Journal of Information System, Applied, Management, Accounting and Research.
http://journal.stmikjayakarta.ac.id/index.php/jisamar ,
jisamar@stmikjayakarta.ac.id , jisamar2017@gmail.com
e-ISSN: 2598-8719 (Online), p-ISSN: 2598-8700 (Printed) , Vol. 10 No.3 (August 2026)

PERBANDINGAN KINERJA METODE STACKING ENSEMBLE DAN NAÏVE BAYES DALAM KLASIFIKASI PENYAKIT ALZHEIMER BERBASIS DATA KLINIS

Alya Rafina[1*], Lindy Almadiani[2], Zuldarmaini[3], Marrylinteri Istoningtyas[4]

Program Studi Teknik Informatika[1],
Program Studi Teknik Informatika[2],
Program Studi Teknik Informatika[3],
Program Studi Teknik Informatika[4]
Fakultas Ilmu Komputer[1], Fakultas Ilmu Komputer[2],
Fakultas Ilmu Komputer[3], Fakultas Ilmu Komputer[4]
Universitas Dinamika Bangsa[1], Universitas Dinamika Bangsa[2],
Universitas Dinamika Bangsa[3], Universitas Dinamika Bangsa[4]

*Correspondent Author: alyarafina03@gmail.com[1]

Authors Email: alyarafina03@gmail.com[1],
lindyalmadiani1715@gmail.com[2], zdarmaini@gmail.com[3],
Marrylinteri@unama.ac.id [4]

Received: May 28, 2026**. Revised:** July 12, 2026**. Accepted:** July 15, 2026**. Issue Period:** Vol.10 No.3 (2026), Pp. 821-833

Abstrak: Penyakit Alzheimer adalah penyebab utama demensia di seluruh dunia, tetapi metode diagnosis yang menggunakan pencitraan medis seperti MRI memiliki masalah biaya yang tinggi dan akses yang sulit, terutama di rumah sakit atau pusat layanan Kesehatan primer. Penelitian ini berupaya menciptakan dan memperbandingkan efektivitas model klasifikasi yang menggunakan data klinis dengan metode Stacking Ensemble dan Naïve Bayes. Model Stacking Ensemble dibuat dengan menggabungkan Random Forest, XGBoost, dan Support Vector Machine sebagai model dasar, serta Logistic Regression sebagai model meta-learner, sementara Gaussian Naïve Bayes digunakan sebagai model acuan. Dataset yang digunakan berasal dari Kaggle, terdiri dari 2.149 rekaman data pasien dan 32 fitur prediktor setelah dilakukan pra-pemrosesan, lalu dibagi menjadi data latih sebesar 80% dan data uji 20% menggunakan metode stratified split. Evaluasi dilakukan dengan menguji pada data uji dan menggunakan metode 5-fold stratified cross-validation. Hasil pengujian menunjukkan bahwa metode Stacking Ensemble mencapai akurasi sebesar 94,65%, presisi 93,88%, recall 90,79%, serta F1-score 92,31%.Angka tersebut jauh lebih baik dibandingkan metode Naïve Bayes yang hanya menghasilkan akurasi sebesar 77,21%. Hasil validasi silang memperkuat temuan tersebut dengan rata-rata akurasi 95,11% untuk metode Stacking Ensemble. Penelitian ini menunjukkan bahwa metode Stacking Ensemble lebih efektif dan konsisten dalam mengklasifikasikan penyakit Alzheimer menggunakan data klinis dibandingkan dengan model Naïve Bayes yang hanya menggunakan satu pendekatan.

DOI: 10.52362/jisamar.v10i3.2535

Journal of Information System, Applied, Management, Accounting and Research.
http://journal.stmikjayakarta.ac.id/index.php/jisamar ,
jisamar@stmikjayakarta.ac.id , jisamar2017@gmail.com
e-ISSN: 2598-8719 (Online), p-ISSN: 2598-8700 (Printed) , Vol. 10 No.3 (August 2026)

Kata kunci: klasifikasi Alzheimer; stacking ensemble; naïve bayes; data klinis; machine learning

Abstract: *Alzheimer's disease is the main reason people develop dementia around the world. However, diagnostic methods that use medical imaging, like MRI scans, are usually costly and not easily available, especially in hospitals and basic healthcare centers. This study is focused on creating and evaluating different classification models using clinical data, and it compares how well the Stacking Ensemble method and the Naïve Bayes method perform. The Stacking Ensemble model uses Random Forest, XGBoost, and Support Vector Machine as the base models, and Logistic Regression acts as the meta-learner.Gaussian Naïve Bayes is also used as the baseline model. The dataset came from Kaggle and includes 2,149 patient records with 32 predictor features after some cleaning and preparation. The data were split into 80% for training and 20% for testing using a stratified method. The model's performance was checked on the testing data and confirmed using 5-fold stratified cross-validation. The experimental results show that the Stacking Ensemble method got an accuracy of 94.65%, precision of 93.88%, recall of 90.79%, and an F1-score of 92.31%. These results do much better than the Naïve Bayes method, which only reached an accuracy of 77.21%. Moreover, cross-validation showed that the proposed model is reliable, achieving an average accuracy of 95.11% for the Stacking Ensemble. These results show that the Stacking Ensemble approach works better and more reliably than the Naïve Bayes model when classifying Alzheimer's disease based on clinical data.*

Keywords: *Alzheimer classification; stacking ensemble; naïve bayes; clinical data; machine learning*

I. PENDAHULUAN

Penyakit Alzheimer tetap menjadi penyebab paling umum dari demensia di seluruh dunia dan memiliki dampak yang sangat besar terhadap kesehatan masyarakat secara global. Menurut data *World Health Organization* (WHO) tahun 2023, lebih dari 55 juta orang mengalami demensia, dan angka ini diperkirakan akan naik menjadi sekitar 78 juta pada tahun 2030.Penyakit Alzheimer menyumbang antara 60 hingga 70 persen dari semua kasus demensia tersebut. Beban ekonomi yang ditimbulkan setiap tahun mencapai lebih dari 1,3 triliun dolar Amerika, sehingga penting untuk mendiagnosis sejak dini agar bisa melakukan tindakan lebih awal untuk memperlambat perkembangan penyakit dan meningkatkan kualitas hidup pasien [1].

Diagnosis Alzheimer sebelumnya cenderung menggunakan pemeriksaan citra seperti *Magnetic Resonance Imaging* (MRI). Meskipun cara ini cukup berhasil, ada beberapa masalah nyata yang muncul, seperti biaya yang cukup mahal, prosedur yang perlu dilakukan secara langsung, serta ketergantungan yang besar pada kemampuan tenaga medis untuk memahami hasilnya. Kondisi ini mendorong pencarian metode alternatif yang lebih efektif dan hemat biaya, salah satunya dengan menggunakan data medis sederhana seperti usia, hasil tes kognitif, dan riwayat kesehatan untuk mendeteksi penyakit Alzheimer di tahap awal [2].

Dalam beberapa tahun terakhir, pendekatan pembelajaran *ensemble* banyak diterapkan untuk meningkatkan hasil klasifikasi model karena kemampuannya menggabungkan beberapa algoritma sekaligus, sehingga menghasilkan prediksi yang lebih stabil dan akurat. Penelitian [3] mengenalkan kerangka kerja *deep ensemble* yang menggabungkan data *World Health Organization* (MRI) dan data klinis, serta terbukti dapat membedakan pasien sehat, penderita *Mild Cognitive Impairment* (MCI), serta penderita Alzheimer dengan tingkat akurasi yang lebih baik. Temuan ini memperkuat keyakinan bahwa menggabungkan data klinis dengan model *machine learning* yang lebih canggih memiliki kemungkinan besar dalam menciptakan sistem diagnosis neurologis yang lebih efektif.

Namun, sebagian besar penelitian sebelumnya masih berfokus pada model yang menggunakan *World Health Organization* (MRI), sementara penggunaan data klinis yang lebih mudah diperoleh di fasilitas

DOI: 10.52362/jisamar.v10i3.2535

Journal of Information System, Applied, Management, Accounting and Research.
http://journal.stmikjayakarta.ac.id/index.php/jisamar ,
jisamar@stmikjayakarta.ac.id , jisamar2017@gmail.com
e-ISSN: 2598-8719 (Online), p-ISSN: 2598-8700 (Printed) , Vol. 10 No.3 (August 2026)

kesehatan primer masih belum banyak dikaji. Hal ini menciptakan perbedaan dalam penelitian tentang pengembangan model klasifikasi yang lebih fleksibel dan bisa digunakan untuk kelompok dengan ciri-ciri yang berbeda [4]. Penelitian [5] berusaha mengatasi masalah tersebut dengan menggunakan pendekatan pembelajaran *ensemble* berbasis *Random Forest* serta data *longitudinal*, dan berhasil mencapai tingkat akurasi dalam mendeteksi tahap awal Alzheimer lebih dari 90%.Meskipun demikian, penelitian ini belum memperhitungkan secara lengkap variasi dari variabel klinis maupun belum melakukan pengujian pada berbagai kelompok populasi. Selaras dengan itu, studi [6] menekankan pentingnya menggabungkan seleksi fitur dan metode *ensemble* yang berhasil meningkatkan akurasi diagnosis hingga 93%, meskipun volume data yang digunakan masih terbatas..

II. METODE DAN MATERI

2.1. Penelitian Terdahulu

Beberapa penelitian sebelumnya sudah menganalisis penggunaan *machine learning* dan *ensemble learning* untuk mengklasifikasikan penyakit berdasarkan data medis. Penelitian yang dilakukan oleh Djaka dan Nurul [7] membuat model pembelajaran *ensemble* (*Random Forest*, *Logistic Regression*, dan *XGBoost*) untuk mendeteksi dini Sindrom Ovarium Polikistik (PCOS) dan berhasil mencapai akurasi 95% serta skor F1 sebesar 92%, dengan hasil yang lebih baik dibandingkan menggunakan algoritma tunggal seperti *K-Nearest Neighbor* (KNN).

Albani, Kurniawan, dan Tania [8] melakukan penelitian perbandingan terhadap sembilan algoritma dalam memprediksi penyakit jantung.Hasil menunjukkan bahwa *CatBoost* memiliki performa terbaik dengan akurasi 98% dan *F1-score* 0,980.Algoritma lainnya seperti *Random Forest*, *XGBoost*, dan *LightGBM* juga menunjukkan hasil yang stabil. Pada topik yang lebih dekat dengan penelitian ini, Sidik [9] membandingkan beberapa algoritma seperti *Logistic Regression*, *Decision Tree*, *Random Forest*, *Support Vector Machine* (SVM), dan *K-Nearest Neighbor* (KNN) untuk mendeteksi penyakit Alzheimer secara dini. Hasil menunjukkan bahwa *Random Forest* merupakan algoritma yang terbaik dengan akurasi 98% dan *recall* 100%, sedangkan *Support Vector Machine* (SVM) memiliki kinerja paling rendah dengan akurasi hanya 40%.

Rudini dan Ernawan [5] menggunakan metode *soft voting ensemble learning* dengan beberapa algoritma seperti *XGBoost*, *Random Forest*, *AdaBoost*, dan *Gradient Boosting* untuk memprediksi demensia Alzheimer.Awalnya akurasi model berada di kisaran 0,89 hingga 0,92, namun setelah menerapkan teknik Random *Oversampling* (ROS) dan *Grid Search*, akurasi meningkat menjadi 0,93 hingga 0,94. Nurhalizah, Setiawan, dan Ardianto [10] membandingkan lima algoritma yaitu *Naïve Bayes*, *Random Forest*, *ANN*, *XGBoost*, dan *Support Vector Machine* (SVM) menggunakan dataset yang sama dengan penelitian ini yang berisi 2.149 data.Hasilnya menunjukkan bahwa *XGBoost* merupakan model yang terbaik dengan akurasi mencapai 95%, sedangkan *Naïve Bayes* dan *Support Vector Machine* (SVM) memiliki akurasi terendah, yaitu 83%.

Secara umum, penelitian tersebut menunjukkan bahwa algoritma *ensemble* seperti *Random Forest* dan *XGBoost* selalu lebih baik daripada algoritma tunggal dalam mengklasifikasikan data medis. Namun, penelitian sebelumnya dalam kasus Alzheimer [5], [9], [10] biasanya hanya membandingkan algoritma secara terpisah atau menggunakan metode *voting ensemble* yang sederhana, dan belum banyak melakukan eksplorasi terhadap pendekatan *Stacking Ensemble* yang menggabungkan beberapa *base learner* yang berbeda dengan bantuan *meta-learner*. Penelitian ini bertujuan mengatasi celah tersebut dengan menggunakan metode *Stacking Ensemble* yang menggabungkan beberapa algoritma seperti *Random Forest*, *XGBoost*, dan *Support Vector Machine* (SVM) sebagai model dasar, serta *Logistic Regression* sebagai model pembuat keputusan utama.Metode ini dibandingkan langsung dengan *Naïve Bayes* menggunakan data klinis pasien Alzheimer, sehingga memberikan pendekatan baru dalam kombinasi algoritma untuk meningkatkan tingkat akurasi dan kestabilan dalam proses klasifikasi.

2.2.Metode yang Digunakan

Stacking Ensemble adalah metode pembelajaran *ensemble* yang menggabungkan beberapa algoritma klasifikasi agar hasil prediksi lebih baik. Berbeda dengan cara voting biasa, *Stacking Ensemble* menggunakan hasil ramalan dari beberapa model dasar sebagai masukan baru untuk model tingkat kedua, yang disebut *meta-learner*, sehingga keputusan akhir yang dihasilkan lebih stabil dan kurang rentan terhadap *overfitting* dibandingkan model yang hanya menggunakan satu model saja [11].

DOI: 10.52362/jisamar.v10i3.2535

Journal of Information System, Applied, Management, Accounting and Research.
http://journal.stmikjayakarta.ac.id/index.php/jisamar ,
jisamar@stmikjayakarta.ac.id , jisamar2017@gmail.com
e-ISSN: 2598-8719 (Online), p-ISSN: 2598-8700 (Printed) , Vol. 10 No.3 (August 2026)

Dalam penelitian ini, arsitektur *Stacking Ensemble* dibuat dalam dua tingkat. Level 0 terdiri dari tiga model pembelajar dasar, yaitu *Random Forest*, *XGBoost*, dan *Support Vector Machine* (SVM), sedangkan Level 1 menggunakan *Logistic Regression* sebagai model pembelajar meta. *Random Forest* dipilih karena kemampuannya yang baik dalam menghadapi data klinis yang rumit dan ketahanannya terhadap data yang tidak normal [12], [13]. *XGBoost* dipilih karena mampu memproses nilai yang hilang secara otomatis dan memiliki performa komputasi yang efisien pada data berbentuk tabel [14], [15]. *Support Vector Machine* (SVM) dipilih karena kemampuannya dalam membedakan kelas yang saling tumpang tindih dengan menggunakan pendekatan kernel *non-linear*, yang sesuai dengan sifat data klinis Alzheimer [16], [17]. *Logistic Regression* berfungsi sebagai *meta-learner* yang belajar tingkat keyakinan dari setiap model dasar melalui fungsi sigmoid, sehingga dapat menghasilkan keputusan akhir dalam bentuk probabilitas diagnosis Alzheimer atau bukan Alzheimer [18], [19].

Sebagai model acuan, penelitian ini menggunakan *Gaussian Naïve Bayes*, yaitu algoritma klasifikasi yang bekerja berdasarkan *Teorema Bayes* dan mengasumsikan bahwa setiap atribut saling independen [20]. Meskipun asumsi tersebut tidak selalu terpenuhi sepenuhnya dalam data medis yang umumnya saling berkaitan, *Naïve Bayes* tetap bisa digunakan sebagai acuan dasar karena komputasinya cepat dan mudah diimplementasikan, sehingga bisa menjadi penanda objektif untuk mengukur seberapa besar peningkatan kinerja yang dicapai oleh *Stacking Ensemble* [21].

Evaluasi kinerja kedua model dilakukan dengan menggunakan confusion matrix serta empat metrik utama, yaitu akurasi, presisi, *recall*, dan *F1-score* [22], [23], dengan persamaan (1) – (4) seperti berikut:

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

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

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

$$F1\text{-}Score = 2\frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

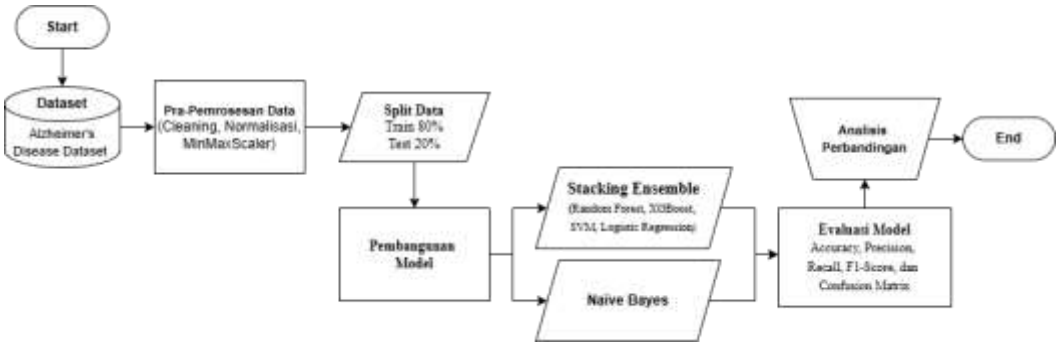
Di mana TP (*True Positive*), TN (*True Negative*), FP (*False Positive*), dan FN (*False Negative*) merupakan bagian utama dari matriks kebingungan [24]. Menggunakan keempat metrik ini sangat penting karena dataset penelitian memiliki distribusi kelas yang tidak seimbang, sehingga recall menjadi indikator penting untuk mengevaluasi kemampuan model dalam mendeteksi kasus Alzheimer secara akurat [25], [26].

2.3. Tahapan Penelitian

Penelitian ini dilakukan dengan tahapan-tahapan yang terlihat pada *flowchart* Gambar 1, mulai dari mengumpulkan dataset, mempersiapkan data sebelum diproses, membagi data, membuat model, mengevaluasi model, sampai melakukan analisis perbandingan hasilnya..

DOI: 10.52362/jisamar.v10i3.2535

Journal of Information System, Applied, Management, Accounting and Research.
http://journal.stmikjayakarta.ac.id/index.php/jisamar ,
jisamar@stmikjayakarta.ac.id , jisamar2017@gmail.com
e-ISSN: 2598-8719 (Online), p-ISSN: 2598-8700 (Printed) , Vol. 10 No.3 (August 2026)



Gambar 1. Alur Metode Penelitian

2.4. Data Penelitian

Dataset yang digunakan adalah *Alzheimer's Disease Dataset* yang disediakan oleh Rabie El Kharoua melalui platform Kaggle. Dataset tersebut terdiri dari 2.149 data pasien dengan 35 atribut yang berbeda. Setelah menghilangkan atribut *PatientID* dan *DoctorInCharge*, digunakan 32 fitur prediktor dan satu variabel target yaitu Diagnosis (0 = Bukan Alzheimer, 1 = Alzheimer). Dataset tidak mengandung missing value ataupun data duplikasi. Ringkasan dataset ditampilkan pada Tabel 1.

Tabel I. Ringkasan Dataset

Karakteristik	***Keterangan***
Sumber	Kaggle (Alzheimer's Disease Dataset)
Jumlah Record	2.149 data pasien
Total Atribut	35 atribut (32 fitur prediktor + 1 target)
Missing Values	0 (tidak ada nilai kosong)
Distribusi Kelas	Non-Alzheimer (0): 1.389 (64,6%) Alzheimer (1): 760 (35,4%)
Rentang Usia	60 – 90 tahun (rata-rata 74,91 tahun)
Tipe Klasifikasi	Biner (Binary Classification)

2.5. Pra-Pemrosesan Data

Pemeriksaan awal terhadap dataset menunjukkan bahwa tidak ada nilai yang kosong atau data yang sama di seluruh 35 kolom, sehingga tidak perlu dilakukan pengisian ulang data. Kolom *PatientID* dan *DoctorInCharge* dihilangkan karena tidak membantu dalam memprediksi, sehingga tersisa 32 fitur yang dapat digunakan untuk prediksi. Fitur kategorikal dalam dataset sudah disajikan dalam bentuk angka, jadi tidak perlu dilakukan proses *encoding* tambahan.

Normalisasi dilakukan pada 12 fitur numerik kontinu yaitu *BMI*, *AlcoholConsumption*, *PhysicalActivity*, *DietQuality*, *SleepQuality*, *CholesterolTotal*, *CholesterolLDL*, *CholesterolHDL*, *CholesterolTriglycerides*, *MMSE*, *FunctionalAssessment*, dan *ADL* menggunakan metode *MinMaxScaler* dengan rentang [0,1], sesuai persamaan (5) berikut:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (5)$$

DOI: 10.52362/jisamar.v10i3.2535

Journal of Information System, Applied, Management, Accounting and Research.
http://journal.stmikjayakarta.ac.id/index.php/jisamar ,
jisamar@stmikjayakarta.ac.id , jisamar2017@gmail.com
e-ISSN: 2598-8719 (Online), p-ISSN: 2598-8700 (Printed) , Vol. 10 No.3 (August 2026)

Di mana X' adalah hasil normalisasi, X adalah nilai asli dari data tersebut, sedangkan Xmin dan Xmax masing-masing merupakan nilai terkecil dan terbesar dari fitur tersebut. Proses scaling diajarkan pada data latih terlebih dahulu, lalu digunakan pada data uji agar informasi tidak bocor ke data uji.

2.6. Pembagian Data

Dataset dibagi dengan metode *Stratified Train-Test Split* dengan perbandingan 80:20, sehingga didapatkan 1.719 data untuk *training* dan 430 data untuk *testing*, dengan jumlah kelas dalam setiap bagian tetap seimbang.

2.7. Pembangunan Model

Penelitian ini membandingkan dua metode pengklasifikasian, yaitu *Stacking Ensemble* dan *Gaussian Naïve Bayes*. Model stacking menggunakan tiga model dasar yaitu *Random Forest*, *XGBoost*, dan *Support Vector Machine* (SVM), serta *Logistic Regression* sebagai model pembuat keputusan utama dengan metode *5-fold cross-validation*. Semua model menggunakan parameter standar tanpa penyesuaian *hyperparameter*.

2.8. Evaluasi Model

Evaluasi dilakukan dengan dua cara, yaitu pengujian *hold-out* pada data uji dan *5-fold stratified cross-validation* pada seluruh dataset. Kinerja model diukur dengan menggunakan beberapa metrik seperti *Accuracy*, *Precision*, *Recall*, *F1-Score*, serta *Confusion Matrix*.

III. PEMBAHASAN DAN HASIL

3.1. Deskripsi Dataset dan Hasil Pra-pemrosesan

Pengecekan terhadap dataset Alzheimer menunjukkan bahwa kolom Diagnosis berfungsi sebagai target klasifikasi dengan dua kategori, yaitu 0 yang menunjukkan bukan Alzheimer dan 1 yang menunjukkan Alzheimer. Kolom *PatientID* berisi angka unik yang mengidentifikasi pasien, sedangkan *DoctorInCharge* hanya memiliki satu nilai yang sama di semua catatan, sehingga kedua kolom tersebut tidak digunakan sebagai bahan untuk memprediksi. Ringkasan hasil pemeriksaan dataset terlihat pada Tabel 2.

Tabel 2. Ringkasan Dataset Hasil Pengecekan

Karakteristik	***Hasil Pengecekan***
Jumlah Record	2.149 pasien
Jumlah kolom awal	35 kolom
Kolom target	Diagnosis
Kelas target	0 = Non-Alzheimer; 1 = Alzheimer
Nilai kosong	0
Data duplikat	0
Kolom yang dihapus	PatientID dan DoctorInCharge
Jumlah fitur prediktor akhir	32 fitur

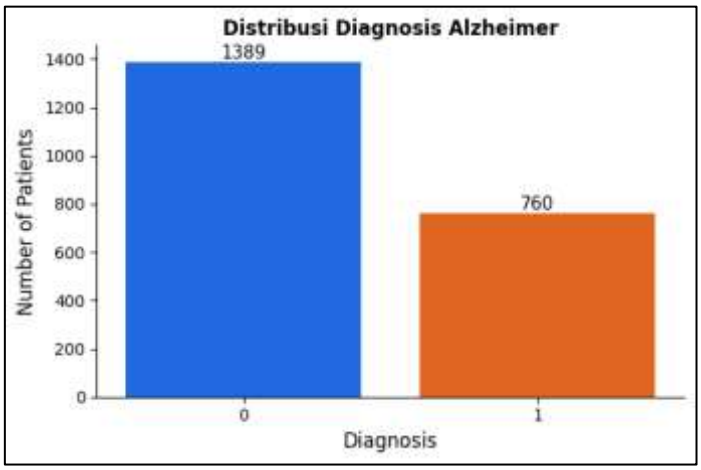
Berdasarkan distribusi target, kelas Non-Alzheimer terdapat 1.389 record atau 64,63%, sedangkan kelas Alzheimer terdapat 760 *record* atau 35,37%, seperti yang ditampilkan pada Tabel 3 dan Gambar 2. Kondisi ini menunjukkan adanya ketidakseimbangan kelas yang tidak terlalu berat, sehingga penilaian kinerja model tidak hanya melihat *accuracy* saja, tetapi juga *precision*, *recall*, dan *F1-Score* pada kelas positif Alzheimer.

DOI: 10.52362/jisamar.v10i3.2535

Journal of Information System, Applied, Management, Accounting and Research.
http://journal.stmikjayakarta.ac.id/index.php/jisamar ,
jisamar@stmikjayakarta.ac.id , jisamar2017@gmail.com
e-ISSN: 2598-8719 (Online), p-ISSN: 2598-8700 (Printed) , Vol. 10 No.3 (August 2026)

Tabel 3. Distribusi Kelas Diagnosis

Kelas	*Label*	*Jumlah*	*Persentase*
Non-Alzheimer	0	1.389	64,63%
Alzheimer	1	760	35,37%
Total	-	2.149	100,00%



Gambar 2. Distribusi Kelas Diagnosis pada Dataset

Tahapan pra-pemrosesan data yang telah dilakukan dirangkum dalam Tabel 4. Tidak ada nilai kosong atau duplikasi data pada dataset, sehingga tidak perlu dilakukan imputasi maupun penghapusan data. Setelah kolom *PatientID* dan *DoctorInCharge* dihapus, dilakukan normalisasi *Min-Max Scaling* pada 12 fitur numerik yang kontinu. Data kemudian dibagi menggunakan metode *stratified split* dengan rasio 80:20, menghasilkan 1.719 data *training* dan 430 data untuk *testing*, seperti yang terlihat pada Tabel 5.

Tabel 4. Tahapan Pra-pemrosesan Data

Tahap	*Tindakan*	*Hasil*
Pengecekan missing value	Mengecek nilai kosong pada seluruh kolom	Tidak ditemukan missing value
Pemeriksaan duplikat	Memeriksa duplikasi data secara keseluruhan	Tidak ditemukan data atau baris yang duplikat
Penghapusan kolom	Menghapus kolom PatientID dan DoctorInCharge	32 fitur prediktor tersisa
Normalisasi	Min-Max Scaling pada 12 fitur bertipe float	Rentang [0,1]
Split data	Stratified train-test split 80%:20%, random_state=42	1.719 data training dan 430 data testing

DOI: 10.52362/jisamar.v10i3.2535

Journal of Information System, Applied, Management, Accounting and Research.
http://journal.stmikjayakarta.ac.id/index.php/jisamar ,
jisamar@stmikjayakarta.ac.id , jisamar2017@gmail.com
e-ISSN: 2598-8719 (Online), p-ISSN: 2598-8700 (Printed) , Vol. 10 No.3 (August 2026)

Tabel 5. Distribusi Data Latih dan Data Uji

Subset	*Jumlah Data*	*Non-Alzheimer (0)*	*Alzheimer (1)*
Training set (80%)	1.719	1.111	608
Testing set (20%)	430	278	152
Total	2.149	1.389	760

3.2. Implementasi Model

Pengujian dilakukan dengan membandingkan model *Stacking Ensemble* sebagai model utama dengan *Gaussian Naive Bayes* sebagai model pembanding, menggunakan fitur yang telah dilakukan pra-pemrosesan sebelumnya dan pembagian data uji yang sama agar perbandingan bisa dilakukan secara adil. Dalam model *Stacking Ensemble*, tiga *base learner* dilatih menggunakan data latih, dan hasil prediksi berupa probabilitas dari masing-masing model tersebut digunakan sebagai input untuk *Logistic Regression*, sehingga menghasilkan prediksi akhir. Pembuatan *meta-feature* menggunakan validasi silang *5-fold* yang dilakukan secara *stratified* dengan *random_state*=42 agar hasilnya dapat direproduksi kembali. Konfigurasi setiap algoritma ditampilkan pada Tabel 6.

Tabel 6. Konfigurasi Model Stacking Ensemble

Level	*Algoritma*	*Peran*	*Parameter*
Level 0	Random Forest	Base learner	n_estimators=100; random_state=42
Level 0	XGBoost	Base learner	n_estimators=100; learning_rate=0,1; random_state=42
Level 0	SVM	Base learner	kernel=rbf; probability=True; random_state=42
Level 1	Logistic Regression	Meta learner	max_iter=1000; random_state=42

Evaluasi dilakukan dengan dua skenario yaitu pengujian *hold-out* menggunakan 430 *record* dari dataset uji, serta pengujian *5-fold stratified cross-validation* yang menggunakan seluruh dataset dengan opsi shuffle=True dan random_state=42. Dalam kedua skenario tersebut, kelas positif ditentukan sebagai Diagnosis=1 (Alzheimer).

3.3. Hasil Pengujian dan Evaluasi Performa

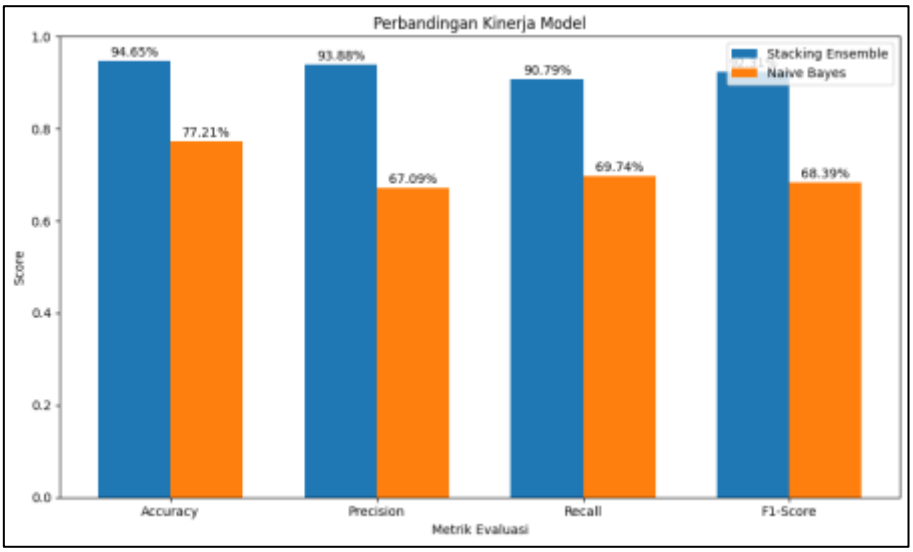
Tabel 7 menunjukkan hasil penilaian kedua model pada 430 data uji. *Stacking Ensemble* mencapai tingkat akurasi sebesar 94,65%, presisi 93,88%, *recall* 90,79%, dan skor F1 sebesar 92,31%. Dalam data uji yang sama, model *Gaussian Naive Bayes* memiliki akurasi sebesar 77,21%, presisi sebesar 67,09%, *recall* sebesar 69,74%, dan skor F1 sebesar 68,39%. Perbandingan metrik evaluasi ditunjukkan pada Gambar 3.

Tabel 7. Perbandingan Metrik Evaluasi pada Data Uji

Model	*Accuracy*	*Precision*	*Recall*	*F1-Score*
Stacking Ensemble	94,65%	93,88%	90,79%	92,31%
Gaussian Naive Bayes	77,21%	67,09%	69,74%	68,39%

DOI: 10.52362/jisamar.v10i3.2535

Journal of Information System, Applied, Management, Accounting and Research.
http://journal.stmikjayakarta.ac.id/index.php/jisamar ,
jisamar@stmikjayakarta.ac.id , jisamar2017@gmail.com
e-ISSN: 2598-8719 (Online), p-ISSN: 2598-8700 (Printed) , Vol. 10 No.3 (August 2026)

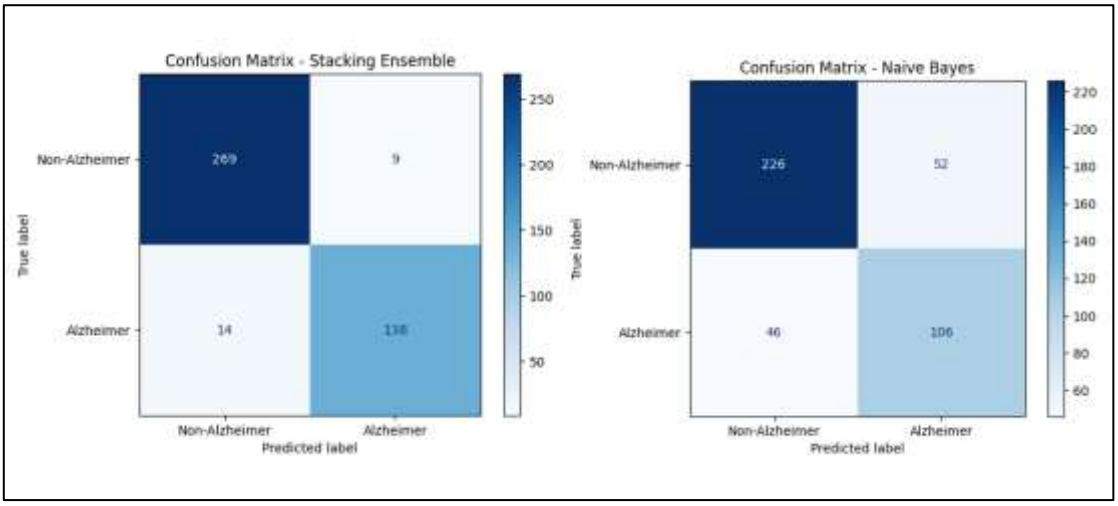


Gambar 3. Perbandingan Metrik Evaluasi pada Data Uji

Dari 152 data Alzheimer yang diuji, metode *Stacking Ensemble* mampu mengenali 138 data secara benar dan melewatkan 14 data, sedangkan *Gaussian Naive Bayes* mampu mengenali 106 data secara benar dan melewatkan 46 data. Dalam kelas Non-Alzheimer, metode *Stacking Ensemble* memberikan 9 hasil positif yang salah, sedangkan metode *Gaussian Naive Bayes* memberikan 52 hasil positif yang salah. Detail dari *confusion matrix* ditunjukkan pada Tabel 8 dan ditampilkan dalam Gambar 4.

Tabel 8. Confusion Matrix pada Data Uji

Model	***TN***	***FP***	***FN***	***TP***
Stacking Ensemble	269	9	14	138
Gaussian Naive Bayes	226	52	46	106



Gambar 3. Perbandingan Metrik Evaluasi pada Data Uji

DOI: 10.52362/jisamar.v10i3.2535

Journal of Information System, Applied, Management, Accounting and Research.
http://journal.stmikjayakarta.ac.id/index.php/jisamar ,
jisamar@stmikjayakarta.ac.id , jisamar2017@gmail.com
e-ISSN: 2598-8719 (Online), p-ISSN: 2598-8700 (Printed) , Vol. 10 No.3 (August 2026)

Validasi silang digunakan untuk mengevaluasi kemampuan model tetap konsisten dalam memberikan hasil yang baik pada lima kelompok data yang berbeda. Rata-rata *accuracy Stacking Ensemble* adalah 95,11% dengan standar deviasi 0,59%, sedangkan *Gaussian Naive Bayes* mencapai rata-rata akurasi sebesar 79,76% dengan standar deviasi 1,00%. Nilai rata-rata *precision*, *recall*, dan *F1-score* metode *Stacking Ensemble* masing-masing mencapai 94,56%, 91,45%, dan 92,97%, yang lebih baik dibandingkan metode *Gaussian Naive Bayes* yang hanya mencapai 72,82%, 68,42%, dan 70,51% terlampir pada tabel 9. Nilai *accuracy* yang berbeda pada setiap *fold* terlihat di Tabel 10, yang menunjukkan konsistensi hasil model *Stacking Ensemble* pada berbagai subset data.

Tabel 9. Hasil 5-Fold Stratified Cross-Validation (Mean ± SD)

Model	*Accuracy*	*Precision*	*Recall*	*F1-Score*
Stacking Ensemble	95,11% ± 0,59%	94,56% ± 0,57%	91,45% ± 1,66%	92,97% ± 0,90%
Gaussian Naive Bayes	79,76% ± 1,00%	72,82% ± 2,27%	68,42% ± 1,72%	70.51% ± 1.19%

Tabel 10. Akurasi Setiap Fold

Model	*Fold 1*	*Fold 2*	*Fold 3*	*Fold 4*	*Fold 5*
Stacking Ensemble	95,58%	94,18%	94,65%	95,58%	95,57%
Gaussian Naive Bayes	78,60%	79,53%	81,16%	78,84%	80,65%

3.4. Pembahasan Temuan Penelitian

Hasil dari penelitian menunjukkan bahwa metode *Stacking Ensemble* berhasil mengklasifikasikan kondisi Alzheimer dengan hasil yang sangat baik. Pada data uji, model mencapai tingkat *accuracy* sebesar 94,65% dan *F1-Score* sebesar 92,31%, yang sesuai dengan hasil validasi silang yang menunjukkan rata-rata *accuracy* 95,11% dan rata-rata *F1-Score* 92,97%. Hasil yang mendekati pada data uji dan rata-rata validasi silang menunjukkan bahwa model memiliki kemampuan untuk beradaptasi dengan berbagai jenis data dan tidak hanya bergantung pada satu bagian data tertentu saja.

Jika dibandingkan dengan *Gaussian Naive Bayes*, metode *Stacking Ensemble* meningkatkan *accuracy* sebesar 17,44 poin persentase, *precision* sebesar 26,79 poin persentase, *recall* sebesar 21,05 poin persentase, dan *F1-score* sebesar 23,92 poin persentase pada data uji. Perbandingan tersebut dapat dilihat pada tabel 11. Hasil ini menunjukkan bahwa tujuan penelitian untuk melakukan perbandingan telah tercapai, yaitu *Stacking Ensemble* bekerja jauh lebih baik daripada model *baseline Gaussian Naive Bayes* pada dataset dan konfigurasi yang terapkan.

Tabel 11. Selisih Performa pada Data Uji

Metrik	*Stacking Ensemble*	*Gaussian Naive Bayes*	*Selisih*
Accuracy	94,65%	77,21%	17,44 poin persentase
Precision	93,88%	67,09%	26,79 poin persentase
Recall	90,79%	69,74%	21,05 poin persentase
F1-Score	92,31%	68,39%	23,92 poin persentase

Dari segi kesalahan klasifikasi, *Stacking Ensemble* menghasilkan 14 *false negative*, jauh lebih sedikit dibandingkan 46 *false negative* yang dihasilkan oleh *Gaussian Naive Bayes* yang menunjukkan penurunan sebanyak 33 *record*. Pada saat yang sama, jumlah *false positive* juga berkurang dari 52 *record* menjadi 9 *record*.

DOI: 10.52362/jisamar.v10i3.2535

Journal of Information System, Applied, Management, Accounting and Research.
http://journal.stmikjayakarta.ac.id/index.php/jisamar ,
jisamar@stmikjayakarta.ac.id , jisamar2017@gmail.com
e-ISSN: 2598-8719 (Online), p-ISSN: 2598-8700 (Printed) , Vol. 10 No.3 (August 2026)

Penurunan kedua jenis kesalahan tersebut menjelaskan peningkatan yang signifikan pada *precision*, *recall*, dan *F1-score* model *Stacking Ensemble*. Hal ini sangat penting dalam konteks medis karena adanya *false negative* bisa menyebabkan penundaan dalam menangani pasien Alzheimer yang sebenarnya mengalami penyakit tersebut.

Berdasarkan hasil uji dan validasi silang, model *Stacking Ensemble* terbukti lebih cocok digunakan untuk klasifikasi penyakit Alzheimer berdasarkan data klinis dibandingkan dengan *Gaussian Naive Bayes* dalam penelitian ini. Meskipun begitu, hasil ini hanya berdasarkan dataset yang digunakan dan belum diuji pada data dari sumber atau populasi lain, sehingga performa yang dicapai belum bisa dianggap sebagai bukti bahwa sistem ini siap digunakan secara klinis tanpa verifikasi tambahan.

IV. KESIMPULAN

Penelitian ini mengembangkan dan membandingkan cara kerja metode *Stacking Ensemble* serta *Naïve Bayes* dalam mengklasifikasikan penyakit Alzheimer menggunakan data klinis. Dataset yang digunakan terdiri dari 2.149 record dan 35 kolom, setelah kolom *PatientID* dan *DoctorInCharge* dihilangkan, pemodelan dilakukan dengan menggunakan 32 fitur prediktor dan target Diagnosis. Data tidak memiliki *missing value* atau duplikasi data, sehingga proses pra-pemrosesan hanya fokus pada normalisasi dengan metode *Min-Max Scaling* dan pembagian data secara *stratified*.

Hasil pengujian pada data uji menunjukkan bahwa metode *Stacking Ensemble* memberikan performa yang sangat baik, dengan tingkat akurasi sebesar 94,65%, presisi 93,88%, *recall* 90,79%, dan skor F1 sebesar 92,31%.Hal ini jauh lebih baik dibandingkan metode *Gaussian Naive Bayes*, yang hanya mencapai akurasi 77,21%, presisi 67,09%, *recall* 69,74%, dan skor F1 sebesar 68,39%. Peningkatan hasil yang dicapai metode *Stacking Ensemble* dibandingkan dengan *Naive Bayes* adalah sebesar 17,44 poin persentase pada akurasi, 26,79 poin persentase pada presisi, 21,05 poin persentase pada recall, dan 23,92 poin persentase pada skor F1. Hasil dari *confusion matrix* juga menunjukkan kelebihan metode *Stacking Ensemble* dengan jumlah *false negative* sebanyak 14 dan *false positive* sebanyak 9, yang jauh lebih rendah dibandingkan *Naive Bayes* yang memiliki 46 *false negative* dan 52 *false positive*. Ini adalah temuan yang penting karena *false negative* yang rendah sangat penting dalam konteks deteksi dini penyakit medis. Hasil validasi silang *5-fold* juga memperkuat temuan tersebut, di mana akurasi rata-rata dari *Stacking Ensemble* mencapai 95,11% dibandingkan dengan 79,76% pada *Gaussian Naive Bayes*, yang menunjukkan bahwa keunggulan metode *Stacking Ensemble* konsisten dan tidak tergantung pada pembagian data tertentu.

Dengan demikian, dapat disimpulkan bahwa metode *Stacking Ensemble* yang menggabungkan Ra*ndom Forest*, *XGBoost*, dan *Support Vector Machine* sebagai *base-learner* dengan *Logistic Regression* sebagai *meta-learner* terbukti lebih baik dan lebih konsisten dibandingkan model tunggal *Naïve Bayes* dalam mengklasifikasikan penyakit Alzheimer berdasarkan data klinis. Temuan ini membantu memajukan sistem skrining awal Alzheimer yang lebih murah, lebih cepat, dan tidak menyakitkan, yang menggunakan data klinis.Selain itu, temuan ini juga menjadi dasar untuk mengembangkan sistem pengambilan keputusan medis berbasis kecerdasan buatan di masa depan.

Karena hasil penelitian ini masih dibatasi dengan menggunakan satu sumber dataset dan belum diuji pada data sumber lain, disarankan untuk penelitian berikutnya agar dapat melakukan analisis kurva ROC-AUC, menentukan *threshold* yang mempertimbangkan dampak klinis dari *false negative* dan *false positive*, membandingkan berbagai strategi dalam menangani ketidakseimbangan kelas, serta menambahkan analisis tambahan.

REFERENSI

[1] M. D. Alfarisi *et al.*, "HUBUNGAN PENCEGAHAN PRIMER DEMENSIA DENGAN KEJADIAN DEMENSIA PADA LANSIA," *PREPOTIF J. Kesehat. Masy.*, vol. 9, no. 1, pp. 1584–1590, Apr. 2025, doi: 10.31004/prepotif.v9i1.42134.

[2] T. E. K. Cersonsky *et al.*, "Using the Montreal cognitive assessment to identify individuals with subtle cognitive decline.," *Neuropsychology*, vol. 36, no. 5, pp. 373–383, Jul. 2022, doi: 10.1037/neu0000820.

DOI: 10.52362/jisamar.v10i3.2535

[3] Nagarathna, Kusuma, and H. Huliyappa, "Ensemble Stacking Method of Classifying the Stages of Alzheimer's Disease by using MRI Dataset," *Int. J. Artif. Intell.*, vol. 11, no. 2, pp. 62–69, Dec. 2024, doi: 10.36079/lamintang.ijai-01102.602.
[4] Y. S. Austin *et al.*, "Klasifikasi Penyakit Alzheimer Dari Scan Mri Otak Menggunakan Convnext," *J. Teknol. Inf. Dan Ilmu Komput.*, vol. 11, no. 6, pp. 1223–1232, Dec. 2024, doi: 10.25126/jtiik.1168117.
[5] E. Rudini and F. Ernawan, "Prediction of Alzheimer's Dementia Using Soft Voting Ensemble Learning with Machine Learning," *IJACI Int. J. Adv. Comput. Inform.*, vol. 1, no. 1, pp. 48–55, Apr. 2025, doi: 10.71129/ijaci.v1.i1.pp48-55.
[6] K. Kirso, "Improving Alzheimer's Disease Prediction Accuracy using Feature Selection, K Fold Cross Validation, and KNN Imputer Techniques," *Telematika*, vol. 18, no. 1, pp. 75–90, Feb. 2025, doi: 10.35671/telematika.v18i1.3055.
[7] T. P. D. Djaka and Nurul Anisa Sri Winarsih, "Analisis Kinerja Ensemble Learning dan Algoritma Tunggal dalam Klasifikasi Sindrom Ovarium Polikistik Menggunakan Random Forest, Logistic Regression, dan XGBoost," *Infotekmesin*, vol. 16, no. 1, pp. 43–51, Jan. 2025, doi: 10.35970/infotekmesin.v16i1.2504.
[8] M. S. Albani, D. Kurniawan, and K. D. Tania, "Komparasi Model Ensemble dan Algoritma Machine Learning Untuk Memprediksi Penyakit Jantung," vol. 7, no. 4, 2026.
[9] J. Margonda and J. Barat, "Perbandingan Kinerja Algoritma Pembelajaran Mesin untuk Deteksi Dini Penyakit Alzheimer," *Comput. Sci.*, vol. 5, no. 1, 2025.
[10] R. Suci and R. Agus, "Comparison of machine learning algorithms for alzheimer's risk classification".
[11] the Alzheimer's Disease Neuroimaging Initiative *et al.*, "A Stacking Framework for Multi-Classification of Alzheimer's Disease Using Neuroimaging and Clinical Features," *J. Alzheimers Dis.*, vol. 87, no. 4, pp. 1627–1636, Jun. 2022, doi: 10.3233/JAD-215654.
[12] M. A. Ramdhani, "Gunung Djati Converence Series," vol. 3, 2021.
[13] Gde Agung Brahmana Suryanegara, Adiwijaya, and Mahendra Dwifebri Purbolaksono, "Peningkatan Hasil Klasifikasi pada Algoritma Random Forest untuk Deteksi Pasien Penderita Diabetes Menggunakan Metode Normalisasi," *J. RESTI Rekayasa Sist. Dan Teknol. Inf.*, vol. 5, no. 1, pp. 114–122, Feb. 2021, doi: 10.29207/resti.v5i1.2880.
[14] Y. Burnwal and Dr. R. C. Jaiswal, "A Comprehensive Survey on Prediction Models and the Impact of XGBoost," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 11, no. 12, pp. 1552–1556, Dec. 2023, doi: 10.22214/ijraset.2023.57625.
[15] P.-A.-H. Pham and N.-D. Hoang, "Metaheuristic optimization of extreme gradient boosting machine for enhanced prediction of lateral strength of reinforced concrete columns under cyclic loadings," *Results Eng.*, vol. 24, p. 103125, Dec. 2024, doi: 10.1016/j.rineng.2024.103125.
[16] H. B. Alwan and K. R. Ku-Mahamud, "A Review on Support Vector Machine Problems and Solutions".
[17] Danis Rifa Nurqotimah, A. Naseh Khudori, and R. Siwi Pradini, "Implementasi Algoritma Support Vector Machine (SVM) Untuk Klasifikasi Penyakit Stroke," *J. Appl. Comput. Sci. Technol.*, vol. 5, no. 2, p. press, Dec. 2024, doi: 10.52158/jacost.v5i2.817.
[18] A. Srinivasan, A. Mishra, and P. Kumar-M, Eds., *R for Basic Biostatistics in Medical Research*. Singapore: Springer Nature Singapore, 2024. doi: 10.1007/978-981-97-6980-3.
[19] D. He, R. Du, S. Zhu, M. Zhang, K. Liang, and S. Chan, "Secure Logistic Regression for Vertical Federated Learning," *IEEE Internet Comput.*, vol. 26, no. 2, pp. 61–68, Mar. 2022, doi: 10.1109/MIC.2021.3138853.
[20] A. C. Müller and S. Guido, *Introduction to machine learning with Python: a guide for data scientists*, First edition. Sebastopol, CA: O'Reilly Media, 2017.
[21] R. S. Nurhalizah, R. Ardianto, and P. Purwono, "Analisis Supervised dan Unsupervised Learning pada Machine Learning: Systematic Literature Review," *J. Ilmu Komput. Dan Inform.*, vol. 4, no. 1, pp. 61–72, Aug. 2024, doi: 10.54082/jiki.168.
[22] Y. Dubey, A. Bhongade, P. Palsodkar, and P. Fulzele, "Efficient Explainable Models for Alzheimer's Disease Classification with Feature Selection and Data Balancing Approach Using Ensemble Learning," *Diagnostics*, vol. 14, no. 24, p. 2770, Dec. 2024, doi: 10.3390/diagnostics14242770.
[23] N. Ismail and S. Lestari, "Mendiagnosis Penyakit Ginjal Kronis Mengunakan Algoritme C4.5," 2023.
[24] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," vol. 5, 2021.

DOI: 10.52362/jisamar.v10i3.2535

[25] D. Musfiroh, U. Khaira, P. E. P. Utomo, and T. Suratno, "Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon: Sentiment Analysis of Online Lectures in Indonesia from Twitter Dataset Using InSet Lexicon," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 24–33, Mar. 2021, doi: 10.57152/malcom.v1i1.20.

[26] Rangga Gelar Guntara, "Visualisasi Data Laporan Penjualan Toko Online Melalui Pendekatan Data Science Menggunakan Google Colab," *ULIL ALBAB J. Ilm. Multidisiplin*, vol. 2, no. 6, pp. 2091–2100, Apr. 2023, doi: 10.56799/jim.v2i6.1578.

DOI: 10.52362/jisamar.v10i3.2535