

Peramalan Jumlah Penderita Jenis Penyakit Utama Di Kota Surabaya Menggunakan Metode *Autoregressive Integrated Moving Average (ARIMA)*

Tri Diana Rimadhani

UPN “Veteran” Jawa Timur

Amalia Anjani Arifiyanti

UPN “Veteran” Jawa Timur

Rizka Hadiwiyanti

UPN “Veteran” Jawa Timur

Alamat: Jl. Rungkut Madya No.1, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur 60294

Korespondensi penulis: dianarimadhani12@gmail.com

***Abstract.** The decline in Covid-19 pandemic cases over the past few months before 2023 not only reduced Indonesia's challenges in dealing with other types of diseases. Surabaya city itself has health issues that need to be addressed, as indicated by an increase in several diseases in 2022. One attempt to handle disease cases is to predict the future case numbers to anticipate actions early on. By using data on the number of cases of Respiratory System Diseases during the year 2022, this data will be used to predict the number of cases in the coming months in 2023 by creating a variation of a univariate ARIMA forecasting model. Therefore, the research results show forecasted data that tend to increase when viewed from the analysis and the plotted graph of the forecast results.*

***Keywords:** forecast, disease, ARIMA*

Abstrak. Penurunan kasus pandemi Covid-19 selama beberapa bulan tahun terakhir sebelum tahun 2023, tidak semata-mata menurunkan tantangan Indonesia dalam menghadapi kasus jenis penyakit lainnya. Kota Surabaya sendiri memiliki masalah kesehatan yang harus ditangani yang ditunjukkan dengan beberapa penyakit mengalami peningkatan di tahun 2022. Salah satu upaya penanganan kasus penyakit adalah mengetahui jumlah kasus di masa depan agar dapat mengantisipasi tindakan sejak awal. Dengan menggunakan data jumlah penderita Penyakit pada Sistem Pernafasan selama periode tahun 2022, data tersebut akan diprediksi jumlah kasusnya dalam beberapa bulan ke depan di tahun 2023 dengan membuat variasi model peramalan metode ARIMA univariat. Sehingga hasil penelitian didapatkan data hasil peramalan yang cenderung meningkat ketika dilihat dari hasil analisis serta grafik plot hasil peramalan.

Kata kunci: peramalan, penyakit, ARIMA

LATAR BELAKANG

Kasus pandemi Covid-19 yang telah menyerang secara global termasuk Indonesia mulai menunjukkan tren penurunan kasus selama beberapa bulan terakhir sebelum memasuki tahun 2023. Tren penurunan kasus tersebut dilaporkan berada di bawah seribu kasus dimana mulai terjadi sejak pertengahan April (CNN Indonesia, 2022). Dengan penurunan kasus pandemi tersebut, tidak semata-mata menurunkan tantangan Indonesia dalam menghadapi kasus jenis penyakit lainnya. Menurut Kementerian Kesehatan Republik Indonesia, Negara Indonesia saat ini sedang mengalami *Triple Burden* atau dapat disebut beban tiga kali lipat

Received Juli 28, 2023; Revised Agustus 22, 2023; Accepted September 02, 2023

Tri Diana Rimadhani, dianarimadhani12@gmail.com

berbagai masalah penyakit. Yang pertama adalah adanya Penyakit Infeksi *New Emerging* dan juga *Re-Emerging* yang diketahui seperti Covid-19. Yang kedua adalah penyakit menular yang belum teratasi dengan baik. Serta yang ketiga merupakan Penyakit Tidak Menular (PTM) yang masih cenderung naik di setiap tahun (Kementrian Kesehatan, 2022).

Sebagai kota terbesar kedua setelah Jakarta, Surabaya yang merupakan ibukota Provinsi Jawa Timur menjadi salah satu dari 4 pusat pertumbuhan utama di Indonesia (Liputan6.com, 2022). Kota Surabaya sendiri memiliki masalah kesehatan yang harus ditangani. Beberapa kasus penyakit menular seperti diare, campak hingga TBC masih menunjukkan angka yang tinggi. Bahkan kasus penyakit TBC di Surabaya yang ditemukan menjadi kasus terbanyak di Jawa Timur pada tahun 2022 (Detik Jatim, 2023).

Dengan adanya penemuan kasus masih cukup tinggi tersebut, salah satu upaya penanganan kasus adalah mengetahui jumlah kasus di masa depan agar dapat mengantisipasi tindakan sejak awal. Maka dari itu untuk mengetahui secara optimal perlu adanya model kuantitatif dengan tujuan untuk menjelaskan kondisi jumlah kasus yang mengalami kenaikan maupun penurunan.

Peramalan adalah sebuah proses memanfaatkan data yang telah didapat untuk memperkirakan kejadian di masa depan dengan menganalisis data dari masa lampau menggunakan teknik atau cara tertentu sehingga dapat dilakukan pencegahan untuk segala sesuatu yang terjadi. Ketepatan ramalan yang telah dibuat dapat menentukan baik atau tidaknya hasil penelitian yang telah dilakukan (Rahayu, 2022).

Dari beberapa penelitian yang telah dilakukan sebelumnya, data berbentuk *time-series* dimana data dikumpulkan dari rentang waktu tertentu dan memiliki jumlah pada setiap rentang waktu tersebut. Data jumlah penyakit yang cenderung memiliki pola musiman serta tren, metode peramalan ARIMA dapat digunakan sebagai metode yang sesuai untuk meramalkan jumlah penderita penyakit utama yang ditemukan di Kota Surabaya. Penelitian yang dilakukan oleh Syaharuddin, Qisty serta Lalu (Syaharuddin, Akmal, & Sucipto, 2022) yang menganalisis perbedaan tingkat akurasi peramalan data *time-series* mendefinisikan ARIMA yang biasa disebut dengan metode *Box-Jenkins* merupakan metode teknik peramalan untuk model *time-series* yang didasarkan pada perilaku variabel yang diamati dan fenomena musiman dimana memiliki periode yang relatif sama. Metode ini biasa digunakan untuk peramalan jangka pendek.

Dalam penelitian ini, dengan menggunakan data jumlah penderita Penyakit pada Sistem Pernafasan selama periode tahun 2022, data tersebut akan diprediksi jumlah kasusnya dalam beberapa bulan ke depan di tahun 2023 dengan membuat variasi model peramalan metode ARIMA univariat. Dari hasil penelitian ini diharapkan mampu memberikan kontribusi bagi Dinas Kesehatan Kota Surabaya dalam pengambilan keputusan untuk upaya penanganan penyakit di Kota Surabaya khususnya untuk Penyakit Pada Sistem Pernafasan.

KAJIAN TEORITIS

Peramalan (Forecasting)

Peramalan adalah suatu bentuk perhitungan yang dilakukan dengan tujuan memperkirakan kejadian yang akan terjadi di masa depan menggunakan referensi data yang berasal dari masa lalu. Metode dan model yang berdasar pada sejarah dari variabel yang akan di prediksi. Istilah univariat sendiri mengimplikasikan bahwa peramalan didasarkan pada sampel dari observasi *time-series* tanpa mempertimbangkan efek dari variabel lainnya (Moosa, n.d.).

Metode Autoregressive Integrated Moving Average (ARIMA)

ARIMA yang biasa disebut dengan metode *Box-Jenkins* merupakan metode teknik peramalan untuk model *time-series* yang dikembangkan George Box dan Gwilym Jenkins yang didasarkan pada perilaku variable yang diamati dan fenomena musiman dimana memiliki periode yang relatif sama. Metode ini biasa digunakan untuk peramalan jangka pendek karena model ini memiliki ketepatan yang kurang baik untuk peramalan jangka panjang karena akan cenderung datar (Sigit & Setiyoargo, 2020).

METODE PENELITIAN

Penelitian ini menggunakan teknik peramalan atau *forecasting* pada data Penyakit Pada Sistem Pernafasan yang ditemukan di Kota Surabaya. Metode peramalan ini menggunakan metode ARIMA (*Autoregressive Integrated Moving Average*). Adapun tahapan yang akan dilalui yaitu meliputi :

Pengumpulan Data

Dalam tahap ini, data penelitian diperoleh dari situs web resmi pemerintah data.go.id yang memiliki dataset berbentuk *csv* berjudul Banyaknya Penyakit Utama yang Ditemukan di Puskesmas Menurut Jenis Penyakit tahun 2022. Adapun studi literatur yang dilakukan berasal dari jurnal terdahulu mengenai peramalan menggunakan metode ARIMA.

Pre-processing Data

Dilakukan persiapan data sebelum dilakukan pembuatan model. Proses data mining diawali dengan *pre-processing* data yang terdiri dari pengumpulan data yang dibutuhkan, dilanjutkan dengan transformasi data untuk mengubah data yang didapatkan menjadi format data yang dapat diproses oleh *data mining* (Meilani & Pardistya, 2020)), Hasil transformasi data digunakan untuk analisis statistik, pemodelan pada *machine learning*, dan visualisasi informasi. Setelah dilakukan *pre-processing* data, kemudian dilanjutkan uji stasioneritas dengan ADF (Augmented Dickey-Fuller) test.

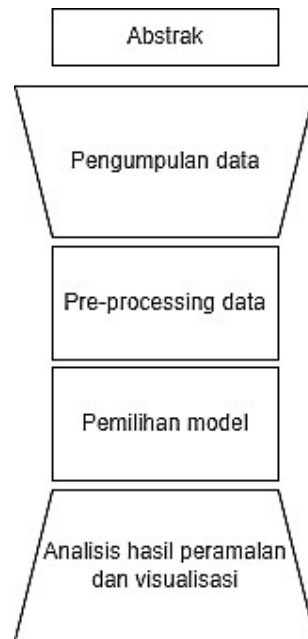
Pemilihan Model

Model ARIMA didapatkan dari pengidentifikasian nilai AR dan MA dengan melakukan plot ACF (*Autocorrelation Function*) dan PACF (*Partial Autocorrelation Function*). Membangun model dengan menganalisis lag berdasarkan plot autokorelasi PACF (*Partial Autocorrelation Function*) dan ACF (*Autocorrelation Function*) untuk mengidentifikasi nilai order p, d, dan q (Hariadi & Sulantari, 2021).

Analisis Hasil Peramalan dan Visualisasi

Model yang didapatkan kemudian diterapkan pada dataset guna memperoleh hasil prediksi peramalan jumlah penderita jenis penyakit utama. Selanjutnya dilakukan visualisasi hasil peramalan.

Dalam penelitian ini, digunakan analisis menggunakan bahasa pemrograman *Python* dimana menggunakan *library* khusus untuk analisis data, melakukan pemodelan *supervised learning* dalam *machine learning* dan juga visualisasi data sederhana. Penelitian ini menggunakan *tools* dari *Google* untuk proses pengolahan data yaitu *Google Collaboratory* dan juga *Excel*.



Gambar 1 Metodologi Penelitian

HASIL DAN PEMBAHASAN

Dalam bab ini, dijelaskan apa yang telah dilakukan sesuai dengan metodologi. Diantaranya pengumpulan data, *pre-processing*, pemilihan model, serta analisis hasil peramalan dan visualisasi.

Pengumpulan Data

Data mentah yang didapat dari *website* resmi pemerintah data.go.id berisi 18 jenis penyakit yang telah diklasifikasikan menurut jenis penyakit utama. Data yang akan digunakan dalam penelitian ini hanya menggunakan data salah satu jenis penyakit yaitu Penyakit Pada Sistem Pernafasan yang diambil dari periode Januari hingga Desember 2022. Pada Tabel 1 berikut merupakan data hasil pembersihan data yang hanya menunjukkan data pada Penyakit Pada Sistem Pernafasan.

Tabel 1 Data Aktual

No.	Periode	Wilayah	Jenis Penyakit	Jumlah Kasus
9	1/1/2022	Surabaya Barat	Penyakit pada sistem pernafasan	443
27	1/1/2022	Surabaya Selatan	Penyakit pada sistem pernafasan	208
45	1/1/2022	Surabaya Barat	Penyakit pada sistem pernafasan	312
63	1/1/2022	Surabaya Barat	Penyakit pada sistem pernafasan	300
81	1/1/2022	Surabaya Selatan	Penyakit pada sistem pernafasan	437
...
1352	12/1/2022	Surabaya Pusat	Penyakit pada sistem pernafasan	426
1354	12/1/2022	Surabaya Timur	Penyakit pada sistem pernafasan	569
1356	12/1/2022	Surabaya Selatan	Penyakit pada sistem pernafasan	761
13581	12/1/2022	Surabaya Selatan	Penyakit pada sistem pernafasan	151
1359	12/1/2022	Surabaya Utara	Penyakit pada sistem pernafasan	864

Data yang telah dikumpulkan dalam Tabel 1 kemudian akan dipisah berdasarkan wilayah yang terdapat di Surabaya. Terdapat 5 wilayah di Surabaya diantaranya Surabaya Barat, Surabaya Pusat, Surabaya Selatan, Surabaya Timur dan juga Surabaya Utara. Setelah didapatkan data pada wilayah masing-masing kemudian data di uji stasioneritas untuk mengetahui data stasioner atau tidak.

Pre-Processing Data

Pre-Processing Data diawali dengan pembagian dataset menjadi 5 berdasarkan wilayah yang terdapat di Surabaya. Setelah itu dilakukan uji stasioneritas karena dalam menganalisis *time-series* untuk model ARIMA diperlukan kestasioneran data. Dengan melihat nilai dari *p-value*, data diidentifikasi kestasionerannya. Berikut merupakan hasil uji ADF pada masing-masing wilayah.

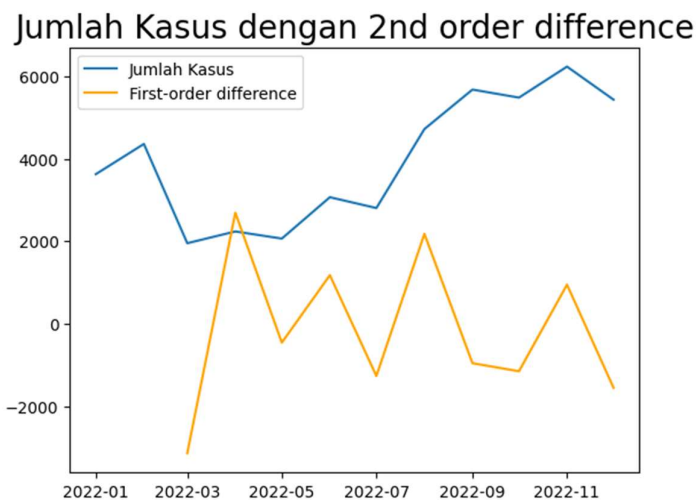
Tabel 2 Hasil Uji Stasioner

Penyakit Pada Sistem Pernafasan			
Wilayah	ADF Statistic	P-Value	Kestasioneran
Surabaya Barat	-2.186	0.2114	Belum Stasioner
Surabaya Pusat	-2.583	0.0966	Belum Stasioner
Surabaya Selatan	-1.68	0.4417	Belum Stasioner
Surabaya Timur	-1.963	0.3028	Belum Stasioner
Surabaya Utara	-1.546	0.5106	Belum Stasioner

Tabel 2 merupakan hasil pengujian ADF *test* untuk mengecek kestasioneran data. Berdasarkan tabel tersebut dihasilkan bahwa seluruh data pada penderita jenis Penyakit Pada Sistem Pernafasan berdasarkan wilayahnya masih belum menunjukkan data stasioner dimana nilai *p-value* > (lebih dari) 5% atau 0,05.

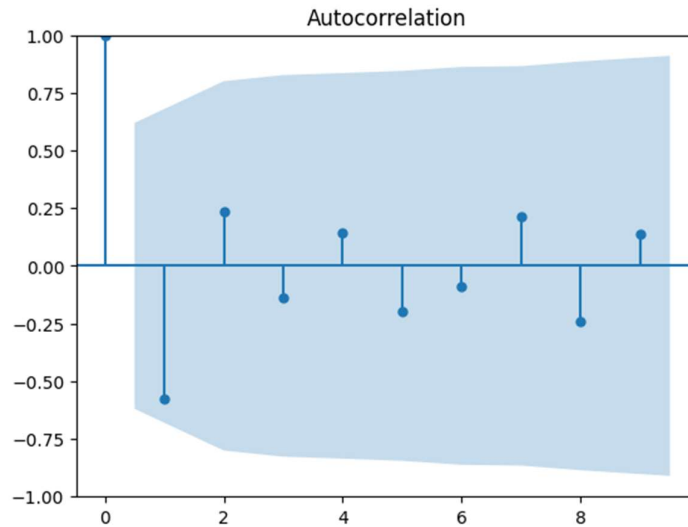
Pemilihan Model ARIMA (p,d,q)

Sebelum melakukan pemilihan model ARIMA (p,d,q) dilakukan diferensiasi pada data yang belum stasioner sehingga data menjadi stasioner. Dalam proses diferensiasi, akan didapatkan order diferensiasi dimana data menjadi stasioner sehingga order tersebut akan digunakan sebagai nilai estimasi d pada model ARIMA (p,d,q).

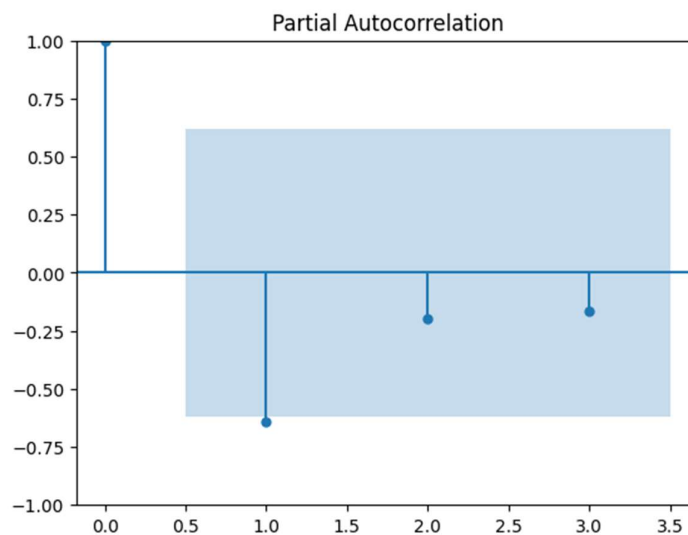


Gambar 2 2nd Difference Data Penyakit Pada Sistem Pernaafasan Wilayah Surabaya Barat

Berdasarkan Gambar 2 dapat dilihat bahwa plot data sudah bergerak disekitar rata-rata setelah dilakukan proses diferensiasi sebanyak 2 kali. Maka dari itu, dapat dikatakan bahwa plot sudah bersifat stasioner dan didapatkan nilai estimasi untuk d adalah d=2. Setelah proses diferensiasi dilakukan maka selanjutnya menentukan nilai parameter untuk p dan q pada model ARIMA(p,d,q) dengan plot ACF (*Autocorrelation Function*) dan PACF (*Partial Autocorrelation Function*).



Gambar 3 Plot ACF dan PACF Penyakit Pada Sistem Pernafasan Wilayah Surabaya Barat (a)



Gambar 4 Plot ACF dan PACF Penyakit Pada Sistem Pernafasan Wilayah Surabaya Barat (b)

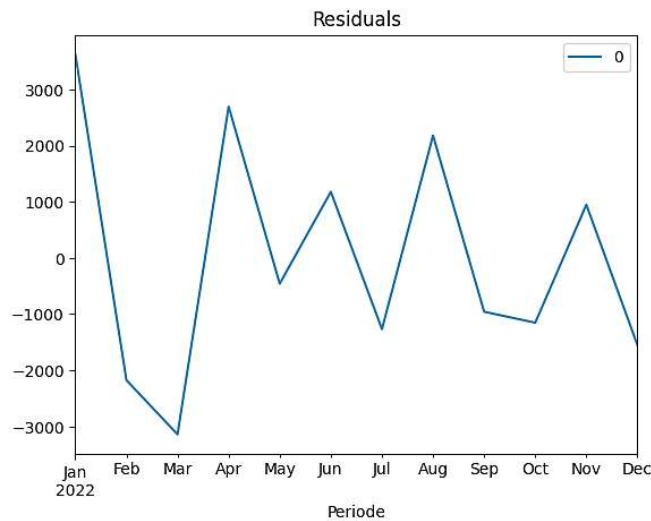
Berdasarkan Gambar 3 ditunjukkan bahwa nilai koefisien atau lokasi lag menurun dengan cepat pada lag 1 dan 0 pada masing-masing plot. Dengan begitu untuk nilai sementara parameter p adalah $p=0$ atau $p=1$ sedangkan untuk nilai sementara parameter q adalah $q=0$ atau $q=1$. Setelah mendapatkan model sementara, dilakukan analisis pada nilai AIC yang kemudian dari model ARIMA sementara yang memiliki nilai AIC paling kecil adalah yang menjadi model ARIMA terpilih. Sehingga didapatkan model ARIMA (p,d,q) terpilih untuk seluruh wilayah ditunjukkan pada Tabel 3.

Tabel 3 Pemilihan Model ARIMA

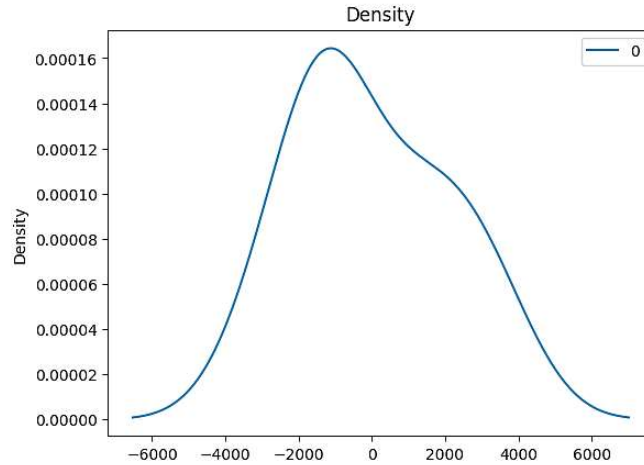
Penyakit Pada Sistem Pernafasan		
Wilayah	Model	AIC
Surabaya Barat	ARIMA(0, 2, 1)	179,908
Surabaya Pusat	ARIMA(1, 2, 1)	181,940
Surabaya Selatan	ARIMA(1, 2, 0)	190,489
Surabaya Timur	ARIMA(0, 2, 0)	186,544
Surabaya Utara	ARIMA(1, 2, 0)	184,825

Analisis Hasil Peramalan dan Visualisasi

Beberapa model ARIMA yang telah terpilih kemudian dianalisis pada *residuals errors* dan persebaran datanya. Pada grafik *residuals errors* Gambar 5 terlihat bahwa *residuals errors* tersebar secara independen serta data terdistribusi mendekati normal pada grafik dengan judul ‘Density’.



Gambar 5 Hasil Residuals Error Penyakit Pada Sistem Pernafasan Wilayah Surabaya Barat (a)



Gambar 6 Hasil Residuals Error Penyakit Pada Sistem Pernafasan Wilayah Surabaya Barat (b)

Selain dianalisis untuk *residuals errors*, model kemudian dievaluasi menggunakan nilai MAPE (*Mean Absolute Percentage Error*) dan juga RMSE (*Root Mean Square Error*). Tabel 4 menunjukkan nilai MAPE (*Mean Absolute Percentage Error*) dan juga RMSE (*Root Mean Square Error*), dimana dapat dilihat untuk nilai RMSE (*Root Mean Square Error*) memiliki nilai yang cukup tinggi. Tingginya nilai RMSE (*Root Mean Square Error*) mengartikan bahwa *error* yang dimiliki oleh model cukup tinggi.

Tabel 4 Evaluasi Model Menggunakan MAPE dan RMSE

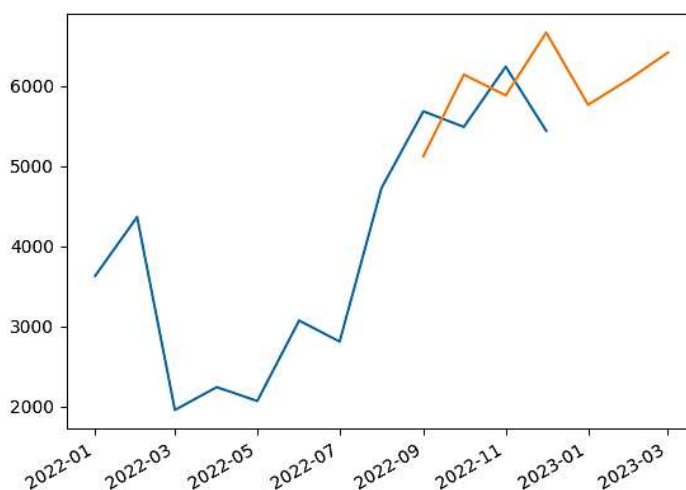
Wilayah	Model	MAPE	RMSE
Surabaya Barat	ARIMA (0, 2, 1)	0.1252	770.4776
Surabaya Pusat	ARIMA (1, 2, 1)	0.1794	1095.479
Surabaya Selatan	ARIMA (1, 2, 0)	0.1932	2235.603
Surabaya Timur	ARIMA (0, 2, 0)	0.2444	2434.477
Surabaya Utara	ARIMA (1, 2, 0)	0.1638	1778

Setelah model sudah dianalisis dan dinilai cocok, model akan diimplementasikan pada data aktual sehingga menghasilkan data hasil peramalan untuk periode 3 bulan kedepan. Periode tersebut adalah periode Januari hingga Maret tahun 2023. Tabel hasil peramalan dapat dilihat pada Tabel 5 berikut.

Tabel 5 Data Hasil Peramalan 3 Periode

Wilayah	Periode Tahun 2023		
	Januari	Februari	Maret
Surabaya Barat	5761	6086	6411
Surabaya Pusat	5220	5454	5762
Surabaya Selatan	6060	3249	438
Surabaya Timur	3991	810	-2371
Surabaya Utara	5663	5663	1557

Untuk grafik dari hasil peramalan dapat dilihat pada Gambar 7 berikut. Jika diamati, grafik tersebut menunjukkan kenaikan jumlah kasus pada 3 periode ditahun 2023.



Gambar 7 Grafik Plot Hasil Peramalan

KESIMPULAN DAN SARAN

Model yang didapatkan untuk diimplementasikan pada data aktual memiliki nilai MAPE (*Mean Absolute Percentage Error*) yang cukup baik, namun untuk nilai RMSE (*Root Mean Square Error*) menunjukkan nilai yang cukup tinggi sehingga model memiliki nilai *error* yang cukup tinggi. Meskipun begitu, untuk *residuals errors* serta persebaran datanya menunjukkan model akan cocok untuk diimplementasikan dengan data aktual. Oleh karena itu didapatkan data hasil peramalan yang cenderung meningkat ketika dilihat pada grafik plot hasil peramalan.

DAFTAR REFERENSI

- CNN Indonesia. (2022). Deretan Penyakit dan Masalah Kesehatan di Indonesia sepanjang 2022. Retrieved from <https://www.cnnindonesia.com/nasional/20221226070250-20-891959/deretan-penyakit-dan-masalah-kesehatan-di-indonesia-sepanjang-2022#>.
- Detik Jatim. (2023). Hingga Maret 2023, 1.691 Kasus TBC Ditemukan di Surabaya. Retrieved from <https://www.detik.com/jatim/berita/d-6655833/hingga-maret-2023-1691-kasus-tbc-ditemukan-di-surabaya>.
- Hariadi, W., & Sulantari. (2021). Application Of ARIMA Model For Forecasting Additional Positive Cases Of Covid-19 In Jember Regency. *Enthusiastic International Journal Of Statistics And Data Science*, 20-27.
- Kementrian Kesehatan. (2022). Masalah dan Tantangan Kesehatan Indonesia Saat Ini. Retrieved from <https://kesmas.kemkes.go.id/konten/133/0/masalah-dan-tantangan-kesehatan-indonesia-saat-ini>.
- Liputan6.com. (2022). 6 Fakta Menarik Surabaya, Kota Pahlawan yang Punya Taman Terbaik di Asia. Retrieved from <https://www.liputan6.com/lifestyle/read/4929979/6-fakta-menarik-surabaya-kota-pahlawan-yang-punya-taman-terbaik-di-asia>.
- Meilani, S., & Pardistya, I. Y. (2020)). Pengaruh Return On Equity (Roe) Dan Current Ratio (Cr) Terhadap Harga Saham. *E-Jurnal Ekonomi Dan Bisnis Universitas Udayana*, 1159-1180.
- Moosa, I. A. (n.d.). *Exchange Rate Forecasting: Techniques and Applications*.
- Rahayu, R. (2022). Penerapan Forecasting Dalam Jumlah Kasus Penyakit Malaria Menggunakan Metode Exponential Smoothing. *Jurnal Informatika Dan Teknologi Informasi*, 98-103.
- Sigit, N., & Setiyoargo, A. (2020). Analisis Peramalan Jumlah Penderita Hipertensi Pada Lansia Di Kabupaten. *Jurnal Rekam Medis dan Informasi Kesehatan*, 7-12.
- Syahrudin, Akmal, Q. S., & Sucipto, L. (2022). Metode ARIMA, ARIMAX, Dan SARIMA: Sebuah Meta-. *Jurnal Informatika Kaputama (JIK)*, 502-509.