International Journal of ChemTech Research
CODEN (USA): IJCRGG ISSN: 0974-4290
Vol.7, No.6, pp 2601-2605, 2014-2015

Simple and Quick Classification of Soil for Sunflower Cultivation using Data Mining Algorithm

Durga Karthik[1], K.Vijayarekha[2]

[1]CSE, SRC,SASTRA University, Kumbakonam. India.
[2]School of EEE, SASTRA University, Thanjavur, India.

Abstract: Thanjavur a district in the state of Tamilnadu, India is gifted with the river Cauvery that flows throughout the district. Due to change in the weather pattern and scanty rainfall the water table in the region has depleted. As rice a primary crop of this region requires more water the need of the hour is to find a crop that require less water and profitable too . Our aim is to identify the suitability of sunflower cultivation in this region as sunflower requires very less water, profitable and well suited for cultivation in tropical region. As the demand for edible oil is ever increasing, India is importing it from other countries. The properties of soils from various regions of Thanjavur are analyzed to find the suitability of sunflower cultivation in this district. Data Mining techniques are employed to classify if the soil is suitable or not.
Keywords: Classification, PCA, JRIP, Naïve Bayes, Multivariate.

Introduction

Thanjavur , the rice bowl of Tamilnadu in India has experienced deficit rainfall during the monsoon , that has led to irrigation problems. The region is located at latitude 10º 08', longitude 78º 48' with an altitude of 59m. Agriculture is the primary occupation but due to changes in weather conditions and due to low profit, the farmers tend to change their occupation. So, it is high time to identify other fast and beneficial crops such as sunflower. This edible plant can withstand very high temperature and to pest attacks. The cultivation of any crop depends on the soil, weather and water. Sunflower an important new plant can be grown in tropical regions that has maximum of 40 degrees, water requirement is much less and hence the soil suitability has to be identified. The soil can be analysed for various chemical properties such as EC, pH, along with micro nutrients N,P,K and macro nutrients Zn, B,Cu,Fe.

The classification [1] techniques can be applied to the soil data, the results of classification helps to determine the cultivation pattern. Sunflower [2] cultivation depends on the various factors of soil. GIS has proved to be important and soil resources[3] are interpreted for Thanjavur region to determine cropping changes. Multivariate analysis [4][5] along with PCA [6] can be applied to speed up the process. Chemometric [7] techniques can be applied to analyze the soil properties.

Materials and Methods

Standard methods were adopted for identification of Ec,pH,N,P,K,Fe,Zn,Cu,B. The soil analysis for the above elements were carried out at Tamil Nadu Rice Research Station, Aduthurai, Thanjavur District. The above parameters were taken for analysis due to the following reasons. The EC determines the texture and fields productivity, pH value indicates if the soil is acidic or basic. The micro nutrients N increases the crop yield , P and K increases the efficiency of fertilizers and yield. The macro nutrients such as Fe is used by plants during photosynthesis and respiration, Mn depletion leads to withering, Zn required for a healthy growth, Cu prevents root damage , protects from fungus and Boron increases seed yield for the sunflower plant. The results of analysis for 50 locations from Thanjavur district is given in Table 1.

Table 1: Soil properties from various locations of Thanjavur District ,Tamilnadu,India.

no	Location (Thanjavur)	EC (DSM^{-1})	Ph	N Kg/acre	p Kg/acre	K Kg/acre	Fe ppm	Mn ppm	Z ppm	Cu ppm	B ppm
1	Timakudi	0.11	7.3	110.8	136.5	135	9.6	0.9	5.4	1.1	0
2	Melatur	0.62	7.2	63	31.9	125	2.2	2.6	1.4	1.3	0
3	Aduthurai	0.66	7.6	56	18.9	125	3	2.6	1.2	1.2	0
4	Agathur	0.68	7.6	70	18.2	125	2.6	2.5	1.3	1.2	0
5	Thirukattupalli	0.78	7.8	77	19.3	125	2.2	2.6	1.3	1.4	0
6	Rajageeri	0.6	7.6	56	22.2	140	3	2.8	1.4	1.4	0
7	Buthalaur	0.7	7.5	63	15.6	115	2.6	2.4	1.2	1.4	0
8	Aavur	0.7	7.6	84	18.2	132	2.6	2.6	1.4	1.6	0
9	Mangudi	0.69	7.6	63	35.1	120	2.8	2.2	1.2	1.6	0
10	Kallur	0.7	7.2	56	20.8	132	3	2.8	1.4	1.4	0
11	Ayyampettai	0.69	7.2	91	18.2	140	2.8	2.7	1.3	1.2	0
12	Orathanadu	0.68	7.8	63	22.2	151	2.1	2.6	1.5	1.4	0
13	Neelathnallur	0.76	7.7	70	24.8	92.5	2.2	2.8	1.6	1.2	0
14	Kandivur	0.72	7.2	56	22.8	125	2.2	2.7	1.2	1.2	0
15	Veerasingampettai	0.6	7.6	63	20.8	132	3	2.8	1.5	1.4	0
16	Thirunallur	0.9	7.6	84	27.3	140	3	2.6	1.2	1.4	0
17	Kantharvakottai	0.77	7.7	77	24.8	132.5	2.2	2.7	1.2	1.2	0
18	Needamangalam	0.76	7.9	77	32.5	85	2.2	2.6	1.2	1.4	0
19	Papanasam	0.7	7.7	77	24.8	152.5	4	2.6	1.2	1.4	0
20	Thirubuvanam	0.77	7.7	70	22.2	140	2	2.8	1.2	1.4	0
21	Peravurani	0.68	7.4	42	18.2	140	2.1	2.6	1.5	1.2	0
22	Thambikottai	0.7	7.6	63	19.5	150	2.1	2.6	1.6	1.4	0
23	Pattukottai	0.68	7.4	49	195	132.5	2.1	2.4	1.2	1.2	0
24	Thirunallur	0.8	7.4	70	26	132.5	2	2	1.3	1.3	0
25	Tharasuram	0.6	7.8	63	18.2	132.5	3	2.5	1.2	1.2	0
26	Chozapuram	0.7	7.6	56	22.2	140	2.6	2.8	1.5	1.6	0
27	Kandithampattu	0.69	7.6	91	13	166	2.8	2.6	1.4	1.4	0
28	Thiruvidaimaruthur	0.71	7.2	49	26	125	2.1	2.4	1.2	1.4	0
29	Kapisthalam	0.7	7.7	84	19.5	157.5	2.7	2.8	1.2	1.4	0
30	Bhudalur	0.11	7.3	63	31.9	125	2.2	2.6	1.4	1.3	0
31	Manojipatti	0.11	7.4	101.5	15	115	2.6	1.2	5.4	1.3	0
32	Vallam	0.62	7.2	105	8	120	2.5	1.3	1.4	0.8	0
33	Anaikarai	0.66	7.6	112	19	125	2.3	1.4	1.2	0.9	0
34	Mannargudi	0.11	7.7	119	17	130	2.5	1.3	1.3	3.1	0
35	Rettipalayam	0.66	7.2	126	16	135	2.6	1.3	1.2	3.6	0
36	Nanjikottai	0.7	7.1	110	8	36	2	4.3	1.1	4.2	0
37	Gurungkulam	0.8	7.4	86	8.5	25	2.3	3.6	0.9	0.9	0
38	kallur	0.6	7.9	48	7	27	15.2	1.6	3.6	3.6	0
39	Karambayam	0.7	7.4	96	6.3	31	3	3.8	4.5	4.5	0
40	Muthupettai.	0.6	7.6	101.5	4	42	11.5	1.9	5.2	5.2	0
41	Marungulam	0.8	7.5	66.5	6	95	2.2	1.6	1.3	0.6	0.2
42	Mattur	0.9	7.4	70	8	100	2.6	1.2	0.4	0.5	1.2
43	Ramapuram	0.1	7.4	91	12	50	2.7	1.2	1.7	0.2	1.3
44	Sengippatti	0.11	7.2	101.5	13	13	4.2	2.5	1.1	0.1	0.1
45	Thittai	0.9	7.5	105	19	180	4.1	1.3	1	0.5	1.1
46	Thuraiyur	0.78	7.4	70	24	95	0.7	2.6	3.6	3.1	0.8
47	Vannarapettai	0.68	7.2	73.5	25	100	0.6	2.8	0.6	4.2	1.2
48	Viramarasampettai.	0.6	7.1	112	11	42	0.9	1.1	3.8	4.6	1.3
49	Pattukottai	0.7	7.6	115.5	15	41	3.9	0.9	3.9	1.1	0.3
50	Thambikottai	0.78	7.3	119	16	165	3.2	2.9	1.1	0.9	1.4

Sunflower is a robust plant that require a bare minimum conditions. The Indian standards for sunflower plantation as per Central Control Laboratory, Kudimianmalai,Tamilnadu is given in Table 2. All 50 soil samples were collected from agricultural fields from various locations of Thanjavur district , Tamilnadu. The farmers in this region cultivate paddy extensively for generations. Occasional and unexpected flood or drought leads to loss and makes their life very difficult. New techniques and latest trends in agriculture can be beneficial as the region is very fertile. A chart showing the Ec, Ph and the micro nutrients from the results is shown in Figure 1. On comparing the results with the above parameters in most of the regions , Ec and Ph values were within the requirement. N,P,K values even if not sufficient can be supplemented and similarly the micro nutrients. As the volume of data is high , dimensionality reduction is required for faster classification.

Table 2: Required Range - Sunflower

Component	Required Range For Sunflower
pH	6.0 - 7.2
EC (dsm^{-1})	Less than 4.8
N (kgs/acre)	16-36
P (kgs/acre)	40-60
K (kgs/acre)	10 - 16
Fe (ppm)	2.5
Zn (ppm)	2
Cu (ppm)	0.4
B (ppm)	0.4

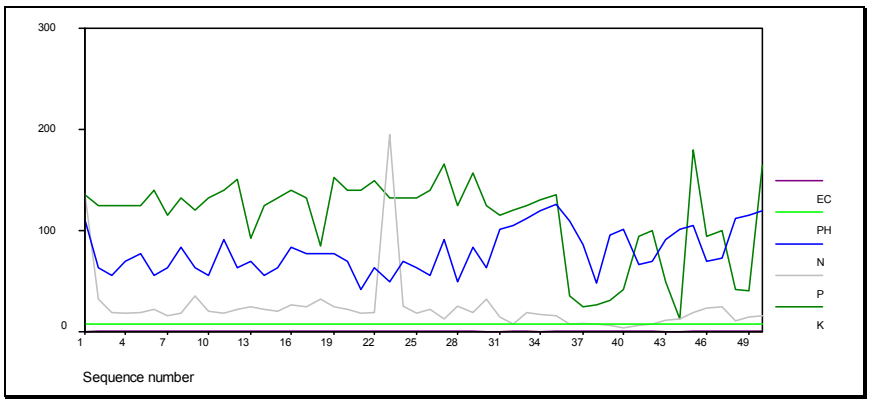


Figure 1: Chart showing Ec, Ph,N,P,K for Thanjavur Region

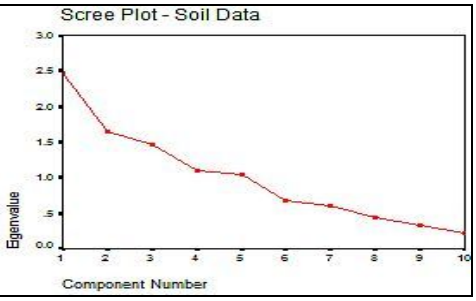
Principal Component Analysis

All 10 parameters are vital for plant growth and yield, to speed up the classification dimensionality reduction can be applied. PCA is one such tool for identifying the principal component and helps in dimensionality reduction. IBM SPSS tool was used to analyze PCA for the soil data. The results of PCA is shown in Table 3. The cumulative percentage of variance can be considered for component identification. Components that has up to 80% of variance should be considered and hence the first five components : Ec, pH, N,P,K can be taken as principal soil components . It is well known if N,P,K is not sufficient in soil ,it can be substituted using fertilizers. So,Ec and pH are taken for rule generation to classify, if the soil is suitable for sunflower cultivation or not.

Table 3: PCA Results and scree plot – Soil Data

Total Variance by PCA -Sunflower Data Explained

Componer	Initial Eigenvalues			action Sums of Squared Loadi		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.471	24.708	24.708	2.471	24.708	24.708
2	1.654	16.545	41.253	1.654	16.545	41.253
3	1.467	14.667	55.920	1.467	14.667	55.920
4	1.103	11.028	66.948	1.103	11.028	66.948
5	1.046	10.456	77.404	1.046	10.456	77.404
6	.672	6.723	84.127			
7	.602	6.023	90.150			
8	.439	4.393	94.543			
9	.334	3.344	97.887			
10	.211	2.113	100.000			



DATA MINING

JRIP – Rule Generation

JRIP algorithm can be used to classify the data, as it can generate a set of mutually exclusion rules .Weka adata mining software provides JRIP algorithm for rule generation. Initially 10 components were analyzed for soils and that was reduced to 5 components by PCA, to further simplify and speed up only EC and pH for the soil data were considered for rule generation. EC values indicates the salinity of the soil, pH values should be maintained for proper growth of plant. The rules generated by JRIP for soil is given below.

Rule 1 :(EC <= 4.8) and (PH >= 5.7) and (PH <= 8) => Suitable=Yes

Rule 2: => Suitable=no .

The JRIP is a learner based algorithm, which has evolved two rules for sunflower cultivation for the regions based on soils. The rule 1 is, EC value should be less than 4.8 and pH should range from 5.7 to 8 to classify *Suitable* as yes. Rule 2 is mutually exclusive, hence if the Rule 1 condition fails, then it is not suitable for sunflower cultivation.

Classification

The two rules were used to classify the soil data withNaïve Bayes algorithm available in Weka software. The results of classification is given below in Figure 2.The naïve based algorithm based on the JRIP rules classified 50 locations were classified either as yes or no and the classification result is shown in Table 4.

The Kappa statistics is near 1 hence the classification can be considered as a perfect agreement. The detailed class wise accuracy is shown in Table 5. that depicts the various accuracy measures for the proposed rules. True positive rate is near one, precision and recall measures for both the rules confirm the accuracy of the rules. Confusion matrix is shown in Table 6.

Figure 2: Naïve Bayes Classifier results for soil data using Weka.

Table 4: Classification Result

Location	EC(DSM-1	Ph	Suitablity
Timakudi,Kumbakonam	0.11	7.3	yes
Melatur, Thanjavur.	0.62	7.2	yes
Aduthurai, Thanjavur.	0.66	7.6	yes
Agathur, Thanjavur	0.68	7.6	yes
Thirukattupalli, Thanjavur.	0.78	7.8	yes
Rajageeri, Thanjavur	0.6	7.6	yes
Buthalaur, Thanjavur.	0.7	7.5	yes
Aavur, Thajavur	0.7	7.6	yes
Mangudi,Kumbakonam	0.69	7.6	yes
Kallur, Thanjavur.	0.7	7.2	yes
Ayyampettai, Thanjavur.	0.69	7.2	yes
Orathanadu, Thanjavur.	0.68	7.8	yes
Neelathnallur, Thanjavur.	0.76	7.7	yes
Kandivur, Thanjavur.	0.72	7.2	yes
Veerasingampettai, Thanjavur.	0.6	7.6	yes
Thirunallur, Thanjavur.	0.9	7.6	yes
Kantharvakottai, Thanjavur.	0.77	7.7	yes
Needamangalam, Thanjavur.	0.76	7.9	yes
Papanasam, Thanjavur.	0.7	7.7	yes
Thirubuvanam, Thanjavur.	0.77	7.7	yes
Peravurani, Thanjavur.	0.68	7.4	yes
Thambikottai, Thanjavur.	0.7	7.6	yes
Pattukottai, Thanjavur.	0.68	7.4	yes
Thirunallur, Thanjavur.	0.8	7.4	yes
Tharasuram, Thanjavur.	0.6	7.8	yes
Chozapuram,Kumbakonam	0.7	7.6	yes
Kandithampattu, Thanjavur	0.69	7.6	yes
Thiruvidaimaruthur, Thanjavur	0.71	7.2	yes
Kapisthalam, Kumbakonam, Th	0.7	7.7	yes
Bhudalur, Thanjavur	0.11	7.3	yes
Manojipatti, Thanjavur.	0.11	7.4	yes
Vallam, Thanjavur.	0.62	7.2	yes
Anaikarai, Thanjavur.	0.66	7.6	yes
Mannargudi, Thanjavur.	0.11	7.7	yes
Rettipalayam, Thanjavur.	0.66	7.2	yes
Nanjikottai, Thanjavur.	0.7	7.1	yes
Gurungkulam, Thanjavur	0.8	7.4	yes
kallur, Thanjavur.	0.6	7.9	yes
Karambayam, Thanjavur.	0.7	7.4	yes
Muthupettai, Thanjavur.	0.6	7.6	yes
Marungulam, Thanjavur.	0.8	7.5	yes
Mattur, Thanjavur.	0.9	7.4	yes
Ramapuram, Thanjavur.	0.1	7.4	yes
Sengippatti, Thanjavur.	0.11	7.2	yes
Thittai, Thanjavur.	0.9	7.5	yes
Thuraiyur, Thanjavur.	0.78	7.4	yes
Vannarapettai, Thanjavur.	0.68	7.2	yes
Viramarasampettai, Thanjavur.	0.6	7.1	yes
Pattukottai, Thanjavur.	0.7	7.6	yes
Thambikottai, Thanjavur.	0.78	7.3	yes

Table 5: Accuracy of classes for Soil Data.

Detailed Accuracy By Class						
TPRate	**FPRate**	**Precision**	**Recall**	**F-Measure**	**ROC Area**	**Class**
0.974	0	1	0.974	0.987	0.998	Yes
1	0.026	0.917	1	0.957	1	No

Table 6. Confusion Matrix for Naïve Bayes

a	b	Classified as
38	1	yes
0	11	no

Scheme:weka.classifiers.bayes.NaiveBayes		
Correctly Classified Instances	49	98 %
Incorrectly Classified Instances	1	2 %
Kappa statistic	0.9436	
Mean absolute error	0.02	
Root mean squared error	0.128	
Relative absolute error	5.7206 %	
Root relative squared error	30.8873 %	
Total Number of Instances	50	
Ignored Class Unknown Instances	1	

Figure 2: Naïve Bayes Classifier results for soil data using Weka.

Conclusion

Data Mining algorithms such as JRIP and Naïve Bayes with PCA has found to be an excellent method to classify the locations based on only two important soil properties. This method has found to be quick and simple for classification. Sunflower requires slightly acidic to an extent alkaline pH value (5.7 TO 8), acidic soil leads to faster aging, uneven crop growth, stunted plant growth and higher alkalinity decreases nutrient uptake by the plant. EC range from low to moderate, if it is higher than 4.8 water use efficiency for grain yield becomes lower. The analyzed regions have a suitable Electrical conductivity and has a good pH range for sunflower cultivation. It has been identified that almost all regions in Thanjavur district are suited for sunflower cultivation. As such the soil is fertile and has normal temperature of 40 ˚C ,hence sunflower cultivation will be beneficial to the farmers. This work can be extended to identify the other suitable crops that can be cultivated in any regions based on their soil properties using JRIP.

References

1. Bhargavi .P,Dr.Jyothi,.S 2009, Applying Naïve Bayes Data Mining Technique for Classification of Agricultural Land Soils, IJCSNS -International Journal of Computer Science and Network Security, Vol.9, No.8.
2. Francois E Leland (1996) ,Salinity Effects on Four Sunflower Hybrids,Agron .J. 88:215-219
3. Punithavathi.J Tamilenthi.R,Baskaran R ,2011, Interpretation of Soil Resources using remote sensing and GIS in Thanjavur district, Tamilnadu,India, Advances in Applied Science Research,2011,2(3):525-535.
4. Karthik D., K. Vijayarekha and K. Manjula ,2012, Multivariate analysis for detecting Adulteration in edible a review. IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012) March 30, 31.272-276.
5. Karthik D., K. Vijayarekha,2012, SATTVA-Statistical Affirmative Testing Tool for Various Adulterants, Research Journal of Applied Sciences, Engineering and Technology 4(24): 5357-5360.
6. Karthik D, K. Vijayarekha, S.Sekar, 2014, Profiling Water Quality Using Multivariate Chemometric Method,Polish Journal of Environmental Studies,Vol 23,No:2.573-576
7. Spanos TH, Simeonov V, S.Tsakovski,D.Thiokas, 2004, Chemometric study of Soil analysis data, 2(2),402-416.

*****