Journal of Analysis and Computation (JAC)
(An International Peer Reviewed Journal), www.ijaconline.com, ISSN 0973-2861
Volume XVII, Issue I, Jan-June 2023

APPLYING DATA MINING TECHNIQUES ON SOIL FERTILITY PREDICTION

Akanksha Sunil Kadam[1], Karan Manohar Gharale[2], Nripesh Kumar Nrip[3], Dr. Ajay Kumar[4]

[1&2]Student,[3]Assistant Professor, Bharati Vidyapeeth (Deemed to be University) Institute of Management Kolhapur Maharashtra India
[4]Assistant Professor Bharati Vidyapeeth (Deemed to be University) Institute of Management and Research, New Delhi.

ABSTRACT

The techniques of data mining are very popular in the area of agriculture. The advancement in Agricultural research has been improved by technical advances in computation, automation, and data mining. Now a days, data mining is being used in a vast areas like healthcare, insurance, marketing, retail, communication, agriculture. The products of data mining system and domain specific data mining application software's are available for trial made use, but data mining in agricultural on soil datasets is a relatively a young and contemporary research domain. Larger volume of data is harvested along the with the crop harvest in agriculture. Inferring the knowledge from huge volume of data is virtually a difficult task in the current scenario. This research uses the data mining techniques for analysis of soil dataset. This data mining algorithms are used for analysing the soil datasets for classification purposes. The various techniques of data mining are used and compared in this research.

Keywords: data mining, classification, regression, Naive Bayes, Soil Fertility, Soil Nutrient

[1] INTRODUCTION:

Data mining includes the utilization of refined data analysis tools to find previously unknown, valid patterns and relationships in huge data sets. These tools can incorporate statistical models, machine learning techniques, and mathematical algorithms, such as neural networks or decision trees. Thus, data mining incorporates analysis and prediction [1]. Soil testing is a practical and common sense means of using reliable chemical analyses to assess soil nutrient levels and determine how fertilizer use can be improved. Note that there are no words like "exact," "precise," or "accurate" in this definition. A key word is "assess," which means "to

evaluate," "to estimate," or "to set a fixed value."[2] The soil testing laboratories are provided with suitable technical literature on various aspects of soil testing, including testing methods and formulations of fertilizer recommendations [3]. Before starting, farming involves a big amount of research. Right from weather to soil efficiency, everything needs to be tested to ensure the productivity of farms. A soil test lets the farmers understand what amount of nutrients are already present in the soil and how much extra is needed. It helps in bringing increased uniformity of nutrient availability across the fields for more uniform crop growth. It also helps in examining soil structure, soil texture, and colour alo ng with the soil's nutrient level and pH content. With help of these test reports, farmers can decide what can be done to improve the soil quality in their farms [4]. 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. Data Mining process involves

a) Collect, clean, and load the data into data warehouse system
b) Stores the data in multidimensional format
c) Provides information access to analysts and decision makers
d) Analyzation of data using different ap plications
e) Presents the data using different patterns

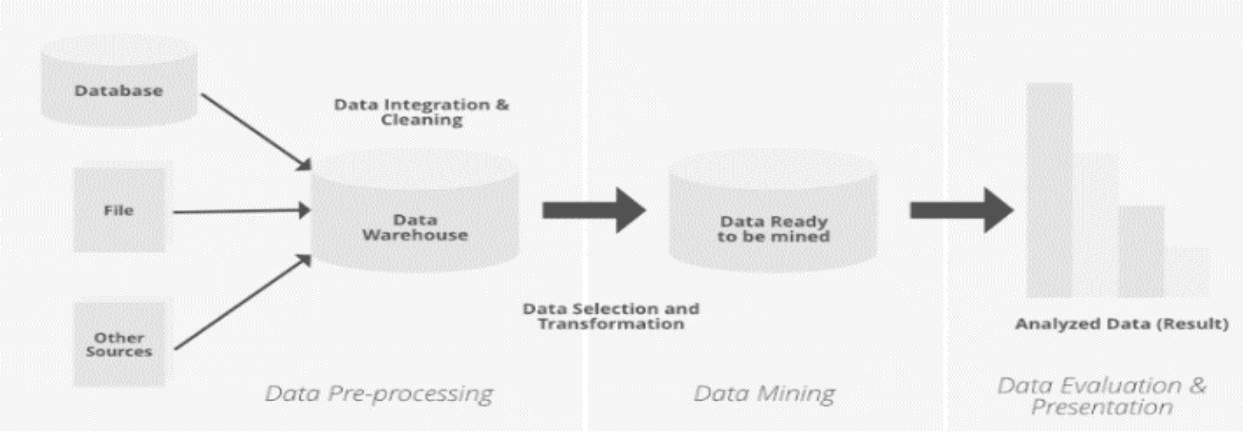


Fig1: Data Mining System Architecture, Source: Prepared by researchers.

[2] RESEARCH METHODOLOGY:

2.1 Dataset Collection:

The pH of the soils of Gaganbawda tehsil of Kolhapur district ranged from 4.86 to 6.94. The soils were strongly acidic to slightly acidic in reaction. Among the soil samples tested, 31 samples (36.48 Per cent) were strongly acidic, 50 samples (58.82 Per cent) were moderately acidic, 4 samples (4.70 Percent) were slightly acidic in nature The acidic reaction of maximum soil of the tehsil might be due to slopy land and undulating topography, high rainfall leading to leaching losses abases from the surface soils and accumulation of iron oxides and sesquioxide. The similar results were also recorded by Mishra et al. (2014). The EC of soils of Gaganbawda tehsil were ranged from 0.10 to 0.32 dSm-1 with average mean value 0.15 dS m-1 (normal). These observations indicate that, all the 100 per cent soils were normal, no saline in nature and suitable for healthy plant growth. The low EC may be due to low temperature, high humidity, light texture of soil, heavy rainfall, high erosion and leaching down of soluble

salts. Singh et al. (2009) [9] also reported the similar finding in Hoshangabad district of Madhya Pradesh.[6]

Table 1: pH, EC and Per cent CaCO3 equivalent status

Particular	pH (1:2.5)	EC (dS m^{-1}) (1:2.5)	Per cent $CaCO_3$ Equivalent
Range	4.86 – 6.94	0.10-0.32	0.37 -2.25
Mean	5.77	0.15	1.31
Category	Strongly acidic 31(36.48%)	Normal 85 (100%)	Barely calcareous 22(25.88%)
	Moderately acidic 50(58.82%)		Slightly calcareous 61(71.77%)
	Slightly acidic 4(4.70%)		Moderately calcareous 2(2.35%)
95% Confidence limit	2.64	0.34	1.88

Source: International Journal of Chemical Studies 2021; 9(1): 596(5)

Table 2: Status of organic carbon and available nitrogen, phosphorus, and potassium status

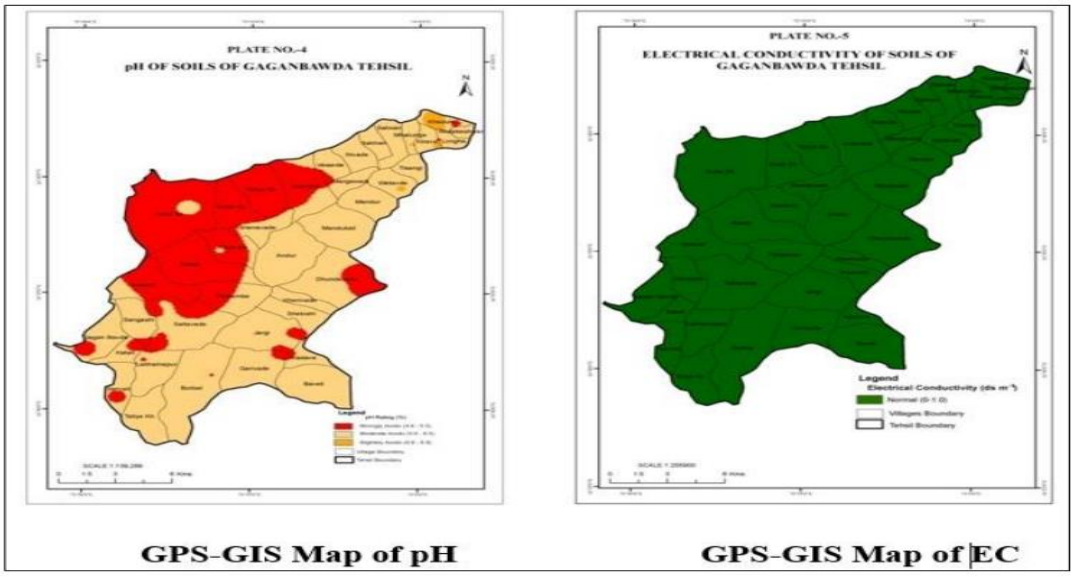
Particular	Organic carbon (%)	Available nutrients (kg ha^{-1})		
		N	P	K
Range	0.58-1.80	151.60-385.72	4.48-14.33	100.80-380.80
Mean	1.20	321.64	11.42	188.95
Very low	-	-	4(4.70%)	-
Low	-	9 (10.59%)	77 (90.60%)	14 (16.47%)
Moderate	2(2.35%)	76 (89.41%)	4 (4.70%)	35 (41.18%)
Moderately High	7 (8.24%)	_	_	32 (37.65%)
High	16 (18.82%)	_	_	2 (2.35%)
Very high	60 (70.59%)	_	_	2(2.35%)
95% Confidence limit	1.22	234.12	9.85	280

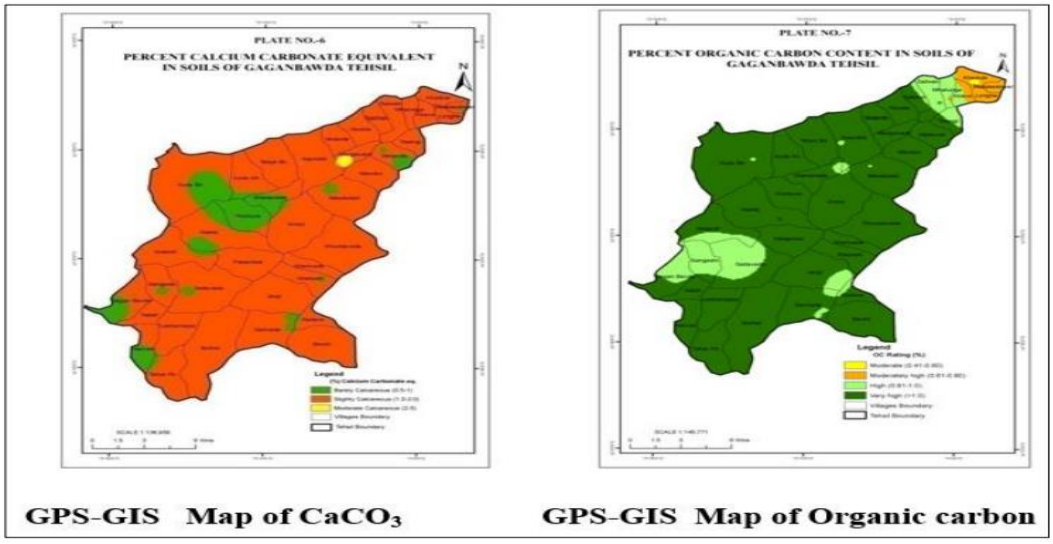
Source: International Journal of Chemical Studies 2021; 9(1): 597(5)

2.2 Literature survey:

Data Mining is applied to analyse large data sets and found useful patterns in the data. Data Mining is used in various fields to recognize patterns which are used in analyzation and prediction. Many studies describe how to analyse agricultural data especially soil information by using classification, regression, correlation, clustering and machine learning. The results of soil analysis on different data sets with a range of Data Mining techniques may useful to farmers to get right insight to perform their activities with less cost and to improve the crop yields, such as by measuring soil properties the farmers can decide what kind of crops to be adopted and use of fertilizers etc. The soil analysis may use in many dimensions such as to protect the environment, diagnosis of crop culture troubles, to identify nutrient deficiencies, energy conversation, and so on. In soil analysis, we can test different properties of soils like pH, organic matter, ammonium N, calcium carbonate equivalency, etc.

[3] Diagrams of GPS GIS maps





[7]

3.1 Usage of Data Mining techniques in various Usage of fields of agriculture:

Data Mining has wide scope within the field of agriculture, particularly on soils to derive information. Numerous techniques of mining and machine learning square measure probably utilized.

i. Predicting trends and behaviour of soils betting on ingredients and climate conditions.
ii. Discovery of soil patterns that are unknown.
iii. Decision trees for crop and soil management.
iv. Artificial Neural Networks for sensing the soils for the adoption of crops.
v. Genetic algorithms used for soil allocation ways.
vi. Nearest neighbour ways for soils classification.

vii. Rule primarily based induction for yield prediction in agriculture for estimation and prediction of farming problems

[4] A COMPARATIVE STUDY OF SOIL CLASSIFICATION:

4.1 Navie Bayes:

A naive Bayes classifier is a simple probabilistic classifier based on applying Bayes theorem with strong (naive) independence assumptions. Depending on the precise nature of the probability model, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. An advantage of the naive Bayes classifier is that it only requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification.[8,9]

4.2 JRip:

This algorithm implements a propositional rule learner, Repeated Incremental Pruning to Produce Error Reduction (RIPPER), which was proposed by William W. Cohen as an optimized version of IREP. In this paper, three classification techniques (naïve Bayes, J48 (C4.5) and JRip) in data mining were evaluated and compared on basis of time, accuracy, Error Rate, True Positive Rate and False Positive Rate. Tenfold cross-validation was used in the experiment. Our studies showed that J48 (C4.5) model turned out to be the best classifier for soil samples[8,9]

[5] CONCLUSION:

In this paper, we have suggested an analysis of the soil data using different algorithms and prediction technique. In spite the fact that the least median squares regression is known to produce better results than the classical linear regression technique, from the given set of attributes, the most accurately predicted attribute was “P” (Phosphorous content of the soil) and which was determined using the Linear Regression technique in lesser time as compared to Least Median Squares Regression. In this paper we have demonstrated a comparative study of various classification algorithms i.e. Naïve Bayes, J48 (C4.5), JRip with the help of data mining tool WEKA. J48 is very simple classifier to make a decision tree, but it gave the best result in the experiment. In future, we contrive to build Fertilizer Recommendation System which can be utilized effectively by the Soil Testing Laboratories. As per the soil sample given to lab for testing and cropping pattern the system will recommend suitable fertilizer.

[6] REFERENCES:

[1]"Data mining techniques", Java T-Point, https://www.javatpoint.com/data-mining-techniques

[2]"passel2.unl.edu", Soil testing is a practical and common sense, https://passel2.unl.edu/view/lesson/f7fb007e892f/1

[3] “Methods Manual-Soil Testing in India”, Department of Agriculture & Cooperation Ministry of Agriculture Government of India, 2011 https://www.scirp.org/(S(351jmbntvnsjt1aadkposzje))/reference/ReferencesPapers.aspx?ReferenceID=1651999

[4]"green.solution.in", (Improves Productiving of Soil), https://green-solution.in/what-is-soil-testing-and-why-is-it-important/

[5] Mapping of macronutrient status in soils of Gaganbawda tehsil of Kolhapur district using GPS-GIS technique P-ISSN: 2349–8528 E-ISSN: 2321–4902 www.chemijournal.com IJCS 2021; 9(1): 595-600 © 2021 IJCS https://www.chemijournal.com/archives/2021/vol9issue1/PartH/8-6-463-258.pdf

[6] Gosavi KD. GPS-GIS based soil fertility maps of Radhanagri Tehsil of Kolhapur district (M. S.) M. Sc. (Agri.) Thesis, M.P. K.V., Rahuri 2016. https://www.semanticscholar.org/paper/Mapping-of-macronutrient-status-in-soils-of-tehsil-Waghmare-Pawar/752a8d5a31e1b64776ac1c942a88f081663bfd1b

[7] K. Samundeeswari, Dr.K.Srinivasan, "Datamining Techniques in Agriculture Prediction of Soil Fertility" ,IJSER,2017 https://www.researchgate.net/publication/325116873_A_Survey_of_using_Data_Mining_Techniques_for_Soil_Fertility

[8] Data Mining Technique to Analyse Soil Nutrients based on Hybrid Classification Analysis of
Soil Nutrients Naïve Bayes, Decision Tree and Hybrid Classification. Manjula and S. Djodiltachoumy
https://www.semanticscholar.org/paper/DATA-MINING-TECHNIQUE-TO-ANALYZE-SOIL-NUTRIENTS-ON-M.-Djodiltachoumy/72021586d4a8b51143968f0cb806f2bd3d51e85f

Author[s] brief Introduction

1. **Akanksha Sunil Kadam:** Student of MCA-II (Sem-III), Bharati Vidyapeeth (Deemed to be University) Institute of Management Kolhapur.
   Mail: akankshakadam3614@gmail.com
2. **Karan Manohar Gharale:** Student of MCA-II (Sem-III), Bharati Vidyapeeth (Deemed to be University) Institute of Management Kolhapur.
   Mail: karangharale3908@gmail.com
3. **Prof. Nripesh Kumar Nrip:** Assistant Professor, Dept. of Computer Application, Bharati Vidyapeeth (Deemed to be University) Institute of Management Kolhapur.
   Email: nripesh.nrip@bharatividyapeeth.edu
4. **Dr. Ajay Kumar**
   Assistant Professor, Dept. of Computer Application, Bharati Vidyapeeth (Deemed to be University) Institute of Management and Research, New Delhi.
   Email: ajay.kumar@bharatividyapeeth.edu