ISSN (Online) 2395-2717

International Journal of Engineering Research in Electrical and Electronic Engineering (IJEREEE)

Vol 9, Issue 5, May 2022

Estimation of Generated Electricity for the Solar Power Plant by Using Polynomial Regression

[1] Ali Riza OZER, [2] Omer Faruk BAY

[1] [2] Electrical and Electronic Eng., Gazi University, Turkey.
Corresponding Author Email: [1] ali.riza.ozerr@gmail.com, [2] omerbay@gazi.edu.tr

Abstract*— The estimation of electrical energy that can be supplied from the renewable energy sources becomes a weighty matter regarding to increasing the share of these sources in the electrical energy sector all over the world. The installed power of power plants in Turkey has reached 100334 MW. Depending on the installed power, progress has been made in the electricity transmission and distribution sector over the years. Thus, there exist 21 distribution regions in electricity distribution in Turkey. Some distribution regions have reached 25 MWs in terms of solar installed capacity. While estimating the electrical energy demand for any distribution region, the total electrical energy that can be supplied by the solar power plants (SPP) in that region should be taken into account. In accordance with the legislation in Turkey, the authorized electricity retail sales company in each distribution region has to enter the daily estimation of the electrical energy, to be obtained from SPPs, located in its own region, on the web site of EPIAS (Energy Markets Operation Joint Stock Company). In this study, the estimation of electricity generation of the 990 kW solar power plant belonging to Osmaniye Special Provincial Administration and located in the town of Cevdetiye for the days 24.08.2021 and 25.08.2021 was carried out with the polynomial regression algorithm. Approximately 85% performance metric was got from the polynomial regression algorithm.

***Index Terms*—Mean absolute percentage error (MAPE), -Score, SPP (Solar power plant) and polynomial regression.**

I. INTRODUCTION

In this study, an electricity generation forecast, for the days of 24.08.2021 and 25.08.2021 was carried out by using the generated electricity data of 990 kW Osmaniye Special Provincial Administration solar power plant for the time period of 2020-2021 (Fig. 1.). The study covers a very short time interval as only 1-year electricity generation data is acquired. At the same time, no private organization has shared the electricity generation data of their own power plants, in accordance with the law on the protection of personal data. The data, about generation of electrical energy, of Osmaniye Special Provincial Administration between August 2020 and August 2021 was be able to obtained from this institution.

Fig.1. The SPP of Osmaniye Special Provincial Administration (Its coordinates are 37.129621, 36.213784)

If dataset is comprehensive as much as 5-years, this estimation could have turned into a root study capable of predicting more than 2 days. The main goal of this study is to predict electricity power generation of 2-days for August 2021. Another objective of the study is to determine whether the investments, planned for future, is feasible for many state institutions such as Osmaniye Special Provincial Administration. For instance, whenever Osmaniye Special Provincial Administration desire to get another solar power plant investment near by the available photovoltaic power system, the return on investment will be specified by the data about electrical energy, generated by the existing 990 kW solar power plant.

The study is comprised of five sections. In the introduction, general description for Osmaniye Special Provincial Administration SPP (Fig. 1.) and objective of the study are presented. In the second stage, information, about renewable energy sources is handled. In the third chapter general information about the SPP which was funded by the Osmaniye Special Provincial Administration, hourly dataset about the generated energy and the data about weather forecasting are submitted. In the fourth part, dataset which is used for the prediction method (Polynomial regression) is given. In the fifth section, prediction is realized by using polynomial regression algorithm. In the chapter of discussion, the findings are evaluated. In the part of results, results of the prediction are presented and general opinions about the study are given.

II. OVERALL ASESSMENT OF RENEWABLE ENERGY SOUCES IN TURKEY

Renewable energy in Turkey can be examined in 5 groups: Hydraulic, solar, biomass, geothermal and wind. The electricity generation shares of these renewable energy sources for 2021 are presented in Fig. I., given in below.

ISSN (Online) 2395-2717

International Journal of Engineering Research in Electrical and Electronic Engineering (IJEREEE)

Vol 9, Issue 5, May 2022

Table 1. The share of renewable energy sources [1]

Hydraulic	17.670%
Wind	9.843%
Solar	0.493%
Geothermal	3.219%
Biomass	1.879%
Conventional Resources	66.896%

In Fig. I. above, the supply of electricity generation in 2021 from conventional sources is exactly twice the supply from renewable sources. In 2021, the share of solar energy is not even 1%. The most important share in renewable electricity generation is hydraulic.

The second largest renewable energy supply after hydraulics is from wind. After the wind, geothermal and biomass come respectively. Today, renewable energy is following an increasing trend in electrical energy supply. As can be seen from Fig. II., insufficient hydroelectric potential has been obtained from hydroelectric power plants in Turkey due to climate change and drought. Therefore, the contribution of renewable energy to electricity generation was 40% in 2020, and this rate declined to 33.103% in 2021 [1].

The share of renewable energy in 2021 decreased by 7% compared to 2020. The biggest factor in this decrease is drought. Because of the drought, hydroelectric power plants supplied less electricity. For this reason, the share of hydroelectric power plants in 2020 was 26%, and this rate was 17% in 2021. From 2020 to 2021, the share of hydroelectric power plants in electricity generation decreased by 9%; however, the share of wind, solar, geothermal and biomass increased by 2% in total. In other words, a net decrease of 7% was recorded in renewable energy in 2021 compared to 2020 (Fig. II.). According to Fig. II. almost no change can be seen for geothermal energy in the years of 2020-2021.

Electrical energy generation in 2021 is shown on the basis of resources in Fig. 2., in below [1].

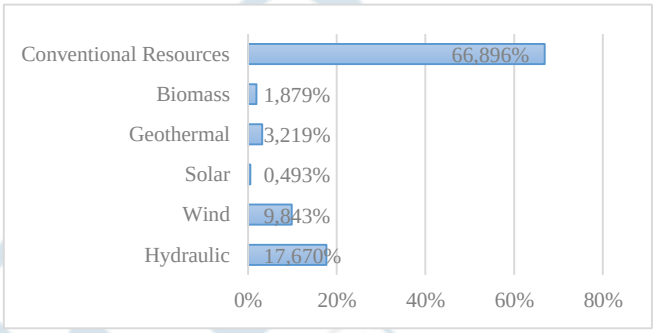

Fig. 2. Electrical energy supply in terms of sources (2021) [1]

Table 2. The percentage difference of electricity supply in terms of renewable energy sources from 2020 to 2021[1]

Sources	*Hydraulic (%)*	*Wind (%)*	*Solar (%)*	*Geothermal (%)*	*Biomass (%)*	*Total (%)*
2020	26,81832381	8,437067123	0,145072832	3,210028466	1,400836272	40,011328 5
2021	17,6695821	9,842640004	0,493485638	3,218841576	1,879307816	33,103857 1
Difference from 2020 to 2021 (%)	-9,14874171	1,40557288	0,348412806	0,00881311	0,478471544	-6,9074714

III. GENERAL DESCRIPTION OF THE SOLAR POWER PLANT, THE DATASET ABOUT ELECTRICAL ENERGY SUPPLY & WEATHER FORECASTING

General Description of the SPP

The solar power plant (SPP), belonging to Osmaniye Special Provincial Administration was established in Cevdetiye town of Osmaniye. Its installed power is 990 kW and it consists of 3640 panels on a plot of 28586,04 m^2 [2]. The electricity generation data of the said SPP are shown in Fig. 3. Fig. 3. (a) and Fig. 3. (b) are both showing the electrical power between 2020-august and 2021-august.

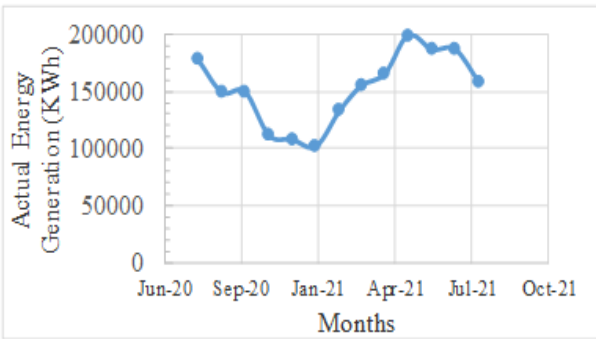

Fig. 3. Electricity generation from August of 2020 to August of 2021

ISSN (Online) 2395-2717

International Journal of Engineering Research in Electrical and Electronic Engineering (IJEREEE)

Vol 9, Issue 5, May 2022

Meteorological Data

Meteorological data were taken from the website named "Solcast". The meteorological dataset acquired from the website is hourly and consists of the following parameters: Global Horizontal Radiation (GHI, W/m^2), direct normal radiation (DNI, W/m^2), direct horizontal radiation (EBH, W)./m^2), diffuse horizontal radiation (DHI, W/m^2), spherical oblique radiation (GTI, W/m^2), cloud opacity (%), Albedo (Whiteness, %), zenith (°)), azimuth angle (direction angle, °), temperature (°C), wind speed (m/s), wind direction (°), relative humidity (%), surface pressure (hPa), precipitation water (PWAT, kg / m^2), snow depth (SWDE, cm), dew point (DWPT, °C). The aforementioned dataset was downloaded in ".csv" format at 60-minute intervals. The parameters of this data given above are processed as inputs in the Python program.

IV. THE POLYNOMIAL REGRESSION IN ESTIMATION

The polynomial regression is used in this estimation study as a method. The estimation study was coded in Python programming language and the program was developed in Anaconda Navigator. In the study, used libraries are that "matplotlib", "pandas", "numpy", "scikit-learn", "seaborn", "datetime". In regression analysis dependent parameters are predicted corresponding to one or more independent variables. Regression analysis is to determine the value of the dependent variable depending on the value of the independent variable [3-4]. Regression analysis using one independent variable is expressed as univariate regression analysis. The curve equation for univariate polynomial regression can be obtained from the following expression (1): [5-6].

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \cdots + \beta_k {x_i}^k + \varepsilon_i$$
$$i = 1,2,3, \dots, N \qquad (1)$$

In this expression each parameters are expressed in below [6].

y_i: Dependent variable i. observed value,
x_i: The independent variable i. observed value,
N: Number of units,
β_0: Regression constant,
$\beta_1, \beta_2, \beta_3, \dots, \beta_k$: Regression coefficients,
k: Degree of independent variable,
ε_i: Error term of the model

Polynomial regression is not an additive method. The additive methods are curvilinear, exponential, multiplicative and etc. Regression is basis on the assumption that the independent variable parameter has an effect on the dependent variable parameter [6-7]. At this point, discussion of quadratic and cubic polynomial regression models is convenient. These models are discussed one by one as follows:

Quadratic Regression

The quadratic regression is a model that deals with the cause-effect relationship between two variables [6,8]. The predicted values of the variables β_0, β_1 and β2 taken from the sample, considered, are b_0, b_1 and b_2, and when the error term ε, recognized as the estimation size, is acquired from the sample "e". While term of n indicates the number of sample units. The curve equation of the regression model depending single parameter can be shown in the following expression below (2) [6,9,10]:

$$y_i = b_0 + b_1 x_i + b_2 x_i^2 + e_i \mid i = 1,2,3, \dots, N \qquad (2)$$

Graphical form of the quadratic regression model is shown in Fig. 4.

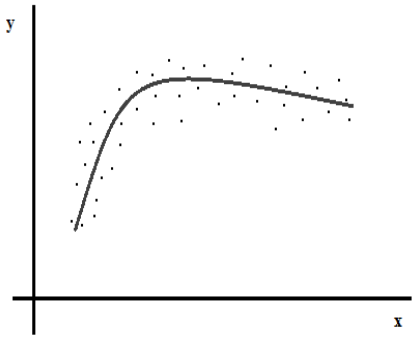

Fig. 4. Quadratic regression model [11]

Cubic Regression

Cubic Regression Model: The third-order regression model based on the cause-effect relationship between two variables is called the cubic regression model [6, 8]. The curve equation of the univariate cubic regression model can be determined from the expression (3) given below [6, 9, 10]:

$$y_i = b_0 + b_1 x_i + b_2 x_i^2 + b_3 x_i^3 + e_i$$
$$i = 1,2,3, \dots, N \qquad (3)$$

Graphical form of the quadratic regression model is shown in Fig. 5.

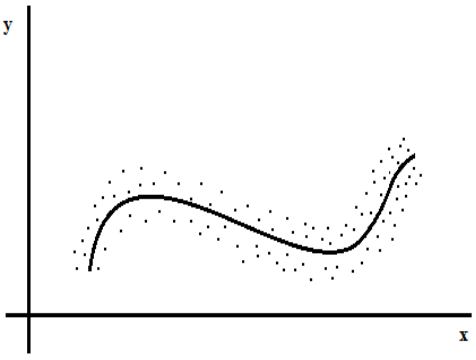

Fig. 5. Cubic regression model [11]

MAPE (Mean Absolute Percentage Error) is a generally used to compute accuracy of predictions. MAPE is computed by the following formula [12]:

ISSN (Online) 2395-2717

International Journal of Engineering Research in Electrical and Electronic Engineering (IJEREEE)

Vol 9, Issue 5, May 2022

$$MAPE = \frac{1}{n} * \sum_{i=1}^{n} \left| \frac{A_i - F_i}{A_i} \right| \quad (4)$$

A_i: *Real Data*
F_i: *Predicted Data*
n: *Number of Data*

According to the quotation from Lewis's Demand Forecasting and Inventory Control book; Models with MAPE value below 10% are accepted as "very good", models between 10% and 20% recognized as "good", and models between 20% and 50% recognized as "acceptable" [13].

R-squared (R^2) score is to compute the performance of any regression models [14-15]. After training the regression model, the excellence of fitting of any trained models can be verified by using R^2 - Score. R^2 - Score specifies the scatterness of data points, that surround the regression line. However this performance metric can be defined to as the coefficient of determination. The R^2- Score gets any value between 0% - 100% any time. 0% score represents that the dependent variable have no variability around the estimation values, 100% rate means that the dependent variable has all the variability around the values estimated by present regression model. The high R^2- Score implies the success of the trained model, however low R^2- Score represents the imperfection of the model [16].

V. ESTIMATION PROGRAM DEVELOPMENT

How linearly 'KWh', 'AirTemp', 'CloudOpacity', 'Dhi', 'Dni', 'Ebh', 'Ghi', 'GtiTracking', 'PrecipitableWater', 'SnowWater', 'SurfacePressure', '10Wind' , 'WindSpeed10m' and clock data ('S_0', 'S_1', 'S_2', 'S_3', 'S_4', 'S_5', 'S_6', 'S_7', 'S_8', 'S_9', 'S_10' ', 'S_11', 'S_12', 'S_13', 'S_14', 'S_15', 'S_16', 'S_17', 'S_18', 'S_19', 'S_20', 'S_21', 'S_22', 'S_23') parameters changing 1.0 value (in the light-colored part) on the given scale in Fig. 11. show a linear change with the other input parameters. The data on the scale with a color between 0.4 and 0 (in the black-colored part) show non-linear changing with the other parameters. Because the dataset is restricted, all of these parameters are inputted to the model to get it more complex and prevent its overfitting.

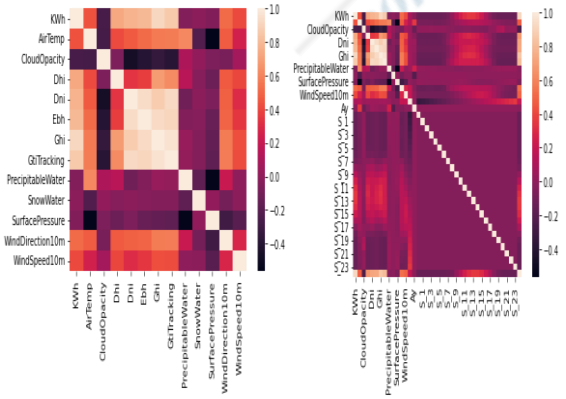

Fig. 6. Feature selection maps for the model

Hourly data is encoded as 'S_0', 'S_1', 'S_2', 'S_3', 'S_4', 'S_5', 'S_6',

'S_7', 'S_8', 'S_9', 'S_10', 'S_11', 'S_12', 'S_13', 'S_14', 'S_15', 'S_16', 'S_17', 'S_18', 'S_19', 'S_20', 'S_21', 'S_22', 'S_23 '. This is expressed as "Data Encoding". Data encoding is to reflect hourly data on the columns (As being input to the model) as 'Yes or No'. "Yes" means logical-1, "No" means logical-0. By this way, hourly data took place in columns is located in rows as "1" or "0". Therefore hourly data becomes an input variable. These hourly data is numbered from 0 to 23 as constituting 24 hours and taken place in the columns of the model. Thus hourly data is converted to digital form such that the current time is '1' for each 1-hour time interval and all the remaining 23 hours are '0'. In fact, this data is transformed to binary data matrix and processed as an input parameter of the model.

In Fig. 7., it is verified that whether the parameters of 'CloudOpacity', 'Dhi', 'Ghi' changes linearly with the 'KWh'. Ghi and Dhi have the most linear relationship with the parameter of 'KWh', is observed. However, as stated in the previous paragraph, all parameters here are used in the training of the model.

All parameters given in the first paragraph are inputted to the polynomial regression algorithm. If the data set was large enough, some of these features could be subtracted. However, seeing the probability of overfitting or underfitting of the model; all of given inputs are used to add complexity to the model.

In this study from Polynomial Regression, the graphics in Fig. 7. – Fig. 12. are obtained. The numeral results are given in Fig. III. MAPE (Mean Absolute Percentage Error) values are shown in Fig. IV. As can be seen from Fig. IV. MAPE values are found as 14.58%, 16.78% and 18.23% respectively from 2^{nd} to 4^{th} degree of the Polynomial Regression.

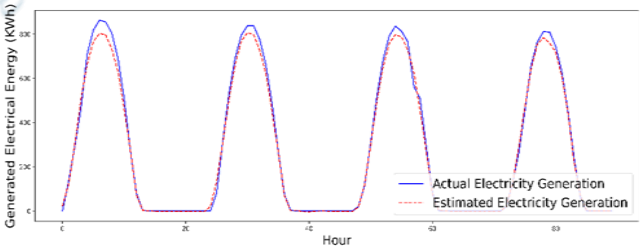

Fig. 7. Fitting of 2^{nd} degree polynomial regression model (R^2- Score =96.87%)

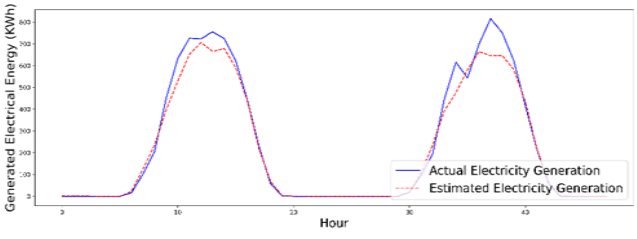

Fig. 8. Estimated electricity generation versus actual electricity generation via 2^{nd} degree polynomial regression model (MAPE = 14.58%)

ISSN (Online) 2395-2717

International Journal of Engineering Research in Electrical and Electronic Engineering (IJEREEE)

Vol 9, Issue 5, May 2022

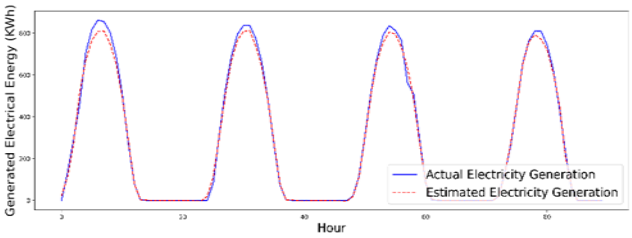

Fig. 9. Fitting of 3rd degree polynomial regression model (R^2- Score =97.22%)

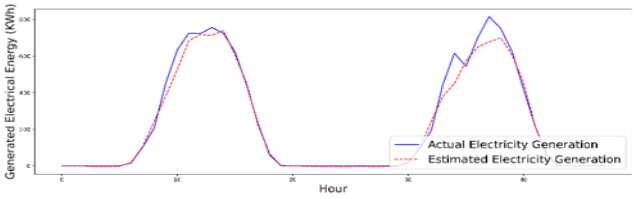

Fig. 10. Estimated electricity generation versus actual electricity generation via 3rd degree polynomial regression model (MAPE = 16.78%)

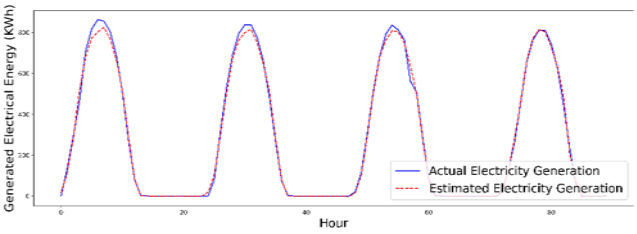

Fig. 11. Fitting of 4th degree polynomial regression model (R^2- Score =97.63%)

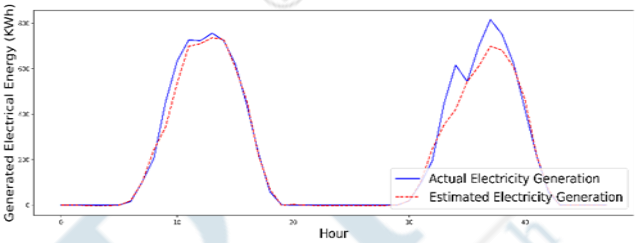

Fig. 12. Estimated electricity generation versus actual electricity generation via 4th degree polynomial regression model (MAPE = 18.23%)

Table. 3. Prediction and actual values of electrical energy generation from polynomial regression

Date	*Prediction from 4th Order Polynomial Regression*	*Prediction from 3rd Order Polynomial Regression*	*Prediction from 2nd Order Polynomial Regression*	*Actual Electrical Energy Supply*
2021-08-24 06:00:00	19.31742	18.72353	23.43884973	15.75
2021-08-24 07:00:00	101.9878	105.5136	125.5189617	102.375
2021-08-24 08:00:00	243.3769	243.537	235.9862681	206.325
2021-08-24 09:00:00	342.6648	381.16	397.2466138	452.025
2021-08-24 10:00:00	535.7927	527.2697	529.4794396	633.15
2021-08-24 11:00:00	698.4738	684.1009	651.7983449	726.075
2021-08-24 12:00:00	709.5371	717.3397	706.1987597	722.925
2021-08-24 13:00:00	735.6228	712.9393	665.1173939	756
2021-08-24 14:00:00	729.5866	739.1406	678.8793214	724.5
2021-08-24 15:00:00	608.587	605.2016	589.8321152	622.125
2021-08-24 16:00:00	459.3358	452.1904	439.2489714	442.575
2021-08-24 17:00:00	221.6508	222.0656	219.3022707	229.95
2021-08-24	70.12632	67.34992	68.48540295	58.275

ISSN (Online) 2395-2717

International Journal of Engineering Research in Electrical and Electronic Engineering (IJEREEE)

Vol 9, Issue 5, May 2022

18:00:00				
2021-08-25 06:00:00	19.51666	17.04642	18.47992114	18.9
2021-08-25 07:00:00	103.1024	101.1767	109.5479919	96.075
2021-08-25 08:00:00	247.8359	244.3056	235.8476671	195.3
2021-08-25 09:00:00	352.6551	380.4014	392.2144418	445.725
2021-08-25 10:00:00	420.7302	449.8164	474.6459907	615.825
2021-08-25 11:00:00	543.0222	574.8928	581.0942389	543.375
2021-08-25 12:00:00	611.0407	648.5073	664.9358493	699.3
2021-08-25 13:00:00	698.3483	677.2099	645.3599195	815.85
2021-08-25 14:00:00	681.1626	699.7653	646.3184008	751.275
2021-08-25 15:00:00	606.7025	604.6988	581.872807	622.125
2021-08-25 16:00:00	458.8839	449.9445	433.086613	415.8
2021-08-25 17:00:00	214.4803	214.6715	214.6117946	214.2
2021-08-25 18:00:00	68.01085	65.02374	65.98894995	56.7

Table 4. MAPE values of the model in terms of polynomial degrees

Polynomial Degree	***2.***	***3.***	***4.***
MAPE (%)	14.58	16.78	18.23

VI. CONCLUSION

The best solutions are obtained in the hours of "06:00, 07:00 and 17:00", such that predicted values are closest the actual energy values for the days of 24.08.2021 and 25.08.2021. The least successes between actual and estimated values are seen at "10:00" for both two days. The most deviation between real and predicted values take place at "13:00" in the day of 25.08.2021. All of these can be seen from Fig. III., given in above by looking carefully to it.

As can be seen from Fig. IV. given in above, MAPE values increases with respect to the polynomial degree. The best estimation is carried out from 2^{nd} degree polynomial regression, its MAPE is 14.58%.

These results can be assumed to be "Good" according to Lewis's book of "Demand Forecasting and Inventory Control", because the results of 14.58%, 16.78% and 18.23% are in range of 10% - 20%.

If more rows, in other words examples are acquired, less MAPE values can be given from the model. This is a basic study, which includes insufficient dataset for other estimation studies. This point refers importance of the study. The estimation program may be utilized for other solar applications, moreover used for wind energy applications.

"Degree" is a hyperparameter for polynomial regression. Arranging degree of the polynomial regression is so crucial for performance metrics (MAPE and R2-Score).

If new data is added to the dataset, the results may be under 10% MAPE. If dataset is enriched, deep learning algorithms are applied to the dataset, in this case again MAPE may be decreased under 10%.

Here it can be deduced that, success of any prediction by using polynomial regression is directly depending upon the degree and size of the dataset. More dataset generally optimizes MAPE and R^2-Score, but more degree of the polynomial regression does not yield better solutions forever.

ISSN (Online) 2395-2717

International Journal of Engineering Research in Electrical and Electronic Engineering (IJEREEE)

Vol 9, Issue 5, May 2022

REFERENCES

[1] https://seffaflik.epias.com.tr/transparency/uretim/gercekleseen-uretim/gercek-zamanli-uretim.xhtml, Accessed on: Jan. 2, 2022

[2] http://www.osmaniyeozelidare.gov.tr/gunes-enerji-santrali-yapimi, Accessed on: Jan. 2, 2022

[3] D. A. Freedman, "Statistical Models: Theory and Practice," Univercity of California, Cambridge University Press, New York, The USA, pp. 424, [Online]. Available: https://books.google.com.tr/books?hl=tr&lr=&id=fW_9BV5Wpf8C&oi=fnd&pg=PR1&dq=Freedman+DA.+Statistical+Models:+Theory+and+Practice,+Cambridge+University+Press,+New+York,+2005,+424.&ots=2iMaSCGWPG&sig=URHSO0O-wHClQGIweTsUxuOreJc&redir_esc=y#v=onepage&q&f=false

[4] R. Alpar, "Uygulamalı Çok Değişkenli İstatistiksel Yöntemler," ed., Detay Press, Ankara, 2017, pp. 840.

[5] J. Fan, I. Gijbels, "Local, Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability 66," CRC Press, 1996, pp. 336.

[6] J. O. Rawlings, S. G. Pantula, D. A. Dickey, "Applied Regression Analysis (2 nd ed.): A Research Tool," Springer Science & Business Media, 1998, pp. 657. [Online]. Available: http://web.nchu.edu.tw/~numerical/course1012/ra/Applied_Regression_Analysis_A_Research_Tool.pdf

[7] S. Aggrey, "Comparison of three nonlinear and spline regression models for describing chicken growth curves," Poultry Science Association, vol. 81, no.12, pp. 1782-1788, 2002, to be published. DOI: 10.1093/ps/81.12.1782

[8] G.A.F. Seber, C.J. Wild, "Nonlinear Regression," John Wiley&Sons, New Jersey, 2003, pp. 768, to be published. DOI: 10.1002/0471725315

[9] B.C. Lee, D.M. Brooks, "Regression modeling strategies for microarchitectural performance and power prediction," Harvard Computer Science Group Technical Report, USA, Rep. TR-08-06, Mar. 2006, [Online]. Available: https://dash.harvard.edu/bitstream/handle/1/24829620/tr-08-06.pdf?sequence=1&isAllowed=y

[10] N.R. Draper , H. Smith, "Applied Regression Analysis," ed., John Wiley&Sons, New York, 2014.

[11] B. Varol, "Parçalı Regresyon İle Polinom Regresyon Analizlerinin Karşılaştırması", Msc thesis, Adnan Menderes University, Health Sciences Institute, Aydın, 2017.

[12] https://www.geeksforgeeks.org/how-to-calculate-mean-absolute-percentage-error-in-excel/, Accessing dare: 18.05.2022

[13] F. Aslay, U. Ozen, "Estimating soil temperature with artificial neural networks using meteorological parameters," Journal of Polytechnic, vol. 16, no. 4, pp. 139-145, 2013, to be published DOI: 10.2339/2013.16.4, 139-145.

[14] J. Lupón, H. K. Gaggin, M. de Antonio, M. Domingo, A. Galán, E. Zamora, J. Vila, J. Peñafiel, A. Urrutia, E. Ferrer, N. Vallejo, J. L. Januzzi, and A. Bayes-Genis, ''Biomarker-assist score for reverse remodeling prediction in heart failure: The ST2-R2 score,'' Int. J. Cardiol., vol. 184, pp. 337–343, Apr. 2015.

[15] J.-H. Han and S.-Y. Chi, ''Consideration of manufacturing data to apply machine learning methods for predictive manufacturing,'' in Proc. 8th Int. Conf. Ubiquitous Future Netw. (ICUFN), Jul. 2016, pp. 109–113.

[16] F. Rustam, A. A. Reshi, A. Mehmood, S. Ullah, B-W. On, W. Aslam, A. G. S. Chio, "COVID-19 Future forecasting using supervised machine learning models,", IEEE Access, vol. 8, pp. 101489 – 101499, May. 2020, to be published. DOI: 10.1109/ACCESS.2020.2997311