ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 08 August 2024

Optimizing Crop Selection using Machine Learning and Environmental Factors

[1] Krishna Teja Yalla, [2] Sai Sitharam Sashank Akundi, [3] Usha Kiran Alluri, [4] Vuda Sreenivasa Rao
[5] Srinivasa Reddy Mallidi

[1][2][3][4][5] Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, India
Email: [1] krishnatejayalla@gmail.com, [2] seetharamshashak4@gmail.com, [3] alluriushakiran2019@gmail.com,
[4] vsreenivasarao@gmail.com, [5] srinivasareddymallidi22@gmail.com

Abstract*— The optimization of crop yield has arisen as a major problem in modern agriculture. because of the growing demand for food and the necessity for efficient resource management. Accuracy and precision are hampered by the limits associated with conventional agricultural production and prediction techniques impacting crop optimization. The practice of projecting a crop's future output is known as crop yield prediction. while considering the dependent factors into account. While the methods of Machine Learning (ML) are widely used to forecast yield and their potential to make accurate assessments and possess an adequate understanding of crop attributes usually leads to more accurate forecasts and qualifies them for prediction. Using factors of environment such as Temperature, PH, Rainfall, Humidity and nutrient levels (N, P, K) plays a vital role in evaluating crop health and growth phases. This research explores the serviceability of algorithms for crop optimization by considering accuracy, among other criteria. The study's conclusion notes that it expects further developments in ML techniques and existing technology to produce comprehensive, affordable solutions for better estimates of crop and environmental states, enabling well-informed decision-making. This study examines the many ML approaches used in crop optimization and offers a thorough evaluation of the approaches' accuracy.

***Index Terms*— Machine Learning, Crop Selection, Environmental factors, Crop prediction**

I. INTRODUCTION

Agriculture, being the backbone and plays a key role where economic growth of a country like India is considered is facing the challenge of meeting the increasing global requirement for food. In a situation where crop yield rates are steadily declining, A clever system is required. that can address this issue of declining crop output. With the delicate interplay of environmental specifications like soil nutrients, temperature, humidity, PH, and rainfall, the need for precision farming becomes paramount. ML relies heavily on Feature Selection and Classification. Feature selection entails choosing the best attributes from a dataset. It entails selecting an appropriate subset of the original attributes from a larger set that complies with a benchmark that has been like class separability or classification performance.

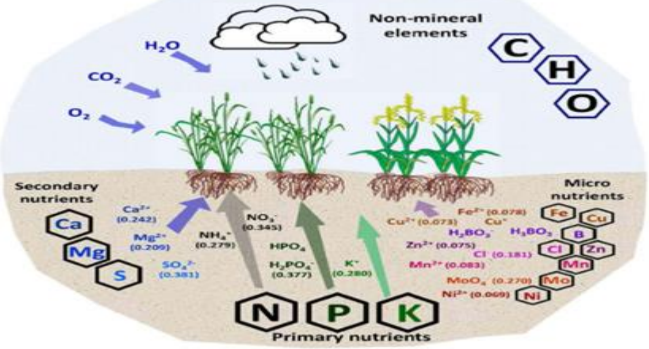


Fig [1]: Environmental factors affecting crop conditions [10]

The key variables under consideration include Nitrogen (N), Phosphorus (P), Potassium (K), temperature, humidity, PH levels, and rainfall patterns. These elements have a notable impact on plant development and crop yield. Scientists are using ML to improve the precision and promptness of crop optimization. To assess how well various ml and related methodologies are analyzed and anticipate crop conditions.

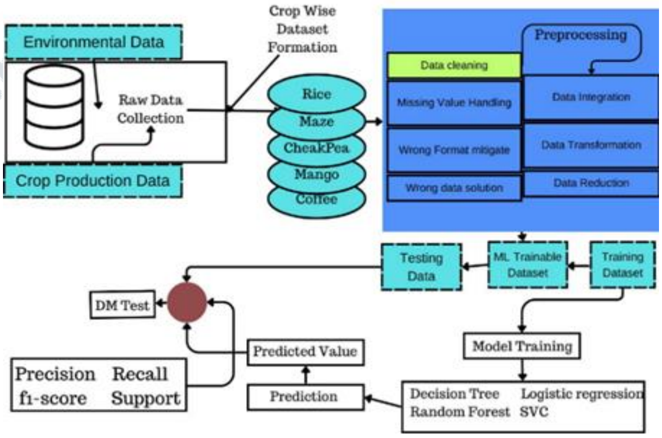


Fig [2]: Flowchart for Crop Optimization by using ML.

Identifying the benefits and drawbacks of different ml algorithms and their relevance in specific agricultural scenarios. Because of their ability to recognize patterns in big datasets and adjust to changing conditions, ml techniques offer a potential alternative for enhancing crop sustainability. By aligning crop choices with environmental conditions, farmers can optimize resource utilization, reduce environmental impact, and maximize overall productivity. The findings of this study have the potential to change agricultural techniques, opening the path for a more resilient and productive future. Ensuring optimal crop selection remains critical for maximizing agricultural production.

Optimizing crop selection is a comprehensive approach that aims to maximize agricultural efficiency, economic returns, and environmental sustainability. This process requires various factors, including local climate and weather conditions, soil characteristics, water availability, and the specific demands of the market. It plays a vital role in the success of agricultural practices. Crop selection determines the profitability, productivity, and sustainability in farming operations. Traditionally, farmers choose crops based on their experience, local knowledge, and basic environmental indicators to decide which crops to cultivate. However, these traditional methods lacked the precision and accuracy needed to maximize yields and minimize risks. Evaluating data and providing predictions depend on the correlations between environmental and climatic factors.ml offers a transformative solution to the limitations of traditional crop selection methods. ML and data analytics improve predictive modelling, allowing farmers to anticipate and mitigate future issues.

II. RELATED WORKS

The Decision_Tree_Algorithm[1] is applied. Decision_Trees have a tree structure similar to a flowchart. Three nodes make up the decision_tree: the internal_node, the leaf_node, and the root _node. The root_node is the highest element. The internal_nodes are those that are located between the leaf_node, which serves as the structure's terminal element. Every leaf_node has a class label, every branch shows the test's result, and every internal_node indicates a test on an attribute. Rules are framed and branched out from nodes and sub-nodes until a decision is made, forming the tree. The problem of growing a decision_tree from available data has been studied by many researchers from various fields, such as pattern recognition, ML.

Different approaches are used to create decision rules for decision_tree as decision_tree classifiers. Based on quality parameters[2] like information gain, gain ratio, gini index, etc., the nodes are chosen from the top level. Various algorithms are employed, including ID3, CART, and CHAID. Gain Ratio is used by the C4.5 Decision_Tree to build the tree; the element with the highest gain ratio is designated as the root_node, and the dataset is divided according to the values of the root element. Once more, each sub-node's information gain is computed separately, and the procedure is continued until the prediction is fulfilled. Decision_tree classifiers are frequently used in data processing, prediction, management of missing information, and variable selection. because they lack ambiguity, are simple to use, and remain reliable even when there are missing values. It is possible to employ continuous or discrete variables as independent or target variables.

Support Vector Machine (SVM) algorithm is employed[3]. A potent supervised ML approach for regression and classification problems is called Support_Vector_Machine. Fundamentally, SVM looks for the best hyperplane in an n-dimensional space to divide data points into distinct groups. The data points that are closest to the hyperplane and determine its orientation and position are referred to as "support vectors". To improve the model's robustness and generalization SVM maximizes the margin between support vectors of distinct classes. When classes cannot be separated linearly, SVM maps the input data into a higher-dimensional space where a hyperplane can successfully divide the classes using a method known as the kernel trick.

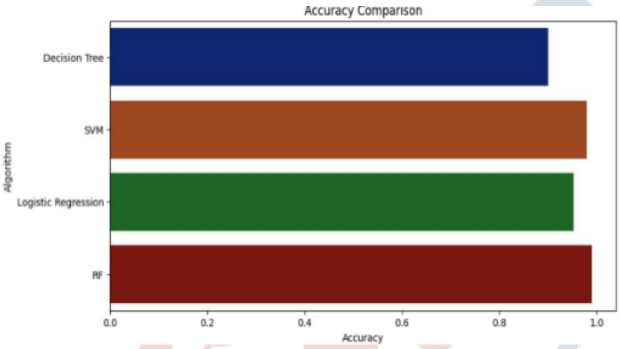
Vector machine classifiers are used in numerous fields for many reasons[4]. Initially, classification problems can be solved using SVM because it is built upon solid mathematical foundations and strong theoretical background knowledge. Secondly, datasets with many features work well with SVMs, which are usually good performers in high-dimensional spaces. Thirdly, whether a classification problem is linear or non-linear can be effectively addressed by SVM through the use of different kernel functions. Fourthly, SVM after being trained on small-scale data exhibits robust generalization ability that makes it capable of classifying new points accurately. The goal here is to choose such a hyperplane that would correctly separate positive and negative views maximizing distance between them (margin). It means if we have two groups not intersecting each other at all then everything should be fine but when they do overlap things get more complicated – especially if those classes cannot

The Random Forest Algorithm is employed[5]. The most well-known and potent supervised machine learning algorithm, Random Forest, can handle both classification and regression tasks. It works by building a large number of decision trees during training and producing class outputs, which are the mean prediction (regression) or the mode of the classes (classification) of the individual trees. In a forest with more trees, the prediction is more reliable. The many objectives of the Random Forest algorithm boost the accuracy of training and testing speed. The random forest_classifier works similarly to a corps and is made up of a huge number of unique trees. The random forest_classifier divides each unique decision tree into a class prediction, and our model's prediction is determined by the class with the most votes[6]. The random forest_classifier is an algorithm that relies on crowd intelligence and is both easy and powerful. The random forest_algorithm is utilized to categorize data. The training sets are used. The original dataset is created and then utilized to construct the decision_tree. A random forest is generated after training sets are constructed using a decision_tree.

Logistic_regression is a tool for capturing the non linear connections between predictor_variables and the likelihood of an event[7]. In agriculture the link between factors and crop suitability can be quite intricate. Logistic_regression effectively models these relationships by fitting a sigmoid curve to the data enabling predictions in cases where the relationship is not straightforward it is main task. The coefficients derived from regression models offer insights into how each predictor variable influences the outcome. In our project this interpretability is crucial for understanding which environmental factors play a role, in determining crop suitability. This understanding can help farmers make informed decisions to improve their land management practices and boost crop yield.

By using the above-mentioned ML techniques. we performed operations on the dataset crop recommendation a csv data file taken from Kaggle and we obtain accuracy of these ML techniques as shown in below table.

S.NO	ML ALGORITHM	ACCURACY
1	DECISION_TREE	90.000
2	LOGISTIC_REGRESSION	95.227
3	SVC	96.954
4	RANDOM FOREST	97.090



III. PROBLEM STATEMENT

The existing methods can have restricted feature ranking, be economically impractical for small-scale farms, have long processing periods, and be inefficient. To tackle these problems, the proposed method combines all ML optimization techniques, including logistic regression, random forest, and decision trees. A yield prediction and data analysis integrated ensembled model that is optimized is given. Crop prediction accuracy is raised and precision farming techniques are enhanced by adopting a comprehensive approach

IV. METHODOLOGIES

The accuracy obtained by ML techniques such as decision tree classifiers, logistic regression, support vector machines, and Random Forests is prioritised in this research report. The dataset utilized in the study was obtained from Kaggle. Modifying models in agricultural contexts by ensembling techniques ensures their dependability and usefulness. Stacking is the most well-known ensembling strategy for classifier representation in data classification. Stacking is a technique for improving model performance by predicting many nodes while building a new model. The gathering of the dataset is an essential activity in this step. Data preparation makes sure that missing data is avoided and that the missing values are filled up using the right methods. use the stacking algorithm It chooses the characteristics that let you to create several distinct learners, which you then utilize to create an intermediate prediction—one for every learnt model. Next, a new model is added that gains knowledge from the intermediate predictions made for the same target.It is stated that this last model is layered on top of the others. It enhances overall performance and frequently produces a model that is superior to all individual intermediate models.

Data Collection and Preprocessing

Prior to executing the stacking process on the dataset, Kaggle provides crop recommendations. This dataset needs to undergo data pretreatment. During this process, categorical variables are converted into numerical values, the mean of the corresponding features fills in the missing values, and the pertinent features that enhance model performance after stacking are chosen.

S.NO	ML Algorithm	Accuracy
1	Decision tree classifier	90.000
2	Logistic regression	95.227
3	SVC	96.954
4	Random Forest	97.090
5	Stacking	99.045

Ensembling Using Stacking

Stacking is a method used to improve prediction capability of a model. It does so by combining many models' outputs to produce one final prediction that usually has higher accuracy than if any single model were operating alone.The precision of any model depends on the features it uses. Some factors that can influence might include different species of crops, types of soils, humidity levels, rainfall amounts and temperatures among others. Models are categorized into two; base models and meta models which are trained using training data from base models thus generating unique predictions in them whereby the input data is combined with base models' predictions to give the final prediction in meta models.An accurate result is achieved through stacking algorithm.

1	Stacking	99.045

Algorithm For Stacking

1.Divide your data: Split your data into two parts - one for training models and the other for testing their performance.

2. Train different models: Train different models using various algorithms like decision trees, logistic regression, support vector machines, and random forest. Each model

learns from the training data to make predictions.

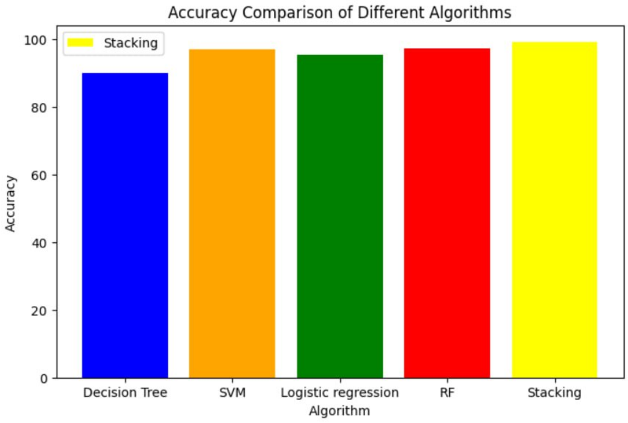
3. Make predictions: Let each trained model make predictions on the test data. So, you'll have predictions from each model for every data point in the test set.

4. Combine predictions: Take all these predictions from different models and put them together. Think of it as gathering opinions from different experts.

5. Train a meta model: Now, you have a bunch of predictions from different models. Train another model (this is called the meta model) using these predictions as input. This meta model learns to combine the predictions from the base models to make a final decision.

6. Final prediction: When you have new data, use your base models to make predictions, then use those predictions as input for the meta model to get the final prediction

V. RESULTS



We assessed single ML models such as Decision Tree, Logistic Regression, Support Vector Machine (SVM), and Random Forest on our dataset prior to applying the stacking algorithm. The initial findings indicated that these models did not perform well in terms of accuracy since their accuracies were lower than what we had expected. Therefore, we used a stacking algorithm to boost accuracy.

The stacking improved the dataset's accuracy to 99.045 which is better than that of individual models by a large margin.

This shows that the stacking algorithm greatly utilizes the strengths shown by different individual models thus yielding higher accuracy rates in choosing crops for cultivation. In addition, it also overcomes weak points found in each model through combination with other diverse modeling approaches thereby generating more reliable forecasts.

VI. CONCLUSION

Rainfall, temperature, humidity, soil nutrients (N, P, and K levels), and pH must be combined using Ml algorithms in order for the process to be finished. As a result, choosing which crop to grow where on a farm will be easier. By employing data obtained from these variables in previously unheard-of ways, farmers may make more educated choices regarding the crops they grow, resulting in higher yields and resource conservation for sustainable agriculture. Furthermore, even the technology that is currently in use may alter its tactics whenever unpredictable weather patterns change due to ML systems constant adaptation to their surroundings. This would help farmers make informed crop selections based on location and time through dealing skills.

VII. FUTURE ENHANCEMENT

This analysis exhibits that a range of attributes are used in the chosen articles, depending on the extent of the study and the data's accessibility. While crop optimization is examined in every paper, the features vary. The magnitude, crop, and geological location of the investigations also vary. To identify the top-performing model, models with more and fewer features should be evaluated. These enhancements could lead to a rise in the application and accuracy of crop optimization. Here are some concepts for the next work

- Use Additional Data Sources.
- Resolution in Space and Time.
- Advanced Engineering of Features.
- Adaptive and Online Learning Models.

REFERENCES

[1] Yan-yan SONG, Ying LU, Decision tree methods: applications for classification and prediction Shanghai Archives of Psychiatry, 2015, Vol. 27, No. 2

[2] Nikita Patel, Saurabh Upadhyay, Study of Various Decision Tree Pruning Methods with their Empirical Comparison in WEKA, International Journal of Computer Applications (0975 – 8887) Volume 60– No.12, December 2012

[3] S.Veenadhari, Dr Bharat Misra, Dr CD Singh, Machine learning approach for forecasting crop yield based on climatic parameters, 2014 International Conference on Computer Communication and Informatics (ICCCI -2014), Jan. 03 – 05, 2014, Coimbatore, INDIA

[4] Mayank Champaneri, Darpan Chachpara, Chaitanya Chandvidkar, Mansing Rathod, Crop Yield Prediction Using Machine Learning, International Journal of Science and Research (IJSR) ISSN: 2319-7064

[5] Dr. V. Geetha, A. Punitha, M. Abarna, M. Akshaya, S.Illakiya, AP.Janani, An Effective Crop Prediction Using Random Forest Algorithm, IEEE ICSCAN 2020

[6] Shriya Sahu, Meenu Chawla, Nilay Khare, An Efficient Analysis Of Crop Yield Prediction Using Hadoop Framework Based On Random Forest Approach, International Conference on Computing, Communication and Automation (ICCCA2017)

[7] Himani Bhavsar, Mahesh H. Panchal, A Review on Support Vector Machine for Data Classification, ISSN: 2278 – 1323 International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 1, Issue 10, December 2012

[8] Hyun-Chul Kim, Shaoning Pang, Hong-Mo Je, Daijin Kim*, Sung Yang Bang, Constructing support vector machine ensemble, Pattern Recognition 36 (2003) 2757 – 2767

[9] Ashanira Mat Deris, Azlan Mohd Zain, Roselina Sallehuddin, Overview of Support Vector Machine in Modeling Machining Performances, procedia Engineering 24(2011) 308-312

[10] Xiaonan Zou, Yong Hu, Zhewen Tian, Kaiyuan Shen, Logistic Regression Model Optimization and Case Analysis, 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT)

[11] Abdallah Bashir Musa, Logistic Regression Classification for Uncertain Data, Research Journal of Mathematical and Statistical Sciences ISSN 2320 Vol. 1-6, February 2 (2), (2014)

[12] Zan Yang, Dan Li, Application of Logistic regression with Filter in Data Classification, Proceedings of the 38th Chinese Control Conference July 27-30,2019, Guangzhou, China

[13] Suresh, N., et al. "Crop yield prediction using random forest algorithm." 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), pp. 279-282, 2021, doi: 10.1109/ ICACCS51430.2021.9441871

[14] Bahzad Taha Jijo, Adnan Mohsin Abdulazeez, Classification Based on Decision Tree Algorithm for Machine Learning, January 2021 · Journal of Applied Science and Technology Trends 2(1):20-28

[15] Ajay Kumar Bhardwaj 1, Geeta Arya, Raj Kumar, Lamy Hamed, Hadi Pirasteh-Anosheh, Poonam Jasrotia, Prem Lal Kashyap and Gyanendra Pratap Singh Switching to nanonutrients for sustaining agroecosystems and environment: the challenges and benefits in moving up from ionic to particle feeding Bhardwaj et al. Journal of Nanobiotechnology (2022) 20:19

[16] Abhishek kumar, Crop Growth Recommendations: Optimal Conditions for Higher Yields,kaggle.

IFERP Explore Your Research Journey...