JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 4, Juli 2025, Hal 864-870
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

Peningkatan Kinerja Klasifikasi Naive Bayes dengan Fitur Adaboost pada Penyakit Diabetes

Eni Rohaini[1], Nurhayati[2], Riza Pahlevi[3], Gunardi[4] , Jasmir[5]
[1,2,3]Teknik Informatika, Universitas Dinamika Bangsa, Jambi, Indonesia
[4]Sistem Informasi, Universitas Dinamika Bangsa, Jambi, Indonesia
[5]Sistem Komputer, Universitas Dinamika Bangsa, Jambi, Indonesia
Email: enirohaini0104@gmail.com, nurhayatihalim75@gmail.com, rizapahlevi@unama.ac.id, gun4rdi.sj@gmail.com, ijay_jasmir@yahoo.com
Email Penulis Korespondensi: enirohaini0104@gmail.com

Abstrak
Diabetes adalah gangguan metabolisme kronis yang ditandai dengan tingginya kadar gula (glukosa) dalam darah. Diabetes memerlukan perhatian dan manajemen yang berkelanjutan untuk mencegah komplikasi yang dapat memengaruhi kualitas hidup penderita. Dalam menghadapi kondisi ini, deteksi dini diabetes menjadi krusial untuk mencegah komplikasi yang lebih serius. Pemanfaatan teknologi dianggap sebagai solusi untuk mengurangi kesalahan estimasi. Dalam penelitian ini, digunakan metode Naïve Bayes berkolaborasi dengan fitur Adaptive Boosting (AdaBoost). Tujuan dari penelitian ini adalah untuk meningkatkan kinerja metode klasifikasi Naive Bayes dengan menggunakan fitur AdaBoost. Dataset yang digunakan diambil dari situs Kaggle. Temuan menunjukkan bahwa fitur Adaboost mempu meningkatkan nilai evaluasi kinerja klasifikasi naive bayes. Algoritma naïve bayes mengalami peningkatan akurasi sebesar 4,94%, presisi sebesar 4,22%, recall sebesar 6,6% dan f1-score 5,42%.
Kata Kunci: Naïve Bayes, Fitur AdaBoost, Diabetes, Peningkatan Akurasi, Pembelajaran Mesin

Abstract
Diabetes is a chronic metabolic disorder characterized by high levels of sugar (glucose) in the blood. Diabetes requires continuous attention and management to prevent complications that can affect the quality of life of sufferers. In dealing with this condition, early detection of diabetes is crucial to prevent more serious complications. The use of technology is considered a solution to reduce estimation errors. In this study, the Naïve Bayes method was used in collaboration with the Adaptive Boosting (AdaBoost) feature. The purpose of this study was to improve the performance of the Naive Bayes classification method using the AdaBoost feature. The dataset used was taken from the Kaggle site. The findings show that the Adaboost feature is able to increase the evaluation value of the naive bayes classification performance. The naïve bayes algorithm experienced an increase in accuracy of 4.94%, precision of 4.22%, recall of 6.6% and f1-score of 5.42%.
***Keywords**: Naïve Bayes, AdaBoost features, Diabetes, accuracy improvement, machine learning*

1. PENDAHULUAN

Penyakit Diabetes adalah penyakit yang menjadi sebab berbagai masalah kehidupan seperti menyebabkan kebutaan, gagal ginjal dan serangan jantung serta kematian [1], [2]. Masih ada beberapa kasus yang terlambat dalam mengetahui gejala awal, sehingga saat diperiksa sudah dalam kondisi serius. Oleh karena itu, dalam menanggapi kasus seperti ini, sistem deteksi dini penyakit diabetes sangat diperlukan untuk mencegah penyakit yang lebih kompleks. Pemanfaatan teknologi dapat digunakan untuk meminimalkan kesalahan estimasi. Dalam dunia kedokteran, teknologi memungkinkan para ahli atau dokter untuk menginterpretasikan penyakit dengan cepat dan akurat [3].

Proses diagnosis penyakit oleh tenaga medis memiliki peran yang sangat penting, tidak hanya dalam hal akurasi yang tinggi tetapi juga kecepatan dalam menentukan penyakit yang dialami pasien. Kesalahan atau keterlambatan dalam diagnosis dapat berdampak pada kesalahan dalam pengobatan dan penanganan pasien. Oleh karena itu, khususnya dalam diagnosis penyakit yang berkaitan dengan organ pernapasan, dibutuhkan pendekatan yang memiliki tingkat akurasi tinggi serta proses yang cepat[4][5][6].

Ilmu pengetahuan teknologi dapat membantu kebutuhan medis dengan memanfaatkan penggunaan *machine learning* dalam hal ini adalah *Naive Bayes* [7] [7][8]. *Naïve Bayes* adalah algoritma pembelajaran mesin yang sering digunakan [9]. Ilmuwan Inggris Thomas Bayes menciptakan metode probabilitas dan kategorisasi statistik yang dikenal sebagai *Naïve Bayes*. Disisi lain disebut juga juga sebagai sebagai *Teorema Bayes* karena membuat prediksi tentang masa depan berdasarkan pengalaman masa lalu. *Naive Bayes* digunakan untuk menyelesaikan masalah prediksi berupa klasifikasi. Selain itu, *Naive Bayes* memiliki jumlah data latih (*training*) lebih kecil dibandingkan dengan data asli, sehingga mudah menentukan dalam perkiraan parameter yang diperlukan pada alur proses klasifikasi dengan nilai akurasi yang tinggi [10][11][12]. Pemanfaatan naive bayes *machine learning* diharapkan mampu memudahkan dalam melakukan pengklasifikasian penyakit diabetes dengan efektif, sehingga seseorang dapat diketahui penyakit diabetes lebih awal agar

JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 4, Juli 2025, Hal 864-870
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

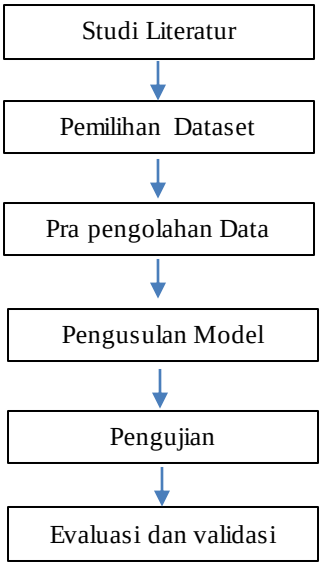
dapat menekan angka kematian penyakit diabetes di Indonesia. Berdasarkan uraian yang dijelaskan, penelitian ini mengimplementasikan metode *Naïve Bayes classifier* untuk mengklasifikasi penyakit diabetes.

Dengan meluasnya ketersediaan data dalam format digital dan meningkatnya kebutuhan untuk mengaksesnya secara fleksibel, maka proses klasifikasi telah menjadi tugas yang fundamental dan penting. Salah satu tantangan utama dalam proses klasifikasi adalah dimensi tinggi dari ruang fitur [13]. yang seringkali terdiri dari puluhan ribu fitur dalam beberapa domain. Sebagian besar dari fitur ini tidak relevan atau tidak bermanfaat dalam konteks klasifikasi, bahkan beberapa di antaranya dapat merugikan akurasi klasifikasi. Selain itu, jumlah besar fitur ini dapat memperlambat proses klasifikasi [14].

Salah satu aspek umum dalam pendekatan klasifikasi adalah pemilihan fitur. Pemilihan fitur dapat meningkatkan efisiensi dan efektivitas pengklasifikasi [15], baik dengan mengurangi jumlah data yang dianalisis maupun mengidentifikasi fitur yang relevan untuk dipertimbangkan dalam proses pembelajaran [16]. Salah satu fitur yang lebih unggul adalah fitur *AdaBoost*. *AdaBoost* merupakan fitur yang konsep dasarnya adalah memilih dan menggabungkan sekelompok pengklasifikasi lemah untuk membentuk suatu klasifikasi yang kuat [17]. Fitur *AdaBoost* ini khusus dirancang untuk tujuan klasifikasi, di mana pembelajarannya ditujukan untuk meningkatkan akurasi setiap algoritma pembelajaran yang lemah [18][19]. *AdaBoost* digunakan secara umum untuk meningkatkan akurasi pembelajaran yang lemah pada tugas klasifikasi sebagian *supervised learning*. Oleh sebab itu, mengacu pada penelitian di atas, maka kami melakukan penelitian untuk meningkatkan akurasi pada metode *naive bayes* dan fitur *AdaBoost* sebagai teknik untuk meningkatkan nilai evaluasi kinerja klasifikasi

2. METODOLOGI PENELITIAN

Metodologi penelitian ini kami susun dalam bentuk urutan kegiatan atau kerangka kerja penelitian, seperti yang telihat dalam bagan dibawah ini:



Gambar 1. Kerangka kerja Penelitan

2.1 Studi Literatur

Dalam hal ini penulis melakukan studi literatur yang berkaitan dengan penyakit diabetes, fitur AdaBoost dan Naïve Bayes.

2.2 Pemilihan Dataset

Data input merupakan bahan yang digunakan dalam penelitian ini untuk diolah menggunakan metode. Penelitian menggunakan data sekunder yang diperolehnya dari situs Kaggle ”https://www.kaggle.com/”. *Pima Indians Diabetes Database*. Pemilik dataset diabetes yaitu *National Institutes of Diabetes and Digestive and Kidney Diseases*. *Dataset Pima Indians Diabetes Database* disumbangkan oleh *Vincent Sigillito Research Center, RMI Group Leader Applied Phisics Laboratory The John Hopkins University*. Di dalam dataset *Pima indians Diabetes Database* berisi data pasien sebanyak 768 pasien perempuan. Masing- masing pasien setidaknya sudah berusia 21 tahun keatas dari keturunan suku Indian Pima. Suku Indian Sungai Gila di Arizona Selatan, AS, telah menjadi rumah bagi suku Indian Pima sejak tahun 1965. Pada usia 35 tahun, suku ini memiliki risiko diabetes terbesar yang tercatat di seluruh dunia (50%). Terdapat sembilan fitur atribut pada penyakit diabetes perempuan di *Pima Indian* dengan 8 variabel dan 1 label outcome. Pada data

JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 4, Juli 2025, Hal 864-870
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

ini terbagi menjadi data latih dan data uji. Untuk jumlah keseluruhan data ini terdiri dari 768 baris record dan 8 fitur. Pada label outcome berisikan 268 data pasien positif diabetes dan 500 data pasien negative. Fitur yang digunakan terdapat pada tabel 1.

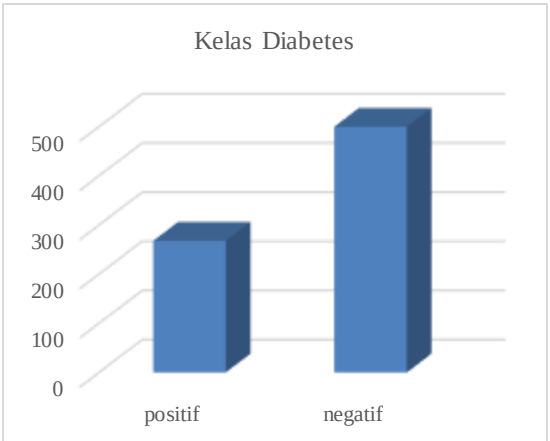
Tabel 1 Fitur Dataset

No	Fitur	Keterangan
1	Pregnancies	Fitur masa kehamilan
2	Glucose	Fitur kadar Glukosa dalam darah
3	BloodPressure	Fitur pengukuran tekanan darah
4	SkinThickness	Fitur ketebalan kulit
5	Insulin	Fitur kadar Insulin dalam darah
6	BMI	Fitur indeks massa tubuh
7	DiabetesPedigreeFunction	Fitur persentase Diabetes
8	Age	Fitur usia

Pada Tabel 2 merupakan cuplikan dataset yang digunakan, dataset penyakit diabetes memiliki 8 fitur.

Pregnancies	*Glucose*	*Blood Pressure*	*Skin thickness*	*Insulin*	*BMI*	*DPF*	*Age*
6	148	72	35	0	33,6	0,627	50
1	85	66	29	0	26,6	0,351	31
8	183	64	0	0	23,3	0,672	32
1	89	66	23	94	28,1	0,167	21
0	137	40	35	168	43,1	2,288	33
5	116	74	0	0	25,6	0,201	30
3	78	50	32	88	31	0,248	26
10	115	0	0	0	35,3	0,134	29
2	197	70	45	543	30,5	0,158	53
8	125	96	0	0	0	0,232	54
4	110	92	0	0	37,6	0,191	30
10	168	74	0	0	38	0,537	34
10	139	80	0	0	27,1	1,441	57

Gambar 2 menjelaskan kelas diabetes yang terdiri dari 2 kelas yaitu kelas positif dan kelas negatif. Jumlah kelas positif yaitu 268 dan kelas negatif yaitu 500.



Gambar 2. Grafik kelas dataset Diabetes

JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 4, Juli 2025, Hal 864-870
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

2.3 Pra Pengolahan Data

Pada tahap ini dilakukan beberapa tahap pra pengolahan data seperti data dibersihkan kemudian dilakukan tranformasi data sebelum dilakukan pembuatan model.

2.4 Pengusulan Model

Data yang dianalisis kemudian dikelompokkan ke dalam variabel-variabel yang saling berhubungan, dilanjutkan dengan pembuatan model yang sesuai dengan karakteristik data tersebut. Selain itu, pembagian data menjadi data latih (training data) dan data uji (testing data) juga diperlukan dalam proses pengembangan model.

2.5 Pengujian

Langkah selanjutnya adalah pengujian, pengujian dilakukan untuk menentukan tingkat akurasi suatu algoritma. Kinerja algoritma dapat dievaluasi dan divalidasi menggunakan *confusion matrix*. *Confusion matrix* adalah alat ukur berbentuk matriks 2x2 yang digunakan untuk menghitung tingkat ketepatan klasifikasi pada dataset. Evaluasi model klasifikasi didasarkan pada pengujian yang mengidentifikasi prediksi benar dan salah, dengan hasil pengujian tersebut disusun dalam bentuk *confusion matrix*. [28]

Untuk mendapatkan klasifikasi yang optimal, metrik evaluasi memiliki fungsi yang sangat penting dalam tahapan klasifikasi. Metrik evaluasi merupakan seperangkat kriteria atau tolok ukur yang digunakan untuk mengukur kualitas kinerja suatu sistem, model, proses, atau entitas. Dengan adanya metrik ini, kita dapat menilai secara kuantitatif maupun kualitatif seberapa efektif suatu sistem atau model dalam mewujudkan targetnya. Berbagai bidang seperti pembelajaran mesin, analisis data, rekayasa perangkat lunak, dan bisnis memanfaatkan metrik evaluasi secara luas untuk mengukur dan menyempurnakan efektivitas suatu proses dan hasilnya.

Dalam konteks pembelajaran mesin, metrik evaluasi sangat penting untuk menilai kinerja model pada tugas-tugas tertentu. Tugas-tugas yang berbeda (misalnya, klasifikasi, regresi, pengelompokan) mungkin memerlukan metrik yang berbeda pula. Beberapa metrik evaluasi umum dalam pembelajaran mesin adalah akurasi, presisi, recall, spesifisitas dan f1-score

2.6 Evaluasi dan Validasi.

Dalam sebuah penelitian, dilakukan evaluasi terhadap model yang digunakan untuk mengukur tingkat akurasi model tersebut.

2.7. *Adaptive Boosting* (*Adaboost*)

Sebagai metode reduksi dimensi yang paling umum, Seleksi Fitur (FS) berfokus pada pembersihan data dari gangguan (noise) serta fitur-fitur yang tidak penting atau berlebihan. Tujuan utama FS adalah meningkatkan performa model. Proses seleksi fitur melibatkan pemilihan subset fitur yang paling relevan dari sekumpulan fitur awal berdasarkan kriteria seleksi yang ditetapkan. Tindakan ini berperan signifikan dalam mengurangi beban komputasi dengan menyingkirkan fitur-fitur yang tidak memberikan nilai tambah. Teknik seleksi fitur seringkali menjadi langkah persiapan penting sebelum melatih model pembelajaran, dan seleksi fitur yang efektif dapat menghasilkan akurasi yang lebih tinggi, waktu pelatihan yang lebih singkat, dan model yang lebih sederhana. Perlu dipahami bahwa seleksi fitur berbeda dengan ekstraksi fitur, meskipun keduanya bertujuan mengurangi dimensi data. Seleksi fitur mempertahankan fitur asli yang dianggap penting, sementara ekstraksi fitur menciptakan fitur-fitur baru melalui transformasi data asli yang dianggap lebih baik dalam menangkap pola informasi.

AdaBoost merupakan salah satu dari beberapa variasi algoritma boosting [19]. Adaboost, suatu bentuk pembelajaran ensemble yang sering digunakan dalam algoritma boosting, dapat diintegrasikan dengan algoritma classifier lain untuk meningkatkan kinerja klasifikasi secara keseluruhan. Ide intuitifnya adalah bahwa menggabungkan berbagai model terbukti bermanfaat. Adaboost dan variasinya telah menemukan keberhasilan di berbagai domain, karena landasan teoritisnya yang kuat, prediksi yang tepat, dan implementasi yang mudah.

2.8. *Naïve Bayes* (*NB*)

NB menonjol sebagai salah satu model pengklasifikasi probabilistik yang paling mudah dipahami [9]. Asumsi naif ini dapat menawarkan kompromi yang seimbang antara kinerja dan pengeluaran komputasi. Bo Tang, Steven Kay dan Haibo He [20] menemukan bahwa Naïve Bayes juga dapat bekerja dengan baik ketika fitur-fitur saling bergantung. Selain itu, Netzer et al berpendapat bahwa model generatif yang dihasilkan mudah ditafsirkan dan dijelaskan. Selain itu, sebagai pengklasifikasi generatif, *Naïve Bayes* dapat dianggap cocok untuk ukuran sampel yang lebih kecil karena regularisasi yang melekat, mengurangi kemungkinan *overfitting* jika dibandingkan dengan pengklasifikasi diskriminatif[22]. Metode ini menawarkan keuntungan berupa kebutuhan data pelatihan yang sedikit dalam penentuan parameter klasifikasi. Asumsi bahwa variabel-variabel saling independen menyederhanakan proses, karena klasifikasi hanya memerlukan varians setiap variabel dalam suatu kelas, menghilangkan kebutuhan akan matriks kovarians yang lebih kompleks. Jadi Naive Bayes adalah algoritma klasifikasi yang berharga dan banyak digunakan yang dikenal karena kesederhanaan, kecepatan, dan efektivitasnya dalam berbagai aplikasi, terutama yang melibatkan data teks. Meskipun asumsi "naif"-nya mungkin tampak

JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 4, Juli 2025, Hal 864-870
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

seperti batasan yang signifikan, namun sering kali bekerja dengan sangat baik dalam praktik. Memahami prinsip, variasi, kelebihan, dan kekurangannya sangat penting untuk menerapkannya secara efektif pada masalah pembelajaran mesin.

3. HASIL DAN PEMBAHASAN

Bagian ini menjelaskan tentang hasil dan pembahasan yang berasal dari eksperimen yang dilakukan sesuai dengan kerangka penelitian yang diuraikan pada bagian sebelumnya. Eksperimen tersebut berkisar pada penilaian data menggunakan metode *naive bayes* dan fitur *AdaBoost*. Dengan validasi split 80:20. Pengujian yang dilakukan dalam penelitian ini meliputi pengujian naive bayes dengan fitur *AdaBoost*.

Tabel 1. Confusion Matrix Naive Bayes sebelum dan sesudah menggunakan AdaBoost

Metrik	Sebelum	Sesudah
TP	291	320
FP	89	76
FN	99	74
TN	289	298

Tabel 1 di atas menunjukkan Confusion Matrix untuk metode Naive Bayes sebelum dan sesudah dioptimalkan dengan AdaBoost. Confusion Matrix ini membantu dalam mengevaluasi performa model klasifikasi berdasarkan empat metrik Utama yaitu : True Positive (TP), False Positive (FP), False Negative (FN) dan True Negative (TN). Pada perubahan Perubahan True Positive (TP) nilai yang dioperoleh sebelum AdaBoost adalah 291 dan sesudah AdaBoost adalah 320. Ini berarti bahwa terjadi peningkatan. Model yang telah dioptimalkan dengan AdaBoost lebih mampu mengklasifikasikan sampel positif dengan benar, sehingga jumlah True Positive meningkat. Ini berarti lebih banyak sampel yang seharusnya positif berhasil diklasifikasikan dengan benar.

Pada Perubahan False Positive (FP) nilai yang diperoleh sebelum AdaBoost adalah 89 dan sesudah AdaBoost adalah 76, yang berarti terjadi penurunan. Jumlah False Positive berkurang, yang berarti model setelah menggunakan AdaBoost lebih jarang melakukan kesalahan dalam mengklasifikasikan sampel negatif sebagai positif. Ini menunjukkan bahwa model menjadi lebih presisi dalam mengenali kelas positif.

Pada perubahan False Negative (FN) Nilai yang dioperoleh sebelum AdaBoost adalah dan sesudah AdaBoost adalah 74, berarti terjadi penurunan. False Negative yang lebih rendah menunjukkan bahwa model lebih sedikit salah dalam mengklasifikasikan sampel positif sebagai negatif. Ini berarti model menjadi lebih sensitif terhadap kelas positif, sehingga lebih banyak sampel positif berhasil dideteksi.

Pada perubahan True Negative (TN) diperoleh nilai sebelum AdaBoost adalah 289 dan sesudah AdaBoost adalah 298, berarti terjadi kenaikan. Kenaikan nilai True Negative menunjukkan bahwa model lebih baik dalam mengidentifikasi sampel negatif dengan benar, yang berarti lebih sedikit kesalahan dalam klasifikasi kelas negatif.

AdaBoost mampu meningkatkan True Positive (TP) dan True Negative (TN), sehingga model lebih baik dalam mengklasifikasikan sampel dengan benar. False Positive (FP) dan False Negative (FN) menurun, yang berarti model lebih jarang melakukan kesalahan dalam prediksi. Model setelah menggunakan AdaBoost menjadi lebih akurat, lebih presisi, dan lebih sensitif dalam mengklasifikasikan data.

Kami menemukan bahwa fitur Adaboost yang dikorelasikan dengan metode Naïve Bayes cenderung memiliki proporsi hasil yang lebih tinggi dibandingkan dengan menggunakan metode Naïve Bayes tanpa fitur. Sebelum menggunakan fitur ini, dapat dilihat hasil komposisi nilai matrixnya. Jika melihat hasil tersebut, nilai False Positive dan False Negative masih tinggi. Hal ini juga dapat terjadi karena kemampuan Naïve Bayes dalam mengolah data kompleks masih belum optimal. Setelah menggunakan fitur tersebut, hasil False Positive dan False Negative mengalami penurunan dibandingkan dengan nilai False Positive dan False Negative sebelumnya. Penurunan yang paling signifikan adalah False Positive, sehingga mengakibatkan peningkatan nilai performa klasifikasi. AdaBoost meningkatkan akurasi, precision, recall, dan F1-score. Peningkatan terbesar terlihat pada False Negative (FN) yang menurun cukup signifikan. Naïve Bayes yang ditingkatkan dengan AdaBoost menjadi lebih andal dalam mengenali kelas positif sambil mengurangi kesalahan klasifikasi negatif. Pada proses ini, pengaruh Adaboost dapat dibuktikan.

Setelah dilakukan pengujian *naïve bayes* dengan fitur *Adaboost*, untuk melihat gambaran perbedaan hasil tersebut dapat dilihat pada tabel 2. Dari tabel 2 dapat dilihat bahwa hasil pengujian naïve bayes tanpa menggunakan fitur adaboost diperoleh nilai akurasi sebesar 75,52%; presisi 76,57%; recall sebesar 74,61%; dan f1 score sebesar 75,58%. Setelah menggunakan adaboost, nilai akurasi meningkat sebesar 80,46%; presisi menjadi 80,80%; recall menjadi 81,21% dan f1-score menjadi 81,01%. Artinya algoritma naïve bayes mengalami peningkatan akurasi sebesar 4,94%, presisi sebesar 4,22%, recall sebesar 6,6% dan f1-score 5,42%. Dari matriks konfusi di atas, diperoleh masing-masing nilai evaluasi kinerja klasifikasi yaitu presisi, akurasi recall dan f1-score seperti pada tabel 2 dan gambar 3 berikut:

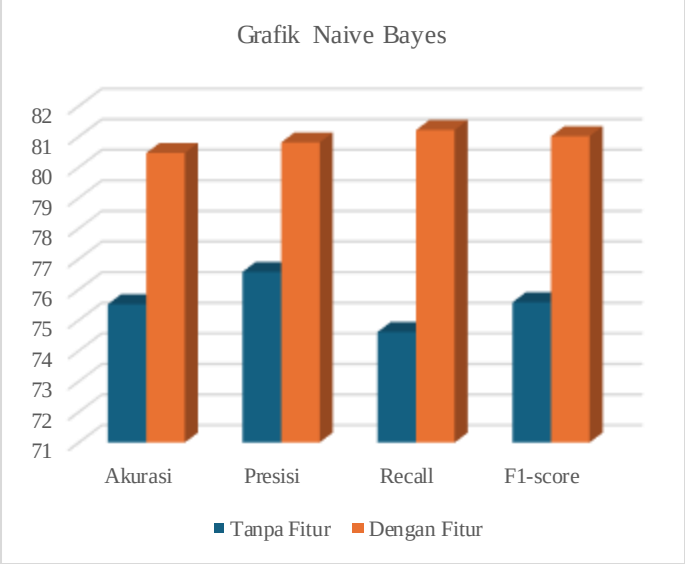
Tabel 2. Perbandingan Evaluasi Kinerja Naïve Bayes

Evaluation	NB (%)	NB + AdaBoost (%)

JURNAL SISTEM INFORMASI TGD
Volume 4, Nomor 4, Juli 2025, Hal 864-870
P-ISSN : 2828-1004 ; E-ISSN : 2828-2566
https://ojs.trigunadharma.ac.id/index.php/jsi

Accuracy	75,52083333	80,46875
Precision	76,57894737	80,80808
Recall	74,61538462	81,21827
F1-Score	75,58441558	81,01266

Gambar 3 adalah representasi dari tabel 2 yang menunjukkan secara nyata jarak perbedaan hasil klasifikasi naïve bayes menggunakan fitur AdaBoost dengan klasifikasi naïve bayes tanpa menggunakan fitur Adaboost pada penyakit diabetes.



Gambar 3. Perbandingan Evaluasi Kinerja Naïve Bayes

Berdasarkan percobaan ini, kami telah berhasil membuktikan pengaruh AdaBoost dalam meningkatkan nilai evaluasi kinerja klasifikasi Naive Bayes. Dari pembahasan di atas, dapat dilihat bahwa metode Naive Bayes di atas mengalami peningkatan setelah menggunakan adaboost. Namun nilai keseluruhannya tidak terdistribusi secara merata. Nilai recall precision dan f1-score accuracy mengalami perbedaan yang cukup tinggi, hal ini terjadi karena perbedaan nilai false positive dan false negative yang terlalu signifikan.

4. KESIMPULAN

Penggunaan metode AdaBoost sebagai teknik optimasi terhadap algoritma Naive Bayes terbukti memberikan peningkatan signifikan pada performa model klasifikasi. Berdasarkan hasil evaluasi Confusion Matrix, terjadi peningkatan pada nilai True Positive (TP) dan True Negative (TN), serta penurunan pada nilai False Positive (FP) dan False Negative (FN). Hal ini menunjukkan bahwa model menjadi lebih akurat dalam mengklasifikasikan data positif maupun negatif secara benar, serta lebih jarang melakukan kesalahan klasifikasi. Secara kuantitatif, model Naive Bayes yang telah dioptimalkan dengan AdaBoost mengalami peningkatan performa. Dari eksperimen yang dilakukan terlihat hasil pengujian naïve bayes mengalami peningkatan akurasi sebesar 4,94%, presisi sebesar 4,22%, recall sebesar 6,6% dan f1-score 5,42%. Peningkatan terbesar terlihat pada nilai Recall, yang mencerminkan kemampuan model dalam mengenali sampel positif secara lebih sensitif dan efektif. Penurunan signifikan pada False Positive juga menandakan bahwa model menjadi lebih presisi dalam mengidentifikasi kelas positif. Dengan demikian, integrasi AdaBoost terhadap Naive Bayes secara keseluruhan mampu memperkuat kemampuan klasifikasi model, membuatnya lebih andal dalam menangani data yang kompleks dan meningkatkan kualitas prediksi.

UCAPAN TERIMAKASIH

Terima kasih disampaikan kepada Yayasan Dinamika Bangsa dan Universitas Dinamika Bangsa melalui Lembaga Penelitan dan Pengabdian Masyarakat atas dukungan sarana dan bantuan dana penelitian.

DAFTAR PUSTAKA

[1] American Diabetes Association, "Classification and Diagnosis of Diabetes: Standards ofMedical Care in Diabetes," *Diabetes Care*, vol. 45, no. Suppl, pp. 17–38, 2022.
[2] Z. T. Bloomgarden, "Diabetes and COVID-19," *J. Diabetes*, vol. 12, no. 4, pp. 347–348, 2020, doi: 10.1111/1753-0407.13027.
[3] Y. Zhu, B. Kong, R. Liu, and Y. Zhao, "Developing biomedical engineering technologies for reproductive medicine," *Smart*

Med., vol. 1, no. 1, 2022, doi: 10.1002/smmd.20220006.

[4] I. W. T. Wahyudi and I. G. A. G. A. Kadyanan, "Implementasi Logistic Regression dalam Sistem Diagnosa Penyakit Diabetes dengan KNN," *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 11, no. 4, p. 743, 2023, doi: 10.24843/jlk.2023.v11.i04.p12.

[5] H. Nizar, A. S. Shafira, J. Aufaresa, M. A. Awliya, and U. Athiyah, "Perbandingan Metode Logika Fuzzy Untuk Diagnosa Penyakit Diabetes," *Explor. Sist. Inf. dan Telemat.*, vol. 12, no. 1, p. 37, 2021, doi: 10.36448/jsit.v12i1.1763.

[6] D. Diana Dewi, N. Qisthi, S. S. S. Lestari, and Z. H. S. Putri, "Perbandingan Metode Neural Network Dan Support Vector Machine Dalam Klasifikasi Diagnosa Penyakit Diabetes," *Cerdika J. Ilm. Indones.*, vol. 3, no. 09, pp. 828–839, 2023, doi: 10.59141/cerdika.v3i09.662.

[7] L. K. Foo, S. L. Chua, and N. Ibrahim, "Attribute weighted naïve bayes classifier," *Comput. Mater. Contin.*, vol. 71, no. 1, pp. 1945–1957, 2022, doi: 10.32604/cmc.2022.022011.

[8] Y. J. Choi, G. Farnadi, B. Babaki, and G. Van Den Broeck, "Learning fair naive bayes classifiers by discovering and eliminating discrimination patterns," *AAAI 2020 - 34th AAAI Conf. Artif. Intell.*, pp. 10077–10084, 2020, doi: 10.1609/aaai.v34i06.6565.

[9] D. Berrar, "Bayes' theorem and naive bayes classifier," *Encycl. Bioinforma. Comput. Biol. ABC Bioinforma.*, vol. 1–3, no. 2018, pp. 403–412, 2018, doi: 10.1016/B978-0-12-809633-8.20473-1.

[10] T. Septiani Nurfauzia Koeswara, M. Sukrisno Mardiyanto, and M. Abdul Ghani, "Penerapan Particle Swarm Optimization (PSO) Dalam Pemilihan Atribut Untuk Meningkatkan Akurasi Prediksi Diagnosispenyakit Hepatitis Dengan Metode Naive Bayes," *J. Speed – Sentra Penelit. Eng. dan Edukasi*, vol. 12, no. 1, pp. 1–10, 2020.

[11] A. Samosir, M. S. Hasibuan, W. E. Justino, and T. Hariyono, "Komparasi Algoritma Random Forest, Naïve Bayes dan K-Nearest Neighbor Dalam klasifikasi Data Penyakit Jantung," *Pros. Semin. Nas. Darmajaya*, vol. 1, no. 0, pp. 214–222, 2021, [Online]. Available: https://jurnal.darmajaya.ac.id/index.php/PSND/article/view/2955

[12] W. A. Amanda, "Klasifikasi Pasien Gagal Jantung Menggunakan Metode Naive Bayes Dengan Penerapan Diskritisasi," no. February, p. 6, 2021.

[13] A. W. Baur, M. Breitsprecher, and M. Bick, "Catching Fire : Start-Ups in the Text Analytics Software Industry," pp. 1–14, 2014.

[14] X. Deng, Y. Li, J. Weng, and J. Zhang, "Feature selection for text classification: A review," *Multimed. Tools Appl.*, vol. 78, no. 3, pp. 3797–3816, 2019, doi: 10.1007/s11042-018-6083-5.

[15] K. Kurniabudi, A. Harris, and E. Rosanda, "Optimalisasi Seleksi Fitur Untuk Deteksi Serangan Pada IoT Menggunakan Classifier Subset Evaluator," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 4, p. 885, 2022, doi: 10.30865/jurikom.v9i4.4618.

[16] J. Jasmir, S. Nurmaini, R. F. Malik, and B. Tutuko, "Bigram feature extraction and conditional random fields model to improve text classification clinical trial document," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 19, no. 3, pp. 886–892, 2021, doi: https://doi.org/10.12928/TELKOMNIKA.v19i3.18357.

[17] A. Lestari, "Increasing Accuracy of C4.5 Algorithm Using Information Gain Ratio and Adaboost for Classification of Chronic Kidney Disease," *J. Soft Comput. Explor.*, vol. 1, no. 1, pp. 32–38, 2020, doi: 10.52465/joscex.v1i1.6.

[18] W. Wang and D. Sun, "The improved AdaBoost algorithms for imbalanced data classification," *Inf. Sci. (Ny).*, vol. 563, pp. 358–374, 2021, doi: 10.1016/j.ins.2021.03.042.

[19] Y. Wang and L. Feng, "Improved Adaboost Algorithm for Classification Based on Noise Confidence Degree and Weighted Feature Selection," *IEEE Access*, vol. 8, pp. 153011–153026, 2020, doi: 10.1109/ACCESS.2020.3017164.

[20] B. Tang, S. Kay, and H. He, "Toward Optimal Feature Selection in Naive Bayes for Text Categorization," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 9, pp. 2508–2521, 2016, doi: 10.1109/TKDE.2016.2563436.

[21] M. Elhadad, D. Gabay, and Y. Netzer, "Automatic Evaluation of Search Ontologies in the Entertainment Domain using Text Classification".

[22] A. Choi, "Structured Features in Naive Bayes Classification," pp. 3233–3240.