Jurnal Infortech
Volume 7 No. 1 Juni 2025
E-ISSN: 2715-8160

Penerapan Algoritma *Naive Bayes* Pada Penerimaan Santri Baru Pendidikan Pondok Pesantren

Yopi Handrianto[1*], Muadz Zidan AlWafi[2]

[1,2]Universitas Bina Sarana Informatika
e-mail: [1]yopi.yph@bsi.ac.id, [2]muadzzidan123@gmail.com

Diterima	Direvisi	Disetujui
11-03-2025	05-05-2025	12-06-2025

Abstrak - Penerimaan santri baru merupakan tahapan penting dalam sistem pendidikan di sekolah formal maupun di pondok pesantren. Di beberapa pondok pesantren proses penerimaan santri baru sering kali dilakukan secara manual, sehingga rentan terhadap kesalahan data dan penilaian subjektif, yang menyebabkan hasil dari tes seleksi tidak akurat. Seperti yang terjadi di pesantren Shalahuddin Al-Ayyubi Kabupaten Bekasi. Penelitian ini bertujuan untuk membantu pengelola pesantren menetapkan kelulusan tes seleksi santri baru dengan menerapkan algoritma naïve bayes dalam proses penerimaan santri baru, supaya hasil dari tes seleksi bisa lebih akurat dan terkomputerisasi. Penelitian menggunakan pendakatan kuantitatif dengan memanfaatkan data awal sebanyak 47 record, yang dibagi menjadi data testing, 80% untuk data pelatihan dan 20% untuk data pengujian. Hasil penelitian menunjukkan bahwa algoritma naïve bayes mampu memberikan akurasi sebesar 88.89%, dengan precision sebesar 66,67%, recall sebesar 100% dan nilai AUC sebesar 1,000. Dari hasil prediksi, 7 santri dinyatakan diterima dan 2 santri tidak diterima. Terbukti bahwa algoritma Naïve Bayes memiliki tingkat akurasi yang tinggi serta mampu memberikan hasil prediksi yang valid dan terukur.

Kata Kunci: Penerimaan Santri Baru, Pesantren, Algoritma Naïve Bayes

Abstract *- The admission of new students is an important stage in the education system in formal schools and in Islamic boarding schools. In some Islamic boarding schools, the process of accepting new students is often carried out manually, so it is prone to data errors and subjective assessments, which causes the results of the selection test to be inaccurate. As happened at the Shalahuddin Al-Ayyubi Islamic boarding school, Bekasi Regency. This research aims to help pesantren managers determine the passing of the new student selection test by applying the naïve bayes algorithm in the process of accepting new students, so that the results of the selection test can be more accurate and computerized. The study used quantitative training by utilizing 47 initial data, which was divided into testing data, 80% for training data and 20% for testing data. The results show that the naïve bayes algorithm is able to provide an accuracy of 88.89%, with a precision of 66.67%, a recall of 100% and an AUC value of 1,000. From the prediction results, 7 students were declared accepted and 2 students were not accepted. It is proven that the Naïve Bayes algorithm has a high level of accuracy and is able to provide valid and measurable prediction results*

Keywords: Admission of New Students, Islamic Boarding School, Naïve Bayes.

PENDAHULUAN

Pesantren Shalahuddin Al-Ayyubi Bekasi menghadapi tantangan dalam proses seleksi santri baru karena masih menggunakan metode manual. Hal ini menyebabkan subjektivitas dalam penilaian dan berpotensi terjadi kesalahan data. Oleh karena itu, penerapan teknologi dalam proses seleksi santri menjadi solusi yang dapat meningkatkan efisiensi dan akurasi.

Sistem yang dapat mengotomatisasi proses penerimaan siswa baru menjadi semakin penting seiring dengan kemajuan teknologi informasi (Sunarto & Rosyani, 2023), Penggunaan metode analisis berbasis algoritma dapat menjadi solusi untuk meningkatkan efisiensi dan akurasi dalam proses penerimaan santri. *Naive Bayes* adalah metode klasifikasi yang mengacu pada teorema Bayes dan menganggap setiap atribut bersifat independen, memungkinkan setiap atribut memberikan kontribusi terhadapa nilai pada variabel kelas secara terpisah (Rachman & Handayani, 2021). Kelebihan dari Algoritma *Naïve Bayes* yaitu mudah diimplementasikan, memerlukan jumlah data sedikit yang dibutuhkan untuk klasifikasi dan lebih cepat

dalam perhitungan (Arfanda et al., 2021). Algoritma ini dapat digunakan untuk menganalisis berbagai kriteria penerimaan, seperti nilai tes, prestasi akademik, kehadiran dalam tes wawancara, hingga latar belakang keluarga. Dengan menerapkan Algoritma *Naïve Bayes*, pondok pesantren dapat menyaring calon santri dengan lebih akurat dan meminimalisir subjektivitas dalam proses seleksi.

Data mining merupakan salah satu konsentrasi ilmu di bidang statistika dan komputer dengan tujuan akhir melakukan ekstraksi data dari sebuah set data menggunakan metode yang cerdas, dan merubah data tersebut dalam struktur yang komprehensif untuk penggunaan lebih lanjut (Wahyuddin et al., 2016).

Terdapat berbagai teknik dalam *data mining* yang dimanfaatkan untuk menggali pola serta pengetahuan dari sekumpulan data, diantaranya : (Rahayu et al., 2018).

a. Klasifikasi (*Classification*).

Melalui metode ini, data diklasifikasikan ke dalam kelompok atau kategori tertentu berdasarkan atribut atau karakteristik yang dimilikinya. Contoh teknik yang digunakan antara lain *Decision Tree*, *Artificial Neural Network*, dan algoritma klasifikasi lainnya.

b. Regresi (*Regression*).

Regresi dimanfaatkan untuk membentuk model hubungan antara variabel independen dan dependen. Teknik ini berfungsi untuk memperkirakan nilai kontinu berdasarkan pola hubungan linier maupun non-linier di antara variabel-variabel tersebut.

c. Klastering (*Clustering*).

Klastering merupakan proses pengelompokan data ke dalam klaster berdasarkan kemiripan atribut atau fitur tertentu. Contoh metode yang termasuk dalam teknik ini adalah algoritma *K-Means* dan *Hierarchical Clustering.*

d. Asosiasi (*Association*).

Teknik ini dimanfaatkan untuk mengidentifikasi keterkaitan atau asosiasi antar item atau variabel dalam suatu himpunan data. Algoritma populer seperti *Apriori* dan *Eclat* digunakan untuk mengekstraksi aturan asosiasi dari data transaksional.

Naïve Bayes adalah metode prediktif yang memanfaatkan prinsip probabilitas sederhana berdasarkan teorema Bayes, dengan asumsi adanya independensi kuat antar variabel.. Menurut Bayes (terutama *Naïve Bayes*), independensi yang kuat pada fitur berarti bahwa ada atau tidaknya fitur lain dalam data yang sama tidak mempengaruhi fitur pada data tersebut. (Muttaqin et al., 2023)

RapidMiner, yang kini dikenal sebagai Altair Ai Studio, adalah perangkat lunak *opensource* yang digunakan sebagai solusi untuk melakukan analisis prediktif, *data mining* dan *text mining*. Untuk membatu pengguna membuat pilihan terbaik, *RapidMiner* menggunakan berbagai teknik prediksi dan deskripsi. *RapidMiner* menyediakan lebih dari 500 operator untuk keperluan *data mining*, yang mencakup fungsi *input, output*, pra-pemrosesan data, serta visualisasi (Dennis, 2013).

METODOLOGI PENELITIAN

Tahapan pada penelitian ini sebagai berikut:

1. Identifikasi Masalah

Mengidentifikasi permasalahan yang ingin diselesaikan. Permasalahan yang ditangani dalam penelitian ini berhubungan dengan proses penerimaan santri baru yang masih dilakukan secara manual, sehingga memerlukan sistem yang lebih efektif untuk membantu pengambilan keputusan.

2. Studi Literatur

Tahap ini melibatkan kajian terhadap literatur atau penelitian-penelitian sebelumnya yang relevan dengan topik penerimaan santri baru dan teknik yang akan digunakan, yaitu algoritma *Naïve Bayes.*

3. Pengumpulan Data

Pengumpulan data dilakukan dengan mengumpulkan informasi serta data yang diperlukan untuk membangun model penerimaan santri menggunakan algoritma *Naïve Bayes*.

4. Pengolahan Data

Setelah data terkumpul, proses pengolahan dilakukan untuk membersihkan data, mentransformasi data, membagi data menjadi data pelatihan dan data pengujian agar sesuai dengan kebutuhan analisis menggunakan algoritma *Naïve Bayes.*

5. Menentukan Hasil Akurasi Data

Mengukur performa model *Naïve Bayes* yang telah dibuat dalam memprediksi calon santri yang layak diterima. Akurasi model dihitung untuk menilai seberapa baik model dapat bekerja dalam proses seleksi santri baru berdasarkan data yang ada.

6. Kesimpulan

Tahap terakhir adalah menyimpulkan hasil penelitian berdasarkan hasil analisis dan evaluasi yang telah dilakukan. Kesimpulan ini mencakup penjelasan tentang efektivitas model *Naïve Bayes* dalam membantu proses penerimaan santri serta saran untuk pengembangan sistem yang lebih baik dimasa depan.

Populasi dalam penelitian ini adalah seluruh calon santri yang mendaftar di pesantren Shalahuddin Al-Ayyubi selama periode 2023-2025. Data populasi ini mencakup data-data pendaftar seperti:

1. Nilai tes tulis (tes psikologi, akademik).
2. Nilai tes Al-Quran (hafalan, bacaan dan tajwid).
3. Hasil wawancara.

HASIL DAN PEMBAHASAN

Hasil dari penelitian mencangkup penerapan algoritma naive bayes untuk menentukan penerimaasn santri baru di pondok pesantren Shalahuddin Al-Ayyubi, pengolahan data dilakukan dengan menggunakan s*oftware* Altair Ai Studio 2024.1.0, berikut data yang akan digunakan pada penelitian yang dilakukan (lihat Gambar 1).

Data yang sudah terkumpul dilakukan pre-processing data melalui beberapa tahapan, yaitu:

1. *Data Cleaning*

Pada tahap *data cleaning*, terdapat beberapa faktor yang dapat memengaruhi kualitas data, seperti nilai yang hilang (*missing value*), data duplikat, dan outliers. Namun, berdasarkan hasil pemeriksaan dataset ini, tidak ditemukan adanya nilai yang hilang, data duplikat, maupun outliers.

2. *Transformation Data*

Tahap ini bertujuan untuk mengubah jenis format file atau tipe data dari masing-masing atribut pada dataset ini.

3. *Split Data*

Pembagian data menjadi dua bagian yaitu data pelatihan dan data pengujian. Pada dataset ini menggunakan rasio 80% data pelatihan dan 20% data pengujian..

Data yang sudah melewati tahap pengolahan data/pre-processing data selanjutnya diterapkan ke dalam proses model *Naïve Bayes* dengan menggunakan beberapa operator seperti terlihat pada Gambar 2.

Evaluasi model menggunakan confusion matrix akan menghasilkan matrix yang terdiri dari *True Positive, True Negative, False Positive Dan False Negative*. Matrix tersebut dapat dilihat pada Gambar 3.

Berdasarkan Gambar 3 menunjukkan bahwa jumlah *True Positive (TP)* sebanyak 2, *False Negative (FN)* sebanyak 0, *False Positive (FP)* sebanyak 1 dan *True Negative (TN)* sebanyak 6. Dari data tersebut nilai *Accuracy, Precision* dan *Recall* dapat dihitung secara manual sebagai berikut:

a. *Accuracy* = (TP+TN)/(TP+TN+FP+FN) + (2 +6)/(2+6+1+0) = (2 + 6)/(2+6+1+0) = 8/9 = 0,8889 = 88.89%
b. *Precision* = TP/(TP+FP) + (2)/(2+1) = 2/3 = 0,6667 = 66.67%
c. *Recall* = TP/(TP+FN) + (2)/(2+0) = 2/2 = 1 = 100.00%

Untuk menilai kinerja algoritma *Naive Bayes*, digunakan grafik *Receiver Operating Characteristic* (ROC) dan nilai *Area Under Curve* (AUC). ROC adalah grafik yang menunjukan hubungan antara Rate Positif Benar (TPR) dan Rate Negatif Positif (FPR) pada berbagai nilai ambang (Lihat Gambar 4).

Terlihat pada Gambar 4 bahwa kurva hampir sempurna mendekati sudut kiri atas grafik. Hal ini menunjukkan bahwa model memiliki tingkat akurasi yang sangat tinggi dalam membedakan antara data calon santri yang LULUS dan TIDAK LULUS.

Nilai AUC (*Area Under Curve*) yang diperoleh adalah 1.000, yang berarti bahwa model memiliki kemampuan prediksi yang sempurna, Dengan kata lain:

1. Model mampu mengidentifikasi seluruh calon santri yang benar-benar LULUS (*True Positive*) tanpa memberikan prediksi salah terhadap calon yang TIDAK LULUS (*False Positive*).
2. Nilai AUC = 1.000 menunjukkan bahwa algoritma Naïve Bayes bekerja dengan optimal pada dataset yang digunakan.

Dengan data awal sebanyak 47 *record* yang kemudian dilakukan pembagian untuk data testing sebanyak 9 *record* terbukti algoritma *Naïve Bayes* dapat menghasilkan akurasi yang sangat baik sebesar 88.89%. Yaitu 7 santri diterima dan 2 santri tidak diterima.

KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan, algoritma *Naïve Bayes* terbukti mampu memberikan hasil yang sangat baik dalam menentukan penerimaan santri baru. Dengan data awal sebanyak 47 *record*, dilakukan pembagian untuk data testing sebanyak 9 *record*. Algoritma *Naïve Bayes* menghasilkan akurasi sebesar 88.89%, *Precision* sebesar 66.67%, *Recall* sebesar 100.00%, dan nilai AUC sebesar 1.000. Dari hasil prediksi, 7 santri diterima dan 2 santri tidak diterima.

Algoritma *Naïve Bayes* memiliki tingkat akurasi yang tinggi serta mampu memberikan hasil prediksi yang valid dan terukur sehingga dapat dijadikan rekomendasi untuk pengambilan keputusan dalam proses peneriman santri baru dilingkungan pondok pesantren.

Implementasi sistem prediksi berbasis algoritma *Naïve Bayes* diharapkan dapat diintegrasikan kedalam sistem informasi pesantren agar pengambilan keputusan dapat dilakukan secara *real-time* dan terdokumentasi dengan baik. Selain itu, penelitian ini dapat dijadikan sebagai dasar untuk pengembangan sistem rekomendasi santri baru berbasis kecerdasan buatan dimasa mendatang.

REFERENSI

Aprilla Dennis. (2013). Belajar data mining dengan rapidminer. *Innovation and Knowledge Management in Business Globalization: Theory & Practice, Vols 1 and 2*, *5*(4), 1–5. http://esjournals.org/journaloftechnology/archive/vol1no6/vol1no6_6.pdf%5Cnhttp://www.airccse.org/journal/nsa/5413nsa02.pdf

Arfanda, I., Ramdhan, W., & Yusda, R. A. (2021). Naive bayes dalam menentukan penerima bantuan langsung tunai. *Digital Transformation Technology*, *1*(1), 9–16. https://doi.org/10.47709/digitech.v1i1.1091

Muttaqin, Widiyanto, wahyu wijaya, Mandias, Muhammad Munsarif Green Ferry Stenly Richard Pungus, A. W., Wiranti Kusuma Hapsari, S. A. H., Aslam Fatkhudin, Pasnur, E. F. B., & Mochammad Anshori, Suryani, N. S. (2023). *Pengenalan data mining* (R. Watriantos (ed.); 1st, April 2 ed., Issue July). Penerbit Yayasan Kita Menulis. https://repository.upy.ac.id/4946/1/FullBook Pengenalan Data Mining.pdf

Rachman, R., & Handayani, R. N. (2021). Klasifikasi algoritma naive bayes dalam memprediksi tingkat kelancaran pembayaran sewa teras umkm. *Jurnal Informatika*, *8*(2), 111–122. https://doi.org/10.31294/ji.v8i2.10494

Rahayu, P., Sudipa, I. G. I., Suryani, Surachman, A., Ridwan, A., Darmawiguna, I. G. M., Sutoyo, M., Slamet, I., Harlina, S., & May Sanjaya, I. M. (2018). *Buku ajar data mining* (Vol. 1, Issue January 2024). https://www.researchgate.net/publication/377415198_BUKU_AJAR_DATA_MINING

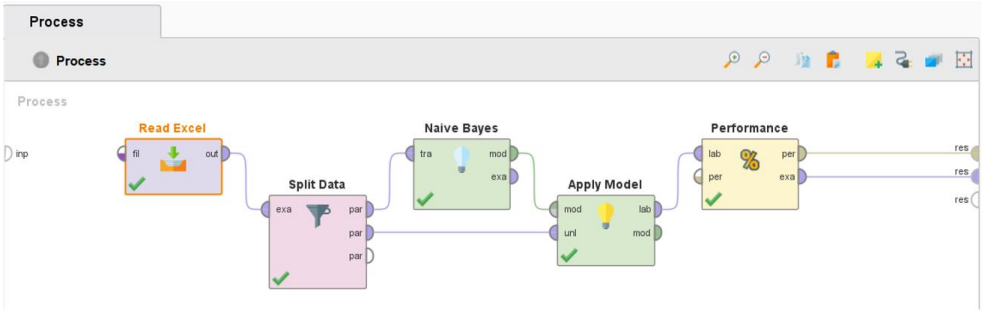
Sunarto, J. B., & Rosyani, P. (2023). Perancangan sistem informasi penerimaan peserta didik baru di mi al-hidayah arco berbasis web. *LOGIC: Jurnal Ilmu Komputer …*, *25*(2), 73–79. https://journals.upi-yai.ac.id/index.php/TEKINFO/article/view/3936/2988

Wahyuddin, Rismayani, Sihotang, jay I., Aisa, S., Gunawan, H., Masturoh, N. T. S., Radiyah, U., Harlina, Z. G. S., & Muslihi, M. T. (2016). Data warehouse dan data mining. In *GAES-PACE Book Publisher* (Issue January). Penerbit Yayasan Kita Menulis. https://www.researchgate.net/publication/377494554_Data_Warehouse_dan_Data_Mining

	A	B	C	D	E	F	G	H	I
1	NO	NAMA	TES TULIS AGAMA	TES TULISPSIKOLOG	TES WAWANCARA	TES HAFALAN	TES TAJWID	TES TILAWAH	HASIL
2	1	KUMARA JIBRANI AL-GHIFARI	0	0	0	0	0	0	TIDAK LULUS
3	2	MAULANA LUTHAN SYAFIQ	40	35	80	80	75	75	LULUS
4	3	DEWA NAUFAL RAFA BAZLA	35	25	85	75	70	70	LULUS
5	4	MUHAMMAD LABIB	40	40	85	75	75	75	LULUS
6	5	SATRIA PRABASWARA	40	45	90	70	70	80	LULUS
7	6	MUHAMMAD REYVAN ARSHAVIN	20	25	70	60	50	50	TIDAK LULUS
8	7	NAUFAL FAJRI NUZULA	45	40	90	75	75	75	LULUS
9	8	AHMAD HISYAM NAUFAL	40	50	80	85	80	80	LULUS
10	9	ABDURRAHMAN RAFI	35	40	80	75	70	85	LULUS
11	10	DAWAM MUNAJAT	25	40	85	60	65	65	LULUS
12	11	MUHAMMAD IBNU ABBAS	25	30	75	60	60	65	LULUS
13	12	M. IBNU KHALIS HABSORO	30	35	80	70	70	70	LULUS
14	13	ILYAS IBNU AENUROFIQ	40	40	75	75	75	70	LULUS
15	14	ABDUS SAMI	40	45	85	80	75	75	LULUS
16	15	SYIHAB AL-HAFIZ	40	40	75	75	75	70	LULUS
17	16	ADLY MUHAMMAD DZIKRI	40	40	80	75	70	70	LULUS
18	17	ZAIS ALKHOIR	30	20	70	55	55	60	TIDAK LULUS
19	18	MUHAMMAD NIZAM AL GHIFARI	20	20	70	50	60	50	TIDAK LULUS
20	19	SATRIA MAHARDIKA HIDAYAT	40	40	85	80	75	75	LULUS
21	20	HAIKAL FAJAR SYA'BAN	45	50	85	80	80	75	LULUS
22	21	MUHAMMAD LAFIFUDIN MUTASHAM	40	30	80	70	60	70	LULUS
23	22	ZIYAD ALKHIYAR	20	25	70	70	50	50	LULUS
24	23	AZZAM FATHIN AL FARISI	35	35	80	70	75	75	LULUS
25	24	GHOZY FATHIN AL VARO	40	45	80	70	70	75	LULUS
26	25	MUHAMMAD MAHAREZA DWI PUTRA	45	50	85	70	70	70	LULUS
27	26	ARMEN HALIM NOER	50	40	75	80	75	75	LULUS
28	27	FAUZAN ABDURRAZAK	45	50	85	60	70	70	LULUS
29	28	ARSAHNDY ALFARIZI	40	45	80	70	70	75	LULUS
30	29	MUHAMMAD LINGGA TRY SYAHPUTRA	20	20	55	55	60	50	TIDAK LULUS
31	30	IBRAHIM IRSYAD ALTANSYAH	20	20	70	50	70	55	TIDAK LULUS
32	31	MUHAMMAD AZZAM KHOIRULLAH	20	20	55	55	50	70	TIDAK LULUS
33	32	AMMAR	15	20	60	70	50	50	TIDAK LULUS
34	33	MUHAMMAD ABDULLAH HANIF	40	45	70	80	70	70	LULUS
35	34	MUHAMMAD RASYID ALBAIHAQI	30	35	70	70	70	75	LULUS
36	35	DZAKY ACHMAD	20	20	50	70	55	55	TIDAK LULUS
37	36	SAYYID ABDULLAH AZZAM	15	20	70	55	60	50	TIDAK LULUS
38	37	MUHAMMAD SYAMIL ABDULLAH	30	20	50	50	70	50	TIDAK LULUS
39	38	LUTHFI SAKHI ZAIDAN	30	15	70	50	55	50	TIDAK LULUS
40	39	AHMAD SYAHRUL RAMADHAN	40	30	75	70	60	70	LULUS
41	40	REI FATHAN ATTAYA ZAMBIEK	45	40	80	75	70	70	LULUS
42	41	UMAR SOBRUN JAMILAN	30	25	75	70	70	60	LULUS
43	42	MUHAMMAD AZZAM	30	40	70	80	70	70	LULUS
44	43	ANAS NUR IMANSYAH	40	40	75	70	60	60	LULUS
45	44	MUHAMMAD FADLAN	35	45	80	70	70	70	LULUS
46	45	ALTHAF FAUZI RAHMAN	45	40	75	70	75	75	LULUS
47	46	AHMAD IBN FAUZI	50	45	80	75	80	75	LULUS
48	47	RIDHO HADI NUGRAHA	40	45	75	70	75	75	LULUS
49									

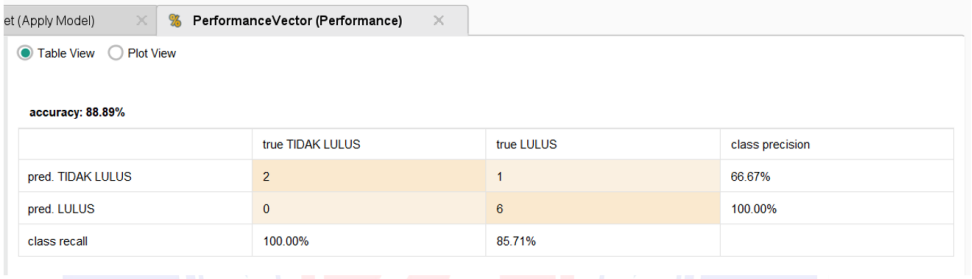
Sumber: Hasil Penelitian (2024)

Gambar 1. Data Penelitian



Sumber: Hasil Penelitian (2024)

Gambar 2. *Process Model Naïve Bayes*



	true TIDAK LULUS	true LULUS	class precision
pred. TIDAK LULUS	2	1	66.67%
pred. LULUS	0	6	100.00%
class recall	100.00%	85.71%	

Sumber: Hasil Penelitian (2024)

Gambar 3. *Confusion Matrix*

Sumber: Hasil Penelitian (2024)

Gambar 4. Kurva ROC