International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-12, October 2019

Feature Selection towards Soil Classification in the context of Fertility classes using Machine Learning

Janmejay pant, Pushpa Pant, Ashutosh Bhatt, Himanshu Pant, Nirmal Pandey

Abstract: ***Soil is recognized as one of the most valuable entity which is responsible for sustaining life on the earth. It is clear that Indian economy is largely dependent on agriculture. In India, for a large section of the population the sole or major source of earnings is agriculture. There are many factors which are responsible in yield production and also affect agriculture. Soil is one of them and plays a crucial role in yield production. Soil nutrients are important aspects that contribute to soil fertility. In this paper we classify soil in the form of fertility index level using machine learning. We induced rule model on training data and apply this model on test data for making predictions for fertility level classes. In this paper we have used Rough Set method to classify the data based on fertility level class and calculate the accuracy of the used classifier.***

Keywords: Feature Selection, Fertility Index, Rough Set, Soil nutrients, Reduct, Rule

I. INTRODUCTION

Soil nutrients are a crucial property that contributes to the soil fertility and other environment factors [1]. In agriculture it is important to enhance yield or crop production which is dependent on soil management. Enhancing the yield production depends and is directly affected by soil parameters or nutrients such as pH value, organic matters, electrical conductivity, P, K etc. Agriculture is fully dependent on soil quality. But due to heavy agriculture production the quality of soil may decrease and the nutrients present in soil may lost. In recent years machine learning has been very helpful in studying the elimination of nutrients in soil. Nowadays soil classification and prediction problems are easily handled by Machine Learning techniques [2]. Various Machine Learning techniques were used to predict the soil nutrients, soil type and soil moisture [2], [3].

Revised Manuscript Received on October 30, 2019.
***** Correspondence Author
Janmejay Pant*, Dept. of Computer Science, Graphic Era Hill University, Bhimtal, India., Email: jpant@gehu.ac.in
Puspha Pant, Dept. of Geography, M.B. Govt. P.G College, Haldwani, India, Email: pushpapant01@gmail.com
Ashutosh Bhatt, Dept. of Computer Science , BIAS, Bhimtal, India, Email:ashutoshbhatt123@gmail.com
Himanshu Pant, Dept. of Computer Science, Graphic Era Hill University, Bhimtal, India, Email: himanshupantbareilly@gmail.com
Nirmal Pandey, Dept. of Computer Science and Applications, Amrapali Group of Institute Haldwani, India, Email:nirmal.bias@gmail.com

Soil fertility is affected by many factors like air, water, organic matter and nutrients. Due to some factors, like use of fertilizers, pesticides, insecticides, cultivation etc, and the fertility of soil is regularly deceasing in recent times [4]. In India there is deficiency of primary nutrients in agricultural soil. In this regard it would be better to utilize the different type of land such as saline, alkaline wastelands by analyzing different soil fertility parameters such as fertility index and soil depletion index [4]. .

II. DATA SET DESCRIPTION

In order to classify soil data based on fertility level, the used data set was collected from one of the soil testing laboratories in Almora district (Uttarakhand).We have used Rosetta tool for generating the rules through which we can further classify our data set and we make prediction for the test data. Based on the soil properties the fertility of soil is defined in three different levels. These are Level 1, Level 2 and Level 3.

We have collected the 53 samples of soil from Almora district (Uttarakhand) by random sampling. The complete data is taken from three different sites corresponding to three different varieties of oak trees growing there, namely Banj oak, Tlionj oak and Kharsu oak. Some samples are distributed and some are undistributed. The primary data was sent to soil testing laboratory for measuring the properties of soil. The data set has 11 independent attributes and 1 is dependent or target attribute which is fertility level. Table 1 describes data collected from each soil sample.

Table I: Attribute Description

Attribute	Description
Depth	in cm from sample is collected
sand	Sand particles in soil ,%
slit	Slit particles in soil ,%
clay	clay particles in soil ,%
moisture	Moisture present in soil,%
pH	pH value of soil
C	Organic carbon,%
N	Nitrogen,%
OM	Organic Matter,%
C:N	Carbon and Nitrogen Ratio in soil
Fertility Level	Dependent class

III. MATERIALS AND METHODS

Machine Learning method is used for classification. For this the following concepts are used:

A. Feature Selection

FS technique [5], [6] is used to obtain a reduced representation of the data set that is smaller in volume in comparison to the original one. The reduced features also maintain the integrity of the original data. Now the feature selection technique is expected to identify most significant soil properties specifying the soil fertility [7].

The feature selection method is implemented to extract the most significant soil property for soil fertility level class. In this paper we have used one machine learning techniques that is Rough Set [5], [7], [8]. Feature Selection has four major steps [7]-

(i) Generation Procedure
(ii) Evaluation Subset
(iii) Stopping Criterion
(iv) Validation Procedure.

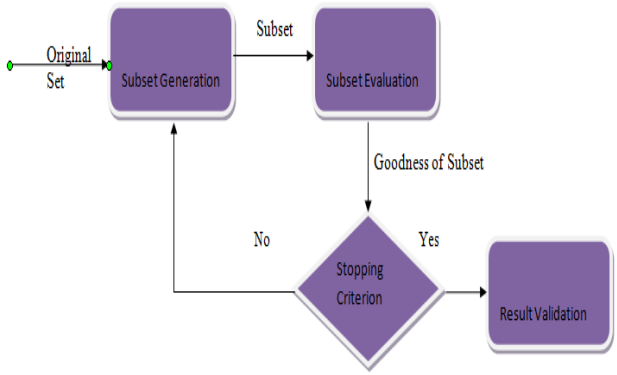


Fig. 1.The feature selection implementation [8].

In this paper, after extracting the important features we have developed an automated system for generating the rules which help in soil classification based on soil fertility.

B. Rough Set Theory

RST [12], [14 is one of the most important mathematical tools to deal with uncertainty which may arises due to granularity in domain of discourse or universe. Rough set is also useful to deal vagueness of data. It is a mathematical framework for analyzing tabular data. There are various basic concepts defined by Z. Pawlak [12], [14]. Some of them are:

Information System (IS)

The Appropriate data which is basically used for research work and is arranged in a tabular format is known as information System (IS) [9]. An information system can be formally defined as: Let A=(A1, A2, A3…, Ak) be a non empty finite set of attributes and U={(a1,a2,…,.ak)} be non empty finite set of k-tuples , known as the objects. V (Ai) denotes the set of values for the attributes Ai [9]. Then information system is defined as an ordered pair I(U, A) such that for all i=1,2,…k there is a function fi [9], [5] . This means, every object in the set U has an attribute value for every element in the set A. The set U is called the universe of the information system [8], [9], [12]. Thus, we can say that Information System is a data table. Let us assume that I= (U, A, C, D) [5] is a decision table or decision system. Here, U is universe of finite set, A is an attribute of data table I and C and D are subsets of A. C is a conditional subset of A and D is decision attribute. Va is the value of set of a, the elements of U are objects [5], [8], [9], [12].

Indiscernibility Relation

Let I = (U, A) be an information system where U= {(a1…ak)} is the non empty finite set of k tuples known as the object and A= {A1…Ak} is a non empty finite set of attributes [10],[12]. Let be a subset of attributes [10], [12]. Then the set of P-indiscernible objects is defined as the set of objects having the same set of attribute values [10], [12].

The RST treats data in terms of equivalence classes this means set of objects that are indiscernible with respect to attributes [8], [12].

Reducts

The objects of an information system can be divided into a group of similar types, i. e, equivalence classes [8], [10]. An Indiscernibility relation can do this task. However, the entire set of attributes, A, may not be necessary to preserve the Indiscernibility among the set of indiscernible object [8], [9]. We can say that there may a subset that is sufficient to maintain the classification based on Indiscernibility. A minimal set of attributes required to preserve the Indiscernibility relation among the objects of an information system is called a reduct. Let Say, given an information system a reduct [8], [9], [11] is a minimum group of attributes such that [8], [9].

Decision Rules

After getting reducts, rules may be generated from reducts for classification of the objects. IF-THEN rules [9], [12] are constructed by reading of the values for each attribute in the reduct and associating them with one or more decision classes. The THEN-part will only include one decision class unless the decision class is rough with respect to the attributes in the reduct. Rules are evaluated according to how general they are and how specific they are. How general they are in term of coverage and how specific they are in terms of accuracy.

C. Fundamental Characteristics of Rough Set Theory[8]

1. This theory doesn't require any prelude and additional information about data. Other similar theories need the same information as preliminary, for example, statistics needs probability, fuzzy set theory needs grade of membership as prior information. But Rough set theory does not require these types of information as prior.
2. For extracting the hidden patterns form the data rough set provides efficient and intelligent methods, algorithms and tools.
3. By using rough set we can reduce the original data to find the minimal sets which contain the identical knowledge as in the initial data, i.e. we can reduce the dimension or size of original data without losing the information which is derived by the original data. For this we can use feature selection.
4. It allows calculating the importance of data.
5. Rough set allows producing the set of decision rules in automatic way.

IV. METHODOLOGY

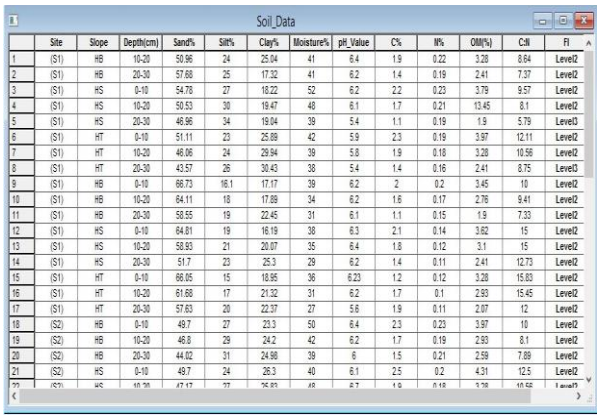
In this paper we have tried to describe how Rough Set theory is useful for feature selection of soil data.

The whole data, which is stored in tabular form, is split in two parts one is training set and the other is test set. The division of data set in two parts is useful for classification and prediction. The training set is classified using rough set to find reducts and rules. Once we get rules, we then apply the rules in test data to make some predictions and classify the test data. Researchers are using Rough set theory as a classifier to extract the relevant attributes to achieve the accurate and exact classification results. Feature selection is basically finding the reducts from data to reduce the volume of the data without losing the information.

In this paper we find the reducts, generate rules and obtain classification results for the soil sample based on fertility levels.

V. EXPERIMENT AND RESULTS

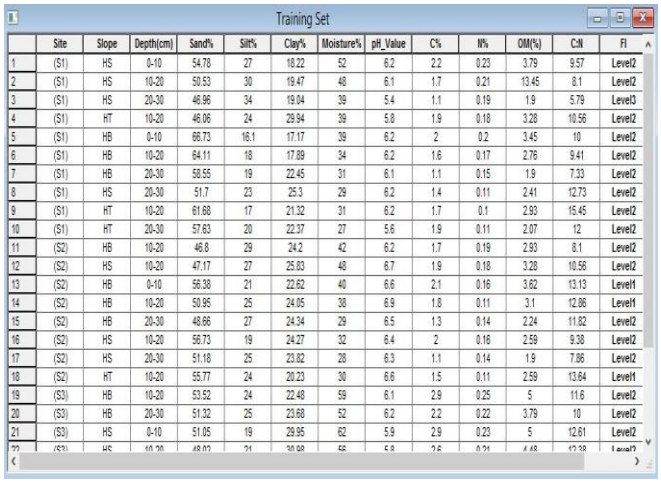
We have used rough set in ROSTTA tool to apply our soil samples of different locations of hilly areas in Almora district in Uttarakhand. As we have already discussed in methodology section, we split the data in training and testing parts. For the data we induced a model that is rule model on one part of the data, i.e., induce rule model on training part. We used the same rule model to classify the objects in remaining data, i.e., used the generated rule model on test part. In this section we describe the data used in our experiments to generate results using Rough Set theory.


Soil_Data

	Site	Slope	Depth(cm)	Sand%	Silt%	Clay%	Moisture%	pH_Value	C%	N%	OM(%)	C:N	FI
1	(S1)	HB	10-20	50.96	24	25.04	41	6.4	1.9	0.22	3.28	8.64	Level2
2	(S1)	HB	20-30	57.68	25	17.32	41	6.2	1.4	0.19	2.41	7.37	Level2
3	(S1)	HS	0-10	54.78	27	18.22	52	6.2	2.2	0.23	3.79	9.57	Level2
4	(S1)	HS	10-20	50.53	30	19.47	48	6.1	1.7	0.21	13.45	8.1	Level2
5	(S1)	HS	20-30	46.96	34	19.04	39	5.4	1.1	0.19	1.9	5.79	Level3
6	(S1)	HT	0-10	51.11	23	25.89	42	5.9	2.3	0.19	3.97	12.11	Level2
7	(S1)	HT	10-20	46.06	24	29.94	39	5.8	1.9	0.18	3.28	10.56	Level2
8	(S1)	HT	20-30	43.57	26	30.43	38	5.4	1.4	0.16	2.41	8.75	Level3
9	(S1)	HB	0-10	66.73	16.1	17.17	39	6.2	2	0.2	3.45	10	Level2
10	(S1)	HB	10-20	64.11	18	17.89	34	6.2	1.6	0.17	2.76	9.41	Level2
11	(S1)	HB	20-30	58.55	19	22.45	31	6.1	1.1	0.15	1.9	7.33	Level2
12	(S1)	HS	0-10	64.81	19	16.19	38	6.3	2.1	0.14	3.62	15	Level2
13	(S1)	HS	10-20	58.93	21	20.07	35	6.4	1.8	0.12	3.1	15	Level2
14	(S1)	HS	20-30	51.7	23	25.3	29	6.2	1.4	0.11	2.41	12.73	Level2
15	(S1)	HT	0-10	66.05	15	18.95	36	6.23	1.2	0.12	3.28	15.83	Level2
16	(S1)	HT	10-20	61.68	17	21.32	31	6.2	1.7	0.1	2.93	15.45	Level2
17	(S1)	HT	20-30	57.63	20	22.37	27	5.6	1.9	0.11	2.07	12	Level2
18	(S2)	HB	0-10	49.7	27	23.3	50	6.4	2.3	0.23	3.97	10	Level2
19	(S2)	HB	10-20	46.8	29	24.2	42	6.2	1.7	0.19	2.93	8.1	Level2
20	(S2)	HB	20-30	44.02	31	24.98	39	6	1.5	0.21	2.59	7.89	Level2
21	(S2)	HS	0-10	49.7	24	26.3	40	6.1	2.5	0.2	4.31	12.5	Level2

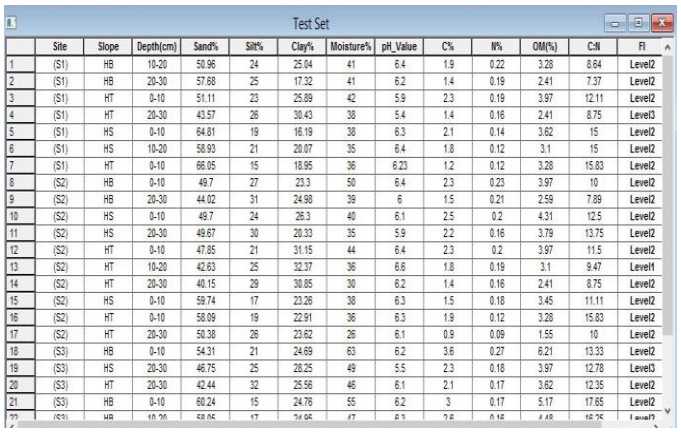
Fig. 2 Demonstration of Soil Data

As stated above we have collected 53 soil samples from the different location of hilly area in Almora district, Uttarakhand. A snapshot of this data set is in Fig 2. In this data the different sites from where we collected the samples were renamed: Banjoak site renamed as S1, Tlionj oak site renamed as S2 and Kharsu oak site renamed as S3. The above data is split out in two parts with spilt factor 0.5. This means 50 percent of data will spilt in training part and rest 50 percent will split into test part.


Training Set

	Site	Slope	Depth(cm)	Sand%	Silt%	Clay%	Moisture%	pH_Value	C%	N%	OM(%)	C:N	FI
1	(S1)	HS	0-10	54.78	27	18.22	52	6.2	2.2	0.23	3.79	9.57	Level2
2	(S1)	HS	10-20	50.53	30	19.47	48	6.1	1.7	0.21	13.45	8.1	Level2
3	(S1)	HS	20-30	46.96	34	19.04	39	5.4	1.1	0.19	1.9	5.79	Level3
4	(S1)	HT	10-20	46.06	24	29.94	39	5.8	1.9	0.18	3.28	10.56	Level2
5	(S1)	HB	0-10	66.73	16.1	17.17	39	6.2	2	0.2	3.45	10	Level2
6	(S1)	HB	10-20	64.11	18	17.89	34	6.2	1.6	0.17	2.76	9.41	Level2
7	(S1)	HB	20-30	58.55	19	22.45	31	6.1	1.1	0.15	1.9	7.33	Level2
8	(S1)	HS	20-30	51.7	23	25.3	29	6.2	1.4	0.11	2.41	12.73	Level2
9	(S1)	HT	10-20	61.68	17	21.32	31	6.2	1.7	0.1	2.93	15.45	Level2
10	(S1)	HT	20-30	57.63	20	22.37	27	5.6	1.9	0.11	2.07	12	Level2
11	(S2)	HB	10-20	46.8	29	24.2	42	6.2	1.7	0.19	2.93	8.1	Level2
12	(S2)	HS	10-20	47.17	27	25.83	48	6.7	1.9	0.18	3.28	10.56	Level2
13	(S2)	HB	0-10	56.38	21	22.62	40	6.6	2.1	0.16	3.62	13.13	Level1
14	(S2)	HB	10-20	50.95	25	24.05	38	6.9	1.8	0.11	3.1	12.86	Level1
15	(S2)	HB	20-30	48.66	27	24.34	29	6.5	1.3	0.14	2.24	11.82	Level2
16	(S2)	HS	10-20	56.73	19	24.27	32	6.4	2	0.16	2.59	9.38	Level2
17	(S2)	HS	20-30	51.18	25	23.82	28	6.3	1.1	0.14	1.9	7.86	Level2
18	(S2)	HT	10-20	55.77	24	20.23	30	6.6	1.5	0.11	2.59	13.64	Level1
19	(S3)	HB	10-20	53.52	24	22.48	59	6.1	2.9	0.25	5	11.6	Level2
20	(S3)	HB	20-30	51.32	25	23.68	52	6.2	2.2	0.22	3.79	10	Level2
21	(S3)	HS	0-10	51.05	19	29.95	62	5.9	2.9	0.23	5	12.61	Level2

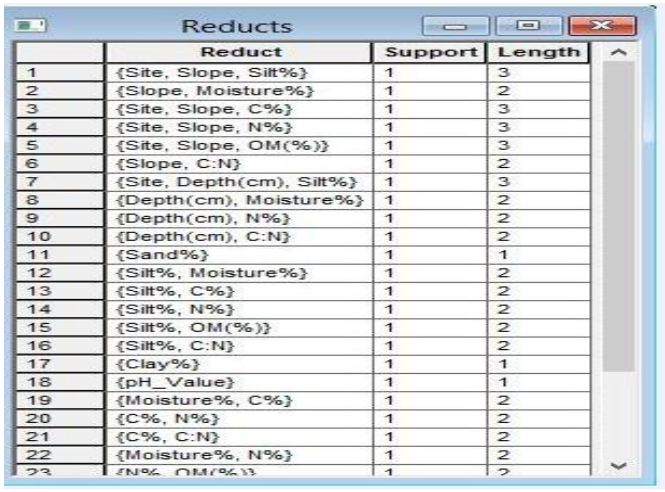
Fig. 3 Training Set


Test Set

	Site	Slope	Depth(cm)	Sand%	Silt%	Clay%	Moisture%	pH_Value	C%	N%	OM(%)	C:N	FI
1	(S1)	HB	10-20	50.96	24	25.04	41	6.4	1.9	0.22	3.28	8.64	Level2
2	(S1)	HB	20-30	57.68	25	17.32	41	6.2	1.4	0.19	2.41	7.37	Level2
3	(S1)	HT	0-10	51.11	23	25.89	42	5.9	2.3	0.19	3.97	12.11	Level2
4	(S1)	HT	20-30	43.57	26	30.43	38	5.4	1.4	0.16	2.41	8.75	Level3
5	(S1)	HS	0-10	64.81	19	16.19	38	6.3	2.1	0.14	3.62	15	Level2
6	(S1)	HS	10-20	58.93	21	20.07	35	6.4	1.8	0.12	3.1	15	Level2
7	(S1)	HT	0-10	66.05	15	18.95	36	6.23	1.2	0.12	3.28	15.83	Level2
8	(S2)	HB	0-10	49.7	27	23.3	50	6.4	2.3	0.23	3.97	10	Level2
9	(S2)	HB	20-30	44.02	31	24.98	39	6	1.5	0.21	2.59	7.89	Level2
10	(S2)	HS	0-10	49.7	24	26.3	40	6.1	2.5	0.2	4.31	12.5	Level2
11	(S2)	HS	20-30	49.67	30	20.33	35	5.9	2.2	0.16	3.79	13.75	Level2
12	(S2)	HT	0-10	47.85	21	31.15	44	6.4	2.3	0.2	3.97	11.5	Level2
13	(S2)	HT	10-20	42.63	25	32.37	36	6.6	1.8	0.19	3.1	9.47	Level1
14	(S2)	HT	20-30	40.15	29	30.85	30	6.2	1.4	0.16	2.41	8.75	Level2
15	(S2)	HS	0-10	59.74	17	23.26	38	6.3	1.5	0.18	3.45	11.11	Level2
16	(S2)	HT	0-10	58.09	19	22.91	36	6.3	1.9	0.12	3.28	15.83	Level2
17	(S2)	HT	20-30	50.38	26	23.62	26	6.1	0.9	0.09	1.55	10	Level2
18	(S3)	HB	0-10	54.31	21	24.69	63	6.2	3.6	0.27	6.21	13.33	Level2
19	(S3)	HS	20-30	46.75	25	28.25	49	5.5	2.3	0.18	3.97	12.78	Level3
20	(S3)	HT	20-30	42.44	32	25.56	46	6.1	2.1	0.17	3.62	12.35	Level2
21	(S3)	HB	0-10	60.24	15	24.76	55	6.2	3	0.17	5.17	17.65	Level2

Fig. 4 Test Set

Once the data is divided into two parts, we can extract reducts and induce Classification rules on training set by Johnson's algorithm and standard voting method respectively.


Reducts

	Reduct	Support	Length
1	{Site, Slope, Silt%}	1	3
2	{Slope, Moisture%}	1	2
3	{Site, Slope, C%}	1	3
4	{Site, Slope, N%}	1	3
5	{Site, Slope, OM(%)}	1	3
6	{Slope, C:N}	1	2
7	{Site, Depth(cm), Silt%}	1	3
8	{Depth(cm), Moisture%}	1	2
9	{Depth(cm), N%}	1	2
10	{Depth(cm), C:N}	1	2
11	{Sand%}	1	1
12	{Silt%, Moisture%}	1	2
13	{Silt%, C%}	1	2
14	{Silt%, N%}	1	2
15	{Silt%, OM(%)}	1	2
16	{Silt%, C:N}	1	2
17	{Clay%}	1	1
18	{pH_Value}	1	1
19	{Moisture%, C%}	1	2
20	{C%, N%}	1	2
21	{C%, C:N}	1	2
22	{Moisture%, N%}	1	2
23	{N%, OM(%)}	1	2

Fig. 5 Extracted Reducts

For feature selection we get 28 reducts followed by their support and length.

Rules

	Rule	LHS Support	RHS Support	RHS Accuracy	LHS Coverage	RHS Coverage	RHS Stability	LHS Length	RHS Length
1	pH_Value(6.4) => FI(Level2)	2	2	1.0	0.074074	0.086957	1.0	1	1
2	pH_Value(6.3) => FI(Level2)	2	2	1.0	0.074074	0.086957	1.0	1	1
3	pH_Value(5.9) => FI(Level2)	1	1	1.0	0.037037	0.043478	1.0	1	1
4	Moisture%(52) AND C%(2.2) => FI(Level2)	2	2	1.0	0.074074	0.086957	1.0	2	1
5	Moisture%(48) AND C%(1.7) => FI(Level2)	1	1	1.0	0.037037	0.043478	1.0	2	1
6	Moisture%(39) AND C%(1.9) => FI(Level2)	1	1	1.0	0.037037	0.043478	1.0	2	1
7	Moisture%(39) AND C%(2) => FI(Level2)	1	1	1.0	0.037037	0.043478	1.0	2	1
8	Moisture%(34) AND C%(1.6) => FI(Level2)	1	1	1.0	0.037037	0.043478	1.0	2	1
9	Moisture%(31) AND C%(1.1) => FI(Level2)	1	1	1.0	0.037037	0.043478	1.0	2	1
10	Moisture%(29) AND C%(1.4) => FI(Level2)	1	1	1.0	0.037037	0.043478	1.0	2	1
11	Moisture%(31) AND C%(1.7) => FI(Level2)	1	1	1.0	0.037037	0.043478	1.0	2	1
12	Moisture%(27) AND C%(1.9) => FI(Level2)	1	1	1.0	0.037037	0.043478	1.0	2	1
13	Moisture%(42) AND C%(1.7) => FI(Level2)	1	1	1.0	0.037037	0.043478	1.0	2	1
14	Moisture%(48) AND C%(1.9) => FI(Level2)	1	1	1.0	0.037037	0.043478	1.0	2	1
15	Moisture%(29) AND C%(1.3) => FI(Level2)	1	1	1.0	0.037037	0.043478	1.0	2	1
16	Moisture%(32) AND C%(2) => FI(Level2)	1	1	1.0	0.037037	0.043478	1.0	2	1
17	Moisture%(28) AND C%(1.1) => FI(Level2)	1	1	1.0	0.037037	0.043478	1.0	2	1
18	Moisture%(59) AND C%(2.9) => FI(Level2)	1	1	1.0	0.037037	0.043478	1.0	2	1
19	Moisture%(62) AND C%(2.9) => FI(Level2)	1	1	1.0	0.037037	0.043478	1.0	2	1
20	Moisture%(56) AND C%(2.6) => FI(Level2)	1	1	1.0	0.037037	0.043478	1.0	2	1
21	Clay%(24.2) => FI(Level2)	1	1	1.0	0.037037	0.043478	1.0	1	1

Fig. 6 Generated Sorted Rules

713 classification rules have been generated. These rules are very important for classifying the test set of our soil data. Based on these rules we can predict the exact fertility level class for the test part objects. In the above Fig 6 the generated rules are sorted in by IF and THEN part of the rule. There are 713 rules that have been generated. The following statistics can be defined for each rule.

- *LHS Support:* the total number of objects in the training set matching the IF part.
- *RHS Support:* Number of objects in the training set matching the IF-part and the THEN-part (LHS and RHS support is the same unless the THEN-part contains several decisions).
- *RHS Accuracy:* RHS support divided by LHS support (Accuracy is 1.0 unless the THEN-part contains several decisions).
- *LHS Coverage:* LHS support divided by the number of objects in the training set.
- *RHS Coverage:* RHS Support divided by the number of objects in the decision class listed in the THEN part of the rule.
- *RHS Stability:* Not applicable for the Johnson algorithm (always 1.0).
- *LHS Length:* Number of attributes in the IF-part of the rule.
- *RHS Length:* Number of decisions in the THEN-part of the rule.

The above rules can be used to classify the objects or tuples of test set. Standard voting is used as classification method. In the present study we predict the fertility index class of test set based on the rules.

The confusion matrix [7], [8], [12] in Fig.7 shows the overall accuracy (i.e. 0.961538), as well as the sensitivity [8], [13] and accuracy for each class. For example, the Level 2 decision class has a sensitivity of 1.0 (i.e. of 23+0+0 = 23 objects actually belonging to Level 2, 23 was correctly classified as Level 2: 23/23 = 1.0) and an accuracy of 0.95 (i.e. of 23+1 = 24 objects predicted to Level 2, 23 were actually belonging to this class: 23/24 = 0.95).

Classification Rule

		Predicted			
		Level2	Level3	Level1	
Actual	Level2	23	0	0	1.0
	Level3	1	1	0	0.5
	Level1	0	0	1	1.0
		0.958333	1.0	1.0	0.961538
ROC	Class	Undefined			
	Area	3.402820e+038			
	Std. error	3.402820e+038			
	Thr. (0, 1)	3.402820e+038			
	Thr. acc.	3.402820e+038			

Fig. 7 Classification Rule

Through the confusion matrix [7], [8] designed above we can define classifier's performance in terms of accuracy for the given soil data. In our experiments on soil samples the accuracy of rough set is approx 96% which defines 96 percent of tuples are correctly classify by the classifier in terms of fertility level class that is the target class for the data samples.

VI. CONCLUSION

. In this paper a feature selection is used towards soil classification in terms of soil fertility level classes using machine learning. The target class fertility level is considered in three different levels based on the soil properties. Rough set theory is applied for classification and prediction results and finally calculated the performance of classifier in terms of accuracy where predicted values and actual values also play an important role to calculate the performance of classifier

VII. FUTURE SCOPE

In current research machine learning is widely used in soil science. We can extend this assignment using other machine learning techniques like ANN, Fuzzy etc. Deep learning is another important tool that can be used for implementing a classification model to predict fertility.

ACKNOWLEDGMENT

The first author is thankful to Prof R P Pant, Dept. of Mathematics, DSB Campus, Kumaun University Nainital for his continuous guidance and valuable suggestations.

REFERENCES

1. Hao Li, Weijia Leng, Yibing Zhou and Zhilong Xiu.," Evaluation Models for Soil Nutrient Based on Support Vector Machine and Artificial Neural Networks",Hindawi Publishing Corporation,The Scientific World Journal,2014, Volume. Article ID 478569. 7 pages.
2. M.S. Suchithra, Maya L. Pai, "Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters", Information Processing in Agriculture, 2019,
3. https://doi.org/10.1016/j.inpa.2019.05.003.
4. Reashma SJ, Pillai AS,"Edaphic factors and crop growth using machine learning—A review", International conference on intelligent sustainable systems (ICISS), 2017 IEEE 2017, p. 270–74.
5. Nikhita Awasthi, Abhay Bansal,"Application of Data Mining Classification Techniques on Soil Data Using R",International Journal of Advances in Electronics and Computer Science,2017,ISSN: 2393-2835,Volume-4. Issue.
6. Janmejay Pant, Kamlesh Padaliya, Himanshu Pant, "Rough Set Approach for Feature Selection in IDS",International Journal of Innovations & Advancement in Computer Science,2015, Volume 4. Special Issue September. ISSN 2347 – 8616.

7. Janmejay Pant, Amit Juyal , Shivani Bahuguna,"Soft set, a soft Computing Approach for Dimensionality Reduction", International Journal of Innovative Science, Engineering & Technology,2015,Vol. 2, Issue 4.
8. Marzieh Mokarram, Mehran Shaygan, George ch. Miliaresi,"Balancing Soil Parameters and Farmers Budget by Feature Selection and Ordered Weighted Averaging" Geographia Technica,2018,Vol.13, Issue 1,pp 73 to 84.
9. Janmejay Pant, R.P Pant, Amit Juyal, Himanshu Pant,"Rule Generation Methods for Medical Data (Heart Disease) Using Rough Set Approach",International Conference on Advances in Engineering Science Management & Technology .Availableat SSRN: https://ssrn.com/abstract=3384652.
10. Samir Roy, Udit Chaudhary."Introduction to Soft Computing: Neuro-Fuzzy and Genetic Algorithms", pp 145-171.
11. Priyanka Suyal, Janmejay Pant, Akhilesh Dwivedi & Manoj Chandra Lohani,"Performance Evaluation of rough set based classification models to intrusion detection system".2nd International Conference on Advances in Computing, Communication, & Automation (ICACCA), 2016.
12. Janmejay Pant, Amit Juyal, Himanshu Pant, Akhilesh Dwivedi, "A Real Time Application of Soft Set in Parameterization Reduction for Decision Making Problem ".International Journal of Electrical and Computer Engineering (IJECE), 2017, Vol. 7, No. 1, pp. 324~329 ISSN: 2088-8708, DOI: 10.11591/ijece.v7i1.pp324-329.
13. Pawlak, Z., "Rough Set", In International Journal of Computer and Information Sciences,Vol. 11, pp.341-356, 1982.
14. Jensen, J., "Combining rough set and fuzzy sets for feature selection". Ph.D Thesis from Internet, 2005.
15. Maryam Zavareha and Viviana Maggioni,"Application of Rough Set Theory to Water Quality Analysis: A Case Study", 2018, MPDI. www.mdpi.com/journal/data

AUTHORS PROFILE

Janmejay Pant is pursuing his Doctorate in Information Technology from Kumaun University Nainital and done M.tech in Information Technology from Graphic Era deemed to be University, Dehradun. Currently he is working as Assistant Professor in Dept. of computer science at Graphic Era Hill University, Bhimtal Campus. He has total Academic teaching experience of more than 8 years with more than 25 publications in reputed National and International SCOPUS, UGC Approve Journals and conferences. His research area includes Machine Learning, Soft Computing, and Data Mining and Deep Learning. He is an active member of CSI. He secured First Rank in Uttarakhand State Eligibility Test (USET) in Computer Science and Applications in 2018. He also received Research Award at University Level in September 2019.

Pushpa Pant obtained her Doctorate in Geography from Kumaun University Nainital in 1992. She has Post Graduate teaching experience of more than 18 years with more than 10 publications in reputed journals, including SCI and SCOPUS indexed journals, and conferences. Currently she is working as Assistant Professor in M.B. Govt. P.G College Haldwani, Uttarakhand. She was awarded Senior Research Fellowship (SRF) by Council of Scientific and Industrial Research (CSIR) during her doctoral research work (July 1990 – June 1993) and Post Doctoral Fellowship by Indian Council of Social Sciences Research (ICSSR) during 1999 – 2001.

Ashutosh Bhatt obtained his Doctorate in Computer Science from Kumaun University Nainital in 2009. He has total Academic teaching experience of more than 16 years with more than 40 publications in SCI, SCOPUS journals and conferences. His research area includes Deep Learning, Machine Learning, Soft Computing and Data Science. Currently he is working as Associate Professor in BIAS Bhimtal, Uttarakhand. He is also associated with many renowned National/International Journals as Lead Guest Editor/reviewer/editorial board member. He is life member of CSI(Computer Society of India) and senior member of IEEE. Presently he is also serving as District Coordinator of Uttarakhand Science Education & Research Centre (USERC). He had served as State Student Coordinator of Region I Uttarakhand of CSI (for three year) and also served as Secretary of Uttarakhand ACM Professional Chapter.

Himanshu Pant is pursuing Doctorate from Graphic Era Hill University. He received his MCA from UPTU Lucknow. His areas of interest are Machine Learning, Soft computing and data mining. He has published various papers in the different journals. Currently he is working as Assistant professor in Graphic Era Hill University Bhimtal. He has an experience of around 07 years of teaching

Nirmal Pandey is an Assistant Professor at Faculty of Computer Science and Applications in Amrapali Group of Institute Haldwani, Nainital. Presently he is pursuing Ph.D. from Department of Information Technology, Kumaun University, and Nainital. He has done his Master degree of Computer Applications from Birla Institute of Applied Sciences, Bhimtal; Nainital in year 2000. His research area includes Machine Learning, Soft computing. He has 6 years of software industry experience specialization in core Java and 12 years of professional teaching experience.