Review on Crop Recommendation System

Prof. Dr. Tripti Sharma, HOD, CSE(DS)

Aditi Jain, Akshay Singh, Vaibhav Chauhan
Department of CSE, CSE(DS), CSE(AIML)
Inderprastha Engineering College, Uttar Pradesh, India

Abstract: *This review paper critically examines ten diverse research papers on crop recommendation systems, conducting a comprehensive comparative study on the datasets and methodologies employed. The selected studies, spanning various geographical regions and agricultural contexts, provide valuable insights into the evolving landscape of precision agriculture. The review systematically analyzes the datasets utilized in each study, considering factors such as geographic location, climatic conditions, and soil characteristics. Furthermore, the paper scrutinizes the methodologies, encompassing machine learning algorithms, statistical models, and data integration techniques, to assess their efficacy in predicting optimal crop recommendations. By synthesizing findings from these ten studies, this review aims to identify common trends, challenges, and advancements in crop recommendation systems. The comparative analysis serves to highlight best practices, potential limitations, and areas for future research, contributing to the refinement and advancement of precision agriculture strategies for sustainable and efficient crop management.*

Introduction

Farming in India holds paramount significance as it constitutes the backbone of the country's economy, providing livelihoods to a significant portion of the population. The diverse agro-climatic zones allow for the cultivation of a wide array of crops, contributing to the nation's food security. However, Indian farmers grapple with numerous challenges, particularly in deciding the appropriate crops to grow and determining the optimal amount of fertilizers to use.

The choice of crops is intricate, influenced by climatic conditions, soil quality, and market demand. Miscalculations can lead to financial setbacks. Additionally, the excessive or insufficient use of fertilizers can impact both crop yield and soil health, exacerbating environmental concerns. In this context, implementing a crop recommendation system powered by machine learning emerges as a transformative solution.

Machine learning algorithms, analyzing historical data on weather patterns, soil types, and crop performances, can offer precise recommendations. Such systems can guide farmers in selecting crops attuned to their specific region, ensuring optimal yields. Moreover, by advising on the appropriate amount and type of fertilizers, these systems promote sustainable farming practices. Integrating technology into agriculture not only enhances productivity but also aids in sustainable resource management, addressing the challenges faced by Indian farmers and fortifying the agricultural sector for the future.

Literary Survey

India is a country where a majority rely on agriculture as their occupation. The repeated cultivation of the same crops without considering soil conditions and the use of same fertilizers lead to reduced crop yields, soil acidification, and top-layer damage. To address these issues, we need to use automated means of predicting the best crops for cultivation for a particular area and region. By utilizing machine learning algorithms, we can help farmers in making informed decisions [1].

In addition to suggesting the most suitable crop for a specific piece of land, we can also provide information on the required content and quantity of fertilizers, as well as the necessary seeds for cultivation. Machine learning algorithms such as SVM were used for rainfall prediction and Decision Tree for crop prediction. The system further recommends the most suitable crop, provides information on required nutrients and seeds, and displays the market price and approximate yield for the recommended crop [1].

Precise Farming uses advanced technologies such as IOT, Data Mining, Data Analytics, Machine Learning to collect the data, train the systems and predict the results. With the help of technologies Precise farming helps to reduce manual labor and increase productivity. It guides an individual for smart farming [2],[10].

Ten classification algorithms were used in [2] to find the best model for future prediction. These include Naïve Bayes, Logistic Regression, SVM, Decision Tree, Bagging Classifier, XG Boost Classifier, Random Forest Classifier, Ada Boost Classifier, LGBM Classifier, Gradient Boosting and KNN. The training and testing ratio was kept to be equal i.e. 50:50. Six algorithms obtained a score of more than 90%. The Random Forest model hyper tuned with Randomized CV was selected as the best model since its accuracy is 95.45%.

Data is the most valuable. Hence, more information can be obtained apart from using them for prediction. Hence more information can be obtained apart from using them for prediction. The work [2] not only recommends the crops and also uses the data to obtain various information that would provide a detailed view about the predicted crops this includes specifying the Growing Degree Days such as heat units, amount of heat needed for the crop growth and the amount of nitrogen, phosphorous and potassium content need to be supplied for the growth per 200 lb. fertilizer. It also uses a formula to calculate the GDD for the crops.

The weed identification module [2] uses Resnet 152v2 pre-trained model for prediction with 4 hidden layers containing 1024,512,512 and 32 neurons respectively with ReLu as activation function. The output layer contains just one neuron with Softmax activation function for classifying the images.

Not only predicting the image is important but also recommending Herbicides for the predicted weeds would be very useful for the user. Herbicides are generally used for killing the weeds. Manually removing the weeds are mostly impossible for a large farm. Hence herbicides are optimal way for destroying the weeds from the land. To find a suitable herbicide the chemical component to kill them is identified from agricultural websites. Based on the chemical components the herbicide having it has an active component is being picked. For all crops same herbicide cannot be used. It could harm the growth of the crops as well as affect the soil fertility. Hence crop specific herbicides are predicted using Random Forest Classifier [2].

The input parameters for Random Forest Classifier are crop name and the name of the weed. The crop name must be entered by the user. The name of the weed is identified by the Resnet152V2 fine tuning model when the image is been uploaded by the user. The Random Forest Classifier classifies the crops based on its highest probability herbicides are listed [2].

The paper [3] compares various supervised learning algorithms like KNN, Decision Tree, and Random Forest on the dataset containing 22 varieties of crops. It uses Decision Tree and Random Forest Classifier and evaluates the model's performance under two criterions- Entropy and Gini Index.

It concluded that the crop prediction dataset showed the best accuracy with Random Forest Classifier both in Entropy and Gini Criterion with 99.32%. In contrast, K-Nearest Neighbour with k = 5 has the lowest accuracy among the three with 97.04%, and the accuracy of Decision Tree Classifier is in between KNN and Random Forest Classifier. When comparing the accuracy value, Decision Tree Gini criterion gave a better accuracy of 98.86% compared to Decision Tree Entropy Criterion.

Modern farms and agricultural operations are taken place more totally different than those many decades agone, primarily due to advancements in technology, as well as sensors, devices, machines, and knowledge technology. Today’s agriculture habitually uses refined technologies like robots, temperature and wetness sensors, aerial pictures, and GPS technology, and lots of complex IOT devices. These advanced devices in agriculture enable businesses and farmers to be additional profitable, efficient, safer, and more environmentally friendly. Remote sensors, cameras, and alternative connected devices will gather data twenty-four hours per day over a complete farm or land. These will monitor plant health, soil condition, temperature, humidity, etc. the quantity of information these sensors will generate is overwhelming. this enables farmers to achieve a far better an improved understanding of state of matters on the bottom through advanced technology which will inform them additional regarding their situation more accurately and quickly. The environmental data that is gathered by remote sensors are processed by algorithms and statistical data which will be understood and helpful to farmers for decision makings and keep track of their farms. The more inputs and statistical data collected, and higher the algorithmic rule is at predicting the outcomes [4].

The paper [4] focuses on addressing the challenges faced in Sri Lanka, where despite having manual agricultural knowledge, there is a lack of automated systems to detect environmental factors and suggest optimal crops for cultivation.

It integrates various technologies, including Arduino microcontrollers for environmental data collection such as temperature, humidity, pH and sunlight etc, machine learning techniques such as Naïve Bayes and Support Vector Machine. The automatic processing of environmental factors eliminates the need for specialist guidance and minimizes maintenance costs. With an accuracy exceeding 95%, the system is deemed suitable for Both rural and urban areas in Sri Lanka [4].

Using the feedbacks collected from the farmer the accuracy of the predictions is sharpened by neglecting the invalid data. As an example, if the farmers are continuously providing the negative feedback for cultivating strawberry in Galle, the system itself learned, and the accuracy of the final output is increased which means in future it will not suggest strawberry to grow in Galle [4].

The paper [5] begins by highlighting the crucial role agriculture plays in providing employment and income in rural areas. Despite its importance, the paper notes that the yield per hectare in India is lower than global standards. The authors identify reasons for the high suicide rate among marginal farmers and present their paper as a study offering a solution to address these issues.

The system involves connecting farmers through a smartphone app, utilizing GPS technology for user identification and location. Farmers specify the area and soil type, and machine learning algorithms, including Support Vector Machine (SVM), Artificial Neural Network (ANN), Random Forest (RF), Multivariate Linear Network (MLN), and a combination of regression and KNN, are employed to predict crop yields. The Random Forest algorithm demonstrated the highest accuracy at 95%. Additionally, the system recommends the use of chemical fertilizers to enhance output [5].

An ensemble learning approach to enhance prediction accuracy. It uses parameters such as depth, texture, pH, soil colour, permeability, drainage, water holding, and erosion. The ensemble technique employed is Majority Voting, with base learners including Support Vector Machine, Naïve Bayes, Multi-layer Perceptron, and Random Forest [6],[9].

The paper [7] highlights the challenges observed in Maharashtra when the price of onions doubled in August. Observing the shoot in the price, many of the farmers in the state decided to grow onions on their farm, in the hope of making exorbitant profits. While this resulted in abundant supply in certain regions of Maharashtra, many other regions suffered failed crop output due to unfavorable conditions for growing onions. A subsequent shortage again in the following months had harsh ramifications on the lives of common man, as middleclass households could no longer afford onion- a frequently used commodity in their kitchen. This example just goes on to show that a farmer's decision about which crop to grow is generally clouded by his intuition and other irrelevant factors like making instant profits, lack of awareness about market demand, overestimating a soil's potential to support a particular crop, and so on. A very misguided decision on the part of the farmer could place a significant strain on his family's financial condition. Perhaps this could be one of the many reasons contributing to the countless suicide cases of farmers that we hear from media on a daily basis. In a country like India, where agriculture and related sector contributes to approximately 20.4 per cent of its Gross Value Added (GVA), such an erroneous judgment would have negative implications on not just the farmer's family, but the entire economy of a region. For this reason, we have identified a farmer's dilemma about which crop to grow during a particular season, as a very grave one.

The system is divided into 2 sub-modules – one for crop prediction and the other for rainfall prediction. For crop prediction four machine learning algorithms were used. These were decision tree, random forest, KNN and Neural Network. The neural network obtained the highest accuracy i.e. 91%. The rainfall predictor achieved an overall accuracy of 71%.

The overall system [8] was divided into three modules as – Profit Analysis (which estimates the profit on growing a crop in a given state), Crop Recommender (recommends the top 5 crops that can be grown in an area) and Crop Sustainability Predictor (which gives the probability of growing the crop in a given area). Hence, providing an insight into the overall agricultural activities and increasing the productivity.

In the paper [8] seven machine learning algorithms were used: Decision Tre, K Nearest Neighbour, K Nearest Neighbour with cross validation, Linear Regression Model, Naive Bayes, Neural Network, Support Vector Machine. Finally, Neural Network was used as it obtained the highest accuracy.

Comparison between papers

Types of crops	Dataset used	Models used	Accuracy obtained	References
Various crops	Combined data from various sources Govt Websites, VC Form Mandya, APMC website for crop prediction, Previous year rainfall data for rainfall prediction	SVM classifier using RBF Kernel for rainfall prediction, Decision Tree for crop recommendation	—	[1]
Rice, banana, jute, cotton, coconut, mango, papaya,orange, muskmelon, grapes, maize, pomegranate, blackgram, moongbean, mothbeans, pigeonbeans, kidneybeans, chickpeas,coffee, apple, lentil, watermelon	https://www.kaggle.com/ atharvaingle/crop-recommendation-dataset	KNN, Naïve Bayes, LR, SVM, DT Classifier, Bagging Classifier, RF, AdaBoost Classifier, Gradient Boost Classifier, XGBoost Classifier, LBGM Classifier	95.45% with RF hypertuned with Randomized CV	[2]
Rice, banana, jute, cotton, coconut, mango, papaya,orange, muskmelon, grapes, maize, pomegranate, blackgram, moongbean,	https://www.kaggle.com/ atharvaingle/crop-recommendation-dataset	KNN, DT Gini, DT Entropy, RF Gini, RF Entropy	99.32% with RF Classifier both in Gini and Entropy Criterion.	[3]

mothbeans, pigeonbeans, kidneybeans, chickpeas,coffee, apple, lentil, watermelon				
Bean, Lettuce, Carrot, Cabbage	Initial data set is collected from the Department of Agriculture Sri Lanka, other agriculture books, Agricultural web sites, and other reports and research papers	Naïve Bayes, SVM	92% overall accuracy	[4]
Various crops	___	SVM, ANN, RF, KNN	95% with RF	[5]
Groundnut, pulses, cotton, vegetables, banana, paddy, sorghum, sugarcane, coriander	Polytest Laboratories soil testing lab, Pune, Maharashtra, India, Marathwada University	Ensemble with base learners such as SVM, Naïve Bayes, MLP, RF.	__	[6]
Bajra, jowar, maize, rice, wheat, barley, cotton, groundnut, gram, jute, other pulses, potato, ragi, tur, rapeseed and mustard, sesame, soybean, sugarcane, sunflower, tobacco	India Agriculture and Climate Data Set	DT, KNN, RF, Neural Network	91% with Neural Network	[7]
Bottle Gourd, groundnut, Jowar, khesari, Orange, potato, Raddish, sanhampp, rice, wheat, cotton, sugarcane, tea,	Various datasets from Govt. Website and Kaggle.com (Yield Dataset, Cost of Cultivation Dataset, Modal Price of crops dataset, Soil Nutrient Content Dataset,	Decision Tree, K Nearest Neighbour, K Nearest Neighbour with cross validation, Linear Regression,	89.88% with Neural Network	[8]

coffee, cashew, rubber	Rainfall Temperature Dataset)	Naive Bayes, Neural Network, Support Vector Machine		
Cotton, Sugarcane, Wheat, Rice	Open source dataset from data.gov.in	Ensemble using base learners SVM, Naïve Bayes	99.91%	[9]
Various crops	___	Naïve Bayes, RF	96.89%	[10]

References

[1] Mahendra N, Dhanush Vishwakarma, Nischitha K, Ashwini, Manjuraju M. R, 2020, Crop Prediction using Machine Learning Approaches, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 08 (August 2020)

[2] Senthil Kumar Swami Durai, Mary Divya Shamili, Smart farming using Machine Learning and Deep Learning techniques, Decision Analytics Journal, Volume 3, 2022, 100041, ISSN 2772-6622, https://doi.org/10.1016/j.dajour.2022.100041.

[3] Madhuri Shripathi Rao *et al* 2022 *J. Phys.: Conf. Ser.* **2161** 012033 “Crop Prediction using Machine Learning”

[4] Bandara, P.; Weerasooriya, T.; T.H., R.; Nanayakkara, W.; M.A.C, D.; M.G.P, P. Crop Recommendation System. Int. J. Comput. Appl. 2020, 175, 22–25, https://doi.org/10.5120/ijca2020920723.

[5] R. Jadhav and D. P. Bhaladhare, "A Machine Learning Based Crop Recommendation System: A Survey," Journal of Algebraic Statistics, vol. 13, no. 1, pp. 426–430, May 2022.

[6] Rajak, Rohit Kumar, Ankit Pawar, Mitalee Pendke, Pooja Shinde, Suresh Rathod, and Avinash Devare. "Crop recommendation system to maximize crop yield using machine learning technique." International Research Journal of Engineering and Technology 4, no. 12 (2017): 950-953.

[7] Doshi, Z., Nadkarni, S., Agrawal, R., & Shah, N.K. (2018). AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms. *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, 1-6.

[8] A, P., Chakraborty, S., Kumar, A., & Pooniwala, O.R. (2021). Intelligent Crop Recommendation System using Machine Learning. *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, 843-848.

[9] N. H. Kulkarni, G. N. Srinivasan, B. M. Sagar and N. K. Cauvery, "Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique," *2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*, Bengaluru, India, 2018, pp. 114-119, doi: 10.1109/CSITSS.2018.8768790.

[10] et al., A. (2021). Crop Recommendation on Analyzing Soil Using Machine Learning. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(6), 1784–1791. Retrieved from https://turcomat.org/index.php/turkbilmat/article/view/4033