Comparison of Machine Learning Algorithms for the Power Consumption Prediction

- Case Study of Tetouan city –

Abdulwahed Salam, Abdelaaziz El Hibaoui
Faculty of Sciences
Tetouan, Morocco
Abdelmalek Essaadi University
Sallam.ye@yahoo.com, hibaoui@uae.ma

Abstract— **Predicting electricity power consumption is an important task which provides intelligence to utilities and helps them to improve their systems' performance in terms of productivity and effectiveness. Machine learning models are the most accurate models used in prediction. The goal of our study is to predict the electricity power consumption every 10 minutes, and/or every hour with the determining objective of which approach is the most successful. To this end, we will compare different types of machine learning models that recently have gained popularity: feedforward neural network with backpropagation algorithm, random forest, decision tree, and support vector machine for regression (SVR) with radial basis function kernel. The parameters associated with the comparative models are optimized based on Grid-search method in order to find the accurate performance. The dataset that is used in this comparative study is related to three different power distribution networks of Tetouan city which is located in north Morocco. The historical data used has been taken from Supervisory Control and Data Acquisition system (SCADA) every 10 minutes for the period between 2017-01-01 and 2017-12-31. The results indicate that random forest model achieved smaller prediction errors compared to their counterparts.**

Keywords—Energy Prediction, Artificial Neural Networks, Random Forest, Decision Tree, Support Vector Regression, Linear Regression

I. INTRODUCTION

Over time, increased consumption of electric power and increased attention to its details, such as electrical forecasting, have been concentrated by researchers. Also, Electricity companies have been spending a lot of money and effort to control and manage their electric power effectively. Therefore, it becomes necessary to absorb all new and available methods and to choose the best in proportion to the nature and quality of services provided by energy companies. One of those things which needs to be considered is how to know the energy produced and consumed in order to balance production and consumption, to decrease the cost of production, and to control future planning?

Knowing the real production and consumption of power is the first step of making a good electrical system. To save resources and reduce costing, power utilities are required to balance between produced power and customers' consumption. The relationship is described by the following equation:

$$production - (consumption + ELT) = v$$

Where ELT is the energy lost in the transportation and v is real number. According to the value of v we distinguish between the three following cases:

- If v is a large positive number, that means there is a quantity of energy produced but not used. In general, this exceeded energy is lost. The problem is up to now, the extra power needs to be stored or the production needs to be reduced.
- If v is a large negative number, that means the consumption is larger than production. In this case the problem blackout can be occurred. So, new resources of energy are required to handle this situation.
- If v is positive and v is near to 0, that means it is a stable electrical system and there is a harmony between the production and consumption.

The goal of power companies is to keep the value of v like the third case. They have to make a balance between production and consumption. To this end, they need a strong system which stands for accurate prediction.

The electric power consumption has increased in the last decade due to the growing of economic development and population. Specifically, the annual electricity power consumption is increasing in the industrial and domestic sectors. There are many factors that determine the energy consumption, we cite here weather, population, price of electricity and consumer behavior. The variability of those factors makes the energy prediction more difficult.

Considering this complexity, researches were concentrated on finding the most accurate prediction models to predict the real demand of energy consumption. So, many algorithms and models have been proposed to offer solutions to the power prediction. In general, those algorithms are classified into three categories: statistical, engineering, and artificial intelligence. In literature, several researchers use artificial intelligence algorithms to make a prediction model especially that one's of machine learning algorithms.

Pınar Tüfekc[1] examined machine learning to predict full load electrical power output of a base load operated combined cycle power plant. Evangelia Xypolytou et al. [2] studied Short-term electricity consumption forecast with artificial neural networks a case study of office buildings.

978-1-7281-1182-7/18/$31.00 ©2018 IEEE

Muhammad Waseem Ahmad et al. [3] compared feed-forward back-propagation artificial neural network with random forest to predict power consumption of a hotel in Madrid, Spain. M. ErdemGünay[4] predicted annual gross electricity demand by artificial neural networks case of Turkey. Murat Kankal et al [5] studied the performance of an artificial neural network for modeling electricity energy demand in Turkey. K.P.Amber et al [6] compared five artificial intelligent system techniques to predict electricity power consumption of a building located in London. Fazil Kaytez et al. [7] compared the regression analysis, neural networks and least squares support vector machines for predicting the electricity energy consumption of Turkey. Henrique Pombeiro et al. [8] compared Linear regression vs. fuzzy modeling vs. neural networks models to predict electricity consumption in an institutional building. Subodh Paudel et al. [9] predicted energy consumption of low energy building based on support vector machine. Hamid R. Khosravani et al. [10] compared prediction models for energy consumption based on neural networks of a bioclimatic building.

Those previous cited models may be appropriate for some cases but it cannot be generalized.

In this study, we will compare the well-known machine learning algorithms in order to increase the efficiency and revenues of the electrical generating and distribution networks companies, and to assist them planning their capacity and operations to supply all consumers with the required energy reliably. The availability of historical data allows us to use the supervised models such as decision tree, support vector machine for regression, artificial neural network (feedforward neural network with backpropagation algorithm) and random forest. Those models will also be compared to the linear regression method. The comparative study is based on historical consumption energy data of Tetouan city for the period between 2017-01-01 and 2017-12-31. The historical data used are for Quads, Boussafou and Smir distribution networks. It was taken from the SCADA system of the regional distribution company of drinking water and electricity (AMENDIS). This data is exclusive and have not been used before our research. Due to the dependence of prediction models on the input variables to obtain the best output, we will include the weather data that is taken for the same period of power consumption. Moreover, we will use the attributes of date and time as independent variables, study the impact of those factors on the prediction, and determine the importance of each factor on power consumption.

The rest of this paper is organized as follows: Section II exposes a technical overview of different machine learning algorithms in comparison. Section III presents the case study and the description of datasets. Section IV is devoted to the methodology used. Experiments and their results are covered in Section V. And finally, Section VI concludes the paper and dresses some perspectives.

II. Overview of Machine Learning Algorithms

In this section, we will provide a brief description of five different machine learning methods: linear regression, decision tree, random forest, support vector machine for regression with radial basis function kernel, and artificial neural network. Here, we use the common notations to describe the algorithms.

A. Linear Regression

Linear regression is considered as one of the simplest approaches and it can be used as a baseline performance measure. It is based on linear relation between the dependent and independent variables[11]. It is defined by this equation:

$$f(x) = a_0 + a_1 x_1 + a_1 x_1 + \cdots + a_n x_n \quad (1)$$

Where $x_1, x_{2,\dots,} x_n$ are the available inputs and $a_0, a_1, \dots, a_n$ are the functional weights.

B. Decision Tree

The decision tree is a commonly used machine learning method [12]. It is a kind of classification or regression and it utilizes a tree structure to separate a set of data into several predefined classes giving the characterization, generalization and classification of given datasets[13].

Its goal is to predict the target variable value by learning simple decision rules deduced from the data and it shows how the target variable can be forecasted by predictor variables set.

There are many types of decision tree generation such as ID3[12], C4.5[14] and classification and regression trees (CART)[15]. In this work, we implemented CART, along with a Scikit-learn: Machine Learning in Python[16]. CART is nonparametric procedure to predict continuous dependent variable which utilizes a binary tree to divide the predictor space into subsets recursively[17]. According to our study, we found the performance of this model better than its family counterpart.

C. Random Forest

Random forest is classified as an ensemble approach which combines the performance of numerous decision tree algorithms to predict the variable value. It was proposed to improve the accuracy of decision tree and it has different construction for regression and classification.

A number of N regression trees are built by random forest and the result is the average. While N trees are grown, a regression predictor is defined as:

$$f_{rf}^{N}(x) = \frac{1}{N} \sum_{n=1}^{N} T_{ree}(x) \quad (2)$$

$$X = \{x_1, x_2, \dots \dots, x_p\}$$

Where x is p-dimensional vector of inputs and $T_{ree}(x)$ is referred to decision tree.

Selecting randomly a set of trees in the forest is accomplished to make new training set. The set of unselecting trees is known as out of bag samples[18]. Random features are selected in each split node of a decision tree instead of all features. This process is repeated in order to create a random forest [19]. Aggregation of each individual prediction trees makes the prediction of the random forest and this aggregation prediction gets better performance than the individual prediction of trees [20]. Moreover, random forest provides an estimation of the relevant important features and how each feature affects the prediction [21]. In sum, simplicity, velocity, interpretability, accuracy and ease of use are the most important properties of random forest.

D. Support Vector Machine

Support vector machine was introduced in the late 1960s and it has not got significant consideration until recent years. SVM is a type of supervised method to achieve the classification of multidimensional and it was originally invented as a linear classification then to a non-linear classifier. Lately, it was used to solve regression problems[22] which is based on the concept of support vectors and called support vector regression (SVR).

It is defined as:

$$y = f(x) = w^T\Phi(x) + b \quad (3) \qquad (3)$$

Where Φ is any nonlinear function to map input to output:

$$\Phi: x \to \Phi(x) \in R^H$$

The best solution is detected by minimizing the following function:

$$\frac{1}{2}\|w\|^2 + \mathrm{C}\sum_{i=1}^{n}\xi_i + \zeta_i \quad (4)$$

$$subject\ to \begin{cases} y_i - w^T\Phi(x) - b \le \epsilon + \xi_i \\ w^T\Phi(x) + b - y_i \le \epsilon + \zeta_i \\ \xi_i, \zeta_i \ge 0 \end{cases}$$

Where C is constant to control the penalty factor which is used to balance between smoothness and data fitting. ξ and ζ are slack variables to optimize the problems and ϵ is the loss function which is used to estimate the accuracy of prediction. One advantage of SVR is finding a unique solution to minimize the convex function and it depends on providing C and ϵ [23][24].

E. Artificial Neural Network

Artificial neural network imitates the work of the brain. It is a technology that is currently widely used due to its ability to solve complex issues. Also, the artificial neural network is the most common method to develop nonlinear problems of regression and classification. Many types of networks are available in literature. Here, we concentrated on the most used one, namely: feedforward neural network. It learns through training not through programming and it collects the knowledge by identifying the relationships of data. Feedforward neural network basically consists of at least three layers, an input layer which receives the data and processes it to the hidden layers, the hidden layers connect the input layer to the output layer through connections and the output layer in our case (regression) combines of one output and there are no connections back from the output layer to the hidden layer or from the hidden layer to the input layer. The values are processed through transfer function from the input layer to the hidden layer and multiplied by the connection values. Also, the values are forwarded from the hidden layers to the output layer in the same way. Mathematically, it can be described by the following equation:

$$Y_o = f_o\left(\sum_{i=1}^{n_{hl}} w_i f_i\left(\sum_{j=1}^{n_{hl-1}} w_{ij} f_j\left(\sum_{k=1}^{n_{hl-2}} w_{jk} f_k(..) + b_j\right) + b_i\right) + b_o\right) \quad (5)$$

Where Y_o, f, l, n_{hl}, w, b_i represent neural network output, the activation function, the number of hidden layers, the number of neurons in the hidden layer, the weight of connections, and the bias of the neuron respectively. The transfer function is called activation function and there are many types. Sigmoid[25], Hyperbolic Tangent Function(Tanh)[26], Rectified Linear Unit (ReLU)[27], Exponential linear Unit (ELU)[28], Scaled Exponential linear Unit (SELU)[29] and Swish[30] activation functions were used in this work.

Feedforward neural network learn through different types of learning rules, but backpropagation is the most used algorithm. To reduce errors, learning rule is used with optimization algorithms to find the best parameters and compare the predicted output value with the real value and the errors' feedback in order to adjust the weight of connections. This step is repeated until it reaches the minimum number of errors or number of epochs.

III. CASE STUDY

Tetouan is a city located in the north of Morocco which occupies an area of around 10375 km² and its population is about 550.374 inhabitants, according to the last Census of 2014, and is increasing rapidly, approximately 1.96% annually. Since it is located along the Mediterranean Sea, its weather is mild and rainy in the winter, hot and dry during the summer months. The power consumption data was collected from Supervisory Control and Data Acquisition System (SCADA) of Amendis which is a public service operator and in charge of the distribution of drinking water and electricity since 2002. The purpose of the electricity distribution network is to serve low and medium voltage consumers in Tetouan regions. For this purpose, the delivery and distribution of electrical energy from the point of delivery to the end user, the customer, is ensured by Amendis. The energy which is distributed comes from the National Office of Electricity and Drinking Water. After transforming the high voltage (63 kV) to medium voltage (20 kV), it is allowed to transport and distribute the energy. The distribution network is powered by 3 source stations, namely: Quads, Smir and Boussafou.

The data which is used in this study was the historical data of power consumption which was collected every 10 minutes for the period between 2017-01-01: 00:00:00 and 2017-12-31: 23:50:00. It is a unique dataset, and it does not have any missing data. It is consisted of the date, time and the consumption of the three distribution networks. Figure 1 shows the output consumption of the three distribution networks for the whole year of 2017 at each hour.

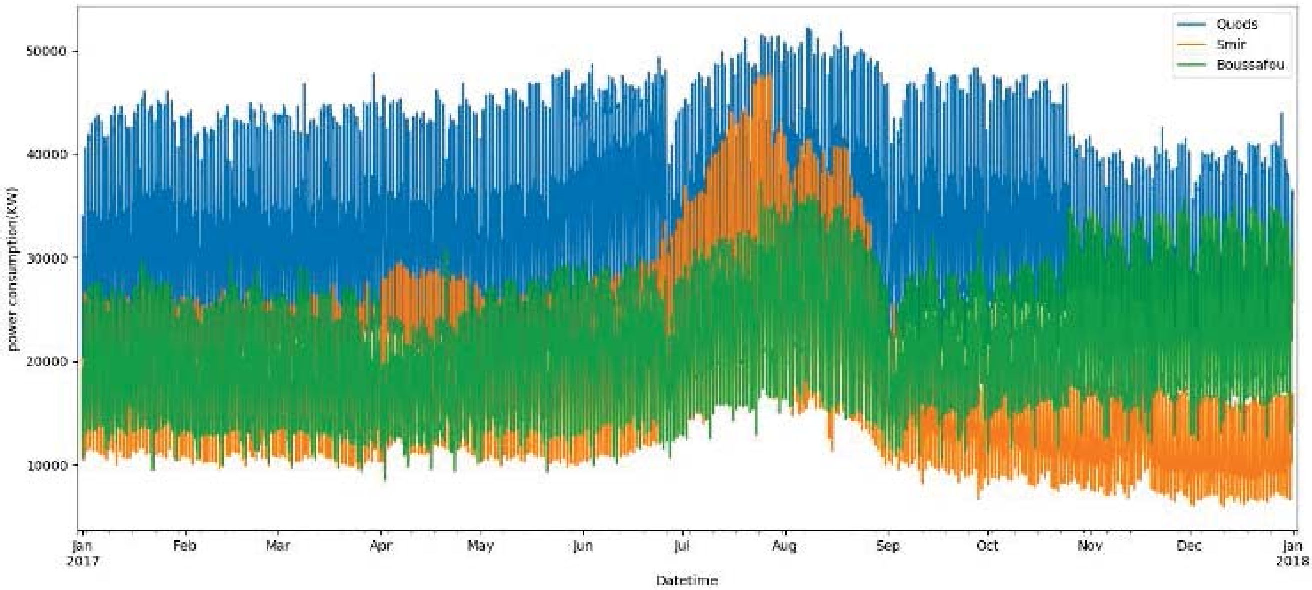

Fig 1. Power consumption for the year of 2017 at each hour for three distribution

There are similarities and differences between the three distributions. The increasing of power consumption in the summer is similar and that because of the hot weather and vacation time (the number of visitors grow up the population). But the difference is the reduced power consumption of Quads and Smir distribution in November and December compared to the power consumption of Boussafou distribution at the same months.

Different attributes of date and time are used as the inputs for the prediction models. Month, day of month, hour, day of year, week of year, day of week, quarter and minute are the independent variables and their correlation to the dependent variable which is shown in Figure 2.

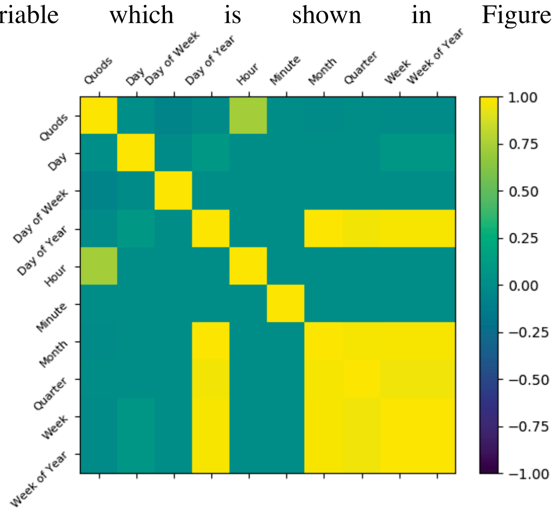

Fig 2 Correlation relationship between power consumption of Quads distribution network and calendar variables

Due to the people behavior, the consumption of power is changed in working days compared to weekends. Usually, on working days the consumption is decreased in household and increased in factories, commercial and public establishments and the opposite happened on weekends. Our dataset is aggregated data and it is not determined for specific types of building to know the general effect of working days or weekends. Figure 3 shows the consumption of the week days and it shows slight electricity power consumption is used on Sunday compared to the other days.

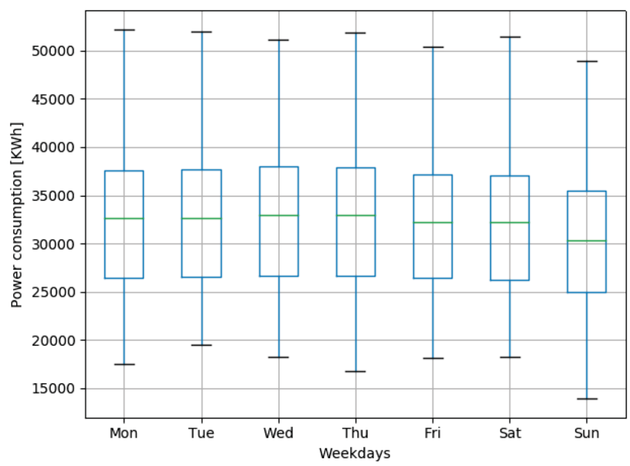

Fig 3. Box plot comparison of electricity consumption among week days

With no doubt, there are many factors which affect the power consumption such as weather, income, population, electricity price and etc. One of the factors that used in this work is weather and its data was gathered from sensors. Those sensors are located both in the airport at the center of the city and in Faculty of Science. This data is collected in the period between 2017-01-01 and 2017-12-31 every 5 minutes. We reformed the data to be in every 10 minutes like the power consumption data by resampling the data and taking the average of two reading. The feathers of weather used in our study are: temperature, humidity, wind speed, diffuse flows and general diffuse flows.

Table I shows the weather properties and the correlation between the power consumption of Quads distribution and the corresponding weather over the full period of the dataset

TABLE I WEATHER PROPERTIES AND COEFFICIENT OF CORRELATION BETWEEN THE INPUT VARIABLES AND THE OUTPUT VARIABLE

	Count	Mean	STD*	Min	Max	Correlation
Quads	52560	32330	7133.05	13895.7	52204.4	1
Temperature	52560	18.81	5.82	3.247	40.010	0.440221
Humidity	52560	68.26	15.55	11.34	94.8	-0.287421
Wind Speed	52560	1.96	2.34	0.05	6.483	0.167444
Diffuse flows	52560	75.03	124.21	0.011	936.00	0.080274
Global Diffuse	52560	182.67	264.41	0.004	1163.00	0.187965

*STD: Standard derivation

In Table I and Figure 2, we showed the power consumption correlation to the calendar and weather attributes. Also, we applied feature selection to data in order to determine the importance of predictive variables and get rid of the unimportant features. There are many ways to perform that, one of them is random forest which identifies the true predictor of a large number of candidates[31]. It is shown in Figure 4 that all variables are important but hour and temperature are the most valuable.

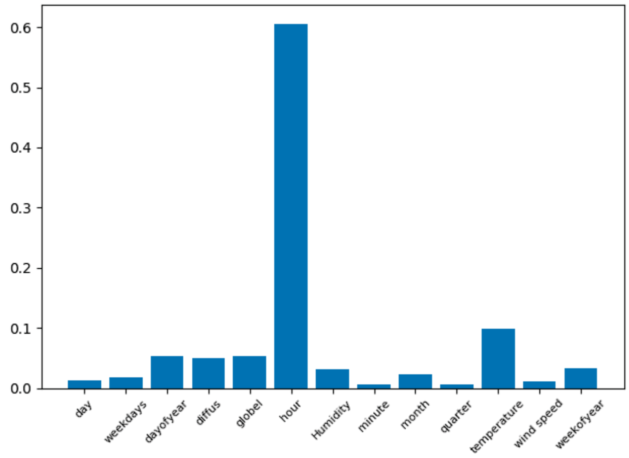

Fig 4 Variable importance for Quads distribution dataset for minutely consumption

IV. METHODOLOGY

In this study, we used five known types of algorithms that used in prediction. Despite their advantages and the accuracy of the algorithms, these models require an accurate selection of the arrangement parameters in order to achieve the best performance. We utilized grid search method to find the best parameter for the models. It is classified as exhaustive method for the best parameter values. The grid search method is recommended to be used along with cross-validation in order to obtain best values [33]. It has to explore each parameter by setting sort of values at first. Then, the method will show the score for each parameter value to be considered which one will be selected. This method is applicable when the required maximum of parameters is known[32]. In this work, the calculations have been implemented using Python 3.6, with base algorithm from Keras[34] and scikit-learn[16].

As it mentioned in the case study section, we got the data from different resources. After extracting, transforming and loading data from the resources, we normalized the data as a result of depending some models' performance on normalization such as neural network. We transformed the data into the values between zero and one by using Min-Max Normalization which is one of the best used techniques. This normalization is achieved by:

$$X = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (6)$$

We used grid search to find the best parameters for algorithms. Random Forest algorithm depends on several hyperparameters. Selecting the appropriate values of these parameters is an important step to get the most accurate result and there is no rule to be determined and followed. The most important parameters are number of trees in the forest, number of features to consider at each split, max depth of each tree in the forest, the required minimum number of samples to split and the minimum number of samples demanded to be at a leaf node. Those parameters are optimized by cross-validation and grid search method. A rang of parameter values were selected and trained on them. The number of trees parameter was tested on set of the values {10, 20, 30, 50, 100, 200, 300}. The number of features was tested on {1, 2, 3, 4, 5, 6, 7, 8, 9} etc. The grid search method shows the score for each parameter value to be considered as chosen values. The best values which are gotten by the grid search method were *30, 7, None, 2* and *1* for the number of trees and the number of features, max depth of the tree, min samples split and min sample leaf parameters respectively.

In decision tree, different parameters need to be set and examined to compare the result with other algorithms. The most important parameter of decision tree to be selected were tested in sets by grid search were: the depth of the tree, the minimum number of samples required to split an internal node, the minimum number of samples required to be at a leaf node and the number of features to consider when looking for the best split. The best result of these parameters which were gotten by grid search were *None, 10, 10* and *9* for the parameters above respectively.

Support vector regression is characterized by usage of kernel functions. Radial basis function is used in this work according to its lower error compared to polynomial, linear and sigmoid kernels as it is reported by [35]. Cost and gamma are kernel parameters and are required to be optimized. To assess the accuracy of support vector regression, we tested the model on different parameter companions. The values of cost parameter were in the set {1, 10, 100, 1000} and the values of gamma were in the set {0.01, 0.001, 0.0001}. The best performance result was when the cost equals to 10 and when gamma equals to 0.01.

The most challenge in feedforward neural network is how to define the number of hidden layers and the number of neurons in each hidden layer. Also, choosing the suitable activation function is another challenge. A lot of studies tried to figure it out but no real rule can define that. In this work, the number of factors is optimized by grid search. The number of hidden layers and neurons are selected by grid search. For one hidden layer we specified the number of neurons by the following function [36]:

$$N_h = N + 1 \qquad (7)$$

Where N is the number of input data.

In one hidden layer, we also tested the neural network on {4,6,8,9,10,11,12,13,16,18,20,25,30} neurons according to [37].

Moreover, the grid search tested accuracy of another number of neurons in a deep network with two hidden layers. Each hidden layer consists of 30 and 20 layer respectively[37]. The model of nine hidden layer consists of 200, 160, 120, 80, 60, 40, 30, 20 neurons. For this model, each layer was also examined by the grid search. Another factor to be optimized by grid search is the activation function. Sigmoid, Tanh, ELU, ReLU, SELU and Swish activation functions have been considered in the grid as a result of their variety and popularity. The optimization algorithms which were used in the grid search to be optimized are stochastic gradient descent (SGD) [38] and Adam[39]. One hidden layer with 10 neurons is selected by grid search, SELU activation is also chosen and the Adam is implemented as the preferable optimizer. We manually set the number of epochs to 100. The initial learning rate is seted to 0.001. The initialization of training is Glorot uniform initialization[26]. We used no dropout and 0.9 momentum.

V. Experiments and Result

The original dataset was collected over 10 minutes and our study is examined for the prediction of 10 minutes and one hour power consumption periods to give the utilities the ability of decision making. All independent inputs are used in the experiment according to their effect which is explained and showed in the analysis of parameters.

We applied performance criterion to evaluate the models. We utilized two different measures: Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) which are defined as:

$$RMSE = \sqrt{\frac{\sum_{k=1}^{N}\left(Y_{f,K}-Y_{r,K}\right)^2}{N}} \qquad (8)$$

$$MAE = \frac{\sum_{k=1}^{N}\left|Y_{f,K}-Y_{r,K}\right|}{N} \qquad (9)$$

Where Y_f is the predicted values, Y_r is the actual values and $\bar{Y}$ is the average.

In order to evaluate the five models on datasets, 10 minutes consumption and an hour consumption, datasets are divided into a train set and test set. Each algorithm is trained by using 75% of the data and 25% for testing. The test set is usually used to judge the models, but we also used training set to show the ability of learning. For comparison purposes, we compared the median of 9 implementation in all models of dataset. All parameters that have optimized by search grid before were for the comparison of dataset of 10 minutes power consumption.

The experimental results for the prediction of the 10 minutes period are presented in Table II. From the two performance criterions, it is shown that random forest model achieved the best results for the four examinations. Also, it is noticed that feedforward neural network achieved a close result to the random forest in Quads distribution network.

TABLE II. RSME AND MAE COMPARISON OF ALGORITHMS IN 3 DISTRIBUTIONS FOR 10 MINUTES POWER CONSUMPTION

Algorithm	Quads Distribution				Smir Distribution				Boussafou Distribution				Aggregated Distribution			
	RSME		MAE		RSME		MAE		RSME		MAE		RSME		MAE	
	Train	*Test*	*Train*	*Test*	*Train*	*Test*	*Train*	*Test*	*Train*	*Test*	*Train*	*Test*	*Train*	*Test*	*Train*	*Test*
RF	671.7	**3174.7**	472.8	2663.5	214.1	**2336.9**	135.6	1939.6	594.5	**3227.8**	420.5	2475.9	482.3	**4481.1**	318.5	3595.3
DT	840.2	4613.9	550.7	3962.3	306.7	2849.8	179.3	2396.3	611.7	3543.7	405.9	2759.5	790.6	5957.3	490.0	4835.5
SVR	4092.3	3898.7	3192.3	3046.0	4205.5	5584.3	3298.6	4680.8	3334.4	3981.3	2671.6	3066.6	10821.8	9647.2	8505.6	7692.8
FFNN	2562.2	3203.6	1945.7	2601.2	3815.8	4877.8	2976.9	4007.1	2731.5	3745.6	2119.5	2965.2	6487.6	7045.9	4985.02	5583.9
LR	4404.2	3925.5	3522.1	3112.2	4068.4	4949.9	3213.4	4033.4	3142.8	5785.7	2504.1	4647.2	10687.4	10152.5	8450.9	8110.3

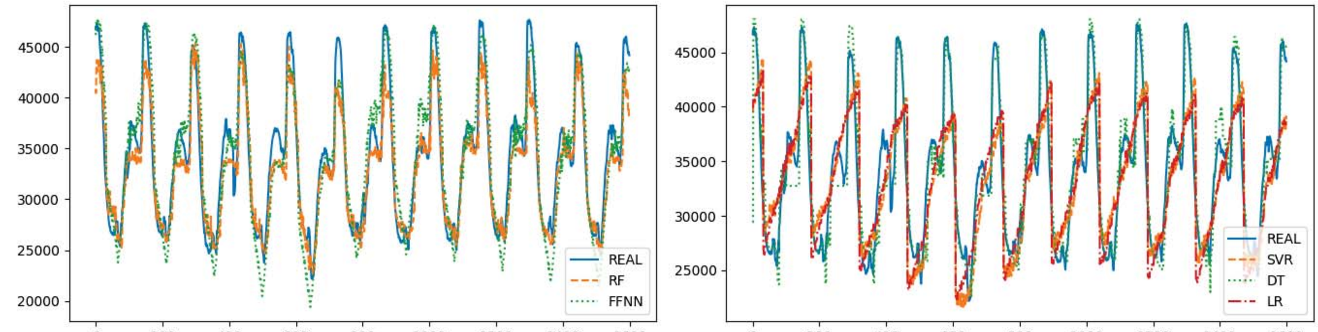

Fig 5. Actual vs predicted electricity power consumption of comparative models for the Quads distribution of every 10 minutes

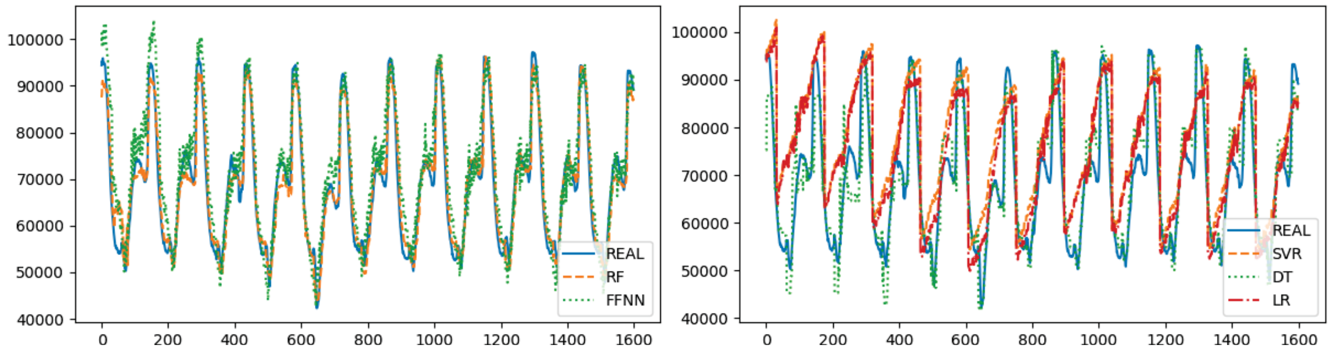

Fig 6. Actual vs predicted forecast of comparative models for the aggregation power consumption every 10 minutes

Power utilities need to have the prediction of different time periods such as hours, days, weeks, months and sometimes years for decision making and plans. In this work, we used the prediction of 10 minutes and one hour periods and it can be applied to different time periods. For the hourly prediction, we reduced the number of parameters that became ineffective such as minutes. As the value of power consumption and the parameters were changed, we needed to optimize comparative algorithms again by the same optimizer method (grid search) for the same sets of parameters. Table III shows the optimized parameters of the comparative models and it is obviously different from distribution to another. Table IV shows the result of the models in each distribution and the aggregation of the three distribution. The results present that also the random forest still performs the best achievement.

TABLE III OPTIMIZINING COMPARITIVE MODEL PARAMETERS FOR EVERY HOUR POWER CONSUMPTION BY USING GRID SEARCH

Model	Quads Distribution Parameter	Smir Distribution Parameter	Boussafou Distribution Parameter	Aggregated Distribution Parameter
RF	Num of features = 3, min samples split = 2, Num of Trees = 50, max depth of the tree = None, min sample leaf = 1	Num of features = 7, min samples split = 3, Num of Trees = 10, max depth of the tree = None, min sample leaf = 1	Num of features = 7, min samples split = 3, Num of Trees = 10, max depth of the tree = None, min sample leaf = 10	Num of features = 5, min samples split = 2, Num of Trees = 100, max depth of the tree = None, min sample leaf = 1
DT	Num of features = 5, min samples split = 2, max depth of the tree = None, min sample leaf = 10	Num of features = 7, min samples split = 3, max depth of the tree = None, min sample leaf = 10	Num of features = 9, min samples split = 2, max depth of the tree = None, min sample leaf = 10	Num of features = 9, min samples split = 3, max depth of the tree = None, min sample leaf = 3
SVR	C= 10, gamma= 0.01	'C': 1, 'gamma': 0.01	'C': 1000, 'gamma': 0.01	C= 1, gamma= 0.01
FFNN	Activation = ReLU, optimizer = SGD, batch size = 100, layers=one, neurons=25, number of epochs = 100, learning rate = 0.001	Activation = SELU, optimizer = Adam, batch size = 350, layers=one, neurons=4, number of epochs = 100, learning rate = 0.001	Activation = SELU, optimizer = Adam, batch size = 250, layers=one, neurons=8, number of epochs = 100, learning rate = 0.001	Activation = SELU, optimizer = Adam, batch size = 250, layers=one, neurons=4, number of epochs = 100, learning rate = 0.001

TABLE IV. RSME AND MAE COMPARISON OF ALGORITHMS IN 3 DISTRIBUTION NETWORKS FOR THE ONE HOURLY POWER CONSUMPTION

Model	Quads Distribution				Smir Distribution				Boussafou Distribution				Aggregated Distribution			
	RSME		MAE		RSME		MAE		RSME		MAE		RSME		MAE	
	Train	***Test***	***Train***	*Test*	***Train***	*Test*	***Train***	***Test***	***Train***	***Test***	***Train***	*Test*	***Train***	***Test***	***Train***	*Test*
RF	3185.8	**21109.7**	2286.7	15442.0	3602.2	**14700.9**	2342.3	11955.2	5669.8	**19504.1**	4079.2	15777.6	4960.9	**28769.3**	3493.7	24033.3
DT	8218.56	26706.6	5879.6	23216.4	7364.0	16301.5	4716.7	13392.9	5766.6	20272.6	4091.4	16332.3	7487.5	38016.7	4724.6	30354.9
SVR	24954.1	26746.2	19707.4	21291.8	24094.4	29986.9	18886.2	24206.8	19320.6	23827.4	15435.9	18262.8	62601.5	56235.9	49188.3	44758.6
FFNN	19166.6	21127.3	14511.1	15622.3	20115.4	19845.6	15182.1	15235.3	20679.2	21873.3	17081.5	17161.9	46393.3	49175.1	36238.2	38693.8
LR	25961.7	23455.2	20798.6	18643.9	24018.5	29528.9	18985.4	24096.9	18528.9	34486.8	14767.1	27776.1	62840.1	59939.8	49712.8	47962.1

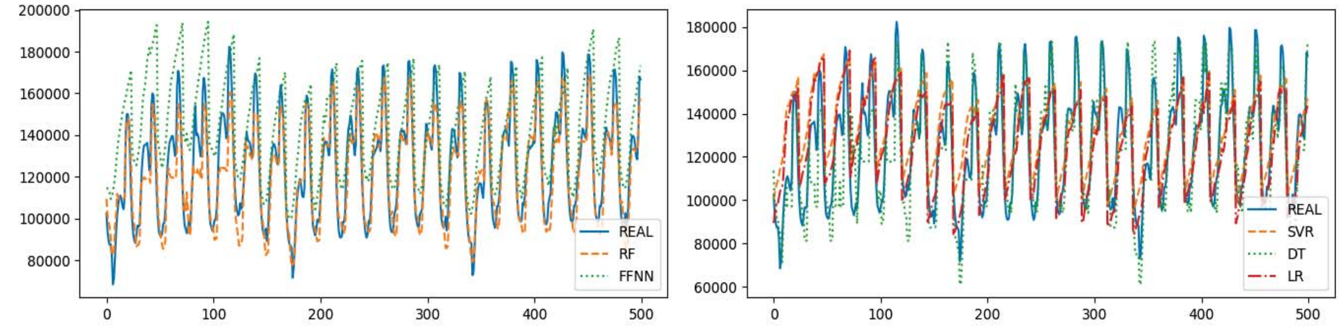

Fig 7. Actual vs predicted electricity power consumption of comparative models for the Boussafou distribution network every 10 minutes

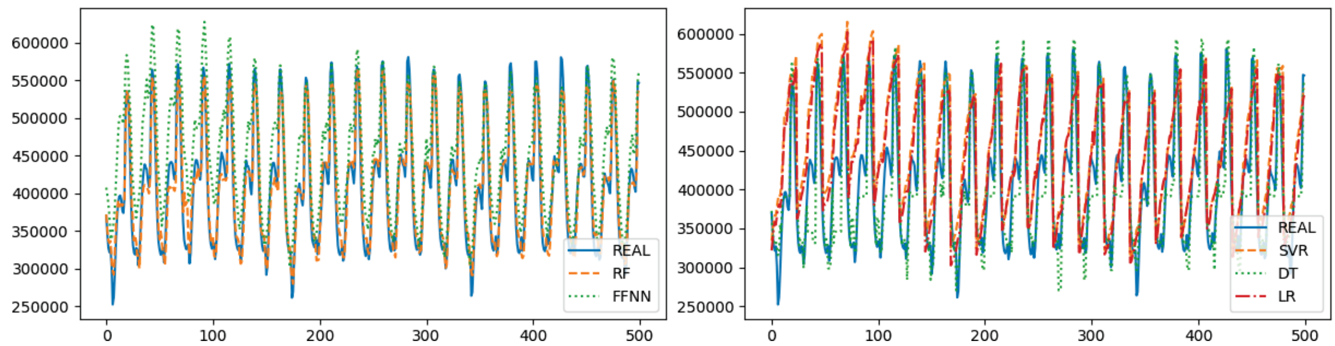

Fig 8. Actual vs predicted electricity power consumption of comparative models for the hourly aggregation of the three distributions

VI. CONCLUSION

Accurate prediction of power consumption represents a necessary part of electricity management for sustainable, productive and effective systems. In this paper, linear regression, decision tree, random forest, feedforward neural network and supper vector machine for regression were used to predict the power consumption of three distribution networks in Tetouan city. The results of those algorithms were compared to determine which one gives the best performance in term of energy forecasting. The dataset that we utilized in this work is exclusive and have not been used before and is used to predict the power consumption of 10 minutes, and one-hour periods. Calendar and weather predictive variables were included. It was shown that hour and temperature were the most predictive prominent variables. We optimized the comparative models by grid search to figure out the best parameters of the models. The results indicate that the random forest model outperformed other models for the prediction of electricity power consumption of Tetouan city. As a perspective of this work, we hope to apply the same study to different Morocco's power supplier and distribution companies including that one's of renewable energy. Likewise, we plan to give a financial study and measure the economic impact.

ACKNOWLEDGMENT

We would like to thank the Amendis company for supplying us the data of power consumption, also the physics department of the Faculty of Science of Tetouan for weather data. Likewise, thanks to Google Colaboratory, Keras. Scikit-learn and TensorFlow developments teams for their facilities.

REFERENCES

[1] P. Tüfekci, "Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods," *Int. J. Electr. Power Energy Syst.*, vol. 60, pp. 126–140, Sep. 2014.

[2] E. Xypolytou, M. Meisel, and T. Sauter, "Short-term electricity consumption forecast with artificial neural networks — A case study of office buildings," in *2017 IEEE Manchester PowerTech*, 2017, pp. 1–6.

[3] M. W. Ahmad, M. Mourshed, and Y. Rezgui, "Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption," *Energy Build.*, vol. 147, pp. 77–89, Jul. 2017.

[4] M. E. Günay, "Forecasting annual gross electricity demand by artificial neural networks using predicted values of socio-economic indicators and climatic conditions: Case of Turkey," *Energy Policy*, vol. 90, pp. 92–101, Mar. 2016.

[5] M. Kankal and E. Uzlu, "Neural network approach with teaching–learning-based optimization for modeling and forecasting long-term electric energy demand in Turkey," Neural Comput. Appl., vol. 28,no. S1, pp. 737–747, Dec. 2017.

[6] K. P. Amber, R. Ahmad, M. W. Aslam, A. Kousar, M. Usman, andM.S. Khan, "Intelligent techniques for forecasting electricityconsumption of buildings," Energy, vol. 157, pp. 886–893, Aug. 2018.

[7] F. Kaytez, M. C. Taplamacioglu, E. Cam, and F. Hardalac,"Forecasting electricity consumption: A comparison of regressionanalysis, neural networks and least squares support vector machines,"Int. J. Electr. Power Energy Syst., vol. 67, pp. 431–438, May 2015.

[8] H. Pombeiro, R. Santos, P. Carreira, C. Silva, and J. M. C. Sousa,

"Comparative assessment of low-complexity models to predict electricity consumption in an institutional building: Linear regression vs. fuzzy modeling vs. neural networks," *Energy Build.*, vol. 146, pp. 141–151, Jul. 2017.

[9] S. Paudel *et al.*, "A relevant data selection method for energy consumption prediction of low energy building based on support vector machine," *Energy Build.*, vol. 138, pp. 240–256, Mar. 2017.

[10] H. Khosravani *et al.*, "A Comparison of Energy Consumption Prediction Models Based on Neural Networks of a Bioclimatic Building," *Energies*, vol. 9, no. 1, p. 57, Jan. 2016.

[11] N. Fumo and M. A. Rafe Biswas, "Regression analysis for prediction of residential energy consumption," *Renew. Sustain. Energy Rev.*, vol. 47, pp. 332–343, Jul. 2015.

[12] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, Mar. 1986.

[13] Z. Yu, F. Haghighat, B. C. M. Fung, and H. Yoshino, "A decision tree method for building energy demand modeling," *Energy Build.*, vol. 42, no. 10, pp. 1637–1646, Oct. 2010.

[14] S. L. Salzberg, "C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993," *Mach. Learn.*, vol. 16, no. 3, pp. 235–240, Sep. 1994.

[15] L. Breiman, *Classification and Regression Trees*, 1st Editio. Routledge, 2017.

[16] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[17] M. A. Razi and K. Athappilly, "A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models," *Expert Syst. Appl.*, vol. 29, no. 1, pp. 65–74, Jul. 2005.

[18] R. Jiang, W. Tang, X. Wu, and W. Fu, "A random forest approach to the detection of epistatic interactions in case-control studies," *BMC Bioinformatics*, vol. 10, no. Suppl 1, p. S65, Jan. 2009.

[19] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[20] M. J. Kane, N. Price, M. Scotch, and P. Rabinowitz, "Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks," *BMC Bioinformatics*, vol. 15, no. 1, p. 276, Aug. 2014.

[21] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random Forests for land cover classification," *Pattern Recognit. Lett.*, vol. 27, no. 4, pp. 294–300, Mar. 2006.

[22] H. Drucker, H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support Vector Regression Machines," *Adv. NEURAL Inf. Process. Syst. 9*, vol. 9, pp. 155--161, 1997.

[23] R. E. Edwards, J. New, and L. E. Parker, "Predicting future hourly residential electrical consumption: A machine learning case study," *Energy Build.*, vol. 49, pp. 591–603, Jun. 2012.

[24] V. Rodriguez-Galiano, M. Sanchez-Castillo, M. Chica-Olmo, and M. Chica-Rivas, "Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines," *Ore Geol. Rev.*, vol. 71, pp. 804–818, Dec. 2015.

[25] A. A. Minai and R. D. Williams, "On the derivatives of the sigmoid," *Neural Networks*, vol. 6, no. 6, pp. 845–853, Jan. 1993.

[26] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks." pp. 249–256, 31-Mar-2010.

[27] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical Evaluation of Rectified Activations in Convolutional Network," May 2015.

[28] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)," Nov. 2015.

[29] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-Normalizing Neural Networks," Jun. 2017.

[30] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for Activation Functions," Oct. 2017.

[31] R. Genuer, J.-M. Poggi, and C. Tuleau-Malot, "Variable selection using random forests," *Pattern Recognit. Lett.*, vol. 31, no. 14, pp. 2225–2236, Oct. 2010.

[32] M. Ataei and M. Osanloo, "Using a Combination of Genetic Algorithm and the Grid Search Method to Determine Optimum Cutoff Grades of Multiple Metal Deposits," *Int. J. Surf. Mining, Reclam. Environ.*, vol. 18, no. 1, pp. 60–78, Jan. 2004.

[33] C.-J. Lin, "A Practical Guide to Support Vector Classification Motivation and Outline," 2003.

[34] F. Chollet and others, "Keras." 2015.

[35] R. Zuo and E. J. M. Carranza, "Support vector machine: A tool for mapping mineral prospectivity," *Comput. Geosci.*, vol. 37, no. 12, pp. 1967–1975, Dec. 2011.

[36] K. G. Sheela and S. N. Deepa, "Review on Methods to Fix Number of Hidden Neurons in Neural Networks," *Math. Probl. Eng.*, vol. 2013, pp. 1–11, Jun. 2013.

[37] S. Karsoliya, "Approximating Number of Hidden layer neurons in Multiple Hidden Layer BPNN Architecture," *Int. J. Eng. Trends Technol.*, 2012.

[38] T. Zhang and Tong, "Solving large scale linear prediction problems using stochastic gradient descent algorithms," in *Twenty-first international conference on Machine learning - ICML '04*, 2004, p. 116.

[39] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *3rd Int. Conf. Learn. Represent. San Diego, 2015*, Dec. 2014.