The relevant data mining algorithm for predicting the quality of production of olive in Granada region influenced by the climate change

Maroi TSOULI FATHI
Abdelmalek Essaâdi University, Faculty of Sciences and Techniques of Tangier, LMA Laboratory
Morocco
Maroi.tsouli@gmail.com

Mostafa EZZIYYANI
Abdelmalek Essaâdi University, Faculty of Sciences and Techniques of Tangier, LMA Laboratory
Morocco
ezziyyani@gmail.com

Loubna CHERRAT
Chouaib Doukkali University, Facultyu of Sciences El Jaddida, LAROSE Laboratory
Morocco

Sandra SENDRA
University of Granada Department of Signal Theory, Telematics and Communications
Spain
ssendra@ugr.es

Jaime Lloret
Polytechnic University of Valencia, Integrated Management Coastal Research Institute
Spain

ABSTRACT

The field of agriculture is extremely sensitive to the change of the climate, the variations intra and inter-seasonal cause the increase in the temperatures and the variations on the modes of precipitation which decreases the seasonal crop yields and increases the probability of bad short-term harvests and a reduction of the long-term production. However, this relation between climate change and agriculture are not yet foreseeable for the future, it will be thus interesting to make a predictive study which will allow the climatic analysis of data followed by a Macroclimatic study of data in order to establish the connection between climate change and agricultural production, and suggested afterward plans of adaptation to this change. In this study, we will carry out a comparative study, between the various methodology and tools of analysis of data of data mining to choose the algorithms that will adapt the best for our predictive analysis which will allow us to determine the threat of the impact of the climate change on the production of certain agricultural crops in Spain.

in this article, we will establish the relationship between climate change and agriculture is the purpose and use of data mining algorithms to establish a predictive analysis on the impact of climatic change on the olive cultivation in Spain.

KEYWORDS

Spain, Agriculture, climate change, data mining Algorithm, data analysis, Kernel-SVM, Decision tree, Naive Bayesian Classifier, Random Forest, Logistic-regression.

1 INTRODUCTION

The uncontrolled growth of greenhouse gas emissions such as Carbon dioxide (CO2), Methane (CH4) and Nitrous oxide (N2O). And warming the planet, resulting in the melting of glaciers, this is what causes the increase in precipitation, the multiplication of extreme weather phenomena, as well as the lagged seasons. The acceleration of climate change, coupled with global Population and income growth is in the first place a threat to food security. Climate change firstly affects the yield of irrigated crops, leading to severe declines in the production of certain agricultural crops. Spain is one of the most endangered countries, considered to be an agricultural country and the second agricultural country

with a large agricultural area that can be exploited. Among the main Spanish crops, citrus is the world's largest producer of oranges and the world's largest producer of olive oil [7], but this natural environment is threatened by several climatic handicaps such as low rainfall. Weather can cause significant variations in crop production from one year to the next, leading the state to become involved in climate risk management through insurance schemes. Spanish agriculture is particularly sensitive to climate change. A quarter of the total area of Spain is already threatened by desertification

2 STUDY OF THE CLIMATE IMPACT ON OLIVE CULTIVATION

2.1 Context:

As a choice of study, in this article it was chosen to study the region of Andalusia in Spain. Because it is an agricultural region known by the cultivation of the olive trees, it is also the agricultural culture chosen to make the test of prediction, it should be mentioned that the region of Andalusia contains a Mediterranean type climate which has undergone several changes caused by climatic change.

• Locations: Andalusia Spain

• Culture selected for study: olive trees

The choice of culture and chosen taking into consideration a non-irrigated crop develop in the Granada region is a culture that takes an important part in the economy of Spain [¬7].

"Spain is the leading olive producing country (31% of world production) and Andalusia alone accounts for 19% of world production. It is 1,400,000 hectares of olive trees, or 33% of the cultivated area in the world, which make this landscape so special.''

Figure 1: **Region of study**

2.2 The culture of the olive:

The olive tree is a Mediterranean tree to consider among the fruit trees, robust and very resistant to heat, it can reach fifteen to twenty meters height for 8 to 12 m spreading. and live several centuries. Because growth is slow, characterized also by a rather rounded and ample silhouette and persistent foliage.

Its flowering period is from April to June according to the regions and the varieties its flowers are of a white color after the flow comes the fruit in the case of the olive tree the fruit makes its appearance from June to August. The fruit of oblong shape in drupe, at first tender green, becomes violet then black. Under favorable climate, it produces 15 to 50 kg of olives after 5 years of cultivation at best and after 12 years at the most.

for a good yield, the olive tree needs certain cultural requirements namely hardiness according to the type of variety the tree also needs a good exposure to the sun because the summer period is necessary for the good development of the fruit, the soil is also an element not to forget the olive tree needs a light soil and filters the water quickly but it does not prevent it so easily to develop in dry soil, poor, stony and some limestone suitable.

2.3 Climate study on the region of Andalusia (Mediterranean Climate)

The climate of Spain is for the greater part of the territory a Mediterranean climate. Average annual temperatures are high and rainfall is relatively low (and very insufficient in summer.

The Mediterranean climate is a type of temperate climate or "warm temperate" or "subtropical west facade" which is characterized by hot, dry summers and mild, wet winters.

The term "Mediterranean" is explained by its characteristic presence around the Mediterranean Sea, but other regions of the world possess the same climatic conditions. These are the western façades of the continents, between 30 and 40 degrees latitude (California, central Chile, Cape area in South Africa, South and West Australia) [7].

This variant of the Mediterranean climate of Spain concerns the coastal regions of the Mediterranean Sea and the south of the peninsula (Andalusia).

It is characterized by a very hot and dry summer, low rainfall over the year and almost non-existent in summer, drought is a feature of the Mediterranean climate.

Our climate study will then take place in the region of Andalusia more precisely in the city of Granada which enjoys a warm temperate climate. The winter in Granada is characterized by much more precipitation than in summer. According to Köppen and Geiger, the climate is classified as Csa. The average annual temperature in Granada is 15.5 ° C. It falls on average 450 mm of rain per year

	Janvier	Février	Mars	Avril	Mai	Juin	Juillet	Août	Septembre	Octobre	Novembre	Décembre
Température moyenne (°C)	7	8.7	11.1	13.3	16.4	21.9	25.5	25.2	21.5	16.2	11.2	7.8
Température minimale moyenne (°C)	2.3	3.4	5.8	7.6	10.4	14.6	17.4	17.5	14.6	10.5	6.3	3.5
Température maximale (°C)	11.8	14	16.4	19	22.5	29.2	33.7	33	28.4	22	16.2	12.1
Température moyenne (°F)	44.6	47.7	52.0	55.9	61.5	71.4	77.9	77.4	70.7	61.2	52.2	46.0
Température minimale moyenne (°F)	36.1	38.1	42.4	45.7	50.7	58.3	63.3	63.5	58.3	50.9	43.3	38.3
Température maximale (°F)	53.2	57.2	61.5	66.2	72.5	84.6	92.7	91.4	83.1	71.6	61.2	53.8
Précipitations (mm)	54	49	54	51	37	11	5	6	20	45	55	63

Figure 2: **Example of temperature and precipitation condition of Mediterranean's climate**

The difference in precipitation between the driest month and the wettest month and 58 mm (Fig.3). As well as the difference between the lowest and highest temperature of the year, the difference is 18.5 ° C (Fig.4).

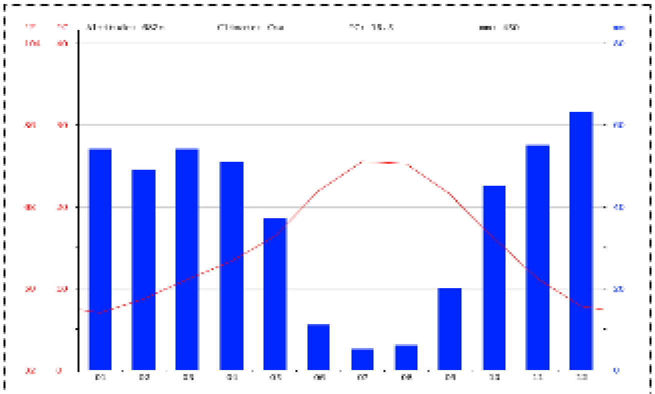

Figure 3: **Graphical representation of one year's data of Mediterranean's climate**

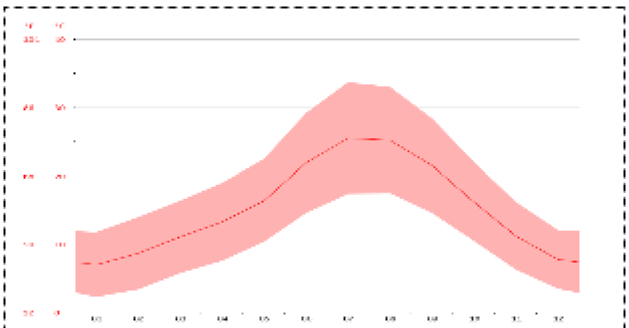

Figure 4: **Temperature curve of one year's data of Mediterranean's climate (Grenade Example)**

2.4 Problems caused by climate change linked to the cultivation of olive trees

Each irrigated agricultural crop needs some climatic conditions to develop to the best and yield the best yield.
The idea is therefore to study for each crop the most appropriate type of climate and the characteristics of this climate before climate change. Foreseeing is what this crop will be adapted to the new climatic conditions or not concerning our study the culture of the olive tree is adapted to the Mediterranean climate, it does not like the humidity and tolerates the frost badly. Olive trees in wetlands are much less developed than others. So, we will look for the conditions necessary for the cultivation of the olive tree such as:

-Temperature: The olive tree fears the cold. Negative temperatures can be dangerous, if it occurs at the time of flowering. The olive tree, on the other hand, is able to withstand high temperatures. This adaptation to drawing water from rooting allows it to withstand temperatures on the order of + 40 ° C.

-Rainfall: Less than 350 mm rainfall irrigation can't be economically viable.

-Atmospheric humidity: It can be useful insofar as it is not excessive (+ 60%) nor constant as it promotes the development of diseases and parasites.

-Altitude: The cultivation altitude of the olive tree depends on the altitude. The limits to be exceeded are 700 to 800 m for the slopes exposed to the north and 900 to 1000 m for the slopes exposed to the south.

-Other climatic factors:

• Fog: It is harmful because it causes the flowers to fall (sagging)

• Snow: It causes branches to break

• Hail: It destroys the young twigs

As a conclusion, the olive tree is a tree adapted to the Mediterranean climate, it does not like the humidity and tolerates the frost badly. Olive trees in wetlands are much less developed than others, this year in the region of Andalusia many trees suffered from cold (yellowed / blackened) leaves. due to lower temperatures during the winters

The limit of olive cultivation often corresponds to the zone of extension of this climate. This tree is damaged by frost if the temperature drops below -10 ° C during a winter rest period and can withstand several months of drought.

3 RELATED WORK

Datamining has recently influenced the field of agriculture several research and works have been inspired by datamining algorithms to solve the problems of the situation so far complex in the field of agriculture namely the influence of climate on seasonal crops, estimating the impact of climate on agricultural production...

In this section, we present related work done in the field of agriculture and weather forecasting systems using datamining algorithms such as neural networks.

Kulshrestha [14] studied the effect of artificial neural network meteorological parameters on the yield of aonla (Indian fruit) under different fertilizer treatments the results show that the April temperature range has a high correlation coefficient with higher yield as well as extreme temperature variations adversely affect the fruit yield of AONLA.

Y.Radhika and Shashi [15] favored weather forecasting using daily temperature at a given location to predict the temperature of the day following the same location in this process and achieve through application of the vector machines (SVM) which is based on the prediction at maximum daily temperatures, to classify over an interval of n previous days called the order of the input. The system and prove to be efficient is observed for different bays from 2 to 10 days using the optimal values of the kernel.

Santhosh Baboo and I. Kadar Shereef [16] present an application of the Neural Rear Propagation Network (BPN) for time prediction. Their proposed idea is tested using the real-time data set. The results are compared with the practical functioning of the meteorological service and these results confirm that the real-time processing of meteorological data indicates that BPN weather forecasts have shown an improvement not only on guidance predictions from numerical models but also on Local weather forecasts.

S. Kotsiantis et al. [17] focused on studying the effectiveness of data mining techniques in estimating minimum, maximum and mean temperature values. A number of experiments were conducted by them with well-known regression algorithms using temperature data from the city of Patras in Greece. The performance of their algorithms was evaluated using standard statistical indicators, such as the correlation coefficient, root mean square error, etc.

4 METHODOLOGIES

The methodology to adapt in our predictive analysis is to consist of three steps:

-Weather Monitoring: Analyse climate data bases to predict droughts, floods to make decisions for resource optimization and see if the climate is always adaptable for the cultivation period of certain crops.

-Production Supervision: Analyse the production of certain crops and see the history of past years, comparing the production of each year with the climatic data.

-Prediction of production according to climate forecasts: Consider the production of agricultural crops as a manageable input according to climatic conditions.

4.1 Collecting data

As we see in this data Figure 5 , we extract information that we call climate data the history we are going to use is between 1986 and 2016 we will work with four variables: precipitation in mm, maximum temperature (temp.max. C °) mean temperature (mean temp.C °) and minimum temperature (temp.min.C °).

```
STATION,"NAME","DATE","EMXP","PRCP","TMAX","TMIN"
SPE00120098,"GRANADA, SP","1996-01","0.73","1.22","73.2","60.4"
SPE00120098,"GRANADA, SP","1996-02","2.17","2.70","69.9","60.0"
SPE00120098,"GRANADA, SP","1996-03","0.47","1.13","72.0","59.2"
SPE00120098,"GRANADA, SP","1996-04","0.03","0.05","73.1","62.3"
SPE00120098,"GRANADA, SP","1996-05","0.03","0.04","77.0","65.5"
SPE00120098,"GRANADA, SP","1996-06","0.02","0.02","79.1","67.3"
SPE00120098,"GRANADA, SP","1996-07","0.02","0.02","81.8","69.6"
SPE00120098,"GRANADA, SP","1996-08","0.01","0.01","81.6","71.1"
SPE00120098,"GRANADA, SP","1996-09","0.70","1.33","81.4","70.3"
SPE00120098,"GRANADA, SP","1996-10","0.00","0.00","80.6","69.0"
SPE00120098,"GRANADA, SP","1996-11","0.08","0.13","76.9","65.6"
SPE00120098,"GRANADA, SP","1996-12","0.38","1.40","74.0","61.1"
SPE00120098,"GRANADA, SP","1997-01","0.22","0.33","71.1","59.8"
SPE00120098,"GRANADA, SP","1997-02","0.02","0.02","74.0","62.9"
SPE00120098,"GRANADA, SP","1997-03","1.14","1.42","75.9","62.5"
SPE00120098,"GRANADA, SP","1997-04","0.17","0.24","75.2","63.3"
SPE00120098,"GRANADA, SP","1997-05","0.02","0.02","76.2","65.5"
SPE00120098,"GRANADA, SP","1997-06","0.01","0.01","79.8","68.6"
SPE00120098,"GRANADA, SP","1997-07","0.00","0.00","80.9","70.1"
SPE00120098,"GRANADA, SP","1997-08","0.00","0.00","81.2","71.2"
SPE00120098,"GRANADA, SP","1997-09","0.19","0.35","81.8","71.5"
SPE00120098,"GRANADA, SP","1997-10","0.49","1.11","80.4","69.6"
```

Figure 5: **Climate data extracted from NOAA**

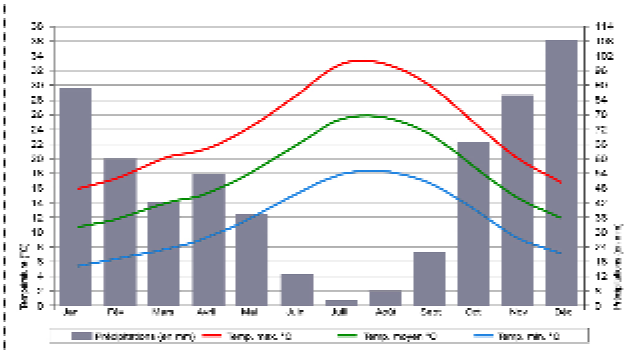

Figure 6: **Graphical representation of the data extracted from NOAA**

5 RESULT

5.1 Data mining ALGORITHMS and techniques for the predictive analysis tested

Datamining and a data analysis tool allows the exploration of data according to a process of extraction of important and useful information from large sets, to arrive thereafter presented in a summarized, organized and analysis. The aim of datamining and thus repartitioning on four simple techniques: prediction, identification, classification and optimization [¬6] Beginning with the identification this data mining function serves to identify and make the relationship between the data. While the predictive model has the role of finding a model that identifies data classes or the concept whose data is organized, the objective being to be able to use this model to predict the label class Whose class is unknown [2]. While the classification and clustering technique is designed for classification Unknown samples using information provided by a set of Samples classified. This set is generally regarded as a Learning set. If a learning set is not available, there is no Prerequisite Knowledge on the data to be classified. In that case, grouping techniques were used to divide a set of unknowns in clusters [1]

	Precision-prediction	**Accuracy**
Kernel-SVM	85%	0.727273
Decision tree	99%	0.363636
Random Forest	97%	0.181818
Naive-bayes	75%	0.545455
Logistic-regression	80%	0.714286

Table 1: Results of different classification algorithms built on our model. Note that, our dataset contains The history of 20 years of climate data input of real active.

The idea of this analysis is to find the best algorithm that gives the best prediction. For example, here the comparison between four classification and prediction algorithms is: Kernel-SVM, Decision tree, Random Forest, Naive-Bayes and Logistic-regression.

the test is carried out on real data of the region of Andalusia one uses two low data a climate one while the other contains statistics on the means of production of the olive tree in the region of Andalusia to compare the production of olive trees taking into account climatic parameters, each line is the line corresponding to one year, our test history starts from 1996 and ends in 2016 that we already have functionality as in Features in figure without target function, and prediction is the predicted value for each line).

Thus, if the prediction is less than 0, we will say that the production of this year will not be good because the climate does not correspond to the desirable exigency of where the production will be interfering, otherwise it is a good production.

In general, machine learning classifiers derive these performance results and each algorithm posed accuracy result obtained is represented as an array and graph.

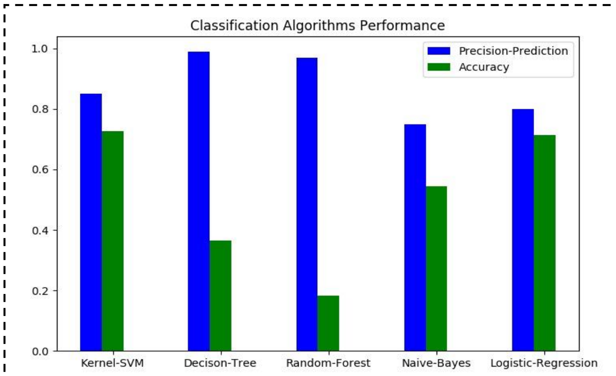

Figure 7: **Comparison of the performances of every algorithm used to do tests made for the prediction**

As expected, the decision tree gave us the best results, the decision tree is built on the entire dataset using all feature sets, Random Forest also gave good results Because it is an example of random forests, only a subset of rows is used randomly and many features are randomly selected to agree and a decision tree is built on that subset. SVM shows that it is also a good classifier.

6 CONCLUSIONS

In this article, the objective is to create a new technique to predict the adaptation of the olive tree to climatic changes in the Granada region by considering the use of different data mining algorithms [3]. The idea is to perform A structural analysis which consists of extracting rules of normal climate and applying them to the climate which has undergone a change in climate to compare this change with the climatic rules which are adapted to the cultivation of the olive tree and to leave by Future, behavioral and identity-based forecasting. Our future work will consist in doing the same analysis on other cultures in other regions based on the same context which consists of the collection of data to and to foresee to convert the entries into discrete binary values as early and using naïve-Bayes and decision trees, based on its datamining techniques [13], the Farmer field could plant different crops in different environments according to simple predictions.

REFERENCES

[1] Patel, H., & Patel, D. (2014). A Brief survey of Data Mining Techniques Applied to Agricultural Data. International Journal of Computer Applications, 95(9), 6–8. https://doi.org/http://dx.doi.org/10.5120/16620-6472
[2] Mucherino, A., Papajorgji, P., & Pardalos, P. (2009). Survey of Data Mining Techniques Applied To Agriculture. Operational Research, 9(2), 121–140. http://doi.org/10.17148/IJARCCE.2016.5263
[3] Esmita Gupta ME Student, A. P. (2014). Process Mining A Comparative Study. International Journal of Advanced Research in Computer and Communication Engineering, 3(11), 17–23. https://doi.org/10.17148/ijarcce
[4] Koppen, C. S. (1961). Climat Climat sec, 1–3.
[5] Sur, R., & Aride, Z. (n.d.). Carte de la.Platt, J. C. (1998). Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines. Advances in Kernel Methods, 185–208. https://doi.org/10.1.1.43.4376
[6] Big Data et machine learning.pdf. (n.d.).
[7] HUILES D OLIVE - OLIVE OILS. (2016), (November), 2016.
[8] Olaiya, F. (2012). Application of Data Mining Techniques in Weather Prediction and Climate Change Studies, (February), 51–59. https://doi.org/10.5815/ijieeb.2012.01.07
[9] Kotsiantis, Chapman, L., & Thornes, J. E. (2003). The use of geographical information systems in climatology and meteorology, 3, 313–330.
[10] Iglesias, C., Torres, J. M., & Nieto, P. J. G. (2014). Turbidity Prediction in a River Basin by Using Artificial Neural Networks : A Case Study in Northern Spain, 319–331. https://doi.org/10.1007/s11269-013-0487-9
[11] Goyal, M. K., Burn, D. H., & Ojha, C. S. P. (2012). Evaluation of machine learning tools as a statistical downscaling tool : temperatures projections for multi-stations for Thames River Basin , Canada, 519–534. https://doi.org/10.1007/s00704-011-0546-1
[12] Luo, Q., & Yu, Q. (2012). Developing higher resolution climate change scenarios for agricultural risk assessment : progress , challenges and prospects, 557–568. https://doi.org/10.1007/s00484-011-0488-4

[13] Ahmed, K., Shahid, S., Haroon, S. Bin, & Xiao-jun, W. (2015). Multilayer perceptron neural network for downscaling rainfall in arid region : A case study of Baluchistan , Pakistan, (6), 1325–1341.
[14] Kulshrestha, M., Shekh, A. M., & Pandey, V. (2010). a Study on Effect of Weather Parameters By Artificial Neural Networks on Yield of Aonla (Indian Gooseberry) Under Different Fertilizers Treatments, 4, 232–235.
[15] Y.Radhika and M.Shashi, Atmospheric Temperature Prediction using Support Vector Machines, International Journal of Computer Theory and Engineering, Vol. 1, No. 1, April 2009 1793-8201.
[16] Dr. S. Santhosh Baboo and I.Kadar Shereef. An Efficient Weather Forecasting System using Artificial Neural Network. International Journal of Environmental Science and Development, Vol. 1,No. 4, October 2010 ISSN: 2010-0264.
[17] Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: A review of classification and combining techniques. Artificial Intelligence Review, 26(3), 159–190. https://doi.org/10.1007/s10462-007-9052-3