

Penerapan konsep data mining dengan Metode Seasonal ARIMA dalam Peramalan Produksi Padi

Jefri Esna Thomas Radjabaycolle^{*1}, Devi Valentino Waas², Victor Eric Pattiradjawane³, Emanuella M. C. Wattimena⁴, Doms Upuy⁵, Citra Fathia Palembang⁶

^{1,2,3,4,5,6}Program Studi Ilmu Komputer, Universitas Pattimura, Ambon

e-mail: ^{*1}jefri.radjabaycolle@lecturer.unpatti.ac.id, ²devi.waas@lecturer.unpatti.ac.id,

³victoreric82@gmail.com, ⁴emanuella.wattimena@lecturer.unpatti.ac.id,

⁵doms.upuy@lecturer.unpatti.ac.id, ⁶citra.palembang@lecturer.unpatti.ac.id

Abstrak

Pendekatan data mining menjadi elemen kunci dalam analisis data skala besar untuk mendapatkan informasi yang berguna dalam berbagai bidang, termasuk sektor pertanian. Melalui penerapan teknik data mining, penelitian ini bertujuan untuk mengeksplorasi, mengolah, dan menganalisis data produksi padi di kecamatan Denpasar Selatan dan kecamatan Denpasar Timur, Provinsi Bali, guna mendukung pengambilan keputusan yang lebih akurat. Metode yang digunakan dalam penelitian ini adalah Seasonal ARIMA (SARIMA), yang secara khusus dirancang untuk menangani pola musiman pada data time series. Penelitian ini menitikberatkan pada proses pengolahan data, pemilihan model prediktif yang tepat, dan evaluasi kinerja model. Model SARIMA yang dipilih untuk kedua kecamatan yaitu SARIMA $(0,0,1)(1,1,1)^6$ untuk kecamatan Denpasar Selatan dan SARIMA $(0,0,0)(1,1,0)^6$ untuk kecamatan Denpasar Timur. Hasil penelitian menunjukkan bahwa model SARIMA mampu memberikan prediksi produksi padi yang baik, sehingga dapat menggambarkan tren kenaikan atau penurunan produksi pada periode berikutnya. Temuan ini menunjukkan bahwa penerapan metode SARIMA yang didukung oleh teknik data mining dapat menjadi alat bantu yang efektif untuk analisis data produksi padi.

Kata kunci— Produksi Padi, Data Mining, Peramalan, Seasonal Arima

1. PENDAHULUAN

Berdasarkan data dari Badan Pusat Statistika (BPS) Provinsi Bali tahun 2022 yang dikutip dalam keputusan Menteri Dalam Negeri Nomor 100.1.1-6117 Tahun 2022, Kota Denpasar memiliki luas wilayah 125,87 km², dengan pemanfaatan lahan pertanian berupa sawah sebesar 1.871 hektar dan lahan bukan sawah sebesar 510 hektar pada tahun 2022 berdasarkan data dari Dinas Pertanian Kota Denpasar. Luas lahan sawah ini sebagian besar dimanfaatkan untuk pertanian produksi padi. Produksi padi di Kota Denpasar berdasarkan data dari Dinas Pertanian Kota Denpasar pada tahun 2020 sebesar 27.912 ton, pada tahun 2021 sebesar 25.337,62 ton dan pada tahun 2022 sebesar 28.236 ton. Karena produksi padi di Kota Denpasar setiap tahun selalu mengalami perubahan, berdasarkan permasalahan tersebut maka dilakukan sebuah penelitian peramalan produksi padi untuk mengetahui gambaran apakah hasil produksi padi di wilayah Kota Denpasar akan mengalami kenaikan ataupun penurunan pada periode selanjutnya. Penelitian ini memanfaatkan data produksi padi dari Dinas Pertanian Kota Denpasar sebagai data acuan dalam melakukan peramalan hasil produksi padi di Kota Denpasar dimasa mendatang.



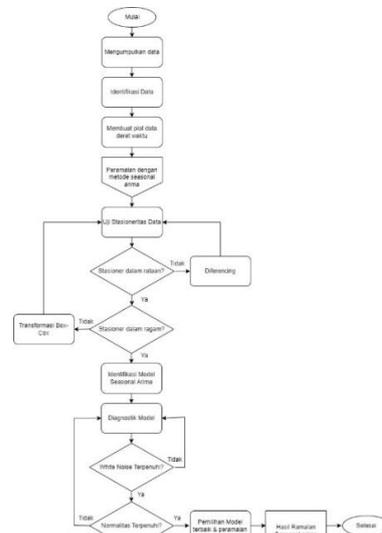
Gambar 1. Plot Data Produksi Padi Kota Denpasar Tahun 2019-2022

Berdasarkan gambar 1. terlihat bahwa grafik data produksi padi memiliki pola musiman karena kecenderungan data *time series* yang berulang setiap periode [1]. Data musiman didefinisikan sebagai data dengan pola yang berulang-ulang dalam selang waktu yang tetap [2]. Runtun waktu musiman mempunyai karakteristik yang ditunjukkan oleh adanya korelasi beruntun yang kuat pada jarak semusim. [3]. Peramalan untuk data yang memiliki pola musiman dikembangkan dengan berbagai metode, berdasarkan penelitian yang dilakukan oleh [4] metode *Seasonal ARIMA* memiliki kelebihan relatif mudah untuk digunakan dalam memprediksi data runtun waktu berpola musiman.

2. METODE PENELITIAN

2.1. Tahapan Penelitian

Tahapan pada penelitian ini mengikuti flowchart pada Gambar 1 berikut ini:



Gambar 2. Flowchart Penelitian

Pada gambar2. menunjukan tahapan penelitian peramalan produksi padi dengan menggunakan metode *Seasonal ARIMA*

Tahapan pada penelitian ini umumnya mengikuti langkah-langkah sebagai berikut: [5]

1. Uji Normalitas

Uji normalitas digunakan untuk mengetahui kenormalan suatu data, dimana kenormalan suatu data juga dapat mengindikasikan bahwa data tersebut stasioner dalam varians.

2. Stasioneritas dalam mean

Data dikatakan stasioner dalam mean apabila fluktuasi data berada di sekitar suatu nilai rata-rata yang konstan.

3. Differencing

Differencing merupakan bentuk transformasi untuk menstasionerkan data runtun waktu yang tidak stasioner dalam mean. Identifikasi model dilakukan setelah dipastikan bahwa data telah stasioner. Selanjutnya ditentukan model sementara dengan menentukan nilai $(p,d,q)(P,D,Q)^s$. Penentuan nilai tersebut dilakukan dengan mengamati grafik fungsi ACF (*Autocorelation Function*) dan PACF (*Partial Autocorelation Function*). Nilai AR (p) dapat ditentukan dengan memperhatikan grafik PACF non-musiman, nilai MA (q) dapat ditentukan dengan memperhatikan grafik fungsi ACF, sedangkan d merupakan banyaknya proses *differencing* yang di lakukan. Pemilihan model ARIMA dapat dipilih dengan beberapa kriteria dapat dilihat pada tabel 1 sebagai berikut

Tabel 1. Kriteria pemilihan model ARIMA

	Model	ACF	PACF
Non Musiman	AR (p)	Menurun secara perlahan setelah lag p (<i>dies down</i>)	Terputus setelah lag p (<i>cut off</i>)
	MA (q)	Terputus setelah lag q (<i>cut off</i>)	Menurun secara perlahan setelah lag q (<i>dies down</i>)
	ARMA (p,q)	Menurun secara perlahan setelah lag p,q (<i>dies down</i>)	Menurun secara perlahan setelah lag p,q (<i>dies down</i>)
	ARIMA (p,d,q)	Menurun secara perlahan setelah lag (<i>dies down</i>) setelah <i>differencing</i>	Menurun secara perlahan setelah lag (<i>dies down</i>) setelah <i>differencing</i>
Musiman	SAR (P)	Menurun secara perlahan setelah lag P^s (<i>dies down</i>)	Terputus setelah lag P^s (<i>cut off</i>)
	SMA (Q)	Terputus setelah lag Q^s (<i>cut off</i>)	Menurun secara perlahan setelah lag Q^s (<i>dies down</i>)
	SARMA (P,Q)	Menurun secara perlahan setelah lag $(P,Q)^s$ (<i>dies down</i>)	Menurun secara perlahan setelah lag $(P,Q)^s$ (<i>dies down</i>)
	SARIMA (P,D,Q)	Menurun secara perlahan setelah lag (<i>dies down</i>) setelah <i>differencing</i>	Menurun secara perlahan setelah lag (<i>dies down</i>) setelah <i>differencing</i>

4. Uji Diagnostik

Uji diagnostik digunakan untuk melihat apakah residual dari model bersifat white noise atau tidak. White noise menunjukkan bahwa suatu data bersifat random dan stasioner, dimana ini merupakan syarat untuk melakukan peramalan [6]

5. Pemilihan Model terbaik dan peramalan

Pemilihan model terbaik dilihat berdasarkan nilai MAPE terkecil. Nilai MAPE diperoleh melalui perbandingan data out sample dengan data hasil peramalan dari setiap model dan dicari nilai error nya.

3. HASIL DAN PEMBAHASAN

3.1. Data Bulanan Produksi padi Kota Denpasar

Tabel 2. Data Produksi Padi Kecamatan Denpasar Utara dan Denpasar Timur

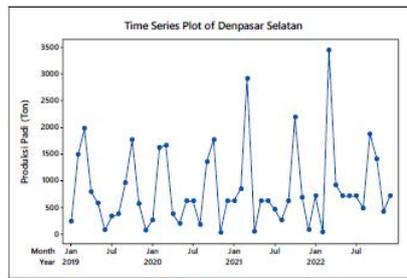
Tahun	Bulan	Denpasar Utara	Denpasar Timur
2019	Januari	423	192
	Februari	252	484
	Maret	1156	993
	April	2535	1564
	Mei	95	460
	Juni	658,42	227
	Juli	123	116
	Agustus	319	91
	September	492	449
	Oktober	1674	1005
	November	529	1387
	Desember	303	235
	<i>Mean</i>	658,42	600,25
	dst	dst	dst
2022	Januari	169	447
	Februari	70	212
	Maret	2316	999
	April	1113	829
	Mei	421	369
	Juni	169	264
	Juli	747	165
	Agustus	302	620

September	951	746
Oktober	2134	1123
November	784	492
Desember	535	958
<i>Mean</i>	747	602

3.2. Perhitungan Metode Seasonal ARIMA

Proses perhitungan metode seasonal arima akan dikerjakan untuk kecamatan Denpasar selatan dan kecamatan Denpasar Timur. Sebagai contoh akan tunjukkan langkah-langkah pengerjaan dari kecamatan Denpasar Selatan.

1. Identifikasi plot data



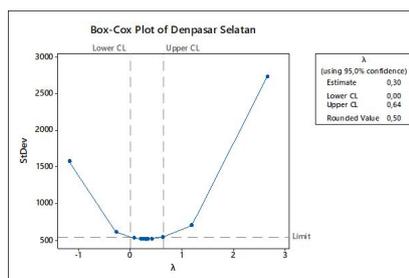
Gambar 3. Plot data produksi padi kecamatan Denpasar Selatan transformasi 1

Gambar 3 menunjukkan plot data deret waktu produksi padi kecamatan Denpasar Selatan periode Januari 2019 – Desember 2022 yang memiliki 48 data. Plot tersebut menunjukkan adanya pola data musiman dilihat dari adanya data naik turun pada periode tertentu yang menandakan bahwa data tersebut berpola musiman. [7]. Selanjutnya data dibagi menjadi dua bagian yaitu data *in sample* dan *out sample*. Data *in sample* digunakan untuk membentuk model yang kemudian digunakan untuk melakukan peramalan, sedangkan data *out sample* digunakan untuk mengevaluasi hasil peramalan dari model yang diperoleh dan untuk membandingkan dengan hasil peramalan [8].

2. Uji Stasioneritas Data

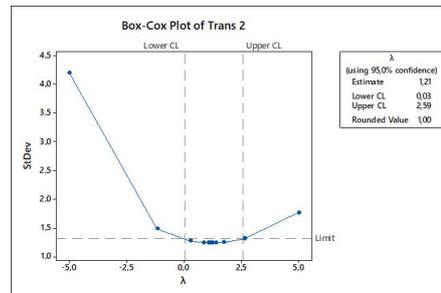
a. Stasioner terhadap varian

Berikut ini merupakan hasil uji kestasioneran data terhadap varian dengan menggunakan transformasi Box-Cox. Data dikatakan stasioner jika nilai *Rounded Value* atau $\lambda = 1,00$ [9]



Gambar 4. Plot data Box-Cox produksi padi Denpasar Selatan

Pada Gambar 4 diperoleh nilai pada *Rounded Value* atau $\lambda = 0,50$, sehingga perlu dilakukan transformasi 2 pada data karena belum stasioner

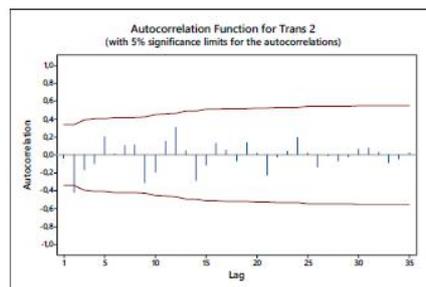


Gambar 5. Plot Box-Cox produksi padi Denpasar selatan transformasi 2

Pada gambar 5. Diperoleh nilai *Rounded Value* atau $\lambda = 1,00$. Hal ini menunjukkan bahwa data telah stasioner terhadap varian setelah dilakukan transformasi kedua.

b. Stasioner terhadap rata-rata

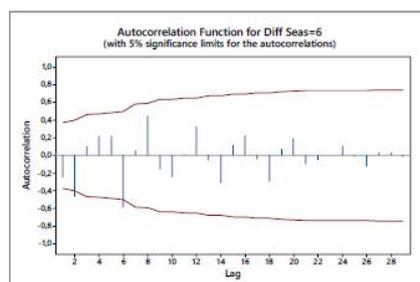
Data dikatakan stasioner terhadap rata-rata dapat dilihat dari plot ACF (*Autocorrelation Function*), lag pada plot ACF menunjukkan nilai autokorelasi pada data. [10]



Gambar 6. Plot ACF Non Musiman

Pada gambar 6. Memperlihatkan plot ACF data produksi padi Denpasar Selatan setelah dilakukan *differencing* musiman yang menunjukkan jika terdapat 1 lag yang keluar dari *confidence interval* (garis selang kepercayaan), sehingga pada plot ACF non musiman dikatakan sudah stasioner terhadap rata-rata. Maka tidak perlu dilakukan differencing sehingga nilai (d) pada model ARIMA adalah 0.

Selanjutnya dilakukan proses differencing musiman untuk mengidentifikasi model musiman. Proses differencing musiman dilakukan dengan melakukan pergeseran data sebanyak 6 periode kebelakang. Plot ACF setelah differencing musiman dapat dilihat pada Gambar 7.



Gambar 7. Plot ACF differencing musiman

Pada gambar 4.32 memperlihatkan plot ACF data produksi padi Denpasar Selatan setelah dilakukan *differencing* musiman yang menunjukkan jika terdapat lag yang melewati *confidence interval* yaitu lag ke-6 yang artinya lag tersebut signifikan atau stasioner terhadap rata-rata, sehingga nilai (D) pada model ARIMA adalah 1.

3. Identifikasi model sementara

Setelah dilakukan uji stasioner terhadap varian dan rata-rata, selanjutnya dilakukan identifikasi model sementara yang dilihat berdasarkan plot ACF maupun PACF yang telah stasioner yang dinotasikan sebagai ARIMA $(p,d,q)(P,D,Q)^s$, dimana (p) merupakan AR non musiman, (q) merupakan MA non musiman, (d) merupakan *differencing* pada non musiman. Sedangkan (P) merupakan AR musiman, (Q) merupakan MA musiman, dan (D) merupakan *differencing* pada musiman dan (s) adalah perlambangan dari periode musimannya.

Berdasarkan plot ACF dan PACF non musiman dan musiman diperoleh model ARIMA sementara yang mungkin yaitu ARIMA $(1,0,1)(1,1,0)^6$, ARIMA $(1,0,1)(1,1,1)^6$, ARIMA $(1,0,1)(2,1,0)^6$, dan ARIMA $(1,0,1)(2,1,1)^6$ dan model lainnya yang didapat melalui *trial and error* pada *software* minitab18 yaitu ARIMA $(1,0,0)(1,1,0)^6$ dan ARIMA $(0,0,1)(1,1,1)^6$.

4. Uji Signifikan parameter model

Pada uji signifikansi parameter model diperoleh 2 model sementara yang memiliki parameter signifikan yaitu model ARIMA $(1,0,0)(1,1,0)^6$ karena memiliki nilai *p-value* $0,022 < 0,05$ dan nilai *p-value* $0,000 < 0,05$ sehingga model telah lolos uji signifikan. Pada model ARIMA $(0,0,1)(1,1,1)^6$ memiliki nilai *p-value* $0,000 < 0,05$, nilai *p-value* $0,013 < 0,05$ dan nilai *p-value* $0,002 < 0,05$ yang artinya model ini juga telah lolos uji signifikan. Selanjutnya kedua model tersebut dilakukan uji asumsi residual.

5. Uji Asumsi Residual

Dalam menentukan model ARIMA yang terbaik, harus dipilih model yang seluruh parameternya signifikan, kemudian juga memenuhi 2 asumsi residual yaitu berdistribusi normal dan white noise [11]. Setelah dilakukan pengujian, harus memenuhi syarat white noise dengan nilai *p-value* yang lebih besar dari 0,05. Model dikatakan memenuhi asumsi White noise jika nilai *p value* lebih dari 0,05 [12]. Hasil uji white noise pada tabel 3.

Tabel 3. Hasil Uji White Noise

Model	Lag	Chi-Square	DF	p-value	Keterangan
$(0,0,0)(1,1,0)^6$	12	13,39	10	0,174	White Noise
	24	19,58	22	0,609	White Noise
$(0,0,0)(0,1,1)^6$	12	13,89	9	0,126	White Noise
	24	25,45	21	0,228	White Noise

Selanjutnya dilakukan uji distribusi normal pada kedua model sementara dan diperoleh hasil pada model ARIMA $(1,0,0)(1,1,0)^6$ telah berdistribusi normal dengan *p-value* $0,15 > 0,05$, sedangkan model ARIMA $(0,0,1)(1,1,1)^6$ telah berdistribusi normal dengan *p-value* $0,15 < 0,05$

6. Perhitungan nilai error

Selanjutnya dilakukan pemilihan model terbaik dengan melihat presentasi nilai eror terkecil dengan menggunakan MAPE. MAPE digunakan karena dapat memberikan gambaran seberapa akurat prediksi model jika dibandingkan dengan nilai actual yang sebenarnya [13]. Semakin kecil kuadrat kesalahan, maka model yang digunakan akan semakin baik untuk dilakukan peramalan. Suatu model mempunyai kinerja sangat bagus apabila nilai MAPE berada di bawah 10%, dan mempunyai kinerja bagus apabila nilai MAPE berada pada rentang 10% hingga 20% [14].

Dengan menggunakan rumus $MAPE = \left(\frac{\sum APE}{n} \right)$ [15] diperoleh perbandingan kedua model ARIMA yang ditunjukkan pada table berikut 4.

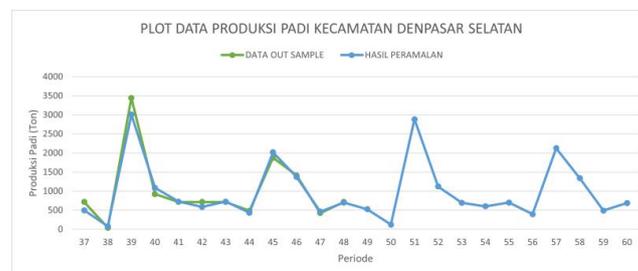
Tabel 4. Perbandingan model ARIMA

Model	MAPE
ARIMA (1,0,0)(1,1,0) ⁶	18,85%
ARIMA (0,0,1)(1,1,1) ⁶	17,38%

Berdasarkan tabel 4, model yang dipilih menjadi model yang terbaik adalah ARIMA (0,0,1)(1,1,1)⁶ dengan nilai MAPE sebesar 17,38%

7. Hasil Peramalan

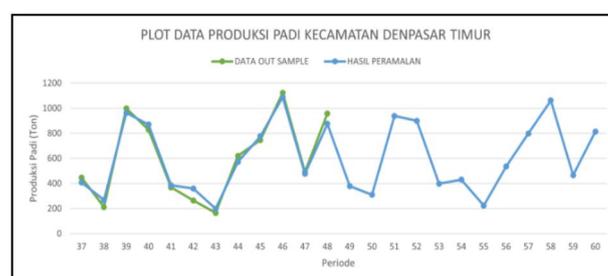
a. Hasil peramalan metode seasonal ARIMA (0,0,0)(1,1,0)⁶ pada produksi pada kecamatan Denpasar utara



Gambar 8. Plot perbandingan data out sample dan hasil peramalan model ARIMA (0,0,1)(1,1,1)⁶

Pada gambar 8 menunjukkan perbedaan plot data hasil peramalan mendekati plot data *out sample* karena pada model ARIMA (0,0,1)(1,1,1)⁶ menghasilkan nilai MAPE 17,38% yang artinya hasil peramalan baik. Selanjutnya dengan langkah perhitungan yang sama, dilakukan perhitungan untuk kecamatan Denpasar timur.

b. Hasil peramalan produksi padi kecamatan Denpasar Timur menggunakan metode Seasonal ARIMA (0,0,0)(1,1,0)⁶



Gambar 9. Plot Perbandingan data out sample dan hasil peramalan model ARIMA (0,0,0)(1,1,0)⁶

Pada gambar 9. menunjukkan perbedaan plot data hasil peramalan mendekati plot data *out sample* karena pada model ARIMA (0,0,0)(1,1,0)⁶ menghasilkan nilai MAPE 10,91% yang artinya hasil peramalan baik.

4. KESIMPULAN

1. Penelitian ini berhasil mengimplementasikan metode Seasonal ARIMA untuk peramalan produksi padi di kecamatan Denpasar Selatan dan Denpasar Timur, Provinsi Bali. Hasil penelitian menunjukkan bahwa model SARIMA yang dioptimalkan untuk masing-masing kecamatan, yaitu Seasonal ARIMA(0,0,1)(1,1,1)⁶ untuk kecamatan Denpasar selatan dan model seasonal ARIMA (0,0,0)(1,1,0)⁶ untuk kecamatan Denpasar Timur dapat memberikan hasil prediksi yang cukup akurat. Model ini dapat menggambarkan tren produksi padi, baik yang cenderung naik maupun turun pada periode berikutnya.
2. Penerapan metode SARIMA dalam penelitian ini membuktikan bahwa pendekatan ini efektif dalam menangani data time series yang memiliki pola musiman. Dengan hasil prediksi yang akurat, pemerintah daerah dan para pelaku pertanian dapat memanfaatkan informasi ini untuk pengambilan keputusan terkait pengelolaan atau strategi peningkatan produksi padi.

5. SARAN

1. Dapat melakukan analisis terhadap faktor-faktor eksternal yang dapat mempengaruhi produksi padi, seperti iklim, kebijakan pemerintah, dan dinamika pasar. Dengan memasukkan variabel eksternal ke dalam model, hasil peramalan dapat lebih menggambarkan kondisi sebenarnya.
2. Dapat melakukan perbandingan metode seasonal ARIMA dengan metode lainnya seperti LSTM (Long Short-Term Memory), Exponential Smoothing dan lainnya. Hal ini akan membantu mengevaluasi kelebihan dan kekurangan masing-masing metode serta memilih metode yang paling sesuai dengan karakteristik data produksi padi.
3. Peramalan produksi padi dapat juga dilakukan untuk kecamatan Denpasar barat, kecamatan Denpasar utara dan kota Denpasar.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Dinas Pertanian Kota Denpasar yang telah memberi dukungan informasi berupa data penelitian yang sangat membantu jalannya penelitian ini.

DAFTAR PUSTAKA

- [1] Anis Mahfud Al'afi, Widiarti, Dian Kurniasari, Mustofa Usman. 2020. Peramalan data Time series Seasonal Menggunakan Metode Analisis Spektral. Jurnal Siger Matematika, Vol. 01, No. 01
- [1] Lukmaini, S., Nugraheni, K., dkk. 2023. Peramalan Inflasi Provinsi Kalimantan Timur Tahun 2016-2022 Menggunakan Metode Seasonal Autoregressive Integrated Moving Average (SARIMA). Prosiding Seminar Nasional Matematika, Statistika, dan Aplikasinya. Terbitan III
- [2] Anggraeni, A. S., Chandra Utama, R., dkk. 2022. Penghalusan Eksponensial dan

- Dekomposisi Saham Apple.inc. Jurnal Sintak. Vol 1. No 1.
- [3] Serly cania., dkk. 2023. Penerapan Model Seasonal Autoregresif Integrated Moving Average (SARIMA) pada Jumlah Penumpang Kereta Api di Sumatera Barat. JOSTECH: journal of science and technology.
- [4] Prianda, B. G., dan Widodo, E. 2021. "Perbandingan Metode Seasonal Arima dan Extreme Learning Machine pada Peramalan Jumlah Wisatawan Mancanegara Ke Bali". BAREKENG: Jurnal Ilmu Matematika dan Terapan, 15(4), 639–650.
- [5] Lin Fadliani., dkk. 2021. Peramalan dengan Metode SARIMA pada Data Inflasi dan Identifikasi Tipe Outlier (studi kasus: Data Inflasi Indonesia Tahun 2008-2014). Statistika, Vol. 9, No. 2.
- [6] Dadang Ruhiat., dkk. 2022. Forecasting Data Time Series Berpola Musiman Menggunakan Model SARIMA (Studi Kasus: Sungai Cipeles-Warungpeti). Jurnal Riset Matematika dan Sains Terapan. Volume 2, Nomor 1.
- [7] Paramita., dkk. 2020. Aplikasi Metode Bayesian Model Averaging (BMA) dengan pendekatan Markov Chain Monte Carlo (MCMC) untuk Peramalan Curah Hujan di Stasiun Meteorologi Kota Makasar. VARIANSI: Journal of statistics and Its Application on teaching and Research. Vol. 2, No. 3.
- [8] Sonia Faradilla dan Agus Suharsono. 2023. Peramalan Penjualan Produk Baja dan Besi di PT MSU dengan Pendekatan Metode ARIMA dan Single Moving Average. Jurnal Sains dan Seni ITS. Vol. 12, No. 1.
- [9] Reni Yuliyanti., dkk. 2022. Peramalan Penduduk Menggunakan Model ARIMA. Jurnal Kajian dan Terapan Matematika. Volume 8, Edisi 2.
- [10] Wahidah Alwi., dkk. 2023. Peramalan Produksi Padi Menggunakan Metode SARIMA di kabupaten Bone. Jurnal Matematika dan Statistika serta Aplikasinya. Vol. 11, No. 2.
- [11] Ari Dirgantara., dkk. 2020. Peramalan Inflasi di Kota Surakarta Menggunakan Model autoregressive Integrated Moving Average (ARIMA). Prosiding Seminar Edusainstech FMIPA UNIMUS.
- [12] Ida Nabilah dan Indra Ranggadara. 2020. Mean Absolute Percentage Error untuk Evaluasi Hasil Prediksi Komoditas Laut. Journal of Information System. Vol 5, No. 2.
- [13] Achmad David Mico, dkk. 2022. Peramalan Penjualan Batu Gamping pada UD Eko Jaya Menggunakan Single Exponential Smoothing dan Double Exponential Smoothing. Jurnal Cafeteria. Vol. 3, No. 2.
- [14] Chika Syifa Audinasyah dan Solehudin. 2024. Sistem Forecasting Perencanaan Produksi dengan Metode Single Exponential Smoothing pada Home Industry Tempe Putera Sejahtera. Jurnal Ekonomi dan Manajemen Teknologi (EMT). Vol. 8, No. 3.