GIS-Enhanced Crop Yield Modeling with Machine Learning

Venkatesh S.D.[1], Chitra K.[1], Harilakshami V.M.[1]

[1]Dayananda Sagar Academy Of Technology And Management, Karnataka, India

Email: venkateshsd3103@gmail.com, chitra-mca@dsatm.edu.in, harilakshmivm001@gmail.com

Abstract

India, with its vast population and agrarian society, faces challenges in agricultural practices. Many farmers continue to grow the same crops repeatedly without experimenting with new varieties. To address these issues, we have developed a system using machine learning algorithms aimed at helping farmers. Our system recommends the most suitable crops for specific lands based on soil content and weather conditions. It also provides information on the appropriate type and number of fertilizers and the necessary seeds for cultivation. By using our system, farmers can diversify their crops, potentially increase their profit margins, and reduce soil pollution.

Keywords

Classification Algorithms, Decision Tree, KNN, Machine Learning in Agriculture.

Introduction

Agriculture is a key occupation in India, essential for its economic and developmental progress. With over 60% of land dedicated to farming, adopting new agricultural technologies is crucial. Traditional crop prediction relied on farmers' experience, often leading to soil nutrient deficiencies and reduced yields. To address this, we developed a machine learning-based system that recommends the best crops based on soil and weather conditions. Utilizing algorithms like Support Vector Machine (SVM) and Decision Tree, the system analyzes data such as rainfall, temperature, humidity, and ph. It advises on suitable crops, necessary nutrients, seed requirements, yield estimates, and market prices, enhancing farmers' profitability and soil health.

The traditional methods of crop prediction and yield estimation in India rely heavily on farmers' experience, often leading to the repeated cultivation of the same crops and inadequate fertilizer use. These results in reduced yields, soil acidification, and topsoil damage. There is a critical need for a technology-driven solution to recommend optimal crops and nutrient requirements based onprecise soil and weather data. This project aims to develop a machine learning-based system that analyzes environmental parameters to suggest the most suitable crops, appropriate fertilizer amounts, and provide yield and market price forecasts, thereby improving agricultural productivity and sustainability.

Submission: 13 October 2024; **Acceptance:** 29 November 2024

Literature Review

Kushwaha and Bhattachrya discuss methods for predicting crop yield and recommending suitable crops to enhance farmers' profits and the agricultural sector's quality (Kushwaha and Bhattachrya, 2015). The paper utilizes large volumes of soil and weather data, referred to as big data, processed on the Hadoop platform using agro algorithms. By analyzing this extensive data repository, the system predicts the most suitable crops for specific conditions and suggests ways to improve crop quality, ultimately aiding in better agricultural decision-making and productivity.

Zingade and his team members explores the agriculture is vital to India's economy, contributing 20% to the GDP (Zingade et a.., 2020). This project aims to modernize agriculture using machine learning, specifically the Naïve Bayes algorithm, to help farmers decide which crops to grow based on soil, weather, temperature, and rainfall, maximizing profit with minimal investment.

Girish explores methods for predicting crop yield and rainfall using machine learning techniques (Girish, 2018). The study examines various approaches such as linear regression, SVM, KNN, and decision tree algorithms. Through their analysis, they find that SVM exhibits the highest efficiency among these methods for predicting rainfall. The research highlights SVM's capability to accurately forecast rainfall patterns, providing valuable insights for agricultural planning and management.

Patil and team explore various machine learning methods aimed at enhancing crop yield (Patil et al., 2020). The research delves into artificial intelligence techniques, including machine learning algorithms and big data analysis tailored for precision agriculture. Specifically, the paper details the development of a crop recommender system utilizing approaches such as KNN, ensemble-based models, and neural networks. These methods are designed to leverage data-driven insights to recommend optimal crops based on specific environmental and agricultural conditions, thereby improving productivity and sustainability in agriculture.

Koti and Supriya have discussed some of the problems faced by the farmers, how artificial intelligence is used in the agricultural sector, need for weather predictions and proposed algorithm for irrigation using machine learning techniques (Koti and Supriya, 2021).

Methodology

In the existing system, an agro-based country's economic growth heavily relies on agriculture. As the population grows, so does the dependency on agriculture, impacting the country's economic progress. In this context, the crop yield rate becomes crucial. To boost crop yield, various biological methods (such as improving seed quality, crop hybridization, and using strong pesticides) and chemical methods (like applying fertilizers, urea, and potash) are employed. Additionally, crop sequencing techniques are essential to enhance the overall yield throughout the season. One existing approach is the Crop Selection Method (CSM), which demonstrates how farmers can achieve higher yields by strategically selecting crops (Katarya et al., 2020). For the proposed system, the classification Algorithms is to enhance crop yields and collect the different datasets which are required for the project. Figure 1 shows the demonstration of model training. The data set must have the following attributes such as soil ph. value, humidity, temperature rainfall, label and components like N(Nitrogen), P(Phosphorous), K(Potassium).

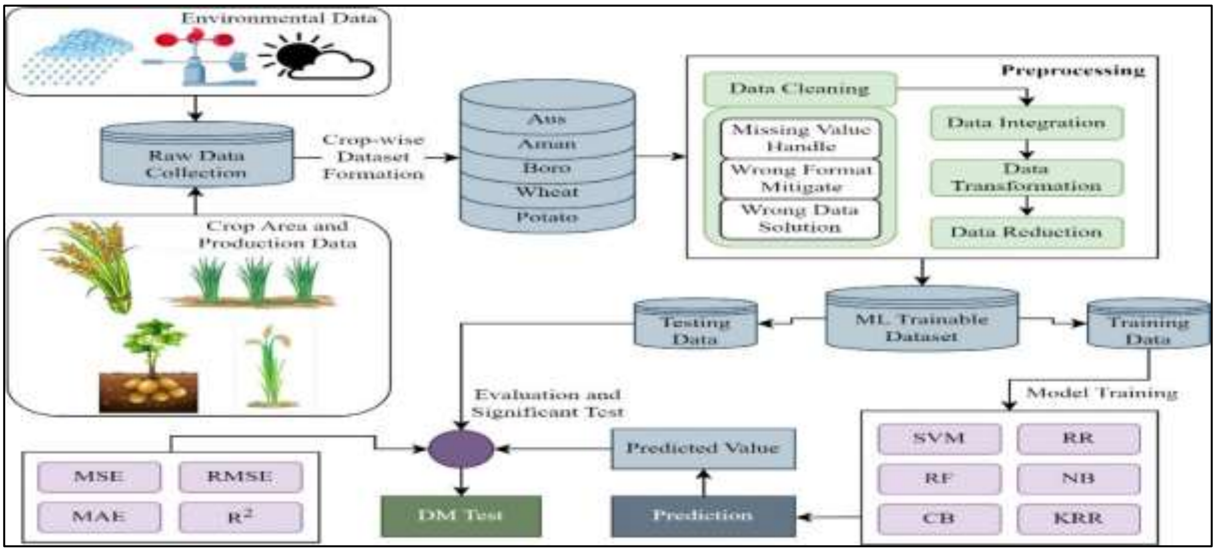

Figure 1: Demonstration of model training

The dataset needs pre-processing to remove redundant attributes and noisy data. Initially, we perform data cleaning to identify and exclude irrelevant factors from crop prediction. Attributes that either have the same values for all entries or are unrelated to the prediction task are removed. During exploratory data analysis, categorical factors are converted into binary values (0 and 1) to indicate their presence or absence. These binary values help in the classification process.

The decision tree method selects the best root nodes and continues splitting the tree based on attributes until all elements belong to the same class. This versatile supervised learning algorithm handles both categorical and continuous dependent variables and is mainly used for classification problems. It divides the population into two or more homogeneous sets based on the most significant attributes, making the groups as distinct as possible. The decision tree algorithm identifies the best splits for different features, helping to select the most suitable crop. Its feature selection methodology makes it ideal for predicting suitable crops. Here are the selection attributes used by the decision tree classifier.

Results and Discussion

In evaluating crop prediction using machine learning, two algorithms, K-Nearest Neighbors (KNN) and Decision Trees, were employed to assess accuracy and efficiency. KNN, leveraging instance-based learning, calculates predictions based on similarities to existing data points, making it effective for datasets with clear patterns. On the other hand, Decision Trees utilize hierarchical structures to partition data based on feature values, aiming to optimize information gain or node purity at each split. While KNN tends to offer high accuracy in scenarios with well- defined data patterns, Decision Trees excel in interpretability and handling of irrelevant features. The choice between these models hinges on the dataset's complexity and the need for interpretability versus raw predictive power in crop maintenance prediction tasks. Figure 2 shows the density distributions of nitrogen, phosphorus, and potassium ratios.

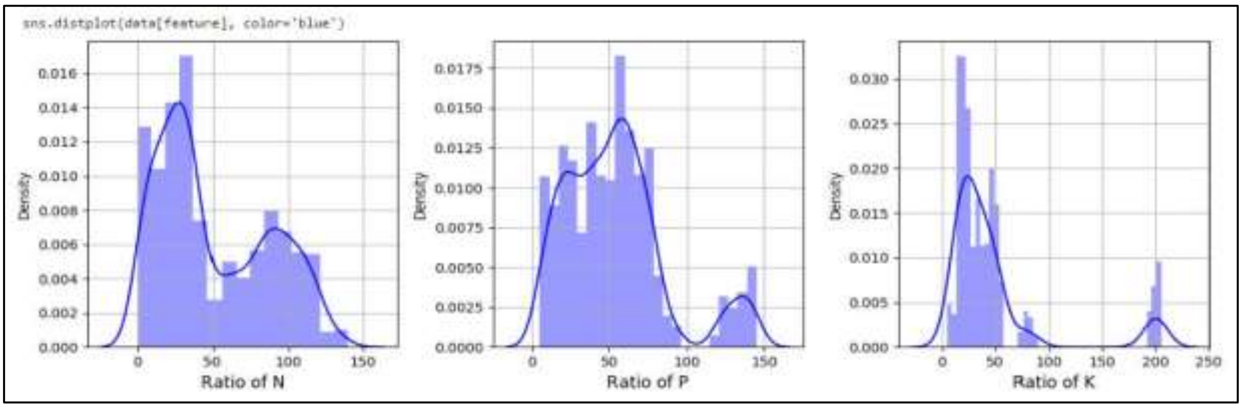

Figure 2: Density Distributions of Nitrogen, Phosphorus, and Potassium Ratios

The above density plots provide a visual summary of how the ratios of Nitrogen, Phosphorus, and Potassium are distributed within the dataset. The bimodal and multimodal patterns indicate that there might be distinct subgroups within the data, each with different characteristic ratios of these elements. Visualization helps in understanding the underlying variability of these key features in the dataset. Figure 3 shows the scatter plot of temperature versus humidity for different crops and distribution of rainfall across crops.

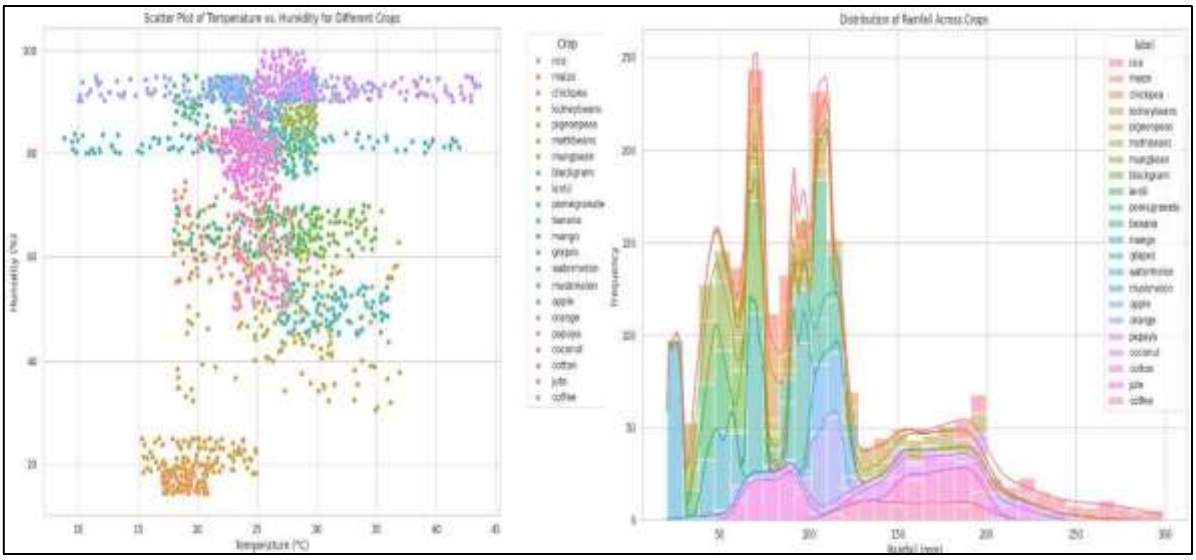

Figure 3: Scatter Plot of Temperature vs. Humidity for Different Crops and Distribution of Rainfall Across Crops

The scatter plot shows the relationship between temperature and humidity for various crops. Eachpoint represents a data instance, with different colors indicating different crops. The distribution and clustering of points can reveal how different crops perform under specific temperature and humidity conditions. The plot helps in understanding which crops thrive temperature and humidity ranges. Table 1 shows the comparison of Models for crop prediction.

Table 1: Comparison of Models for crop prediction**.**

Feature	K-Nearest Neighbors	Decision Tree
Handling Features	Sensitive to irrelevant features	Can handle irrelevant features
Scalability	Less scalable	More scalable
Prediction Time	Higher	Lower than KNN
Training Time	Low	Higher than KNN
Accuracy	98%	95%

In the context provided, K-Nearest Neighbors (KNN) is noted to have a higher accuracy of 98% compared to a Decision Tree model. This difference arises due to their distinct modeling approaches. KNN is an instance-based learning method where predictions are made based on similarities to known data points, making it straightforward and effective for datasets with clear patterns. In contrast, Decision Trees construct hierarchical structures that partition the data basedon feature values, aiming to maximize information gain or purity at each split.

Conclusion

This innovative crop recommendation system employs versatile classifier models, capable of adapting to various crop types. The system not only identifies ideal crops for cultivation but also aids in pinpointing optimal sowing, growth, and harvesting periods by analyzing yield patterns. While Decision Trees may falter when dealing with datasets exhibiting high fluctuations, KNN demonstrates superior performance in such instances. Furthermore, integrating classification algorithms like KNN and Decision Trees yields enhanced outcomes compared to relying on a single classifier. The empirical results underscore this advantage, with K-Nearest Neighbors (KNN) attaining a remarkable accuracy of 98% and Decision Tree achieving a commendable accuracy of 95%. This highlights the potential of utilizing a combination of classifiers to improve the accuracy and robustness of crop recommendation systems. By leveraging the strengths of different algorithms, farmers can make more informed decisions about crop selection and cultivation practices, ultimately leading to increased agricultural productivity.

Acknowledgements

We express our sincere gratitude to Dayananda Sagar Academy of Technology and Management (DSATM) for their resources and support in conducting our research on the "GIS-Enhanced Crop Yield Modeling with Machine Learning" We also extend our deepest thanks to our families, especially our mother, for their unwavering love, support, and financial assistance, which have been crucial in completing this project. We are truly grateful to everyone who contributed to the success of this research.

References

Aggarwal, S., & Kumar, A. (2019). A smart irrigation system to automate irrigation process using IOT and artificial neural network. In *2019 2nd International Conference on Signal Processing and Communication (ICSPC)* pp. 310-314. https://doi.org/10.1109/ICSPC46172.2019.8976631

Girish L. (2018). Crop Yield and Rainfall Prediction in Tumakuru District using Machine Learning". https://www.ijream.org/papers/NCTFRD2018015.pdf

Katarya, R., Raturi, A., Mehndiratta, A., & Thapper, A. (2020). Impact of machine learning techniques in precision agriculture. In 2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE) (pp. 1-6). https://doi.org/10.1109/ICETCE48199.2020.9091741

Koti, M.S., Supriya, R. (2021). Analysis of Farm Data Using Artificial Intelligence. In: Raj, J.S., Iliyasu, A.M., Bestak, R., Baig, Z.A. (eds) Innovative Data Communication Technologies and Application. Lecture Notes on Data Engineering and Communications Technologies, vol 59. Springer, Singapore. https://doi.org/10.1007/978-981-15-9651-3_18

Kushwaha, A. K., & Bhattachrya, S. (2015). Crop yield prediction using Agro Algorithm in Hadoop. International Journal of Computer Science and Information Technology & Security (IJCSITS), 5(2), 271-274. https://api.semanticscholar.org/CorpusID:11500659

Patel, K. G., & Patil, M. S. (2022). Artificial intelligence in agriculture. *International Journal for Research in Applied Science and Engineering Technology*, *10*(2), 624-627.

Patil, P., Panpatil, V., & Kokate, S. (2020). Crop prediction system using machine learning algorithms. Int. Res. J. Eng. Technol.(IRJET), 7(02). https://www.irjet.net/archives/V7/i2/IRJET-V7I2163.pdf

Priyani, S., Hakani, R., & Jhala, D. K. (2023). Smart Agriculture system Using Machine Learning. *Researchgate. Net*. http://dx.doi.org/10.13140/RG.2.2.30302.77121

Samui, P., Mandla, V. R., Krishna, A., & Teja, T. (2011). Prediction of rainfall using support vector machine and relevance vector machine. Earth Science India, 4. https://www.researchgate.net/publication/268439113_Prediction_of_Rainfall_Using_

Support_Vector_Machine_and_Relevance_Vector_Machine

Zingade, D. S., Buchade, O., Mehta, N., Ghodekar, S., & Mehta, C. (2017). Crop prediction system using machine learning. Int. J. Adv. Eng. Res. Dev. Spec. Issue Recent Trends Data Eng, 4(5), 1-6.