

Penerapan Metode Support Vector Machine untuk Memprediksi Kelulusan Siswa di SMA Bina Marga

Nur Irvan Rizqi^{1*}, Prasasti Karunia Farista², Agussalim³

¹Prodi Magister Teknologi Informasi UPN “Veteran” Jawa Timur

²Prodi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jawa Timur, Surabaya,

³Departemen Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jawa Timur, Surabaya, Jawa Timur, Indonesia

Email: 1,*25066020006@student.upnjatim.ac.id, 2prasasti.karunia.fasilkom@upnjatim.ac.id, 3 agussalim.si@upnjatim.ac.id

Email Penulis Korespondensi: 125066020006@student.upnjatim.ac.id

Abstrak– Kelulusan siswa merupakan indikator penting dalam menilai keberhasilan pendidikan, yang dipengaruhi oleh berbagai faktor seperti nilai akademik, aktivitas ekstrakurikuler, kemampuan interpersonal, kondisi sosial ekonomi, serta dukungan orang tua. Prediksi kelulusan dapat membantu sekolah dalam mendeteksi siswa yang berisiko tidak lulus dan memberikan intervensi yang tepat waktu. Penelitian ini bertujuan untuk membangun model prediksi kelulusan siswa dengan menggunakan metode Support Vector Machine (SVM), yang dikenal efektif dalam memisahkan data ke dalam kategori tertentu secara optimal. Data yang digunakan mencakup aspek akademik dan non-akademik. Proses pemodelan meliputi pembersihan data, seleksi fitur, optimasi parameter, serta evaluasi performa menggunakan metrik akurasi, presisi, dan recall. Hasil menunjukkan bahwa model SVM mampu memberikan prediksi kelulusan yang akurat berdasarkan data yang didapat yakni nilai precision sebesar 97%, recall sebesar 90%, dan akurasi sebesar 97,16% menunjukkan bahwa model memiliki tingkat kinerja yang sangat baik. Dengan demikian, model ini dapat dikategorikan sebagai model dengan performa excellent dan layak digunakan sebagai alat bantu dalam menentukan kelulusan siswa secara akurat dan andal...

Kata Kunci: SVM, Prediksi, Kelulusan, Sekolah

Abstract– Student graduation is an important indicator in assessing the success of education, influenced by various factors such as academic achievement, extracurricular activities, interpersonal skills, socioeconomic conditions, and parental support. Graduation prediction can help schools identify students at risk of not graduating and provide timely interventions. This study aims to develop a student graduation prediction model using the Support Vector Machine (SVM) method, which is known to be effective in optimally classifying data into specific categories. The data used include both academic and non-academic aspects. The modeling process involves data cleaning, feature selection, parameter optimization, and performance evaluation using accuracy, precision, and recall metrics. The results show that the SVM model can accurately predict student graduation, with a precision of 97%, recall of 90%, and accuracy of 97.16%, indicating excellent model performance. Therefore, this model can be categorized as an excellent-performing model and is feasible to be used as a reliable decision-support tool in determining student graduation accurately and effectively..

Keywords: SVM, Prediction, Graduation, School

1. PENDAHULUAN

Pendidikan merupakan faktor krusial yang menentukan tingkat kemajuan suatu bangsa. Untuk mewujudkan cita-cita nasional sebagaimana tercantum dalam Undang-Undang Dasar, yaitu mencerdaskan kehidupan bangsa, diperlukan sistem pendidikan yang berkualitas. Pendidikan memiliki peran penting dalam meningkatkan mutu sumber daya manusia, karena melalui pendidikan yang baik, potensi individu dapat dikembangkan secara optimal[1]. Salah satu indikator keberhasilan proses pendidikan adalah tingkat kelulusan siswa. Namun, kelulusan siswa tidak hanya dipengaruhi oleh faktor akademik, melainkan juga oleh berbagai aspek lain. Berdasarkan hasil wawancara dengan pihak sekolah, dapat disimpulkan bahwa faktor-faktor yang memengaruhi kelulusan siswa meliputi prestasi akademik, partisipasi dalam kegiatan ekstrakurikuler, serta aspek-aspek pendukung lainnya[2].

Kelulusan siswa merupakan salah satu indikator utama dalam menilai keberhasilan suatu sekolah. Memprediksi kelulusan siswa menjadi hal yang penting karena dapat membantu pihak sekolah dalam mengidentifikasi siswa yang berisiko tidak lulus atau berpotensi putus sekolah. Dengan demikian, sekolah dapat memberikan intervensi dini untuk meningkatkan performa akademik mereka[3]. Selain itu, hasil prediksi ini juga memberikan informasi berharga bagi pihak sekolah dalam mengalokasikan sumber daya secara lebih efektif, misalnya dengan memberikan dukungan tambahan kepada siswa yang membutuhkan[4].

Hasil wawancara dengan pihak sekolah menunjukkan bahwa beberapa faktor seperti nilai akademik, tingkat keaktifan dalam kegiatan sekolah, dan kemampuan interpersonal memiliki kontribusi nyata terhadap keberhasilan siswa dalam proses pembelajaran. Temuan ini sejalan dengan penelitian Taufik (2020), yang menyatakan bahwa kegagalan siswa sering kali berkaitan dengan kurangnya pencapaian pada indikator akademik dan non-akademik. Beberapa penelitian sebelumnya telah membuktikan efektivitas metode Support Vector Machine (SVM) dalam memprediksi kelulusan siswa. Misalnya, penelitian yang dilakukan oleh[5]

berhasil menerapkan SVM untuk memprediksi kelulusan mahasiswa tingkat sarjana dengan tingkat akurasi mencapai 92,6%. Selanjutnya, [6] menggunakan SVM untuk memprediksi performa akademik siswa sekolah menengah dan memperoleh hasil yang akurat, sehingga dapat digunakan sebagai alat bantu dalam pengambilan keputusan pihak sekolah. Penelitian lain oleh [7] juga menerapkan SVM untuk memprediksi risiko siswa putus sekolah, sehingga membantu meningkatkan kinerja belajar mereka. Sementara itu, [8] menerapkan metode yang sama untuk memprediksi kelulusan mahasiswa di Universitas Yarmouk dengan mempertimbangkan faktor-faktor seperti nilai ujian masuk, jenis kelamin, dan latar belakang pendidikan. Hasil penelitian tersebut menunjukkan tingkat akurasi sebesar 91,2%, yang menegaskan kemampuan SVM dalam menghasilkan prediksi yang andal.

Berdasarkan tinjauan penelitian terdahulu, sejumlah studi telah menerapkan algoritma Support Vector Machine (SVM) untuk memprediksi performa akademik. Shahiri et al. (2015) menunjukkan bahwa SVM mampu mencapai akurasi hingga 92,6% dalam memprediksi kelulusan mahasiswa. Vaheed (2022) juga membuktikan efektivitas SVM dalam memprediksi performa akademik siswa sekolah menengah. Namun demikian, terdapat beberapa kesenjangan yang belum sepenuhnya dijawab oleh penelitian sebelumnya, antara lain minimnya penelitian yang secara khusus memfokuskan pada prediksi kelulusan siswa SMA di Indonesia dengan konteks data akademik dan non-akademik yang terintegrasi, kurangnya model prediksi yang dapat digunakan langsung oleh pihak sekolah sebagai alat bantu dalam memberikan intervensi dini terhadap siswa yang berisiko tidak lulus. Tidak banyak penelitian yang mengombinasikan eksplorasi data, analisis korelasi, dan pengujian model SVM secara menyeluruh dalam satu alur penelitian yang sistematis untuk konteks prediksi kelulusan.

Oleh karena itu, penelitian ini berkontribusi dalam mengatasi kesenjangan tersebut melalui pengembangan model prediksi kelulusan berbasis SVM dengan dataset dari SMA Bina Marga, yang mencakup faktor akademik dan non-akademik. Evaluasi model dilakukan menggunakan metrik akurasi, presisi, dan recall untuk memastikan keandalan prediksi. Dengan demikian, hasil penelitian ini diharapkan dapat menjadi referensi dan alat bantu pengambilan keputusan bagi sekolah dalam meningkatkan strategi pembelajaran dan bimbingan akademik.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian



Gambar 1. Tahapan Penelitian

Tahapan penelitian pada gambar 1 dimulai dengan pengumpulan data dari SMA Bina Marga sebagai dasar analisis. Selanjutnya dilakukan pemahaman data untuk mengenali struktur dan karakteristik setiap variabel yang digunakan. Tahap berikutnya adalah pembersihan data, yaitu menghapus data duplikat serta menangani nilai yang hilang agar data siap digunakan[9]. Setelah itu dilakukan *analisis eksploratif* (EDA) guna memahami pola dan distribusi data melalui visualisasi, diikuti dengan analisis korelasi untuk menilai hubungan antarvariabel yang memengaruhi kelulusan. Data yang telah bersih kemudian melalui proses persiapan data, meliputi normalisasi dan transformasi agar sesuai untuk pemodelan[10]. Tahap pemodelan menggunakan algoritma *Support Vector Machine* (SVM) dilakukan untuk membangun model prediksi. Model yang dihasilkan kemudian dievaluasi menggunakan confusion matrix guna menilai tingkat akurasi, presisi, dan recall[11]. Terakhir, tahap

pengujian dilakukan dengan data baru untuk memastikan model memiliki kemampuan prediksi yang baik dan dapat diandalkan.

2.2 Sumber Data dan Bahan Penunjang Penelitian

Untuk mendukung validitas dan reliabilitas penelitian, bahan-bahan penunjang data yang digunakan dalam studi ini meliputi:

1. Data Primer berupa data akademik dan non-akademik siswa SMA Bina Marga, mencakup nilai tugas, nilai UTS, nilai UAS, keaktifan ekstrakurikuler, kemampuan interpersonal, serta indikator lain yang disediakan pihak sekolah.
2. Dokumen Kurikulum dan Standar Penilaian Sekolah, yang digunakan untuk memahami konteks penilaian akademik serta kriteria kelulusan siswa.
3. Literatur Primer seperti jurnal ilmiah dan prosiding terkait penggunaan SVM dalam prediksi pendidikan, eksplorasi data, dan analisis korelasi.
4. Literatur Sekunder berupa buku ajar data mining, exploratory data analysis, dan metodologi penelitian kuantitatif sebagai dasar pengembangan metode analisis.
5. Perangkat dan Tools Pendukung, seperti Google Colab, Python (scikit-learn, pandas, numpy), dan visualisasi data untuk mendukung proses analisis dan pemodelan.

Penambahan bahan-bahan ini diperlukan untuk memperkuat landasan penelitian dan menunjukkan bahwa proses pengolahan data dilakukan berdasarkan sumber yang valid dan dapat dipertanggungjawabkan.

2.3 Data Mining

Data mining merupakan suatu metode dalam bidang ilmu komputer yang banyak digunakan untuk menggali pengetahuan dari kumpulan data yang besar. Proses ini bertujuan untuk menemukan pola-pola atau hubungan tertentu yang tersembunyi di dalam basis data, sehingga dapat menghasilkan informasi baru yang bermanfaat dalam pengambilan keputusan. Umumnya, data mining banyak diterapkan dalam bidang *machine learning* dan statistika[12]. Secara umum, data mining dapat diartikan sebagai proses analisis terhadap sekumpulan data yang kompleks untuk memperoleh wawasan atau pengetahuan yang bernilai, yang dapat membantu meningkatkan efisiensi, mengurangi biaya operasional, atau mencapai kedua tujuan tersebut sekaligus. Tahapan awal dalam data mining meliputi pembersihan data (*data cleaning*), integrasi data (*data integration*), dan seleksi data (*data selection*)[13]. Pada tahap pembersihan data, dilakukan penghapusan data yang tidak lengkap, perbaikan kesalahan, serta penghilangan noise atau outlier. Setelah itu, data dari berbagai sumber digabungkan menjadi satu himpunan data yang terintegrasi. Tahap berikutnya adalah seleksi data, yaitu memilih data yang relevan untuk digunakan dalam proses analisis selanjutnya.

2.4 Google Colab

Google Colaboratory, atau yang lebih dikenal dengan *Google Colab*, merupakan platform berbasis cloud yang disediakan oleh Google dan dapat digunakan secara gratis. Alat ini berfungsi sebagai lingkungan pemrograman interaktif yang memungkinkan pengguna menjalankan kode Python secara langsung melalui peramban web tanpa perlu melakukan instalasi perangkat lunak tambahan[14]. *Google Colab* dibangun di atas *Jupyter Notebook*, yang banyak digunakan oleh peneliti data, akademisi, dan pengembang untuk melakukan eksperimen, analisis data, serta pembelajaran mesin (*machine learning*). Keunggulan utama dari *Google Colab* meliputi akses gratis ke sumber daya komputasi seperti GPU dan TPU, integrasi penuh dengan *Google Drive*, dukungan kolaborasi secara *real-time*, ketersediaan berbagai pustaka dan paket siap pakai, serta kemudahan integrasi dengan platform *GitHub*[15].

2.5 Support Vector Machine (SVM)

Support Vector Machine (SVM) pertama kali diperkenalkan oleh Vladimir N. Vapnik dan Alexey Ya. Chervonenkis pada tahun 1963. SVM merupakan salah satu metode dalam machine learning yang menggunakan konsep ruang vektor berdimensi tinggi untuk melakukan klasifikasi data[16]. Algoritma ini bekerja berdasarkan prinsip optimisasi matematis guna menemukan fungsi linear terbaik yang mampu memisahkan data ke dalam kelas tertentu secara optimal. Secara umum, SVM termasuk dalam algoritma pembelajaran terawasi (*supervised learning*), yang bertujuan mencari hyperplane terbaik sebagai batas pemisah antara dua atau lebih kelas data dengan jarak maksimum di antara keduanya. Hyperplane berfungsi sebagai garis atau bidang pembatas yang memisahkan kelompok data, sedangkan titik-titik data yang berada paling dekat dengan hyperplane disebut sebagai support vectors[17].

Dalam konteks klasifikasi, semakin besar jarak atau margin antara hyperplane dan support vectors, maka semakin baik pula kemampuan model SVM dalam melakukan generalisasi terhadap data baru. Oleh karena itu, SVM banyak digunakan untuk berbagai aplikasi prediksi dan klasifikasi, termasuk dalam pemodelan performa atau perilaku tertentu seperti prediksi kelulusan siswa maupun estimasi waktu kedatangan pelanggan [18]. Dalam kasus data yang bersifat *non-linear*, algoritma *Support Vector Machine* (SVM) menggunakan pendekatan

yang disebut dual formulation, yang memanfaatkan konsep *kernel trick* untuk memetakan data ke ruang fitur berdimensi lebih tinggi agar dapat dipisahkan secara linear[19]. Secara matematis, bentuk dual dari fungsi objektif SVM dinyatakan sebagai:

$$\max \sum_i \alpha_i - (1/2) \sum_i \sum_j \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (1)$$

dengan batasan:

$$0 \leq \alpha_i \leq C \quad \sum_i \alpha_i y_i = 0 \quad (2)$$

Keterangan:

1. α_i merupakan Lagrange multiplier yang digunakan dalam proses optimisasi,
2. y_i adalah label kelas dari data ke- i ,
3. $K(x_i, x_j)$ adalah fungsi kernel yang merepresentasikan hasil dot product antara dua data dalam ruang fitur berdimensi tinggi.

Jenis kernel yang umum digunakan dalam SVM antara lain:

1. Linear kernel: $K(x_i, x_j) = x_i \cdot x_j$
2. Polynomial kernel: $K(x_i, x_j) = (\gamma x_i \cdot x_j + r)^d$
3. Gaussian RBF kernel: $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$

Melalui pendekatan ini, SVM berupaya menemukan hyperplane optimal yang mampu memisahkan dua kelas data dengan margin maksimum, yaitu jarak antara hyperplane dengan titik data terdekat dari masing-masing kelas. Semakin besar margin yang diperoleh, semakin baik kemampuan model dalam melakukan generalisasi terhadap data baru[20].

2.6 Pengujian Model

Teknik analisis data dalam penelitian ini menggunakan pendekatan kuantitatif, yang berlandaskan pada prinsip-prinsip matematika dan pengolahan data numerik[21]. Analisis dilakukan dengan memanfaatkan data berupa nilai rata-rata tes potensi akademik serta nilai rata-rata siswa pada semester 1 dan 2, yang kemudian diuji menggunakan algoritma Support Vector Machine (SVM) untuk membangun model prediksi kelulusan. Aturan atau pola yang dihasilkan dari proses pemodelan tersebut selanjutnya dievaluasi menggunakan metode *Confusion Matrix*. *Confusion Matrix* merupakan alat analisis yang digunakan untuk mengukur sejauh mana model klasifikasi mampu mengenali dan memprediksi data dengan benar, melalui perbandingan antara hasil prediksi model dengan data aktual yang ada[22].

3. HASIL DAN PEMBAHASAN

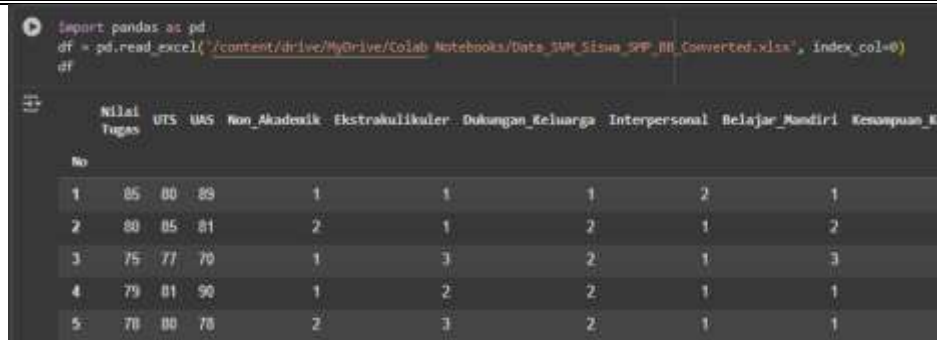
Dalam penelitian ini, data yang digunakan berasal dari SMA Bina Marga dengan total 116 data siswa. Setiap data terdiri atas 10 atribut, yang meliputi 9 atribut sebagai variabel prediktor dan 1 atribut sebagai variabel target (hasil). Penelitian ini bertujuan untuk membangun model yang dapat memprediksi status kelulusan siswa, yaitu mengklasifikasikan apakah seorang siswa dinyatakan lulus atau tidak lulus berdasarkan kombinasi faktor-faktor akademik dan non-akademik yang dimilikinya.

Nilai Tugas	UTS	UAS	Non_Akademik	Ekstrakurikuler	Dukungan_Keluarga	Interpersonal
85	80	89	Tinggi	Tinggi	Tinggi	Sedang
80	85	81	Sedang	Tinggi	Sedang	Tinggi
75	77	70	Rendah	Rendah	Sedang	Rendah
79	81	90	Rendah	Sedang	Sedang	Tinggi
78	80	78	Sedang	Rendah	Sedang	Tinggi
65	63	64	Sedang	Tinggi	Rendah	Rendah
78	77	80	Tinggi	Sedang	Tinggi	Tinggi
90	98	80	Rendah	Rendah	Rendah	Sedang

Gambar 2. Data Mentah

3.1. Penyusunan Model Data

Model data mining dalam penelitian ini dikembangkan menggunakan platform *Google Colab*. Algoritma *Support Vector Machine* (SVM) diimplementasikan melalui perangkat lunak tersebut. Proses pembentukan model dimulai dengan tahap pembacaan data (*Read Excel*), di mana data pelatihan (*training data*) disimpan dalam sebuah file bernama *Data_SVM_Siswa_SMP_BB_Converted.xlsx*. File tersebut tersimpan di *Google Drive* milik peneliti, kemudian diakses melalui *Google Colab* dengan membuat file notebook baru dan menautkan link dari *Google Drive* untuk memanggil data tersebut ke dalam lingkungan kerja Colab.



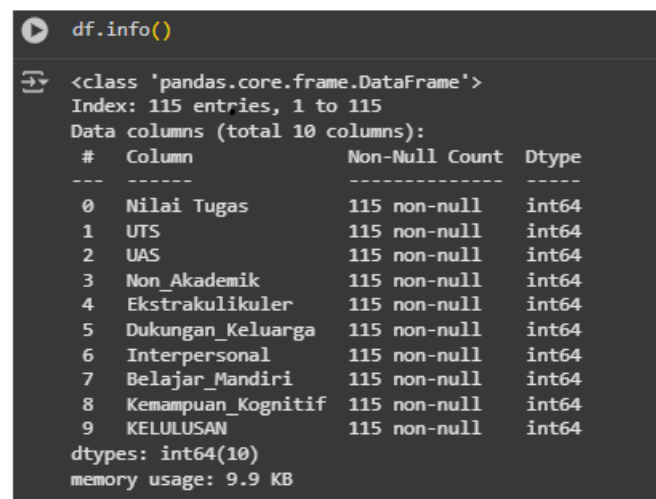
No	Nilai Tugas	UTS	UAS	Non_Akademik	Ekstrakurikuler	Dukungan_Keluarga	Interpersonal	Belajar_Mandiri	Kemampuan_Kognitif
1	85	80	89	1	1	1	2	1	
2	80	85	81	2	1	2	1	2	
3	75	77	70	1	3	2	1	3	
4	79	81	90	1	2	2	1	1	
5	78	80	78	2	3	2	1	1	

Gambar 3. Import Data dari *Google Drive* ke *Colab*

Pada tahap ini, data akan di tampilkan sesuai dengan yang di import dari *google drive*.

3.2. Data Understanding

Proses ini dilakukan untuk menampilkan dan memahami data yang terdapat dalam tabel.



```

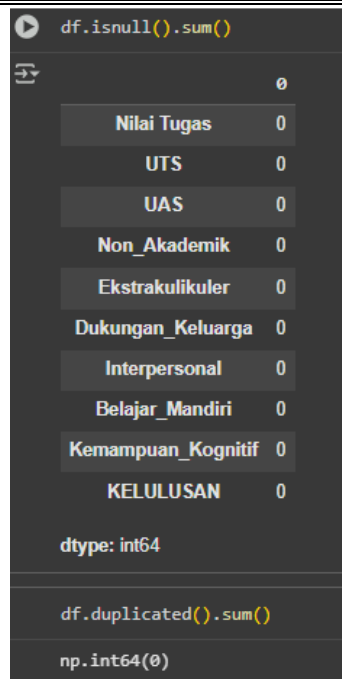
df.info()
<class 'pandas.core.frame.DataFrame'>
Index: 115 entries, 1 to 115
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   Nilai Tugas           115 non-null    int64
1   UTS                   115 non-null    int64
2   UAS                   115 non-null    int64
3   Non_Akademik          115 non-null    int64
4   Ekstrakurikuler       115 non-null    int64
5   Dukungan_Keluarga     115 non-null    int64
6   Interpersonal          115 non-null    int64
7   Belajar_Mandiri       115 non-null    int64
8   Kemampuan_Kognitif     115 non-null    int64
9   KELULUSAN             115 non-null    int64
dtypes: int64(10)
memory usage: 9.9 KB
    
```

Gambar 4. Data Understanding

Berdasarkan gambar 4, terdapat sepuluh atribut yang digunakan dalam menentukan kelulusan siswa, dan seluruh atribut tersebut memiliki tipe data berupa *integer*.

3.3. Cleaning Data

Cleaning Data merupakan proses pembersihan data dari duplikasi atau data yang sama, sehingga data yang dihasilkan bersifat unik dan tidak berulang. Pada tahap ini, dilakukan penghapusan terhadap data yang terduplikasi atau tercatat ganda.



```
df.isnull().sum()
```

	0
Nilai Tugas	0
UTS	0
UAS	0
Non_Akademik	0
Ekstrakurikuler	0
Dukungan_Keluarga	0
Interpersonal	0
Belajar_Mandiri	0
Kemampuan_Kognitif	0
KELULUSAN	0

dtype: int64

```
df.duplicated().sum()
```

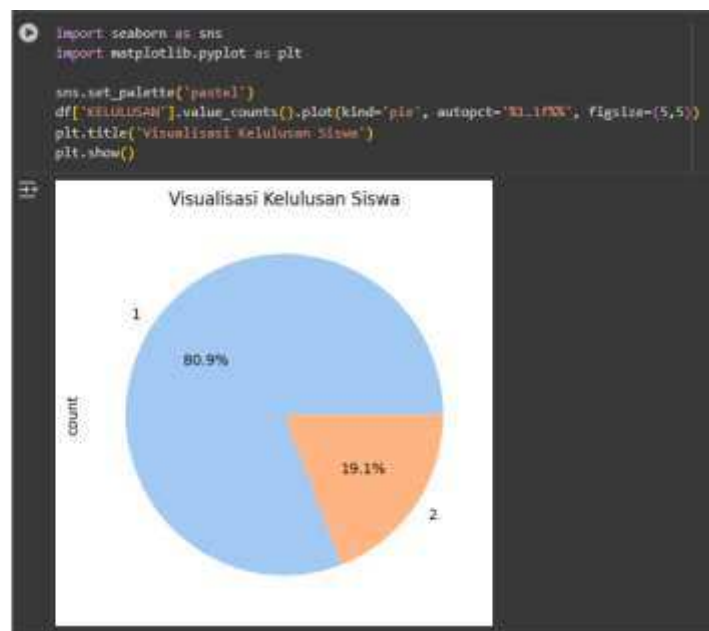
```
np.int64(0)
```

Gambar 5. Cek Duplicate Data

Berdasarkan gambar 5, dapat diketahui bahwa dari total 116 data yang tersedia, tidak terdapat data yang terduplikasi, sehingga hasil pemeriksaan menunjukkan nilai 0.

3.4. Exploratory Data Analysis (EDA)

Pada tahap ini dilakukan proses visualisasi data terkait kelulusan siswa dengan menggunakan diagram berbentuk pie chart atau diagram lingkaran.

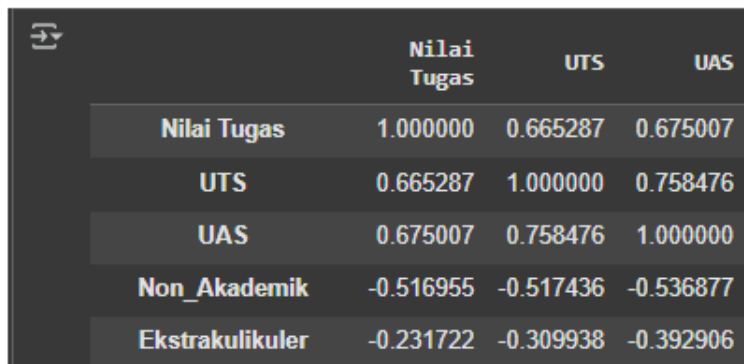


Gambar 6. Exploratory Data Analysis (EDA)

Berdasarkan gambar di atas, dapat diketahui bahwa persentase siswa yang lulus mencapai 80%, sedangkan yang tidak lulus sebesar 19,1%. Hal ini menunjukkan bahwa jumlah siswa yang lulus jauh lebih besar dibandingkan dengan siswa yang tidak lulus.

3.5. Nilai Korelasi

Nilai korelasi merupakan ukuran yang menunjukkan tingkat hubungan antara satu atribut dengan atribut lainnya. Jika nilai korelasi mendekati 1, maka hubungan antaratribut dikategorikan kuat, sedangkan jika nilainya menjauh dari 1 hingga -1, maka hubungan yang terbentuk tergolong lemah.



	Nilai Tugas	UTS	UAS
Nilai Tugas	1.000000	0.665287	0.675007
UTS	0.665287	1.000000	0.758476
UAS	0.675007	0.758476	1.000000
Non_Akademik	-0.516955	-0.517436	-0.536877
Ekstrakurikuler	-0.231722	-0.309938	-0.392906

Gambar 7. Nilai Korelasi

Berdasarkan gambar di atas, dapat disimpulkan bahwa nilai korelasi antara atribut tugas, UTS, dan UAS tergolong tinggi karena mendekati nilai 1. Namun, tingkat korelasi dengan atribut ekstrakurikuler, non-akademik, dan interpersonal relatif rendah.

3.6. Preparation Data

Preparation Data atau *Penyiapan Data* merupakan tahap yang sangat penting dalam proses data mining, yang bertujuan untuk mempersiapkan data mentah agar siap digunakan dalam proses pemodelan dan analisis. Tahapan ini mencakup berbagai kegiatan, antara lain pembersihan data (*data cleaning*), integrasi data (*data integration*), transformasi data (*data transformation*), reduksi data (*data reduction*), serta pengambilan sampel data (*data sampling*)[23].

Pada tahap pembersihan data, dilakukan penghapusan data yang tidak konsisten, perbaikan kesalahan, serta penanganan terhadap missing values. Sementara itu, integrasi data bertujuan untuk menggabungkan data dari berbagai sumber menjadi satu kesatuan dataset yang terintegrasi[24]. Tahap transformasi data meliputi proses normalisasi, pemetaan atribut numerik, perubahan format data, serta penerapan berbagai operasi matematika maupun statistik untuk menghasilkan data yang lebih siap digunakan dalam analisis lanjutan.



```

1)
X = df.drop(columns=['KELULUSAN'])
y = df['KELULUSAN']

print('X :', X.shape)
print('y :', y.shape)

X : (115, 9)
y : (115,)

2)
from sklearn.preprocessing import StandardScaler

# Select only numerical columns before scaling
X_numeric = X.select_dtypes(include=['int64', 'float64'])

scaler = StandardScaler()
scaler.fit(X_numeric)
X_scaled = scaler.transform(X_numeric)

3)
from sklearn.model_selection import train_test_split

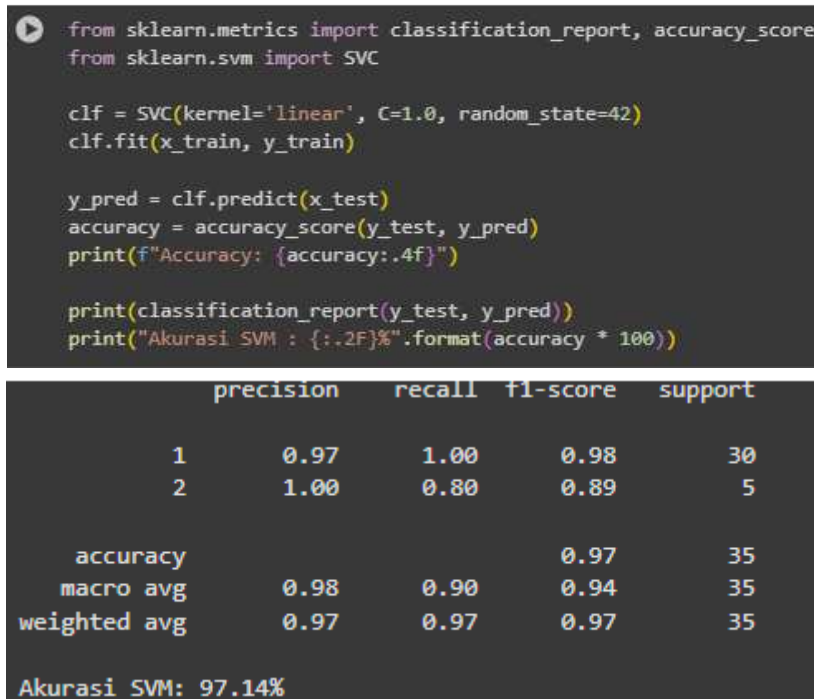
x_train, x_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2, random_state=42)

```

Gambar 8. Preparation Data

3.7. Modeling dan Evaluasi

Pada model ini digunakan *Confusion Matrix* sebagai alat untuk memvalidasi tingkat akurasi dari hasil prediksi. Nilai akurasi yang tinggi menunjukkan bahwa metode yang diterapkan memiliki kinerja yang baik, sehingga dapat direkomendasikan untuk digunakan dalam menyelesaikan permasalahan prediksi kelulusan siswa.



Gambar 9. Confusion Matrix

Berdasarkan gambar 9, diperoleh nilai precision sebesar 98%, recall sebesar 90%, dan accuracy sebesar 97,14%. Hasil tersebut menunjukkan bahwa model memiliki kinerja yang sangat baik dan dapat dikategorikan sebagai data dengan tingkat keakuratan yang *excellent*.

3.8. Testing

Testing merupakan tahap krusial dalam proses data mining yang bertujuan untuk memastikan bahwa model yang telah dibangun memiliki kemampuan generalisasi yang baik serta mampu menghasilkan prediksi yang akurat terhadap data baru yang belum pernah digunakan sebelumnya.

```

import pandas as pd
new_data = {'Nilai Tugas': [20], 'UTS' : [25], 'UAS' : [26], 'Ekstrakurikuler' : [2], 'Non_Akademik' : [1], 'Dukungan_Keluarga' : [2],
            }
new_data = pd.DataFrame(new_data)
new_data

```

	Nilai Tugas	UTS	UAS	Ekstrakurikuler	Non_Akademik	Dukungan_Keluarga	Interpersonal	Belajar_Mandiri	Kemampuan_Kognitif
0	20	25	26	2	1	2	1	2	2


```
scaled_new_data = scaler.transform(new_data)
y_pred_new = clf.predict(scaled_new_data)

print("\nHasil Prediksi Data Baru:", y_pred_new)
if y_pred_new[0] == 0:
    print("❌ Prediksi: Tidak Lulus (0)")
else:
    print("✅ Prediksi: Lulus (1)")

➡ Akurasi: 0.5
Precision: 1.0
Recall: 0.5
Confusion Matrix:
[[0 0]
 [1 1]]

Hasil Prediksi Data Baru: [0]
❌ Prediksi: Tidak Lulus (0)
```

Gambar 10. Testing Data

Berdasarkan gambar di atas, dilakukan pengujian akurasi model dengan memasukkan data baru ke dalam algoritma SVM. Data uji tersebut terdiri atas nilai tugas sebesar 20, UTS sebesar 25, UAS sebesar 26, ekstrakurikuler bernilai 2, non-akademik bernilai 1, dan seterusnya sesuai dengan data pada gambar. Hasil prediksi dari data tersebut menunjukkan nilai 0, yang menandakan bahwa siswa tersebut tidak lulus.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, model Support Vector Machine (SVM) terbukti mampu memprediksi kelulusan siswa dengan tingkat akurasi sebesar 97,14%, precision sebesar 98%, dan recall sebesar 90%. Nilai-nilai ini menunjukkan bahwa model memiliki performa yang sangat kuat dalam mengklasifikasikan siswa ke dalam kategori lulus dan tidak lulus. Analisis korelasi menunjukkan bahwa variabel akademik seperti nilai tugas, UTS, dan UAS memiliki pengaruh yang kuat terhadap prediksi kelulusan. Sebaliknya, faktor non-akademik seperti kegiatan ekstrakurikuler dan kemampuan interpersonal memberikan kontribusi yang lebih rendah. Temuan ini dapat menjadi dasar bagi pihak sekolah untuk lebih memfokuskan intervensi kepada siswa yang memiliki nilai akademik rendah. Selain memberikan hasil prediksi, model SVM yang dibangun juga dapat digunakan sebagai alat bantu bagi sekolah dalam melakukan deteksi dini terhadap siswa yang berisiko tidak lulus. Dengan demikian, penelitian ini memberikan manfaat praktis sekaligus kontribusi akademik terhadap pengembangan model prediksi pendidikan. Penelitian ini masih memiliki keterbatasan, yaitu jumlah dataset yang relatif kecil dan belum mencakup variabel kontekstual lain seperti kondisi keluarga atau motivasi belajar. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan dataset yang lebih besar serta menambahkan variabel-variabel baru untuk meningkatkan performa model.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang tulus kepada pihak pemberi dana penelitian atas dukungan dan kepercayaannya sehingga penelitian ini dapat terlaksana dengan baik. Bantuan finansial yang diberikan sangat berperan dalam mendukung seluruh proses penelitian hingga selesai. Ucapan terima kasih juga disampaikan kepada Bapak Dr. Eng. Agussalim, S.Pd., M.T., dan Ibu Karunia Farista Ananto, S.Kom., M.Kom., M.IM., atas bimbingan, arahan, serta masukan berharga yang sangat membantu dalam penyusunan dan penyempurnaan penelitian ini. Penulis menyadari masih terdapat keterbatasan dalam penelitian ini, sehingga kritik dan saran yang membangun sangat diharapkan untuk peningkatan penelitian di masa mendatang..

REFERENCES

- [1] P. Aprilianti, E. Mulazamah, N. Faizah, and D. T. Setiyoko, "PERAN PENDIDIKAN DALAM MEMBENTUK CITA-CITA WARGA NEGARA YANG BERKUALITAS," *J. ILMU Pendidik.*, vol. 6, no. 1, pp. 123–134, 2025.
- [2] A. Taufik, "Analisis Indikator kegagalan siswa dalam menempuh pendidikan di sekolah," *J. Ilm. Pendidik. Dan Pembelajaran*, vol. 4, no. 3, pp. 537–545, 2020.
- [3] A. Fatunnisa and H. Marcos, "Prediksi Kelulusan Tepat Waktu Siswa SMK Teknik Komputer Menggunakan Algoritma Random Forest," *J. Manaj. Inform. JAMIKA*, vol. 14, no. 1, pp. 101–111, 2024.
- [4] N. Nikmah, D. S. B. Tumeko, and N. A. N. Murniati, "Dampak Perencanaan Berbasis Data terhadap Peningkatan Hasil Belajar Siswa," *Pendas J. Ilm. Pendidik. Dasar*, vol. 9, no. 04, pp. 341–353, 2024.
- [5] A. M. Shahiri, W. Husain, and N. A. Rashid, "A review on predicting student's performance using data mining techniques," *Procedia Comput. Sci.*, vol. 72, pp. 414–422, 2015.
- [6] Sk. Vaheed, R. Pratap Singh, P. Nayak, and Ch. Mallikarjuna Rao, "Student's Academic Performance Prediction Using Ensemble Methods Through Educational Data Mining," in *Smart Intelligent Computing and Applications, Volume 1*, vol. 282, V. Bhateja, S. C. Satapathy, C. M. Travieso-Gonzalez, and T. Adilakshmi, Eds., in Smart Innovation, Systems and Technologies, vol. 282., Singapore: Springer Nature Singapore, 2022, pp. 215–224. doi: 10.1007/978-981-16-9669-5_20.
- [7] M. M. E. Khoudier *et al.*, "Prediction of student performance using machine learning techniques," in *2023 5th Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, IEEE, 2023, pp. 333–338. Accessed: Oct. 21, 2025. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/10296766/>
- [8] G. Mestres *et al.*, "Vascular access surgery can be safely performed in an ambulatory setting," *J. Vasc. Access*, vol. 20, no. 2, pp. 195–201, Mar. 2019, doi: 10.1177/1129729818794356.
- [9] R. Rustiyana *et al.*, *Data Mining: Algoritma dan Penerapannya*. PT. Sonpedia Publishing Indonesia, 2025. Accessed: Oct. 21, 2025. [Online]. Available:
- [10] F. Elfaladonna, I. G. T. Isa, D. Sartika, and A. M. Putra, *Buku Ajar Dasar Exploratory Data Analysis (EDA)*. Penerbit NEM, 2024. Accessed: Oct. 21, 2025. [Online]. Available:
- [11] G. Gusrianty, F. Fenly, D. Jollyta, E. Erlin, R. N. Putri, and D. Oktariana, "Penerapan Linear Discriminant Analysis Untuk Meningkatkan Kinerja Algoritma Support Vector Machine," *J. Inform. J. Pengemb. IT*, vol. 10, no. 4, pp. 895–906, 2025.
- [12] R. F. Putra *et al.*, *Data Mining: Algoritma dan Penerapannya*. PT. Sonpedia Publishing Indonesia, 2023. Accessed: Oct. 21, 2025. [Online]. Available:
- [13] Y. Asriningtias and R. Mardhiyah, "Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa," *J. Inform.*, vol. 8, no. 1, pp. 837–848, 2014.
- [14] F. Wilyani, Q. N. Arif, and F. Aslimar, "Pengenalan Dasar Pemrograman Python Dengan Google Colaboratory," *J. Pelayanan Dan Pengabd. Masy. Indones.*, vol. 3, no. 1, pp. 08–14, 2024.
- [15] N. S. N. Az-zahrani, H. K. A. Elo, F. Salim, A.-Z. A. Ramadhani, C. Meysyanti, and L. N. A. Purwantiningsih, *Python untuk Analisis Data*. SIEGA Publisher, 2025. Accessed: Oct. 21, 2025. [Online]. Available:
- [16] Y. Kristian, "Analisa Citra Wajah Bayi Untuk Deteksi Nyeri Dan Tangis Menggunakan Multi Stage Classification Dan Deep Learning," PhD Thesis, Institut Teknologi Sepuluh Nopember, 2018. Accessed: Oct. 21, 2025. [Online]. Available: <https://repository.its.ac.id/59669/2/2213301008-Disertation.pdf>
- [17] B. H. Putra, "Penerapan Algoritma Support Vector Machine Pada Ekstraksi Informasi Daftar Riwayat Hidup," PhD Thesis, Universitas Komputer Indonesia, 2019. Accessed: Oct. 21, 2025. [Online]. Available: <https://elibrary.unikom.ac.id/id/eprint/1474/>
- [18] N. Fitriyah, B. Warsito, and I. M. Di Asih, "Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (SVM)," *J. Gaussian*, vol. 9, no. 3, pp. 376–390, 2020.
- [19] R. Swastika, S. Mukodimah, F. Susanto, M. Muslihudin, and S. I. P. Adab, *IMPLEMENTASI DATA MINING (Clustering, Association, Prediction, Estimation, Classification)*. Penerbit Adab, 2023. Accessed: Oct. 21, 2025. [Online]. Available:
- [20] S. Fachrurrazi, "Penggunaan Metode Support Vector Machine (SVM) Untuk Mengklasifikasi Dan Memprediksi Angkutan Udara Jenis Penerbangan Domestik Dan Penerbangan Internasional Di Banda Aceh," PhD Thesis, Universitas Sumatera Utara, 2011. Accessed: Oct. 21, 2025. [Online]. Available: <https://repositori.usu.ac.id/handle/123456789/39194>
- [21] I. Machali, "Metode penelitian kuantitatif (panduan praktis merencanakan, melaksanakan, dan analisis dalam penelitian kuantitatif)." Fakultas Ilmu Tarbiyah dan Keguruan Universitas Islam Negeri (UIN) Sunan ..., 2021. Accessed: Oct. 21, 2025. [Online]. Available: <https://digilib.uin-suka.ac.id/id/eprint/50344/>

- [22] F. R. Valerian, M. Syarief, and D. A. Fatah, "Klasifikasi tingkat obesitas menggunakan metode gbm dan confusion matrix," *JATI J. Mhs. Tek. Inform.*, vol. 9, no. 2, pp. 2242–2249, 2025.
- [23] L. A. Supriyono *et al.*, *Buku Ajar Big Data dan Data Mining: Konsep, Metodologi, dan Aplikasi*. PT. Sonpedia Publishing Indonesia, 2025. Accessed: Oct. 21, 2025. [Online]. Available:
- [24] T. Gori, A. Sunyoto, and H. Al Fatta, "Preprocessing Data dan Klasifikasi untuk Prediksi Kinerja Akademik Siswa," *J. Teknol. Inf. Dan Ilmu Komput.*, vol. 11, no. 1, pp. 215–224, 2024.