

Prediksi Hasil Belajar Peserta Didik Menggunakan Model Multiple Linier Regression

¹Firmansyah, ^{2*}Agus Yulianto

¹Universitas Bina Sarana Informatika, ²Universitas Nusa Mandiri
Jakarta, Indonesia

firmansyah.fmh@bsi.ac.id, agus.asy@nusamandiri.ac.id

*Penulis Korespondensi

Diajukan : 01/09/2022

Diterima : 22/09/2022

Dipublikasi : 02/10/2022

ABSTRAK

Hasil belajar merupakan prestasi yang dicapai oleh peserta didik dengan mengikuti proses akademik seperti pengajaran, diskusi hingga ujian. Permasalahan institusi pendidikan saat ini adalah tidak dapat mengantisipasi dan memprediksi hasil belajar di masa depan. Dengan memprediksi hasil belajar, institusi dapat melakukan langkah preventif terhadap proses belajar pada periode berikutnya. Untuk dapat menentukan kebijakan dalam proses belajar mengajar, maka dibuat prediksi hasil belajar menggunakan metode multiple linier regression dengan menggunakan variabel kehadiran dan peringkat sebagai variabel dependen dan hasil belajar sebagai variabel independen. Multiple linier regression terbukti dapat memprediksi hasil belajar peserta didik berdasarkan variabel kehadiran dan peringkat. Dari nilai koefisien regresi dapat disimpulkan bahwa jika kehadiran turun sebanyak 1 maka persentase kehadiran juga akan turun sebesar 0.03%, sedangkan jika peringkat turun sebanyak 1 maka peringkat siswa akan berkurang sebanyak 0.53.

Kata Kunci: *multiple linier regression*, prediksi, prediksi hasil belajar, *linier regression*

I. PENDAHULUAN

Pendidikan dimulai dari pengajar memberikan pelajaran di kelas, diskusi dan serangkaian proses hingga akhirnya peserta didik mendapatkan hasil belajar. Hasil belajar merupakan prestasi yang dicapai oleh peserta didik dengan mengikuti proses akademik seperti pengajaran, diskusi hingga ujian. Banyak faktor yang mempengaruhi hasil belajar seperti kualitas pendidik, kemampuan guru memahami siswa, penguasaan materi, komunikatif dan memiliki kepribadian yang dewasa (Dakhi, 2020).

Untuk menciptakan hasil belajar yang baik maka harus diketahui terlebih dahulu faktor-faktor yang menentukan hasil belajar peserta didik. Hasil belajar dapat dipengaruhi oleh pengajar, peserta didik dan infrastrukturnya. Salah satu yang dapat dilihat adalah dari sisi peserta didik itu sendiri seperti hasil belajar di masa lalu, tingkat kehadiran, kedisiplinan bahkan status social (Gunawan et al., 2021).

Permasalahan institusi pendidikan saat ini adalah tidak dapat mengantisipasi dan memprediksi hasil belajar di masa depan. Dengan memprediksi hasil belajar, institusi dapat melakukan langkah preventif terhadap proses belajar pada periode berikutnya.

II. STUDI LITERATUR

1. Penelitian Terkait

Penelitian yang terkait dengan prediksi hasil belajar dengan berbagai metode sudah dilakukan oleh para peneliti lainnya seperti memprediksi pertumbuhan usaha mikro kecil dan menengah (UMKM) menggunakan regresi linier berganda dengan 2 variabel dependent yaitu rata-rata suku

bunga dan inflasi (Firmansyah & Yulianto, 2021). Prediksi nilai akademik mahasiswa menggunakan metode naïve bayes, klasifikasinya berdasarkan mahasiswa yang kuliah sambil bekerja, berdasarkan jadwal kerja mahasiswa dan berdasarkan waktu kuliah (Annisa & Sasongko, 2020). Aplikasi untuk dapat memprediksi nilai siswa sekolah dasar menggunakan metode radial basis function dan juga jaringan syaraf tiruan (Afriani et al., 2019). Prediksi hasil belajar siswa secara daring selama pandemic COVID-19 menggunakan algoritma C4.5, atribut yang digunakan yaitu absensi, tugas, ulangan harian dan ujian. Penelitian ini memprediksi hasil belajar siswa apakah memuaskan atau tidak memuaskan (Fitriani et al., 2021).

2. Penilaian Hasil Belajar

Hasil belajar merupakan perubahan tingkah laku sebagai hasil belajar dalam definisi yang lebih luas mencakup bidang kognitif, efektif, dan psikomotorik (Nurdin Mansur, 2015). Secara sederhana, yang dimaksud dengan hasil belajar adalah kemampuan yang diperoleh peserta didik setelah melalui kegiatan belajar. Secara praktis, hasil belajar juga dimaksudkan untuk mengungkapkan kemampuan siswa dalam bentuk angka-angka (Achdiyat & Utomo, 2017) dimana angka tersebut adalah hasil penilaian terhadap kemampuan peserta didik setelah menjalani proses pembelajaran. Dan dimaksudkan untuk mengetahui kemampuan peserta didik (Azizah & Widjajanti, 2019).

3. Regresi Linier

Regresi linier diperkenalkan pertama kali oleh Sir Francis Galton pada tahun 1894. Regresi linier adalah model statistik yang diaplikasikan ke dalam set data untuk mendapatkan hasil relasi antar variabel (Kumari & Yadav, 2018). Regresi linier memiliki variabel X dan Y, dimana variable X merupakan variabel dependen dan variabel Y adalah variabel independen. Prediksi terhadap nilai variabel dependen dapat dilakukan apabila variabel independennya diketahui (Firmansyah & Yulianto, 2021). Secara sederhana, multiple linier regression adalah metode statistik untuk memprediksi atau melihat pengaruh variabel dependen terhadap variabel independen. Rumus untuk regresi linier berganda menggunakan persamaan di bawah ini :

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$y = a + b.x$$

Keterangan :

a, b = konstanta

x = variable dependent

y = variable independent

n = jumlah data

III. METODE

A. CRISP-DM (Cross-Industry Standard Process for Data Mining)

CRISP-DM (Cross-Industry Standard Process for Data Mining) digunakan sebagai kerangka kerja dalam menerapkan proses data mining (Larose, 2006) adapun prosesnya yaitu :

1. Business Understanding Phase
2. Data Understanding Phase
3. Data Preparation Phase

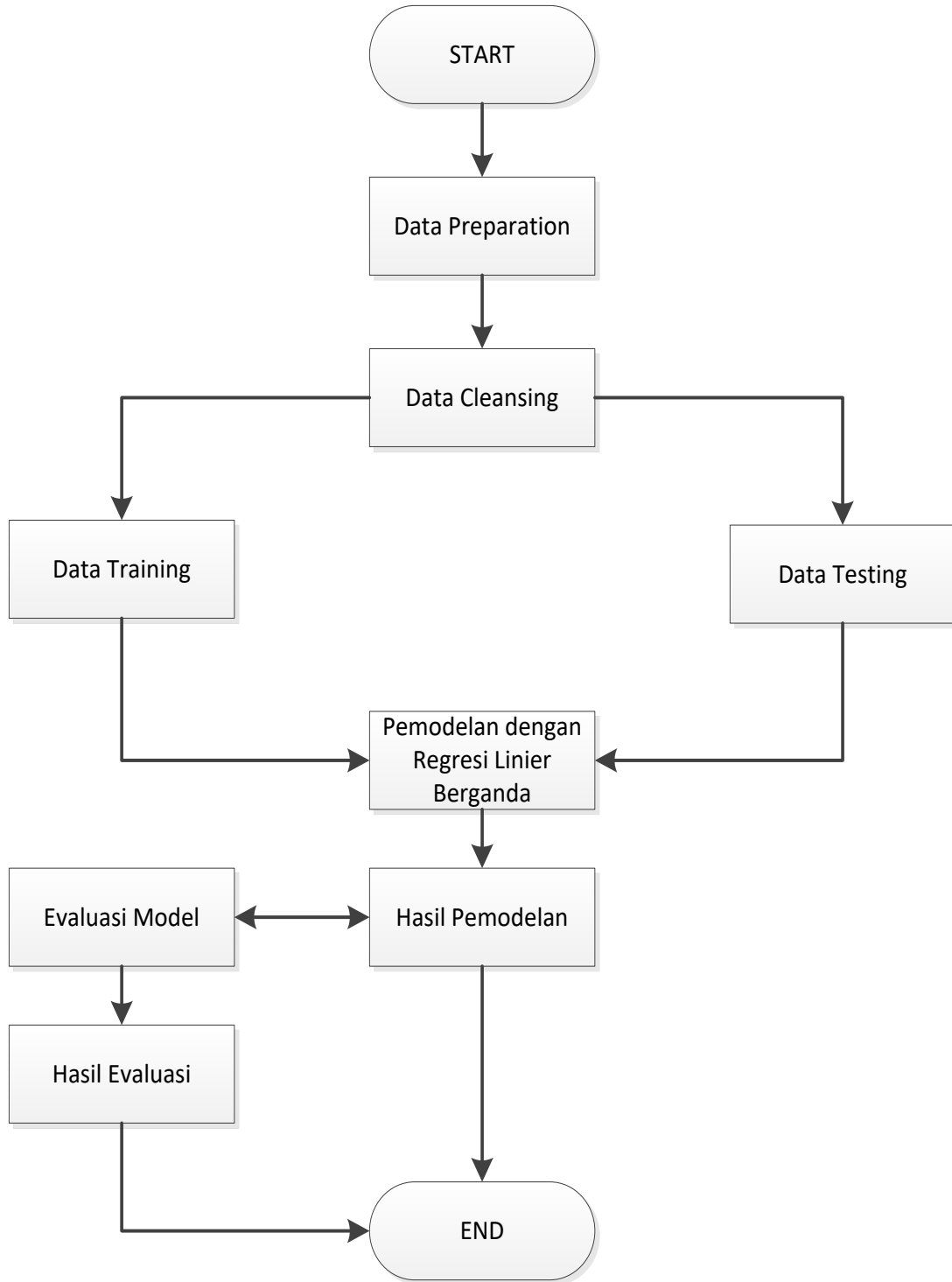
4. Modelling Phase
5. Evaluation Phase
6. Deployment Phase



Gambar 1 CRISP-DM Process

B. Flowchart Pemodelan Regresi Linier Berganda

Flowchart digunakan untuk menggambarkan alur proses mulai dari data preparation hingga pemodelan. Flowchart untuk menggambarkan proses pemodelan menggunakan metode regresi linier berganda ini seperti di bawah ini :



Gambar 2 Proses Pemodelan

IV. HASIL DAN PEMBAHASAN

1. Business Understanding Phase

Hasil belajar merupakan salah satu indikator prestasi dari peserta didik sehingga perlu adanya model algoritma yang dapat memprediksi hasil belajar peserta didik. Tujuannya adalah sebagai acuan instansi dalam melakukan langkah-langkah preventif yang dapat mempengaruhi hasil belajar.

2. Data Understanding Phase

Sumber data merupakan data primer hasil belajar peserta didik yang diambil secara acak dari sekolah dasar negeri di wilayah Jakarta Barat untuk periode akademik tahun 2019-2020.

3. Data Preparation Phase

Total data yang diambil berjumlah 300 baris dengan 2 atribut independen (X) dan 1 atribut dependen (Y). Detailnya dibawah ini :

Tabel 1 Klasifikasi Variabel

Variable name		Type data
Kehadiran (X1)		Numerik (persentase)
Peringkat (X2)		Numerik (persentase)
Hasil Belajar (Y)		Numerik

Dari total data sebanyak 300, dilakukan cleaning redundansi terhadap data sehingga data training berjumlah 248 baris.

4. Modelling Phase

Data training dibuatkan dalam bentuk spreadsheet dimana ada 2 variabel dependent (X1 dan X2) dan variabel independent (Y) dengan periode tahun 2019 sampai dengan 2020. Setiap kolom dihitung jumlah dan rata-rata menggunakan tabel di bawah ini :

Tabel 2 Perhitungan variabel X dan Y

NO URUT	KEHADIRAN (X1)	PERINGKAT (X2)	HASIL BELAJAR (Y)
1	100	12	84
2	100	3	93
3	98	22	78
4	98	2	93
5	98	16	83
6	98	20	79
7	98	8	88
8	98	4	92
9	98	1	94
10	96	17	83
...
248	96	20	83
Jumlah	23,984	2,665	21,024
Rata-rata	96.35	10.75	84.77

Langkah berikutnya adalah menghitung masing-masing variabel seperti di bawah ini :

Tabel 3 Perhitungan variabel bagian 1

NO URUT	$y_i = Y_i - \bar{Y}$	$x_{1i} = X_{1i} - \bar{X}_1$	$x_{2i} = X_{2i} - \bar{X}_2$
1	-0.77	3.65	1.25
2	8.23	3.65	-7.75
3	-6.77	1.65	11.25
4	8.23	1.65	-8.75
5	-1.77	1.65	5.25
6	-5.77	1.65	9.25
7	3.23	1.65	-2.75
8	7.23	1.65	-6.75
9	9.23	1.65	-9.75
10	-0.77	3.65	1.25
...
248	-1.77	-0.35	9.25
SUM	20,939.23	23,797.65	2654.25

Tabel 4 Perhitungan variabel bagian 2

NO URUT	y_i^2	X_{1i}^2	X_{2i}^2
1	0.60	13.35	1.57
2	67.66	13.35	60.00
3	45.89	2.73	126.65
4	67.66	2.73	76.49
5	3.15	2.73	27.60
6	33.34	2.73	85.64
7	10.41	2.73	7.54
8	52.21	2.73	45.51
9	85.12	2.73	94.98
10	3.15	0.12	39.11
...
248	3.15	0.12	85.64
SUM	10,943.35	2,178.18	8,743.35

Tabel 5 Perhitungan variabel bagian 3

NO URUT	$y_i X_{1i}$	$y_i X_{2i}$	$X_{1i} X_{2i}$
1	-2.83	-0.97	4.58
2	30.05	-63.72	-28.30
3	-11.20	-76.24	18.61

4	13.60	-71.94	-14.46
5	-2.93	-9.32	8.69
6	-9.55	-53.43	15.30
7	5.33	-8.86	-4.54
8	11.95	-48.75	-11.15
9	15.25	-89.91	-16.11
10	0.62	-11.10	-2.17
...
248	0.62	16.42	3.21
SUM	-107.58	-5,462.23	-491.15

Setelah menghitung variabel, maka selanjutnya adalah menghitung nilai koefisien b1, b2 dan b0 dimana b1 dan b2 merpresentasikan X1 dan X2 sedangkan b0 adalah intercept, untuk menghitung b1 menggunakan persamaan berikut :

$$\hat{b}_1 = \frac{(\sum x_1 y)(\sum x_2^2) - (\sum x_2 y)(\sum x_1 x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

$$\hat{b}_1 = \frac{(-107.58 * 8743)(-5462.23 * -491.15)}{(2178.18 * 8743) - (-491.15)^2}$$

$$\hat{b}_1 = -0.19$$

Untuk menghitung b2 menggunakan persamaan berikut :

$$\hat{b}_2 = \frac{(\sum x_2 y)(\sum x_1^2) - (\sum x_1 y)(\sum x_1 x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

$$\hat{b}_2 = \frac{(-5462 * 2178.18)(-107.58 * -491.15)}{(2178.18 * 8743) - (-491.15)^2}$$

$$\hat{b}_2 = -0.64$$

Untuk menghitung nilai b0 menggunakan persamaan berikut :

$$\hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X}_1 - \hat{b}_2 \bar{X}_2$$

$$\hat{b}_0 = 84.77 - (-0.19 * 96.35) - (-0.64 * 10.75)$$

$$\hat{b}_0 = 110.17$$

Kemudian menghitung R square dengan persaman berikut :

$$R^2 = \frac{\hat{b}_1(\sum x_2 y) + \hat{b}_2(\sum x_1 y)}{\sum y^2}$$

$$R^2 = \frac{(-0.19) * (-107.58) + (-0.64) * (-5462.23)}{10943.35}$$

Menentukan nilai Y yang nantinya akan menjadi persamaan untuk prediksi.

$$Y = \hat{b}_0 + \hat{b}_1X_1 + \hat{b}_2X_2$$

$$Y = 93.59 + (-0.03 * X1) + (-0.56 + X2)$$

Persamaan Y di atas merupakan persamaan yang bisa dijadikan sebagai prediksi, cukup memasukkan nilai X1 dan X2. Jika kita masukkan nilai kehadiran adalah 86% dan peringkat siswa adalah 2 maka prediksi nilai siswa adalah 90.31, detailnya seperti di bawah ini :

$$Y = 93.59 + (-0.03 * X1) + (-0.56 + X2)$$

$$Y = 93.59 + (-0.03 * 86) + (-0.56 + 2)$$

$$Y = 90.31$$

Di bawah ini merupakan statistik regresi dan koefisien regresi yang direpresentasikan dalam tabel :

Tabel 6 Statistik Regresi

<i>Regression Statistics</i>	
Multiple R	0.86
R Square	0.73
Adjusted R Square	0.73
Standard Error	2.02
Observations	248

Tabel 7 Koefisien Regresi

<i>Coefficients</i>	
Intercept	93.59
Kehadiran	-0.025
Peringkat	-0.561

5. Evaluation Phase

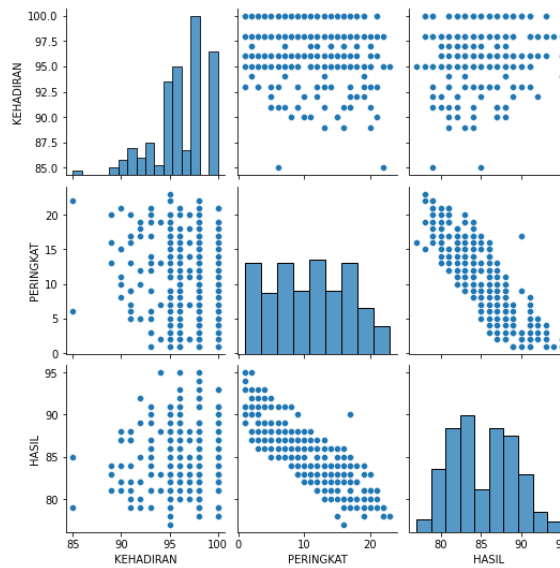
Untuk pengujian akurasi menggunakan model RMSE (*Root Mean Square Error*), MSE (*Mean Square Error*) dan MAPE (*Mean Percentage Absolute Error*). Hasil perhitungannya dalam tabel di bawah ini :

Tabel 8 Pengujian Model

RMSE	5.237
MAPE	0.322
MSE	27.432

6. Deployment Phase

Implementasi menggunakan Python pada Jupyter Notebook dengan menampilkan grafik scatter plot seperti di bawah ini :



Gambar 3 Scatter plot

Python juga dapat menampilkan hasil pengujian seperti dibawah ini :

OLS Regression Results				
Dep. Variable:	HASIL	R-squared:	0.319	
Model:	OLS	Adj. R-squared:	0.314	
Method:	Least Squares	F-statistic:	57.42	
Date:	Wed, 03 Aug 2022	Prob (F-statistic):	3.55e-21	
Time:	06:35:43	Log-Likelihood:	-773.83	
No. Observations:	248	AIC:	1554.	
Df Residuals:	245	BIC:	1564.	
Df Model:	2			
Covariance Type: nonrobust				
	coef	std err	t	P> t [0.025 0.975]
const	110.1707	11.552	9.537	0.000 87.417 132.924
x1	-0.1927	0.119	-1.621	0.106 -0.427 0.042
x2	-0.6356	0.059	-10.708	0.000 -0.752 -0.519
Omnibus:	505.420	Durbin-Watson:	1.837	
Prob(Omnibus):	0.000	Jarque-Bera (JB):	344664.597	
Skew:	-12.553	Prob(JB):	0.00	
Kurtosis:	183.899	Cond. No.	3.20e+03	

Gambar 4 Hasil pengujian menggunakan Python

V. KESIMPULAN

Multiple linier regression terbukti dapat memprediksi hasil belajar peserta didik berdasarkan variabel kehadiran dan peringkat. Dari nilai koefisien regresi dapat disimpulkan bahwa jika kehadiran turun sebanyak 1 maka persentase kehadiran juga akan turun sebesar 0.03%, sedangkan jika peringkat turun sebanyak 1 maka peringkat siswa akan berkurang sebanyak 0.53. Hasil pengujian model menggunakan MAPE masih dalam kategori sangat baik karena masih dalam nilai dibawah 10% yaitu 0.322 %. *Multiple linier regression* cukup akurat dalam memprediksi hasil belajar, model ini juga dapat digunakan untuk memprediksi permasalahan lainnya seperti prediksi penjualan, prediksi strategi promosi dan lain-lain.

VI. REFERENSI

- Achdiyat, M., & Utiomo, R. (2017). *Kecerdasan Visual-Spasial, Kemampuan Numerik dan Prestasi Belajar Matematika. 1*(1), 155.
- Afriani, R., Hotlan Sitorus, S., & Ristian, U. (2019). Aplikasi Prediksi Nilai Siswa Sekolah Dasar Menggunakan Metode Radial Basis Function Berbasis Web (Studi Kasus : Sdn 19 Sungai Raya). *Coding: Jurnal Komputer Dan Aplikasi*, 07(03), 85–96.
- Annisa, R., & Sasongko, A. (2020). Prediksi Nilai Akademik Mahasiswa Menggunakan Algoritma Naïve Bayes. *JST (Jurnal Sains Dan Teknologi)*, 9(1), 1–10. <https://doi.org/10.23887/jstundiksha.v9i1.19488>
- Azizah, I. N., & Widjajanti, D. B. (2019). Keefektifan pembelajaran berbasis proyek ditinjau dari prestasi belajar, kemampuan berpikir kritis, dan kepercayaan diri siswa. *Jurnal Riset Pendidikan Matematika*, 6(2), 233–243. <https://doi.org/10.21831/jrpm.v6i2.15927>
- Dakhi, A. S. (2020). Peningkatan Hasil Belajar Siswa. *Jurnal Pendidikan Indonesia*, 1(3), 350–361. <https://doi.org/10.36418/japendi.v1i3.33>
- Firmansyah, & Yulianto, A. (2021). *Prediksi Pertumbuhan Jumlah Unit Usaha Mikro Kecil dan Menengah (UMKM) Menggunakan Metode Regresi Linier Berganda*. 6(April), 1–12.
- Fitriani, Y., Defit, S., & Nurcahyo, G. W. (2021). Prediksi Hasil Belajar Siswa Secara Daring pada Masa Pandemi COVID-19 Menggunakan Metode C4.5. *Jurnal Sistim Informasi Dan Teknologi*, 3, 120–127. <https://doi.org/10.37034/jsisfotek.v3i3.54>
- Gunawan, Kustiani, L., & Hariani, L. S. (2021). Faktor Yang Mempengaruhi Hasil Belajar Siswa. *Mimbar Ilmu*, 26(2), 193. <https://doi.org/10.23887/mi.v26i2.35688>
- Kumari, K., & Yadav, S. (2018). Linear Regression Analysis Study. *Journal of the Practice of Cardiovascular Sciences*, 4(1), 33. https://doi.org/10.4103/jpcs.jpcs_8_18
- Larose, D. T. (2006). *Data Mining Methods and Models*. Johns Wiley & Sons.
- Nurdin Mansur. (2015). Pencapaian Hasil Belajar Ditinjau Dari Sikap Belajar Mahasiswa. *Lantanida Journal*, 3(2), 107–115.