# Dynamic prediction of energy and power usage cost using linear regression-machine learning analysis

To cite this article: Rani Fathima Kamal Basha *et al* 2021 *J. Phys.: Conf. Ser.* **1921** 012067

View the article online for updates and enhancements.

# Dynamic prediction of energy and power usage cost using linear regression-machine learning analysis

**Rani Fathima Kamal Basha[1], Bharathi.M.L[2] and Kanagaraj Venusamy[3]**

[1,3]Engineering Department, University of Technology and Applied Sciences-Al Mussanah, Al Muladdha, Sulatante of Oman.
[2]Department of EEE, Sathyabama Institute of Science and Technology, Chennai, India.

*E-mail:* rani@act.edu.om

**Abstract:** As a known fact, energy usage and demand exponentially rises year after year, hence forth power based companies are apparently looking out for a forecasting approach with better approximations. Based on the usage history at the customer level with the emergence of machine learning, and its association with various prediction and decision making fields. This paper aims to use a machine learning algorithm to predict the cost levied on the customer proportional to the usage. The efficacy of this model is compared to the results obtained with the mathematical computations. It is evident that the accuracy is 95% with reference to the multilinear regression algorithm

## 1. INTRODUCTION

Energy optimization is one of top areas of research interest.as a matter of fact the demand and productivity mismatch is increasing linearly, though various economic strategies are followed in residential complexes and commercial structures to optimize power which is obtained from fossil fuels, renewable and nuclear resources. If a Proper algorithm to anticipate the cost is to be deduced, measurement and estimation is the foremost task based on the International guidelines. Predicting the utilization of power and the cost related to the usage [1]-[3] involves tedious calculation and numerical computations. The Proposed method is a combination of machine learning techniques, probability and statistical concepts. The dynamic fast changing power consumption pattern can be accommodated using a machine learning linear regression and SVM model.

This emerging concepts can be incorporated in smart grid and smart city Planning. Flexibility is a most valuable asset to particularly in power distribution and electricity retail segments.by using an adapting algorithm based on the demand and the production. The cost incurred and the fee levied on the customers can be changed. This can be associated with the [4]-[6]. The uber transport where the price is dependent on supply and demand. The objective of the user is to save money and time if the requirements are not narrowed down. Neural Network reinforced bearing is adopted [7]-[8]. The research on forecasting the usage is conducted meticulously based on the market characteristics and on par with the database. The tariff algorithm is generated and the same is been downloaded in the smart meters in residence. These meters are in turn connected to the cloud server and cost is generated on day basis which can be visible by the user and the utility can be varied accordingly.

## 2. MACHINE LEARNING MODELS

Machine learning is a category of artificial intelligence that facilitates nearer prediction of the output or

feature, without making use of traditional computational strategy. The objective is to train the machine with a set of predefined codes which enables the machine to take decisions based on the previous learning. Apparently a dynamic programming algorithm can replace a human decision on a particular task. In few situations this scenario has proved to be much better than human programmers in sensitive issues [9]-[10].Machine learning is a theory where you load immense data to the computer and instruct the machine to correlate the data in a particular pattern. These patterns are famously known as machine learning models or algorithms that can range from a simple linear equation to multi constraint complex equation, which is nearly impose the human calculation. The machine or the computer will utilize the model to learn the patterns in the features and output.

### 2.1 Supervised Learning
In this category the labelled data [11] is used to train the model based on the previous experience. This can be related to the very relevant prediction to know whether a person may be affected to covid 19 based on the temperature of the person symptoms, suffered, duration or occurrence and the previous medical history of the persons; the system learns the pattern and suggests the predictions. The illustration (Figure.1) is given in the below.
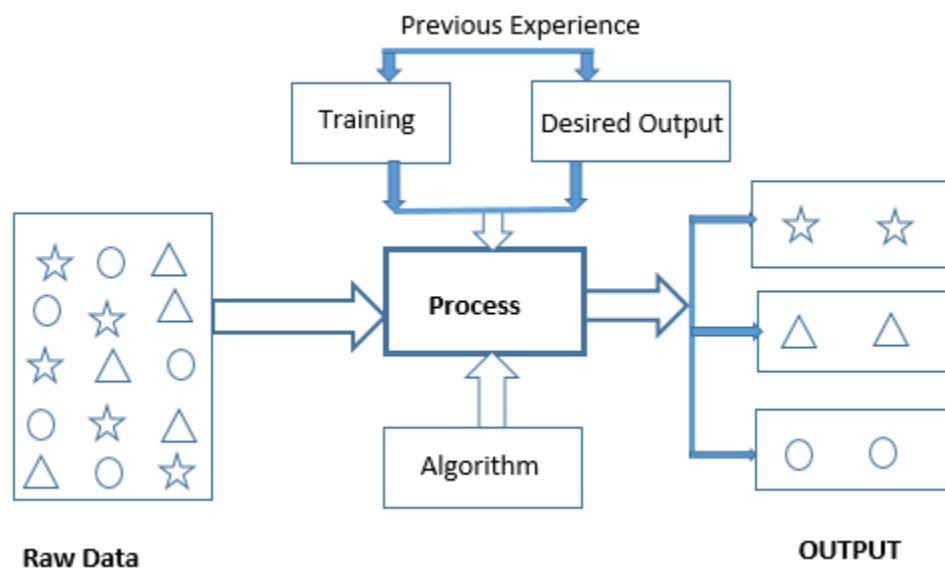


**Fig.1** Illustration of Supervised Learning

### 2.2 Unsupervised learning
The utility of unsupervised learning (Figure.2) is not as wide spread or prominent as supervised learning. Indeed, its used for a very narrowed application. In this category the machine is not assisted with concrete finite set of data and the outcomes cannot be predicted [12].
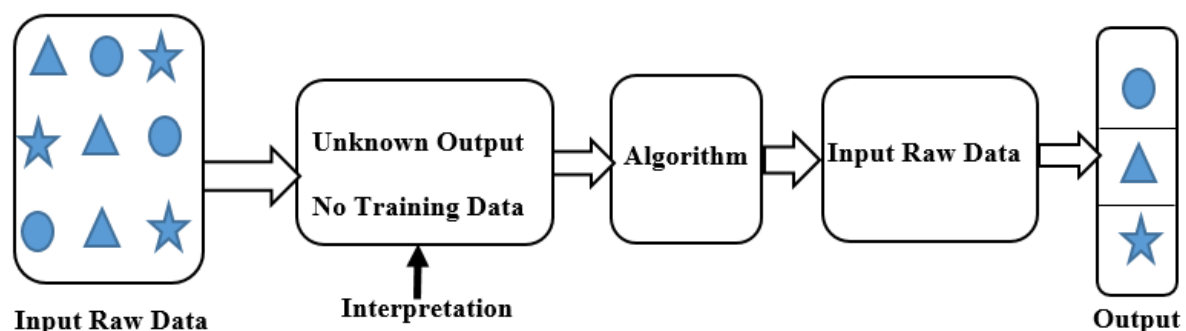


**Fig.2** Illustration of Unsupervised Learning

*2.3 Reinforcement learning*

Reinforcement learning (RL) is a sort of Machine Learning (Figure.3) which is completely about taking appropriate action to exploit reward in a specific state. It is applied by many software and machines to identify the finest probable behaviour or route it should take in a particular state. RL means to establish or encourage a pattern of behaviour [13].
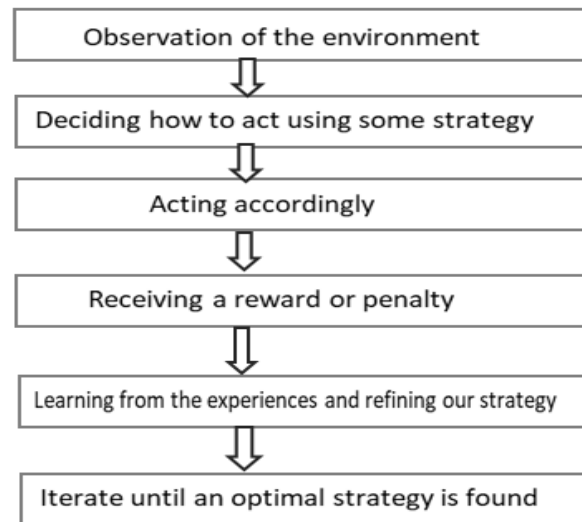


**Fig.3** Illustration of Reinforcement Learning

## 3. THEORY OF LINEAR REGRESSION

In statistics, one of the reliable and basic approach to build a relationship between two or more scalar responses (dependent and independent variable), They are basically classified into two categories namely simple linear regression and multiple linear regression [14-[15]. These are selected based on the number of features of the respective problem. It is a linear analysis hence not suitable for non-linear problems. This regression pattern is used extensively for various practical approaches. The versatility of the algorithm is based on a simple assumption that all the constraints involved in the problem are linear in nature. This could easily fit in a solution than a nonlinear problem. The set of equations [1],[2] and [3] Represent the steps involved to evaluate the coefficient of determination (R).The value of R gives the accuracy measure of the solution trained by the neurons to the actual mathematical calculation.

Mean Square error =Actual - Predicted Value

$$Y=b0+b1+\in \quad\quad\quad (1)$$

Where   b0=Y interrupt
        b1=Slope
        $\in$=Error
        X =Independent Variable
        Y=Dependent Variable
        Objective: To identify the best fit line
        Accuracy-R square. (Co efficient of determination)

$$R^2 = \frac{\sum(y_p - \bar{y})^2}{\sum(y - \bar{y})^2} \quad\quad\quad (2)$$

Where   $y_p$=Predicted Value
        $\bar{y}$=Mean Value
        y=Actual Value
        R= 0 to 1 (normal range)

Gradient decent

$$m = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})^2} \qquad (3)$$

## 4. THEORETICAL CALCULATION OF POWER AND COST INCURRED

The day to day necessities such as heating, cooling, illuminations and much more required energy consumption which is customizable and reflects on the cost incurred The new revolutionized technology makes the appliance usage time based on service provider optimization and algorithm. For instance, high usage cost during peak time and minimal cost during the off peak hours. Firstly, the wattage of the device used and the time of usage is observed and the same is being calculated for 30 days roughly. The cumulative cost is figured out depending on the service provider's energy plan, the price is calculated as shown in Table 1, 2 and 3.

**Table .1** Residential load power calculation

| Name of the Appliances | Power rating | No of Appliances | Total power | Time duration | Total energy watt-hr | kWh / day | 30 days |
|---|---|---|---|---|---|---|---|
| Tube light | 36 | 5 | 180 | 8 | 1440 | | |
| Fan | 70 | 5 | 350 | 12 | 4200 | | |
| Fridge | 200 | 1 | 200 | 24 | 4800 | | |
| Air conditioner | 1000 | 1 | 1000 | 5 | 5000 | | |
| Laptop | 50 | 1 | 50 | 2 | 100 | | |
| Television | 70 | 1 | 70 | 2 | 140 | 19.08 | 572.4 |
| Geyser | 1000 | 1 | 1000 | 1 | 1000 | | |
| Mixer | 400 | 1 | 400 | 1 | 400 | | |
| Washing Machine | 500 | 1 | 500 | 1 | 500 | | |
| Microwave oven | 1500 | 1 | 1500 | 1 | 1500 | | |
| Total units | | | | | 19080 | | |

**Table .2** Charges levied by the service provider

| LT1 Residential(1Phase) | Unit | 0-100 | 101-300 | 301-500 | 501-1000 | >1000 |
|---|---|---|---|---|---|---|
| Fix charges (Indian Rupee. Rs. 50) | Amount(Rs) | 3 | 6.73 | 9.75 | 10.50 | 11.50 |
| | FAC | 0.37 | 0.62 | 0.80 | 0.84 | 0.89 |

**Table.3** Calculation of cost

| Units | 384.3 | 254 | 300 | 200 | 100 | 150 | 350 | 250 | 400 | 500 |
|---|---|---|---|---|---|---|---|---|---|---|
| Price (Indian Rupee) | 2465 | 1336.4 | 1646 | 1273 | 300 | 636.5 | 2133.5 | 1309.5 | 2651 | 3596 |

## 5. RESULTS AND DISCUSSION

Based on the performance evaluation metrics it is parametrized evidently by results that the error between the actual estimation and the computer learned estimation is minimized. The regression technique is aimed to generate a feasible plane, whose equation gives the solution more accurately. In linear regression method the final coefficient of determination in Figure.9 substantiate that the error is minimized and the most apt solution is obtained. Figure.4 to Figure.6 represents the python code for

linear regression technique. The look up table of the power usage data in "kWh" for 30 days and the relevant cost is imported as a csv file. It can be located from the code, the training of the model is done with 80% of the data from the imported data and the testing is done using the balance 20%.The graph in Figure.8. Shows a straight line which passes through maximum possible nodes and the equation representing the line, proves to be a viable solution. In Figure.7. The regression graph for the testing the data is shown.Figure.9 shows the regression (R value) which is 98.6%. This conclude the accuracy in prediction to be very high.

```
[ ]  import pandas as pd
     import matplotlib.pyplot as plt

 ●   data = pd.read_csv("power_data.csv")   ## import the data and convert into dataframe
     data.head()

         no of units    cost

     0          384.3  2465.0

     1          254.0  1336.4

     2          300.0  1646.0

     3          200.0  1273.0

     4          100.0   300.0


[ ]  data.isnull().sum()

     no of units    0
     cost           0
     dtype: int64
```

**Fig.4**. Python code of power data

```
[ ]  #x = data[['no of units',"abc","def"]].values
     x = data['no of units'].values
     y = data['cost'].values

 ●   x.shape

     (10,)

[ ]  y.shape

     (10,)

[ ]  ### shape of the array (no.of samples,1)
     x = x.reshape(len(x),1)
     y = y.reshape(len(y),1)

[ ]  x.shape

     (10, 1)

[ ]  y.shape

     (10, 1)
```

**Fig.5** Python code for samples

```
[ ]  len(xtrain)

     8

[ ]  len(xtest)

     2

[ ]  xtest

     array([[100.],
            [350.]])

 ▶   #### Build the model
     from sklearn.linear_model import LinearRegression
     model = LinearRegression()

[ ]  #### Train the model
     model.fit(xtrain,ytrain)

     LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)

[ ]  #### Predictions
```
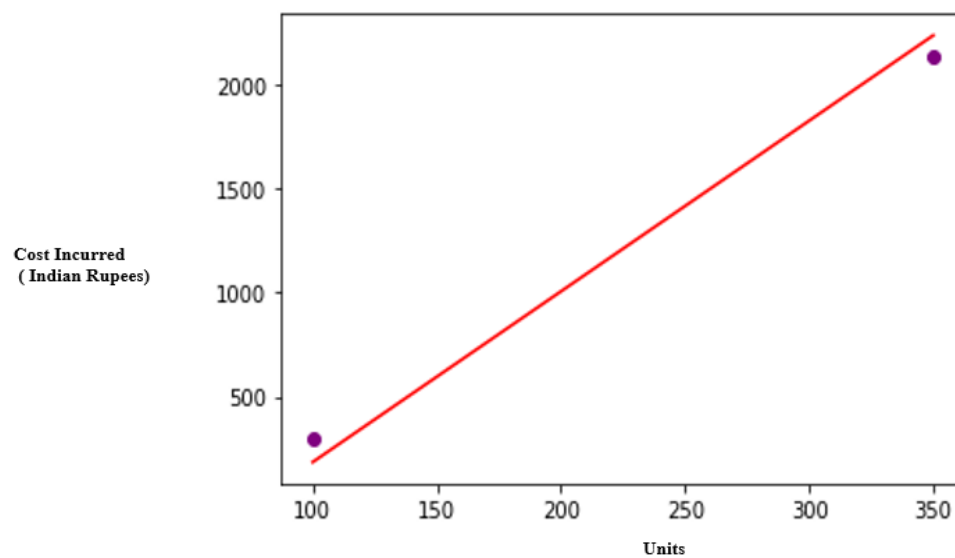
**Fig.6** Python code on linear regression



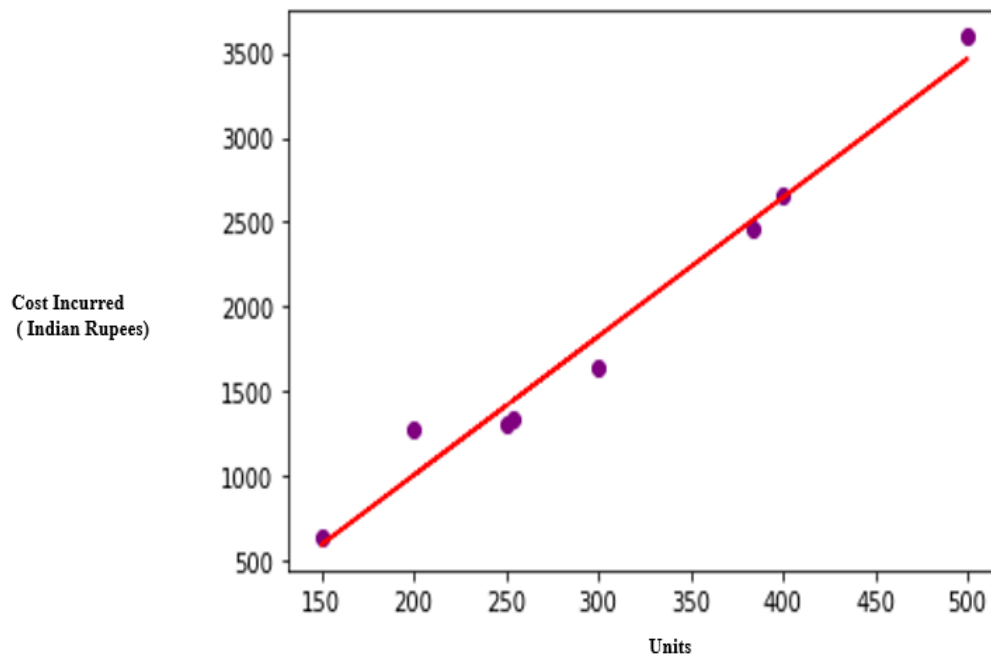**Fig .7** Line of regression of testing data

**Fig.8** line of regression of training data



**Fig .9** Accuracy metrics

## 6. CONCLUSION

Machine learning and its allied studies will be ruling the future in all tenures But the lack of skilled labour is to be considered as it's an emerging field. Dynamic calculation of cost proportional to the highly non-linear load is a challenging task but an appropriate and adaptable algorithm is making it a plug and play affair and claims to be more accurate with the results obtained through mathematical calculations. The coefficient of determination or the Accuracy metric component proves the validity of the theory. The text of your paper should be formatted as follows:

## REFERENCES

[1]    Y. Kwon et al., "Mantis: Efficient Predictions of Execution Time, Energy Usage, Memory Usage and Network Usage on Smart Mobile Devices," in *IEEE Transactions on Mobile Computing*, vol. **14**, no. 10, pp. 2059-2072, 1 Oct. 2015.

[2]    K. Kouzelis, Z. H. Tan, B. Bak-Jensen, J. R. Pillai and E. Ritchie, "Estimation of Residential Heat Pump Consumption for Flexibility Market Applications," in *IEEE Transactions on Smart Grid*, vol. **6**, no. 4, pp. 1852-1864, July 2015..

[3]    https://github.com/tylerhaun

[4]    Z. Cao, L. Liu and O. Markowitch, "Comment on "Highly Efficient Linear Regression Outsourcing to a Cloud"," in *IEEE Transactions on Cloud Computing*, vol. **7,** no. 3, pp. 893-893, 1 July-Sept. 2019.

[5]   B. Sravani and M. M. Bala, "Prediction of Student Performance Using Linear Regression," 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2020, pp. 1-5

[6]   C. M. Wu, P. Patil and S. Gunaseelan, "Comparison of Different Machine Learning Algorithms for Multiple Regression on Black Friday Sales Data," 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 2018

[7]   Barak, S. and S.S. Sadegh, Forecasting energy consumption using ensemble ARIMA–ANFIS hybrid algorithm. International Journal of Electrical Power &Energy Systems, 2016. 82: p. 92-104.

[8]   Abdoli, B., et al., Evaluation of the air-borne ultrasound on fluidized bed drying of shelled corn: Effectiveness, grain quality, and energy consumption. Drying 150 Technology, 2018, 36(14): p. 1749-1766.

[9]   Ascione, F., et al., Artificial neural networks to predict energy performance and retrofit scenarios for any member of a building category: A novel approach. Energy, 2017

[10]  Azar, J., et al., An energy efficient IoT data compression approach for edge machine learning. Future Generation Computer Systems, 2019. 96: p. 168-175.

[11]  Bhinge, R., et al., Toward a Generalized Energy Prediction Model for Machine Tools. Journal of Manufacturing Science and Engineering, Transactions of the ASME, 2017, 139(4).

[12]  Chou, J.S. and D.S. Tran, Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. Energy, 2018: p. 709-726.

[13]  Yiqiang Chen, Qiang Yang, Jie Yin and Xiaoyong Chai, "Power-efficient access-point selection for indoor location estimation," in IEEE Transactions on Knowledge and Data Engineering, vol. 18, no. 7, pp. 877-888, July 2006.

[14]  T. Petrovic, K. Echigo and H. Morikawa, "Detecting Presence From a WiFi Router's Electric Power Consumption by Machine Learning," in IEEE Access, vol. 6, pp. 9679-9689, 2018.

[15]  S. Shan, B. Cao and Z. Wu, "Forecasting the Short-Term Electricity Consumption of Building Using a Novel Ensemble Model," in IEEE Access, vol. 7, pp. 88093-88106, 2019, doi: 10.1109/ACCESS.2019.2925740.