

MACHINE LEARNING ALGORITHMS FOR CROP YIELD PREDICTION: A SURVEY

K. PORCHILAMBI

Ph.D Research Scholar, PG & Research Department of Computer Science,
Vivekanandha College of Arts and Sciences for Women (Autonomous), Elayampalayam.

Dr. P. SUMITRA

Assistant Professor PG & Research Department of Computer Science,
Vivekanandha College of Arts and Sciences for Women (Autonomous), Elayampalayam.

Abstract: Agriculture is full of uncertainty due to climate change, ground water deficiency, rainfall and evolution of new pests. Crop yield prediction in agriculture is a very big dilemma and there is huge dataset where farmers find difficult to predict the yield and seed selection. In today's situation due to increase in the population the production of grains and agricultural products needs to be increased simultaneously to meet the demands of the people. Olden farming techniques need to be combined with the modern technology to enrich the results. Several Environment and economic factors like climate, rain and pesticide are unpredictable that affects the crop growth which in turn affects the productivity. Statistical and mathematical tools can be used to quantify the yield using past data. The result can assist farmers in crop choice and give an insight of the productivity that result in increased profit. Data set has been collected from Tamil Nadu and statistical websites for analysis and Machine learning algorithms and their role in agriculture were analyzed. This paper focus on the Machine learning algorithms like Naïve Bayesian, Support vector machine, Neural networks, Decision tree, K Nearest Neighbor in crop yield prediction.

Keywords: Machine learning, K-Nearest neighbor, Naïve Bayes, Support Vector Machine, Decision tree, Regression tree.

I. INTRODUCTION

India is an agricultural country. India's economy is determined by agricultural products export and import. Agriculture plays a vital role in Indian economy. Due to uncertainty in the crop yield there is a great fall in the economic status. The major crops of India are Rice, Wheat, Pulses and Grains. Day by day the population of India is growing and the crops productivity need to be increased to feed the population. Government policies and researchers were putting great effort to motivate formers and provide useful information to them. The productivity of crop is highly influenced by the factors like water management, soil nutrients and pesticides. The irrigation is a key factor for the crop growth. Irrigation is reliant on ground water, seasonal rainfall and canal water. Due to the scarcity of water the crop growth and productivity is mainly affected. The Climate change also plays a vital role. Today due to global warming the climate is unpredictable.

The nutrients of soil like zinc, copper, magnesium need to be maintained for crop growth. Proper pesticides usage prevents the loss of crop. Today there is a variety of information available to gain knowledge about agriculture. These information need to be processed for making appropriate decision in crop selection and yield prediction. The joint venture of statistics and Information technology will help the formers to take suitable decision in crop plantation. Machine learning (ML) techniques like K-Nearest Neighbor, Naïve Bayesian classification, Neural networks, Decision Tree (DT), DT rules, Support vector machine (SVM) and Random forest are suitable for dealing with these problems. These algorithms can be used to analyze the raw data and to predict a feature forecast of the data sets.

II. LITERATURE REVIEW

Machine learning is used in different fields like Medicine, Banking, Web based applications and Agriculture for prediction, classification and clustering. In this paper Machine learning applications in the field of agriculture has been reviewed and discussed. Machine learning algorithms is suitable for predicting large data set and certain unpredictable situations prevailing in Crop prediction can be solved using these algorithms.

A. Neural Networks

Artificial Neural network (ANN) works similar to human brain and its structure is based on the neuron cells of brain. A typical ANN contains three layers, input layer, one or more hidden layers and an output layer.

The input layers X_i are associated with weights W_{ij} and passed to the hidden layer. A transfer function f_i is applied to the weighted values to produce the output Y . The weights are adjusted to match the observed and output of ANN. Bias b_i is added with the hidden layer. Neural networks learn through examples and the knowledge is stored in hidden layer Multilayer perceptron and back propagation models are used to predict the outcome of the model.

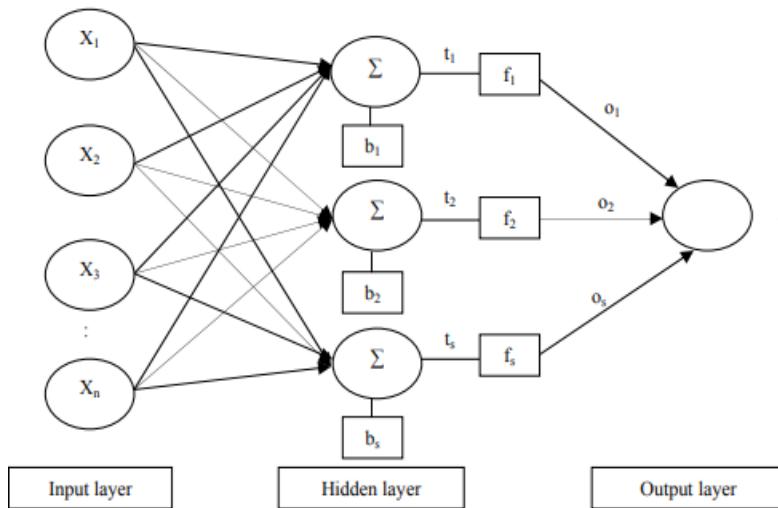


Fig 1- Neural Network with three layers

B. Support Vector Machine

Support Vector Machine (SVM) is a classifier method that performs classification tasks by constructing hyperplanes in a multidimensional space that separates cases of different class labels. SVM supports both regression and classification tasks and can handle multiple, continuous and categorical variables.

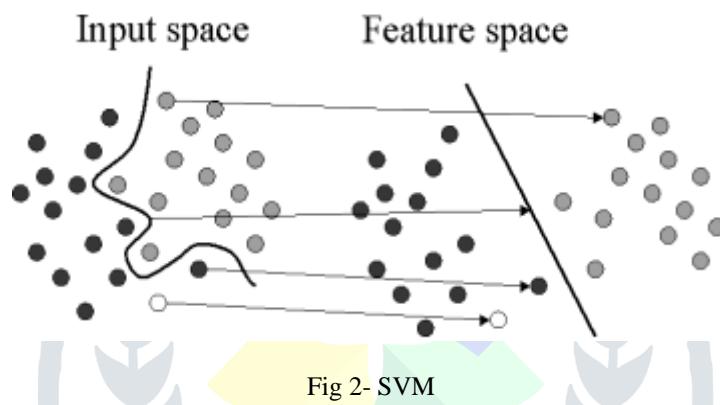


Fig 2- SVM

C. K-Nearest Neighbor

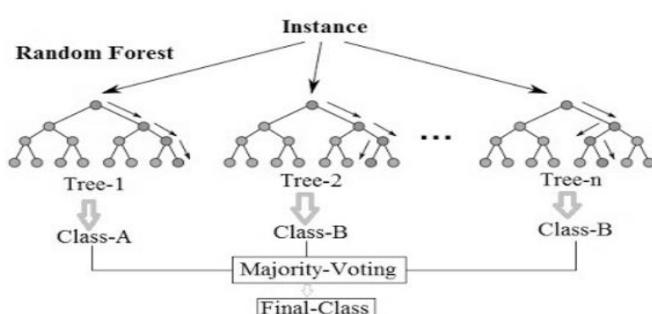
The K-Nearest neighbor (KNN) is a classification model that assigns the new input to the nearest neighbor based on the distance factor. Euclidean distance is calculated between the neighbors and the data are assigned based on the closeness to the clusters. The data within the cluster are similar to each other and they are dissimilar to other clusters.

Algorithm:

1. Load the crop yield data
2. Initialize the value of k (randomly)
3. For getting the predicted class, iterate from 1 to total number of training data points
4. Calculate the distance between test data and training data. (using Euclidean, Chebychev)
5. Sort the calculated distances in ascending order based on distance values
6. Get top k rows from the sorted array
7. Get the most frequent class of these rows
8. Return the predicted class

D. Random forest

Random forest is a supervised learning method, which build forest of decision trees and use the prediction of decision trees to make better prediction.



It is an Bagging method used to find the efficient method and implement. Random forest is suitable for classification and Regression

E. Naïve Bayes

The Naïve bayes works on the basis of Bayes theorem.

$$P(C/X) = P(X/C) P(C) / P(X) \quad \text{----- (1)}$$

- $P(c/x)$ is the posterior probability of *class* (*c*, *target*) given *predictor* (*x*, *attributes*).
- $P(c)$ is the prior probability of *class*.
- $P(x/c)$ is the likelihood which is the probability of *predictor* given *class*.
- $P(x)$ is the prior probability of *predictor*.

Based on the posterior probaility the future of data can be predicted. Naïve bayes work well for large data set.

F. Decision tree

Decision tree is a predictive model which works by checking condition at every level of the tree and proceeds towards bottom of the tree where various decisions are listed. The condition depends on the application and the outcome might be in terms of decision. The factors like Rainfall, Climate, Pesticides, soil type and groundwater can be used as input to Prediction algorithm and the resultant will be in terms of the prediction of productivity.

III. METHODOLOGY

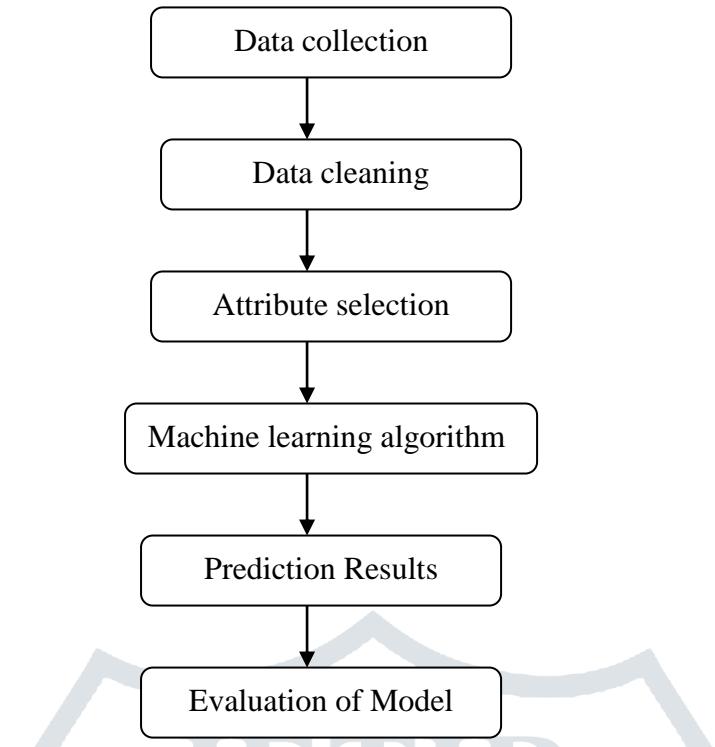
Machine learning suits for agriculture field and several researches have been taken place to improve the productivity using available resources like water, seeds, pesticide and so on. The most influencing factors of agriculture were listed below in the table.

Table I. Attributes for crop prediction

S.no	Attribute	Values	Description
1	Soil type	Black, Red, Alluvial, Clay soil	The nature of soil
2	Rainfall	Low, Medium, High	Amount of Rainfall during cultivation time
3	Season	Kaar, Advanced Kaar, Pishanam	Time of cultivation
4	Fertilizer	Dung, Urea, Potassium, Sulphide	Type of fertilizers used in field
5	Soil composition	Nitrogen, phosphorous, potassium	The minerals ratio specified using terms Low, high, Medium
6	Temperature	Low, Medium, High	The temperature level during cultivation
7	Seed Quality	Medium, Good	Same varieties of seeds with different quality
8	Crop rotation	Rice, Ragi, Sugarcane, Groundnut	Crops used throughout the year
9	Sowing procedure	Manual, Machine	Method of sowing the seeds

Steps of algorithm:

1. Collect the data from different sources and clean the data.
2. The attribute selection is mandatory to improve the efficiency of the algorithm. The most influencing attribute is selected.
3. The data is pre-processed and fed to the Machine learning algorithm.
4. The best predictive Machine learning algorithm is selected and applied.
5. These steps are carried out until accurate results are obtained.



Conclusion

Agriculture is a field full of uncertainty. Indian farmers still adopt the traditional models and they lack in Modern technologies. Machine Learning techniques can be implemented in the plantation field and the relationship between different influencing factors can be predicted and the yield can be improved.

The findings from this paper are

- 1) It is visible that Machine learning algorithm can be used to improve the crop yield and the Hybrid method can be devised to fit the requirement and demand of the farmers.
- 2) The study shows that all the authors have researched about the crop yield and less concentrated area is crop selection using Machine learning. Quality crop can yield quality crops as well as increase in quantity.

The future work of this paper may be based on Crop selection and to improve the crop yield rate using Machine learning methods.

IV. REFERENCES

- [1] Utkarsha P Et al, "Evaluation of Modified K-means clustering algorithm in crop prediction, International journal of Advanced computer research, vol. 4, no. 3, September 2014, Pp: 799-807
- [2] Sheens Agra, Sachin Ahuja, "Machine learning and its application: A Review", IEEE Press, 2017, pp: 57-60, International conference on Big data analytics and computational intelligence
- [3] Siti Khairunniza-Bejo Et al, "Application of Artificial Network in predicting crop yield: A review", Journal of food science and Engineering, vol.4, January 2014, pp: 1-9.
- [4] Vaneesbeer Singh Et al, "Analysis of soil and prediction of crop yield(Rice) using Machine learning approach", International journal of advanced research in computer science, vol. 8, no. 5, May-June 2017, Pp: 1254-1259
- [5] Narayanan Balakrishnan and Dr.Govindarajan Muthukumarasamy, "Crop production-Ensemble Machine learning model for prediction", International journal of computer science and engineering, vol. 5, Issue 7, July 2016, Pp: 148-153
- [6] Aadithya Et al, "Fuzzy logic based hybrid recommender of maximum yield crop using soil, weather and cost", ICTACT journal on soft computing, vol. 6, Issue 4, July 2016, Pp: 1261-1269
- [7] Nur Athirah Ashaary Et al, "Neural network application in the change of reservoir water level stage forecasting", Indian journal of science and technology, vol. 8(13), July 2015, Pp: 1-6
- [8] S.Veenadhari Et al, "Machine learning approach for forecasting crop yield based on climatic parameters", IEEE press, Jan 2017
- [9] Subhadra Mishra Et al, "Applications of Machine learning techniques in agricultural crop production: A review paper", Indian journal of science and technology, vol. 9(38), DOI10.17485/ijst/2016/v9i38/95032, October 2016
- [10] Prajakta prashant bhangale Et al, "Improved crop yield prediction using neural network", IJARIIE, vol. 3(2) 2017, Pp: 3094-3101
- [11] Dr.A. R. Pon Periyasamy, P. Sharmila Devi, "Density based clustering technique on Crop prediction", International journal of Advanced research in Computer science and software engineering, vol. 7, Issue 4, April 2017, pp. 246-250.
- [12] Rupinder singh, Gurpreet singh, "Wheat crop yield assessment using Decision Tree Algorithm", International journal of Advanced research in Computer science, vol.8, No.5, May-June 2017, Pp: 1809-
- [13] Hetal Patel and Dharmendra Patel, "A comparative study on various Data Mining with special reference to Crop yield prediction", International journal of Science and Technology, vol 9(22), June 2016.
- [14] Raorane A.A, Kulkarni R.V, "Data Mining: An effective tool for yield estimation in the agricultural sector", International journal of Emerging trends and Technology in Computer Science, vol.1, Issue 2, July-August 2012, Pp: 75-79

- [15] E.Manjula, S.Djodiltachoumy, "A Model for prediction of Crop yield", International journal of computational Intelligence and Informatics, vol. 6, o. 4, March 2017, Pp: 298-305
- [16] Niketa Gandhi Et al,"Rice crop yield prediction in India using support vector machines", IEEE press 2016,
- [17] SML Venkata Narasimhamurthy, AVS pavan Kumar, "Rice crop yield forecasting using random forest algorithm", International journal for research in Applied science and Engineering technology, vol.5, Issue 10, october 2017, Pp:1220-1225
- [18] S.Veenandhari Et al, "Soyabean productivity Modelling using Decision tree algorithm", International journal of computer applications, vol.27, No. 7, August 2011,Pp:11-15
- [19] Anitha Arumugam, "A predictive modeling approach for improving paddy crop productivity using data mining techniques", Turkish journal of Electrical engineering and computer sciences, Doi: 10.3906/elk-1612-361, Pp: 4777 -4787

