International Journal of Research in Engineering, Science and Management
Volume 4, Issue 8, August 2021
https://www.ijresm.com | ISSN (Online): 2581-5792

A Machine Learning Approach for Crop Prediction and Crop Yield Prediction

M. N. Charish Patel[1*], M. N. Kruthi[2], K. S. Shirisha[3], H. C. Karthik[4], M. J. Lahari[5]
[1,3,4]*Student, Department of Computer Science and Engineering, Sri Venkateswara College of Engineering, Bengaluru, India*
[2]*Student, Department of Computer Science and Engineering, RV College of Engineering, Bengaluru, India*
[5]*Student, Department of Information Science and Engineering, S.J.C. Institute of Technology, Chikkaballapur, India*

***Abstract*: Agriculture is the major source of food supply for the man-kind. In the light of decreasing crop production currently, deciding the right crop for the harvestable land has become the crucial criteria of agriculture. Therefore, we have pro-posed a method which helps in suggesting the most suitable crop for a specific land and predict its yield based on the analysis of geographical and climatic parameters using ma-chine learning techniques. In our project, we have imple-mented Random Forest regressor, decision tree regressor and gradient boosting regressor algorithms for crop yield prediction model and random forest classifier and decision tree classifier for crop prediction model. Algorithms were trained with proper training data set initially and later test-ed with test dataset. Then all the algorithms were compared based on certain metrics and the best algorithms were cho-sen for further implementation.**

***Keywords*: Machine Learning, Classification algorithm, Decision Tree regressor, Gradient Boosting regressor, Random Forest, Crop prediction, Yield prediction.**

1. Introduction

Agriculture is one of the major revenues producing sectors and acts a backbone of economy in many countries, as India is majorly a country of villages. Machine learning is a sub-field under Artificial Intelligence that uses statistical methods, to provide ability to the computer to learn from data. Recently, researches are being conducted to implement and adopt technology in crop cultivation, weather prediction, yield prediction, insect detection, fertilizer detection etc. This project deals with predicting the crop which can be cultivated on particular land and determine the yield of the crop using Machine learning algorithms and techniques using many external factors like climate, type of soil etc.

Effective utilization of agricultural land has become the crucial factor for successful farming but, finding out the best suited crop for their land becomes tedious for normal small-scale farmers. Hence, machine learning aims on the development of programs based on various algorithms that can learn and work when exposed to new data with relatively less cost. Training these models with data will give us desired output in an efficient way to predict the crop and yield of the crop we need to train the model with all the factors which is required. In the production of agriculture products like fruits, Vegetable, spices etc. are produced in a good quantity and quality only when it meets with certain factors like climate, fertilizers, water availability and ph and acidity contain in it etc. By keeping all these factors in mind, we have to work on building a model which helps in predicting the crop and yield in the condition and location of person growing crops. Here by taking all the factors affecting a crop production we built a model by training it with the pre-framed data set. This model can help in increasing the crop production and yield of crop considering all the factors affected. Many all over the world have been trying to increase the rate of productivity, we can use this model as reference to how crop productivity can be increased in particular area of land and in different climatic condition.

2. Literature Survey

Pavan Patil, Virendra Panpatil, Shrikant Kokate [1], By using Classification Algorithms, Decision Tree, KNN for prediction of crop by analyzing requisite data. Decision tree shows poor performance than native bayes when dataset is having more variations.

Saeed Khaki , Lizhi Wang [2], Neural networks were able to learn nonlinear Relationships between environmental conditions, and make accurate predictions of yields in new locations. There was drawback of black box property in this approach.

Marizel, B. Villanueva, Ma. Louella M. Salenga [3], Computer Vision, Convolutionary neural networks, MATLAB.

For Bitter Melon Crop Yield Prediction. Neural network turned out to be very effective. Huge Data was required for training the model in order to make the model effective.

Ashwini Rao, Janhavi, Abhishek Gowda, Manjunatha [4], K means clustering, Support vector machine algorithm for soil classification and crop prediction. Computational complexity of SVM does not depend on dimensionality of input data Less prone to overfitting

Karandeep Kaur [5]. Machine Learning: Applications in Indian Agriculture. Artificial neural networks, K-nearest neighbors Decision Trees were used to analyze agriculture in India where large amount of data sets were used as a prerequisite.

*Corresponding author: charishpatelmn3500@gmail.com

Fabrizio Balducci, Donato Impedovo Giuseppe Pirlo [6], Machine Learning Applications on Agricultural Datasets for Smart Farm Enhancement. Feed-forward neural network Linear regression decision tree were used to train the model.

Ferentinos, Konstantinos. P [11] implanted the system by training the neural networks and uses five basic CNN technologies AlexNet, AlexNetOWTBn, GoogLeNet, Overfeat, Visual Geometry Group.

3. Methodology

The proposed Method uses Data set and Algorithms to produce the desired output.

It follows as:

1. Data set
2. Algortithms:
    a) Gradient boosting regressor
    b) Decision tree regressor/classifier
    c) Random Forest classifier/regressor

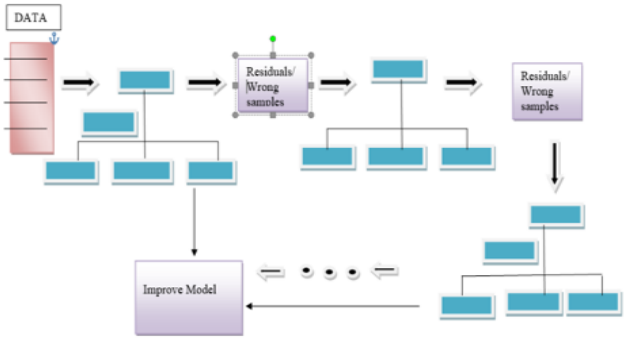

Fig. 1. Gradient boosting working

A. *Data Sets*

FAO (Food and Agricultural Organization) website is the source of datasets for the project. Dataset includes features such as country, area harvested, yield per hectare, average rainfall per year, average temperature, relative humidity, pH, number of fertilizers and pesticides used.

FAOSTAT is an organization of USA which provides free access to all statistics and data of food around the countries (Which includes stocks, agriculture market, Agriculture pests etc.) and territories. Every factored data which is required to create this model is available on this website.

B. *Algorithm*

In this model we use three algorithms i.e. gradient boosting regressor, Decision tree and random forest classifier. Gradient boosting is a machine learning technique that can be used for regression. It produces an ensemble weak prediction model in which typically decision trees act as weak learners. It builds a model in stage-wise fashion and generalizes them by allowing optimization of an arbitrary differentiable loss function. Decision tree algorithm is a supervised machine learning algorithm. Decision tree algorithm belongs to the family of predictive modelling approaches that are used in statistics, data mining and machine learning. Random forest is a supervised machine learning algorithm for classification and regression. Random forest creates multiple decision trees at training time and outputs the mode of the classes for classification problems and mean prediction for regression problems of the individual trees.

1) Gradient boosting regressor

Here we used this algorithm to create a weak model since gradient boosting takes loss function and minimize it. The loss function is a measure indicating how good are model's coefficients are at fitting the underlying data. Logically, loss function depends on what we are trying to optimize, the key is to minimize the error in the model. Gradient boosting also takes error value from previous model to reduce the error in the model. This model learns from the previous mistake and taking it into the factor it analyses the data further to give efficient output. Loss functions is the major factor in this method since it reacts from the error being noted and taking it into the factor it starts increasing the efficiency of the model.

2) Decision Tree regressor

It uses a decision tree as a predictive model to traverse from observations about an item represented by branches to conclusions about the item's target value represented by leaf nodes. Decision tree algorithm can be used for solving both regression and classification problems. The goal of using decision tree algorithm is to create a training model that can predict target class or value for input variable by simple decision rules inferred from prior training data. In prediction process of decision tree, we start from root of the tree, values of the root and record's attribute are compared, on the basis of results of comparison we follow the branch that corresponds to that value and jump to the next node. Decision tree which has categorical target variable is called as categorical variable decision tree and the one with continuous target variable is known as continuous variable decision tree.

3) Random Forest classifier

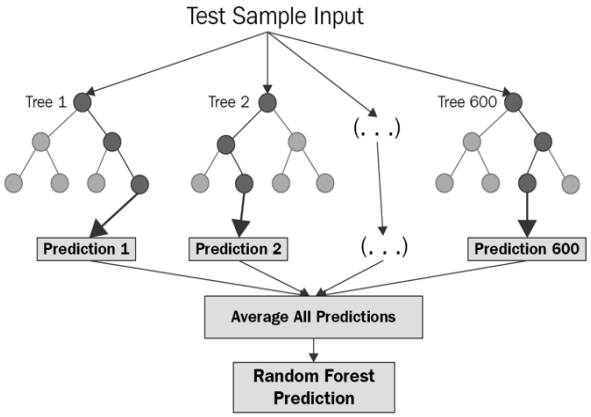

Fig. 2. Random Forest test data analyzing view

Random forest builds an ensemble of decision trees that are trained by bagging method and the trees are randomized by precision techniques, which substantially increases the performance of model. We have implemented random forest in both crop prediction and yield prediction as it can be used for both classification and regression problems. It uses averaging to overfit the model and analyze the data based on it.co-relation between the model is low which is the key for the model. This model is probably more powerful than the others. Every tree in this classifier protects itself from the errors. Since it takes the

recurring average of the decision trees makes it effective and accurate among the models in many conditions. The machine learns from past experience and tries to make accurate decisions in further steps. Random forest creates multiple decision trees at training time and outputs the mode of the classes for classification problems and mean prediction for regression problems of the individual trees.

4. Experimental Result and Discussion

R-Squared/R^2 also called as coefficient of determination is a statistical measure in a regression model that determines the proportion of variance in the dependent variable that can be explained by the independent variable. In other words, R^2 shows how well the data fit the regression model. R^2 value lies between 0 and 1, the more R^2 value approaches 1, more is the goodness of fit of the regression model.

R^2 value obtained for algorithms in crop yield prediction model are as follows:

Decision tree regressor: 0.97429

Gradient boosting regressor: 0.97427

Random forest regressor: 0.87807

Hence, decision tree regressor algorithm is implemented in crop yield prediction mode.

The accuracy of a classification algorithm is one way to measure how well the algorithm classifies a data point correctly. Accuracy is the number of correctly predicted data points by the model out of all the data points. More formally, it is defined as the number of true positives and true negatives divided by the number of true positives, true negatives, false positives, and false negatives. A true positive or true negative is a data point that is correctly classified as true or false by the algorithm, respectively. on the other hand, false positive or false negative is a data point that the algorithm incorrectly classified.

Accuracy of algorithms used in crop prediction model are as follows:

Decision tree classifier: 87.567

Random forest classifier: 90.967

Hence, random forest classifier forest algorithm is chosen for crop prediction model.

Images of GUI are as follows:

Fig. 3. GUI

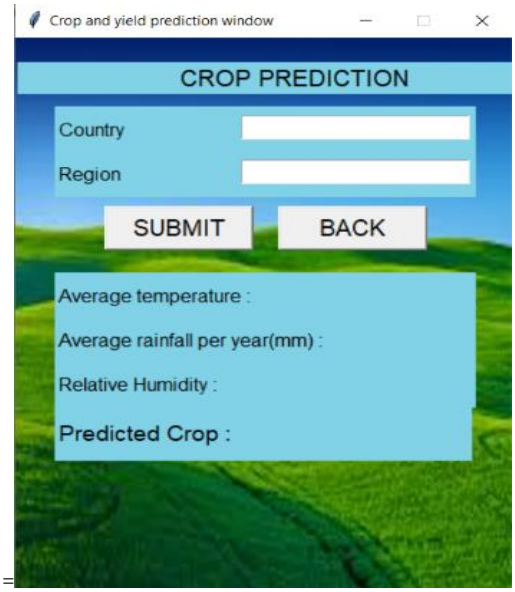

Fig. 4. GUI of Crop Prediction

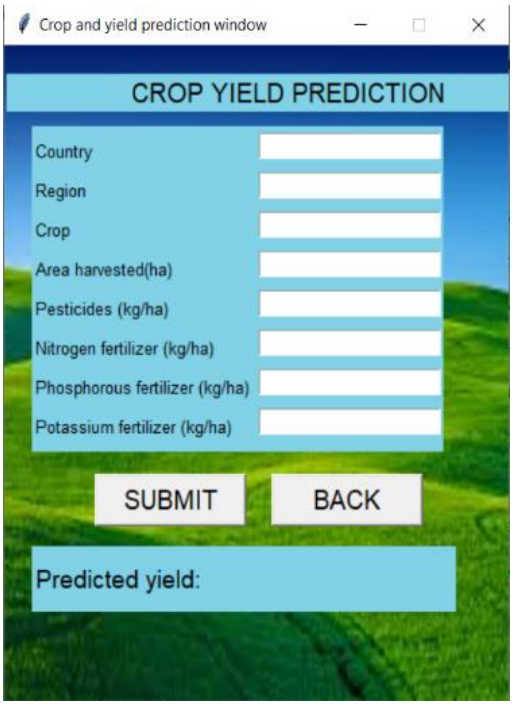

Fig. 5. GUI crop yield prediction

5. Conclusion and Future Work

In current world, there is a drastic change in Agri-Tech currently. Not most of the farmers are using latest tech gadgets in their farms. We often see IoT related agriculture in several journals but none of them are properly being adopted in Indian farms. There is a huge gap between technology and farmers in current trend. Many start-ups have emerged to bridge this gap between the technology and the farmers. Now, even many MNCs are investing in Agri-Tech. Demand for food is exponentially increasing due to rise in population. People thoughts about heavy machinery in farms era is now being replaced by smart technology such as Internet of Things, Artificial Intelligence and Machine Learning.

Today, the agricultural sector has grown into a highly competitive and globalized industry, where farmers have to consider local climatic and geographic aspects in order to guarantee economic survival and sustainable production. Feeding this growing world population asks for continuous increases in food production, but harvestable land remains a limited resource. Today, agriculture is growing as a highly competitive and globalized industry. Farmers should guarantee their production sustainability for the economic survival of agriculture field, hence there arises a matter of taking geographical and climatic features into account for sustainable production. Feeding this exponentially growing world population demands continuous exponential increase in food production as well, but this is mainly hindered by the limited availability of crop land. So, agriculture should find out the way for continuous and sustainable increase in productivity and efficiency on all its stages to cope with these challenges, Agriculture requires a continuous and sustainable increase in productivity and efficiency on all levels of agricultural production.

Presently our farmers are not effectively using technology and analysis, so there may be a chance of wrong selection of crop for cultivation that will reduce their income. To reduce those type of loses we have developed a farmer friendly system with GUI, that will predict which would be the best suitable crop for particular land. This makes the farmers to take right decision in selecting the crop for cultivation so that agricultural sector will be developed by innovative idea. In the present model temperature, rainfall data is being taken from openweathermap.org, a weather API which can be improved by GPS interfacing, so that climatic condition data of any geographical region can be obtained easily and accurately. Voice input feature can be included which makes the model more flexible to be used by farmers.

Acknowledgment

The Authors wish to thank SRINIVASA G, Assistant Professor, Dept. of CSE, Sri Venkateswara College of engineering Bangalore for providing assistance and valuable suggestion for the work. Review and comments from him were very helpful. Authors are also thankful for support he lended which was very helpful and essential in bringing out this work.

References

[1] Pavan Patil, Virendra Panpatil, Shrikant Kokate, “Crop Prediction System using Machine Learning Algorithms,” International Research Journal of Engineering and Technology.

[2] Saeed Khaki, Lizhi Wang, “Crop Yield Prediction Uing Deep Neural Networks”, Industrial Engineering Department, Iowa State University, Ames, Iowa, USA.

[3] Marizel, B. Villanueva, Ma. Louella M. Salenga, “Bitter melon crop Yield Prediction using Machine Learning Algorithm”, International Journal of Advanced Computer Science and Applications

[4] Ashwini Rao, Janhavi, Abhishek Gowda, Manjuntha, “Machine Learning in Soil Classification and Crop Detection,” International Journal for Scientific Research and Development.

[5] Karandeep Kaur, “Machine Learning: Applications in Indian Agriculture,” International Journal of Advanced Re-search in Computer and Communication Engineering.

[6] Fabrizio Balducci, Donato Impedovo Giuseppe Pirlo, “Machine Learning Applications on Agricultural Datasets for Smart Farm Enhancement”, Dipartimento di Informatica, Università degli studi di Bari Aldo Moro, 70125 Bari, Italy.

[7] S. R. Rajeswari, Parth Khunteta, Subham Kumar, Amrit Raj Singh, Vaibhav Pandey, “Smart Farming Prediction Using Machine Learning,” International Journal of Innovative Technology and Exploring Engineering.

[8] karishma Mohiuddin, Mirza Mohtashim Alam,” A Short Review on Agriculture Based on Machine Learning and Image Processing,” Acta Scientific Agriculture.

[9] T. Venkat Narayana Rao, “Prediction of Soil Quality Using Machine Leaning Techniques,” International Journal of Scientific & Technology Research, Volume 8, Issue 11.

[10] Victor Mokaya, “Future of Precision Agriculture in India using Machine learning and Artificial Intelligence,” International Journal of Computer Sciences and Engineering.

[11] Ferentinos, Konstantinos P, “Deep learning models for plant disease detection and diagnosis,” Computers and Electronics in Agriculture, vol. 145, pp. 311-318, 2018.