

Survey of Estimation of Crop Yield using Agriculture Data

Anu priya K R

Dept of computer Science and Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India

Abstract - Data mining techniques are increasingly being used in finding useful pattern from large data sets .Huge amount of data is being produced day by day from various sources and are stored in data sets .half of the people in India depends on agriculture for their maintenance .Data mining plays a major role in estimation of crop production depending upon the various data sets such as soil verify, rainfall, humidity, temperature, cities etc .This paper focuses on analysing agriculture data ,pre-processing the data to remove the noise and find the optimal parameter to increase the crop yield which is more accurate and optimal .This paper also makes an effort to understand the different complications to eliminate the noise from the initial dataset ,know the problems of various algorithms used in earlier approaches.

Key Words: Data cleaning; Big data; Data preprocessing
Optimal yield; Noise elimination.

1. INTRODUCTION

Big data deal with large amount of data which cannot be accommodated by a normal computer in terms of storage, processing and computation and data analysis technique [1].As the size of data is increasing every moment which causes complexity in terms of volume, variety, velocity, veracity as a result it requires more storage, computation and speed should be more[2].

Data mining on this big data bit a way of removing noise from the initial dataset which is called pre-processing using some of the pre-processing techniques. Rapid fluctuation are common in market in such scenario it is difficult for the farmers to choose the type of crop to be grown in land or estimate the yield. Data mining is the procedure of using huge data set to infer important hidden knowledge.

Data mining is divided into 7 methods:

- Data Cleaning
- Data integration
- Data selection
- Data transformation
- Data mining
- Pattern estimation
- Knowledge display

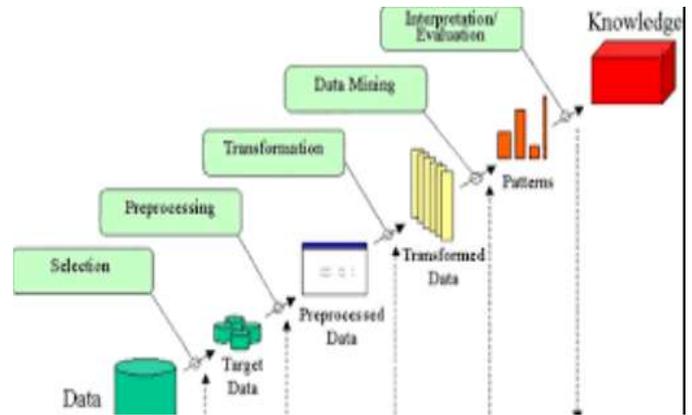


Figure 1.shows the different stages in data mining

Along with this in today's applications mostly all of them involve crowd sourcing and semantic input for improved data analysis in such an environment the tracing and analysing data provenance of the data discovery and integration of several parallel and distributed computing generated data is the scope for implementing newer exploratory data analysis and interpretation methods where in the, integration of heterogeneous data, and developing new models for massive data computation is possible.

The main purpose of data cleaning is to manipulate and transform raw data so that the information content enfolded in the data set can be exposed, or made more easily accessible. The quality of data is very pivotal in many organizations as it maintains the basic characteristics of cleaned data. So it's necessary to obtain a proper sample data for knowledge extraction. Preprocessing can produce a smaller data set than the original, which allows us to improve efficiency in the data mining process. This performance includes data reduction techniques such as feature selection, sampling, instance selection and discretization.

There are many methods for cleaning noisy and inconsistent data like:

- Filter-based Method
- Imputation Method
- Hybrid method
- Wrapper method
- Global imputation of stray and non-missing attribute

- Embedded Method

Data mining involves six common classes of tasks

- Detection of anomalies (detection of outliers / change / deviation): documentation of rare data records, which may be stimulants or data errors that require further study.
- Learning association rules (dependence modeling): observing for relationships between variables. For example, a shop can collect data on customer buying habits. Using the rules of the learning association, the shop can regulate which products are regularly bought together and use this information for marketing purposes. This is sometimes mentioned to as an study of the market basket.
- Grouping: the task of determining groups and structures in the data that in one way or another are "similar", without using structures known in the data.
- Classification: it is the task of simplifying the known structure to apply it to innovative data. For example, an email program might try to categorize a message as "legitimate" or "spam".
- Regression: try to find a function that models the data with the lowest error, to estimate the relationships between data or data sets.
- Summary: provides a more solid illustration of the data set, including viewing and reporting.

LITERATURE SURVEY

Mohammad Motiur Rahman Naheena Haq, Rashedur M Rahman have done an analysis on forecasting the rice yield of Bangladesh. The study has shown that we can have good relationship designed among the rice yield and the climatic variable of a given area within certain threshold. This study has helped the people concerned with agriculture such as the farmer etc to make decisions to diminish their loss in the following years.

Jayaram Hariharakrishnan, Mohanavalli .S, Srividya, Sundhara Kumar K.B carried survey on various data cleaning methods such as Filter method, Imputation method and Wrapper method. It focused on how efficient this data cleaning techniques where in terms of blank spaces and missing values. It compares different data mining techniques and suggest which data cleaning technique is best

Ratul Dey, Sanjay Chakraborty , this paper help us understand how incremental DBSCAN clustering algorithm and Convex-Hull can be used to predict the weather of the upcoming days. In this they have shown that these algorithms are suitable for active database where the weather data keeps changing regularly. The accuracy of the techniques used is calculated based upon the hit and the miss times.

S.Kanaga Suba Raja,Rishi .R,Sundarsan.E,Srijit focuses on try to predict crop yield and value by analyzing patterns in past data.The technique used is Sliding Window and Non linear Regression to predict several features that influence the agriculture production such as rainfall,Temperture etc..The proposed system helps to decrease the harm faced by the farmers and would recover the crop yield.

Jharna Majumdar, Sneha Naraseeyappa and Shilpa Ankalaki, the survey tells us that data mining techniques are very necessary approach for estimation of various crop yields. This paper focuses on the analysis agricultural dataset and find optimal solution for maximizing the crop production. PAM, CLARA etc are some of the different data mining techniques that are been used.

III. PROPOSED SYSTEM

Several data mining techniques are carried on the input dataset to accomplish the finest performance yielding method. Clustering method can obtain better results. DBSCAN algorithm can be used to estimate the crop yield by considering various datasets and forming the clusters for making decisions.

DBSAN is resistant to noise and can handle clusters of various shapes and sizes. The results obtained are more optimal and efficient hence farmers can make better decision.

IV. CONCLUSIONS

A good clustering algorithm should be able to forecast the crop yield which is more effective and optimal compared to other clustering algorithms. There are numerous clustering algorithms for predicting crop yield based upon the various agricultural data set.

Thus executing the above system would further lead to the agricultural growth of our country. The DBSCAN algorithm helps us to forecast best range of best temperature, worst temperature and rain fall to achieve higher production of different crops. Clustering methods are associated using quality metrics. Rendering to the analyses of clustering quality metrics, DBSCAN gives the better clustering quality

REFERENCES

1. "A survey on pre-processing and postprocessing techniques in data mining." Tomar, Divya, and Sonali A1;arwal International Journal of Database Theory & Application 7.4 (2014).
- 2."Enhancing data analysis with noise removal." Xiong, Hui, et al, IEEE Transactions on Knowledge and Data Engineering 18.3 (2006): 304- 319.
3. " Survey of Pre-processing Techniques for Mining Big Data" JayaramHariharakrishnan*,Mohanavalli.S*,Srividya*, Sundhara Kumar K.B** Department of Information

Technology, SSN College of Engineering Kalavakkam, Tamil Nadu, India.IEEE[2017].

4. "Analysis of agriculture data using data mining techniques: application of big data" Jharna Majumdar* , Sneha Naraseeyappa and Shilpa Ankalaki [2017]

5. "Demand Based Crop Recommender system for Farmers" S.Kanaga Suba Raja,Rishi .R,Sundarsan.E,Srijit.V Department of Information Technology , Easwari Engineering College, Chennai, India.[2017]

6. "Convex-Hull & DBSCAN Clustering to Predict Future Weather" Ratul Dey , Sanjay Chakraborty Computer Science & Engineering ,Institute of Engineering & Management[2015]

7. "Application of Data Mining Tools for Rice Yield Prediction on Clustered Regions of Bangladesh ".Mohammad Motiur Rahman , Naheena Haq, Rashedur M Rahman Electrical and Computer Engineering Department North South University Dhaka, Bangladesh [2014]