International Journal on Emerging Technologies 12(1): 13-18(2021)

ISSN No. (Print): 0975-8364
ISSN No. (Online): 2249-3255

Prediction of Olive Oil Productivity using Machine Learning Decision Tree Algorithm

Mohammad Zaid Alkelani[1] and Mohammed Awad[2]
[1]Department of Computer Science, Arab American University, Palestine
[2]Department of Computer Systems Engineering, Arab American University, Palestine,

(Corresponding author: Mohammad Zaid Alkelani)
(Received 05 October 2020, Revised 10 November 2020, Accepted 07 December 2020)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: The production of olive oil is of great importance in the Palestinian economy, it considers an important consumable resource and one of the most important elements of food security. On another hand, it's one of the main sources of income as it contributes to about 13% of the value of the annual agricultural production. It is noted that the production of olive oil in Palestine varies from year to year and varies from one governorate to another due to the climatic factors that contribute to fluctuating production based on rain and other climate elements. Applying machine learning and data mining techniques to analyze data in agriculture is a promising approach for the prediction of agricultural yield management. Most of the challenges that we faced in this study focused on the availability of recorded data for olive oil productivity for 25 years, which we were able to collect in cooperation with many sources. In this paper, we endeavored to study the impact of climatic factors on olive oil productivity using the decision tree algorithm. We aim to determine the most important climatic parameters that play a major role in determining the level of the past and the future production of olive oil. The decision tree algorithm achieved excellent results in specifying the factors that the most important and which influence the productions. Our results showed that this method will be awesome for the classification of climatic parameters and prediction of the future olive oil quantity. These study findings will serve the interest of researchers, decision-makers in predicting future yields of the olive oil.

Keywords: Decision tree, C5.0 algorithm, Climatic factors, Data mining, Machine learning, productivity.

I. INTRODUCTION

The olive trees are the backbone of the Palestinian agricultural sector, as it is considered a priority for rain-fed traditional crops in Palestine. It is also considered one of the pillars of agriculture in Palestine since ancient eras, as it is an important element in the economy and has been used in many matters such as industry, lighting, medicine, and food. Olive trees occupy high proportions of the total areas planted with fruitful trees, which are estimated about (1.263) thousand dunums in Palestine, equivalent to about (85.3%) of these areas [1]. The cultivation of olives in the West Bank covers large areas of its land, due to the ability of olive trees to withstand the different natural conditions in terms of temperature, amount of rain, or rugged land and soil fertility, as the area of cultivated land in Palestine is about (987) thousand dunums [1]. The demand for olive oil increases in the global markets year after year, and the area that is grown also increases despite the presence of a surplus in olive oil at the local level in some years, and despite this, it is expected that there will be a shortage in the coming years as a result of the increasing demand in addition to the fluctuation in production from one year to the next. Palestinian olive oil is considered one of the finest types of oils around the world because it is extracted through the use of mechanical methods, and therefore the oil has a flavor and taste that distinguishes it from other oils that are extracted by solvents. This distinction is due to several characteristics starting from the type of cultivated olives, where the oil is extracted from the finest types of olive fruits around the world, the method of harvesting where the manual method is used, using the hand to pick the fruits, the date of the harvest where the fruits are harvested on specific dates from the Ministry of Agriculture according to the date of fruiting and the quantities of rains during the year, the period between harvest and presses is very small, due to the availability of many olive presses in Palestine, and therefore the fruit is not subject to a difference in color or corruption, Palestinian olive oil is organic where pesticides or chemical fertilizers are not used in most areas where there are olive trees. Some recent studies have shown that there is a positive or negative relationship between olive oil production quantities and various climate elements such as rainfall, temperatures, and other weather factors [2].

Thus, due to the existence of important relationships between the various elements of climate and the multiple processes in the agricultural sector, so-called agricultural metrology and agricultural climatology were born, which is a science that specializes in studying the impact of climatic factors that play important roles in everything related to crops [2]. Any production of different crops requires certain climatic conditions to grow perfectly and achieve the best returns and the best production [3]. Thus, there must be studies related to crops and lead to the detection of the best climatic characteristics that are suitable for this crop and achieve the desired goal. The main climatic elements that play an important role in agriculture are temperature, humidity, wind, and rain. Olive trees need low temperatures and high temperatures within specific periods of the year, they bear the cold and need it to produce buds in the winter period, while the olive trees need heat and a relatively long summer to increase the rates of oil and accelerate fruit ripening.

Average relative humidity in the range of (60%) and mild climatic conditions during the flowering stage help to produce well, as long as that this percentage does not exceed because it leads to the emergence of diseases and parasites [3]. Olive trees also need wind in the process of transporting pollen between the different varieties.

The rain factor is considered one of the most important factors in olive oil productivity in Palestine as the production quantities vary and differ over the years due to the dependence on this factor mainly, Therefore, the relationship between rain and olive yield is a direct relationship, the higher the rainfall rates, the higher the productivity. Consequently, the olives need to grow and be viable to a percentage of rain is between 200 to less than 400 mm per year in the case of supplemental irrigation and they need a rate of 400 mm and more rain per year in the rain-fed mode [4]. In recent years, the subjects of agricultural forecasting have received great attention from researchers and those interested, due to the varied needs such as reports related to fluctuation and rotation of production, inventory identification, future financial and strategic planning [5], discovering unknown patterns, etc.

Therefore, most of the criteria related to future production, especially the biological cycle of plants and crops, will be directly related to climatic conditions, different weather conditions, and climate elements.

It is a big challenge, and it is a vital need to find creative techniques to understand and detect the relations or discover the hidden patterns that play more roles between the climatic elements and future crop outputs in Palestine. Straightforward statistical calculations don't adequate or not efficient and may be unable to provide forecasting and expectations that are convenient in respect of characteristics of the diverse regions [6]. One of the main ideas behind studying the effect of climate on the production of olive oil in Palestine is to exploit data mining techniques and machine learning applications that have not previously been used on olive crops to discover relationships and analyze patterns between the mix of elements that are in the form of dependent and independent variables, to provide a new prediction method of agriculture in the Palestinian context.

By exploiting the meteorological data then we become able to forecast the impact of several meteorological parameters on the crops using the decision tree approach [7]. In this study, we will introduce a familiar artificial intelligent model based on the supervised learning manner using a machine learning algorithm so-called decision tree algorithms, to predict the most important climate elements that influence the future production of olive oil in Jenin city - Palestine which depends on the study of historical climatic data and the production of olive oil over the past years.

The actual and real data of climate, production of olive oil that was aggregated from the Palestinian meteorological department and Palestinian central bureau of statistics over past years, will be applied for training the model (Jenin city as a case study). Every weather item for every climate value was Assigned to produce one case and also the production of olive oil, this case can be High, Low and represent the average of each parameter over 24 years to classify the data values to fit the decision tree algorithm using C5.0 Package that will be used in our work. Thus, the model will predict the classification and find the correlation between climatic factors and olive oil crop productivity. The arranging of our paper is as follows. In section II, we present the related work. Section III, presents the preliminary and background concepts. Section IV, Information theory and the basics of the idea of a decision tree, Section V, methodology, Section VI, presents the results obtained based on the applied model and discussion, the conclusions and future work present in section VII.

II. LITERATURE REVIEW

Machine learning applications and data mining techniques have contributed to solving many agricultural problems in recent years, and many researchers and interested in climate issues and their impact on different crops have contributed to finding solutions to complex problems with the help of various artificial intelligence techniques. In this section, we review the near related work that was investigated in the agriculture domains and the impact of climate elements using artificial intelligence models such as data mining and other techniques.

Veenadhari, *et al.,* [7] the authors attempted to explore the impact of climatic parameters on soybean productivity by employing the decision tree induction method. The study aimed to evaluate the influence of climatic elements such as rainfall, evaporation, temperature, and humidity on the soybean yield, the authors also tried to obtain beneficial information and facts from secondary data on climatic factors. The decision tree method showed that there exists a strong relationship between climatic factors and soybean productivity, in the other hand and based on decision tree results, the relative humidity is the first factor that influenced the productivity of soybean followed by temperature and rainfall respectively, also they concluded that the decision tree induction method is very useful in forecasting the factors and elements which control the degree of rising or weak soybean productivity based on specific weather conditions and parameters.

Shakoor *et al.,* [8] two types of supervised machine learning algorithms have been carried out called the decision tree and k-nearest neighbors algorithm to Predict the agricultural production output, the authors aimed according to these algorithms to detect the patterns that exist in a given dataset that include the temperature and rainfall elements readings for six crops for twelve years, the authors aimed at this study to offer a learning agent that can assist in making decisions to develop agriculture and make it more efficient and increase profits by using technology, also the study focused on 6 different crops, the given dataset includes the details of crop yields, min and max temperature, rainfall, and other parameters. According to the result analysis, the decision tree achieved the lowest value of error than the k-nearest neighbor algorithm in most cases.

A software tool was implemented for forecasting the effect of climatic parameters on crop productivity [9]. The authors used a decision tree algorithm based on the C4.5 Package. The study aims to detect the more influencing climatic factors on the yields for certain crops in selected districts. The model has applied a while twenty years dataset on four types of plants, Soybean, Paddy, Maize, Wheat. The accuracy of the applied model was graded between 76 to 90 percent and the overall forecast accuracy of about 82.00 percent for the adopted model.

A framework to help Sugarcane farmers was developed to estimate and detect any failure in yield and decrease the losses at the early periods [10]. In this study, the authors designed a framework based on the use of a decision tree manner to classify and predict sugarcane productivity. The authors relied on three variables called climate-related, agronomic-related, and weather disturbance as input factors. The framework is capable to pre-process and combine the inputs of data from various sources add to that it offers an excellent and

precise prediction of crop productivity, also determines the most important variables that affect productivity.
Three machine learning algorithms have been implemented and used to classify and forecast the wheat crop yield [11]. This study discussed the influence of humidity and rainfall on wheat crops. This study aims to identify and detect the most important rules related to wheat crop productivity in higher accuracy and the lowest errors to help the farmers to expect the future productivity of the wheat crop. Bayesian and decision tree classifiers have been used to predict wheat crop productivity. Six scenarios have been run out, rainfall and relative humidity as the input variables with wheat crop yield based on Bayesian and decision trees algorithms. Among three classifiers (J48, Random Tree, and Naïve Bayes) used the results show that the Random Tree algorithm is the best and most appropriate for wheat crop productivity prediction according to the selected dataset.
Aviv and Lundsgaard-Nielsen [12] developed and designed a model to predict the variants that would contribute and assist in obtaining high production of soybean by utilizing soil properties and remote sensing. They employed and applied one of the machine learning methods Represented by a decision tree algorithm to detect the predictive Features for 21 factors. The researchers concluded that through this mode.

III. BACKGROUND

Data mining is the process of finding useful data, correlations, and patterns within a vast amount of datasets by employing various techniques like classification, clustering, and regression to predict the outcomes [13]. These outcomes will bring a lot of benefits surely by cut expenses and raises profit, reduce risks, and neutralization of blurry conditions. Recently data mining has contributed to the domain of agriculture in various studies and researches that employing data mining algorithms to address the complex problems in the field of agriculture and mainly the impact of climate conditions on seasonal crops and find the relations between weather conditions and production [3]. A large set of algorithms has been developed related to data mining depending on the nature of the problem and categorize types. A decision tree is one of the famous machine learning algorithms used in wide ranges. Thus it is a structure that applying to partition huge of a dataset into smaller subsets of records by running a sequence of easy rules. This manner can be used to construct the classification and prediction, Leaf node represents a decision and every branch should have a value [14]. This algorithm is called Hunt's algorithm and it is a greedy and recursive algorithm. The main objective of the decision tree is to build a training model to predict the class by learning decision rules that can be learned from actual data that represent the training data.
We should aware that decision trees consist of some important terminology, the first is the root node which represents the whole various dataset or apart and this can be divided into several identical sets, the second is the splitting which represents a procedure of separate the nodes into other sub-nodes, third is the decision node which represents the process of splitting a sub-node into other sub-nodes, fourth is the branch or sub-tree that represent a subsection of the whole tree, and finally the leaf or terminal node that represents the nodes that cannot be split and it represents the decisions and Fig. 1 show some explanation [15].

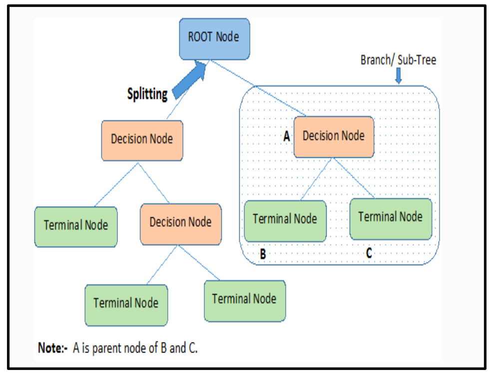

Fig. 1. Basic Decision Tree [15].

A decision tree could be a good machine learning choice for many reasons, the most important idea that the output is discrete, so we don't need an estimate or a probability, also it good when no large data is available, add to that when the data is noisy, thus it is low risk, don't hungry for data, very reliable, fast to train, and compact. And on the other hand, when classes are disjoint and the node doesn't share anything with other nodes and also other reasons.

IV. INFORMATION THEORY AND DECISION TREE

Shannon in 1948 published his famous paper entitled A Mathematical Theory of Communication to explain how much the information can be quantified with ultimate accuracy. This Theory considered is fundamental to the Interactive Dichotomizer-3 (ID3), C4.5, and C5.0 algorithm [16]. According to Shannon, the Entropy term represent the best-applied method, thus, in the beginning, it determines the amount of information that is given according to the event, hence the more probable the event is, the more information is provided [16]. Officially, and in the general meaning let us restrict things into binary, thus when we have a set of training sample (S) and have positive samples and also negative samples to represent two values or two situations for classification, so which mean that we know then the probability of positive samples represented by the cardinality of the set of positive samples divided by the cardinality of all training dataset, and the probability of negative samples represented by the cardinality of the set of negative samples divided by the cardinality of all training dataset as the following mathematical terminology.

$$P_{\oplus} = \frac{|S_{\oplus}|}{|S|} \qquad (1)$$

$$P_{\ominus} = \frac{|S_{\ominus}|}{|S|} \qquad (2)$$

$P_{\oplus}$, $P_{\ominus}$: the probability of positive and negative samples, $S_{\oplus}$, $S_{\ominus}$: positive and negative samples, S: training data set, Cardinality has been several elements in a set or number of members in set or list or vector, etc.
Hence, when we see this terminology we immediately think about entropy. In a more specific explanation, if we have a probability distribution $P = (p_1, p_2, p_3, p_n)$ and a training sample (*S*) then the information that we obtain by this certain distribution called the entropy of (*P*) [16], and following expression presented it:

$$Entropy\ (P) = -\sum_{i=1}^{n} P_{c=i} * log_2\,(P_{c=i}) \qquad (3)$$

$P_{c=i}$: is the probability of I^{th} classes.
Since any decision tree is constructed according to the principle of top-down, beginning from a root to split the

data into addition sets to obtain distinct groups which include the homogeneous elements, then the best way is to calculate the entropy. To construct a tree we need to measure the Gain when we add a specific attribute or some feature and which one of them adds value to our classification and which one of them is the principal component to put a tree together, the Gain is the expected reduction in entropy upon sorting on an attribute, thus we need to calculate the Gain and following expression presented it:

$$Gain(S, A) = Entropy(S) - \sum_{v \in VA} \frac{|Sv|}{|S|} Entropy(Sv) \quad (4)$$

Where S: is the data set, A: is the attribute, S_v: is the number of elements in a set and it is a subset of (S), V_A: is all values of all attributes. All these measurements help us to grow a decision tree when we have many attributes to select the best attribute to generate the most compact branching.

V. METHODOLOGY

There are several versions of decision trees; one of the new successor versions is the C5.0 algorithm. This algorithm is the extension of C4.5 and this last one is also an extension of the Interactive Dichotomizer-3 (ID3). It is a classification algorithm that is proper to deal with large data set. Additional improvements have been made than C4.5 about memory issues, efficiency, functionality, and speed running [17]. This algorithm represents the modern generation in the enhanced domain of Machine Learning Algorithms based on decision trees [18]. If we compared the C5.0 algorithm with older versions like C4.5 and ID3 according to data type, so we will discover very important features that added and developed, for instance, 1- the date, 2- timestamp, and 3- resolving the problem of missing value by adding a new data type called 'not applicable'.

One of the important advantages of this algorithm is that it saves a lot of time in computing and getting more reliable and accurate results by reducing features that are not necessary and are not important in cases that include large amounts of features or attributes and are not useful for the required classification by using such as post-pruning to avoid overfitting for instance, and thus it offers strong and effective rule sets to designated decision trees. Other improvements make it superior like supporting the sampling and cross-validation, and various trees are created and put together to enhance the predictions [18].

The core of the C5.0 algorithm is to divide samples into another subset based on the large information gain values. The subset of samples that are fetched according to the previous partitioning will be split several times. This process frequently runs until the subset cannot be split into another new subset. This splitting happens in general, according to certain fields [19]. Any samples set that has not any benefit or support the model accuracy will be excluded lastly.

In the agricultural fields, the use of climatic conditions and meteorology is one of the best ways to help farmers and agricultural authorities in obtaining future information about the productivity of crops and the conditions of their success, for example, the amounts of rain and temperature, play a vital role in the growth of plants and trees, including olives. From this methodology, we can enhance future forecasts and make it easier for those involved in the agricultural sector to develop strategies and make appropriate decisions in the near term. Therefore, to provide the best recommendations and future scenarios, we would like to have multiple characteristics that play the largest role in determining the quantities of production according to different climate elements and production quantities over the past years. Based on the data mining techniques, we can take advantage of these techniques to discover the rules from available data that are previously obtained.

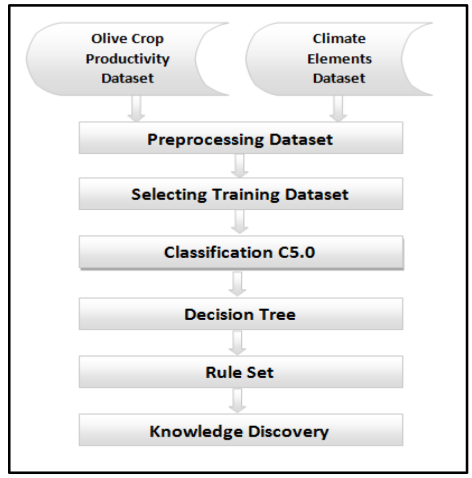

Fig. 2. Methodlogy Diagram.

Primarily, for a training data set, we use the famous decision tree algorithm named C5.0. Fig. 2 presents the block diagram of the model.

The steps applied in the study:

1) Initially, we assign the dataset of the olive productivity and weather conditions as training data for algorithm processing. 17 records assigned to the training set and 8 records assigned to the testing site.
2) Adopt the olive productivity as a target and the other climatic element attributes applied as inputs.
3) Employ the decision tree - C5.0 algorithm in a dataset to create the root and leaf for the tree by running the splitting procedure.
4) Execute a similar procedure for all attributes until the final split and hence construct the decision tree.
5) Some pruning was carried out on the tree to obtain the major accuracy.

Our proposed model to generate a C5.0 decision tree (some features have gotten from [17]) depends on the following steps:

Input:

➢ Dataset, D, training sets rows with related classes.
➢ Attribute_ list, set of selected factors [climate...].
➢ Attribute_ splitting _C5.0_manner, it is a process to assign the data rows into a certain class by using the large information gained either at the subset level or hole training data set.

Output: [C5.0 decision tree]

Procedure:

❖ *Produce a node N*
❖ *IF all rows interiors the (D) from the same class (C), THEN*
❖ *Assign (N) such as a leaf node to the specific class (C)*
❖ *IF attribute_ list = null, THEN*
Assign (N) such as a leaf node to the majority class in (D)
❖ *Carry out Attribute_ splitting _C5.0_manner(D, Attribute_ list) to locate the best splitting_ norm*
– Assign node (N) to splitting_ norm
– IF attribute_ splitter are discrete and multiple divides are available THEN
– Attribute_ list ←Attribute_ list | Attribute_ splitter
– According to every result $_{(y)}$ of splitting_ norm

– *Assign $D_{(y)}$ to become like a set of data rows in (D) regarding result $_{(y)}$*
– *IF $D_{(x)}$ is null THEN append a leaf to majority class-related (D) for a node (N)*
– *Else, append node produced by C5.0 decision tree ($D_{(y)}$, Attribute_ list) for the node (N)*
– *Return (N)*

VI. EXPERIMENTAL RESULTS AND DISCUSSION

In this paper, to establish and find the most influential climate parameters that play the greatest role in determining the level of olive oil production, it was necessary to collect sets of data related to climatic elements in the Jenin Governorate in Palestine, as well as the quantities produced from olive oil over the past 25 years. The actual data was aggregated from 1995 until 2019 to both olive oil productivity, quantity, and the different climate factors. The historical data for the climatic elements recorded were obtained from the Palestinian Meteorological Department, and historical data for the quantities of olive oil production was obtained from the Palestinian Central Bureau of Statistics.
The Climatic factors adopted in our analysis are average rainfall, the average maximum temperature, average relative humidity, and average evaporation with the related class of olive oil productivity to training the C5.0 decision tree algorithm, Table 1 shows our data set collected.

Table 1: Different climate factors and Olive oil productivity.

Year	Rainfall mm	Temperature ċ	Relative Humidity %	Evaporation mm/day	Oil Productivity	
1995	56.7	19.7	65	5.7	4579.308 →	**High**
1996	48.2	26.1	65	5.7	2668.075 →	**Low**
1997	78.2	25.3	72	5.4	1323.184 →	**Low**
1998	52.3	26.4	60	5.5	3900 →	**High**
1999	37.1	26.5	64	5.6	707 →	**Low**
2000	81.4	26.3	66	5.9	5908 →	**High**
2001	22.2	26.6	64	5.5	1034 →	**Low**
2002	55.6	26.2	61	5.6	6371 →	**High**
2003	84.5	26.4	64	5.9	2908 →	**High**
2004	59.8	25.6	65	5.5	4712 →	**High**
2005	47.4	25.6	65	5.4	1481 →	**Low**
2006	62	25.6	65	5.8	6811.5 →	**High**
2007	50.7	25.9	65	5.8	1576.6 →	**Low**
2008	49	26.3	63	5.6	4866.4 →	**High**
2009	84.6	26.2	66	5.4	1584.6 →	**Low**
2010	52.5	28	64	5.8	7183.8 →	**High**
2011	55.8	25.5	69	5.2	3369.636 →	**Low**
2012	92.9	25.7	68	7.4	7835.18 →	**High**
2013	60	26.4	66	5.7	4375.996 →	**High**
2014	49	26.6	66	5.6	6645.622 →	**High**
2015	66.1	26.5	66	5.6	5522.743 →	**High**
2016	54.9	27	64	5.8	5035.646 →	**High**
2017	29.2	26.8	65	5.7	4371.7 →	**High**
2018	84.8	26.9	69	5.5	4871.074 →	**High**
2019	81.2	26.5	67	5.6	8997 →	**High**

From the average of olive oil production, we can classify the yields into high and low values as the last column in the previous table. The classification based on the C5.0 decision tree algorithm for factors that influence the olive oil yield was carried out by using the equation Entropy (P) that was presented previously for each average rainfall, the average maximum temperature, average relative humidity, and average evaporation with the related class of olive oil productivity individually and Gain values also calculated based on other presented equation.
From the C5.0 decision tree algorithm analysis, we find that the influence of evaporation has the largest Gain of 1.0 followed by rainfall that has less and Gain of 0.88 and last, is the maximum temperature has of 0.29 as depicted in Fig. 3.
To predict the final production of olive oil, given the dataset of all continuous climate parameter values and also the categorical values of olive oil production, we applied the C5.0 decision tree algorithm for all yields of years. Fig. 4 shows a decision tree constructed for olive oil crop productivity.

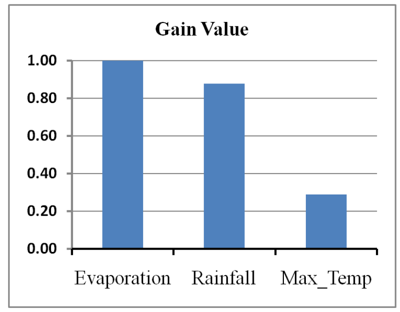

Fig. 3. The gain obtained for olive oil yield.

Interestingly, we discovered the root node is evaporation as a major impact factor on the crop yield. This means that this parameter is the most important element and it is linked strongly to the final crop yield. Also, we discovered that other branch/sub-tree ore decision nodes like rainfall and maximum temperature form important parameters as a final result and predicted decisions.
This information on this model is very useful for both farmers and the authorities of agriculture to predict the productivity of olive oil crops in the early time of the year and to allow for them to make a future perception of olive oil yield and for the strategic planning for productions also.

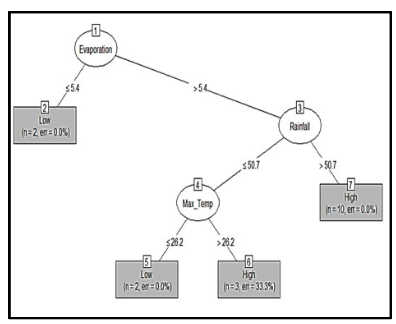

Fig. 4. Decision tree constructed for olive oil crop productivity.

Below are the precise rules that were generated from the tree that is useful to expect the climatic elements responsible for either high or low olive oil crop produced according to certain circumstances.
Rules:
Rule 1:IF Evaporation <= 5.4 THEN
[Low Productivity]
Rule 2:IF Evaporation > 5.4 Rainfall > 50.7 THEN
[High Productivity]
Rule 3:IF Evaporation > 5.4 Rainfall < = 50.7 AND Maximum Temperature > 26.2 THEN

[High Productivity]
Rule 4:IF Evaporation > 5.4 Rainfall < = 50.7 AND Maximum Temperature <= 26.2THEN
[Low Productivity]

This experimental result shows the Possibilities of the applied decision tree model, especially the C5.0 algorithm since it achieves the generalization correlations and accuracy of about 90%. According to the outcome of our analysis, it is precise and reliable where the error was below 10%. This error to expect a close value was due to a lack of more training data set where all datasets did not exceed 25 historic records and this in itself is an achievement and excellence.

VII. CONCLUSION

The present research applied the C5.0 decision tree to estimate the impact of different climatic parameters on the olive oil crop productivity. This study clarifies the modern perspective to use this method for extracting new helpful knowledge from existing raw data on weather conditions. The C5.0 decision tree demonstrates that there are an existing relationship and the high correlation between olive oil productivity and the various climatic factors that influence olive oil productivity through presented rules and prediction accuracy. The most important conclusions of the study are:

(i) The C5.0 decision tree algorithm analysis points out that the production of olive oil was significantly affected by evaporation followed by rainfall and maximum temperature.

(ii) The C5.0 decision tree proves very fast in executing and has a large possibility in representations and extracting the knowledge and pattern discovery.

(iii) Regarding high or low olive oil productivity, the C5.0 decision tree very useful in predicting the responsible climate factors and controls productivity based on strong generated rules.

Moreover, our study can help balance the local market for the olive oil sector, estimating future needs from exports and imports, setting future strategic plans, and achieving profits in the near term.We will seek to have new information sources of data related to other products as soon as is available to use them in predicting using the modern machine learning methodology and data mining techniques. Our model will be enhanced and publicize in different sectors with collaboration with the Palestinian Ministry of Agriculture, moreover, other implementations will be developed in predictions and classification to contribute the scientific research in our country and to invest our time and lives.

ACKNOWLEDGMENTS

The authors acknowledge with thanks the ministry of agriculture for providing us with the Olive Oil Productivity in the last years.

Conflicts of Interest. The authors declare no conflict of interest.

REFERENCES

[1]. Palestinian Central Bureau of Statistics (2010) Agricultural Census, Final Results - Palestinian Territory. Ramallah, Palestine.

[2]. Musa, Fatima, & Ahmad Omar (2008). Effect of Climate on the Rate of Olive Productivity in the West Bank. Ph.D. dissertation.

[3]. Fathi, M. T., Ezziyyani, M., Cherrat, L., Sendra, S., & Lloret, J. (2017). The relevant data mining algorithm for predicting the quality of production of olive in Granada region influenced by the climate change. In Proceedings of the Mediterranean Symposium on Smart City Application (pp. 1-6).

[4]. Abdullah, Samir & Hussein, Imad (2016). Toward enhancing the competitiveness of the Palestinian Olive sector.https://www.mas.ps/files/server/20172601130056-2.pdf (accessed 15 Jan 2020).

[5]. Fornaciari, M., Orlandi, F., & Romano, B. (2005). Yield forecasting for olive trees: a new approach in a historical series (Umbria, Central Italy). *Agronomy Journal, 97*(6), 1537-1542.

[6]. Awad, M., & Zaid-Alkelani, M. (2019). Prediction of Water Demand using Artificial Neural Networks Models and Statistical Model. *International Journal of Intelligent Systems & Applications, 11*(9).

[7]. Veenadhari, S., Mishra, B., & Singh, C. D. (2011). Soybean productivity modelling using decision tree algorithms. *International Journal of Computer Applications, 27*(7), 11-15.

[8]. Shakoor, M. T., Rahman, K., Rayta, S. N., & Chakrabarty, A. (2017). Agricultural production output prediction using supervised machine learning techniques. In *2017 1st International Conference on Next Generation Computing Applications* (NextComp) (pp. 182-187). IEEE.

[9]. Veenadhari, S., Misra, B., & Singh, C. D. (2014). Machine learning approach for forecasting crop yield based on climatic parameters. In *2014 International Conference on Computer Communication and Informatics* (pp. 1-5). IEEE.

[10]. Vidyavathi, B. M. (2015). Ameliorated methodology for the design of sugarcane yield prediction using decision tree. *Compusoft, 4*(7), 1882.

[11]. Singh, R. (2017). Wheat Yield Assessment using Decision Tree Algorithms. *International Journal of Advanced Research in Computer Science, 8(*5).

[12]. Aviv, T., & Lundsgaard-Nielsen, V. (2017). Ensemble of cubist models for soy yield prediction using soil features and remote sensing variables. In Proceedings of the *23rd ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada* (pp. 13-17).

[13]. Gera, M., & Goel, S. (2015). Data mining-techniques, methods and algorithms: A review on tools and their validity. *International Journal of Computer Applications, 113*(18).

[14]. Pandya, R., & Pandya, J. (2015). C5. 0 algorithm to improved decision tree with feature selection and reduced error pruning. *International Journal of Computer Applications, 117*(16), 18-21.

[15]. Decision Tree Algorithm, Explained - KDnuggets. https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html (accessed 25 Jan 2020).

[16]. Hssina, B., Merbouha, A., Ezzikouri, H., & Erritali, M. (2014). A comparative study of decision tree ID3 and C4. 5. *International Journal of Advanced Computer Science and Applications, 4*(2), 13-19.

[17]. Revathy, R., & Lawrance, R. (2017). Comparative analysis of C4. 5 and C5. 0 algorithms on crop pest data. *International Journal of Innovative Research in Computer and Communication Engineering, 5*(1), 50-58.

[18]. Bujlow, T., Riaz, T., & Pedersen, J. M. (2012). A method for classification of network traffic based on C5. 0 Machine Learning Algorithm. In *2012 international conference on computing, networking and communications (ICNC)* (pp. 237-241). IEEE.

[19]. Patil, N., Lathi, R., & Chitre, V. (2012). Comparison of C5. 0 & CART classification algorithms using pruning technique. *International Journal of Engineering Research and Technology*, *1*(4), 1-5.

How to cite this article: Alkelani, M. Z. and Awad, M. (2021). Prediction of Olive Oil Productivity using Machine Learning Decision Tree Algorithm. *International Journal on Emerging Technologies, 12*(1): 13–18.