7 Crop recommendation using machine learning

Paramveer Kaur[a] *and Brahmaleen Kaur Sidhu*

Department of Computer Science and Engineering, Punjabi University, Patiala, Punjab, India

Abstract

Agriculture serves as the cornerstone of India's economic expansion, constituting the primary income source for a significant proportion of its populace, encompassing both those directly engaged in agricultural activities and those who depend on it indirectly for their livelihoods. Therefore, it is essential for farmers to make the correct option possible when cultivating any crop so that the farmer can make maximum profit from the agriculture field. To make the agriculture sector profitable, one of the technologies that may be used in this age of rapid technological improvement is known as machine learning (ML). In this research paper, various ML algorithms, such as logistic regression (LR), decision trees (DT), LightGBM, and random forest (RF), have been utilized to analyze a dataset. The primary objective is to predict the most suitable crop based on soil attributes such as (nitrogen, phosphorous, potassium) NPK content, humidity, temperature, soil pH level, and rainfall. Out of this random forest and LightGBM comes with great accuracy whereas decision tree and logistic regression have less accuracy. In addition, ML algorithms will likely find applications in a variety of agricultural subfields in the near future, including the diagnosis of plant diseases, the selection of soil types, and the forecasting of retail pricing.

Keywords: Machine learning, crop recommendation, decision tree, random forest, logistic regression, LightGBM

I. Introduction

The importance of agriculture to the Indian economy and to human life cannot be ignored. It serves as one of the main professions that are necessary for human existence. India's population relies heavily on income from agriculture and related industries. Around 82% of farmers are classified as small and marginal, underscoring the central importance of agriculture as the main source of income for 70% of rural households (Chavva, n.d.). The financial health of the agricultural sector is tightly intertwined with the success of every harvest, which, in turn, is influenced by a wide range of variables, including weather patterns, soil quality, fertilizer use, and market prices. Due to climate change it has become difficult for farmers to choose appropriate crop for particular season. Price of crops given by the government to farmers also the effective factor to grow any crop. According to National Crime Records Bureau (NCRB) statistics, the farmer suicides rate has remained high between 2014 and 2020. In 2014, 56 farmers have committed suicide, and by 2020, the number has risen to 5,500 (Affarirs, n.d.).

The choice of crop to be sown depends on availability of resources like soil, water, seed, manure, fertilizer and market profit. However, climate conditions and soil properties are consider as natural parameters for the success of any crop grown by farmer. Due to variations in soil, water, and air quality throughout the year, it is difficult to predict how best to use different types of fertilizer and what crop to be grown. The rate of agricultural output is falling continuously in this situation (Pande et al., 2021). The issues can be solved by using the advanced technologies to improve the agriculture sector. The farmer cannot predict the weather but the technology can predict the climate like rainfall, humidity and temperature based on past data and help famers in real manner.

Precision agriculture (PA) is a method of farm management that takes into account the specific conditions of particular fields and crops via the collection, organization, and analysis of data (António Monteiro, 2021). It is observed that in recent years precision agriculture technologies have helped the farmers as well as the environment by suggesting the required amount of fertilizers, pesticides and water for crops. Farmers have got more profit by these technologies in terms of money also.

In the realm of computer technology, the most recent advancements encompass block chain, internet of things (IoT), deep learning, machine learning (ML) and cloud computing, which are useful for solving difficult problems in various fields like health, biochemistry, agriculture, cybercrime, robotics, banking, meteorology, medicine and robotics (Vishal Meshram, 2021). However in this paper, the focus is only on ML in agriculture. ML algorithms (Sharma et al., 2018) which are support vector machines, Naïve Bayes, neural network, decision tree (DT), K-Nearest Neighbor, XG-boost (e-Xtreme Gradient Boosting), multi-variate linear regression, linear regression (LR), chi-square automatic interaction detection (CHAID) and sliding window non-linear regression are helpful at various stages of crop grown cycle and provides maximum accuracy. In this proposed system out of

[a]paramveer1067@gmail.com

this four, ML models are deployed to make accurate crop recommendations.

II. Related work

Pudumalar et al. (2016) addressed precision agriculture. This study proposes an ensemble model with majority voting technique utilizing RT, Naïve Bayes, CHAID, and K-nearest neighbor (KNN) as learners to effectively and correctly suggest a crop for site-specific parameters. The study (Kanaga Suba Raja et al., 2017) analyses historical data to predict a farmer's crop output and price. Sliding window non-linear regression predicts agricultural output depending on rainfall, temperature, market prices, land area, and crop yield. Zeel Doshi et al. (2018) developed a soil dataset-based crop recommendation system for only four crops. The ensemble model uses random forest (RF), Naive Bayes, and linear support vector machines base learners. The majority voting technique is employed in the combination approach because it is the most accurate. The author uses Big Data analytics and ML to create an AgroConsultant, an system that assists Indian farmers choose the best crop based on sowing season, farm location, soil properties, and environmental factors like temperature and rainfall (Doshi et al., 2018). Rainfall predictor, another approach created by academics, predicts annual precipitation. The system uses DT, KNN, RF, and neural networks.

Shilpa Mangesh Pande et al. (2021) provide farmers a simple yield projection tool. Farmers utilize a smartphone app to connect to the internet. GPS locates users. Enter location and soil type. ML systems can identify the most profitable crops and predict agricultural yields for user-selected crops. SVM, RF, MLR, ANN, and KNN are used to predict agricultural production. The research (A et al., 2021) suggests a way to assist farmers pick crops by considering planting time, soil qualities including type of soil, pH value, and nutrient content, meteorological aspects like rainfall, temperature, and state location. The suggested system has been developed using linear regression as well as neural network. Another study (Gosai et al., 2021) forecasts the best crop based on N, P, K, pH of soil, humidity, temperature, and rainfall. Decision trees, SVM, Nave Bayes, Support vector machine, LR, RF, and XGBoost were utilized to develop suggested system, and the maximum accuracy was of XGBoost. Distribution analysis, majority voting, correlation analysis and ensembling are used to create 22 crop recommendations (Kulkarni et al., 2018). A three-level technique solves crop recommendations. Chhikara et al. (2022) propose a ML-based crop recommender system that can accurately forecast the yield of 22 different crop types, addressing farmers' crop management issues like crop selection, yield, and profit. Researchers employ decision trees, Naive Byes, SVM, LR, RF, and Xgboost. Pradeepa Bandara et al. developed a crop recommendation system for Sri Lanka (Bandara et al., 2020). The study provides a theoretical as well as a conceptual platform for a recommendation system using Arduino microcontrollers, ML approaches such as Naive Bayes (multi-nomial) and SVM and unsupervised ML algorithms that are K-Means Clustering and Natural Language Processing (NLP) (sentiment analysis). Avinash Kumar et al. (2019) addressed crop selection and disease issues. SVMs classification model, decision tree model, and logistic regression model were used to create this recommendation system.

III. Objectives

The aim of the proposed work is to implement ML algorithms for developing crop recommendation system and it is based on chemical properties of soil and weather condition and to evaluate the proposed system.

IV. Background techniques

A. Logistic regression

It is a ML algorithm primarily applied to classification problems, operates on the foundation of predictive analysis rooted in probability (Rymarczyk et al., 2019). Notably, the LR model adopts a more intricate cost function than linear regression. This cost function, often referred to as the "Sigmoid function" or "logistic function," replaces the linear function. Due to the fundamental premise of logistic regression, the output range of the cost function is inherently confined to the interval [0, 1]. This constraint stems from the nature of logistic regression, where it inherently models the probability of an event occurring, rendering linear functions unsuitable for capturing its nuances and characteristics.

B. Decision tree

When it comes to representing models for use in data classification, decision trees are among the most popular approaches (Jijo and Abdulazeez, 2021). DT stand as versatile assets in numerous domains, spanning machine learning, image processing, and pattern recognition. Their core role revolves around the task of classification and, as a result, they find wide application as classifiers within the field of data mining. These decision trees are architecturally composed of interconnected nodes and branches, each node representing a collection of attributes within discrete classification categories. Each branch within the tree signifies a potential value associated with the respective node. Decision trees earn considerable favor for their proficiency in swiftly and accurately managing large datasets.

C. Random forest

In the field of ML, random forest, a supervised learning technique, has demonstrated considerable success. It's versatile and can handle various tasks like classification and prediction. What sets random forest apart is that it operates like a team of decision trees collaborating to solve problems. Rather than rely on a one decision tree, it combines the results of many trees, each trained on different parts of the data. This cooperative approach enhances accuracy. Essentially, the more trees in this "forest," the better the performance, and it's less prone to errors (Dabiri et al., 2022; Gera et al., 2021).

D. LightGBM

Tree-based learning algorithms are at the foundation of LightGBM, a gradient boosting framework (Tang et al., 2020). It has the many benefits because of its decentralized and efficient design such as increased efficiency and accelerated training time. It uses less memory and allows for multi-GPU and distributed learning. LightGBM has ability to process large amount of data as well as gives enhanced precision.

V. Methodology

The proposed methodology has been shown in the Figure 7.1 to implement an accurate crop recommendation system. The detailed process of developing recommender system has been provided which includes various stages data collection, data pre-processing, model development and training and testing.

In the first stage dataset used for proposed approach was retrieved from Github repository https://github.com/gabbygab1233/Crop-Recommender/blob/ main/Crop_recommendation.csv. The dataset is combination of rainfall, climate and fertilizers dataset which is collected from Indian data available for agriculture sector. Table 7.1 shows the description of features that has been used. It has 8 features and 2200 records.

The preparation of the collected raw data to make it appropriate for use in building the ML model is the next step. This is the initial and most crucial step in the creation of an ML model. The data that is available on different sources is not clean always. Data must be cleaned and formatted before any action can be performed on it. Pre-processing the data is necessary for this purpose.

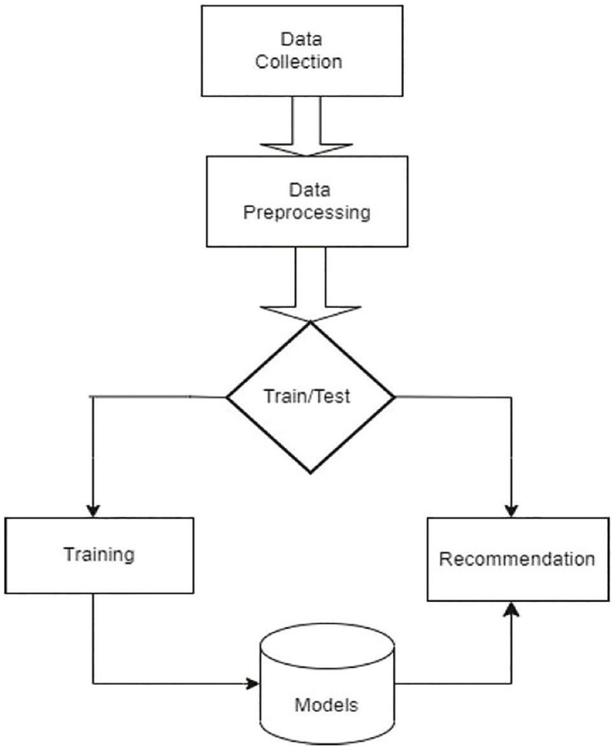

Figure 7.1 Flow chart of proposed methodology

Table 7.1 Feature description

Attribute	Attribute description
Nitrogen	Proportion of nitrogen amount in soil
Phosphorus	Ratio of phosphorus content in soil
Potassium	Ratio of potassium amount in soil
Rainfall	Rainfall (mm)
Humidity	Relative humidity (%)
pH level	Soil's pH value
Temperature	Temperature (degree celsius)

Table 7.2 Accuracy of algorithms

Algorithm/classifier	Predicted accuracy
LightGBM	0.99
Decision tree	0.98
Random forest	0.99
Logistic regression	0.95

VI. Results

Table 7.2 displays the performance of machine learning approaches employed in the development of a crop recommendation system, including LR, DT, RF, and LightGBM. The accuracy of mentioned algorithms has been assessed using evaluation parameters such as precision, recall, F1-score, and support.

Precision
It indicates the percentage of accurate predictions executed by the model. It specifically assesses the classifier's capacity to correctly discern instances as either positive or negative. The precision parameter is determined by dividing the accurately predicted positive counts by the sum of all predictions, encompassing both correct and incorrect classifications.

$$\text{Precision} = \frac{TP}{TP+FP}$$

Recall
It denotes the percentage of positive cases successfully identified by the model. It quantifies a classifier's ability to effectively locate all instances that belong to the designated target class. The recall metric is computed by dividing the count of correct predictions for the target class by the total number of predictions made, encompassing both accurate and inaccurate classifications.

$$\text{Recall: } \frac{TP}{TP+FN}$$

F1-score
It signifies the proportion of positive predictions that are accurate. This metric is computed from the combination of both precision and recall, and it falls within a range from 1.0 – indicating the best performance to 0.0 – representing the worst. Given that F1-scores encompass both precision and recall in their computation, they are considered more informative than simple accuracy measurements. When assessing and comparing classifier models, it is advisable to use the weighted average of F1-scores rather than relying solely on overall accuracy.

$$\text{F1-score} = \frac{2*Precision*Recall}{Precision+Recall}$$

Support
It refers to the frequency of actual occurrences of a specific class within the dataset. It serves as an indicator of how evenly or unevenly the training data is distributed. Disparities in support could potentially suggest that the reported scores by the classifier may not be as reliable as desired. It is important to note that support remains consistent across various models; its primary role is to aid in the analysis of the testing methodology.

VII. Discussion

The agriculture sector plays pivotal role in driving a nation's economic development. To enhance the profitability and productivity of this sector, it is imperative to explore innovative solutions. In this research paper, a crop recommendation system powered by ML technologies is introduced. The system is designed to help farmers in making accurate decision when choosing the crop varieties based on the intricate interplay of soil attributes and weather conditions.

VIII. Conclusion

The research paper introduces a crop recommendation system designed for farmers to make optimal crop choices through predictive analytics. This system takes into account critical parameters, such as soil characteristics (NPK content, pH, and humidity), along with meteorological factors (temperature and rainfall). To develop this system, ML approaches, including LR, DT, RF along with LightGBM, are employed. This algorithms have been applied on collected dataset. Dataset was normally distributed with negligible outliers and this is analyzed after pre-processing steps. LightGBM and RF exhibit the highest accuracy among various algorithms. Looking forward, enhancing the system's performance is achievable through the continuous updating of datasets, ensuring its alignment with evolving agricultural conditions. Moreover, the system's scope can be expanded to include crop disease detection which can assist farmers in maximizing benefits in the agriculture sector.

References

Chavva, Mr Konda Reddy. India at a glance. n.d. https://www.fao.org/india/fao-in-india/india-at-a-glance/en/ (accessed june 15, 2023).

Gera. T., Singh, J., Mehbodniya, A., Webber, J. L., Shabaz, M., and Thakur, D. (2021). Dominant feature selection and machine learning-based hybrid approach to analyze android ransomware. *Sec. Comm. Netw.*, 2021, 1–22.

Affarirs, M. O. H. NCRB on data portal. [Online]. Available: https://data.gov.in/search?title=suicide.

Pande, Shilpa Mangesh, Prem Kumar Ramesh, ANMOL ANMOL, B. R. Aishwarya, KARUNA ROHILLA, and KUMAR SHAURYA. (2021). Crop recommender system using machine learning approach. *In 2021 5th international conference on computing methodologies and communication (ICCMC),* 1066–1071. IEEE.

António Monteiro, S. S. P. G. (2021). Precision agriculture for crop and livestock. *MDPI/Animals*, 1–18.

Meshram, Vishal, Kailas Patil, Vidula Meshram, Dinesh Hanchate, and S. D. Ramkteke. (2021). Machine learning in agriculture domain: A state-of-art survey. *Artificial Intelligence in the Life Sciences*, 1, 1–11. doi: https://doi.org/10.1016/j.ailsci.2021.100010

Sharma, Pinkashia, and Jaiteg Singh. (2018). Machine learning based effort estimation using standardization. *In 2018 International Conference on Computing, Power* *and Communication Technologies (GUCON),* 716–720. IEEE.

Pudumalar, S., E. Ramanujam, R. Harine Rajashree, C. Kavya, T. Kiruthika, and J. Nisha. (2017). Crop recommendation system for precision agriculture. *In 2016 Eighth International Conference on Advanced Computing (ICoAC),* 32–36. IEEE.

Jijo, B. T. and Abdulazeez, A. M. (2021). Classification based on decision tree algorithm for machine learning. *J. Appl. Sci. Technol. Trends,* 02(1), 20–28.

Rymarczyk, T., Kozlowski, E., Klosowski, G., and Niderla, K. (2019). Logistic regression for machine learning in process tomography. *Sensors*, 1–19.

Dabiri, Hamed, Visar Farhangi, Mohammad Javad Moradi, Mehdi Zadehmohamad, and Moses Karakouzian. (2022). Applications of Decision Tree and Random Forest as Tree-Based Machine Learning Techniques for Analyzing the Ultimate Strain of Spliced and Non-Spliced Reinforcement Bars. *Applied Sciences*. 12(10), 4851. 1–13 doi: https://doi.org/10.3390/app12104851.

Tang, Mingzhu, Qi Zhao, Steven X. Ding, Huawei Wu, Linlin Li, Wen Long, and Bin Huang. (2020). An improved lightGBM algorithm for online fault detection of wind turbine gearboxes. *Energies*, 13(4): 807. 1–16. doi: https://doi.org/10.3390/en13040807.

Raja, S. Kanaga Suba, R. Rishi, E. Sundaresan, and V. Srijit. (2017). Demand based crop recommender system for farmers. *In 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR),* 194–199. IEEE. doi: 10.1109/TIAR.2017.8273714.

Doshi, Zeel, Subhash Nadkarni, Rashi Agrawal, and Neepa Shah. (2018). AgroConsultant: intelligent crop recommendation system using machine learning algorithms. *In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA),* 1–6. IEEE. doi: 10.1109/ICCUBEA.2018.8697349.

Kulkarni, Nidhi H., G. N. Srinivasan, B. M. Sagar, and N. K. Cauvery. (2018). Improving crop productivity through a crop recommendation system using ensembling technique. *In 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS),* 114–119. IEEE. doi: 10.1109/CSITSS.2018.8768790

Priyadharshini, A., Swapneel Chakraborty, Aayush Kumar, and Omen Rajendra Pooniwala. (2021). Intelligent crop recommendation system using machine learning. *In 2021 5th international conference on computing methodologies and communication (ICCMC),* 843–848. IEEE. doi: 10.1109/ICCMC51019.2021.9418375

Gosai, D., Raval, C., Nayak, R., Jayswal, H., and Patel, A. (2021). Crop recommendation system using machine learning. *Int. J. Sci. Res. Comp. Sci. Engg. Inform. Technol., 554–569.*

Ray, Rakesh Kumar, Saneev Kumar Das, and Sujata Chakravarty. (2022). Smart Crop Recommender System-A Machine Learning Approach. *In 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence),* 494–499. IEEE. doi: 10.1109/Confluence52989.2022.9734173

Chhikara, Sonam, and Nidhi Kundu. (2022). Machine Learning based Smart Crop Recommender and Yield Predictor. *In 2022 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS),* 474–478. IEEE. doi: 10.1109/ICCCIS56430.2022.10037678

Bandara, P., Weerasooriya, T., R. T. H., Nanayakkara, W., D. M. A. C., and P. M. G. P. (2020). Crop recommendation system. *Int. J. Comp. Appl.,* 175(22), 22–25.

Kumar, Avinash, Sobhangi Sarkar, and Chittaranjan Pradhan. (2019). Recommendation system for crop identification and pest control technique in agriculture. *In 2019 International Conference on Communication and Signal Processing (ICCSP),* 0185–0189. IEEE, 2019.doi: 10.1109/ICCSP.2019.8698099