[Belabed*et al., 5(9): September, 2016]
IC™ Value: 3.00
ISSN: 2277-9655
Impact Factor: 4.116

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

INPUT ANALYSIS USING SOIL CLASSIFICATION TECHNIQUES AND PREDICTING ITS EVOLUTION IN MOROCCO ORIENTAL TO CONVERSION WHEAT/OLIVE AT HORIZON 2020

Imane Belabed*, TALIBI ALAOUI , Abdelmajid Belabed*, Mamie Matoir, Bouziani Naoual
* University Mohammed premier; faculty of sciences,Morroco

DOI: 10.5281/zenodo.154196

ABSTRACT

Agronomic and pastoral research in East Morocco depend an evolution in time and space of the soil which is an important factor otherwise, limiting plant cover.

Thus, a chemical and physicochemical soil study was assisted by advanced techniques such as data mining. Today, data exploitation on a global scale are used in a large number of vast agribusiness operating areas. Products of computer operating systems and specific domain data extrapolation are applicable across disciplines, however they are still relatively new on agricultural . Hence a lack of bibliographic data is recorded.

This research aims to analyze soil data using classification techniques. It focuses on soil classification using various algorithms.

KEYWORDS: Data mining, Classification, Regression, Soil tests, Coverage rate, Olive-wheat reconversion

INTRODUCTION

Currently the Data Mining is an area of recent and crucial approach in the world of bioinformatics research and agro-informatics. The techniques are useful to raise a significant and usable knowledge that can be perceived by many researchers. Data mining consists of various methodological programs that are mainly produced and used for real decisions of the sectors involved **[1].** These techniques are well prepared for their respective areas of knowledge. The use of statistical analysis is both tedious and precise.

A soil test consist a selection of samples and determine its nutrient, composition and other characteristics. Measure fertility and indicate deficiencies needed to address it **[2]**. Appropriate techniques of soil testing, including testing methods and formulations to recommend appropriate fertilizer**. [4]** This helps farmers to decide the extent of fertilizer and manure to be applied at various stages of the growth cycle of the culture and have agriculture for better conversion.

In this work we used the classification of variables to search for underlying structures in the data, it also helps to identify redundant groups of variables reflecting the same type of information; to separate the orthogonal groups of variables, reporting additional information. This gives us valuable information on data architecture **[5].**

MATERIALS AND METHODS

Field of study

We chose to study Oriental Morocco soil data to identify groups of variables, components such soils and avoid duplication where correlations are distinguished. The aim is to reduce the number of variables, focusing on synthetic variables to finally get to know which group of variables explain the better the yield of wheat and olive trees in eastern Morocco with a view to evaluate a methods adopted by the Green Morocco's strategy in the 2020 horizon is the conversion of wheat in olive crops.

We worked on 74 plots of eastern Morocco with 12 variables that are (SA, TA, C, N, MO, AR, CA, CT, E, K, P, and pH).

Table 1: Symbols soil study

C	**Carbon**
N	**Nitrogen**
MO	**Organic Matter**
SA	**salinity**
LI	**Limon**
AR	**CLay**
CT	**Total carbon**
CA	**Active carbon**
PH	**pH**
P	**Phosphorus**
K	**Potassium**
E	**saturation extract diluted to 1/5**

RESULTS AND DISCUSSION

Methodology and results

In this article we have chosen to work with the data analysis methodologies unsupervised namely the classification variables around latent components to identify groups of variables, components such soils and avoid duplication **[3].**

The goal is to reduce the number of variables, focusing on synthetic variables to finally get to know which group of variables explain the better the yield of wheat and olive, in this study. We explain later in this article adopted the algorithms associated with their results

Classification of variables around latent components

This approach is based on the following concept: to represent a group of variables, the "average" variable in some way, we use a latent variable that is the first factor of the principal component analysis **[3].**
On this basis, a variable group is represented by the first axis of the PCA (Principal Component Analysis) ; we can decline the variables of classification strategies. We have used in this work three algorithms of classification variables: hierarchical clustering, classification with the k-means algorithm and classification with VARCLUS algorithm

The algorithm of Hierarchical Classification

The algorithm takes a hierarchical approach. The process of building is the hierarchical ascending classification. . At each step, we merge the two groups generating the smallest loss of variability explained and quantified by the difference between the sum of their own first values and the neat first value of the group formed.

Cluster summary

Table 2:List of variable by cluster

Cluster	# Members	Variation Explained	Proportion Explained
1	2	1,8302	0,9151
2	8	5,6867	0,7108
3	2	1,3230	0,6615
Total		8,8399	0,7367

The above table (Table 2) illustrates a division into three classes with a variability of 73.6 %

ISSN: 2277-9655
Impact Factor: 4.116

Cluster correlations – Structure

Table 3:Correlations variables with all classes

Attribute	# membership	Cluster 1	Cluster 2	Cluster 3
std_Ar (%)_1	1	0,0886	0,7959	0,3574
std_Li (%)_1	1	0,4247	0,8576	0,4548
std_Sa (%)_1	1	-0,2941	-0,9526	-0,4573
std_CT (%)_1	1	0,9566	0,1749	0,2572
std_CA (%)_1	1	0,9566	0,2427	0,1841
std_E1/5 (mg/l)_1	1	0,3306	0,4703	0,8133
std_pH_1	0	0,0995	-0,5828	-0,3046
std_C (%)_1	1	0,1811	0,9371	0,6089
std_N (%) _1	1	0,1643	0,9504	0,6417
std_MO_1	1	0,1767	0,9377	0,6113
std_P_1	1	0,0446	0,4253	0,8133
std_K (ppm)_1	0	0,1452	0,6403	0,1403

The above table (Table 3) leads us to conclude that the PH factor and K (ppm) do not belong to any group since their low correlations with any cluster es and mumbership sucks

conclusions

• Salinity Sa of The soil is negatively correlated with Ar, Li, C, N, MO more the soil is rich in Ar, Li, C, N, MO more salinity is low and vice versa

• More CA is high more CT is also the same case for E1 / 5 (mg / l) and P

Correlation between these classes and the yield of wheat and olives

- Wheat Yield:

Table 4:Correlation between the classes and the variable " yield of wheat "

t	X	r	r	t	Pr(>|t|)
Wheat Yield	VCHca_1_1	-0,0908	0,0082	-0,7739	0,4415
Wheat Yield	VCHca_1_2	0,7056	0,4979	8,4490	0,0000
Wheat Yield	VCHca_1_3	0,4467	0,1996	4,2367	0,0001

We notes that the performance of wheat is correlated 70% with the second class which includes (C N MO AR LI SA) that is more C, N, MO, AR, LI are high more SA is low more the wheat production is high.

- the olive tree yield:

Table 5:Correlation between the classes and the variable " yield olive"

Y	*X*	*r*	*r*	*t*	*Pr(>|t|)*
Olive tree yield	*VCHca_1_1*	*0,2141*	*0,0458*	*1,8598*	*0,0670*
Olive tree yield	*VCHca_1_2*	*0,7813*	*0,6105*	*10,6222*	*0,0000*
Olive tree yield	*VCHca_1_3*	*0,6183*	*0,3822*	*6,6746*	*0,0000*

also note that the yield of the olive tree is correlated 78% with the second class that includes (C, N, MO, AR, TA, SA) that is to say more C, N MO, AR, LI are high more the SA is low more wheat production is high, but 61.8% correlated with the third class which groups E and P. To get good results we performed the same steps with the algorithm VARCLUS.

Varclus algorithm
VARCLUS is a top down approach, in fact, exploration stops when there are no more relevant subdivisions
In this work we obtained the same results as the algorithm of hierarchical classification because one performs a bottom-up approach (hierarchical classification) and the other descending (VARCLUS).

K-Means algorithm
K-means for the classification variables is a variant of re-allocation method, adapted to variables. It is still based on the principle of latent components. We set the number of groups that are formed randomly at first. The variables are iteratively assigned to the nearest group, under the square of the correlation with the first factor, until there is convergence. The criterion is to maximize the total variability explained.
Note: In this algorithm the two variables were eliminated K and pH from the analysis since in both algorithms we have seen they do not belong to any group, also because we have rolled the algorithm by keeping them but they have distorted our results.

Cluster summary

Table 6:List of variable by cluster

Cluster	# Members	Variation Explained	Proportion Explained
1	6	5,0079	0,8347
2	2	1,8302	0,9151
3	2	1,3230	0,6615
Total		8,1611	0,8161

With this algorithm the variability explains 81.6% against 73.6% with both algorithms can be previously illustrated maybe is due to the elimination of two variables K and pH.

Cluster correlations – Structure

Table 7:Correlations variables with all classes

Attribute	# membership	Cluster 1	Cluster 2	Cluster 3
std_Ar (%)_1	1	0,7807	0,0886	0,3574
std_Li (%)_1	1	0,8664	0,4247	0,4548
std_Sa (%)_1	1	-0,9470	-0,2941	-0,4573
std_CT (%)_1	1	0,1994	0,9566	0,2572
std_CA (%)_1	1	0,2650	0,9566	0,1841
std_E1/5 (mg/l)_1	1	0,4979	0,3306	0,8133
std_C (%)_1	1	0,9552	0,1811	0,6089
std_N (%) _1	1	0,9612	0,1643	0,6417
std_MO_1	1	0,9564	0,1767	0,6113
std_P_1	1	0,4405	0,0446	0,8133

With this algorithm, we get the same conclusions as the hierarchical classification and Varclus

- Salinity SA of The soil is negatively correlated with Ar, Li, C, N, MO more the soil is rich in Ar, Li, C, N, MO more salinity is low and vice versa
- More CA is high more CT is also, Also the case E1 / 5 (mg / l) and P

Correlation between these classes and the yield of wheat and olives tree

- Wheat yield :

Table 8:Correlation between the classes and the variable " yield of wheat "

Y	X	r	r	t	Pr(>|t|)
Wheat yield :	VCKMeans_1_1	0,6890	0,4747	8,0659	0,0000
Wheat yield :	VCKMeans_1_2	-0,0908	0,0082	-0,7739	0,4415
Wheat yield :	VCKMeans_1_3	0,4467	0,1996	4,2367	0,0001

We have found that wheat yield is 68.9% correlated with the first class that includes (C, N, MO, AR, TA, SA) that is C, N, MO, AR, LI more high SA is more low wheat production is high, but in this case there's no significant correlation with the second class which includes E1 / 5 (mg / l) and P is the order of 44.6%.

- Olive tree yield :

Table 9:Correlation between the classes and the variable " yield olive"

Y	X	r	r	t	Pr(>|t|)
Olive tree yield :	VCKMeans_1_1	0,7846	0,6155	10,7362	0,0000
Olive tree yield :	VCKMeans_1_2	0,2141	0,0458	1,8598	0,0670
Olive tree yield :	VCKMeans_1_3	0,6183	0,3822	6,6746	0,0000

We also note that the yield of the olive tree is correlated 78.4% with the first class that includes C, N, MO, AR, LI, SA is to say. More C, N MO, AR LI is high SA is more low wheat production is high, but 61.8% correlated with the third class which groups E and P

CONCLUSION AND PERSPECTIVE

To sum up, without any ambiguity that the olive and wheat yield depend on the same elements of the soil which are C, N, MO, AR, LI and SA with a slight difference in favor of the production of the olive tree which is dependent on the E and P with a correlation r = 0, 61

For the conversion of wheat crops in olive it can be assure at the level of the soil they can give the same output but must also take into account other parameters such as rainfall, investment KDH in each culture to really evaluate the effectiveness of this strategy. Another complementary study will come into perspective for supply to our vision of contributing to decisions of this conversion in 2020.

ACKNOWLEDGEMENTS

Kadda Salma , PhD Student, University Mohammed premier; faculty of sciences,Morroco

REFERENCES

[1] A. Kumar & N. Kannathasan, (2011), “A Survey on Data Mining and Pattern Recognition Techniques for Soil Data Mining “, IJCSI .International Journal of Computer Science Issues, Vol. 8, Issue 3,

[2] “Soil test”, Wikipedia, February 2012

[3] CLASSIFICATION DE VARIABLES – APPLICATION À LA BASE PERMANENTE DES ÉQUIPEMENTS Brigitte GELEIN (*), Olivier SAUTORY (**)

[4] "Methods Manual-Soil Testing in India",Department of Agriculture & Cooperation Ministry of Agriculture Government of India, 2011

[5] Vigneau E., Qannari E.M., Sahmer K., Ladiray D., "Classification de variables autour de composantes latentes", Revue de statistique appliquée, 2006, vol 54, n°1, p. 27-45.

[6] S. Cunningham & G. Holmes, (1999),"Developing innovative applications in agriculture using data mining", Department of Computer Science University of Waikato Hamilton, New Zealand, Technical Report

[7] R. Vamanan & K. Ramar, (2011), "Classification Of Agricultural Land Soils A Data Mining Approach", International Journal on Computer Science and Engineering, ISSN: 0975-3397, Vol. 3

[8] P. Bhargavi & S. Jyothi, (2011), "Soil Classification Using Data Mining Techniques: A Comparative Study", International Journal of Engineering Trends and Technology

ANNEXES

Annexe 1: Study Data

Sites	Ar (%)	Li (%)	Sa (%)	CT (%)	CA (%)	E1/5 (mg/l)	pH	C (%)	N (%)	C/N	MO	P (ppm)	K (ppm)
Bouarfa	7.2	12.1	80.4	6.5	0	0.09	8.49	0.32	0.05	6.4	0.56	11.4	350
	7.7	13	79.3	6.5	0	0.09	8.49	0.24	0.04	6	0.42	11.4	400
	6.7	10	83.9	5.5	0	0.1	8.19	0.21	0.03	7	0.37	11.4	400
	10.7	17.4	71.8	7.4	0	0.1	8.26	0.29	0.04	7.25	0.5	11.4	420
	6	10.9	82.2	5.5	0	0.08	8.51	0.2	0.03	6.66	0.34	11.4	300
	10.8	13.8	77.3	6.5	0	0.1	8.49	0.32	0.04	8	0.55	11.4	370
	16.1	11.8	68.2	9.8	8	0.14	7.74	0.17	0.03	5.66	0.3	11.4	380
	17.2	17.8	64.2	9.8	8	0.12	8.2	0.29	0.04	7.25	0.5	11.4	460
	12.5	9.6	76.7	6.5	0	0.1	8.41	0.24	0.03	8	0.42	11.4	400
	10.2	9.9	77.3	2.5	0	0.12	8.59	0.16	0.03	5.33	0.27	11.4	540
	9.6	9.9	77.1	5.5	0	0.09	7.98	0.19	0.03	6.33	0.34	11.4	360
	6	9.1	84.5	8.9	6.5	0.09	7.85	0.17	0.03	5.66	0.3	11.4	250
Hamorzag	19.1	11	68.6	7.6	0	0.14	8.33	.036	0.05	7.2	0.62	11.4	420
	10.3	16.3	72.4	5.5	0	0.51	7.85	0.5	0.05	10	0.86	11.4	440
	9.8	9.2	79.2	5.5	0	0.11	8.2	0.18	0.03	6	0.3	2.3	310
	12.3	16	66.1	9.2	5	0.11	7.93	0.56	0.06	9.33	0.97	11.4	540

	13.8	15.8	65.4	7.6	0	0.13	8.15	0.38	0.05	7.6	0.65	11.4	530
	14.8	16.4	66.9	11	7.5	0.13	8.22	0.43	0.06	7.16	0.73	11.4	520
Iche	17.2	9.7	73.1	5.5	0	0.47	8.03	0.2	0.04	5	0.34	11.4	510
	11.4	11	74.1	4	0	0.11	8.26	0.22	0.04	5.5	0.39	2.3	400
	17.6	15.3	64.6	10.1	7	0.12	8.33	0.24	0.04	6	0.42	11.4	530
	14	29.3	54.5	7.6	0	0.1	8.33	0.39	0.05	7.8	0.67	11.4	540
	9.7	15.6	74.6	5.5	0	0.09	8.27	0.21	0.04	5.25	0.36	2.3	380
Figuig	14	13.9	70.8	5	0	0.12	7.65	0.26	0.04	6.5	0.45	2.3	470
	14.7	21.9	59.8	8.5	6.5	0.11	8.45	0.37	0.06	6.16	0.65	11.4	490
	7.6	25.5	64.8	9.2	5	0.09	8.3	0.26	0.04	6.5	0.45	11.4	390
	11.9	15	71.4	9.2	5	0.1	8.39	0.34	0.05	6.8	0.59	11.4	430
	14.9	4.6	76.5	5	0	0.09	8.05	0.21	0.05	4.2	0.37	11.4	450
	9	13.2	76.1	6.7	0	0.11	8.35	0.26	0.04	6.5	0.45	11.4	440
	19.5	28.6	53	23	13.5	0.29	8.17	0.91	0.11	8.3	1.57	11.4	690
	9.3	11.6	74	14.9	7.5	0.21	8.37	0.47	0.06	7.8	0.81	11.4	570
	1	27	73	12.3	7	0.21	8.4	0.47	0.06	7.8	0.81	11.4	610
	1.2	40.9	60.6	18.7	11	0.26	8.4	0.64	0.08	8	1.1	11.4	690
	14.3	12.6	71.6	18.1	9.5	0.25	8.43	0.58	0.07	8.3	1	11.4	650
Tandrara	8.2	27.9	65.2	22.1	10.5	0.29	8.56	0.6	0.08	7.5	1.03	11.4	640
	16.1	16.9	66.5	11	8.5	0.14	8.24	0.49	0.06	8.16	0.85	11.4	480
	19.9	22.5	53.6	16.9	9	0.12	7.74	0.6	0.08	7.5	1.04	11.4	520
	6.8	14.1	77.2	12.7	4.5	0.11	7.55	0.21	0.03	7	0.36	11.4	290
	9	8.8	80.5	10.1	4	0.11	8.55	0.15	0.03	5	0.26	2.3	230
Oujda	13.7	22.8	57.7	8.2	6	0.27	7.96	1.07	0.11	9.7	1.84	2.3	370
	18.4	32.7	42.5	11.4	10	0.14	8.28	1.83	0.17	10.7	3.16	11.4	590

Est -Ain beni Mathar	9.6	19.2	67.1	10.4	10	0.13	8.39	0.6	0.07	8.5	1.04	11.4	520
	14.6	24.6	54.1	10.4	9	0.16	7.85	0.84	0.08	10.5	1.45	11.4	550
Matarka	19.1	25.1	56	11.2	9.5	0.14	8.13	0.64	0.08	8	1.11	11.4	550
	9.9	20.8	63.9	7.9	0	0.16	8.45	0.6	0.07	8.5	1.04	16	620
	10.7	10.9	76	7	0	0.13	8.05	0.59	0.06	9.8	1.02	11.4	450
Nord-Ain	6.3	23	64.9	13.1	11	0.11	8.32	0.86	0.09	9.5	1.48	11.4	500
Beni	19.4	25.9	50	8.2	6	0.12	7.86	0.53	0.06	8.8	0.91	11.4	460
Mathar	9.5	22.8	63.9	24.4	10.5	0.12	8.4	0.7	0.08	8.7	1.21	20.6	610
Debdou	30	31	33.4	0.8	0	0.13	7.78	1.75	0.17	10.2	3.02	20.6	760
	29.9	38.7	27.2	12.6	0	0.13	7.9	1.95	0.17	11.4	3.36	11.4	710
Trarid	21.1	20.3	52.7	11.4	8.5	0.20	8.35	0.61	0.07	8.7	1.06	11.4	630
	16.3	23.4	55.6	9.5	8	0.22	8.33	0.62	0.07	8.8	1.07	16	580
	16.2	19.7	62.1	14	11	0.36	8.32	0.76	0.08	9.5	1.31	2.3	540
	25.2	22.7	50.3	9.9	8	0.43	8.22	0.63	0.08	7.8	1.09	11.4	600
	15.8	31.4	52.7	9.1	8	0.28	7.84	0.56	0.07	8	0.97	11.4	660
	27.5	20.5	47	9.9	8	0.14	8.39	0.7	0.08	8.7	1.21	11.4	560
	18.4	21.7	58.6	9.9	8	0.18	8.04	0.58	0.06	9.6	0.99	11.4	550
Berkane	30.2	29.4	34.8	1.6	1.2	0.24	7.46	1.33	0.14	9.5	2.29	34.3	620
	32	30.6	32.1	0.8	0	0.2	7.52	1.54	0.16	9.6	2.66	11.4	750
Nador	22.1	29.2	42.1	3.2	0	0.23	7.66	1.34	0.17	7.9	2.32	20.6	1080
	24	24.2	45	5	0	0.2	7.9	0.94	0.12	7.8	1.61	11.4	570
	24.5	26.1	43.1	4.1	2.3	0.19	7.78	1.04	0.12	8.7	1.79	2.3	710
	18.7	34.5	43.6	2.4	0	0.21	7.56	1.72	0.18	9.6	2.96	108.5	780
	27	32.9	37	5.8	0	0.87	7.92	1.92	0.21	9.1	3.32	114	660
	24.6	35.5	35.2	18.6	15	0.43	7.82	1.92	0.19	10.1	3.32	45.8	470

Taourirt	30	30.1	37.4	7.8	0	1.96	7.67	2.05	0.2	10.2	3.54	2.3	430
	30.1	40	26.8	25.3	23.5	1.13	7.88	0.79	0.09	8.8	1.37	11.4	210
	21	25.6	47.3	31.2	12.5	1.57	7.87	1.09	0.1	10.9	1.88	2.3	530
Layoune	11.2	28.8	55.6	17.7	10	0.22	8.14	1.73	0.15	11.5	2.99	2.3	530
	19.5	29	45.6	13	8.5	0.25	8.02	0.77	0.09	8.6	1.32	25.2	840
Driouche	32.2	36.2	25	19	12.5	0.38	8.11	1.09	0.13	8.4	1.88	16	1440
	26.7	30.7	35	14.4	8	0.28	7.81	1.25	0.16	7.8	2.15	34.3	1070
	18	35	41.4	16.8	10	1.13	8	2.13	0.23	9.3	3.68	238.4	580

Annexe 2 : data about the yield of the wheat and olive

Sites	Rendement T/Ha olivier	rendement T/Ha Blé
Bouarfa	**0**	**0**
	0	**0**
	0	**0**
	0.69	**28**
	0	**0**
	0.69	**28**
	0	**0**
	0	**0**
	0.69	**28**
	0	**0**
	0	**0**
	0	**0**
	0	**0**
	0	**0**

Hamorzag	**0**	**0**
	0	**0**
	0	**0**
	0	**0**
Iche	**0.5**	**28**
Figuig	**0**	**0**
	0	**0**
	0.69	**28**
	0	**0**
	0.69	**28**
	0	**0**
	0	**0**
	0.69	**28**
	0	**0**
	0	**0**
Tandrara	**0**	**0**
	0	**0**
	0	**0**
	0	**0**
	0	**0**
	0	**0**
	0	**0**
	0	**0**
	0	**0**
	0	**0**

Oujda	1.20	10.9
	1.29	7.0
Est -Ain beni Mathar	0	7.0
	0	7.1
Matarka	0	0
	0	0
	0	0
Nord-Ain Beni Mathar	0	0
	0	0
	0	0
Debdou	3.81	24
	3.72	26
Trarid	0	0
	0	0
	0	0
	0	0
	0	0
	0	0
	0	0
Berkane	2.54	57
	2.59	59
Nador	3.71	52
	2.92	52
	3.00	52
	3.70	55

	3.70	**55**
Taourirt	**6.60**	**41**
	6.16	**43**
	6.00	**41**
	5.59	**38**
Layoune	**4.59**	**41**
	4.51	**38**
Driouche	**4.00**	**42**
	4.10	**32**
	4.35	**29**