



Prediksi Nilai Akhir Semester Siswa Menggunakan Algoritma Random Forest

Slamet Kusworo¹, Nugroho Adhi Santoso^{2*}, Rifki Dwi Kurniawan³
¹⁻³STMIK YMI Tegal, Indonesia

Alamat: Jalan Pendidikan No. 1, Pesurungan Lor, Margadana, Kota Tegal.

Korespondensi penulis: nugrohoadhisantoso@stmik-tegal.ac.id*

Abstract. *Students' end-of-semester grades show the results of the learning process that has been carried out for 1 semester, therefore accurate evaluation is very important to determine the extent to which learning objectives have been achieved properly. This research aims to create an appropriate prediction model to project students' final semester grades before the assessment period so that it can help teachers to identify at-risk students and provide appropriate treatment. This research uses the random forest algorithm and uses a dataset of student grades. The Random Forest model was successfully obtained through training and model evaluation using test data, showing an accuracy of 93.33%. The use of the random forest algorithm to predict students' final semester grades can be the right way to produce high accuracy for the benefit of educators in handling their students.*

Keywords: *Final grade prediction, Random forest algorithm, Learning evaluation*

Abstrak. Nilai akhir semester siswa menunjukkan hasil dari proses pembelajaran yang telah dilakukan selama 1 semester, oleh karena itu evaluasi yang akurat sangat penting untuk mengetahui sejauh mana tujuan pembelajaran telah tercapai dengan baik. Penelitian ini bertujuan untuk menciptakan model prediksi yang tepat untuk memproyeksikan nilai akhir semester siswa sebelum periode penilaian sehingga dapat membantu guru untuk mengidentifikasi siswa yang berisiko dan memberikan penanganan yang tepat. Penelitian ini menggunakan algoritma random forest dan menggunakan Dataset nilai siswa. Model Random Forest berhasil diperoleh melalui pelatihan dan evaluasi model menggunakan data uji, menunjukkan akurasi sebesar 93,33%. Penggunaan algoritma random forest untuk memprediksi nilai akhir semester siswa dapat menjadi cara yang tepat untuk menghasilkan akurasi yang tinggi untuk kepentingan pendidik dalam menangani siswanya.

Kata kunci: Prediksi Nilai Akhir, Algoritma random forest, Evaluasi pembelajaran

1. LATAR BELAKANG

Evaluasi hasil belajar siswa merupakan komponen Penting dalam sistem pendidikan, berperan dalam menilai pencapaian akademik dan efektivitas pembelajaran (Yoga Siswa, 2023) Nilai akhir semester merupakan indikator utama keberhasilan proses pembelajaran. Evaluasi yang akurat terhadap nilai ini sangat penting untuk mengetahui sejauh mana siswa memahami materi yang telah diajarkan dan apakah tujuan pembelajaran telah tercapai. Dengan demikian, kemampuan untuk memprediksi nilai akhir semester sebelum periode penilaian dapat membantu guru untuk mengidentifikasi siswa yang berisiko dan memberikan intervensi yang tepat (Pekuwali, 2020). Kesalahan dalam memprediksi nilai akhir semester dapat menyebabkan keputusan yang salah dalam penentuan kebijakan pembelajaran selanjutnya. Menurut BPS, tingkat putus sekolah untuk tingkat SMA/SMK pada tahun 2022 mencapai 1,45%, hal ini disebabkan karena rendahnya pencapaian akademik siswa. (Bunda & Junaidi, 2021). Oleh karena itu, penelitian tentang prediksi nilai akhir semester menggunakan algoritma

machine learning menjadi sangat penting untuk meningkatkan keberhasilan pembelajaran (Ramadhani et al., 2022).

Penelitian prediksi nilai akhir siswa menggunakan berbagai algoritma klasifikasi dilatarbelakangi oleh kebutuhan untuk mengoptimalkan proses pembelajaran dan evaluasi di dunia pendidikan. Dalam upaya meningkatkan kualitas pendidikan, pendidik perlu alat yang dapat memprediksi kinerja siswa secara lebih akurat, sehingga dapat memberikan perhatian khusus kepada siswa yang berisiko rendah performanya (Wijayaningrum et al., 2023). Penggunaan *algoritma* klasifikasi, seperti *Decision Tree*, *Naive Bayes*, dan *Support Vector Machine*, menawarkan pendekatan yang berbasis data untuk memodelkan hubungan antara berbagai faktor yang memengaruhi prestasi akademik siswa, seperti nilai ujian, absensi, interaksi di kelas, dan faktor sosial-ekonomi (Nalatissifa et al., 2021).

2. KAJIAN TEORITIS

Beberapa penelitian telah dilakukan dengan menggunakan *algoritma klasifikasi* dengan tujuan untuk mengidentifikasi faktor-faktor yang memengaruhi prestasi akademik siswa dan mengembangkan model prediksi untuk mengantisipasi potensi keberhasilan atau kegagalan siswa di masa depan (Nalatissifa et al., 2021). Contohnya penelitian yang dilakukan oleh Supangat & Sulistyawan untuk memprediksi Nilai Akhir Matakuliah Mahasiswa Menggunakan *Metode K-Means Clustering* dan *Naive Bayes Classifier* (Sulistyawan et al., 2021). Berikutnya lagi penelitian dengan judul “Perbandingan Metode Klasifikasi *Naive Bayes* Dan *K-Nearest Neighbor* Dalam Memprediksi Prestasi Siswa” (Winantu & Khatimah, 2023), Penelitian ini membandingkan metode klasifikasi *Naive Bayes* dan *K-Nearest Neighbor* dalam memprediksi prestasi siswa.

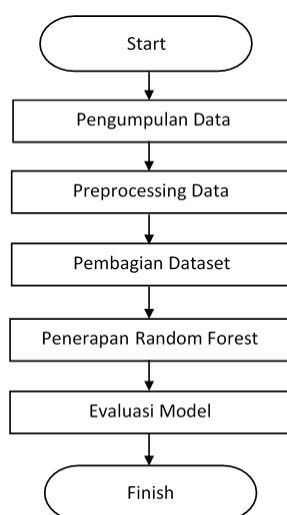
Akan tetapi penggunaan algoritma klasifikasi seperti *Naive Bayes* dan *K-Nearest Neighbor* memiliki kekurangan dan kelemahan masing-masing. seperti sensitif terhadap fitur yang tidak *relevan* sehingga dapat menurunkan akurasi prediksi, serta memerlukan kompleksitas komputasi yang tinggi (Wijayaningrum et al., 2023). Hal ini mendorong penelitian untuk mencari *algoritma* klasifikasi yang lebih dapat diandalkan dalam memprediksi prestasi siswa. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model prediksi nilai akhir siswa dengan menggunakan algoritma *Random Forest*, yang diharapkan dapat memberikan akurasi yang lebih baik dibandingkan dengan metode lain.

Random Forest beroperasi dengan cara menciptakan sejumlah pohon keputusan selama proses pelatihan, dimana hasil dari setiap pohon keputusan akan dirata-ratakan untuk meningkatkan ketepatan prediksi. *Metode Random Forest* (RF) juga dikenal sebagai teknik

yang mampu meningkatkan akurasi hasil, karena dalam proses pembentukan simpul anak untuk setiap node dilakukan secara acak. tehnik ini digunakan untuk membangun pohon keputusan yang terdiri dari *root node*, *internal node*, dan *leaf node* dengan cara memilih *atribut* dan data secara acak sesuai aturan yang ditetapkan. *Root node* adalah simpul yang berada di puncak, seringkali disebut sebagai akar dari pohon keputusan (Putri Choirunisa, 2020). Penerapan *Random Forest* dalam meramalkan nilai akademik memberikan kesempatan bagi model untuk memperhitungkan berbagai elemen yang berpengaruh terhadap performa siswa. Faktor – faktor seperti hasil ujian, frekuensi kehadiran dan keterlibatan dalam aktivitas kelas dapat berkontribusi pada peningkatan ketepatan dalam memprediksi nilai akhir semester.

3. METODE PENELITIAN

Penelitian ini Mempunyai sejumlah tahapan penelitian yang dibuat dengan baik untuk mendapatkan hasil yang akurat dan relevan. Urutan langkah-langkah penelitian ini mencerminkan alur sistematis yang diikuti selama proses penelitian. Tahapan penelitian dapat dilihat pada gambar1 sebagai berikut:



Gambar 1. Tahapan Penelitian

Pegumpulan Data

Pada tahapan ini dimulai dengan identifikasi dan seleksi sumber data yang relevan, yaitu nilai tugas, nilai ulangan harian ,ujian tengah semester (UTS), ujian akhir semester (UAS), kehadiran, dan faktor lain yang mempengaruhi kinerja akademik siswa. Data tersebut dikumpulkan melalui sistem informasi akademik. Data nilai tersebut diperoleh dari mata pelajaran informatika pada siswa kelas XI pada tahun ajaran 2023 / 2024 dengan jumlah sampel data di ambil dari 100 Siswa.

Processing Data

Langkah awal pada proses ini adalah pembersihan data, yang meliputi penanganan data yang hilang atau tidak valid melalui metode imputasi atau penghapusan, selain memverifikasi konsistensi data untuk memastikan bahwa nilai yang digunakan bebas dari perbedaan. Setelah ini, normalisasi atau standardisasi dilakukan pada fitur dengan skala yang bervariasi, seperti kehadiran dan nilai ujian, untuk menjamin bahwa model tetap tidak bias terhadap fitur yang memiliki skala yang lebih besar.

Pembagian DataSet

Hasil penelitian ini diharapkan dapat memberikan manfaat praktis dalam meningkatkan akurasi prediksi nilai akhir siswa. Dengan prediksi yang lebih akurat, sekolah dapat lebih cepat mengidentifikasi siswa yang membutuhkan bantuan atau intervensi khusus serta merancang strategi pembelajaran yang lebih efektif dan berbasis data. Lebih lanjut, penerapan algoritma random forest dalam konteks ini dapat berkontribusi pada pengembangan sistem penilaian pendidikan berbasis teknologi yang lebih efisien, sehingga pada akhirnya meningkatkan kualitas pendidikan di Indonesia.

Penerapan Random Forest

Pada tahapan ini diterapkan *algoritma* random forest untuk menciptakan model prediksi yang bertujuan untuk memprediksi nilai akhir semester siswa. Random forest merupakan metode pembelajaran *ansambel* yang terdiri dari banyak pohon keputusan, di mana setiap pohon dilatih secara mandiri menggunakan subset data pelatihan yang acak. Proses ini dilakukan melalui teknik *bagging* (*bootstrap* agregasi), yang memungkinkan penentuan sampel data pelatihan secara acak untuk setiap pohon keputusan yang dibangun. Selain itu, untuk menekan korelasi antar pohon dan meningkatkan keanekaragaman, pemilihan fitur juga dilakukan secara acak pada setiap belahan pohon. Dengan cara ini, setiap pohon keputusan yang dihasilkan memiliki struktur berbeda walaupun dilatih dengan data serupa. Keberagaman antar pohon-pohon ini memungkinkan model *Random Forest* untuk mengurangi *overfitting*, yaitu kondisi di mana model terlalu cocok dengan data pelatihan dan kinerjanya menurun pada data yang belum pernah dilihat sebelumnya, serta meningkatkan generalisasi model pada data yang lebih luas (Liu et al., 2012).

Setelah setiap pohon keputusan dilatih, setiap pohon akan memunculkan prediksi yang berbeda berdasarkan *subset* data yang digunakan. Untuk tugas pengelompokan, prediksi akhir model *Random Forest* ditentukan melalui pengambilan suara terbanyak, yaitu dengan

memilih kelas dengan prediksi terbanyak dari satu pohon. Dalam kasus regresi, prediksi akhir dihitung dengan mengambil rata-rata prediksi yang diberikan oleh setiap pohon. Hasil akhir ini mewakili kesepakatan seluruh pohon yang terlibat dalam model

Berikut ini Langkah-langkah proses *Random Forest* secara ringkas dan sederhana

1. **Dataset Bootstrap:** *Dataset D* disampling ulang dengan pengembalian (*Bootstrap*) untuk membentuk T subset, D_1, D_2, \dots, D_T . Pada setiap node pohon, dipilih secara acak m fitur ($m < M$) untuk menentukan *split* terbaik, guna mengurangi korelasi antar pohon.
2. **Pembangunan Pohon:** Pohon keputusan dibangun secara independen pada setiap *subset* D_i . *Split* terbaik ditentukan menggunakan fungsi impuritas, seperti *Gini Index*:

$$Gini = 1 - \sum_{k=1}^K p_k^2$$

atau *Entropi*:

$$Entropy = - \sum_{k=1}^K p_k \log_2(p_k)$$

di mana p_k adalah *probabilitas* kelas k . Pohon dibuat maksimal tanpa proses *pruning*.

3. **Prediksi Ensemble:** Prediksi akhir diperoleh dengan menggabungkan hasil dari semua pohon. Untuk klasifikasi, digunakan voting mayoritas:

$$\hat{y} = \text{Mode}(\{y_1, y_2, \dots, y_T\})$$

Sedangkan untuk regresi, digunakan rata-rata prediksi:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T y_t$$

Dengan pendekatan ini, *Random Forest* memberikan model yang akurat dan *robust* terhadap *overfitting*.

Evaluasi Model

Model *Random Forest* yang dibangun kemudian dievaluasi menggunakan data pengujian untuk mengukur akurasi dan kemampuan prediksi nilai akhir semester siswa. Hasil evaluasi model dapat mencakup metrik seperti akurasi *F1-score*, yang memberikan gambaran tentang seberapa efektif model dalam memprediksi nilai akhir siswa berdasarkan data yang ada. Penentuan hasil akhir prediksi dengan menggunakan suara mayoritas atau rata-rata ini memastikan bahwa model memiliki tingkat prediktif yang tinggi dalam berbagai jenis tugas, baik untuk klasifikasi maupun regresi (Ihzaniah et al., 2023).

4. HASIL DAN PEMBAHASAN

Pengumpulan data

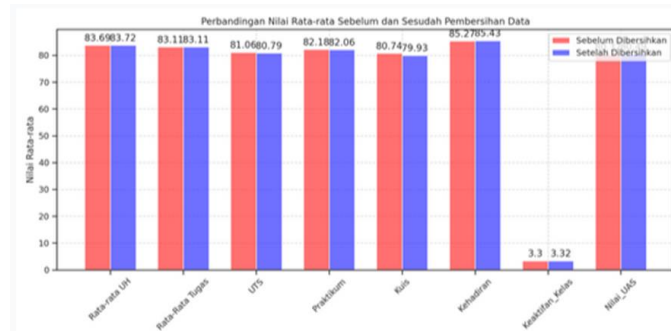
Dataset yang digunakan dalam penelitian ini terdiri dari 100 siswa dengan berbagai atribut terkait performa akademik dan karakteristik pribadi. Dataset mencakup 20 atribut, termasuk atribut prediktor seperti rata-rata ulangan harian, rata-rata tugas, nilai UTS, partisipasi dalam diskusi, serta atribut target berupa nilai UAS (Nilai Akhir Semester). Berikut ini kami tampilkan beberapa bagian isi dataset yang ditampilkan hanya sebagian saja. Dataset siswa ditunjukkan pada gambar 2 sebagai berikut:

NO	Nama Siswa	Rata-rata UH	Rata-Rata Tugas	UTS	Praktikum	Kuis	Kehadiran	Keterlambatan	Bimbel	Nilai_UAS
1	ABDUL SOFAN SOFIAN	82	84	80	85	81	78	1	0	82
2	AGIN FAUZA BIL JANAT	80	82	77	82	78	74	1	0	79
3	ARUM MULYANI	77	81	73	78	74	87	1	1	75
4	ARUM SEKAR SARI	83	84	81	86	82	91	0	1	83
5	AULIA NUR HADIANI	85	84	81	86	82	96	0	0	83
6	BASALFA MEYTA KASIH	83	83	82	87	84	83	0	1	75
7	DIANDRA AYUDIA SUPARNO	79	81	73	78	74	91	0	0	75
8	GENDIZ NURAENI	86	84	82	87	83	87	1	0	84
9	INAYIAH SEKAR AYU	86	84	81	86	82	87	1	0	83
10	LISA LAORA YULIANTI	87	82	81	86	83	87	1	1	84
11	MARTHA NURVANI	89	82	80	86	83	100	0	0	84
12	MELISA SEPTI RAMADHANI	84	85	77	79	83	91	0	0	84
13	MUHAMMAD DANANG SAPUTI	82	83	73	83	81	74	2	0	82
14	NABILA AURA ARDIFAH	82	81	81	84	82	100	0	1	83
15	NAJWA HANDANU NAFTINSI	86	83	81	84	82	91	1	1	83
16	NAVA ALIFYA PUTRI	83	83	73	84	82	91	1	1	83
17	NUR LAELATUL NASPIYAH	82	81	73	84	82	96	0	0	83
18	NUR LAJLI KOMARIYAH	82	83	82	83	81	83	2	0	82
19	PUAN SALMA	82	84	81	83	81	96	1	1	82
20	REVA KELVIANTO	82	85	82	84	82	96	1	0	83

Gambar. 2. Dataset siswa

Processing Data

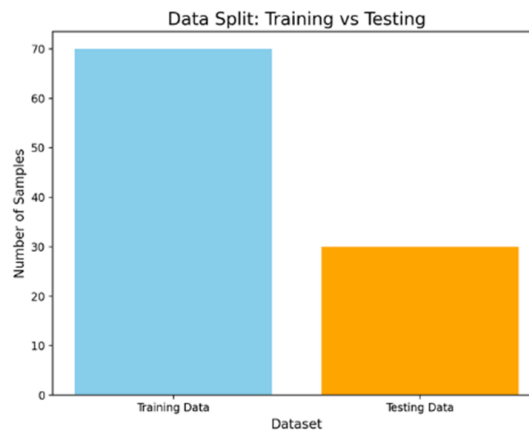
Proses selanjutnya adalah pembersihan data sebelum data diolah, ini merupakan langkah fundamental dalam memastikan kualitas dataset sebelum diterapkan pada algoritma prediktif, seperti Random Forest. Grafik berikut menggambarkan perbandingan distribusi data sebelum dan setelah pembersihan, menunjukkan peningkatan konsistensi dan representasi data. Langkah ini dilakukan untuk memastikan bahwa dataset yang digunakan bebas dari bias akibat nilai kosong atau duplikasi, sehingga dapat mendukung akurasi model prediktif secara optimal, berikut ini disajikan dataset sebelum dan sesudah data dibersihkan untuk diolah dengan Algoritma Random Forest pada gambar 3.



Gambar 3. Grafik perubahan dataset yang telah dibersihkan

Grafik di atas memperlihatkan distribusi data untuk semua atribut sebelum dan setelah pembersihan menggunakan visualisasi boxplot. Sebelum pembersihan (ditunjukkan dengan warna merah), distribusi data pada beberapa atribut menunjukkan adanya nilai kosong atau tidak valid, yang tercermin dari rentang yang tidak terdefinisi atau median yang tidak konsisten. Setelah pembersihan (ditunjukkan dengan warna biru), distribusi data menjadi lebih stabil dan representatif, dengan median yang lebih terpusat dan rentang yang jelas untuk setiap atribut. Peningkatan ini mencerminkan keberhasilan proses pembersihan dalam mengatasi kekosongan data, mengurangi bias, dan memastikan kesiapan dataset untuk analisis prediktif yang andal. Hal ini menunjukkan bahwa data telah diolah sesuai dengan standar metodologi penelitian yang baik untuk mendukung pengambilan keputusan berbasis data.

Pembagian Dataset



Gambar 4. Grafik pembagian *dataset*

Keterangan Gambar:

- Grafik batang yang berwarna biru menunjukkan porsi 70 % dari data set yang dijadikan sebagai data *training*
- Grafik batang yang berwarna oranye menunjukkan porsi 30 % dari data set yang dijadikan sebagai data *testing*

Dataset terdiri dari beberapa atribut input, seperti nilai ulangan harian, nilai tugas, nilai UTS, nilai praktikum, serta variabel tambahan seperti partisipasi diskusi, keaktifan bertanya, jenis kelamin, umur, jarak rumah, waktu belajar, dan keikutsertaan dalam bimbingan belajar (bimbel). Atribut target yang menjadi fokus penelitian adalah nilai akhir semester (Nilai UAS). Dataset dibagi Menjadi 70% dan 30 % menggunakan metode stratified sampling, yang memastikan distribusi kategori pada atribut target (Nilai UAS) tetap proporsional di kedua

subset (data pelatihan dan data pengujian). Hal ini penting untuk menjaga representasi distribusi data, terutama jika data target memiliki kelas yang tidak seimbang.

Penerapan random forest

Dataset dilatih dan di uji pada tahapan ini untuk menghasilkan nilai prediksi yang baik dan akurat , disini dijabarkan tahapan tahapan yang dilakukan.

Pelatihan data :

Proses pelatihan dilakukan dengan menggunakan *Algoritma Random Forest*, yang dirancang untuk bekerja secara optimal dengan data tabular. Model dilatih menggunakan parameter berikut:

- ***n_estimators***: Sebanyak 100 pohon digunakan untuk memastikan stabilitas prediksi.
- ***max_depth***: Kedalaman maksimum pohon ditentukan melalui eksperimen untuk mencegah overfitting.
- ***criterion***: Kriteria pemisahan yang digunakan adalah "*gini impurity*" untuk mengukur kualitas pemisahan.

Proses pelatihan dimulai dengan membangun beberapa pohon keputusan secara acak berdasarkan subset data pelatihan. Setiap pohon dilatih dengan menggunakan kombinasi fitur yang berbeda untuk meningkatkan diversitas model.

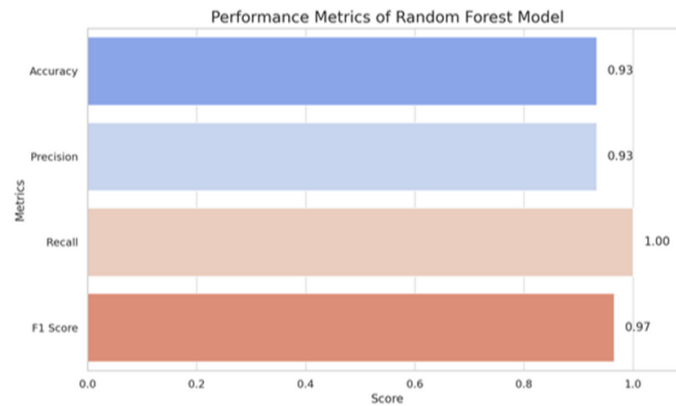
Pengujian Model :

Evaluasi pengujian model dilakukan dengan mengukur kemampuan generalisasi model pada data pengujian yang belum terlihat selama pelatihan. Model menghasilkan prediksi nilai akhir semester (nilai UAS) untuk setiap siswa, yang kemudian dibandingkan dengan nilai aktual untuk menilai performa Akurasi, *Precision*, *Recall*, dan *F1 Score* digunakan sebagai metrik evaluasi, yang secara kolektif menunjukkan bahwa model memiliki kinerja yang optimal dengan tingkat kesalahan minimal. Pengujian juga menegaskan pentingnya atribut seperti nilai UTS, tugas, dan partisipasi diskusi dalam memprediksi nilai akhir.

Evaluasi Model

Sebagai bagian dari penelitian ini, kami mengevaluasi performa model Random Forest dalam memprediksi nilai UAS siswa dengan parameter akademik, yaitu, tugas atau TGS, Ulangan Harian atau UH, UTS, dan Praktikum. Untuk memastikan apa sejauh mana performa sudah optimal, metrik evaluasi utamanya adalah Akurasi, *Precision*, *Recall*, dan *F1 Score*.

Visualisasi performa model kami berdasarkan masing-masing metrik, dari mana model dapat memberi hasil yang prediksi yang benar dan konsisten, dapat dilihat pada Gambar 5 berikut ini:



Gambar 5. Akurasi, Precision, Recall, dan F1 Score

Grafik pada Gambar 5. menyediakan representasi visual dari metrik evaluasi yang digunakan untuk mengukur representasi model *Algoritma Random Forest*. dengan prediksi berbagai fitur yang telah ditentukan sebelumnya untuk mengukur nilai akurasi pada Nilai Ujian Akhir Semester siswa. Metrik ini menyertakan Akurasi, Precision, Recall, dan F1 Score, yang mewakili perspektif terpisah dari prestasi model.

Akurasi : Model memiliki akurasi setara dengan 93,33%, mengekspresikan bahwa rata-rata prediksi UAS nilai ini wadah untuk uji nilai UAS sebagian besar siswa.

Precision : Precision memiliki representasi setara dengan 93,33%, menunjukkan bahwa model mampu menyediakan prediksi UAS tumpul dalam pengukuran olahraga ini, yang berarti prediksi kesalahan berkurang.

Recall : Recall adalah 100%, mengekspresikan bahwa model menerima kemampuan untuk menangkap semua kasus di atas tingkat akurasi, juga tidak kehilangan informasi penting sehubungan dengan nilai UAS siswa.

F1 Score : F1 Score setara 96,55%, mewakili titik koordinasi optimal antara Precision dan recall. Ini mengekspresikan bahwa model memiliki representasi stabil dari prediksi yang dilakukan.

Melalui visualisasi ini dapat dinyatakan bahwa model algoritma berbasis *Random Forest* memiliki tingkat keberhasilan sangat baik dalam Prediksi Nilai UAS siswa berdasarkan beberapa nilai mencakup nilai Ulangan harian, nilai tugas, UTS, praktikum dan nilai pendukung lain. Oleh karena itu, ini berarti bahwa model Random Forest adalah pendekatan model yang sangat bagus untuk dijadikan referensi pemodelan dalam bidang apapun.

5. KESIMPULAN DAN SARAN

Tujuan dari penelitian ini adalah memprediksi nilai akhir semester siswa di SMA PGRI 1 Taman Pematang menggunakan algoritma Random Forest. Dataset yang digunakan mencakup faktor akademik dan non-akademik seperti nilai ulangan, tugas, UTS, partisipasi diskusi, keaktifan bertanya, atribut demografis dan lain-lain. Seperti yang telah disebutkan, model yang dibuat dengan pembagian dataset 70 dan 30 dan mampu memperoleh akurasi 93,33%, precision 93,33%, recall 100%, dan f1-score 96,55%. Model dengan performa yang baik dan akurat memungkinkan guru untuk mengetahui faktor-faktor siswa yang dapat beresiko mendapat nilai rendah lebih dini dan memungkinkan untuk melakukan intervensi. Implementasi model ini memberikan manfaat yang sangat penting, termasuk efektivitas pengajaran, pengambilan keputusan berbasis data, dan peningkatan mutu pendidikan di SMA PGRI 1 Taman.

DAFTAR REFERENSI

- Bunda, A. P., & Junaidi, J. (2021). Penyebab rendahnya hasil belajar peserta didik kelas X IIS mata pelajaran sosiologi semester ganjil tahun ajaran 2020/2021 di SMAN 10 Padang. *Jurnal Kajian Pendidikan Dan Pembelajaran*, 2(4), 297–306.
- Ihzanah, L. S., Setiawan, A., Wijaya, R. W. N., Matematika, P. S., Kristen, U., Wacana, S., Tengah, J., & Berganda, R. L. (2023). *Jambura Journal of Probability and Statistics*, 4(1), 17–29.
- Liu, Y., Wang, Y., & Zhang, J. (2012). New machine learning algorithm: Random forest. In *Lecture Notes in Computer Science* (Vol. 7473, pp. 246–252). Springer. https://doi.org/10.1007/978-3-642-34062-8_32
- Nalatissifa, H., Gata, W., Diantika, S., & Nisa, K. (2021). Perbandingan kinerja algoritma klasifikasi Naive Bayes, Support Vector Machine (SVM), dan Random Forest untuk prediksi ketidakhadiran di tempat kerja. *Jurnal Informatika Universitas Pamulang*, 5(4), 578. <https://doi.org/10.32493/informatika.v5i4.7575>
- Pekuwali, A. A. (2020). Prediction of student learning outcomes using the Naive Bayesian Algorithm (Case Study of Tama Jagakarsa University). *IOP Conference Series: Materials Science and Engineering*, 823(1). <https://doi.org/10.1088/1757-899X/823/1/012056>
- Putri, C. (2020). Implementasi Artificial Intelligence untuk memprediksi harga penjualan rumah menggunakan metode Random Forest dan Flask. *UII.Ac.Id*, 1–101.
- Ramadhani, A. N., Ardiansyah, R. G., & Latifah, U. (2022). Penilaian alat hasil belajar untuk siswa sekolah dasar SDN Sindangsari 1 Desa Sindangsari. *Alsyst*, 2(2), 292–302. <https://doi.org/10.58578/alsyst.v2i2.303>

- Sulistiyawan, M. R., Studi, P., Informatika, T., & Teknik, F. (2021). Pemodelan prediksi tingkat kelulusan mahasiswa dengan pendekatan algoritma naïve bayes. *Jurnal Informatika*, 405–414.
- Wijayaningrum, V. N., Putri, I. K., Kirana, A. P., Mubarak, M. R., Harahap, D. M., & Hamesha, B. R. (2023). Analisis performa seleksi atribut untuk menentukan potensi mahasiswa putus studi. *Jurnal Informatika Polinema*, 9(2), 237–244. <https://doi.org/10.33795/jip.v9i2.1300>
- Winantu, A., & Khatimah, C. (2023). Perbandingan metode klasifikasi Naive Bayes dan K-Nearest Neighbor dalam memprediksi prestasi siswa. *INTEK: Jurnal Informatika Dan Teknologi Informasi*, 6(1), 58–64. <https://doi.org/10.37729/intek.v6i1.3006>
- Yoga Siswa, T. A. (2023). Komparasi optimasi Chi-Square, CFS, Information Gain dan ANOVA dalam evaluasi peningkatan akurasi algoritma klasifikasi data performa akademik mahasiswa. *Informatika Mulawarman: Jurnal Ilmiah Ilmu Komputer*, 18(1), 62. <https://doi.org/10.30872/jim.v18i1.11330>