

Perbandingan metode ARIMA dan ARIMAX dalam Memprediksi Jumlah Wisatawan Nusantara di Pulau Bali

Faiq Riestiansyah¹, Devina Damayanti², Miranda Reswara^{3*}, Ronny Susetyoko⁴

^{1,3}Jurusan Sains Data Terapan Politeknik Elektronika Negeri Surabaya

¹faiqfawwazain@gmail.com

²devinadyanti@gmail.com

^{3*}d4.r4.k0@gmail.com

⁴ronny@pens.ac.id

Abstrak—Indonesia memiliki berbagai potensi pemanfaatan yang berbeda tergantung dari sumber daya alamnya seperti bahan tambang, lahan pertanian, pariwisata dan lain-lain. Untuk meningkatkan pendapatan pada sektor pariwisata diperlukan data peramalan jumlah wisatawan yang berkunjung ke Pulau Bali. Data hasil peramalan tersebut dapat menjadi acuan untuk pengembangan dan pengoptimalisasian hal yang perlu diperbaiki di sektor kepariwisataan ini. Tujuan dari dilakukannya penelitian ini adalah untuk mengetahui perbandingan hasil prediksi terhadap Jumlah Wisatawan Nusantara yang berkunjung ke Pulau Bali. Salah satu model yang sering digunakan untuk masalah peramalan adalah model ARIMA. Model ARIMA yang juga disebut Runtut Waktu Box-Jenkins ini hanya cocok digunakan untuk kasus peramalan jangka pendek, karena jika digunakan untuk peramalan jangka panjang, model ini biasanya akan cenderung menghasilkan grafik time series datar. Setelah melakukan kedua pemodelan (ARIMA dan ARIMAX) selanjutnya membandingkan performa kedua model tersebut dalam melakukan prediksi Jumlah Wisatawan Nusantara yang berkunjung ke Pulau Bali dalam waktu tertentu dengan melihat error (RMSE) dari masing - masing model. Semakin rendah nilai RMSE maka semakin baik model tersebut bekerja dalam melakukan prediksi. Harapannya hasil dari penelitian ini dapat dimanfaatkan oleh siapapun yang memiliki kepentingan dalam pengembangan sektor pariwisata di Pulau Bali.

Kata kunci—ARIMA, ARIMAX, RMSE, Peramalan, Deret Waktu

Abstract—Indonesia has various potential uses that differ depending on its natural resources such as mining materials, agricultural land, tourism and others. To increase income in the tourism sector, data for forecasting the number of tourists visiting the island of Bali is needed. The forecasted data can be used as a reference for the development and optimization of things that need to be improved in the tourism sector. The purpose of this research is to compare the prediction results to the number of Indonesian tourists visiting the island of Bali. One model that is often used for forecasting problems is the ARIMA model. The ARIMA model which is also called the Box-Jenkins Time Series is only suitable for short-term forecasting cases, because if it is used for long-term forecasting, this model will usually tend to produce flat time series graphs. After doing the two models (ARIMA and ARIMAX) then compare the performance of the two models in predicting the number of Indonesian tourists visiting the island of Bali in a certain time by looking at the error (RMSE) of each model. The lower the RMSE value, the better the model works in making predictions. It is hoped that the results of this research can be used by anyone who has an interest in the development of the tourism sector on the island of Bali

I. PENDAHULUAN

Indonesia memiliki puluhan bahkan ribuan pulau yang tersebar dari Sabang sampai Merauke [4]. Setiap pulau yang ada di Indonesia memiliki berbagai potensi untuk pemanfaatan sumber daya alam seperti bahan tambang, lahan pertanian, pariwisata dan lain sebagainya yang dapat dikembangkan menjadi sumber ekonomi bagi Indonesia [6]. Salah satu pulau yang memiliki potensi pengembangan dan menjadi sumber pendapatan pada sektor pariwisata terbanyak di Indonesia adalah Pulau Bali. Perkembangan ekonomi suatu daerah yang memiliki angka penghasilan di bidang pariwisata akan dilihat dari jumlah wisatawan yang datang berkunjung dan juga pengeluaran pendanaan untuk berwisata ke daerah tersebut [9]. Untuk meningkatkan pendapatan pariwisata di perlukan Data Peramalan Jumlah Wisatawan yang berkunjung ke Pulau Bali. Data hasil peramalan tersebut dapat menjadi acuan untuk pengembangan dan pengoptimalisasian hal yang perlu diperbaiki di sektor kepariwisataan ini. Tujuan dilakukannya penelitian ini adalah untuk mengetahui Perbandingan Hasil Prediksi metode ARIMA dan ARIMAX terhadap Jumlah Wisatawan Nusantara yang berkunjung ke Pulau Bali. Harapannya hasil dari penelitian ini dapat dimanfaatkan oleh siapapun yang memiliki kepentingan dalam pengembangan sektor pariwisata di Pulau Bali.

II. METODOLOGI PENELITIAN

A. Analisis Prediktif

Analisis prediktif adalah metode analisis untuk mengolah sekumpulan data agar bisa memprediksi kondisi di masa yang akan datang. Metode ini biasa digunakan oleh para business intelligence untuk melihat kondisi bisnis berdasarkan data historis, yang di proses dengan berbagai model dan mesin pengolah canggih. Predictive analytics di lakukan untuk melihat kondisi di waktu yang akan datang dan berbagai kemungkinan yang dapat terjadi. Dengan begini, perusahaan bisa mempersiapkan hal - hal buruk yang mungkin terjadi sekaligus mempersiapkan solusinya. Dengan demikian, perusahaan akan lebih siap menghadapi tantangan bisnis di masa yang akan datang.

B. Analisis Time Series / Deret Waktu

Time series atau deret waktu merupakan pengamatan satu atau beberapa variabel yang diambil secara beruntun terhadap interval waktu yang tetap (Wei, 2006). Pada tahun 1976 George Box dan Gwilyn Jenkins memperkenalkan analisis time series untuk pertama kalinya. Analisis time series adalah salah satu prosedur statistika yang digunakan pada peramalan kejadian di masa depan. Analisis time series menggunakan data yang terpaut oleh waktu, sehingga korelasi antara

kejadian saat ini dengan periode waktu sebelumnya akan terjadi. Selain berhubungan antara waktu, time series juga terdapat kemungkinan adanya hubungan antara dimensi lain seperti wilayah ataupun dimensi lain yang saling berkaitan.

Salah satu modifikasi yang ingin diimplementasikan pada C. Prediksi

Prediksi adalah sebuah peramalan mengenai apa yang akan terjadi di masa depan dalam jangka waktu pendek (Taylor, Celuch, & Goodwin., 2004) [8]. Secara umum, ada dua jenis prediksi yaitu kualitatif dan kuantitatif. Prediksi kualitatif merupakan prediksi yang bersifat subjektif, hal ini karena didasarkan pada pengalaman empiris, intuisi pengambilan keputusan dan emosi manusia. Sedangkan, prediksi kuantitatif merupakan prediksi yang bersifat objektif sebab didasarkan pada data aktual dan diolah menggunakan metode tertentu.

D. Normalisasi

Normalisasi atau normalisasi data adalah suatu cara mengubah data ke menjadi nilai yang mudah dipahami (Patel & Mehta, 2011) [8]. Normalisasi bertujuan untuk mendapatkan data dengan ukuran yang lebih kecil yang mewakili data yang asli tanpa kehilangan karakteristik sendirinya Keseluruhan data akan dinormalisasi dengan fungsi linier (Parto & Sahu, 2015).

E. Model ARIMA

Salah satu model yang sering digunakan untuk masalah peramalan adalah model ARIMA. Model ARIMA yang juga disebut Runtut Waktu Box-Jenkins ini hanya cocok digunakan untuk kasus peramalan jangka pendek, karena jika digunakan untuk kasus peramalan jangka panjang, model ini biasanya akan cenderung menghasilkan grafik time series datar. Pada model ARIMA ini terdapat istilah - istilah berikut ini:

1) Autoregressive Model (AR)

Bentuk Model Autoregressive atau biasa disebut ARIMA (p, 0, 0) dinyatakan sebagai berikut:

$$Y_t = \alpha_0 + \theta_1 Y_{t-1} + \dots + \theta_p Y_{t-p} + et$$

Dengan :

Y_t = nilai observasi pada saat t

α_0 = konstanta

θ_p = parameter autoregressive ke p

et = nilai galat saat t

2) Moving Average (MA)

Bentuk umum Moving Average atau biasa disebut ARIMA (0,0,q) dinyatakan sebagai berikut :

$$Y_t = \theta_0 + \theta_1 et-1 + \dots + \theta_q et-q$$

Dengan :

θ_0 = konstanta

θ_q = parameter moving average ke q

$et-p$ = nilai galat pada saat $t - k$

3) Autoregressive Moving Average (ARMA)

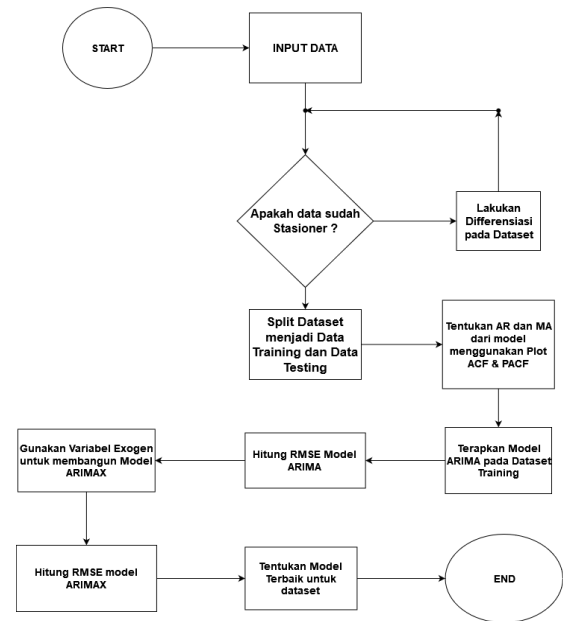
Model umum untuk kombinasi proses AR (1) dan MA (1) atau disebut ARIMA (1,0,1) dinyatakan sebagai berikut :

$$Y_t = \alpha_0 + \theta_1 et-1 + \dots + \theta_q et-q + \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p}$$

4) Autoregressive Integrated Moving Average (ARIMA) Apabila non-stasioner ditambahkan pada campuran proses ARMA, maka memenuhi model umum ARIMA (p, d, q) terpenuhi [1]. Persamaan untuk kasus sederhana ARIMA (p,1,q) adalah sebagai berikut :

$$Y_t = (1 + \alpha_1) y_{t-1} + \dots + (1 + \alpha_p) y_{t-p} + et + \theta_1 et-1 + \dots + \theta_q et-q$$

Metodologi yang digunakan pada penelitian ini adalah sebagai berikut.



Gambar 1. Menjelaskan tentang metodologi penelitian dengan menggunakan algoritma.

III. HASIL DAN PEMBAHASAN

A. Preprocessing Dataset

Sebelum memulai proses pembangunan model ARIMA dan ARIMAX, pertama difokuskan terlebih dahulu pada dataset yang dimiliki.

TABEL I
JUMLAH SEKUMPULAN DATA UNTUK KUNJUNGAN PARIWISATA YANG DATANG KE PULAU BALI.

Tahun	Jumlah.Kunjungan.WisNus	Tingkat.Inflasi
2018-01-30	743456	3.25
2018-02-28	655719	3.18
2018-03-30	762622	3.40
2018-04-30	777287	3.41
2018-05-30	682521	3.23
2018-06-30	1156151	3.12

Seperti yang terlihat, bahwa dataset memiliki 3 variabel. Namun yang harus diperhatikan di sini adalah variabel Jumlah Kunjungan Wisata Nusantara dan Tingkat Inflasi saja karena variabel Tahun hanya bekerja sebagai index agar pada saat ingin melakukan visualisasi gambar akan lebih relevan. Perlu di ingat juga dalam melakukan Time Series Analysis maka hal pertama adalah mengubah format dataset yang semula awalnya data frame menjadi format Time Series dengan menggunakan Tsibble.

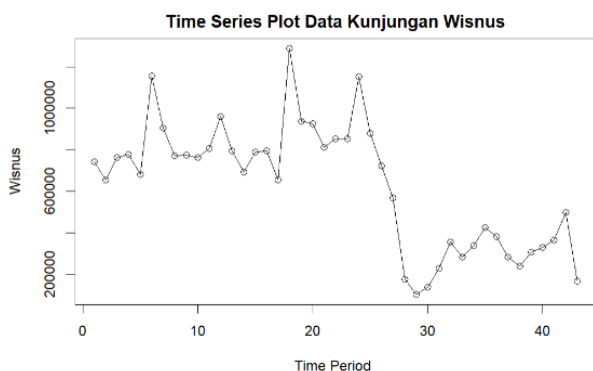
TABEL II
DATA FRAME SUDAH DIUBAH MENJADI DATA TIME SERIES

Tahun	Jumlah.Kunjungan.WisNus	Tingkat.Inflasi
2018 Jan	743456	3.25
2018 Feb	655719	3.18
2018 Mar	762622	3.40
2018 Apr	777287	3.41
2018 May	682521	3.23
2018 Jun	1156151	3.12

Gambar di atas menunjukkan dataset yang sudah di konversi menjadi format Time Series dengan index per bulan.

B. Uji Stasioneritas Data

. Setelah melalui proses *preprocessing* dataset, langkah selanjutnya adalah mencoba memenuhi asumsi dalam pemodelan ARIMA yaitu data bersifat stasioner. Data yang bersifat stasioner adalah saat suatu data tidak mengalami kenaikan atau penurunan atau rata – rata dan varians pada suatu dataset dan setiap lag akan bersifat tetap atau sama. Untuk melihat apakah dataset yang dimiliki sudah bersifat Stasioner atau belum, di sini akan tampilkan Plot Data Time Series.



Gambar 2. Data Time Series yang berbentuk Grafik Plot.

Dapat di lihat pada grafik di atas, Data Time Series terlihat masih belum Stasioner karena data cenderung mengalami penurunan setelah lag 25. Untuk lebih memastikan bahwa data ini belum Stasioiner, lakukan Uji Statistik Dickey-Fuller Test pada fitur tersebut.

Augmented Dickey-Fuller Test

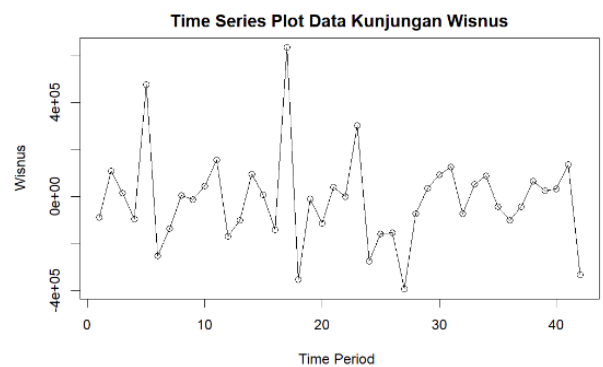
```
data: train$Jumlah.Kunjungan.WisNus
Dickey-Fuller = -2.6839, Lag order = 3, p-value = 0.3036
alternative hypothesis: stationary
```

Gambar 3. Uji statistik Dickey-Fuller Test

Dapat disimpulkan bahwa hasil dari Uji Dickey-Fuller Test tersebut memiliki p-value sebesar 0.30 atau lebih dari 5% (0.05) yang berarti data tersebut memang belum Stasioner.

C. Diferensiasi Data

Dari hasil pengamatan Plot Time Series dan hasil Uji Dickey-Fuller Test di atas yang menyatakan data belum Stasioner maka diperlukan langkah untuk membuat dan menjadi Stasioner, proses yang perlu dilakukan adalah proses Differencing. Proses Differencing adalah proses yang digunakan untuk menghilangkan ketergantungan deret pada waktu atau juga bisa disebut keuntungan temporal. Ketergantungan ini termasuk struktur tren dan musim. Hasil dari data yang sudah melalui proses Diferensiasi satu kali berikutnya ditampilkan berupa grafik berikut.



Gambar 4. Data Time Series Grafik Plot setelah melalui proses Diferensiasi.

Augmented Dickey-Fuller Test

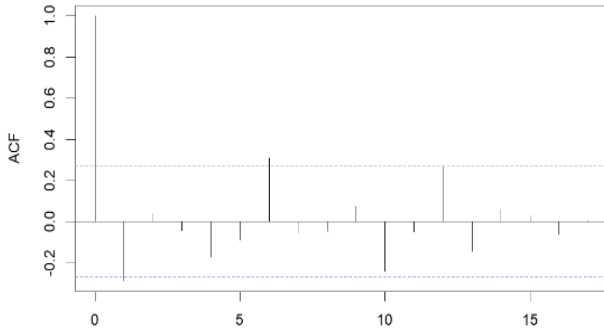
```
data: data.diffed
Dickey-Fuller = -3.8254, Lag order = 3, p-value = 0.02725
alternative hypothesis: stationary
```

Gambar 5. Augmentasi menggunakan metode Dickey-Fuller Test

Dan terbukti, setelah dilakukan Uji Dickey-Fuller Test pada data yang sudah dilakukan diferensiasi menunjukkan data sudah stasioner sehingga dapat dilanjutkan ke proses berikutnya.

D. Membangun Model ARIMA

Dalam menerapkan model ARIMA (p, d, q) Langkah pertama yang harus dilakukan adalah menentukan nilai p dan q. Untuk mencari nilai p dan q tersebut perlu di tampilkan grafik ACF dan PACF. Berikut adalah grafik plot ACF yang dapat dijadikan acuan.



Gambar 6. Grafik Plot ACF

Berdasarkan plot ACF di atas dapat disimpulkan bahwa pada lag 1 terjadi perubahan nilai yang signifikan sehingga kita dapat menggunakan 1 sebagai nilai MA (q). Maka untuk asumsi awal kita memiliki ARIMA (0, 1, 1) dan ARIMA (1, 1, 1), bisa dilakukan estimasi model untuk membandingkan signifikansi kedua model tersebut.

```
Call:
arima(x = data_arima$Jumlah.Kunjungan.WisNus, order = c(0, 1, 1))

Coefficients:
      ma1
    -0.3157
s.e.    0.1431

sigma^2 estimated as 3.256e+10: log likelihood = -716.72, aic = 1437.44

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ma1 -0.31575    0.14312  -2.2062  0.02737 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Gambar 7. Code ARIMA

```
Call:
arima(x = data_arima$Jumlah.Kunjungan.WisNus, order = c(1, 1, 1))

Coefficients:
      ar1      ma1
    0.3537  -0.6527
s.e.    0.3930    0.3312

sigma^2 estimated as 3.228e+10: log likelihood = -716.53, aic = 1439.06

z test of coefficients:

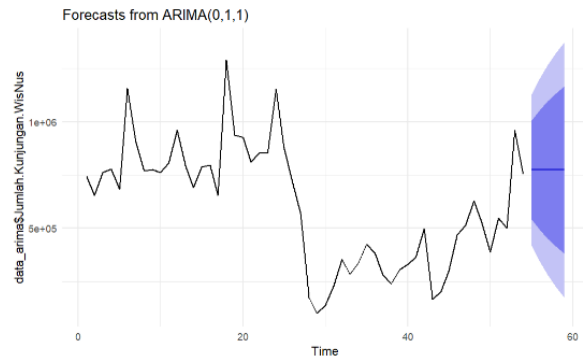
      Estimate Std. Error z value Pr(>|z|)
ar1  0.35368    0.39303  0.8999  0.36818
ma1 -0.65272    0.33116 -1.9710  0.04872 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Gambar 8. Code ARIMAX

TABEL III
PERBANDINGAN HASIL ANTARA NILAI ARIMA DAN ARIMAX.

Model	P-Value	Keterangan
ARIMA (0,1,1)	0.027	Signifikan
ARIMA (1,1,1)	0.368	Kurang Signifikan

E. Peramalan dengan Model ARIMA



Gambar 9. Grafik Forecasts ARIMA

Dapat dilihat dari grafik peramalan di atas bahwa hasil peramalan cenderung datar jika menggunakan model ARIMA.

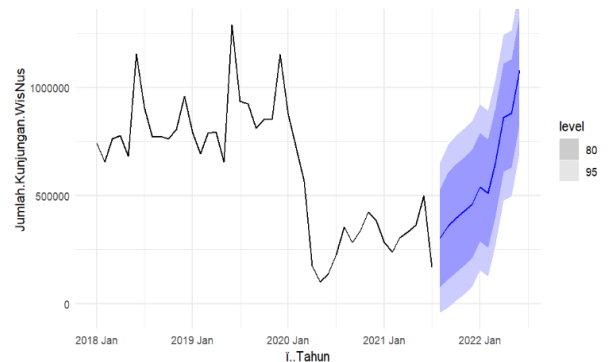
F. Membangun Model dan Peramalan dengan ARIMAX

Dalam penerapan Model ARIMAX, hanya perlu satu variabel / fitur tambahan yang biasa dinamakan variabel eksogen. Variabel tersebut akan digunakan untuk pertimbangan dalam melakukan prediksi kedepannya. Pada permasalahan ini, mencoba menggunakan tingkat inflasi sebagai variabel eksogen tersebut.

i. Tahun <S3: yearmonth>	Jumlah.Kunjungan.WisNus <int>	Tingkat.Inflasi <dbl>
2021 Aug	202187	1.59
2021 Sep	298950	1.60
2021 Oct	468826	1.66
2021 Nov	513482	1.75
2021 Dec	629590	1.87
2022 Jan	527447	2.18
2022 Feb	389690	2.06
2022 Mar	547726	2.64
2022 Apr	500740	3.47
2022 May	960692	3.55

Gambar 10. Hasil Peramalan

Gambar di atas adalah hasil Peramalan Jumlah Kunjungan Wisatawan Nusantara dengan tingkat inflasi Indonesia sebagai variabel eksogen.



Gambar 11. Prediksi Grafik ARIMAX

Dari grafik di atas dapat dilihat bahwa prediksi dengan menggunakan ARIMAX menghasilkan grafik yang lebih masuk akal dibandingkan grafik prediksi dengan ARIMA yang cenderung datar.

G. *Perbandingan Model ARIMA vs ARIMAX*

Setelah melakukan kedua pemodelan (ARIMA dan ARIMAX) selanjutnya dapat membandingkan performa kedua model tersebut dalam melakukan prediksi Jumlah Wisatawan Nusantara yang berkunjung ke Pulau Bali dalam waktu tertentu dengan melihat error (RMSE) dari masing masing model. Semakin rendah nilai RMSE maka semakin baik model tersebut bekerja dalam melakukan prediksi.

RMSE
178757.2

Gambar 12. Hasil RMSE

"RMSE model ARIMAX: 171266.36"

Gambar 13. Hasil RMSE Model ARIMAX

TABEL IV
Hasil RMSE ARIMA dan ARIMAX

Model	RMSE	MAE	MAPE
ARIMA (0,1,1)	178757	128288.5	29.933
ARIMA (1,1,1)	178007.9	123941.2	29.78018
ARIMAX (1,0,0)	171266	135816.52	27.76

IV. KESIMPULAN

Berdasarkan hasil perbandingan dan prediksi data Wisatawan di Pulau Bali tahun 2018 - 2022 dengan menggunakan Metode ARIMA dan ARIMAX dapat disimpulkan sebagai berikut:

1. Jumlah pengunjung Wisatawan di Pulau Bali pada tahun 2018 – 2022 memiliki variasi yang berbeda – beda.

2. Hasil ramalan jumlah pengunjung wisatawan di Pulau Bali pada tahun 2018 - 2022 dengan menggunakan model ARIMA (0,1,1) dapat dilihat dari grafik peramalan di atas bahwa hasil peramalan cenderung datar.
3. Untuk hasil ramalan jumlah pengunjung wisatawan di Pulau Bali pada tahun 2018 – 2022 dengan menggunakan model ARIMAX dapat dilihat pada grafik di atas bahwa prediksi dengan menggunakan ARIMAX menghasilkan grafik yang lebih masuk akal dibandingkan grafik prediksi dengan ARIMA yang cenderung datar.
4. Hasil setelah mencari prediksi menggunakan error (RMSE) dari masing – masing model tersebut terlihat bahwa semakin nilai RMSE rendah maka akan semakin baik model tersebut bekerja dalam melakukan prediksi.

REFERENSI

- [1] Nurfitri & Shantika, M. 2020. PEMODELAN DATA RUNTUN WAKTU DENGAN ARIMAX. (Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster)).
- [2] Tirta & Moh, Y. 2018. Pemilihan Model Peramalan Terbaik Menggunakan Model Arima dan Winters Untuk Meramalkan Indeks LQ45
- [3] B. B. O. R. E. Putri and Noeryanti, "Pemodelan Time Series Dalam Peramalan Jumlah Pengunjung Objek Wisata Di Kabupaten Gunung Kidul Menggunakan Metode ARIMAX Efek Variasi Kalender," J. Stat. Ind. dan Komputasi, vol. 4, no. 1, pp. 81–88, 2019.
- [4] N. P. N. Hendayanti and M. Nurhidayati, "Perbandingan Metode Seasonal Autoregressive Integrated Moving Average (SARIMA) dengan Support Vector Regression (SVR) dalam Memprediksi Jumlah Kunjungan Wisatawan Mancanegara ke Bali," J. Varian, vol. 3, no. 2, pp. 149–162, 2020, doi: 10.30812/varian.v3i2.668.
- [5] Hendayanti, N. P. N., Suniantara, I. K. P., & Nurhidayati, M. (2019). Penerapan Support Vector Regression (Svr) Dalam Memprediksi Jumlah Kunjungan Wisatawan Domestik Ke Bali. Jurnal Varian, 3(1), 43-50. <https://doi.org/10.30812/varian.v3i1.506>
- [6] A. Pratama, R. C. Wihandika, and D. E. Ratnawati, "Implementasi Algoritme Support Vector Machine (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa," J. Pengemb. Teknol. Inf. dan Ilmu Komput. e-ISSN, 2548, 964X, vol. 2, no. 4, pp. 1704–1708, 2018.
- [7] L. Surtiningsih, M. T. Furqon, and S. Adinugroho, "Prediksi Jumlah Kunjungan Wisatawan Mancanegara Ke Bali Menggunakan Support Vector Regression dengan Algoritma Genetika," J. Pengemb. Teknol. Inf. dan Ilmu Komput., vol. 2, no. 8, pp. 2578–2586, 2018.
- [8] Hadiriyanto, I., Darsyah, M. Y., & Semarang, U. M. (2018). Peramalan Jumlah Wisatawan Mancanegara di Provinsi Bali dengan Menggunakan ARIMA dan Winter. Prosiding Seminar Nasional Mahasiswa Unimus, 1, 405–411.
- [9] M. Ibrahim, H. R. P. Negara, and ..., "Prediction of Land Area Harvest, Production, Rice Productivity: A Accuracy Analysis of ARIMA Methods," ... Biosyst. J., 2021.
- [10] F. Fejriani, M. Hendrawansyah, L. Muharni, S. F. Handayani, and Syaharuddin, "Forecasting Peningkatan Jumlah Penduduk Berdasarkan Jenis Kelamin menggunakan Metode Arima," Jurnal Kajian, Penelitian dan Pengembangan Pendidikan, vol. 8, no. 1 April. pp. 27–36, 2020.
- [11] R. Perwira Negara, "Computational Modeling of ARIMA-based G-MFS Methods: Long-term Forecasting of Increasing Population," Int. J. Emerg. Trends Eng. Res., vol. 8, no. 7, pp. 3665–3669, 2020, doi 10.30534/ijeter/2020/126872020.