IE 360 Project

Solar Power Forecasting

Prepared By

Adrian Efe Liakof

Berke Aktürk

Nusret Kaan Polat

Instructor

Mustafa Gökçe Baydoğan

1.Introduction

In this project, the aim is to provide hourly solar power prediction of KIVANC 2 GES (Güneş Enerjisi Santrali) for the next day by using the time series data starting from 1st of February 2022 to last available data of the production which is two days before the forecasted day.

This report will start by building an ARIMA model as the base comparison model. Then, we will be building time series linear regression models with different regressors. After building all the models, we will be comparing the models according to their WMAPEs and use the best model to make a prediction.

2.Preparing the Data

a.Arranging the Data: There two data sets to be used in the project where one of them is containing the daily production amount data and the other one contains the contains the regressors: related cloud cover, downward shortwave radiation (DSWRF), relative humidity and temperature data for six different locations. We started by turning the long format regressor data to wide format. Then, we merged the regressor data set with the production data set.

b.Filling the Missing Data: Some production data are missed so to prevent the models from facing NAs, we added the production data from 24 hours before to every NA position. We started with a for loop starting from the first date so if there are two consecutive days with NA, they will be assigned the same production number.

c.Adding Dates and Simple Regressors: We added index and date information columns. Then, we added the Max Capacity column which is done by looking at the 24 hours data starting from 2 days before. This made us lose the data of two days but because the training data we have is long enough, this loss does not damage the models.

d.Deleting Hours with No Production: Between 8 p.m. and 5 a.m. , there is no significant production so we deleted the rows within these hours. By default, these hours are assigned “0” at the forecast phase

e.Adding Average of Regressors and Capacity Regressors: Because there is regressor data for 6 different locations and there is no significant superiority of a location, we added average of these locations for each regressor as a new regressors. We thought adding average of the regressors directly would not be logical since capacity differs significantly from period to period and hour to hour so we added Max Capacity*Each Regressors as new columns to the data. We also added the 42 hours before production data, 2 hours before temperature and 2 hours before humidity data as lagged regressors since the effect of temperature and humidity is not instantaneous in real time, it takes some time for solar panels to get hot or cold.

3.Model Creation Phase

a.ARIMA Model I:

The first ARIMA model is created by using the outputs given by the auto.arima function. According to the auto.arima function, the best parameter selection was (p,d,q) = (4,1,2). An arima model which uses parameters mentioned above created and it yielded the performance measures mentioned below:

Hourly Production

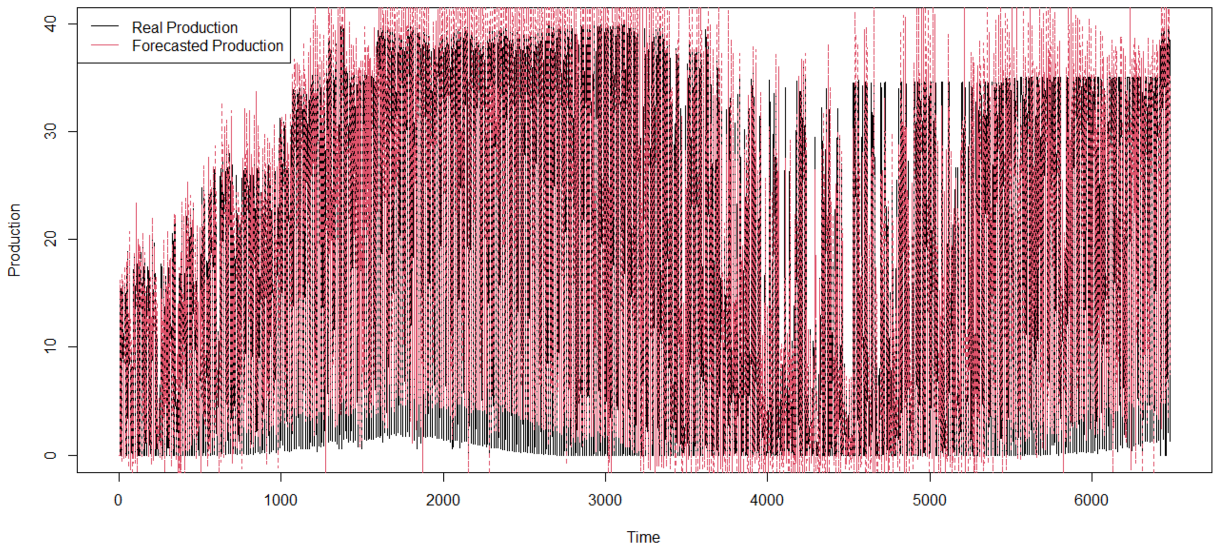


Figure 1: Hourly production rates forecasted with ARIMA (p,d,q) = (4,1,2)

WMAPE : 0.2215882

b.ARIMA Model II:

After building the first model and performing some examinations about the performance of the model, we decided to extract the first 1200 data to have a steady time series by eliminating the nonstationary period of the production data. As it can be seen from the picture above, the capacity of production is continuously increasing until the 1200th index.

Hourly Production

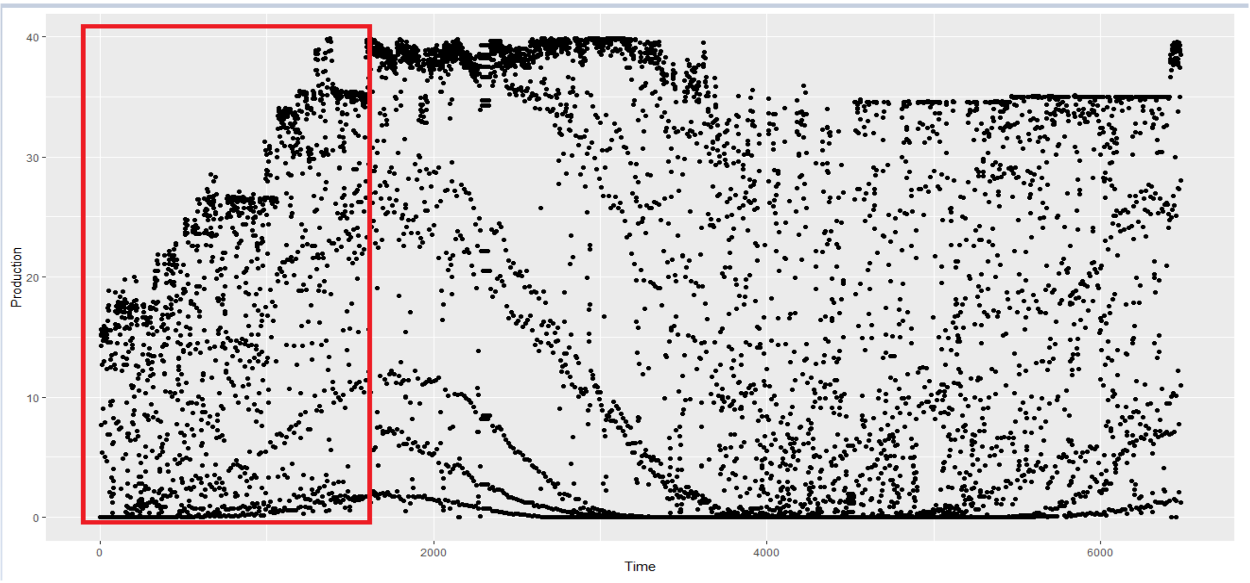


Figure 2: Hourly production rates

After eliminating the first part of the data, the same model is built by using the same configuration of the previous ARIMA model. The outputs of the second ARIMA model is as follows:

Hourly Production

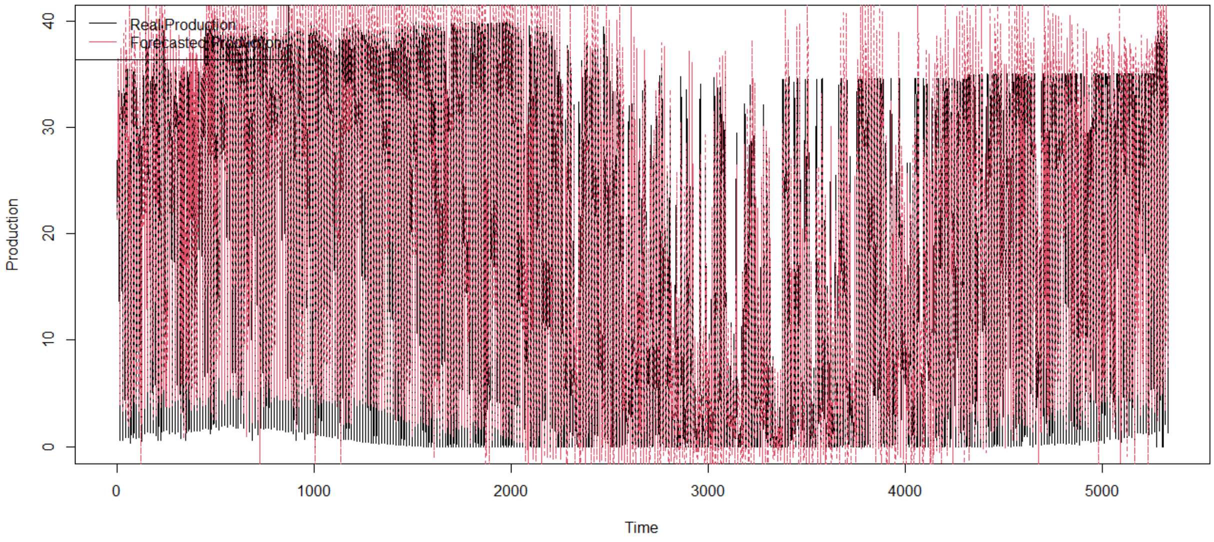


Figure 3: Hourly production rates forecasted with ARIMA (p,d,q) = (4,1,2) excluding first 1000 data

WMAPE : 0.2208339

c.ARIMA Model III:

After observing the fact that removing the nonstationary part of the data has a positive effect on the performance of the ARIMA model, we decided to extend the removal window to obtain more stationary training data and use the last 1150 data to train the model. Moreover, we have used our theoretical backgrounds to find parameters for the SARIMA model instead of using the auto.arima function. Therefore, the autocorrelation and partial autocorrelation plots have been taken into consideration. The autocorrelation and partial autocorrelation functions are as follows:

ACF Values of ARIMA Model III

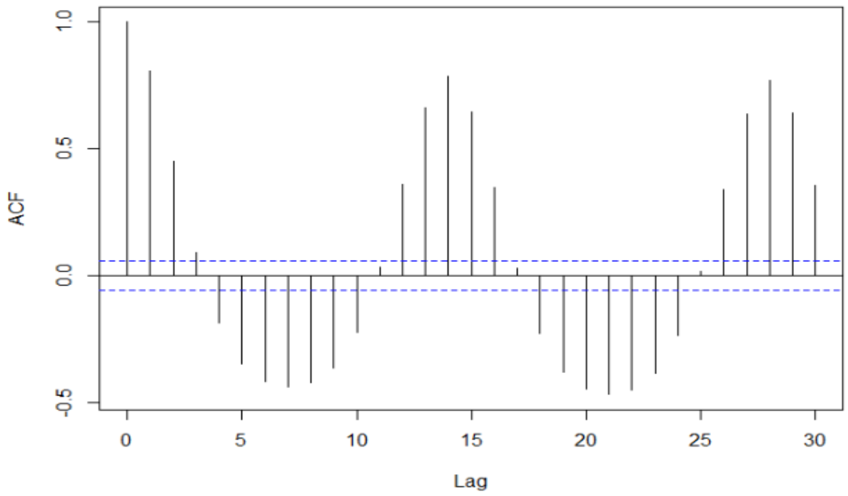


Figure 4: ACF of ARIMA Model III

ACF Values of ARIMA Model IV

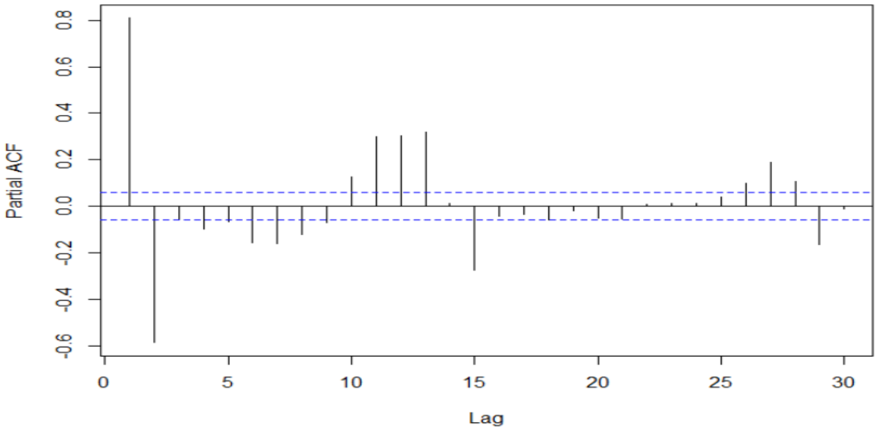


Figure 5: ACF of ARIMA Model IV

When the autocorrelation and partial autocorrelation graphs are investigated, it can be seen that the autocorrelation function is sinusoidal and there is a significant spike at lag 3 on the partial autocorrelation function. By using this information, we decided to use (p,d,0) = (3,1,0) parameters for our SARIMA model. To model the effect of the daily seasonality, we decided to use lag 14 for the seasonal part of the SARIMA model. Lag 14 is selected since, during the model creation phase, we discarded specific hours which always yields 0 productions. Therefore, lag 14 corresponds to the production of the same hour of the day before the current day. After the adjustments mentioned above, a significant improvement in the performance of the model is achieved. The outputs of the third model are as follows:

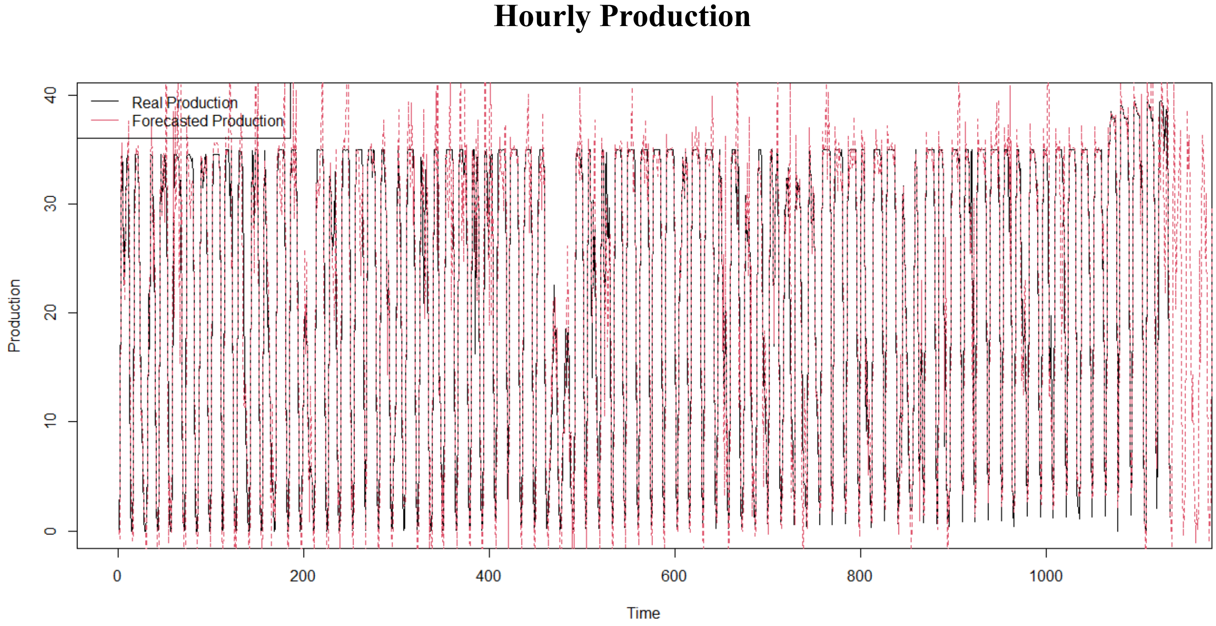


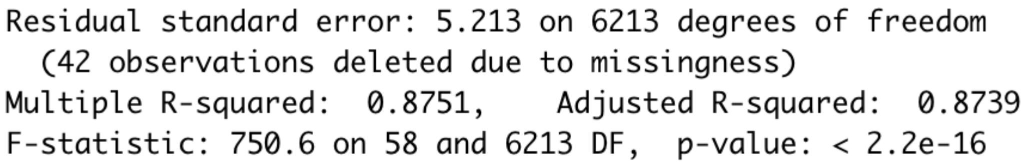
Figure 6: SARIMA model results

WMAPE = 0.1223735

However, in our forecasting environment, we do not have data for yesterday so we can not use lag 14 in our model. Therefore the model having approx. 0.12 wmape is not valid to use for forecasting.

d.Regression Model I:

The first model was created by using the all variables to see which variables can affect the model overall. After we get high level intuition about the regressors, we will determine the regressors to be eliminated. Because the output is so long, it will be added to the appendix page. When we look at the output, the attributes representing hours as factors came to be significant. From this, it is thought that using hourly models will be more logical to understand the trend and effects of other regressors. Also, lag 42 and regressors which show the multiplications of regressors and maximum capacities are significant as well. Lag 42 is used to solve some seasonality in the data. Multiplications will be explained in the model II.

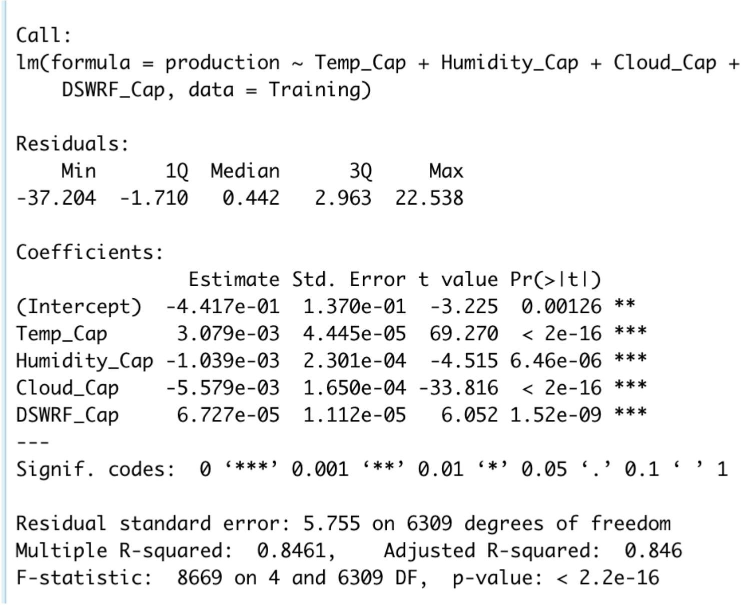


```
Residual standard error: 5.213 on 6213 degrees of freedom
  (42 observations deleted due to missingness)
Multiple R-squared:  0.8751,    Adjusted R-squared:  0.8739
F-statistic: 750.6 on 58 and 6213 DF,  p-value: < 2.2e-16
```

Figure 7: Results of 1st regression model

e.Regression Model II:

As we discussed before we have added the multiplications of averages of variables and maximum capacities to examine their effect on the production. In the first stage, we thought to add only these regressors to the model since we thought they both disclose the effect of capacities and variables and relations between them.



```
Call:
lm(formula = production ~ Temp_Cap + Humidity_Cap + Cloud_Cap + 
    DSWRF_Cap, data = Training)

Residuals:
    Min      1Q  Median      3Q     Max 
-37.204  -1.710   0.442   2.963  22.538 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -4.417e-01  1.370e-01  -3.225  0.00126 ** 
Temp_Cap      3.079e-03  4.445e-05  69.270  < 2e-16 ***
Humidity_Cap -1.039e-03  2.301e-04  -4.515 6.46e-06 ***
Cloud_Cap    -5.579e-03  1.650e-04 -33.816  < 2e-16 ***
DSWRF_Cap     6.727e-05  1.112e-05   6.052 1.52e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.755 on 6309 degrees of freedom
Multiple R-squared:  0.8461,    Adjusted R-squared:  0.846 
F-statistic:  8669 on 4 and 6309 DF,  p-value: < 2.2e-16
```

Figure 8: Results of 2nd regression model

As we can see in the output given above, the model has reached 0.8461 adjusted R squared by using only these 4 regressors. Then we have checked the distribution of residuals to see how we can improve the model.

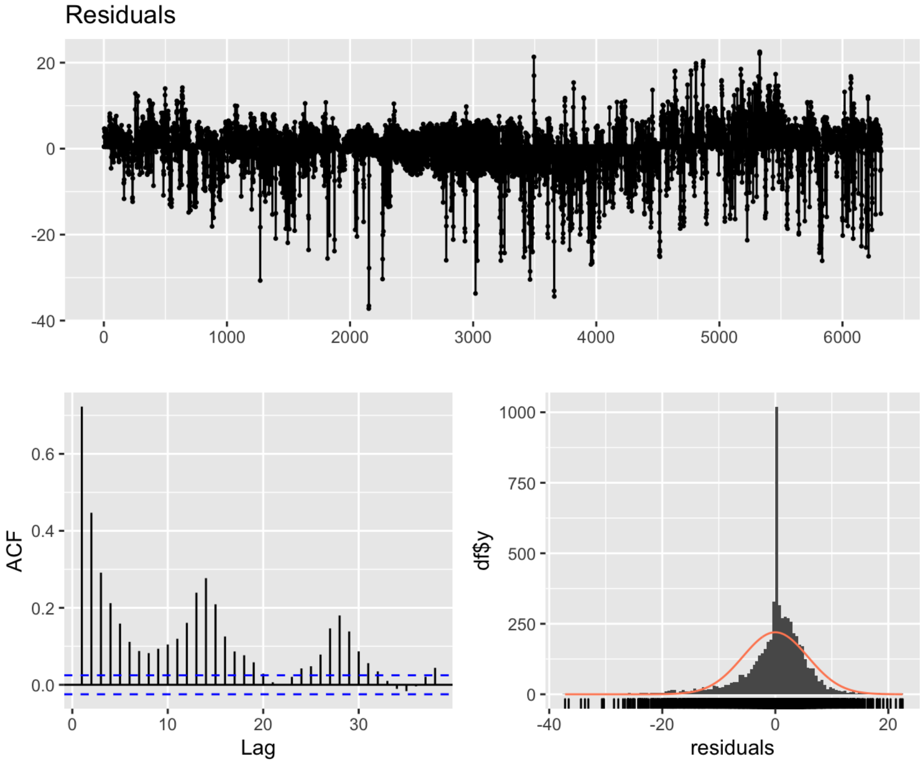


Figure 9: Residual graphs of 2nd regression model

When we examine the residual graphs, the seasonality can be observed easily in the ACF graph. The variance does not seem to be constant in the first graphs so the model can be improved by adding some new regressors and lag values to remove seasonality.

f.Regression Model III:

For the third model, we thought to use some important regressors which have already explained some of other regressors in the previous models and we decided to use only the last part of the original data for training since in the previous part of data, there is an increasing trend which we do not see in the last months. We have used only the flat section starting from mid-Feb to May.

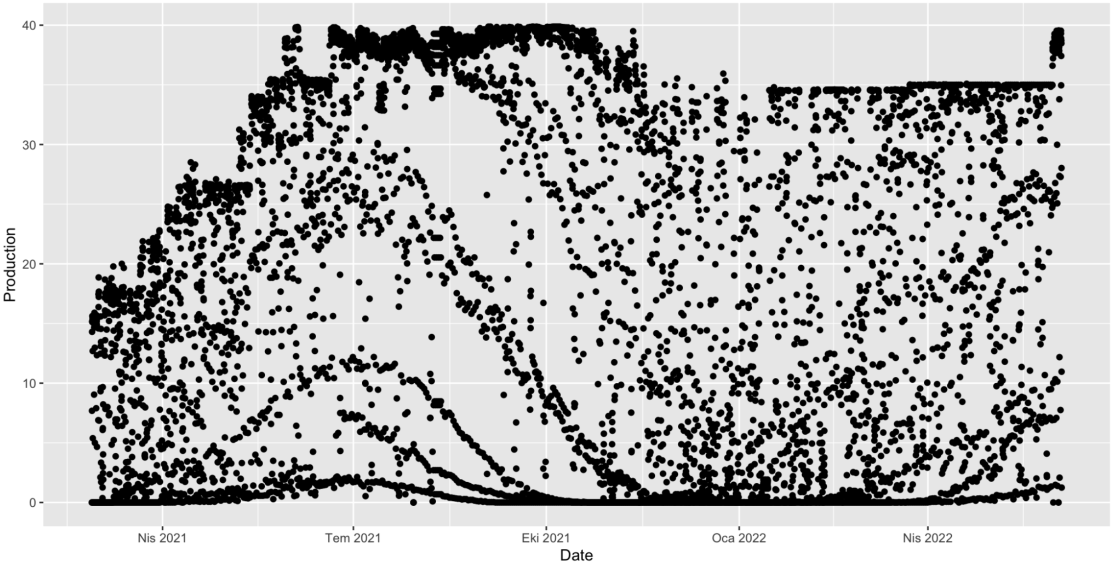
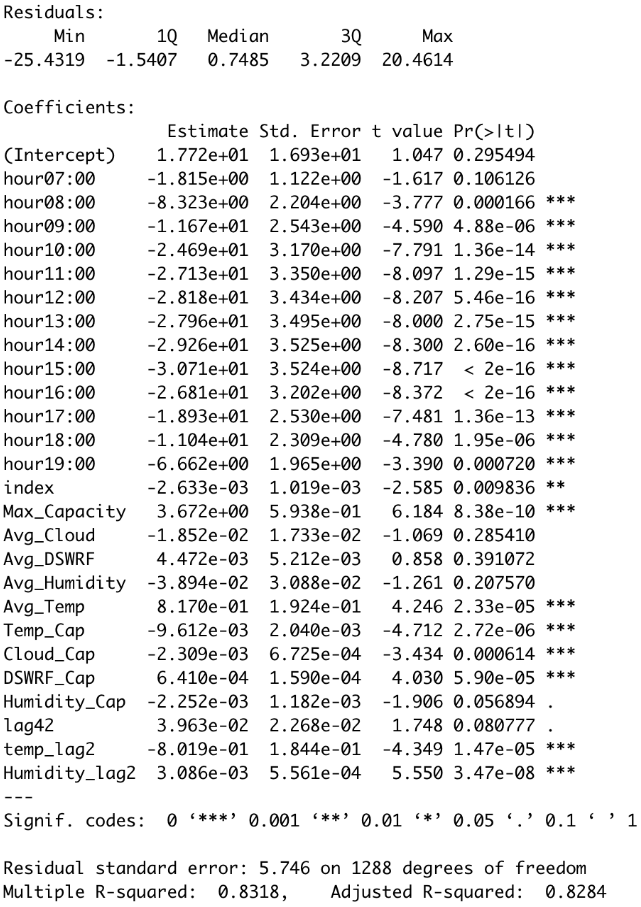


Figure 10: The distribution of production over days



```
Residuals:
     Min       1Q   Median       3Q      Max
-25.4319  -1.5407   0.7485   3.2209  20.4614

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.772e+01  1.693e+01   1.047 0.295494
hour07:00     -1.815e+00  1.122e+00  -1.617 0.106126
hour08:00     -8.323e+00  2.204e+00  -3.777 0.000166 ***
hour09:00     -1.167e+01  2.543e+00  -4.590 4.88e-06 ***
hour10:00     -2.469e+01  3.170e+00  -7.791 1.36e-14 ***
hour11:00     -2.713e+01  3.350e+00  -8.097 1.29e-15 ***
hour12:00     -2.818e+01  3.434e+00  -8.207 5.46e-16 ***
hour13:00     -2.796e+01  3.495e+00  -8.000 2.75e-15 ***
hour14:00     -2.926e+01  3.525e+00  -8.300 2.60e-16 ***
hour15:00     -3.071e+01  3.524e+00  -8.717  < 2e-16 ***
hour16:00     -2.681e+01  3.202e+00  -8.372  < 2e-16 ***
hour17:00     -1.893e+01  2.530e+00  -7.481 1.36e-13 ***
hour18:00     -1.104e+01  2.309e+00  -4.780 1.95e-06 ***
hour19:00     -6.662e+00  1.965e+00  -3.390 0.000720 ***
index         -2.633e-03  1.019e-03  -2.585 0.009836 **
Max_Capacity   3.672e+00  5.938e-01   6.184 8.38e-10 ***
Avg_Cloud     -1.852e-02  1.733e-02  -1.069 0.285410
Avg_DSWRF      4.472e-03  5.212e-03   0.858 0.391072
Avg_Humidity  -3.894e-02  3.088e-02  -1.261 0.207570
Avg_Temp       8.170e-01  1.924e-01   4.246 2.33e-05 ***
Temp_Cap      -9.612e-03  2.040e-03  -4.712 2.72e-06 ***
Cloud_Cap     -2.309e-03  6.725e-04  -3.434 0.000614 ***
DSWRF_Cap      6.410e-04  1.590e-04   4.030 5.90e-05 ***
Humidity_Cap  -2.252e-03  1.182e-03  -1.906 0.056894 .
lag42          3.963e-02  2.268e-02   1.748 0.080777 .
temp_lag2     -8.019e-01  1.844e-01  -4.349 1.47e-05 ***
Humidity_lag2  3.086e-03  5.561e-04   5.550 3.47e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.746 on 1288 degrees of freedom
Multiple R-squared:  0.8318,    Adjusted R-squared:  0.8284
```

Figure 11: Results of 3rd regression model

Although this model has lower adjusted R^2 value compared to previous models, we thought to use it since it is easy to interpret if we do not have a deep performance gap. As we discussed earlier, we will try to use this model as a prior model for hourly model creation.

g.Regression Model IV:

In the final model phase, we treated each hour independently and built a model for each hour. We deleted hours between 8 p.m. and 5 a.m. since they return 0 for most of the data so for the remaining 14 hours of the day, we built 14 different models. This final models use the following regressors to build models.:

-Average of the each regressor(Temp, DSWRF,Humidity,Cloud Cover)
-Max Capacity
-Max Capacity*(Average of the each regressors)
-3 days lag of the production data since we can use 3 days' before data to make a prediction of the next day
-2 hour lag of the temperature and humidity since it takes time for the panels to get hot and wet.

By building models for each hour, we improved the predictions significantly since each hour had different characteristics including the max capacity. Following summary tables and residual tables are for 3 models with hour 9,12 and 17..

Hour 9:

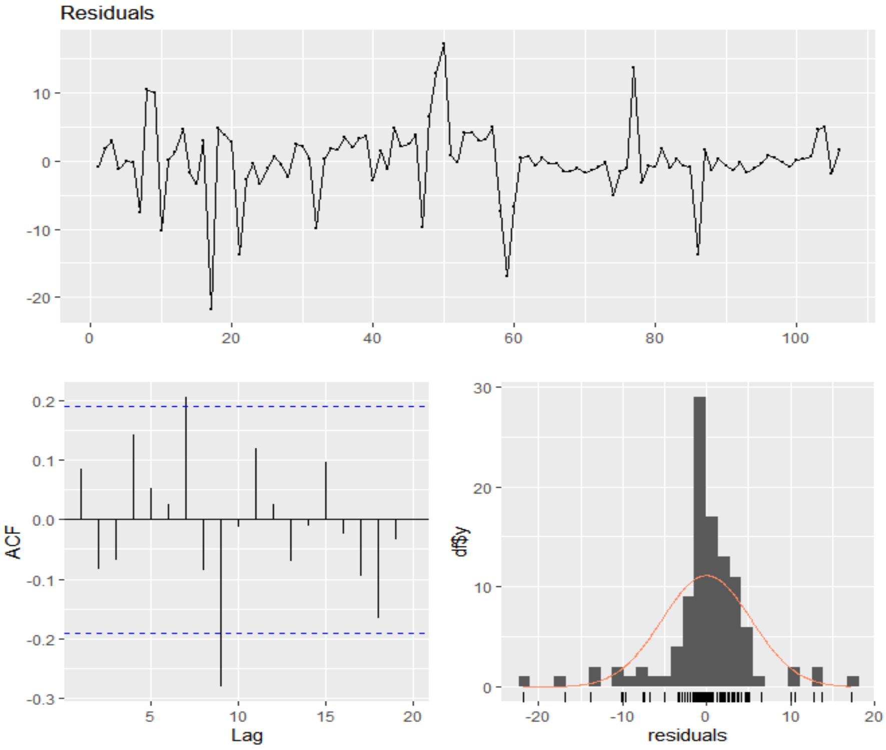
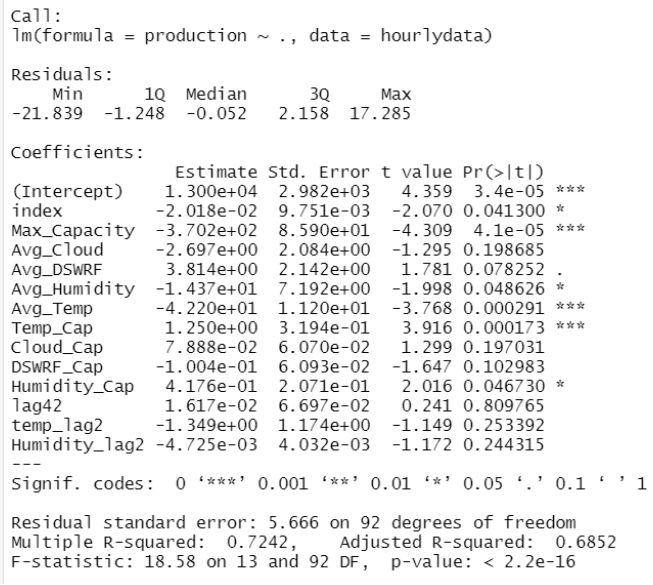


Figure 12 : Residuals of hourly model for 09.00



```
Call:
lm(formula = production ~ ., data = hourlydata)

Residuals:
    Min      1Q  Median      3Q     Max 
-21.839  -1.248  -0.052   2.158  17.285 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)    1.300e+04  2.982e+03   4.359  3.4e-05 ***
index         -2.018e-02  9.751e-03  -2.070 0.041300 *  
Max_Capacity  -3.702e+02  8.590e+01  -4.309  4.1e-05 ***
Avg_Cloud     -2.697e+00  2.084e+00  -1.295 0.198685    
Avg_DSWRF      3.814e+00  2.142e+00   1.781 0.078252 .  
Avg_Humidity  -1.437e+01  7.192e+00  -1.998 0.048626 *  
Avg_Temp      -4.220e+01  1.120e+01  -3.768 0.000291 ***
Temp_Cap       1.250e+00  3.194e-01   3.916 0.000173 ***
Cloud_Cap      7.888e-02  6.070e-02   1.299 0.197031    
DSWRF_Cap     -1.004e-01  6.093e-02  -1.647 0.102983    
Humidity_Cap   4.176e-01  2.071e-01   2.016 0.046730 *  
lag42          1.617e-02  6.697e-02   0.241 0.809765    
temp_lag2     -1.349e+00  1.174e+00  -1.149 0.253392    
Humidity_lag2 -4.725e-03  4.032e-03  -1.172 0.244315    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.666 on 92 degrees of freedom
Multiple R-squared:  0.7242,	Adjusted R-squared:  0.6852 
F-statistic: 18.58 on 13 and 92 DF,  p-value: < 2.2e-16
```

Figure 13: Results of hourly model for 09.00

WMAPE = 0.1103007

Hour 12:

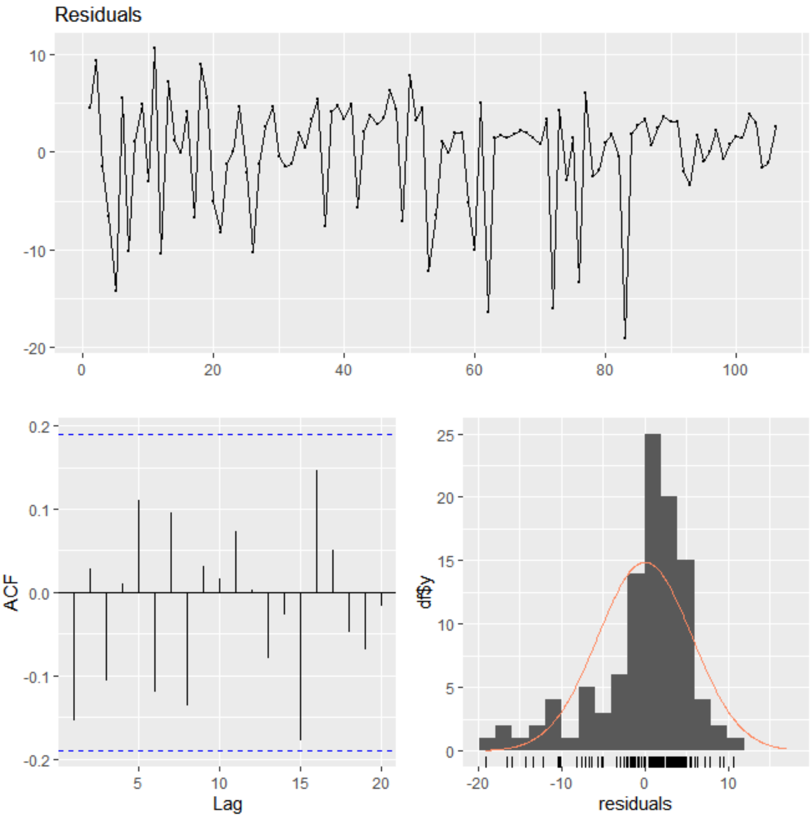
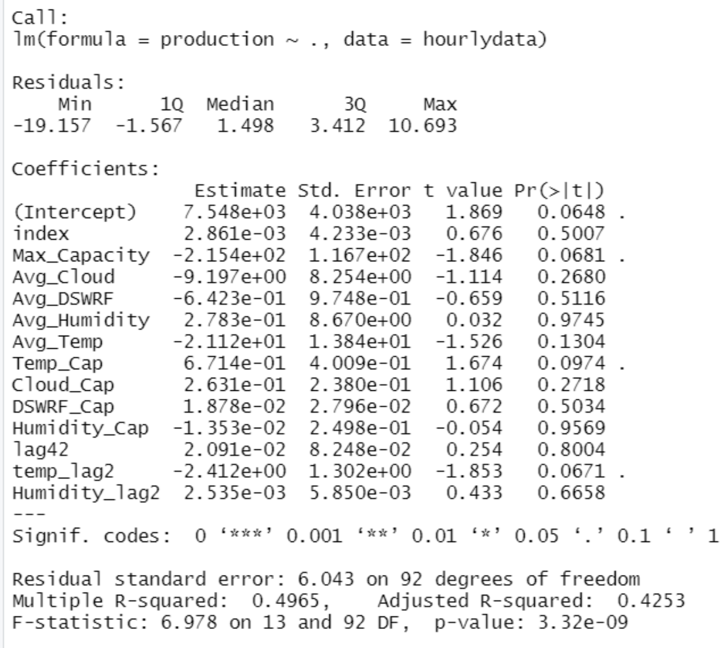


Figure 14 : Residuals of hourly model for 12.00


```
Call:
lm(formula = production ~ ., data = hourlydata)

Residuals:
    Min      1Q  Median      3Q     Max 
-19.157  -1.567   1.498   3.412  10.693 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)  
(Intercept)    7.548e+03  4.038e+03   1.869   0.0648 .
index          2.861e-03  4.233e-03   0.676   0.5007  
Max_Capacity  -2.154e+02  1.167e+02  -1.846   0.0681 .
Avg_Cloud     -9.197e+00  8.254e+00  -1.114   0.2680  
Avg_DSWRF     -6.423e-01  9.748e-01  -0.659   0.5116  
Avg_Humidity   2.783e-01  8.670e+00   0.032   0.9745  
Avg_Temp      -2.112e+01  1.384e+01  -1.526   0.1304  
Temp_Cap       6.714e-01  4.009e-01   1.674   0.0974 .
Cloud_Cap      2.631e-01  2.380e-01   1.106   0.2718  
DSWRF_Cap      1.878e-02  2.796e-02   0.672   0.5034  
Humidity_Cap  -1.353e-02  2.498e-01  -0.054   0.9569  
lag42          2.091e-02  8.248e-02   0.254   0.8004  
temp_lag2     -2.412e+00  1.302e+00  -1.853   0.0671 .
Humidity_lag2  2.535e-03  5.850e-03   0.433   0.6658  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.043 on 92 degrees of freedom
Multiple R-squared:  0.4965,	Adjusted R-squared:  0.4253 
F-statistic: 6.978 on 13 and 92 DF,  p-value: 3.32e-09
```

Figure 15: Results of hourly mode for 12.00

WMAPE = 0.137473

Hour 17:

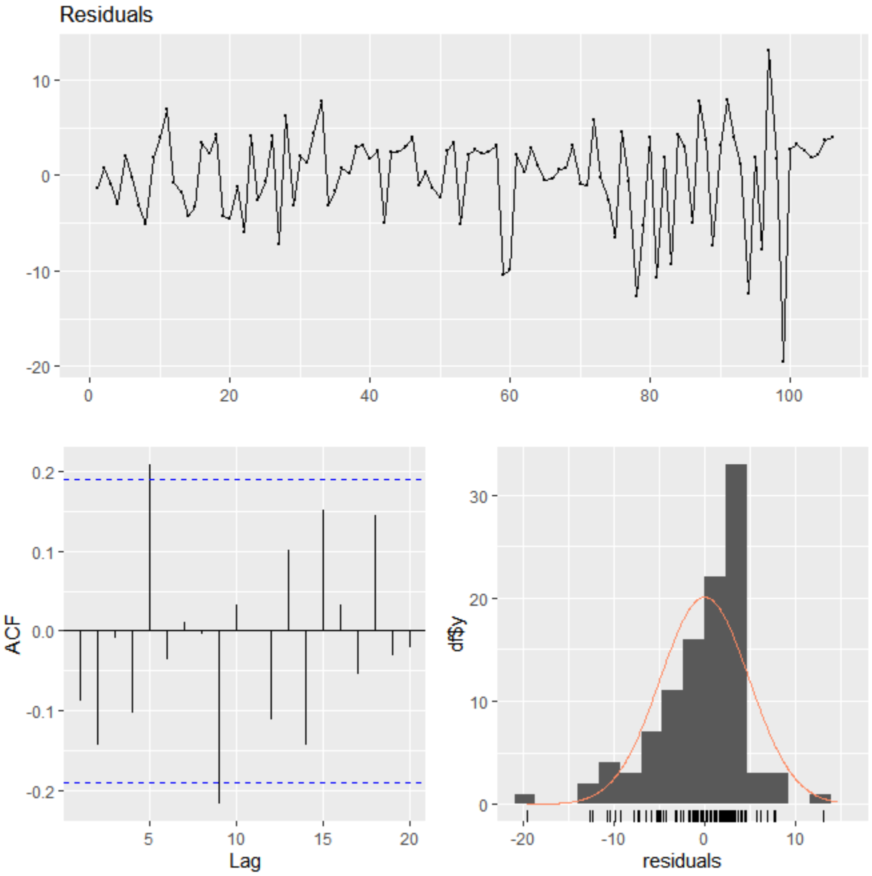
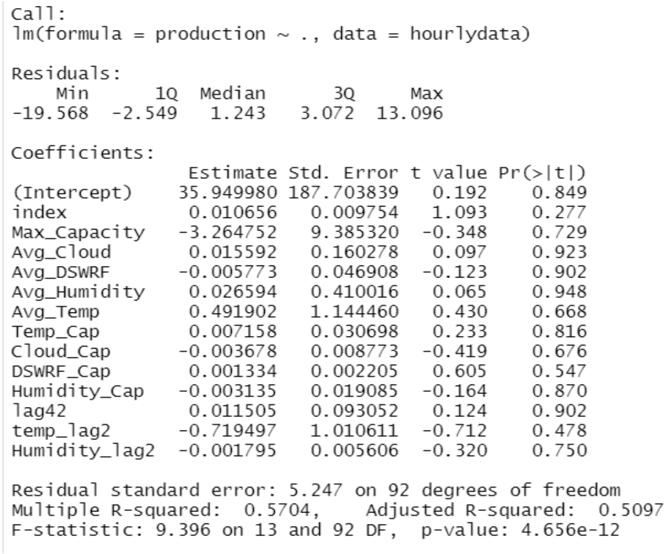


Figure 16 : Residuals of hourly model for 17.00



```
Call:
lm(formula = production ~ ., data = hourlydata)

Residuals:
    Min      1Q  Median      3Q     Max 
-19.568  -2.549   1.243   3.072  13.096 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    35.949980 187.703839   0.192    0.849
index           0.010656   0.009754   1.093    0.277
Max_Capacity   -3.264752   9.385320  -0.348    0.729
Avg_Cloud       0.015592   0.160278   0.097    0.923
Avg_DSWRF      -0.005773   0.046908  -0.123    0.902
Avg_Humidity    0.026594   0.410016   0.065    0.948
Avg_Temp        0.491902   1.144460   0.430    0.668
Temp_Cap        0.007158   0.030698   0.233    0.816
Cloud_Cap      -0.003678   0.008773  -0.419    0.676
DSWRF_Cap       0.001334   0.002205   0.605    0.547
Humidity_Cap   -0.003135   0.019085  -0.164    0.870
lag42           0.011505   0.093052   0.124    0.902
temp_lag2      -0.719497   1.010611  -0.712    0.478
Humidity_lag2  -0.001795   0.005606  -0.320    0.750

Residual standard error: 5.247 on 92 degrees of freedom
Multiple R-squared:  0.5704,	Adjusted R-squared:  0.5097 
F-statistic: 9.396 on 13 and 92 DF,  p-value: 4.656e-12
```

Figure 17: Results of hourly mode for 17.00

WMAPE = 0.2592808

By looking at the residuals and summary tables, we can say that:
-Most of the variables explain each other which we can understand by looking at the significance of the variables.
-Residuals are nearly normally distributed and there is no significant autocorrelation for any lag since we included the lag as a regressor to the model.
-This data is not strongly regressor dependent, just by building models for each hour, we can obtain accurate results without including many regressors since most of the time, production is equal to max capacity. Time series analysis characteristic of the data is much stronger than its linear regression characteristic.

4. Evaluation Phase

To test the four models and see how they forecast, we divided the model into a training test and test set. We used the days between March 1st and May 24th as proposed in the project description. We used a back testing approach for each model, we trained the model again and again for each day to forecast which is a real representation of how a forecast can be done. For example, to test the 25th of March, we used the dates between 1st of January and 23rd of March to train a model and we forecasted 25th of March with that date's regressors. This approach is logical in the sense that it learns each newly added data to forecast the newest observation and it does not include a data point to train a model to forecast that date. This prevents data from entering both the training set and test set which is the real case with predicting an unknown day.

For each model, earlier days have significantly large WMAPEs because we trained our models with days starting from 1st of January 2022 so for the earlier days, there is not much data to train a linear regression model. This is the main reason we also checked the only WMAPEs of the month May to make a conclusion since days of the month May have enough data to accurately model the case.

Table 1: WMAPE values for days starting from 1 Mar to 24 May

Day	Model 1	Model 2	Model 3	Model 4
1	0.3145	0.4500	0.5545	0.4332
2	0.5609	0.6354	0.7400	0.7986
3	0.7095	0.8779	1.8488	0.7351
4	0.2228	0.3017	0.1056	0.1474
5	0.3010	0.2843	0.2087	0.1663
6	1.5031	1.1857	2.2608	2.0372
7	0.4296	0.4124	0.3436	0.4155
8	0.2705	0.2354	0.1517	0.1321
9	0.3819	0.4634	0.4261	0.3645
10	0.2562	0.2797	0.4132	0.4937
11	0.3009	0.2330	0.5215	0.2927
12	0.3733	0.2389	0.1781	0.2641

13	0.6676	0.4219	0.3370	0.3571
14	0.2298	0.1146	0.1561	0.1304
15	0.2525	0.1822	0.1583	0.1897
16	0.3464	0.3715	0.2990	0.3092
17	0.2864	0.2694	0.3147	0.2318
18	0.3446	0.2323	0.2597	0.1991
19	0.3600	0.1513	0.1119	0.1168
20	0.2749	0.1035	0.0698	0.0906
21	0.2951	0.1712	0.2142	0.1996
22	0.4028	0.2252	0.2314	0.1137
23	0.2672	0.1684	0.1469	0.1027
24	0.3740	0.2824	0.3052	0.2421
25	0.1794	0.0895	0.0935	0.1458
26	0.3401	0.3842	0.2915	0.4188
27	0.1786	0.0876	0.1450	0.1308
28	0.1511	0.0830	0.1533	0.1005
29	1.0416	0.9832	0.9766	1.0209
30	0.1454	0.1447	0.2067	0.1093
31	0.1554	0.1549	0.1210	0.2051
32	0.0979	0.1515	0.3145	0.1322
33	0.0932	0.1272	0.1906	0.0909
34	0.1130	0.1528	0.1719	0.0518
35	0.1522	0.1363	0.1140	0.1760
36	0.3818	0.3687	0.3637	0.2211
37	0.1017	0.1056	0.1382	0.0858
38	0.1651	0.1555	0.1892	0.2267
39	0.2880	0.2685	0.2656	0.3382
40	0.1354	0.1780	0.1500	0.1351
41	0.1656	0.2353	0.1844	0.0991
42	0.2133	0.2459	0.2096	0.1748
43	0.1999	0.1095	0.1235	0.1419
44	0.1703	0.0888	0.0970	0.0965
45	0.1397	0.0915	0.0774	0.0946
46	0.1206	0.0751	0.0926	0.0926
47	0.1130	0.1315	0.1180	0.1571
48	1.3325	1.1973	1.1770	1.0656
49	1.5505	1.4168	1.0208	0.7299
50	0.1149	0.1521	0.1320	0.1198

51	0.2242	0.2309	0.2622	0.1988
52	0.1910	0.2774	0.2013	0.1415
53	0.0937	0.1215	0.1188	0.1196
54	0.0834	0.1034	0.1665	0.1260
55	0.0622	0.1014	0.1650	0.1012
56	0.0697	0.1125	0.1480	0.0834
57	0.0885	0.0982	0.1179	0.0711
58	0.1054	0.1899	0.1651	0.1375
59	0.0832	0.0747	0.0740	0.0531
60	0.0794	0.0877	0.0888	0.0571
61	0.3406	0.3160	0.2825	0.2479
62	0.1788	0.1806	0.1564	0.2057
63	0.4015	0.3542	0.3706	0.3100
64	0.2749	0.2754	0.2141	0.2284
65	0.1507	0.1651	0.1828	0.1583
66	0.4135	0.3195	0.3507	0.5066
67	0.3906	0.3952	0.4530	0.3485
68	0.2771	0.3191	0.3104	0.1859
69	0.1410	0.1347	0.1092	0.1758
70	0.1619	0.1587	0.1549	0.2279
71	0.1091	0.0907	0.0945	0.0976
72	0.5066	0.4687	0.4664	0.3739
73	0.1507	0.1319	0.1195	0.1445
74	0.1111	0.1243	0.1104	0.1111
75	0.5356	0.4606	0.5009	0.3266
76	0.6957	0.6372	0.6181	/80.4550
77	0.1321	0.1443	0.1173	0.1666
78	0.3430	0.2696	0.2798	0.2340
79	0.1023	0.1553	0.1292	0.1522
80	0.1997	0.2359	0.2100	0.2217
81	0.1340	0.1127	0.1069	0.1082
82	0.1029	0.0794	0.0903	0.0820
83	0.1465	0.1551	0.1841	0.2143
84	0.2141	0.2322	0.1502	0.1693
85	0.1200	0.0890	0.1177	0.0884
Avg	0.2939	0.2707	0.2925	0.2536
Avg of May	0.2589	0.2465	0.2401	0.2272

By comparing the WMAPEs of the different models, we can conclude that Model 4 gives the lowest WMAPE for both all the days and only the month of May. This is logical since each hour has different capacities and characteristics, and Model 4 builds different models for each hour of the day. (Model 4 is referring to 14 models which are the models for 14 hours of the day with significant production).

5. Conclusion:

In this project, the aim was to provide hourly solar power predictions of KIVANC 2 GES (Güneş Enerjisi Santrali) for the next day. We were given two data sets where one of them was containing related regressors, the other one containing production data for each hour between the dates 1st of January 2021 and 5th of June 2022. We have built several models to examine the relation between regressors and the production amounts. At every stage of building a new model, we have tried to improve models by considering the insights and information coming from the results of previous models.

In the first part, we have analyzed the structure of data and we have decided which part to use in the regression models. After that, we have thought about the logical relation between regressors and production amounts to create new regressors to be added. After building several models and comparing their WMAPEs with backtesting approach between dates 1st of March 2022 and 24th of May 2022, we decided a final model to use as follows:

- Exclude hours between 8 p.m. and 5 a.m since they are nearly always 0.
- Build different models for each hour since.
- Include average of each regressor since locations' effects on the model are very close to each other.
- Include max capacity and (Max Capacity)*(Average of Each Regressor) for each hour.
- Add lag 42 to the model since we can use the last 2 days before data to forecast tomorrow.
- Add lag 2 of temperature and humidity to the model.

This final model was the most successful one in terms of lowest WMAPE which is the model we also used to make a forecast on competition phase. These all models showed that the most important feature for this data set is max capacity and hour information since production reaches max capacity in most of the hours and regressors have a comparably small effect on the total production. Time series analysis models (ARIMA) give sufficiently good forecasts which also support the idea that regressors not too strongly significant.