Applied of feed-forward neural network and facebook prophet model for train passengers forecasting

Resa Septiani Pontoh¹, S Zahroh¹, H R Nurahman¹, R I Aprillion¹, A Ramdani¹, D I Akmal¹
¹Department of Statistics, Universitas Padjadjaran, Indonesia

Corresponding email: resa.septiani@unpad.ac.id

Abstract. Train is public transportation that is widely used by people in Indonesia. Due to the high level of comfort at a low price relatively. Based on released data by PT KAI, the number of train passengers has increased in almost every holiday season, thereby, it is suspected that there are seasonal patterns with fixed and random periods. In 2013, the government issued a policy related to infrastructure development, then, it caused the number of passengers significantly increasing in the following year. Thus, we need a model that can accommodate these patterns to forecast the number of train passengers accurately. The use of neural network methods such as Feed Forward Neural Network (FFNN), nowadays, becomes popular in facing big data including unexpected fluctuation on the data. Additionally, recently, Facebook announced an accurate method of forecasting, called Prophet model, for data which have trend, seasonality, holidays, and missing data. Hence, the forecast for monthly train passengers on this research is modelled by FFNN and Facebook Prophet. The result shows that Prophet performs better than FFNN. However, the difference in the value of MAPE is not too large.

1. Introduction
Public transportation is a type of transportation provided by the government to travel together with a predetermined route and time. The poorly coordinated system of several types of public transportation causes public transportation to be less attractive to the communities. Some problems faced by people in Indonesia are unscheduled arrival and departure times, unstable price, unclear routes or stops, and so on. Public transportation, such as buses and trains, is usually more popular for the public because it has fixed schedules and prices. However, the use of trains, nowadays, becomes popular than other kinds of public transportation since trains have their own paths or railways so that users do not feel any impact of congestion that often occurs in several major cities in Indonesia.

The train is commonly used by people in Indonesia, especially in Java. Its high level of comforts and having relatively low prices are factors driving the high number of train users. In addition, there are several types of trains in Indonesia followed by several types of classes. Based on data released by PT Kereta Api Indonesia (KAI) (Persero), the number of train passengers increases when entering the holiday season, both during religious holidays and school holidays. Related to this, in 2013, the government issued a policy related to infrastructure development. PT KAI increased the departure schedules and the number of carriages to defeat the issue of expanding train passengers. Therefore, this obviously brought the expanding pattern of train passengers.

Based on the previous information, it indicates there are trends and seasonal patterns in the data of the number of passengers followed by fixed or random periods.
Figure 1 shows that the data patterns of the monthly number of train passengers demonstrate a seasonal pattern each year and there is also a trend occurred since 2013. Additionally, some outliers occur due to the flood in passengers during the holiday season. Over-capacity train services will make users disappointed and switch to other modes of transportation while also making train service providers lose potential income from potential passengers who are not accommodated. On the other hand, under-capacity will make railroad service providers bear the additional burden due to unloaded carriages [1].

The surge in train passengers that occur during the holiday season causes train tickets always to be sold out before departure, accordingly, we need a forecasting method to find out the number of train passengers so the number of carriages and departure schedules that can be adjusted according to the number of passengers.

The number of train passenger data is included in the type of time series data derived from non-linear processes, therefore, the forecast model commonly used, as Autoregressive Integrated Moving Average (ARIMA) cannot provide a good model for obtaining accurate forecasting results in the non-linear time series data. For non-linear data, currently, ARIMA is not appropriate. Predicted results do not follow the actual data patterns and tend to have constant or flat results. The neural network model is one of the forecasting methods that supported mathematical models which will allow complicated non-linear relationships [2]. Its ability to face unexpected fluctuation has made the NN, to be phenomenal. Feed Forward Neural Network (FFNN) is the most well-known neural network model for forecasting [3], it typically consists of three layers of the neuron, those are input layer, hidden layer, and output layer. All information proliferates alongside the associations toward the path from the system contributions to the system yields, subsequently the term feed-forward. Several sectors have applied this method, including daily river flow [4], lynx data [5], electricity price [6], and passenger flows on metro lines [7].

Another popular method for forecasting is Facebook Prophet composed by Facebook’s Data Science Team supposed to own balanced instinctive parameters while not knowing the main points of the fundamental model. Prophet has three fundamental model segments. It takes into account the patterns of trend, seasonal, and holidays [8]. The outcome of the analyst-in-the-loop approach that humans and machines work automatically. The Prophet begins by demonstrating a time series using indicated parameters analyzing, producing forecasts, and evaluating them. Once the poor performance is recognized or a problem happens, Prophet surfaces these problems to the analyst to change them to understand what turned out badly and how to adjust the model depends on the criticism. Prophet is optimized for the business forecast, however, they claimed that Prophet makes it much straightforward to form an affordable and accurate forecast. Prophet has been applied outside the business forecast, such as air pollution [9], bitcoin [10], and website traffic [11] forecast. Hence, the forecast for the number of
train passengers will be used in this model because the patterns of train passengers are following the characteristics of the Prophet's model itself.

This research aims to obtain forecasting models for train passengers in the Java island region using the FFNN and Facebook Prophet Model. It is expected that those models might provide high accuracy so it can be used to predict the number of train passengers in the next periods because it refers to [1], effective capacity management is the key to the success of railroad service providers although it is not easy. The method to be used in this research, Facebook Prophet Model, as a novelty in the number of train passengers forecast.

2. Research Methodology
2.1. Data Sources and Research Variables
The dataset that will be employed is the data of train passengers collected monthly in the Java Island region from January 2006 to August 2019. Data are obtained from the official website of the Central Statistics Agency (BPS). As explained earlier, data on the number of train passengers contains a seasonal pattern with a fixed and random period that is influenced by holiday factors. Furthermore, there is a pattern of trends caused by changes in government policy. The method that will accommodate these patterns is the Facebook Prophet Model.

2.2. Method
2.2.1. Feed Forward Neural Network (FFNN). Artificial Neural Network was first introduced by McCulloch and Pits in 1943 [12]. ANN is a partner science framework that works in a similar technique with natural neural systems that are accepted to be profoundly exact [13]. Afterwards, [14] focused on machineability in perceptrons for single-layer feed-forward systems. FFNN is the most prestigious neural system model for utilization of time arrangement forecast. It thinks about a solitary concealed layer with a few hubs. A sum of repeats frameworks are fitted, each with starting loads that showed up at the midpoint of handling figures. The yields inform forecast for its future qualities, while the input nodes are the previous lagged observations. Hidden nodes with material non-direct exchange capacities process information conveyed by the input nodes.

A neural network will have several layers and units in each layer. Figure 2 shows the structure of the FFNN. This network has consisted of four units among the input layer (layer A), three units among the hidden layer (layer B), and a unit among the output layer (layer C). If a unit is among the input layer, it is an equivalent kind of inputs. If a unit is in consecutive layers, it is a constant kind of inputs as a result of the range of units in the previous layer. Each connection between input-to-unit and unit-to-unit is modified by weight. Also, each unit has an additional input that a relentless value of one is assumed. The weight that modifies this additional input is termed the bias.
All of the information propagates to the connections in the direction from the network inputs to the network outputs, hence the term feed-forward. Figure 3 shows an associate model of a unit in which weights and biases, likewise as all various network connections, are neglected for reasons of clearness. Once the network is running, each node in the hidden layer execute the computation in Eq. 1 on inputs and relocate the result \( O_c \) to the units in the consecutive layer.

\[
O_c = h_{\text{hidden}}(\sum_{p=1}^{P} i_{c,p} w_{c,p} + b_c) \quad \text{wherever} \quad h_{\text{hidden}}(x) = \frac{1}{1+e^{-x}}.
\]  

Eq. 1 is activation function of hidden nodes, wherever \( O_c \) is the output of the presently hidden layer node \( c \), \( P \) is the number of units among the antecedent hidden layer or the number of network inputs, \( i_{c,p} \) is input to unit \( c \) of antecedent hidden layer node \( p \) or the network input \( p \), \( w_{c,p} \) is the weight that modifies the link from input \( p \) to node \( c \), and \( b_c \) is the bias. In Eq. 1, \( h_{\text{hidden}}(x) \) is the sigmoid activation function. To avoid saturation of the activation function, the training data should be scaled consequently. In the same way, weight and bias are initialized to fittingly scaled data before training. Each output layer unit carries out the calculation in Eq. 2 on its inputs and transmits the result \( O_c \) to a network output.

\[
O_c = h_{\text{output}}(\sum_{p=1}^{P} i_{c,p} w_{c,p} + b_c) \quad \text{wherever} \quad h_{\text{output}}(x) = x.
\]  

Eq. 2 is the activation function of associate nodes in the output layer, wherever \( O_c \) is the output of the recent output nodes \( c \), \( P \) is the variety of nodes in the antecedent hidden layer, \( i_{c,p} \) is input to node \( c \) from the previously hidden nodes \( p \), \( w_{c,p} \) is the weight that modifies the link from node \( p \) to node \( c \), and \( b_c \) is the bias \( h_{\text{output}}(x) \) is a linear activation function.

2.2.2. Facebook Prophet Model. Prophet is created by Facebook's Data Science Team in 2017 for making business conjectures [8]. Tune the parameters, as a rule, are exhausting for completely programmed forecast systems and to combine helpful suspicions or heuristics is in some cases excessively burdensome. A period arrangement model, nonetheless, configurable by non-specialists who may have little information concerning time arrangement models and have space information concerning the data-producing process.

![Figure 4](image-url). Schematic view of the analyst-in-the-loop approach to prediction at scale (Source: [8])

As we can see in Figure 4, analyst-in-the-loop utilizing a flexible specification by demonstrating the time arrangement that incorporates a direct human understanding. At that point demonstrate estimates for this model and evaluate gauge execution. Once there are issues that warrant human mediation or terrible showing, the expert can analyze the forecast and without a doubt changed the model.

The Prophet model uses an analyzable time arrangement model [16] with three primary model segments: trend, seasonality, and holidays.

\[
y(t) = g(t) + s(t) + h(t) + \epsilon_t.
\]
where $g(t)$ is the trend function that models non-periodic changes, $s(t)$ is the seasonality that represents periodic changes (weekly and yearly), $h(t)$ are the impacts of the excursion that happen on in all likelihood sporadic calendars, and $\epsilon_t$ is an error term which is not defined by the model. Through time as a regressor, Prophet is attempting to suit numerous linear and nonlinear elements of time as parts [11].

Two trend models that comprise Facebook applications: a non-linear saturating growth model and a piecewise linear model. A nonlinear model is usually modelled utilizing the logistic growth model, that in its simplest type is

$$g(t) = \frac{C}{1 + \exp(-k(t-m))}, \quad (4)$$

where $C$ is the carrying capability, $k$ is the rate of growth, $m$ is an offset parameter. Once the rate $k$ is adjusted, the offset parameter should even be adjusted to attach the endpoints of segments. The piecewise logistic growth model is then:

$$g(t) = \frac{C(t)}{1 + \exp(-(k + a(t)^T \delta_j)(t - (m + a(t)^T \gamma)))}, \quad (5)$$

where $\delta$ and $\gamma$ could be a vector rate adjustment defines the modification inside the rate that happens at the time $s_j$. The change points because of a development, which ends up within the rate of growth can be modified and so the trend model is

$$g(t) = (k + a(t)^T \delta_j)\gamma(t - (m + a(t)^T \gamma)), \quad (6)$$

where $k$ is the rate of growth, $m$ is an offset parameter, $\delta_j$ is the rate adjustment, and $\gamma$ is set to $-s_j \delta_j$ to create the function continuously. In automatic change points choice, $\delta_j \sim Laplace(0, \tau)$.

To fit the projected model with seasonality effects and forecast supported it, it uses a Fourier series that provides a versatile model. Seasonal effects may be portrayed as within the following equation:

$$s(t) = \sum_{n=1}^{N} a_n \cos \left( \frac{2\pi nt}{P} \right) + b_n \sin \left( \frac{2\pi nt}{P} \right), \quad (7)$$

where $P$ could be a regular amount.

Holidays and events usually are not a periodic pattern, thereby their impacts do not seem to be well modelled by a sleek cycle. Prophet permits the analyst to produce past and future custom list events. A window around such days is taken into account individually and further parameters are fitted to model the effect of holidays and events.

2.2.3. Model Evaluation. The model accustomed forecast is tested to urge the predicted accuracy value of the model. As a rule, the selection of simplest model criteria frequently utilized is Mean Absolute Percentage Error (MAPE) in light of the fact that the worth is as a rate with the goal that it is fitting to be acquainted with measure the exactness of a model. A model becomes the best model if it has a small MAPE value. MAPE is used to measure the distinction between the value of the forecast results with the actual value as a percentage with the subsequent equation:

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{X_t - \hat{X}_t}{X_t} \right| \times 100\%, \quad (8)$$

where $n$ is the number of observations, $X_t$ is the observed value, and $\hat{X}_t$ is the predicted value. There is a MAPE standard category for forecasting evaluation [17]. These categories refer to Table 1.

| Table 1. MAPE standard category for forecasting evaluation |
|-----------------|-----------------|
| MAPE            | Forecasting Criterion |

5
| Accuracy of forecasting | \(<10\%\) | Accuracy of forecasting is very good |
|-------------------------|---------|----------------------------------|
|                        | \(10-20\%\) | Accuracy of forecasting is good |
|                        | \(20-50\%\) | Accuracy of forecasting is enough good |
|                        | \(>50\%\) | Accuracy of forecasting is not good |

### 3. Result and Discussion

In forecasting, the first important step is to overview the patterns of the data. Based on Figure 1, the data patterns have a changing trend that appeared in 2013 and also 2018. Moreover, the data shows a monthly seasonal pattern per year. In this research, forecasting the number of train passengers in the Java Island region presents two different models, the Feed Forward Neural Network (FFNN) and Facebook Prophet Model. One of the advantages of FFNN is its ability to learn the data pattern while doing the training process so that the error could be minimized. On the other hand, Prophet is able to detect changes automatically in the trend and seasonal patterns.

Descriptive statistical analysis is firstly done before doing modelling using the Prophet model. The purpose of descriptive statistical analysis in this research is to have preliminary knowledge of the selected variables. The descriptive statistics of the monthly train passengers in the Java Island region are summarized in Table 2.

**Table 2. Descriptive Analysis Results**

| Statistic | Passengers (people) |
|-----------|---------------------|
| count     | 166                 |
| mean      | 43459830            |
| std       | 15400460            |
| min       | 21518000            |
| max       | 76606000            |

Based on the descriptive data analysis from January 2006 to October 2019 result in Table 2, the minimum and maximum value of 166 monthly data of train passengers are 21518000 and 76606000 people. The averages of the monthly train passengers are 43459830 people with a standard deviation of 15400460 people, where it can be said that the data have high variation.

Firstly, in FFNN, parameter initialization was carried out. In this research, we applied one hidden layer feed-forward networks and NNAR \((p,k)\) notation to inform, in the hidden layer, there are \(p\) lagged inputs and \(k\) nodes. The number of networks to fit with different random starting weights is set to be 10. These are then averaged when producing forecasts with the maximum iteration is set to be 150. The best model obtained from an average of 10 networks is NNAR\((13,7)\) with 106 weights and \(\sigma^2\) estimated as 5296661.

By this model, we can iteratively simulate future values. The distribution for all future values is built based on the fitted neural network. Figure 5 is a 5-possible future value simulation of the number of train passengers and each sample accommodates the next 12 months after the observed data. Therefore, by used the best model, which performed by software and produces a Mean Absolute Percentage Error (MAPE) value of 4.27%.
On the other hand, in the process of analysis using Prophet, the column that showing time must be named "ds" and the variable column to be predicted must be named "y". The process is performed by software and produces a MAPE value of 3.36%.

| Model   | MAPE  |
|---------|-------|
| FFNN    | 4.27% |
| Prophet | 3.36% |

Hereafter, Prophet model will be used to forecasts the number of train passengers in the Java Island region. Plots showing actual and predictive data and forecasting data for the next 12 months from November 2019 to October 2020 by the Prophet model can be seen in Figure 6.

Figure 6 mentions that the Prophet model can forecast train passengers follow the patterns of the actual data, although there are outliers in the data. For the future predict, the forecasts produced by the
Prophet model fluctuates following the previous seasonal pattern. This proves that the Prophet model might learn all the patterns, both trends and seasonal or holidays, not only last observation. The point forecast, lower limits, and upper limit with a significance 5% shown in Table 3.

Table 4. Predictive value of the train passengers (thousand people)

| Month | Point Forecast | Lower Limit | Upper Limit |
|-------|----------------|-------------|-------------|
| Nov-19| 76356.72       | 73837.98    | 78572.06    |
| Dec-19| 73198.02       | 70723.58    | 75665.92    |
| Jan-20| 74996.43       | 72724.38    | 77095.99    |
| Feb-20| 72533.38       | 70260.44    | 75038.24    |
| Mar-20| 70041.55       | 67881.78    | 72408.14    |
| Apr-20| 77833.96       | 75474.44    | 80117.27    |
| May-20| 74035.61       | 71696.02    | 76586.9     |
| Jun-20| 79847.69       | 77431.78    | 82455.44    |
| Jul-20| 77292.16       | 75058.01    | 79683.37    |
| Aug-20| 71114.41       | 68840.16    | 73497.46    |
| Sep-20| 82363.39       | 79891.38    | 84485.32    |
| Oct-20| 75021.45       | 72603.45    | 77543.68    |

Based on Table 4, we can see that the values of forecast are not around its average value, but rather fluctuating around its maximum value. This is because the data on the number of train passengers contains a positive trend so that its value continues to increase significantly each year.

4. Conclusion

Based on the analysis that has been done to predict the number of train passengers in the Java Island region we can conclude that the Facebook Prophet Model performs better than FFNN. MAPE of 4.27% is obtained for the FFNN and 3.36% for the Facebook Prophet Model. Consequently, it can be said that both models have a high degree of accuracy because of its fairly small prediction error rate. The number of train passengers for the next 12 months by the Prophet model can be seen in Table 4 and the plot shows in Figure 6. In Figure 6, it can be seen that the Prophet model might produce a value that fluctuated following the previous data pattern. Additionally, several studies split data into training and testing data before modeling of the Prophet model. It can be applied for further research if is use a lot of data dataset.

The method used is a mathematical model which is an alternative to the forecasting method commonly used, such as the Autoregressive Integrated Moving Average (ARIMA), that does not have any parameters. Moreover, for further research, it is recommended to be able to compare this model with a Seasonal Autoregressive Integrated Moving Average Exogenous (SARIMAX) which can predict seasonal patterned data with several outliers that occur due to holiday effects.

References

[1] Dingler M H, Lai Y C and Barkan C P L 2009 Transp. Res. Rec. 2117 33-40
[2] Hyndman R J and Athanasopoulos G 2018 Forecasting: Principles and Practice (Melbourne: OTexts)
[3] Zahroh S, Hidayat Y, Pontoh R S, Santoso A, Sukono, and Talib A T 2019 Proc. Int. Conf. on Industrial Engineering and Operations Management (Riyadh) (Riyadh: IEOM Society) p 406-412
[4] Yaseen Z M, El-Shafie A, Afan H A, Hameed M, Mohtar W H M W and Hussain A 2016 *Neural. Comput. Appl.* **27** 1533–1542
[5] Kajitani Y, Hipel K W, and Mcleod A I 2005 *J. Forecast.* **4** 105-117
[6] Anamika, Peesapati R, and Kumar N 2018 *IEEE Syst. J.* **12** 3075-3084
[7] Gallo M, Luca G D, D’Acierno L, and Botte M 2019 *J. Sens* **19** 3424
[8] Taylor S J and Letham B 2018 *Am. Stat.* **72** 37-45
[9] Yenidoğan I, Çayir A, Kozan O, Dağ T, Arslan Ç 2018 *Int. Conf. on Computer Science and Engineering (Sarajevo)* (Sarajevo: IEEE Xplore Digital Library) p 621-624
[10] Subashini A, Sandhiya K, Saranya S, and Harsha U 2019 *Int. Res. J. Multidisciplinary Technovation* **1** 56-63
[11] Samal K K R, Babu K S, Das S K, and Acharaya 2019 *Proc. Int. Conf. on Information Technology and Computer (Singapore)* (New York: Association for Computing Machinery Digital Library) p 80-85
[12] McCulloch W S and Pits W H 1943 *Bulletin of Mathematical Biophysics* **5** 115-133
[13] Fausset L 1994 *Fundamental of Neural Networks: Architectures, Algorithms, and Applications* (New Jersey: Prentice-Hall)
[14] Rosenblatt, F 1962 *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms* (Washington D.C: Spartan)
[15] Plummer E A 2000 *Time Series Forecasting with Feed-Forward Neural Networks: Guidelines And Limitations* Department of Computer Science and The Graduate School of The University of Wyoming Thesis p 4-9
[16] Harvey A and Peters S 1990 Estimation procedures for structural time series models *J. Forecast.* **9** 89-108
[17] Lewis C 1982 *Industrial and Bussiness Forecasting Methods* (London: Butterworth)