# IE 360 Project Report
## Spring 2022

*Group 6*

**Ahmet Yiğit Doğan - 2018402105**
**Emre Burak Baş - 2018402096**
**Sercan Böçkün - 2019402123**

## 1. Table of Contents

## 2. Introduction

Sustainable energy resources are crucial for civilization to keep prevailing on Earth. After a century of developments majorly fueled by unrenewable and carbon based resources, investments to produce sustainable resources are boosting. Turkey located on the parallels of 36-42 is a relatively promising country in terms of its total sunlight exposure time during a year. “KIVANC 2 GES” is a facility located in Gülnar, Mersin is one of the major solar energy producers. Managing the total production, transporting the energy and legal regulations are factors these facilities are dealing with. In order to advance the planning of production, the facility desires to know the next day’s probable hourly productions in terms of “*MwH”.* The dataset provided by the company comprises the prior hourly energy productions starting from the date of “*02/01/2021*” till the “*24/05/2022*”. Given a date, the lastly updated productions will be the last hour data of 2 days ago. The predictions will be made for the dates of “*May 25-June 3, 2022*”.

The hourly production of energy from the panels is a multivariate problem. Direct exposure to sunlight is one of the major variables to fulfil the capacity of panels. Secondly the condition of panels such as the coverage of dust, condition of electronics and materials, legal restrictions, financial and logistical constraints etc. can majorly affect the production at any given hour and time. Since the presence of direct sunlight is desired weather forecasts for the prediction days would be quite logical to make use of. A comprehensive weather dataset is provided which consists of hourly cloud coverage, temperature, humidity and shortwave radiation flux.

The final weather data consists of 9 locations’ 4 weather features, date and hour which in total makes 38 columns and 11808 rows of data. Weather data and the production datasets are merged accordingly with the dates.

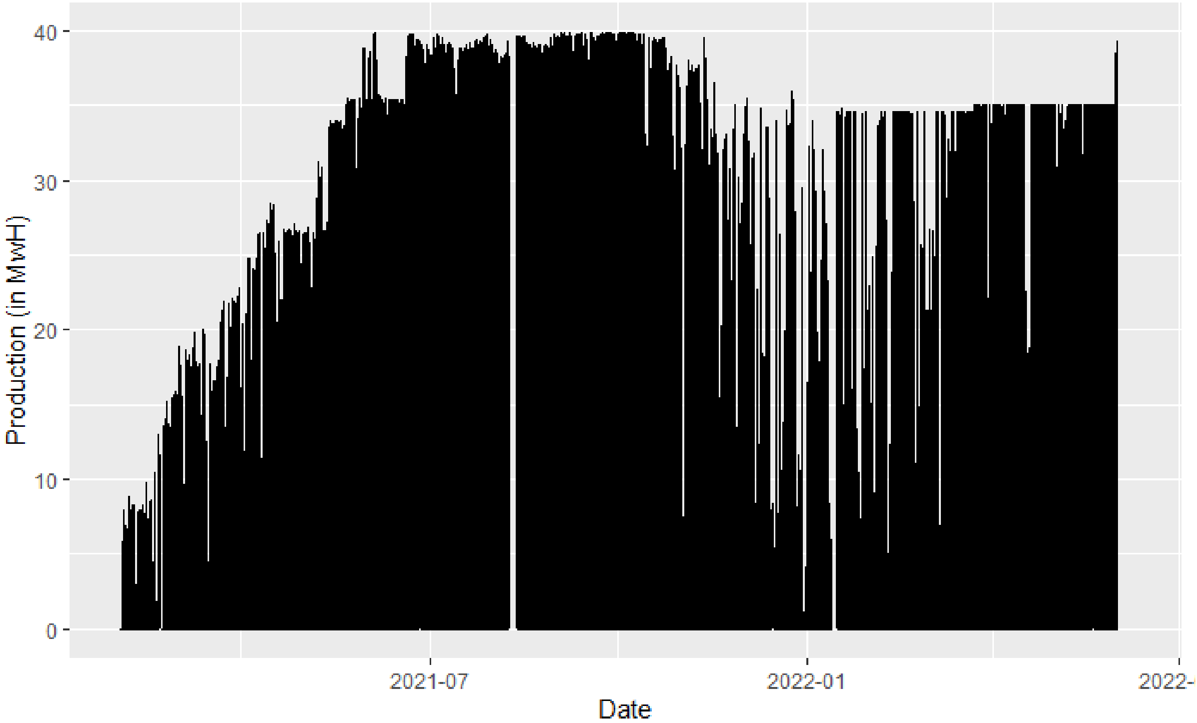

**Figure 1.** Hourly energy productions

The above figure shows the levels of energy production of the facility. One can easily detect the maximum production thresholds of 40 and 35 in different time intervals. Secondly the increasing trend of production at the first months of data indicates a capacity building period. These features of the data require manual approaches such as editing the predictions. Furthermore since Turkey is not located exactly at the equator line, the number of hours of sun exposure is different during a year. In winter the sun does not rise at 5 am in the morning while in summer it does. Similarly at 7 pm the sun has already set in winter times meanwhile the sun is still present at the same hour in summer times. These critical hours will be contemplated differently than the other hours. Night hours which in other words are the hours that sun is never present at will be discovered from the data and prediction inputs for these hours will be manually entered zero. Lastly the other hours will be fitted into the classic regression model with some factors. Detailed description of the models to be presented at the next chapter.

## 3. Modelling Approach

To obtain a basic understanding of the data set, production rates can be checked on an hourly basis, the below table displays the yearly average production rates of each hour.

**Table 1.** Average production rates (hourly, at two significance)

| Hour | Production | Hour | Production |
|---|---|---|---|
| 0 | 0,00 | 12 | 28,15 |
| 1 | 0,00 | 13 | 27,78 |
| 2 | 0,00 | 14 | 26,58 |
| 3 | 0,00 | 15 | 24,63 |
| 4 | 0,00 | 16 | 20,06 |
| 5 | 0,03 | 17 | 10,89 |
| 6 | 1,26 | 18 | 3,17 |
| 7 | 9,21 | 19 | 0,33 |
| 8 | 20,66 | 20 | 0,00 |
| 9 | 26,41 | 21 | 0,00 |
| 10 | 28,01 | 22 | 0,00 |
| 11 | 28,23 | 23 | 0,00 |

These average values imply an important characteristic of the data set, that is the suitability of production rates to be inspected in groups. Since the averages reveal three obviously separate groups of time periods, it is fair to implement a divide-and-conquer approach and break down the problem into sub-problems.

The first set of hours, from midnight to 4 AM and from 20:00 to midnight every day, has an average production rate of 0 at two significance, which means that no production, or negligible production, has been seen in these hours by now. Therefore, there is no need to create a forecasting model for this part, and it is fair to assign 0 to the next day's predictions for these time periods.

The second set is the regular hours where production average fluctuates between 9 and 20, which are expected yearly values for the prolific part of day. This period takes place between 7 AM and 5 PM.

The remaining part of the day consists of the hours that have average production rates below 9. At this point, the assumption is that these hours are affected by the season dramatically. To check the validity of this assumption visually, a yearly line chart for these specific hours can be drawn.

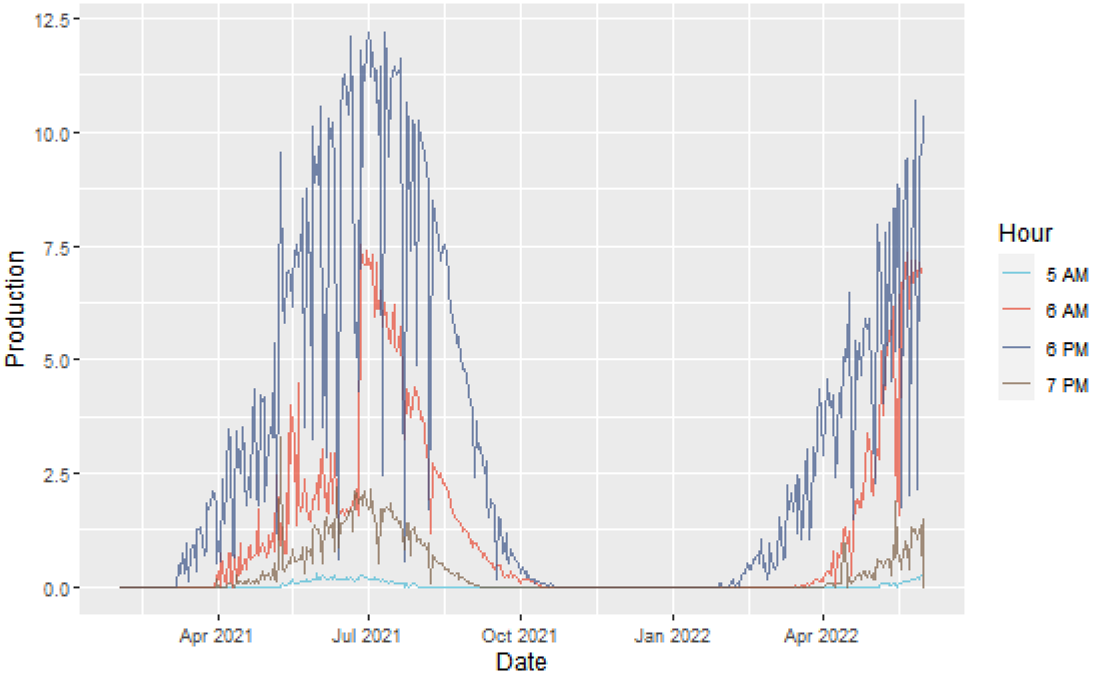

**Figure 2.** Production (in MwH) at critical hours

The strong seasonal behaviour of these hours can be seen in the plot. Actually, it is expected that production rates are significantly correlated with weather for every time period of the day. What separates these four hours from the regular ones is that they yield consecutive production rates of 0 at a certain period of the year. This observation will constitute the pillar of their models.

By some trial and error and heuristic approach, 6 PM is omitted from this list, since in the first days of the prediction period the regular hours model yielded more reasonable results for this hour. From this point on, 5 AM, 6 AM and 7 PM will be addressed as *"critical hours"*, and their models will employ different approaches compared to regular hours.

### a. Models for Critical Hours

In addition to the sharp transitions between 0-hours and critical hours, and also between critical hours and regular hours, in terms of production rates, significant increases and decays occur during intra-critical hour transitions. Thus, the problem can be splitted further by creating one model for each critical hour.

The common approach for all three critical hours is determining their “*active periods*”. What is meant by saying active periods is the part of the year in which at these hours production has been seen, and consequently, expected to be seen. To detect these periods, the months or weeks when the investigated critical hour have at least one non-zero production value are filtered. This operation has been done on a monthly basis for 5 AM, and on a weekly basis for 6 AM and 7 PM. To pick these time periods as basis, first and last non-zero observations of each critical hour in 2021 in Figure 2 is considered.

The outcome of this filtering is given to the linear model as a predictor. The predictor includes the following logic: If the corresponding row has a timestamp that lies in the active period, the column of the new predictor takes the value of its month number (week number for 5 AM). Otherwise, it takes the value “*No*”.

After adding the categorical “*is in nonzero period?*” variable, an extra 3 lag production column is inserted, which cleans the autocorrelations significantly. The final linear model of each critical hour includes all the raw weather variables and these two additional variables.

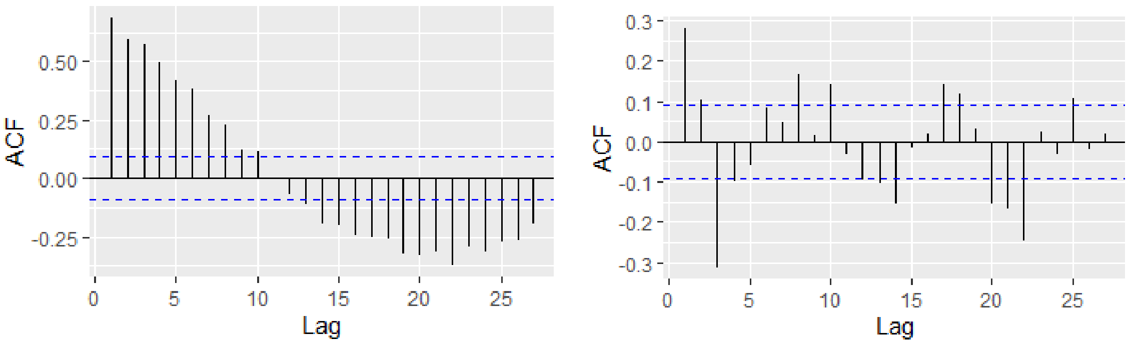

**Figure 3.** Left: Autocorrelation function of 5 AM model before adding lag 3 as a predictor, Right: ACF after the inclusion of lagged production values

### b. Model for Regular Hours

The hours from 6 AM to 7 PM will be modelled as a group. Considering the case of earth rotating around itself, the angle of sun rays is differing hour by hour in a given day. So even though the sun is present at both 1 PM and 5 PM, the production is not the same. This case will be handled by adding the hour data to the rows as a factor variable. Secondly, earth’s rotation around the sun and its tilted axis causes another phenomena which is the maximum peak the sun can reach in a day. In winter months the maximum angle of the sun is lower than the summer months. That’s also the reason for summer to be present in the north hemisphere while winter is present in the south hemisphere. This effect will be modelled by adding the month variable as a factor to the model. The residuals of this basic model constructed with weather, hour and month variables had an Adjusted $R^2$ score of 0.72. However the autocorrelation of residuals were very high at the lags of 12 (24 hours before).

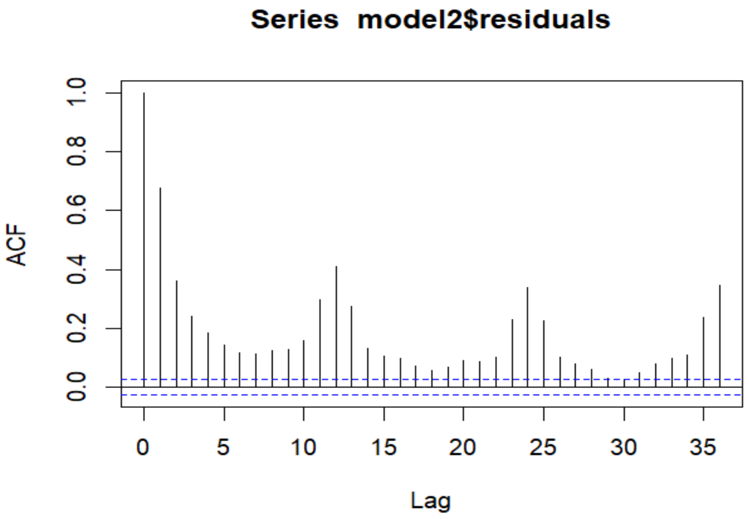

**Figure 4.** Autocorrelation of Residuals of the Basic Model

Using the most previous day’s available production data (36 lagged prod. data) of the same hour, both the performance of the model improved and the autocorrelation of the residuals reduced significantly. Looking at the past performances of the residuals it has been detected that the predictions for the 7 AM had very similar residuals. The predictions for that hour have been edited manually by summing the prediction with the most recent prediction’s residual for the same hour.

c. Predictions

As stated in the introduction section, due to regulations there is a maximum production threshold of 35. This behaviour has been consistently in play since the late 2021. In order to boost the prediction precision, outputs higher or slightly lower than 35 will be rounded to 35. This restriction is relaxed after encountering the first value above 35 on May 30. Secondly, because of intercept or other variables the model gives outputs lower or slightly higher than 0. If there’s a genuine belief that these values indicate none production, predictions will be rounded up or down to zero. Thirdly as mentioned before, at some hours the model is performing slightly poorer. Remember that there were lagged production data in variables. By making predictions for these lagged hours in the past and obtaining residuals, these residuals of the previous predictions are utilised in the current prediction process by summing them up with the model's prediction output.

### d. Results

Following the mentioned approaches the predictions and the real observations occurred as the following. Submitted predictions and real production data can be investigated at *Appendix 2*.

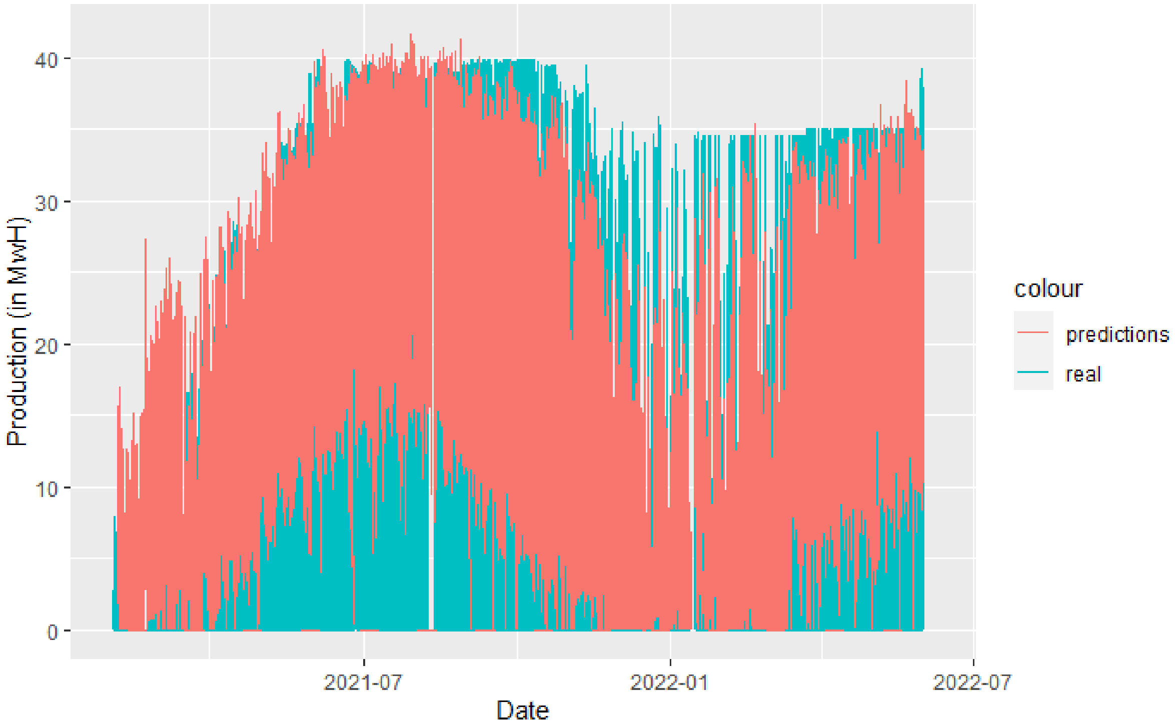

**Figure 5.** Predictions vs. Real Data

## 4. Conclusion and Future Work

Although the model performed well in the prediction period between May 25 and June 3, it can be enhanced by implementing more advanced analysis. First off, critical values are determined at the significance level 2. When the significance level is increased during the averaging of the production values on an hourly basis, it can be seen that an average production of 0.0013 is present for 8 PM, implying very rare nonzero production values have been recorded at that time. Therefore, the set of critical hours can be extended in a more elaborate work.

As an alternative way, the problem can be approached with a "time series analysis" viewpoint. After the data exploration, rolling mean, variance, and autocorrelation at lag 1 series and KPSS unit root test results show us the obvious non-stationarity of the production data both visually and statistically. Since the data given is an hourly production data, it is more statistically appropriate to use 24-lag differencing. This lag decision is seen to be performing well when the plot of the 24-lagged differences versus time index, rolling mean, variance and autocorrelation, and KPSS unit test root test results are examined.

When the “auto.arima()” function, provided by the “forecast” package is used, both with regressors (differenced weather data) and without regressors, the final model included 1 autoregressive term and 1 moving average term. No differencing is applied by the function, so it seems like the 24-lag differencing has been enough and no further differencing is necessary.

## 5. Code

One can access the R codes of the partial model by clicking here, and ARIMA model here.

## 6. Appendices

### a. Appendix 1: R codes for Plotting Figures

#### i. Code for Figure 1

```
ggplot(data, aes(x = datetime,y = production))      +
        geom_line()                                  +
        xlab("Date")                                 +
        ylab("Production (in MwH)")
```

#### ii. Code for Table 1

```
round(aggregate(data$production, list(data$hour), FUN = mean), 2)
```

#### iii. Code for Figure 2

```
critical_hours %>%
        ggplot( aes(    x       = datetime,
                        y       = production,
                        group   = hour,
                        color   = hour))                        +
        scale_color_manual(
                name = "Hour",
                labels = c(     "5 AM", "6 AM",
                                "6 PM", "7 PM"          ),
                values = c(     "#4DBBD5B2", "#E64B35B2",
                                "#3C5488B2", "#7E6148B2"))      +

        xlab("Date")                                            +
        ylab("Production")                                      +
        geom_line()
```

#### iv. Code for Figure 3

```
checkresiduals(model_five_am$residuals)
```

#### v. Code for Figure 4

```
checkresiduals(model_regular$residuals)
```

vi. Code for Figure 5

```
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 (in MwH)")
```

b. Appendix 2: Submitted Predictions vs. Real Data

May.25

| | Pred. | Real | | Pred | Real |
|---|---|---|---|---|---|
| 0 | 0,00 | 0,00 | 12 | 35,00 | 35,00 |
| 1 | 0,00 | 0,00 | 13 | 31,53 | 33,95 |
| 2 | 0,00 | 0,00 | 14 | 31,80 | 30,40 |
| 3 | 0,00 | 0,00 | 15 | 31,02 | 15,11 |
| 4 | 0,00 | 0,00 | 16 | 30,69 | 13,80 |
| 5 | 0,12 | 0,18 | 17 | 24,37 | 19,74 |
| 6 | 5,94 | 6,67 | 18 | 7,51 | 8,08 |
| 7 | 27,90 | 25,33 | 19 | 0,66 | 1,20 |
| 8 | 35,00 | 34,99 | 20 | 0,00 | 0,00 |
| 9 | 35,00 | 35,00 | 21 | 0,00 | 0,00 |
| 10 | 35,00 | 34,96 | 22 | 0,00 | 0,00 |
| 11 | 35,00 | 35,00 | 23 | 0,00 | 0,00 |

May.26

| | Pred. | Real | | Pred | Real |
|---|---|---|---|---|---|
| 0 | 0,00 | 0,00 | 12 | 35,00 | 35,00 |
| 1 | 0,00 | 0,00 | 13 | 33,75 | 35,00 |
| 2 | 0,00 | 0,00 | 14 | 32,31 | 35,00 |
| 3 | 0,00 | 0,00 | 15 | 28,78 | 33,58 |
| 4 | 0,00 | 0,00 | 16 | 27,40 | 30,36 |
| 5 | 0,12 | 0,18 | 17 | 16,58 | 25,63 |
| 6 | 6,70 | 7,19 | 18 | 5,43 | 10,68 |
| 7 | 27,29 | 26,54 | 19 | 1,28 | 1,25 |
| 8 | 34,46 | 35,00 | 20 | 0,00 | 0,00 |
| 9 | 35,00 | 35,00 | 21 | 0,00 | 0,00 |
| 10 | 35,00 | 35,00 | 22 | 0,00 | 0,00 |
| 11 | 35,00 | 35,00 | 23 | 0,00 | 0,00 |

May.27

| | Pred. | Real | | Pred | Real |
|---|---|---|---|---|---|
| 0 | 0,00 | 0,00 | 12 | 35,00 | 35,00 |
| 1 | 0,00 | 0,00 | 13 | 34,35 | 35,00 |
| 2 | 0,00 | 0,00 | 14 | 33,55 | 35,00 |
| 3 | 0,00 | 0,00 | 15 | 31,00 | 32,48 |
| 4 | 0,00 | 0,00 | 16 | 26,97 | 4,09 |
| 5 | 0,12 | 0,21 | 17 | 16,15 | 2,29 |
| 6 | 6,74 | 6,80 | 18 | 8,64 | 2,14 |
| 7 | 26,65 | 25,45 | 19 | 0,87 | 1,07 |
| 8 | 35,00 | 35,00 | 20 | 0,00 | 0,00 |
| 9 | 35,00 | 35,00 | 21 | 0,00 | 0,00 |
| 10 | 35,00 | 35,00 | 22 | 0,00 | 0,00 |
| 11 | 35,00 | 35,00 | 23 | 0,00 | 0,00 |

May.28

| | Pred. | Real | | Pred | Real |
|---|---|---|---|---|---|
| 0 | 0,00 | 0,00 | 12 | 35,00 | 35,00 |
| 1 | 0,00 | 0,00 | 13 | 32,43 | 35,00 |
| 2 | 0,00 | 0,00 | 14 | 26,77 | 35,00 |
| 3 | 0,00 | 0,00 | 15 | 19,54 | 32,60 |
| 4 | 0,00 | 0,00 | 16 | 19,66 | 26,20 |
| 5 | 0,12 | 0,15 | 17 | 19,72 | 20,97 |
| 6 | 5,25 | 6,99 | 18 | 9,50 | 9,47 |
| 7 | 25,51 | 26,05 | 19 | 0,88 | 1,23 |
| 8 | 35,00 | 35,00 | 20 | 0,00 | 0,00 |
| 9 | 35,00 | 35,00 | 21 | 0,00 | 0,00 |
| 10 | 35,00 | 35,00 | 22 | 0,00 | 0,00 |
| 11 | 35,00 | 35,00 | 23 | 0,00 | 0,00 |

May.29

| | Pred. | Real | | Pred | Real |
|---|---|---|---|---|---|
| 0 | 0,00 | 0,00 | 12 | 35,00 | 35,00 |
| 1 | 0,00 | 0,00 | 13 | 35,00 | 35,00 |
| 2 | 0,00 | 0,00 | 14 | 35,00 | 34,84 |
| 3 | 0,00 | 0,00 | 15 | 32,22 | 35,00 |
| 4 | 0,00 | 0,00 | 16 | 28,70 | 24,06 |
| 5 | 0,15 | 0,18 | 17 | 22,62 | 25,23 |
| 6 | 6,93 | 7,11 | 18 | 9,83 | 9,50 |
| 7 | 26,93 | 26,05 | 19 | 0,87 | 1,34 |
| 8 | 33,62 | 35,00 | 20 | 0,00 | 0,00 |
| 9 | 35,00 | 35,00 | 21 | 0,00 | 0,00 |
| 10 | 35,00 | 35,00 | 22 | 0,00 | 0,00 |
| 11 | 35,00 | 35,00 | 23 | 0,00 | 0,00 |

May.30

| | Pred. | Real | | Pred | Real |
|---|---|---|---|---|---|
| 0 | 0,00 | 0,00 | 12 | 35,00 | 38,12 |
| 1 | 0,00 | 0,00 | 13 | 35,00 | 37,86 |
| 2 | 0,00 | 0,00 | 14 | 35,00 | 37,95 |
| 3 | 0,00 | 0,00 | 15 | 32,96 | 37,92 |
| 4 | 0,00 | 0,00 | 16 | 28,38 | 37,22 |
| 5 | 1,75 | 0,25 | 17 | 20,00 | 24,94 |
| 6 | 7,32 | 6,90 | 18 | 10,29 | 10,01 |
| 7 | 25,05 | 24,70 | 19 | 1,18 | 0,00 |
| 8 | 35,00 | 35,00 | 20 | 0,00 | 0,00 |
| 9 | 35,00 | 34,96 | 21 | 0,00 | 0,00 |
| 10 | 35,00 | 36,60 | 22 | 0,00 | 0,00 |
| 11 | 35,00 | 38,49 | 23 | 0,00 | 0,00 |

May.31

| | Pred. | Real | | Pred | Real |
|---|---|---|---|---|---|
| 0 | 0,00 | 0,00 | 12 | 35,00 | 37,96 |
| 1 | 0,00 | 0,00 | 13 | 35,00 | 37,96 |
| 2 | 0,00 | 0,00 | 14 | 35,00 | 37,86 |
| 3 | 0,00 | 0,00 | 15 | 35,00 | 37,93 |
| 4 | 0,00 | 0,00 | 16 | 28,45 | 37,64 |
| 5 | 0,18 | 0,22 | 17 | 19,77 | 26,77 |
| 6 | 7,51 | 6,93 | 18 | 9,66 | 10,33 |
| 7 | 26,05 | 25,60 | 19 | 1,01 | 1,48 |
| 8 | 35,00 | 39,00 | 20 | 0,00 | 0,00 |
| 9 | 35,00 | 39,27 | 21 | 0,00 | 0,00 |
| 10 | 35,00 | 38,84 | 22 | 0,00 | 0,00 |
| 11 | 35,00 | 38,58 | 23 | 0,00 | 0,00 |

Jun.01

| | Pred. | Real | | Pred | Real |
|---|---|---|---|---|---|
| 0 | 0,00 | 0,00 | 12 | 35,00 | 38,23 |
| 1 | 0,00 | 0,00 | 13 | 35,00 | 38,21 |
| 2 | 0,00 | 0,00 | 14 | 35,00 | 38,20 |
| 3 | 0,00 | 0,00 | 15 | 35,00 | 29,98 |
| 4 | 0,00 | 0,00 | 16 | 28,28 | 37,92 |
| 5 | 0,20 | 0,25 | 17 | 25,88 | 25,07 |
| 6 | 7,84 | 7,09 | 18 | 9,05 | 10,13 |
| 7 | 26,80 | 25,92 | 19 | 1,45 | 1,39 |
| 8 | 35,00 | 39,15 | 20 | 0,00 | 0,00 |
| 9 | 35,00 | 39,53 | 21 | 0,00 | 0,00 |
| 10 | 35,00 | 39,15 | 22 | 0,00 | 0,00 |
| 11 | 35,00 | 38,74 | 23 | 0,00 | 0,00 |

Jun.02

|  | Pred. | Real |  | Pred | Real |
|---|---|---|---|---|---|
| 0 | 0,00 | 0,00 | 12 | 37,08 | 38,21 |
| 1 | 0,00 | 0,00 | 13 | 36,96 | 37,64 |
| 2 | 0,00 | 0,00 | 14 | 35,80 | 38,14 |
| 3 | 0,00 | 0,00 | 15 | 35,00 | 37,86 |
| 4 | 0,00 | 0,00 | 16 | 31,26 | 33,79 |
| 5 | 0,20 | 0,00 | 17 | 21,31 | 27,37 |
| 6 | 7,14 | 0,00 | 18 | 10,34 | 12,18 |
| 7 | 26,68 | 0,00 | 19 | 1,26 | 2,08 |
| 8 | 35,00 | 39,28 | 20 | 0,00 | 0,00 |
| 9 | 35,54 | 39,56 | 21 | 0,00 | 0,00 |
| 10 | 36,25 | 38,90 | 22 | 0,00 | 0,00 |
| 11 | 38,31 | 38,93 | 23 | 0,00 | 0,00 |

Jun.03

| | Pred. | Real | | Pred | Real |
|---|---|---|---|---|---|
| 0 | 0,00 | 0,00 | 12 | 37,34 | 38,90 |
| 1 | 0,00 | 0,00 | 13 | 36,81 | 34,96 |
| 2 | 0,00 | 0,00 | 14 | 36,28 | 37,42 |
| 3 | 0,00 | 0,00 | 15 | 35,92 | 38,71 |
| 4 | 0,00 | 0,00 | 16 | 35,30 | 38,48 |
| 5 | 0,18 | 0,21 | 17 | 26,90 | 28,03 |
| 6 | 7,83 | 7,75 | 18 | 10,02 | 10,96 |
| 7 | 26,05 | 27,28 | 19 | 1,44 | 1,25 |
| 8 | 38,82 | 39,34 | 20 | 0,00 | 0,00 |
| 9 | 39,49 | 39,53 | 21 | 0,00 | 0,00 |
| 10 | 38,21 | 39,53 | 22 | 0,00 | 0,00 |
| 11 | 37,99 | 39,30 | 23 | 0,00 | 0,00 |