Jurnal Ilmiah Informatika (JIF)

| ISSN (Print) 2337-8379 | ISSN (Online) 2615-1049 |

Jurnal online di akses di http://ejournal.upbatam.ac.id/index.php/jif

Prediksi Penerimaan Mahasiswa Baru Universitas Singaperbangsa Karawang dengan Random Forest

Mochammad Fitra Pamungkas[1], Betha Nurina Sari[2], Iqbal Maulana[3]

[123] Universitas Singaperbangsa Karawang, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361, Indonesia

INFORMASI ARTIKEL

Sejarah Artikel:
Diterima Redaksi: 09-07-2025
Revisi Akhir: 18-08-2025
Diterbitkan *Online*: 10-09-2025

KATA KUNCI

New student admission

Prediction

Random Forest

SNBT

KORESPONDENSI

E-mail: mochammad.fitra18209@student.unsika.ac.id

A B S T R A C T

The selection of study programs and the selection process for new student admissions are crucial stages that have an impact on the smooth running of studies and students' future careers. However, this process is often still done subjectively, thus potentially causing a mismatch between interests, abilities, and the chosen study program. This research aims to implement the Random Forest algorithm, analyze the factors that influence acceptance, and determine the performance of the prediction model generated by the Random Forest algorithm in assisting new student admissions for the National Selection Based on Test (SNBT) pathway at Universitas Singaperbangsa Karawang. This research uses the Knowledge Discovery in Database (KDD). The data used was 815 new student data of Universitas Singaperbangsa Karawang, which included UTBK scores, school background, and choice of prospective student study programs. In the Transformation and Data Mining stages, 4 data splitting scenarios were carried out, namely 90:10, 80:20, 70:30, and 60:40. The best performance with the Random Forest model is generated by the 90:10 data splitting scenario with an accuracy value of 0.945, precision of 0.937, recall of 0.943, f1-score of 0.940 and AUC value of 0.983.

1. PENDAHULUAN

Berdasarkan data [1], jumlah siswa yang melanjutkan ke tingkat pendidikan tinggi berada dibawah 28%, terdapat lebih banyak siswa yang tidak melanjutkan ke tingkat perguruan tinggi karena banyak faktor. Dari data tersebut, siswa yang dapat melanjutkan ke tingkat pendidikan tinggi harus memanfaatkan kesempatan tersebut dengan baik. Salah satunya dengan memilih perguruan tinggi impian yang sesuai dengan minat dan bakat.

Perguruan tinggi juga memiliki peran krusial dalam mencetak sumber daya manusia yang unggul. Salah satu aspek strategis yang sangat menentukan dalam hal ini adalah proses penerimaan mahasiswa baru (PMB), yang juga mencerminkan antusiasme dan persepsi masyarakat terhadap institusi pendidikan tinggi tersebut. Semakin tinggi minat calon mahasiswa untuk mendaftar, semakin menunjukkan reputasi dan kualitas pendidikan yang dimiliki oleh perguruan tinggi tersebut. Oleh sebab itu, PMB menjadi momentum penting yang dinantikan setiap tahun oleh berbagai perguruan tinggi negeri di Indonesia.

Dalam meningkatkan jumlah mahasiswa baru, perguruan tinggi negeri terlibat dalam persaingan yang cukup ketat, sehingga jumlah penerimaan mahasiswa dapat mengalami naik turun setiap tahunnya di berbagai wilayah. Di sisi lain, proses seleksi mahasiswa baru menjadi semakin kompleks karena perbedaan jenis tes dan kriteria penilaian dari masing-masing perguruan tinggi.

Salah satu bentuk seleksi yang memiliki pengaruh besar dalam proses penerimaan adalah Seleksi Nasional Berdasarkan Tes (SNBT). SNBT merupakan sistem seleksi yang dilaksanakan oleh perguruan tinggi melalui ujian tertulis atau daring, dengan tujuan menilai kesiapan dan kemampuan akademik calon mahasiswa. Ujian SNBT mencakup beberapa aspek, antara lain potensi akademik, kemampuan dasar dalam sains dan teknologi, bahasa Inggris, serta keterampilan berpikir.

Agar calon mahasiswa dapat menentukan pilihan universitas dan program studi yang sesuai dengan hasil SNBT, dibutuhkan prediksi peluang kelulusan mereka. Salah satu pendekatan yang dapat dimanfaatkan adalah dengan menggunakan algoritma Random Forest, yaitu algoritma klasifikasi berbasis probabilitas

dan statistik yang dapat memprediksi kemungkinan di masa depan berdasarkan data historis. Dengan menggabungkan teknologi informasi dan metode Random Forest, diharapkan dapat membantu calon mahasiswa dalam menentukan jurusan dan perguruan tinggi yang paling sesuai dengan minat dan kemampuannya.

Penelitian sebelumnya telah mengidentifikasi beberapa metode yang digunakan dalam prediksi untuk mendukung pengambilan keputusan. Beberapa metode yang digunakan diantaranya meliputi *Iterative Dichotomizer 3, Naïve bayes*, dan algoritma klasifikasi lainnya. Metode *random forest* lebih sering menunjukkan kinerja lebih baik dibandingkan ID3 dan metode klasifikasi sederhana lain, namun penggunaan *random forest* masih belum banyak digunakan terutama dalam prediksi penerimaan mahasiswa baru. Oleh sebab itu penelitian bertujuan untuk melengkapi kekurangan pengetahuan tersebut dan berkontribusi pada perkembangan teknologi informasi dalam bidang pendidikan.

2. TINJAUAN PUSTAKA

2.1 Prediksi

Prediksi merupakan suatu tindakan untuk meramal sesuatu yang dapat terjadi pada masa mendatang berdasarkan data yang sudah ada. Menurut [2] Prediksi merupakan proses sistematis yang dilakukan untuk memperkirakan tentang sesuatu yang paling mungkin akan terjadi di masa depan berdasar informasi yang ada pada masa lalu dan sekarang. Prediksi tidak harus memberikan jawaban pasti kejadian yang akan datang, melainkan berusaha untuk mencari jawaban yang sedekat mungkin akan terjadi.

2.2 Perguruan Tinggi

Perguruan tinggi merupakan institusi pendidikan yang menyelenggarakan pendidikan tinggi serta menjadi tempat berlangsungnya kegiatan pengajaran, penelitian, dan pengabdian kepada masyarakat. Berdasarkan Undang-Undang Nomor 12 Tahun 2012 tentang Pendidikan Tinggi, perguruan tinggi adalah satuan pendidikan yang dapat berbentuk akademi, politeknik, sekolah tinggi, institut, atau universitas yang menyelenggarakan pendidikan akademik, vokasi, dan/atau profesi [3].

Sebagai pusat pengembangan ilmu pengetahuan dan teknologi, perguruan tinggi juga berfungsi sebagai agen perubahan (*agent of change*) dalam masyarakat. Perannya semakin penting dalam menghadapi tantangan global dan perkembangan teknologi informasi. Oleh karena itu, proses penerimaan mahasiswa baru sebagai bagian dari manajemen akademik harus dilakukan secara transparan, objektif, dan berbasis data, guna memastikan seleksi calon mahasiswa yang potensial dan sesuai dengan visi institusi. [4].

2.3 Penerimaan Mahasiswa Baru

Penerimaan mahasiswa baru merupakan proses seleksi dan penyeleksian calon mahasiswa yang dilakukan oleh perguruan tinggi untuk menentukan siapa saja yang layak diterima menjadi mahasiswa berdasarkan kriteria tertentu. Proses ini bertujuan untuk menjaring calon mahasiswa yang memiliki potensi akademik, kompetensi, dan kesiapan untuk mengikuti pendidikan tinggi.

Menurut Peraturan Mendikbud No. 6 Tahun 2020 tentang Penerimaan Mahasiswa Baru Program Sarjana pada Perguruan Tinggi Negeri[5], penerimaan mahasiswa baru program sarjana dilakukan melalui tiga jalur, yaitu:

1. Seleksi Nasional Berdasarkan Prestasi (SNBP)
2. Seleksi Nasional Berdasarkan Tes (SNBT)
3. Seleksi Mandiri oleh Perguruan Tinggi Negeri

2.4 *Train-Test Split*

Dalam pembangunan model prediksi, dataset umumnya dibagi menjadi data latih dan data uji melalui proses *train-test split*. Tujuannya adalah untuk melatih model pada sebagian data dan menguji kemampuannya pada data yang belum pernah dilihat, guna mengevaluasi generalisasi model dan mencegah *overfitting*. Rasio umum yang digunakan adalah 70:30, 80:20, dan 90:10, tergantung pada ukuran dataset dan kebutuhan validasi model secara objektif [6].

2.5 Orange data mining

Orange Data Mining adalah sebuah perangkat lunak *open-source* yang digunakan untuk menganalisis data dengan pendekatan visual. Dikembangkan oleh Bioinformatics Laboratory di University of Ljubljana, Dirancang untuk kemudahan penggunaan, Orange data mining memungkinkan pengguna membangun alur kerja analisis data dengan menyeret dan menghubungkan komponen-komponen (*widget*) melalui antarmuka grafis yang intuitif. Orange menyediakan berbagai fitur analisis seperti klasifikasi, regresi, *clustering*, asosiasi, dan visualisasi data, serta mendukung berbagai algoritma machine learning, termasuk *Random Forest, Naive Bayes, k-Nearest Neighbor*, dan lain-lain.

Keunggulan utama dari Orange adalah kemampuannya dalam menyederhanakan proses data preprocessing, pelatihan model, evaluasi performa, hingga interpretasi hasil, yang sangat bermanfaat bagi peneliti dan mahasiswa dalam mengolah data secara efisien dan terstruktur. Orange data mining juga mendukung integrasi dengan Python bagi pengguna yang ingin mengembangkan fungsionalitas lebih lanjut [7].

2.6 Algoritma Random Forest

Random forest pertama kali diperkenalkan oleh Breiman yang menggunakan pengacakan untuk membuat *Decision Tree*. Algoritma Random Forest merupakan pengembangan dari metode *Classification and Regression Tree* (CART) dengan menggabungkan teknik *bootstrap aggregating* (*bagging*) serta pemilihan fitur secara acak (*random feature selection*) [8]. Random Forest adalah metode klasifikasi yang bekerja dengan membangun sejumlah pohon keputusan. Teknik ini mampu meningkatkan akurasi dengan membangkitkan simpul anak pada setiap node secara berulang, kemudian memilihnya secara acak. [9].

Kemudian hasil klasifikasi dari setiap pohon diakumulasikan dan dipilh hasil klasifikasi yang paling banyak muncul[10]. Struktur metode ini terdiri atas *root node, internal node, dan leaf node. Root node* merupakan simpul paling atas yang berperan sebagai akar dari pohon keputusan. *Internal node* berfungsi sebagai simpul percabangan dengan satu input dan minimal dua

output. Sementara itu, *leaf node* atau *terminal node* adalah simpul akhir yang hanya memiliki satu input tanpa *output*. Proses pembentukan pohon keputusan diawali dengan perhitungan nilai entropy untuk menentukan tingkat ketidakmurnian atribut, serta nilai *information gain* sebagai dasar pemilihan atribut terbaik. Perhitungan entropy mengacu pada persamaan (1), sedangkan nilai information gain dihitung menggunakan persamaan (2).

$$Entropy(Y) = -\sum_i p\,(c|Y) \log 2p(c|Y) \quad (1)$$

Keterangan :

Y = Himpunan kasus

P(c|Y) = Proporsi nilai Y terhadap kelas c.

$$Informaton\ Gain(Y,a) = Entropy(Y) - \sum_{v \epsilon Values(a)} \frac{|Yv|}{Ya} Entropy(Y_v) \quad (2)$$

Keterangan :

Values(a) = Nilai yang mungkin dalam himpunan kasus a.

Y_v = Subkelas dari Y dengan kelas v yang berhubungan dengan kelas a.

Y_a = Semua nilai yang sesuai dengan a.

2.7 *Confusion Matrix*

Confusion matrix merupakan sebuah metode untuk evaluasi model klasifikasi yang berbentuk tabel matriks. *Confusion Matrix* ini umumnya digunakan untuk menilai kinerja dari *machine learning* dan dapat digunakan sebagai alat visual untuk mengevaluasi hasil klasifikasi [11].

Tabel 1. *Confusion Matrix*

		Nilai Aktual	
		Positif	Negatif
Nilai Prediksi	Positif	TP	FP
	Negatif	FN	TN

2.8 *Area Under Curve* (AUC)

Area Under Curve merupakan salah satu metrik yang umum digunakan untuk mengevaluasi performa model klasifikasi dalam machine learning. Metrik ini mengacu pada luas area di bawah kurva ROC (*Receiver Operating Characteristic*), yang menunjukkan hubungan antara true positive rate dan false positive rate pada berbagai ambang batas. Nilai AUC berada antara 0 hingga 1, dengan nilai yang mendekati 1 menunjukkan bahwa model mampu membedakan kelas positif dan negatif dengan sangat baik. Sebaliknya, nilai AUC yang mendekati 0,5 menunjukkan bahwa performa model hampir setara dengan tebakan acak. AUC sangat bermanfaat ketika data yang digunakan tidak seimbang, karena mampu menilai kinerja model secara menyeluruh tanpa bias terhadap proporsi kelas.

Menurut [12] yang dikutip oleh [13] Kelompok nilai AUC dapat diidentifikasi seperti pada Tabel 2.

Tabel 2. Klasifikasi Nilai AUC

Nilai AUC	**Klasifikasi**
0,90 – 1,00	Sangat baik
0,80 – 0,90	Baik
0,70 – 0,80	Cukup
0,60 – 0,70	Kurang
0,5 – 0,60	Keliru

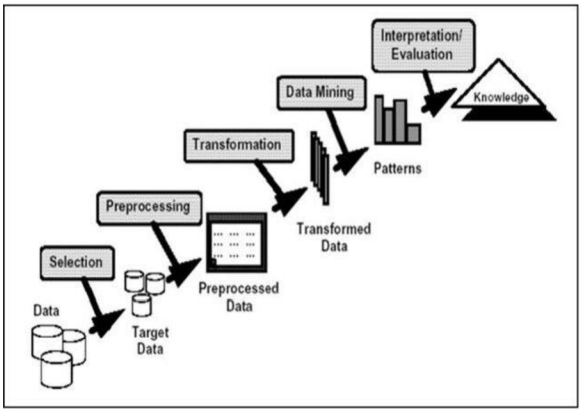
2.9 Penelitian Terdahulu

Dalam beberapa penelitian yang sudah dilakukan sebelumnya untuk dapat mengatasi masalah dengan berbagai macam metode. Penelitian yang dilakukan oleh [14] dimana hasil dalam penelitian untuk memprediksi penerimaan mahasiswa baru di Universitas Lambung Mangkurat didapatkan nilai akurasi sebesar 89,3%. Penelitian selanjutnya dicoba oleh [15] didapatkan dari hasil penelitian, mampu melakukan prediksi jumlah santri baru dengan tingkat akurasi yang sangat tinggi mencapai 97,73%. Kemudian penelitian yang dilakukan oleh [16] didapatkan hasil menganalisa data prediksi penerimaan salah satu perguruan tinggi di Medan yaitu Universitas XYZ dengan nilai akurasi sebesar 98%.

3 METODOLOGI

Penelitian ini menggunakan *Knowledge Discovery in Database*. *Knowledge Discovery in Databases* (KDD) proses dari menggunakan metode data mining untuk mencari informasi yang berharga, pola – pola yang ada dalam data, yang melibatkan algoritma untuk mengidentifikasi pola pada data.

Gambar 1. Tahapan KDD (*Knowledge Discovery in Database*)



1. *Data selection*
   Dalam tahapan pertama ini dilakukan permintaan data calon mahasiswa baru Universitas Singaperbangsa Karawang tahun 2024 kepada rektorat bidang akademik yang meliputi nama, hasil ujian masuk, kode prodi, program studi tujuan, asal kota, latar belakang pendidikan, dan daya tampung.
2. *Preprocessing*
   Tahap preprocessing melibatkan sejumlah proses yang bertujuan menyiapkan data agar siap digunakan pada tahap Data Mining. Proses ini dilakukan untuk memenuhi kebutuhan pengolahan data serta meningkatkan kualitas model dalam melakukan klasifikasi pada tahap Data Mining.
3. *Transformation*
   Transformasi data dilakukan untuk mengubah data ke dalam bentuk yang sesuai dengan kebutuhan analisis. Selanjutnya, tahap pemilihan fitur dilakukan untuk memilih fitur atau atribut yang paling relevan dengan kelas yang akan diprediksi.
4. *Data mining*
   Pada tahap ini, algoritma random forest akan digunakan untuk melakukan klasifikasi prediksi penerimaan pada data calon mahasiswa baru yang telah melewati tahap

preprocessing dan transformation. Perhitungan prediksi penerimaan akan dilakukan dan data akan diklasifikasikan menggunakan label yang telah diberikan.

5. Evaluation

Pada tahap terakhir, evaluasi, dilakukan untuk mengevaluasi model yang telah dibuat untuk mengukur akurasi yang dihasilkan saat melakukan klasifikasi dengan algoritma Random Forest. Metode *confusion matrix, accuracy, precision, recall, f1 score* dan AUC digunakan untuk mengevaluasi keberhasilan model.

4 HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini adalah data calon mahasiswa baru Universitas Singaperbangsa Karawang tahun 2024. Data yang terkumpul sebanyak 824 calon mahasiswa baru dari semua jurusan yang ada di Universitas Singaperbangsa Karawang. Data ini berisi nama, nilai snbt, kode prodi, jurusan tujuan, asal kota, asal sekolah dan daya tampung. Dari data tersebut yang akan digunakan sebagai atribut yaitu nama, nilai snbt, jurusan tujuan, asal sekolah, dan daya tampung. Sedangkan jurusan sebagai label. Selengkapnya bisa dilihat pada gambar berikut.

nama	nilai_snbt	jurusan_program_studi	asal_sekolah	daya_tampung
ABDULLAH AHMAD ASH-SHIDIQI	679.49	Agribisnis (S1)	SMA Swasta	60
ANDITA TIARA MAHARANI DUARSA	680.28	Agribisnis (S1)	MA Swasta	60
ARFAN MUHAMMAD RASYAD	511.31	Agribisnis (S1)	SMA Negeri	60
…	…	…	…	…

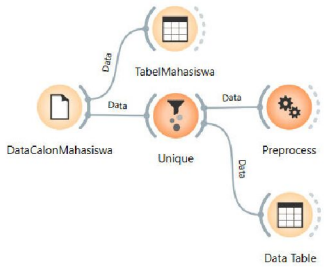
Gambar 2. Hasil Data Seleksi

Setelah data terkumpul, dilakukan pelabelan menggunakan atribut *nilai_snbt* berdasarkan *passing grade* tiap jurusan, di mana nilai di atas *passing grade* diberi label "peluang lolos" dan nilai di bawahnya diberi label "peluang tidak lolos".

nama	nilai_snbt	jurusan_program_studi	asal_sekolah	daya_tampung	passing_grade	peluang
Abdullah Ahmad Ash-Shidiqi	679.49	Agribisnis (S1)	SMA Swasta	60	613.79	lolos
Andita Tiara Maharani Duarsa	680.28	Agribisnis (S1)	MA Swasta	60	613.79	lolos
Arfan Muhammad Rasyad	511.31	Agribisnis (S1)	SMA Negeri	60	613.79	tidak lolos
…	…	…	…	…	…	..

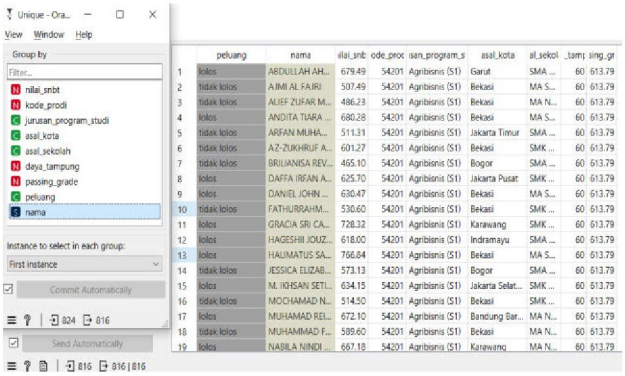
Gambar 3. Hasil Pelabelan Data Calon Mahasiswa Baru

Selanjutnya dilakukan tahap preprocessing yang bertujuan menyiapkan data untuk data mining dengan meningkatkan kualitasnya, dimulai dari data cleaning untuk menghapus kesalahan, duplikasi, data tidak lengkap, dan ketidakkonsistenan.



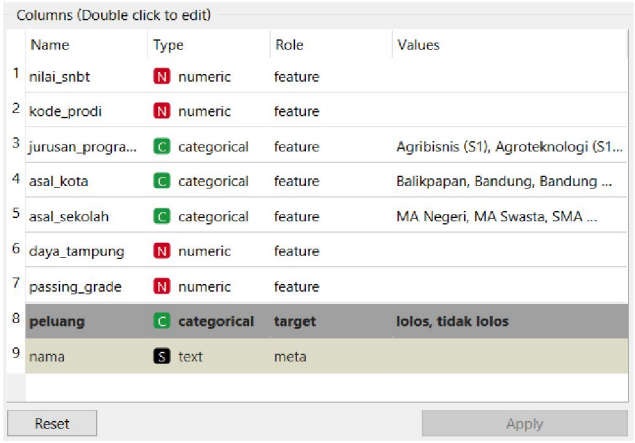
Gambar 4. Implementasi Proses Data Cleaning dalam Orange Data Mining

Hasil data cleaning menunjukkan tidak ada data tidak lengkap, namun ditemukan 8 duplikat berdasarkan atribut nama dari 824 data, sehingga tersisa 816 data bersih.



Gambar 5. Hasil data cleaning di Orange Data Mining

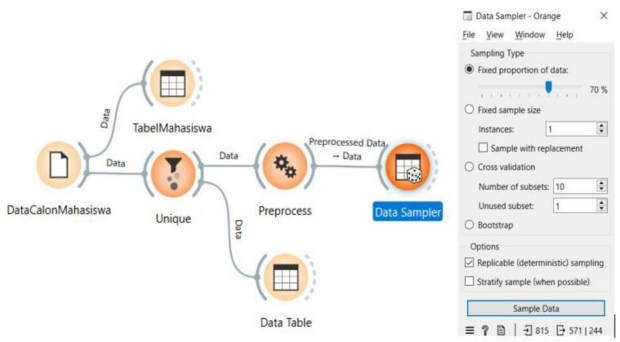
Dalam tahap transformasi dilakukan pengubahan data mentah menjadi format yang dapat diproses oleh algoritma random forest. Proses transformasi memastikan bahwa seluruh input data berada dalam format yang tepat dan ideal untuk proses klasifikasi random forest.



Gambar 6. Proses Pemilihan Tipe Input Pada Data Calon Mahasiswa Baru

Pada Gambar 6 atribut nama diberi tipe *text,* atribut jurusan_program_studi, asal_sekolah, asal_kota dan peluang dibeli tipe *categorical,* dan atribut nilai_snbt, kode_prodi, daya_tampung, dan passing_grade diberi tipe *numeric.*

Tahap *data mining* menggunakan algoritma *Random Forest* diawali dengan *data splitting* menjadi data latih dan data uji dalam empat skenario rasio: 90:10, 80:20, 70:30, dan 60:40.



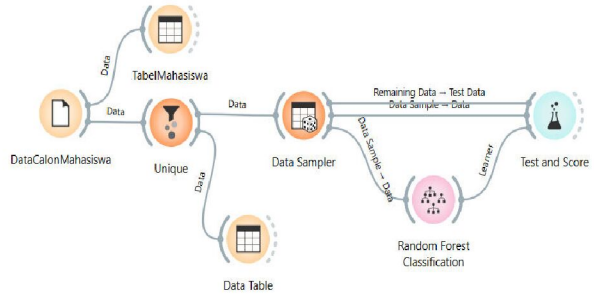
Gambar 7. Implementasi Proses *Split Data* di *Orange Data Mining*

Pada Gambar 7 data yang telah bersih dihubungkan dengan data sampler. Pada data sampler dipilih sampling type sesuai dengan skenario pengujian yang telah ditentukan. Jumlah data ulasan pada empat skenario pengujian dari hasil proses data splitting dapat dilihat pada Tabel 3.

Tabel 3. Hasil proses data splitting

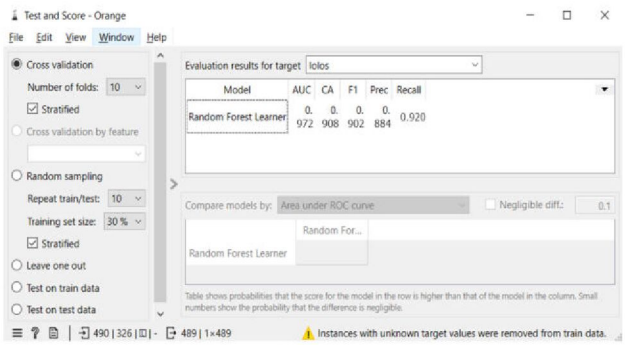
Keterangan	60:40	70:30	80:20	90:10
Data *Training*	490	572	653	735
Data *Testing*	326	244	163	81

Setelah melakukan data splitting, data akan diklasifikasi dengan menggunakan algoritma random forest. Proses implementasi algoritma random forest ini menggunakan empat pembagian data yang telah dilakukan sebelumnya.

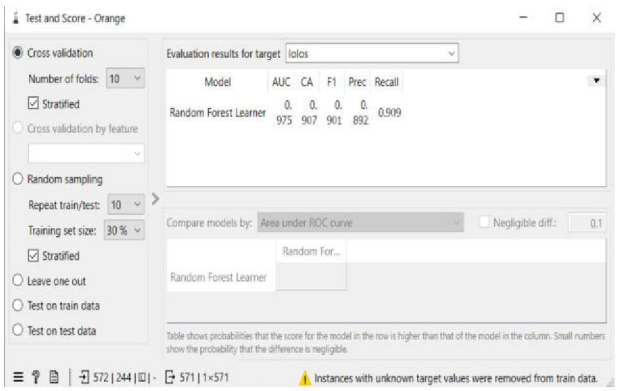


Gambar 8. Implementasi Algoritma *Random Forest* Dalam Orange Data Mining

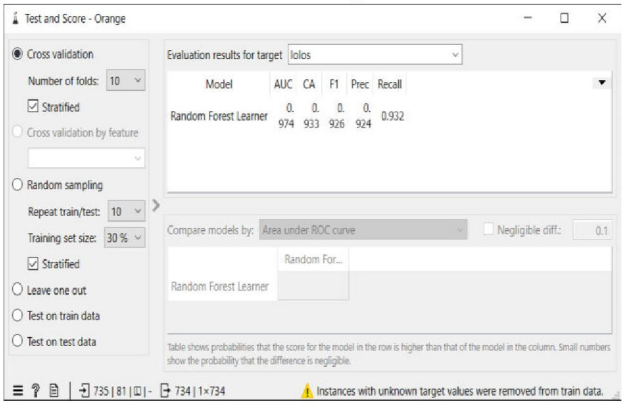
Pada Gambar 8 data yang sudah terbagi menjadi data training dan data splitting dihubungkan ke algoritma random forest dan hasil dari klasifikasi dilakukan dengan cross validation dengan 10 fold dan dapat dilihat dalam test and score.



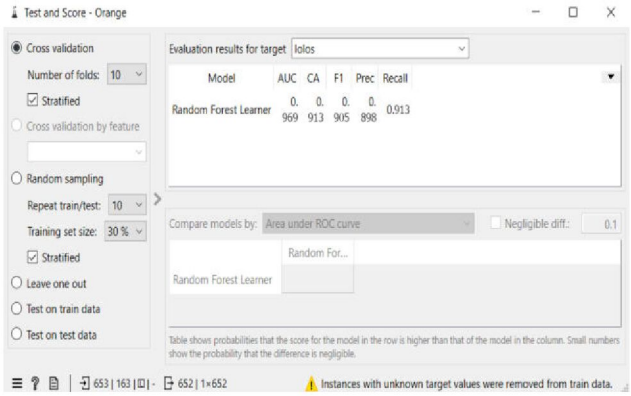
Gambar 9. Hasil dari klasifikasi pembagian data 60:40 dengan cross validation dengan 10 fold



Gambar 10. Hasil dari klasifikasi pembagian data 70:30 dengan cross validation dengan 10 fold



Gambar 11. Hasil dari klasifikasi pembagian data 80:20 dengan cross validation dengan 10 fold

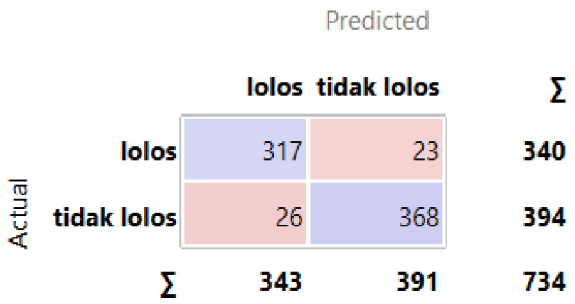


Gambar 12. Hasil dari klasifikasi pembagian data 90:10 dengan cross validation dengan 10 fold

Tabel 4. Hasil klasifikasi algoritma random forest dengan pembagian data

Pembagian data	AUC	CA	F1	Prec	Recall
60:40	0.972	0.908	0.902	0.884	0.920
70:30	0.975	0.907	0.901	0.892	0.909
80:20	0.969	0.913	0.905	0.898	0.913
90:10	0.974	0.933	0.926	0.924	0.932

Pada Tabel 4 diketahui hasil klasifikasi dengan penggunaan data dengan pembagian 90:10 menghasilkan nilai akurasi tertinggi sebesar 93.3%. Pada pembagian data lain hasil akurasi juga mendapatkan nilai yang tinggi dikarenakan sifat dari random forest ini sendiri dapat menghasilkan akurasi yang tinggi.



Gambar 13. *Confusion matrix* pembagian data 90:10

Berdasarkan tabel tersebut, terlihat bahwa algoritma *Random Forest* mampu memberikan performa yang baik dalam proses klasifikasi pada ketiga skenario, ditunjukkan oleh tingginya nilai accuracy, precision, recall, f1-score, dan AUC. AUC mengindikasikan bahwa seluruh skenario pembagian data berada pada kategori "Sangat Baik".

5 KESIMPULAN DAN SARAN

Berdasarkan uji yang telah dilakukan, prediksi penerimaan mahasiswa baru dengan algoritma *random forest* melalui *Orange data mining*. Proses klasifikasi menggunakan empat skenario pembagian data (90:10, 80:20, 70:30, 60:40), skenario terbaik diperoleh pada rasio 90:10 antara data training dan testing, yang menghasilkan nilai accuracy sebesar 0.933, precision sebesar 0. 924, recall sebesar 0. 932, f1-score sebesar 0.928 dan nilai AUC sebesar 0.974 yang mengindikasikan bahwa data berada pada kategori "Sangat Baik".

UCAPAN TERIMA KASIH

Pada kesempatan kali ini penulis menyampaikan ucapan terima kasih kepada:

a. Seluruh civitas akademika Universitas Singaperbangsa Karawang.
b. Dosen Pembimbing yang telah memberikan arahan serta bimbingan.

DAFTAR PUSTAKA

[1] Badan Pusat Statistik Provinsi Jawa Barat, "Angka Partisipasi Kasar (APK) Provinsi Jawa Barat menurut Jenjang Pendidikan," Badan Pusat Statistik. Accessed: Jun. 30, 2025. [Online]. Available: https://jabar.bps.go.id/id/statistics-table/2/OTI2IzI=/angka-partisipasi-kasar-apk-provinsi-jawa-barat-menurut-jenjang-pendidikan.html

[2] I. Nawangsih, I. Melani, S. Fauziah, and A. I. Artikel, "pelita teknologi prediksi pengangkatan karyawan dengan metode algoritma c5.0 (studi kasus pt. Mataram cakra buana agung," *Jurnal Pelita Teknologi*, vol. 16, no. 2, pp. 24–33, 2021.

[3] Republik Indonesia, *Undang-Undang Republik Indonesia Nomor 12 Tahun 2012 tentang Pendidikan Tinggi*. Indonesia, 2012.

[4] Direktorat Jenderal Pendidikan Tinggi, "Kerangka Kualifikasi Nasional Indonesia dan Pendidikan Tinggi," Jakarta, 2020.

[5] Kementerian Pendidikan dan Kebudayaan Republik Indonesia, *Peraturan Mendikbud No. 6 Tahun 2020 tentang Penerimaan Mahasiswa Baru Program Sarjana pada Perguruan Tinggi Negeri*. 2020.

[6] Aurélien Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (2nd ed.)*, 2nd Edition. O'Reilly Media, 2019.

[7] J. Demšar *et al.*, "Orange: Data Mining Toolbox in Python Tomaž Curk Matija Polajnar Laň Zagar," 2013.

[8] L. Breiman, "Random Forests," *Mach Learn*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[9] S. Devella, Y. Yohannes, and F. N. Rahmawati, "Implementasi Random Forest Untuk Klasifikasi Motif Songket Palembang Berdasarkan SIFT," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 7, no. 2, pp. 310–320, Aug. 2020, doi: 10.35957/jatisi.v7i2.289.

[10] F. Amalia Kurniawan and A. Prima Kurniati, "Analisis Dan Implementasi Random Forest Dan Classification Dan Regression Tree (Cart) Untuk Klasifikasi Pada Misuse Intrusion Detection System Tugas Akhir-2011 Fakultas Teknik Informatika Program Studi S1 Teknik Informatika." [Online]. Available: www.tcpdf.org

[11] S. A. Putri, N. Selayanti, M. Kristanaya, M. P. Azzahra, M. G. Navsih, and K. M. Hindrayani, "Penerapan Machine Learning Algoritma Random Forest Untuk Prediksi Penyakit Jantung," *Seminar Nasional Sains Data*, vol. 2024, [Online]. Available: https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction.

[12] F. Gorunescu, *Data Mining*, vol. 12. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. doi: 10.1007/978-3-642-19721-5.

[13] Y. Handayani, T. Hidayat, D. Novitaningrum, A. Rahman Ismail, U. Selamat, and J. Tengah, "Perbandingan Algoritma Logistic Regression Dan Naïve Bayes Classifier Dalam Identifikasi Penyakit Liver," 2025. [Online]. Available: http://jurnal.goretanpena.com/index.php/JSSR

[14] V. Karina, "Application Of Random Forest Method To Prediction Of Student Candidates Accepted In The Snmptn Pathway (Case Study At Universitas Lambung Mangkurat)," 2024.

[15] L. Hidayah and M. I. Rosadi, "Penerapan Algoritma Random Forest Untuk Memprediksi Jumlah Santri Baru," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3S1, Oct. 2024, doi: 10.23960/jitet.v12i3S1.5237.

[16] M. Rianto and R. Yunis, "Analisis Runtun Waktu Untuk Memprediksi Jumlah Mahasiswa Baru Dengan Model Random Forest," *Paradigma - Jurnal Komputer dan Informatika*, vol. 23, no. 1, Mar. 2021, doi: 10.31294/p.v23i1.9781.

BIODATA PENULIS

Mochammad Fitra Pamungkas
Merupakan Mahasiswa Program studi Informatika Fakulats Ilmu Komputer di Universitas Singaperbangsa Karawang.

Betha Nurina Sari
Merupakan Dosen Fakultas Ilmu Komputer di Universitas Singaperbangsa Karawang

Iqbal Maulana
Merupakan Dosen Fakultas Ilmu Komputer di Universitas Singaperbangsa Karawang