Crop yield prediction using Agro Algorithm in Hadoop

Ashwani Kumar Kushwaha
M.Tech (Software Technology)
VIT University, Vellore, India,
14ashwani@gmail.com

SwetaBhattachrya
Assistant Professor, SITE
VIT University, Vellore, India,
sweta.b@vit.ac.in

ABSTRACT - Agriculture is one of the important sources of survival and one of the most important factors in the economic growth of the country. Researchers in the field of the agriculture have studied and implemented different well-organized mechanisms which would predict and increase the crop yieldsand make agriculture highly profitable.In this paper we have tried to predict the crop yield, suggest the best crop thereby improve the quality and profitability of the agriculture sector by processing huge volume of data often called as Big Data using Hadoop platform. The paper also focuses on the soil type and helps to find which particular crop would be suitable for a particular type of soil. In case of crop yield soil plays an important role andsoil information is predicted by considering the weather details of the previous years. Hence the study will predictthe suitability of a crop for a particular climatic condition and the possibilities of improving the crops quality by using weather and disease related data sets. Predicting the quality of cropsis quite complexand we need to take care of multiple approaches. This research paper help to identify mechanisms to get good quality and improved crop yields using a new algorithm named as"Agro algorithm" implemented in Hadoop platform and uses Hadoop framework to handle large amount of data sets.

Keywords: Agro algorithm, Crop quality, Crop yield, Big Data, Hadoop platform.

I. INTRODUCTION

Since the creation and evolution of human beings in the planet, it has always been a struggle to manage the assets to meet the fundamental necessities [5]. The fundamental asset are soil, water and air as said in the Upanishads, Vedas and in ancient Hindu literature and our primary responsibility is to manage these resources efficiently and keepthem clean. In the last century there has been a phenomenal increase in the number of both humans and animals. The fast growing urbanization and industrialization in few decades is rapidly consuming all the natural resources and theyare degrading much faster than we had ever imagined. There has been an immense gap between the demand and supply of food and edible resources and as of today we do not have proper solution or mechanism to tackle the crisis due to issues like weather, temperature and other factors which are beyond our control.India has the ability to achieve extraordinary increase in the crop yield production with the help of expansion of irrigation and technological innovation in agriculture. There are multiple strategies which can be adoptedto improve the number and quality of crops. In comparison to other developed countries attaining the essential growth is an extreme challenge in India due to the lack of maintenance of the resources on which the production systems are dependent[2]. The successful usage of quality farming relies upon various elements, counting the degree to which conditions inside a field are known and dealt with, the sufficiency of information proposal and the level of user control. The different types of quality agriculture incorporate a wide exhibit of themes including variability of the soil asset base, climate, plant genetics, crop diversity, hardware execution and most physical, substance and natural inputs utilized for crop production. The quality of agriculture must fit the needs and abilities of the rancher and must be profitable.The present study and research paper uses a new algorithm entitled as "Agro algorithm" to predict the quality crop yields by determining the soil information and informs whichcrop cultivation would be better in a certain type of soil. The information of soil can be determined by the weather information based on theweather related datasets of previous yearscontaining the minimum temperature, maximum temperature, Solar radiation (cal/cm2/day), sun shine (hours), rain fall (mm), rainy days, evaporation (mm), relative humidity (percentage), wind velocity (kmph). Also data sets related to cropdiseases are taken into consideration i.e. whichclimatic conditionhave better or poor chances of disease prone or disease free crop productions.The paper implements the Agro algorithm for big data usingHadoop platformto deal with large amount of datasets. We get the data from the Hadoop Distributed File System (HDFS) and read it line by line and perform normalization for

the given data by taking the statistical average mean of data. Then it is classified as our classifier is a simple statistical-based learning scheme. Finally we classify data based on disease and data extracted from the classifier and we predict the quality of soil and which crop to yield. We select the month in which we plan to sow the crop and then we provide the previous year information analyse the data sets of that particular month and then classify the data based ondisease and data extracted from the classifier and finally predict the soil and crop. The prediction of the soil is representedwith a pie-chart with the respective percentages of the prediction of the crop. Based on the predictions, the crop and soil are divided in to five categories such as “very good”, “good”, “average”, “bad” and “very bad”.

II. LITERATURE SURVEY

There exist numerous methods and proposed mechanism/models for the prediction of crop yield with innovative ways of analysing and classifying datasets. But they hardly discuss the issues and methods of handling large and complex datasets are.Hence classifying large datasets remains a very difficult and complicated task with an additional expectation of enhanced performance makes it more challenging. The study by *H. Guo et.al*proposed multi relation cluster by many view creation without improvement in the original dataset usingMRC Algorithm [6]. The basic idea is to use clustering techniques which in turn uses data mining approach at different levels.Finally Viewed validation algorithm was used to validate the view but this approach but was not helpful for very large and difficult relational datasets [1].A mode clustering approach for relational databasesby grouping of probabilistic learning and reasoning for linear scaling was proposed by Ben Taskar et.al. Fine tuning clusters and model selection was done by training datasets.

The concept of cluster domain expertise is extremely necessary for modelling but becomes extremely difficult to achieve and implement in real time in case of automatic model construction[3].Different efforts have been made to survey the use of data mining methods in the field of farming. Soft computing and advanced technical methods havebeen applied in the field of farming such as artificial neural networks, the k nearest neighbour, the k-means, support vector machines and ID3 algorithms. The application of Data mining techniques in horticulture is a moderately new approach and provides new insight in animal management, forecasting and in the prediction of agricultural crops.

III. KNOWLEDGE REQUIRED BEFORE CROP PLANTATION

In order to predict improved quality of crop one should know have complete idea about the crop in terms of horticulture, plant biology and other relevant areas. This would provide idea about all the details related to the basic growth requirements of the crops and required environmental support.
The following basic stepsare some of the guidelines required in qualityfarming [5]:

A. Assessing variability

This is the first and one of the most important steps in quality farming. The different procedures for accessing spatial variability are promptly available and have been connected widely in quality agriculture.

B. Managing variability

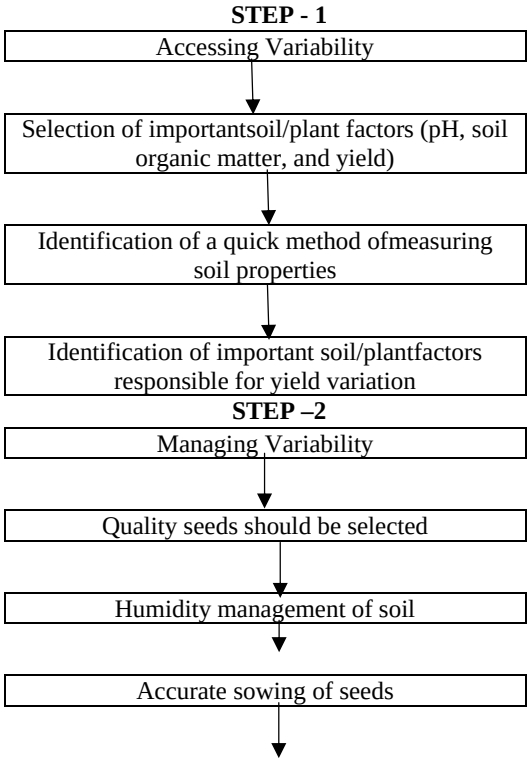
This is the second step in quality farming;greater the spatial need of an adaptable soil, higher is the possibility for quality management.

C. Evaluation

The following farming issues need to be rigorously analysed and evaluatedin order to achieve optimized and improved quality of crop production.

- Environment
- Economicsand
- Technology transfer

The sequence of steps to be following are shown in the following Figure No.1



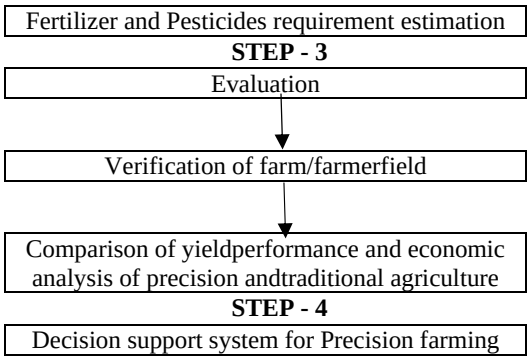


Figure1: Sequence of Steps for quality crop productions

IV. PROPOSED WORK

A. Analysis Algorithm

```
Get Data from DFS
Varval,total.avg
For each val then
Begin
ReadFile()
End
For each total then
Begin
Total+=val
End
Avg=total/N;
```

We get the data from the Hadoop Distributed File System (HDFS) and read it line by line, and perform normalization for the data by taking statistical average mean of data.

B. Classification Algorithm

```
Varmean,sum
Var pi= 1068966896 / 340262731;
Get NormalizedData from DFS
For each row then
Begin
Sum+=value
End
Return sqrt((sumSq-((sum*sum)/n))/(n-1))
```

Our classification algorithmused is a statistical-based learning system having the following qualities:

- It works astonishingly well
- Fastlearning and fast training
- Smallruntime memory required
- Simple and perfect

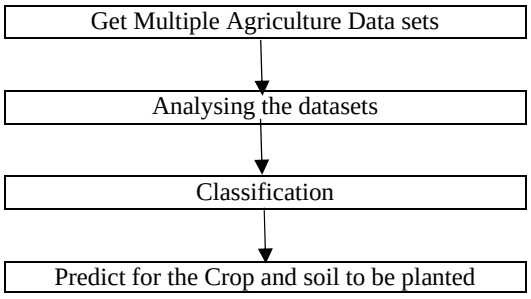
The learning can be completed via statistical modelling,. Their attributes aresimilarly importantwhereas statistically independent. This implies that learning about the estimation of a specific trait doesn't let us know anything about the estimation of another quality.

C. Prediction Algorithm

```
Var like, what,kclass
Get ClassifiedData from DFS
If like then
Begin
Likelihood(like,kclass)
End
Function likelihood(like,kclass)
Begin
Case CROP:
        putInCrop(like);
break;
Case SOIL :
        putInSoil(like);
break;
Case DISEASE:
        putInDisease(like);
break;
default:
returnkclass;
End
```

Here we classify data based on disease and the data extracted from the classifier and this helps to predict the soil and crop.

V. PROPOSED ARCHITECTURE



VI. OBSTACLES

There are numerousissues adopting quality farming in all the countries dependent on farming across the globe. These issuesdiffer in terms of regions but some issues are specific to the Indian conditions as mentionedbelow [6]:

- Institutional constraints,land possession and infrastructure
- Observations and Culture of the users
- Technical gaps and knowledge

- Small farm size are smaller
- Lack of Success stories are very less
- Local mechanical expertise are very few
- Market failures and Heterogeneity of harvesting systems
- Local mechanical expertise are very few
- Quality, costsand data availability

VII. CONCLUSION

This paper proposes how to make agriculture well organized by predicting and thus improve the crop yields by using soil information. The paper introduces a new Agro algorithm which is used to predict the suitability of a crop for a particular soil type and enhances the overall quality of agricultural production. This also helpsthe farmers to select a particular crop to sow depending on the climatic condition and provides necessary information to choose the best weather to do quality farming. The paper uses big data using Hadoop platform which helps to deal with the large amount of datasets in agricultural domain.

VIII. FUTURE WORK

There exists different types of diseases affecting a particular type of crop and it isoften difficult to predict which type of disease might occur and its cause, frequency and tentative season of period of occurrence. The future study would incorporate classification of the different types of disease for a particular crop and help in predictions related to this area which would enhance crop production and also eliminate wastage.

IX. ACKNOWLEDGMENT

The authors of this paper would like to express appreciation and thanks towardsall the farmers of our country who toil and work tremendously hard ignoring all the natural, social and economic hardships on order to provide food and edible resources for the citizens of our country. The efforts and approaches mentioned in the paper would be considered satisfying and successful if the study could make any positive impact on their continuous struggle to yield the best crops andease their continuous fight for achieving quality agricultural production.

REFERENCES.

[1] Taskar,B., Segal, E.,Koller, D., “Probabilistic clustering in relational data”, IJCAI-01, pp. 870--876, 2001.
[2] M. GunasundariAnanthara, Dr. T. Arunkumar, Ms. R. Hemavathy “CRY – An improved Crop Yield Prediction model using Bee Hive Clustering Approach for Agricultural data sets”Proceedings of the 2013 International Conference onPattern Recognition, Informatics and Mobile Engineering, pp. 473 – 479, February 21-22, 2013
[3] Yethiraj N G. “Applying data mining techniques in the field of agriculture and allied sciences”, International Journal of Business Intelligents ISSN: 2278-2400, Vol 01, Issue 02, December 2012
[4] Takoi K. Hamrita, Jeffrey S. Durrence, George Ve L LIdis, “Precision Farming Practices” IEEE Industry Applications Magazine, Mar- Apr, 2009
[5] U.K. Shanwad, V.C. Patil and H. Honne Gowda “Precision Farming: Dreams and Realities for Indian Agriculture”, Map India, 2014
[6] H. Guo, H. L. Viktor “Multirelational Cluster: AMultiple ViewApproach” Springer Knowledge Information System, Vol.17, Issue-3,pg.no-287–312, 2008.