

Recommendation System For Selecting Junior High School Based On Elementary School Students' Grades Using Random Forest Classifier (Case Study At MI From Kecamatan Asemrowo, Surabaya)

Sistem Rekomendasi SMP Berdasarkan Nilai Raport SD Menggunakan Metode Random Forest

Studi Kasus: MI Kecamatan Asemrowo, Surabaya

Tria Monica Yulita Putri¹, Bernadus Anggo Seno Aji², Yohanes Setiawan³

¹Teknologi Informasi, Institut Teknologi Telkom Surabaya, Indonesia

^{2,3}Institut Teknologi Telkom Surabaya, Indonesia

¹tria.monica@student.ittelkom-sby.ac.id, ^{2*}bernadus.seno@ittelkom-sby.ac.id,

³yohanes@ittelkom-sby.ac.id

*: *Penulis korespondensi (corresponding author)*

Informasi Artikel

Received: December 2020

Revised: January 2021

Accepted: January 2021

Published: February 2021

Menggunakan style info

Abstract (menggunakan style abstract)

Objective: This study aims to help recommend students whether they can choose the right state junior high school based on grades and distance. To implement a public junior high school recommendation system based on report cards and distance using the Random Forest method with the website.

Design/method/approach: Using the Random Forest Method in training data and data testing with hyperparameter tuning functions and confusion matrix calculations to get results for accuracy, precision, recall and f1 score.

Results: According to the confusion matrix calculation's outcomes, the results had an accuracy of 70%. The results of the Random Forest Recommendation If Students are Predicted to Enter Public Middle Schools contain a list of Public Middle Schools and information on the distance of the school not more than 2.5 km from the student's home address to the school's Public Middle School with the Surabaya City New Student Admissions (PPDB) zoning rules.

Authenticity/state of the art: This research has different recommendation criteria for public junior high schools, namely based on report cards and distance, something prior research has never attempted.

Abstrak (menggunakan style abstrak)

Keywords: State Junior High School;
Prediction; Recommendation; Random
Forest (RF)

Kata kunci: SMP Negeri; Prediksi;
Rekomendasi; Random Forest (RF)

Tujuan: Penelitian ini bertujuan untuk membantu merekomendasikan siswa apakah dapat memilih SMP Negeri yang tepat berdasarkan nilai raport dan jarak. Untuk mengimplementasikan sistem rekomendasi SMP Negeri berdasarkan nilai raport dan jarak menggunakan metode *Random Forest* dengan *website*.

Perancangan/metode/pendekatan: Menggunakan Metode *Random Forest* dalam data *training* dan data *testing* dengan fungsi *hyperparameter tuning* dan perhitungan *confusion matrix* untuk mendapatkan hasil akurasi, presisi, *recall* dan *f1 score*.

Hasil: Berdasarkan hasil perhitungan *confusion matrix* maka didapatkan hasil akurasi sebesar 70%. Hasil rekomendasi *random forest* apabila siswa-siswa diprediksi masuk SMP Negeri berisi daftar SMP Negeri dan informasi tentang jarak sekolah yang tidak lebih dari 2,5 km dari alamat rumah siswa ke SMP Negeri dengan aturan zonasi Penerimaan Peserta Didik Baru (PPDB) Kota Surabaya.

Keaslian/ state of the art : Penelitian ini memiliki kriteria rekomendasi SMP Negeri yaitu berdasarkan Nilai Raport dan Jarak, yang mana belum pernah dilakukan oleh penelitian-penelitian terdahulu.

1. Pendahuluan

Sekolah MI di Kecamatan Asemrowo Surabaya masih memiliki tingkat kelulusan siswa yang masuk SMP Negeri yang masih rendah. Sekolah MI di Kecamatan Asemrowo Surabaya dalam merekomendasikan siswa lulusannya untuk masuk ke SMP Negeri rata-rata masih secara manual, yaitu dengan cara konsultasi ke wali kelas. Lalu, wali kelas akan merekomendasikan SMP Negeri dari sisi ranking kelas siswa tersebut. Akan tetapi, metode tersebut masih kurang efektif karena memakan waktu dan tidak efisien. Maka dibutuhkannya sebuah sistem untuk membantu orang tua dan wali kelas dalam memilih SMP Negeri serta karena adanya batasan dalam memilih SMP Negeri yang ingin dituju hanya 2 pilihan. Maka sistem ini dibuat untuk membantu merekomendasikan siswa dalam memilih SMP yang sesuai dengan nilai raport dan jarak rumah siswa ke sekolah, sehingga peluang siswa tersebut untuk masuk ke SMP Negeri lebih besar.

Penelitian ini bertujuan membantu merekomendasikan siswa apakah dapat memilih SMP Negeri yang tepat dan sesuai dengan nilai raport dan jarak. Banyak Masyarakat pada tempat studi kasus yang diambil dalam penelitian ini berpendapat bahwa Sekolah Menengah Pertama Negeri merupakan sekolah yang tidak mengeluarkan biaya sebanyak sekolah swasta dan siswa yang mampu melanjutkan Pendidikan di Sekolah Menengah Pertama Negeri adalah mereka yang memiliki standar kualitas yang baik. Maka dari itu, kebanyakan orang tua yang kurang dalam hal *financial* terkadang memilih untuk tidak menyekolahkan anaknya apabila anaknya tidak masuk ke Sekolah Menengah Pertama Negeri. Hal tersebut menjadi urgensi. Sistem yang

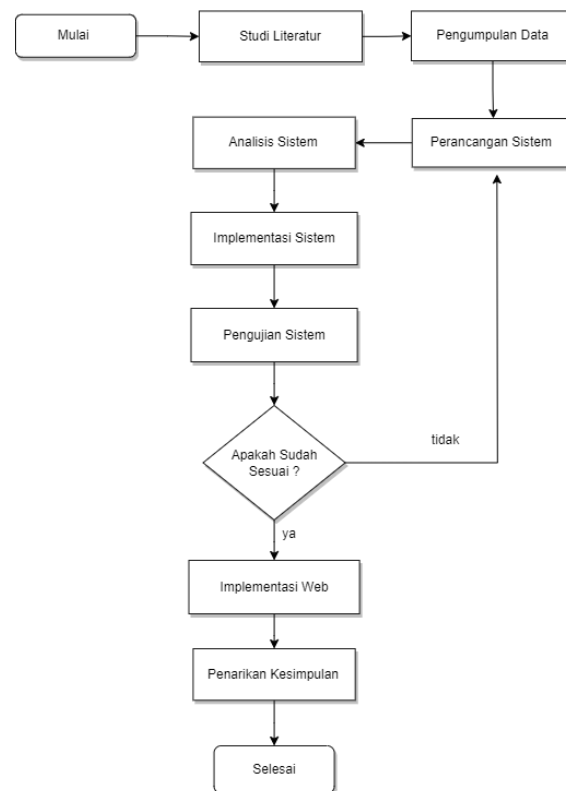
dibangun merupakan sistem rekomendasi SMP Negeri berbasis web berdasarkan nilai raport dengan menggunakan metode *Random Forest*.

Beberapa penelitian yang membahas sistem rekomendasi sekolah telah dilakukan, yaitu Sistem Rekomendasi Sekolah Menengah Atas bagi Calon Siswa di Kota Bandar Lampung, penelitian ini mengadopsi Metode SAW (*Simple Additive Weighting*) berbasis Sistem Informasi Geografis penelitian ini memanfaatkan Metode SAW untuk melakukan penjumlahan bobot persentase berdasarkan prestasi ujian nasional dan *passing grade* sekolah[1]. Selanjutnya, penelitian membahas tentang Pengumpulan Data dan Prakiraan Penerimaan di Seluruh SMP Negeri Kota Cimahi dengan Penerapan Metode *Random Forest*. Penelitian tersebut membuktikan dengan menggunakan *random forest* terbukti mencapai hasil akurasi yang sangat tinggi dalam meramalkan nilai masuk sekolah. Hasil tingkat akurasi yang dicapai mencapai 99,15% [2]. Dan yang terakhir penelitian membahas Perbandingan Prestasi Algoritma C4.5, Pohon *Gradient Boosting*, *Random Forest*, dan Pembelajaran Mendalam dalam Pengolahan Data Pendidikan dengan menggunakan metode *Random Forest* menghasilkan akurasi sebesar 80,31% [3].

Berdasarkan penelitian-penelitian sebelumnya, dapat diambil sebuah keputusan penelitian ini menggunakan *Random Forest* (RF) untuk memprediksi nilai siswa dengan merekomendasikan siswa masuk ke sekolah SMP Negeri sesuai dengan nilai raport, karena metode *Random Forest* (RF) memiliki karakteristik yang cocok sesuai dengan atribut-atribut yang telah ditentukan dan terbukti kuat dan akurat sehingga dapat meningkatkan hasil akurasi.

2. Metode yang digunakan

Penelitian ini menggunakan Python untuk mengimplementasikan metode *Random Forest* (RF) dalam memproses data. Pengolahan data melibatkan pemisahan data uji dan data latihan yang menghasilkan persentase sebagai hasil akhir. **Gambar 1** menunjukkan diagram alir metodologi penelitian.



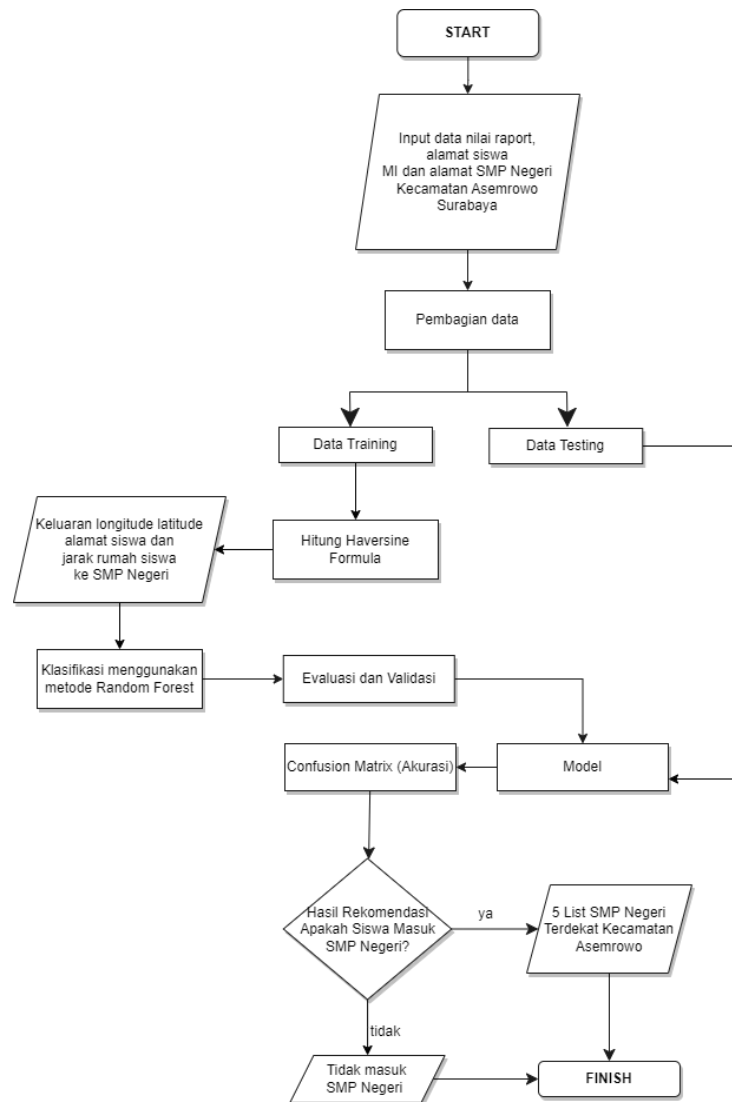
Gambar 1. Diagram Alir dari Metodologi Penelitian

Tahap pertama, Studi Literatur adalah pemahaman teori dan konsep dasar diperoleh melalui studi literatur, termasuk pencarian informasi relevan dan istilah kunci terkait. Pengumpulan Data adalah data alumni siswa MI diambil dari tahun akademik 2018-2022 di Kecamatan Asemrowo Surabaya, mencakup nilai Matematika, Bahasa Indonesia, dan Bahasa Inggris dari semester kedua kelas enam, serta alamat dan lokasi SMP Negeri. Perancangan Sistem adalah Tahap perancangan sistem melibatkan rancangan detail sistem dan desain *dashboard* web[4]. Analisis Sistem adalah sebuah komponen sistem dianalisis untuk mengidentifikasi masalah, peluang, dan kebutuhan, guna memberikan rekomendasi perbaikan. Implementasi Sistem adalah data nilai raport siswa diolah dengan Metode *Random Forest* (RF) sesuai rencana sistem. Pengujian Sistem adalah pengujian yang meliputi data pelatihan, dan pengukuran akurasi menggunakan *Confusion Matrix* serta metrik seperti *Accuracy*, *Precision*, *Recall*, *Specificity*, dan *F1 Score*[5]. Setelah pengujian sistem berhasil maka tahap selanjutnya, Implementasi Web adalah hasil rekomendasi diintegrasikan dalam tampilan website. Dan tahap terakhir yaitu penarikan Kesimpulan adalah kesimpulan ditarik berdasarkan hasil penelitian dan evaluasi metode serta sistem yang digunakan.

2.1. Gambaran Analisis Sistem

Dalam analisis sistem, terdapat panduan rekomendasi SMP Negeri berdasarkan standar nilai raport Matematika, Bahasa Indonesia, dan IPA semester 2 kelas 6. Data siswa dan daftar SMP Negeri diproses menggunakan metode *Random Forest* (RF) sebagai *follow-up* untuk siswa MI di Asemrowo, Surabaya. RF memberi peringatan dini dan prediksi penerimaan siswa di SMP

Negeri. **Gambar 2** menunjukkan diagram alur sistem (*flowchart*) langkah-langkah pengolahan data dengan metode *Random Forest*.

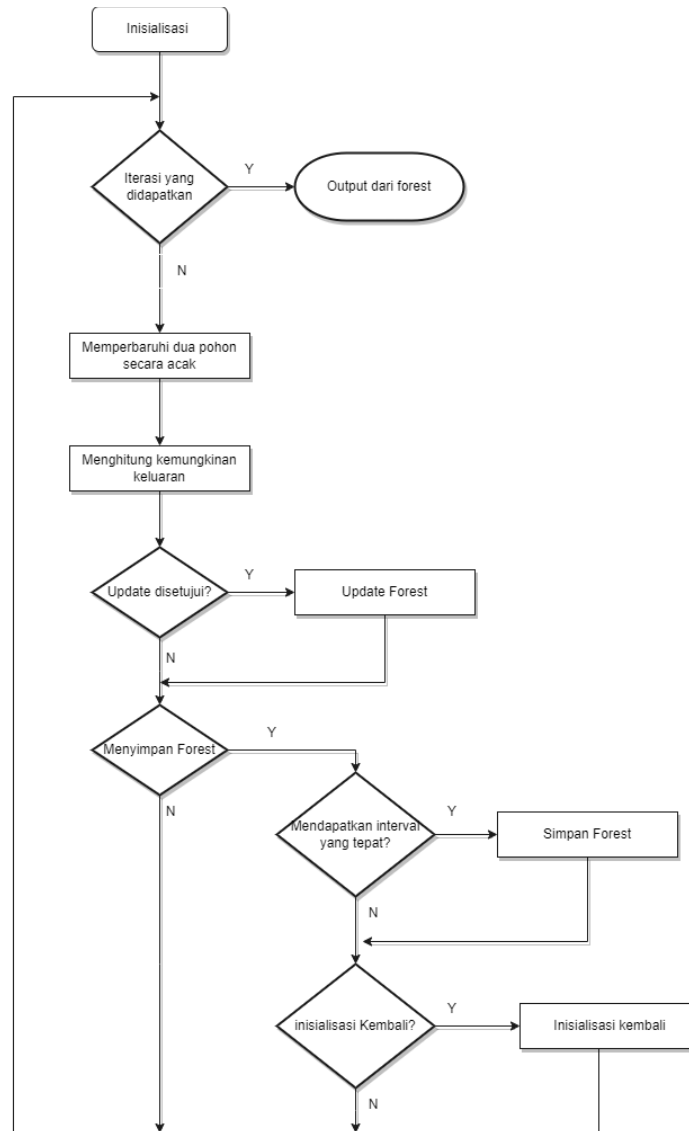


Gambar 2. Diagram Alur Sistem

Penjelasan dari diagram alur sistem **Gambar 2**. Yang pertama yaitu Input Data, Data diperoleh dari dua MI di Kecamatan Asemrowo Surabaya, termasuk nilai raport (Matematika, IPA, Bahasa Indonesia) dan alamat siswa tahun 2018-2022, serta alamat 5 SMP Negeri. Selanjutnya pembagian Data, data dibagi menjadi pelatihan dan pengujian, untuk melatih dan mengevaluasi model[6]. Lalu, hitung *Haversine Formula* (Mencari *Longitude, Latitude, Distance*), perhitungan *haversine* digunakan untuk mencari longitude, latitude, dan jarak antara alamat rumah siswa dan SMP Negeri. Selanjutnya yaitu dilakukan evaluasi model, model dievaluasi melalui pengujian dan penilaian performa menggunakan *confusion matrix*[7]. Selanjutnya tahap akhir didapatkannya hasil rekomendasi, evaluasi dan validasi menghasilkan rekomendasi siswa untuk masuk SMP Negeri, dengan 5 SMP Negeri terdekat sebagai output jika memenuhi kriteria.

2.2. Gambaran Perancangan Sistem

Pada Proses *pre-processing*, pelatihan data, dan pengujian data untuk sistem rekomendasi SMP akan dijelaskan lebih detail pada tahap perancangan sistem. **Gambar 3** menunjukkan flowchart yang akan digunakan untuk menunjukkan bagaimana ketika algoritma random forest bekerja[8].



Gambar 3. Flowchart Random Forest (RF)

- Pre-processing* digunakan untuk membersihkan data mentah menjadi variabel yang ditetapkan sebelumnya untuk data latih[9].
- Data *training* digunakan untuk mengenali pola klasifikasi *Random Forest*. Penggantian rasio data training dan testing dilakukan untuk optimalitas. Pembagian data testing dan data training menggunakan rasio 30 : 70.

- c. *Hyperparameter Tuning Random Forest* digunakan untuk membangun model random forest, parameter ditentukan sesuai referensi[10]. Tabel 1 berisi parameter dan nilai-nilai yang digunakan untuk memilih parameter terbaik.

Tabel 1. Nilai Parameter Tuning

Parameter Tuning	Nilai Parameter Tuning
<i>max_depth</i>	Range (1,5)
<i>min_samples_split</i>	2, 5, 10
<i>max_leaf_nodes</i>	Range (5,20)
<i>min_samples_leaf</i>	2,5,10
<i>n_estimators</i>	Range (5,30)
<i>max_samples (bootstrap samples)</i>	True
<i>max_features</i>	Sqrt

- d. *Performance Measurement*

Confusion Matrix mengklasifikasikan perhitungan, adalah ukuran kinerja untuk klasifikasi dengan dua kelas atau lebih[11]. Tabel *Matrix* berisi empat kombinasi nilai aktual dan prediksi: *True Positif*, *True Negatif*, *False Positif*, dan *False Negatif*. Ini digunakan untuk mengukur kinerja melalui *Accuracy*, *Precision*, *Recall (Sensitivity)*, *Specificity*, dan *F1 Score*[12].

- **Accuracy:** Perbandingan antara prediksi yang tepat (baik positif maupun negatif) dengan seluruh data.

$$Accuracy = \frac{(TP+TN)}{(TP+FP+FN+TN)} \quad (1)$$

- **Precision:** Perbandingan antara prediksi positif yang tepat dengan seluruh hasil yang diprediksi positif.

$$Precision = \frac{TP}{(TP+FP)} \quad (2)$$

- **Recall (Sensitivity):** Perbandingan antara prediksi positif yang tepat dengan keseluruhan data yang memang benar positif.

$$Recall (Sensitivity) = \frac{TP}{(TP+FN)} \quad (3)$$

- **Specificity:** Merupakan perbandingan antara prediksi yang benar sebagai negatif dengan seluruh data yang memang negatif.

$$Specificity = \frac{TN}{(TN+FP)} \quad (4)$$

- **F1 Score:** *F1 Score* merupakan perbandingan rata-rata presisi dan *recall* yang dibobotkan adalah hasil perbandingan tertimbang antara rata-rata presisi dan *recall*.

$$F1 Score = 2 * \frac{Recall*Precision}{Recall+Precision} \quad (5)$$

Keterangan:

TP = Jumlah data positif yang diprediksi dengan benar.

TN = Jumlah data negatif yang diprediksi dengan benar

FP = Jumlah data negatif yang diprediksi sebagai positif.

FN = Jumlah data positif yang diprediksi sebagai negatif.

Dengan mengukur semua metrik, maka dapat membantu mengetahui sejauh mana sistem berhasil dalam berbagai aspek, seperti memberikan rekomendasi yang tepat serta dapat membantu untuk mengidentifikasi area yang perlu ditingkatkan untuk meningkatkan kualitas rekomendasi yang diberikan kepada siswa.

3. Hasil dan Pembahasan

3.1. Hasil Pengumpulan Data

3.1.1. Data Yang Digunakan

Data nilai raport dan alamat siswa alumni (2018-2022) diperoleh dari dua MI di Kecamatan Asemrowo, Surabaya: MI Pangeran Diponegoro dan MI Tarbiyatus Shiblyan. Data berformat xlsx dengan 394 data. Data tersebut meliputi: tahun angkatan alumni, nilai raport dengan mata Pelajaran Bahasa Indonesia, Matematika, Ilmu Pengetahuan Alam (IPA), nama siswa (yang telah disamarkan), alamat rumah siswa dan list lulusan SMP Negeri yang telah di input oleh pihak sekolah.

Setelah memperoleh data asli, dilakukan *preprocessing* yang terdiri dari perhitungan jarak koordinat menggunakan teknik *geocoding*, mengubah alamat menjadi *longitude* dan *latitude*, serta menghitung jarak dalam KM dengan *calculate distance*. Atribut yang tak relevan seperti nama siswa, alamat, dan SMP Negeri dihapus.

3.1.2. Hasil Pre Processing Data

Setelah pengumpulan, seleksi atribut, dan pembersihan data, peneliti memperoleh 394 data sebagai dasar penelitian. Data telah melalui *preprocessing* untuk persiapan analisis. Selanjutnya, data dibagi menjadi data pelatihan dan pengujian untuk melatih dan menguji model. Berikut adalah tampilan dataset setelah *preprocessing*:

Tabel 2. Contoh sampel data setelah di pre-processing

rata_rata_mtk	rata_rata_bhs_indo	rata_rata_ipa	latitude_rumah	longitude_rumah	jarak ke SMP_Negeri 42	jarak ke SMP_Negeri 50	jarak ke SMP_Negeri 25	jarak ke SMP_Negeri 3	jarak ke SMP_Negeri 4	target
69	77	73	7,2523067	112,7136755	0,949940365217835	2,219886604	2,471391282	2,410673193	2,533097526	0
66	76	73	7,2537807	112,7044179	1,793557427	1,281431869	2,046946421	3,458704407	3,569022805	0

76	80	76	7,2500608	112,7176665	0,539633348395862	2,651916035	2,90318882	2,175494428	2,317411786	0
72	80	81	7,2467543	112,7070249	1,237210203	2,067973866	2,849329816	3,32252547	3,462677719	1
73	79	82	7,2579574	112,697072	2,726252168	0,468803783440214	1,728643376	4,267099898	4,359306594	1

Setelah *preprocessing* data selesai, langkah berikutnya adalah dilakukan proses split data, terdapat 314 data yang digunakan sebagai data *training* dan 79 data yang digunakan sebagai data *testing*. Pemisahan ini dilakukan dengan rasio 70% data *training* dan 30% data *testing*. Dalam penelitian ini, proses pembelajaran model atau algoritma yang di terapkan digunakan untuk data *training*. Kemudian, data *testing* akan digunakan untuk menguji performa model yang telah dilatih. Berikut merupakan tabel contoh sampel data *training* dan data *testing* setelah dilakukan split data:

- Ringkasan Data *Training* dapat dilihat pada **Tabel 3**

Tabel 3. Data *Training*

rata_r ata mtk	rata_r ata bhs_i ndo	rata_rata ipa	latitude_ru mah	longitude_ rumah	jarak ke SMP_ Negeri 42	jarak ke SMP_ Negeri 50	jarak ke SMP_Ne geri 25	jarak ke SMP_Ne geri 3	jarak ke SMP_ Negeri 4	target
89	89	91	-7,2522868	112,7051266	1,265169125	1,788988445	2,318010588	2,967432254	3,0881744	1
85	84	81	-7,2532223	112,7016401	2,030270428	1,144448073	2,114311357	3,769255085	3,880808759	0
88	87	88	-7,2522868	112,7051266	1,265169125	1,788988445	2,318010588	2,967432254	3,0881744	1
87	87	87	-7,2523579	112,7141485	1,14120408	1,9481629	2,317438932	2,700820579	2,819386217	1
81	79	82	-7,2500095	112,7060738	1,632629899	1,594428698	2,463218704	3,539358605	3,664998954	0

- Dan contoh sampel Data *Testing* dapat dilihat pada **Tabel 4**

Tabel 4. Data *Testing*

rata_rata_mtk	rata_rata_bhs_indo	rata_rata_ipa	latitude_rumah	longitude_rumah	jarak ke SMP_Negeri 42	jarak ke SMP_Negeri 50	jarak ke SMP_Negeri 25	jarak ke SMP_Negeri 3	jarak ke SMP_Negeri 4	target
75	79	78	7,2565847	112,7074428	1,736475838	1,367213721	1,790653271	3,116896883	3,214478116	0
82	79	84	7,2464863	112,707658	1,036358786	2,251353686	2,968378273	3,170596183	3,315504428	0
90	89	91	7,2490772	112,7047109	1,395699551	1,78466534	2,555081598	3,334063588	3,464027088	1
88	78	91	7,2507189	112,7023103	1,985008534	1,390078231	2,429487806	3,881807279	4,003596307	1
89	89	89	7,2522868	112,7051266	1,265169125	1,788988445	2,318010588	2,967432254	3,0881744	1

3.1.3. Atribut yang digunakan

Dalam penelitian ini, enam atribut digunakan: nilai raport dengan nilai rata-rata dari mata pelajaran Bahasa Indonesia, Matematika, Ilmu Pengetahuan Alam, *longitude*, *latitude* alamat rumah siswa, hasil perhitungan jarak rumah siswa ke sekolah SMP Negeri dan target 0 untuk tidak diterima dan 1 untuk diterima.

3.2. Hasil Evaluasi dan Analisis Model *Random Forest*

3.2.1. Hasil *Hyperparameter Tuning Random Forest*

Tujuan dari proses ini adalah mencari kombinasi parameter terbaik yang akan digunakan dalam model *Random Forest*, proses ini dilakukan dengan menggunakan data *training*. **Tabel 5** berikut menunjukkan hasil *Hyperparameter Tuning*.

Tabel 5. Hasil Parameter *Tuning*

<i>max_depth</i>	<i>min_samples_split</i>	<i>max_leaf_nodes</i>	<i>min_samples_leaf</i>	<i>n_estimators</i>	<i>max_samples (bootstrap samples)</i>	<i>max_features</i>	Akurasi
4	2	9	5	28	TRUE	sqrt	80,58%
4	5	9	5	28	TRUE	sqrt	80,58%
4	10	9	5	28	TRUE	sqrt	80,58%
4	2	7	5	21	TRUE	sqrt	79,93%
4	5	7	5	21	TRUE	sqrt	79,93%

Hasil *Hyperparameter Tuning* menunjukkan parameter terbaik: ***Bootstrap: True, max_depth: 4, max_features: sqrt, max_leaf_nodes: 9, min_samples_leaf: 5, min_samples_split: 2, n_estimators: 28***. Akurasi terbaik model *random forest* adalah **80,58%**, menunjukkan kemampuan rekomendasi yang tinggi. Akurasi menggambarkan kesesuaian hasil dengan data actual.

3.2.2. Akurasi Model *Random Forest*

Penilaian model *Random Forest* menggunakan *Confusion Matrix* termasuk *True Positif, True Negatif, False Positif, dan False Negatif. Performance* diukur dengan *Accuracy, Precision, Recall, Spesifisitas, dan F1 Score*. Akurasi hasil model, melalui *Confusion Matrix*, dinyatakan dalam persentase untuk menggambarkan keberhasilannya dalam memberikan rekomendasi. **Tabel 6** berikut menunjukkan hasil *Confusion Matrix*.

Tabel 6. Hasil *Confusion Matrix*

Confusion Matrix		Nilai Aktual	
		Direkomen dasikan masuk SMP Negeri	Tidak Direkomendasi kan masuk SMP Negeri
Nilai Prediksi	Direkomendasikan masuk SMP Negeri	28	8
	Tidak Direkomendasikan masuk SMP Negeri	15	28

Tabel 6 menunjukkan hasil dari *Confusion matrix* dengan dimensi 2x2 yang mencakup empat elemen yang digunakan untuk menggambarkan hasil klasifikasi, yakni sebagai berikut :

- ***True Positive (TP)***: Menjelaskan bahwa siswa yang seharusnya diterima berdasarkan kriteria rekomendasi, model berhasil merekomendasikan 28 siswa dengan benar yang sesuai dengan kriteria rekomendasi random forest yaitu input nilai raport dan alamat siswa.
- ***True Negative (TN)***: Menjelaskan bahwa dari semua siswa yang seharusnya ditolak berdasarkan kriteria rekomendasi, model berhasil mengidentifikasi dan merekomendasikan dengan benar 28 siswa yang sebenarnya tidak memenuhi kriteria rekomendasi random forest.
- ***False Positive (FP)***: Menjelaskan bahwa model telah merekomendasi 8 siswa yang sebenarnya tidak memenuhi kriteria, tetapi model memprediksi mereka sebagai siswa yang harus diterima. Dengan kata lain, terdapat kesalahan pada model pada saat rekomendasi dilakukan.
- ***False Negative (FN)***: Menjelaskan bahwa terdapat 15 siswa yang sebenarnya memenuhi kriteria untuk masuk ke SMP Negeri, tetapi model tidak merekomendasikannya.

Berikut pada **Tabel 7** merupakan hasil keseluruhan dari *Accuracy, Precision, Recall (Sensitivitas), Spesifitas* dan *F1 Score*

Tabel 7. Hasil Keseluruhan Perhitungan *Performance*

Metode Klasifikasi	Accuracy (%)	Precision (%)	Recall (Sensitivity) (%)	Specificity (%)	F1 Score (%)
Random Forest	70%	77%	65%	77%	70%

Berdasarkan tabel 7, terdapat hasil akurasi dari 79 siswa yang digunakan model berhasil mencapai akurasi sebesar 70% yang dimana semua prediksi dilakukan sesuai dengan nilai raport siswa secara aktual. *Presisi* dari Dari 79 siswa yang direkomendasikan oleh model untuk masuk SMP Negeri, didapatkan hasil sebesar 77% sesuai dengan nilai aktual. *Spesifisitas* Dari 314 siswa yang digunakan didapatkan hasil 77% model berhasil mengidentifikasi dengan benar kasus yang sebenarnya negatif (tidak layak direkomendasikan). *Recall (Sensitivity)* 65% menunjukkan bahwa dari semua siswa yang sebenarnya memiliki potensi untuk direkomendasikan, model hanya berhasil mengidentifikasi sekitar 65% dari total siswa tersebut. dan *F1 score* sebesar 70% mencerminkan keseimbangan antara *presisi* dan *recall* dalam evaluasi model yang dimana *presisi* 77% sejauh mana model memberikan rekomendasi yang akurat dan *recall* 65% sejauh mana model mengidentifikasi sejumlah besar calon siswa yang layak direkomendasikan.

3.2.3. Hasil Rekomendasi Menggunakan *Random Forest*

Bagian ini menyajikan hasil rekomendasi dengan metode *Random Forest*. Metode ini mempertimbangkan nilai raport siswa (Matematika, IPA, Bahasa Indonesia) dan jarak ke SMP Negeri (SMP Negeri 42, 50, 25, 3, dan 4).

Tabel 8 menunjukkan hasil percobaan sistem rekomendasi siswa-siswa yang diprediksi masuk ke SMP Negeri berdasarkan nilai raport dan jarak.

Tabel 8. Hasil Rekomendasi *Random Forest* siswa-siswa yang diprediksi masuk ke SMP Negeri berdasarkan nilai raport dan jarak.

No	Nilai Raport			Alamat Rumah Siswa		Hasil Rekomendasi
	Rata-rata Matematika	Rata-rata Bahasa Indonesia	Rata-rata IPA	Longitude	Latitude	Random Forest
1	95	95	88	- 7,2685368	112,70099 52	SMP Negeri 50 dan SMP Negeri 25 Surabaya
2	76	79	79	-7,250667	112,70478 46	SMP Negeri 25, SMP Negeri 42 dan SMP Negeri 50 Surabaya
3	84	90	85	- 7,2506703	112,71487 71	SMP Negeri 42, SMP Negeri 3 dan SMP Negeri 25 Surabaya
4	91	91	90	- 7,2533431	112,70207 83	SMP Negeri 42, SMP Negeri 3 dan SMP Negeri 4 Surabaya
5	91	78	83	- 7,2467543	112,70702 49	SMP Negeri 42 dan SMP Negeri 25 Surabaya

Tabel 9 merupakan siswa-siswa yang diprediksi tidak dapat masuk SMP Negeri berdasarkan nilai raport dan jarak.

Tabel 9. Hasil Rekomendasi *Random Forest* apabila siswa-siswa yang diprediksi tidak dapat masuk SMP Negeri berdasarkan nilai raport dan jarak

No	Nilai Raport			Alamat Rumah Siswa		Hasil Rekomendasi Random Forest
	Rata-rata Matematika	Rata-rata Bahasa Indonesia	Rata-rata IPA	Longitude	Latitude	
1	69	77	73	-7,2523067	112,7136755	MOHON MAAF SISWA TIDAK DITERIMA DI SMP NEGERI
2	69	78	77	-7,2538007	112,7048919	MOHON MAAF SISWA TIDAK DITERIMA DI SMP NEGERI
3	66	78	71	-7,2521323	112,70096	MOHON MAAF SISWA TIDAK DITERIMA DI SMP NEGERI
4	68	79	79	-7,2460718	112,7197091	MOHON MAAF SISWA TIDAK DITERIMA DI SMP NEGERI
5	68	81	78	-7,2499758	112,7078413	MOHON MAAF SISWA TIDAK DITERIMA DI SMP NEGERI

3.3. Hasil Pengujian Sistem

3.3.1. Hasil Implementasi Web

Hasil akurasi dan rekomendasi dari metode *Random Forest* diimplementasikan ke dalam website menggunakan *framework Flask* dan *Python*. Pembuatan API untuk mengolah rekomendasi dengan *Random Forest* memanfaatkan *library Python joblib*. *Joblib* membantu menyimpan dan memuat model dalam aplikasi web, memastikan kemudahan akses dan penggunaan.

Berikut dibawah ini merupakan tampilan - tampilan hasil implementasi *website* :

1. Tampilan awal (Home)

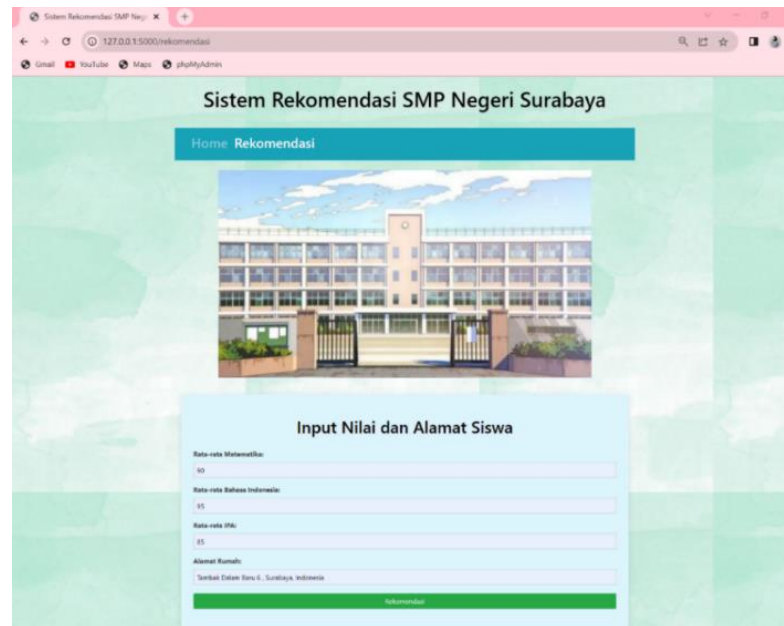


Gambar 4. Tampilan Pertama dari antarmuka Sistem Rekomendasi SMP Negeri Surabaya

Gambar 4 menampilkan antarmuka pertama (Halaman Utama) dari situs web Sistem Rekomendasi SMP Negeri Surabaya. Terdapat menu Rekomendasi yang digunakan untuk input nilai dan alamat siswa yang kemudian didapatkan hasil rekomendasi SMP Negeri sesuai dengan

jarak terdekat antara alamat siswa dengan alamat SMP Negeri tersebut sehingga siswa mendapat gambaran dalam memilih SMP Negeri.

2. Tampilan Rekomendasi



The screenshot shows a web browser window with the URL 127.0.0.1:5000/rekomendasi. The page title is 'Sistem Rekomendasi SMP Negeri Surabaya'. Below the title is a navigation bar with 'Home' and 'Rekomendasi'. A large image of a school building is displayed. Below the image is a form titled 'Input Nilai dan Alamat Siswa'. The form contains four input fields: 'Rata-rata Matematika' (value: 80), 'Rata-rata Bahasa Indonesia' (value: 85), 'Rata-rata IPA' (value: 85), and 'Alamat Rumah' (value: 'Sembak Dalam Baru I, Surabaya, Indonesia'). A green 'Rekomendasi' button is at the bottom of the form.

Gambar 5. Tampilan Rekomendasi Untuk Input Nilai dan Alamat Siswa

Pada **Gambar 5** merupakan tampilan menu Rekomendasi pada *website* Sistem Rekomendasi SMP Negeri Surabaya. Terdapat input nilai dan alamat siswa. Pada tampilan rekomendasi ini, siswa dapat memasukkan nilai rata-rata mata pelajaran Matematika, Bahasa Indonesia, dan Ilmu Pengetahuan Alam (IPA), serta alamat rumah siswa. Di bagian bawah halaman, terdapat tombol berwarna hijau yang dapat digunakan untuk memulai proses rekomendasi SMP Negeri.

3. Tampilan Hasil Rekomendasi Apabila Siswa-siswa yang diprediksi masuk ke SMP Negeri berdasarkan nilai raport dan jarak



The screenshot shows the 'Hasil Rekomendasi' page. The page title is 'Sistem Rekomendasi SMP Negeri Surabaya'. Below the title is a navigation bar with 'Home' and 'Rekomendasi'. A large image of a school building is displayed. Below the image is a section titled 'Hasil Rekomendasi'. The text reads: 'Rekomendasi SMP Negeri Terdekat Yang Dapat Siswa Pilih Yaitu:'. Below this text are two bullet points: '• SMP_Negeri_42 (Jarak: 0.382501950646204)' and '• SMP_Negeri_25 (Jarak: 2.3458613256469184)'.

Gambar 6. Contoh Tampilan Hasil Rekomendasi Apabila Siswa-siswa yang diprediksi masuk ke SMP Negeri berdasarkan nilai raport dan jarak

Pada **Gambar 6** merupakan contoh tampilan ketika merekomendasikan SMP Negeri. Terdapat kotak di bagian bawah header yang menunjukkan hasil rekomendasi siswa-siswa yang diprediksi masuk ke SMP Negeri berdasarkan nilai raport dan jarak dengan menampilkan beberapa list SMP Negeri include dengan keterangan jarak SMP Negeri tersebut yang kurang dari 2,5 km dari alamat rumah siswa ke SMP Negeri tersebut sesuai dengan peraturan zonasi yang berlaku selama proses Penerimaan Peserta Didik Baru (PPDB) Kota Surabaya.

4. Tampilan Hasil Rekomendasi Apabila Siswa-siswa yang diprediksi tidak dapat masuk ke SMP Negeri berdasarkan nilai raport dan jarak



Gambar 7. Contoh Tampilan Hasil Rekomendasi apabila siswa-siswa yang diprediksi tidak dapat masuk SMP Negeri berdasarkan nilai raport dan jarak.

Gambar 7 menampilkan yang serupa dengan **Gambar 6**, yang menunjukkan contoh proses rekomendasi SMP Negeri. Pada tampilan hasil rekomendasi ini, akan ditampilkan hasil rekomendasi apabila siswa-siswa yang diprediksi tidak dapat masuk SMP Negeri berdasarkan nilai raport dan jarak.

3.4. Hasil Analisis Sistem

Setelah penerapan model *Random Forest*, sistem dievaluasi untuk mengukur kinerja dan efektivitasnya. Hasil analisis dari beberapa faktor yang dievaluasi yaitu, kecepatan respons sistem Pengujian menunjukkan waktu respons sistem rata-rata di bawah 2 detik untuk menghasilkan rekomendasi SMP Negeri. Kecepatan ini memberikan pengalaman pengguna yang baik dan meminimalkan waktu tunggu. Lalu akurasi rekomendasi SMP Negeri, Performa model dievaluasi dengan *confusion matrix* dan metrik seperti *Accuracy*, *Precision*, *Recall (sensitivity)*, *Specificity*, dan *F1 score*. Hasil pengujian menunjukkan model *Random Forest* mencapai akurasi sekitar 70%. Ini memberikan keyakinan pada hasil rekomendasi sistem.

4. Kesimpulan dan Saran

Metode *Random Forest* dapat digunakan untuk merekomendasikan SMP Negeri yang sesuai. Dari 394 data, metode *random forest* menghasilkan Akurasi 70%, *Precision* 77%, *Specificity* 77%, *Recall* 65% dan *F1 Score* 70%. Hasil tersebut menunjukkan bahwa bahwa model mampu memberikan rekomendasi yang akurat dan relevan. Sistem Rekomendasi berbasis web berhasil dikembangkan untuk rekomendasi SMP Negeri. Dengan mematuhi pedoman zonasi yang berlaku selama proses Penerimaan Peserta Didik Baru (PPDB) Kota Surabaya, sistem ini mempertimbangkan jarak antara tempat tinggal siswa dan SMP Negeri yang direkomendasikan.

Adapaun beberapa saran yang diusulkan sebagai pengembangan penelitian selanjutnya adalah penyempurnaan algoritma *Random Forest* dalam proses pemilihan fitur, parameter model supaya mencapai hasil yang lebih baik secara keseluruhan. Memperbanyak sekolah atau data yang diolah supaya menghasilkan hasil yang lebih akurat. Dapat mempertimbangkan faktor-faktor lain yang relevan seperti prestasi akademik sebelumnya, minat siswa, dan lingkungan social yang dapat meningkatkan akurasi dan keberlanjutan rekomendasi. Dilakukan pengembangan sistem dengan menambahkan fitur-fitur pada web supaya dapat lebih mudah diakses serta dipahami oleh guru, pendidik, dan lainnya.

Daftar Pustaka

- [1] P. P. SARI, “Sistem Rekomendasi Sekolah Menengah Atas Untuk Calon Siswa Di Kota Bandar Lampung Menggunakan Metode SAW (Simple Additive Weighting) Berbasis Sistem Informasi Geografis,” *Angew. Chemie Int. Ed.* 6(11), 951–952., no. Mi, pp. 5–24, 2019.
- [2] G. Pratama, M. Nasrun, S. Si, A. Siswo, and R. Ansori, “Pengumpulan Data dan Prediksi Masuk Di Semua SMP Negeri Kota Cimahi Menggunakan Metode Random Forest.”
- [3] S. Mutrofin, M. Mughniy Machfud, D. H. Satyareni, R. Venantius, H. Ginardi, and C. Fatchah, “Komparasi Kinerja Algoritma C4.5, Gradient Boosting Trees, Random Forest, Dan Deeo Learning Pada Kasus Educational Data Mining”, doi: 10.25126/jtiik.2020732665.
- [4] R. A. Juperta, “Pembangunan Aplikasi Pemesanan Pada Tukang Sayur Keliling Disertai Fitur Rekomendasi Masakan Harian Memanfaatkan Fcm Dan Lbs Berbasis Android,” pp. 45–130, 2019.
- [5] R. G. Ramli and Y. Sibaroni, “Klasifikasi Topik Twitter menggunakan Metode Random Forest dan Fitur Ekspansi,” vol. 9, no. 1, pp. 79–92, 2022.
- [6] M. L. Suliztia, “Penerapan Analisis Random Forest Pada Prototype Sistem Prediksi Harga Kamera Bekas Menggunakan Flask,” *Fak. Mat. Dan Ilmu Pengetah. Alam*, pp. 1–107, 2020.
- [7] I. A. Rahmi, F. M. Afendi, and A. Kurnia, “Metode AdaBoost dan Random Forest untuk Prediksi Peserta JKN-KIS yang Menunggak,” *Jambura J. Math.*, vol. 5, no. 1, pp. 83–94, 2023, doi: 10.34312/jjom.v5i1.15869.
- [8] Q. Li and G. D. Clifford, “Secondary Analysis of Electronic Health Records,” *Second. Anal. Electron. Heal. Rec.*, pp. 1–427, 2016, doi: 10.1007/978-3-319-43742-2.
- [9] P. Oktaviani¹, M. T. , Ibnu Asror, S.T., ², and Dr. Moch. Arif Bijaksana³, “Analisis Implementasi Sistem OLAP dan Klasifikasi Ketepatan Waktu Lulus dan Undur Diri Mahasiswa Teknik Informatika Universitas Telkom Menggunakan Random forest,” vol. 5, no. 14, pp. 63–65, 2018, doi: 10.15900/j.cnki.zylf1995.2018.02.001.
- [10] H. I. Rhys, *Machine Learning with R, the tidyverse, and mlr*. 2020.
- [11] M. I. Putra, “Sistem Rekomendasi Kelayakan Kredit Menggunakan Metode Random Forest pada BRI Kantor Cabang Pelaihari,” *J. Tek. Inform. dan Sist. Inf.*, vol. 13, no. 1,

p. 61, 2019.

- [12] J. Jondri and K. M. Lhaksana, "Prediksi Retweet Berdasarkan Feature User-based Menggunakan Metode Klasifikasi Random Forest," *eProceedings ...*, vol. 8, no. 5, pp. 11183–11191, 2021, [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/15633%0Ahttps://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/15633/15346>