International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

Crop Prediction based on Soil Classification using Machine Learning with Classifier Ensembling

Vrushali C. Waikar[1], Sheetal Y. Thorat[2], Ashlesha A. Ghute[3], Priya P. Rajput[4], Mahesh S. Shinde[5]

[1-4]*Student, M. E. S. College of Engineering Pune, Maharashtra, India*
[5]*Professor, Dept. of Computer Engineering, M. E. S. College of Engineering Pune, Maharashtra, India*

Abstract: *Agriculture is the heart of many countries and soil is the main important element of agriculture. There are different soil kinds and each kind has different features for different crops. In this field, now a day's different methods and models are used to increase the quantity of the crops. So the main purpose of this system is to create a model that helps farmers to know which crop should take in a particular type of soil. In this system, we are using machine learning techniques which help to suggest the crops according to soil classification or soil series. The model only suggests soil type and according to soil type it can suggest suitable crops. In this, different classifiers are used and according to that the model suggests the crop.*

Key Words: Soil Series, Machine Learning, Data Mining, Chemical Features, Prediction

1. INTRODUCTION

Agriculture is the main source of the Indian Economy. India is also known as a country of farmers. In India 50% man force is involved in agriculture activities. In Agriculture the soil is the main and basic thing [1]. But now also the farmers are using the traditional method. Because of the traditional method farmers did not get satisfactory results means the quantity of crops is not increasing. To increase the quantity of crops need good quality of soil. So soil testing is done. Soil testing is important rather than the main task of farming. The production and quality of crops totally depend on the soil. Soil testing is essential because it gives information on all nutrients which present in the soil such as Ca (Calcium), K (Potassium), and N (Nitrogen). Farmers in India, especially many regions in Maharashtra state faces drought due to which their crop and yielding are getting degraded. They don't have any idea about the availability of nutrients in their field. They use their own experience to plow the crop which has very less success ratio. Due to less success ratio they are unable to pay their loan amount sanctioned for their crop. In unsuccessful for their repayment of the loan amount they attempt to suicide. It is the main reason for increasing suicide cases [2].

To help the farmers to decide the crop to be plow for their benefits we motivated to build this system. This dataset consists of the available nutrient for farmers' soil and rainfall for a particular region. Based on nutrients value, our system predicts soil type. According to the soil type system predicts a list of crops that can grow in a particular soil. Hence the yield of the crop increases, as well as the farmer, earn more money with this new method. We create the system with the help of advanced technology. We use machine learning to create the system. Machine learning concentrates on the creation of computer programs that can access data and use it to learn from that. Machine learning allows building models from sample data and give the ability to take decision automatically according to past experiences [2].

2. LITERATURE REVIEW

In this section, we review some of the significant works done in the agriculture field for crop prediction. In [3], the authors concentrated on the use of applications of data mining techniques in the agricultural field. As data mining is a new rising technology so authors also study and examined the problem of forecasting agricultural productivity. The authors discussed the main objective of this work was finding the desired data models that give high accuracy and high generality in terms of the yield forecasting capabilities. For these different types of data mining techniques were judged on different data sets by authors.

In [4], the authors proposed a method named Crop Selection Method (CSM). The authors specify the proposed method to help to find out the crop selection problems and also help to raise the net yield rate of crop over seasons and help to get maximum economic growth. The authors discussed the different influencing parameters which can be used for crops by different forecasting models. The authors also specify the machine learning and different methods of machine learning. The

proposed Crop Selection Method classified crops as seasonal crops, whole year crops, short time plantation crops, long time plantation crops.

In [5], the authors suggested a smart way to forecast the crop yield and also suggested the ideal climatic factors to maximize the crop yield. In this the multivariate polynomial regression, support vector machine regression, and random forest models were used to forecast the crop yield per acre. The suggested method also uses yield and weather data which was collected by the author from the United States Department of Agriculture. The author also used the Mean Absolute Error (MAE), Root MeanSquared Error (RMSE), median absolute error and R-Square values to compare between multivariate polynomial regression, support vector machine regression and random forest.

In [6], the authors used the sliding window non-linear regression method to suggest crop yield and price by evaluating patterns from the past data. The authors studied several districts of the state of Tamilnadu. The authors proposed a system designed in such way that it suggests the best crop choices for a farmer to do farming. The designed system had done demand level classification. The demand level classification means the demand for the crops is predicted by classifying the dataset of change in the market prices of crops. In the proposed system the authors also did the text to speech conversion.

In [7], authors have proposed an ideal system called AgroConsultant, which aims to help the Indian farmers to decide about which crop to grow depending on the season, geographical location, and soil characteristics. For making such a system the authors used a machine-learning algorithm like the random forest, K-Nearest Neighbors (K-NN), Decision Tree, Neural Network. The proposed system also includes the map visualization feature and rainfall predictor.

In [8], the authors created an innovative structure named as eXtensible Crop Yield Prediction Framework (XCYPF). This framework provides the selection of crop and also dependent and independent variables and datasets for crop yield forecasting. The authors discussed how this framework is adaptable and expandable. The authors used rainfall data and surface temperature for crop yield forecasting for rice and sugarcane crops. The authors used the new methodology which combines the use of vegetation indices.

In [9], authors designed a crop yield prognosis model (CRY) which works on an adaptive cluster approach. It uses the Bee Hive modeling approach to study and classify the crop based on crop growth pattern and yield. The authors also described the Bee Hive Cluster which was used because it describes the agricultural datasets which help to decide crop growth and the Bee Hive Cluster is considered as heterogeneous data cluster which maintained as repositories. The authors studied the Bee Hive which had better performance than others. This Bee Hive algorithm is used for getting patterns and the graphs were plotted to describe yield variation.

In [10], authors built a cloud-based agricultural framework which helps the Indian farmers and agricultural industries to get useful information about agriculture. This framework provides soil classification and crop yield forecasting. The authors use a hybrid support vector machine for soil classification and customized artificial neural network to forecast wheat yield. Also the authors used Amazon S3 to store agricultural data and they used the Heroku cloud.

3. METHODOLOGY

Here we have built soil classification and crop prediction system with machine learning methods. We have used two datasets. The first dataset is a soil dataset and a crop dataset. Soil dataset has attributes like PH, electrical conductivity, nitrogen, phosphorus, potassium, soil type and crop dataset has attributes like soil type and crop list. Soil dataset is divided into two sets namely training set and testing set. 70% of data has been used for training the model and 30% of the data used for testing the performance of the model.

We have designed a web-based system to predict the crop based on soil classification. The main aim of us is to build a user-friendly system based on soil parameters. Data preprocessing is a data mining technique that is used to convert the raw data into a useful format. There are several data mining techniques. We have done data preprocessing by data cleaning, data integration, data transformation, data reduction. Data cleaning can be performed by removing noisy data and correct inconsistency in data. Data is taken from multiple resources. Data integration is used to merge the data from multiple resources into consistent data. Data transformation is the process of normalizing the data. Data reduction can reduce the size of data by aggregation, eliminating redundant features.

The five algorithms we have used are Support Vector Machine, Bagged Tree, Adaboost, Naive Bayes, and Artificial Neural Network. We are increasing the accuracy of our system by using an ensemble method. Ensemble model generates more accurate results than the base model. The ensemble method is a combination of several base models to predict a more accurate result and Fig.1 describes the ensemble method. Each base model predicts the class label, from which we get the majority voting and suggest crops to the farmer, which should be sown in the field.

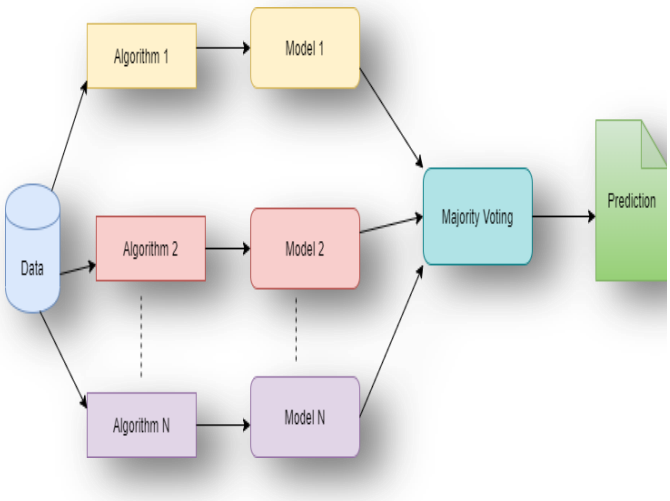

Fig - 1: Ensemble Method

The overview of the system:

I. Input :

The prediction of crop is dependent on soil parameters such as PH, electrical conductivity, nitrogen, phosphorus, potassium, soil type to predict crop accurately. The farmer provides the soil parameters to the system.

II. Data Preprocessing:

Data obtained from various sources are in the form of raw data. Data is generally in raw form, we want to convert it to useful form by using data preprocessing. It Consist of redundant, incomplete, inconsistent data. So in Data preprocessing raw data is converted into normalized form.

III. Classification Algorithm

Data classification takes place in two steps, The first step is to create a classification model and the second step where the model is used to predict the class label for given data. The five classification algorithms used by us are Support Vector Machine, Bagged Tree, Adaboost, Naive Bayes, and Artificial Neural Network.

A. Support Vector Machine (SVM):

Support Vector Machine (SVM) is a supervised machine learning algorithm. The Support Vector Machine (SVM) is used for classification and regression but the Support Vector Machine (SVM) is mostly used for the classification [11]. The Support Vector Machine (SVM) algorithm provides a high accuracy rate so that we also used the Support Vector Machine algorithm. In this technique every data item is represented as a point in an N-dimensional space and hyper plane is constructed that separates the data points into different classes and then this hyper plane is used for classification. The hyper plane will divide the datasets into different classes as positive and negative.

B. Artificial Neural Network (ANN):

The Artificial Neural Network (ANN) is a Neural Network which is a computing system. The Artificial Neural Network (ANN) is a simulated form of biological neural networks that present in the brains of humans and animals. The artificial neurons are the basic unit of Artificial Neural Network (ANN). The artificial neuron is a collection of connected units. In the biological brain the synapses can transmit the signal, just like that in Artificial Neural Network the artificial neuron can receive and process the signals. In Artificial Neural Network, the artificial neuron is considered as a node and the connection between two nodes is considered as an edge. The neurons and edges have some weights. The weight can increase or decrease according to the strength of the signal. The neurons consist of some threshold which is used to fire the neurons. If the total signal crosses the threshold then we said the neuron is fired means the signal is sent to other neurons [12].

C. Naive Bayes:

Naive Bayes algorithm is a classification technique in machine learning and it is also called as probabilistic classifiers. The Naive Bayes algorithm is a supervised machine learning algorithm. The Naive Bayes uses the Bayes' theorem with a strong independence belief between the features. The Naive Bayes is the simplest Bayesian Network Model [13].

The Bayes' theorem is stated as [14]:

$$P(h|d)=(P(d|h) * P(h)) / P(d)$$

Where,

P(h|d) is a probability of hypothesis h given the data d.

P(d|h) is a probability of data d given that the hypothesis h was true

P(h) is a probability of hypothesis h being true

P(d) is a probability of the data

D. AdaBoost:

AdaBoost is best used to drive the efficiency of decision making on binary classification problems. The AdaBoost is one type of boosting algorithm. The boosting is used to increase the performance of the system. More recently it may be raise to as discrete AdaBoost because it is used for classification rather than regression. AdaBoost can be used to improve the performance of any machine learning algorithm. It can be used to learn weak classifiers. These are models that achieve the most accurate accuracy of a random opportunity on a classification problem. It can be used in conjunction with many other types of learning algorithms to improve their performance. Initially, all weights are set in equal proportions, but the weight of each component classified incorrectly is increased so that the previous classifier who did not predict gained more weight on the next iteration [15].

E. Bagged Tree:

We have used a bagged tree which also known as bagged decision tree ensemble classifier. Bagging is a method through which the performance of the system can be improved statistically. Bagging generates a set of models which trained on a random sampling of the data. The suggestions from the models are combining to produce the final forecasting of crop [1].

IV. Majority Voting:

Ensemble techniques that are used to create multiple models and then combine them to produce an accurate result. Every model predicts the class label for each instance and depending upon the majority that class label is selected.

V. Result:

A Farmer's decision generally clouded by his intuitions about which crops to grow in the field. Crops are suggested to the farmer based on the type of soil which helps the farmer to increase Economic situation. These steps describe the system architecture. The Fig.-2 described the whole system architecture.

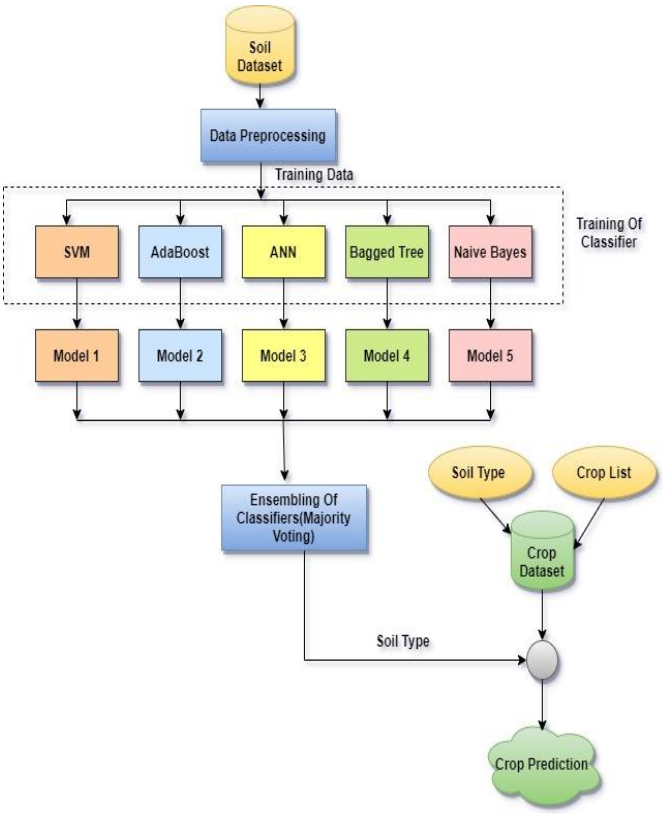

Fig.-2: System Architecture

Result Analysis:

The system that we have built is based on the two datasets as well as the different algorithms are used to increase the performance of the system. We recognized the result from the applying algorithms. The accuracy of each algorithm is given in the following table

Table -1: Result

Algorithm	Accuracy (%)
Artificial Neural Network(ANN)	92
Support Vector Machine(SVM)	98.5
Bagged Tree	88
Naive Bayes	76
AdaBoost	66

4. CONCLUSION AND FUTURE SCOPE

After analyzing many past systems on crop prediction, this system suggest crop based on soil classification with ensembling classifiers system has been created. The Artificial Neural Network (ANN), Bagged Tree, Naive Bayes, Adaboost, and Support Vector Machine (SVM) algorithms are combined to improve the accuracy of the system which gives the list of appropriate crop according to the soil type. In the future with advanced classification algorithms and techniques, the accuracy of the system can be increased with various datasets. The system can predict the crop, based on soil parameters. In the future, the location recommendation module can be added according to the crop suggestion means according to suggest crop the appropriate location will be suggested.

REFERENCES

[1] Sk Al Zaminur Rahman, S.M. Mohidul Islam, Kaushik Chandra Mitra," Soil Classification using Machine Learning Methods and Crop Suggestion Based on Soil Series",2018 21st International Conference of Computer and Information Technology (ICCIT), 21-23 December, 2018.

[2] S. Panchamurthi. M.E.,M.D. Perarulalan,A. Syed Hameeduddin,P. Yuvaraj,"Soil Analysis and Prediction of Suitable Crop for Agriculture using Machine Learning", International Journal for Research in Applied Science & Engineering Technology(IJRASET),ISSN:2321-9653;IC Value:45.98;SJ Impact Factor:6.887,Volume 7 Issue III,Mar 2019.

[3] D Ramesh,B Vishnu Vardhan,"Data Mining Techniques and Applications to Agricultural Yield Data," International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 9, September 2013.

[4] RakeshKumar ,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), Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, T.N., India. 6 - 8 May 2015. pp.138-145.

[5] AyushShah,AkashDubey,Vishesh Hemnani,Divye Gala and D. R. Kalbande," Smart Farming System: Crop Yield Prediction Using Regression Techniques,"Springer Nature Singapore Pte Ltd.2018H.Vasudevan

etal.(eds.),Proceedings of International Conference on Wireless Communication, Lecture Notes on Data Engineering and CommunicationsTechnologies 19.
[6] S. Kanaga Suba Raja,Rishi R.,Sundaresan E.,Srijit V.,"Demand Based Crop Recommender System for Farmers",2017 IEEE International Conference on Technological Innovations in ICT For Agriculture and Rural Development(TIAR2017), 978-1-5090-4437-5/17/$31.00.
[7] ZeelDoshi, RashiAgrawal, SubhashNadkarni,Prof. Neepa Shah,"AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms,"2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA),978-1-5386-5257-2/18/$31.00.
[8] Aakunuri Manjula,Dr.G.Narsimha," XCYPF: A Flexible and Extensible Framework for Agricultural Crop Yield Prediction,"IEEE Sponsored 9th International Conference on Intelligent Systems and Control (ISCO)2015, 978-1-4799-6480-2/15/$31.00.
[9]M.Gunasundari,Dr.T.Arunkumar,Ms.R.Hemavathy,"CR Y-An improved Crop Yield Prediction model using Bee Hive Clustering Approach for Agricultural data sets," Proceedings of the 2013 International Conference on Pattern Recognition,Informatics and Mobile Engineering, February 21-22, 978-1-4673-5845-3/13/$31.00.
[10]K.Aditya Shastry and H.A.Sanjay,"Cloud-Based Agricultural Framework for Soil Classification and Crop Yield Prediction as a Service,"Springer Nature Singapore Pte Ltd. 2019 N. R. Shetty et al. (eds.), Emerging Research in Computing, Information, Communication and Applications, Advances in Intelligent Systems and Computing 882.
[11] Support Vector Machine Available At - https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/
[12] Artificial Neural Network Available At - https://en.wikipedia.org/wiki/Artificial_neural_network
[13] Naive Bayes Available At - https://en.wikipedia.org/wiki/Naive_Bayes_classifier
[14] Naive Bayes Available At - https://machinelearningmastery.com/naive-bayes-for-machine-learning/
[15] Adaboost Available At - https://machinelearningmastery.com/boosting-and-adaboost-for-machine-learning/