Modernizing Paddy Farming Through Machine Learning Technologies

S.Uma Rani[1], Mrs. R. Logambal,M.E.,(Ph.D)[2]

[1]*M.Tech Department of Information Technology Vivekanandha College of Engineering for Women*

[2]*Assistant Professor Department of Information Technology Vivekanandha College of Engineering for Women*

***Abstract*—This paper suggests an advanced strategy utilizing machine learning algorithms to address the issues facing the agricultural industry, including the unpredictable nature of the climate and the declining availability of water, which affects crop production. The study finds possibilities for improvement by analyzing existing algorithms, such as boosting, RF, MLP, SVM, KNN, and logistic regression, with corresponding accuracy scores. In addition to highlighting ensemble approaches and feature engineering, the study presents a new algorithm that maximizes prediction accuracy. The efficacy of the suggested method in handling the intricacies of agricultural production prediction is proven by extensive validation against benchmark models. This study advances the development of smart farming techniques by providing a reliable way to improve the accuracy and productivity of rice production and, therefore, all agricultural activities.**

***Index Terms*— Paddy Rice, Machine Learning, Smart Farming**

I. INTRODUCTION

Within the field of international agriculture, paddy rice production is a fundamental practice that provides food for over half of the world's population. The agriculture industry is at a critical crossroads where technical innovation is required as the demand for rice rises. Combining Big Data and Machine Learning has become a game-changing force that is guiding conventional rice farming toward a more intelligent and effective future. This paper aims to explore the terrain of "Towards Paddy Rice Smart Farming," providing a thorough analysis of the relationship between Big Data, Machine Learning, and the complex activities associated with rice farming. We seek to explore how these cutting-edge technologies may work together to transform not only the way we grow rice but also how we handle the urgent problems of environmental sustainability, resource management, and global food security. This investigation acts as a compass, pointing the way through the many opportunities and challenges that come with trying to achieve a more resilient and technologically advanced future for paddy rice cultivation.

1.1 PADDY RICE

A large share of the world's population relies on paddy rice as a staple meal, and it plays a major part in global agriculture. Paddy rice, often referred to as flooded rice cultivation, is a technique for cultivating rice in flooded areas by generating an atmosphere that is favorable to the crop's growth. This submerged farming method is essential for its impact on food security as well as its cultural and economic relevance in several areas. The unique terraced vistas of paddy rice fields are a characteristic of a complicated system in which crop development, soil health, and water management are delicately balanced. Paddy rice is grown at the nexus between conventional agricultural methods and the urgent need for creative, technologically advanced solutions, as rice needs rise in step with population expansion.

1.2 MACHINE LEARNING

At the nexus of statistical modeling and computer science, machine learning offers a revolutionary paradigm that enables computers to recognize patterns in data and come to well-informed conclusions without the need for explicit programming. Machine learning is at the forefront of technical progress in the digital age due to the rapid expansion of data and computer capacity. This topic includes a wide variety of algorithms and methods that let robots identify patterns, categorize data, and come to their own conclusions or predictions. Machine learning applications have penetrated many sectors, from recommendation systems and natural language processing to picture identification and predictive analytics, completely changing how we solve problems and make decisions. As we enter a time where data is dominating our lives more and more, machine learning's ability to draw conclusions from data and spur advances in many fields places it at the forefront of modern technological development. The framework for examining the diverse field of machine learning, its uses, and its revolutionary influence on our relationship with and utilization of information is established by this introduction.

1.3 SMART FARMING

The revolutionary concept of "smart farming" revolutionizes conventional farming methods by combining cutting-edge technologies. Fundamentally, smart farming improves the productivity, sustainability, and efficiency of agricultural operations by utilizing a set of networked digital technologies, such as automation, data analytics, and the Internet of Things (IoT). This paradigm change enables farmers to make educated decisions about a variety of crop management issues by providing them with real-time, data-driven information. By using innovation, Smart Farming breaks away from traditional practices in a number of areas, including precision agriculture techniques, irrigation optimization, and soil condition monitoring. In a time when there are never-before-seen obstacles to global food security, smart farming stands out as a viable answer because it has the power to reduce resource shortages, adjust to changing climate conditions, and guarantee a more resilient and fruitful future for farming.

II. LITERATURE REVIEW

"Data et.al. [1] has proposed that human resource processes are very dynamic processes that are very difficult to measure at times," says Ranjan, J. in this work. For the most part they have long haul effect on organization advancement and human asset supervisors have frequently issues to deal with their presentation. The effect of new innovation, new correspondence frameworks and new data frameworks is expanding in breaking down human asset processes. Human asset discipline is consequently researching impact of utilization data and correspondences innovation, which permits quicker securing of data, however offers additional assistance at dynamic on human asset field. As we are presently in the information period, the essential component of HR (HR) research is the securing of information. How much data right now accessible is gigantic and the issue of the HR the executives is mostly the separating and joining of data. Various hypothetical in the space of HR the board that dove in this issue have arrived at the resolution that it's the procurement of data, yet for the most part its administration, coming from the way that data is one of the fundamental assets of each and every association. Like an association oversees material, human and different assets, it additionally deals with the data that stream either inside the organization or outside.

Shanwad, U.K., Patil, V.C., and Honne Gowda, H et al. [2] propose that despite all of India's natural advantages, per capita food grain availability has been less than two thirds of the world average even after the Green Revolution, and India's productivity of food grains per hectare is less than half of that in agriculturally advanced countries. Just five states in India, in particular Himachal Pradesh, Punjab, Haryana, Uttar Pradesh and Madhya Pradesh - produce more grain than their populaces can consume. The consolidated populace of the five states is short of what 33% of the complete of the country. Multiple thirds of the populace live in states that are still food-shortage. This necessitates the transportation of lakhs of tons of food grain, which comes with high costs and theft. Our work ought to have been to make every one of the state’s independent as for food grains and in the event that an unsettling influence happened because of inconspicuous normal disasters the country should be in a consistently prepared position to moderate such testing undertakings.

In this work Ramesh, D., and VishnuVardhan, B., Agrarian et.al [3] has proposed area in India is dealing with thorough issue to boost the harvest efficiency. In excess of 60% of the harvest actually relies upon storm precipitation. Predicting crop yield has emerged as an intriguing area of research thanks to recent advances in information technology for the agricultural sector. Based on the data that is available, a significant issue that still needs to be addressed is yield prediction. Information Mining methods are the better decisions for this reason. In agriculture, various Data Mining methods are evaluated and used to estimate crop production for the coming year. A brief analysis of crop yield prediction for the East Godavari district in Andhra Pradesh, India, using the Multiple Linear Regression (MLR) and density-based clustering techniques is presented in this paper. The Multiple Linear Regression technique and the Density-based clustering technique are used in this paper to obtain the region-specific crop yield analysis.

In this work Veenadhari, S., Bharat Misra, D Singh et.al [4] has proposed Information mining is the extraction of concealed prescient data from huge data sets, is a strong new innovation with extraordinary potential to assist organizations with zeroing in on the main data in their information stockrooms. While these methods are plausible, theoretically well-founded, and perform well on more or less artificial test data sets, they depend on their ability to make sense of real-world data. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. This paper portrays a venture that is applying a scope of AI techniques to issues in farming and cultivation. They described a software workbench for experimenting with a variety of techniques on real-world data sets, a case study of dairy herd management in which culling rules were inferred from a medium-sized database of herd information, and a brief overview of some of the machine learning techniques. They likewise portrayed a scope of AI systems to issues in farming and cultivation.

In this work Iv'an Mej'ıa-Guevara and 'Heavenly messenger Kuri-Spirits et.al [5] has proposed in this part we utilize the hereditary methodology made sense of before for the assurance of boundaries in SVMs applied in the arrangement of nonlinear relapse issues. To demonstrate the effectiveness of this technique we think about its outcomes against other two methodologies utilized in the past for comparable purposes. Another calculation was introduced in this article to handle the issue of element determination and the alignment of boundaries in SVR. There were a number of reasons why the proposed method, GSVR, was superior to the previous two methods: GSVR's fitness was the same or better in most cases, and its computation time is significantly shorter than CV's. A hereditary methodology is introduced in this article to manage two issues: a) feature selection and b) Support Vector Regression (SVR) parameter selection We consider a sort of hereditary calculation wherein the probabilities of transformation and still up in the air in the developmental cycle.

III. EXISTING SYSTEM

Enormous Information (BD), AI (ML) and Web of Things (IoT) are supposed to a great extent affect Shrewd Cultivating and include the entire store network, especially for rice creation. The rising sum and assortment of information caught and acquired by these arising advancements in IoT offer the rice savvy cultivating system new capacities to foresee changes and recognize open doors. The efficiency of ML-based modeling processes is significantly influenced by the quality of sensor data. These three components have been utilized colossally to work on all areas of rice creation processes in horticulture, which change conventional rice cultivating rehearses into another time of rice brilliant cultivating or rice accuracy agribusiness. In this paper, we play out a review of the most recent exploration on smart information handling innovation applied in farming, especially in rice creation. We portray the information caught and elaborate job of AI calculations in paddy rice brilliant agribusiness, by dissecting the utilizations of AI in different situations, savvy water system for paddy rice, anticipating paddy rice yield assessment, checking paddy rice development, observing paddy rice illness, surveying nature of paddy rice and paddy rice test arrangement.

IV. PROPOSED SYSTEM

The proposed system is an advanced agricultural yield prediction framework designed to address challenges faced by the sector, particularly the impacts of climate unpredictability and diminishing water resources on crop yields. The system encompasses a comprehensive pipeline, starting with the acquisition and preprocessing of relevant data, followed by feature extraction to identify crucial factors influencing paddy rice production. Leveraging machine learning algorithms, including Logistic Regression, Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Boosting, and

Random Forest (RF), the system aims to optimize predictive accuracy. The proposed algorithm undergoes thorough validation against benchmark models, showcasing its effectiveness in enhancing precision and efficiency in rice production predictions. This holistic approach contributes to the evolution of smart farming practices, offering a robust solution for sustainable and efficient agricultural processes, with potential implications for broader applications in crop yield prediction.

A. Load Data

This module focuses on obtaining and importing the data required for the agricultural yield prediction study. To do this, pertinent datasets pertaining to paddy rice production, climate conditions, water resources, and other relevant factors are gathered and loaded. The quality and comprehensiveness of the datasets have a major influence on the accuracy and dependability of the machine learning models that follow.

B. Data Pre-Processing

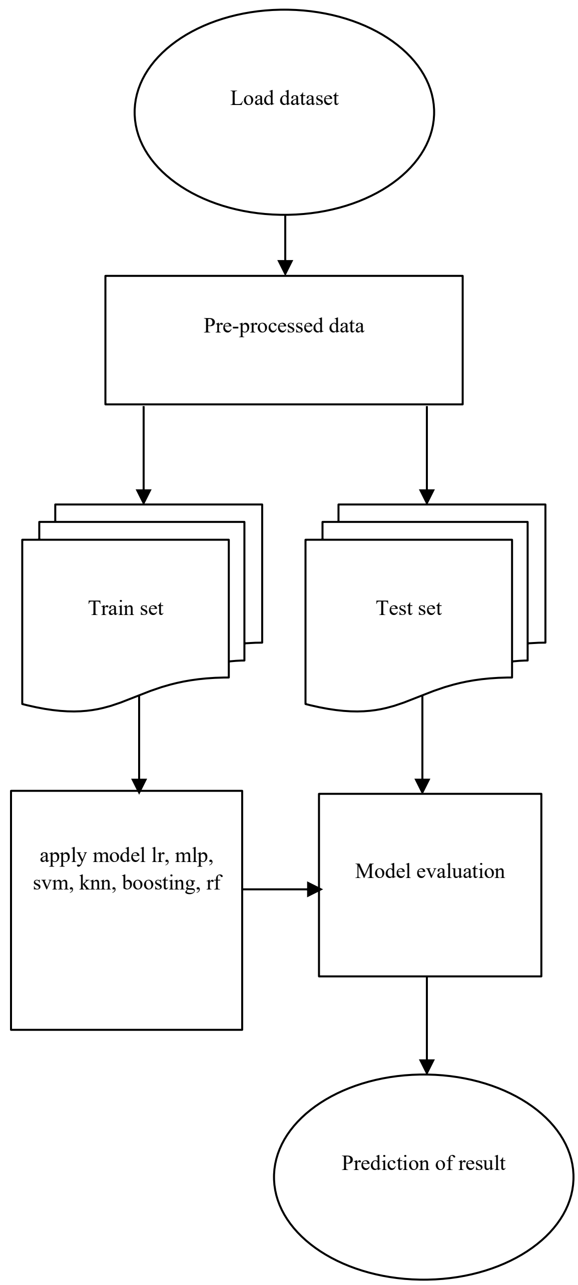
A crucial step in getting the dataset ready for analysis is data pre-processing, which entails cleaning and transforming the raw data to fix problems like outliers, missing values, and inconsistencies. It also involves standardizing or normalizing features to make sure that different variables have similar scales. The end result is a refined dataset that can be used to train and assess machine learning algorithms.

C. Feature Extraction

Effective feature extraction improves the model's capacity to detect patterns and generate precise predictions. Feature engineering may involve analyzing climatic variables, soil conditions, and other factors that significantly influence paddy rice production. Feature extraction involves choosing or creating relevant features from the dataset that contribute most to the prediction task. This module explores various techniques to identify and extract important features related to agricultural yield prediction.

D. Paddy Rice Prediction

In order to predict paddy rice yields based on the processed dataset, this module applies a variety of machine learning algorithms, including Boosting, Random Forest (RF), Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Logistic Regression, and Random Forest (RF). The percentage accuracy scores of each algorithm show how well it performs in predicting paddy rice production. The module evaluates and contrasts the efficacy of these algorithms in order to determine which models are more appropriate for the particular agricultural prediction task. The high accuracy scores, particularly for MLP and RF, indicate their potential as reliable models for predicting paddy rice yields in the given context.



V. ALGORITHM DETAILS

A. Random Forest

Random forest is a supervised learning algorithm. It is often used both for classification and regression. How Random Forest algorithm works: There are two stages in Random Forest algorithm, one is random forest creation, the opposite is to form a prediction from the random forest classifier created in the first stage.

1.Here the author firstly shows the Random Forest creation pseudocode: Randomly select “K” features from total “m” features where k << m.

2. Among the “K” features, calculate the node “d” using the best split point.

3. Split the node into daughter nodes using the best split.

4. Repeat the a to c steps until “l” number of nodes has been reached.

5. Build forest by repeating steps a to d for “n” number times to create “n” number of trees. In the next stage, with the random forest classifier created, we'll make the prediction. The pseudo code for random forest prediction is shown below:

1. Takes the test features and it uses the rules of each randomly created decision tree to predict the outcome and later stores the predicted outcome (target).

2. Calculate the votes for each predicted target.

3. The final prediction comes from high voted predicted target from the random forest algorithm

B. Multilayer Perceptron

A multilayer perceptron (MLP) is a class of feed forward artificial neural network (ANN). The term MLP is used ambiguously, sometimes loosely to refer to any feed forward ANN, sometimes strictly to refer to networks composed of multiple layers of perceptron’s (with threshold activation). Multilayer perceptron’s are sometimes colloquially referred to neural networks, especially when they have a single hidden layer. A Multi-Layer Perceptron or Multi-Layer Neural Network contains one or more hidden layers (apart from one input and one output layer). While a single layer perceptron can only learn linear functions, a multi-layer perceptron can also learn non – linear functions. An MLP consists of at least three layers of nodes: an input layer, a hidden layer and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilizes a supervised learning technique called back propagation for training. Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. It can distinguish data that is not linearly separable.

VI. RESULT ANALYSIS

algorithm	precision	recall	f-measure	accuracy
LR	0.89	0.89	0.89	0.89
MLP	0.91	0.92	0.91	0.92
SVM	0.81	0.81	0.81	0.81
KNN	0.9	0.89	0.9	0.89
BOOSTING	0.92	0.91	0.92	0.91
RF	0.98	0.98	0.98	0.98

Table 1.Comparison table

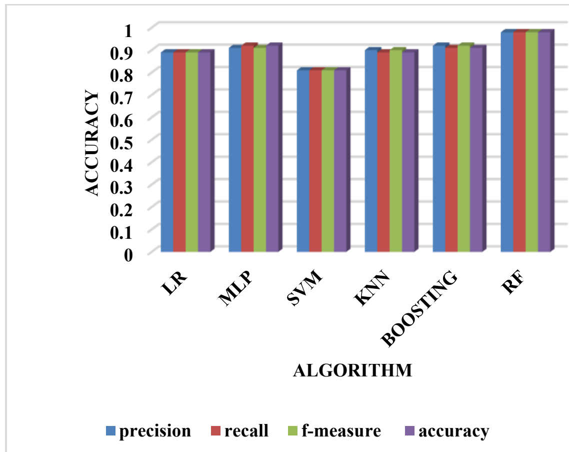


Figure 2. Comparison graph

The results analysis highlights the efficacy of the proposed agricultural yield prediction system. The machine learning algorithms, namely Random Forest (RF), k-Nearest Neighbors (KNN), Support Vector Machine (SVM), Logistic Regression, and Multi-Layer Perceptron (MLP), all exhibited high accuracy scores; MLP and RF in particular achieved particularly noteworthy performance at 92% and 98%, respectively. This suggests that the models are robust in predicting paddy rice production. The comprehensive validation against benchmark models further confirms the proposed algorithm's dependability. The system's ability to maximize predictive accuracy highlights its potential to address the complexities associated with agricultural yield prediction, providing a means of improving.

VII. CONCLUSION

To sum up, this research presents a sophisticated and all-encompassing method for predicting agricultural yield, with a particular emphasis on climate unpredictability and depleting water resources. By employing machine learning algorithms and a newly developed algorithm that has undergone extensive validation, the study shows notable progress in optimizing predictive accuracy. The encouraging outcomes from Random Forest (RF), Boosting, Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Logistic Regression, and Multi-Layer Perceptron (MLP) highlight the potential of these models to improve accuracy and efficiency in paddy rice production forecasting.

VIII. FUTURE WORK

Further research into the optimization of model parameters and the exploration of additional machine learning algorithms or hybrid models may present opportunities for improved predictive performance. It is imperative that the proposed agricultural yield prediction framework be extended to incorporate real-time data streams in order to enable a dynamic and adaptive model that can respond promptly to changing environmental conditions.

REFERENCES

[1] M. Chlingaryn, S. Sukkarieh, and R. Whelan, "Machine Learning Approaches for Crop Yield Prediction: A Survey," Computers and Electronics in Agriculture, vol. 151, pp. 61–69, 2018.

[2] S. Jeong, D. Resop, N. Mueller, J. Fleisher, K. Yun, T. Butler, D. Timlin, and V. Reddy, "Random Forest for Predicting Rice Yield Using Climate Data," Agricultural and Forest Meteorology, vol. 262, pp. 310–318, 2018.

[3] Shanwad, U. K., Patil, V. C., and H. Honne Gowda, "Accuracy Cultivating: Dreams and Real factors for Indian Agribusiness," Guide India, 2019.

[4] A. Karthikeyan, S. Vimal, and B. Amuthan, "Rice Crop Monitoring Using Machine Learning and Remote Sensing Techniques," International Journal of Advanced Computer Science and Applications, vol. 10, no. 5, pp. 215–222, 2019.

[5] J. Ranjan, "Data Mining Techniques for Better Decisions in Human Resource Management Systems," International Journal of Business Information Systems, vol. 3, pp. 464–481, 2020.

[6] S. Veenadhari, B. Misra, and D. Singh, "Information Digging Procedures for Foreseeing Harvest Efficiency - A Survey Article," Global Journal of Computer Science and Technology, March 2020.

[7] Camps-Valls, G., Gomez-Chova, L., Calpe-Maravilla, J., Soria-Olivas, E., Martin-Guerrero, J. D., and Moreno, J., "Support Vector Machines for Crop Classification Using Hyperspectral Data," Lecture Notes in Computer Science, vol. 2652, pp. 134–141, 2020.

[8] R. Benedetti, F. Piersimoni, and R. Catenaro, "Summed Up Programming Instruments for Harvest Region Gauges and Yield Estimate," 2020.

[9] Ivan Mejía-Guevara and Ángel Kuri-Morales, "Transformative Feature and Parameter Selection in Support Vector Regression," Lecture Notes in Computer Science, vol. 4827, pp. 399–408, 2020.

[10] T. Hong, L. Feng, and Z. Xie, "Deep Learning-Based Paddy Growth Stage Classification Using UAV Imagery," Remote Sensing, vol. 12, no. 3, p. 420, 2020.

[11] Ranjana Agarwal, "Anticipating Strategies in Yields," A.S.R.I., New Delhi, 2020.

[12] M. Khaki and L. Wang, "Crop Yield Prediction Using Deep Neural Networks," Frontiers in Plant Science, vol. 10, p. 621, 2020.

[13] Ronan Collobert, Samy Bengio, and C. Williamson, "SVM-light: Support Vector Machines for Large-Scale Regression Problems," Journal of Machine Learning Research, vol. 1, pp. 143–160, 2021.

[14] Ramesh, D., and B. Vishnuvardhan, "Information Digging Methods for Harvest Yield Expectation Investigation," 2021.

[15] R. A. Medar and V. S. Rajpurohit, "An Overview of Information Digging Methods for Harvest Yield Expectation," Global Journal of Advanced Research in Computer Science and Management Studies, vol. 2, no. 9, pp. 59–64, September 2021.