



INNOVATIVE: Journal Of Social Science Research

Volume 3 Nomor 4 Tahun 2023 Page 10065-10079

E-ISSN 2807-4238 and P-ISSN 2807-4246

Website: <https://j-innovative.org/index.php/Innovative>

## Penerapan Data Mining Untuk Memprediksi Penerimaan Peserta Didik Baru Jalur Prestasi Akademik Di SMA Negeri 13 Jakarta Dengan Menggunakan Algoritma Random Forest

Rasiban<sup>1✉</sup>, Samuel Praja Raymond Maruli<sup>2</sup>

Teknik Informatika, STIKOM Cipta Karya Informatika

Email: [rasiban@gmail.com](mailto:rasiban@gmail.com)<sup>1✉</sup>, [Samuel.praja97@gmail.com](mailto:Samuel.praja97@gmail.com)<sup>2</sup>

### Abstrak

Indonesia sebagai salah satu negara berkembang memiliki tujuan yaitu meningkatkan kualitas pendidikan. Salah satu faktor penentu kualitas pendidikan adalah penerimaan peserta didik baru di sekolah. SMA Negeri 13 Jakarta, yang memiliki daya tampung terbatas, banyak diminati calon siswa yang ingin melanjutkan pendidikan ke tingkat SMA. Menggunakan algoritma Random Forest, penelitian ini bertujuan untuk memprediksi sekolah-sekolah tingkat SMP yang memiliki peluang tertinggi diterima di SMA Negeri 13 Jakarta melalui jalur prestasi akademik. Prediksi yang akurat dapat membantu pengambilan keputusan dan perencanaan pendidikan, menjadi acuan bagi orang tua siswa yang berencana mengirimkan anaknya ke SMA Negeri 13 Jakarta melalui jalur prestasi akademik. Dengan tingkat akurasi prediksi sebesar 90.67%, model mampu memberikan hasil prediksi yang akurat dan tingkat kesalahan yang rendah sebesar 9.33%, menandakan kualitas model yang baik dalam melakukan prediksi. Nilai kappa sebesar 0.883 juga menunjukkan kualitas prediksi yang tinggi, dengan tingkat kesepakatan yang baik. Berdasarkan hasil analisis, beberapa SMP dengan peluang tertinggi untuk diterima melalui jalur prestasi akademik di SMA Negeri 13 Jakarta adalah SMP Negeri 30 Jakarta, SMP Negeri 173 Jakarta, SMP Negeri 95 Jakarta, SMP Negeri 129 Jakarta, dan SMP KK Marsudirini.

Kata Kunci: *Data Mining, Random Forest, Prediksi, Penerimaan Peserta Didik Baru*

### Abstract

Indonesia, as one of the developing countries, has the goal of improving the quality of education. One of the factors determining the quality of education is the admission of new students into schools. SMA Negeri 13 Jakarta, which has limited capacity, is in great demand by prospective students who want to continue their education at the high school level. Using the Random Forest algorithm, this study aims to predict the junior high schools that have the highest chance of being admitted to SMA Negeri 13 Jakarta through the academic achievement pathway. Accurate predictions can help with decision-making and educational planning, becoming a reference for parents who plan to send their children to SMA Negeri 13 Jakarta through the academic achievement pathway. With a prediction accuracy rate of 90.67%, the model is able to provide accurate prediction results with a low error rate of 9.33%, indicating good model quality in making predictions. The kappa value of 0.883 also indicates high prediction quality with a good level of agreement. Based on the results of the analysis, some of the junior high schools with the highest chance of being accepted through the academic achievement pathway at SMA Negeri 13 Jakarta are SMP Negeri 30 Jakarta, SMP Negeri 173 Jakarta, SMP Negeri 95 Jakarta, SMP Negeri 129 Jakarta, and SMP KK Marsudirini.

Keywords: *Data Mining, Random Forest, Prediction, New Student Admission.*

### PENDAHULUAN

Pendidikan adalah salah satu aspek penting dalam pembangunan suatu negara. Indonesia sebagai salah satu negara berkembang memiliki tujuan yang sama yaitu meningkatkan kualitas pendidikan. Salah satu faktor penentu kualitas pendidikan adalah penerimaan peserta didik baru di sekolah. Oleh karena itu, setiap tahunnya pemerintah Indonesia selalu mengadakan penerimaan peserta didik baru (PPDB) secara online untuk mengisi kursi kosong di sekolah-sekolah. SMA Negeri 13 Jakarta adalah salah satu sekolah menengah atas di Jakarta Utara yang memiliki daya tampung terbatas dan banyak diminati oleh calon siswa yang ingin melanjutkan pendidikan di tingkat SMA. Dalam proses penerimaan peserta didik baru, di SMA Negeri 13 Jakarta memiliki beberapa jalur penerimaan, salah satunya adalah jalur prestasi akademik. Pada jalur ini, calon siswa yang memiliki prestasi akademik yang baik di SMP asalnya akan memiliki kesempatan lebih besar untuk diterima di SMA Negeri 13 Jakarta.

Data mining adalah suatu teknologi yang dapat mengolah dan menganalisis data dari berbagai sumber untuk menemukan pola atau hubungan yang tersembunyi dalam data (Saikin & Kusri, 2019). Dalam kasus ini, data mining akan digunakan untuk memprediksi SMP

mana saja yang memiliki peluang tertinggi dalam penerimaan peserta didik baru jalur prestasi akademik di SMA Negeri 13 Jakarta. Pada dasarnya, prediksi dilakukan berdasarkan data historis yang dianalisis menggunakan metode tertentu untuk mendapatkan gambaran tentang keadaan di masa yang akan datang, sehingga dapat membantu proses pengambilan keputusan (Fahreza, 2022). Metode yang akan digunakan adalah algoritma Random Forest.

Algoritma Random Forest mampu menghasilkan prediksi yang akurat dengan membagi dataset menjadi banyak pohon keputusan dan mengambil hasil dari mayoritas prediksi dari setiap pohon keputusan tersebut (Latifah et al., 2019). Penggunaan algoritma Random Forest dapat mempercepat proses analisis data, sehingga dapat memprediksi hasil seleksi penerimaan peserta didik baru jalur prestasi akademik dengan lebih efisien.

Dalam beberapa penelitian sebelumnya, algoritma Random Forest mampu mengklasifikasikan dengan baik dan memiliki performa akurasi yang baik. Dalam penelitian (Pratama et al., 2019), peneliti mengusulkan untuk mengumpulkan data dan memprediksi masuk SMP Negeri di kota Cimahi menggunakan Random Forest. Metode ini juga mempertimbangkan faktor pendukung calon siswa seperti zonasi dan jalur prestasi. Akurasi performansi metode Random Forest pada penelitian ini mencapai lebih dari 90%. Dalam penelitian (Saleh et al., 2022), peneliti membangun sebuah sistem klasifikasi kelulusan mahasiswa menggunakan algoritma Random Forest dengan fitur-fitur yang telah ditentukan. Nilai akurasi meningkat seiring dengan menaikkan penggunaan jumlah pohon, yaitu dengan nilai akurasi 90.00% dengan menggunakan 50 pohon.

Pada penelitian (Triwidianti et al., 2021), peneliti melakukan penelitian dengan menggunakan tiga metode pada data mining untuk memprediksi prestasi siswa tingkat pendidikan menengah kejuruan dengan metode Decision Tree, Naive Bayes, dan Random Forest. Hasil penelitian menunjukkan bahwa metode yang terbaik dari penelitian ini adalah metode Random Forest dengan tingkat akurasi 97.3%.

Peneliti dalam penelitian (Mu'alim & Hidayati, 2022) menemui sebuah kendala yang dihadapi MAN Sintang untuk menentukan jurusan siswa di MAN Sintang, yaitu waktu yang lama dan kurang akurat untuk mendapatkan jurusan yang sesuai dengan kemampuan siswa dikarenakan cara perhitungan yang masih manual. Oleh sebab itu, peneliti membuat sebuah sistem klasifikasi penjurusan siswa berupa aplikasi berbasis website. Hasil pengujian yang diperoleh menggunakan confusion matrix menghasilkan nilai accuracy sebesar 94.38%, precision sebesar 95.23% dan recall sebesar 94.25%.

Dalam penelitian (Rianto & Yunis, 2021), peneliti menggunakan model Random Forest untuk memprediksi jumlah mahasiswa baru di Universitas XYZ pada periode 2010-2019. Hasil

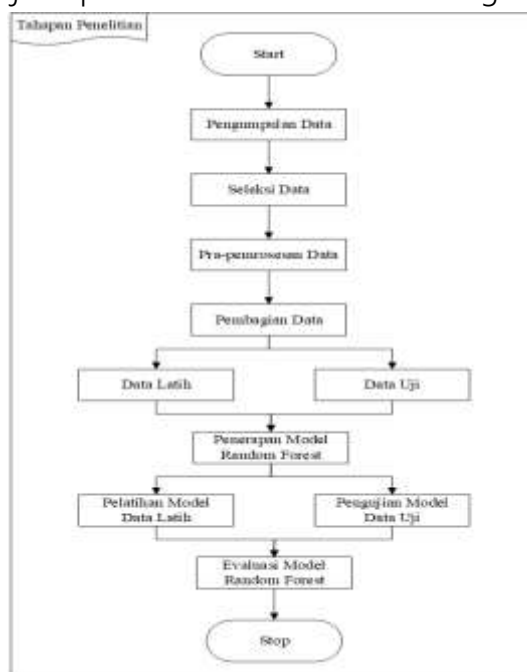
penelitian menunjukkan bahwa model yang dihasilkan memiliki tingkat akurasi sebesar 99.8%, Parameter terbaik dari model yang digunakan adalah maxnodes 100 dan ntree 900, dan terdapat trend penurunan jumlah mahasiswa untuk beberapa tahun ke depan. Sedangkan dalam penelitian (Linawati et al., 2020), peneliti mengusulkan teknik data mining yang efisien untuk memprediksi prestasi akademik mahasiswa dengan teknik klasifikasi dari data mining yaitu Decision Tree C4.5 dan Random Forest. Random Forest menghasilkan akurasi 92.4%. Sedangkan, hasil penggunaan algoritma Decisison Tree C4.5 menghasilkan akurasi 87.1%.

Penelitian ini diharapkan memberikan kontribusi signifikan dalam bidang pendidikan, khususnya terkait penerimaan peserta didik baru melalui jalur prestasi akademik. Selain itu, diharapkan penelitian ini dapat membantu siswa dan orang tua dalam mempersiapkan diri dalam memilih SMP yang dapat meningkatkan peluang siswa tersebut diterima di SMA Negeri 13 Jakarta melalui jalur prestasi akademik. Hasil penelitian ini dapat menjadi pedoman bagi sekolah dalam meningkatkan kualitas pendidikan, terutama di tingkat menengah pertama, dengan tujuan meningkatkan peluang siswa diterima di sekolah unggulan melalui jalur prestasi akademik. Selain itu, diharapkan penelitian ini dapat memberikan kontribusi pada penelitian serupa dalam bidang data mining di Indonesia, terutama dalam penggunaan algoritma Random Forest untuk memprediksi peluang berdasarkan data yang tersedia.

## METODE PENELITIAN

### Tahapan Penelitian

Berikut merupakan perancangan pada penelitian untuk memprediksi penerimaan peserta didik baru melalui jalur prestasi akademik di SMA Negeri 13 Jakarta.



## Gambar 1. Tahapan Penelitian

### Pengumpulan Data

Data yang digunakan penelitian ini diperoleh dari Dinas Pendidikan Provinsi DKI Jakarta melalui situs PPDB online dan situs arsipnya. Data merupakan data hasil PPDB online dari tahun ajaran 2018/2019 hingga tahun ajaran 2023/2024 dengan total sampel data sebanyak 800 data. Dalam tahun ajaran 2018/2019, data yang digunakan adalah data hasil seleksi PPDB online jalur Domisili Dalam DKI Tahap 1 Lokal, Tahap 2 Umum dan Tahap 3 Umum.

### Seleksi Data

Proses seleksi data, atau yang dikenal sebagai data selection, melibatkan analisis data yang relevan dari dataset karena tidak semua data diperlukan dalam proses data mining. Data yang dipilih dan diseleksi dari dataset akan dijadikan objek analisis. Dalam penelitian ini, data yang terpilih mencakup nomor daftar sebagai id masing-masing calon siswa pendaftar yang diterima, nilai akhir, asal sekolah calon siswa pendaftar yang diterima, dan hasil seleksi penerimaan calon siswa pendaftar.

### Pra-pemrosesan Data

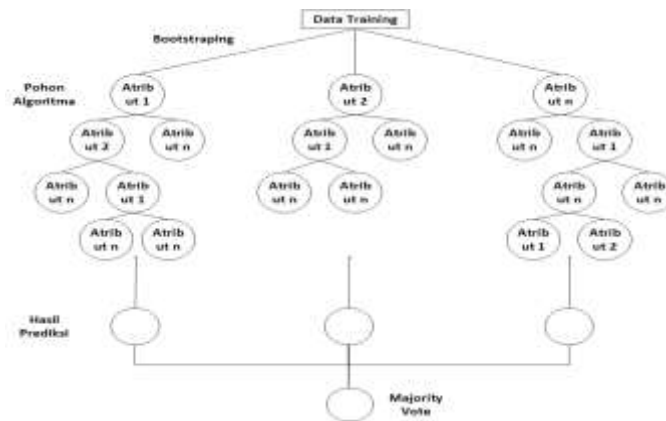
Dalam penelitian ini, yang dilakukan merupakan penyesuaian data serta normalisasi atau standarisasi pada atribut yang memiliki skala yang berbeda untuk memastikan semua atribut memiliki pengaruh yang seimbang dalam pembentukan model. Dari atribut yang ada, peneliti melakukan penyesuaian data menjadi bentuk yang disesuaikan untuk proses data mining.

### Pembagian Data

Dari 800 data, akan dibagi menjadi dua bagian, yaitu data latih dan data uji. Pembagian data dilakukan dengan perbandingan data latih sebesar 70%, yang berjumlah 575 data, dan data uji sebesar 30%, yang berjumlah 225 data.

### Penerapan Model Random Forest

Menerapkan algoritma Random Forest sebagai metode utama untuk membangun model prediksi. Random Forest adalah metode ensemble yang menggabungkan beberapa pohon keputusan (decision trees) untuk melakukan prediksi. Setiap pohon dalam Random Forest dibangun dengan memilih subset acak dari data latih dan subset acak dari atribut yang akan digunakan dalam pemilihan keputusan. Proses pembentukan hutan dilakukan dengan menggabungkan hasil dari pohon-pohon keputusan yang terbentuk.



Gambar 2. Kumpulan hutan dari Decision Tree yang dibuat oleh Random Forest

### Evaluasi Model Random Forest

Penerapan model diterapkan terlebih dahulu pada data latih, untuk menguji performa model yang telah dilatih. Lalu, model yang telah dilatih, diaplikasikan pada data uji. Setelah model dibangun, evaluasi model dilakukan dengan menghitung metrik evaluasi kinerja model untuk polynominal classification, seperti accuracy, classification error, dan nilai Kappa. Dengan analisis hasil pengujian dan validasi, dapat diukur keakuratan dan kehandalan prediksi model Random Forest dalam memprediksi siswa SMP dengan peluang tertinggi dalam diterima di SMA Negeri 13 Jakarta melalui jalur prestasi akademik.

### HASIL DAN PEMBAHASAN

Pengujian metode Random Forest dalam penelitian ini menggunakan tools RapidMiner Studio. Pembahasan serta tampilan dari tools RapidMiner Studio adalah sebagai berikut:

#### Atribut dan tipe atribut

Informasi data atribut dan tipe atribut pada RapidMiner yang digunakan pada dataset penelitian ini dapat dilihat pada Tabel 1. Dan contoh dataset yang digunakan dalam penelitian ini bisa dilihat dalam Gambar 3.

Tabel 1. Atribut dan Tipe Atribut pada Dataset

Nama Atribut	Tipe Atribut	Jenis Atribut
no_daftar	Real	Special (id)
asal_sekolah	Polynominal	Special (label)
nilai_akhir	Real	Regular
hasil_seleksi	Polynominal	Regular
accepted_per_year	Integer	Regular
accepted_all	Integer	Regular
avg_accepted_per_year	Real	Regular
avg_accepted_all	Real	Regular
avg_nilai_akhir_pt	Real	Regular

Row No.	nr_dokter	asal_kesehatan	umur_dokter	jenis_kelamin	accepted_p...	accepted_ab	img_accepted...	img_accepted...	img_misi_ab...
1	2001830041T	SMP Negeri 3	83	MPa	115	317	0.448	0.306	85.795
2	2001830041T	SMP Negeri 3	83	MPa	115	317	0.448	0.306	85.795
3	2001830041T	SMP Negeri 3	82.638	MPa	115	317	0.448	0.306	85.795
4	2001830041T	SMP Negeri 3	82	MPa	115	317	0.448	0.306	85.795
5	2001830041T	SMP Negeri 1	81.750	MPa	1	1	0.004	0.001	81.750
6	2001830041T	SMP Negeri 1	81.750	MPa	115	317	0.448	0.306	85.795
7	2001830041T	SMP Negeri 1	81.838	MPa	115	317	0.448	0.306	85.795
8	2001830041T	SMP Negeri 1	81.838	MPa	115	317	0.448	0.306	85.795
9	2001830041T	SMP Negeri 1	81.500	MPa	115	317	0.448	0.306	85.795
10	2001830041T	SMP Negeri 1	81.500	MPa	115	317	0.448	0.306	85.795
11	2001830041T	SMP Negeri 1	81.200	MPa	115	317	0.448	0.306	85.795
12	2001830041T	SMP Negeri 2	81.138	MPa	8	28	0.038	0.031	87.281
13	2001830041T	SMP Negeri 2	81.138	MPa	7	15	0.027	0.018	87.083
14	2001830041T	SMP Negeri 2	81	MPa	5	18	0.028	0.024	86.952
15	2001830041T	SMP Negeri 2	81	MPa	14	28	0.055	0.035	86.408
16	2001830041T	SMP Negeri 2	80.888	MPa	1	1	0.004	0.001	80.888

Gambar 3. *Dataset*

## Select Attributes

Memilih atribut yang akan digunakan dalam proses perhitungan *Random Forest* yang digunakan dalam membentuk pohon-pohon keputusan.



Gambar 4. Select Attributes

## Pembagian Data

Melakukan pembagian *dataset* menjadi data latih dan data uji, dengan rasio 0.7:0.3.



Gambar 5. Pembagian rasio *dataset*

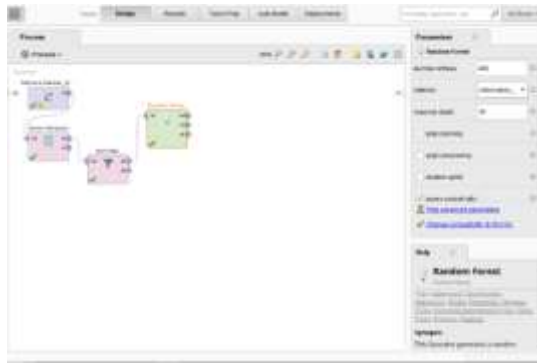
## Penerapan Model *Random Forest*

Berikut langkah penerapan model *Random Forest* dengan bantuan *tools* RapidMiner.

- a) Pada *tab Operators*, lalu cari *Random Forest*, drag dan drop ke *view Process*, hubungkan *output partition 1* dari operator *Split Data* (data latih) ke *input operator Random Forest*.
- b) Tentukan jumlah pohon acak yang akan dihasilkan melalui proses *bootstrapping* pada *tab Parameters* pada bagian *number of trees*. Semakin banyak pohon yang digunakan, semakin lama waktu yang dibutuhkan untuk melatih model. Dalam banyak kasus, jumlah pohon (*number of trees*) yang umumnya dianjurkan untuk *Random Forest* adalah sekitar 100 hingga 500 pohon. Pilih jumlah pohon yang memberikan kinerja yang baik tanpa mengorbankan waktu pelatihan yang berlebihan. Dalam penelitian ini, penulis menggunakan jumlah pohon sebanyak 400 pohon, hal ini dikarenakan *trade-off* antara waktu pelatihan dan kinerja model dengan 400 pohon sangat baik, memiliki tingkat akurasi yang sama dengan yang dihasilkan 500 pohon, akan tetapi memiliki waktu pelatihan yang relatif lebih singkat jika dibandingkan dengan 500 pohon. Setelah selesai menentukan jumlah pohon, langkah selanjutnya ialah memilih *criterion* yang ingin digunakan.
- c) *Criterion* menentukan cara mengukur kualitas pemisahan pada setiap *node* dalam pohon keputusan. Pada dasarnya, *criterion* akan memperkirakan seberapa baik pemisahan atribut tersebut dalam menghasilkan *subset* data yang homogen atau murni, di mana setiap *subset* tersebut terdiri dari satu kelas data atau kelas yang sangat mirip. Secara umum, *Random Forest* menggunakan "*Gini impurity*" (*gini\_index*) atau "*Entropy*" (*information\_gain*) sebagai kriteria default untuk pemilihan atribut. Pada penelitian ini, penulis menggunakan *criterion information\_gain*, hal ini dikarenakan kriteria ini bekerja dengan baik ketika target atau label memiliki banyak kelas (*polynominal*). *Entropy* mengukur ketidakpastian dalam *dataset*, dan dengan memiliki banyak kelas, perhitungan *entropy* akan memberikan informasi yang baik tentang seberapa baik atribut tertentu dapat memisahkan data menjadi kelompok-kelompok yang homogen.
- d) Setelah memilih *criterion*, langkah selanjutnya adalah menentukan *maximal depth*. Parameter ini mengontrol seberapa dalam pohon keputusan dapat tumbuh selama proses pembangunan model. Penulis menetapkan kedalaman maksimal yang digunakan dalam penelitian ini adalah 10, angka tersebut dipilih dikarenakan untuk memberikan keseimbangan antara kompleksitas model dan performa. Pada operator ini, *Random Forest* akan diaplikasikan untuk mengelola



dan melatih data latih, yang nantinya hasil pelatihan tersebut bisa diaplikasikan pada data uji.

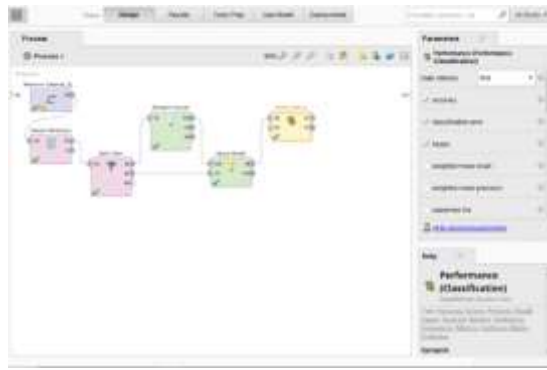


Gambar 6. Penerapan Model *Random Forest*

#### Evaluasi Model *Random Forest*

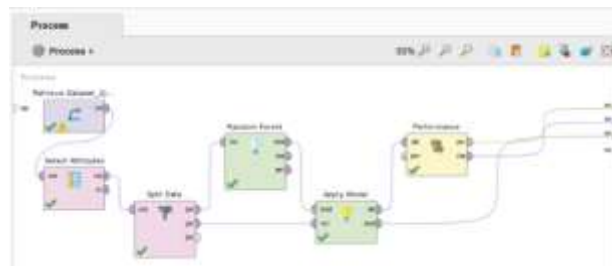
Berikut adalah langkah-langkah untuk mengevaluasi model *Random Forest* menggunakan *tools* RapidMiner.

- a) Pada *tab Operators*, cari *performance (classification)*, drag dan drop ke *view Process*, hubungkan *output labelled* pada operator *Apply Model* ke *input labelled* operator *Performance*. Pada *tab Parameters*, karena dalam memprediksi label asal\_sekolah yang merupakan *polynomial classification*, maka penulis memilih kriteria *accuracy*, *classification error* dan *kappa*. Operator ini berguna untuk mengevaluasi performa statistik dalam tugas klasifikasi. Hasil dari operator ini berupa daftar nilai kriteria performa untuk tugas klasifikasi tersebut. *Accuracy* adalah metrik yang paling sederhana dan umum digunakan dalam mengukur persentase prediksi yang benar dari keseluruhan prediksi yang dilakukan oleh model. Sedangkan, *classification error* adalah kebalikan dari *accuracy* dan mengukur persentase prediksi yang salah dari keseluruhan prediksi yang dilakukan oleh model. *Kappa* adalah metrik evaluasi yang sangat berguna dalam *polynomial classification*, terutama ketika ada ketidakseimbangan kelas dalam data. *Kappa* mengukur sejauh mana model dapat mencapai tingkat kesepakatan yang lebih baik daripada yang diharapkan secara acak.



Gambar 7. Tampilan operator *Performance (Performance (Classification))* pada *view Process*

b) Dan yang terakhir, hubungkan *output performance* pada operator *Performance (Performance (Classification))* ke *input result 1*, dan *output example set* ke *input result 2* untuk mengetahui hasil dari data uji beserta *performance*-nya. Hubungkan *output model* *Apply Model* ke *result 3* untuk melihat hasil data uji.



Gambar 8. Tampilan *operators* pada *view Process*

## Analisis Hasil Prediksi

### *Accuracy criterion*

Akurasi adalah metrik yang mengukur seberapa akurat model dalam memprediksi kelas yang benar dari semua prediksi yang dilakukan. Rumus akurasi adalah:

$$\text{Akurasi} = \frac{\text{Jumlah prediksi benar}}{\text{Total data}}$$

Dalam penelitian ini, hasil penghitungan akurasi untuk data uji adalah 90.67%, yang berarti model benar-benar memprediksi 90.67% data dengan benar.

	True SGP	True SGP	True SGP	True SGP	True SGP	True SGP
pred. SGP	88	0	0	0	0	0
pred. SGP	0	0	10	0	0	0
pred. SGP	0	0	5	0	0	0
pred. SGP	0	0	0	4	0	0
pred. SGP	0	0	0	0	5	0

Gambar 9. Hasil *accuracy*

### *Classification error criterion*

*Error* klasifikasi adalah kebalikan dari akurasi. Metrik ini mengukur seberapa banyak prediksi yang salah dibandingkan dengan jumlah total prediksi. Rumus error klasifikasi

adalah:

$$\text{Error Klasifikasi} = \frac{\text{Jumlah prediksi salah}}{\text{Total data}}$$

Dalam penelitian ini, error klasifikasi yang dimiliki saat menguji data uji adalah 9.33%, yang berarti model salah memprediksi 9.33% data.



classification\_error: 9.33%

	true SMP...	true SMP...	true SMP...	true SMP...	true SMP...	true
pred. SM...	95	0	0	0	0	0
pred. SM...	0	0	0	0	0	0
pred. SM...	0	0	5	0	0	0
pred. SM...	0	0	5	4	0	0
pred. SM...	0	0	0	0	5	0

Gambar 10. Hasil *classification error*

### Kappa criterion

*Kappa*, juga dikenal sebagai *Cohen's kappa*, adalah metrik yang mengukur tingkat kesepakatan antara prediksi model dengan hasil sebenarnya, dan mengatasi masalah ketidakseimbangan kelas. Nilai kappa berkisar dari -1 hingga 1, dengan nilai 1 menunjukkan kesepakatan sempurna, 0 menunjukkan kesepakatan yang sama dengan kebetulan, dan nilai negatif menunjukkan kesepakatan yang kurang baik. Semakin tinggi nilai kappa, maka model klasifikasi dapat dipercayai untuk melakukan prediksi pada data yang belum pernah dilihat sebelumnya. Rumus kappa adalah:

$$Kappa = \frac{P_o - P_e}{1 - P_e}$$

Di mana  $P_o$  adalah *observed agreement*, yaitu persentase kesepakatan antara prediksi model dan nilai sebenarnya pada *dataset*. Sedangkan,  $P_e$  adalah *expected agreement*, yaitu persentase kesepakatan yang diharapkan secara acak. Dalam penelitian ini, nilai *kappa* yang dicapai dalam pengujian data uji adalah 0.883, yang berarti model memiliki tingkat kesepakatan yang sangat baik dalam memprediksi data pada data uji.



kappa: 0.883

	true SMP...	true SMP...	true SMP...	true SMP...	true SMP...	true
pred. SM...	95	0	0	0	0	0
pred. SM...	0	0	0	0	0	0
pred. SM...	0	0	5	0	0	0
pred. SM...	0	0	0	4	0	0
pred. SM...	0	0	0	0	5	0

Gambar 11. Hasil *kappa*

### Hasil Prediksi

Berdasarkan hasil keputusan yang dibuat *Random Forest*, maka menghasilkan hasil prediksi untuk *label/asal\_sekolah*. SMP Negeri 30 Jakarta merupakan nilai atribut yang paling banyak muncul, yaitu sebanyak 95 kali dengan fraksi sebesar 0.4222 dalam *dataset*. Diikuti

dengan SMP Negeri 173 Jakarta sebanyak 21 kali dengan fraksi sebesar 0.0933 dalam *dataset*. SMP Negeri 95 Jakarta di posisi ketiga, dengan kemunculan dalam *dataset* sebanyak 15 kali dengan fraksi sebesar 0.0667. Untuk hasil keseluruhan, dapat dilihat pada pada Tabel 2.

Tabel 2. *Nominal Values*

Index	Nominal value	Absolute Count	Fraction
1	SMP Negeri 30 Jakarta	95	0.4222
2	SMP Negeri 173 Jakarta	21	0.0933
3	SMP Negeri 95 Jakarta	15	0.0667
4	SMP Negeri 129 Jakarta	9	0.0400
5	SMP KK Marsudirini	8	0.0356
6	SMP Negeri 244 Jakarta	7	0.0311
7	SMPK Tunas Bangsa Sunter	6	0.0267
8	SMP Negeri 231 Jakarta	5	0.0222
9	SMP Negeri 277 Jakarta	5	0.0222
10	SMP Barunawati 3	4	0.0178
11	SMP Hang Tuah 1	4	0.0178
12	SMP Hang Tuah 3	4	0.0178
13	SMP Negeri 114 Jakarta	4	0.0178
14	SMP Negeri 84 Jakarta	4	0.0178
15	SMP Islam Al-Azhar Kelapa Gading	4	0.0178
16	SMP Kristen 4 Penabur	2	0.0089
17	SMP Labschool Rawamangun	2	0.0089
18	SMP Negeri 143 Jakarta	2	0.0089
19	SMP Negeri 151 Jakarta	2	0.0089
20	SMP Negeri 170 Jakarta	2	0.0089
21	SMP Negeri 55 Jakarta	2	0.0089
22	SMP Strada Santo Fransiskus Xaverius I	2	0.0089
23	SMP Strada Santo Fransiskus Xaverius II	2	0.0089
24	MTs. Al-Muhajirin	1	0.0044
25	MTs. Negeri 5 Jakarta	1	0.0044
26	SMP Islam Al-Azhar 12 Rawamangun	1	0.0044
27	SMP Negeri 112 Jakarta	1	0.0044
28	SMP Negeri 123 Jakarta	1	0.0044
29	SMP Negeri 137 Jakarta	1	0.0044
30	SMP Negeri 221 Jakarta	1	0.0044
31	SMP Negeri 289 Jakarta	1	0.0044
32	SMP Negeri 53 Jakarta	1	0.0044
33	SMP Negeri 62 Jakarta	1	0.0044
34	SMP Negeri 74 Jakarta	1	0.0044
35	SMP Santa Cilia	1	0.0044
36	SMP Santo Yakobus	1	0.0044
37	SMP Strada Santo Fransiskus Xaverius III	1	0.0044
38	MTs. Al-Jihad	0	0.0
39	MTs. Istiqlal	0	0.0
40	MTs. Negeri 31 Jakarta	0	0.0
41	MTs. Negeri 39 Jakarta	0	0.0
42	MTs. Umdatur Rasikhien Jakarta	0	0.0

Index	Nominal value	Absolute Count	Fraction
43	Pondok Pesantren Terpadu Al-Multazam	0	0.0
44	SMP Advent Anggrek Jakarta	0	0.0
45	SMP Al-Irsyad Al-Islamiyyah	0	0.0
46	SMP Barunawati 2	0	0.0
47	SMP Binaul Ummah Kuningan	0	0.0
48	SMP Darul Ma'arif	0	0.0
49	SMP IT Ash-Shiddiq	0	0.0
50	SMP International Islamic Secondary School (IISS)	0	0.0
51	SMP Islam Amalina	0	0.0
52	SMP Jubilee	0	0.0
53	SMP Kanaan Global School	0	0.0
54	SMP Kristen IPEKA Sunter	0	0.0
55	SMP Kristen Penabur Kota Wisata	0	0.0
56	SMP Marie Joseph	0	0.0
57	SMP Mekar Tanjung	0	0.0
58	SMP National Global School	0	0.0
59	SMP Negeri 116 Jakarta	0	0.0
60	SMP Negeri 118 Jakarta	0	0.0
61	SMP Negeri 121 Jakarta	0	0.0
62	SMP Negeri 136 Jakarta	0	0.0
63	SMP Negeri 140 Jakarta	0	0.0
64	SMP Negeri 152 Jakarta	0	0.0
65	SMP Negeri 162 Jakarta	0	0.0
66	SMP Negeri 163 Jakarta	0	0.0
67	SMP Negeri 18 Jakarta	0	0.0
68	SMP Negeri 193 Jakarta	0	0.0
69	SMP Negeri 213 Jakarta	0	0.0
70	SMP Negeri 216 Jakarta	0	0.0
71	SMP Negeri 217 Jakarta	0	0.0
72	SMP Negeri 226 Jakarta	0	0.0
73	SMP Negeri 23	0	0.0
74	SMP Negeri 261 Jakarta	0	0.0
75	SMP Negeri 266 Jakarta	0	0.0
76	SMP Negeri 269 Jakarta	0	0.0
77	SMP Negeri 270 Jakarta	0	0.0
78	SMP Negeri 279 Jakarta	0	0.0
79	SMP Negeri 34 Jakarta	0	0.0
80	SMP Negeri 4 Bekasi	0	0.0
81	SMP Negeri 4 Jakarta	0	0.0
82	SMP Negeri 43 Jakarta	0	0.0
83	SMP Negeri 5 Jakarta	0	0.0
84	SMP Negeri 61 Jakarta	0	0.0
85	SMP Negeri 65 Jakarta	0	0.0
86	SMP Negeri 75 Jakarta	0	0.0
87	SMP Negeri 94 Jakarta	0	0.0
88	SMP Negeri 99 Jakarta	0	0.0
89	SMP Paskalis 3	0	0.0
90	SMP Saint John's Catholic School	0	0.0
91	SMP Sekolah HighScope Indonesia Kelapa Gading	0	0.0
92	SMPIT Miftahul Ulum	0	0.0

Index	Nominal value	Absolute Count	Fraction
93	SMP Islam Al-Azhar 6 JP	0	0.0

## SIMPULAN

Berdasarkan hasil implementasi dan pengujian, dapat disimpulkan bahwa Random Forest dalam penelitian ini memiliki tingkat akurasi prediksi sebesar 90.67%, mengindikasikan kemampuan yang tinggi dalam memprediksi penerimaan peserta didik baru. Dengan classification error hanya sebesar 9.33%, model ini memiliki kinerja yang baik dan tingkat kesalahan yang rendah. Selain itu, nilai kappa sebesar 0.883 mencerminkan kualitas prediksi yang sangat baik dengan tingkat kesepakatan yang tinggi. Hasil analisis juga menunjukkan bahwa SMP Negeri 30 Jakarta, SMP Negeri 173 Jakarta, SMP Negeri 95 Jakarta, SMP Negeri 129 Jakarta, dan SMP KK Marsudirini memiliki peluang tertinggi untuk diterima melalui jalur prestasi akademik di SMA Negeri 13 Jakarta. Hasil penelitian ini tentu memiliki kekurangan dan saran-saran yang dapat menjadi pertimbangan untuk pengembangan lebih lanjut atau perbaikan dalam model prediksi yang diimplementasikan. Saran-saran tersebut mencakup pertimbangan untuk menggunakan atribut tambahan yang relevan dalam memprediksi penerimaan peserta didik baru guna meningkatkan kualitas prediksi, evaluasi berkala terhadap model yang telah dikembangkan untuk menjaga kinerja optimal, eksplorasi algoritma data mining lain selain Random Forest untuk tugas prediksi klasifikasi multikelas (polynomial classification) seperti penelitian ini, dan pengujian model pada lingkungan sekolah lain dengan karakteristik serupa atau berbeda untuk memvalidasi dan menguji aplikabilitasnya.

## DAFTAR PUSTAKA

- Aisyah, S., Sembiring, A. C., & Sitanggang, D. (2023). ASSOCIATION RULE. PUBLISH BUKU UNPRI PRESS ISBN, 1(1), 1-53.
- Fahreza, A. (2022). Penerapan Data Mining dengan Metode Single Moving Average dalam Pengolahan Data Penerimaan Siswa Baru. *Proceeding Seminar Nasional Ilmu Komputer*, 2(1), 25–34.
- Id, I. D. (2021). Machine Learning: Teori, Studi Kasus dan Implementasi Menggunakan Python (Vol. 1). Unri Press.
- Iqbal, M., Nawawi, H. M., Saelan, M. R. R., Maulana, M. S., & Mustopa, A. (2023). Optimasi

- Hyperparameter Multilayer Perceptron Untuk Prediksi Daya Beli Mobil. *Jurnal Manajemen Informatika dan Sistem Informasi*, 6(1), 73-81.
- Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P. (2022). Customer churn prediction system: a machine learning approach. *Computing*, 1-24.
- Latifah, A. L., Shabrina, A., Wahyuni, I. N., & Sadikin, R. (2019). Evaluation of Random Forest model for forest fire prediction based on climatology over Borneo. *2019 International Conference on Computer, Control, Informatics and Its Applications (IC3INA)*, 4–8.
- Linawati, S., Nurdiani, S., Handayani, K., & Latifah, L. (2020). Prediksi Prestasi Akademik Mahasiswa Menggunakan Algoritma Random Forest Dan C4. 5. *Jurnal Khatulistiwa Informatika*, 8(1), 491514.
- Karmita, S., Arief Bramanto, W. P., Gaffar, A. O., & Wiguna, A. S. (2018). Prediksi Jumlah Calon Mahasiswa Baru Menggunakan Fuzzy Time Series-Time Invariant. In *Prosiding Seminar Ilmu Komputer dan Teknologi Informasi Vol (Vol. 3, No. 1)*.
- Mehtab, S., & Sen, J. (2022). Stock price prediction using machine learning and deep learning algorithms and models. *Machine Learning in the Analysis and Forecasting of Financial Time Series*, 235-303.
- Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of machine learning*. MIT press.
- Mu'alim, F., & Hidayati, R. (2022). Implementasi Metode Random Forest Untuk Penjurusan Siswa Di Madrasah Aliyah Negeri Sintang. *JUPITER (Jurnal Penelitian Ilmu Dan Teknologi Komputer)*, 14(1), 116–125.
- Pratama, G., Nasrun, M., & Ansori, A. S. R. (2019). Pengumpulan Data Dan Prediksi Masuk Di Semua Smp Negeri Kota Cimahi Menggunakan Metode Random Forest. *EProceedings of Engineering*, 6(2).
- Rizqifaluthi, H., & Yaqin, M. A. (2018). Process mining Akademik Sekolah menggunakan rapidminer. MATICS: *Jurnal Ilmu Komputer dan Teknologi Informasi (Journal of Computer Science and Information Technology)*, 10(2), 47-51.
- Rerung, R. R. (2018). Penerapan data mining dengan memanfaatkan metode association rule untuk promosi produk. *J. Teknol. Rekayasa*, 3(1), 89.
- Rianto, M., & Yunis, R. (2021). Analisis Runtun Waktu Untuk Memprediksi Jumlah Mahasiswa Baru Dengan Model Random Forest. *Paradigma*, 23(1).
- Saikin, S., & Kusriani, K. (2019). Model Data Mining Untuk Karakteristik Data Traveller Pada Perusahaan Tour and Travel. *Jurnal Manajemen Informatika Dan Sistem Informasi*, 2(2), 61–68.
- Saleh, J., Adrian, A., & Sanger, J. (2022). SISTEM KLASIFIKASI KELULUSAN MAHASISWA DENGAN ALGORITMA RANDOM FOREST. *Jurnal Ilmiah Realtech*, 18(1), 10–14.
- Triwidianti, J., Alfian, F. Y., & Prasajo, M. (2021). Perbandingan Metode Data Mining Untuk Prediksi Prestasi Siswa Tingkat Pendidikan Menengah Kejuruan Pada Sekolah Menengah Kejuruan Negeri (SMKN 1) Gadingrejo Pringsewu Lampung. *Prosiding Seminar Nasional Darmajaya*, 1, 126–133.