

PENERAPAN ALGORITMA RANDOM FOREST UNTUK MEMPREDIKSI JUMLAH SANTRI BARU

Laelatul Hidayah^{1*}, Muhammad Imron Rosadi²

^{1,2}Universitas Yudharta Pasuruan; Jl. Yudharta No. 07 (Pesantren Ngalah) Sengonagung Purwosari Pasuruan, Jawa Timur 67162; 0343-611186.

Received: 22 Agustus 2024

Accepted: 5 Oktober 2024

Published: 12 Oktober 2024

Keywords:

New students; Random Forest; Predictions.

Correspondent Email:

laelatulhidayahela@gmail.com;

Abstrak. Setiap tahun ajaran baru, setiap instansi pendidikan baik formal maupun nonformal melaksanakan proses penerimaan siswa secara rutin, termasuk Pondok Pesantren Ngalah yang membuka penerimaan santri baru dengan 2 gelombang. Namun, pelaksanaan penerimaan santri baru tersebut masih menghadapi kendala seperti jumlah santri baru yang melebihi kuota asrama yang tersedia. Oleh karena itu, dibutuhkan suatu prediksi jumlah santri baru untuk membantu pihak pesantren dalam mengoptimalkan proses penerimaan santri baru. Penerapan algoritma *random forest* pada penelitian ini digunakan untuk memprediksi jumlah santri baru. Hasil nilai akurasi, MAE dan MSE digunakan untuk mengukur kinerja dari algoritma *random forest* dan menentukan nilai parameter terbaik pada prediksi jumlah santri baru. Hasil penelitian ini menunjukkan bahwa algoritma *random forest* memiliki nilai akurasi yang tinggi yaitu sebesar 97,73% dalam memprediksi jumlah santri baru dan nilai MAE serta MSE yang sangat rendah, masing-masing sebesar 0,05 dan 0,18.

Abstract. Every new school year, every educational institution, both formal and non-formal, carries out a routine student admission process, including the Ngalah Islamic Boarding School which opens the admission of new students with 2 batches. However, the implementation of the admission of new students still faces obstacles such as the number of new students who exceed the available dormitory quota. Therefore, a prediction of the number of new students is needed to assist the pesantren in optimizing the process of admitting new students. This study aims to predict the number of new students using the Random Forest algorithm. Accuracy, MAE and MSE measurements are used to measure the performance of the proposed algorithm and determine the best parameter values in predicting the number of new students. The results of this study show that the random forest algorithm has a high accuracy value in predicting the number of new students, which is 97.73% and low MAE and MSE values, which are 0.05 and 0.18, respectively..

1. PENDAHULUAN

Pondok pesantren adalah bentuk pendidikan tradisional di mana siswa tinggal bersama dan belajar di bawah bimbingan kiai. Pondok juga memiliki asrama tempat tinggal santri[1]. Salah

satu Pesantren yang ada di Jawa Timur adalah Pesantren Ngalah. Pesantren Ngalah ini memiliki 16 asrama yang terdiri dari 7 asrama putra, 8 asrama putri, dan 1 asrama anak – anak.

Setelah peneliti melakukan sesi wawancara dengan pihak panitia PPDB (Penerimaan Peserta Didik Baru) Pondok Pesantren Ngalah, dapat disimpulkan bahwa Pondok Pesantren Ngalah membuka penerimaan santri baru dengan 2 gelombang yang biasanya dilakukan saat awal tahun ajaran baru. Jumlah santri yang mendaftar di Pondok Pesantren Ngalah juga mengalami peningkatan dan penurunan setiap tahunnya. Hal ini tentu mempengaruhi kesiapan pihak pesantren dalam hal pemenuhan kebutuhan santri baru terutama tentang pemenuhan tempat tinggal santri (asrama). Oleh sebab itu, untuk mempersiapkan pesantren dalam memenuhi kebutuhan semua santri baru, diperlukan prediksi jumlah santri baru yang akan masuk.

Dengan perkembangan teknologi seperti sekarang ini, kita bisa memanfaatkan Machine Learning dalam berbagai bidang termasuk untuk memprediksi jumlah santri baru di Pondok Pesantren Ngalah. Adapun beberapa algoritma Machine Learning untuk prediksi yaitu Logistic Regression (LR), Support Vector Machine (SVM), Decision tree, Random Forest (RF), Naive bayes, K-Nearest Neighbour (KNN), dan Artificial Neural Network. Diantara algoritma Machine Learning tersebut, algoritma Random Forest menunjukkan nilai akurasi yang paling unggul[2].

Penelitian terkait prediksi yaitu oleh[3] yang menggunakan model random forest untuk memprediksi jumlah mahasiswa baru dengan hasil akurasi mencapai 98 % dengan nilai MSE dan MAE dibawah 5 %.

Untuk tujuan dari penelitian ini yaitu agar kita mengetahui nilai akurasi dari penerapan algoritma random forest untuk memprediksi jumlah santri baru di Pondok Pesantren Ngalah Pasuruan.

2. TINJAUAN PUSTAKA

2.1. Prediksi

Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil[4].

Menurut Kamus Besar Bahasa Indonesia (KBBI), prediksi adalah hasil

dari kegiatan memprediksi atau meramal atau memperkirakan nilai pada masa yang akan datang dengan menggunakan data masa lalu. Adapun beberapa metode analisis untuk prediksi yaitu regresi logistik (logistic regression), random forest, dan artificial neural network. Dari ketiga metode tersebut, nilai akurasi terbaik dihasilkan oleh metode Random Forest dengan capaian akurasi sebesar 79,70% [5].

2.2. Data Mining

Data mining adalah gabungan dari sejumlah disiplin ilmu komputer yang didefinisikan sebagai proses penemuan pola-pola baru dari kumpulan data yang sangat besar, meliputi metode-metode irisan dari artificial intelligence, machine learning, statistics, dan database systems[6]. Penerapan metode data mining sangat membantu untuk menganalisis suatu permasalahan[7].

Secara garis besar data mining dibagi menjadi 2 kategori utama, yaitu :

- a. Deskriptif, yaitu proses menemukan karakteristik penting dari suatu basis data
- b. Prediktif, yaitu proses menemukan pola dari suatu data dengan menggunakan beberapa variabel[8].

2.3. Machine learning

Machine learning didefinisikan sebagai bidang studi yang memberikan komputer kemampuan belajar tanpa diprogram secara eksplisit. *Machine learning* digunakan untuk mengajari mesin cara menangani data dengan lebih efisien. Machine learning berbagai algoritma yang berbeda untuk menyelesaikan masalah data[9].

Machine learning juga mampu menangani data yang kompleks, memberikan efisiensi analitis dan hasil yang objektif[10]

2.4. Random Forest

Algoritma *Random Forest* adalah penyempurnaan lebih lanjut dari algoritma Classification and Regression Tree (CART). Algoritma

Random Forest dikembangkan oleh Breiman yang menggunakan pengacakan untuk membuat pohon keputusan. Output – output dari pohon keputusan ini dikumpulkan menjadi satu kemudian ditentukan penyelesaian masalahnya.

Pengacakan ini dilakukan dengan 2 cara. Pertama, dataset adalah sampel dengan penggantian (bootstrap sampling) lalu yang kedua terjadi pada node keputusan. Pada setiap node, sejumlah prediktor dipilih kemudian algoritma akan menguji semua ambang batas yang mungkin untuk semua variabel dan memilih ambang batas variabel kombinasi yang menghasilkan pemisahan terbaik[11].

Algoritma *random forest* memiliki beberapa kelebihan, termasuk kemampuan untuk meningkatkan akurasi dalam kasus data yang hilang, kemampuan untuk menahan anomali, dan kemampuan untuk penyimpanan data yang efisien. Selain itu, algoritma Random Forest memiliki proses seleksi fitur yang dapat mengambil fitur terbaik, meningkatkan kinerja pada model klasifikasi. Karena fitur seleksi ini, algoritma Random Forest pasti dapat bekerja dengan baik pada data yang besar dengan parameter yang kompleks.[12].

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N \hat{y}_i \quad (1)$$

Keterangan :

N = Jumlah sampel

\hat{y} = prediksi akhir

\hat{y}_i = Prediksi dari pohon ke-i

2.5. Evaluasi Model

Ada banyak parameter yang bisa digunakan untuk mengevaluasi kinerja dari algoritma *random forest* diantaranya MAE (*Mean Average Error*) dan MSE (*Mean Square Error*). Evaluasi MAE (*Mean Average Error*) ini merupakan nilai rata-rata kesalahan dari hasil prediksi dengan data sebenarnya. Sedangkan evaluasi MSE (*Mean Square Error*) mengatur kesalahan

prediksi yang besar karena kesalahan-kesalahan itu dikuadratkan [13].

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

Keterangan :

n : Jumlah sampel

y_i : Nilai target sebenarnya

\hat{y}_i : Nilai prediksi

3. METODE PENELITIAN

Untuk melakukan penelitian dengan baik, diperlukan prosedur yang runtut dan sistematis, seperti:

3.1. Tahapan Prosedur Penelitian

Prosedur penelitian adalah serangkaian langkah yang sistematis dan terstruktur untuk melakukan penelitian dari awal hingga akhir sehingga hasilnya valid dan bisa diandalkan. Didawah ini adalah tahapan prosedur penelitian pada penelitian ini :

1.Studi Literatur

Mengkaji penelitian – penelitian terdahulu yang relevan dengan masalah prediksi jumlah santri baru guna mendukung penelitian ini dan meningkatkan pemahaman peneliti terhadap permasalahan yang ada. Pengetahuan ini diambil dari berbagai jurnal nasional maupun internasional.

2.Pengumpulan Data

Peneliti melakukan pengumpulan data langsung dari Pondok Pesantren Ngalah selaku objek dari penelitian ini. Pengumpulan data ini harus dilakukan dengan benar dan tepat karena ini sangat berpengaruh terhadap hasil penelitian.

3.Pemrosesan Data

Peneliti memproses data dengan menggunakan metode Random Forest.

4.Melakukan Pengujian

Pada tahap ini data hasil prediksi akan dihitung nilai akurasi serta nilai MAE dan MSE.

3.2. Desain Penelitian

Desain penelitian dibuat untuk mengetahui gambaran dari penelitian yang akan dilakukan. Sistem algoritma yang

digunakan untuk memprediksi jumlah santri baru pada penelitian ini adalah algoritma Random Forest. Berikut adalah proses dari algoritma yang akan digunakan :

1. Data yang digunakan adalah data historis pendaftaran santri baru 2 tahun terakhir di Pondok Pesantren Ngalah
2. Proses utama dari penerapan algoritma ini adalah hasil dari perhitungan *Random Forest*

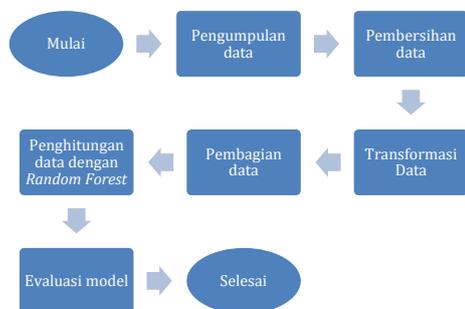
3.3. Analisa

Setelah mendapatkan berbagai informasi yang dibutuhkan dari studi literatur, tahap selanjutnya yaitu tahap analisa. Berikut adalah rincian analisisnya :

3.3.1. Kebutuhan Data

Pada tahap ini, peneliti mengumpulkan berbagai data yang dibutuhkan. Setelah data terkumpul, data akan diolah dengan metode Random Forest. Adapun data yang digunakan adalah data historis penerimaan santri baru di Pondok Pesantren Ngalah.

3.3.2. Tahapan Prediksi



Gambar 3. 2 Alur sistem prediksi *Random forest*

Berikut adalah analisa tahapan prediksi jumlah santri baru dengan aloritma *Random Forest* :

1. Peneliti mengumpulkan data berupa data historis penerimaan santri baru tahun 2022 dan 2023
2. Data dibersihkan dengan menghapus data yang tidak diperlukan
3. Mengubah seluruh data menjadi data numerik

4. Membagi data menjadi dua yaitu data latih dan data uji dengan rasio 90:10, 80:20, 70:30, 60:40, dan 50:50.
5. Menggunakan algoritma Random Forest untuk melatih model pada data latih. Prediksi akhir didapat dengan mengambil rata – rata prediksi dari semua pohon keputusan.
6. Evaluasi model dengan melihat nilai akurasi dan nilai MSE serta MAE.

3.3.3. Dataset

Dataset yang digunakan yaitu data historis penerimaan santri baru di Pondok Pesantren Ngalah pada tahun 2022 dan 2023.

Tabel 3.3.1 Contoh *dataset*

Nama	Xxxxxxx
JK	Laki - Laki
Tempat Lahir	Pasuruan
Tanggal Lahir	08-07-2010
Id yayasan	230024
Asrama	K
Formal	Mts
RT/RW	0
Dusun	Ledug
Kelurahan	Ledug
Kecamatan	Prigen
Kabupaten	Pasuruan
Provinsi	Jawa Timur
Kode pos	67157
Nama Ayah	Yyyyyy
Nama ibu	Zzzzzz

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Setelah mengumpulkan data dengan meminta data historis penerimaan santri baru langsung kepada sekretaris pusat Pondok Pesantren

Ngalah, penulis berhasil mendapatkan sebanyak 2.214 data dari tahun 2022 dan 2023. Dari data yang telah diperoleh, penulis dapat menyimpulkan bahwa ada 7 atribut dari total 17 atribut yang mempengaruhi prediksi jumlah santri baru, yaitu jenis kelamin, asrama, formal, kelurahan, kecamatan, kabupaten dan provinsi.

4.2. Pengolahan Data

4.2.1. Pembersihan Data

Dari perolehan dataset tersebut kemudian dilakukan pembersihan data dengan langkah - langkah sebagai berikut :

1. Mengganti nama kolom untuk membuat kolom lebih mudah dimengerti dan diakses
2. Menghapus baris pertama yang merupakan header untuk memastikan bahwa data yang sebenarnya dimulai dari baris kedua kebawah.
3. Mengisi nilai kosong dengan modus (nilai yang paling sering muncul) untuk kolom kategorikal karena pengisian dengan mode memastikan tidak ada nilai yang tertinggal dan mempertahankan representasi umum dari kategori
4. Mengisi nilai kosong dengan median untuk kolom numerik karena pengisian nilai kosong dengan median untuk kolom numerik tidak mengubah distribusi secara signifikan, sehingga tetap menjaga representasi data.

4.2.2. Transformasi Data

Seluruh data yang ada pada *dataset* dikonversi menjadi data numerik karena algoritma *random forest* hanya dapat bekerja dengan data numerik.

4.2.3. Pembagian Data

Untuk memahami bagaimana jumlah data pelatihan mempengaruhi kinerja model, data ini dibagi menjadi dua, yaitu data latihan dan data uji, dengan rasio perbandingan 90:10, 80:20, 70:30, 60:40, dan 50:50.

4.2.4. Perhitungan Data

Data dihitung nilai prediksinya dengan menerapkan algoritma *random forest*.

```
y_pred = rf.predict(X_test)
print("Accuracy:", accuracy_score(y_test, y_pred))
print("Classification Report:\n", classification_report(y_test, y_pred))
```

Gambar 4.2.1 Source code perhitungan data dengan *random forest*

4.3. Evaluasi Model

Setelah melakukan penerapan algoritma *random forest* dengan rasio pembagian data latihan dan data uji yang berbeda, maka selanjutnya adalah evaluasi model untuk melihat hasil perbandingan dari perbedaan rasio pembagian data yang telah dilakukan dan menentukan mana rasio yang terbaik. Evaluasi model ini dilihat dari nilai akurasi dan nilai MAE serta MSE. Berikut hasil evaluasi model dengan rasio pembagian data yang berbeda - beda :

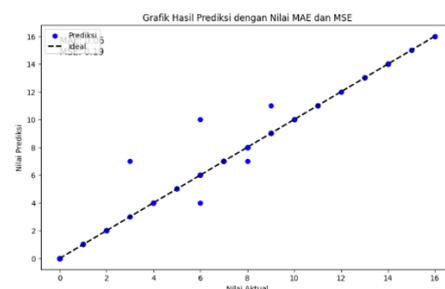
1. Rasio 90 : 10

Accuracy: 0.9773755656108597

Gambar 4.3.1 hasil akurasi rasio 90 : 10

Mean Squared Error (MSE): 0.18552036199095023
Mean Absolute Error (MAE): 0.058823529411764705

Gambar 4.3.2 nilai MAE dan MSE rasio 90 : 10



Gambar 4.3.3 Grafik nilai MAE dan MSE rasio 90 : 10

Gambar diatas menunjukkan hasil akurasi dan nilai MSE dan MAE dari algoritma *random forest* dengan rasio pembagian data sebesar 90 : 10. Gambar tersebut menunjukkan bahwa algoritma pohon random memiliki nilai akurasi yang sangat tinggi sebesar 97,73%. Selain itu, nilai MAE dan MSE algoritma masing-masing sebesar 0,0588

dan 0,1855. Ini menunjukkan bahwa algoritma *random forest* cukup baik dengan tingkat kesalahan prediksi yang relatif rendah.

Berikut adalah 10 data acak yang merupakan perbandingan data aktual dan data prediksi dengan rasio 90 : 10 :

	Aktual	prediksi	Benar
122	16	16	True
1905	9	11	False
213	10	10	True
754	9	9	True
1198	0	0	True
58	1	1	True
1031	0	0	True
1624	8	8	True
1568	0	0	True
2199	6	6	True

Gambar 4.3.4 Perbandingan data aktual dan data prediksi rasio 90 : 10

Dari 10 data acak diatas, model berhasil memprediksi 9 data tepat dan 1 data yang salah prediksi yaitu data ke-1905, dimana pada data aktual data ke-1905 masuk ke kelas 9 (PPN-I) sedangkan model memprediksi data tersebut masuk ke kelas 11 (PPN-K).

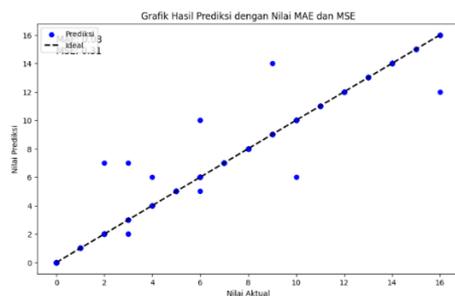
2. Rasio 80 : 20

Accuracy: 0.9751131221719457

Gambar 4.3.5 Hasil akurasi rasio 80 : 20

Mean Squared Error (MSE): 0.30995475113122173
Mean Absolute Error (MAE): 0.07918552036199095

Gambar 4.3. 6 Nilai MAE dan MSE rasio 80 : 20



Gambar 4.3. 7 Grafik nilai MAE dan MSE rasio 80 : 20

Gambar diatas menunjukkan hasil akurasi serta nilai MSE dan MAE dari algoritma *random forest* dengan rasio pembagian data sebesar 80 : 20. Dari gambar diatas bisa dilihat nilai akurasi dari algoritma yang diterapkan sangat tinggi, yaitu mencapai 97,51%. Hal ini menunjukkan bahwa algoritma *random forest* sangat akurat dalam

melakukan prediksi dengan hanya sekitar 2,49% prediksi yang salah.

Dan untuk nilai MAE dan MSE nya masing – masing sebesar 0,0791 dan 0,3099. Hal ini menunjukkan bahwa algoritma *random forest* memiliki performa yang baik dengan kesalahan prediksi yang relatif rendah.

Berikut adalah 10 data acak yang merupakan perbandingan data prediksi dan data aktual dengan rasio 80:20 :

	Aktual	prediksi	Benar
2059	0	0	True
1096	4	4	True
1474	7	7	True
432	0	0	True
951	0	0	True
1412	12	12	True
758	2	2	True
410	6	10	False
1899	0	0	True
2175	0	0	True

Gambar 4.3.8 Perbandingan data aktual dan data prediksi rasio 80 : 20

Dari 10 data acak tersebut, model hanya melakukan 1 kesalahan prediksi pada data ke-410, dimana pada data aktual data tersebut masuk ke kelas 6 (PPN-F) sedangkan model memprediksi data tersebut masuk ke kelas 10 (PPN-J).

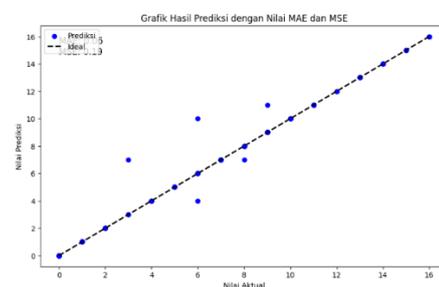
3. Rasio 70 : 30

Accuracy: 0.975867269984917

Gambar 4.3.9 Nilai akurasi rasio 70 : 30

Mean Squared Error (MSE): 0.19457013574660634
Mean Absolute Error (MAE): 0.058823529411764705

Gambar 4.3.10 Nilai MAE dan MSE rasio 70 : 30



Gambar 4.3.11 Grafik Nilai MAE dan MSE rasio 70 : 30

Gambar diatas menunjukkan hasil akurasi dan nilai MSE dan MAE dari algoritma *random forest* dengan rasio pembagian data sebesar 70 : 30. Algoritma *random forest* memiliki

akurasi yang sangat tinggi, yaitu 97,58%, seperti yang ditunjukkan pada gambar.

Selain itu, nilai MAE dan MSE nya masing – masing sebesar 0,0588 dan 0,1945. Ini menunjukkan bahwa algoritma *random forest* cukup baik dengan tingkat kesalahan prediksi yang relatif rendah dengan rasio pembagian data latih dan data uji sebesar 70 : 30.

Berikut adalah 10 data acak yang merupakan perbandingan data aktual dan data prediksi dengan rasio 70 : 30 :

	Aktual	prediksi	Benar
2059	0	0	True
1096	4	4	True
1474	7	7	True
432	0	0	True
951	0	0	True
1412	12	12	True
758	2	2	True
410	6	10	False
1899	0	0	True
2175	0	0	True

Gambar 4.3. 12 Perbandingan data aktual dan data prediksi rasio 70 : 30

Dari 10 data acak diatas, model juga hanya melakukan 1 kesalahan prediksi yang sama seperti rasio 80 : 20 yaitu pada data ke-410, dimana pada data aktual data tersebut masuk ke kelas 6 (PPN-F) sedangkan model memprediksi data tersebut masuk ke kelas 10 (PPN-J).

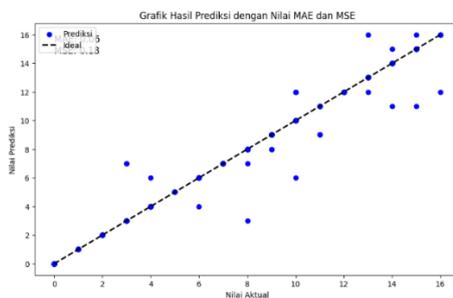
4. Rasio 60 : 40

Accuracy: 0.9751131221719457

Gambar 4.3.13 Nilai akurasi rasio 60:40

Mean Squared Error (MSE): 0.18212669683257918
Mean Absolute Error (MAE): 0.05995475113122172

Gambar 4.3.14 Nilai MAE dan MSE rasio 60 : 40



Gambar 4.3. 15 Grafik nilai MAE dan MSE rasio 60 : 40

Untuk rasio pembagian data latih dan data uji sebesar 60 : 40, algoritma *random forest* menghasilkan nilai akurasi sebesar 97,51 %. Hal ini

menunjukkan bahwa algoritma *random forest* hanya melakukan kesalahan prediksi sebesar 2,49%. Selain itu, untuk nilai MAE dan MSE nya masing-masing hanya sebesar 0,0599 dan 0,1821. Hal ini juga menunjukkan bahwa algoritma *random forest* memiliki tingkat kesalahan yang relatif rendah dalam memprediksi jumlah santri baru.

Berikut adalah 10 data acak yang merupakan perbandingan data aktual dan data prediksi dengan rasio 60 : 40 :

	Aktual	prediksi	Benar
1246	0	0	True
1552	1	1	True
2133	0	0	True
1039	0	0	True
1213	0	0	True
495	7	7	True
1699	0	0	True
279	7	7	True
759	13	12	False
2128	0	0	True

Gambar 4.3.16 Perbandingan data aktual dan data prediksi rasio 60 : 40

Dari 10 data acak diatas dapat dilihat bahwa model memprediksi 9 data yang tepat dan 1 data yang salah yaitu data ke-759, dimana pada data aktual data tersebut masuk ke kelas 13 (PPN-M) sedangkan model memprediksi data tersebut masuk ke kelas 12 (PPN-L).

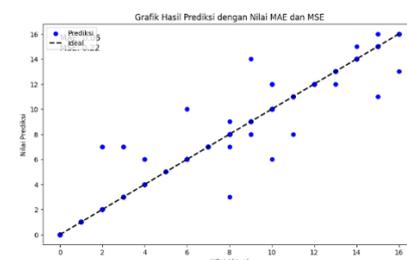
5. Rasio 50 : 50

Accuracy: 0.9773550724637681

Gambar 4.3.17 Nilai akurasi rasio 50:50

Mean Squared Error (MSE): 0.22192028985507245
Mean Absolute Error (MAE): 0.0625

Gambar 4.3.18 Nilai MAE dan MSE rasio 50:50



Gambar 4.3. 19 Grafik nilai MAE dan MSE rasio 50:50

Untuk rasio pembagian data sebesar 50 : 50, algoritma *random forest* menghasilkan nilai akurasi sebesar

97,73 %. Selain itu, untuk nilai MAE dan MSE nya masing-masing hanya sebesar 0,0625 dan 0,2219. Hal ini juga menunjukkan bahwa algoritma *random forest* memiliki tingkat kesalahan yang relatif rendah dalam memprediksi jumlah santri baru.

Berikut adalah 10 data acak yang merupakan perbandingan data aktual dan data prediksi dengan rasio 50 : 50 :

	Aktual	prediksi	Benar
2059	0	0	True
1096	4	4	True
1474	7	7	True
432	0	0	True
951	0	0	True
1412	12	12	True
758	2	2	True
410	6	10	False
1899	0	0	True
2175	0	0	True

Gambar 4.3. 20 Perbandingan data aktual dan data prediksi rasio 50:50

Dari 10 data acak diatas, model juga hanya melakukan 1 kesalahan prediksi yang sama seperti rasio 80 : 20 dan 70 : 30 yaitu pada data ke-410, dimana pada data aktual data tersebut masuk ke kelas 6 (PPN-F) sedangkan model memprediksi data tersebut masuk ke kelas 10 (PPN-J).

5. KESIMPULAN

- Pembersihan dan pengonversian dataset menjadi data numerik merupakan komponen penting dalam prediksi menggunakan algoritma *random forest*.
- Pembagian data latih dan data uji berbeda, rasio pembagian data latih dan data uji pada prediksi algoritma *random forest* tidak berdampak signifikan pada hasil prediksinya. Ini menunjukkan bahwa kerja algoritma *random forest* tetap konsisten dan akurat.
- Rasio pembagian data terbaik sebesar 90:10 memiliki hasil akurasi sebesar 97,73%, dan nilai MAE dan MSE masing-masing sebesar 0,0588 dan 0,1855, masing-masing. Dengan kata lain, algoritma *random forest* cocok untuk memprediksi jumlah murid baru.
- Implementasikan algoritma *Random forest* pada *dataset* yang lebih

kompleks lagi untuk membuktikan performa yang sangat baik dari algoritma *random forest*.

UCAPAN TERIMA KASIH

Penulis sangat berterima kasih kepada orang tua, keluarga, dosen yang mengajar Teknik Informatika, dan semua pihak terkait yang telah mendukung penelitian ini.

DAFTAR PUSTAKA

- P. Wijaya, R. W. Sembiring, and S. S, "Analisis Metode Backpropagation Memprediksi Penerimaan Santri/Wati di Pondok Pesantren Modern Al-Kautsar," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.*, vol. 6, no. 1, p. 140, 2021, doi: 10.30645/jurasik.v6i1.278.
- S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 1, pp. 1–16, 2019, doi: 10.1186/s12911-019-1004-8.
- M. Rianto and R. Yunis, "Analisis Runtun Waktu Untuk Memprediksi Jumlah Mahasiswa Baru Dengan Model Random Forest," vol. 23, no. 1, 2021.
- Z. Zulfauzi and M. N. Alamsyah, "Penerapan Algoritma Naive Bayes Untuk Prediksi Penerimaan Mahasiswa Baru Studi Kasus Universitas Bina Insan Fakultas Komputer," *J. Teknol. Inf. Mura*, vol. 12, no. 02, pp. 156–165, 2020, doi: 10.32767/jti.v12i02.1096.
- H. K. Pambudi, P. G. A. Kusuma, F. Yulianti, and K. A. Julian, "Prediksi Status Pengiriman Barang Menggunakan Metode Machine Learning," *J. Ilm. Teknol. Infomasi Terap.*, vol. 6, no. 2, pp. 100–109, 2020, doi: 10.33197/jitter.vol6.iss2.2020.396.
- S. P. Dewi, N. Nurwati, and E. Rahayu, "Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor," *Build. Informatics, Technol. Sci.*, vol. 3, no. 4, pp. 639–648, 2022, doi: 10.47065/bits.v3i4.1408.
- M. I. Rosadi, C. B. Sanjaya, and L. Hakim, "Klasifikasi Diabetic Retinopathy Menggunakan Seleksi Fitur Dan Support Vector Machine," *J. Resist. (Rekayasa Sist. Komputer)*, vol. 1, no. 2, pp. 109–117, 2018, doi: 10.31598/jurnalresistor.v1i2.312.
- N. Noviyanto, "Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian Penderita COVID-19 Berdasarkan Negara di Benua Asia," *Paradig. - J. Komput. dan Inform.*, vol. 22, no. 2, pp. 183–188, 2020, doi: 10.31294/p.v22i2.8808.
- B. Mahesh, "Machine Learning Algorithms - A Review," *Int. J. Sci. Res.*, vol. 9, no. 1, pp. 381–

- 386, 2020, doi: 10.21275/ART20203995.
- [10] H. Al Azies, F. A. Rohmatullah, H. B. Rochmanto, and D. Putri, "TOWARDS OPTIMIZATION : DATA-DRIVEN APPROACH K-MEDOIDS CLUSTERING ALGORITHM FOR REGIONAL EDUCATION QUALITY," vol. 12, no. 3, 2022.
- [11] E. D. Attanasi and T. C. Coburn, "Random Forest," pp. 1182–1185, 2023, doi: 10.1007/978-3-030-85040-1_265.
- [12] S. Devella, Y. Yohannes, and F. N. Rahmawati, "Implementasi Random Forest Untuk Klasifikasi Motif Songket Palembang Berdasarkan SIFT," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 7, no. 2, pp. 310–320, 2020, doi: 10.35957/jatisi.v7i2.289.
- [13] Sabar Sautomo and Hilman Ferdinandus Pardede, "Prediksi Belanja Pemerintah Indonesia Menggunakan Long Short-Term Memory (LSTM)," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 99–106, 2021, doi: 10.29207/resti.v5i1.2815.