

Crop Recommendation System to Maximize Crop Yield using Machine Learning Technique

Rohit Kumar Rajak¹, Ankit Pawar², Mitalee Pendke³, Pooja Shinde⁴, Suresh Rathod⁵,
Avinash Devare⁶

¹²³⁴⁵⁶Dept. of Computer Engineering, Sinhgad Academy of Engineering, Maharashtra, India

Abstract - Agriculture in India plays a major role in economy and employment. The common difficulty present among the Indian farmers are they don't opt for the proper crop based on their soil necessities. Because of this productivity is affected. This problem of the farmers has been solved through precision agriculture. This method is characterized by a soil database collected from the farm, crop provided by agricultural experts, achievement of parameters such as soil through soil testing lab dataset. The data from soil testing lab given to recommendation system it will use the collect data and do ensemble model with majority voting technique using support vector machine (SVM) and ANN as learners to recommend a crop for site specific parameter with high accuracy and efficiency.

Key Words: Precision agriculture, recommendation system, ensemble model, SVM, ANN, RANDOM TREE, NB-classifier.

1. INTRODUCTION

India is characterized by small farms. Over 75% of total land capitals within the country are less than 5 acres. Most crops are rain nourished, with just about 45% of the land irrigated. As per some estimations, about 55% of total population of India depends on farming. In the US, because of heavy mechanization of agriculture, it is about 5%.

India is one of the biggest producers of agricultural products and still has very less farm productivity. Productivity needs to be increased so that farmers can get more pay from the same piece of land with less labour. Precision agriculture provides a way to do it. Precision farming, as the name implies, refers to the applying of precise and proper total of comment like pee, fertilizers, soil etc. at the proper time to the crop for increasing its productivity and increasing its yields. Not all precision agriculture systems offer best results. But in agriculture it is important that the recommendations made are accurate and precise because in case of errors it may lead to heavy material and capital loss. Many researches are being carried out, in order to attain an accurate and efficient model for crop prediction. Ensembling is one such technique that is included in such research works. Among these various machine learning techniques that are being used in this field; this paper proposes a system that uses the voting method to build an efficient and accurate model.

2. LITERATURE SURVEY

The paper [1] states the requirements and planning needed for developing a software model for precision farming is discussed. It deeply studies the basics of precision farming. The author's start from the basics of precision farming and move towards developing a model that would support it. This paper describes a model that applies Precision Agriculture (PA) principles to small, open farms at the individual farmer and crop level, to affect a degree of control over variability. The comprehensive objective of the model is to deliver direct advisory services to even the smallest farmer at the level of his/her smallest plot of crop, using the most accessible technologies such as SMS and email. This model has been designed for the scenario in Kerala State where the average holding size is much lower than most of India. Hence this model can be positioned elsewhere in India only with some modifications.

The paper [2] makes a qualified cogitation of assortment algorithms and their performance in yield prediction in precision husbandry. These algorithms are implemented in a data set collected for several years in yield prediction on soya bean crop. The algorithms used for yield prediction in this paper are Support Vector Machine, Random Forest, Neural Network, REPTree, Bagging, and Bayes. The conclusion drawn at the end is that bagging is the best algorithm for yield prediction among the above stated algorithms since the error deviation in bagging is minimum with a mean absolute error of 18985

The paper [3] shows the importance of crop selection and the factors deciding the crop selection like production rate, market price and government policies are discussed. This paper proposes a Crop Selection Method (CSM) which solves the crop selection problem and improves net yield rate of the crop. It suggests a series of crop to be selected over a season considering factors like weather, soil type, water density, crop type. The predicted value of influential parameters determines the accuracy of CSM. Hence there is a need to include a prediction method with improved accuracy and performance.

The paper [4] aims to solve the crucial problem of selecting the classifiers for the ensemble learning. A method to select a best classifier set from a pool of classifiers has been proposed. The proposal aims to achieve higher accuracy and performance. A method called SAD

was proposed based on accuracy and classification performance. Using Q statistics, the dependency between most relevant and accurate classifiers is identified. The classifiers which were not chosen were combined to form the ensemble. This measure is supposed to ensure higher performance and diversity of the ensemble. Various methods such as SA (Selection by Accuracy), SAD (Selection by accuracy and Diversity) and NS (No selection) algorithm were identified. Finally it is inferred that SAD works better than others.

The paper [5] proposes various classification methods to classify the liver disease data set. The paper emphasizes the need for accuracy because it depends on the dataset and the learning algorithm. Classification algorithms such as Naïve Bayes, ANN, ZeroR and VFI were used to classify these diseases and compare the effectiveness, correction rate among them. The performance of the models were compared with accuracy and computational time. It was concluded that all the classifiers except naive bayes showed improved predictive performance. Multilayer perceptron show the highest accuracy among the proposed algorithms.

The paper [6] tries to solve the problem of food insecurity in Egypt. It proposes a framework which would predict the production, and import for that particular year. It uses Artificial Neural Networks along with Multi-layer perceptron in WEKA to build the prediction. At the end of the process we would be able to visualize the amount of production import, need and availability. Therefore it would help to make decisions on whether food has to be further imported or not.

The soil datasets in paper [7] are analyzed and a category is predicted. From the predicted soil category the crop yield is identified as a Classification rule. Naïve Bayes and KNN algorithms are used for crop yield prediction. The future work stated is to create efficient models using various classification techniques such as support vector machine, principal component analysis.

3. METHODOLOGY

3.1 Dataset Collection

The dataset containing the soil specific attributes which are collected from Polytest Laboratories soil testing lab, Pune, Maharashtra, India. In addition, similar sources of general crop data were also used from Marathwada University. The crops considered in our model include groundnut, pulses, cotton, vegetables, banana, paddy, sorghum, sugarcane, coriander. The number of examples of each crop available in the training dataset is shown. The attributes considered were Depth, Texture, Ph, Soil Color, Permeability, Drainage, Water holding and Erosion.

The above stated parameters of soil play a major role in the crop's ability to remove water and nutrients from the

soil. For crop growth to their possible, the soil must provide acceptable environment for it. Soil is the anchor of the roots. The water holding capacity determines the crop's ability to absorb nutrients and other nutrients that are changed into ions, which is the form that the plant can use. Texture determines how porous the soil is and the comfort of air and water movement which is essential to prevent the plants from becoming waterlogged. The level of acidity or alkalinity (Ph) is a master variable which affects the availability of soil nutrients. The activity of microorganisms present in the soil and also the level of exchangeable aluminum can be affected by PH. The water holding and drainage determine the infiltration of roots. Hence for the following reasons the above stated parameters are considered for choosing a crop.

3.2 Crop Prediction using Ensembling technique

Ensemble is a data mining model also known as the Model Combiners that combine the power of two or more models to attain better prediction, efficiency than any of its models could achieve alone. In our system, we use one of the most familiar Ensembling technique called Majority Voting technique. In the voting technique any number of base learners can be used. There should be at least two base learners. The learners are chosen in a way that they are capable to each other yet being complimentary also. Higher the competition higher is the chance of better prediction. But it is necessary for the learners to be complimentary because when one or few members make an error, the probability of the remaining members correcting this error would be high. Each learner builds itself into a model. The model gets trained using the training data set provided. When a new data has to be classified, each model predicts the class on its own. Finally, the class which is predicted by majority of the learners is voted to be the class label of the new sample.

3.3 Learners Used in the Model

3.3.1 Support Vector machine

Support vector machines (SVM) is set of supervised learning strategies used for classification, regression and outlier's discovery. it's a classification technique. Here, we have a tendency to plot every information item as some extent in n-dimensional house (where n is variety of options you have) with the worth of every feature being the worth of a selected coordinate. it's a classification technique. during this algorithmic rule, we have a tendency to plot every information item as some extent in n-dimensional house (where n is variety of options you have) with the worth of every feature being the worth of a selected coordinate. A Support Vector Machine (SVM) is discriminative classifier correctly bounded by a separating hyperplane. In alternative words, given labeled coaching information (supervised learning), the algorithmic rule outputs associate degree best hyperplane that categorizes new examples. Support vector simple machine (SVM) may

be a set of supervised learning strategies used for classification, regression and outlier's uncovering.

3.3.2 NAÏVE Bayes

It is not single algorithm, but a clan of algorithmic rules. All naive Bayes socio-economic classifiers adopts that the value of a particular feature is independent of the value of any other feature, given the class variable. Naive Thomas Bayes classifier could be a straightforward probabilistic classifier that works supported applying theorem (from Bayesian statistics) with robust naive independence assumptions. it's a classification technique supported Bayes' theorem with associate degree assumption of independence between predictors. In straightforward terms, a Naive Thomas Bayes categoryifier assumes that the presence of a specific feature in an exceedingly class is dissimilar to the presence of the other feature. as an example, a fruit could also be thought of to be associate degree apple if it's red, round, and concerning a pair of inches in diameter. Even if these features depend on each other or upon the existence of the other features, a naive Bayes classifier would consider all of these properties to independently contribute to the probability that this fruit is an apple. These Learners predict the class label for each of the training data set. The class label that is predicted by the majority of the models is voted through the majority voting technique and the class label of the training data set is decided. From the ensembled models the rules are generated.

3.3.3 Multi-layer Perceptron (Artificial Neural Network)

Multi Layer perceptron (MLP) could be a feed forward neural network with several layers between input and end product layer. Feed forward implies that knowledge period in one direction from input to output layer. MLPs square measure typically used for pattern classification, recognition, anticipation and approximation. Associate ANN relies on a set of connected units or nodes referred to as artificial neurons (analogous to biological neurons in associate animal brain). every association (synapse) between neurons will transmit an indication from one to a different. The receiving (postsynaptic) somatic cell will method the signal(s) then signal neurons connected to that. Artificial neural networks (ANNs) or connectionist systems square measure computing systems galvanized by the biological neural networks that represent animal brains

In Neural Networks some nodes use a nonlinear activation operate that was developed to model the frequency of action potentials, or firing, of biological neurons. for instance, in image recognition, they could learn to spot pictures that contain cats by analyzing example pictures that are manually labelled as "cat" or "no cat" and exploitation the results to spot cats in different pictures. they are doing this with none a priori information

concerning cats, e.g., that they have fur, tails, whiskers and cat-like faces. Instead, they evolve their own set of relevant characteristics from the learning material that they process.

3.3.4 Random Forest

Random forests square measure associate ensemble learning methodology for classification, regression and different tasks, that operate by building a mess of call trees at coaching time and outputting the category that's the mode of the categories or mean prediction of the individual trees. Random call forests correct for call tree custom of over fitting to their coaching set. the primary rule for random call forests was created by Tin Kam Ho victimization the random mathematical space methodology, which, in Ho's formulation, could be a thanks to implement the "stochastic discrimination" approach to classification projected by Eugene Kleinberg. associate addition of the rule was developed by Leo Breiman and Adele bargainer, and "Random Forests" is their trademark. The addition combines Breiman's "bagging" plan and random choice of options, introduced 1st by Ho and later severally by Amit and Geman so as to construct a group of call trees with controlled variance. though random forests are naturally designed to figure solely with third-dimensional information, it's been shown that one may use them for random objects with the employment of solely pairwise similarities between objects.

Random Forest could be a trademark term for associate ensemble of call trees. Random Forest is assortment of call trees (so called Forest). To classify a replacement object supported attributes, every tree offers a categoryification and that we say the tree votes for that class.

3.4 Rules induced from the Model

The rule below demonstrates an example of the proposed recommendation system.

IF ph is mild alkaline
AND depth is above 90
AND water holding capacity is LOW
AND drainage is moderate
AND erosion is LOW
THEN PADDY

The IF part of the rule states the soil specifications needed for the cultivation of the recommended crop which is specified in the THEN PART of the rule.

4. CONCLUSION

Our work would help farmers to increase productivity in agriculture, prevent soil degradation in cultivated land, and reduce chemical use in crop production and efficient use of water resources. Our future work is aimed at an

improved data set with large number of attributes and also implements yield prediction.

REFERENCES

- [1] Satish Babu (2013), 'A Software Model for Precision Agriculture for Small and Marginal Farmers', at the International Centre for Free and Open Source Software (ICFOSS) Trivandrum, India.
- [2] Anshal Savla, Parul Dhawan, Himtanaya Bhadada, Nivedita Israni, Alisha Mandholia, Sanya Bhardwaj (2015), 'Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture', Innovations in Information, Embedded and Communication systems (ICIIECS).
- [3] Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh (2015), 'Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique', International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM).
- [4] Liying Yang (2011), 'Classifiers selection for ensemble learning based on accuracy and diversity' Published by Elsevier Ltd. Selection and/or peer-review under responsibility of [CEIS].
- [5] A.T.M Shakil Ahamed, Navid Tanzeem Mahmood, Nazmul Hossain, Mohammad Tanzir Kabir, Kallal Das, Faridur Rahman, Rashedur M Rahman (2015), 'Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in Bangladesh', (SNPD) IEEE/ACIS International Conference.
- [6] Aymen E Khedr, Mona Kadry, Ghada Walid (2015), 'Proposed Framework for Implementing Data Mining Techniques to Enhance Decisions in Agriculture Sector Applied Case on Food Security Information Center Ministry of Agriculture, Egypt', International
- [7] Monali Paul, Santosh K. Vishwakarma, Ashok Verma (2015), 'Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach', International Conference on Computational Intelligence and Communication Networks