Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 10, Nomor 1, Februari 2025, pp 453-467
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

Analisis Metode *Support Vector Machine* (SVM) Dan *Random Forest* Terhadap Perilaku Dan Prestasi Siswa Berbasis Kurikulum Merdeka Di SMK Negeri 1 Stabat

Ibrahim[1], Muhammad Iqbal[2], Khairul[3]
[1,2,3]Magister Teknologi Informasi, Universitas Pembangunan Panca Budi, Medan, Indonesia
Email: ibrahim31@guru.smk.belajar.id[1], muhammadiqbal@dosen.pancabudi.ac.id[2], khairul@dosen.pancabudi.ac.id[3]

Abstract

This study analyzes the influence of student behavior on academic achievement at SMKN 1 Stabat, using a quantitative approach with simple linear regression analysis on 245 tenth-grade students. The results show a positive and significant influence, with a coefficient of determination (R^2) value of 0.272, indicating that student behavior contributes 27.2% to academic achievement. Discipline and responsibility are the most influential behavioral aspects, where students with high discipline tend to be more consistent in learning and completing assignments. The implications of this study highlight the importance of character building and positive behavior development within the school environment through collaboration among teachers, parents, and school administrators. Although not the sole determinant of success, good behavior such as discipline, responsibility, and active learning has been proven to enhance student learning outcomes. Therefore, efforts to improve the quality of education should consider not only cognitive aspects but also the development of constructive behavior through guidance programs and the strengthening of intrinsic motivation.

Keywords: *Achievement, machine learning, random forest, SVM, regression*

Abstrak

Penelitian ini menganalisis pengaruh perilaku siswa terhadap prestasi belajar di SMKN 1 Stabat, menggunakan pendekatan kuantitatif dengan analisis regresi linear sederhana pada 245 siswa kelas X. Hasilnya menunjukkan adanya pengaruh positif dan signifikan dengan nilai koefisien determinasi (R^2) sebesar 0,272, yang berarti perilaku siswa berkontribusi 27,2% terhadap prestasi belajar. Kedisiplinan dan tanggung jawab menjadi aspek perilaku yang paling berpengaruh, di mana siswa dengan kedisiplinan tinggi cenderung lebih konsisten dalam pembelajaran dan penyelesaian tugas. Implikasi penelitian ini menekankan pentingnya pembentukan karakter dan perilaku positif di lingkungan sekolah melalui kerja sama antara guru, orang tua, dan pihak sekolah. Meskipun bukan satu-satunya faktor penentu keberhasilan, perilaku yang baik seperti disiplin, tanggung jawab, dan keaktifan dalam belajar terbukti mampu meningkatkan hasil belajar siswa. Oleh karena itu, upaya peningkatan mutu pendidikan perlu memperhatikan tidak hanya aspek kognitif, tetapi juga pembinaan perilaku konstruktif melalui program bimbingan dan penguatan motivasi intrinsik

Kata kunci*: Prestasi, machine learning, random forest, svm, regresi*

1. PENDAHULUAN

Kurikulum Merdeka yang diperkenalkan oleh Kementerian Pendidikan dan Kebudayaan Indonesia bertujuan untuk memberikan fleksibilitas kepada satuan pendidikan dalam merancang proses pembelajaran yang berorientasi pada pengembangan kompetensi siswa. Implementasi kurikulum ini menekankan pada pembelajaran yang berpusat pada siswa, dengan harapan dapat meningkatkan prestasi akademik sekaligus membentuk karakter yang unggul [6]. Algoritma

machine learning yang digunakan untuk klasifikasi dan regresi. Keunggulan SVM terletak pada kemampuannya dalam menangani data berdimensi tinggi dan menghasilkan prediksi yang akurat. SVM bekerja dengan mencari hyperplane terbaik yang memisahkan data ke dalam kelas-kelas[9]. Dalam konteks pendidikan, SVM dapat digunakan untuk memprediksi perilaku siswa berdasarkan data historis. Misalnya, data absensi, nilai tugas, dan aktivitas siswa dapat dianalisis menggunakan SVM untuk menentukan apakah seorang siswa berisiko rendah atau tinggi dalam hal prestasi akademik. SVM memiliki beberapa kernel yang dapat digunakan untuk menangani data yang tidak dapat dipisahkan secara linear, seperti kernel polynomial, radial basis function (RBF), dan sigmoid. Pemilihan kernel yang tepat penting untuk memastikan model SVM dapat menangkap pola data dengan baik. Salah satu keunggulan SVM adalah kemampuannya menangani data non-linear dengan menggunakan kernel trick. Dengan pendekatan ini, data dipetakan ke ruang dimensi yang lebih tinggi sehingga lebih mudah dipisahkan. Namun, SVM juga memiliki kelemahan, seperti waktu komputasi yang tinggi untuk dataset besar dan kebutuhan akan parameter tuning yang optimal. Parameter seperti C (regularization) dan gamma (untuk kernel RBF) harus disesuaikan untuk mendapatkan performa terbaik [1]. Random Forest adalah algoritma pembelajaran mesin berbasis ensemble yang digunakan untuk klasifikasi, regresi, dan tugas prediktif lainnya. Algoritma ini dikembangkan dengan menggabungkan beberapa pohon keputusan (decision tree) untuk menghasilkan model yang lebih stabil, akurat, dan tahan terhadap overfitting [7]. Random Forest memberikan hasil akhir dengan menggabungkan prediksi dari semua pohon. Pada tugas klasifikasi, prediksi akhir didasarkan pada suara mayoritas (majority voting) dari pohon-pohon. Sementara itu, pada tugas regresi, hasil akhir adalah rata-rata prediksi dari semua pohon [8].

2. METODOLOGI PENELITIAN

2.1. Sampel

Sampel yang digunakan dalam penelitian ini diambil dari SMK Negeri 1 Stabat dalam kurun waktu semester ganjil tahun 2023 – 2024, yang menggunakan teknik pengumpulan data *Kuesioner* yang diberikan kepada guru SMK Negeri 1 Stabat, data yang dapat merupakan data yang relevan dan lengkap untuk dilakukan analisis. Data yang digunakan terdiri dari:

1. Data nilai siswa SMK Negeri 1 Stabat.
2. Data yang digunakan 80% sebagai data uji (Data Train).
3. Data yang digunakan 20% sebagai data test (Data Test).

2.2. Variabel Penelitian

Variabel – variabel yang digunakan dalam penelitian ini adalah sebagai berikut:

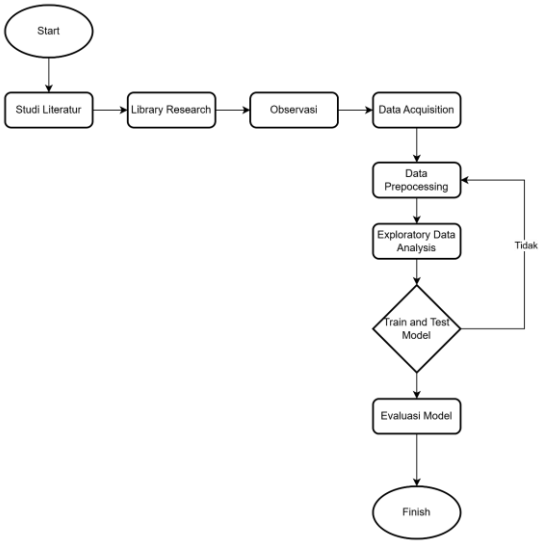
Tabel 1. Tabel Variabel Penelitian

No.	Nama	Kelas	Jurusan	Total_MB	Rata-rata_MB	Total_KD	Rata-rata_KD	Nilai_Prestasi
1	AMANDA NAYSILLA	X	AKUNTANSI	39	3,9	36	3,60	83,91
2	ANNISA	X	AKUNTANSI	42	4,2	40	4,00	83,82

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 10, Nomor 1, Februari 2025, pp 453-467
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

No.	Nama	Kelas	Jurusan	Total_MB	Rata-rata_MB	Total_KD	Rata-rata_KD	Nilai_Prestasi
	ZAHRA							
3	AULIA RIFKA PUTRI	X	AKUNTANSI	46	4,6	46	4,60	85,45
4	AZIZA ZAINAL	X	AKUNTANSI	38	3,8	43	4,30	85,27
5	BALQIS AYUNI MARHANI	X	AKUNTANSI	46	4,6	46	4,60	83,45

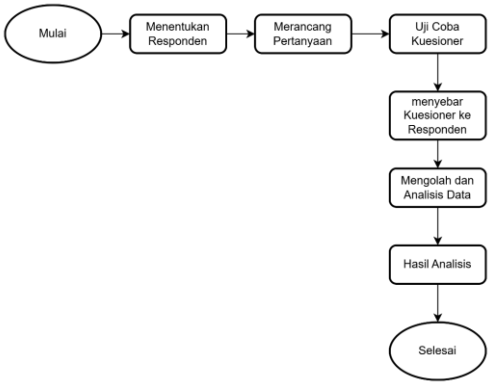
2.3. Kerangka Penelitian



Gambar 2. *flowchart* kerangka penelitian

2.4. Pengumpulan Data Primer

Data primer diperoleh langsung dari subjek penelitian, yaitu siswa kelas X SMK Negeri 1 Stabat yang telah melakukan pengisian *Kuesioner*, yang dapat dilihat pada gambar *flowcart diagram* 3 berikut:



Gambar 3. *Flowchart* Pengumpulan data Primer

2.5. Teknik Analisis data

Teknik yang digunakan dalam penelitain ini meliputi:

1. *Exploratory* Data *Analysis* (EDA)

a. Analisis deskriptif dilakukan untuk mengenali karakteristik dari data yang dimiliki.
b. Visualisasi data untuk melihat tren, pola serta anomali dalam dataset.

2. Pengujian Dataset
   a. Pembagian Dataset menjadi data *train* (data uji) dan data *test* (data latih).
   b. Menentukan variabel X(Variabel bebas) dan variabel Y(variabel terikat).
   c. Menggunakan model *SVM* dan *Random forest* dalam melakukan proses training dan testing.

3. Pemodelan Algoritma
   a. Algoritma SVM(*Support Vector Machine*) adalah sebuah model *machine learning* yang dapat melakukan analisis pengklasifikasi dan melakukan regresi terhadap dataset.

$$f(x_d) = \sum_{i=1}^{Nsv} a_{iyi}(yx_i^T x + r)^p + b \quad (1)$$

Keterangan:
Nsv = jumlah *Support Vector.*
a = *alpha*, pengali *Lanrange.*
i =1,2,3,...,Nsv.
y = label / kelas dari data.
$(yx_i^T x + r)^p + b$ = persamaan *polynominal linear.* (2)

   b. Model *Random Forest* termasuk dalam teknik *ensemble* yang menyusun sejumlah pohon keputusan, kemudian mengestimasi hasil dengan menghitung rata-rata dari prediksi tiap pohon.

$$\hat{y} \sum_{m=1}^{M} h_m(x) \quad (3)$$

Keterangan:
M = jumlah pohon keputusan.
$h_m(x)$ = hasil prediksi dari masing-masing pohon keputusan.

4. Evaluasi Model

Pada penelitian ini *Evaluation Matrix* yang digunakan adalah MSE(*Mean Squared Error*), MAE(*Mean Absolute Error*), RMSE(*Root Mean Square Error*) dan R^2(*Skor R-kuadrat*).

a) MSE (*Mean Squared Error*)
MSE adalah metrik evaluasi yang mengukur rata-rata kuadrat selisih antara nilai aktual dan nilai prediksi.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (5)$$

Keterangan:
y_i =nilai aktual.
$\hat{y}_i$ =nilai prediksi.
n =jumlah data.

```
mse = mean_squared_error(y_test, y_pred)
print(f"MSE: {mse:.2f}")
```

Gambar 4. Implementasi MSE di *google colab*

b) RMSE (*Root Mean Square Error*)

RMSE adalah matrik untuk menghitung akar kuadrat dari rata-rata selisih kuadrat antara nilai aktual dan prediksi. Lebih sensitif terhadap kesalahan besar.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (6)$$

Keterangan:

y_i = nilai aktual.

$\hat{y}_i$ = nilai prediksi.

n = Jumlah total data atau observasi.

$\sum_i^n=1$ = Penjumlahan dari semua error kuadrat.

$(y_i - \hat{y}_i)^2$ = Kuadrat dari error, untuk menghilangkan nilai negatif dan memberi penalti lebih besar untuk error besar.

```
rmse = np.sqrt(mse)
print(f"RMSE: {rmse:.2f}")
```

Gambar 5. Implementasi RMSE di *google colab*

c) MAE (*Mean Absolute Error*)

MAE adalah matrik untuk mengukur rata-rata dari selisih absolut antara nilai sebenarnya dan hasil prediksi.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (7)$$

Keterangan:

y_i = nilai aktual.

$\hat{y}_i$ = nilai prediksi.

```
mae = mean_absolute_error(y_test, y_pred)
print(f"MAE: {mae:.2f}")
```

Gambar 6. Implementasi MAE di *google colab*

d) *R-squared* (R^2)

R^2 atau koefisien determinasi adalah metrik yang menunjukkan seberapa baik model regresi menjelaskan variasi data aktual.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (8)$$

Keterangan:

SS_{res} = Jumlah kuadrat dari selisih antara nilai aktual dan prediksi.

SS_{tot} = jumlah kuadrat dari selisih rata-rata nilai aktual.

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 10, Nomor 1, Februari 2025, pp 453-467
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

```
r2 = r2_score(y_test, y_pred)
print(f"R-squared: {r2:.2f}")
```

Gambar 7. Implementasi R^2 di *google colab*

2.6 Perhitungan Manual Algoritma SVM Dan Random Forest

2.6.1. Sampling Data

Sampel data perilaku dan prestasi siswa kelas x Smk Negeri 1 Stabat

Tabel 2. Tabel samel data perilaku dan pestasi

Kelas	Jurusan	Total_MB	Rata-rata_MB	Total_KD	Rata-rata_KD	Nilai_Prestasi
X	AKUNTANSI	39	3,9	36	3,60	83,91
X	AKUNTANSI	42	4,2	40	4,00	83,82
X	AKUNTANSI	46	4,6	46	4,60	85,45
X	AKUNTANSI	38	3,8	43	4,30	85,27
X	AKUNTANSI	46	4,6	46	4,60	83,45
X	AKUNTANSI	46	4,6	45	4,50	87,18
X	AKUNTANSI	47	4,7	48	4,80	85,36

2.6.2. Perhitungan evaluation matrik SVM (*Support Vector Machine*)

Menghitung MSE, RMSE, MAE, dan R^2 untuk SVMData Sampel (Diambil dari Tabel 2):

Tabel 3. Tabel evaluation matrik SVM

Total_MB	Rata-rata_MB	Total_KD	Rata-rata_KD	Nilai_Prestasi (Aktual)
39	3.9	36	3.60	83.91
42	4.2	40	4.00	83.82
46	4.6	46	4.60	85.45

Asumsi Model Linear Sederhana:
Persamaan prediksi:
y^=w1·Total_MB+w2·Rata-rata_MB+w3·Total_KD+w4·Rata rata_KD+by^=w1·Total_MB+w2·Rata-rata_MB+w3·Total_KD+w4·Rata-rata_KD+b
Langkah Manual (Pendekatan Linear Regression):

a. Hitung Koefisien (w) dan Bias (b):
   Dengan asumsi nilai koefisien sederhana (misal: w1=0.1w1=0.1, w2=1w2=1, w3=0.05w3=0.05, w4=2w4=2, b=50b=50):
b. Prediksi untuk Sampel 1:
   y^1= (0.1·39) +(1·3.9)+(0.05·36) +(2·3.60) +50=3.9+3.9+1.8+7.2+50=66.8
c. Prediksi untuk Sampel 2:
   y^2= (0.1·42)+(1·4.2) +(0.0540) +(2·4.00) +50=4.2+4.2+2+8+50=68.4
d. Prediksi untuk Sampel 3:
   y^3= (0.1·46) +(1·4.6)+(0.05·46)+(2·4.60)+50=4.6+4.6+2.3+9.2+50=70.7

Hasil Evaluasi:

$$MSE = \frac{748.09}{3} = 249.36$$

$$MAE = \frac{47.28}{3} = 15.76$$

RMSE = √(Total Kuadrat Error / n)

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 10, Nomor 1, Februari 2025, pp 453-467
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

= √(748.09 / 3)
= √249.36
= 15.79
R^2 (Koefisien Determinasi)
Langkah-langkah:
Hitung Rata-Rata Nilai Aktual (ȳ):
ȳ = (83.91 + 83.82 + 85.45) / 3 = 84.39
Total Sum of Squares (SStot)
$(83.91 - 84.39)^2 = 0.2304$
$(83.82 - 84.39)^2 = 0.3249$
$(85.45 - 84.39)^2 = 1.1236$
SStot = 0.2304 + 0.3249 + 1.1236 = 1.6789
Residual Sum of Squares (SSres)
$SSres = \sum(y_i - \hat{y}_i)^2$= 748.09 (Sudah di Hitung Sebelumnya)
R^2=1-$\frac{SS_{res}}{SS_{tot}} = 1 - \frac{748.09}{1.6789}$=1 - 445.42=-444.42

Tabel 4. Hasil perhitungan MAE,RMSE,MSE, R^2

Sampel	Aktual (y)	Prediksi (ŷ)	Error (y - ŷ)	Kuadrat Error
1	83.91	66.8	17.11	292.75
2	83.82	68.4	15.42	237.78
3	85.45	70.7	14.75	217.56

2.6.3. Perhitungan *evaluation* matrik *Random Forest*

Menghitung MSE, RMSE, MAE, dan R^2 untuk Random Forest.
Asumsi Pembuatan 2 Pohon Keputusan Sederhana:
Jika Total_MB ≤ 40 → Prediksi = 83.91 (nilai aktual sampel 1).
Jika Total_MB > 40 → Prediksi = 85.45 (nilai aktual sampel 3).
Pohon 2:
Aturan:
Jika Rata-rata_KD ≤ 4.0 → Prediksi = 83.82 (nilai aktual sampel 2).
Jika Rata-rata_KD > 4.0 → Prediksi = 85.45 (nilai aktual sampel 3).
Hasil Evaluasi
Perhitunagn MSE = 0.674/3 = 0.225
Perhitungan MAE = 0.674/3 = 0.225
Perhitungan RMSE = RMSE = = = 0.474
Perhitungan R2 (Koefisien Determinasi)
Langkah:
1.Hitung Rata Rata Nilai Aktual ($\bar{y}$):
$\bar{y} = \frac{83.91 + 83.82 + 85.45}{3} = \frac{253.18}{3} \approx 84.39$
2.Hitung Total Sum of Squares (SStot):
SStot = $SS_{tot} = \sum(y_i - \hat{y}_i)^2$
Sampel 1: (83.91 – 84.39)2 = (-0.48)2 = 0.2304
Sampel 2: (83.82 – 84.39)2 = (-0.57)2 = 0.3249
Sampel 1: (85.45 – 84.39)2 = (1.06)2 = 1.1236

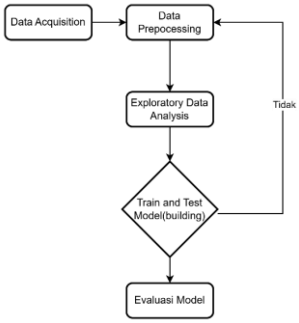
Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 10, Nomor 1, Februari 2025, pp 453-467
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

SStot = 0.2304 + 0.3249 + = 1.1236 = 1.6789
Perhitungan Residual Sum Of Squares (SSres)
$SSres = \sum(y_i - \hat{y}_i)^2$= 0.674 (sudah di hitung sebelumnya)
Perhitungan R2 :
R^2=1-$\frac{SS_{res}}{SS_{tot}} = 1 - \frac{0.674}{1.6789}$=1-0.401=0.599

3. HASIL DAN PEMBAHASAN

3.1. Tahapan Pengolahan Data

Pada penelitian ini tahapan pengolahan data dapat dilihat pada gambar 1 yang dimana proses pengolahan data dimulai dari pengumpulan data (Data *Acquisition*), pengolahan data (data *Prepocessing*), visualisasi data (*Exploratory* data analisis), pembagian data menjadi data train dan test (data splitting) dan evaluasi model.



Gambar 8. *flowcart* pengolahan data

3.2. Data Acquisition

Tahap data *acquisition* bertujuan untuk mengumpulkan informasi yang dibutuhkan dalam penelitian. Pada penelitian ini data didapat melalui data koesioner yang dilakukan di smk negeri 1 stabat, data yang didapat terdiri dari 10 kolom dan 245 baris. Seperti dapat dilihat pada gambar berikut:

	Unnamed: 0	No	Nama	Kelas	Jurusan	Total_MB	Rata_Rata_MB	Total_KD	Rata_Rata_KD	Nilai_Prestasi
0	0	1	AMANDA NAYSILLA	X	AKUNTANSI	39	3.9	36	3.6	83.91
1	1	2	ANNISA ZAHRA	X	AKUNTANSI	42	4.2	40	4.0	83.82
2	2	3	AULIA RIFKA PUTRI	X	AKUNTANSI	46	4.6	46	4.6	85.45
3	3	4	AZIZA ZAINAL	X	AKUNTANSI	38	3.8	43	4.3	85.27
4	4	5	BALQIS AYUNI MARHANI	X	AKUNTANSI	46	4.6	46	4.6	83.45
...	...	...	...	...	...	...	...	...	...	...
240	240	241	VIEA LVIERA	X	MPLB	49	4.9	46	4.6	88.09
241	241	242	VIRA SABRINA	X	MPLB	48	4.8	45	4.5	89.00
242	242	243	YAZNIL ARFAH	X	MPLB	45	4.5	46	4.6	86.73
243	243	244	YOLANDA PRATIWI	X	MPLB	47	4.7	47	4.7	87.09
244	244	245	ZAHARA	X	MPLB	41	4.1	42	4.2	87.55

245 rows × 10 columns

Gambar 9. Dataset Perilaku Dan Prestasi

3.3. Data *Prepocessing*

Tahap *prepocessing* bertujuan untuk mempersiapkan dataset terkait pembangunan model dan pelatihan model *machine learning*, tahapan ini juga bertujuan meningkatkan kualitas dari dataset agar mendapatkan hasil yang maksimal. Tahapan yang dilakukan sebagai berikut:

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 10, Nomor 1, Februari 2025, pp 453-467
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

a) Menghapus baris yang hilang

Pada proses ini dilakukan penghapusan baris yang hilang atau *Null* agar tidak mengganggu hasil validitas analisis model. Proses tersebut dapat dilihat pada Gambar 10 berikut:

```
# Menghapus Baris yang Hilang
df_cleaned = df.dropna(subset=['Nama', 'Kelas','Total_MB',
                               'Rata_Rata_MB','Total_KD','Rata_Rata_KD','Nilai_Prestasi'])
```

Gambar 10. *Source Code* Menghapus Baris Yang Hilang

b) Menghapus baris duplikat

Pada proses ini dilakukan penghapusan baris yang duplikat atau yang sama pada dataset bertujuan agar tidak ada data yang sama sehingga membuat hasil dari analisis menjadi bias. Proses tersebut dapat dilihat pada Gambar 11 berikut:

```
# Menghapus Baris yang duplikat
df = df.drop_duplicates()
df.head()
```

Gambar 11. *Source Code* menghapus baris yang duplikat

c) Menjumlahkan nilai 0

Pasa proses ini dilakukan penjumlahan nilai 0 bertujuan untuk melihat berapa banyak baris pada dataset yang memiliki nilai 0 yang selanjutnya jika terdapat banyak nilai 0 akan dilakukan sampling data tersebut namun jika tidak terdapat nilai 0 maka sampling data tidak perlu dilakukan. Proses tersebut dapat dilihat pada Gambar 12 berikut:

```
#Menjumlahkan Nilai 0
df.isna().sum()
```

	0
Unnamed: 0	0
No	0
Nama	0
Kelas	0
Jurusan	0
Total_MB	0
Rata_Rata_MB	0
Total_KD	0
Rata_Rata_KD	0
Nilai_Prestasi	0

dtype: int64

Gambar 12. *Source code* menjumlahkan nilai 0

d) Menghapus kolom yang tidak digunakan

Pada proses ini kolom yang tidak digunakan dalam proses pengolahan model machine learning akan dilakukan penghapusan. Proses tersebut dapat dilakukan dengan menggunakan source code drop kolom, seperti dapat dilihat pada Gambar 13 berikut:

```
#terdapat kolom yang tidak diperlukan dalam pelatihan model
df.drop(['Unnamed: 0', 'No','Nama','Kelas','Jurusan'], axis=1, inplace=True)
df.head()
```

	Total_MB	Rata_Rata_MB	Total_KD	Rata_Rata_KD	Nilai_Prestasi
0	39	3.9	36	3.6	83.91
1	42	4.2	40	4.0	83.82
2	46	4.6	46	4.6	85.45
3	38	3.8	43	4.3	85.27
4	46	4.6	46	4.6	83.45

Gambar 13. *Source code* menghapus kolom

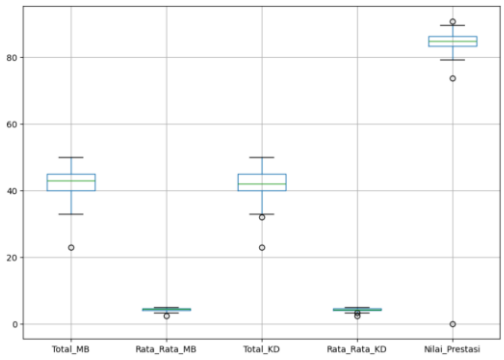
Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 10, Nomor 1, Februari 2025, pp 453-467
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

3.4. *Exploratory Data Analysis*

Pada tahapan *Exploratory Data Analysis* (EDA) dilakukan untuk mengamati serta mengidentifikasi adanya anomali dalam dataset perilaku dan prestasi siswa. Tahapan yang dilakukan sebagai berikut:

a) Melihat outlier pada varibel dengan *boxplot* diagram

boxplot diagram yang digunakan untuk melihat penyebaran data dari lima variabel, yaitu Total_MB, Rata_Rata_MB, Total_KD, Rata_Rata_KD, dan Nilai_Prestasi. Dari diagram tersebut, terlihat bahwa sebagian besar data berada dalam rentang normal, namun ada beberapa nilai yang menyimpang atau disebut outlier. Misalnya, pada variabel Total_MB dan Total_KD, terdapat beberapa data yang nilainya jauh lebih rendah dari yang lain. Hal yang sama juga terjadi pada Rata_Rata_MB dan Rata_Rata_KD meskipun skala nilainya lebih kecil. Sementara itu, pada variabel Nilai_Prestasi, nilai mayoritas cukup tinggi, tetapi ada satu nilai yang sangat rendah dibandingkan yang lainnya. Proses tersebut dapat dilihat pada gambar 7 berikut:

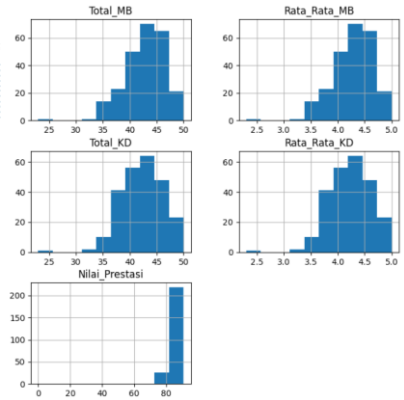


Gambar 14. *Boxplot* diagram

b) Melihat penyebaran data

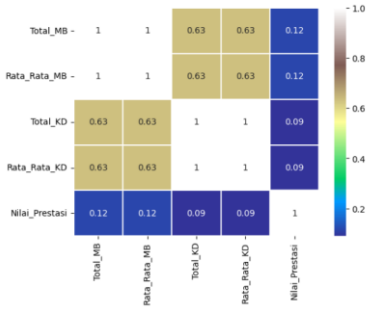
Pada proses ini menggunakan *histogram* dari lima variabel, yaitu *Total_MB, Rata_Rata_MB, Total_KD, Rata_Rata_KD*, dan Nilai_Prestasi. *Histogram* ini digunakan untuk melihat bagaimana data tersebar. Dari diagram, terlihat bahwa data *Total_MB, Rata_Rata_MB, Total_KD,* dan *Rata_Rata_KD* memiliki bentuk distribusi yang hampir normal, artinya sebagian besar nilai berada di tengah dan hanya sedikit data yang berada di nilai ekstrem. Sementara itu, untuk Nilai_Prestasi, data cenderung menumpuk di nilai tinggi, yaitu antara 80 sampai 90, dengan sedikit sekali data di nilai rendah. Ini menunjukkan bahwa mayoritas peserta memiliki nilai prestasi yang sangat baik. Proses tersebut dapat dilihat pada Gambar 15 berikut:

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 10, Nomor 1, Februari 2025, pp 453-467
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik



Gambar 15. *Histogram* penyebaran data

c) Melihat hubungan variabel dengan heatmap korelasi
Pada tahapan *heatmap* korelasi yang menunjukkan hubungan antar variabel, yaitu variabel *Total_MB* berhubungan dengan variabel *Total_KD* dan *Rata – Rata KD* dengan nilai 0.63, *Rata – Rata_MB* berkorelasi dengan *Total_KD dan Rata – Rata KD* dengan nilai 0.63 dan variabel Nilai_Prestasi memiliki hubungan korelasi dengan *Total_MB* dan *Rata – Rata_MB* dengan nilai 0.12, proses tersebut dapat dilihap pada Gambar 16 berikut:



Gambar 16. Tabel *Heatmap* Variabel

3.5. Model Building

Pada tahapan ini dataset yang telah melewati tahapan *prapocessing* memasuki tahapan model *splitting* bertujuan untuk mengolah dataset agar mendapatkan hasil analisis sehingga dapat diketahui peforma dari model *SVM* dan *Random Forest* terhadap dataset perilaku dan prestasi siswa Smk Negeri 1 Stabat. Tahapan yang dilakukan sebagai berikut:

a) Membuat *Label Encoding*
Pada tahapan *Label Encoding* pada kolom *'Total_MB' dan 'Total_KD'* menggunakan *LabelEncoder* dari *library scikit-learn*. Proses ini mengubah data kategorikal pada kolom tersebut menjadi representasi numerik, dimana setiap kategori unik dipetakan ke angka yang berbeda. Hal ini dilakukan agar data dapat diproses oleh model machine learning yang umumnya bekerja lebih baik dengan data numerik. Proses tersebut dapat dilihat pada Gambar 17 berikut:

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 10, Nomor 1, Februari 2025, pp 453-467
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

```
# Label encoding
le_Total_MB = LabelEncoder()
le_Total_KD = LabelEncoder()
df['Total_MB'] = le_Total_MB.fit_transform(df['Total_MB'])
df['Total_KD'] = le_Total_KD.fit_transform(df['Total_KD'])
```

Gambar 17 *Source code label encoding*

b) Membuat variabel X dan Y
Pada tahapan ini variabel X merupakan variabel bebas meliputi kolom *Total_MB, Rata_Rata_MB*, *Total_KD* dan *Rata_Rata_KD*, sedangkan variabel Y yang merupakan variabel terikat adalah *Nilai_Prestasi*. Proses tersebut dapat dilihat pada Gambar 18 berikut:

```
# Feature and target
X = df.drop(columns=["Nilai_Prestasi"])
y = df["Nilai_Prestasi"]
```

Gambar 18. *Source Code* pembagian variabel X dan Y

c) Splitting data
Pada tahapan ini dataset dilakukan pembagian menjadi data train dan data test, pada penelitian ini data train sebanyak 80% dan data test sebanyak 20% dengan *random state*, proses ini dapat dilihat pada Gambar 19 berikut:

```
# Split data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Gambar 19. *Splitting* data

d) Training *Svm* dan *Random Forest*
Pada tahapan ini data dilakukan training model dengan menggunakan *libaray sklearn.metrics* MAE,MSE, RMSE dan R^2, untuk mendapatkan hasil prediksi. Proses tersebut dapat dilihat pada Gambar 20 berikut:

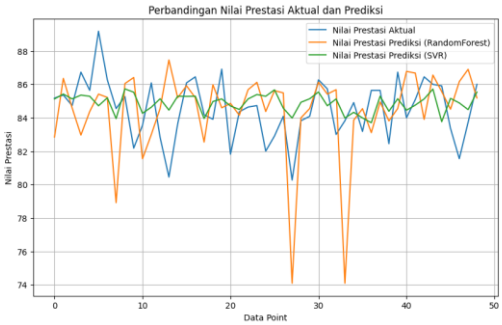
```
# Train model
#import library algoritma
from sklearn import svm
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score # Import appropriate metrics

svmr =svm.SVR()
svmr.fit(X_train,y_train)

y_pred_svmr = svmr.predict(X_test)

model = RandomForestRegressor(random_state=42)
model.fit(X_train, y_train)

# Predict
y_pred = model.predict(X_test)
```

Gambar 20. *Training* Model *Svm* Dan *Random Forest*

3.6. Model Evaluasi

Pada tahapan ini merupakan untuk mengetahui hasil dari train dan test dari kedua algoritma yaitu SVM dan random forest dengan menggunakan fitur MAE,MSE, RMSE dan R^2, maka dari proses tersebut dapat dilihat hasil dari keduanya dengan menggunakan fitur yang sama, diketahui hasil dari model svm untuk *Mean Squared Error*(MSE): 3.09, *Root Mean Squared Error*(RMSE): 1.76, Mean Absolute Error(MAE): 1.35 dan R-squared(R^2): 0.00 dan hasil dari algoritma *Random Forest* untuk Mean Squared Error(MSE): 8.25 *Root Mean Squared*

Error(RMSE): 2.87 *Mean Absolute Error*(MAE): 2.15 dan R-squared(R^2): -1.66, yang dapat dilihat pada Gambar 21 berikut:



Gambar 21. Diagram Prediksi *Svm* Dan *Random Forest*

4. SIMPULAN

Berdasarkan hasil analisis perilaku dan prestasi siswa di SMK Negeri 1 Stabat menggunakan algoritma Support Vector Machine (SVM) dan Random Forest disimpulkan bahwa:

a) Tahapan Pengolahan Data
Proses pengolahan data dilakukan mulai dari pengumpulan data (Data *Acquisition*), pra-pemrosesan (Data *Preprocessing*), eksplorasi data (*Exploratory Data Analysis*), pembagian data (Data Splitting), pembangunan model (Model Building), hingga evaluasi model (Model *Evaluation*).

b) Kualitas Data
Setelah dilakukan proses pembersihan seperti menghapus data hilang, duplikat, dan kolom yang tidak relevan, data yang digunakan memiliki kualitas yang baik untuk proses analisis lebih lanjut.

c) Eksplorasi Data
Hasil EDA menunjukkan terdapat outlier pada beberapa variabel, distribusi data sebagian besar mendekati normal, serta terdapat korelasi moderat antara beberapa variabel, namun hubungan korelasi terhadap Nilai Prestasi tergolong lemah.

d) Perbandingan Model
1. Model SVM menghasilkan nilai Mean Squared Error (MSE) sebesar 3.09, Root Mean Squared Error (RMSE) sebesar 1.76, Mean Absolute Error (MAE) sebesar 1.35, dan R-squared (R2) sebesar 0.00.
2. Model Random Forest menghasilkan nilai Mean Squared Error (MSE) sebesar 8.25, Root Mean Squared Error (RMSE) sebesar 1.76, Mean Absolute Error (MAE) sebesar 2.15, dan R-squared (R2) sebesar -1.66.

e) Kinerja Model Dari hasil evaluasi, model SVM menunjukkan performa yang lebih baik dibandingkan Random Forest pada dataset ini, walaupun secara keseluruhan nilai R-squared (R2) sangat rendah, menunjukkan

bahwa kedua model kurang mampu menjelaskan variasi dari Nilai Prestasi siswa.

DAFTAR PUSTAKA

[1] Adolph, R. (2016). Implementasi Dan Problematika Merdeka Belajar Muhajir.

[2] Anis, Y., Listiyono, H., Wahyudi, E. N., & Mulyani, S. (2024). Pelatihan Tracer Study Siswa SMK-01 Tonjong Brebes Untuk Mengukur Lulusan Pendidikan Vokasi. 02, 93–99.

[3] Artikel, I. (2025). Perancangan Sistem Informasi Manajemen KSP Kencana Bakti untuk Meningkatkan Efisiensi Pengelolaan Data Anggota dan Simpan Pinjam. 6(1), 634–641.

[4] Artikel, R., Suhanjoyo, B. W., Suteja, B. R., & Toba, H. (2023). Deteksi Tindak Kecurangan Penjualan di Perusahaan Distribusi Menggunakan Machine Learning Fraud Detection in Sales of Distribution Companies Using Machine Learning. 9, 300–312.

[5] Deva, C., & Wahyuni, E. (n.d.). Penerapan Algoritma Supervised Machine Learing dalam Optimalisasi Model Klasifikasi Risiko Kardiovaskular Serangan Jantung. 1–19.

[6] Dwi, J., Amrullah, R., Prasetya, F. B., Rahma, A. S., Setyorini, A. D., Salsabila, A. N., Nuraisyah, V., & Jember, U. (2024). Efektivitas Peran Kurikulum Merdeka terhadap Tantangan Revolusi Industri 4 . 0 bagi Generasi Alpha. 4, 1313–1328.

[7] Farhan, N. M., & Setiaji, B. (2023). Indonesian Journal of Computer Science.IndonesianJournalofComputerScience,12(2),284–301. http://ijcs.stmikindonesia.ac.id/ijcs/index.php/ijcs/article/view/3135

[8] Fathirachman Mahing, N., Lazuardi Gunawan, A., Foresta Azhar Zen, A., Abdurrachman Bachtiar, F., & Agung Wicaksono, S. (2023). Klasifikasi Tingkat Stress dari Data Berbentuk Teks dengan Menggunakan Algoritma Support Vector Machine (SVM) dan Random Forest. Jurnal Teknologi InformasiDanIlmuKomputer,10(7),1527–1536. https://doi.org/10.25126/jtiik.1078010

[9] Gardner, C. (2020). Classifying Imbalanced Financial Fraud Data Utilizing Enhanced Random Forest Algorithm Classifying Imbalanced Financial Fraud Data Utilizing Enhanced Random Forest Algorithm.

[10] Hadi, S., Sholihah, Q., & Warsiman, W. (2022). Pembelajaran Inovatif Pendidikan Karakter Pada Mata Kuliah Bahasa Indonesia Meningkatkan Kualitas Sikap, Minat, dan Hasil Belajar Siswa. Briliant: Jurnal Riset Dan Konseptual, 7(4), 905. https://doi.org/10.28926/briliant.v7i4.1148

[11] Hidayat, I., Iqbal, M., Marlina, L., Putera, A., & Siahaan, U. (2025). Analysis Of Heart Failure Prediction. 4(1).

[12] Iqbal, M., Hamdani, Hasibuan, M. S., & Nababan, A. A. (2022). Neuro Network Techniques of Telemetry Multivariate Time Series Processing and Their Applications in Industry. Journal of Theoretical and Applied Information Technology, 100(9), 2696–2702.

[13] Iqbal, M., Nazar, N., Erliana, C. I., Irwansyah, D., Kurniasih, N., Susilo, E., Iskandar, A., Suryani, R., Supriyono, Sudarmaji, Permana, E. P., Erwinsyah, A., Sujinah, Haimah, & Sudarsana, I. K. (2019). Prototype of Learning Applications for Modern Cryptographic Techniques Using RC4 Algorithms to Support Computer Security Courses. Journal of Physics:ConferenceSeries, 1363(1). https://doi.org/10.1088/1742-6596/1363/1/012068

[14] Iqbal, M., Zarlis, M., Tulus, & Mawengkang, H. (2019). Meta-Heuristic Development in Combinatorial Optimization. Journal of Physics: Conference Series, 1255(1). https://doi.org/10.1088/1742-6596/1255/1/012091

[15] Jie, H. (2022). Student Achievement and Cognitive Behavior Prediction Using Data Mining Techniques. 4(1), 66–78.

[16] Khairul, Machrizal, R., Dimenta, R. H., Rambe, B. H., Hanum, F., & Limbong, C. H. (2021). The population dynamics of Helostoma temminckii in the swampy waters of Barumun River, South Labuhan Batu Regency, Indonesia. AACL Bioflux, 14(2), 635–642.

[17] Khairul, Wijaya, R. F., Siahaan, A. P. U., Aryza, S., Hulu, F. N., Rusiadi, Aspan, H., Putra Nasution, M. D. T., Rossanty, Y., Napitupulu, D., & Arisandi, D. (2018). Effect of matrix size in affecting noise reduction level of filtering. International Journal of EngineeringandTechnology(UAE),7(3),1272–1275. https://doi.org/10.14419/ijet.v7i3.11333.