

Pemodelan Klasifikasi Siswa Berprestasi dengan Random Forest: Studi Kasus pada Bimbingan Belajar

Sopiyan Apandi

Program Studi Teknik Informatika, Universitas Pamulang, Indonesia

Article Info

Article history:

Received Dec 22, 2024

Revised May 19, 2025

Accepted May 27, 2025

Keywords:

Data Mining

Machine Learning

Random Forest

Student Performance

ABSTRACT

Academic achievement is a key indicator for evaluating the effectiveness of the learning process. However, tutoring institutions often face challenges in systematically identifying students who require early intervention. This study aims to develop a student performance classification model based on periodic evaluation results using the Random Forest algorithm. The dataset consists of 112 student records from the 2017 to 2018 academic year, with 67 records used for training and 45 for testing. Each record includes scores in core subjects such as Mathematics, Natural Sciences, Social Sciences, and English. The Random Forest model was applied to classify students into high-performing and low-performing categories. The testing results showed that the model could identify high-performing students with a maximum confidence score of 0.933. The model achieved an accuracy of 100%, with both precision and recall also reaching 100%. These findings indicate that machine learning approaches can serve as reliable decision-support tools for tutoring institutions in monitoring and improving student performance proactively.

Copyright © 2025 Universitas Indraprasta PGRI.
All rights reserved.

Corresponding Author:

Sopiyan Apandi,
Teknik Informatika,
Universitas Pamulang,
Tangrang Selatan, Indonesia.
Email: dosen02601@unpam.ac.id

1. PENDAHULUAN

Pendidikan merupakan fondasi penting dalam membentuk individu yang mandiri, bertanggung jawab, dan produktif di tengah masyarakat. Salah satu indikator keberhasilan pendidikan adalah prestasi akademik, yang mencerminkan kemampuan siswa dalam memahami materi pelajaran, serta menjadi tolok ukur bagi kualitas pembelajaran (Putri & Syahril, 2022). Dalam upaya meningkatkan prestasi tersebut, banyak siswa mengikuti Lembaga Bimbingan Belajar (Bimbel) sebagai bentuk pendidikan nonformal yang menyediakan materi tambahan dan evaluasi berkala.

Meskipun hasil evaluasi siswa di Bimbel tercatat secara sistematis, pemanfaatannya masih terbatas pada pelaporan nilai akhir tanpa adanya analisis lebih lanjut. Padahal, data tersebut menyimpan potensi besar untuk memberikan gambaran tentang kemajuan siswa dan mendukung pengambilan keputusan yang strategis. Permasalahan utama yang diangkat dalam penelitian ini adalah belum adanya sistem berbasis analitik yang dapat mengklasifikasikan tingkat prestasi siswa berdasarkan hasil evaluasi secara otomatis dan akurat.

Penelitian sebelumnya telah mengeksplorasi metode klasifikasi seperti *Decision Tree* untuk menentukan tingkat prestasi belajar siswa di lingkungan sekolah formal (Apandi, Nurhidayatulloh, & Subariah, 2024), serta menganalisis efektivitas pembelajaran daring di Bimbel menggunakan *Naive Bayes* (Subariah, Apandi, & Nurhidayatulloh, 2023). Pendekatan lain dalam penentuan mahasiswa berprestasi telah menggunakan kombinasi *FP-Growth* dan SAW untuk meningkatkan akurasi dan efisiensi klasifikasi berbasis data (Ridwan & Gunawan, 2024). Namun, penelitian-penelitian tersebut masih terbatas pada pendekatan model tunggal dan belum menyorot penggunaan algoritma ansambel seperti *Random Forest* yang dikenal memiliki akurasi lebih

tinggi dan kemampuan menangani variabel dalam jumlah besar (Saputra, Sari, & Ningsih, 2022). Selain itu, belum banyak studi yang secara khusus mengkaji penerapan data mining terhadap data evaluasi berkala siswa di lembaga bimbingan belajar.

Oleh karena itu, penelitian ini bertujuan untuk mengklasifikasikan prestasi siswa berdasarkan data hasil evaluasi berkala di Lembaga Bimbingan Belajar menggunakan algoritma *Random Forest*. Hasil klasifikasi ini diharapkan dapat menjadi dasar bagi pengambilan keputusan yang lebih terarah dalam meningkatkan kualitas pembelajaran dan pencapaian akademik siswa

2. METODE

2.1. Data Mining

Menurut Vercellis (2009), kegiatan *data mining* merupakan proses berulang yang ditujukan untuk analisis *database* besar dengan tujuan mengekstrak informasi dan pengetahuan yang akurat dan berpotensi bermanfaat bagi *knowledge workers* yang terlibat dalam pengambilan keputusan dan pemecahan masalah. Sedangkan menurut Han, Kamber, dan Pei (2012), *data mining* adalah proses menemukan pola yang menarik dan pengetahuan dari data yang berjumlah besar. Kemudian, menurut Connolly dan Begg (2015), *data mining* adalah suatu proses ekstraksi atau penggalian data yang belum diketahui sebelumnya, namun dapat dipahami dan berguna dari *database* yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting. Menurut Santosa (2007), *Data Mining* sering disebut KDD (*Knowledge Discovery in Database*). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis, untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar.

Berdasarkan definisi yang dijelaskan oleh para ahli di atas, dapat disimpulkan bahwa *data mining* adalah mengekstrak informasi dan pengetahuan pada *database* atau data dalam jumlah besar dengan cara menemukan pola yang menarik sehingga dapat digunakan dalam pengambilan keputusan bisnis yang penting.

2.2. Klasifikasi

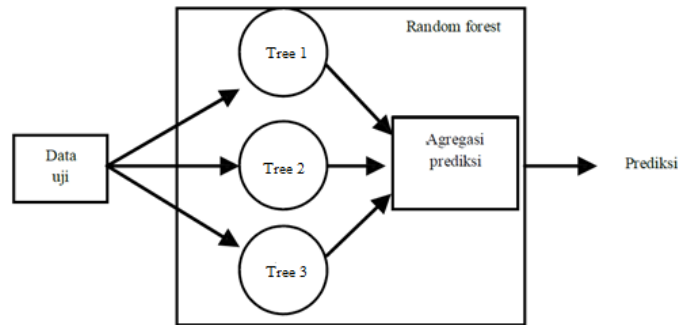
Klasifikasi adalah salah satu pembelajaran yang paling umum pada data mining. Klasifikasi didefinisikan sebagai bentuk analisis data untuk mengekstrak model yang akan digunakan untuk memprediksi label kelas. Kelas dalam klasifikasi merupakan atribut dalam satu set data yang paling unik yang merupakan variabel bebas dalam statistik. Klasifikasi data terdiri dari dua proses yaitu tahap pembelajaran dan tahap pengklasifikasian. Tahap pembelajaran merupakan tahapan dalam pembentukan model klasifikasi, sedangkan tahap pengklasifikasian merupakan tahapan penggunaan model klasifikasi untuk memprediksi label kelas dari suatu data. Contoh sederhana dari teknik data mining klasifikasi adalah pengklasifikasian hewan berdasarkan atribut jumlah kaki, habitat, dan organ pernafasannya yang akan diklasifikasikan ke dalam dua label kelas yaitu unggas dan ikan. Label kelas unggas adalah data yang memiliki jumlah kaki dua, habitatnya di darat, dan organ pernafasannya menggunakan paru-paru, sedangkan label kelas ikan adalah data yang memiliki jumlah kaki nol (tidak memiliki kaki), habitat di air, dan organ pernafasannya menggunakan insang. Banyak algoritma yang dapat digunakan dalam pengklasifikasian data, namun dalam penelitian ini hanya akan menggunakan algoritma *random forest* untuk mengklasifikasikan siswa berprestasi di Bimbel. (Iawe, Kamber, & Pei, 2012).

Pendekatan ini menggunakan teori probabilitas untuk membuat model-model klasifikasi berdasarkan kejadian-kejadian di masa lalu yang bisa untuk menempatkan suatu instans baru ke dalam kelas atau kategori yang paling mungkin. Dengan demikian sebenarnya banyak algoritma yang dapat digunakan untuk klasifikasi dan prediksi hasil belajar siswa dan pada penelitian ini akan digunakan metode klasifikasi *random forest* dengan harapan mendapatkan akurasi yang lebih tinggi dari metode sebelumnya untuk mengklasifikasikan hasil belajar siswa (Firdaus, 2016).

Penelitian ini menggunakan pendekatan kuantitatif dengan metode klasifikasi menggunakan algoritma Random Forest (RF). Tahapan dalam penelitian ini mengacu pada proses *Cross-Industry Standard Process for Data Mining (CRISP-DM)*, yang terdiri dari enam tahap: *business understanding, data understanding, data preparation, modeling, evaluation, dan deployment* (Wirth & Hipp, 2000). Dalam konteks klasifikasi mahasiswa berprestasi, pendekatan klusterisasi juga telah digunakan untuk menganalisis keberhasilan studi sejak awal masa perkuliahan, seperti ditunjukkan oleh Painem et al. (2023) yang menerapkan algoritma *K-Means* dalam mengelompokkan mahasiswa berdasarkan potensi keberhasilannya di masa awal studi.

2.3. Random Forest

Random forest adalah algoritma klasifikasi dan regresi yang merupakan bagian dari kelompok *ensemble learning*, proses prediksi *random forest* Model *base classifier* yang dipakai adalah *decision tree* sehingga membentuk *forest* atau hutan. *Random forest* dipopulerkan oleh Leo Breiman (Anugerah & Tjokorda, 2017).



Gambar 1. Proses Prediksi Random Forest

Dalam *random forest*, pemilihan atribut dalam setiap kali sebuah *node* akan dipecah dan diambil secara acak. Pertama-tama setiap *tree* diberi sampel *data training* dengan menggunakan metode *bagging*, dan tiap *tree* dibangun menggunakan metode yang sama untuk membangun CART. Perbedaan yang mencolok terdapat pada proses pemilihan *splitting criterion*. Alih-alih mempertimbangkan seluruh atribut yang ada pada data, *random forest* hanya mempertimbangkan *subset* dari keseluruhan atribut (biasanya jumlah atribut yang diseleksi ditentukan oleh *user*). Atribut yang akan diseleksi ini didapatkan secara acak. Seperti CART, pembangunan *tree* akan berhenti ketika data sudah homogen atau batas jumlah data minimum terlewati. Namun terdapat variasi *random forest* yang menentukan besar kedalaman *tree* maksimum, algoritma sederhana *random forest* dapat dilihat pada Gambar 1.

Dengan proses pemilihan atribut yang acak pada pembangunan *tree*, *random forest* dapat menangani data dengan dimensi tinggi dengan baik dibanding *model classifier* lainnya. Berbeda dengan *decision tree* biasa, *overfitting* diatasi dengan menjaga variansi model *tree* dalam *forest*. Dalam *random forest* umumnya tidak terdapat *leaf pruning* (penghilangan *node leaf*), tetapi variasi *random forest* seperti *hough forest* dalam implementasinya menerapkan *leaf pruning* untuk menghilangkan *leaf node* dengan probabilitas rendah (Anugerah & Tjokorda, 2017).

Random Forest (RF) adalah klasifikasi yang terdiri dari beberapa pohon keputusan. Setiap pohon keputusan dibangun dengan menggunakan vektor acak. Pendekatan umum yang digunakan untuk menyisipkan vektor acak dalam pembentukan pohon adalah memilih nilai F acak, seperti F atribut fitur masukan untuk dibagi pada setiap *node* di pohon keputusan yang akan dibentuk. Dengan memilih nilai acak F, seharusnya tidak harus memeriksa semua atribut yang ada, hanya melihat F dipilih atribut. Parameter digunakan untuk mengatur kekuatan RF dalam pemilihan nilai F dan jumlah *tree* yang akan dibangun. Jika nilai F terlalu kecil, pohon tersebut memiliki kecenderungan untuk memiliki korelasi yang sangat kecil dan itu berlaku sama untuk sebaliknya (Breiman, 2001)..

Dataset yang digunakan berisi informasi hasil evaluasi siswa di Lembaga Bimbingan Belajar dari tahun ajaran 2017–2018, mencakup 112 data siswa. Atribut-atribut yang dianalisis meliputi nilai rata-rata dari empat mata pelajaran utama (Matematika, Bahasa Inggris, IPA, dan IPS), serta status prestasi akhir siswa yang menjadi label klasifikasi berprestasi tinggi atau kurang.

Tabel 1. Atribut Data

No	Nama Atribut	Tipe Data	Deskripsi
1	Nilai_Matematika	Numerik	Nilai ujian matematika
2	Nilai_BahasaInggris	Numerik	Nilai ujian Bahasa Inggris
3	Nilai_IPA	Numerik	Nilai ujian IPA
4	Nilai_IPS	Numerik	Nilai ujian IPS

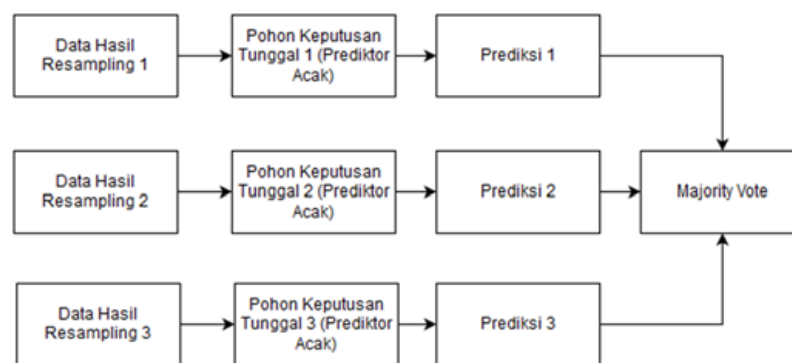
No	Nama Atribut	Tipe Data	Deskripsi
5	Status_Prestasi	Kategorikal	Label: "Berprestasi" atau "Kurang"

Data ini kemudian dibagi menjadi 67 data untuk pelatihan dan 45 data untuk pengujian. Proses *data mining* dilakukan menggunakan *software RapidMiner*, yang mendukung pemrosesan data, *training model*, dan evaluasi kinerja secara visual. Algoritma *Random Forest* yang digunakan memiliki mekanisme *bagging*, yaitu membuat beberapa *decision tree* dari subset acak data pelatihan. Pemilihan fitur dilakukan secara acak untuk mengurangi korelasi antar pohon (Breiman, 2001).

Dalam *Random Forest*, jumlah atribut acak yang digunakan untuk split ditentukan oleh:

$$F = \text{Log}_2 (M + 1)$$

Dari rumus tersebut, M adalah jumlah total fitur dalam *dataset*. Misalnya, untuk 4 atribut (Matematika, Inggris, IPA, IPS), maka $F = \text{Log}_2 (4 + 1) = 2.32$ dibulatkan ke bawah menjadi 2. Setiap pohon dibentuk dari *subset bootstrap* dari data latih dan menghasilkan keputusan melalui majority voting dari seluruh pohon.



Gambar 2. Algoritma Sederhana *Random Forest*

Random Forest (RF) adalah metode klasifikasi dan regresi berdasarkan agregasi sejumlah pohon keputusan, seperti pada Gambar 2. Secara khusus, merupakan sebuah *ensemble* (kumpulan) dari pohon yang dibangun dari kumpulan data pelatihan dan yang divalidasi secara internal untuk menghasilkan nilai respon yang dari nilai *predictor* (Maulana, 2017).

Berikut adalah Tahapan algoritma *random forest*:

1. Pilih nilai n yang menunjukkan jumlah pohon
2. Menghasilkan sampel n *bootstrap* dengan teknik *training set*.
3. Pada setiap *node* pada *tree*, pilih nilai F yang diperoleh dari persamaan.
4. Ambil set sebanyak atribut F yang akan menjadi kandidat atribut untuk membelah setiap *node*. Atribut yang menjadi simpul berikutnya ditentukan berdasarkan kriteria tertentu berdasarkan algoritma pohon keputusan yang dipilih. Selama proses pembentukan pohon, nilai F adalah konstan.
5. RF terus terbentuk tanpa pemangkasan apapun. Hal ini ditunjukkan untuk menghilangkan bias dalam persentase hasil prediksi.
6. Hasil prediksi yang diperoleh dari model frekuensi yang paling sering muncul dari masing-masing pohon keputusan di RF (Mambang dan Byna, 2017).

Tahapan akhir adalah Evaluasi kinerja model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *confusion matrix*. Hasil klasifikasi dari model *Random Forest* menunjukkan seberapa akurat dan presisi hasil perhitungan, yang menunjukkan bahwa model dapat mengklasifikasikan siswa dengan sangat baik.

Gambar 3. *flow chart* tahapan penelitian

2.4. Rapid Miner

Rapid Miner adalah perangkat lunak yang bersifat terbuka (*open source*). Rapid Miner merupakan solusi untuk melakukan suatu analisis terhadap *data mining*, *text mining* dan analisis prediksi. Rapid Miner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. Rapid Miner memiliki kurang lebih 500 operator *data mining*, termasuk operator untuk *input*, *output*, *data preprocessing* dan visualisasi. Rapid Miner merupakan *software* yang berdiri sendiri untuk analisis data dan sebagai mesin *data mining* yang dapat diintegrasikan pada produknya sendiri. Rapid Miner ditulis dengan menggunakan bahasa java sehingga dapat bekerja pada semua sistem operasi (Siregar & Puspabhuana, 2017).

3. HASIL DAN PEMBAHASAN

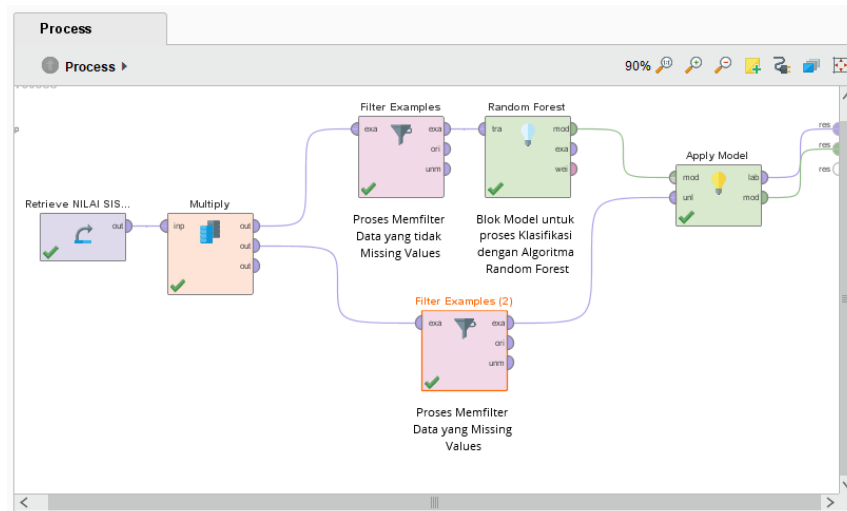
3.1. Data Penelitian

Data yang digunakan pada penelitian ini diperoleh langsung dari bagian admin Bimbel. Untuk melakukan analisis klasifikasi siswa berprestasi, penelitian ini menggunakan *history data* siswa tahun ajaran 2017/2018. Contoh data yang diperoleh dari Bimbel ditunjukkan pada Gambar 4 berikut:

NAMA	MAT	IND	ENG	IPA	IPS	JUMLAH
BAYU MALIK IBRAHIM	85,3	79,5	97,8	74,1	58,8	395,5
CHARLUNNISA PUTRI	76,9	79,5	78,9	74,1	60,9	370,3
FADILLAH SODATUN JAN	60,0	34,1	57,8	67,2	67,1	286,2
DANISA MAYLA PUTRI	74,1	76,4	78,2	60,7	53,8	343,2
EKA PRATIWI	71,8	66,9	52,7	67,8	57,4	316,6
ELFRIDA CALLYSTA S	43,6	83,4	70,9	48,2	39,6	285,7
GALLUH SEKAR AYU	83,6	95,3	83,6	83,9	75,1	421,5
GANIS AGHARANI	62,4	76,4	61,8	62,5	55,6	318,7
HAFIDHUL AZMI	74,1	76,4	63,6	71,4	60,9	346,4
KHUSNUL LATIFUNISA	53,0	62,2	54,5	57,1	57,4	284,2
KUSUMA MAHARTIKA SET	71,8	48,0	63,6	58,9	64,5	306,8

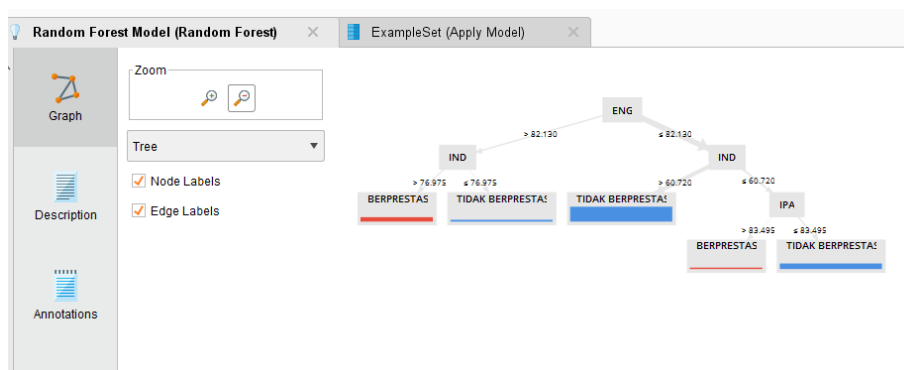
Gambar 4. Ilustrasi Data Sampel Siswa Bimbel

Proses awal yaitu menentukan *data training* 60% dan *data testing* 40%. Kemudian data tersebut diklasifikasi dengan menggunakan algoritma *random forest*. Setelah itu akan dilakukan pengujian terhadap data, langkah awal pengujian adalah mengatur operator yang ada pada Rapid Miner. Berikut Gambar 5 terlihat data siswa yang akan diproses.



Gambar 5. Penerapan Metode Terhadap *Data Testing*

Tahapan ini merupakan tahap penerapan metode *random forest* dengan melakukan *training* terlebih dahulu terhadap metode dengan *data training* untuk selanjutnya dilakukan *testing* dan menganalisis hasil dari proses yang telah dilakukan. Pada Gambar 6 dapat dilihat *result model random forest*.



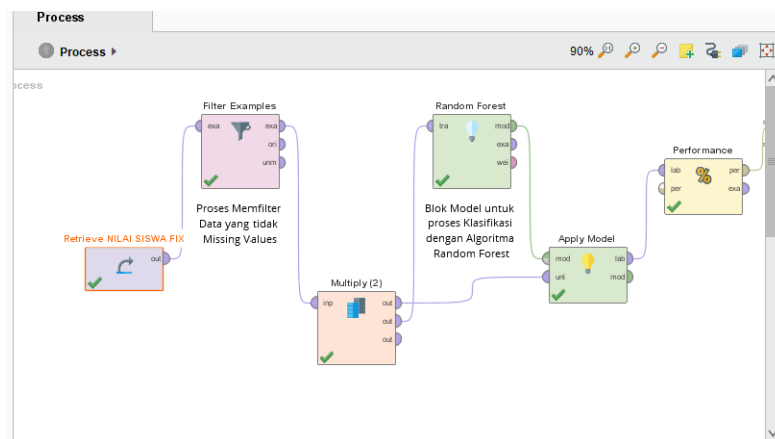
Gambar 6. *Result Model Random Forest*

Gambar tersebut adalah salah satu pohon keputusan yang dihasilkan dari penerapan model algoritma *random forest* dari data nilai siswa Bimbel berupa prediksi prestasi siswa dari pohon keputusan yang telah terbentuk. Pada Gambar 7 dapat dilihat *result apply model random forest*.

KATEGORI	prediction(KATEGORI)	confidence(TIDAK BERPRESTASI)	confidence(BERPRESTASI)	NAMA	MAT	IND
?	BERPRESTASI	0.175	0.825	BINTANG HL...	79.500	67.100
?	TIDAK BERPRESTASI	0.995	0.005	NADIN NUR ...	69.300	64.900
?	TIDAK BERPRESTASI	0.995	0.005	QUEEN AQIL...	69.300	73.700
?	TIDAK BERPRESTASI	0.995	0.005	ARUNA SALA...	78.900	63.600
?	BERPRESTASI	0.075	0.925	BAYU MALIK...	85.300	79.500
?	TIDAK BERPRESTASI	0.995	0.005	FADILLAH S...	60	34.100
?	TIDAK BERPRESTASI	0.965	0.035	HILMI NAWA...	68.500	68.100
?	TIDAK BERPRESTASI	0.995	0.005	MARYANA PU...	60	75
?	TIDAK BERPRESTASI	0.995	0.005	MUHAMAD RI...	0	75
?	TIDAK BERPRESTASI	0.995	0.005	REVA AULIA ...	70.600	79.500
?	TIDAK BERPRESTASI	0.704	0.296	SUSILAWATI ...	83.200	90.900
?	TIDAK BERPRESTASI	0.943	0.057	TRENDY FRI...	81.100	61.300
?	TIDAK BERPRESTASI	0.995	0.005	ADITYA WIJA...	0	78.700
?	TIDAK BERPRESTASI	0.993	0.007	ALBIVI NEVA ...	88.200	78.700
?	TIDAK BERPRESTASI	0.995	0.005	ALFITO DAFF...	43.600	0
?	TIDAK BERPRESTASI	0.995	0.005	BAGAS RAYS...	76.500	52.700

Gambar 7. Result Apply Model Random Forest

Hasil dari penerapan metode algoritma *random forest* salah satunya adalah prediksi prestasi siswa berdasarkan data nilai siswa yang sebelumnya di *training* dan prediksi tersebut diambil berdasarkan *voting* terbaik dari banyak *tree* yang dihasilkan oleh *random forest*. pada Gambar 8 dapat dilihat penerapan operator *performance*.



Gambar 8. Penerapan Operator Performance

Operator *performance* pada tools Rapid Miner digunakan untuk memastikan apakah prediksi yang didapatkan dari metode algoritma *random forest* tersebut bisa dipercaya dan bisa digunakan sebagai acuan untuk mengambil suatu keputusan. Pada Gambar 9 dapat dilihat hasil dari penerapan operator *performance* yang berupa *result performance*.

pred.	true TIDAK BERPRESTASI	true BERPRESTASI	class precision
pred. TIDAK BERPRESTASI	56	0	100.00%
pred. BERPRESTASI	0	10	100.00%
class recall	100.00%	100.00%	

Gambar 9. Result Performance

Dari hasil penerapan metode algoritma *random forest* dapat diprediksi bahwa ada 56 siswa tidak berprestasi dan 10 siswa berprestasi dengan *accuracy*: 100%. Dari *result* di atas dapat disimpulkan bahwa hasil

prediksi memiliki *precision* 100% dan *recall* 100% yang artinya hasil tersebut sempurna dan dapat digunakan sebagai acuan untuk membantu pengambilan suatu keputusan.

3.2 Pembahasan

Hasil eksperimen penerapan algoritma Random Forest berhasil mengklasifikasikan siswa Bimbel menjadi kelompok berprestasi (14 dari 112 siswa) dan tidak berprestasi. Prediksi untuk siswa berprestasi memiliki tingkat keyakinan (*confidence*) yang relatif tinggi, dengan nilai minimum 0.825. Secara keseluruhan, rentang *confidence* model berkisar antara 0.005 hingga 0.933. Analisis nilai rata-rata siswa berprestasi menunjukkan keunggulan pada mata pelajaran Bahasa Inggris (85.38) dan Matematika (83.66), serta area yang perlu diperhatikan pada IPS (70.47) dan IPA (78.96). Temuan ini mengindikasikan bahwa algoritma Random Forest mampu mengidentifikasi pola kinerja siswa dengan tingkat keyakinan yang terukur, memberikan dasar bagi pemahaman karakteristik akademik siswa di Bimbel.

3.3 Implikasi Penelitian

Penelitian menggunakan algoritma *random forest* pada data nilai siswa Bimbel menghasilkan 4 siswa dengan predikat berprestasi yaitu:

Tabel 2. *Prediction Random Forest*

NAMA	MAT	IND	ENG	IPA	IPS	JUMLAH	PREDICTION
BAYU MALIK IBRAHIM	85,3	79,5	97,8	74,1	58,8	395,5	BERPRESTASI
BINTANG HIDAYAT	79,5	67,1	81,9	76,9	81,0	386,4	BERPRESTASI
MUHAMMAD FAUZAN RAMA	89,79	80,07	88,72	73,98	73,43	405,99	BERPRESTASI
SHALWA RIDANI A	85,71	86,71	85,90	71,97	66,79	397,08	BERPRESTASI

Tabel 3. *Result Data Confidence*

NAMA	<i>Confidence</i> (Tidak Berprestasi)	<i>Confidence</i> (Berprestasi)
BAYU MALIK IBRAHIM	0.07505772005772006	0.92494227994228
BINTANG HIDAYAT	0.17505772005772008	0.82494227994228
MUHAMMAD FAUZAN RAMA	0.09283549783549783	0.9071645021645023
SHALWA RIDANI A	0.06683549783549785	0.9331645021645022

Hasil klasifikasi siswa berprestasi dan tingkat *confidence* yang dihasilkan oleh algoritma *Random Forest* dapat menjadi acuan yang signifikan bagi pihak Bimbel dalam pengambilan kebijakan terkait proses belajar siswa. Dengan identifikasi siswa berpotensi yang didukung oleh nilai *confidence* yang tinggi, Bimbel dapat merancang program pengembangan bakat yang lebih terfokus. Selain itu, pemahaman mengenai mata pelajaran dengan rata-rata nilai rendah pada kelompok siswa berprestasi dapat mendorong kebijakan intervensi dini atau program pengayaan yang lebih spesifik. Secara keseluruhan, penelitian ini menunjukkan potensi implementasi *machine learning* sebagai alat bantu pengambilan keputusan berbasis data, yang dapat meningkatkan efektivitas program dan layanan Bimbel dalam mendukung perkembangan akademik siswa.

4. PENUTUP

Berdasarkan analisis menggunakan algoritma *random forest* dalam klasifikasi siswa berprestasi menggunakan data nilai siswa pada Bimbel, maka diperoleh kesimpulan bahwa penelitian ini berhasil menerapkan algoritma *Random Forest* untuk mengklasifikasikan siswa Bimbel menjadi kelompok berprestasi dan tidak berprestasi. Hasil klasifikasi menunjukkan identifikasi 14 siswa berprestasi dari 112 siswa dengan tingkat keyakinan (*confidence*) yang relatif tinggi (minimal 0.825). Dampak signifikan dari temuan ini adalah tersedianya metode identifikasi siswa berpotensi yang terukur dan berbasis data, yang secara langsung mendukung tujuan awal penelitian dalam menyediakan informasi berharga bagi pihak Bimbel. Dengan pemahaman yang lebih baik mengenai karakteristik akademik siswa berprestasi, Bimbel kini memiliki dasar yang lebih kuat untuk merancang kebijakan dan intervensi yang lebih efektif dalam meningkatkan kualitas pembelajaran.

5. SARAN

Saran yang diharapkan dalam pengembangan penelitian selanjutnya agar lebih baik adalah: Sebaiknya data yang digunakan dalam menganalisis prestasi siswa tidak hanya diambil dari periode 2017/2018 saja agar hasil yang didapatkan semakin bervariasi; Bandingkan algoritma *random forest* dengan algoritma lain agar hasil dari algoritma *random forest* tersebut dapat diketahui kekurangan dan kelebihanannya. Sehingga penelitian kedepannya dapat menggunakan algoritma yang lebih baik.

DAFTAR PUSTAKA

- [1] Apandi, S., Nurhidayatulloh, & Subariah, R. (2024). Implementasi metode *Decision Tree* dalam menentukan tingkat prestasi belajar siswa di SMP Negeri 83 Jakarta. *Prosiding Seminar Nasional Informatika dan Sistem Informasi, 1*, 8–17.
- [2] Breiman, L. (2001). *Random forests*. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- [3] Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Waltham, MA: Elsevier Inc.
- [4] Hidayat, R., & Nugroho, A. (2021). Penerapan metode SAW untuk sistem pendukung keputusan penentuan siswa berprestasi. *Jurnal Teknologi dan Sistem Komputer*, 9 (1), 42–49.
- [5] Kurniawan, A., & Prasetyo, E. (2020). Perbandingan algoritma C4.5, Random Forest, dan Naive Bayes dalam prediksi kelulusan mahasiswa. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 7 (4), 545–552.
- [6] Larose, D. T., & Larose, C. D. (2015). *Data mining and predictive analytics*. Wiley.
- [7] Mambang, & Byna, A. (2018). Analisis perbandingan algoritma C4.5, Random Forest, dan CHAID decision tree untuk klasifikasi tingkat kecemasan ibu hamil. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(9), 2957–2966.
- [8] Painem, P., Soetanto, H., & Solichin, A. (2023). Analisis keberhasilan studi awal mahasiswa menggunakan klusterisasi K-Means. *Faktor Exacta*, 16(3), 219–230. <https://doi.org/10.30998/faktorexacta.v16i3.19539>
- [9] Putri, D. M., & Syahril, S. (2022). Analisis faktor yang mempengaruhi prestasi belajar siswa menggunakan data mining. *Jurnal Informatika dan Komputer*, 10(1), 23–30.
- [10] Ramadhani, M., & Supriyadi, E. (2021). Implementasi algoritma decision tree untuk klasifikasi prestasi akademik mahasiswa. *Jurnal Ilmiah Teknik Informatika*, 10 (2), 101–110.
- [11] Rahman, F., & Firdaus, M. I. (2016). Penerapan data mining metode Naive Bayes untuk prediksi hasil belajar siswa sekolah menengah pertama (SMP). *AI-Ulum Sains dan Teknologi*, 2(1), 76–78.
- [12] Rahmawati, E., Saputra, A., & Rahayu, D. (2021). Prediksi kelulusan mahasiswa menggunakan algoritma Random Forest. *Jurnal Teknologi dan Sistem Komputer*, 9(2), 112–119.
- [13] Ridwan, W., & Gunawan, W. (2024). Penentuan mahasiswa berprestasi menggunakan algoritma FP-Growth dan SAW. *Faktor Exacta*, 17(3), 306–313. <https://doi.org/10.30998/faktorexacta.v17i3.23936>
- [14] Santosa, B. (2007). *Data mining terapan dengan Matlab*. Yogyakarta: Graha Ilmu.
- [15] Saputra, M. A., Sari, R. P., & Ningsih, H. (2022). Implementasi algoritma Random Forest untuk klasifikasi data akademik mahasiswa. *Jurnal Ilmu Komputer Terapan*, 6(3), 145–153.
- [16] Siregar, A. M., & Puspabhuana, A. (2017). *Data mining: Pengolahan data menjadi informasi dengan RapidMiner*. Surakarta: CV Kekata Group.
- [17] Subariah, R., Apandi, S., & Nurhidayatulloh. (2023). Analisis efektivitas pembelajaran daring pada Lembaga Bimbingan Belajar Nurul Fikri menggunakan metode Naive Bayes. *Jurnal ESIT (E-Bisnis, Sistem Informasi, Teknologi Informasi)*, 2, 21–30.
- [18] Tan, P. N., Steinbach, M., & Kumar, V. (2018). *Introduction to data mining* (2nd ed.). Pearson.
- [19] Vercellis, C. (2009). *Business intelligence: Data mining and optimization for decision making*. Chichester, UK: John Wiley & Sons.
- [20] Wirth, R., & Hipp, J. (2000). CRISP-DM: *Towards a standard process model for data mining*. In *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, 29–39.