Agricultural Data Mining

Vaishnavi.S[1], Shobana.M [2], Chokkalingam.S[3], Balamurugan.K[4]

[1,2] Assistant Professor, Department of Computer Science and Engineering, SNS College of Technology, Coimbatore, India
[3,4] III year CSE, SNS College of Technology, Coimbatore, India

Abstract— **Agriculture in today's world cannot be run through like it was done by forefathers by word of mouth. Climatic changes of today's world leads to unnatural things to happen such as less rainfall, poor weather and high temperature. It is difficult for farmers to cope up with climatic changes. Data Mining helps agriculture to give good decisions. Using data mining system, agricultural recommendations can be given to the farmers. It provides knowledge about crop to sow, fertilizers needed, water management. This proposed system aims to provide accurate decisions using C5 algorithm.**

Keywords— **Data Mining, Agriculture, Association Rule mining, Prediction**

I. INTRODUCTION

Agriculture plays a vital role in India's economy. It plays as backbone of Indian economy. India is an agricultural country with second highest land area. 54.6% of the population is engaged in agriculture and allied activities (census 2011) and it contributes 17.4% to the country's Gross Value Added (current price 2014-15, 2011-12 series). It contributes 10-16% GDP (Gross Domestic Product) to the Indian economy [1]. The current GDP value from agriculture is displayed in the chart below.

Fig 1 GDP from agriculture (2015 to 2016).

II. AGRICULUTURE

People who cultivate crops keep on facing huge problems such as weather, rainfall, flood, drought and other natural factors. Now days, Agriculture is uncared for its rise which makes farmers to lose their hope in agriculture leading to suicide of farmers. There is a need to assist farmers to help them in agriculture. Impact of IT sector techniques on agriculture is reaching high level day by day.

Plant growth depends on major factors or parameters such as Weather, Rainfall, humidity, soil type, crop variety etc.

Weather condition of country changes from month to month or in a quarterly basis. For example, Coimbatore district taking the district as a whole, January and February can be considered as pleasant months, From March, the climate becomes oppressively hot and this lasts till June. Towards the end of May, the south west monsoon rains bring instant relief to a greater part of the district, at about this period and during June, July and August the gusts of wind cools large parts of the district.

Rainfall varies by cms. Suitable crops have to sow during certain rainfall. Soil types play a vital role. Major types of soil [5] are Loamy, Sand, and Clay. For example, in clay soil, cotton cultivation will give high yield; [2] alluvial soil is suitable for rice, wheat, sugarcane, jute; black soil for groundnut, millets; laterite soils for cashew, rubber, coconut; red and yellow soils for ragi, potato, milet.

Agriculture requires large amount of water so water management is very important constrain in agriculture field. Water management is based on the type of soil. Selection of suitable pesticides and fertilizers along with the unit for the agricultural land is important.

Effect of these parameters on crops is highly unrevealed to many farmers. Less knowledge leads to low yield of crops and many other problems. During past decades, our ancestors suggested about the variety of crop to be cultivated in consideration of agricultural parameters by word of mouth. This can be done in an efficient way by using IT techniques in this century.

III. DATA MINING

Idea to give accurate results to farmers can be done by using the most popular recommending technique called Data Mining. It extracts useful, hidden data

from large dataset. It converts information to knowledge. It searches a large database and give a simple analysis to use.

Mining provides useful solutions for research in agricultural sector leading to find less sales of production of crops, effective irrigation methods. Many queries or doubts of farmers can be cleared.
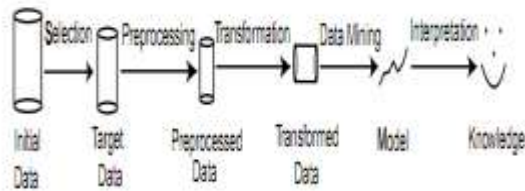

Fig 2 Mining Process

Data mining techniques such as Association Rule Mining, Classification, Clustering and Prediction makes accurate results.

A. Association Rule Mining

It gives frequent patterns of a dataset. It finds the association or relation between items in a dataset

B. Classification

Based on a previous dataset, items are categorized.

C. Clustering

Items are grouped as clusters.

D. Prediction

It is used for future results based on current data set items.

Impact of data mining functionalities in agricultural sector is still a solving issue. With the help of weather the rainfall can be predicted. The level of rainfall will be estimated and provided to the formers for effective cropping and cultivation. By knowing the amount of fertilizer and pesticides to be given for crops high yield can be gathered.

IV. LITERATURE SURVEY

Major Data Mining techniques applied on Agricultural data include k-means, bi clustering, k nearest neighbor, Neural Networks (NN) Support Vector Machine (SVM), Naive Bayes Classifier and Fuzzy c-means etc.

Tripathy et al., [15] describes about the system to manage the pesticides for crops using data mining techniques. These concentrate on providing historical data, current and recommended pest and pesticide information and given the pest models up to the farm

level. Association Rule Mining, formerly proposed for the market basket data, has prospective applications in many areas. Spatial data, such as Remote Sensed Imagery (RSI) data, is a promising area of application.

Qin Ding et al., [17] proposed an efficient approach to derive association rules from spatial data using Peano count tree (P-tree) structure. It provides lossless compression of spatial data. It provides efficient pruning techniques of crops.

ZeluJia et al., [19] introduced mining techniques for spatial data. It uses decision tree algorithm applying to agriculture land grading. The idea is to join spatial data mining/decision tree techniques with expert system techniques and combine them to find an intelligent agriculture land grading information system.

In the field of agriculture, data mining helps the farmers to select the right crop based on the quality of soil and weather condition. It also assists in taking decisions to determine the fertilizer level.

V. PROPOSED SYSTEM

Main objective of this system is to recommend crops depending on agricultural factors. It improves production of crops by considering available resources of land, availability of water, electricity, fertility of land. Proposed system covers following steps.
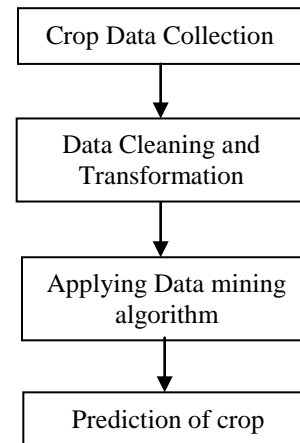

Fig 3 Framework of proposed system

Database of crops is collected from different sources of regions. Database consists information about Soil type, crop types, relation between soil and crop, weather situations and rainfall variations.
System is organized in below manner.
1. Input is got from the farmer
2. It is processed with data mining algorithms

3. Output gives the suitable crop to grow.

Assume,
Soil Type-S
Crop name-C
Weather Data-W
Month – M

Rainfall can be predicted with current weather forecasting conditions. Suitable mining algorithm has to be chosen for predicting the crop.

CROP PREDICTION= SUM OF S+W+C+M

M indirectly gives the rainfall strategy.

Advantages of Proposed systems:
It focuses on four steps:
- Monitoring Current Weather patterns.
- defining the soil types based on the set of features,
- Finding the best decisions, and
- Estimating the effectiveness of suggested decisions.

C5 algorithm is used here to predict the crop yield. C5 algorithms give some of uses such as:
- The proposed system handles training data with missing values of attributes. So, the prediction will be more accurate and effective.
- Handling differing data typed features.
- Prediction probabilities based on the historical dataset
- Pruning the decision tree after its creation
- Improves the performance of the weak classifiers by modifying the process on it.

VI.CONCLUSION

Agriculture processing is tough to learn, when the farmer doesn't know anything about the agricultural activities. This toughness is due to the high dimensional and dynamic data. Agricultural activities are varying according to the location, soil and weather details. This kind of diversitical data can be combined and used by applying data mining algorithm. C5 algorithm shows more accuracy than C4.0. It gives fast decision. In future work, crop disease can be merged with this by making farmers to predict the disease earlier.

REFERENCES

[1]. http://www.tradingeconomics.com/india/gdp-from-agriculture.
[2]. http://www.gardenguides.com/128581-types-soil-crops-grown-india.html
[3]. http://www.tnau.ac.in/dr/zonepdf/WesternZone.pdf
[4]. http://tnhorticulture.tn.gov.in/horti/coimbatore
[5]. http://krishikosh.egranth.ac.in/bitstream/1/2034269/1/NBSS-42.pdf
[6]. Piatetsky-Shapiro, Gregory. *Advances in knowledge discovery and data mining*. Eds. Usama M. Fayyad, Padhraic Smyth, and RamasamyUthurusamy. Vol. 21. Menlo Park: AAAI press, 1996.
[7]. Han, Jiawei, Jian Pei, and MichelineKamber. *Data mining: concepts and techniques*. Elsevier, 2011.Chen, Ming-Syan, Jiawei Han, and Philip S. Yu. "Data mining: an overview from a database perspective." *IEEE Transactions on Knowledge and data Engineering* 8.6 (1996): 866-883.
[8]. Kass, Gordon V. "An exploratory technique for investigating large quantities of categorical data." *Applied statistics* (1980): 119-127.
[9]. Grajski, Kamil A., et al. "Classification of EEG spatial patterns with a tree-structured methodology: CART." *IEEE transactions on biomedical engineering* 12 (1986): 1076-1086.
[10]. Quinlan, J. Ross. "Induction of decision trees." *Machine learning* 1.1 (1986): 81-106.