www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

DATA-DRIVEN AGRICULTURAL SOLUTIONS: MACHINE LEARNING FOR PRECISION CROP RECOMMENDATIONS AND EMERGING TRENDS

Dr.Rajesh Banala[1]
Associate. Professor
Department of CSE(DS)
Tkr College of Engineering & Technology
rajeshb@tkrcet.com

S. Anuhya [2]
B.Tech(Scholar)
Department of CSE(DS)
Tkr College of Engineering & Technology
anuhya.r.satty@gmail.com

T.Akshita[3]
B.Tech(Scholar)
Department of CSE(DS)
Tkr College of Engineering & Technology
toduriakshitha@gmail.com

N.Santhosh[4]
B.Tech(Scholar)
Department of CSE(DS)
Tkr College of Engineering &Technology
santhoshnayak1818@gmail.com

V.Yuvraj[5]
B.Tech(Scholar)
Department of CSE(DS)
Tkr College Engineering & Technology
yuvrajnkl19@gmail.com

ABSTRACT :

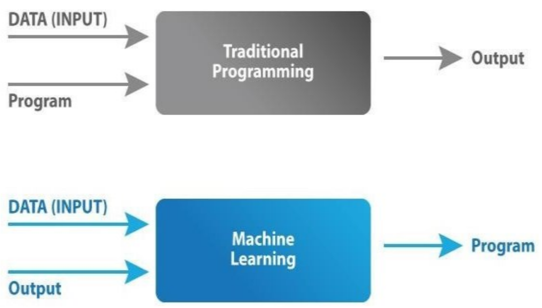
The global population is expected to reach 9.7 billion by 2050. Traditional crop recommendations, dependent on expert knowledge, often lack scalability and precision. Leveraging machine learning (ML) offers a revolutionary approach to optimizing crop selection by analyzing factors like soil properties and climatic conditions. This paper evaluates the performance of seven ML algorithms in recommending crops and identifying pest and disease risks. By automating these tasks, the system reduces reliance on manual expertise while improving accuracy and efficiency. Extensive testing of models using historical datasets has consistently delivered accuracy rates above 95%, with a peak. Farmers of all scales can benefit from this innovative approach to modern agriculture.

This study explores the effectiveness of crop recommendation systems by evaluating the performance of seven distinct machine learning algorithms. Leveraging a combination of

Our findings reveal that the proposed approach consistently achieves remarkable accuracy rates, exceeding 95% across all models, with the highest accuracy peaking at 99.5%. These results underscore the transformative potential of machine learning in modern agriculture, promising significant advancements in crop yield, resource management, and overall farm profitability. Through this innovation, farmers of varying scales stand to benefit from a data-driven approach, fostering resilience and sustainability in agriculture.
Keywords: Crop suggestion, Nitrogen-Phosphorus Potassium (NPK), Humidity, Rainfall, pH, Machine Potassium (NPK), Humidity, Rainfall, pH, Machine Learning (ML), Decision Tree (DT), Support Vector Learning (ML), Decision Tree (DT), Support Vector Machine (SVM), Logistic Regression (

I. INTRODUCTION

Machine learning is a transformative technology that enables computers to learn and adapt without manual programming. Its application in agriculture is particularly impactful, as it addresses the critical challenge of selecting appropriate crops to optimize yield and profitability. Factors such as soil properties, climatic conditions, and market trends are vital in determining crop success. Poor crop choices can result in reduced growth, vulnerability to pests, and lower market demand, ultimately affecting farmers' livelihoods. Machine learning-based crop recommendation systems provide data-driven insights to help farmers make informed decisions, minimizing risks and improving outcomes. These systems analyze a wide range of agricultural data sourced from weather stations, satellites, and IoT sensors. By examining soil quality, rainfall patterns, and temperature variations, machine learning algorithms uncover hidden patterns and correlations that traditional methods often miss. This enables farmers to identify optimal crops for their specific conditions, improving resource efficiency and reducing risks. Additionally, market trend analysis incorporated into these systems ensures that farmers grow crops with high demand, enhancing profitability.



One of the key advantages of these systems is their ability to optimize farming practices beyond crop selection. Machine learning supports irrigation management, pest control, and fertilizer use, enabling farmers to use resources judiciously. By integrating climate resilience strategies, these systems help agriculture adapt to environmental challenges such as changing weather patterns and limited water availability. This contributes to sustainable practices that ensure long-term agricultural viability while protecting natural ecosystems.

Developing a crop recommendation system involves multiple steps, including data collection, preprocessing, and model training. Features such as soil pH, temperature, and humidity are used to train machine learning models, while advanced techniques like feature engineering enhance the system's accuracy. By evaluating and refining these models, researchers can ensure reliable predictions that farmers can trust. These systems empower farmers with actionable insights, reducing dependency on traditional expertise and trial-and- error approaches.

The integration of machine learning in agriculture marks a significant step forward in addressing global food security challenges. By boosting crop yields, conserving resources, and improving profitability, these systems prepare farmers to meet the demands of a growing population sustainably. As technology advances, machine learning will continue to play a pivotal role in revolutionizing agriculture, ensuring productivity and resilience for years to come.

II. RELATED WORK

Machine learning gives computers the ability to learn without being explicitly programmed. In other words, machine learning is turning things or data into numbers and finding patterns in those numbers. The identified patterns help in predicting output for new data points. The fundamental difference between traditional programming and machine learning is shown in Figure 1. Traditional programming and machine learning are two different approaches to solving problems. Traditional programming involves writing code that defines the steps that the software should take to solve the problem. On the other hand, machine learning involves training a model on data so that the model can learn to solve the problem on its own. Machine learning algorithms are primarily categorized into three types based on how machine.

III. METHODOLOGY

This section presents an overview of the methodology pictorially in Figure 6 that we have used to train various models. First, we iterated all the following steps with all the selected machine learning algorithms listed in Section III

1) nput Data: Because the quality and quantity of the data significantly impact a model's accuracy, we ensured the data was clean and well-labeled. As shown in Figure 6, the input to the system is a combination of soil and environmental characteristics. Table II shows a sample of raw data we used to train and test our models

2) Preprocessing: We cleaned the data, removed outliers, and transformed the data into a format that your machine learning algorithm could understand. We primarily removed all null and duplicate records, segregated features from the label column, creating new features from existing feature (also called feature engineering), and described & plotted all the data to ensure no outliers.

3) Choose a machine learning algorithm: In each iteration, we chose one of the seven algorithms we had decided to use. For every selected algorithm, we iterated steps from preprocessing to testing or validating the model to tune the model.

4) Model Configurations: To achieve higher test and cross validation accuracy, we used various configurations such as activation function, epoch, decision tree depth, and a number of nearest neighbors. Figure V shows some other important configurations. Note that we have to keep in mind the performance of the model also; for example, increasing the decision tree's depth could cause to degrade the performance of the mode; Similarly, increasing the value of the number of time data should be fed to the neural network model will also highly impact the performance of the model.

5) Training Models: This is where the machine learning algorithm learns from the data prepared in the "Prepossessing" step.

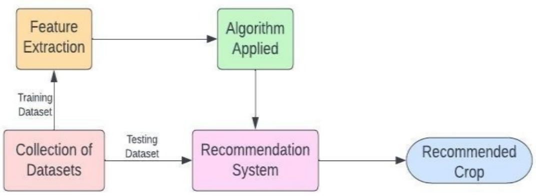
6) Testing Accuracy of the Model: We evaluate the accuracy of the created model against the test data. In addition, we measured the cross-validation accuracy of t to the "Model Configuration" step. In someinstances, we experimented with the feature engineering approach. Suppose the model's accuracy and performance are good at this step.

IV. IMPLEMENTATION DETAILS

Implementing an intelligent urban farming system involves integrating advanced technologies like IoT, data processing, and machine learning to optimize crop production within limited urban spaces. The process begins with establishing a robust data collection infrastructure. IoT-enabled sensors are strategically placed throughout the farming environment to gather real-time information on key factors such as sunlight exposure, soil moisture, temperature, humidity, and CO_2 levels. These sensors continuously monitor the microclimate, ensuring that critical data is captured for analysis. The collected data is then transmitted using wireless protocols like LoRa or Zigbee, which are known for their efficiency and long- range capabilities, to a central cloud server or edge computing device.

Once the data is collected, it undergoes extensive processing and preprocessing to ensure quality and relevance. This involves cleaning the data to remove anomalies and outliers that could skew the analysis. Key features, such as average daily sunlight hours, soil nutrient levels, and water consumption rates, are extracted and standardized. This normalization process ensures that all data inputs are consistent, allowing machine learning models to perform accurately. Properly processed data provides a solid foundation for training machine learning algorithms to recognize patterns and predict outcomes effectively.

Machine learning models play a pivotal role in the system's success. Supervised learning algorithms, such as decision trees and gradient boosting machines (GBMs), classify ideal crops based on environmental conditions and predict optimal watering schedules. Support vector machines (SVMs) analyze temperature and humidity patterns to recommend climate control settings. Additionally, unsupervised learning models, -means clustering, group plants with similar growth needs, while principal component analysis (PCA) reduces data complexity by identifying the most critical growth factors. These models collectively ensure that the farming system is adaptive and responsive to varying conditions.



For deployment, a cloud-based dashboard is developed to provide an interactive platform where real-time sensor data and machine learning predictions are visualized. Farmers can access actionable insights, such as watering times and nutrient recommendations, via this dashboard. To enhance accessibility, a mobile application offers remote monitoring and control capabilities, sending alerts and updates directly to farmers' devices. Additionally, edge computing devices are deployed on-site to process data locally, reducing latency and enabling faster decision-making. This integration of cloud and edge technologies ensures seamless operation and continuous optimization.

Automation is a key component of the intelligent farming system. An automated irrigation system adjusts water supply based on real-time soil moisture data and predicted crop needs, significantly improving water efficiency. Climate control mechanisms, powered by predictive models, regulate LED lighting, ventilation, and humidity to create an optimal environment for plant growth. These automated systems reduce manual labor and ensure that crops receive the precise care they need, enhancing overall productivity.

The system's performance is evaluated using metrics such as prediction accuracy and resource efficiency. Regularly updating and retraining machine learning models with new data helps adapt to seasonal changes and different crop types. This continuous learning process ensures that the system remains effective and responsive to evolving urban farming challenges. Through these implementation steps, intelligent urban farming leverages technology to address spatial constraints, optimize resource usage, and contribute to sustainable food production in metropolitan areas.

V. PROPOSED SYSTEM

In our framework, we have proposed a procedure that is separated into various stages as appeared in Figure 1.

The five phases are as per the following:
1) Collection of Datasets
2) Pre-processing (Noise Removal)
3) Feature Extraction
4) Applied Machine Learning Algorithm
5) Recommendation System
6) Recommended Crop

Flow of the Proposed System

As demonstrated in the figure, the methodology to extract the sentiment conAs demonstrated in the figure, the methodology to extract the sentiment contains the several steps that are described below:\

***(1)** Data Collection:*

The dataset [27] consists of parameters like Nitrogen (N), Phosphorous (P), Potassium (K), PH value of soil, Humidity, Temperature and Rainfall. The datasets have been obtained from the Kaggle website. The data set has 2200 instance or data that have taken from the past historic data. This dataset include twenty two different crops such as rice, maize, chickpea, kidneybeans, pigeonpeas, mothbeans, mungbean, blackgram, lentil, pomegranate, banana, mango, grapes, watermelon, muskmelon, apple, orange, papaya, coconut,cotton, jute, and coffee.

The dataset is separated in Train and Test sets in which 80% of the whole dataset is taken as Train and 20% as Test dataset.

Dataset - Link:
https://www.kaggle.com/datasets/atharva ingle/crop recommendation-dataset

***(2)** Pre-Processing (Noise Removal):*

For the successful application pre-processing is required. The data which is acquired from different resources are sometime in raw form. It may contain some incomplete, redundant, inconsistent data. Therefore in this step such redundant data should be filtered. Data should be normalized

[5]. We also use Power BI to remove peak/downfall, local min-max, outliers, and junk values.

***(3)** Feature Extraction:*

This step is focus on identifying and using most relevant attribute from the dataset. Through this process irrelevant and redundant information is removed for the application of
Classifiers.

we can use to foresee class or value of target variables by learning decision rules deduced from previous data (training data). The Decision tree can be described by two distinct types, namely decision nodes and leaves. The leaves are the results or the

end results. Each node in the tree acts as a test case for some attribute, and each edge descending from that node corresponds to one of the possible answers to the test case. This process is recursive in nature and is repeated for every sub-tree rooted at the new nodes.

VI. Literature Survey

These systems take into consideration a wide range of characteristics, including meteorological conditions, soil type, terrain, temperature and rainfall, crop market price, and crop length when making recommendations for agricultural crops. According to our findings, the following papers were used as sources for our investigation and analysis. Users may make better decisions on what crops to plant according to a strategy presented by Professor Rakesh Shirsath and his co-authors in their research

[1]. Each user's personal information is stored in a subscription based system. Registration has been completed for a farmer. The system includes a module that gathers data on previous crops grown and recommends a suitable crop for planting based on this knowledge. With the help of artificial neutrals networks, we complete the whole process. As a final step, the developer has included a method for receiving input from the farmer so that any issues with the system's functioning may be corrected as

soon as possible. There is a lot of data in knowledge databases, according to Ji-chun Zhao and JianxinGu

[2]. Modules such as knowledge engineer, domain expert, HMI, inference engine, and knowledge base are all instances of users. An effective knowledge foundation for problem resolution is built via the decision system's data collection mechanism. The article uses a number of Hadoop modules to extract characteristics. This is NoSQL, Hive, and Mahout-based system that uses unstructured data and HDFS storage. Wheat and other crops yield numbers were recently released. Crops were not given any consideration. A location detection module, data collecting, and analysis are all part of RSF, according to the research

[3]. Database for agricultural cultivation, database for physiographic data, and module for analysis and storage. As part of the application's location- based mapping feature, a user's present location is compared to a list of similar places and the crops that grow there. A similarity matrix is used to provide recommendations for the user. The location detecting module uses Google Maps API services to get the user's current location so that it can find similar locations. The system, on the other hand, does not ask for human feedback to improve

The process. Majority Voting Technique (MVT) is an ensemble strategy suggested by S.Pudumalar and co-authors in paper [4]. Many models are used to increase forecast accuracy in this method. Using ensemble methods such as Nave Bayes and Random Trees, KNN, and CHAID, it is possible to assure that no matter how bad one approach performs, the others will still be accurate. The majority vote procedure guarantees that the final prediction is correct since models are more likely to give accurate predictions.

VII. Conclusion and Future Work

In conclusion, this research paper has presented crop recommendation models to predict the best crops to grow using multiple advanced machine learning algorithms and a deep neural network. The technique is scalable and easily adapt to new data and regions or countries. The results of this study have several positive implications for the agricultural industry. First, the technique can be used by farmers to make more informed decisions about what crops to grow. Second, the method can be used by governments to develop policies that support the agricultural sector. Third, the method can be used by businesses to create new products and services that support the agricultural industry; Fourth, it will help keep the agricultural goods prices stable. Next, we thoroughly presented agricultural challenges and some interesting future ideas to venture into. Overall, this research has made a significant contribution to the field of agriculture. The technique is scalable, accurate, and easy to use, making it a valuable tool for farmers, governments, and businesses.The open attitude determines the degree and scope of information sharing .Big data analysis technology can effectively improve the crop yield production is updation.The project proposes a novel intelligent system for agricultural crop price prediction. The key idea is to use ensemble of classifiers for prediction. The usage of ensemble of classifiers paves a path way to make a better decision on predictions due to the usage of multiple classifiers. Further, a ranking process is applied for decision making in order to select the classifiers results. This system is used to predict the cost of the crop rate for further. The solution will benefit farmers to maximize productivity in agriculture,reduce soil degradation in cultivated fields, and reduce fertilizer use in crop production by recommending the right crop by considering various attributes. This would provide a comprehensive prediction on the basis of geographical, environmental and economic aspects

he system can be enhanced further to add following functionality:

1. The main future work's aim is to improved data set with larger number of attributes.
2. We need to build a model, which can classify between healthy and diseased crop leaves and also if the crop has any disease, predict which disease is it.
3. To build website and mobile app for easy to use

VIII. ACKNOWLEDGMENTS

I would like to extend my heartfelt gratitude to Dr.Rajesh Banala, my mentor at TKR College of Engineering Technology, for his unwavering guidance, expertise, and continuous support throughout the course of this research. His insightful feedback, constructive criticism, and dedication were pivotal to the successful completion of this project.

IX. REFERENCES

[1] A. L. Samuel, "Some studies in machine learning using the game of checkers," *IBM Journal of Research and Development*, vol. 3, no. 3, pp. 210–229, 1959.

[2] H. Wang, Z. Lei, X. Zhang, B. Zhou, and J. Peng, "Machine learning basics," *Deep learning*, pp. 98–164, 2016.

[3] G. Bonaccorso, *Machine learning algorithms*. PacktPublishing Ltd, 2017.

[4] A. Calzadilla, T. Zhu, K. Rehdanz, R. S. Tol, and C. Ringler, "Climate change and agriculture: Impacts and adaptation options in south africa," *Water Resources and Economics*, vol. 5, pp. 24–48, 2014.

[5] T. Partap, "Hill agriculture: challenges and opportunities," *Indian Journal of Agricultural Economics*, vol. 66, no. 902-2016-67891, 2011.

[6] T. O. Ayodele, "Types of machine learning algorithms," *New advances in machine learning*, vol. 3, pp. 19–48, 2010.

[7] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*. John Wiley & Sons, 2013, vol. 398.

[8] R. E. Wright, "Logistic regression." 1995.

[9] B. Charbuty and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," *Journal of Applied Science and echnology Trends*, vol. 2, no. 01, pp. 20–28, 2021.

[10] M. R. Segal, "Machine learning benchmarks and random forest regression," 2004.

[11] K. Taunk, S. De, S. Verma, and A. Swetapadma, "A brief review of nearest neighbor algorithm for learning and classification," in *2019 nterna*

and ControlSystems (ICCS). IEEE, 2019, pp. 1255–1260. [12] S. Chen, G. I. Webb, L. Liu, and

X. Ma, "A novel selective naïve bayes algorithm,"

Knowledge-Based Systems, vol. 192, p. 105361,

2020.

[13] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B.Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their applications*, vol. 13, no. 4, pp. 18–28, 1998.

[14] I. Steinwart and A. Christmann, *Support vector machines*. Springer Science & Business Media, 2008.

[15] K. Gurney, *An introduction to neural networks*. CRC press, 1997.

[16] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on engineering and Technology (ICET)*, 2017, pp. 1–6.

[17] S. Sharma, S. Sharma, and A. Athaiya, "Activation functions in neural networks," *Towards Data Sci*, vol. 6, no. 12, pp. 310–316, 2017.

[18] J. Dean and S. Ghemawat, "Mapreduce: Simplified data processing

on large clusters," in *OSDI'04: Sixth Symposium on Operating System Design and Implementation*, San Francisco, CA, 2004, pp. 137–150.

[19] D. Dahiphale, R. Karve, A. V. Vasilakos, H. Liu, Z. Yu, A. Chhajer, J. Wang, and C. Wang, "An advanced mapreduce: Cloud mapreduce, enhancements and applications," *IEEE Transactions on Network and service Management*, vol. 11, no. 1, pp. 101–115, 2014.

[20] R. Kiveris, S. Lattanzi, V. Mirrokni, V. Rastogi, and S. Vassilvitskii, "Connected components in mapreduce and beyond," in *SOCC 2014*, 2014. [Online]. Available:

[21] D. Dahiphale, "Mapreduce for graphs processing: New big data algorithm for 2-edge connected components and future ideas," *IEEE Access*, vol. 11, pp. 54 986–55 001, 2023.

[22] A. Oikonomidis, C. Catal, and A. Kassahun, "Deep learning for crop yield prediction: a systematic literature review," *New Zealand Journal of Crop and Horticultural Science*, vol. 51, no. 1, pp. 1–26, 2023.

[23] P. A, S. Chakraborty, A. Kumar, and O. R. Pooniwala, "Intelligent crop recommendation system using machine learning," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, 2021, pp. 843–848.

[24] Z. Doshi, S. Nadkarni, R. Agrawal, and N. Shah, "Agroconsultant: Intelligent crop recommendation system using machine learning algorithms," in *2018 Fourth International*