Boğaziçi University

IE 360: Statistical Forecasting and Time Series

Project Report

05.06.2024 - Spring 2024

Uygar Şahin - 2022402291

Eylül Rana Saraç - 2019402126

Ayşe Sena Yeşilova - 2020402195

CONTENTS

ABSTRACT

Forecasting is a very important step in every business that tries to obtain a benefit at the end of the process. The importance of forecasting increases, especially in business processes where the intended benefit is economic benefit. The aim of the project is to make the energy forecast for Edikli Solar Power Plant as accurately as possible. The accuracy of this estimation and its closeness to real values are very critical as it provides information about the efficiency of the power plant. In addition, this estimation not only shows the efficiency of the power plant, but also acts as a precaution against possible problems that may occur in the future.

1. INTRODUCTION

Nowadays, solar energy has a very important place among energy production methods, both because it is renewable and because of its minimal damage to the environment. Turkey is a country that receives sufficient sunlight at the required angle and intensity to establish an efficient solar power plant.

In this project report, it is explained how short-term energy level predictions of the Edikli Solar Power Plant are made. In order to make the 24-day forecasting process accurate, the data given in the project file was taken as the basis.

1.1 Variables Related to the Forecast Process

***DSWRF_surface (Downward Shortwave Radiation)*:** Panels in solar power plants produce energy according to the intensity of the light falling on them. The angle at which the Sun's rays fall on the Earth changes this density and also changes the shortwave radiation flux formed there. The DWSRF_surface variable is the value of the average downward shortwave radiation flux falling on the Earth, and it is highly related to the forecasting process of the energy level of the solar panel.

USWRF_top_of_atmosphere, USWRF_surface, DLWRF_surface: While the rays from the Sun are reflected from the surface of the Earth, the rays from the Sun continue to come to the Earth. As a result of this continuous radiation exchange, a certain amount of energy remains almost constant between the atmosphere and the Earth, preventing the Earth's climate and temperature from changing with high variance and making it more stable. This energy accumulation plays an important role in forecasting solar energy production.

TCDC_low.cloud.layer,TCDC_middle.cloud.layer,TCDC_high.cloud.layer, TCDC_entire.atmosphere: The clouds formed in the sky were divided into layers and reflected in the data. Since the amount of cloud accumulated in these layers affects the density at which the Sun's rays pass, it also affects the radiation flux value created on the Earth and therefore the energy value in the solar power plant.

CSNOW_surface: This is a binary variable which gets values 0 or 1. When this variable takes the value 1, it indicates that there is snowfall in the region where the power plant is located. During snowfall, areas of the solar panel may be covered with snow, which reduces the level of energy that can be produced.

TMP_surface: This variable shows the temperature in the region where the power plant is located. Temperature gives an idea about the season in the time period we will make forecasting. Since Turkey is a country located in the northern hemisphere, the time period in which the sun is effective decreases in winter and increases in summer. Due to this effect, it becomes important in which season we make our forecast based on temperature. Although the effectiveness of the Sun increases in summer, high temperature reduces the efficiency of solar energy panels. This temperature-efficiency relationship must be considered, too.

1.2 Summary of the Approach

The variables defined above are given as hourly data in the project file. Using this data, a regression model was used to predict the energy that Edikli Solar Power Plant will produce. In order for the regression model to work more accurately and to model the data more easily, the dataset is divided into several categories based on past energy production amounts. While making predictions, in addition to past production values, variables such as radiation flux, cloud accumulation, snow, etc. given in the dataset were added to the model and their effects were taken into account. The production data and the values of the variables in the dataset were placed in a regression model and the energy production forecast for the coming days was made based on this data-variable relationship.

2. APPROACH

2.1 Categorization of Dataset

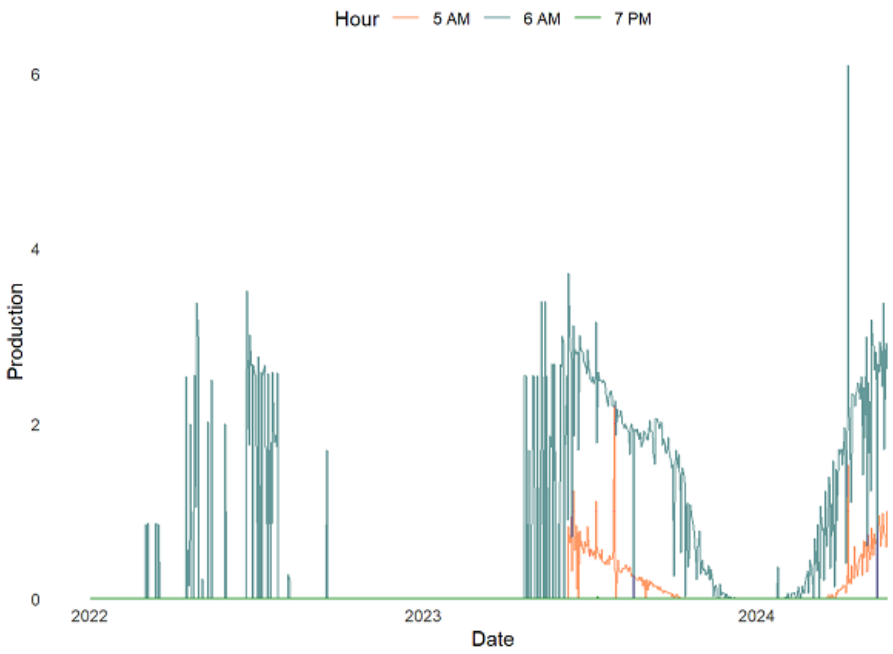
Since our subject in the project we will be estimating is a solar power plant, there are hours when the rays from the sun cannot be converted into energy. Energy production during these hours is given as zero in the dataset.

The given dataset is divided into two categories at the beginning: hours in which production is zero and hours in which production is greater than zero. Thanks to this categorization, the part that will support our model and the redundant parts are separated when making predictions. In addition, reducing the size of the given large data makes it easier for us when applying the model and can be considered as a precaution against possible errors.

The table indicating the average production amount of each hour in the given dataset is given in Appendix Table 1. According to this table, on the hours when production was zero, no production appeared throughout the dataset. Therefore, these hours give the category of hours in which production is zero, which is one of the categories we divided. It is not necessary to apply the model during these hour intervals when the production is zero or very close to zero (if the 3 digits after the comma are also considered zero), these days are estimated to be zero in the future as well.

Then, hours when production was greater than zero were also divided into two categories. This categorization was made according to whether production was affected by seasonality at that time. Since the sunrise time in the winter months is later than in the summer months, production is seasonally 0 at 5 am, 6 am, and 7 pm in some months. We called this classification of hours "critical hours". The behavior of 5am and 6am can be seen from Figure 1. Although production rates at 7 pm is always too low all year round, we addressed it as a critical hour because there are some production between 0.01 – 0.03 for a few months.

Figure 1. Production rate at critical hours



2.2 Modeling for Critical Hours

In addition to the abrupt shifts in production rates between zero-hour and critical hours, as well as between critical hours and regular hours, substantial fluctuations also occur during the transitions within critical hours. Therefore, it is possible to break down the problem further by developing a separate model for each critical hour.

The approach we took for the critical hours was to identify the weeks or months when the production was not zero. To determine them, the weeks or months that have at least one non-zero production rate are filtered for each critical hour. Then for each critical hour, the column created as is_in_nzp of the predictor took the value of that specific month or week if the timestamp of that row was in a non-zero week/month, and took "No" otherwise. Then a three-day lag column is added to the predictor to reduce the autocorrelations of the residuals. The most developed linear model for critical hours consists of raw weather data, 3 lag production, and is_in_nzp factor variable.

2.3 Modeling for Remaining Hours

The regular hours when the sun is effective for production for all months are between 6am to 7pm. The production is affected by seasonality as well. This production value difference varies depending on the cloud density during the day, the angle of incidence of the sun rays, and the intensity of the sun rays. There is also monthly seasonality because the angle of incidence of sunlight in winter is lower than in summer.

To address this issue, we added the month and hour as factor variables to the predictor of the linear model. After that, we used a 36-hour lagged production value (closer available production data) for each hour to reduce the autocorrelations of the residuals.

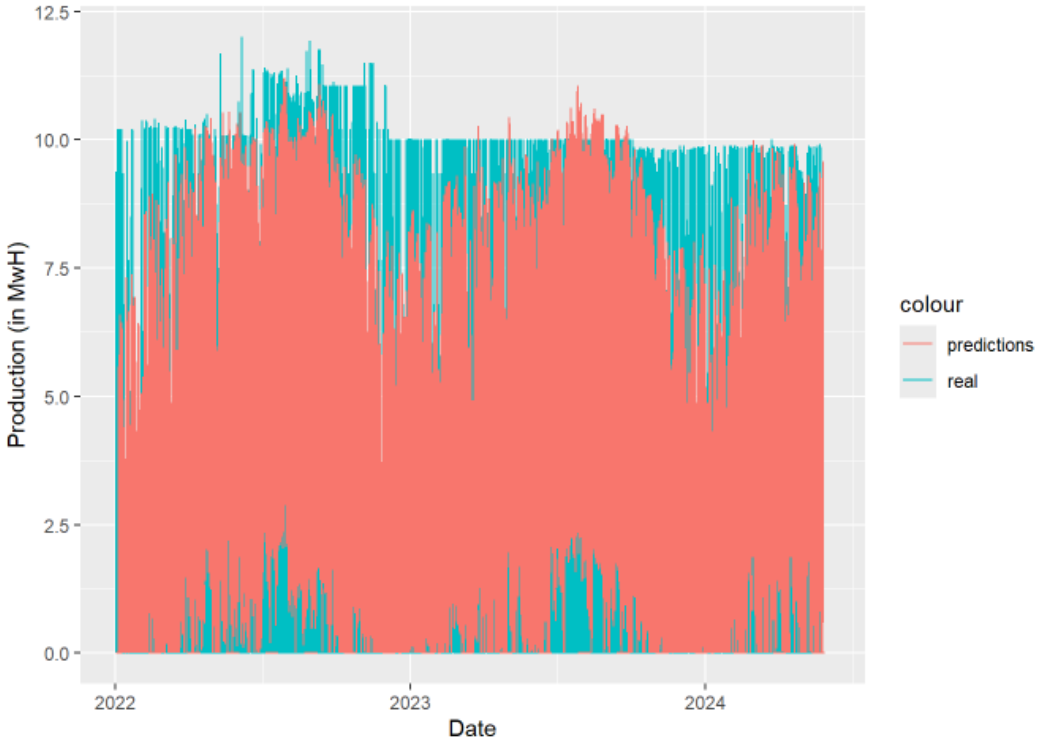
2.4 Predictions

In order to make accurate predictions, it is necessary to identify outlier values and prevent them from disrupting the accuracy of the model. Since the lower bound values in the given dataset are zero, the negative values produced by our model's predictions have been rounded directly to zero. For the upper limit, the maximum value of production in the given dataset is calculated as 12. Therefore, values greater than 12 were determined as outliers and these values were rounded to 12 manually. Furthermore, by summing up the previous residuals and the latest prediction values, the error is reduced, since the data set contains lagged values.

3. RESULTS

The relationship between our estimated values and actual values is given in the figure below.

Figure 2. predictions vs real data



4. CONCLUSIONS AND FUTURE WORK

In order to increase the accuracy of the model, it is necessary to identify where the errors in the model originate and focus on these points. Carefully examining and analyzing adjusted R2 values at the point of detecting these errors can be one of the first steps that play an important role in the development of the model. On the other hand, although operations were carried out by dividing our dataset into several categories, categorizing and analyzing the dataset in more detail can help increase the accuracy of our model.

Although the rounding method used when making predictions takes us to approximately correct values, it is an obstacle that must be overcome in order for the model to reach exactly correct values. In order to make more precise predictions, modeling systems that provide sharper results should be developed instead of these rounding values.

In addition, in order to ensure that the model works better, parameters such as independence of errors, equal variance of errors, and normality of errors should be constantly monitored in more detail and how the modifications we make in the model affect these parameters should be closely examined. Changes in these values can be used as a basis to observe the accuracy of the results created by the modifications made.

5. CODE LINK

You can access the R code here: https://github.com/BU-IE-360/spring24-ayseyesilova/blob/main/htmlproject.html

6. APPENDICES

Appendix Table 1. *Hourly Average Production Table*

Hour	Production	Hour	Production
0	0	12	7,893
1	0	13	7,453
2	0	14	6,379
3	0	15	4,561
4	0	16	2,361
5	0,057	17	0,837
6	0,588	18	0,068
7	2,755	19	0
8	5,434	20	0
9	7,349	21	0
10	7,929	22	0
11	8,005	23	0

Appendix 2. *R codes for plots*

2.1) code for figure 1:

```
critical_hours %>%
  ggplot(aes(x = datetime, y = production, group = factor(hour), color = factor(hour)))
+
  geom_line() +
  scale_color_manual(name = "Hour",
                     labels = c("5 AM", "6 AM", "7 PM"),
                     values = c("#E64B35B2", "#3C5488B2", "#3E8543")) +
  xlab("Date") +
  ylab("Production") +
  theme_minimal() +
  theme(legend.position = "top") +
  scale_y_continuous(expand = c(0, 0))
```

2.2) code for figure 2:

```
ggplot(data_pr_p, aes(x = datetime)     )   +

   geom_line(aes(y = production,
```

```
            color='real',
            group = 1)          )   +

        geom_line(aes(y = predicted,
            color = 'predictions',
            group = 1)          )   +

        xlab("Date")                    +

        ylab("Production")
```

Appendix 3. *code for Table 1*

```
round(aggregate(available_data$production, list(available_data$hour), FUN = mean), 4)
```

Appendix 4. *competition phase predictions vs real data*

May 13. wmape = 0.376

	pred	production		pred	production
0	0	0	12	6.441994	5.34
1	0	0	13	5.974214	4.04
2	0	0	14	4.702902	5.89
3	0	0	15	2.850898	5.32
4	0	0.06	16	1.183686	3.62
5	0.482478	0.76	17	0.071519	1.3
6	1.857501	2.69	18	0	0.16
7	2.095748	5.75	19	0.000391	0
8	4.503184	9.24	20	0	0
9	5.996955	8.88	21	0	0
10	6.475628	7.85	22	0	0
11	6.605879	8.36	23	0	0

May 14.

Wmape = 0.2876

	prediction	production		prediction	production
0	0	0	12	7.447437	7.78
1	0	0	13	7.533618	9.64
2	0	0	14	6.585047	9.7
3	0	0	15	5.101065	7.82
4	0	0.07	16	3.571312	4.55
5	0.460272	0.96	17	2.253216	1.33
6	1.789544	2.59	18	1.554308	0.14
7	2.236762	4.14	19	0	0

8	4.793639	7.6	20	0	0
9	6.49799	8.83	21	0	0
10	8.101983	7.44	22	0	0
11	7.735065	5.84	23	0	0

May 15.

Wmape
=0.267214233

	pred	production		pred	production
0	0	0	12	7.896115	7.92
1	0	0	13	7.407491	7.18
2	0	0	14	6.397065	6
3	0	0	15	4.676544	3.11
4	0	0.08	16	2.251177	2.5
5	0.129406	0.84	17	1.000671	0.83
6	1.318975	2.94	18	0.416241	0.28
7	2.59205	6.58	19	0	0
8	5.064411	9.13	20	0	0
9	6.726767	9.82	21	0	0
10	7.524906	9.87	22	0	0
11	7.945432	9.82	23	0	0

May 16.

Wmape = 0.383436118

	pred	production		pred	production
0	0	0	12	6.661105	5.52
1	0	0	13	5.723915	8.89
2	0	0	14	5.066857	9.5
3	0	0	15	3.5833	5.32
4	0	0.04	16	1.397876	2.88
5	0.553701	0.6	17	0.18377	1.27
6	2.060903	2.89	18	0	0.16
7	4.010704	6.45	19	0.000383	0
8	6.917018	8.39	20	0	0
9	8.386849	8.39	21	0	0
10	7.904611	2.82	22	0	0
11	7.943125	4.96	23	0	0

May 17.
Wmape =
0.293000517

	pred	production		pred	production
0	0	0	12	6.109484	6.71
1	0	0	13	5.89224	7.89
2	0	0	14	5.112067	5.08
3	0	0	15	3.193958	6.78
4	0	0.04	16	1.229083	3.31

5	0.675678	0.76	17	0	1.5
6	1.97266	2.43	18	0	0.18
7	2.465602	4.23	19	0.00069	0
8	4.994101	5.46	20	0	0
9	6.34896	9.42	21	0	0
10	5.937381	8.64	22	0	0
11	6.064114	4.9	23	0	0

May 18.

Wmape = 0.20563075

	pred	production		pred	production
0	0	0	12	8.245396	9.64
1	0	0	13	7.684366	9.78
2	0	0	14	6.586797	9.28
3	0	0	15	4.68667	7.17
4	0	0.15	16	3.004325	4.16
5	0.628387	0.98	17	1.549599	1.31
6	2.47772	2.83	18	0.863723	0.2
7	3.923074	5.81	19	0	0
8	6.884595	8.53	20	0	0
9	8.559062	9.8	21	0	0
10	8.90775	9.83	22	0	0
11	8.742355	9.83	23	0	0

May 19.
Wmape =
0.252986956

	pred	production		pred	production
0	0	0	12	7.730968	9.65
1	0	0	13	7.733312	9.85
2	0	0	14	6.808157	9.57
3	0	0	15	4.681926	7.61
4	0	0	16	2.298585	4.24
5	0.4888	0.39	17	0.821137	0.97
6	2.36727	2.32	18	0	0.04
7	3.846577	5.74	19	0	0
8	6.532751	8.72	20	0	0
9	7.866512	9.8	21	0	0
10	7.477199	9.88	22	0	0
11	7.831771	9.83	23	0	0

May 20.

Wmape = 0.17438146

	pred	production		pred	production
0	0	0	12	7.266623	5.58
1	0	0	13	6.829662	7.28
2	0	0	14	5.019711	4.94
3	0	0	15	3.692417	3.76
4	0	0.1	16	1.3103	3.42
5	0.583898	0.75	17	0	1.28
6	2.013345	1.71	18	0	0.23
7	3.270185	2.82	19	0.000769	0
8	5.571276	4.43	20	0	0
9	7.428387	7.07	21	0	0
10	7.677379	8.79	22	0	0
11	7.045852	8	23	0	0

May 21.

Wmape = 0.331346782

	pred	production		pred	production
0	0	0	12	8.504114	3.95
1	0	0	13	7.712857	7.76
2	0	0	14	6.386	8.1
3	0	0	15	4.309956	4.67
4	0	0.09	16	1.740808	3.55
5	0.732606	0.72	17	0.282177	1.49
6	2.303036	2.85	18	0	0.21
7	4.461056	5.88	19	0	0
8	7.344085	8.42	20	0	0
9	9.076203	7.73	21	0	0
10	8.841764	5.26	22	0	0
11	8.788144	5	23	0	0

May 22.

Wmape = 0.136971993

	pred	production		pred	production
0	0	0	12	9.317034	9.91
1	0	0	13	9.022901	9.83
2	0	0	14	8.069349	9.39
3	0	0	15	6.220214	6.46
4	0	0.07	16	3.751894	3.12
5	0.651151	0.59	17	1.774148	0.97
6	2.462231	2.92	18	1.000062	0.08
7	3.791875	6.24	19	0	0
8	7.024197	8.94	20	0	0

9	9.002616	9.86	21	0	0
10	9.335519	9.85	22	0	0
11	9.409566	9.83	23	0	0

May 23.

Wmape = 0.130223103

	pred	production		pred	production
0	0	0	12	8.620565	9.4
1	0	0	13	8.3493	8.19
2	0	0	14	6.760826	8.25
3	0	0	15	4.967235	4.99
4	0	0.07	16	3.126791	2.8
5	0.589074	1	17	1.433401	0.86
6	1.6171	2.62	18	0.791431	0.09
7	4.177831	5.9	19	0	0
8	6.959804	8.5	20	0	0
9	9.002536	9.77	21	0	0
10	9.249831	9.85	22	0	0
11	9.171733	9.68	23	0	0

May 24.

Wmape = 0.592470623

	pred	production		pred	production
0	0	0	12	6.465	7.21
1	0	0	13	6.2724	5.74
2	0	0	14	4.8272	2.26
3	0	0	15	2.7084	2.09
4	0	0	16	1.0743	2.81
5	0.3029	0.02	17	0	1
6	2.2676	0.02	18	0	0.07
7	3.8041	2.49	19	0.000779	0
8	6.2922	2.43	20	0	0
9	7.4475	4.89	21	0	0
10	6.8265	2.77	22	0	0
11	6.732	5.21	23	0	0

May 25.

Wmape = 0.562930568

	pred	production		pred	production
0	0	0	12	5.520195	6.4
1	0	0	13	5.247586	8.2
2	0	0	14	4.665723	5.75
3	0	0	15	3.231674	5.47
4	0	0.01	16	1.568266	3.78

5	0.501372	0.6	17	0.061463	0.69
6	1.647536	0.9	18	0	0.24
7	2.328096	0.82	19	0.000109	0
8	4.599471	1.71	20	0	0
9	5.787595	3.57	21	0	0
10	5.964773	2.4	22	0	0
11	5.691669	2.65	23	0	0

May 26.

Wmape = 0.263451449

	pred	production		pred	production
0	0	0	12	8.667136	9.41
1	0	0	13	7.909837	7.54
2	0	0	14	6.910459	8.06
3	0	0	15	4.701973	5.21
4	0	0	16	2.395781	3.64
5	0.707	0.27	17	0.993747	0.77
6	1.9807	1.42	18	0.393488	0.04
7	4.677564	3.25	19	0	0
8	7.55711	5.95	20	0	0
9	9.333889	7.75	21	0	0
10	9.252906	5.68	22	0	0
11	9.100762	5.81	23	0	0