

Experimental Analysis of Agricultural Data using Data Mining

Sakshi Sankhe¹ Kanchan Sharma² Kaushalya Shettiyar³ Sagar Bhavsar⁴

^{1,2,3,4}Department of Information Technology

^{1,2,3,4}Theem College of Engineering (Mumbai University), India

Abstract— In the agriculture field farmers and agribusinesses have to make numerous decisions every day and complexities involve the various factors impacting them. Data mining techniques are the approach necessary for fulfilling effective solutions for this problem. Data Mining is a technique that focuses on large datasets to extract information for the prediction and discovery of hidden patterns. Agriculture mainly depends on climate or weather, agricultural topography, etc. The current study presents the numerous data mining/refining techniques and their role in the context of soil fertility, nutrient analysis. A decision tree is a good way for classification in data mining. C4.5, Classification and Regression Trees (ID3) are two mostly used decision tree algorithms for classification. ID3 algorithm produces misclassification, the main drawback of the C4.5 algorithm is that errors when the domain of the target attribute is very large. This paper therefore presents a modified decision tree algorithm to reduce limitations. The model is tested with the data set of soil samples. The test proves that the modified decision tree algorithm has higher classification accuracy when compared to C4.5 and ID3 algorithms. Classification of soil is that the separation of soil into classes or groups each having similar characteristics and potentially similar behavior. Classification of soil is needed so that farmers can know the type of soil and can plough the crops depending on the type of soil.

Keywords: Data mining, agriculture, data analysis

I. INTRODUCTION

Data mining is the technique by analyzing the interesting and unknown patterns in a large volume of datasets from outlining the useful information from various perspectives. Extracted knowledge from different data sets is used for better decision-making in critical situations. Data Mining techniques will be used for prediction the longer-term trends of agricultural processes. Data mining techniques are of two types, one is descriptive which considers the existing data and another is predictive which depends on probability for future analysis.

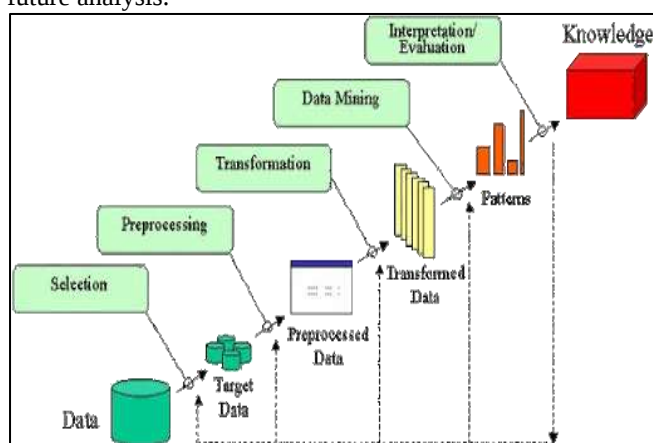


Fig. 1: Typical Data Mining Processing

The Data Mining process involves

- (1) Collect, clean, and load the data into the data warehouse system
- (2) Stores the data in multidimensional format
- (3) Provides information access to analysts and decision-makers
- (4) Analysis of data using different applications
- (5) Presents the data using different Patterns

II. BACKGROUND

Today, India ranks second worldwide in farm output. Agriculture is the widest economic sector and it plays a crucial role in the overall socio-economic of India. The analysis of comparing organic and conventional agriculture yield data showed that currently, individual crops of organic yields are on average 80% of conventional yields. The analysis of 362 datasets also showed a high difference of the yield gap of organic agriculture standard deviation is 21%. Some of this variation seems systematic. E.g. soybean, some other pulses, rice, and corn score higher than 80%, and wheat, barley, and potato scoring lower than 80%. Most regions have relative yields fairly on the point of general average.

III. PROBLEM DEFINITION

In the current scenario, there's no such recommendation system available online. And also farmer has to physically go and get a recommendation based on their soil report by quality executives. In this project, we've got suggested an analysis of the soil information using Decision Tree algorithms and prediction technique. By using the Decision tree algorithm, we recommend farmers about their field soil quality and suggest crops that are suitable to grow in that soil. The use of information technology in agriculture can change the scenario of decision making and farmers can yield in a better way.

IV. OBJECTIVE

The objective of the proposed work is to analyze the agriculture data using data mining techniques were to characterize and classify the soil to provide more details about the quality information of soil. An additional objective of this study was to seek out approximate crops that can be growing in that field. This study was to characterize and classify the soil to provide more details about the subsurface morphological information.

An additional objective of this study was to seek out an approximate value of soil loss from the Lower Moshe Irrigation Scheme using Universal soil loss.

V. LITERATURE SURVEY

We have undergone certain research papers based on the Experimental Analysis Of agricultural Data Using Data Mining and found out a certain methodology and key

findings. During the process, we have also identified certain research gaps which we can overcome while implementing our project.

Ref no.	Paper Title	Author	Year Of Publication	Key Findings	Research Gaps
1	Analysis Of agriculture data Using Data Mining	Lucinda Edwards & Carolina Eriksson	2017	Various data mining techniques are implemented on the input data to assess the best performance yielding method. The present work used data mining techniques PAM, CLARA and DBSCAN to obtain the optimal climate requirement of wheat like optimal range of best temperature, worst temperature and rain fall to achieve higher production of wheat crop.	Clustering is considered as an unsupervised classification process. A large number of clustering algorithms have been developed for different purposes. Clustering techniques can be categorised into Partitioning clustering, Hierarchical clustering, Density-based methods, Grid-based methods and Model based clustering methods.
2	Analysis Of Soil Behaviour and Prediction of Crop Yield using data mining approach	Supriya Dm., Karnataka	2016	The researcher express that large amount of data which is collected and stored for analysis. Making appropriate use of these data often leads to considerable gains in efficiency and therefore economic advantages.	The researchers implemented K-Means algorithm to forecast the pollution in the atmosphere, the K Nearest Neighbour is applied for simulating daily precipitations and other weather variables and different possible changes of the weather scenarios are analyzed using Support Vector Machines.
3	A Survey On Predictive Analysis in Agricultural soil data to predict the best fitting crop	A. Agrawal, J. Basak, V. Jam R. Kothari	2018	Soil classification was measured serious to study due to depending upon the fertility class of the soil domain knowledge experts determines which crops should be taken on that particular soil and which fertilizers should be used for the same.	suggested an analysis of the soil data using different algorithms and prediction technique.
4	Data Mining Technique to Predict the Accuracy Of the Soil fertility	Dr. S. Hari, Dr. Jayashree	2017	The research studies on application of data mining techniques in the field of agriculture. Some of the techniques, such as ID3 algorithms, the k-means, the k nearest neighbour, artificial neural networks and support vector machines applied in the field of agriculture were presented.	This article explores the applications of data mining techniques in the field of agriculture and allied sciences. Historical crop yield information is important for supply chain operation of companies engaged in industries that use agricultural produce as raw material.
5	Data Mining in Agriculture prediction soil fertility	Dr. Shrinivashan, Dr. Mailthali M.	2017	A soil test is the analysis of a soil sample to determine nutrient content, composition and other characteristics. Tests are usually performed to measure fertility and indicate deficiencies that need to be remedied.	In this paper demonstrated acomparative study of varied classification algorithms i.e. Naive bayes, J48 (C4.5), JRip with the assistance of data mining tool J48 is incredibly easy classifier to form a decision tree.

Fig. 2: Table of Literature Review

VI. METHODOLOGY

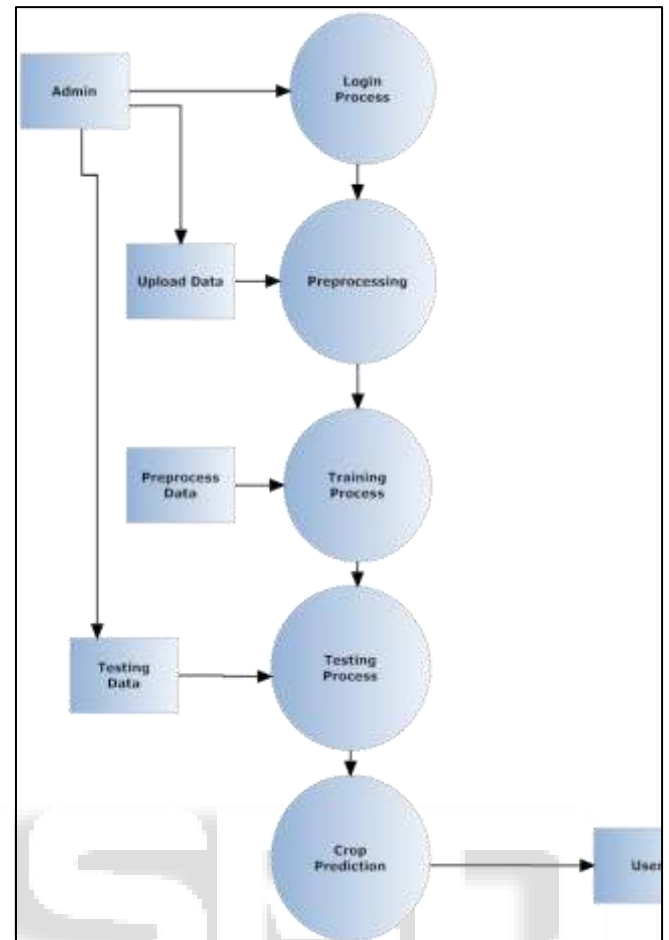


Fig. 3: Methodology Block Diagram for Analysis of Agricultural Data using data mining

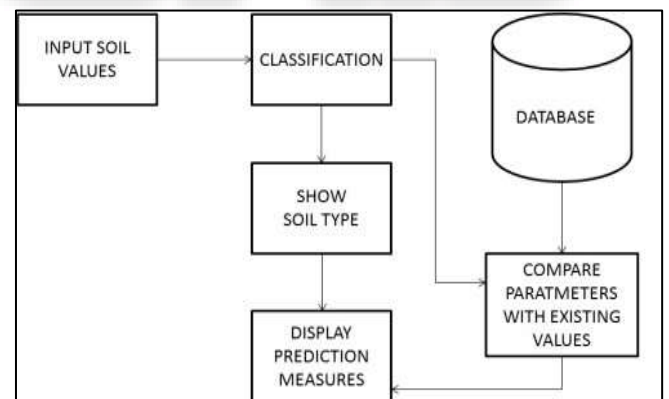


Fig. 4: Soil Classification

VII. ALGORITHM IMPLEMENTED

A. Decision Tree Algorithm

These algorithms find alternative ways to separate the information into partitions. It is the foremost widely known supervised learning technique. In this project, we've got suggested an analysis of the soil information using Decision Tree algorithms and prediction technique. By using the Decision tree algorithm, we recommend farmers about their field soil quality and suggest crops that are suitable to grow in that soil.

B. Regression Tree Algorithm

The Regression Tree Algorithm can be used to find one model those results in good predictions for the new data. These methods work by creating multiple diverse regression models, taking different samples of the initial data set, and so combining their outputs. It models a target prediction value based on independent variables linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). We have used the algorithm to predict rainfall based on the previous year's rainfall data.

VIII. RESULT AND DISCUSSION

This project is basically related to the use of information technology in agriculture which can change the plot of decision making and farmers can yield in a better way. For decision-making on overall issues related to the agriculture field; data mining plays a vital role. The survey discussed the role of data mining in terms of the agriculture field. We have also discussed various varieties of soils, several processing techniques in agriculture and soil containment.

IX. CONCLUSION

Usage of technology during agriculture can change the standard and scenario of deciding making and farmers can yield in a better and convenient way. Data mining plays a crucial and lead role in decision-making on several issues related to the agriculture field. It discusses the role of data mining in the agriculture field and their related work in context to agriculture. It also discusses different data mining applications in solving the different agricultural problems. In this project, we've got suggested an analysis of the soil information using Decision Tree algorithms and prediction technique. By using the Decision tree algorithm, we recommend farmers about their field soil quality and suggest crops that are suitable to grow in that soil. Agriculture is of the utmost important area especially in a mellowing country like India. The use of information technology in agriculture can change the scenario of decision-making and farmer scan yield in a better way. For decision-making on overall issues related to the agriculture field; data mining plays a vital role. The survey discussed the role of data mining in terms of the agriculture field. We have also discussed various kinds of soils, several processing techniques in agriculture and soil containment.

X. FUTURE SCOPE

Due to technology become the main role in the field of agriculture there are a lot of methods and approaches to easier agricultural-related tasks. The system "crop prediction using data mining technology" is developed and tested successfully and satisfies all the requirements of the client. The goals that are achieved by the developed system are:

- Simplified and reduced the manual work.
- Large volumes of data can be stored.
- It provides a smooth workflow. We can add a device to get values directly from the soil testing lab to the server. We can add a module if any queries are there, the staff can directly interact with the administrator very easily.

- It helps farmers in providing the historical crop yield record with a forecast reducing the danger management.

ACKNOWLEDGEMENT

The success of this project required a lot of guidance and support from many people and I am extremely privileged to have completed the project successfully. I would wish to thank everyone for his/her guidance. We sincerely thank our principal, Raquel Shah Sir and HOD Tanzeem Sayed Sir for always encouraging us to do our best. We are obliged to our project guide Prof. Krunal Pimple Sir, who supported and constantly supervised us throughout the project and helped us in not only completing this project but also provided us with an ample amount of knowledge that was beneficial. We are thankful for and fortunate enough to get constant encouragement, support, and guidance from all Teaching staff of the IT Department who helped us in completing our project work.

REFERENCES

- [1] Rajeswari and K. Arunesh, "Analysing Soil Data using mining Classification Techniques," Indian Journal of Science and Technology, vol. 9(19), May 2016.
- [2] Jharna Majumdar, Sneha Naraseeyappa and Shilpa Ankalaki Analysis of agriculture data using data mining techniques: application of big data
- [3] Ramesh D, Vishnu Vardhan B. Data mining techniques to agricultural yield data. In an International journal of research in computer and communication engineering. 2013; 2(9).
- [4] Motiur Rahman M, Haq N, Rahman RM. Application of data mining tools for yield in rice prediction on clustered regions of Bangladesh. IEEE. 2014; 2014:8–13.
- [5] Rahmah N, Sitanggang IS. Determination of optimal epsilon value on DB SCAN algorithm for clustering data on peat land hotspots in Sumatra. IOP conference series: earth and environmental. Science. 2016; 31:012012.
- [6] Mila M (2003) Comparing clustering. In: Proceedings of COLT 2003.
- [7] Surabhi Chouhan, Divakar Singh, and Anju Singh, "A Survey and Analysis of Various Agricultural Crops Classification Techniques," 25-30, February 2016.
- [8] Jharna Majumdar, Sneha Naraseeyappa & Shilpa Ankalaki Journal of Big Data volume 4, Article number: 20 (2017)