Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 13, Issue 1, January 2022: 2097-2105

Comparative Analysis of ML Algorithms and Smart UI for Optimum Crop Yield Prediction

Prof O.V.P.R Siva Kumar[a], Kudikala Sri Vaishnavi[b], Md. Adnan[c], Sai Ganesh P[d]

[a]Professor, Department of Electronics and Communication Engineering, Geethanjali College of Engineering and Technology,Hyderabad
[b]Student, Department of Electronics and Communication Engineering, Geetanjali College of Engineering and Technology
[c]Student, Department of Electronics and Communication Engineering, Geethanjali College of Engineering and Technology
[d]Student, Department of Electronics and Communication Engineering, Geethanjali College of Engineering and Technology

Abstract

A key viewpoint for finding practical and effective solutions to crop yield problems is Machine Learning (ML). Using supervised learning, machine learning (ML) may predict a result from a set of predictors. We build an appropriate function from a set of variables that will map the input variable to the desired output in order to get the desired outcomes. Crop yield prediction comprises predicting crop production from historical data that includes elements related to various crops, such as temperature, humidity, pH, and rainfall. It provides us with a general concept of the best-forecasted crop that will be grown in the field under various weather circumstances. If crop productivity is projected based on numerous qualities by utilizing different datasets, the farmer community can gain significantly. The machine learning techniques Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression are used to make these predictions (LR). Which strategy achieves the highest crop forecast value is examined. The best crop yield model prediction is determined to be provided by the RF algorithm when the accuracy scores of the three applied methods are compared. To make it simpler to enter attribute values and obtain the corresponding prediction, we further built a user interface (UI).

Keywords: support vector machines, logistic regression, user interface, random forest, and prediction

1. Introduction

To survive the changing conditions of the Indian economy, the agriculture sector requires significant improvement. Along with the latest advancements in farming machines and technologies, useful information about various factors pertaining to better crop yield prediction is also important. This information is being gathered by the use of remote sensors, satellite images, surveys, etc. Farmers should have easy access to this information, as well as the knowledge of subject experts and researchers, in order to capitalize on its potential value. This can be achieved by implementing Machine Learning techniques. Machine learning techniques that are simple to implement in agriculture can help us improve the agricultural sector as a whole. Along with all of the developments in farming gear and technologies, it's also vital to have access to useful and credible knowledge on a range of issues.

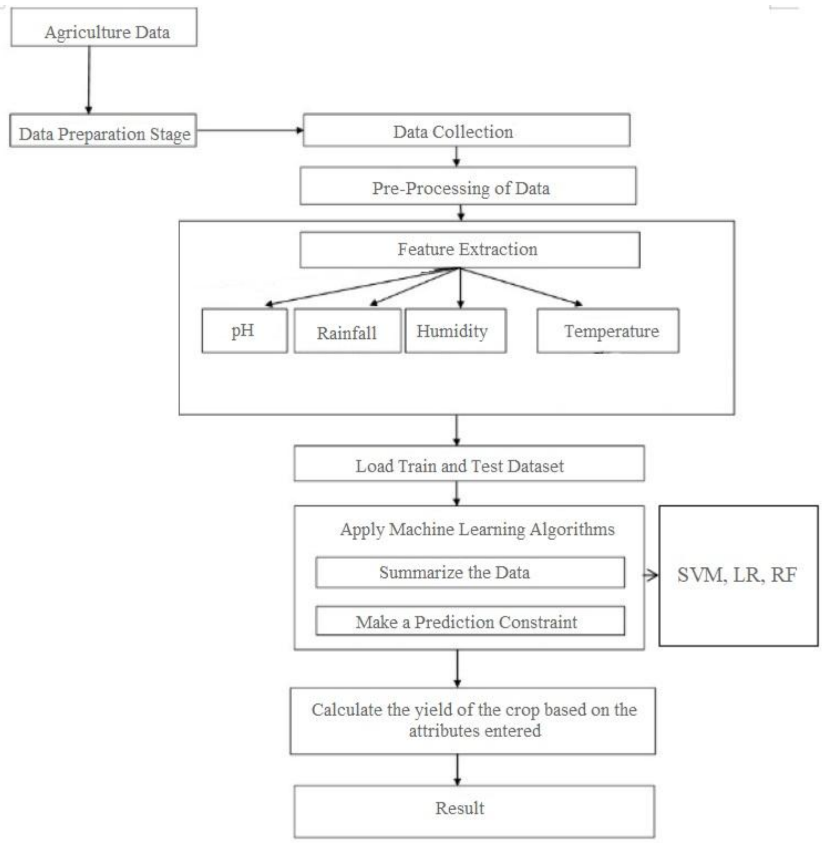
The goal of this study is to put the crop selection technique into practice in order to help farmers solve problems. For example, a poor decision by the farmer could put his family's financial situation under even more strain, ultimately resulting in serious loss. As a result, we understand the stress that a farmer experiences when deciding which crop to plant on his land. The most pressing challenge at the moment is to create a recommendation system that can anticipate the sort of crop that can be grown on a given plot of land, as well as how much of it can be grown. With this goal in mind, we decided to create a system that analyses soil factors such as **N**, **P**, and **K** (Nitrogen, Phosphorus, and Potassium), as well as **pH** levels and predicts the best crop for a certain region.

2.Methodology

A machine learning system is one that can learn from previous experiences(examples) and improve itself without being programmed explicitly. The breakthrough is the concept that a machine can learn from data alone and produce accurate results. By combining data using statistical tools, machine learning predicts outcomes. The output is subsequently used by corporate to provide actionable insights. Data mining and Bayesian predictive modeling are both closely related to machine learning. The machine takes input data and uses an algorithm to generate responses.

In this paper, we trained the dataset with three machine algorithms which are- **Random Forest**, **Logistic Regression**, and **Support Vector Machine** to find out which algorithm achieves the best accuracy and we also proposed a crop production yield predictor to recommend a suitable crop according to the given set of inputs. In this project, we use a regression model. The output obtained while predicting the yield is a continuous system. Whereas to give the recommendation of the crop the obtained output must be the name of the crop. As a result, a classification model is used for crop recommendations. We train and test the dataset with a Random Forest algorithm with which maximum accuracy is achieved.

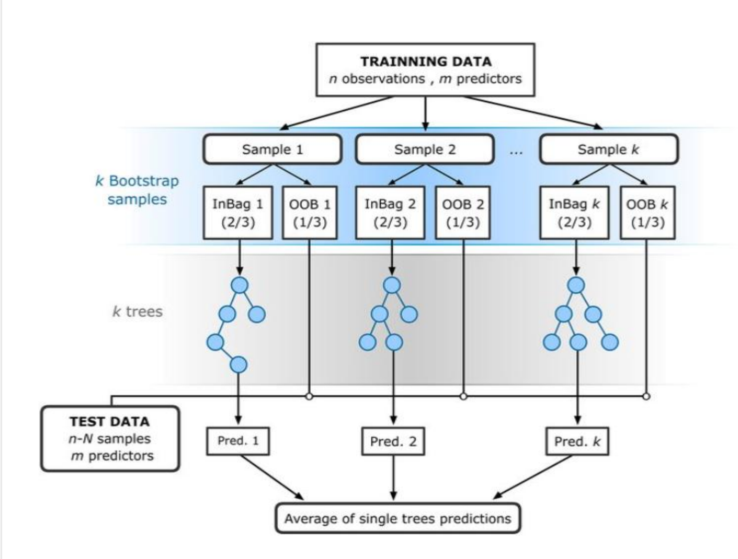
Figure.1 System level Flow Chart



2.1 Random Forest

Each decision tree has a significant variation, but when we mix them all simultaneously, the variance is reduced because each decision tree was perfectly trained using that particular sample of data. As a result, the outcome is dependent on numerous decision trees rather than just one. The majority vote classifier is used to decide the outcome of a classification challenge. The mean of all the results serves as the final result of a regression problem. Aggregation is the name of this section.

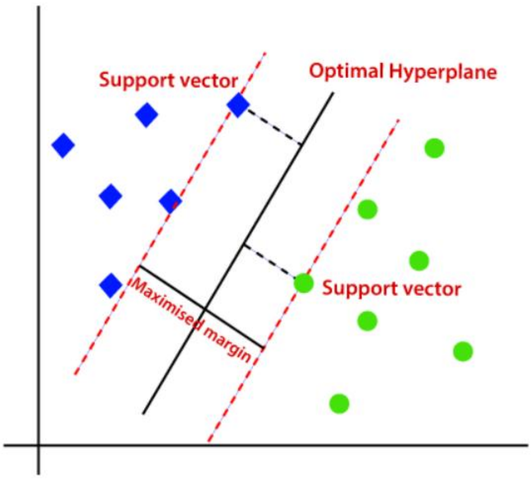
Figure.2 Random Forest Flow Chart [10]



2.2 Support Vector Machine

A supervised machine learning method called SVM can be used to categorize and forecast data. Even though we refer to them as regression issues, classification is the most appropriate solution. The N-dimensional hyperplane that best classifies the input points is what the SVM algorithm looks for. The hyperplane's size is determined by the feature quantity. The hyperplane is just a line if there are just two characteristics in the input. The hyperplane becomes a two-dimensional plane if there are three input features. When there are more than three features, it is difficult to imagine.The Method is such, we can identify a linear decision surface (“hyperplane”) that can divide classes and has the maximum distance (i.e., largest “gap” or “margin”) between border-line elements (i.e., “support vectors”)

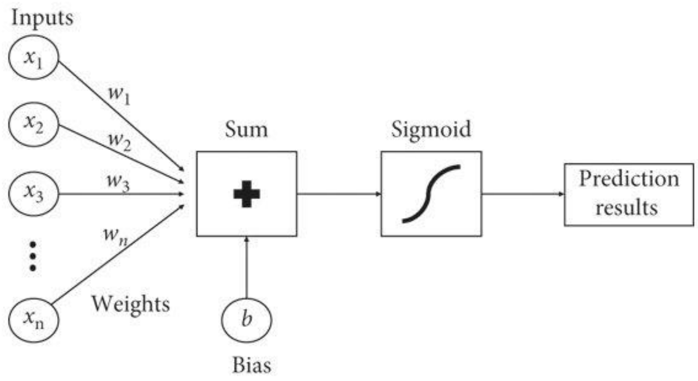
Figure.3 Representation of SVM Algorithm



2.3 Logistic Regression

Only when a decision threshold is introduced into the equation does logistic regression become a classification approach. A key component of logistic regression is the threshold value, which is established by the classification issue itself. The threshold value decision is heavily influenced by the precision and recall levels. Ideally, both precision and recall should be 1, but this is rarely the case.

Figure.4 Flow Chart of LR Algorithm [11]

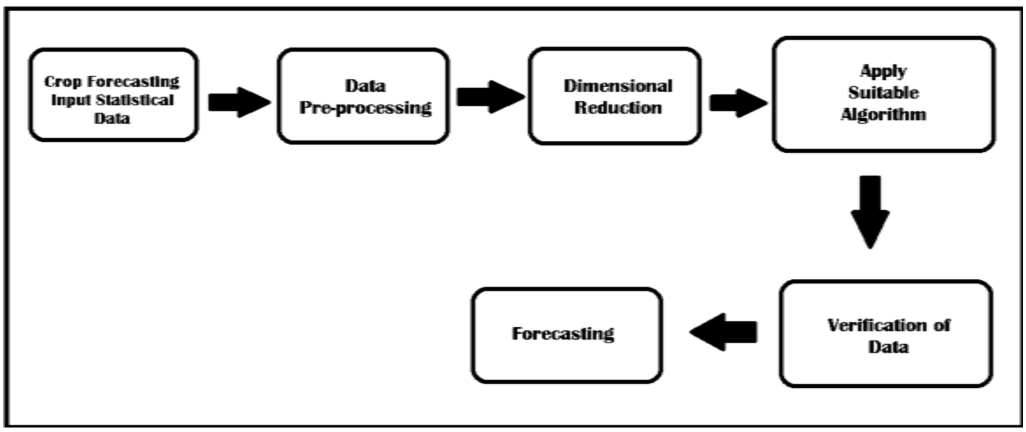


3.Problems Faced in Existing Models

The performance of the existing **K-NN models**, which were utilized for classification and yield prediction, was diminished by nonlinear and highly adaptable K-NN difficulties. They were operated in a locality model that increased the input vector's dimensionality and caused classification confusion. To overcome those issues, we used ML algorithms, which are more efficient in dealing with large data and are more accurate.

4.Proposed System

Figure.4 Flow of Implementation



This scenario is mostly concerned with weather forecasting and crop prediction. These elements assist farmers in cultivating the best food crops and raising the appropriate animals in accordance with environmental components. Farmers can also adapt to climate change by moving planting dates, selecting cultivars with various growth durations, or changing crop rotations. Agricultural statistics numeric data is collected for experimental analysis. Clustering-based techniques and supervised algorithms are utilized to organize the acquired statistical data. Furthermore, relevant classification approaches such as Support Vector Machine (SVM) and neural networks are used to improve outcome categorization.

4.1 Dataset and Pre-Processing

We required a collection containing both falsified and genuine profiles. Humidity, temperature, pH, and rainfall are among the attributes covered in the dataset. The dataset is separated into two parts: training data and testing data. The training dataset is used to train classification algorithms, and the testing dataset is used to determine the algorithm's efficiency. The dataset is used to generate a training dataset that contains 80% of both profiles (authentic and falsified) and a testing dataset that contains 20% of both profiles.

4.2 Feature Selection

Algorithms for classification are used to choose features. There is more discussion of the categorization algorithm. If an attribute does not depend on another attribute and improves the effectiveness of categorization, it is chosen as a feature.After choosing the attributes, the dataset of profiles that have already been identified as false or real is required for the classification algorithm's training. We used a dataset of 3,100 inputs with 3,100 publicly accessible variables, including humidity, temperature, pH, rainfall, and crop name.

4.3 Classification

Classification is the process of categorizing data objects according to the qualities or attributes they share. A classifier is an algorithm that examines the qualities of each data object and assigns it a class. We use Support Vector Machine, Random Forest, and Logistics regression in this research. Support Vector Machine is an elegant and durable technique for classification on huge datasets, similar to social network datasets with millions of profiles.

5. User Interface (UI)

We have implemented a simple web UI using HTML to make it easier for the farmers to enter in the respective values of their farmland to predict which crop grows the best in those conditions based on historical data available to the system through datasets (offline or online). The range of the data can be set through the datasets made available. The UI in the future can be made available to be interacted with in multiple languages to allow farmers to access the same data in the language they feel comfortable in.

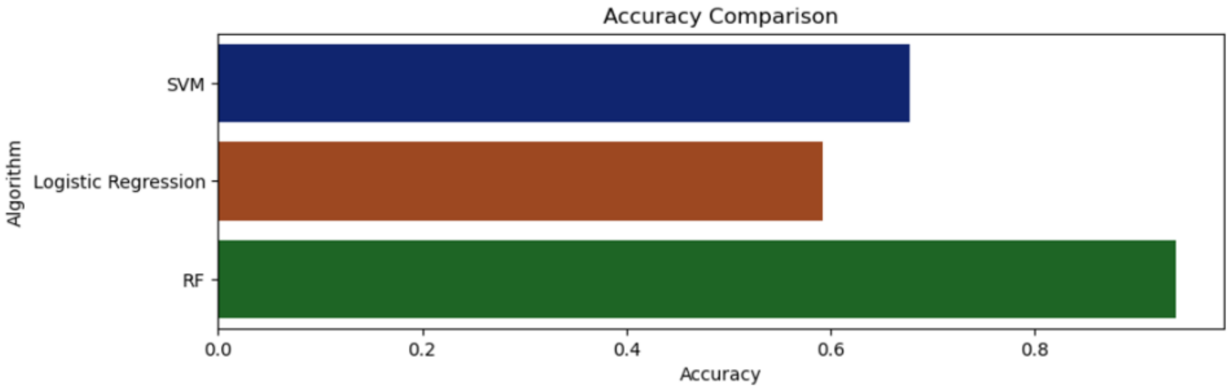
6. Results

We have successfully compared the three ML algorithms and found out the most accurate predicting algorithm to be RF algorithm, also an UI has been successfully implemented to take input values and give predicted crop based on its training and historical data as its output.

Table.1 Comparison of performance measures of the three algorithms

AlgorithmUsed	Accuracy	Precision	Recall	f1-Score
SupportVector Machine	67.74%	85%	68%	71%
LogisticRe gression	59.19%	59%	61%	58%
RandomForest	93.87%	94%	94%	94%

Figure.5 Accuracy Comparison of the Algorithms



7.Conclusion

The current study showed how ML techniques may be utilized to estimate agricultural productivity using meteorological input features. Due to poor crop selection and inadequate use of technology and analysis, our farmers run the risk of having to pay for cultivation, which will lower their profits. We developed a farmer-friendly technique to lessen the frequency of such failures. A graphical user interface (GUI) that forecasts the ideal crop for a specific circumstance for a plot of land, together with details on the nutrients that must be provided, the seeds that must be utilized, the anticipated yield, and the market price for cultivation. To decide which crops are the most productive under certain conditions, the proposed method considers soil N, P, K, and pH levels.

By providing a list of all accessible crops, the system assists farmers in choosing which crop to cultivate in their region. The findings demonstrate that by applying the Random Forest algorithm, we can anticipate crop yields with the greatest degree of accuracy. With the smallest model, the Random Forest algorithm produces the most crop yield models. It is suitable for forecasting large crop yields in agricultural planning. As a result, this method helps farmers find novel crops that haven't been grown before as well as the most lucrative ones. The Internet of Things (IoT) might be used to collect real-time soil values, which would enhance this method in the future. On the farm, sensors can be installed to gather information on the soil's current state.

References (APA)

1. P. Priya, U. Muthaiah, M. Balamurugan. "*Predicting yield of the crop using machine learning algorithm*." International Journal of Engineering Science Research Technology.
2. J.Jeong, J.Resop, N.Mueller and team." *Random forests for global and regional crop yield prediction*." PLoS ONE Journal.
3. Narayanan Balkrishnan and Dr. GovindarajanMuthukumarasamy. *Crop production Ensemble Machine Learning model for prediction.* International Journal of Computer Science and Software Engineering (IJCSSE).
4. S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh. "*Machine learning approach for forecasting crop yield based on climatic parameters*". International Conference on Computer Communication and Informatics (ICCCI).
5. Alif, I. F. Shukanya, and T. N. Afee, "*Crop prediction based on geographical and climatic data using machine learning and deep learning*", Doctoral dissertation, BRAC University) 2018
6. Ecochem Online. (2009). "*Soil Health and Crop yields*." Last modified January 28th 2009.Retrieved on March 4th 2009 from http://ecochem.com/healthy_soil.html

7. M. Alagurajan, and C. Vijayakumaran, "*ML Methods for Crop Yield Prediction and Estimation: An Exploration*" International Journal of Engineering and Advanced Technology, vol. 9 no. 3, 2020
8. Suresh, N. Manjunathan, P. Rajesh, and E. Thangadurai, "*Crop Yield Prediction Using Linear Support Vector Machine*," European Journal of Molecular & Clinical Medicine, vol. 7, no. 6, pp. 2189-2195, 2020.
9. D. Ramesh and B. Vardhan, "*Analysis of crop yield prediction using data mining techniques*", International Journal of Research in Engineering and Technology, vol. 4, no. 1, pp. 47-473, 2015
10. J. Reddy and M. R. Kumar, "*Crop Yield Prediction using Machine Learning Algorithm*," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 1466-1470, doi: 10.1109/ICICCS51141.2021.9432236.
11. Picture credit to Victor F. Rodriguez-Galiano
12. V. Pandith, H. Kour, S. Singh, J. Manhas, and V. Sharma, "*Performance Evaluation of Machine Learning Techniques for Mustard Crop Yield Prediction from Soil Analysis,*" Journal of Scientific Research of Banaras Hindu University, vol. 64, no. 2, 2020
13. Rao, Pooja & Gowda, Sanju& R J, Prathibha. (2019). "*Paddy Yield Predictor Using Temperature, Rainfall, Soil pH, and Nitrogen*", 10.1007/978-981-13-5802-9_23.
14. AnakhaVenugopal, Aparna S, Jinsu Mani, Rima Mathew, Vinu Williams, 2021, "*Crop Yield Prediction using Machine Learning Algorithms*", INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NCREIS – 2021 (Volume 09 – Issue 13),
15. Klompenburg, Thomas &Kassahun, Ayalew&Catal, Cagatay. (2020). "*Crop yield prediction using machine learning: A systematic literature review*". Computers and Electronics in Agriculture. 177. 105709. 10.1016/j.compag.2020.105709.
16. Gopal, P.S. &Bhargavi, R. (2019). "*A novel approach for efficient crop yield prediction*". Computers and Electronics in Agriculture. 165. 104968. 10.1016/j.compag.2019.104968.
17. Pandith, Vaishali&Kour, Haneet& Singh, Surjeet&Manhas, Dr& Sharma, Vinod. (2020). "*Performance Evaluation of Machine Learning Techniques for Mustard Crop Yield Prediction from Soil Analysis*". Journal of scientific research. 64. 10.37398/JSR.2020.640254.
18. R. Ghadge, J. Kulkarni, P. More, S. Nene, and R. L. Priya, "*Prediction of crop yield using machine learning,*" Int. Res. J. Eng. Technology, vol. 5, 2018.