Analisis Klasifikasi Siswa Berprestasi Pada Sekolah Menengah Atas (SMA) Negeri 14 Medan Menggunakan Metode Random

Wira Mindo Pangaribuan[1], Delima Sitanggang[2]
[1 2] Universitas Prima Idonesia

ABSTRACT

Penelitian ini bertujuan untuk mengklasifikasikan siswa berprestasi di SMA Negeri 14 Medan menggunakan metode Random Forest. Identifikasi siswa berprestasi yang akurat dan cepat menjadi tantangan penting di lingkungan pendidikan, terutama dengan meningkatnya volume data akademik dan non-akademik. Penelitian menggunakan data 288 siswa dengan 10 kategori penilaian meliputi nilai ujian, sikap, dan absensi. Metode Random Forest diimplementasikan dengan tahapan eksplorasi data, pembersihan dataset, penyeimbangan kelas menggunakan random oversampling, dan evaluasi model. Hasil penelitian menunjukkan bahwa model Random Forest mencapai akurasi 82,76% dengan nilai presisi 86% dan recall 90% untuk kelas siswa berprestasi. Matriks kebingungan mengonfirmasi model mampu mengidentifikasi 36 dari 40 siswa berprestasi dengan benar. Penelitian ini memberikan kontribusi dalam pemanfaatan machine learning untuk sistem identifikasi siswa berprestasi yang objektif serta mendukung pengambilan keputusan strategis di lingkungan sekolah.

Keywords: Klasifikasi, Siswa Berprestasi, Random Forest, Machine Learning, Random Oversampling, Sekolah.

1. INTRODUCTION

Pendidikan merupakan elemen krusial dalam pengembangan sumber daya manusia yang berkualitas. Di jenjang Sekolah Menengah Atas pemberian penghargaan kepada siswa berprestasi merupakan langkah strategis untuk mengenali dan mendorong siswa yang telah menunjukkan kemampuan akademik dan non-akademik (Bianto et al., 2020; Hidayat et al., 2024). Sektor pendidikan kini diharapkan mampu bersaing dengan memanfaatkan kemajuan teknologi informasi yang dapat mendukung peningkatan daya saing dan operasional sehari-hari serta pengambilan keputusan strategis. Pada dasarnya keberhasilan siswa dinilai berdasarkan penilaian materi teori dan praktik, serta kehadiran dan ketidakhadiran siswa di kelas (Setyani & Sipayung, 2023; Widaningsih & Yusuf, 2022).

Peningkatan volume data yang signifikan dalam setiap tahun mendorong pemanfaatan data untuk menghasilkan pengetahuan dan informasi baru melalui sistem berbasis koleksi data. Produksi dan pemanfaatan informasi serta pengetahuan ini menuntut pengembangan metode yang lebih efisien (Rikson Maruwahal Sijabat et al., 2025). Oleh karena itu perlunya pemanfaatan machine learning dengan teknik klasifikasi untuk mengidentifikasi siswa yang prestasi atau tidak di lingkungan kelas sebagai cara untuk menciptakan iklim edukatif yang kondusif bagi peningkatan prestasi siswa (Parinduri, 2025; Saputra & Nataliani, 2021).

Penelitian sebelumnya yang dilakukan oleh (Kristeni Maria et al., n.d.) Dalam melakukan perbandingan terhadap 2 metode untuk menentukan siswa berprestasi di SMAN 4 Palangka Raya, dengan membandingikan dua metode Profile Matching dan Grey Relational Analysis. Metode pertama Grey Relational Analysis mendapatkan dengan hasil nilai accuration 85,71%, precision 85%, recall 85%, specificity 86,36%, dan f-score 85,00 dan Metode kedua Profile Matching mendapat nilai accuration 50%, precision 45%, recall 47,37%,

specificity 52,17%, dan f-score 46,16%. Ditarik kesimpulan bahwa bahwa metode Grey Relational Analysis pilihan yang lebih baik untuk pengambilan keputusan dalam menentukan siswa berprestasi di SMAN 4 Palangka Raya. Kedua metode sudah baik dalam melakukan klasifikasi siswa berprestasi dengan nilai akurasi sudah diatas 80%, namun masih perlu dilakukan peningkatan akurasi dan kecepatan dalam klasifikasi menggunakan pendekatan lain seperti metode Random Forest.

Random Forest adalah algoritma machine learning berbasis ensemble yang efektif untuk klasifikasi, regresi, dan tugas prediktif, dikembangkan dengan mengombinasikan banyak pohon keputusan untuk menghasilkan model yang lebih stabil, akurat, dan tahan overfitting (Prestasi Akademik Di SMA Negeri et al., n.d.). Algoritma ini memadukan prediksi dari setiap pohon, di mana untuk klasifikasi, keputusan akhir diambil berdasarkan suara mayoritas, sedangkan untuk regresi, hasilnya adalah rata-rata prediksi dari semua pohon (Ibrahim & Iqbal, n.d.).

Pentingnya penentuan siswa berprestasi yang akurat dan cepat didalam dunia pendidikan menjadi alasan utama dilakukan penelitian ini dengan memanfaatkan machine learning dengan pendekatan metode random forest pada Sekolah Menengah Atas (SMA) Negeri 14 Kota Medan. Berdasarkan pemaparan latar belakang masalah peneliti tertarik untuk melakukan penelitian dengan judul "Analisis Klasifikasi Siswa Berprestasi Pada Sekolah Menengah Atas (SMA) Negeri 14 Medan Menggunakan Metode Random Forest"

2. METHOD

A. Jenis Penelitian

Penelitian ini menerapkan metode kuantitatif dan eksploratif-prediktif untuk mengklasifikasikan siswa berprestasi. Caranya adalah dengan menggunakan teknik data mining, khususnya algoritma Random Forest, untuk membangun dan menguji model klasifikasi berdasarkan data akademik dan non-akademik siswa

B. Metode Random Forest

Algoritma Random Forest terbukti kompetitif dalam berbagai aplikasi klasifikasi, termasuk pengukuran kinerja dan evaluasi data kompleks, karena kemampuannya menggabungkan dua atau lebih pohon keputusan (Decision Tree). Setiap pohon keputusan akan memberikan label untuk suatu sampel, dan label yang paling banyak dipilih dari keseluruhan pohon akan menjadi keluaran akhir (Bilqisth & Ikhsanuddin, 2025; Is et al., 2025; Mrg & Hasibuan, 2024).

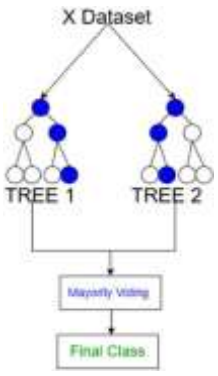


Figure 1 Ilustrasi Pohon Random Forest [13]

C. Klasifikasi

Klasifikasi dalam machine learning adalah jenis pembelajaran terawasi (supervised learning). Tujuannya adalah menciptakan model atau algoritma yang mampu memprediksi kelas atau label data baru, berdasarkan pola yang telah dipelajari dari data pelatihan (*Klasifikasi*, n.d.).

D. Kerangka Penelitian

Kerangka penelitian ini dirancang untuk memandu pelaksanaan penelitian dengan menguraikan metodologi yang akan digunakan untuk mengatasi masalah yang diteliti. Kerangka kerja ini dapat dilihat lebih lanjut pada Gambar 2 dibawah:

Gambar 2 Kerangka Penelitian

1. Identifikasi Masalah

   Pada tahap ini, peneliti menentukan dan merumuskan masalah yang ingin diteliti atau dipecahkan. Identifikasi masalah melibatkan pengamatan fenomena, peninjauan literatur awal, dan penentuan celah pengetahuan yang perlu dijawab. Masalah harus spesifik, terukur, dapat dicapai, relevan, dan terikat waktu.

2. Studi Literatur

   Setelah masalah teridentifikas peneliti akan melakukan penelusuran dan tinjauan literatur yang relevan. Tahap ini meliputi pengumpulan, evaluasi, dan sintesis informasi dari berbagai sumber (buku, jurnal ilmiah, artikel, tesis, dll.) yang berkaitan dengan topik penelitian.

3. Pengumpulan Data

   Pada tahap ini peneliti mengumpulkan data yang diperlukan untuk menjawab masalah penelitian, data yang dikumpulkan meliputi nilai akademik, catatan kehadiran, data partisipasi ekstrakurikuler, dan lain-lain. Metode pengumpulan data bisa berupa survei, observasi, wawancara, atau pengambilan data sekunder dari database sekolah.

4. Analisis Data Random Forest

   Ini adalah tahap inti di mana data yang telah dikumpulkan akan dianalisis menggunakan algoritma Random Forest. Pada tahap ini data akan diproses, melalui tahapan preprocessing (pembersihan, normalisasi), kemudian model Random Forest akan dilatih dan diuji dengan pemrograman python.

5. Hasil

   Pada tahap ini, peneliti akan menarik kesimpulan berdasarkan temuan dari analisis Random Forest. Hasil harus disajikan secara jelas dan objektif dalam bentuk tabel, grafik, atau narasi

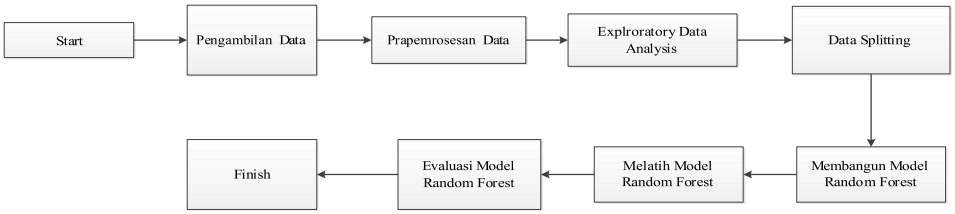
3. RESULTS AND DISCUSSION

A. Implementasi dan Analisis Data

Dengan memanfaatkan kemampuan Random Forest sebagau algoritma machine learning yang telah terbukti kuat dalam melakukan klasifikasi dan ekstraksi fitur, penelitian dilakukan untuk membangun model yang dapat melakukan klasifikasi siswa berprestasi di Sekolah Menengah Atas (SMA) Negeri 14 Medan. Penelitian ini juga bertujuan untuk memberikan wawasan tentang metode Random Forest yang optimal untuk tugas identifikasi siswa berprestasi di lingkungan pendidikan.

B. Tahapan Penolahan Data

Untuk mendapatkan hasil yang akurat dalam pengolahan data klasifikasi siswa berprestasi, penelitian ini menerapkan alur pengolahan data yang sistematis. Alur ini dirancang untuk memastikan bahwa data yang digunakan memiliki kualitas yang baik dan optimal untuk model yang akan dibangun. Gambar 3 dapat dilihat sebagai representasi alur pengolahan data dalam penelitian ini.



Gambar 3 Proses Pengolahan Data

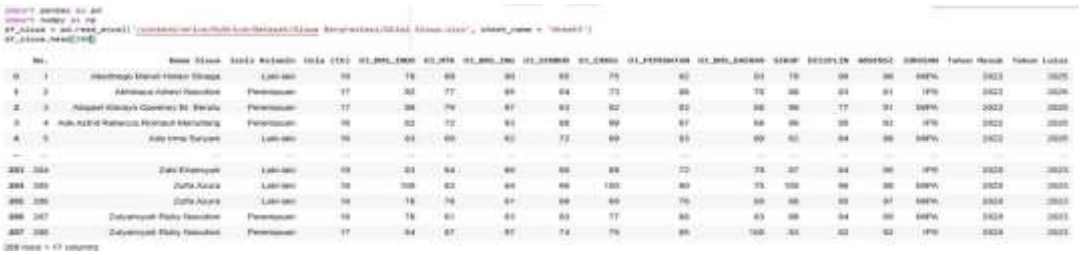
C. Pengambilan Data

Data yang digunakan dalam penelitian ini berasal dari SMA Negeri 14 Medan. Total data yang dikumpulkan adalah 1000 data nilai siswa yang mencakup 10 kategori nilai. Data tersebut terdiri dari nilai Ujian Bahasa Indonesia, Ujian Matematika, Ujian Bahasa Inggris, Ujian Seni Budaya, Ujian Ekstrakurikuler, Ujian Peminatan, Ujian Bahasa Daerah, Sikap Disiplin, dan Absensi.

D. Pemrosesan Data

Pemrosesan data akan dimulai dengan memasukkan dataset kedalam google colaboratory dilanjutkan dengan melakukan Eksplorasi Data (EDA) serta melakukan cleaning data (penghapusan column yang tidak diperlukan)

Data yang akan diolah dimasukkan kedalam google drive dan dikoneksikan langsung menggunakan Google Colaboratory dengan library Pandas :

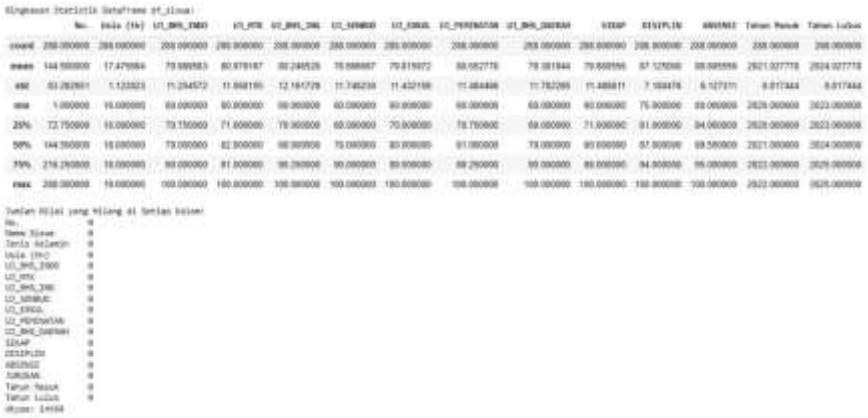


Gambar 4 Dataset Siswa

Dataset yang akan diolah dapat dilihat pada gambar 4 dengan rincian data berisi 288 baris data dan 17 kolom.

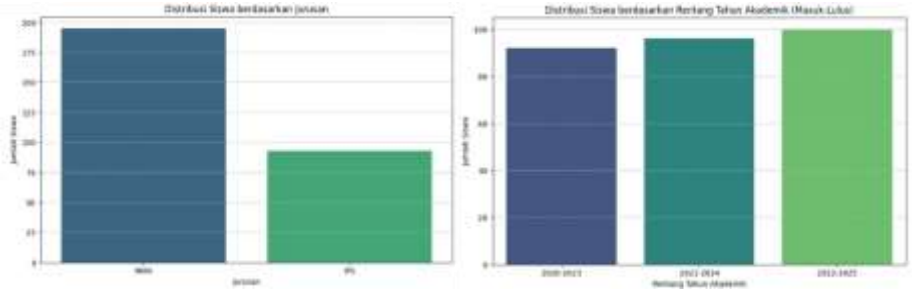
E. Eksplorasi Data (EDA)

Melakukan analisis awal untuk memahami struktur data, ringkasan statistik, distribusi nilai, dan memeriksa nilai yang hilang



Gambar 5 Analisis Deskriptif Dataset

Dataset yang digunakan sudah cukup bagus karena tidak ada data yang kosong(null) maupun berisi nilai 0

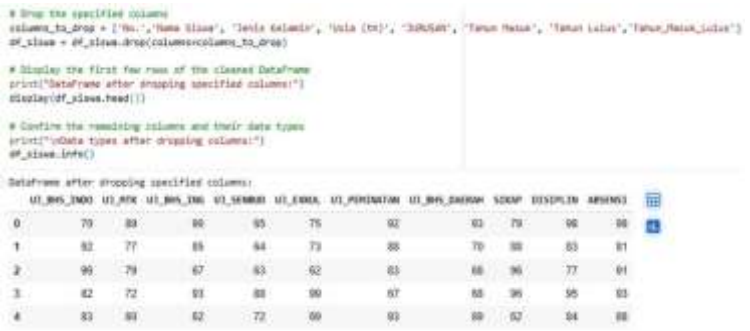


Gambar 6 Distribusi Jurusan dan Rentang Tahun

Berdasarkan gambar 6 dapat dilihat bahwa pada 3 tahun terakhir terdapat peningkatan jumlah siswa dengan total jurusan MIPA yang lebih banyak diminati oleh siswa

F. Pembersihan Dataset

Pembersihan dataset ini dilakukan dengan menghapus kolom-kolom yang dianggap tidak relevan atau tidak diperlukan sebagai fitur (variabel independen) untuk melatih model prediksi performa siswa.



Gambar 7 Penghapusan Kolom Data Yang Tidak Digunakan

Kolom yang dihapus terdiri dari nama, nomor, jenis kelamin, usia, jurusan, dan tahun masuk/lulus karena tidak digunakan langsung sebagai input model prediktif

G. Pembuatan Kolom Performa

Pemisahan fitur (X) dan variabel target (y) perlu dilakukan menentukan variabel target yang akan diprediksi. Pada tahap ini membuat kolom target Performa berdasarkan nilai siswa yang dapat dilihat pada gambar 7. Hasil dari pembuatan kolom berdasarkan nilai avarage pada seluruh siswa dapat dilihat pada gambar 7 dibawah :

Rows of X:

	U3_BHS_INDO	U3_MTK	U3_BHS_ING	U3_SENBUD	U3_EKKUL	U3_PEMINATAN	U3_BHS_DAERAH	SIKAP	DISIPLIN	ABSENSI
0	70	89	90	85	75	92	83	79	98	98
1	92	77	85	84	73	88	70	88	83	81
2	96	79	87	83	82	83	88	96	77	91
3	82	72	93	88	96	87	88	96	96	93
4	83	80	82	72	89	93	89	82	84	88
...	...	...	...	...	...	...	...	...	...	...
283	83	84	86	90	89	72	79	87	84	95
284	100	83	84	96	100	90	75	100	96	88
285	78	78	81	99	89	76	80	88	96	97
286	78	81	83	93	77	88	83	89	84	80
287	84	87	97	74	75	85	100	83	82	92

288 rows × 10 columns

Rows of y:

	Performa
0	1
1	1
2	0
3	1
4	0
...	...
283	1
284	1
285	1
286	0
287	1

288 rows × 1 columns

Gambar 7 Variabel X dan Y

Setelah dilakukan pembuatan kolom performa yang sebagai varibel Y maka perlu dilihat berapa siswa yang berprestasi dan tidak berprestasi. Jumlah siswa yang berprestasi dan tidak berprestasi dapat kita lihat pada gambar 8 dibawah ini:

Gambar 8 Jumlah Data Performa Siswa

Pada gambar 8 dapat dilihat bahwa adanya ketidak seimbangan data antara siswa berprestasi dengan yang tidak berprestasi sehingga perlu dilakukan tahap penyembiangan data untuk mengindari hasil akurasi yang Imbalance pada salah satu kategori data berprestasi maupun tidak berprestasi.

H. Penyembiangan Dataset Performa

Kondisi di mana satu kelas memiliki jumlah data yang jauh lebih banyak daripada kelas lain disebut ketidakseimbangan kelas (class imbalance). Ini bisa menjadi masalah karena model machine learning cenderung belajar lebih baik dari kelas mayoritas ('Berprestasi') dan mungkin kesulitan dalam memprediksi kelas minoritas ('Tidak Berprestasi').

Untuk mengatasi masalah ketidakseimbangan kelas pada data yang diolah dilakukan tahap random oversampling. Random Oversampling adalah teknik untuk mengatasi ketidakseimbangan kelas dengan cara menduplikasi secara acak dari kelas minoritas (kelas 'Tidak Berprestasi' atau 0) pada data pelatihan, hingga jumlahnya seimbang dengan kelas mayoritas ('Berprestasi' atau 1).

```
# Apply Random Oversampling
ros = RandomOverSampler(random_state=42)
X_train_ros, y_train_ros = ros.fit_resample(X_train, y_train)

print("Shape of training data after Random Oversampling:")
print("X_train_ros shape:", X_train_ros.shape)
print("y_train_ros shape:", y_train_ros.shape)
print("\nDistribution of target variable after Oversampling:")
print(y_train_ros.value_counts())
```


```
Jumlah Siswa Berdasarkan Performa Setelah Random Oversampling (Pelatihan):
Performa
0    160
1    160
Name: count, dtype: int64
```

Gambar 9 Proses dan Hasil Random Oversampling

Data yang telah dilakukan proses Random Oversampling menghasilkan jumlah 160 data siswa untuk kelas berprestasi dan 160 tidak berprestasi.

I. Data Spliting

Setelah dilakukan penyeimbangan kelas pada data selanjutnya adalah melakukan pembagian data latih dan data uji.

```
from sklearn.model_selection import train_test_split

# Assuming X and y are already defined from previous steps
# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42, stratify=y)

print("Shape of X_train:", X_train.shape)
print("Shape of X_test:", X_test.shape)
print("Shape of y_train:", y_train.shape)
print("Shape of y_test:", y_test.shape)
```

```
Shape of X_train: (230, 10)
Shape of X_test: (58, 10)
Shape of y_train: (230,)
Shape of y_test: (58,)
```

Gambar 10 Pembagian Data Uji dan Data Latih

Proses pembagian data dibagi menjadi 80% data latih dan data latih 20% data uji yang dapat kita lihat pada gambar 10 diatas

J. Pembangunan Model Random Forest

Model yang digunakan adalah Random Forest karena terbukti kompetitif dalam berbagai aplikasi klasifikasi, termasuk pengukuran kinerja dan evaluasi data kompleks, karena kemampuannya menggabungkan dua atau lebih pohon keputusan (Decision Tree).

```
from sklearn.ensemble import RandomForestClassifier

# Assuming X_train_ros and y_train_ros are available from the oversampling step

# Train a Random Forest model on the oversampled data
model_ros = RandomForestClassifier(random_state=42)
model_ros.fit(X_train_ros, y_train_ros)

print("Membangaun Model Random Forest.")
```

```
Membangaun Model Random Forest.
```

Gambar 11 Pembangunan Model Random Forest

Implementasi model prediktif dimulai dengan mengimpor modul Random Forest Classifier dari pustaka sklearn.ensemble, selanjutnya sebuah objek model Random Forest Classifier diinisialisasi dengan menetapkan nilai random_state sebesar 42 bertujuan memastikan reproduktifitas hasil pelatihan. Model yang telah diinisialisasi ini kemudian dilatih (fitting) menggunakan dataset pelatihan yang telah diseimbangkan melalui proses Random Oversampling yang direpresentasikan oleh variabel fitur (X_train_ros) dan variabel target (y_train_ros).

K. Melatih Model Random Forest

Setelah proses pelatihan model Random Forest dengan data yang telah diatasi ketidakseimbangan kelasnya menggunakan metode random oversampling, langkah selanjutnya adalah mengevaluasi kinerja model menggunakan data uji (test data). Evaluasi ini bertujuan untuk mengukur seberapa baik model dapat menggeneralisasi dan membuat prediksi yang akurat pada data.

```
from sklearn.metrics import accuracy_score

# Predict on the test data
y_pred_ros = model_ros.predict(X_test)

# Calculate the accuracy score
accuracy_ros = accuracy_score(y_test, y_pred_ros)*100

# Print the accuracy score
print(f"Accuracy Score Random Forest: {accuracy_ros:.4f}")
```

```
Accuracy Score Random Forest: 82.7586
```

Gambar 12 Melatih Model Random Forest

Kinerja model diukur menggunakan metrik Akurasi (Accuracy Score), yang dihitung berdasarkan perbandingan antara label kelas yang sebenarnya dari data uji (y test) dengan label kelas yang diprediksi oleh model (y pred_ros). Penjelasan untuk gambar 12 dapat jabarkan sebagai berikut :

1. Import Metrik: Baris from sklearn.metrics import accuracy_score mengimpor fungsi accuracy_score dari pustaka Scikit-learn (sklearn), yang merupakan alat standar untuk evaluasi model dalam machine learning.
2. Prediksi Data Uji: Baris y_pred_ros = model_ros.predict(X_test) menunjukkan proses di mana model yang telah dilatih (model_ros) digunakan untuk memprediksi label kelas pada fitur data uji (X_test), menghasilkan vektor prediksi (y_pred_ros).
3. Perhitungan Akurasi: Baris accuracy_ros = accuracy_score(y_test, y_pred_ros) * 100 menghitung Akurasi Model. Fungsi accuracy_score membandingkan nilai sebenarnya (y test) dan nilai prediksi (y pred_ros) dan mengembalikannya dalam bentuk proporsi. Nilai ini kemudian dikalikan dengan 100 untuk mendapatkan hasil dalam format persentase.
4. Pencetakan Hasil: Baris print(f"Accuracy Score Random Forest: {accuracy_ros:.4f}") mencetak hasil akurasi ke layar dengan format empat angka desimal.

Angka pada gambar 12 menunjukkan bahwa Model Random Forest yang telah dilatih berhasil memprediksi label kelas dengan benar pada 82.76% dari total seluruh sampel data uji. Nilai akurasi sebesar 82.76% ini mengindikasikan bahwa model memiliki kinerja klasifikasi yang cukup baik dan mampu membedakan antara kelas-kelas target pada data uji secara efektif.

L. Evaluasi Model Random Forest

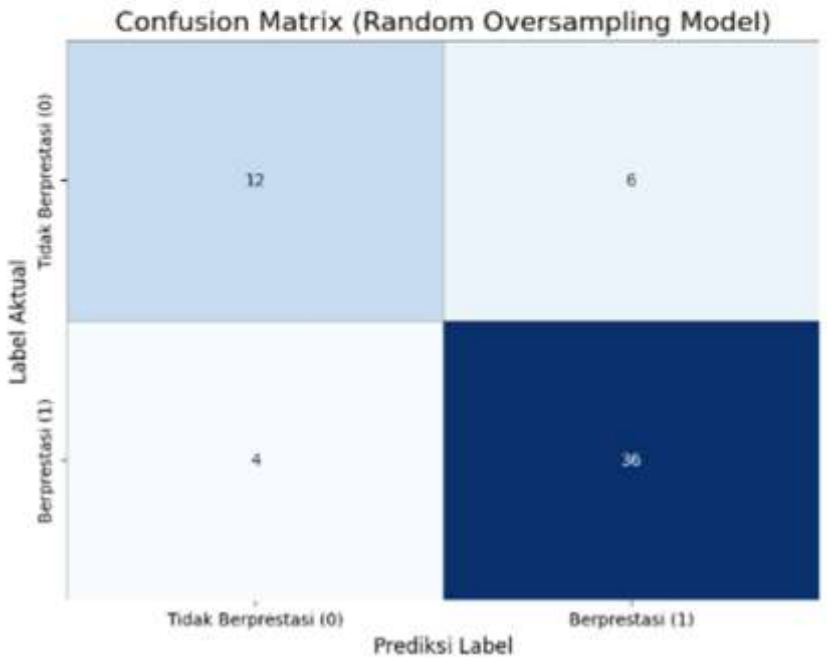
Setelah model Random Forest dilatih menggunakan data yang di-oversampling (Random Oversampling / ROS), evaluasi tidak hanya berhenti pada Akurasi saja. Untuk memahami kinerja model secara komprehensif, terutama dalam menangani ketidakseimbangan data dilakukan analisis menggunakan Matriks Kebingungan (Confusion Matrix) dan Laporan Klasifikasi (Classification Report).

```
Confusion Matrix (Random Oversampling Model):
[[12  6]
 [ 4 36]]

Classification Report (Random Oversampling Model):
                  precision    recall  f1-score   support

Tidak Berprestasi       0.75      0.67      0.71        18
      Berprestasi       0.86      0.90      0.88        40

         accuracy                           0.83        58
        macro avg       0.80      0.78      0.79        58
     weighted avg       0.82      0.83      0.82        58
```



Gambar 13 Confussion Matrix Model Random Forest

Matriks Kebingungan (Confusion Matrix)

Matriks kebingungan adalah alat visual yang penting untuk mengukur performa algoritma klasifikasi. Matriks ini menyajikan jumlah kasus prediksi yang benar dan salah pada setiap kelas. Hasil Matriks Kebingungan untuk Model Random Forest (ROS) disajikan sebagai berikut:

Berdasarkan gambar 3.12 di atas, Matriks Kebingungan menunjukkan rincian prediksi sebagai berikut:

1. True Negative (TN) - Diprediksi 'Tidak Berprestasi' dan Aktualnya 'Tidak Berprestasi': 12 sampel (Kiri Atas). Ini adalah jumlah mahasiswa yang benar diprediksi tidak berprestasi (Kelas 0).
2. False Positive (FP) - Diprediksi 'Berprestasi' tetapi Aktualnya 'Tidak Berprestasi': sampel (Kanan Atas). Ini adalah Kesalahan Tipe I, di mana mahasiswa yang sebenarnya tidak berprestasi keliru diprediksi sebagai berprestasi (Kelas 1).
3. False Negative (FN) - Diprediksi 'Tidak Berprestasi' tetapi Aktualnya 'Berprestasi': sampel (Kiri Bawah). Ini adalah Kesalahan Tipe II, di mana mahasiswa yang sebenarnya berprestasi keliru diprediksi sebagai tidak berprestasi (Kelas 0).

4. True Positive (TP) - Diprediksi 'Berprestasi' dan Aktualnya 'Berprestasi': 36 sampel (Kanan Bawah). Ini adalah jumlah mahasiswa yang benar diprediksi berprestasi (Kelas 1).

Jumlah total data uji (support) adalah 12+6+4+36=58 sampel. Akurasi global model (82.76% dari hasil sebelumnya) diperoleh dari 5812+36×100=82.76%.

Laporan Klasifikasi (Classification Report)

Selain Matriks Kebingungan, metrik kinerja yang lebih rinci disajikan dalam Laporan Klasifikasi. Laporan ini memberikan nilai Presisi (Precision), Rekall (Recall), dan F1-Score untuk masing-masing kelas, yang sangat penting untuk mengevaluasi dampak dari teknik oversampling.

Interpretasi Metrik Kunci:

1. Presisi (Precision): Menunjukkan keakuratan prediksi positif.
    a. Kelas 'Tidak Berprestasi' (0): Presisi 0.75 berarti dari semua sampel yang diprediksi tidak berprestasi, 75% di antaranya benar-benar tidak berprestasi.
    b. Kelas 'Berprestasi' (1): Presisi 0.86 berarti dari semua sampel yang diprediksi berprestasi, 86% di antaranya benar-benar berprestasi.
2. Rekall (Recall) / Sensitivitas: Menunjukkan kemampuan model untuk menemukan semua kasus positif yang relevan.
    a. Kelas 'Tidak Berprestasi' (0): Rekall 0.67 berarti model hanya berhasil mengidentifikasi 67% dari total mahasiswa yang sebenarnya tidak berprestasi (dihitung dari 12/(12+6)).
    b. Kelas 'Berprestasi' (1): Rekall 0.90 berarti model berhasil mengidentifikasi 90% dari total mahasiswa yang sebenarnya berprestasi (dihitung dari 36/(36+4)).
3. F1-Score: Rata-rata harmonik antara Presisi dan Rekall. Nilai ini adalah metrik terbaik untuk imbalanced data karena menyeimbangkan kedua metrik.
    a. Kelas 'Tidak Berprestasi' (0): 0.71
    b. Kelas 'Berprestasi' (1): 0.88

4. CONCLUSION

Secara keseluruhan model Random Forest yang dikombinasikan dengan Random Oversampling menunjukkan akurasi tinggi (82.76%). Model sangat kuat dalam mengidentifikasi mahasiswa Berprestasi (Kelas 1), terbukti dari Rekall (0.90) dan F1-Score (0.88) yang tinggi. Penerapan Random Oversampling berhasil meningkatkan kemampuan model untuk mendeteksi kedua kelas, sebagaimana dicerminkan oleh rata-rata F1-Score (Macro Avg) sebesar 0.79, menunjukkan bahwa model tidak sepenuhnya bias ke kelas mayoritas. Meskipun Random Oversampling (ROS) berhasil meningkatkan akurasi dan recall pada kelas mayoritas (Berprestasi), kinerja pada kelas minoritas (Tidak Berprestasi) masih menunjukkan recall yang relatif rendah (0.67).

5. REFERENCES

Bianto, M. A., Kusrini, K., & Sudarmawan, S. (2020). Perancangan Sistem Klasifikasi Penyakit Jantung Mengunakan Naïve Bayes. *Creative Information Technology Journal*, *6*(1), 75. https://doi.org/10.24076/citec.2019v6i1.231

Bilqisth, S. C., & Ikhsanuddin, R. M. (2025). Analisis Perbandingan Akurasi Klasifikasi Kepuasan Siswa Terhadap Kinerja Guru. *Jurnal Teknik Elektro Dan Informatika*, *20*, 37–44.

Hidayat, R. N., Santoso, B., & Sumirat, L. P. (2024). Sistem Pendukung Keputusan Pemilihan Siswa Berprestasi Menggunakan Metode Simple Additive Weighting dan Weighted Product. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *5*(1), 379–390. https://doi.org/10.57152/malcom.v5i1.1787

Ibrahim, S., & Iqbal, M. (n.d.). Analisis Metode Support Vector Machine (SVM) Dan Random Forest Terhadap Perilaku Dan Prestasi Siswa Berbasis Kurikulum Merdeka Di SMK Negeri. *Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)*, *10*, 453–467. https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

Is, C. E., Syafitri, D., Saputri, C., Sulistianingsih, N., & Hidjah, K. (2025). *Klasifikasi Gaya Belajar Siswa dengan Random Forest dan XGBoost Classification of Learning Styles of Junior High School Students Using Random Forest and XGBoost Algorithm*. *7*(1), 27–36. https://doi.org/10.30812/bite.v7i1.4913

Klasifikasi. (n.d.). Retrieved July 22, 2025, from https://www.kaggle.com/code/idhamananta/klasifikasi

Kristeni Maria, F., Chandra Saputra, A., Hendrik Timang, J., & Palangka Raya, K. (n.d.). Perbandingan Metode Profile Matching dan Grey Relational Analysis untuk Menentukan Siswa Berprestasi di SMA Negeri 4 Palangka Raya Berbasis Website. *JOINTECOMS (Journal of Information Technology and Computer Science) p-ISSN: 2798-284X*, *5*(1), 2798–3862. https://doi.org/10.47111/jointecoms.v5i1

Mrg, R. A., & Hasibuan, M. S. (2024). Best Student Classification using Ensemble Random Forest Method. *Sistemasi*, *13*(3), 1188. https://doi.org/10.32520/stmsi.v13i3.4101

Parinduri, S. K. (2025). Analisis Algoritma Fuzzy Mamdani Penentuan Siswa Berprestasi Pada Sdn 095127 Huta II Maligas Permai. *Jurnal Inovasi Artificial Intelligence & Komputasional Nusantara (JIKOMNUS)*, *2*(1).

Prestasi Akademik Di SMA Negeri, J., Dengan Menggunakan Algoritma Random Forest Rasiban, J., & Praja Raymond Maruli, S. (n.d.). Penerapan Data Mining Untuk Memprediksi Penerimaan Peserta Didik Baru. *INNOVATIVE: Journal Of Social Science Research*, *3*, 10065–10079.

Rikson Maruwahal Sijabat, R., Parlindungan Simanjuntak, R., & Pardingotan Sipayung, S. (2025). Perbandingan Metode SAW dan Weighted Product dalam Pemilihan Siswa Berprestasi. *Jurnal Sains Dan Teknologi*, *7*(1), 13–23. https://doi.org/10.55338/saintek.v7i1.1905

Saputra, E. A., & Nataliani, Y. (2021). Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means. *Journal of Information Systems and Informatics*, *3*(3). http://journal-isi.org/index.php/isi

Setyani, I. A., & Sipayung, Y. R. (2023). Sistem Pendukung Keputusan Menentukan Siswa Berprestasi dengan Metode SAW (Simple Addtive Weighting). *Jurnal Sistem Komputer Dan Informatika (JSON)*, *4*(4), 632. https://doi.org/10.30865/json.v4i4.6179

Widaningsih, S., & Yusuf, S. (2022). Penerapan Data Mining Untuk Memprediksi Siswa Berprestasi Dengan Menggunakan Algoritma K Nearest Neighbor. *Jurnal Teknik Informatika Dan Sistem Informasi*, *9*(3). http://jurnal.mdp.ac.id