PERFORMANCE COLLATION OF CROP YIELD PREDICTION USING SUPERVISED MACHINE LEARNING ALGORITHMS

[1]Raksha Jayaram, [2]Prof. Deepika K

[1]Student, [2]Assistant Professor
[1]Department of Master of Computer Applications,
[1]RV College of Engineering®, Bengaluru, India.

Abstract: Food is a basic need for human life. Every human deserves at least a meal per day for his survival. Agriculture plays a major role in feeding the nation with adequate food. Agriculture in India is a component of the primary sector and is considered as the backbone of Indian economy. India is one of the thickly populated countries with more than half of its population engaged in agriculture. Despite agriculture being the prominent occupation in India, the outcome is disappointing. The farmers strive to get the estimated crop yield but the age-old methodologies used are a hindrance. Accurate crop yield prediction will bring about a drastic change in the lives of the farmers. Thereby contributing to the economic growth of the nation. The proposed system uses supervised machine learning algorithms to predict the crop yield accurately. This paper uses Decision Tree, k-Nearest Neighbor and Artificial Neural network algorithms to predict the crop yield and compare their performance.

Index Terms **- Artificial Neural Networks, crop yield, Decision Tree, k-Nearest Neighbor, prediction, supervised machine learning.**

I. INTRODUCTION

India is having a substantial cultivable land with more than half of the nation toiling as farmers. Although agriculture is considered as the backbone of Indian economy, food insecurity is one of the prevailing concerns in India [1][2]. Farmers are not receiving the expected crop yield. Although framers are engaged in agricultural activities throughout the year, the yield obtained is inadequate. The main reason for unanticipated crop yield is the inaccurate prediction of the crop yield. A farmer predicts the yield based on his past experience with the crop [3]. The experience of a farmer with the crop can be considered meagre in contrast with agricultural data collected over years throughout the nation. The traditional methodology adopted by the farmer to predict the crop yield results in inaccurate prediction. As a result, farmers incur loss and fail to produce adequate crop yield to fulfill the basic needs of the country.

Agricultural produce significantly depends on the climatic factors such as rain, soil, seed, temperature, etc. The climatic factors and the productivity depending on these factors should be studied in order to predict the crop yield accurately [4]. The crop data collected over years should be analyzed [5]. Since India is a developing nation a pricey solution will not help in resolving the problems of Indian agriculture. The use of supervised machine learning algorithms to predict the crop yield will be a simple and effective solution.

The supervised machine learning algorithms help in predicting the crop yield accurately [6]. The algorithms used for the study are Decision Tree, k-Nearest Neighbor and Artificial Neural Networks. The algorithms differ in terms of accuracy as the model development process differs. The proposed system emphasizes on the comparative study of the algorithms in order to consider the algorithm which predicts the crop yield most accurately.

II. RELATED WORK

The strategies which help in solving the farming problems are discussed. It emphasizes the crop selection methodology. The use of machine learning techniques in agriculture improves the crop yield rates. Different methods of machine learning techniques are examined to obtain the performance accuracy [7].

Physical parameters such as rainfall and temperature are considered for the study. It proposes a solution to crop yield prediction with the development of a tool which gives predictions based on the individual crops. An android application is developed to query the results of the machine learning analysis [8].

A solution is proposed using data analytics for crop yield prediction. Table, data visualization tool is used to combine two data sets with the parameter year. K-Means clustering algorithm is used to store big data in clusters. Apriori algorithm is used to count the features. Naïve Bayes algorithm is used to find the exact crop yield [9].

Farming practices across different growing locations and conditions are considered. The accuracy of the performance estimation model in different feature sets are analyzed. The machine learning model is trained with district level data. This model is applied to decompose district level performance data to taluk level performance. The mean error for taluk level is 6% and maximum error is 25% [10].

The paper proposes a method called crop selection problem. The proposed method addresses crop selection affected by the factors such as soil type, water density, etc. It uses statistical and machine learning techniques to maximize the net yield rate of the crops [11].

III. PROPOSED METHODOLOGY

The proposed system uses Decision Tree, k-Nearest Neighbor and Artificial Neural Network supervised machine algorithms for the study. The featured dataset containing the crop data is divided into training data and test data. The training data is given as the input to the algorithm. The algorithm develops a model learning from the training data. The model developed is tested using the test data. The score generated during the testing of the model is the performance measure of the algorithm. The model is considered for predicting the crop yield when a considerable score is attained. The system compares the performances of the algorithms. It suggests the user as to which model performs better aiding the user to choose the most apt algorithm for predicting the crop yield.

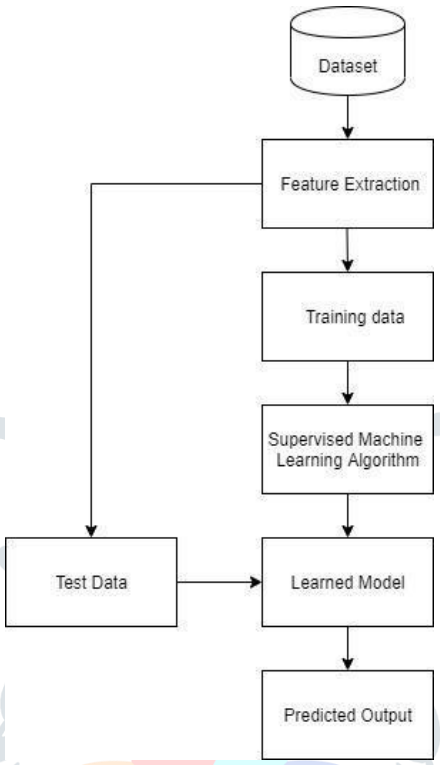

Fig. 3.1. Block diagram of the proposed system

The crop data is collected from various sources. The collected data is pre-processed and feature extraction is done. The features considered are state, year, crop, area, rainfall and production. The features are standardized as per the requirement of the algorithm. The standardization is applied on required features individually. The average of the selected feature is calculated. Standardization is carried using the equation 3.1. The dataset is divided into training data and test data. In the proposed system, 20% of the dataset is considered as the test data for testing the model developed.

$$z = \frac{x_i - \mu}{\sigma} \quad (3.1)$$

Equation 3.1. Standardization formula used for feature extraction

Where,

z = standardized value

x_i = observation for which z is calculated

μ = mean

$\boldsymbol{\sigma}$ = standard deviation

The training data is given as the input to the decision tree algorithm. The algorithm trains the model using the training set. The process starts from the root node and ends at the leaf node. The Mean Squared Error is used as the splitting criteria for nodes. Equation 3.2 is used to calculate the Mean Squared Error. The model is then tested for correctness. When the decision tree learned model generates a considerable score, it is used to predict the crop yield for the given crop data.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2 \quad (3.2)$$

Equation 3.2. Mean Squared Error calculation

Where,

MSE = Mean Squared Error

N = Number of predictions generated from the data points

Y_i = vector of observed values for the variables being predicted

$\tilde{Y}_i$ = predicted values

k-Nearest Neighbor takes the training data as the input. The algorithm considers the similarity measure for developing the model. It calculates the average of the numerical target of the K nearest neighbors. The test data is used to test the model developed. The user inputs the crop data in order to receive the predicted crop yield.

$$\text{Euclidian distance} = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$$

$$\text{Manhattan distance} = \sum_{i=1}^{k}|x_i - y_i|$$

$$\text{Minkowski distance} = \left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{1/q} \quad (3.3)$$

Equation 3.3. Distance metrics

Where,

k = number of dimensions

x_i, y_i = data points

q = order of the norm

The artificial neural network algorithm is given with the training data. The training data is used to train the model. The algorithm considers 5 dimensions in the input layer. The activation function is calculated and the hidden layer processes it. The processing continues until the single output value is obtained. The crop yield is predicted for the input crop data.

IV. CHALLENGES

The crop data is collected from different sources. The data considered should be genuine to yield appropriate prediction. The score generated from the models differ based on the dataset given for training the model. Achieving a considerable score is a challenging task.

V. RESULTS

The results obtained from the implementation of Decision Tree, k-Nearest Neighbor and Artificial Neural Network supervised machine learning algorithms is illustrated in this section.

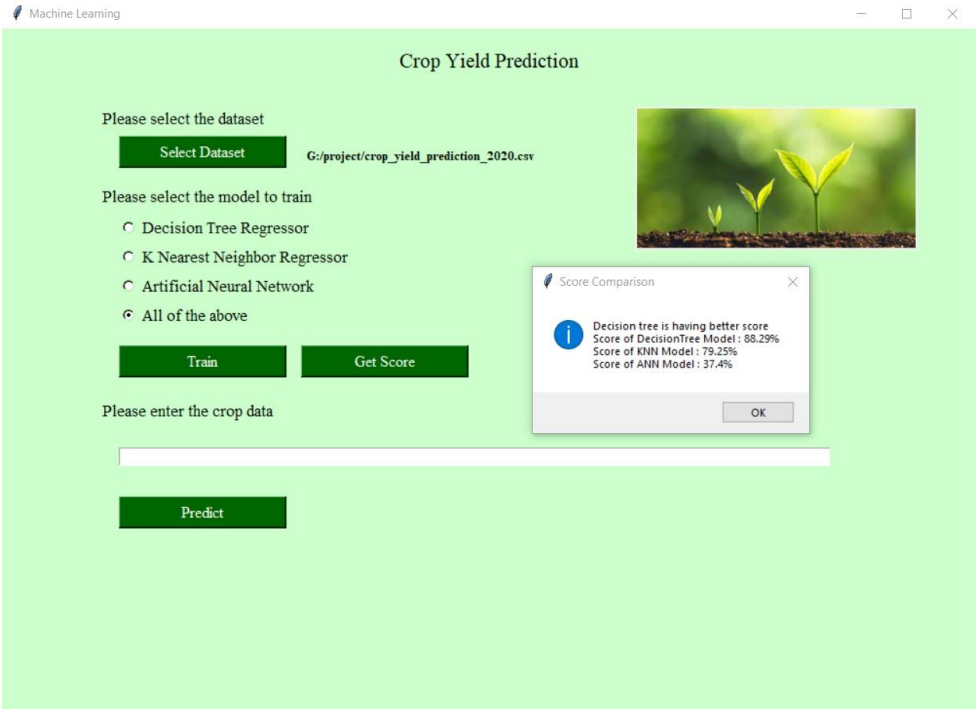

Fig. 5.1 Comparison of the supervised machine learning models

The models are trained with the provided training data. The scores generated from the models are compared. The user is suggested to with the model which provides better accuracy.

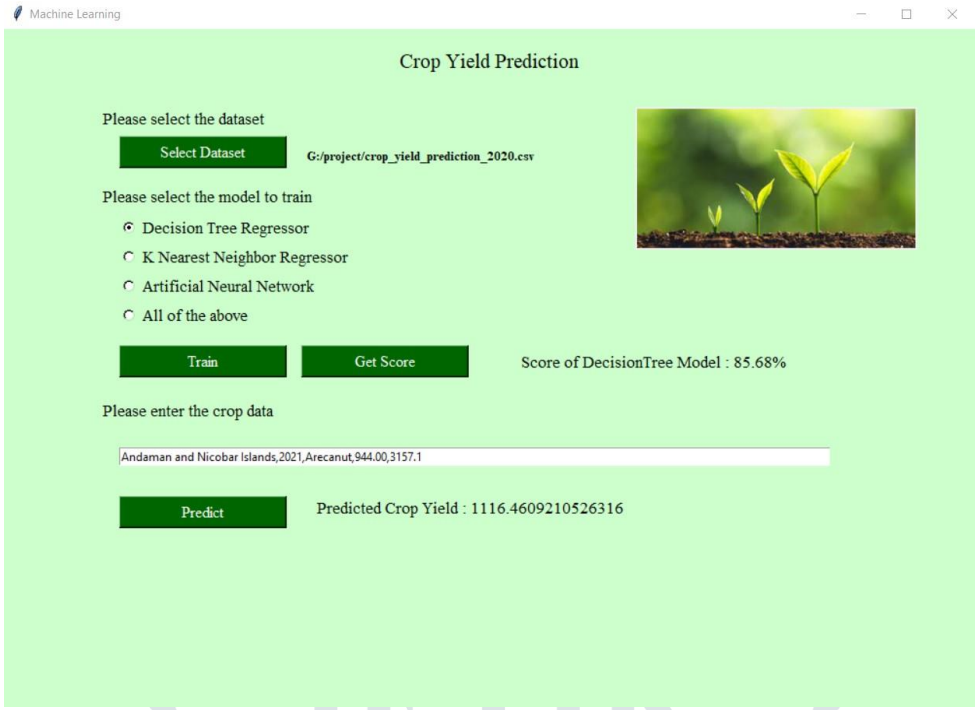

Fig. 5.2. Crop yield prediction using Decision Tree model

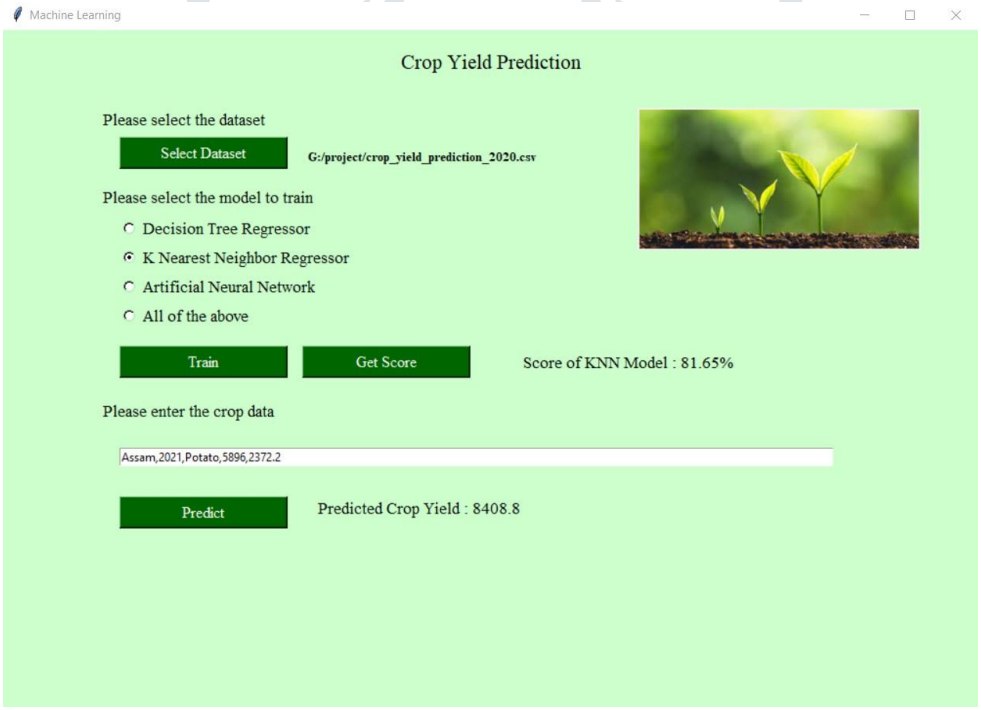

Fig. 5.3. Crop yield prediction using k-Nearest Neighbor model

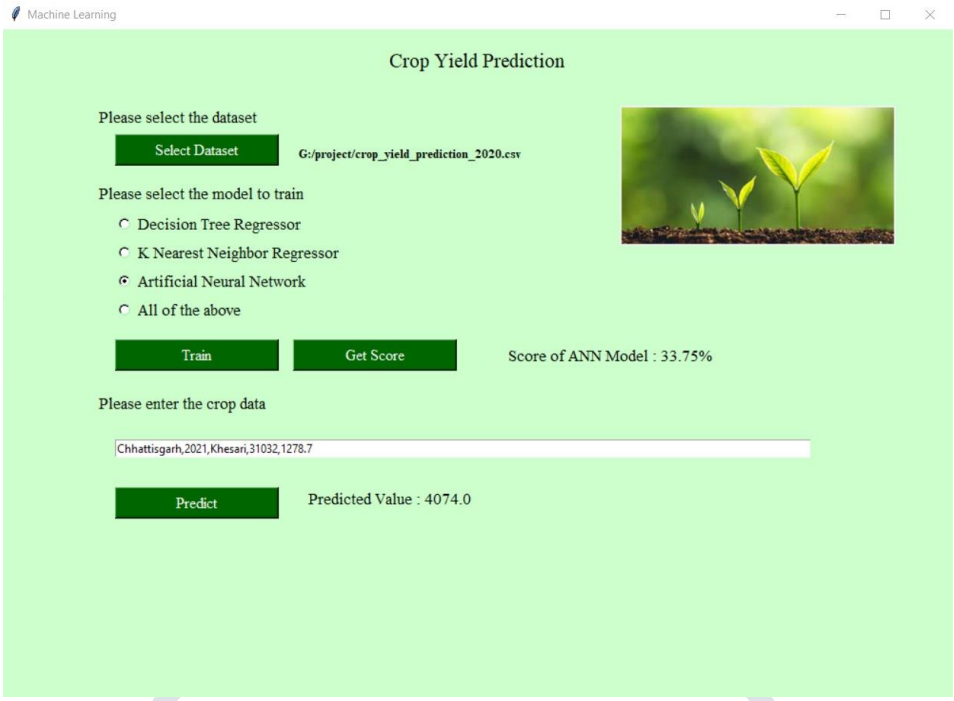

Fig. 5.4. Crop yield prediction using Artificial Neural Network

The user selects the Comma Separated Values file containing the crop data as the input dataset. The algorithm of the user's choice is selected. Clicking the train button trains the model and gives a message to the user stating the training is completed. After the successful completion of the training, the score is generated when the user clicks the get score button. The user enters the crop data for which the yield prediction has to be made and clicks on the predict button. The crop yield predicted value is displayed.

VI. CONCLUSION

The use of the proposed system will aid the farmers in predicting the crop yield accurately. The user of the system can carry out a comparative study in order to find which algorithm gives better accuracy. The algorithm which gives a better score should be considered to predict the crop yield. The predicted crop yield will help the farmers in planning the cost and resources required. The use of the proposed system gives a simple solution with significant outcomes. The system will help India become a food secure country. The adequate produce will contribute to economic growth of the country.

REFERENCES

[1] J. N. Jere and P. Ndayizigamiye, "Analysing the Relationship between the Use of ICTs and Food Insecurity in Smallholder Farmer Households," 2018 IST-Africa Week Conference (IST-Africa), Gaborone, 2018, pp. 1-62.

[2] A. S. Vaidya, G. Makkena, V. Srihari, M. B. Srinivas and S. K. Rao, "A sustainable solution for monitoring malnutrition in children in developing countries," 2013 IEEE Global Humanitarian Technology Conference: South Asia Satellite (GHTC-SAS), Trivandrum, 2013, pp. 170-174.

[3] M. T. Shakoor, K. Rahman, S. N. Rayta and A. Chakrabarty, "Agricultural production output prediction using Supervised Machine Learning techniques," 2017 1st International Conference on Next Generation Computing Applications (NextComp), Mauritius, 2017, pp. 182-187.

[4] S. Veenadhari, B. Misra and C. Singh, "Machine learning approach for forecasting crop yield based on climatic parameters," 2014 International Conference on Computer Communication and Informatics, Coimbatore, 2014, pp. 1-5.

[5] M. Paul, S. K. Vishwakarma and A. Verma, "Analysis of Soil Behavior and Prediction of Crop Yield Using Data Mining Approach," 2015 International Conference on Computational Intelligence and Communication Networks (CICN), Jabalpur, 2015, pp. 766-771.

[6] A. K. Mariappan and J. A. Ben Das, "A paradigm for rice yield prediction in Tamilnadu," 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR), Chennai, vol. 2, 2017, pp. 18-21.

[7] R. Medar, V. S. Rajpurohit and S. Shweta, "Crop Yield Prediction using Machine Learning Techniques," 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), Bombay, India, 2019, pp. 1-5.

[8] Meeradevi and H. Salpekar, "Design and Implementation of Mobile Application for Crop Yield Prediction using Machine Learning," 2019 Global Conference for Advancement in Technology (GCAT), Bengaluru, India, 2019, pp. 1-6.

[9] S. V. Bhosale, R. A. Thombare, P. G. Dhemey and A. N. Chaudhari, "Crop Yield Prediction Using Data Analytics and Hybrid Approach," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-5.

[10] R. B. Guruprasad, K. Saurav and S. Randhawa, "Machine Learning Methodologies for Paddy Yield Estimation in India: a Case Study," IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 2019, pp. 7254-7257.

[11] R. Kumar, M. P. Singh, P. Kumar and J. P. Singh, "Crop Selection Method to maximize crop yield rate using machine learning technique," 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), Chennai, 2015, pp. 138-145.

[12] R. L. F. Cunha, B. Silva and M. A. S. Netto, "A Scalable Machine Learning System for Pre-Season Agriculture Yield Forecast," 2018 IEEE 14th International Conference on e-Science (e-Science), Amsterdam, 2018, pp. 423-430.

[13] T. Siddique, D. Barua, Z. Ferdous and A. Chakrabarty, "Automated farming prediction," 2017 Intelligent Systems Conference (IntelliSys), London, 2017, pp. 757-763.

[14] V. Mehta, C. Jain, K. Kanchan and V. Sawant, "A Machine Learning Approach to Foretell the Probability of a Crop Contracting a Disease," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-7.

[15] S. Bhanumathi, M. Vineeth and N. Rohit, "Crop Yield Prediction and Efficient use of Fertilizers," 2019 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2019, pp. 0769-0773.

[16] J. G. N. Zannou and V. R. Houndji, "Sorghum Yield Prediction using Machine Learning," 2019 3rd International Conference on Bio-engineering for Smart Technologies (BioSMART), Paris, France, 2019, pp. 1-4.

[17] S. Kulkarni, S. N. Mandal, G. S. Sharma, M. R. Mundada and Meeradevi, "Predictive Analysis to Improve Crop Yield using a Neural Network Model," 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, 2018, pp. 74-79.

[18] U. Inyaem, "Construction Model Using Machine Learning Techniques for the Prediction of Rice Produce for Farmers," 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC), Chongqing, 2018, pp. 870-874.

[19] A. Suresh, P. Ganesh Kumar and M. Ramalatha, "Prediction of major crop yields of Tamilnadu using K-means and Modified KNN," 2018 3rd International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2018, pp. 88-93.

[20] R. P. Paswan and S. A. Begum, "ANN for prediction of Area and Production of Maize crop for Upper Brahmaputra Valley Zone of Assam," 2014 IEEE International Advance Computing Conference (IACC), Gurgaon, 2014, pp. 1286-1295.