

Analisis Runtun Waktu Untuk Memprediksi Jumlah Mahasiswa Baru Dengan Model Arima

Ayu Ulfa Jamila¹, Bella Merlin Siregar², Roni Yunis³

^{1,2,3}Program Studi Sistem Informasi, STMIK Mikroskil Medan
e-mail: 172113993@students.mikroskil.ac.id¹
e-mail: 172112084@students.mikroskil.ac.id²
e-mail: roni@mikroskil.ac.id³

Abstrak - Peramalan penerimaan mahasiswa baru sangatlah penting bagi perguruan tinggi agar keputusan yang diambil perguruan tinggi dapat lebih optimal. Penelitian ini bertujuan untuk menemukan model terbaik agar dapat meramalkan jumlah mahasiswa baru untuk 5 tahun ke depan melalui analisis runtun waktu. Model yang digunakan dalam penelitian ini yaitu Model ARIMA (Autoregressive Integrated Moving Average) dengan dataset penerimaan mahasiswa baru Universitas XYZ periode 2010 sampai dengan 2019. Tahap pelaksanaan penelitian mengacu pada metode OSEMN yaitu Obtain data, Scrubbing data, Explore data, Modeling data, dan Interpreting data. Hasil analisis menunjukkan bahwa model terbaik yaitu model ARIMA (2,1,1). Model ARIMA yang dihasilkan menunjukkan nilai MAPE sebesar 7.066147 dengan akurasi sebesar 93%. Hasil peramalan menunjukkan bahwa terjadinya trend penurunan jumlah mahasiswa baru secara merata untuk 5 tahun kedepan.

Kata kunci: Peramalan, Jumlah Mahasiswa Baru, ARIMA, MAPE

Abstract - Forecasting new student admissions is very important for universities so that decisions made by universities can be more optimal. This study aims to find the best model in order to predict the number of new students for the next 5 years through time series analysis. The model used in this study is the ARIMA (Autoregressive Integrated Moving Average) Model with the new student admissions dataset for XYZ University for the period 2010 to 2019. The research implementation stage refers to the OSEMN method, namely Obtain data, Scrubbing data, Explore data, Modeling data, and Interpreting data. The analysis results show that the best model is the ARIMA model (2,1,1). The resulting ARIMA model shows a MAPE value of 7.066147 with an accuracy of 93%. The forecast results show that there is a trend of decreasing the number of new students evenly for the next 5 years.

Keywords: Forecasting, Number of New Students, ARIMA, MAPE

PENDAHULUAN

Penerimaan mahasiswa baru adalah kegiatan utama dalam sebuah perguruan tinggi yang rutin dilakukan setiap tahunnya (Niswatin, 2016). Mahasiswa merupakan salah satu unsur yang penting dalam kegiatan pembelajaran di perguruan tinggi. Kini sudah banyak ditemukan perguruan tinggi yang berkualitas sehingga menyebabkan persaingan antara perguruan tinggi satu dengan yang lain. Maka diperlukannya tindakan untuk meningkatkan kualitas perguruan tinggi yang dapat dilakukan semaksimal mungkin.

Berbagai bentuk strategi dan cara sudah dilakukan oleh perguruan tinggi untuk mencapai target yang diinginkan. Tetapi beberapa strategi tersebut tidak tepat sasaran sehingga target, minat, dan kualitas calon mahasiswa baru tidak sesuai harapan. Untuk mengatasi hal tersebut salah satu hal yang bisa dilakukan perguruan tinggi adalah melakukan peramalan. Peramalan (*Forecasting*) merupakan kegiatan dalam memprediksi kejadian besar ataupun

runtun waktu yang akan datang berdasarkan hasil data melalui pengujian keadaan di masa lalu sehingga dapat mempersiapkan tindakan untuk ke depannya (Achmanda, 2018).

Untuk melakukan peramalan dengan analisis runtun waktu dalam penelitian ini metode yang digunakan adalah ARIMA (*Autoregressive Integrated Moving Average*). ARIMA merupakan Salah satu metode peramalan yang sedang berkembang dan umum digunakan saat ini dan ARIMA memiliki kelebihan yaitu bersifat fleksibel, dan tingkat keakuratannya efektif sehingga tepat digunakan dalam peramalan jangka pendek dan hanya membutuhkan data historis dalam peramalannya (Muhammad et al., 2017).

Penelitian terdahulu terkait tentang peramalan diantaranya Wici Irawan (2019) "Peramalan Harga Saham PT. Unilever Tbk. Dengan Menggunakan Model ARIMA " dan Siska Candra Ningsih dan Padrul Jana (2018) "Pemodelan Penderita HIV/AIDS dengan Metode ARIMA ". Mengingat pentingnya untuk mengetahui jumlah penerimaan

mahasiswa Universitas XYZ di masa depan maka dilakukan peramalan jumlah penerimaan mahasiswa baru menggunakan metode ARIMA (Irawan, 2019) (Ningsih & Jana, 2018).

METODOLOGI PENELITIAN

1. Bahan Penelitian

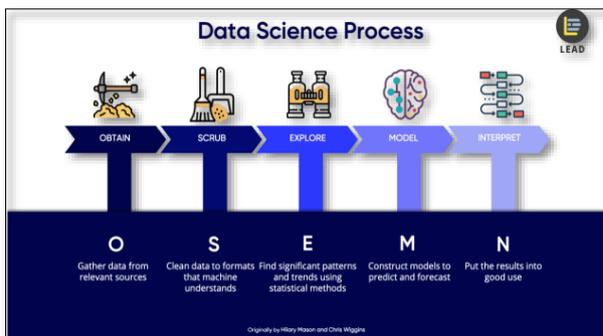
Data yang digunakan dalam penelitian ini diperoleh dari data penerimaan mahasiswa baru periode 2010 hingga 2019 Universitas XYZ. Data ini didapatkan dari bagian PSI Universitas XYZ dalam format *file* CSV dan datanya terdiri dari 9613 baris dan 11 kolom.

Tabel 1. Atribut dataset penerimaan mahasiswa baru

Atribut	Isi Data
Tahun.Akademik	2017/2018
Tanggal.Daftar	31-08-2017
Program.Studi	SISTEM INFORMASI
Jumlah.Mhs	2
Jumlah.Grade.A	1
Jumlah.Grade.B	0
Jumlah.Grade.C	1
Jumlah.Grade.D	0
Kode.Sekolah	10220210
Nama.Sekolah	SMKN 9 MEDAN
Nama.Kota	MEDAN

2. Tahap Penelitian (OSEMN)

Proses OSEMN merupakan model organisasi penelitian yang standar dan dapat diterima secara luas di bidang *Data Science*. Proses OSEMN juga dapat memecahkan masalah dengan *Data Science/Analisis* dalam skala besar (Dineva & Atanasova, 2018). Tahapan OSEMN dapat dilihat pada Gambar 1 (Lau, 2019).



Sumber: (Lau, 2019)

Gambar 1. Proses OSEMN

a. Obtain/Data Collection

Dalam *Data Science* hal yang perlu dilakukan adalah mengumpulkan data. Dalam penelitian ini data yang

diperoleh berasal dari data penerimaan mahasiswa baru dalam format *file* CSV. Jika menggunakan *software* R, di dalamnya terdapat paket khusus yang dapat membaca data dari sumber data ini langsung ke program *Data Science* (Lau, 2019).

b. Scrubbing Data

Dalam memperoleh suatu data tidak jarang data yang diperoleh terdapat data yang nilainya hilang, inkonsistensi, kesalahan, karakter yang aneh, dan kolom yang kurang menarik. Maka diperlukan *scrubbing* data ataupun membersihkan data untuk memudahkan dalam melakukan analisis data jumlah penerimaan mahasiswa baru (Lau, 2019). Operasi umum pada *scrubbing* data yaitu (Janssens, 2019):

- 1) Memfilter baris data
- 2) Mengekstrak pada kolom tertentu
- 3) Mengekstrak kata-kata
- 4) Mengubah nilai
- 5) Mengatasi nilai yang hilang
- 6) Mengubah data dari satu format ke format lainnya

c. Explore Data (Menjelajahi Data)

Menemukan, menyusun, serta memperluas adalah suatu proses yang sangat penting dalam menjelajahi pengumpulan data. Observasi kumpulan data set dapat membantu memilih pendekatan yang terbaik dalam penelitian analisis. Maka diperlukannya pemahaman pada suatu data yang unik, seperti nilai yang tidak biasa. Pendekatan ini digunakan dalam melakukan proses generalisasi data yang diperoleh dari *dataset* penerimaan mahasiswa baru Universitas XYZ dan untuk meringkas suatu karakteristik utama (Dineva & Atanasova, 2018).

d. Modeling Data

Tahap model data merupakan tahap yang paling menarik dan Hal pertama yang perlu dilakukan dalam pembuatan model data adalah mengurangi dimensi suatu kumpulan data serta memilih yang relevan yang berkontribusi pada hasil prediksi (Lau, 2019). Dalam penelitian ini menggunakan regresi dan prediksi dalam memperkirakan nilai di masa depan menggunakan *dataset* jumlah penerimaan mahasiswa baru dengan metode ARIMA yang diimplementasikan menggunakan *software* R (Lau, 2019).

ARIMA (*Autoregressive Integrated Moving Average*) merupakan suatu metode peramalan yang memanfaatkan data di masa lalu dan sekarang melalui variabel dependen yang menghasilkan peramalan jangka pendek yang akurat. Metode ARIMA diperkenalkan oleh Box dan Gwilym Jenkins tahun 1976 yang merupakan musiman dari Box-Jenkins (Indrasietianingsih et al., 2017). ARIMA merupakan statistik yang cocok untuk meramalkan secara cepat, bersifat fleksibel, dan tingkat keakuratannya efektif sehingga tepat digunakan dalam peramalan jangka pendek dan

hanya membutuhkan data historis dalam peramalannya (Muhammad et al., 2017).

1. Identifikasi Model ARIMA

Mengidentifikasi model merupakan hal yang penting dalam *Analysis Time Series*. Di tahap ini, kestasioneran datanya di uji baik dalam variansi dan nilai rata-rata. Maka dilakukan proses identifikasi model orde AR dan MA dalam grafik ACF (*Auto Correlation*) dan PACF (*Partial Auto Correlation*) (Aminuddin, 2018).

Tabel 2. Pola ACF dan PACF

Model	ACF	PACF
R (p)	Menurun secara eksponensial	Terpotong sesudah lag ke- p
MA (q)	Terpotong sesudah lag ke- q	Menurun secara eksponensial
ARMA (p, q)	Menurun secara eksponensial sesudah lag ke ($q - p$)	Menurun secara eksponensial sesudah lag ke ($p - q$)

2. Model ARIMA

Model ARIMA merupakan model *univariate time series* yang menggabungkan antara model AR (*Autoregressive*) dan model MA (*Moving Average*) pada data yang tidak stasioner (As'ad et al., 2017).

a. AR (*Model Autoregressive*)

AR adalah model yang menjelaskan variabel terikat dipengaruhi oleh variabel itu sendiri pada periode di masa lalu. Model ini dinyatakan sebagai AR(p) atau ARIMA ($p, 0, 0$) dan bentuk persamaan modelnya adalah (Aminuddin, 2018):

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + w_t \quad (1)$$

Keterangan,

Z_t : nilai variabel X pada waktu ke- t ,
 ϕ_p : parameter *autoregressive* ke- p ,
 w_t : nilai *error* pada saat ke- t .

b. MA (*Model Moving Average*)

MA (*Model Moving Average*) adalah model yang menjelaskan secara eksplisit hubungan ketergantungan antara nilai kesalahan yang berurutan. Model ini dinyatakan sebagai MA (q) atau model ARIMA ($0, 0, q$) dan bentuk persamaan modelnya adalah (Aminuddin, 2018):

$$Z_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \dots + \theta_q w_{t-q} \quad (2)$$

Dengan,

Z_t : nilai variabel X pada waktu ke- t ,
 θ_q : parameter *moving average* ke- q ,
 w_t : nilai *error* pada saat ke- t .

c. Model ARMA

Model ARMA merupakan gabungan antara model AR dan MA. Bentuk fungsi ARMA yaitu (p, q) atau model ARIMA ($p, 0, q$) dan bentuknya persamaan modelnya adalah (Susanto, 2016):

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \dots + \theta_q w_{t-q} \quad (3)$$

Keterangan,

Z_t : nilai variabel Z pada waktu ke- t ,
 ϕ_i : koefisien regresi ke- $i, i = 1, 2, 3, \dots, p$
 ϕ_p : parameter AR ke- p ,
 θ_q : parameter MA ke- q ,
 θ_i : parameter model MA ke- $i, i = 1, 2, 3, \dots, q$
 w_t : nilai *error* pada saat ke- t ,
 $w_t, w_{t-1}, w_{t-2}, \dots, w_{t-q}$: *error* pada saat $t, t-1, t-2, \dots, t-q$ dan w_t diasumsikan *White Noise* dan normal.

d. Model ARIMA

Model ARIMA (p, d, q) yang dimana orde p menyatakan operator AR, orde d menyatakan hasil dari *differencing*, dan orde q menyatakan operator dari MA (Aminuddin, 2018). Bentuk umum pada model ARIMA adalah (Suseno, 2017):

$$\phi(B)Z_t = \theta(B)w_t \quad (4)$$

Keterangan,

Z_t : nilai variabel X pada waktu ke- t ,
 B : operator *backward shift*,
 w_t : nilai *error* pada saat ke- t .
 Persamaan modelnya adalah:

$$Z_t = (1 + \phi_1)Z_{t-1} + (\phi_1 - \phi_2)Z_{t-2} + \dots + (\phi_p - \phi_{p-1})Z_{t-p} + w_t - \theta_q w_{t-1} - \dots - \theta_q w_{t-q} \quad (5)$$

Keterangan,

Z_t : nilai variabel X pada waktu ke- t ,
 θ_q : parameter *moving average* ke- q ,
 ϕ_p : parameter *autoregressive* ke- p ,
 w_t : nilai *error* pada saat ke- t .

3. Estimasi Parameter

Estimasi parameter dapat dilakukan dengan menggunakan beberapa cara yaitu, *least square*, *unconditional least square*, *moment*, dan *maximum likelihood* (Susanto, 2016).

4. Pengujian Signifikansi Parameter

Pengujian signifikansi parameter dapat dilakukan dengan uji t . Hipotesis yang digunakan pada pengujian model AR (p) adalah (Susanto, 2016):

$H_0 : \phi_i = 0, \forall_i$
 $H_1 : \phi_i \neq 0, \exists_i$ untuk $i = 1, 2, \dots, p$
 Statistik uji parameternya adalah:

$$t = \frac{\hat{\theta}_i}{SE(\hat{\theta}_i)} \quad (6)$$

Pengujian signifikansi parameter pada model MA (q) adalah (Susanto, 2016):

$$H_0 : \theta_i = 0, \forall_i$$

$$H_1 : \theta_i \neq 0, \exists_i \text{ untuk } i = 1, 2, \dots, q$$

Statistik uji parameternya adalah:

$$t = \frac{\hat{\theta}_i}{SE(\hat{\theta}_i)} \quad (7)$$

5. Pengujian Asumsi Residual

Untuk mendapatkan model ARIMA yang terbaik, ada beberapa asumsi yang harus dipenuhi pada residual yaitu *White Noise* dan uji distribusi normal (Susanto, 2016).

a. Uji *White Noise*

Dalam pengujian *White Noise* dapat menggunakan uji *Box-Pierce Modified* atau *Ljung-Box*.

b. Uji Asumsi Distribusi Normal

Uji asumsi distribusi normal dilakukan untuk mengetahui residual berdistribusi normal atau tidak.

6. Forecast (Peramalan)

Untuk melakukan peramalan fungsi yang dapat digunakan adalah fungsi *forecast()* untuk 5 tahun ke depan dengan model yang terbaik. Tahap dalam peramalan terdiri dari 3 macam yaitu *training*, *testing*, dan *forecasting* untuk 5 tahun ke depan dengan fungsi *autolayer()*. Data *training* adalah peramalan yang menggunakan data asli, data *testing* adalah peramalan yang tidak menggunakan data asli, sedangkan *forecasting* adalah hasil peramalan. Langkah-langkah *forecasting* pemodelan data ARIMA di *software R* (Raymond, 2018):

a. *Exploratory Data Analysis*

Menjelajahi dan memvisualisasikan *data frame* dengan menggunakan fungsi *ggplot()*. Selanjutnya mengubah *data frame* menjadi *time series* serta menampilkan plot data penerimaan mahasiswa baru dengan menggunakan fungsi *plot()*

b. *Decomposing the Data* (dekomposisi data)

Proses dekomposisi data dilakukan untuk memeriksa adanya suatu *trend*, *seasonal*, dan *random* komponen.

c. *Test Stationary* (Tes Stasioneritas)

Time series dapat disebutkan stasioner dengan syarat *mean*, *variance*, dan *covariance* bukan merupakan fungsi waktu. Tes stasioneritas dapat dilakukan dengan menggunakan fungsi *adf.test()*.

d. Identifikasi Model

Setelah data stasioner selanjutnya adalah melihat plot ACF (*Autocorrelation*) fungsi yang digunakan adalah fungsi *acf()*. Untuk melihat plot PACF (*Partial Autocorrelation*) fungsi yang digunakan adalah *pacf()*.

e. Membuat model ARIMA terbaik

Dalam melakukan perhitungan model ARIMA fungsi yang dapat digunakan adalah fungsi *auto.arima()*. Pemilihan model terbaik dapat dilihat melalui nilai AIC dan SBC/BIC pada hasil dari proses membuat model.

f. Estimasi Model

Untuk mengetahui apakah model signifikan atau tidak dengan menggunakan fungsi *coefstest()*.

g. Uji asumsi residual/diagnostik

Untuk melakukan pemeriksaan apakah model ARIMA sudah *fit* dengan *residual* fungsi yang dapat digunakan adalah fungsi *ggtsdiag()* dan *packages* yang diperlukan adalah *packages ggfortify*.

7. Interpreting Data (Menafsirkan Data)

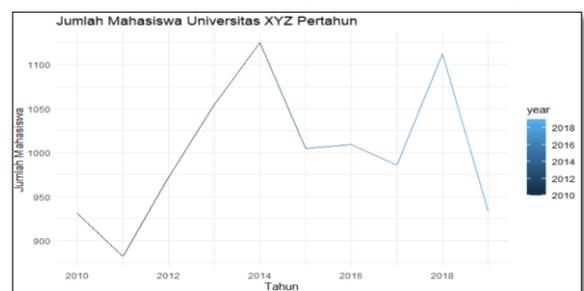
Interpretasi data merupakan tahap terakhir dan tahap paling penting di model OSEMN (Lau, 2019). Interpretasi hasil penelitian penting untuk memahami efektivitas penelitian dan mendeskripsikan secara jelas hasil perbandingan penelitian sendiri dengan penelitian lain (Dineva & Atanasova, 2018).

HASIL DAN PEMBAHASAN

Hasil analisis dari penelitian ini dilakukan dengan menggunakan bantuan *software R 4.0.2*. *Software* ini digunakan untuk menganalisis dan meramalkan jumlah penerimaan mahasiswa baru dengan menggunakan model ARIMA. *Dataset* yang digunakan dalam penelitian sebanyak 9611 data yaitu data jumlah penerimaan mahasiswa baru Universitas XYZ dari tahun 2010 sampai 2019.

1. *Exploratory Data Analysis*

Dataset jumlah mahasiswa baru yang sudah bersih dapat digunakan untuk memvisualisasikan data dalam berbagai macam bentuk plot yang termasuk dalam langkah awal dari *Exploratory Data Analysis* untuk melihat bagaimana perbandingan jumlah mahasiswa baru di masa lampau dan di masa depan.



Gambar 2. Plot data penerimaan mahasiswa baru per tahun

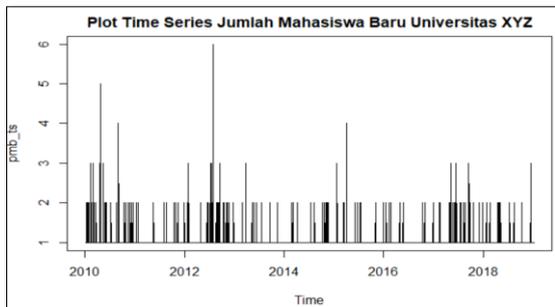
Berdasarkan Gambar 2. terlihat bahwa plotnya merupakan plot data yang fluktuatif. Plot data tersebut diambil dari jumlah penerimaan mahasiswa

baru Universitas XYZ per tahun dalam jangka 10 tahun yaitu mulai dari tahun 2010 hingga 2019. Pada plot penerimaan mahasiswa baru keseluruhan Universitas XYZ terdapat penurunan pada tahun 2011 kemudian mengalami peningkatan yang melonjak hingga tahun 2014, lalu turun kembali hingga tahun 2017 dan mengalami peningkatan lagi di tahun 2018 kemudian turun kembali di tahun 2019. Jumlah mahasiswa tertinggi terjadi pada tahun 2014 dengan jumlah 1125 dan jumlah mahasiswa terendah terjadi pada tahun 2011 dengan jumlah 882. Jumlah mahasiswa baru yang naik turun menyebabkan ketidakstabilan dan tidak konstan terhadap suatu nilai tertentu pada jumlah mahasiswa baru.

2. Forecast jumlah mahasiswa baru Universitas XYZ

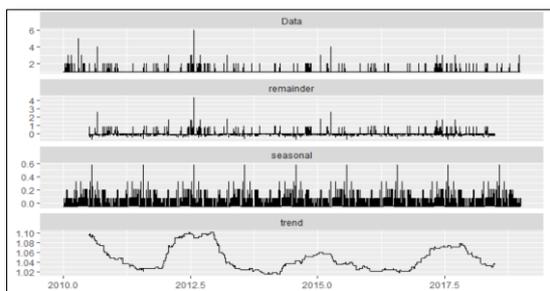
a. Plot data

Langkah awal dalam pemodelan ARIMA adalah memplot data. Berdasarkan Gambar 3. plot data *time series* jumlah penerimaan mahasiswa baru keseluruhan Universitas XYZ menunjukkan bahwa plot data *time series* sudah terlihat stasioner dalam *mean* dan *variance* maka asumsi metode ARIMA telah terpenuhi. Data *time series* diperoleh dari data *training* yang mana jumlah data yang digunakan untuk periode *training* sebanyak 8709 data dan data *testing* sebanyak 902 data.



Gambar 3. Plot data time series Universitas XYZ

b. Decomposing the Data (dekomposisi data)

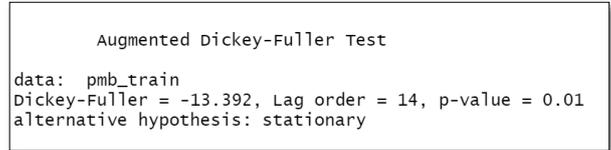


Gambar 4. Plot data, remainder, seasonal, dan trend

Berdasarkan Gambar 4. sudah terlihat plot data mahasiswa baru, pola *seasonal*, dan *trend* yang mengalami kenaikan dan penurunan yang tidak stabil, dan komponen *random* di bagian “remainder”

dari *dataset* jumlah mahasiswa baru keseluruhan Universitas XYZ.

c. Test Stationary (Tes Stasioneritas)



Gambar 5. Tes stasioneritas

Berdasarkan Gambar 5. didapatkan kesimpulan yang sama juga diperoleh dari hasil uji ADF (*Augmented Dickey Fuller*) yang mana nilai *p-value* = 0.01 yang nilainya lebih kecil dari nilai *alpha* yaitu 5% (0,05) maka H_0 ditolak, sehingga artinya data tersebut sudah stasioner dengan demikian data *time series* dapat digunakan untuk membuat model ARIMA.

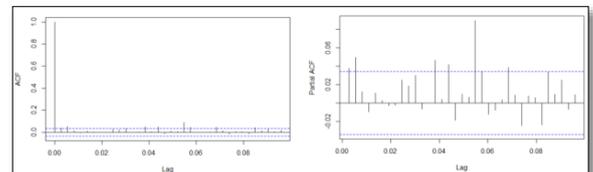
Berikut adalah *summary* dari data:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.000	1.000	1.000	1.051	1.000	6.000

Gambar 6. Summary data

Berdasarkan Gambar 6. dapat dilihat dalam kurun waktu 10 tahun, rata-rata jumlah mahasiswa mendaftar setiap harinya berjumlah 1 mahasiswa.

d. Identifikasi Model ARIMA



Gambar 7. Plot ACF dan PACF

Berdasarkan Gambar 7. plot ACF menunjukkan bahwa terdapat 2 lag yang keluar yang artinya mengandung *moving average* atau MA(0) dan MA(1), Sedangkan untuk plot PACF menunjukkan bahwa terdapat 2 lag yang keluar yang artinya mengandung *autoregressive* atau AR(1) dan AR(2).

e. Membuat model ARIMA

```

ARIMA(2,1,2)(1,0,1)[365] with drift : Inf ARIMA(2,1,1)(1,0,1)[365] : Inf
ARIMA(2,1,0) with drift : 2705.65 ARIMA(1,1,1) : 545.7781
ARIMA(1,1,0)(1,0,0)[365] with drift : Inf ARIMA(2,1,0) : 1398.947
ARIMA(1,1,1)(0,0,1)[365] with drift : Inf ARIMA(3,1,1) : 545.7128
ARIMA(0,1,0) with drift : 1705.648 ARIMA(2,1,2) : 545.7413
ARIMA(0,1,0)(1,0,0)[365] with drift : Inf ARIMA(1,1,0) : 1732.863
ARIMA(1,1,0)(0,0,1)[365] with drift : Inf ARIMA(1,1,1) : 546.2057
ARIMA(1,1,0)(1,0,1)[365] with drift : Inf ARIMA(3,1,0) : 1236.675
ARIMA(1,1,0) with drift : 1734.867 ARIMA(3,1,2) : 547.5922
ARIMA(1,1,0)(0,0,1)[365] with drift : Inf Best model: ARIMA(2,1,1)
ARIMA(1,1,0)(1,0,1)[365] with drift : Inf ARIMA(2,1,0) : 1400.952
ARIMA(2,1,0)(1,0,0)[365] with drift : Inf Series: pmb_train
ARIMA(2,1,0)(0,0,1)[365] with drift : Inf ARIMA(2,1,1)
ARIMA(1,1,0)(1,0,1)[365] with drift : Inf Coefficients:
ARIMA(3,1,0) with drift : 1238.681 ar1 ar2 ma1
ARIMA(3,1,0)(1,0,0)[365] with drift : Inf 0.0220 0.0317 -0.9892
ARIMA(3,1,0)(0,0,1)[365] with drift : Inf s.e. 0.0177 0.0177 0.0032
ARIMA(3,1,0)(1,0,1)[365] with drift : Inf sigma^2 estimated as 0.06891: log likelihood=-267.85
ARIMA(4,1,0) with drift : Inf AIC=543.7 AICC=543.71 BIC=568.09
ARIMA(4,1,0)(0,0,1)[365] with drift : Inf
ARIMA(4,1,0)(1,0,1)[365] with drift : 1029.01
ARIMA(5,1,0) with drift : Inf
ARIMA(5,1,0)(1,0,0)[365] with drift : Inf
ARIMA(5,1,0)(0,0,1)[365] with drift : Inf
ARIMA(5,1,0)(1,0,1)[365] with drift : Inf
ARIMA(5,1,1) with drift : 549.7115
ARIMA(3,1,1)(1,0,0)[365] with drift : Inf
ARIMA(3,1,1)(0,0,1)[365] with drift : Inf
ARIMA(5,1,1)(1,0,1)[365] with drift : Inf
ARIMA(4,1,1) with drift : 547.7437
ARIMA(4,1,1)(1,0,0)[365] with drift : Inf
ARIMA(4,1,1)(0,0,1)[365] with drift : Inf
ARIMA(3,1,1)(1,0,1)[365] with drift : 547.5178
ARIMA(3,1,1)(1,0,0)[365] with drift : Inf
ARIMA(3,1,1)(0,0,1)[365] with drift : Inf
ARIMA(3,1,1)(1,0,1)[365] with drift : Inf
ARIMA(2,1,1) with drift : 545.5132
ARIMA(2,1,1)(1,0,0)[365] with drift : Inf
ARIMA(2,1,1)(0,0,1)[365] with drift : Inf

```

Gambar 8. Membuat model ARIMA dengan auto.arima

Dari plot ACF dan PACF di atas didapatkan model ARIMA terbaik dengan menggunakan auto.arima yaitu model ARIMA (2,1,1). Model ini dikatakan *best model* (model terbaik) karena nilai AIC, *likelihood*, dan MAPE lebih kecil dibandingkan dengan nilai model yang lain sehingga model ARIMA (2,1,1) dapat dikatakan model terbaik dan dapat dilakukan peramalan jumlah mahasiswa baru keseluruhan Universitas XYZ dengan model tersebut seperti yang terlihat pada Gambar 8.

f. Estimasi Model ARIMA

```

z test of coefficients:
      Estimate Std. Error   z value      Pr(>|z|)
ar1   0.0220357  0.0177326    1.2427      0.2140
ar2   0.0357494  0.0177245    2.0170      0.0437 *
ma1  -0.9892223  0.0031809   -310.9916 <-0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

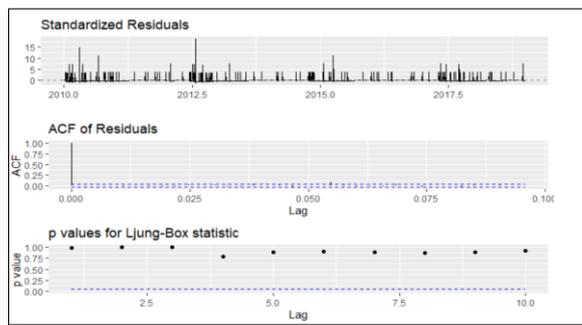
```

Gambar 9. Cek signifikan model ARIMA

Berdasarkan Gambar 9. dapat dilihat bahwa semua estimasi nilai signifikannya *p-value* < 0,05. Sehingga didapatkan model yang signifikan yaitu ARIMA (2,1,1) untuk peramalan 5 tahun ke depan.

g. Uji Diagnostik

Selanjutnya dilakukan uji diagnostik *residual* data. Berikut hasil dari uji diagnostik *residual* model ARIMA (2,1,1):

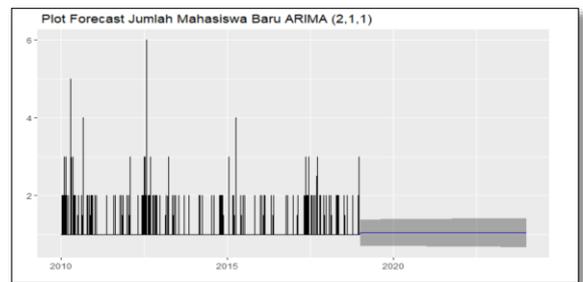


Gambar 10. Uji diagnostik

Berdasarkan Gambar 10. dapat dilihat plot ACF *residual* data merupakan model *white noise* karena tidak terdapat *lag* yang keluar dari garis batas interval. Model terbaik merupakan model yang memenuhi syarat signifikan parameter sehingga model ARIMA (2,1,1) secara *fair* dikatakan *good fit* atau sebagai model terbaik dalam prediksi jumlah mahasiswa baru keseluruhan Universitas XYZ. Sedangkan dilihat dari plot *p-value* ada yang lebih dari $\alpha = 0,05$ yang artinya tidak terdapat autokorelasi pada data sehingga asumsi *non-autokorelasi* terpenuhi. Maka model (2,1,1) cocok digunakan untuk peramalan jumlah mahasiswa baru keseluruhan Universitas XYZ periode 2020-2024.

h. Forecast

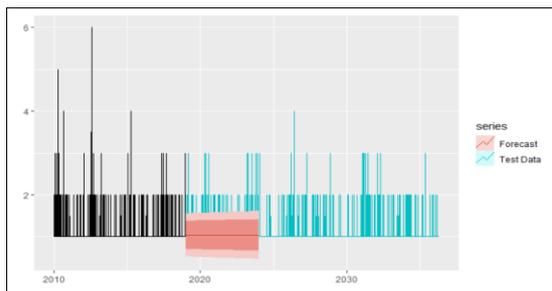
Langkah terakhir dalam analisis *time series* adalah melakukan peramalan untuk 5 tahun ke depan. Maka didapatkan plot *forecasting* mahasiswa baru sebagai berikut:



Gambar 11. Plot hasil forecast Universitas XYZ

Berdasarkan Gambar 11. dapat dilihat bahwa hasil peramalan jumlah mahasiswa baru keseluruhan Universitas XYZ akan mengalami penurunan yang merata. Dapat dilihat bahwa ada *trend* penurunan jumlah mahasiswa baru per hari untuk 5 tahun ke depan. Agar penerimaan mahasiswa baru keseluruhan Universitas XYZ mengalami peningkatan, maka dapat dilakukan antisipasi dengan melakukan perencanaan promosi yang lebih baik lagi sehingga dapat menarik minat mahasiswa untuk mendaftar.

Setelah model diperoleh, selanjutnya yaitu meramalkan menggunakan *training* data yang mana jumlah data yang digunakan untuk periode *training* adalah sebanyak 8709 yaitu data pada tahun 2010-2018 dan untuk periode *testing* sebanyak 902 data yaitu data pada tahun 2018-2019.



Gambar 12. Perbandingan plot train data, test data, dan forecast

Berdasarkan Gambar 12. terlihat plot *train data* berwarna hitam, plot *test data* berwarna hijau, serta pada plot *forecast* berwarna peach. Peramalan jumlah mahasiswa baru dengan model ARIMA (2,1,1) cenderung menurun dengan rata-rata jumlah mahasiswa yang mendaftar setiap harinya yaitu 1 mahasiswa. Jadi dapat disimpulkan bahwa jumlah mahasiswa baru untuk 5 tahun ke depan mengalami penurunan.

Berikut merupakan nilai MAPE dari hasil *forecasting* penerimaan mahasiswa baru keseluruhan Universitas XYZ:

Training set error measures:							
	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	-0.001919297	0.2623445	0.09699969	-2.825751	7.066147	1.018846	-0.0002155975

Gambar 13. Nilai MAPE *forecast* Universitas XYZ
Berdasarkan Gambar 13. didapatkan nilai MAPE 7.066147 dengan nilai akurasi sebesar 93% yang berarti bahwa kesalahan dari model yang digunakan sangat kecil karena nilainya di bawah 10% sehingga hasil peramalan dikatakan sangat akurat.

KESIMPULAN

Berdasarkan hasil analisis peramalan yang telah dilakukan maka dapat disimpulkan bahwa proses analisis *time series* menggunakan metode ARIMA memberikan model ARIMA terbaik untuk peramalan jumlah mahasiswa baru keseluruhan Universitas XYZ yaitu ARIMA (2,1,1).

Hasil peramalan jumlah mahasiswa baru untuk 5 tahun ke depan yaitu mulai dari 2020 hingga 2024 dengan menggunakan metode ARIMA menunjukkan bahwa terjadinya *trend* penurunan jumlah mahasiswa baru secara merata dengan nilai kesalahan atau nilai MAPE sebesar 7.066147 dengan akurasi sebesar 93% yang berarti model bisa digunakan untuk meramalkan jumlah mahasiswa baru dengan sangat akurat.

Dari hasil analisis peramalan disarankan untuk penelitian selanjutnya diharapkan dapat mengkombinasikan metode ARIMA dengan metode *machine learning* yang lain agar tingkat akurasi peramalan menjadi lebih baik lagi.

REFERENSI

- Achmanda, A. D. (2018). Peramalan Jumlah Penderita Penyakit Demam Berdarah Dengue (DBD) Menggunakan Metode ARIMA. *Skripsi Sarjana*.
<http://repositori.usu.ac.id/handle/123456789/12377>
- Aminuddin, Y. F. (2018). Penerapan Filter Kalman Dalam Perbaikan Hasil Prediksi Return Harga Minyak Mentah Dunia Dengan Model ARIMA (Vol. 10, Issue 2).
- As'ad, M., Wibowo, S. S., & Sophia, E. (2017). Peramalan Jumlah Mahasiswa Baru Dengan Model Autoregressive Integrated Moving Average (Arima). *J I M P - Jurnal Informatika Merdeka Pasuruan*, 2(3), 20–33.
<https://doi.org/10.37438/jimp.v2i3.77>
- Dineva, K., & Atanasova, T. (2018). Osemn Process for Working Over Data Acquired By Iot. *Current Trends in Natural Sciences*, 7(13), 47–53.
- Indrasetyaningsih, A., Damayanti, I., & Susanto, T. (2017). Analisis ARIMA Box Jenkins untuk Peramalan Jumlah Kunjungan Wisatawan Mancanegara di Indonesia. *Seminar Nasional Matematika Dan Aplikasinya*, 226–229.
- Irawan, W. (2019). Peramalan Harga Saham PT.Unilever Tbk dengan Menggunakan Metode ARIMA. *Jurnal Matematika UNAND*, 4(3), 80. <https://doi.org/10.25077/jmu.4.3.80-89.2015>
- Janssens, J. (2019). *Data Science at the Command Line*. 08 Oktober. <https://www.bookstack.cn/read/data-science-at-the-command-line/spilt.2.906470b2bf1cf715.md#1.2.1> Obtaining Data
- Lau, D. C. H. (2019). *5 Steps of a Data Science Project Lifecycle*. 3 Januari. <https://towardsdatascience.com/5-steps-of-a-data-science-project-lifecycle-26c50372b492>
- Muhammad, M., Harjono, & Akhsani, L. (2017). Peramalan Mahasiswa Baru FT dan FKIP UM Purwokerto Dengan Model Arima. *Journal Techno*, 18(2), 123–132.
- Ningsih, S. C., & Jana, P. (2018). Pemodelan Penderita HIV / AIDS dengan Metode ARIMA. *Jurnal Penelitian Matematika Dan Pendidikan Matematika*, 3(1), 23–28.
- Niswatin, R. K. (2016). Sistem Seleksi Penerimaan Mahasiswa Baru Menggunakan Metode Weighted Product (Wp). *Seminar Nasional Teknologi Informasi Dan Multimedia 2016*, 31–36.
- Raymond. (2018). *Enlightening Indonesian Data Scientists*. 03 November. <https://www.raymond4ds.com/2018/11/time-series-modelling-arima-step-by-step.html>

- Susanto, Y. (2016). Pemodelan Curah Hujan dengan Pendekatan Model ARIMA, Feed Forward Neural Network dan Hybrid (ARIMA-NN) di Banyuwangi. *Jurnal Sains Dan Seni ITS*, 5(2), 145–150.
- Suseno, S. W. (2017). Penerapan Metode Arima Box-Jenkins Untuk Peramalan Pasien Rawat Jalan Di Rsud Kartini Kabupaten Jepara Berbantuan Eviews. *Skripsi*.
<https://lib.unnes.ac.id/32221/>

PROFIL PENULIS

Ayu Ulfa Jamila, lahir di Tasik Raja pada 18 September 1999. Mahasiswa semester VII Program Studi Sistem Informasi STMIK Mikroskil Medan.

Bella Merlin Siregar, lahir di Kisaran pada 04 April 1999. Mahasiswa semester VII Program Studi Sistem Informasi STMIK Mikroskil Medan.

Roni Yunis, lahir di Bukittinggi pada 19 April 1975. Dosen Tetap di Program Studi Sistem Informasi STMIK Mikroskil Medan.