification, the following machine learning algorithms were implemented and evaluated.

3.4 Model Selection and Evaluation

Model accuracy scores were determined based on their classification performance after the completion of the training phase. **The Random Forest** Classification algorithm gave the highest accuracy at **99.32%**, indicating better prediction than the other models. Both of the ensemble methods, Random Forest and the Bagging Classifier, showed better generalization for crop prediction. Besides, feature importance analysis of the Random Forest model revealed that nitrogen, phosphorus, potassium, and rainfall were some of the most important factors affecting the crop's suitability. It further points in the direction of agricultural best practices and the model's validation in this system.

3.5 Prediction and Deployment Potential

To address the divide between research and application, the machine learning model that was developed through this research is implemented as a web-based decision support system. This system allows users (principally farmers and agricultural professionals) to input parameter values for nitrogen, phosphorus, potassium, temperature, humidity, pH, and rainfall. The program suggests which crop was the best for farmers to grow taking all of the inputs into account using the model. The web application was implemented with Flask (Python) for the backend that connects to the trained model. HTML/CSS/JavaScript were employed to develop the web application frontend in a manner that is user-friendly and responsive. The application demonstrates a high level of usability and supports the principles of smart farming by providing precise and timely recommendations on crops. This arrangement demonstrates an implementation strategy for the study, while allowing the benefits of machine learning to be received by the end-user in a usable and meaningful way.

4. Results and Implementation

During the course of research, we developed a crop-recommendation system based on machine-learning called **Agro-Insight** to help farmers choose the best crop that would give them, the most yield and thus profits. The model was built using various parameters related to soil nutrient levels, weather, and climatic factors such as pH, temperature, humidity, and rainfall. We also compared various machine-learning algorithms and measured their accuracy to choose the best model possible.

The data were fed into a pre-processing phase, where data cleaning was performed to correct inconsistencies, missing values, and outliers, thus maintaining the integrity of the dataset. Finally, the dataset was divided into subsets, 80% samples were used for training & 20% of the samples got used for testing purpose, facilitating an unbiased assessment of the model's predictive efficacy.

After pre-processing, several machine-learning techniques were applied to determine the model which has been effective for crop-recommendation. The evaluated algorithms including Logistic Regression, SVM, KNN, Decision-Trees, Random-Forest, Bagging, Ada-Boost, Gradient-Boosting, Extra Trees.

The results declared that the **Random Forest model** performs excellently in crop classification, showing nearly faultless values in precision, recall and F1-scores for most crop classes. The model has a good prediction accuracy of **99.32%** overall, so the wrong prediction for a certain crop scarcely occurs.

Figure. 2 mentions accuracy 0f various ML models used in this study. Every algorithm was trained with the use of training dataset and evaluated using a testing dataset based on their accuracy in predicting suitable crops below:

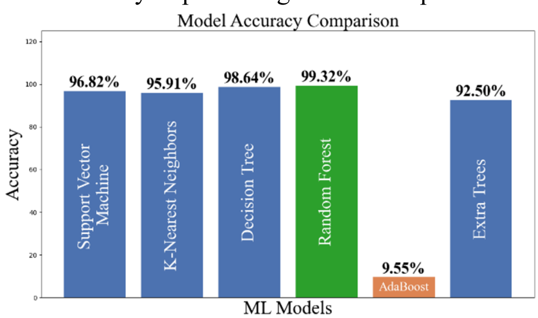
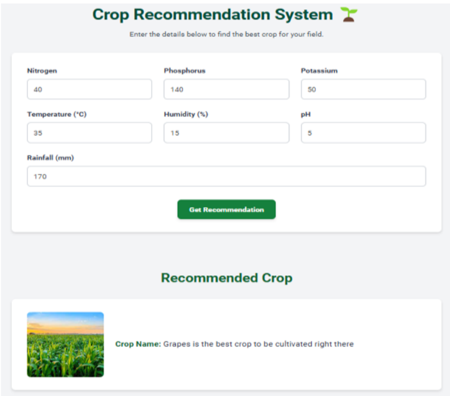


Figure 2: Bar graph showing the accuracies of various machine learning models used in the study.

The model achieved full precision (1.00) and recall (1.00) for commodities like maize, cotton, coconut, papaya, apple, mango, and pomegranate, i.e., the model classifies all instances of these commodities while minimizing false alarms. This means that distinctive and well-trained patterns are present in the dataset for individual crops, which got used for classification by the model.

In most major cases, the Rice (Recall: 0.89) and Jute (Precision: 0.92) were slightly misclassified by the model. The low recall for Rice (0.89) indicates that some Rice instances might have been classified as other crops, leading the model to skip certain conditions in which Rice should be considered for recommendation. By the same token, Jute scored 0.92 in precision, indicating that a very small percentage out of all predictions categorized as Jute were actually other types of crop. The minor inconsistencies might have happened by overlapping feature distributions or similarity of soil and climatic conditions existing between these crops and others.



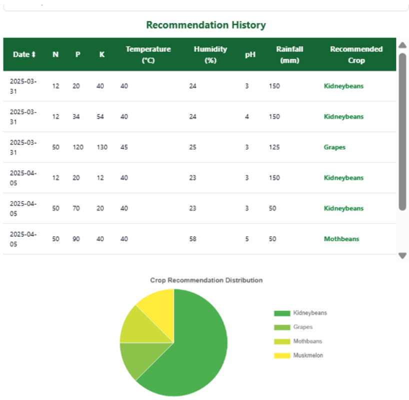


Figure 3 shows Screenshot of the Agro-Insight application interface showing crop recommendation results & Figure 4 contains Screenshot of the Agro-Insight's Recommendation History Page

Figure 3 showcase screenshots of the Agro-Insight application, demonstrating its intuitive layout and ease of data input. Furthermore, the user interface of this website has been designed to enhance usability and provide continuous crop recommendations.

Overall, the model's perfect F1-scores (1.00) for 20 out of 22 crops confirm its reliability. The balanced performance across precision and recall ensures that both false-positives, false-negatives are minimized, making Random Forest, best choice for crop recommendation.

These findings highlight the potential of integrating machine learning into agricultural decision-making, helping farmers improve crop yield and resource management.

Figure 4 represents the **Recommendation History** page, a unique aspect of the developed system which increases the utility of conventional decision support systems by adding personalized record history and visual analytics. While many decision support systems offer the one-time recommendations, this system conserves each user's recommendations and inputs for crop type as per the seven factors. By monitoring these types of applications in a historical fashion, a user can observe what they put in the fusibility and how it aligned or differed from the previous recommendations. In order to add the interpretability, the page provides an interactive crop distribution chart to visually summarize recommended crop(s) and their past inputs. As with other aspects of the system that have focused on enhanced user experience, users will receive a consistent feel of a data-driven, visually representative experience which empowers them to make sustainable decisions over time.

5. Conclusion

To summarize, the study illustrates that machine-learning algorithms are capable of making agricultural decision-making easier via the Agro-Insight crop recommendation system. Equipped with a comprehensive database that bears in mind critical soil properties and climatic factors, we have created strong models capable of providing tailor-made crop recommendations for specific regional conditions. This research explored multiple methods of machine learning, and the best performing models were Random-Forest and Bagging models, which suggest they can suggest suitable crops, and improve agriculture. Results of this study underline the necessity for data-based information about contemporary agriculture, in the face of a much higher global demand for food, and sustainable agricultural practices.

These findings emphasize the demand for data-based insights on contemporary agriculture, across the backdrop of increased global food demand and sustainable agriculture practices. Agro-Insight's mission is to enhance crop production by empowering farmers in their data-driven decisions in real-time to ensure resource efficiency and environmental sustainability. Moreover, the friendly interface of the system is user-oriented to promote the adoption process so that advanced technology is made available to all.

6. Future Scope

Future efforts should investigate a larger dataset with more agricultural areas, additional machine learning methods to provide improved predictive abilities, and the utilization of IoT technologies to create real-time data and enhance predictive abilities. Improving interactive visual tools and multilingual support will further develop the Agro-Insight platform, therefore facilitating access to a larger audience. Overall, this study demonstrates the profound impact, machine-learning technologies can have on agriculture and provides a technological approach to the ever-growing agricultural challenges.

References

N. K., D. Vishwakarma, M. N., A. Ashwini, and M. M.R., ”Crop Prediction using Machine Learning Approaches,” International Journal of Engineering Research & Technology (IJERT), vol. 9, no. 8, Aug. 2020.

M. Patel, A. Rane, and V. Patni, ”Crop Recommendation System using Machine Learning,” International Journal of Scientific Research in Science and Technology, vol. 11, no. 2, pp. 57-63, Aug. 2024. doi: 10.32628/IJSRST2408523.

D. J. Reddy and M. Rudra Kumar, ”Crop Yield Prediction using Machine Learning Algorithm,” in Proceedings of the Fifth International Conference on Intelligent Computing and Control Systems (ICICCS 2021), Anantapur, India: IEEE, 2021.

T. van Klompenburg, A. Kassahun, and C. Catal, ”Crop yield prediction using machine learning: A systematic literature review,” Information Technology Group, Wageningen University & Research, Wageningen, The Netherlands, 2020.

D. Gosai, C. Raval, R. Nayak, H. Jayswal, and A. Patel, ”Crop Recommendation System using Machine Learning,” International Journal of Scientific Research in Computer Science, Engineering and Information Technology, doi: https://doi.org/10.32628/CSEIT2173129.

A. Wani, V. Yadav, R. Todankar, P. Wade, and S. Malik, ”Crop Prediction using IoT & Machine Learning Algorithm,” International Research Journal of Engineering and Technology (IRJET), vol. 9, no. 4, Apr. 2022.

M. H. Widianto, Y. D. Setiawan, B. Ghilchrist, and G. Giovan, ”Smart farming based on IoT to predict conditions using machine learning,” International Journal of Reconfigurable and Embedded Systems (IJRES), vol. 13, no. 3, pp. 595-603, Nov. 2024, doi: 10.11591/ijres.v13.i3.pp595-603.

D. Garg and M. Alam, ”An effective crop recommendation method using machine learning techniques,” International Journal of Advanced Technology and Engineering Exploration, vol. 10, no. 102, 2022, doi: 10.19101/IJATEE.2022.10100456.

Sagar, S.H., Dravya, K.L., Shindhe, R.C., Savanth, H.A., Kumar, P.K.: Crop and Market Price Prediction System Using Machine Learning. Int. Res. J. Mod. Eng. Technol. Sci. 5(5). (2023).

Y. Kharde, S. Mahajan, C. Galande, R. Gondkar, and S. Talole, ”Study of Revolutionary use of Machine Learning and IoT based Model in Agricultural Sector: A Review,” International Journal of Engineering Technology and Management Sciences, vol. 6, no. 6, pp. 1-6, Oct.-Nov. 2022, doi: 10.46647/ijetms.2022.v06i06.043.

P. D. Kadam, R. S. Chavan, A. M. Kulkarni, and S. R. Janrao, ”AgroSys – A Crop Recommendation System,” International Research Journal of Engineering and Technology (IRJET), vol. 8, no. 5, May 2021.

Anguraj, K., Thiyaneswaran, B., Megashree, G., Shri, J. P., Navya, S., & Jayanthi, J. (2021). Crop recommendation on analyzing soil using machine learning. Turkish Journal of Computer and Mathematics Education, 12(6), 1784-1791.

P. A, S. Chakraborty, A. Kumar and O. R. Pooniwala, ”Intelligent Crop Recommendation System using Machine Learning,” 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2021, pp. 843-848, doi: 10.1109/ICCMC51019.2021.9418375.