Implementation of Crop Yield Forecasting Using Data Mining

Pooja M C[1], Sangeetha M[2], Shreyaswi J Salian[3], Veena Kamath[4], Mithun Naik[5]

[1234]Dept of CSE, Srinivas school of engineering, Mangalore. Karnataka, India
[5]Dept of CSE, Srinivas school of engineering, Mangalore. Karnataka, India

--***---

Abstract - *India is basically an agricultural country. In this paper an attempt has been made to study on applications of data mining techniques in the agricultural field. Technique J48of C4.5 algorithm is used. In this module we are building data mining techniques which uses past information like soil type, soil pH, ESP, EC of a particular region to give better crop yield estimation for that region. This model can be used to select the most excellent crops for the region and also its yield there by improving the values and gain of farming also. This aids farmers to decide on the crop they would like to plant for the forthcoming year. Prediction will help the associated industries for planning the logistics of their business.*

*Key Words***:** Data mining, J48 of C4.5 algorithm, productivity, soil type, prediction.

1. INTRODUCTION

Agriculture is the main occupation and the crop production is a difficult phenomenon that is recommended by agriculture input parameters input. Agriculture input parameters changes according to the farmers and the fields. Data mining is the removal of unseen predicted information from big databases. It is the modern technology which gives the information about the data to help companies concentrate on most valuable data. Data mining helps in upcoming trends and character, allowing businesses to gain knowledge-driven decisions. Agriculture and allied activities compose the single largest component of India's gross domestic product, giving nearly 35% of the total and nearly 65% of Indian population depends on this profession. Due to varies of weather factors the agricultural in India are continuously reducing over a decade. The reasons for this were studied mostly using regression analysis. In this paper an attempt has been made to compile the research findings of different researchers who used data. Data mining is the collection of large data to infer important knowledge. (Fig 1) shows that   knowledge of data mining process is divided into seven methods

- a) Data cleaning
- b) Data integration
- c) Data selection
- d) Data transformation
- e) Data mining
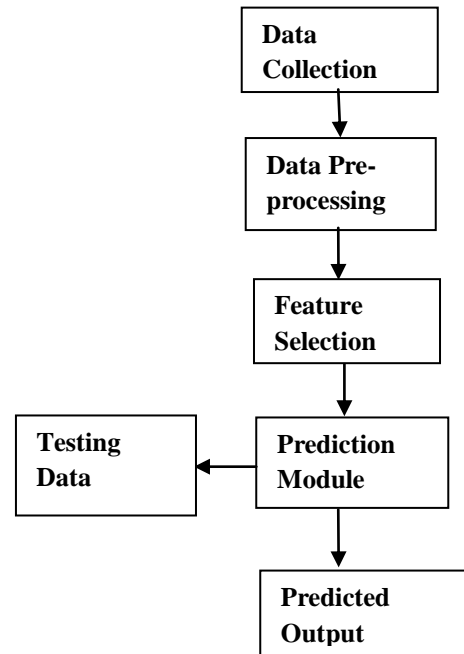- f) Pattern estimation
- g) Knowledge display

Fig -1: Methodology Description

Most of the India's land areas are cultivable. India is the second leading nation in terms of total cultivable land. Agriculture products includes rice, potato, onion, sugarcane, beans, cotton, wheat, tomato etc. India is the biggest producer of sugarcane. Majority of the farmer is not receiving the estimated crop yield due to many reasons. India is the second highest manufacture of wheat and rice. India's agriculture is composed of many number of crops and food staples like rice and wheat. Agricultural products should have the important economic value. The yield of the agriculture depends on the weather conditions. Over sixty percent of the crop depends on the monsoon rainfall. In agriculture sector where farmers have to make innumerable decisions every day and intricate complexities involves the various factors influencing them. In the ancient times, yield of the crop forecast was achieved by the farmer's earlier familiarity on an analyzing the crop. Data volume is large in India[3]. Many problems arise in the crop production such as attack of pests and diseases. Pests and diseases are one of the main reason for the decrease in the crop yield. Due to changes in the weather conditions and lack of facilities is also the reason for reduction in the crop yield. Application of curative measures may decrease the yield loss. To get the better yield of crop one should have the knowledge of the time and the harshness of the outbreak of pests and diseases. Farmer should have the information of crop yield properly otherwise the reduction in the crop yield occurs.

2. IMPLEMENTAION

The proposed system involves a prediction module based on data mining classification algorithm namely J48 used to forecast the yield of major crops based on historical data. It also involves vision to forecast the yield of major crops based on historical data. The data and predicted output are accessible for the farmers through an application. This aids farmers to decide on the crop they would like to plant for the forthcoming year.

3. METHODOLOGY

There are two techniques in data mining they are classification and clustering. Analyzing of the data can be done by classification and prediction. Accuracy of prediction can be increased by classification model. There are three unique methods divided by the data mining algorithm such as supervised, unsupervised, and semi supervised learning. Classification techniques for finding knowledge that are Naive Bayes, J48, Random forest, Artificial Neural Network, Decision tree, Support Vector Machine. This paper is concentrated on J48 of C4.5 algorithm.

J48

The implementation of java called J48 of C4.5 algorithm is used. C4.5 is an algorithm which builds a decision tree using set of labeled data input. This algorithm is the continuation of ID3 algorithm utilized for diminishing of decision trees. This algorithm is frequently used as statistical classifier.

Features of the Algorithm

i.   Both the discrete and continuous are handled by this algorithm. A threshold value is decided by C4.5 for handling continuous attributes. This value divides the data list into those who have their attribute value below threshold and those having more than or equal to it.

ii. This algorithm also handles missing values in the training data.

iii. After the tree is fully constructed, this algorithm performs the pruning of the tree. C4.5 after its construction drives back through the tree challenges to remove branches that are not helping in reaching the leaf nodes.

iv. J48 is the open source Java implementation of the C4.5 algorithm.

v.   Decision tree is based on labeled input data.

vi. Trees generated can be used for classification and for this reason is called a statistical classifier.

vii. Decision tree is well known approach for classification in data mining. viii. The model is tested with test data set of soil samples.

DECISION TREE

Advantage of decision tree is that it is very easy and institutive to explain. Easy to implement and very fast. Disadvantage of decision tree is that they may cause from overfitting.

C4.5 Algorithm

C4.5 is an extension of ID3 algorithm.C4.5 is used for classification and statistical classifier.C4.5 which trains the data same way as ID3.Training the data is a set S=s1,s2,s3…..classified samples. Sample si consists of p dimensional vector(x1,I,x2,i,…….,xp, i),where xj represent attribute values, as well as si falls. At each node of the tree C4.5 chooses the attribute of the data. It splits its samples into subsets in class or other.C4.5 recurs on the low level sub lists.

ID3 algorithm

It has the original set S as the root node. It calculates the entropy H(S) it iterates through very attribute sets. It has the smallest entropy value. To produce the subset of the dataset s is spitted into the selected attributes Algorithm.

4. CROP YIELD PREDICTION

The various types of crop is measured and calculated in metric tons per hectare. The agriculture output would have a crop harvest of 1:4. Crop yield predictions includes an input module. The input consist of soil type, soil pH, seed type. The feature selection used for subset selection of crop details. The crop yield prediction is used to select which crop is suitable for soil type and based on this final prediction s done.

5. COMPARATIVE OF SOIL CLASSIFICATION

Considering local soil classification, should satisfy the needs of local users e.g. farmers, advisors. Here we use J48algorithm.The selected crop will be based on the soil type, pH etc. The final prediction of the crop yield depends on soil type used.
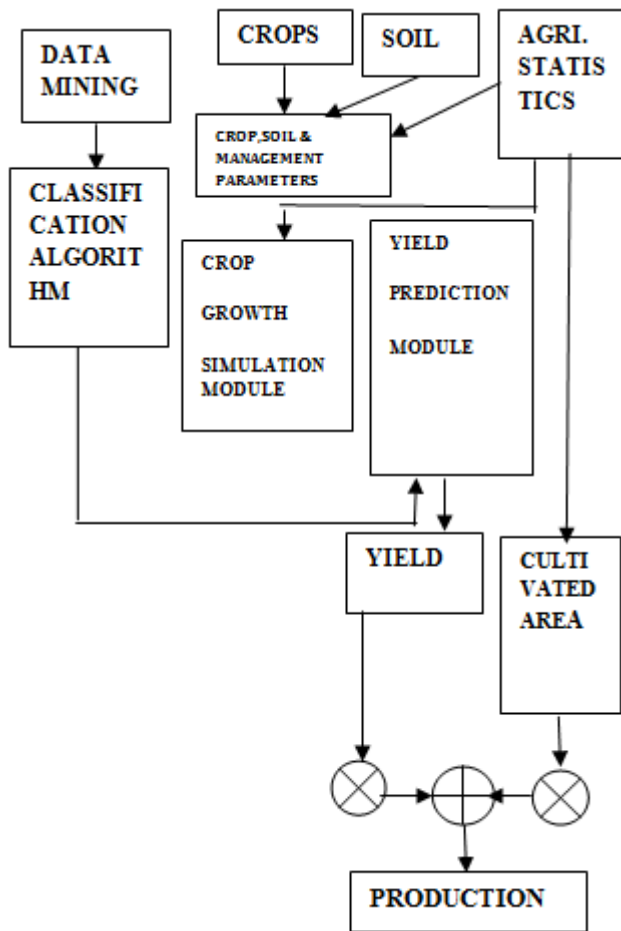
Fig -2: Architecture of the proposed system

The various methods are included in the architecture as shown in fig.2.

6. CONCLUSION

In this paper, we have done analysis of the soil data using J48 algorithm and prediction methods. In this paper we have demonstrated a classification algorithm called J48 (C4.5) using Weka tool . J48 is very simple classifier to make a decision tree, but it gave the best result in the experiment. As per the soil sample given to lab for testing and cropping pattern the system will recommend suitable fertilizer. In future, we contrive to build Fertilizer Recommendation System which can be utilized effectively by the Soil Testing Laboratories.

REFERENCES

[1]   Abdullah, A., Brobst, S., M.Umer M. 2004. "The case for anagri data ware house: Enabling analytical exploration of integrated agricultural data". Proc. of IASTED International Conference on Databases and Applications.

[2]   Abdullah, A., Brobst, S, Pervaiz.I., Umer M.,A. Nisar. 2004. "Learning dynamics of pesticide abuse through data mining". Proc. of Australian Workshop on Data Mining and Web Intelligence, New Zealand, January.

[3]   Abdullah, A., Bulbul.R., Tahir Mehmood. 2005. "Mapping nominal values to numbers by data mining spectral properties of leaves". Proc. of 3rd International Symposium on Intelligent Information Technology in Agriculture. Beijing, China. Oct, 2005.

[4]   Babu, MSP., Ramana Murthy, NV, SVNL Narayana, 2010. "A web based tomato crop expert information system based on artificial intelligence and machine learning algorithms". Int. J. of Comp. Sci., and Information Technologies. Vol. 1(1) . pp. 6-15.

[5]   Basak J., Sudharshan, A., Trivedi D., M.S.Santhanam. 2004. "Weather Data Mining Using Independent Component Analysis". J. of Machine Learning Research 5: pp. 239- 253.

BIOGRAPHIES

Pooja M C is a final year student in computer science department in SSE ,Mangalore.

Sangeetha M is a final year student in computer science department in SSE ,Mangalore.

Shreyaswi J Salian is a final year student in computer science department in SSE ,Mangalore.

Veena Kamath is a final year student in computer science department in SSE ,Mangalore.

Mithun Naik is a asst. professor in SSE Mangalore. He has 1.5 year of experience in computer science department in SSE ,Mangalore.