Load Forecasting in The Context of Global Covid-19 Vaccination Using Facebook Prophet

Kevinaldo Barevan and Abdul Halim

Department of Electrical Engineering, Faculty of Engineering, Universitas Indonesia, Depok, Indonesia

INTRODUCTION

Electrical load forecasting is very important in operations and planning. Historically, electrical load forecasting has been the dominant application in the electric utility business, such as power generation, unit commitments, and allocation of electrical energy reserves. Load forecasting is divided into 3 categories based on the period of short-term load forecasting which is applied in the range of one hour to 1 week, medium-term load forecasting is usually applied for a week to one year, and long-term load forecasting for more than one year [1].

In forecasting electrical loads, there are several parameters that can be used as a reference in knowing changes in load trends. Weather and air temperature are usually the main factors. In addition, a holiday or an event can also be a factor because it is related to community activities that do not escape the use of electricity, for example, the Covid-19 incident. The current events of the Covid-19 pandemic have affected every aspect of life and dramatically changed our behavior patterns. For example, with over 32 million cases in the United States alone, the poorly controlled transmission of Covid-19 has challenged the capacity of healthcare systems and has resulted in more than 580,000 deaths. Until recently, nonpharmaceutical interventions (NPIs) such as physical distancing and wearing masks were the only effective means of controlling the spread of Covid-19. Continuous compliance with these NPIs is variable and difficult to achieve. [2].

Moreover, the economic, social, and health impacts of the COVID-19 pandemic on households and businesses are significant. Many countries around the world are issuing some level of restrictions to contain the spread of Covid-19. Although vaccines have begun to be distributed, the effect of distribution which is expected to restore the mobility of companies and the community is in fact still being discussed regarding its effect on changes in economic conditions. The economic costs of a pandemic can only be reduced through multilateral coordination that ensures equitable access to vaccines, tests, and therapies. Fair allocation of vaccines is the optimal solution from an economic perspective and maximizes global welfare [3].

Methods Facebook Prophet becomes important to evaluate, because this method is a procedure for predicting the data time-series based additive model where trend data can be customized with variable annualized seasonally, weekly, and daily, as well as the effect of the holiday/event. This model works very well with
time-series data which has a strong seasonal effect even if there are multiple seasons on the historical data [4]. In accordance with the data parameters of the Covid-19 event, namely the limitation of mobility and vaccination, it can be deemed appropriate and according to this data to determine its effect on the electrical load with the advantages of the features possessed by the Facebook Prophet.

This paper is structured as follows, Part II will explain the related work and information related to load forecasting. Part III describes the method that presents the formula used for load forecasting with the Facebook Prophet, and part IV summarizes the results of the work that has been done.

STUDY LITERATURE

In this section, some data and methods will be explained. These data will show the relationship between vaccination and mobility. From these two factors, the proposed method will determine the value of load forecasting by considering these factors.

Power Load Forecast

Power system operation and planning rely on forecasting electrical energy demand variables such as hourly load, peak load, and total energy. Based on time, forecasting is categorized into short-term, medium-term, and long-term. Short-term forecasting is where the power system operates on a daily basis. The implementation paradigm is unit commitment, hydrothermal coordination, optimal load flow, demand response, and others. Medium-term forecasts are mostly concerned with fuel import decisions and power unit maintenance scheduling. Long-term forecasting can be used in power system planning [5].

The value of the electrical load in the future must be forecasted so as to maintain the economic value and maintain good power quality. Electrical power load fluctuates naturally depending on several conditions, such as temperature, humidity, pressure, time of day, season, and other factors. These important input variables must be taken into account when designing a power load forecasting system. The electrical energy source control system will use the load forecasting that has been done to determine the number of working generator units. The electrical power produced must be in accordance with the power required to maintain the quality of electricity, such as frequency and voltage [6].

Load Forecasting Method

Another approach that has been carried out [7], is load forecasting using the medium and long term Facebook Prophet and Holt-Winters. Next, the impact of fuel costs is investigated based on five forecasting models to provide insight for policymakers. In addition, in research by doing forecasting even though the context of the research is on the side of the manufacturing industry [8], namely forecasting with the Prophet-SVR hybrid model. Prophet is used to predict seasonal fluctuations and determine SVR input variables, and SVR is used to capture nonlinear patterns. In this study, load forecasting was carried out for factors related to electrical load on weather conditions and in particular vaccine distribution. These factors can help the electrical power planning system so that the quality, reliability, and economic value of electric power can be maintained properly. The stages carried out in this research are data collection, data pre-processing, Facebook Prophet modeling process and data visualization.

Vaccination

Vaccination is seen as the most effective way to stop COVID-19 from spreading. Though the pharmacological efficacy of COVID-19 vaccines has been extensively studied, there is currently little information about the non-pharmaceutical aspects of vaccine inoculations [9]. Research [9] highlights the relevance of non-pharmaceutical COVID-19 treatments (e.g., travel restrictions), as premature removal of such measures could compromise vaccination effectiveness. Vaccination means to immunize a fraction of the population and subsequently remove the targeted population from a social network. Current vaccination strategies identify the targeted population primarily based on social relationships among individuals and can be characterized as contact-based strategies. Many studies have shown that vaccination should be targeted to those individuals who are socially close with the infected person, such as family members or co-workers of the infected person. Other studies have shown that vaccination is a priority for individuals with high social contact or for people with contacts from multiple social classes [10].

With all of these problems, vaccine distribution is expected to be able to restore the economy which affects the mobility of the community and has implications for the condition of electricity demand. The current distribution of vaccines is not a debate in terms of whether it benefits the economy or not because of the uneven distribution [2]. Several indicators of economic activity and mobility have started to improve. Apple Mobility Trend shows increased activity in the United States. Expectations of this improving trend of growth will be maintained as the post-pandemic recovery progresses. This is as shown in Figure 1 regarding the mobility of people in particular in Pennsylvania over time from January 13, 2020 to October 13, 2021 [11].

Figure 1. Pennsylvania Community Mobility

Figure 2 explores how vaccination coverage, efficacy, and delivery time affect mobility [12]. The picture explains the
increase in the cumulative number of cases. The response to Covid-19 is a change in habits (new normal) that accelerates the adoption of digital technology and many of these changes will persist in the long term [13]. Over the next few decades, digital technology is set to make energy systems around the world more connected, intelligent, efficient, reliable, and sustainable. Advances in data, analytics, and connectivity enable a wide range of applications of new digital technologies. Energy systems in the future are expected to be able to identify energy needs and deliver them at the right time, in the right place, and at the lowest cost and in accordance with current conditions [14].

![Cumulative cases](https://via.placeholder.com/150)

Figure 2. Effect of vaccination coverage, efficacy, and delivery on mobility

**Facebook Prophet**

Facebook Prophet is a model for forecasting time series data based on an additive model where non-linear trends according to the type of time series (t) are used starting from yearly, weekly, and daily, in addition, coupled with holiday effects or events. Here are the similarities from the Facebook Prophet.

\[
y(t) = g(t) + s(t) + h(t) + \epsilon(t)
\] (1)

where \(g(t)\) is the Growth Function (such as Linear Growth, Logistic Growth, and flat), \(s(t)\) is the Season effect (for example, weekly or yearly), \(h(t)\) is the Holiday effect, and \(\epsilon(t)\) is the Error of idiosyncratic change that is not accommodated by the model [15].

From equation (1) there are several mathematical functions of Facebook Prophet, one of which is seasonality. Seasonality itself estimates the effects of seasons with a Fourier series. The seasonality equation is shown in equation (2) below.

\[
s(t) = \sum_{n=1}^{N} (a_n \cos \left(\frac{2\pi nt}{p}\right) + b_n \sin \left(\frac{2\pi nt}{p}\right))
\] (2)

where value is a scalable Fourier series value or by default, being the expected regular period of the time series (e.g. \(p = 365.25\) for annual data or \(p = 7\) for weekly data, when scaling the time variable in days), and \& is Fourier series coefficients. The value of the Fourier series itself provides a flexible periodic effect model [15].

Meanwhile, holidays and events enter the list of holidays into the model by assuming that the effect of holidays is independent. Then, add an indicator function that represents whether the time in the data matches the time from the entered vacation list. Additionally, specifying each holiday with the \(k\) parameter. Overall the holiday equation is carried out in the same way as seasonality by generating a regressor matrix [12]. The holiday equation is shown in equations (3) and (4) below.

\[
h(t) = Z(t) k
\] (3)

\[
Z(t) = [1(t \in D_1), ..., 1(t \in D_\ell)]
\] (4)

with \(h(t)\) an effect on vacation (holiday), \(Z(t)\) is a set of value dates past and future for this type of vacation days are included.

**METHOD**

In this section, the methodology used in this research includes several stages. The following is a diagram for the process of this research. In this research process, the researcher uses Python.

![Load Forecasting Flowchart](https://via.placeholder.com/150)

Figure 3. Load Forecasting Flowchart

In this research, the first stage is to collect electrical load data and other parameters from official data sites. Then, process the data so that it is ready for the next process in machine learning. Finally, visualization and analysis are carried out on the results of data processing.

**Data Collection**

There are several types of data collection from different official websites, which include data on electricity loads, weather, mobility restrictions, united and vaccination data as shown in Table 1. The specifications of the parameters used are as follows.

**Electrical Load**

The electrical load used in this study is data from the American PJM [16] the United States, especially in the Pennsylvania area. In this area, the researcher specializes in using consumer data from the Metropolitan Edison (Met-Ed) company. The data
taken is electricity load data (MW) per hour from January 2017 to June 2021.

Weather

Weather in areas where Met-Ed is electrified includes Adams, York, Cumberland, Dauphin, Lebanon, Berks, Lehigh, Lancaster, Chester, Montgomery, Bucks, Northampton, Monroe, and Pike. From this area, data were obtained from the Harrisburg International Airport Station through the Weather Underground website [17].

Mobility Restrictions

Mobility restrictions were obtained from the Pennsylvania government website [18]. Then in this study, the division is divided into 3, namely normal conditions, lockdown, and re-opening.

Vaccination: vaccinations are compiled from the Centers for Disease Control and Prevention website [19] which refers to an article that compiles a global vaccination database. The data taken from this database is the number of vaccine doses per day.

Pre-Processing

There are several steps in pre-processing the data in this study, the first is data normalization.

Resample Data

Resample Data converts some parameters that are still in the form of daily averages to hourly values, such as weather, temperature, and vaccination data.

Encoding

Ordinal encoding of mobility restriction data with encoded labels. The following are the data variables used in this study.

Normalization

Normalization of data by using the maximum absolute scaling of the sklearn library where the data for each feature is scaled to its maximum absolute value.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Data Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electrical Load</td>
<td>As of 1 hour (2017, 2018, 2019, 2020, and 2021)</td>
<td>Each dataset consists of 1 year of training data and 6 months of test data</td>
</tr>
<tr>
<td>Temperature</td>
<td>F (degrees Fahrenheit)</td>
<td>Average day</td>
</tr>
<tr>
<td>Weather</td>
<td>Dew (°F), humidity (%), wind speed (mph), air pressure (Hg), and precipitation (inch)</td>
<td>Average value a day</td>
</tr>
<tr>
<td>Holiday</td>
<td>United States Holidays</td>
<td>US Federal Holiday Calendar from Pandas</td>
</tr>
<tr>
<td>Mobility Restrictions</td>
<td>Normal (0), Re-opening (1), and Lockdown (2)</td>
<td>Encode Labels</td>
</tr>
<tr>
<td>Vaccination</td>
<td>Vaccine dose per day</td>
<td>Average day</td>
</tr>
</tbody>
</table>

Modeling Facebook Prophet

In accordance with the Facebook Prophet’s formulation, the first step in modeling the Facebook Prophet is to import data that has gone through the pre-processing process, then share training data and test data. This data division consists of four datasets each dataset is divided into a data training one year and data test 6 months (approximately 33.33% of the data test) on each dataset. The first dataset is 1-year training data in 2017 and the first 6 months of test data in 2018. The second dataset is 1-year training data in 2018 and the first 6 months of test data in 2019. The third dataset is 1-year training data in 2019 and test data in the first 6 months of 2020. Finally, the fourth dataset is the 1-year training data in 2020 and the first 6 months of test data in 2021.

Then change the name of the date data column and the electrical load. Facebook Prophet expects certain input data, with column $x$ and column $y$ as the base model. This modeling is also carried out by setting the changepoint prior scale and Fourier order parameters as well as adding a regressor from the data variables of temperature, weather, mobility restrictions, and vaccinations and holidays. The holiday used is the US Federal Holiday Calendar from the Pandas library. Finally, in this research, only the fourth dataset, namely the 2020 training data, the 2021 test data, has the addition of the vaccination variable.

Model Evaluation

In evaluating this research model, only the mean absolute percentage error (MAPE) will be used. This type of error calculation is easy to understand because it produces a value in percentage. In addition, the percentage error is absolute, so negative error problems can be avoided. Overall, the smaller the MAPE, the better the forecast [20]. The MAPE equation is shown in equation (5)

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|$$

where $n$ is the sum of the total data, $A_t$ is the value of the actual data, and $F_t$ is the value of forecast or predicted data.

RESULTS AND DISCUSSION

In this study, there were several design variables that were changed, namely entering the input variable changepoint prior scale and the value of the Fourier order. The effect of Fourier order changes does not have a significant effect. Some experienced a decrease in error when the Fourier order value was low, but some other values actually experienced an increase in error even though the value was small. Changepoints on the Prophet are certain points that are optimal points for changing data trends. Meanwhile, the changepoint prior scale itself is used to determine the flexibility of the changepoint effect on the data. Even though in the testing process in the dataset with 2019 training data, the 2020 test data increased when the change point prior scale was getting bigger, but in general the flexibility of the changepoint reduces the error value when the value gets bigger. This is because the changepoint selection itself is done by the Prophet. In addition, the use of the holiday variable also does
not have a significant effect. There are some experiments that give better results, but some don’t.

From the visible results as shown in Table 1 and Table 2, in general, the percent error during training does not have a significant difference even though the error value is lower when the changepoint value of the prior scale is higher. The error results for the dataset with the 2019 test 2020 training data and the dataset with the 2020 test 2021 are much smaller. It is possible that there is an influence from the load profile conditions, mobility restrictions, and the addition of vaccination data parameters. Overall, the high error in the error test in the dataset training 2017 test 2018 and datasets training 2018 test in 2019 could be due to the data training a lot less.

Table 2. Error Load Forecasting Without Holiday Parameters

<table>
<thead>
<tr>
<th>Dataset</th>
<th>No Holiday MAPE (%)</th>
<th>Fourier Order</th>
<th>Changepoint prior scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>Test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jan-Dec 2017</td>
<td>Jan-Jun 2018</td>
<td>9.88</td>
<td>35.87</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9.80</td>
<td>34.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.21</td>
<td>60.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.08</td>
<td>59.18</td>
</tr>
<tr>
<td>Jan-Dec 2018</td>
<td>Jan-Jun 2019</td>
<td>9.77</td>
<td>45.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9.60</td>
<td>42.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.92</td>
<td>26.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.80</td>
<td>26.60</td>
</tr>
<tr>
<td>Jan-Dec 2019</td>
<td>Jan-Jun 2020</td>
<td>10.64</td>
<td>13.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10.53</td>
<td>14.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.41</td>
<td>22.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.31</td>
<td>22.59</td>
</tr>
<tr>
<td>Jan-Dec 2020</td>
<td>Jan-Jun 2021</td>
<td>10.93</td>
<td>15.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10.86</td>
<td>15.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.54</td>
<td>27.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.39</td>
<td>27.03</td>
</tr>
</tbody>
</table>

Furthermore, in the visualization of the results of the processing by the Prophet himself as shown in Figure 4. The x-axis is the date data (ds) and the y-axis is the normalized electrical load data. Where the distribution of data shown in black is training data, the distribution of data shown in green is test data, and blue is the distribution of Prophet data. Of course, the more appropriate the trend from the graph seen in the distribution of blue data with green data, the accuracy of the Prophet will be higher. The following is an example of a visualization of the fourth dataset with 2020 training data and 2021 test data.

Table 3. Error Load Forecasting With Holiday Parameters

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Holiday MAPE (%)</th>
<th>Fourier Order</th>
<th>Changepoint prior scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>Test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jan-Dec 2017</td>
<td>Jan-Jun 2018</td>
<td>9.76</td>
<td>40.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9.72</td>
<td>41.88</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.12</td>
<td>68.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.00</td>
<td>68.98</td>
</tr>
<tr>
<td>Jan-Dec 2018</td>
<td>Jan-Jun 2019</td>
<td>9.78</td>
<td>45.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9.59</td>
<td>42.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.92</td>
<td>25.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.79</td>
<td>26.63</td>
</tr>
<tr>
<td>Jan-Dec 2019</td>
<td>Jan-Jun 2020</td>
<td>10.63</td>
<td>14.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10.53</td>
<td>14.37</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.35</td>
<td>24.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.26</td>
<td>24.28</td>
</tr>
<tr>
<td>Jan-Dec 2020</td>
<td>Jan-Jun 2021</td>
<td>10.89</td>
<td>15.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10.82</td>
<td>14.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.50</td>
<td>27.55</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.35</td>
<td>27.31</td>
</tr>
</tbody>
</table>

From the overall data visualization, Facebook Prophet also visualizes each trend component which can be seen in Figure 5 describes the trend of the overall test data, holiday trends, weekly trends, daily trends, and the addition of regressor variable values. Overall, the trend of electrical loads continues to rise. In the daily trend, the trend has a type of load profile wherein the afternoon to evening it is the largest load of the day. On the weekly trend, on weekdays the electricity load starts to rise and tends to be constant. While the value of holidays is determined by the exact date on each of these holidays which will affect the prediction process by the Facebook Prophet model.
CONCLUSIONS

From the results of the research conducted, the overall process already has a quite small error for training and test data (in the case of Covid-19 with vaccination in 2021, the MAPE value is around 15%). datasets with the effects of the Covid-19 pandemic and vaccination have a much lower value than MAPE results with datasets in normal conditions. This can refer to differences in the parameters used, namely mobility restrictions, and the addition of vaccination data parameters. Research [4] with the historical dataset from January 2015 to May 2020 has MAPE 1.76% for triple seasonality Holt–Winters model, followed by the double seasonality Holt–Winters model with MAPE 1.83%, and Prophet model with multiple regressors with MAPE 2.77%. In research [5], Prophet-SVR has a 9.58% MAPE with a dataset from 2011 to 2019. Compare to the other research, the amount of data used could possibly affect the forecasting process.

REFERENCES


AUTHOR(S) BIOGRAPHY

Kevinaldo Barevan
He earned a Bachelor of Engineering from Universitas Jember (UNEJ) in 2019. Then he continues his study as a postgraduate student of Electrical Engineering with a specialization in Data Engineering and Business Intelligence from Universitas Indonesia.