

## PREDIKSI JUMLAH MAHASISWA BARU MENGGUNAKAN METODE REGRESI LINIER SEDERHANA

N. Almumtazah<sup>1</sup>, N. Azizah<sup>2</sup>, Y. L. Putri<sup>3</sup>, dan Dian C. R. Novitasari<sup>4</sup>

<sup>1,2,3,4</sup> Universitas Islam Negeri Sunan Ampel Surabaya

<sup>1</sup>nilnaalmumtazah@gmail.com, <sup>2</sup>azizahmnurul@gmail.com, <sup>3</sup>yunialaila11@gmail.com, <sup>4</sup>diancrini@uinsby.ac.id

### ABSTRACT

The number of new students at a university during the last 5 years, from 2016 to 2020, has increased and decreased. The data will then be used to predict the number of students in the next 5 years. The purpose of making a prediction is to determine the ratio of available lecturers to the number of new students, to prepare lecture rooms and other facilities. One of the methods using past data to make predictions is the linear regression method. In this study, the independent variable is the period of the academic year, while the dependent variable is the number of new students. The data to be used is data from new students of the faculty of science and technology which consists of 6 study programs with MAPE (*Mean Absolute Percentage Error*) values, namely mathematics (7,2%), marine science (8,76%), biology (5,84%), information systems (6,46%), architecture (7,98%), and environmental engineering (7,52%). Based on the results of the data analysis, a predictive graph of the number of new students in the next 5 years is also obtained with a decreasing linear graph pattern for each study program.

**Keywords** : Linear Regression, New students, Mean Absolute Percentage Error, Prediction

### ABSTRAK

Jumlah mahasiswa baru suatu universitas selama 5 tahun terakhir yaitu mulai tahun 2016 hingga tahun 2020 mengalami kenaikan dan penurunan. Data tersebut selanjutnya akan digunakan untuk memprediksi jumlah mahasiswa 5 tahun mendatang. Tujuan dibuat sebuah prediksi adalah untuk mengetahui rasio dosen yang tersedia dengan jumlah mahasiswa baru, mempersiapkan ruang kuliah dan juga fasilitas lainnya. Salah satu metode dengan penggunaan data masa lampau untuk melakukan prediksi adalah dengan metode regresi linier. Pada penelitian ini yang menjadi variabel bebas adalah periode tahun akademik sedangkan yang menjadi variabel terikat adalah jumlah mahasiswa baru. Data yang akan digunakan merupakan data mahasiswa baru fakultas sains dan teknologi yang terdiri dari 6 program studi dengan nilai MAPE (*Mean Absolute Percentage Error*) yaitu matematika (7,2%), ilmu kelautan (8,76%), biologi (5,84%), sistem informasi (6,46%), arsitektur (7,98%), dan teknik lingkungan (7,52%). Berdasarkan hasil analisis data tersebut juga didapatkan grafik prediksi jumlah mahasiswa baru 5 tahun kedepan dengan pola grafik linier menurun untuk setiap program studi.

**Kata kunci** : Regresi Linier, Mahasiswa Baru, Mean Absolute Percentage Error, Prediksi

## I. PENDAHULUAN

Perguruan tinggi merupakan lembaga lanjutan pendidikan menengah yang ditujukan untuk menjadikan mahasiswa sebagai anggota masyarakat yang dalam penerapan dan pengembangannya memiliki kemampuan akademis dan profesional (Bhakti, Kusdinar and Sunarto, 2020). Adapun orang yang menuntut ilmu di perguruan tinggi baik universitas, sekolah tinggi, institut, akademi, maupun politeknik disebut mahasiswa (Afkarina, Widodo and Furqon, 2019). Sedangkan mahasiswa yang menjalankan kuliah pada tahun pertamanya disebut mahasiswa baru (Eniyati, Santi and Arianto, 2020).

Pada suatu perguruan tinggi negeri mahasiswa baru diterima melalui 3 jalur seleksi yaitu SNMPTN jalur undangan berdasarkan nilai rapor 5 semester-, SBMPTN–jalur seleksi berdasarkan hasil UTBK–, dan mandiri–jalur seleksi melalui serangkaian tes masing-masing perguruan tinggi–(Suwena, 2017). Jumlah mahasiswa baru yang diterima setiap tahunnya akan mempengaruhi proses perkuliahan pada perguruan tinggi tersebut. Salah satu hal yang dapat digunakan sebagai perencanaan saat perkuliahan adalah jumlah mahasiswa baru pada tahun mendatang, hal ini dikarenakan banyaknya jumlah mahasiswa baru akan berkaitan dengan jumlah dosen yang tersedia. Selain itu, penyediaan ruang kuliah dan fasilitas lainnya juga dapat dipersiapkan dengan baik. Oleh sebab itu penting untuk dilakukan prediksi terhadap jumlah mahasiswa baru pada tahun mendatang.

Menurut Muhammad, dkk (2020) prediksi merupakan suatu tindakan untuk memperkirakan keadaan pada masa mendatang berdasarkan data masa lampau. Salah satu metode untuk melakukan prediksi yang dapat digunakan adalah metode regresi linier. Metode regresi linier sendiri terdiri dari 2 jenis yaitu metode regresi linier sederhana dan metode regresi linier berganda. Pada penelitian ini menggunakan metode regresi linier sederhana dengan satu variabel bebas.

Beberapa penelitian sebelumnya telah membuktikan metode ini, seperti penelitian Bhakti, dkk (2020) yang membahas peramalan penerimaan calon mahasiswa menggunakan metode regresi. Pada penelitian ini hasil peramalan diperkuat dengan metode MAPE yang menghasilkan nilai 2,7%. Penelitian dengan menggunakan metode yang sama juga dilakukan Amiruddin dan Ishak (2018) membahas prediksi jumlah mahasiswa persemester. Penelitian ini menghasilkan tingkat error sebesar 4,24% untuk prodi teknik informatika dan 7,69% untuk prodi ilmu hukum. Pada penelitian ini aplikasi yang sudah dibangun memiliki tingkat keakuratan lebih dari 90% sehingga dapat memprediksi jumlah mahasiswa secara tepat. Penelitian selanjutnya dilakukan oleh Mulyani, dkk (2020) dengan metode regresi linier berganda membahas estimasi harga jual mobil bekas. Tujuan dari pembuatan penelitian ini untuk menentukan harga mobil bekas berdasarkan variabel merk, tahun pembuatan, harga beli, dan kondisi mobil. Pada penelitian ini memiliki hasil estimasi dari harga sebesar 2,65%.

Dari beberapa penelitian di atas dapat dikatakan bahwa metode regresi linier mempunyai kemampuan yang baik dalam memprediksi. Sehingga pada penelitian ini digunakan metode regresi linier sederhana dengan harapan dapat melihat prediksi jumlah mahasiswa baru 5 tahun kedepan

serta didapatkan nilai MAPE (*Mean Absolute Percentage Error*) sekecil mungkin agar hasil prediksi memiliki tingkat keakuratan tinggi.

## II. METODE PENELITIAN

### 2.1. Sumber Data dan Variabel Penelitian

Variabel penelitian adalah suatu nilai/sifat/atribut dari objek, kegiatan, atau orang dengan suatu variasi yang ditentukan untuk dikaji dan diambil kesimpulannya oleh peneliti (Sugiyono, 2017). Untuk memperjelas maka diberikan simbol  $X$  sebagai variabel bebas dan simbol  $Y$  sebagai variabel terikat.

#### 1. Variabel bebas

Variabel penyebab atau yang berpengaruh terhadap variabel terikat. Pada penelitian ini diperankan oleh periode tahun akademik.

#### 2. Variabel terikat

Variabel akibat atau yang terpengaruh variabel bebas. Pada penelitian ini diperankan oleh jumlah mahasiswa baru.

Pada penelitian ini data yang digunakan berupa data sekunder, yaitu data yang didapatkan secara tidak langsung atau data yang didapat dari dokumen dan penyampaian orang lain (Rifa'i and Arifin, 2018). Data jumlah mahasiswa baru salah satu universitas di Surabaya fakultas sains dan teknologi 5 tahun terakhir yaitu mulai dari tahun 2016 hingga tahun 2020 telah tersedia sebelumnya sehingga peneliti cukup menghubungi pihak terkait untuk mendapatkan data.

Tabel 1 : Data Jumlah Mahasiswa Baru

Tahun Akademik	Periode Tahun Akademik ( $X$ )	Jumlah Mahasiswa Baru ( $Y$ )					
		Matematika	Ilmu Kelautan	Biologi	Sistem Informasi	Arsitektur	Teknik Lingkungan
2016/2017	1	73	71	72	78	74	73
2017/2018	2	63	60	64	69	67	64
2018/2019	3	54	51	59	73	72	68
2019/2020	4	64	65	56	61	57	59
2020/2021	5	63	53	62	70	73	75
Total	15	317	300	313	351	343	339

### 2.2. Metode Regresi Linier Sederhana

Analisis regresi merupakan perhitungan statistik untuk menguji seberapa erat hubungan antar variabel (Trianggana, 2020). Analisis regresi yang paling sederhana dan sering digunakan adalah regresi linier sederhana. Dalam analisis regresi terdapat satu variabel terikat

yang biasa ditulis dengan simbol  $Y$  dan satu variabel bebas atau lebih yang biasa ditulis dengan simbol  $X$ . Hubungan kedua variabel tersebut memiliki sifat linier sesuai dengan namanya (Amiruddin and Ishak, 2018).

Regresi linier merupakan salah satu perhitungan *time series* metode kuantitatif dimana waktu digunakan sebagai dasar prediksi. Berikut persamaan dasar metode regresi linier sederhana :

$$Y = a + bX \tag{1}$$

dimana :

$Y$  = Variabel terikat

$a$  = Intercept

$b$  = Koefisien variabel  $X$

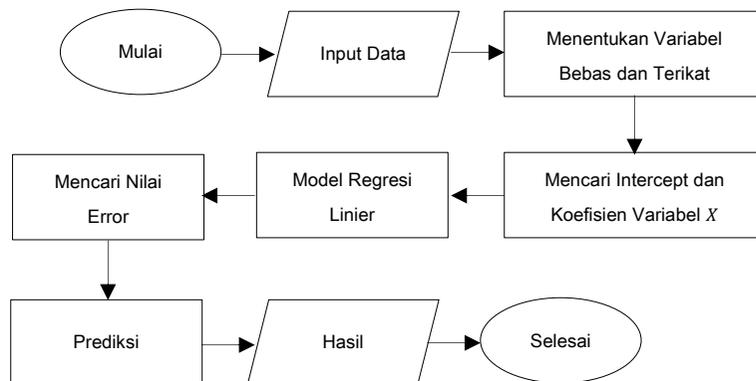
$X$  = Variabel bebas

Selanjutnya digunakan persamaan berikut untuk mencari intercept dan koefisien variabel :

$$a = \frac{(\sum X^2)(\sum Y) - (\sum XY)(\sum X)}{n(\sum X^2) - (\sum X)^2} \tag{2}$$

$$b = \frac{n(\sum XY) - (\sum X)(\sum Y)}{n(\sum X^2) - (\sum X)^2} \tag{3}$$

Adapun tahapan metode regresi linier sederhana adalah sebagai berikut :



Gambar 1 : Diagram Alir Tahapan Penelitian

### 2.3. MAPE (*Mean Absolute Percentage Error*)

MAPE merupakan pengujian hasil sesuai dengan model prediksi yang sudah dilakukan menggunakan data masa lampau. Hudyanti, dkk (2019) menjelaskan untuk menentukan tingkat keakuratan dapat menggunakan metode MAPE dengan menghitung selisih dari data aktual dan data prediksi. Semakin kecil nilai MAPE maka tingkat keakuratan semakin tinggi,

dengan kata lain metode yang digunakan mempunyai kemampuan yang baik dalam memprediksi (Maricar, 2019). Berikut rumus perhitungan MAPE :

$$MAPE = \frac{\sum \frac{|Y - Y'|}{Y} \times 100\%}{n} \quad (4)$$

dimana :

$Y$  = Data aktual

$Y'$  = Data prediksi

$n$  = Jumlah data

### III. HASIL DAN PEMBAHASAN

Pembentukan model regresi linier yang terdiri dari :

a. Menentukan variabel bebas dan variabel terikat.

Penentuan variabel bebas dan terikat telah disebutkan dalam metode penelitian, dimana variabel bebas diperankan oleh periode tahun akademik dengan simbol  $X$  dan variabel terikat diperankan oleh jumlah mahasiswa baru dengan simbol  $Y$ .

b. Mencari intercept dan koefisien  $X$  dengan menggunakan persamaan (2) dan (3).

Sebelum mencari intercept dan koefisien  $X$  maka dilakukan perhitungan  $X^2$  dan  $XY$  menggunakan data pada Tabel 1.

Tabel 2 : Perhitungan  $X^2$  dan  $XY$

$X$	$X^2$	$XY$					
		Matematika	Ilmu Kelautan	Biologi	Sistem Informasi	Arsitektur	Teknik Lingkungan
1	1	73	71	72	78	74	73
2	4	126	120	128	138	134	128
3	9	162	153	177	219	216	204
4	16	256	260	224	244	228	236
5	25	315	265	310	350	365	375
Total	55	932	869	911	1029	1017	1016

Selanjutnya mencari intercept dan koefisien  $X$  dengan menggunakan persamaan (2) dan (3).

Tabel 3 : Mencari Intercept dan Koefisien  $X$

Nilai	Matematika	Ilmu Kelautan	Biologi	Sistem Informasi	Arsitektur	Teknik Lingkungan
Intercept	69,1	69,3	71,0	77,4	72,2	68,1
Koefisien $X$	-1,9	-3,1	-2,8	-2,4	-1,2	-0,1

- c. Membuat model persamaan regresi linier sederhana menggunakan persamaan (1) dan melakukan prediksi untuk mencari nilai  $Y$  (2021/2022) dengan  $X = 6$

Tabel 4 : Persamaan Regresi Linier Sederhana dan Prediksi

Program Studi	Persamaan Regresi	Prediksi
Matematika	$Y = 69,1 + (-1,9)X$	$Y = 69,1 + (-1,9)6 = 57,7 = 58$
Ilmu Kelautan	$Y = 69,3 + (-3,1)X$	$Y = 69,3 + (-3,1)6 = 50,7 = 51$
Biologi	$Y = 71,0 + (-2,8)X$	$Y = 71,0 + (-2,8)6 = 54,2 = 54$
Sistem Informasi	$Y = 77,4 + (-2,4)X$	$Y = 77,4 + (-2,4)6 = 63$
Arsitektur	$Y = 72,2 + (-1,2)X$	$Y = 72,2 + (-1,2)6 = 65$
Teknik Lingkungan	$Y = 68,1 + (-0,1)X$	$Y = 68,1 + (-0,1)6 = 67,5 = 68$

Pengujian MAPE dengan menggunakan persamaan (4) berdasarkan data prediksi yang telah dibuat. Untuk melakukan pengujian MAPE, maka dilakukan prediksi terlebih dahulu dengan menggunakan data pada Tabel 1.

Tabel 5 : Perhitungan Data Prediksi

Tahun Akademik	Data Prediksi ( $Y'$ )					
	Matematika	Ilmu Kelautan	Biologi	Sistem Informasi	Arsitektur	Teknik Lingkungan
2016/2017	67	66	68	75	71	68
2017/2018	65	63	65	73	70	68
2018/2019	63	60	63	70	69	68
2019/2020	62	57	60	68	67	68
2020/2021	60	54	57	65	66	68

Tabel 6 : Perhitungan Selisih

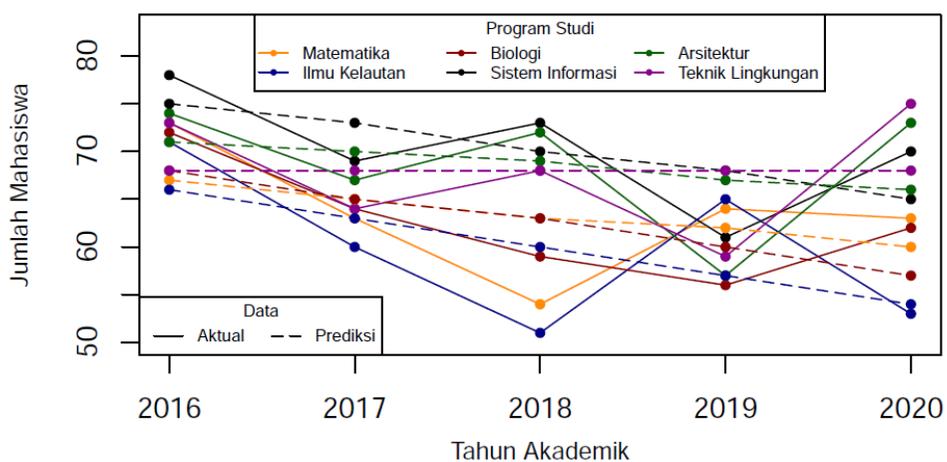
Tahun Akademik	Selisih ( $Y - Y'$ )					
	Matematika	Ilmu Kelautan	Biologi	Sistem Informasi	Arsitektur	Teknik Lingkungan
2016/2017	6	5	4	3	3	5
2017/2018	-2	-3	-1	-4	-3	-4
2018/2019	-9	-9	-4	3	3	0
2019/2020	2	8	-4	-7	-10	-9
2020/2021	3	-1	5	5	7	7

Tabel 7 : Perhitungan Tingkat Error MAPE

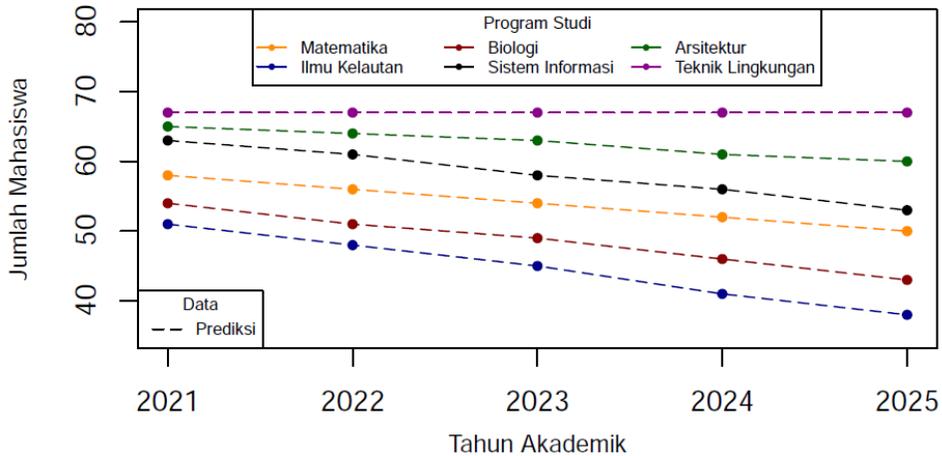
Jumlah Data (n)	Error MAPE $\left(\frac{ Y-Y' }{Y}\right)$					
	Matematika	Ilmu Kelautan	Biologi	Sistem Informasi	Arsitektur	Teknik Lingkungan
1	0,082	0,070	0,056	0,038	0,041	0,068
2	0,032	0,050	0,016	0,058	0,045	0,063
3	0,167	0,176	0,068	0,041	0,042	0
4	0,031	0,123	0,071	0,115	0,175	0,153
5	0,048	0,019	0,081	0,071	0,096	0,093
$\sum \frac{ Y-Y' }{Y}$	0,360	0,438	0,292	0,323	0,399	0,377
$\frac{\sum \frac{ Y-Y' }{Y} \times 100\%}{5}$	7,2%	8,76%	5,84%	6,46%	7,98%	7,52%

Berdasarkan persamaan regresi linier sederhana pada Tabel 4 maka didapatkan hasil prediksi data mahasiswa baru 5 tahun terakhir yang ditampilkan pada Tabel 5. Selanjutnya menghitung selisih antara data aktual ( $Y$ ) dan data prediksi ( $Y'$ ) 5 tahun terakhir. Setelah didapatkan selisih data, selanjutnya akan dilakukan perhitungan tingkat error menggunakan MAPE dan didapatkan hasil sebagaimana pada Tabel 7.

Berdasarkan Tabel 7 didapatkan nilai MAPE untuk prodi matematika, ilmu kelautan, biologi, sistem informasi, arsitektur, dan teknik lingkungan dengan masing-masing sebesar 7,2%, 8,76%, 5,84%, 6,46%, 7,98%, dan 7,52%. Dengan tingkat akurasi masing-masing 92,8%, 91,24%, 94,16%, 93,54%, 92,02%, dan 92,48%. Berikut akan ditampilkan grafik perbandingan data aktual dan data prediksi selama 5 tahun terakhir yaitu tahun 2016 hingga tahun 2020, serta grafik data prediksi 5 tahun kedepan yaitu tahun 2021 hingga tahun 2025.



Gambar 2 : Grafik Data Aktual dan Data Prediksi 5 Tahun Terakhir (2016-2020)



Gambar 3 : Grafik Data Prediksi 5 Tahun Kedepan (2021-2025)

Berdasarkan Gambar 2, terlihat jelas bahwa data hasil prediksi 5 tahun terakhir (2016-2020) dengan data aktual 5 tahun terakhir (2016-2020) kurang akurat. Salah satu kelemahan dari metode regresi linier yang terlihat jelas pada grafik Gambar 3, yaitu apabila nilai  $b$  negatif maka grafik akan terus mengalami penurunan, begitu juga sebaliknya apabila nilai  $b$  positif maka grafik akan terus mengalami kenaikan. Sedangkan pada data aktual dapat dilihat pada Gambar 2, grafik mengalami kenaikan dan penurunan.

Dalam penelitian ini data yang digunakan merupakan data *time series*, sehingga pada penelitian selanjutnya dapat menggunakan metode-metode analisis data yang berkarakteristik rangkaian data atau yang biasa disebut *sequence data*, seperti menggunakan metode Jaringan Syaraf Tiruan (JST) (Syarifudin *et al.*, 2021), *Support Vector Regression* (SVR) (Suwanto and Novitasari, 2020), dan *Long Short-Term Memory* (LSTM) (Haq *et al.*, 2021) dalam memprediksi jumlah mahasiswa baru.

#### IV. KESIMPULAN

Berdasarkan pembahasan penelitian di atas dapat disimpulkan, untuk mendapatkan model regresi maka digunakan data mahasiswa baru 5 tahun terakhir yaitu tahun 2016 hingga tahun 2020. Selanjutnya model tersebut digunakan untuk memprediksi jumlah mahasiswa baru 5 tahun mendatang yaitu tahun 2021 hingga tahun 2025. Meskipun dalam perhitungan penelitian ini masih terdapat kekurangan, tetapi dengan menggunakan metode regresi linier juga efektif dalam memprediksi jumlah mahasiswa baru. Dari perhitungan diatas didapat nilai MAPE untuk prodi matematika sebesar 7,2%, prodi ilmu kelautan sebesar 8,76%, prodi biologi sebesar 5,84%, prodi sistem informasi sebesar 6,46%, prodi arsitektur sebesar 7,98%, dan nilai MAPE untuk prodi teknik lingkungan sebesar 7,52%.

## DAFTAR PUSTAKA

- [1] Afkarina, N. K., Widodo, A. W. and Furqon, M. T., Implementasi Regresi Linier Berganda Untuk Prediksi Jumlah Peminat, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(11), 2019, pp. 10462–10467.
- [2] Amiruddin and Ishak, R., Prediksi Jumlah Mahasiswa Registrasi Per Semester Menggunakan Linier Regresi pada Universitas Ichsan Gorontalo, *ILKOM Jurnal Ilmiah*, 10(2), 2018, pp. 136–143.
- [3] Bhakti, Y. S., Kusdinar, A. B. and Sunarto, A. A., Model Peramalan Penerimaan Calon Mahasiswa Menggunakan Metode Regresi, *Progresif: Jurnal Ilmiah Komputer*, 16(2), 2020, pp. 113–120.
- [4] Eniyati, S., Santi, R. C. N. and Arianto, T., Penggunaan Metode Lagrange dalam Peramalan Jumlah Mahasiswa Baru, *Proceeding SENDIU*, 2020, pp. 263–266.
- [5] Haq, D. Z. *et al.*, Long Short-Term Memory Algorithm for Rainfall Prediction Based on El-Nino and IOD Data, *Procedia Computer Science*. Elsevier B.V., 179(2019), 2021, pp. 829–837.
- [6] Hudiyantri, C. V., Bachtiar, F. A. and Setiawan, B. D., Perbandingan Double Moving Average dan Double Exponential Smoothing untuk Peramalan Jumlah Kedatangan Wisatawan Mancanegara di Bandara Ngurah Rai, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(3), 2019, pp. 2667–2672.
- [7] Maricar, M. A., Analisa Perbandingan Nilai Akurasi Moving Average dan Exponential Smoothing untuk Sistem Peramalan Pendapatan pada Perusahaan XYZ, *Jurnal Sistem dan Informatika*, 13(2), 2019, pp. 36–45.
- [8] Muhammad, I. *et al.*, Peramalan Jumlah Mahasiswa Baru Menggunakan Metode Double Exponential Smoothing (Studi Kasus: Mahasiswa Baru Universitas Pattimura Ambon Tahun 2017), *Jurnal Variance*, 2(1), 2020, pp. 27–33.
- [9] Mulyani, E. D. S. *et al.*, Estimasi Harga Jual Mobil Bekas Menggunakan Metode Regresi Linier Berganda, *Jurnal Sistem Informasi dan Teknologi Informasi*, 9(1), 2020, pp. 1–8.
- [10] Rifa'i, A. and Arifin, Z., Metode Regresi Linier Untuk Memprediksi Jumlah Pendaftar Calon Mahasiswa Baru Universitas Islam Indragiri, 4, 2018, pp. 86–92.
- [11] Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*, Alfabeta, 2017, Bandung.
- [12] Suwanto, S. and Novitasari, D. C. R., Forecasting Solar Activities based on Sunspot Number Using Support Vector Regression (SVR), *JPSE (Journal of Physical Science and*

*Engineering*), 5(1), 2020, pp. 6–10.

- [13] Suwena, K. R., Jalur Penerimaan Mahasiswa Baru Bukan Penentu Prestasi Belajar Mahasiswa', *Ekuitas-Jurnal Pendidikan Ekonomi*, 5(2), 2017, pp. 1–10.
- [14] Syarifudin, M. A. *et al.*, Hotspot Prediction Using 1D Convolutional Neural Network, *Procedia Computer Science*. Elsevier B.V., 179(2019), 2021, pp. 845–853.
- [15] Trianggana, D. A., Peramalan Jumlah Siswa-Siswi Melalui Pendekatan Metode Regresi Linear, *Jurnal Media Infotama*, 16(2), 2020, pp. 115–120.