

Prediksi Jumlah Kunjungan Wisatawan Di Jawa Barat Dengan Model ARIMAX Dan SARIMAX Menggunakan Data Google Trends

Ilham Rizki Julianto¹, Indwiarti², Aniq Atiqi Rohmawati³,

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

⁴Divisi Digital Service PT Telekomunikasi Indonesia

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

³aniqatiqi@telkomuniversity.ac.id

Abstrak

Sejak Tahun 1990-an sektor pariwisata di Indonesia merupakan salah satu penyumbang devisa terbesar negara, hal ini dibuktikan dengan survei yang dilakukan lembaga BPS yang menjelaskan bahwa pada Tahun 2017 terjadi peningkatan jumlah kunjungan wisatawan mancanegara di Indonesia sebesar 12,61%. Pada Tahun 2018 melalui kemenparekraf menjelaskan bahwa sektor pariwisata berkontribusi terhadap PDB negeri dengan menyumbang sebesar Rp 229,50 triliun pada devisa negara. Oleh karena itu untuk meningkatkan kualitas sarana dan prasarana wisata, maka penting bagi para pengelola destinasi wisata untuk mengetahui kebutuhan yang diperlukan untuk meningkatkan layanan dan infrastruktur wisata. Pada Penelitian ini akan dilakukan sebuah peramalan dengan menggunakan metode peramalan model *time series* ARIMAX dan SARIMAX serta data kunjungan wisatawan mancanegara Provinsi Jawa Barat dari BPS yang merupakan data utama penelitian dan data dari Google Trends sebagai variabel eksternal dalam bentuk kueri pencarian. Hasil ramalan menggunakan model *time series* ARIMAX dan SARIMAX memberikan hasil yang cukup baik. Kedua model memberikan penurunan nilai MAPE pada kisaran 1-2% untuk model ARIMA dan ARIMAX sedangkan 1-3% pada model SARIMA dan SARIMAX dengan MAPE terendah sebesar 16,61% pada model SARIMAX

Kata kunci: ARIMAX, SARIMAX, Peramalan, Pariwisata, Google Trends

Abstract

Since the 1990s the tourism sector in Indonesia has been one of the country's largest foreign exchange earners, this is proven by a survey conducted by the BPS agency which explains that in 2017 there was an increase in the number of foreign tourist arrivals in Indonesia by 12,61%. In 2018, the Ministry of Tourism and Creative Economy explained that the tourism sector contributed to the country's GDP by contributing Rp 229,50 trillion to the country's foreign exchange. Therefore, to improve the quality of tourism facilities and infrastructure, it is important for tourist destination managers to know the needs to improve tourism services and infrastructure. In this study, a forecast will be carried out using the ARIMAX and SARIMAX time series forecasting methods as well as data on foreign tourist arrivals in West Java Province gathered by BPS agency which is the main research data and data from Google Trends as an external variable in the form of a search query. The forecast results using the ARIMAX and SARIMAX time series models give quite good results. Both models provide a decrease in MAPE values in the range of 1-2% for the ARIMA and ARIMAX models while 1-3% in the SARIMA and SARIMAX models with the lowest MAPE of 16,61% on the SARIMAX model.

Keywords: ARIMAX, SARIMAX, Forecast, Tourism, Google Trends

1 Pendahuluan

Sejak tahun 1990-an, sektor pariwisata telah menjadi salah satu penyumbang devisa terbesar di negara Indonesia. Menurut Badan Statistik Pusat (BPS) di tahun 2018 jumlah kunjungan wisatawan mancanegara mencapai 15.81 juta kunjungan atau naik 12,61% dibandingkan dengan tahun 2017 yang tercatat 14,04 juta kunjungan [1]. WTO (*World Trade Organization*) juga menjelaskan bahwa peningkatan pariwisata global naik sebesar 7% pada Tahun 2017 yang mencapai 1,3 miliar wisatawan [2]. Hal ini menunjukkan pariwisata merupakan sektor ekspor yang mampu mendorong ekonomi negara. Meningkatnya jumlah wisatawan ke Indonesia perlu diantisipasi dengan pembangunan fasilitas yang memadai seperti infrastruktur wisata, fasilitas umum, dan prasarana lainnya. Untuk menentukan hal itu semua harus dilakukan pengambilan keputusan strategis yang mampu mengoptimalkan antara penyediaan jasa wisata dengan kebutuhan wisatawan, serta perlu dilakukan sebuah peramalan terhadap kunjungan wisatawan di masa mendatang untuk meningkatkan pendapatan produk domestik bruto (PDB) negara. Sektor pariwisata berkontribusi besar dalam PDB negara, hal ini dibuktikan dengan peningkatan jumlah devisa dari Rp 175,71 triliun pada tahun 2015 menjadi Rp 229,50 triliun di tahun 2018 [3].

Studi penelitian sebelumnya menjelaskan bahwa jejak data di internet melalui kueri data pencarian (*search query data*) dalam kurun waktu tertentu dianggap memiliki potensi dalam membantu sebuah peramalan [4], [5]. Sidhartha dan Rupesh (2017) menggunakan kueri data pencarian dalam penelitiannya untuk menganalisis perilaku wisatawan terhadap pemilihan destinasi wisata kota Kerala [6]. Kueri data pencarian yang dimaksud pada studi sebelumnya ini mengacu kepada layanan Google Trends. Google Trends merupakan layanan publik milik Google yang berisikan sebuah tren penelusuran dalam kurun waktu tertentu yang diambil dari seberapa sering suatu istilah dicari dan dibandingkan dari berbagai wilayah dan bahasa melalui layanan Google Search [7].

Pada penelitian ini peramalan akan dilakukan dengan menggunakan model *time series* ARIMAX dan SARIMAX. Data utama model yang akan digunakan dalam penelitian ini adalah data kunjungan wisatawan dari Badan Pusat Statistik (BPS). Rentang data BPS yang akan digunakan adalah data kunjungan wisatawan Provinsi Jawa Barat 9 tahun terakhir mulai dari tahun 2010 sampai tahun 2019 serta data dari Google Trends yang akan bertindak sebagai variabel eksogen pada model yang akan dirancang.

2 Studi Terkait

2.1 Peramalan

Peramalan adalah sebuah cara untuk memprediksi hal yang akan terjadi di masa depan. Peramalan terjadi karena adanya kesenjangan antara pemahaman suatu peristiwa dan kebutuhan di masa mendatang terkait peristiwa itu sendiri. Dengan peramalan keputusan strategis dapat diambil untuk melaksanakan perencanaan yang efektif dan efisien terhadap suatu kejadian yang akan datang. Metode peramalan dibagi menjadi dua kategori, yaitu metode kualitatif dan kuantitatif [8].

2.2 Model Deret Waktu

Deret waktu (*Time Series*) adalah serangkaian data berupa nilai pengamatan yang diukur selama kurun waktu tertentu, berdasarkan interval waktu yang tetap. Model ini berupaya untuk meramalkan kondisi yang akan datang dengan menggunakan data historis dan model ini merupakan contoh dari metode peramalan kuantitatif [9].

2.3 Pariwisata

Menurut Peraturan Daerah Kota Bandung Nomor 7 tahun 2012 tentang Penyelenggaraan Kepariwisata, Pariwisata adalah berbagai macam kegiatan wisata dan didukung berbagai fasilitas serta layanan yang disediakan oleh masyarakat, pengusaha, dan pemerintah daerah dan daya tarik wisata adalah segala sesuatu yang memiliki keunikan, keindahan, dan nilai yang berupa keanekaragaman kekayaan alam, budaya, dan hasil buatan manusia yang menjadi sasaran atau tujuan kunjungan wisatawan, serta usaha pariwisata adalah usaha yang menyediakan barang dan/atau jasa bagi pemenuhan kebutuhan wisatawan dan penyelenggaraan wisatawan [10].

2.4 Google Trends

Google Trends merupakan salah satu layanan publik di internet milik Google LLC yang berisi sebuah tren atau popularitas terhadap penelusuran dalam kurun waktu tertentu yang diambil dari seberapa sering suatu istilah dicari dan dibandingkan dari berbagai wilayah dan bahasa melalui layanan Google Search. Google LLC merilis data kueri pencariannya ke publik pada tahun 2006 hingga saat ini. Google Trends menyediakan informasi terkait popularitas kueri pencarian tersebut murni dari kumpulan hasil pencarian para pengguna layanan Google Search secara utuh tanpa ada filterisasi. Selain sebagai sarana sumber informasi pihak google juga memberikan akses dan berharap bahwa layanan Google Trends ini dapat dimanfaatkan untuk berbagai macam kepentingan penelitian [7], [11].

2.5 Model ARIMA dan ARIMAX

Analisis deret waktu satu variabel (*univariate*) adalah analisis deret waktu yang menggunakan satu variabel dalam pemodelan dan peramalan dengan melihat data historis, tanpa perlu mempertimbangkan variabel lain. ARIMA atau *Autoregressive Integrated Moving Average* adalah salah satu jenis dalam model *univariate* yang merupakan penggabungan dua model AR (*Autoregressive Integrated*) dan MA (*Moving Average*). Dalam penggunaan model ini terdapat aspek yang harus diperhatikan, yaitu stasioneritas data. Data dapat dikatakan stasioner apabila karakteristik seperti *mean* (rata-rata) dan *variability* (jarak) pada data penelitian bersifat konsisten [12]. Jika data belum stasioner maka diperlukan *differencing* untuk menstabilkan *mean* dan *variability* atau juga dapat dilakukan dengan mentransformasi logaritma [13]. Secara umum model ARIMA dinotasikan dalam bentuk ARIMA (p, d, q) dengan notasi sebagai berikut [14]:

$$(1 - \phi_p B - \dots - \phi_p B^p) (1 - B)^d y_t = (1 + \theta_1 B + \dots + \theta_q B^q) \varepsilon_t \quad (1)$$

dimana

- p = banyak komponen *autoregressive* (AR)
- d = order integrasi (level *differencing* untuk menstasionerkan data)
- q = banyak komponen *moving average* (MA)
- ϕ_p = parameter *autoregressive*
- θ_q = parameter *moving average*
- B = operator lag

Dan dapat disederhanakan menjadi

$$y'_t = \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (2)$$

Model ARIMAX merupakan model ARIMA juga, berbeda dengan ARIMA model ARIMAX merupakan model *multivariate* atau model yang memperhatikan variabel eksternal atau eksogen.

2.6 Model SARIMA dan SARIMAX

Model SARIMA atau *Seasonal Autoregressive Integrated Moving Average* adalah model ARIMA musiman dari model deret waktu yang didalamnya terdapat variasi musiman. Musiman dapat didefinisikan sebagai sebuah pola yang terjadi secara berulang pada selang waktu tertentu yang sifatnya tetap, dimana ARIMA dengan model ARIMA (p, d, q) sedangkan SARIMA dengan model SARIMA (p, d, q) (P, D, Q) m dimana P, D , dan Q merupakan parameter musiman dari SARIMA namun masih terkait dengan nilai p, d, q non-musiman sedangkan m merupakan nilai periode atau faktor musiman terjadi [15]. Secara umum model SARIMA dinotasikan dalam bentuk SARIMA (p, d, q) (P, D, Q) m dengan notasi sebagai berikut [16]:

$$\phi_p(B)\Phi_P(B^m)(1-B)^d(1-B^m)^D y_t = \theta_q(B)\Theta_Q(B^m) \varepsilon_t \quad (3)$$

dimana

- p = banyak komponen *autoregressive* (AR) non-musiman
- d = order integrasi non-musiman
- q = banyak komponen *moving average* (MA) non-musiman
- P = banyak komponen *autoregressive* (AR) musiman
- D = order integrasi musiman
- Q = banyak komponen *moving average* (MA) musiman
- m = periode musiman
- ϕ_p, Φ_P = parameter *autoregressive*
- θ_q, Θ_Q = parameter *moving average*
- B = operator lag

2.7 Identifikasi Model ARIMA

Model ARIMA terbagi menjadi dua bagian, yaitu bagian musiman (*seasonal*) dan non-musiman. Setiap bagian tersebut terdiri dari komponen AR, MA, dan *order difference*. Komponen tersebut merupakan nilai model awal ARIMA. Penentuan nilai AR, MA, dan *order difference* dilakukan melalui plot *autocorrelation function* (ACF) dan *partial autocorrelation function* (PACF). Plot tersebut memiliki bentuk berupa diagram batang antara koefisien korelasi (ACF/PACF) dengan *lag*-nya [12], [14]. Untuk lebih jelas cara menentukan nilai dari komponen AR dan MA dapat dilihat pada Tabel 1.

Tabel 1. Aturan Penentuan Parameter AR(p) dan MA(q)

No	Model	ACF	PACF
1	AR(p)	<i>Dies down</i> (Turun secara eksponensial)	<i>Cut off</i> setelah lag p
2	MA(q)	<i>Cut off</i> setelah lag q	<i>Dies down</i> (Turun secara eksponensial)
3	AR(p) atau MA(q)	<i>Cut off</i> setelah lag p	<i>Cut off</i> setelah lag q

4	ARMA(p, q)	<i>Dies down</i> (Turun secara eksponensial)	<i>Dies down</i> (Turun secara eksponensial)
---	------------	--	--

Keterangan:

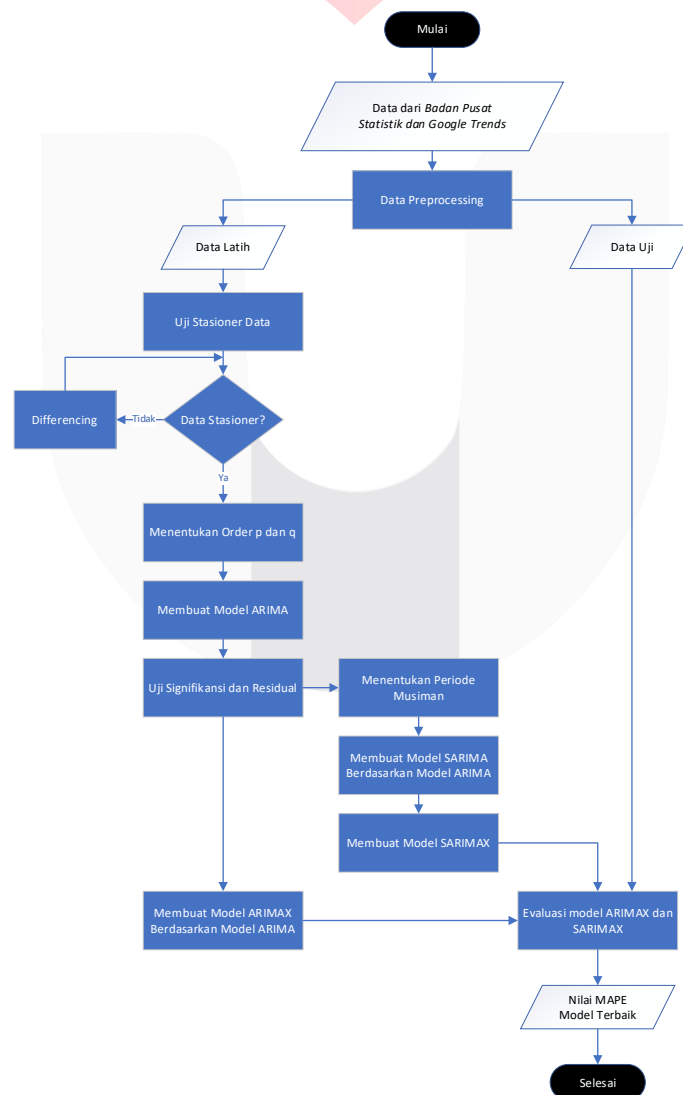
1. *Cut Off* : lag yang tidak signifikan terhadap garis batas signifikansi lag
2. *Dies Down* : lag yang bergerak turun dengan bertambahnya lag (sinusoidal)

2.8 Uji Stasioneritas (*Unit Root Test*)

Uji Stasioneritas digunakan untuk memeriksa jika data memiliki *unit root* dan untuk mengkonfirmasi posisi misi data. Pengujian data dilakukan untuk melihat jika data terintegrasi pada ordo yang sama atau tidak. Uji Stasioneritas dapat dilakukan dengan menggunakan uji ADF (*Augmented Dickey-Fuller*) dan PP (*Phillips-Perron*). Pengujian ini dilakukan dengan membandingkan *t-statistic* yang dihasilkan dengan ambang batas (*critical values*) pada batas (1%, 5%, 10%) dan jika *t-statistic* lebih rendah dari ambang batas, maka data stasioner [17].

3 Sistem yang Dibangun

Untuk memudahkan proses penelitian terhadap peramalan kunjungan wisatawan Provinsi Jawa Barat, maka dibuat sebuah perencanaan yang tergambar ke dalam *flowchart* rancangan model pada Gambar 1. *Flowchart* ini berisikan tahapan pembuatan model mulai dari mempersiapkan dataset, membuat model dasar dan evaluasi model untuk mendapatkan model prediksi yang optimal, sehingga peramalan kunjungan wisatawan mancanegara ke Provinsi Jawa Barat dapat dilakukan dengan baik.



Gambar 1. Flowchart rancangan model ARIMAX dan SARIMAX

3.1 Data Pre-processing

Sebuah metode untuk melakukan pengumpulan dan persiapan data awal yang dimana data akan ditransformasikan atau di encode ke dalam bentuk tertentu (menghilangkan data yang tidak relevan dengan kasus penelitian) sehingga mesin dapat dengan mudah untuk memproses atau menguraikannya. Dengan kata lain fitur atau atribut data dapat dengan mudah untuk ditafsirkan oleh algoritma [18]. pada penelitian ini data yang digunakan adalah sebagai berikut

- a. Data *time series* Kunjungan Wisatawan Provinsi Jawa Barat melalui kueri pencarian Google Trends sebagai data *multivariate* atau eksogen.
- b. Data *time series* Kunjungan Wisatawan Provinsi Jawa Barat yang dapat diakses melalui laman resmi Badan Pusat Statistik sebagai data *univariate*.

Kegiatan ini dilakukan agar data yang diolah sesuai dengan kasus yang sedang diteliti dan untuk memberikan batasan penelitian serta untuk mengurangi error dalam model yang akan dirancang.

3.2 Pembuatan Model ARIMA dan SARIMA

Proses pembuatan model ARIMA dan SARIMA akan meliputi uji stasioner data, menganalisis lag pada nilai ACF dan PACF, menentukan nilai parameter yang sesuai, serta melakukan identifikasi terhadap model yang dibangun. Model tersebut dibuat berdasarkan data *time series* Kunjungan Wisatawan Provinsi Jawa Barat yang diambil melalui Badan Pusat Statistik (BPS).

3.3 Pembuatan Model ARIMAX dan SARIMAX

Proses pembuatan model ARIMAX dan SARIMAX memerlukan model ARIMA dan SARIMA sebagai model utama sebelum data *multivariate* dimasukkan ke dalam model. Data tersebut adalah data kunjungan wisatawan Provinsi Jawa Barat yang didapat melalui Google Trends. Tahapan selanjutnya dalam membangun model ARIMAX dan SARIMAX adalah melakukan identifikasi order (p, d, q) untuk ARIMAX dan (p, d, q) (P, D, Q) untuk SARIMAX serta menentukan nilai musiman dan non-musiman melalui komponen AR dan MA.

3.4 Uji Signifikansi Menggunakan Ljung-Box

Ljung-Box merupakan sebuah metode yang digunakan untuk menguji kelayakan suatu model. Metode ini dilakukan dengan cara menguji *white-noise* pada residual model. jika nilai residual *p-value* memiliki nilai alfa (α) kurang dari 0.05 maka model dapat dikatakan *white-noise* atau model layak digunakan, sebaliknya jika lebih besar dari 0.05 maka model tidak layak untuk digunakan [19].

3.5 Evaluasi Model Menggunakan MAPE (Mean Absolute Percentage Error)

Evaluasi model akan dilakukan menggunakan metode MAPE, metode ini menghitung perbedaan antara data asli dengan data hasil peramalan. Hasil dari perbedaan tersebut diabsolutkan, kemudian dihitung ke dalam bentuk persentase. Dari hasil persentase kemudian didapat nilai mean-nya. Model dengan nilai MAPE berada di bawah 10% maka memiliki kinerja sangat bagus, dan nilai MAPE di antara 10%-20% memiliki kinerja bagus. Berikut ini persamaan dari MAPE, yaitu [20]:

$$M = \frac{100}{n} \sum_{t=0}^n \left| \frac{A_t - F_t}{A_t} \right|$$

dimana

- A_t = nilai aktual pada data t
- F_t = nilai peramalan pada data t
- n = jumlah periode data

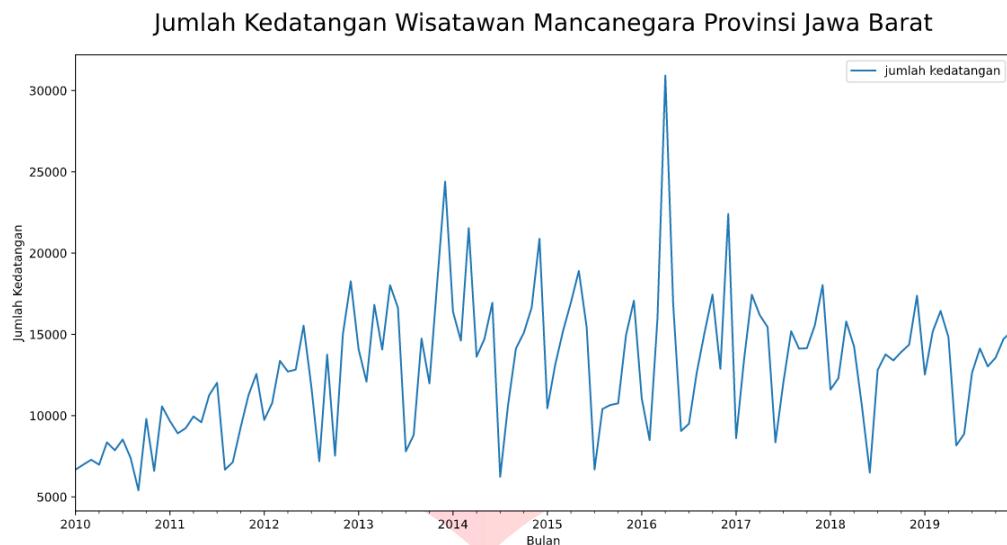
3.6 Melakukan Peramalan

Pada tahap ini akan dilakukan peramalan menggunakan model ARIMAX dan SARIMAX yang sudah dirancang menggunakan data latih kunjungan wisatawan Provinsi Jawa Barat dari BPS terhadap data uji dari Google Trends yang berisi kumpulan kueri pencarian terkait hal yang identik dengan kunjungan wisatawan Provinsi Jawa Barat.

4 Evaluasi

Jumlah kunjungan wisatawan mancanegara Provinsi Jawa Barat dari tahun 2010 - 2019 terlihat cukup fluktuatif. Hal ini dibuktikan dengan kunjungan wisatawan pada transisi Tahun 2015 dan 2016 mengalami

penurunan yang cukup drastis, dan memasuki pertengahan tahun 2016 sampai dengan awal tahun 2017 mengalami peningkatan yang cukup besar lalu pada pertengahan tahun 2017 sampai dengan tahun 2019 jumlah kunjungan wisatawan kembali menurun yang dipastikan bahwa data. Apabila diperhatikan secara seksama pada Gambar 2, grafik kunjungan wisatawan cenderung membentuk sebuah pola atau terindikasi memiliki nilai musiman yang kuat.



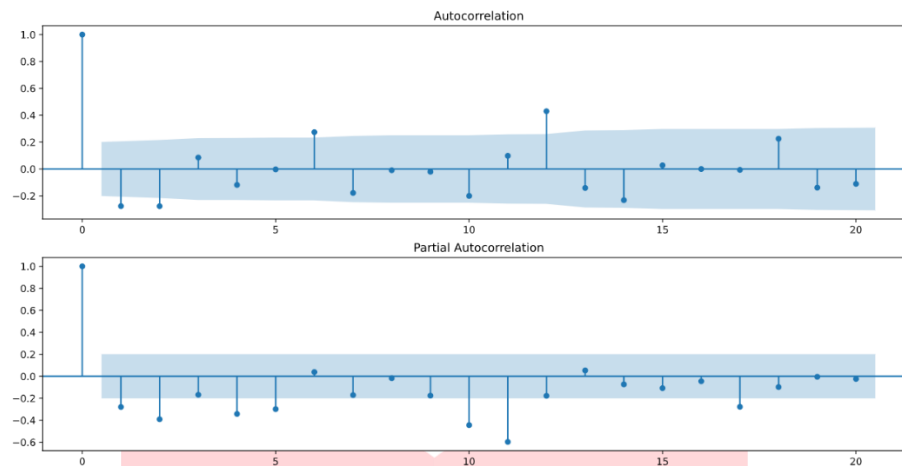
Gambar 2. Grafik Jumlah Kunjungan Wisatawan Mancanegara Provinsi Jawa Barat Periode Tahun 2010-2019

Untuk mendapatkan model ARIMAX dan SARIMAX adalah dengan menentukan model dasar ARIMA terlebih dahulu. Hal pertama yang perlu dilakukan dalam menentukan model ARIMA ialah melakukan uji stasioner terhadap data latih dengan uji *Augmented Dickey-Fuller* atau uji *adfuller* yang dimana nilai *p-value* pada hasil uji *adfuller* harus berada di bawah nilai 0,05 serta nilai *t-statistic* harus lebih rendah dari *critical values* untuk dikatakan stasioner, sebaliknya jika hal tersebut tidak terpenuhi maka harus dilakukan *differencing* terhadap data latih dan nilai ordo *d* pada model bertambah satu setiap kali *differencing* dilakukan. Tabel 2 menjelaskan hasil pengujian stasioner terhadap data latih untuk setiap rasio pembagian data.

Tabel 2. Uji stasioner terhadap rasio pembagian data latih dan data uji

No	Rasio Pembagian data (Data Latih : Data Uji)	Order Difference	<i>p-value</i> (<i>Adfuller</i> test)	<i>t-statistic</i>	<i>critical values</i>	Keterangan
1	60:40	0	2,40e-01	-2,112	1%: -3,544 5%: -2,911 10%: -2,593	Belum Stasioner
		1	1,13e-02	-3,390	1%: -3,246 5%: -2,912 10%: -2,594	Sudah Stasioner
2	70:30	0	3,05e-01	-1,960	1%: -3,525 5%: -2,903 10%: -2,589	Belum Stasioner
		1	2,22e-02	-3,164	1%: -3,127 5%: -2,904 10%: -2,589	Sudah Stasioner
3	80:20	0	1,16e-01	-2,496	1%: -3,511 5%: -2,897 10%: -2,585	Belum Stasioner
		1	4,15e-13	-8,295	1%: -3,511 5%: -2,897 10%: -2,585	Sudah Stasioner

Kesimpulan yang dapat diambil dari Tabel 2 adalah uji stasioner yang dilakukan pada tiap rasio pembagian data memiliki *order d* yang sama, yaitu sebanyak satu kali *differencing*.



Gambar 3. Plot ACF dan PACF pada rasio pembagian data 80:20

Langkah selanjutnya yang harus dilakukan setelah uji stasioner adalah menentukan nilai parameter p dan q . Nilai p atau AR didapat melalui plot PACF sedangkan q atau MR didapat melalui plot ACF dengan memperhatikan letak *lag* yang terjadi *cut off* pada kedua plot ACF dan PACF. Pada Gambar 3 kedua plot baik ACF dan PACF terjadi *cut off* pada lag 1 dan 2. Maka model tentatif yang mungkin dari rasio pembagian data 80:20 adalah ARIMA(0, 1, 1), ARIMA(0, 1, 2), ARIMA(1, 1, 0), ARIMA(1, 1, 1), ARIMA(1, 1, 2), ARIMA(2, 1, 0), ARIMA(2, 1, 1), dan ARIMA(2, 1, 2). Untuk lebih jelasnya dapat dilihat pada Tabel 3.

Tabel 3. Analisis uji signifikansi model tentatif ARIMA

No	Rasio Pembagian data (Data Latih : Data Uji)	Model Tentatif ARIMA (p, d, q)	Komponen	<i>p-value (Ljung-box)</i>	Uji Signifikansi
1	60:40	ARIMA (0, 1, 1)	MA (1)	2,60597E-20	Signifikan
		ARIMA (1, 1, 0)	AR (1)	7,76897E-03	Signifikan
		ARIMA (1, 1, 1)	AR (1)	1,16687E-01	Tidak Signifikan
			MA (1)	2,00117E-25	Signifikan
		ARIMA (2, 1, 0)	AR (1)	3,71867E-04	Signifikan
			AR (2)	5,34418E-03	Signifikan
		ARIMA (2, 1, 1)	AR (1)	3,19500E-03	Signifikan
			AR (2)	1,28000E-04	Signifikan
			MA (1)	2,38588E-01	Tidak Signifikan
2	70:30	ARIMA (0, 1, 1)	MA (1)	9,46098E-31	Signifikan
		ARIMA (0, 1, 2)	MA (1)	2,40000E-05	Signifikan
			MA (2)	5,23000E-04	Signifikan
		ARIMA (1, 1, 0)	AR (1)	2,57400E-03	Signifikan
		ARIMA (1, 1, 1)	AR (1)	7,44168E-02	Tidak Signifikan
			MA (1)	8,17316E-44	Signifikan
		ARIMA (1, 1, 2)	AR (1)	5,86204E-01	Tidak Signifikan
			MA (1)	4,11618E-01	Tidak Signifikan
			MA (2)	8,05600E-02	Tidak Signifikan
		ARIMA (2, 1, 0)	AR (1)	2,81716E-04	Signifikan
			AR (2)	1,88358E-05	Signifikan
		ARIMA (2, 1, 1)	AR (1)	8,14625E-02	Tidak Signifikan
			AR (2)	7,06370E-02	Tidak Signifikan

		ARIMA (2, 1, 2)	MA (1)	5,17064E-22	Signifikan
			AR (1)	7,83046E-01	Tidak Signifikan
			AR (2)	5,67807E-01	Tidak Signifikan
			MA (1)	3,39975E-01	Tidak Signifikan
			MA (2)	3,60545E-01	Tidak Signifikan
3	80:20	ARIMA (0, 1, 1)	MA (1)	4,57703E-46	Signifikan
		ARIMA (0, 1, 2)	MA (1)	1,00059E-09	Signifikan
			MA (2)	2,60509E-03	Signifikan
		ARIMA (1, 1, 0)	AR (1)	8,80000E-05	Signifikan
		ARIMA (1, 1, 1)	AR (1)	1,00846E-01	Tidak Signifikan
			MA (1)	2,32348E-58	Signifikan
		ARIMA (1, 1, 2)	AR (1)	5,10802E-01	Tidak Signifikan
			MA (1)	4,25530E-01	Tidak Signifikan
		ARIMA (2, 1, 0)	MA (2)	1,26835E-01	Tidak Signifikan
			AR (1)	1,14730E-06	Signifikan
		ARIMA (2, 1, 1)	AR (2)	6,37600E-05	Signifikan
			AR (1)	2,03116E-01	Tidak Signifikan
			AR (2)	1,74681E-01	Tidak Signifikan
		ARIMA (2, 1, 2)	MA (1)	2,41598E-21	Signifikan
			AR (1)	4,17585E-14	Signifikan
			AR (2)	1,70794E-04	Signifikan
			MA (1)	1,74989E-69	Signifikan
			MA (2)	1,03445E-20	Signifikan

Dari hasil uji signifikansi terhadap masing-masing model tentatif pada Tabel 3 terdapat beberapa calon/kandidat model tentatif maka akan dilakukan perbandingan nilai error terhadap masing-masing model tersebut.

Tabel 4. Analisis nilai error pada model tentatif ARIMA

No	Rasio Pembagian data (Data Latih : Data Uji)	Model Tentatif ARIMA (p, d, q)	Nilai Error (MAPE)
1	60:40	(0, 1, 1)	21,16%
		(1, 1, 0)	32,62%
		(2, 1, 0)	25,21%
2	70:30	(0, 1, 1)	24,62%
		(0, 1, 2)	25,94%
		(1, 1, 0)	61,81%
		(2, 1, 0)	53,13%
3	80:20	(0, 1, 1)	22,30%
		(0, 1, 2)	24,24%
		(1, 1, 0)	39,76%
		(2, 1, 0)	33,36%
		(2, 1, 2)	17,67%

Dep. Variable:	jumlah kunjungan	No. Observations:	96			
Model:	ARIMA(2, 1, 2)	Log Likelihood	-917.669			
Date:	Tue, 22 Jun 2021	AIC	1845.337			
Time:	22:58:11	BIC	1858.107			
Sample:	01-01-2010	HQIC	1850.497			
	- 12-01-2017					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	1.1032	0.146	7.555	0.000	0.817	1.389
ar.L2	-0.3796	0.101	-3.759	0.000	-0.578	-0.182
ma.L1	-1.8513	0.105	-17.619	0.000	-2.057	-1.645
ma.L2	0.9108	0.098	9.332	0.000	0.720	1.102
sigma2	1.401e+07	3.73e-09	3.75e+15	0.000	1.4e+07	1.4e+07

Dep. Variable:	jumlah kunjungan	No. Observations:	96			
Model:	SARIMAX(2, 1, 2)x(2, 1, 2, 12)	Log Likelihood	-797.596			
Date:	Tue, 22 Jun 2021	AIC	1613.191			
Time:	23:45:22	BIC	1634.961			
Sample:	01-01-2010	HQIC	1621.937			
	- 12-01-2017					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	-0.5558	0.537	-1.034	0.301	-1.609	0.498
ar.L2	-0.1197	0.233	-0.514	0.607	-0.576	0.337
ma.L1	-0.2421	0.523	-0.463	0.643	-1.267	0.783
ma.L2	-0.5095	0.488	-1.044	0.297	-1.466	0.447
ar.S.L12	-0.4599	1.293	-0.356	0.722	-2.993	2.073
ar.S.L24	0.5312	0.367	1.448	0.148	-0.188	1.250
ma.S.L12	0.0592	2.554	0.023	0.982	-4.947	5.065
ma.S.L24	-0.9370	0.656	-1.429	0.153	-2.222	0.348
sigma2	1.341e+07	7.39e-07	1.81e+13	0.000	1.34e+07	1.34e+07

Gambar 4. Hasil Analisis Model ARIMA (2, 1, 2) dan SARIMA (2, 1, 2) (2, 1, 2)₁₂

Pada tabel 4 dapat dilihat bahwa model ARIMA (2, 1, 2) memiliki rata-rata eror keseluruhan yang relatif lebih kecil dari model lainnya. Model ARIMA(2, 1, 2) juga dapat dituliskan ke dalam persamaan sebagai berikut:

$$(1 - \phi_1 B - \phi_2 B^2)(1 - B)^d y_t = (1 + \theta_1 B + \theta_2 B^2) \varepsilon_t$$

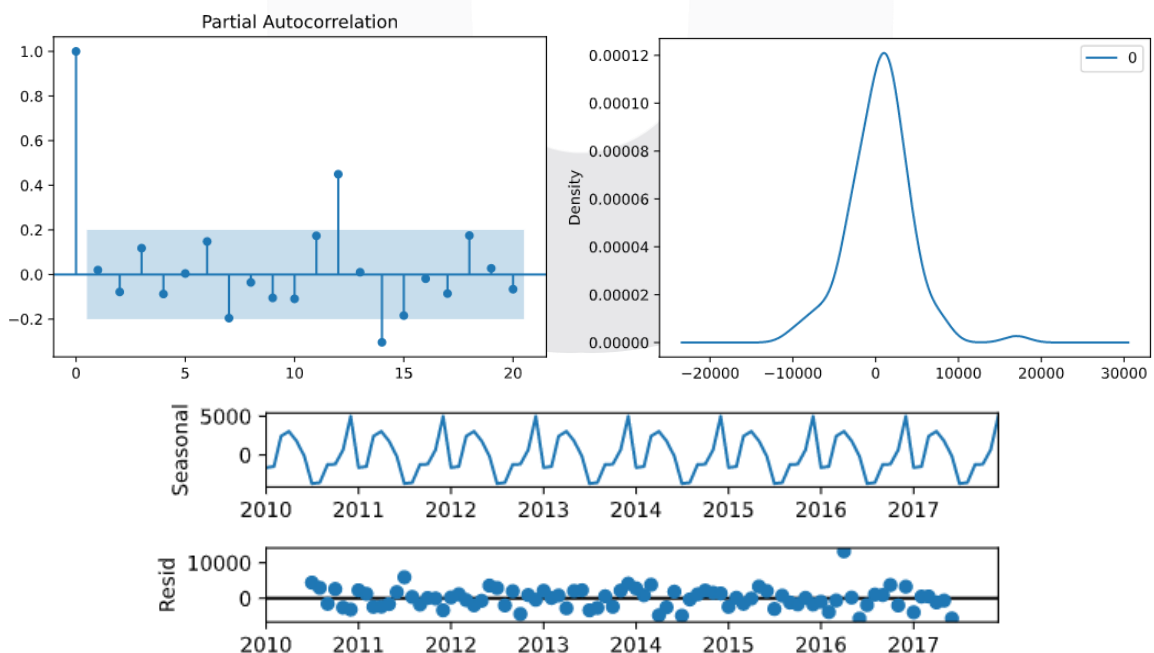
$$(1 - 1,1032B - (-0,3796B)) (1 - B) y_t = (1 + (-1,8513B) + 0,9108B) \varepsilon_t$$

$$(1 - 1,7236B - 0,7236B^2) y_t = (1 - 0,9405B) \varepsilon_t$$

$$y_t = 1,7236y_{t-1} + 0,7236y_{t-2} + \varepsilon_t - 0,9405\varepsilon_{t-1}$$

untuk memastikan bahwa model tersebut layak digunakan perlu dilakukan uji normalitas dan residual pada model tentatif ARIMA (2, 1, 2) dengan persamaan residual sebagai berikut:

$$\varepsilon_t = y_t - 1,7236y_{t-1} - 0,7236y_{t-2} + 0,9405\varepsilon_{t-1}$$

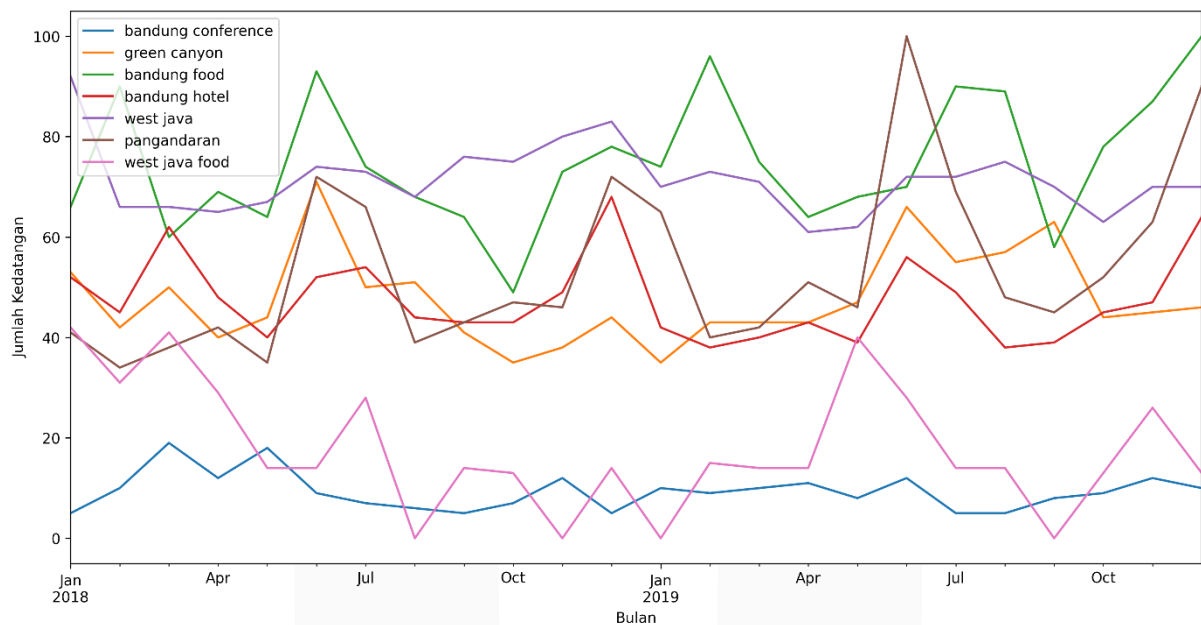


Gambar 5. Uji residual dan normalitas model tentatif ARIMA(2, 1, 2) dengan rasio pembagian data 80:20

Dari hasil uji residual dan normalitas yang ditampilkan pada Gambar 5 menunjukkan bahwa model tentatif ARIMA (2, 1, 2) memiliki nilai residual yang cenderung berdistribusi normal serta sebagian besar nilai lag pada

grafik PACF mendekati sumbu x. Hal ini normal terjadi dan menunjukkan bahwa model tentatif ARIMA (2, 1, 2) layak untuk digunakan. Setelah model ARIMA (p, d, q) didapat, dilanjutkan dengan perancangan model SARIMA. Perancangan model SARIMA pada penelitian ini akan mengacu kepada model dasar ARIMA(2, 1, 2) dengan nilai musiman yang terdapat pada gambar 5. Sehingga model SARIMA yang dibentuk adalah model SARIMA(2, 1, 2)(2, 1, 2)₆ dan SARIMA(2, 1, 2)(2, 1, 2)₁₂.

Setelah model ARIMA dan SARIMA ditentukan selanjutnya adalah mengumpulkan data melalui layanan Google Trends dengan kueri pencarian yang berkaitan dengan objek destinasi wisata di Jawa Barat. Kueri pencarian yang dipilih pada penelitian ini mengacu kepada pencarian populer pada laman tripadvisor.com. Perlu diketahui bahwa kueri pencarian yang didapat melalui layanan Google Trends merupakan data kuantitatif yang dijumlahkan berdasarkan seberapa sering kueri tersebut dicari melalui layanan Google Search dan data tersebut merupakan data dengan frekuensi harian. Dikarenakan data yang dimiliki pada data utama BPS adalah data dengan frekuensi bulanan, data Google Trends juga akan di konversikan ke dalam data dengan frekuensi bulanan, maka didapat data sebagai berikut:



Gambar 6. Kueri Data Pencarian Google Trends

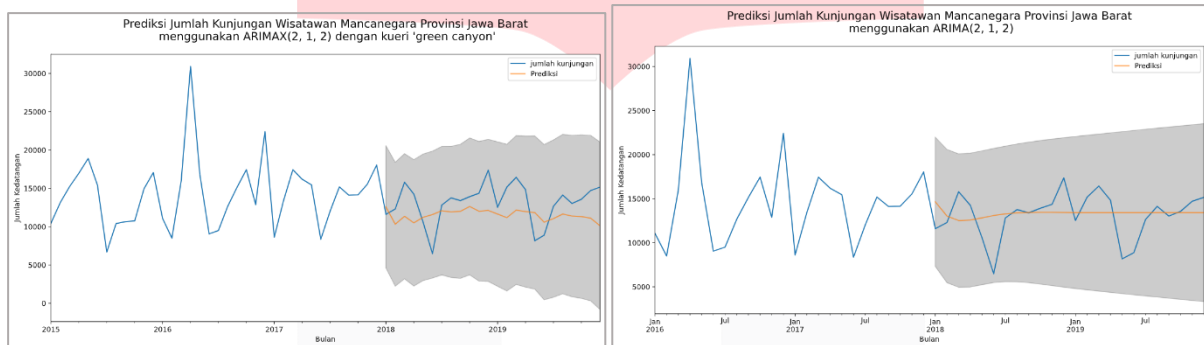
Dengan demikian model siap digunakan untuk peramalan terhadap jumlah kunjungan wisatawan Provinsi Jawa Barat baik pada model ARIMA dan SARIMA serta ARIMAX dan SARIMAX dengan menggunakan data eksternal dari Google Trends. Hasil peramalan dapat dilihat melalui Tabel 5.

Tabel 5. Nilai Mape Hasil peramalan model ARIMA, ARIMAX, SARIMA, dan SARIMAX dengan Data Eksternal Google Trends

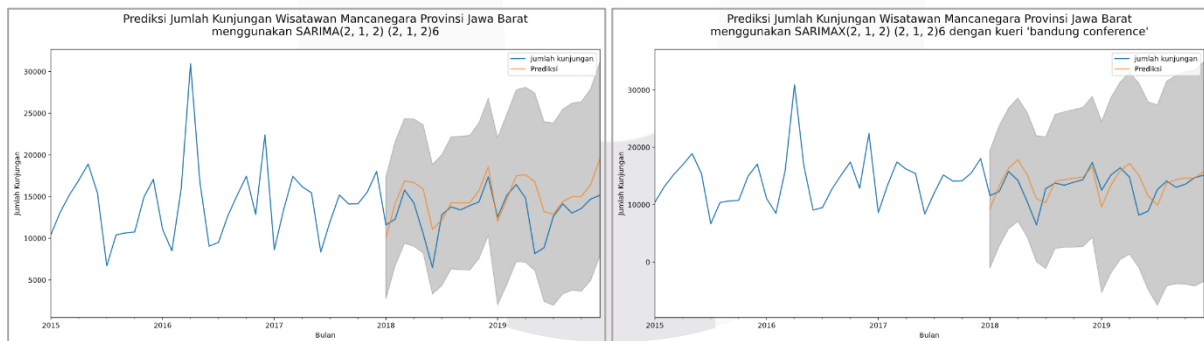
No	Periode Musiman	Model	Kueri Pencarian	MAPE
1	-	ARIMA (2,1,2)	-	17,67%
		ARIMAX (2,1,2)	"bandung conference"	17,83%
			"green canyon"	16,71%
			"bandung food"	17,53%
			"bandung hotel"	18,58%
			"west java"	21,35%
			"pangandaran"	25,09%
			"west java food"	17,39%
2	6 Bulan	SARIMA (2,1,2) (2,1,2) ₆	-	18,05%
		SARIMAX (2,1,2) (2,1,2) ₆	"bandung conference"	17,91%
			"green canyon"	23,88%
			"bandung food"	18,14%

3	12 Bulan	SARIMA (2,1,2) (2,1,2)12	“bandung hotel”	18,13%
			“west java”	23,82%
			“pangandaran”	18,81%
			“west java food”	21,30%
		SARIMAX (2,1,2) (2,1,2)6	-	19,34%
			“bandung conference”	18,58%
			“green canyon”	18,12%
			“bandung food”	19,28%
			“bandung hotel”	18,29%
			“west java”	16,61%
			“pangandaran”	22,93%
			“west java food”	19,71%

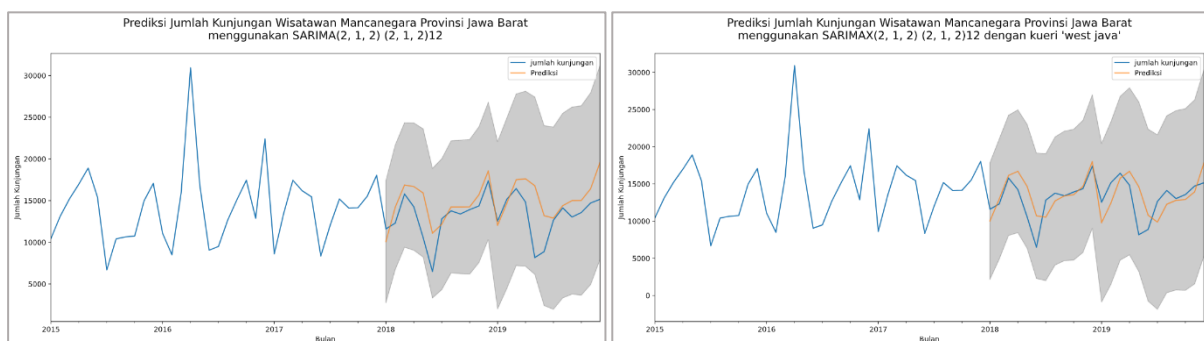
Dilihat dari Tabel 5 model ARIMA dan ARIMAX menunjukkan adanya peningkatan nilai MAPE yang relatif lebih kecil dibandingkan dengan model SARIMA dan SARIMAX, namun model dasar ARIMA memiliki nilai MAPE yang cukup baik dibandingkan dengan model SARIMA. Hal ini juga dapat dilihat melalui grafik prediksi pada Gambar 7, 8, dan 9.



Gambar 7. Grafik prediksi model ARIMA dan ARIMAX



Gambar 8. Grafik prediksi model SARIMA dan SARIMAX periode musiman 6 bulan



Gambar 9. Grafik prediksi model SARIMA dan SARIMAX periode musiman 12 bulan

5 Kesimpulan

Berdasarkan hasil penelitian yang telah dilaksanakan terhadap kedua model ARIMAX dan SARIMAX dalam melakukan peramalan kunjungan wisatawan mancanegara ke Provinsi Jawa Barat. Model SARIMAX (2, 1, 2) (2, 1, 2)₁₂ dengan kueri pencarian “west java” adalah model dengan nilai MAPE terendah sebesar 16,61%. Secara garis besar model ARIMAX pada penelitian ini memberikan nilai penurunan MAPE yang relatif kecil dengan kisaran 1-2% dibandingkan SARIMAX yang berada pada kisaran penurunan MAPE 1-3% setelah penambahan kueri pencairan dari Google Trends. Dari kedua hasil tersebut walaupun penurunan akurasi pada nilai MAPE tidak signifikan tetapi dapat disimpulkan bahwa pemilihan antara model ARIMA dan SARIMA memiliki ragam hasil yang berbeda.

Referensi

- [1] Badan Pusat Statistik Republik Indonesia, “Statistik Kunjungan Wisatawan Mancanegara 2018,” BPS RI, Jakarta, 2018.
- [2] UNWTO, GTERC, “Asia Tourism Trends 2018 Edition – Executive Summary,” UNWTO, GTERC, 2018.
- [3] Kementerian Pariwisata dan Ekonomi Kreatif, “Laporan Kinerja Kementerian Pariwisata Tahun 2019,” Kementerian Pariwisata dan Ekonomi Kreatif, Jakarta, 2019.
- [4] B. Bokelmann dan S. Lessmann, “Spurious patterns in Google Trends data - An analysis of the effects on tourism demand forecasting in Germany,” *Tourism Management*, vol. 75, p. 8, 2017.
- [5] X. Li, B. Pan, R. Law dan X. Huang, “Forecasting tourism demand with composite search index,” *Tourism Management*, vol. 59, p. 63, 2017.
- [6] S. S. Padhi dan R. K. Pati, “Quantifying potential tourist behavior in choice of destination using Google Trends,” *Tourism Management Perspectives*, vol. 24, p. 45, 2017.
- [7] Google LLC, “FAQ about Google Trends data,” Google LLC, [Online]. Available: <https://support.google.com/trends/answer/4365533?hl=en>. [Diakses 20 Januari 2021].
- [8] A. A. Ete, “Peramalan Jumlah Kedatangan Wisatawan Mancanegara Ke Indonesia Menurut Pintu Masuk Menggunakan Singular Spectrum Analysis dan ARIMA,” dalam *Tesis. Matematika Dan Ilmu Pengetahuan Alam, Statistika, Institut Teknologi Sepuluh Nopember*, Surabaya, 2017.
- [9] S. K. Wardani, “Pemodelan Multivariate Time Series Menggunakan Multi Input Transfer Function untuk Meramalkan Curah Hujan di Kota Semarang,” dalam *Skripsi, Matematika Dan Ilmu Pengetahuan Alam, Matematika, Universitas Negeri Semarang*, Semarang, 2017.
- [10] Pemerintah Daerah Kota Bandung, “Peraturan Daerah Kota Bandung Nomor 07 Tahun 2012 Tentang Penyelenggaraan Kepariwisataaan,” Bandung, 2012.
- [11] P. Evita dan I. Ariqoh, “Google Trends Analytics Dalam Bidang Pariwisata,” *Majalah Ekonomi*, vol. 24, p. 232, 2019.
- [12] Badan Pusat Statistik, “Seasonal Adjustment Dan Peramalan Pdb Triwulanan,” Badan Pusat Statistik, 2016.
- [13] W. Alam, M. Ray, R. R. Kumar, K. Sinha, S. Rathod dan K. N. Singh, “Improved ARIMAX modal based on ANN and SVM approaches for forecasting rice yield using weather variables,” *Indian Journal of Agricultural Sciences*, vol. 88, p. 102, 2018.
- [14] R. J. Hyndman dan G. Athanasopoulos, *Forecasting: Principles and Practice 3rd Edition*, Melbourne: OTexts, 2021.
- [15] S. Xua, H. K. Chanb dan T. Zhangb, “Forecasting the demand of the aviation industry using hybrid time series SARIMA-SVR approach,” *Transportation Research Part E*, vol. 122, p. 171, 2019.
- [16] A. U. Ukhra, “Pemodelan Dan Peramalan Data Deret Waktu Dengan Metode Seasonal ARIMA,” *Jurnal Matematika UNAND*, vol. 3, pp. 60 - 61, 2016.
- [17] N. E. Oktavia dan Amri, “ Analisis Kausalitas Antara Inflasi Dan Konsumsi Di Indonesia,” *Jurnal Ilmiah Mahasiswa (JIM) Ekonomi Pembangunan Fakultas Ekonomi dan Bisnis Unsyiah*, vol. 2, p. 169, 2017.
- [18] Y. Asriningtias dan M. Rudhyah, “Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa,” *JURNAL INFORMATIKA*, vol. 8, pp. 837-839, 2014.
- [19] R. A. Pitaloka, Sugito dan R. Rahmawati, “Perbandingan Metode Arima Box-Jenkins dengan ARIMA Ensemble Pada Peramalan Nilai Impor Provinsi Jawa Tengah,” *Journal Gaussian*, vol. 8, p. 1998, 2019.
- [20] B. Putro, M. T. Furqon dan S. H. Wijoyo, “Prediksi Jumlah Kebutuhan Pemakaian Air Menggunakan Metode Exponential Smoothing,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, p. 4681, 2018.

