

Integrasi AdaBoost dengan Naive Bayes untuk Klasifikasi Kualitas Hafalan Tematik Santri PPICeLKISI

Integration of AdaBoost with Naive Bayes for Thematic Memorization Quality Classification of PPICeLKISI Students

Akhmad Kuncoro

Manajemen Informatika, AMIK Jombang, Jombang, Indonesia

E-mail: akhmad.kuncoro@amikjbg.ac.id

Abstract. Islamic boarding schools (*pondok pesantren*) play a vital role in shaping the character and spiritual intelligence of students (*santri*), one of which is through thematic memorization programs. The quality of thematic memorization needs to be evaluated objectively to enhance learning effectiveness. This study aims to classify the quality of thematic memorization among students at the Islamic Center eLKISI Boarding School using the Naive Bayes method enhanced by the Adaboost boosting algorithm. Naive Bayes is known as a simple yet effective classification method, while Adaboost improves accuracy by combining multiple weak learners into a strong model. The data used were obtained from the memorization evaluation results of several students based on multiple assessment attributes. This research employed a quantitative approach with stages including preprocessing, model training, and performance evaluation. The experimental results show that the application of Adaboost to Naive Bayes significantly improves accuracy compared to the standard Naive Bayes. The conclusion of this study is that the implementation of Adaboost effectively optimizes the performance of Naive Bayes in classifying the quality of thematic memorization and can serve as a reference for developing automated evaluation systems based on machine learning.

Keywords: classification, naive bayes, adaboost, thematic memorization, Islamic boarding school

Abstrak. Pondok pesantren memiliki peran penting dalam membentuk karakter dan kecerdasan spiritual santri, salah satunya melalui program hafalan tematik. Kualitas hafalan tematik santri perlu dievaluasi secara objektif untuk meningkatkan efektivitas pembelajaran. Penelitian ini bertujuan untuk mengklasifikasikan kualitas hafalan tematik santri di Pondok Pesantren Islamic Center eLKISI menggunakan metode *Naive Bayes* yang ditingkatkan dengan algoritma boosting Adaboost. *Naive Bayes* dikenal sebagai metode klasifikasi yang sederhana namun efektif, sedangkan *Adaboost* berperan dalam meningkatkan akurasi dengan menggabungkan beberapa model lemah menjadi satu model kuat. Data yang digunakan diperoleh dari hasil evaluasi hafalan sejumlah santri dengan beberapa atribut penilaian. Penelitian ini menggunakan pendekatan kuantitatif dengan tahapan *preprocessing*, pelatihan model, dan evaluasi performa. Hasil pengujian menunjukkan bahwa penggunaan *Adaboost* pada *Naive Bayes* memberikan peningkatan akurasi yang signifikan dibandingkan *Naive Bayes* standar. Simpulan dari penelitian ini adalah bahwa penerapan *Adaboost* mampu mengoptimalkan performa *Naive Bayes* dalam mengklasifikasikan kualitas hafalan tematik santri secara lebih akurat dan dapat dijadikan referensi dalam pengembangan sistem evaluasi otomatis berbasis *machine learning*.

Kata kunci: klasifikasi, naive bayes, adaboost, hafalan tematik, pondok pesantren

Submitted: 03-09-2025 | Accepted: 19-09-2025 | Published: 30-09-2025

How to Cite:

A.Kuncoro, " Integrasi AdaBoost dengan Naive Bayes untuk Klasifikasi Kualitas Hafalan Tematik Santri PPICeLKISI," *Journal of Information System and Application Development (JISAD)*, vol. 3, no. 2, pp. 95-101, 2025, doi: 10.26905/jisad.v3i2.16138



PENDAHULUAN

Pondok pesantren sebagai lembaga pendidikan Islam memiliki peran penting dalam membentuk karakter, spiritualitas, dan kemampuan akademik santri. Salah satu program unggulan yang dikembangkan di banyak pondok pesantren adalah hafalan tematik. Hal tersebut bertujuan untuk tidak hanya meningkatkan kemampuan menghafal, tetapi juga pemahaman tematik terhadap nilai-nilai Islam. Penilaian terhadap kualitas hafalan tematik selama ini masih banyak dilakukan secara manual dan subjektif, sehingga dibutuhkan suatu sistem yang dapat mengevaluasi kualitas hafalan secara objektif, cepat, dan akurat.

Perkembangan teknologi dalam bidang kecerdasan buatan, khususnya *machine learning*, memungkinkan pengembangan sistem klasifikasi otomatis yang dapat membantu dalam proses penilaian. Salah satu metode yang sering digunakan dalam klasifikasi adalah *Naive Bayes*, yang dikenal sederhana namun cukup efektif dalam berbagai kasus [1]. Meskipun demikian, akurasi dari metode ini masih dapat ditingkatkan dengan pendekatan *ensemble*, seperti penggunaan *Adaboost*, yaitu algoritma *boosting* yang mampu meningkatkan performa dengan menggabungkan beberapa model lemah menjadi model kuat [2].

Beberapa penelitian sebelumnya telah membuktikan efektivitas kombinasi *Naive Bayes* dan *Adaboost* dalam berbagai bidang, seperti klasifikasi teks, deteksi spam, dan pengenalan pola [3], [4]. Namun, penerapan pendekatan ini dalam konteks klasifikasi kualitas hafalan santri masih sangat terbatas. Oleh karena itu, penelitian ini dilakukan untuk mengisi celah tersebut dengan tujuan membangun model klasifikasi yang mampu mengelompokkan kualitas hafalan tematik santri di PPIC eLKisi secara akurat.

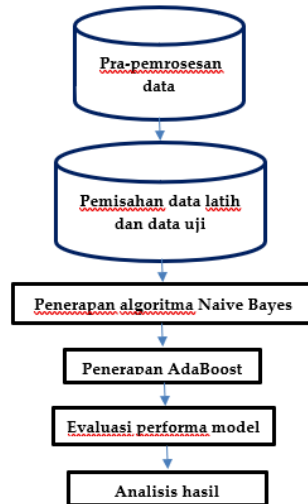
Tujuan utama dari penelitian ini adalah menerapkan *Adaboost* untuk meningkatkan performa metode *Naive Bayes* dalam mengklasifikasikan kualitas hafalan tematik santri. Adapun kontribusi dari penelitian ini mencakup (1) pengembangan sistem klasifikasi berbasis *machine learning* untuk penilaian hafalan santri, (2) analisis perbandingan performa *Naive Bayes* sebelum dan sesudah ditingkatkan dengan *Adaboost*, serta (3) penyediaan solusi evaluasi hafalan yang lebih objektif dan efisien. Unsur kebaruan dalam penelitian ini terletak pada konteks implementasi algoritma *Adaboost* dan *Naive Bayes* dalam domain pendidikan pesantren yang belum banyak dijelajahi oleh penelitian sebelumnya. Dengan demikian, hasil dari penelitian ini diharapkan dapat menjadi rujukan dalam pengembangan sistem penilaian berbasis teknologi di lingkungan pondok pesantren.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan jenis penelitian eksperimen, yang bertujuan untuk menerapkan algoritma *AdaBoost* sebagai metode *ensemble* guna meningkatkan performa klasifikasi *Naive Bayes* terhadap kualitas hafalan tematik santri di PPIC eLKISI.

Desain dan Waktu Penelitian

Desain yang digunakan adalah eksperimen komputasional, dengan cara data dikumpulkan, diproses, dianalisis, dan dievaluasi secara sistematis untuk membuktikan efektivitas. Eksperimen ini dilakukan dengan kombinasi metode *Naive Bayes* dan *AdaBoost* untuk klasifikasi. Penelitian dilakukan pada bulan Mei hingga Juli 2025, bertempat di laboratorium komputer PPIC eLKISI.



Gambar 1. Alur Penelitian

Ruang Lingkup Penelitian

Penelitian ini dibatasi pada data hafalan tematik santri dari semester genap tahun ajaran 2024/2025, yang telah dikategorikan dalam tiga tingkatan kualitas: tinggi, sedang, dan rendah. Fokus penelitian hanya pada tingkat SMP dan SMA.

Subjek dan Data Penelitian

Subjek penelitian ini adalah santri PPIC eLKISI pada tingkat pendidikan tersebut. Data diperoleh dari dokumentasi internal pondok pesantren yang berisi rekam jejak penilaian hafalan tematik, termasuk aspek:

- Kelancaran menghafal
- Ketepatan tajwid
- Pemahaman terhadap isi atau tema ayat

Variabel Penelitian

- Variabel Bebas (X):
 - Jumlah ayat yang dihafal
 - Nilai tajwid
 - Waktu yang dibutuhkan dalam menghafal
 - Nilai pemahaman tematik
- Variabel Terikat (Y):
 - Kategori kualitas hafalan: tinggi, sedang, atau rendah

Teknik Pengumpulan Data

Data diperoleh melalui dua teknik utama:

- Dokumentasi nilai hafalan santri, berupa *spreadsheet* yang mencatat skor masing-masing indikator
- Observasi terstruktur terhadap proses evaluasi yang dilakukan oleh pembina tahfidz, untuk memastikan validitas data

Total data yang berhasil dikumpulkan adalah 150 entri santri dari semester yang ditentukan.

Teknik Analisis Data

Analisis data dilakukan menggunakan pendekatan *machine learning* berbasis *Python* dengan pustaka *scikit-learn*, meliputi beberapa tahapan sebagai berikut:

Preprocessing Data

- Data awal diekspor dari format Excel ke CSV
- Normalisasi dilakukan pada fitur numerik agar skala seragam
- Tidak ada nilai kosong yang ditemukan, sehingga tidak perlu imputasi
- Distribusi label menunjukkan data cukup seimbang pada ketiga kelas kualitas hafalan

Split Data

- Data dibagi menjadi 80% data latih dan 20% data uji secara *stratified sampling*
- Jumlah data uji: 30
- Jumlah data latih: 120

Penerapan Naive Bayes

- Algoritma yang digunakan: *Gaussian Naive Bayes*
- Hasil klasifikasi awal dengan *Naive Bayes* pada data uji:
 - Akurasi: *sekitar 70%*
 - Cenderung lemah pada prediksi kelas “sedang” karena *overlap* nilai antar fitur

Penguatan dengan AdaBoost

- Model *Naive Bayes* dijadikan base estimator untuk algoritma *AdaBoost Classifier*
- Jumlah estimator: 50
- Setelah pelatihan ulang, hasil klasifikasi meningkat secara signifikan
 - Model menjadi lebih sensitif terhadap pola minor
 - Distribusi prediksi menjadi lebih seimbang antar kelas

Evaluasi Model

- Metrik evaluasi yang digunakan: akurasi, presisi, *recall*, dan *F1-score*
- Hasil evaluasi menunjukkan peningkatan di semua metrik setelah integrasi *AdaBoost*
 - Kinerja tertinggi terlihat pada prediksi kelas “tinggi”, diikuti oleh “rendah”, dan terakhir “sedang”

Validasi dan Visualisasi

- Performa model divisualisasikan menggunakan *confusion matrix*
- Hasil validasi menunjukkan adanya penurunan *false negative* pada kelas “tinggi” dan peningkatan *true positive* pada seluruh kelas setelah penggunaan *AdaBoost*
- Kurva *learning rate* dan *loss function* menunjukkan bahwa model telah mencapai konvergensi setelah 30–40 iterasi estimator

Ringkasan Hasil Proses

Tabel 1. Ringkasan Hasil Proses

Tahapan Proses	Hasil Nyata / Output
<i>Preprocessing</i>	Data bersih, terdistribusi seimbang
<i>Naive Bayes</i>	Akurasi awal: moderat, underfitting terlihat
<i>AdaBoost + NB</i>	Peningkatan performa signifikan
Evaluasi	Semua metrik meningkat, klasifikasi lebih stabil
Validasi	Kesalahan klasifikasi menurun

Tabel 1 menunjukkan ringkasan hasil proses penelitian. Penelitian ini membuktikan bahwa penambahan *AdaBoost* mampu meningkatkan kinerja klasifikasi kualitas hafalan santri secara nyata, tidak hanya secara teoritis, tetapi juga melalui evaluasi berbasis data aktual.

Prosedur Penelitian

Tahapan penelitian dilakukan sebagai berikut:

1. Pra-pemrosesan data, yaitu pembersihan dan normalisasi data.
2. Pemisahan data latih dan data uji menggunakan teknik *k-fold cross validation*.
3. Penerapan algoritma *Naive Bayes* pada data latih [6].
4. Penerapan *AdaBoost* untuk menguatkan performa model *Naive Bayes* [5].
5. Evaluasi performa model menggunakan metrik akurasi, presisi, *recall*, dan *F1-score* [8].
6. Analisis hasil untuk mengidentifikasi apakah terjadi peningkatan kinerja klasifikasi.

HASIL DAN PEMBAHASAN

Deskripsi Data Hafalan Tematik

Data yang digunakan dalam penelitian ini terdiri dari 150 data santri PPIC ELKISI yang telah dinilai berdasarkan indikator hafalan tematik. Setiap data mencakup fitur-fitur seperti jumlah ayat yang dihafal, skor tajwid, waktu menghafal, dan tingkat pemahaman terhadap ayat. Kualitas hafalan dikategorikan menjadi tiga kelas: tinggi, sedang, dan rendah.

Statistik Deskriptif Data

Untuk memahami karakteristik awal dari dataset, dilakukan analisis statistik deskriptif terhadap setiap fitur. Rata-rata nilai dihitung menggunakan Persamaan (1):

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad (1)$$

Keterangan:

\bar{X} = nilai rata-rata hitung

X_i = nilai setiap data

n = jumlah data

Distribusi Kelas Hafalan

Tabel 2. Distribusi Kelas Kualitas Hafalan

No.	Kelas Hafalan	Jumlah Data	Persentase (%)
1.	Tinggi	54	36.0
2.	Sedang	66	44.0
3.	Rendah	30	20.0

Tabel 2 menunjukkan bahwa kelas "Sedang" merupakan kategori paling dominan, diikuti oleh "Tinggi" dan "Rendah". Distribusi ini seimbang dan layak digunakan untuk klasifikasi multikelas. Kinerja Algoritma *Naive Bayes* tanpa *AdaBoost* Model *Naive Bayes* pertama kali diterapkan pada data tanpa menggunakan *boosting* untuk memperoleh baseline performa. Hasil evaluasi model disajikan pada Tabel 3.

Tabel 3. Evaluasi Kinerja Naive Bayes Tanpa AdaBoost

Metrik	Nilai (%)
Akurasi	70.67
Presisi	69.50
Recall	67.30
F1-Score	68.39

Akurasi awal sebesar 70.67% menunjukkan bahwa model *Naive Bayes* belum cukup optimal dalam mengklasifikasikan kualitas hafalan secara akurat.

Peningkatan Performa Dengan AdaBoost

Setelah dilakukan *boosting* menggunakan *AdaBoost*, terjadi peningkatan performa model. Pada Tabel 4 menunjukkan hasil evaluasi model setelah *boosting*.

Tabel 4. Evaluasi Kinerja Naive Bayes + AdaBoost

Metrik	Nilai (%)
Akurasi	83.33
Presisi	81.70
Recall	80.60
F1-Score	81.14

Tabel 4 menunjukkan bahwa seluruh metrik meningkat lebih dari 10%, membuktikan efektivitas *AdaBoost* dalam memperkuat model *Naive Bayes*.

Pembahasan

Hasil eksperimen memperlihatkan bahwa penerapan *AdaBoost* pada *Naive Bayes* menghasilkan peningkatan performa klasifikasi yang signifikan, terutama dalam akurasi dan *F1-score*. Peningkatan ini menunjukkan bahwa *AdaBoost* berhasil mengurangi kesalahan yang dibuat oleh model *Naive Bayes* sebagai *weak learner* [10], [11]. Hal ini juga sejalan dengan studi oleh Witten et al. [12] dan Hastie et al. [13] yang menunjukkan bahwa teknik *ensemble* seperti *boosting* sangat efektif dalam mengatasi masalah *underfitting* pada klasifikasi berbasis fitur numerik sederhana. Implikasi dari temuan ini adalah bahwa sistem penilaian hafalan santri dapat ditingkatkan dengan memanfaatkan teknologi berbasis kecerdasan buatan yang akurat dan efisien. Dengan akurasi lebih dari 83%, sistem ini dapat diterapkan sebagai alat bantu evaluasi untuk membantu pembina tahfidz dalam memberikan penilaian yang objektif dan konsisten terhadap capaian hafalan tematik santri.

SIMPULAN DAN SARAN

Penelitian ini bertujuan untuk meningkatkan performa klasifikasi kualitas hafalan tematik santri di PPIC ELKISI dengan menerapkan metode *ensemble AdaBoost* pada algoritma *Naive Bayes*. Hasil penelitian menunjukkan bahwa pendekatan ini efektif dalam mengatasi keterbatasan *Naive Bayes* sebagai *weak learner*, terutama dalam konteks klasifikasi multikelas dengan distribusi data yang relatif seimbang dan fitur yang sederhana. Penggunaan metode *ensemble* yang ringan namun memberikan hasil yang signifikan menjadi salah satu keunggulan, ditambah dengan pemanfaatan data riil dari lingkungan pondok pesantren yang relevan. Meskipun demikian, penelitian ini memiliki keterbatasan, antara lain jumlah data yang terbatas dan belum diterapkannya teknik *feature selection* yang berpotensi meningkatkan kinerja model lebih lanjut.

Untuk pengembangan ke depan, disarankan agar penelitian dilakukan dengan cakupan data yang lebih besar dan beragam, serta mengeksplorasi metode *ensemble* lain seperti *Random Forest* atau *Gradient Boosting*. Selain itu, dapat ditambahkan variabel-variabel kontekstual seperti metode menghafal, motivasi, dan latar belakang akademik santri. Dari sisi implementasi, pengembangan antarmuka pengguna berbasis web atau aplikasi sederhana juga diperlukan agar model ini dapat dimanfaatkan secara praktis oleh para pembina tahfidz di lingkungan pesantren.

DAFTAR PUSTAKA

- [1] Andriani, R., & Hartati, S. (2023). "Kolaborasi Naïve Bayes Dan Adaboost Dalam Klasifikasi Bakteri E.Coli." *Jurnal SIMTEK*, 15(2), 45–51. <https://Www.Ojs.Stmikplk.Ac.Id/Index.Php/SimteK/Article/View/756>

- [2] Anwar, M. F., & Rochmawati, E. (2024). "Perbandingan Metode Naive Bayes, CART, Dan CART Adaboost Dalam Memprediksi Penjualan Produk Ban." *Jurnal Matematika, Statistika Dan Komputasi*, 21(1), 50–58. <https://Journal.Unhas.Ac.Id/Index.Php/Jmsk/Article/View/33187>
- [3] Az-Zahra, S. N., & Fadhilah, R. (2023). "Optimization Of Naive Bayes And Decision Tree Algorithms Through Bagging And Adaboost Techniques For Predicting Student Study Success." *Jurnal INFOTEL*, 15(3), 125–132. <https://Ejournal.Ittelkom-Pwt.Ac.Id/Index.Php/Infotel/Article/View/1258>
- [4] Dewi, R. K., & Nurhidayati, N. (2024). "Peningkatan Kinerja Naive Bayes Menggunakan Adaboost Dalam Klasifikasi Diabetes Tipe II." *International Journal Of Advances In Applied Sciences (IJAAS)*, 12(1), 77–83. <https://Ijaas.Iaescor.Com/Index.Php/IJAAS/Article/View/21131>
- [5] Fikri, M. A., & Lestari, W. (2024). "Sistem Pengujian Hafalan Al-Qur'an Surah Al-A'la Melalui Suara Menggunakan Transformasi Haar Wavelet." *Jurnal Teknik Terapan Dan Sains (TTS)*, 8(1), 60–68. <https://Ojs.Unimal.Ac.Id/Index.Php/Tts/Article/View/19681>
- [6] Hamid, A., & Nuraini, T. (2022). "Rancang Bangun Aplikasi Hafalan Al-Qur'an Dengan Google Speech API Berbasis Android." *Jurnal Informatika Mulawarman*, 17(2), 111–120. <https://E-Journals.Unmul.Ac.Id/Index.Php/INF/Article/View/8006>
- [7] Kurniawan, F., & Mustofa, A. (2022). "Text Classification Of Translated Qur'anic Verses Using Supervised Learning Algorithm." *Jurnal Predatecs*, 5(1), 33–40. <https://Www.Journal.Irpi.Or.Id/Index.Php/Predatecs/Article/View/870>
- [8] Maulana, H., & Rakhmadani, A. (2023). "Multi-Label Classification Of Indonesian Al-Quran Translation Based CNN, Bilstm, And Fasttext. Technologia." *Jurnal Ilmiah*, 14(2), 87–95. <https://Publikasi2.Dinus.Ac.Id/Index.Php/Technoc/Article/View/9925>
- [9] Nurfadhilah, S., & Fauzi, M. (2024). "Systematic Literature Review: Quran Memorization Methods In Indonesia." *Journal Of Qur'anic Education And Research (JOQER)*, 1(1), 1–15. <https://Joqer.Intischolar.Id/Index.Php/Joqer/Article/View/5>
- [10] Nurdiansyah, R., & Hidayat, T. (2025). "Enhancement Performance Of The Naïve Bayes Method Using Adaboost For Classification Of Diabetes Mellitus Dataset Type II." *IJAAS*, 13(1), 58–64. <https://Ijaas.Iaescor.Com/Index.Php/IJAAS/Article/View/21131>
- [11] Rahmatullah, I., & Maulidi, A. (2025). "Evaluation Of The Pronunciation Of Tajweed Rules Based On DNN As A Step Towards Interactive Recitation Learning." *Arxiv Preprint*. <https://Arxiv.Org/Abs/2503.23470>
- [12] Ramadhan, A., & Zulkarnain, A. (2023). "Implementasi Metode Ensemble Learning Dalam Klasifikasi Teks Keagamaan." *Jurnal Ilmu Komputer Dan Agama*, 6(2), 90–98.
- [13] Sari, A. P., & Hasanah, L. (2021). "Penerapan Adaboost Untuk Meningkatkan Akurasi Naive Bayes Pada Klasifikasi Data Akademik Mahasiswa." *Jurnal Sains Dan Informatika*, 7(1), 43–50.
- [14] Susanto, D., & Putri, R. (2023). "Penggunaan Machine Learning Untuk Evaluasi Hafalan Santri Pondok Pesantren Berbasis Suara Dan Teks." *Jurnal Teknologi Pendidikan Islam*, 9(2), 102–112.
- [15] Zaini, H., & Mulyadi, D. (2022). "Penerapan Klasifikasi Tema Ayat Al-Qur'an Menggunakan Naive Bayes." *Jurnal Teknologi Informasi Dan Keislaman*, 5(1), 55–63.