PREDICTION OF CROP USING ENSEMBLING APPROACHES

Dr. A. Anna Lakshmi, R. Aruna, S. Banu Priya, S. Daicy Rani, E. Mahalakshmi

Dr. A. Anna Lakshmi, HoD / CSE Department, Thamirabharani Engineering College,Tirunelveli, Tamil Nadu, India, annalaxmi.raj@gmail.com

Final Year Students, Thamirabharani Engineering College, Tirunelveli, esakkimuthu2418@gmail.com

Abstract:

Agriculture has a significant part in the country's economic development. Climate warming and other environmental changes have posed a serious danger to agriculture. Crop yield prediction remains a challenging task in this domain. There are various parameters that affect the yield of crop like rainfall, temperature, fertilizers, pesticides, pH level, and other atmospheric conditions and parameters. Machine Learning (ML) is a critical component in finding practical and successful solutions to this issue. Crop yield production prediction entails predicting crop yield based on historical data such as weather parameters, soil parameters, and previous crop yield. The purpose of this work is to use the Random Forest method to estimate agricultural yield based on existing data. The models were built using real data from Tamil Nadu, and the models were evaluated using samples. The farmer would be able to forecast the crop output before cultivating the agriculture land with the aid of the prediction. Random Forest, a sophisticated and popular supervised machine learning method, is used to reliably estimate agricultural productivity in the proposed work. In existing system, Naïve Bayes and Logistic Regression Method are used.

Index Terms— Machine Learning (ML), Random Forest (RF), Artificial Neural Network (ANN), crop yield prediction*,* Naïve Bayes, Logistic Regression Method

1. INTRODUCTION

Agriculture is the main occupation for the people of India, covering 60% of the nation land and catering the basic needs of 1.2 billion people [1]. Ancient people cultivate the crops in their own land and so they have been accommodated to their needs. Therefore, the natural crops are cultivated and have been used by many creatures such as human beings, animals, and birds. Nowadays, modern people don't have awareness about the cultivation of crops at the right time and at the right place. There are multiple ways to increase and improve the crop yield and the quality of the crops. For the benefit of the farmers, modernization of agriculture procedures is carried out today. The crop yield or production majorly depends on the weather conditions, environmental changes, rainfall (which at times is uncertain), water management, and the utilization of pesticides. Therefore, farmers are not able accomplish expected yield of crop. Now a days data mining, machine learning as well as deep learning approaches are used by various researchers to enhance and improve the yield of crop and their quality [11,12]. Machine Learning can gain proficiency with the machine without characterized computer programming, so it improves machine execution by distinguishing and portraying the consistency and pattern of drive information. Data mining software is an analytical tool that allows users to analyze data from many different dimensions or angles, categorize, and summarize the relationships identified. Accurate information about the history of crop yield is an important thing for making decisions related to agricultural risk management.

2. LITERATURE SURVEY

1. Riddhi Dange, Jinisha Kande,Kunal Bhosale, Prof. Ankush Hutke, "Predicting Agricultural Produce using Machine Learning Techniques", IJIRT, Volume 8, Issue 12, pp: 329-334, May 2022.

   Description: In this paper, crop yield is predicted by obtaining the coordinates of the crop field on Google Maps API using GPS which is then uploaded in the system and used to calculate the area under cultivation using marked points on maps. Further, the crop name and area under cultivation will be used to predict the crop yield and this information is provided to farmers. Using the Random Forest Classifier, the prediction of the suitable crop for the suitable conditions will be achieved.

2. Nischitha K, Dhanush Vishwakarma, Mahendra N, Ashwini, Manjuraju M.R, "Crop Prediction using Machine Learning Approaches", International Journal of Engineering Research & Technology (IJERT), Vol. 9, Issue 08, pp: 23-26, August-2020.

   Description: The system takes the required input from the farmers such as Temperature, Humidity and pH. This all input data applies to Machine Learning predictive algorithms like Support Vector Machine (SVM) and Decision tree to identify the pattern among data and then process it as per input conditions. The system recommends the crop for the farmer and also recommends the amount of nutrients to be add for the predicted crop.

3. Mythresh A, Lavanya B, Meghana BS, Nisarga B, "Crop Prediction using Machine Learning", International Research Journal of Engineering and Technology (IRJET), Volume: 07, Issue: 06, pp: 1697-1700, June 2020.

   Description: The system takes input from various sources and repositories for weather, soil and crop requirement data and uses Naive-Bayes algorithm to predict the best suitable crop for any given area.

3. EXISTING SYSTEM

A Non-Dispersive Infrared (NDIR) methane gas sensor prototype has achieved a minimum detection limit of 1 parts per million by volume (ppm). In order to decrease the noise level, a single frequency filter algorithm based on Fast Fourier Transform (FFT) is adopted for signal processing. There is strong interest in a new generation of Earth observing satellites equipped with Visible and ShortWave InfraRed (VSWIR) imaging spectrometers. These imaging spectrometers are sensitive to gas absorption features, which allows for the detection and quantitative mapping of methane, carbon dioxide, and water vapor.

4. PROPOSED WORK

The proposed future imaging spectrometers are explicitly designed to map greenhouse gases, given their design and the prior successful work using similar airborne sensors to map methane, the upcoming generation of spaceborne imaging spectrometers has great potential for global mapping of near-surface methane emissions. This project examines the possibility of using these future sensors to monitor emissions from three of the largest anthropogenic methane sectors: oil/gas, waste management, and agriculture.

4.1 Random Forest Classifier:

Random Forest is a supervised learning algorithm. It creates a forest and makes it somehow random. It is a flexible, easy to use machine learning algorithm that produces, even without hyper-parameter tuning, a great result most of the time. It is also one of the most used algorithms, because its simplicity and the fact that it can be used for both classification and regression tasks which form the majority of current machine learning systems.
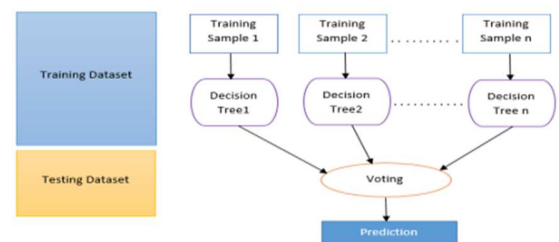

Figure 1. Random Forest Classifier

4.2 Logistic Regression:

Target variable can take more than two value and values should be ordered. The logistic regression is similar to binary logistic regression, except the label (class) is now an integer in {1, 2, … n} where n is number of classes. Scores for all classes are calculated. It is implemented using SoftMax function. Class with most votes is chosen for prediction. For the experimental study, Logistic regression on all parameters for multiclass labels has been implemented in Matlab using function mnrfit and mnrval returns the predicted probabilities for the multinomial logistic regression model with predictors X, and the coefficient estimates B.

4.3 Artificial Neural Network:

It usually involves a large number of processors operating in parallel and arranged in tiers. The first tier receives the raw input information analogous to optic nerves in human visual processing. Each successive tier receives the output from the tier preceding it, rather than the raw input in the same way neurons further from the optic nerve receive signals from those closer to it. The last tier produces the output of the system.

Input Layer: It accepts inputs in several different formats provided by the programmer.

Hidden Layer: It presents in-between input and output layers. It performs all the calculations to find hidden features and patterns.

Output Layer: The input goes through a series of transformations using the hidden layer, which finally results in output that is conveyed using this layer. The artificial neural network takes input and computes the weighted sum of the inputs and includes a bias. This computation is represented in the form of a transfer function.

$$\sum_{i=1}^{n} Wi * Xi + b$$

It determines weighted total is passed as an input to an activation function to produce the output. Activation functions choose whether a node should fire or not. Only those who are fired make it to the output layer. There are distinctive activation functions available that can be applied upon the sort of task we are performing.
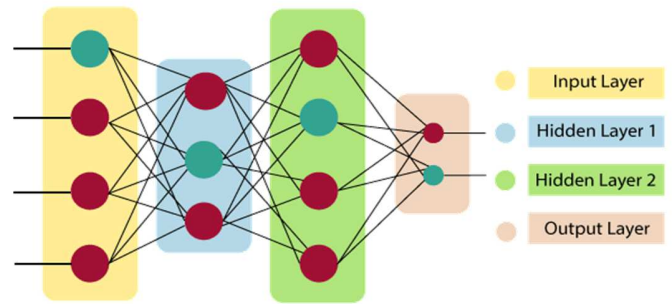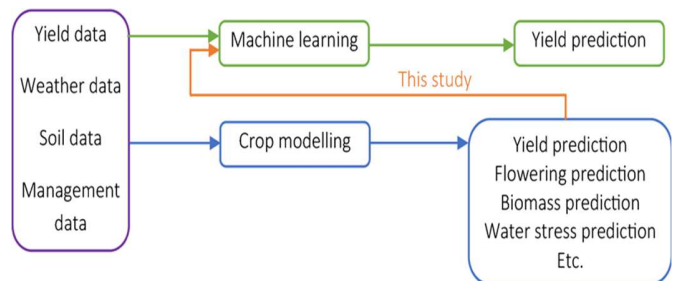

Figure 2. Artificial Neural Network (ANN) Layers

5. METHODOLOGY

The crop modelling parameters include rainfall, temperature, fertilizers, pesticides, pH level, and other atmospheric conditions.

6. LIST OF MODULES

6.1 Dataset Collection:

The data is collected from various sources that are used for crop yield prediction, such as weather data, soil data, and crop management data. Based on the information the ultimate goal would be to predict crop production using ensemble techniques.

Figure 3. Dataset (P, K, N, pH, Rainfall, etc.,)

The dataset contains pH, Nitrogen, Potassium, Calcium, Rainfall, etc., to detect the crop prediction.

6.2 Data preprocessing:

This component involves cleaning and transforming the raw data to ensure that it is accurate, complete, and consistent. This includes tasks such as removing missing values, scaling, and normalization. The final step on data preprocessing is the splitting of training and testing data.

6.3 Feature Selection:

This module involves selecting a set of base models that will be used in the ensemble. Common base models for crop yield prediction include linear regression, decision trees, and neural networks. Then, train each of the base models using the preprocessed data. Once trained the base models, then combine them to form the ensemble model.

6.4 Evaluate the ensemble model:

Evaluate the performance of the ensemble model using a test set of data. The performance of the ensemble method would be evaluated using metrics such as mean squared error, R-squared error or R-squared to assess the accuracy of the model.

7. RESULT

In this project, collected data set will be uploaded and prediction for crop yield will be generated by applying ML techniques. The results depend on the information present in the collected data set. Accurate the information about the parameters in the collected datasets, better the results will be. A bar plot showing the efficiency of all models in crop yield prediction:
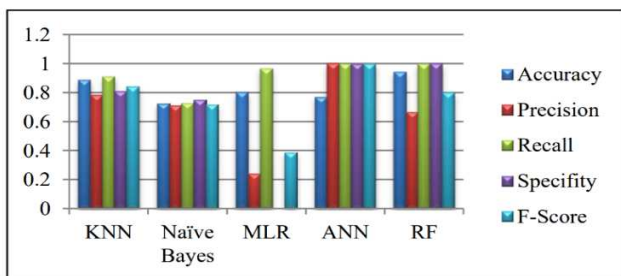

Figure 4. Accuracy comparison of all models

8. CONCLUSION

The effects of global warming on farm output from rain-fed systems were investigated in this research. It is hoped that better planning made possible by more timely information exchange, in this case on predicted yield, would assist to alleviate food poverty. Musanze (a dynamic agricultural area in Rwanda) data has been used to evaluate the accuracy of forecasts for maize and Irish potato yields using three different MLMs: RF, SVR, and PR. The R2, MAE, and RMSE values for both crops reveal that the RF model provides the best fit to the data. The data set included past agricultural yields and weather conditions, with meteorological factors (such as precipitation and temperature) serving as predictors. Furthermore, the relationship between the weather and agricultural yield has been investigated. The discussion section explains how we determined the best values of precipitation and temperature at each stage of crop growth for maximum crop output.

Future work:

Due to their efficacy in gleaning insights from measurable data, MLMs have been widely adopted as foundational technologies for ubiquitous computing. Section 3 demonstrates that RF model outperformed SVR and PR in predicting yields.

9. REFERENCES

[1] Riddhi Dange, Jinisha Kande,Kunal Bhosale, Prof. Ankush Hutke, "Predicting Agricultural Produce using Machine Learning Techniques", IJIRT, Volume 8, Issue 12, pp: 329-334, May 2022.

[2] Nischitha K, Dhanush Vishwakarma, Mahendra N, Ashwini, Manjuraju M.R, "Crop Prediction using Machine Learning Approaches", International Journal of Engineering Research & Technology (IJERT), Vol. 9, Issue 08, pp: 23-26, August-2020.

[3] Mythresh A, Lavanya B, Meghana BS, Nisarga B, "Crop Prediction using Machine Learning", International Research Journal of Engineering and Technology (IRJET), Volume: 07, Issue: 06, pp: 1697-1700, June 2020.

[4] E. Manjula, S. Djodiltachoumy, "A Model for Prediction of Crop Yield", International Journal of Computational Intelligence and Informatics, Vol. 6: No. 4, March 2017.

Special Issue - 2023

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICCIDT - 2023 Conference Proceedings

[5] V. A. Windarni, E. Sediyono, and A. Setiawan, "Using GPS and Google maps for mapping digital land certificate," Informatics and Computing (ICIC), International Conference on, 2016, pp. 422–426.

[6] S. Mishra, P. Paygude, S. Chaudhary and S. Idate, "Use of data mining in crop yield prediction," 2018 2nd International Conference on Inventive Systems and Control (ICISC) IEEE Xplore, pp. 796-802.

[7] Deepak Murugan, Akanksha Garg, Tasneem Ahmed, and Dharmendra Singh, "Fusion of Drone and Satellite Data for Precision Agriculture Monitoring", 2016, 11th International Conference on Industrial and Information Systems (ICIIS).

[8] Dmitrii Shadrin, Andrey Somov, Tatiana Podladchikova, and Rupert Gerzer, "Pervasive Agriculture: Measuring and Predicting Plant Growth using Statistics and 2D/3D Imaging", 2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC).

[9] Prof. D.S. Zingade, Omkar Buchade, Nilesh Mehta, Shubham Ghodekar, Chandan Mehta, "Crop prediction system using machine learning", International Journal of Advance Engineering and Research Development

[10] Vikky Aprelia Windarni, Eko Sediyono, Adi Setiawan, "Using GPS and Google Maps for Mapping Digital Land Certificates", 2016 International Conference on Informatics and Computing (ICIC).

[11] Rushikesh Bhav, Mayuresh Deodhar, Kevin Bhalodia, Prof. Mansing Rathod, "Android Application for Crop Yield Prediction and Crop Disease Detection", International Journal of Innovative Science and Research Technology, Volume 3, Issue 3, March– 2018.

[12] Santosh Reddy,, Abhijeet Pawar, Sumit Rasane, Suraj Kadam, "A Survey on Crop Disease Detection and Prevention using Android Application", IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 4, April 2015.