

NOWCASTING TINGKAT PENGHUNIAN KAMAR HOTEL MENGUNAKAN GOOGLE TRENDS (NOWCASTING OCCUPANCY RATE OF HOTEL ROOM WITH GOOGLE TRENDS)

Ika Ayuningtyas¹, Ika Wirawati²

¹Badan Pusat Statistik Provinsi Kalimantan Timur

²Badan Pusat Statistik

Jln. Kemakmuran No. 4 Samarinda, Kalimantan Timur

E-mail: ika.ayu@bps.go.id

ABSTRAK

Perkembangan penggunaan internet di Indonesia berkembang sangat pesat. Seiring dengan hal tersebut, semakin banyak orang yang memanfaatkan mesin pencari di internet. Salah satu yang paling banyak dimanfaatkan adalah *Google Search* dengan menggunakan kata kunci. Intensitas pencarian dengan berbagai kata kunci tersebut menghasilkan data yang berjumlah besar (*big data*) yang kemudian disajikan dalam bentuk Indeks *Google Trends*. Informasi pada indeks GT ini diprediksi memiliki korelasi dengan aktivitas saat ini sehingga dapat membantu dalam memprediksi rilis data berikutnya. Indeks GT ini sangat membantu dalam memprediksi masa kini yang sering disebut *nowcasting*. Penulis menggunakan data Tingkat Penghunian Kamar Hotel (TPK) hotel di Indonesia dan Indeks GT dengan kata kunci "hotel" selama periode Januari 2011-Juni 2020 yang kemudian diolah menggunakan metode analisis *time series regression* dan ARIMAX (1,1,1). Dari hasil pengolahan diperoleh ukuran keakuratan RMSE dan MAPE dimana metode ARIMAX memiliki nilai RMSE dan MAPE terkecil jika dibandingkan *time series regression*. Model yang terpilih digunakan untuk melakukan prediksi masa kini (*nowcasting*) TPK bulan Juli dan Agustus 2020 dengan hasil masing-masing sebesar 24,16 persen dan 24,49 persen.

Kata kunci: *nowcasting*, TPK, *google trends*, regresi *time series*, ARIMAX

ABSTRACT

The number of internet use in Indonesia has grown rapidly. Consequently, there are more people using search engines on the internet. One of the most popular search engines is Google Search with the common way to search is by inserting the keywords. High intensity of searching with huge various keywords produces large amounts of data (big data) which is presented as a Google Trends Index (GT). The information on the GT index is predicted have a correlation with forecasting current activity, therefore it can help predicting the next released data. This GT index is very helpful in predicting present data which is called nowcasting. The author uses the Hotel Occupancy Level (TPK) data for hotels in Indonesia and the GT Index with the keyword "hotel" during the period January 2011 to June 2020 and then processed using time series regression and ARIMAX analysis methods. The process produced the accuracy measurements RMSE and MAPE for each method and found that ARIMAX method has the smallest RMSE and MAPE values compared with time series regression. The chosen model will be used to predict nowcasting TPK for July and August 2020 which are respectively at 24.16 percent and 24.49 percent.

Keywords: *nowcasting*, TPK, *google trends*, *time series regression*, ARIMAX

PENDAHULUAN

Perkembangan penggunaan internet di Indonesia berkembang sangat pesat. Seiring dengan hal tersebut, semakin banyak orang yang memanfaatkan mesin pencari di internet. Mesin pencari yang cukup banyak digunakan yakni *Google Search*. *Google Search* biasanya digunakan untuk mencari berbagai informasi yang relevan dengan menggunakan kata kunci. Intensitas pencarian dengan berbagai kata kunci tersebut menghasilkan data yang berjumlah besar (*big data*) yang kemudian dicatat oleh Google. Sejak tahun 2009, Google kemudian menyebarkan intensitas pencarian ini melalui *Google Trends* (GT). GT dapat menyediakan laporan harian, mingguan, dan bulanan mengenai volume kata kunci yang terkait dengan berbagai informasi yang relevan. Volume kata kunci tersebut disajikan dalam bentuk indeks yang memiliki nilai antara 0-100. Informasi pada indeks GT ini diprediksi memiliki korelasi dengan aktivitas saat ini sehingga dapat membantu dalam memprediksi rilis data berikutnya. Hal yang harus diperhatikan terkait penggunaannya yaitu indeks GT tidak berpotensi untuk memprediksi masa depan, namun sangat membantu dalam memprediksi masa kini. Suatu prediksi untuk informasi di masa kini disebut *nowcasting*.

Nowcasting merupakan prediksi masa depan yang sangat dekat dan masa lalu yang sangat baru. Terminologi *nowcasting* ini terdiri dari "now" dan "forecasting". Istilah ini sangat populer dalam bidang meteorologi, namun belakangan ini mejadi populer di bidang ekonomi. Teknik *nowcasting* ini telah banyak digunakan untuk memantau keadaan ekonomi secara *real time*.

Banyak indikator yang telah dihasilkan dalam mengukur kondisi ekonomi, namun sayangnya indikator-indikator ini diterbitkan dengan jeda waktu. Tingkat Penghunian Kamar Hotel (TPK) merupakan salah satu *leading indicator* dalam mengukur ekonomi pada sektor pariwisata, khususnya sektor penyediaan akomodasi. TPK merupakan perbandingan banyaknya malam kamar yang terpakai dengan banyaknya malam kamar yang tersedia. Angka ini dirilis oleh Badan Pusat Statistik dengan jeda waktu satu bulan. Dengan jeda waktu tersebut, kondisi ekonomi sektor tersebut pada saat ini tidak dapat diketahui.

METODE

Tulisan ini bertujuan untuk memprediksi TPK menggunakan metode *Time Series Regression* dan ARIMAX pada periode bulan Juli dan Agustus 2020. Data yang digunakan adalah TPK yang dihasilkan oleh Badan Pusat Statistik dan Indeks GT yang diperoleh dari kata kunci "hotel", dengan rincian sebagai berikut:

1. Data TPK yang digunakan adalah periode Januari 2011 sampai dengan Juni 2020, digunakan sebagai variabel respon (Y_t).
2. Indeks GT dikumpulkan pada periode Januari 2011 sampai dengan Agustus 2020, digunakan sebagai variabel bebas (X_t). Data indeks GT periode Januari sampai dengan Juni 2020 digunakan untuk membangun model. Sementara indeks GT periode Juli-Agustus 2020 digunakan untuk melakukan prediksi TPK pada periode yang sama menggunakan model terpilih.

Time Series Regression

Metode pertama yang digunakan yakni analisis regresi *Time Series*. Analisis regresi merupakan salah satu teknik analisis data yang digunakan untuk mengkaji hubungan antara dua variabel atau lebih. Analisis regresi *time series* merupakan analisis regresi di mana variabel respon berautokorelasi yang biasanya terdapat pada data *time series*. Dengan demikian antar variabel respon dapat dibangun suatu hubungan yang dalam analisis ini berbentuk linier.

Jika data time series $Y_t, t = 1, 2, \dots, n$ berautokorelasi maka model regresi antar pengamatan dapat disajikan dalam persamaan:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_k Y_{t-k} + \varepsilon_t \dots\dots\dots(1)$$

dimana:

α, β merupakan koefisien regresi dan ε_t merupakan residual yang mengalami proses independen dan identik serta berdistribusi normal dengan nilai mean 0 dan varians σ^2 .

ARIMAX

Metode kedua yang digunakan dalam melakukan analisis yakni model ARIMAX. Model ARIMA merupakan model yang umum digunakan untuk data *time series* dan bermanfaat dalam melakukan *forecasting*. Dalam suatu penelitian, penambahan berapa variabel yang dianggap memiliki pengaruh yang signifikan terhadap data dilakukan untuk menambah akurasi dalam melakukan *forecasting*. ARIMAX merupakan model ARIMA dengan variabel tambahan. Model ARIMAX dengan orde (p, d, q) dengan differencing sebanyak d dapat ditulis dalam persamaan:

$$Y_t = \beta_1 X_{1,t} + \beta_2 X_{2,t} + \dots + \beta_p X_{p,t} + \frac{\theta_q(B)}{\phi_p(B)(1-B)^d} a_t \dots\dots\dots(2)$$

dimana:

p = orde AR

q = orde MA

Evaluasi Model

Evaluasi model digunakan untuk melakukan pemilihan model terbaik. Dua ukuran keakuratan ini digunakan untuk menentukan metode terbaik, yakni menggunakan nilai RMSE dan MAPE.

Root Mean Square Error (RMSE) merupakan salah satu indeks yang dapat digunakan untuk mengevaluasi ketepatan model. Penghitungan RMSE dapat dituliskan sebagai berikut:

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{n}} \dots\dots\dots(3)$$

dimana n adalah jumlah pengamatan.

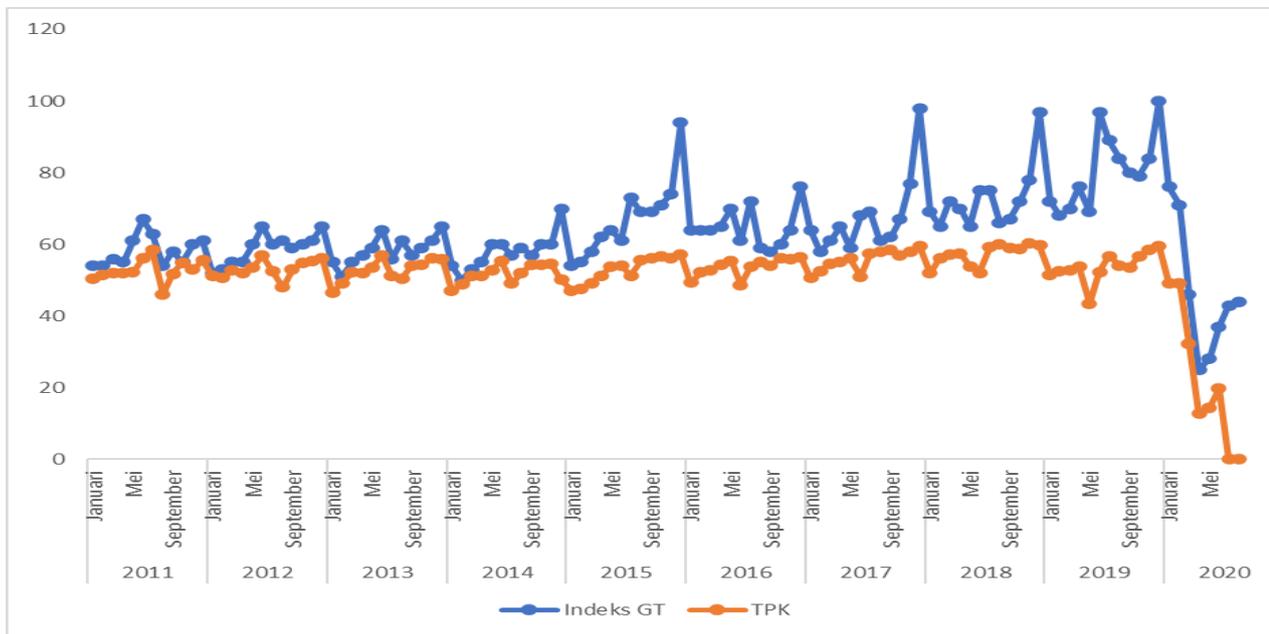
Mean Absolute Percentage Error (MAPE) menyatakan persentase kesalahan hasil peramalan terhadap hasil aktual. MAPE dapat mengindikasikan seberapa besar kesalahan dalam melakukan ramalan. Penghitungan MAPE dapat dituliskan sebagai berikut.

$$MAPE = \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \dots\dots\dots(4)$$

Metode dengan nilai RMSE dan MAPE terkecil akan dipilih sebagai metode terbaik.

HASIL DAN PEMBAHASAN

Plot *time series* pada Gambar 1 menunjukkan pola sebaran data tingkat penghunian kamar hotel bintang di Indonesia dan Indeks GT periode Januari 2011 sampai dengan Juni 2020. Gambar tersebut memperlihatkan adanya kesamaan pola sebaran data antara TPK dan indeks GT. Hal ini cukup untuk menguatkan dugaan bahwa indeks GT yang cenderung lebih cepat tersedia datanya bisa digunakan untuk memprediksi TPK yang rilis datanya membutuhkan jeda waktu.



Gambar 39. Tingkat Penghunian Kamar Hotel Bintang di Indonesia dan Indeks *Google Trends* periode 2011-2020.

Sebagaimana disebutkan sebelumnya, untuk memprediksi TPK bulan Juli-Agustus 2020 maka dibangun model *time series regression* dan ARIMAX menggunakan data TPK dan Indeks GT bulan Januari 2011 sampai dengan Juni 2020.

Pemodelan dengan *Time Series Regression*

Analisis dengan menggunakan regresi *time series* dilakukan dengan menggunakan beberapa kombinasi variabel dengan penjelasan sebagai berikut:

1. Model 1 dibangun dengan menggunakan data indeks GT pada bulan ke- t sebagai variabel bebas.
2. Model 2 dibangun dengan menggunakan data TPK pada bulan sebelumnya ($t-1$) sebagai variabel bebas.
3. Model 3 dibangun dengan menggunakan data data indeks GT pada bulan ke- t dan TPK pada bulan ke- $(t-1)$ sebagai variabel bebas.

Hasil analisis regresi *time series* dengan beberapa kombinasi variabel dinyatakan dalam model sebagai berikut:

Tabel 22. Persamaan Model *Time Series Regression*.

Model Regresi	Persamaan	<i>p-value</i> (model)
Regresi ($y \sim x$)	$Y_t = 27,48838 + 0,38945X_t + \epsilon_t$(5) (4,22e-16) (1,95e-14)	1,953e-14
Regresi ($y \sim y_{t-1}$)	$Y_t = 6,03100 + 0,88052 Y_{t-1} + \epsilon_t$(6) (0,0766) (<2e-16)	< 2,2e-16
Regresi ($y \sim y_{t-1} + x$)	$Y_t = 3,82318 + 0,70432 Y_{t-1} + 0,17912 X_t + \epsilon_t$(7) (0,225) (< 2e-16) (8,56e-06)	< 2,2e-16

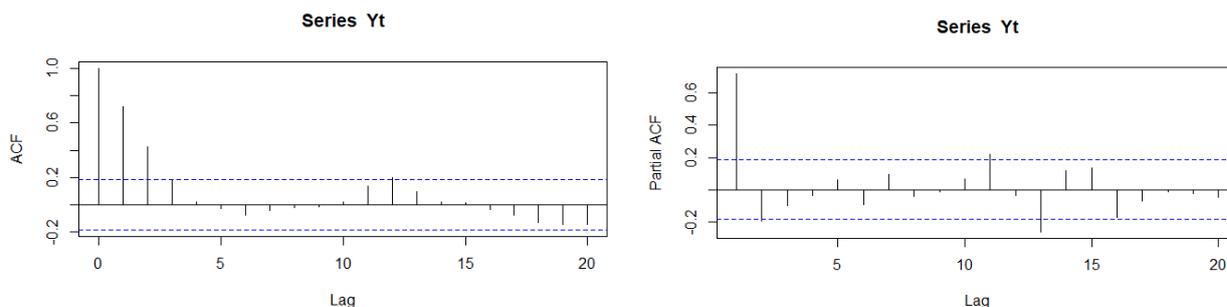
Keterangan:

- Nilai dalam kurung dibawah nilai pendugaan parameter adalah *p-value* untuk masing nilai pendugaan parameter

Berdasarkan hasil analisis regresi *time series* pada Tabel 1, hasil pengujian parameter menunjukkan bahwa seluruh parameter yang digunakan telah signifikan dengan *p-value* yang bernilai kurang dari 5%. Demikian pula pengujian untuk model telah menunjukkan nilai *p-value* yang bernilai kurang dari 5%, sehingga dapat disimpulkan bahwa model telah layak digunakan.

Pemodelan dengan ARIMAX

Analisis menggunakan model ARIMAX dilakukan dengan menambahkan variabel Indeks GT ke dalam model ARIMA. Identifikasi orde pada pemodelan ARIMAX dilakukan dengan membuat plot ACF dan PACF untuk data TPK (Gambar 2). Berdasarkan pola data yang ditunjukkan pada Gambar 2, maka model ARIMAX yang akan digunakan yakni model (1,1,1).



Gambar 2. Plot ACF dan PACF Tingkat Penghunian Kamar Hotel (TPK)

Hasil model ARIMAX dengan variabel Indeks GT sebagai variabel tambahan dinyatakan dalam persamaan berikut:

$$Y_t = 0,3273X_t + \frac{-0,8022 Y_{t-1}}{0,5613 Y_{t-1}} a_t \dots\dots\dots(8)$$

Perbandingan Hasil *Time Series Regression* dan Metode ARIMAX

Tabel 2 menampilkan nilai RMSE dan MAPE dari hasil analisis data menggunakan Time Series Regression dan metode ARIMAX. Dua ukuran keakuratan ini digunakan untuk menentukan metode terbaik. Metode dengan nilai RMSE dan MAPE terkecil akan dipilih sebagai metode terbaik. Perbandingan ukuran keakuratan pada masing-masing metode sebagai berikut:

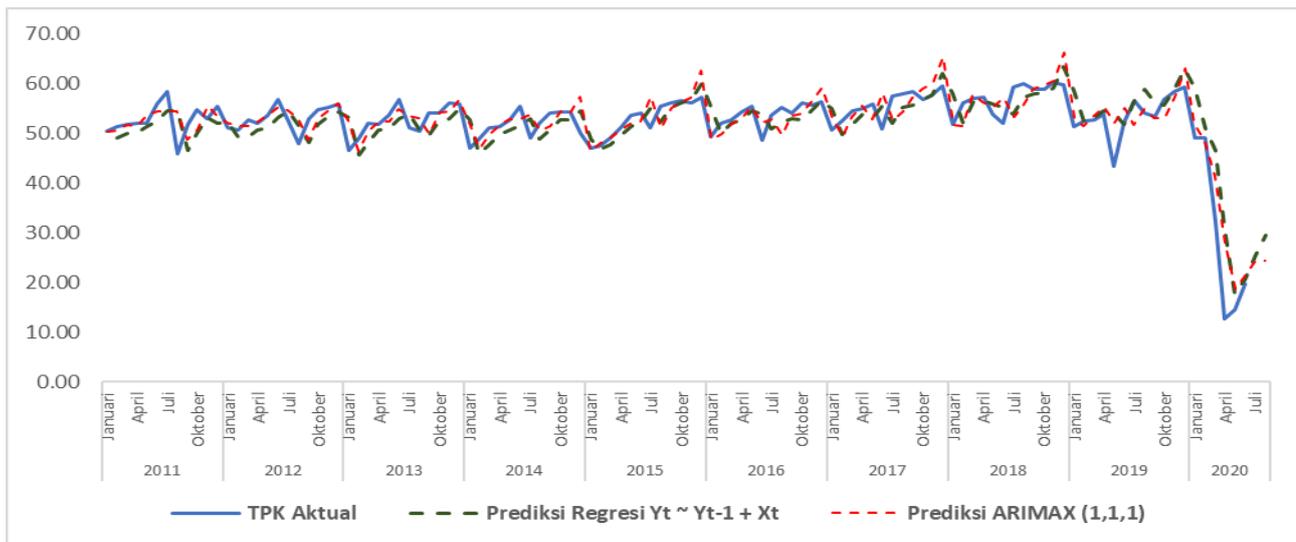
Tabel 2. Ukuran Keakuratan Metode Analisis.

Metode	Ukuran Keakuratan	
	RMSE	MAPE
Regresi ($y \sim x$)	5,54676	10,38284
Regresi ($y \sim y_{t-1}$)	4,38038	7,15405
Regresi ($y \sim y_{t-1} + x$)	4,00160	6,93593
ARIMAX (1,1,1)	3.44732	5,77268

Dari ukuran keakuratan terlihat bahwa metode ARIMAX (1,1,1) dengan variabel bebas Indeks GT memiliki RMSE dan MAPE yang lebih kecil dibandingkan metode regresi *time series*. Dengan demikian metode terbaik untuk memprediksi TPK bulan Juli-Agustus 2020 adalah metode ARIMAX (1,1,1).

Pendugaan Tingkat Penghunian Kamar Hotel

Berdasarkan hasil evaluasi model, model ARIMAX merupakan metode yang terbaik untuk memprediksi angka TPK bulan Juli-Agustus 2020 karena memiliki nilai RMSE dan MAPE yang lebih kecil dibandingkan hasil *time series regression*. Namun karena nilai RMSE dan MAPE untuk model Regresi ($y \sim y_{t-1} + x$) juga tidak jauh berbeda dengan metode ARIMAX, maka kedua metode ini akan digunakan untuk memprediksi TPK bulan Juli-Agustus 2020, dengan memasukkan data Indeks GT bulan Juli-Agustus 2020 sebagai variabel bebas. Gambaran perbandingan hasil prediksi model regresi *time series* dan ARIMAX, diberikan pada Gambar 3.



Gambar 3. Perbandingan Hasil Prediksi Model Regresi ($y \sim y_{t-1} + x$), ARIMAX (1,1,1) dan Data Aktual TPK.

Hasil prediksi TPK bulan Juli-Agustus 2020 diberikan pada Tabel 3. Tabel 3 menunjukkan prediksi masa kini (*nowcasting*) untuk TPK periode bulan Juli dan Agustus 2020 untuk masing-masing metode.

Tabel 3. Prediksi TPK Indonesia bulan Juli dan Agustus 2020 (persen).

Metode	Bulan	
	Juli 2020	Agustus 2020
Regresi ($y \sim y_{t-1} + x$)	25,40044	29,59450
ARIMAX (1,1,1)	24,16142	24,48872

Diprediksi pada bulan Juli 2020, TPK Hotel Bintang di Indonesia antara 24 – 26 persen, sedangkan pada bulan Agustus 2020, diprediksi TPK Hotel Bintang di Indonesia sebesar 24 – 30 persen.

KESIMPULAN

Pada penelitian ini dapat ditunjukkan bahwa Indeks *Google Trends* dapat digunakan untuk membangun model tingkat penghunian kamar hotel. Berdasarkan penelitian ini juga dapat disimpulkan bahwa pemanfaatan Indeks *Google Trends* dalam melakukan prediksi masa kini (*nowcasting*) semakin dimungkinkan mengingat banyaknya pengguna internet yang memanfaatkan mesin pencari *Google Search*. Banyak metode yang dapat digunakan dalam melakukan prediksi masa kini, termasuk metode analisis *time series regression* dan ARIMAX. Metode *time series* yang lain juga perlu untuk dikembangkan mengingat pandemi COVID-19 yang berdampak pada segala bidang kehidupan sehingga pola data pada beberapa indikator penting mengalami perubahan yang ekstrim.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Bapak Anggoro Dwittjahyono, Ibu Yusniar Juliana Nababan dan Bapak Widyo Pura Buana yang telah memberikan kesempatan dan motivasi demi kelancaran penulisan *paper* ini.

DAFTAR PUSTAKA

- Choi, H., & Varian, H. (2012). *Predicting the Present with Google Trends. The Economic Record vol 88, 88(Special Issue), 2-9.*
- Coble, D., & Pincheira, P. (2017). *Now-Casting Building Permits with Google Trends. MPRA Paper No. 76514.*
- Dini, N. S., Haryono, & Suhartono. (2012). *Peramalan Kebutuhan Premium dengan Metode ARIMAX untuk Optimasi Persediaan di Wilayah TBBM Madiun. Jurnal Sains dan Seni ITS Vol. 1 No. 1, 230-235.*
- Fathurahman, M., & Haeruddin. (2011). *Pemodelan Regresi Linier untuk Data Deret Waktu. Jurnal Eksponensial Nomor 2, 35-41.*
- Giannone, D., Reichlin, L., & Small, D. (2008). *Nowcasting: The real-time informational content of macroeconomic data. Journal of Monetary Economics, 665-676.*
- Muditomo, A. (2020). *Model Alternatif Nowcasting di Indonesia. BASKARA Journal of Business and Entrepreneurship, 65-70.*
- Nooraeni, R., Purba, N. S., & Yudho, N. P. (2020). *Using Google Trend Data as an Initial Signal Indonesia Unemployment Rate. Proceeding Contributed Paper Session Volume 3 62nd ISI World Statistic Congress 2019 (pp. 266-273). Kuala Lumpur, Malaysia: ISI World Statistics Congress.*
- Perdana, A. S. (2010). *Perbandingan Metode Time Series Regression dan ARIMAX pada Pemodelan Data Penjualan Pakaian di Boyolali. Surabaya: Institut Teknologi Sepuluh Nopember.*
- Shumway, R. H., & Stoffer, D. S. (2010). *Time Series Analysis and Its Applications: With R Examples (Third Edition). New York: Springer.*
- Zhang, Y., Yu, C. L., Li, H., & Hong, Y. (2018). *Nowcasting China's GDP Using a Bayesian Approach. JMSE, 232-258.*