A Study on Crop Yield Forecasting Using Classification Techniques

R.Sujatha
Research Scholar
Department of Computer Science
Ayya Nadar Janaki Ammal College
Sivakasi
sujathavcet@ gmail.com

Dr.P.Isakki @ Devi
Assistant Professor
Department of Computer Applications
Ayya Nadar Janaki Ammal College
Sivakasi
harish24devi@ gmail.com

Abstract: India is generally an agricultural country. Agriculture is the single most important contributor to the Indian economy. Agriculture crop production depends on the season, biological, and economic cause. The prognosticating of agricultural yield is a challenging and desirable task for every nation. Nowadays, Farmers are struggling to produce the yield because of unpredictable climatic changes and drastically reduce in water resource so; we are creating an agriculture data. This data could be gathered, stored and analyzed for useful information. It is used to promote new advanced methods and approaches such as data mining that can give the information of the previous results to the crop yield estimation. In this paper, we have demonstrated to estimate the crop yield, choose the most excellent crop, thereby improves the value and gain of the farming area using data mining techniques.

Keywords- **Data mining, classification algorithms, Crop details, yield, prediction.**

I.INTRODUCTION

Data mining is the removal of unseen foretelling information from the enormous database, is a great new tool with enormous probable to assist business groups consider closely on the most essential statistics in company data repositories. Data mining associates the use of complicated statistics, analysis tools to find previously unidentified, suitable unseen structure and interaction in the huge dataset [1]. Data Mining is used to explore huge data, sets and launch helpful classifications and patterns in data sets. The aim of the Data mining procedure is to mine the information data from a data collection and change it into an explainable framework for additional use. Data Mining is the procedure of discovering in the unknown past, and a potentially growing model in huge database. The extracted data is used for representation as a perfect for a forecast or classification. Data mining is generally grouped as predictive and descriptive type. But in farming areas, predictive type is essentially used. The data mining techniques are Classification, Association rules, clustering and Regression.
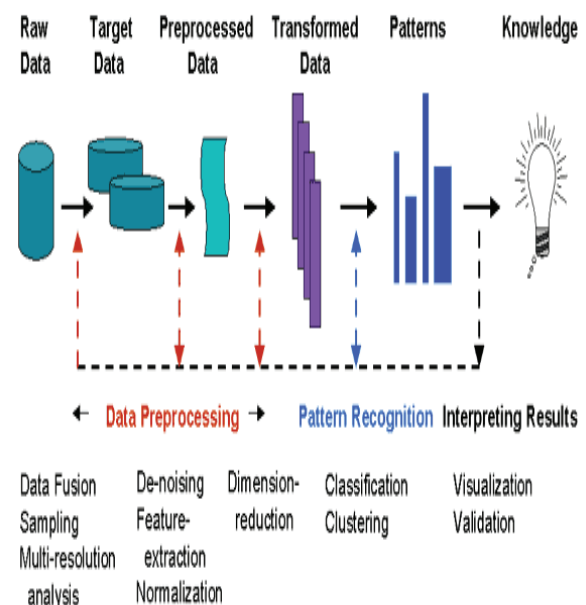

Fig 1: KDD Process in Data mining

Data mining is the procedure of using huge data sets to infer important hidden knowledge. (Fig 1) shows that knowledge discovery data mining process is divided into seven methods:

- ❖ Data cleaning
- ❖ Data integration
- ❖ Data selection

- ❖ Data transformation
- ❖ Data mining
- ❖ Pattern estimation
- ❖ Knowledge display

Over sixty percent of India's land areas are cultivable making it the second leading nation in terms of total cultivable land. Agriculture products of important economic value include rice, wheat, potato, tomato, onion, mangoes, sugarcane, beans, cotton, etc. India ranks among the peak five producers of many agricultural items like coffee, cotton, etc. India ranks the second biggest manufacturer of wheat and rice in the world. Items like milk, many fresh fruits, spices, jute, wheat, rice, etc, where India is the largest manufacturer. India is one of the biggest producers of Sugarcane in the world. In India, the majority of the farmers is not receiving the estimated crop yield due to several reasons.

The agricultural yield mainly depends on the weather environment. Crop yield forecast is a significant farming problem. Every cultivator is concerned with intentional, how much harvest he is about to expect. In the ancient times, crop yield forecast was achieved by farmer's earlier familiarity on an exacting crop. The volume of data is massive in Indian agriculture. The date when turn into information is extremely valuable for a lot of purposes [3]. In addition to providing food and unprocessed material, agriculture also gives employment to the very big percentage of the people.

In crop production and attacks of pests and diseases are the two major aspects which need consideration. Forecasts of crop productivity, previous to harvest are needed for assorted policy agreements relating to distribution, storage, rating, marketing, import- export, etc. pests and diseases are one of the key elements of the reduction in crop yield. Suitable application of curative measures may decrease the yield loss. For the application of these areas, one must have prior knowledge of the time and the harshness of the outbreak of pests and diseases

II. RELATED WORK

Veenadhari et al, described [4] to review the study on the purpose of data mining methods in the area of agriculture. A few of the data mining methods, such as the k-means,ID3 algorithms, the k nearest neighbor, support vector machines, artificial neural networks and applied in the area of agriculture were presented. An exact estimate of crop range and risk assists these business groups in planning, supply chain decision similar to production scheduling.

Narsi Reddy Gayam explained the hypothesis of normality of crop production yields using data gathered from India concerning Soybean and sugar cane. The unacceptable assumption was examined by means of the Lilliefore method united with exhaustive qualitative investigation of the data [8].

The paper [9] described they expect crop yields using data mining techniques. Inputs are given by farmers, using these inputs, easily predict crop yields.

Ranjana Agarwal explained the crop yield prediction methods. This paper also describes crop production problems, weather detail and prediction models [10].

III. METHODOLOGY

A. CLASSIFICATION

The major techniques of data mining are namely classification and clustering. Classification and prediction are two kind of analyzing data which being used by mine models which describes foremost classes of data and prediction of trends in future data. To increase the accuracy of prediction can be obtained by classification model when classifying samples the test set unseen in training are one of the major goals of classification algorithm.

Data mining algorithms divided into three unique methods of learning called supervised, unsupervised and semi supervised learning. A lot of classification techniques for finding knowledge that are Rule Based Classifiers, Bayesian Networks, Nearest Neighbor, Support Vector Machine, Decision Tree, Artificial Neural Network, Rough Sets, Fuzzy Logic, and Genetic Algorithms [4].

a) Naive Bayes

Naive bayes is an algorithm for classification of probability based on bayes theorem using hypothesis of strong autonomy. It's totally depends on the probability model's precise nature. This classification

[Type text]

technique can be trained very strong in a supervised learning setting. A merit of this algorithm is classifier it needs only a minimal quantity of training data for evaluating the constants like means and variances of variables which are importance for classification

b) J48

A java based implementation called J48 of C4.5 algorithm. C4.5 is an algorithm mainly focused to do a decision tree analysis. It builds a decision tree using set of labeled data input and it can be validated counter to invisible labeled test data for quantify how it's being generalized. This Algorithm is continuation of Quinlan's ID3 algorithm which was enhanced as C4.5 by himself. An ID3 algorithm being re utilized for diminishing of decision trees, derivations of rules, value extents and so on. This algorithm frequently being used as statistical classifier

c) Random forest

An algorithm called as Random Forest being used to systematize the huge data with perfection. It's an altogether method of learning for categorization and regression that increases the number of decision trees at that time and performing the category that is individual tree's classifier mode.

d) Artificial Neural Network

Neural networks are also comparable to the biological neuron network system in the sense that functions are implemented together and in parallel units, rather than there being an obvious delineation of sub tasks to which different units are assigned. A popular word called ANN (Artificial Neural Network) intended to neural network models that employed in an intellectual psychology, Artificial Intelligence and statistics of data. Central Nervous system (CNS) is an emulation being used to design the neural network methods that are subject of theoretical and computational neuroscience.

e) Decision tree

It is one of the algorithms which are used for the classification in data mining. A use of mining technique which is related to risk management proves the success of the advanced Geospatial Decision Support System (GDSS). Decision tree is a structure of an initial learning the miniature mainly focused to learn a rule which is common from the instances of observations [5][6].

f) Support Vector Machine

Support Vector Machine is capable of classifying data samples in two disjoint clusters. Support Vector Machine is a set of supervised learning method which is used for regression and classification. It can create a model which speculates a latest example whether fall into the group or not. It is an idea for set of correlated superintended learning techniques that show patterns and rate data being used for regression analysis and classification. It can use a group of input to the prediction for each input, two possible class forms making the non probabilistic binary linear classifier [7].

IV. CROP YIELD PREDICTION

The measurement of crop yield is used for a food grains, legume and is usually calculated in metric tons per hectare. Crop harvest is able to refer the real seed invention from the plant. For model, production of corn yielding four innovative productions of corn would have a crop harvest of 1:4. It is also called to as agricultural output.

An architecture of the crop yield prediction model which includes an input module, which is responsible for taking input from the farmer. The input module contains crop name, land area, soil type, soil pH, pest details, weather, water level, seed type. The feature selection module is responsible for subset selection of an attribute from crop details. The crop yield prediction model used to predict plant growth, plant diseases. After feature selection, the data go to classification rule for grouping similar contents
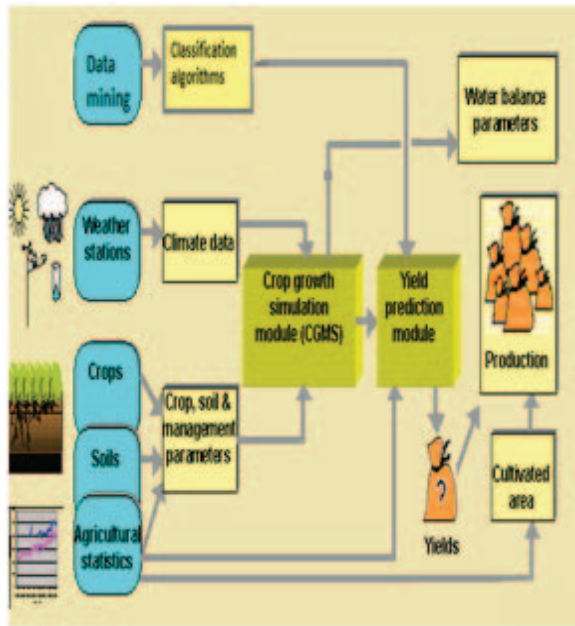
[Type text]

Fig 3: Architecture of the proposed system

Using climate data and crop parameters used to predict crop growth can be predicted. Then prediction rules will be applied to the output of classifying crop details in terms of crop name, pesticide and total yield details.

V.CONCLUSION

This paper describes how improving agriculture efficient by prophesying and improves yields by previous agriculture information. It also used to select a best crop by farmer, to plant depending on the weather situation and provides required information to prefer the suitable season to do excellence farming. This paper presents new research possibilities for the application of new classification methodologies to the problem of yield prediction. Using these techniques, the crop yield can be improvised and increase the income level of the farmer, will be increased.

REFERENCES

[1] Ranjan, J., "Data Mining Techniques for better decisions in Human Resource Management Systems," International Journal of Business Information Systems, IJBIS, vol. 3, pp. 464-481, 2008.

[2] Shanwad, U.K., Patil, V.C., and Honne Gowda, H., "Precision Farming: Dreams and Realities for Indian Agriculture", Map India, 2014.

[3] Ramesh, D., and VishnuVardhan, B., "Analysis Of Crop Yield Prediction Using Data Mining Techniques", 2015.

[4] Veenadhari, S., Bharat Misra, D Singh, "Data mining Techniques for Predicting Crop Productivity – A review article", IJCST, International Journal of Computer Science and technology, march 2011.

[5] Georg Ruß, Rudolf Kruse, Martin Schneider, and Peter Wagner. "Estimation of neural network parameters for wheat yield prediction" In Max Bramer, editor, Artificial Intelligence in Theory and Practice II, volume 276 of IFIP International Federation for Information Processing, pages 109–118. Springer, July 2008.

[6] Iv´an Mej´ıa-Guevara and ´Angel Kuri-Morales. "Evolutionary feature and parameter selection in support vector regression". In Lecture Notes in Computer Science, LNCS, volume 4827, pages 399–408. Springer, Berlin, Heidelberg, 2007.

[7] Ronan Collobert, Samy Bengio, and C. Williamson. "Svm torch: Support vector machines for large-scale regression problems". JMLR, Journal of Machine Learning Research, 1:143–160, 2001.

[8] Roberto Benedetti A, Remo Catenaro A, Federica Piersimoni B, "Generalized Software Tools For Crop Area Estimates And Yield Forecast "2010.

[9] Ramesh A. Medar and Vijay. S. Rajpurohit "A Survey of data mining techniques for crop yield prediction",IJARCSMS International Journal of Advance Research in Computer Science and Management Studies Volume 2, Issue 9, September 2014 pg. 59-64.

[10] "Forecasting Techniques in Crops" by Ranjana Agarwal, I, A.S.R.I., New Delhi.

[Type text]