

Implementasi Model ARIMAX dan Jaringan Saraf Tiruan dalam Memprediksi Kedatangan Wisatawan ke Provinsi Bali

Februari-2021

Implementasi Model ARIMAX dan Jaringan Saraf Tiruan dalam Memprediksi Kedatangan Wisatawan ke Provinsi Bali

Ilham Wahyu Adli¹, Indwiarti², Annisa Aditsania³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹ilhamwahyuadli@students.telkomuniversity.ac.id, ²indwiarti@telkomuniversity.ac.id,

³aaditsania@telkomuniversity.ac.id

Abstrak

Pariwisata adalah berbagai macam kegiatan wisata dan didukung berbagai fasilitas serta layanan yang disediakan oleh masyarakat, pengusaha, pemerintah dan pemerintah daerah. Dengan adanya industri pariwisata pendapatan produk domestik bruto dapat ditingkatkan. Oleh karena itu para pemangku kepentingan harus mengetahui apa saja yang harus ditingkatkan pada sektor pariwisata. Hal ini dapat dilakukan dengan cara melakukan peramalan kedatangan wisatawan dengan bantuan *search query* dari data Google Trend sehingga para pemangku kepentingan dapat mengetahui apa yang diinginkan wisatawan.

Penelitian ini dilakukan menggunakan metode peramalan model ARIMAX dan algoritma jaringan saraf tiruan sebagai pembanding, dengan data BPS kedatangan wisatawan mancanegara ke Provinsi Bali melewati pintu Bandara Ngurah Rai sebagai data primer dan data google trend sebagai variabel eksogen. Dengan adanya peramalan kedatangan wisatawan diharapkan dapat membantu para pemangku kepentingan dalam membangun sektor pariwisata agar lebih baik lagi.

Kedua model dapat memprediksi kedatangan wisatawan dengan baik dengan model terbaik pada penelitian ini adalah model ARIMA dengan error sebesar 12,8% dengan variabel eksogen yang paling berpengaruh adalah *search query* "Bali Beach" yang dapat mengurangi error pada model sebesar 1,4%.

Kata kunci : Peramalan, ARIMA, ARIMAX, Pariwisata, Google Trends, Jaringan Saraf Tiruan

Abstract

Tourism is a variety of tourist activities and is supported by various facilities and services provided by the community, businessmen, government and local governments. With the tourism industry, gross domestic product income can be increased. Therefore, stakeholders must know what to improve in the tourism sector. This can be done by forecasting tourist arrivals with the help of search queries from Google Trend data so that stakeholders can find out what tourists want.

This research was conducted using the ARIMAX model forecasting method and artificial neural network algorithm as a comparison, with BPS data on foreign tourist arrivals to Bali Province through the door of Ngurah Rai Airport as primary data and Google Trend data as exogenous variables. With the forecasting of tourist arrivals, it is hoped that it can help stakeholders in building the tourism sector to be even better. Both models can predict tourist arrivals well with the best model in this study is the ARIMA model with an error of 12.8% with the most influential exogenous variable is the search query "Bali Beach" which can reduce errors in the model by 1.4%.

Keywords: Forecasting, ARIMA, ARIMAX, Tourism, Google Trend, Neural Network

1. Pendahuluan

Organisasi pariwisata dunia atau yang lebih dikenal dengan World Tourism Organization (WTO) mengatakan Asia dan Pasifik memiliki peran penting dalam pariwisata global, baik sebagai pasar masuk maupun keluar. Hal ini dikarenakan setengah dari semua kedatangan wisatawan internasional di Asia dan Pasifik tercatat di tujuan Asia Timur Laut pada 2017 (160 juta), sementara 37% dicatat di Asia Tenggara (120 juta), 8% di Asia Selatan (27 juta) dan 5% di Oceania (17 juta) [1]. Maka dari itu sektor pariwisata telah menjadi salah satu industri yang menguntungkan di dunia.

Untuk meningkatkan pendapatan dari sektor pariwisata, pengambil keputusan harus menentukan barang, jasa, dan sumber daya manusia yang harus dioptimalkan. Oleh karena itu, perlu adanya peramalan akurat mengenai sektor pariwisata sehingga dapat meningkatkan pendapatan produk domestik bruto (PDB) sektor pariwisata. Pada tahun 2017 sektor pariwisata telah memberikan kontribusi terhadap PDB Nasional sebesar 5%, dengan jumlah devisa sebesar 200 Triliun rupiah [2].

Search query data di Google tersedia untuk publik di dalam Google Trends. Hal ini memungkinkan untuk mengakses trend dari suatu *query* pada kurun waktu tertentu. Dalam studi sebelumnya Google Trends Data digunakan untuk melakukan *forecasting* potensi perilaku turis di Kerala, India [3]. Studi yang sama juga telah

dilakukan dan membuktikan bahwa hasil dengan menambahkan data Google Trend berpengaruh dalam melakukan peramalan [4].

Model ARIMA adalah model gabungan dari model *Autoregressive* (AR) dan *Moving Average* (MA). Secara umum model ARIMA dituliskan dengan notasi ARIMA (p,d,q), dimana p menyatakan ordo dari proses AR, d menyatakan ordo dari *differencing* (pembedaan) yang dilakukan agar data stasioner, dan q menyatakan ordo dari proses MA. Model ARIMAX merupakan penambahan parameter eksogenus. *Multilayer Perceptron* (MLP) adalah neural network yang terdiri dari input layer, hidden layer dan output layer.

Autoregressive Integrated Moving Average with exogenous variable (ARIMAX) dipilih sebagai metode pengembangan peramalan dari perilaku kedatangan wisatawan ke Indonesia terutama Bali. Model ARIMAX dengan variabel eksogen Google Trend juga pernah diimplementasikan untuk meramalkan kasus demam berdarah di Surabaya, Indonesia oleh Anggraeni dkk. (2016) dan menunjukkan peningkatan MAPE sebesar 3% [5]. Untuk dilakukan pengukuran pemodelan studi ini akan membandingkan dengan algoritma jaringan saraf tiruan. Studi mengenai pemodelan kedatangan wisatawan menggunakan jaringan saraf tiruan telah dilakukan oleh Palmer dkk. (2006), dengan hasil studi MAPE sebesar 3.44% menunjukkan bahwa jaringan saraf tiruan dapat diimplementasikan pada peramalan kedatangan wisatawan [6]. Studi lainnya telah dilakukan oleh Munandar (2019), mengenai *multivariate input data time series* menggunakan *Multilayer Perceptron* menghasilkan RMSE minimal 8,68% [7]

Data yang akan digunakan dalam penelitian ini adalah data kedatangan turis yang didapatkan dari Badan Pusat Statistik (BPS), dan juga data Google Trend. Data BPS yang akan diambil adalah data 10 tahun periode sebelumnya. Informasi yang akan didapat adalah kedatangan wisatawan mancanegara ke Provinsi Bali melewati pintu Bandara Ngurah Rai. Contoh kata kunci yang disarankan dapat membantu peramalan kedatangan turis antara lain: "hotels", "flights" dan "destination country" [4].

Dengan adanya peramalan pada kedatangan wisatawan mancanegara ke Bali diharapkan pemangku kepentingan dapat mengoptimalkannya pariwisata di Bali dengan menggunakan model terbaik antara ARIMAX atau jaringan saraf tiruan.

2. Studi Terkait

2.1 Data Time Series

Time series adalah urutan pengamatan yang diambil secara berurutan dalam kurun waktu tertentu [8]. Sehingga data time series merupakan kumpulan data dari kurun waktu tertentu untuk menentukan analisis yang akan dilakukan.

2.2 Peramalan

Peramalan adalah suatu teknik analisa perhitungan yang dilakukan dengan pendekatan kualitatif maupun kuantitatif untuk memperkirakan kejadian dimasa depan dengan menggunakan referensi data-data di masa lalu. Peramalan dapat dijadikan sebagai dasar pembuatan keputusan baik itu jangka pendek maupun jangka panjang. Di dalam peramalan hasil yang terbaik adalah hasil yang dapat memberikan kesalahan (*error*) paling sedikit. Peramalan merupakan bantuan penting untuk perencanaan yang efektif dan efisien [9]. Peramalan data time series perlu memperhatikan pola suatu data.

2.3 Google Trend

Google Trends adalah situs web milik Google, LLC yang berisi trend penggunaan kata kunci di pencarian pada mesin pencari Google. Google Trends memberikan akses ke sampel yang sebagian besar tidak difilter terkait permintaan penelusuran sebenarnya yang dibuat ke Google. Sampel tersebut dianonimkan (tidak akan ada pengguna yang teridentifikasi secara pribadi), dikategorikan (topik untuk kueri penelusuran ditentukan), dan digabungkan (dikelompokkan bersama). Ini memungkinkan Google Trend menampilkan minat dalam topik tertentu dari seluruh dunia atau hingga ke geografi tingkat kota [10]. Data dari Google Trend merupakan data *time series* yang memperlihatkan nilai suatu trend pada kurun waktu tertentu.

2.4 Model Algoritma Time Series

Model ARIMA adalah model yang secara penuh mengabaikan independen variabel dalam membuat peramalan, model ini menggunakan nilai masa lalu dan sekarang untuk mendapatkan peramalan jangka pendek yang akurat. Oleh sebab itu model ARIMA (p, d, q) sering kali digunakan dalam pemodelan peramalan seperti yang telah dilakukan pada studi sebelumnya. Model ARIMA (p, d, q) memiliki persamaan umum yaitu [8]:

$$wt = \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (1)$$

Dimana: $\phi_1, \phi_2, \dots, \phi_p$: Parameter Autoregressive
 $\theta_1, \theta_2, \dots, \theta_q$: Parameter Moving Average
 w_t : Nilai prediksi pada periode t, t = 1, 2, 3, ..., n
 a_t : Nilai error yang bersifat acak pada periode t, t = 1, 2, 3, ..., n

Berdasarkan [9] tahapan model ARIMA dibagi menjadi 3 fase yaitu:

1. Identifikasi
 - Data preparasi (mengstasionerkan data)
 - Mengidentifikasi model berdasarkan ACF dan PACF
2. Estimasi dan Percobaan
 - Uji signifikansi model
 - Pengecekan ACF untuk residual
 - Pengecekan normalitas residual
3. Aplikasi Model
 - Menggunakan model untuk peramalan

Model ARIMAX sendiri merupakan model ARIMA dengan penambahan variabel eksogen. Sehingga persamaan dari model ARIMAX didapatkan seperti:

$$w_t = \beta x_t + \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (2)$$

Dimana: β : Parameter eksogen
 x_t : Data eksogen saat t.

2.5 Model Algoritma Jaringan Saraf Tiruan

Artificial Neural Network atau Jaringan Syaraf Tiruan (JST) merupakan kecerdasan buatan yang memiliki karakteristik mirip dengan jaringan syaraf biologi. Implementasi *neuron* dalam JST adalah *perceptron* yang memiliki fungsionalitas yang sama. Secara sederhana *perceptron* terdiri dari 3 bagian yaitu: *input function*, *activation function*, dan *output function*. JST memiliki 2 metode pelatihan yaitu: *supervised* dan *unsupervised*. *Time series forecasting* sendiri merupakan implementasi dari *supervised learning*. Didalam *time series forecasting* arsitektur yang dapat digunakan adalah arsitektur *multilayer perceptron* (MLP) [12]. Kerangka kerja yang akan dipakai dalam peramalan wisatawan adalah:

$$X_t = f \left(\beta_0 + \sum_{j=1}^q \beta_j g(X_{t-1} \varphi_{ij} + \varphi_{0j}) \right) \quad (3)$$

$$\varphi_{ij}, i = 0, 1, \dots, p, j = 1, \dots, q$$

$$\beta_j, j = 0, 1, \dots, q$$

Dimana: f : Fungsi output
 g : Fungsi aktivasi
 p : Banyaknya output
 q : Banyaknya neuron dalam *hidden layer*
 χ_t : Output MLP
 χ_{t-1} : Input MLP
 β_j : Bobot output dengan *hidden layer*
 φ_{ij} : Bobot input dengan *hidden layer*

Input layer pada jaringan saraf tiruan perlu dilakukan normalisasi agar seluruh input berada dalam rentang nilai yang sama [13], untuk mencapai kondisi rentang tersebut digunakan normalisasi min-max. Dimana normalisasi min-max mengubah data menjadi rentang [0-1] dengan rumusnya yaitu:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (4)$$

2.6 Ukuran Evaluasi Peramalan

Untuk mengetahui model yang telah dibangun dapat menjadi acuan para pemangku pembuat keputusan, dibutuhkan teknik yang dapat mengevaluasi suatu model peramalan. *Mean absolute error* (MAPE) adalah salah satu ukuran yang paling banyak digunakan dalam mencari akurasi peramalan, dikarenakan keunggulannya dalam skala-independensi dan interpretabilitas [14]. MAPE didefinisikan seperti:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (5)$$

Dimana: n : Banyaknya Data
 A_t : Data aktual
 F_t : Data hasil peramalan
 t : Data poin waktu ke-t

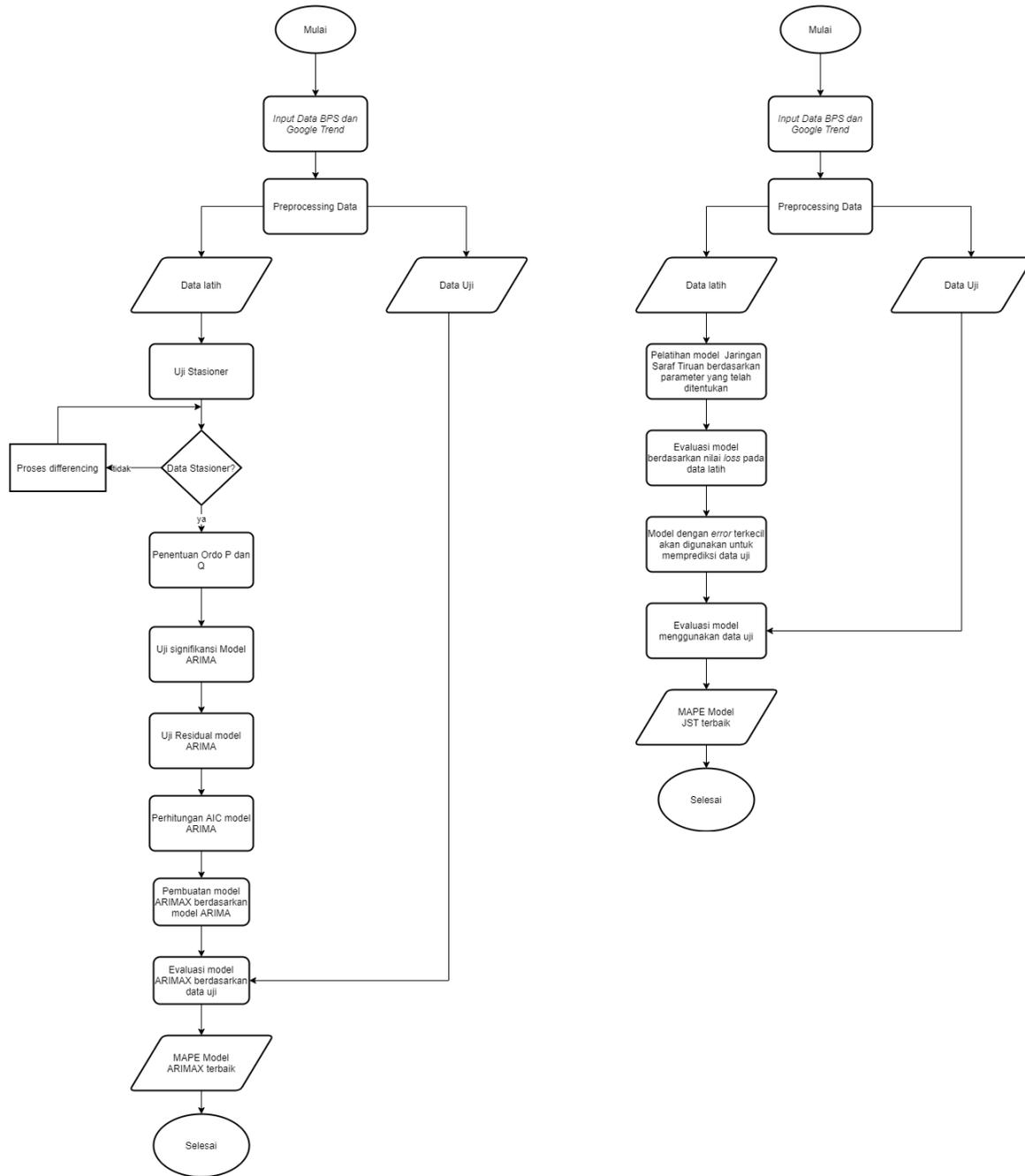
Menurut Lewis (1982) interpretasi nilai MAPE dapat dilihat pada tabel 1 [15].

Tabel 1 Kriteria Evaluasi MAPE

MAPE	Interpretasi
< 10	Peramalan sangat akurat
10 – 20	Peramalan baik
20 – 50	Peramalan masuk akal
> 50	Peramalan tidak akurat

3. Sistem yang Dibangun

Dalam melakukan prediksi kedatangan wisatawan ke Bali, Indonesia, dilakukan beberapa tahapan seperti pada *flowchart* gambar 1. Pada sebelah kiri merupakan urutan pembuatan model ARIMA dan ARIMAX hingga tahap evaluasi model, dan pada sebelah kanan merupakan urutan pembuatan model jaringan saraf tiruan arsitektur *multilayer perceptron*. Setelah mendapatkan nilai evaluasi kedua model selanjutnya adalah membandingkan hasil evaluasi tersebut untuk mencari tahu model mana yang lebih baik dalam melakukan peramalan kedatangan wisatawan mancanegara ke Provinsi Bali.



Gambar 1 Flowchart model ARIMA dan ARIMAX (kiri) dan Jaringan Saraf Tiruan (kanan)

3.1 Preprocessing Data

Pada penelitian ini akan digunakan model peramalan menggunakan Autoregressive Integrated Moving Average with Exogenous input (ARIMAX) dan jaringan saraf tiruan untuk memprediksi kedatangan turis ke Provinsi Bali. Data yang akan digunakan didapatkan dari data BPS dalam kurun waktu 10 tahun yaitu 2008 – 2018, dan juga data Google Trend sebagai variabel eksogenya. Data yang digunakan adalah “Jumlah Kunjungan Wisatawan Mancanegara per bulan ke Indonesia Menurut Pintu Masuk, 2008 - 2018 (Kunjungan) pada pintu masuk Bandara Ngurah Rai”. Beberapa *search query* Google Trend juga telah ditentukan yaitu: “Bali Hotel”, “Bali Culture”, “Bali Flight”, “Bali Beach” dan “Bali Tourism”. Penentuan *search query* didapatkan berdasarkan asumsi-asumsi yang berkaitan dengan kedatangan wisatawan mancanegara yang dipilih dari beberapa faktor yaitu: infrastruktur, kegiatan dan lokasi. Dalam tahapan ini dataset akan dibagi menjadi data latih dan data uji dengan perbandingan 60:40, 70:30, dan 80:20. Dikarenakan data yang akan digunakan merupakan data *time series* yang berpengaruh terhadap urutan waktunya, maka pembagian data latih dan data uji dilakukan dengan cara sekuensial. Dimana data latih akan digunakan sebagai data acuan pembuatan model, sedangkan data uji digunakan untuk melakukan evaluasi pada model yang telah dibangun.

3.2 Pembuatan Model ARIMA dan ARIMAX

Tahap pertama yang dilakukan adalah identifikasi struktur model dengan menentukan kestasioneran data. Kestasioneran data dapat dilihat dari uji Augmented Dickey-Fuller (ADF) dengan cara menghitung nilai P -value-nya. Jika P -value hasil uji ADF memiliki nilai diatas 0,05 maka data tersebut belum mencapai sifat stasioner. Data yang belum stasioner dapat ditransformasi dengan cara *differencing*. *Differencing* dapat dilakukan dengan menghitung nilai perbedaan antara w_t dengan w_{t-1} . Data memasuki proses *differencing* akan berpengaruh pada ordo d model ARIMA, ketika data pertama kali dilakukan proses *differencing* ordo d pada model akan bernilai 1, untuk kedua kali maka ordo 2 akan bernilai 2, dan seterusnya.

Data yang sudah stasioner dalam varians maupun rata-rata akan dianalisis untuk menentukan model sementara (tentatif) ARIMA (p,d,q) yang sesuai. Menetapkan ordo p dan q dapat dilihat dengan mengamati pola Autocorrelation Function (ACF) dan Partial Autocorrelation Function (PACF). Cara menentukan ordo p dan q yang sesuai dapat melihat pada tabel 2.

Tabel 2 Tabel Identifikasi Ordo P dan Q

Model	ACF	PACF
AR(p)	<i>Dies down</i> (Turun secara eksponensial)	<i>Cut off</i> setelah lag p
MA(q)	<i>Cut off</i> setelah lag q	<i>Dies down</i> (Turun secara eksponensial)
ARMA (p,q)	<i>Dies down</i> (Turun secara eksponensial)	<i>Dies down</i> (Turun secara eksponensial)
AR (p) atau MA (q)	<i>Cut off</i> setelah lag p	<i>Cut off</i> setelah lag q

Tahap selanjutnya adalah memeriksa signifikansi parameter model, apakah parameter model yang didapat dari ARIMA sementara signifikan atau tidak. Signifikansi parameter dengan melihat P -value pada tiap ordo yang dihasilkan, jika P -value kurang dari 0,05 maka ordo tersebut signifikan. Kemudian pemeriksaan diagnostik untuk membuktikan model cukup memadai atau sudah baik untuk digunakan dalam peramalan. Pemeriksaan diagnostik yang dilakukan yaitu uji *white noise* dan uji normalitas. Uji *white noise* dapat dilakukan dengan cara melihat persebaran residual plot ACF sedangkan uji normalitas dengan melihat persebaran distribusi residualnya.

Setelah dilakukan uji signifikansi model dan uji diagnostik pada model, akan terlihat model mana yang cukup baik untuk peramalan data time series yang di analisis. Model dikatakan baik apabila model memenuhi ketiga uji tersebut. Langkah berikutnya akan dilakukan estimasi parameter untuk menentukan model terbaik, yaitu dengan melihat nilai AIC (*Akaike Information Criterion*) yang paling kecil.

Setelah mendapatkan model ARIMA terbaik langkah selanjutnya untuk membangun model ARIMAX adalah dengan cara memasukan data variabel eksogen kedalam data *time series*. Setelah data sudah memiliki variabel eksogen peramalan kedatangan wisatawan bisa dilakukan.

3.3 Peramalan Kedatangan Wisata Menggunakan Model ARIMAX

Pada tahap ini, akan dilakukan peramalan terhadap data uji untuk kemudian dievaluasi dengan menggunakan metode evaluasi perhitungan yang telah ditentukan.

3.4 Evaluasi Model ARIMAX

Tahap ini merupakan tahap lanjutan dari peramalan yang telah dilakukan pada data uji, dengan menghitung nilai MAPE. Suatu model akan lebih baik jika memiliki nilai MAPE yang lebih kecil.

3.5 Pembuatan Model *Multilayer Perceptron*

Langkah pembuatan model jaringan saraf tiruan adalah sebagai berikut:

1. Inisiasi parameter model

Tahap ini melakukan inisialisasi awal terhadap nilai bias dan bobot secara acak di dalam model jaringan saraf tiruan.

2. Perhitungan maju

Tahap ini melakukan perhitungan maju dengan nilai bias dan bobot yang telah di tentukan di tahap inisialisasi.

3. Perhitungan mundur

Tahap ini melakukan perhitungan mundur dari hasil keluaran *error* pada tahap perhitungan maju yang akan digunakan parameter dalam pelatihan. Pelatihan akan selesai jika *error* yang didapatkan kurang dari iterasi. Dan digunakan untuk memperbaiki semua bobot.

4. Penyimpanan model

Nilai bias dan bobot akan disimpan sesuai dengan hasil yang diperoleh pada saat proses perhitungan mundur selesai. Dan proses 2-4 akan dilakukan sebanyak iterasi yang telah ditentukan.

5. Pengujian jaringan

Jaringan yang sudah dilatih pada tahapan sebelumnya, sehingga mendapatkan nilai bobot yang telah ditentukan akan digunakan untuk mengolah data uji sehingga hasil keluaran dapat dievaluasi.

Dengan spesifikasi arsitektur *input layer* berjumlah 8, *hidden layer* 2 dengan jumlah *neuron* 6 hingga 8 pada *hidden layer* pertama dan 1 hingga 2 pada *hidden layer* kedua yang akan menghasilkan *output layer* sebanyak 1 yaitu hasil peramalan. Dengan iterasi sebanyak 200 dan 500 diharapkan hasil *bias* dan *weight* model JST dapat menjadi acuan dalam model latih. Model dengan *loss* terkecil dari percobaan parameter diatas akan dijadikan model yang akan digunakan dalam peramalan. Dikarenakan *loss* terkecil merupakan hasil penurunan *error* paling besar pada saat tahap pelatihan model

3.6 Peramalan Kedatangan Wisatawan Menggunakan Model *Multilayer Perceptron*

Pada tahap ini, akan dilakukan peramalan terhadap data uji untuk kemudian dievaluasi dengan menggunakan metode evaluasi perhitungan yang telah ditentukan.

3.7 Evaluasi Model *Multilayer Perceptron*

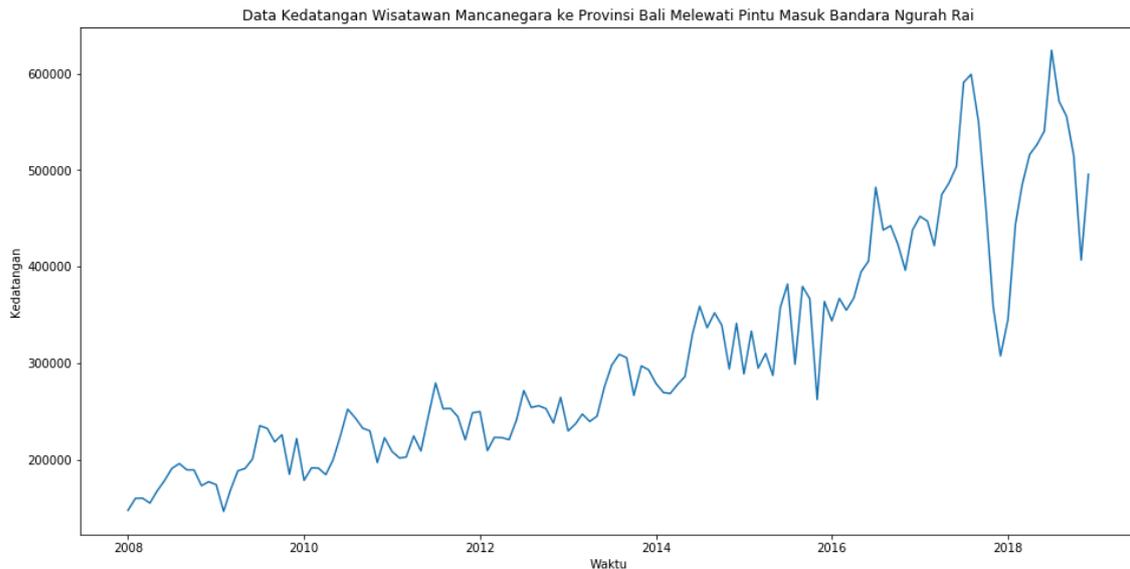
Tahap ini merupakan tahap lanjutan dari peramalan yang telah dilakukan pada data uji, dengan menghitung nilai MAPE. Suatu model akan lebih baik jika memiliki nilai MAPE yang lebih kecil.

3.8 Evaluasi Model ARIMAX dan *Multilayer Perceptron*

Setelah mendapatkan nilai MAPE dari tiap model yang telah dibangun akan dilakukan perbandingan MAPE dari kedua model tersebut, sehingga dapat mengetahui model terbaik yang dapat digunakan pada saat melakukan peramalan selanjutnya.

4. Evaluasi

Kedatangan wisatawan mancanegara dari tahun 2008-2018 melewati Bandara Ngurah Rai terlihat memiliki peningkatan tiap tahunnya. Data tersebut akan dipisah menjadi data latih dan data uji dengan perbandingan 80:20. Data latih adalah 80% data pertama kedatangan wisatawan sedangkan data uji adalah 20% sisanya (20% data terakhir). Dapat dilihat juga pada tahun kisaran 2017 dan 2018 terdapat sebuah *anomaly* dimana kedatangan wisatawan di Bali mengalami penurunan yang sangat drastis. Hal ini dikaernakan adanya erupsi dari Gunung Agung yang mengakibatkan penuruan jumlah kedatangan wisatawan.



Gambar 2 Plot kedatangan wisatawan mancanegara melewati pintu masuk Bandara Ngurah Rai

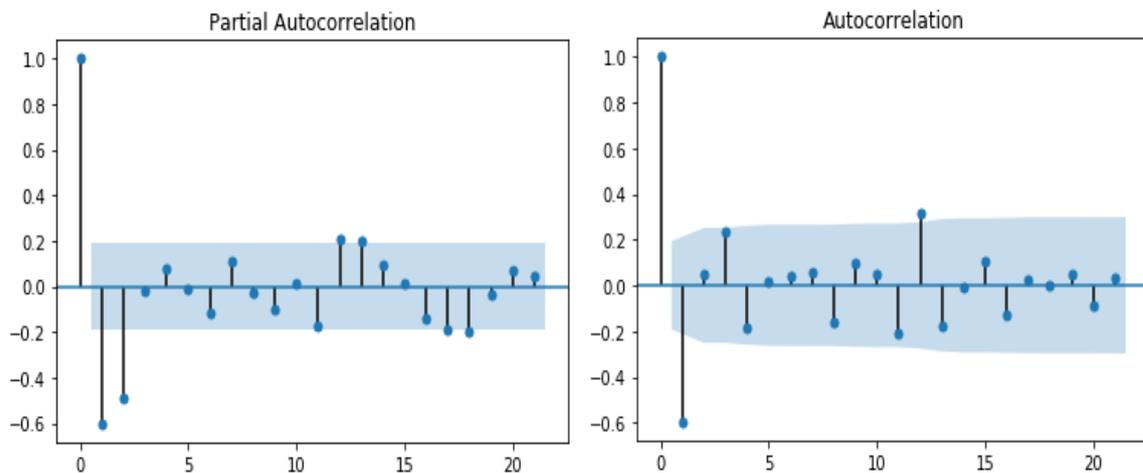
4.1 Model ARIMA

Tahapan awal model ARIMA adalah uji stasioner pada data latih, uji stasioner dapat dilakukan dengan Augmented Dickey–Fuller test, dimana nilai *p-value* harus dibawah 0,05 agar data dapat disebut stasioner. Jika *p-value* belum dibawah 0,05 maka proses *differencing* harus dilakukan. Proses *differencing* akan menentukan ordo *d* pada model ARIMA. Tabel 3 merupakan percobaan Augmented Dickey–Fuller test pada tiap rasio pembagian data.

Tabel 3 Uji stasioner pada data latih tiap rasio pembagian data

Rasio Pembagian Data	<i>differencing</i>	<i>p-value</i> Augmented Dickey–Fuller test	Uji Stasioner
60:40	0	9,97e-01	Belum Stasioner
	1	4,88e-02	Sudah Stasioner
70:30	0	9,94e-01	Belum Stasioner
	1	8,84e-02	Belum Stasioner
	2	2,22e-06	Sudah Stasioner
80:20	0	9,97e-01	Belum Stasioner
	1	1,94e-01	Belum Stasioner
	2	1,09e-27	Sudah Stasioner

Dari tabel 3, tiap rasio pembagian data menghasilkan ordo *d* yang berbeda-beda. Pada rasio 60:40 proses *differencing* hanya dilakukan sebanyak 1 kali, sehingga ordo *d* yang diperoleh adalah 1. Sedangkan untuk rasio 70:30 proses *differencing* dilakukan sebanyak 2 kali, sehingga ordo *d* pada kedua rasio pembagian tersebut adalah 2.



Gambar 3 Plot hasil PACF dan ACF rasio pembagian 80:20

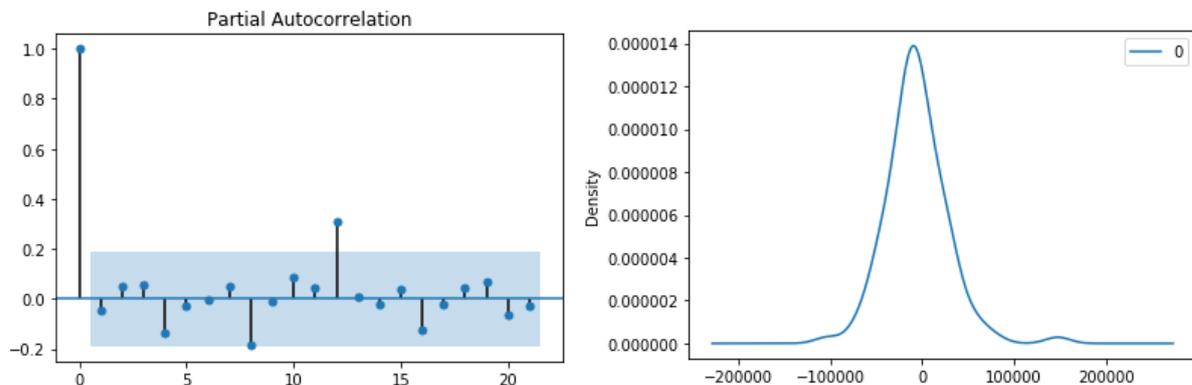
Untuk menentukan model arima selanjutnya adalah menentukan ordo p dan q. Ordo AR atau p dapat dilihat dai plot Partial Autocorrelation Function (PACF) dengan melihat pada lag berapa data cut off. Sedangkan untuk memementukan ordo MA atau q, dapat dilihat dari plot Autocorrelation Function (ACF) dengan melihat pada lag berapa data cut off. Dapat dilihat pada gambar 3 plot PACF data cut off pada lag 1 dan 2 sedangkan untuk plot ACF data cut off pada lag 1. Sehingga model tentatif sekarang adalah ARIMA (1, 2, 1) dan (2, 1, 1).

Setelah mendapatkan model tentatif, selanjutnya adalah melakukan uji signifikansi parameter, yang dapat dilakukan dengan melihat nilai *p-value* dari tiap ordo P dan Q. Nilai *p-value* tersebut harus lebih kecil dibandingkan nilai alpha yang telah ditentukan yaitu 0,05. Hasil dari pengecekan signifikansi parameter dapat dilihat pada tabel 4.

Tabel 4 Uji signifikansi model tentatif ARIMA

Rasio Pembagian Data	ARIMA (p, d, q)	Variabel	<i>p-value</i>	Uji Signifikansi
60:40	ARIMA (1, 1, 1)	AR (1)	9,72761e-01	Tidak Signifikan
		MA (1)	9,36575e-01	Tidak Signifikan
	ARIMA (2, 2, 1)	AR (1)	8,30355e-01	Tidak Signifikan
		AR (2)	8,80225e-01	Tidak Signifikan
		AR (3)	7,80528e-01	Tidak Signifikan
		AR (4)	3,81651e-01	Tidak Signifikan
ARIMA (2, 2, 1)	AR (5)	5,53233e-01	Tidak Signifikan	
	MA (1)	8,75434e-01	Tidak Signifikan	
70:30	ARIMA (1, 2, 1)	AR (1)	5,04535e-03	Tidak Signifikan
		MA (1)	3,79433e-22	Signifikan
	ARIMA (2, 2, 1)	AR (1)	1,86776e-04	Signifikan
AR (2)		6,98508e-02	Tidak Signifikan	
MA (1)		1,53666e-18	Signifikan	
80:20	ARIMA (1, 2, 1)	AR (1)	5,00453e-03	Signifikan
		MA (1)	3,79433e-22	Signifikan
ARIMA (2, 2, 1)	AR (1)	1,86776e-04	Signifikan	
	AR (2)	6,98508e-02	Tidak Signifikan	
	MA (1)	1,53666e-18	Signifikan	

Model tentatif ARIMA yang lulus uji signifikansi adalah ARIMA (1, 2, 1) dengan rasio perbandingan data sebesar 80:20. Selanjutnya moden tentatif akan dilakukan uji *white noise* dan normalitas untuk melihat kelayakan model.



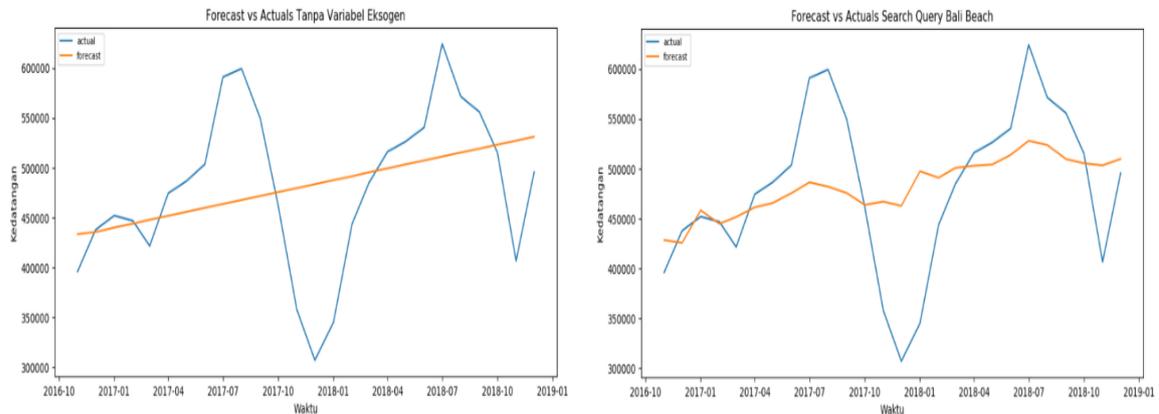
Gambar 4 Uji residual model tentative ARIMA (1, 2, 1) rasio data 80:20

Dapat dilihat dari grafik PACF residual model ARIMA (1, 2, 1) Sebagian besar lag-nya berada di dalam batas PACF, dan dengan persebaran residual yang berkisar di 0 membuktikan bahwa normalitas residual model ARIMA tersebut normal. Dikarenakan tidak ada model tentatif lain maka perhitungan nilai AIC tidak perlu dilakukan. Sehingga model yang akan digunakan untuk memprediksi kedatangan wisatawan adalah ARIMA (1, 2, 1). Setelah mendapatkan ordo ARIMA berdasarkan data latih, proses peramalan dapat dilakukan pada data uji. Hasil peramalan dapat dilihat pada tabel 5.

Tabel 5 Nilai MAPE pada data uji model ARIMA dan ARIMAX

MAPE ARIMA	MAPE ARIMAX Query Bali Hotel	MAPE ARIMAX Query Bali Culture	MAPE ARIMAX Query Bali Flight	MAPE ARIMAX Query Bali Beach	MAPE ARIMAX Query Bali Tourism
12,76679	12,38952	12,76198	12,76291	<u>11,34859</u>	12,77927

Dari gambar 5, dapat dilihat pengingkatan peramalan ketika menggunakan variabel eksogen berupa *search query* “Bali Beach”. Hal ini juga sejalan dengan hasil evaluasi model yang telah dilakukan menggunakan MAPE dimana penurunan MAPE terjadi ketika model ARIMA menggunakan variabel eksogen tambahan yang dapat dilihat di tabel 5.



Gambar 5 Peramalan pada data uji model ARIMA dan ARIMAX *search query* “Bali Beach”

4.2 Model Jaringan Saraf Tiuan

Skenario pelatihan dalam jaringan saraf tiruan telah ditentukan yaitu:

- a. Jumlah input layer sebanyak 8 neuron dengan hasil output layer adalah 1.
- b. Jumlah iterasi sebanyak 200 dan 500.
- c. Jumlah *hidden layer* pertama berkisar 6 hingga 8 neuron, dan pada *hidden layer* kedua 1 atau 2 neuron.

Hasil pelatihan tiap scenario telah dilakukan dengan hasil yang didapatkan dapat dilihat pada tabel 6. Skenario pelatihan yang memiliki nilai *loss* MAPE terkecil akan digunakan dalam peramalan data latih. Pada penelitian kali ini tiap variabel eksogen memiliki nilai yang berbeda untuk tiap neuronnya. Untuk MLP tanpa variabel eksogen jumlah neuron tiap *hidden layer*-nya berturut-turut adalah 8 dan 1 dengan *epoch* 200, untuk *search query* “Bali Hotel” adalah 8 dan 2 dengan *epoch* 500, dan seterusnya. Berdasarkan model ARIMA rasio data yang digunakan adalah 80:20, sehingga pembagian data latih dan data uji pada jaringan saraf tiruan hanya dibatasi pada rasio 80:20

Tabel 6 Hasil *loss* MAPE Normalisasi pada data latih model jaringan saraf tiruan

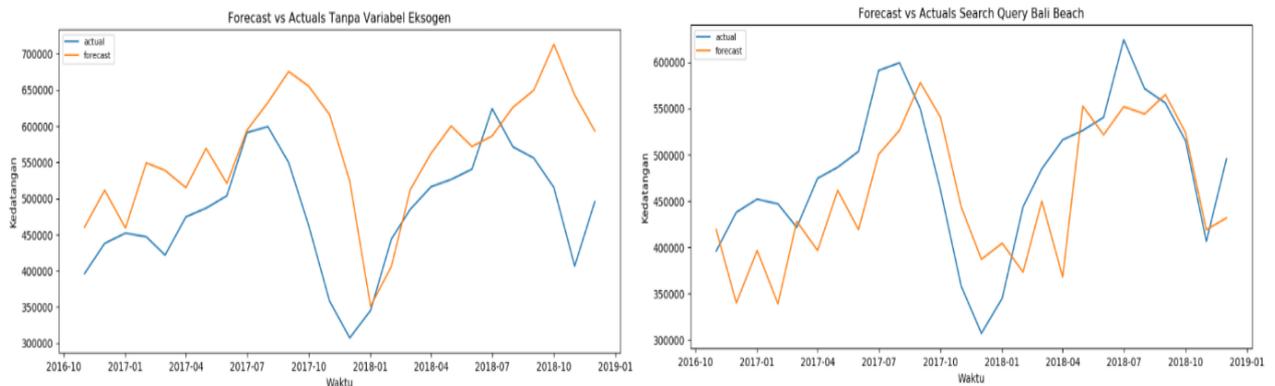
Neuron <i>Hidden layer</i> Pertama	Neuron <i>Hidden layer</i> Kedua	Epoch	MAPE JST	MAPE JST Query Bali Hotel	MAPE JST Query Bali Culture	MAPE JST Query Bali Flight	MAPE JST Query Bali Beach	MAPE JST Query Bali Tourism
6	1	200	39,2850	17,8847	27,9836	<u>21,8140</u>	33,8872	33,9485
6	2	200	40,8869	35,8112	28,5391	29,6221	32,0002	44,8459
7	1	200	44,6292	40,7225	36,2637	44,5565	29,7731	28,4367
7	2	200	38,5275	28,9852	39,4362	44,5415	29,8321	36,6717
8	1	200	<u>29,5778</u>	44,9057	18,6021	30,8449	43,3569	33,5653
8	2	200	41,6180	28,9925	31,0580	24,8266	22,6255	<u>28,4347</u>
6	1	500	40,9397	41,5336	21,8516	30,5269	40,6159	31,1847
6	2	500	38,4408	40,9294	17,8923	26,2069	30,1006	39,1669
7	1	500	41,6941	19,9282	<u>17,1840</u>	38,5931	38,4664	37,4299
7	2	500	38,0378	17,5094	21,6753	40,3759	<u>20,7195</u>	39,8625
8	1	500	43,1401	37,6869	22,1085	27,7758	40,0084	35,8732
8	2	500	42,0103	<u>17,5270</u>	29,4095	35,3647	37,5435	44,9108

Setelah mendapatkan nilai *loss* MAPE pada proses pelatihan, model dengan *loss* terkecil akan digunakan untuk mengevaluasi model menggunakan data uji. Hasil evaluasi peramalan dapat dilihat pada tabel 7.

Tabel 7 Nilai MAPE Denormalisasi pada data uji model jaringan saraf tiruan

MAPE JST	MAPE JST Query Bali Hotel	MAPE JST Query Bali Culture	MAPE JST Query Bali Flight	MAPE JST Query Bali Beach	MAPE JST Query Bali Tourism
20,02043	19,91200	15,19553	15,23622	<u>12,18942</u>	20,43146

Dari gambar 6, dapat dilihat peningkatan peramalan ketika menggunakan variabel eksogen berupa *search query* “Bali Beach”. Hal ini juga sejalan dengan hasil evaluasi model yang telah dilakukan menggunakan MAPE dimana penurunan MAPE terjadi ketika model jaringan saraf tiruan menggunakan variabel eksogen tambahan yang dapat dilihat di tabel 7.



Gambar 6 Peramalan pada data uji model JST dan JST *search query* “Bali Beach”

4.3 Perbandingan Model ARIMAX dengan Model Jaringan Saraf Tiuan

Tabel 8 Perbandingan nilai MAPE pada data uji model ARIMAX dan jaringan saraf tiruan

Model	MAPE Tanpa Variabel Eksogen	MAPE Query Bali Hotel	MAPE Query Bali Culture	MAPE Query Bali Flight	MAPE Query Bali Beach	MAPE Query Bali Tourism
ARIMAX	12,76689	12,38952	12,76198	12,76291	<u>11,34859</u>	12,77927
Jaringan Saraf Tiruan	20,02043	19,91200	15,19553	15,23622	12,18942	20,43146

Berdasarkan tabel 8 kedua model dapat memprediksi kedatangan wisatawan mancanegara ke Provinsi Bali dengan sangat baik, penggunaan variabel eksogen juga mayoritas menjadi data tambahan yang dapat meningkatkan hasil peramalan dengan melihat penurunan nilai MAPE. Kedua model mendapatkan pengurangan MAPE terbesar pada variabel eksogen “Bali Beach” yang berarti wisatawan mancanegara memilih untuk berwisata ke Bali dikarenakan kawasan pantai yang terdapat di Pulau Bali.

5. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan kedua model dapat meramalkan kedatangan wisatwan mancanegara dengan baik. Model terbaik pada penelitian ini adalah model ARIMAX dengan nilai MAPE sebesar 11,35%. Dengan *search query* terbaik berdasarkan data Google Trend adalah *search query* "Bali Beach" yang mendapatkan penurunan MAPE sebesar 1,4% untuk model ARIMA dan 7,8% penurunan menggunakan model jaringan saraf tiruan.

Daftar Pustaka

[1] World Tourism Organization. 2020. Global and Regional Tourism Performance. [Online] Available at: <https://www.unwto.org/global-and-regional-tourism-performance>. [Accessed 03 April 2020].
 [2] Kementerian Pariwisata dan Ekonomi Kreatif. 2019. Rencana Strategis Kementerian Pariwisata dan Ekonomi Kreatif 2018-2019. Jakarta: Kementerian Pariwisata dan Ekonomi Kreatif.
 [3] Padhi, S. S. dan Pati, R. K. 2017. Quantifying potential tourist behavior in choice of destination using Google Trends. *Tourism Management Perspectives*, vol. 24, no. 1, pp. 34-47.
 [4] Bangwayo-Skeete, P. dan Skeete, R. W. 2015. Can Google data improve the forecasting performance of tourist arrivals? *Mixed-data sampling approach*. *Tourism Management*, vol. 46, no. 1, pp. 454-464.

- [5] Anggraeni, W. dan Aristiani, L. A. 2016. Using Google Trend data in forecasting number of dengue fever cases with ARIMAX method case study: Surabaya, Indonesia. 2016 International Conference on Information & Communication Technology and Systems (ICTS), pp. 114-118.
- [6] Palmer, A., Montaña, J. J. dan Sese, A. 2006. Designing an artificial neural network for forecasting tourism time series. *Tourism Management*, vol. 27, no. 5, pp. 781-790.
- [7] Munandar, D. 2019. Multilayer Perceptron (MLP) and Autoregressive Integrated Moving Average (ARIMA) Models in Multivariate Input Time Series Data: Solar Irradiance Forecasting. *International Journal on Advanced Science, Engineering and Information Technology*.
- [8] Box, G. E., Jenkins, G. M., Reinsel, G. C., dan Ljung, G. M., 2015. *Time Series Analysis: Forecasting and Control*, 5th Edition, Hoboken, New Jersey: John Wiley and Sons Inc.,.
- [9] Hyndman, R. J. dan Athanasopoulos, G. 2018. *Forecasting: Principles and Practice*, 2nd edition, Melbourne, Australia: OTexts.
- [10] Google LLC, FAQ about Google Trends data - Trends Help. [Online] Available at: https://support.google.com/trends/answer/4365533?hl=id&ref_topic=6248052 [accessed Feb. 11, 2020]
- [11] Wheelwright, S., Makridakis, S., dan Hyndman, R. J. 1998. *Forecasting: methods and applications*. John Wiley & Sons.
- [12] Claveria, O. dan Torra, S. 2014. Forecasting tourism demand to Catalonia: Neural networks vs. time series models, *Economic Modelling*, vol. 36, pp. 220-228.
- [13] Imthias, A. 2014. Re: Why normalization is necessary in ANN?. [Online] Available at: https://www.researchgate.net/post/Why_normalization_is_necessary_in_ANN/5477692bcf57d797108b4592/citation/download. [Accessed 22 January 2021].
- [14] Kim, S. dan Kim, H. 2016. A new metric of absolute percentage error for intermittent demand forecasts, *International Journal of Forecasting*, vol. 32, no. 3, pp. 669-679.
- [15] Lewis, C. D. 1982. *Industrial and business forecasting methods*. London: Butterworths.