**GROUP 21**

Hasan Alp YILDIZLAR 2018402201

Hayrettin İlbey GÜNGÖR 2018402183

Hamza KALE 2018402204

[06.07.2022]

**PROJECT REPORT**

“Solar Power Forecasting”

IE 360.01

I. INTRODUCTION 3

II. RELATED LITERATURE 3

III. APPROACH 3

IV. RESULTS 4

V. CONCLUSIONS AND FUTURE WORK 4

VI. CODE 8

# **Introduction**

1. **Problem Description**

To create a sustainable world, it is vital to use sustainable energy sources such as solar energy. In order to use solar energy with maximum benefit, it is necessary to evaluate different alternatives. In this project, the solar power plant named KIVANC 2 GES located in Mersin (between 36-37° north latitude and 33-35° east longitude) is inspected. In this study, the average daily amount of energy produced will be compared and the location where energy can be produced with maximum efficiency will be determined via forecasts.

1. **Descriptive Analysis of the Given Data**

The raw data were obtained from KIVANC 2 GES. The first of these data includes 4 different metrics (TEMP, REL\_HUMIDITY, DSWRF, CLOUD\_LOW\_LAYER; will be explained in detail in the 2nd part) for 9 different locations, and includes hour-based information between 2/1/2021 and 6/6/2022. In addition, as a second data, the daily energy production amount between these dates was from KIVANC 2 GES.

When the data is analyzed in general, one of the striking situations is that between 8.00 pm and 04.00 am, the energy production sees approximately 0 value and peak value between 10.00 am and 04.00 pm. This situation shows us that there is a daily seasonality. While it is quite normal to produce more energy during periods of increased sun exposure, it is quite normal for it not to produce energy when the sun is not present. In addition, it is seen that 4 different metrics affect the production amount in different ways. For example, it is seen that the negative effect of cloud amount on energy production is more than the effect of temperature factor on energy production.

1. **Summary of the Proposed Approach**

There are two datasets to be used in the project, one of which includes daily production quantity data, and the other includes regressors: cloud cover, downward shortwave radiation (DSWRF), humidity, and temperature for different locations. We begin by translating long-format regressor data o a large format. Then we combined the regressor dataset with the production dataset.

In this project, there are 2 basic methodologies that we have created to reach the real value of our estimations. The provincial model is a fairly simple linear model. It has 7 different parameters (trend, w\_hour, is\_night, av\_temp, av\_dswrf, av\_cloud, av\_hum). In our second modeling, it was obtained by using ARIMA and adding and subtracting different parameters by trial and error.

# **Related Literature**

TEMP: Temperature is a factor that affects efficiency. It can represent seasonality. Also, we know that high temperatures affect solar panels and decrease their efficiency.

REL\_HUMIDITY: Humidity is another factor that we take into consideration. This variable provides information on rainy or cloudy times that potentially reduce efficiency.

DSWRF: This metric is the short version of the downward shortwave radiation flux. It is known that it is closely related to the production level.

CLOUD\_LOW\_LAYER: The low cloud layer is another factor. We can explain it as total cloud cover data for the low-level types of clouds.

# **Approach**

1. **MODEL 1 (differ~trend+w\_hour+is\_night+av\_temp+ av\_dswrf,+av\_cloud+av\_hum)**

The first model is a simpler linear model and consists of a few basic parameters: trend, w\_hours, is\_night

* **Trend**: There seems to be a small trend that we think is due to global warming, so we added the trend to our model. In the light of this metric, we can calculate the linear increase in the amount of solar energy production.
* **is\_night**: This parameter allows us to distinguish between day and night so that we can easily exclude the 0 values at night in the solar energy production table. In this way, the values between ..am- .. pm is excluded.
* **w\_hour**: This metric provides the opportunity to integrate the day-time breakdown of the given data into the model and to differentiate on the basis of day and hour. Since the weather does not change very rapidly, there is no significant difference in energy production from a few days ago. Because of this situation, we used the lagged variables from a few days ago.
* **av\_temp**:This metric calculates the average of the temperature amounts for 9 grid points
* **av\_dswrf**:This metric calculates the average of the DSWRF values for 9 grid points
* **av\_cloud**:This metric calculates the average of the cloud rates for 9 grid points
* **av\_hum**:This metric calculates the average of the humidity amounts for 9 grid points

For our linear regression models, we first use only trend, is\_night, and w\_hour parameters for differences. Then we add the difference with production data from 24 hours ago which is lag\_24. Despite the fact that this model has the smallest WMAPE value which is 0.02552820758, this model is not a solid model in terms of fast-changing weather conditions because it just depends on the hour of the day and the data from 24 hours ago. There are not any variables about weather conditions.

Our second linear regression model uses trend, is\_night, w\_hour, av\_temp, av\_dswrf, av\_cloud, and av\_hum as inputs. Even though this model has a bigger WMAPE, it gives more reliable results when the weather is unstable.

Finally, we delete the w\_hour parameter from our previous model. Since the model of the temperature variables represent the hourly seasonality, w\_hourly was unnecessary. We can see it from the WMAPE value which did not a lot: 0.04667612068

1. **MODEL 2**

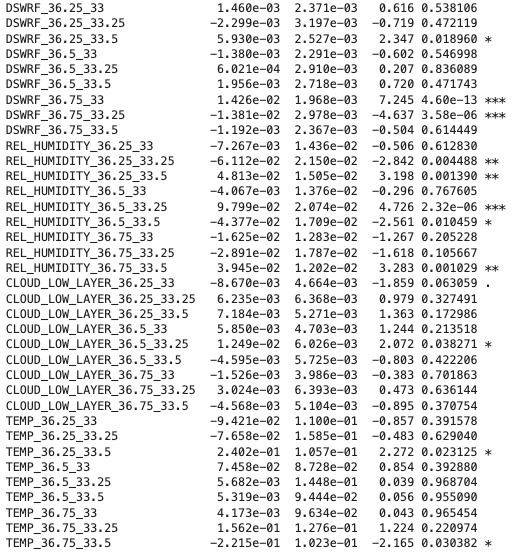
ARIMA was integrated into the model as the 1st model did not match the real values and made predictions with a large error rate. To better realize the data and forecast the future, the autoregressive integrated moving average model is added to the model.

In this way, autocorrelations and partial autocorrelations were studied to determine the lags. Thanks to this approach, the lags variable could be added to minimize forecast errors. Also, by plotting the ACF and PCF, all significant coefficients are determined and the patterns could be discovered. Also, the remaining patterns help us to add new AR and MA terms to improve the factuality of forecasting values.

While creating the first ARIMA included model, we use all variables for the regression matrix. The auto.arima function is used to determine p,d, and q values. The best p, d, and q values for our first ARIMA model are (1,0,1) according to auto.arima function. WMAPE value for this model is 0.04353538333.

In order to improve our ARIMA model, we deleted temperature variables and we have made progress in WMAPE which is 0.04040631568

* ARIMA with all variables
* ARIMA with all variables without temperature variables
* ARIMA with DSWRf variables and some relative humidity variables



**Figure 1: Choosing Crucial Parameters**

As seen in Figure 1, the effect of some parameters on the model is superior to other factors. While trying to optimize our model, we prioritized the more obvious parameters and ignored the others. For this reason, we excluded parameters that don’t have (\*\*\*) and (\*\*) degrees of importance in the model. As a result, except for 4 new parameters for REL\_HUMIDITY and 2 parameters for DSWRF, all other parameters were excluded from model 2.

Note: In the light of the model we created earlier, all parameters of DSWRF are included in the model because DSWRF parameters are of high importance for the model.

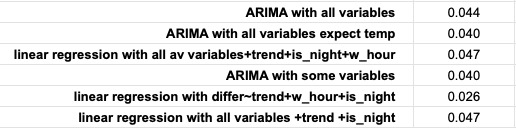
# **Results**

Provide your results and discussion.

The first model we created was weak in terms of algorithm and insufficient in terms of energy production prediction. The error rate for daily forecasts is really high. In this model, which does not include all the methods we explained in the "Approach" part, we realized that the action to be taken is to exclude metrics that will minimize irrelevant metrics and include metrics that increase the forecast success rate.

It can be said that our new model gives more realistic results with less ratio of error, together with the updated metrics and extracted parameters in model 2 we created.

# **Conclusions and Future Work**

In this project, 3 different approaches were applied for each of the 2 different models. The results of these models have tested with the "weighted mean absolute percentage deviation" method. 

**Figure 2: Weighted Mean Absolute Percentage Deviation Results**

"differ~trend+w\_hour+is\_night" gave more realistic results, but since no variable is added to that model, it cannot create a reliable model in rapid weather changes. For this reason, the "ARIMA with DSWRF and some Relhumidity" model was chosen as the model with the best performance.

1. **Code & Resources**
2. Github Link:
3. Kıvanç Enerji Website: https://www.kivancenerji.com.tr/kivanc2ges.html