iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** IV **Month of publication:** April 2019

DOI: https://doi.org/10.22214/ijraset.2019.4171

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue IV, Apr 2019- Available at www.ijraset.com

Classification of Soil using Data Mining Technique

Rajalaxmi Hegde[1], SandeepKumar Hegde[2] Suma Aithal[3], Ranjith R[4], Yarra Kavya[5]
[1, 2]*Assistant Professor,* [3, 4, 5]*Student, Computer Science and Engineering, NMAM Institute of Technology, Karkala, India*

Abstract: ***The rapid elevation in technology has created data in plethora which has both required as well as unwanted. In order to truncate the unwanted data and classify, ample of data mining techniques has been implemented. The main goal of data mining is to collect, preprocess and classify the data based on certain features of soil such as PH, Nitrogen, Oxygen and Carbon percent of soil and predict the crops grown for the specific soil type. Data from various countries are gathered and analyzed and were calculated based on pre-defined Fuzzy Rules which returned the precise result at the end as an output.***
Keywords: Data Mining, Pre-defined Fuzzy Rules, traditional classification, acidity, data gathering, pre-processing, feature extraction, predict, attributes/features, mean annual temperature and precipitate

I. INTRODUCTION

Data Mining is one of the most efficient and non-time consuming method for the larger data sets which is highly applicable for the research field and also for agricultural purposes. Traditional classification involves tables, flowcharts etc. It prescients the type of crop grown on particular type of vegetation and also the type of soil with the acidity. The insightfulness of an acidity is based certain attributes which are being considered during the study when the calculation takes place with the help of fuzzy classification rules. In this project, the data set from humongous number ofcountries are taken from an open source platform() which then are applied to perform the four main steps of classification such as data gathering, pre-processing, feature extraction and application of rules for the selected attributes. The effectiveness is compared using the key factors such as precision and accuracy from the processed data by the fuzzy classification rules. These rules are mainly used to determine the contents of carbon, nitrogen, oxygen, sulpher and phosphorous by the previously determined characteristics such as vegetation, country, cultivation, total nitrogen, total phosphorous, latitude, longitude, biome, and PH. This kind of technique is used topredict the type of soil required for the cultivation of crops which helps researchers to know more about the land and soil in detail and also assists them to suggest the type of fertilizers required for controlling the pests for different regions or the countries.

II. OBJECTIVES

The designed application provides a deep insight of the behavior of the soil in a particular region. It provides a feature extraction method to sort the data set according to the users need and thereby determining the necessary values of certain attributes such as PH, Nitrogen, and Phosphorous etc. This enriches the cultivation and thus avoiding any unnecessary commitments towards a region of land.

III. LITERATURE SURVEY

The classification techniques which are acted on the data sets of the soil we possess are JRip & J48, which is a java implementation technique and is one of the decision tree classification rule and Naïve Bayes techniques. This project carried out by V.Rajeshwari and K.Arunesh considering only the red and black types of soil and forecasting the kappa statistics [1]. B.Murugeshkumar, Dr. K.anandkumar, Dr. A.Bharathi [2] conducted a survey on soil classification with the data mining technologies such as USDA soil taxonomy and the UNESCO system FAO. The survey involved different mining techniques which is relevant to the subject and thus comparing the results of all possible outcomes those algorithms such as Decision tree, K-means clustering and also includes a study of artificial neural network. To identify the fertility prediction of the soilthe team has used Naïve Bayes and also compared its result with the results of J48 algorithm which enriches the cultivation of crops [3]. Madhuri Kommineni, Someswari Perla, Divya bharathi Yedla proposed a survey on the different techniques of soil classification which involved classified soil types such as Red Gravels , Black Cotton, Sandy Alluvial, Saline alkali. The role of the methods for classification is on the improvement of the soil fertility and the nutrient analysis [4].The study of comparing various mining0 techniques is done in a usual research purposes and this researches determine the accuracy of those methods. Bharathi and Dr.S.Jyothi made a study on Fuzzy algorithm and compared it to with the GA Tree which in turn resulted in 99% accuracy for fuzzy classification rules whereas the GA Tree algorithm doesn't meet the part of the requirements [5]. Warangal region had a different set of soil types and this analysis taken place where data mining plays a major role for determining minerals such as Phosphorus (P) for protein synthesis, Nitrogen (N) to produce chlorophyll in plants, Calcium (Ca) for chemical balance, Magnesium (Mg) considered as great plant growth promoter, Iron (Fe) and Humus. By the usage of linear regression the attributes are detected for that particular region [6].

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue IV, Apr 2019- Available at www.ijraset.com

III. PROPOSED SYSTEM

The proposed system concentrates on generating rules using fuzzy classification for the input attributes and its corresponding membership functions. When the working environment is having various dimensions eventually its complexity increases and his complexity can be reduced using fuzzy classification technique. The decision makers can easily focus on the content they are interested on under certain context in the sub-classifications by grouping the attributes. Fuzzy classification is straight forward as it can interpret the results only through the Boolean values either true or false. This is achieved by mapping propositional functions PF of the membership functions with the universe of discourse U and which in turnderives the set of truth values T. In our project we use the formulas to calculate the percentage of nitrogen, carbon and phosphorous in total and predict the soil type. It is mentioned below.

IV. SYSTEM ARCHITECTURE

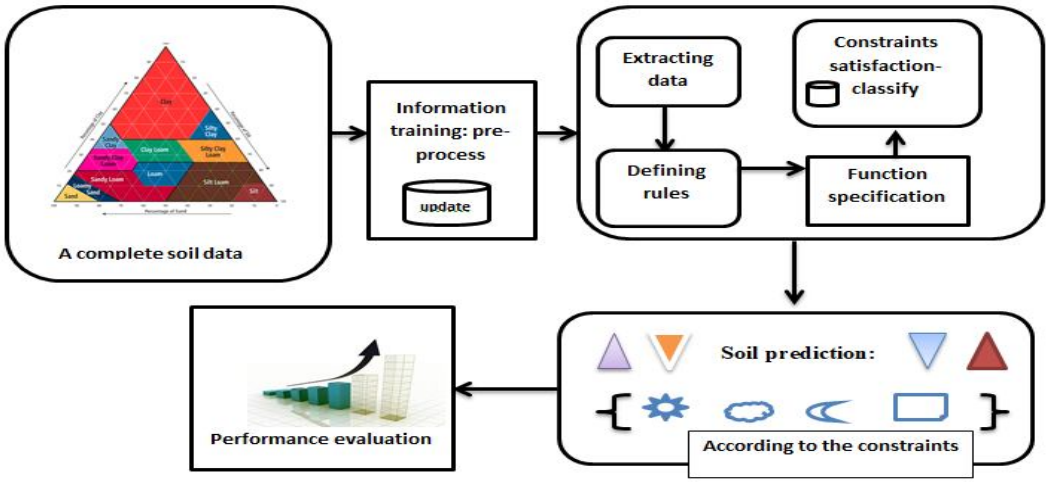


Figure 1. Data Flow Diagram of the proposed work

A. *Modules Description*

1) *Soil Data Collection:* The mixture minerals and organic matter are primarily considered as the factors on which the different kind of soils are collected by also considering the amount of water and air in the mixture of soils that are collected. As the location changes the deposition of the mixture of minerals and organic matters will differ in their composition and it changes constantly. Input can be gathering as text file or excel file format to train the further progress. According to the weather, location area soil has been categorized by making all values segregate to analyzing work.
2) *Input Data Pre-Processing:* Pre-process is the process of training the dataset for the purpose of interrupting into the core work. Training the data will get the process of removing the missing content, irrelevant information from the valuable data. The entire data updating soil values has been discovering in the basis of machine learning techniques further moving on for classification work.
3) *Fuzzy Based Classification:* The soil data is classified by defining its corresponding membership functions. An expression which contains one or more variables with some values, then the expression becomes a fuzzy propositional function. Thus, Fuzzy set is made by grouping of the individuals and processing those individuals based on their characteristics. Fuzzy usually shows the values either exactly true or exactly false but the extended fuzzy technique supports the concept of partial truth.
4) *Predicting Soils:* The fuzzy classification techniques decide a set of rules which is an important task and it has to be designed for the training data set without a knowledge of the problem which is supposed to be classified. Values can be easily predicted by classified soil forms. After process the Classification, soil data has been successfully updated into the database for the further references.
5) Performance Evaluation: This modules enables users get to know graphically the value of entire progress. By showing the mathematical evaluation for the entire classification work by concluding the better performances. Reducing time on current working instances will show the integrity of project.Time consumption and accuracy has been screening with comparing the other classification development as improvement.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue IV, Apr 2019- Available at www.ijraset.com

V. OUTCOMES

The soil data set sample is taken from the open source platform which consists of the soil features from large number of countries. The sample has variety of characteristics such as latitude, longitude, elevation, country, MAT, MAP, and also has 623 instances of soil from different countries. The screen shot of the gathered data is as shown below.

	country	biome	latitude	longitude	elevation	vgtn	MAT	MAP	carbon	nitrogen	phos	org_carb	tot_nit	org_phos	Ph		date	upperdep	depth	note	meth
2	Australia	Bare soil	-32.35	116.07	_	Restored	_	_	5.33	0.5	_	2241.67	45	_	5.97	_	0	10	0~10	_	od
3	China	Bare soil	29.72	117.98	147-350	copper mi	16.2	1335.9	6.88	1.09	_	684.17	125.71	_	4.79	_	0	20	0~20	_	FE
4	China	Bare soil	29.72	117.98	147-350	copper mi	16.2	1335.9	13.2	1.48	_	721.67	131.43	_	4.89	_	0	20	0~20	_	FE
5	China	Bare soil	29.72	117.98	147-350	copper mi	16.2	1335.9	17.96	1.92	_	752.5	137.14	_	4.97	_	0	20	0~20	_	FE
6	China	Bare soil	29.72	117.98	147-350	copper mi	16.2	1335.9	19.37	1.96	_	1037.5	146.43	_	4.99	_	0	20	0~20	_	FE
7	China	Bare soil	29.72	117.98	147-350	copper mi	16.2	1335.9	53.88	4.75	_	1073.33	149.29	_	5.02	_	0	20	0~20	_	FE
8	China	Bare soil	29.72	117.98	147-350	copper mi	16.2	1335.9	60.7	5.01	_	1027.5	154.29	_	5.49	_	0	20	0~20	_	FE
9	China	Bare soil	29.72	117.98	147-350	copper mi	16.2	1335.9	74.29	5.99	_	1159.17	146.43	_	5.69	_	0	20	0~20	_	FE
10	China	Bare soil	29.72	117.98	147-350	copper mi	16.2	1335.9	81.63	6.47	_	1177.5	153.57	_	5.76	_	0	20	0~20	_	FE
11	India	Bare soil	24.08	83.62	300	Xanthium	20	1263	20.58	_	0.74	_	55.71	12.58	7.8	_	0	10	0~10	_	FI
12	India	Bare soil	24.18	83.01	299-416	Mine spoi	_	_	27.67	1.93	0.39	_	55	10.97	7.9	_	0	10	0~10	_	FI
13	United kir	Bare soil	51.98	-0.62	_	_	_	_	14.28	_	_	1041.67	92.86	_	7.05	1990: July	0	23	0~23	Low-meta	FE
14	United kir	Bare soil	51.98	-0.62	_	_	_	_	7.33	_	_	1350	114.29	_	7.05	1990: July	0	23	0~23	High-met	FE
15	Canada	Boreal For	55.97	-120.47	720	spruce: pi	_	_	100.58	6	_	3650	200	_	_	2002: Sept	0	3	0~3	mature fo	FE

Figure 2. Snapshot of the gathered data set and its attributes

The pre-processing is mainly a process in which the undefined value is considered as 0.0 for the calculation to be easier on the further implementation and the vegetation which are not defined are considered to be an apostrophe. The unwanted data from the data set is removed and only certain features are taken into the consideration. The screenshot the above process is shown at the bottom.

Cou...	Biome	Latit...	Long...	Elev...	Vege...	MAT	MAP	Soil_...	Soil_...	Soil_...	Soil_...	Total...	Total...	pH	Date	Upp...	Low...	Depth	Note	Meth...
Austr...	Crop...	-34.42	150.67	0.0	not d...	0.0	0.0	36.33	0.0	0.0	1000	0.0	0.0	0.0	0.0	0	10	0~10	0.0	FE
Austr...	Crop...	-36.33	145.7	0.0	not d...	0.0	0.0	22.67	0.0	0.0	1666...	0.0	0.0	0.0	0.0	0	10	0~10	0.0	FE
Austr...	Crop...	-38.2	147.32	0.0	not d...	0.0	0.0	19.5	0.0	0.0	4666...	0.0	0.0	0.0	0.0	0	10	0~10	0.0	FE
Austr...	Crop...	0.0	0.0	0.0	not d...	0.0	0.0	18	0.0	0.0	1000	0.0	0.0	0.0	0.0	0	10	0~10	0.0	FE
Austr...	Crop...	0.0	0.0	0.0	not d...	0.0	0.0	38.42	0.0	0.0	2833...	0.0	0.0	0.0	0.0	0	10	0~10	0.0	FE
Austr...	Crop...	-35.88	140.05	0.0	not d...	0.0	0.0	46.42	0.0	0.0	1500	0.0	0.0	0.0	0.0	0	10	0~10	0.0	FE
Austr...	Crop...	-28.2	152.1	0.0	not d...	0.0	685	0.0	4.59	0.0	1458...	94.29	0.0	7.88	1988...	0	10	0~10	no fe...	FI
Austr...	Crop...	-28.2	152.1	0.0	not d...	0.0	685	0.0	4.79	0.0	1475	100.71	0.0	7.72	1988...	0	10	0~10	23 kg...	FI
Austr...	Crop...	-28.2	152.1	0.0	not d...	0.0	685	0.0	4.66	0.0	1475	102.86	0.0	7.6	1988...	0	10	0~10	69 kg...	FI
Austr...	Crop...	-28.2	152.1	0.0	not d...	0.0	685	0.0	5.18	0.0	1441...	93.57	0.0	7.69	1988...	0	2.5	0~2.5	no fe...	FI
Austr...	Crop...	-28.2	152.1	0.0	not d...	0.0	685	0.0	5.18	0.0	1450	100.71	0.0	7.44	1988...	0	2.5	0~2.5	23 kg...	FI

Figure 3.Snapshot of the dataset after feature extraction

Fuzzy logic represents the non-statistical uncertainties using the pre–defined rules. The Fuzzy Logic captures the uncertainties associated with thinking and reasoning of a human. In this kind of technique where, the rules are applied to determine the pH value of the soil which is cognitive process similar to a human mind by training the data by the fuzzy rules.

The fuzzy rules are implemented by the java programming language. Fuzzy classification rules allow readable and interpretable rule bases and considering it as best suited for classification problem with uncertainty. The fuzzy classification techniques decide a set of rules which is an important task and it has to be designed for the training data set. In the project, value of carbon biomass to that of an organic carbon is rationed to get the actual carbon percentage of the sample data set. Similarly for the nitrogen and phosphorous percentage estimation that returns the label of acidity in soil which could range from ultra-acidic to base level.

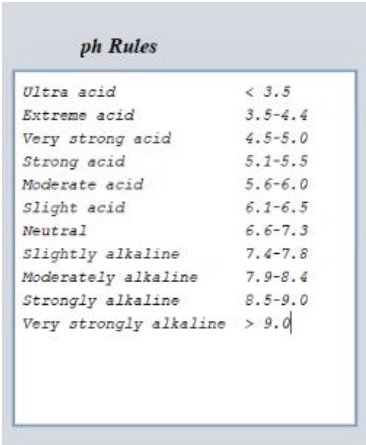


Figure 4. Fuzzy classification pH rules

The graph is plotted based on the precision and recall method where the percentage of previously estimated features such as carbon, nitrogen and phosphorous are considered. The percentage is expressed as the ratio of the total number of the retrieved records that are relevant to the context to the number of relevant records in the database. From the study, it is seen that, recall value is less if there is no much relevant record to be retrieved from the database. For different data sets, different precision and recall values are obtained in the form of percentages. In our project, the present data set has the precision value of 7% and recall value as 75% which is highly accurate. The performance of implemented fuzzy classification algorithm is screenshot below.

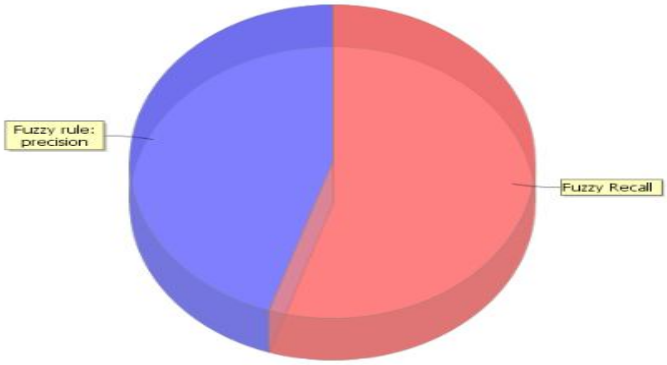


Figure 5. The Graph plotted after calculating the precision and recall according to the fuzzy rules.

VI. RESULT AND CONCLUSION

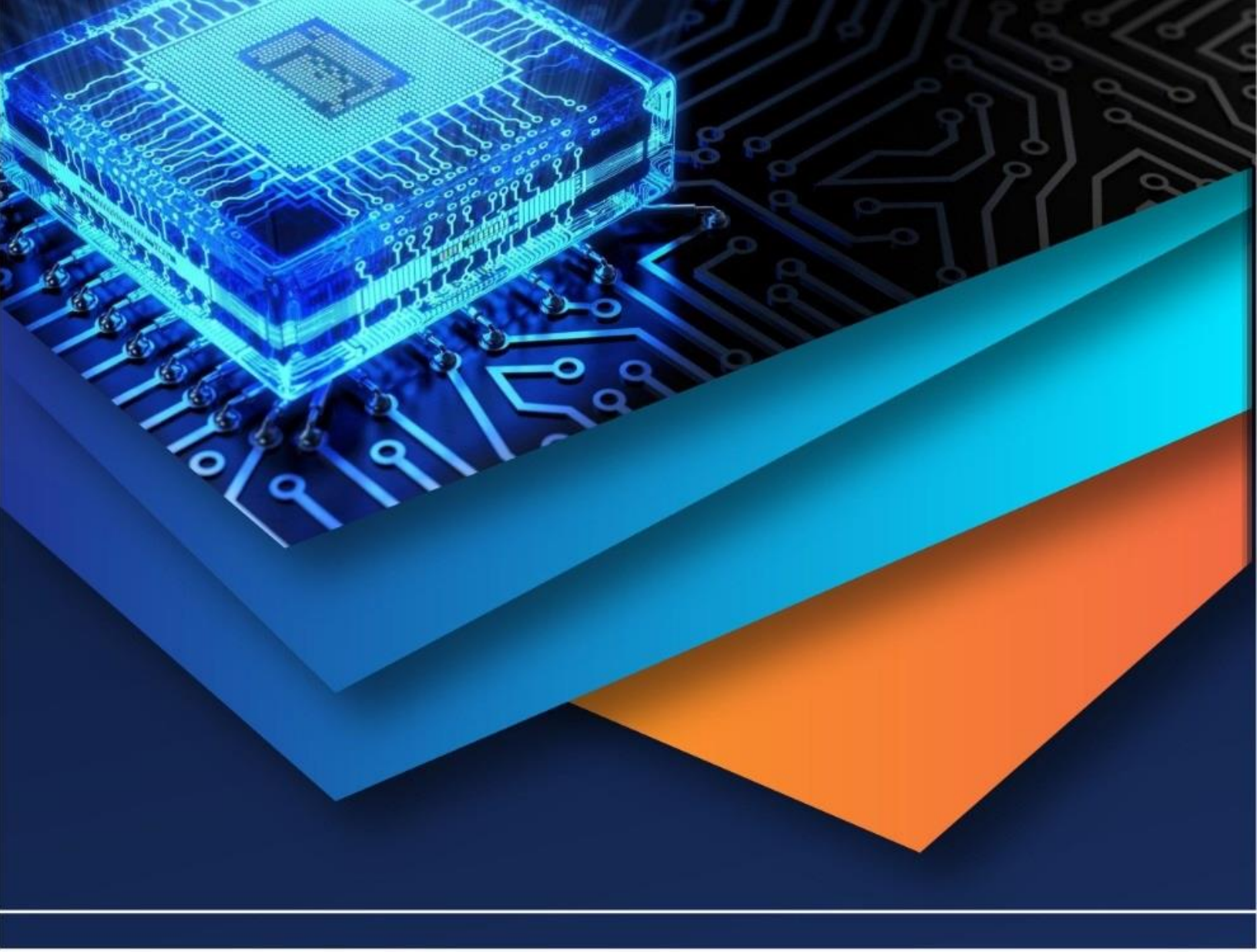
In this project, we have calculated the percentage of nitrogen, carbon and phosphorous using the fuzzy rules which then estimates the performance of the classification algorithm in the form of precision and recall. The rules could only be applied for the supervised learning data set. This helps to classify the types of soil on the basis of certain features of soil such as biomass of carbon, nitrogen, phosphorus, biomes and vegetation. This is benevolent for the researchers to help them estimate the type of soil required for cultivation of crops on the different varieties of land.

VII.FUTURE WORK

In this paper, the type of soil for cultivation of varieties of crops is determined using fuzzy pre-defined rules and the performance is estimated as well. But in future, we are planning to work on the project to help the soil testing laboratories with prediction of agricultural fertilizers for the crops to grow efficiently.

REFERENCES

[1] V.Rajeshwari and K. Arunesh, Analyzing Soil Data using Data Mining Classification Techniques,Indian Journal of Science and Technology, Vol 9(19), DOI: 10.17485/ijst/2016/v9i19/93873, May 2016
[2] B. Murugesh Kumar, Dr. K Ananda Kumar, Dr. A Bharathi A survey on soil classification methods using data mining techniques, International Journal of Current Trends in Engineering & Research (IJCTER) e-ISSN 2455–1392 Volume 2 Issue 7, July 2016
[3] S.S.Baskar, L.Arockiam, S.Charles, Applying Data Mining Techniques on Soil Fertility Prediction, International Journal of Computer Applications Technology and Research Volume 2– Issue 6, 660 - 662, 2013, ISSN: 2319–8656
[4] Madhuri Kommineni, Someswari Perla, Divya Bharathi Yedla, A Survey of using Data Mining Techniques for Soil Fertility, International Journal of Engineering & Technology, 7 (2.7) (2018) 917-918
[5] P. Bhargavi and Dr. S. Jyothi, Soil Classification Using Data Mining Techniques: A Comparative Study,International Journal of Engineering Trends and Technology- July to Aug Issue 2011
[6] VelidePhanikumar and Lakshmi Velide, Data mining plays a key role in soil data analysis of Warangal region, Internatio Journal of Scientific and Research Publications, Volume 4, Issue 3, March 2014



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY