

# Crop Recommendation Using Machine Learning Techniques

Hrushikesh Gargote<sup>1</sup>, Vaibhav Vyas<sup>2</sup>, Allwin Noble<sup>3</sup>, Akshay Chorge<sup>4</sup>  
Modern Education Society's College of Engineering

**Abstract-** India ranks second in the world in terms of population. Due to this there is a huge demand for food and a concern towards the food security of India. The most common problem faced by Indian farmer is selection of right crop. Selection of the right crop increases the productivity and thus enables farmers to increase their overall yield. This common problem of farmers can be solved by Crop Recommendation (CR) systems. In the Crop Recommendation systems we use the historical data of soil, weather and market prices and analyze the data using different machine learning algorithms and recommend the farmers the right crop based on their specific parameters. This increases the yield of crops and avoids wrong choices. In this survey paper the problem is solved by proposing a list of various crop recommendation systems which analyze the various parameters by using machine learning algorithms and can predict the best suitable crop for sowing.

**Keywords-** Crop recommendation(CR), machine learning.

## I. INTRODUCTION

Due to rapid increase in urbanization in India the land resources for agriculture has been reducing significantly also there is not much use of advanced technology in agriculture as seen in other countries. Due to this, agricultural production isn't increasing in proportion to the growing population. So a Crop Recommendation (CR) system has to be designed considering various factors that affect the crop growth. The factors which affect the yield of crop are climate, soil composition and also the quality of seeds. The weather conditions like rainfall, wind, thunder storm, etc majorly affect the yield. The soil compositions like nitrogen, phosphorous, zinc, etc. also affects the crop health. Agriculture is a planned activity where goal is to generate an efficient working model for crop prediction. This paper proposes an array of method called Crop Recommendation (CR) system to increase net yield of crops.

Predictions made by system should be accurate and precise because in case of false prediction it may lead to heavy losses for the farmer.

## II. RELATED WORK

Karandeep Kaur discussed has discussed about various applications of data mining techniques in agriculture [1]. The various applications stated by her includes crop selection and crop yield prediction, weather forecasting, smart irrigation system and crop disease prediction. According to him selection of the appropriate crop that should be sown plays a very important role in increasing crop yield. Also Indian farmers are mostly dependent on seasonal rains for irrigation so according to him data mining techniques must be used both for proper forecasts as well as developing smart irrigation systems to make use of available water efficiently.

The paper [2] states that the common problem which exists among the Indian farmers is that they don't choose the right crop based on their soil requirements. Due to this they have a decrease or a setback in their productivity. In this paper the problem is solved by proposing a recommendation system which uses an ensemble model which incorporates voting technique using algorithms like Random tree, CHAID, K-NN, Naive Bayes as learners to recommend a crop according to site specific parameters with an increased accuracy and efficiency. The results show that this model has an accuracy of around 88% which can be considered trustworthy in recommending the right crop.

The paper [3] elaborates that if the land resource is limited, then selection of crop is a problem. This paper puts forward a method named Crop Selection Method to solve problems regarding selection of the right crop and maximizing net yield rate of crop. According to this paper there are 2 factors that influence yield: Seed Quality and Crop Selection Management. Crop Selection Method is used to select a sequence of crops to be planted over season. It improves net yield rate of crops to be planted over season. It takes crop, sowing to plantation days and predicted yield rate for season as inputs and finds a sequence of crops whose production per day are maximum over season. The disadvantage of this method is that this method CSM depends on the predicted value of the influenced parameters.

The paper [4] uses two datasets of mushroom and soybean. Here the datasets are first subjected to data preprocessing techniques, the PSO search algorithm that acts as an attribute selector, to remove noise is used on the datasets as a part of data preprocessing. The resultant datasets are given to known classifiers like the Naive Bayes, Decision Table and find inference in the observations. These machine learning techniques in combination with PSO search algorithm acts as a major and positive factor in enhancing performance of the model. The results show that the total performance of the classifier can be increased by using of the attribute selector in most of the tested cases. Another algorithm the ZeroR classifier, when this classifier is applied to the datasets, there is an increment in the accuracy parameter for both datasets. It demonstrates that machine learning techniques with feature reduction or data reduction methods is a positive factor in improving and enhancing agricultural production efficiently.

The paper [5] elaborates a web based recommender system for farmers. For the Crop recommender system Random Forest Algorithm was used. A rating system was made based on three factors namely year of cultivation, market price and output/area ratio of the resultant crop. And based on highest rating a crop was recommended. Crop rotation recommendation was implemented using FP tree algorithm through which a farmer could know which sequence of crops could yield better profits. A fertilizer recommender is also implemented. Once a crop is recommended by its required levels of nutrients are validated with the current level of nutrients present in soil and accordingly the appropriate fertilizers along with price are suggested to the farmer.

Niketa Gandhi, Leisa J. Armstrong, Owaiz Petkar proposes Decision Support Systems (DSS) prototype to predict the rice crop yield for Maharashtra state of India and also visualize the historic data. The DSS contains two tabs. One tab is for predicting the yield based on parameters like precipitation, temperature and reference crop evapotranspiration. The other tab for is for data visualization. There are three ways to visualize the data that is year wise, parameter wise and combination with respect to yield. This enabled farmer to know about the future yield and make smart decisions accordingly.

In this paper [7][8] have tested efficiency of various algorithms like SVM[7] and ANN[8] for Rice crop yield prediction. WEKA which is free and open source tool was used for this. For testing 10-fold cross validation was used. Using SVM [7] the classifier was able to produce an accuracy of a good 78.76%. Multilayer perception was used as artificial neural network [8]. In this research the algorithm achieved the

accuracy of 97.54%, sensitivity of 96.33% and specificity of 98.12%. The inefficiency of linear methods was overcome by non-linear methods like ANN which produced a more accurate prediction. This study showed that other classifiers like Naïve Bayes, Bayes Net and Multilayer Perception performed better than SVM.

The paper [9] states the use of clustering approach based on densities which can be used to predict future weather conditions. Here too like paper [4] preprocessing is done on datasets to improve their performance and efficiency. The preprocessing approach used here is the Convex Hull algorithm. This algorithm is applied on the dataset before feeding the clustering algorithm used in this paper with datasets. The database used here is of Kolkata and is of the pollutant data. The paper focuses on weather prediction using air molecules and partial wind flow.

The benefits stated are that it helps in predicting per day forecast and also predicts long term nature of water. The accuracy of this model is at least 74.5%.

The paper [10] proposes a model for weather forecasting using deep Learning methods such as: Recurrent Neural Networks, Conditional Restricted Boltzman Machine and CNN. The dataset which was considered for this experiment was the ENSO dataset provided by International Institution such as National Weather Service Center for Environmental Prediction Climate (NOAA). In their research they found out those deep learning methods provided accurate classification and prediction for many time series problems.

The paper [11] gives an approach to soil moisture prediction and also its collection. In this paper they have models generated by machine learning algorithms like the support Vector, the relevance vector machine on top of which they have created a moisture prediction framework which would be specific to a given site. The created framework is able to predict soil moisture the required number of days (n) ahead based on soil and environmental attributes. This model can forecast soil moisture with a time lapse of at least 2 weeks. The paper states that soil moisture is one of the key factors in the cycle of growth of crops. The model predicts soil moisture n days ahead. The model predicts soil moisture values based on meteorological parameters including temperature, humidity, wind speed, solar radiation, and precipitation together with soil moisture values from the previous day. The dataset which is used in this paper is from the Illinois Climate Network.

### III. PROPOSED MACHINE LEARNING TECHNIQUES.

## A. Naïve Bayes :-

Naïve bayes is a supervised classification algorithm which is based on the implementation of the Bayes theorem. Bayes theorem calculates the probability of an event and predicts its occurrence with respect to some prior knowledge about that event.

For classification it assumes that one particular feature is independent from all other features used to describe an event. Even though all features of an event are dependent on each other the algorithm considers how an individual feature contributes to that particular event. . So it is one of the best algorithms for building a recommender system where one feature is considered to predict multiple class values.

## B. ANN:-

Artificial Neural Network(ANN) is a method based on or it can be said as inspired from the biological concept of the neurons working in the human brain. Each neuron is called a node. The nodes take data as input and performs computational operations on that data. The result is passed to the next set of neurons. The result or the output of each node is called its activation or node value. The links present between each pair of neurons is assigned a particular weight .According to the weight the strength of the link is computed. Weights are constantly reassigned to links and the best possible link between them is selected. There are two types of Topologies of Artificial Neural Networks. One of them is the Feed-Forward topology and another is the Feedback topology. In feed-forward topology the information flows in a unidirectional way whereas in feedback topology the presence of loops is allowed between nodes.

## C. PSO:-

PSO search: - The PSO or the Particle Swarm Optimization is a method by which a problem can be optimized trying to improve a candidate solution in iterations with regard to a given measure of quality. In this method according to a simple mathematical formula over particle aspects of position and velocity the problem can be tackled by having some candidate particles, displacing these particles in the search space. Each particle is guided to a position best known in the search space. These positions are then marked as good positions as found by other candidate particles. This then moves the swarm towards the expected best solution.

## D. Convex Hull:-

The Convex Hull algorithms are a type of algorithms that have a range of applications in mathematics in computer science. Known algorithms computing convex hull are as follows:

- 1.Gift Wrapping or Jarvis March.
- 2.Graham Scan
- 3.Quick Hull
- 4.Monotone chain or Andrew's algorithm
- 5.Incremental convex hull algorithm
- 6.Chan's Algorithm

The idea is to remove those points that would not be part of the convex hull. The proposed method is to find two points with lowest and highest x-coordinates and two points with highest and lowest y-coordinates.

## E. K-nearest Neighbor:-

KNN is non-parametric pattern detection process used for mainly classification and regression. The KNN can be used for classification or regression. KNN is slothful learning or occurrence based learning where each of the computation is deferred until classification. KNN algorithm is the simplest in all machine learning algorithm. Euclidean distance is commonly used distance metric for continuous variables.

$$d(\mathbf{p}, \mathbf{q}) = d(\mathbf{q}, \mathbf{p}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2}$$

$$= \sqrt{\sum_{i=1}^n (q_i - p_i)^2}.$$

Here p and q are two points in Euclidean n dimensional space.

## F. Random Forest:-

Random forest is also called as random decision forest algorithm which is collection of learning practice for classification, regression and further tasks operated by constructing a massive amount of decision trees. At training time and outputting the group that is the style of the classification or regression of the individual trees. In meticulous, trees that are grown incredibly profound be likely to be taught highly uneven patterns, they over fit their training sets. Random forests is a means of averaging numerous profound decision trees, skilled on different parts of the similar training set, with the objective of dropping the variations. Random forests can be used to position the significance of variables in a classification or regression problem in a normal way. As an element of their construction, random forest predictors logically guide to a variation measure between the observations. Individual can moreover

characterize a random forest contrast measure between unlabeled data.

#### G. Support vector machine:-

Support vector machine (SVM) is a supervised model in machine learning. It is used for regression and classification.

SVM can perform non-linear classification. SVM uses a clustering algorithm known as support vector clustering. It is used when data is not labelled. It transforms data and on this it finds the optimal boundary between possible outputs by using kernel trick. It does complex data transformation and then separate data on the basis of labels. By using non-linear classification we can capture more complex relationship. It is based on the theory of decision planes that defines the decision boundaries. It separates the set of objects into their respective groups.

Linear kernel is the equation for prediction for input  $f(x) = B(0) + \sum(a_i * (x, x_i))$  here input is  $(x)$  and each support vector is  $(x_i)$ . It is used for accuracy, works on smaller cleaner dataset; it works on subset of training points. It is used for text classification.

#### F. CHAID:-

Chi-square automatic interaction detection is a decision tree model. It is used for prediction and classification. CHAID also detect the interaction between variables. It is highly visual and also it is easy to interpret. It is having non-parametric multiple regression. Here we have can visually see split variables and associated factor. By identifying target variable it makes the classification. It makes two part initial and parent nodes. It does not require data to distribute. Here the data must be merged and the chi-square must be tested. There are following components of the decision tree in CHAID  
 Root node :-It has the dependent and targeted variable. Here it has targeted and predictor variables.  
 Parent's node: - All high, medium and low are the part of the parents. node.  
 Child node: - All independent variable are part of this node. It is just below the parent's node.  
 Terminal node: - It is the last part of the CHAID. Here dependent variables come first and fewer variables come after hence it is terminal node. It is always useful when the the data set has various variables, and then it has to summarized.

### III. CONCLUSION

India is a nation in which agriculture plays a prime role. In prosperity of the farmers, lies the prosperity of our nation. Thus the works proposed in this paper would help farmers in sowing the right seed depending on the soil requirements to increase productivity and acquire profit out of such a methods. Thus the farmer's can plant the right crop increasing his yield and also increasing his profit .

### REFERENCES

- [1] Rakesh Kumar , M.P. Singh , Prabhat 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).
- [2] S.Pudumalar, E.Ramanujam, R.Harine Rajashreen, C.Kavyan, T.Kiruthikan, J.Nishan, “Crop Recommendation System for Precision Agriculture” 2016 IEEE Eighth International Conference on Advanced Computing (IcoAC).
- [3] Karandeep Kaur, “*Machine Learning: Applications in Indian Agriculture*” international Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 4, April 2016.
- [4] Soumya Sahoo, Sushruta Mishra, Bijayalaxmi Panda, Nachiketa Jena “ Building a New Model for Feature Optimization in Agricultural Sectors” 2016 International Conference on Computing for Sustainable Global Development (INDIACom).
- [5] Kiran Shinde , Jerrin Andrei , Amey Oke , “ Web Based Recommendation System for Farmers” International Journal on Recent and Innovation Trends in Computing and Communication Volume: 3 Issue: 3.
- [6] Niketa Gandhi, Leisa J. Armstrong, Owaiz Petkar, “ Proposed Decision Support System (DSS) for Indian Rice Crop Yield Prediction” 2016 IEEE International Conference on Technological Innovations in ICT For Agriculture and Rural Development (TIAR 2016).
- [7] Niketa Gandhi, Owaiz Petkar, Leisa J. Armstrong, “ Rice Crop Yield Prediction Using Artificial Neural Networks”, 2016 IEEE International Conference on Technological Innovations in ICT For Agriculture and Rural Development (TIAR 2016)
- [8] Niketa Gandhi, Owaiz Petkar, Leisa J. Armstrong, Amiya Kumar Tripathy, “ Rice Crop Yield Prediction in India using Support Vector Machines” 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE).

- [9] Ratul Dey, Sanjay Chakraborty, “ Convex-Hull & DBSCAN Clustering to Predict Future Weather”, 978-1-4799-6908-1/15/\$31.00 ©2015 IEEE.
- [10] Afan Galih Salman, Bayu Kanigoro, Yaya Heryadi “ Weather Forecasting using Deep Learning Techniques” ICACSI 2015.
- [11] Zhihao Hong , Z. Kalbarczyk , and R. K. Iyer “A Data--Driven Approach to Soil Moisture Collection and Prediction ”, 978-1-5090-0898-8/16/\$31.00 ©2016 IEEE .