International Journal of Research in Engineering, Science and Management
Volume-1, Issue-6, June 2018
www.ijresm.com

Implementation of Machine Learning Algorithms for Crop Recommendation Using Precision Agriculture

K. R. Akshatha[1], K. S. Shreedhara[2]

[1]*Student, Department of Studies in Computer Science & Engineering, University BDT College of Engg., Davanagere, India*
[2]*Professor, Department of Studies in Computer Science & Engineering, University BDT College of Engg., Davanagere, India*

***Abstract*—Data mining is the practice of examining and deriving purposeful information from the data. Data mining finds its application in various fields like finance, retail, medicine, agriculture etc. Data mining in agriculture is used for analyzing the various biotic and abiotic factors. Agriculture in India plays a predominant role in economy and employment. The common problem existing among the Indian farmers are they don't choose the right crop based on their soil requirements. Due to this they face a serious setback in productivity. This problem of the farmers has been addressed through precision agriculture. Precision agriculture is a modern farming technique that uses research data of soil characteristics, soil types, crop yield data collection and suggests the farmers the right crop based on their site specific parameters. This reduces the wrong choice on a crop and increase in productivity. In this paper, this problem is solved by proposing a recommendation system through an ensemble model with majority voting technique using Random tree, K-Nearest Neighbor and Naive Bayes as learners to recommend a crop for the site specific parameters with high accuracy and efficiency.**

Index Terms—Precision agriculture, Recommendation system, Ensemble model, Majority Voting technique, Random tree, K-Nearest Neighbor and Naïve Bayes

I. Introduction

India is one among the oldest countries which is still practicing agriculture. But in recent times the trends in agriculture has drastically evolved due to globalization. Various factors have affected the health of agriculture in India. Many new technologies have been evolved to regain the health. One such technique is precision agriculture. Precision agriculture is budding in India Precision agriculture is the technology of "site-specific" farming. It has provided us with the advantage of efficient input, output and better decisions regarding farming. Although precision agriculture has delivered better improvements it is still facing certain issues. There exist many systems which propose the inputs for a particular farming land. Systems propose crops, fertilizers and even farming techniques. Recommendation of crops is one major domain in precision agriculture. Recommendation of crops is dependent on various parameters. Precision agriculture aims in identifying these parameters in a site-specific manner in order to resolve issues regarding crop selection. The "site-specific" technique has improved the results yet there is a need to supervise the results of such systems. Not all precision agriculture systems provide accurate 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 research works is 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.

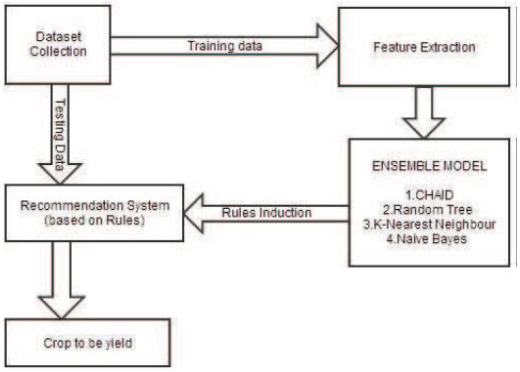

Fig. 1. Architecture of proposed system

In the proposed scheme, Ensemble is a data mining model also known as the Committee Methods or Model Combiners that combine the power of multiple models to acquire greater 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 has to be at least two base learners. The learners are chosen in a way that they are competent 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 sample 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. This method is implemented in Rapid miner tool depicts the process implemented in rapid miner.

II. Literature Survey

The paper [1] states the requirements and planning needed for developing a software model for precision farming is

discussed. It deeply analyses 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 deployed elsewhere in India only with minor modifications. The paper [2] makes a comparative study of classification algorithms and their performance in yield prediction in precision agriculture. 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, REP Tree, 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. 7864. The paper [3] states the necessity for crop yield prediction and its help in a nation's strategic policy making in agriculture. A framework extensible Crop Yield Prediction Framework (XCYPF) is developed. It facilitates flexible inclusion of various techniques towards crop yield prediction. A tool was also developed that would help people to predict crop yield for various crops with dependant and independent variables.

III. Proposed System

The dataset comprising the soil specific attributes which are collected and tested at soil testing lab. In addition, similar online sources of general crop data were also used. The crops considered in our model include millet, groundnut, pulses, cotton, vegetables, banana, paddy, sorghum, sugarcane, coriander. Figure 1 gives an analysis of the dataset. The number of instances of each crop available in the training dataset is depicted. The attributes considered where 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 extract water and nutrients from the soil. For crop growth to their fullest potential, the soil must provide a satisfactory 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. Soil texture which affects the soil's ability to hold onto nutrients. 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 penetration of roots. Hence for the following reasons the above stated parameters are considered for choosing a crop.

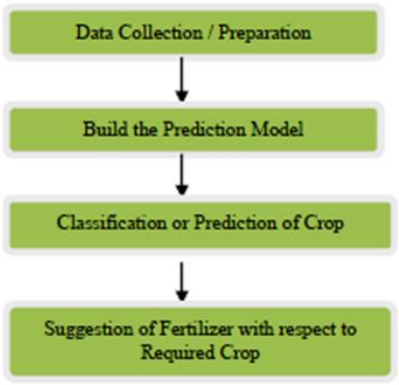

Fig. 2. Data flow diagram

Learners Used in the Model:

A. K-nearest Neighbor

K-Nearest Neighbor can be used for both classification and regression. K-Nearest Neighbors is a non-complex algorithm which stores all the available cases and classifies new cases based on some similarity measure .The sample set is classified based upon the "closeness" that is the distance measure such as Euclidean distance or Manhattan distance.

B. Naïve Bayes:

Naive Bayes classifier is a simple probabilistic classifier which works based on applying Bayes' theorem (from Bayesian statistics) with strong naive independence assumptions. Naive Bayes is a technique for constructing classifier models which assign class labels to problem instances which are represented as vectors of feature values, where the class labels are drawn from some finite set. It is not just a single algorithm for training such classifiers, but a family of algorithms based on a common principle. All naive Bayes classifiers assumes that the value of a particular feature is independent of the value of any other feature, given the class variable. 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.

IV. Conclusion

India is a nation in which agriculture plays a prime role. In prosperity of the farmers, prospers the nation. Thus our work would help farmers in sowing the right seed based on soil requirements to increase productivity and acquire profit out of such a technique. Thus the farmers can plant the right crop increasing his yield and also increasing the overall productivity of the nation. Our future work is aimed at an improved data set with large number of attributes and also implements yield prediction.

REFERENCES

[1] S. Babu, "A software model for precision agriculture for small and marginal farmers," *2013 IEEE Global Humanitarian Technology Conference: South Asia Satellite (GHTC-SAS)*, Trivandrum, 2013, pp. 352-355.

[2] A. Savla, N. Israni, P. Dhawan, A. Mandholia, H. Bhadada and S. Bhardwaj, "Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture," *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, Coimbatore, 2015, pp. 1-7.

[3] A. Manjula and G. Narsimha, "XCYPF: A flexible and extensible framework for agricultural Crop Yield Prediction," *2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO)*, Coimbatore, 2015, pp. 1-5.

[4] Y. Sanghvi, H. Gupta, H. Doshi, D. Koli, A. Ansh and U. Gupta, "Comparison of Self organizing maps and Sammon's mapping on agricultural datasets for precision agriculture," *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, Coimbatore, 2015, pp. 1-5.

[5] R. Kumar, M. P. Singh, P. 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)*, Chennai, 2015, pp. 138-145.