

Analisis Perbandingan PCA-KNN dan SVM untuk Prediksi Risiko Diabetes

Rinda Desfourteen^{a,1,*}, Nadia Damayanti^{b,2}, Rendra Gustriansyah^{c,3}

^{a,b,c} Program Studi Teknik Informatika, Fakultas Ilmu Komputer dan Sains, Universitas Indo Global Mandiri, Indonesia

¹ 2022110132@students.uigm.ac.id*; ² 2022110140@students.uigm.ac.id; ³ rendra@uigm.ac.id

* Penulis Korespondensi

ABSTRAK

Diabetes merupakan penyakit kronis yang sering terlambat terdiagnosis akibat gejala awal yang tidak spesifik, sehingga deteksi dini penting untuk mencegah komplikasi serius. Penelitian ini bertujuan menganalisis dan membandingkan performa kombinasi *Principal Component Analysis* dengan *K-Nearest Neighbor* (PCA-KNN) dan *Support Vector Machine* (SVM) dalam prediksi risiko diabetes. Dataset yang digunakan berasal dari Kaggle dengan 768 entri dan delapan atribut medis. Tahap pra-proses mencakup imputasi median untuk nilai nol, normalisasi *Z-score*, serta reduksi dimensi menggunakan PCA pada model KNN yang menghasilkan lima komponen utama dengan varian kumulatif >80%. Nilai *k* optimal ditentukan melalui *10-Fold Cross Validation* dengan hasil terbaik pada *k*=16. Hasil evaluasi menunjukkan PCA-KNN mencapai akurasi 76,47%, sensitivitas 90,00%, dan spesifisitas 50,94%, lebih baik dibanding KNN standar. Sementara itu, SVM memperoleh akurasi 72,73% dengan spesifisitas tinggi (84,00%) namun sensitivitas rendah (51,85%). Temuan ini mengindikasikan bahwa PCA-KNN lebih sesuai untuk skrining awal karena sensitivitas tinggi, sedangkan SVM dapat digunakan pada tahap konfirmasi berkat spesifisitas yang lebih baik.

Riwayat Artikel

Diterima 14 Juni 2025

Diperbaiki 11 September 2025

Diterbitkan 25 Oktober 2025

Kata Kunci

PCA

KNN

SVM

Prediksi Diabetes

Klasifikasi



This is an open-access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license

1. Pendahuluan

Diabetes merupakan kelompok penyakit metabolik yang mencerminkan kondisi hiperglikemia akibat gangguan pada produksi insulin, penurunan respons sel terhadap insulin, atau terjadi secara bersamaan. Kondisi hiperglikemia yang terjadi secara berkepanjangan dapat membahayakan, mengganggu fungsi, dan berpotensi menyebabkan kegagalan organ progresif, terutama pada jantung, pembuluh darah, ginjal, mata, dan sistem saraf [1]. Berdasarkan hasil Survei Kesehatan Indonesia (SKI) tahun 2023, sebanyak 11,7% penduduk Indonesia yang berusia di atas 15 tahun teridentifikasi memiliki diabetes mellitus melalui pengukuran kadar glukosa darah. Dari jumlah tersebut, 50,2% merupakan penderita diabetes tipe 2, menjadikannya jenis diabetes yang paling banyak diderita di Indonesia [2]. Gejala diabetes sering kali tidak tampak atau disalahartikan dengan penyakit lain, sehingga banyak penderita baru menyadari penyakitnya ketika telah terjadi komplikasi serius. Oleh karena itu, deteksi dini sangat penting agar penanganan dapat dilakukan secara cepat dan tepat.

Salah satu pendekatan yang banyak digunakan dalam bidang data mining untuk prediksi penyakit adalah *K-Nearest Neighbor* (KNN). KNN termasuk metode non-parametrik yang bekerja dengan cara membandingkan jarak suatu data uji terhadap data-data dalam pelatihan, lalu menentukan kelas berdasarkan mayoritas tetangga terdekat. Metode ini sangat populer karena memiliki keunggulan dari sisi kesederhanaan implementasi, tidak memerlukan proses pelatihan yang kompleks, serta memiliki performa yang kompetitif dalam klasifikasi data medis dan diagnosis penyakit [3][4]. Namun, ketika jumlah fitur atau dimensi data cukup besar, performa KNN dapat menurun akibat fenomena *curse of dimensionality*. Untuk mengatasi hal tersebut, digunakan *Principal Component Analysis* (PCA) sebagai metode reduksi dimensi yang bertujuan menyederhanakan representasi data tanpa kehilangan informasi penting. PCA membantu mengurangi kompleksitas perhitungan dan



meningkatkan akurasi model klasifikasi, terutama pada dataset kesehatan yang memiliki banyak variabel [5][6].

Selain KNN, algoritma lain yang banyak digunakan dalam prediksi penyakit adalah *Support Vector Machine* (SVM). Algoritma ini termasuk teknik pembelajaran mesin yang berfungsi menentukan *hyperplane* terbaik agar dapat membedakan data menjadi dua atau lebih kategori dengan jarak pemisah yang maksimal. Algoritma ini efektif mengatasi data non-linier melalui fungsi kernel dan unggul dalam generalisasi, ketahanan terhadap *overfitting*, serta akurasi kompetitif pada klasifikasi medis [7][8].

Beberapa penelitian terdahulu telah memanfaatkan algoritma KNN, SVM, serta pendekatan lainnya seperti PCA dalam konteks klasifikasi dan prediksi di bidang medis. Salah satunya adalah penelitian yang membandingkan metode SVM, KNN, *Decision Tree*, dan *Ensemble Majority Voting* pada dataset diabetes dari Kaggle. Hasil penelitian tersebut menunjukkan bahwa KNN dengan *K-Fold Cross Validation* memperoleh akurasi lebih tinggi, yaitu 87,61%, dibandingkan dengan SVM yang hanya mencapai 77,64% [9]. Namun, KNN yang digunakan masih berupa versi standar tanpa integrasi teknik reduksi dimensi seperti PCA, yang berpotensi menyederhanakan kompleksitas data medis berdimensi tinggi.

Sementara itu, terdapat juga penelitian yang menerapkan algoritma SVM dengan kernel *Radial Basis Function* (RBF) berbasis metode *Forward Selection* untuk melakukan prediksi penyakit diabetes melitus tipe 2. Berdasarkan hasil pengujian, model SVM tersebut berhasil memperoleh akurasi sebesar 91,2%, dengan *precision* 93,0%, *recall* 94,3%, dan *F1-score* 93,7% [10]. Penelitian ini berhasil menunjukkan efektivitas SVM dalam klasifikasi diabetes, namun masih terbatas pada penggunaan satu algoritma saja, tanpa adanya perbandingan dengan metode lain, seperti KNN dengan PCA.

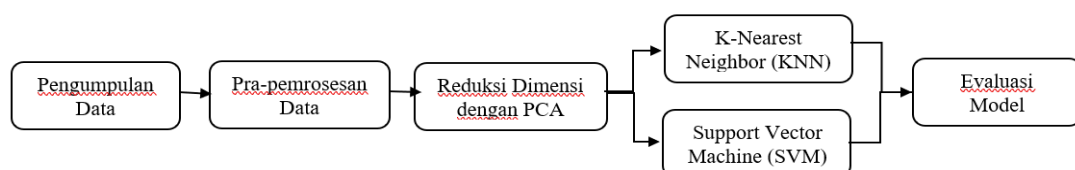
Di sisi lain, terdapat pula penelitian yang menerapkan kombinasi PCA untuk reduksi dimensi dan KNN untuk klasifikasi pada dataset kanker paru-paru. Hasil evaluasi menunjukkan bahwa model PCA-KNN mampu memberikan performa yang sangat baik dengan akurasi mencapai 98% pada nilai $K = 9$. Penelitian ini membuktikan efektivitas PCA dalam menyederhanakan kompleksitas data dan meningkatkan performa KNN dalam klasifikasi medis [11]. Namun, penerapan tersebut masih terbatas pada kasus kanker paru-paru serta tidak melibatkan algoritma lain seperti SVM.

Penelitian sebelumnya telah mengeksplorasi penerapan KNN, SVM, maupun PCA dalam klasifikasi penyakit, akan tetapi sebagian besar masih terbatas pada penggunaan algoritma secara tunggal atau penerapan pada kasus medis selain diabetes saja. Belum banyak penelitian yang secara khusus membandingkan performa PCA-KNN dengan SVM dalam konteks prediksi risiko diabetes. Melalui penelitian ini, penulis bertujuan untuk menganalisis serta membandingkan kedua metode tersebut guna memperoleh gambaran yang lebih komprehensif mengenai efektivitasnya dalam prediksi risiko diabetes.

2. Metode

2.1 Kerangka Penelitian

Penelitian ini bertujuan untuk memprediksi risiko diabetes dengan menerapkan kombinasi PCA dan KNN, serta membandingkan kinerjanya dengan SVM menggunakan dataset diabetes yang tersedia pada platform Kaggle. Langkah-langkah penelitian ini divisualisasikan dalam bentuk diagram alur pada Gambar 1.



Gambar 1 Model Alur Penelitian

2.2 Pengumpulan Data

Dalam penelitian ini, data yang dipakai berasal dari dataset diabetes yang tersedia di Kaggle dengan atribut medis yang terkait dengan prediksi risiko diabetes. Dataset ini memiliki 768 baris data, delapan atribut, dan satu label. Delapan atribut terdiri dari *Pregnancies*, *Glucose*, *BloodPressure*, *SkinThickness*, *Insulin*, *BMI*, *DiabetesPredigree Function*, *Age*, dan label *Outcome*. Tabel 1 berikut ini menyajikan deskripsi masing-masing atribut dalam dataset.

Tabel 1 Deskripsi dari Atribut [11]

No	Atribut	Deskripsi
1	<i>Pregnancies</i>	Banyaknya jumlah kelahiran
2	<i>Glucose</i>	Tingkat gula dalam darah yang diperoleh dua jam pasca uji toleransi gula oral.
3	<i>Blood Pressure</i>	Tekanan darah diastolik (mmHg).
4	<i>Skin Thickness</i>	Tebal lipatan kulit pada area trisep (mm).
5	<i>Insulin</i>	Kadar insulin dalam serum selama dua jam.
6	<i>BMI</i>	Perbandingan antara massa tubuh dalam kilogram dan hasil kuadrat tinggi badan dalam meter yang dikenal sebagai Body Mass Index (BMI)
7	<i>Diabetes Predigree Function</i>	Riwayat penyakit diabetes pada saudara yang mempunyai keturunan dengan penderita.
8	<i>Age</i>	Usia pasien.
9	<i>Outcome</i>	Penanda klasifikasi yang merepresentasikan kondisi penderita, dengan nilai YES menandakan positif diabetes dan NO menandakan negatif diabetes.

2.3 Pra-Pemrosesan Data

Tahap mencakup pembersihan dan normalisasi data. Data yang memiliki nilai hilang atau tidak valid pada atribut-atribut seperti *Glucose*, *BloodPressure*, *SkinThickness*, *Insulin*, dan *BMI* akan diisi menggunakan nilai *median* dari masing-masing atribut tersebut. Penggunaan *median* dipilih untuk menghindari pengaruh outliers atau nilai ekstrem yang dapat memengaruhi distribusi data [12].

Setelah menangani nilai hilang, normalisasi data dilakukan agar seluruh fitur memiliki standar skala yang konsisten dalam proses perhitungan. Proses ini krusial sebab KNN cukup peka terhadap perbedaan skala data. Untuk itu, dilakukan normalisasi *Z-score* pada dataset. Metode *Z-score* digunakan untuk menstandarkan data dengan cara membuat rata-ratanya menjadi nol dan deviasi standarnya bernilai satu. [13]. Adapun rumus sebagai berikut:

$$Z = \frac{X - \mu}{\sigma}$$

Dimana:

- X = nilai data mentah,
- μ = nilai rata-rata keseluruhan data,
- σ = nilai simpangan baku.

2.4 Reduksi Dimensi dengan PCA

Reduksi dimensi adalah teknik yang digunakan untuk menemukan pola dalam data. Teknik ini dapat menyederhanakan data tanpa kehilangan informasi penting serta meningkatkan kinerja klasterisasi [14]. Penelitian ini menggunakan metode PCA untuk reduksi dimensi data. Dengan menggunakan PCA, kita dapat secara optimal memilih jumlah komponen utama yang cukup untuk menjelaskan sebagian besar variansi dalam data [15]. Adapun perhitungan PCA adalah sebagai berikut:

$$X_{Pca} = W^T \cdot X$$

Dimana:

- X = data input,
- W = matriks vektor eigen (komponen utama),
- X_{pca} = data yang telah direduksi dimensinya.

2.5 K-Nearest Neighbor (KNN)

KNN adalah algoritma klasifikasi tanpa parameter tetap yang menentukan kategori dari suatu data dengan melihat kelas di antara k data terdekatnya dalam ruang fitur [16]. Untuk menentukan nilai k yang optimal dalam model ini, diterapkan metode *K-Fold Cross-Validation*. Setelah jumlah k yang optimal ditentukan, KNN digunakan untuk mengklasifikasikan data dengan mengukur kedekatan antara titik data menggunakan jarak Euclidean [17]. Rumus untuk perhitungan jarak Euclidean ditunjukkan sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Dengan:

- x dan y = dua titik data,
- x_i dan y_i = nilai atribut pada dimensi ke- i ,
- n = jumlah dimensi.

2.6 Support Vector Machine (SVM)

Support Vector Machine (SVM) termasuk teknik dalam bidang *machine learning* yang banyak diterapkan pada permasalahan pengelompokan maupun prediksi nilai. Konsep dasarnya adalah menemukan bidang pemisah terbaik yang mampu membedakan data ke dalam dua kelompok dengan jarak batas yang maksimal [18]. Secara matematis, *hyperplane* optimal pada kasus linear dapat dirumuskan sebagai berikut:

$$f(x) = \text{sign} \left(\sum_{i=1}^n \alpha_i y_i K(x_i, x) + b \right)$$

Dengan:

- x = vector data uji,
- x_i = vector data latih,
- $y_i \in \{-1, +1\}$ = label kelas data latih,
- α_i = bobot hasil optimasi,
- $K(x_i, x)$ = peran kernel dalam memetakan informasi ke ruang berdimensi lebih tinggi,
- b = bias.

2.7 Evaluasi Model

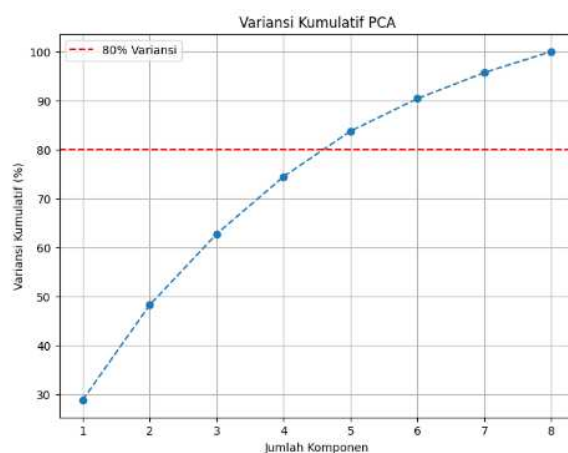
Setelah model KNN dilatih dengan data yang telah direduksi dimensinya menggunakan PCA, serta SVM dibangun dengan data yang sama, keduanya dievaluasi menggunakan beberapa metrik klasifikasi. Metrik penilaian performa yang dipakai antara lain akurasi, sensitivitas, spesifisitas, *balanced accuracy*, dan *Kappa*. Selain itu, untuk membandingkan performa antara dua model, digunakan uji statistik *McNemar* yang menguji signifikansi perbedaan hasil klasifikasi antar model terhadap data yang sama.

3. Hasil dan Pembahasan

Dataset mentah yang digunakan dalam penelitian ini terdiri dari 768 entri data pasien, yang mencakup delapan variabel prediktor medis: *Pregnancies*, *Glucose*, *BloodPressure*, *SkinThickness*,

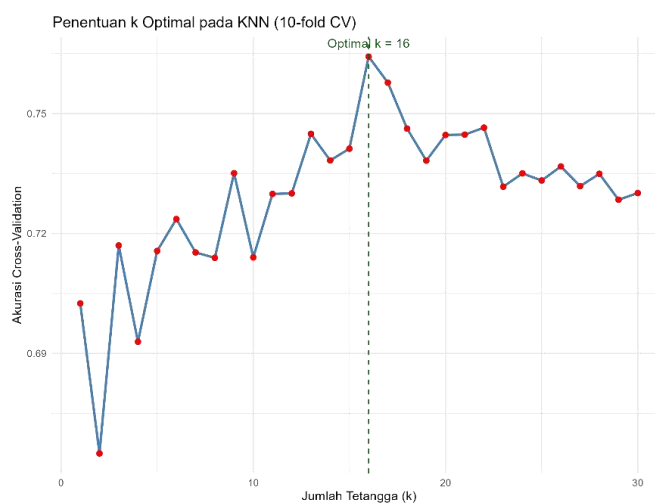
Insulin, BMI, *DiabetesPedigree Function*, dan *Age*, serta satu variabel target klasifikasi yaitu *Outcome*. Dataset ini telah melalui tahap praproses data untuk memastikan kualitas dan validitasnya. Salah satu tahap penting dalam praprosesan adalah penanganan nilai nol yang tidak realistis secara medis dan dianggap sebagai representasi dari nilai yang hilang implisit.

Berdasarkan hasil pemeriksaan, ditemukan sejumlah nilai nol pada beberapa atribut, yaitu *Glucose* sebanyak lima nilai, *BloodPressure* sebanyak 35 nilai, *SkinThickness* sebanyak 227 nilai, *Insulin* sebanyak 374 nilai, dan *BMI* sebanyak 11 nilai. Seluruh nilai nol pada atribut-atribut tersebut digantikan dengan nilai *median* untuk menghindari bias akibat *outlier* dan menjaga representasi distribusi data yang lebih stabil. Setelah proses imputasi, dilakukan proses standarisasi menggunakan metode *Z-score*, dengan tujuan menyamakan skala antar fitur agar perhitungan jarak dalam algoritma KNN dan SVM dapat dilakukan dengan efektif. Selanjutnya, dilakukan reduksi dimensi menggunakan metode PCA untuk metode KNN. PCA diterapkan terhadap delapan atribut medis yang telah dinormalisasi dan menghasilkan lima dimensi utama yang menangkap lebih dari 80% keragaman informasi dalam dataset, sebagaimana ditampilkan pada Gambar 2.



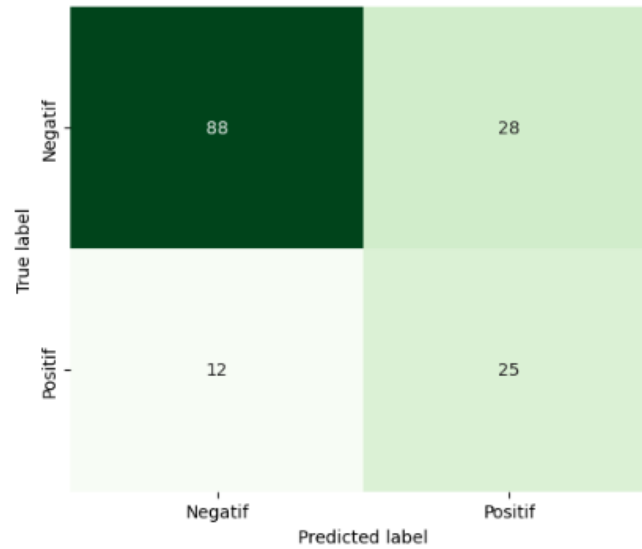
Gambar 2 Variansi Kumulatif PCA

Untuk membangun model klasifikasi, digunakan algoritma KNN. Sebelum melakukan pelatihan akhir, nilai k terbaik dilakukan dengan proses validasi silang menggunakan metode *K-Fold Cross Validation*. Berdasarkan hasil evaluasi terhadap akurasi dari berbagai nilai k , diperoleh bahwa nilai $k = 16$ memberikan akurasi tertinggi, sehingga dipilih sebagai parameter optimal dalam model KNN ini. Informasi tersebut ditunjukkan secara visual pada Gambar 3, di mana titik puncak kurva akurasi tercapai saat k bernilai 16.

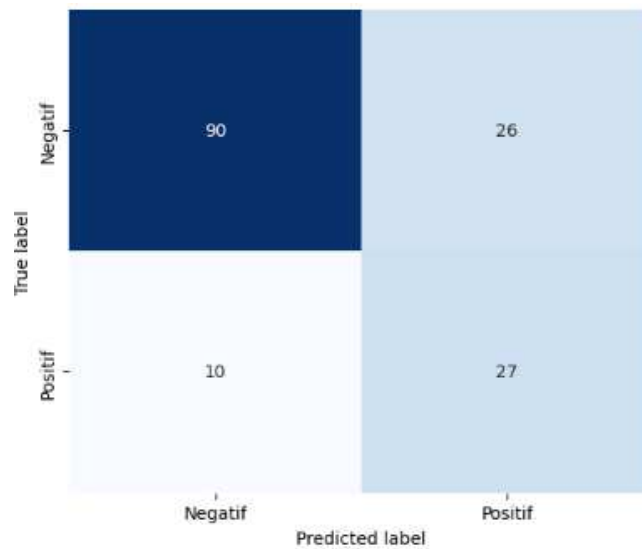


Gambar 3 Akurasi KNN Berdasarkan Nilai k (10-Fold CV)

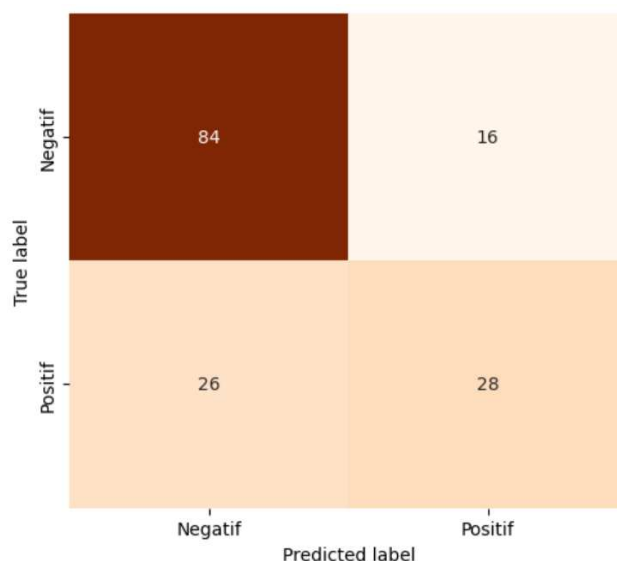
Model KNN kemudian diuji dalam dua skenario, yaitu menggunakan data mentah yang telah dinormalisasi (tanpa PCA) dan data yang telah direduksi dimensinya menggunakan PCA. Sebagai pembandingan, model SVM dikembangkan dengan menggunakan kernel RBF yang secara luas dikenal efektif dalam memetakan hubungan *non-linear* pada data medis. Berbeda dengan KNN yang diuji dengan dan tanpa PCA, model SVM hanya dilatih pada data hasil standarisasi tanpa reduksi dimensi, sehingga fokus analisis tetap berada pada perbedaan karakteristik algoritme. Hasil klasifikasi pada masing-masing model divisualisasikan dalam bentuk *confusion matrix* pada Gambar 4, Gambar 5, dan Gambar 6, sedangkan hasil evaluasi performa model berdasarkan metrik akurasi, sensitivitas, spesifisitas, *balanced accuracy*, *Kappa*, dan *McNemar's p-value* disajikan pada Tabel 2.



Gambar 4 *Confusion Matrix* KNN



Gambar 5 *Confusion Matrix* PCA dan KNN



Gambar 6 Confusion Matrix SVM

Tabel 2 Evaluasi Performa Model

Metrik	KNN Standar	PCA dan KNN	SVM
Akurasi	73,86%	76,47%	72,73%
Kappa	0,3786	0,4407	0,3745
Sensitivitas	88,00%	90,00%	51,85%
Spesifisitas	47,17%	50,94%	84,00%
Balanced Accuracy	67,58%	70,47%	67,93%
McNemar's p-value	0,0177	0,0124	0,1641

Hasil penelitian memperlihatkan bahwa setiap model memiliki karakteristik performa yang berbeda dalam prediksi risiko diabetes. Model KNN standar menghasilkan akurasi 73,86% dengan sensitivitas tinggi (88,00%), yang berarti cukup andal dalam mendeteksi pasien dengan diabetes. Namun, rendahnya spesifisitas (47,17%) menyebabkan banyak individu sehat salah terklasifikasi sebagai penderita (*false positive*), sehingga berpotensi menimbulkan *overdiagnosis*, kecemasan, dan biaya medis yang tidak perlu. Integrasi PCA pada KNN mampu meningkatkan akurasi menjadi 76,47% dan sensitivitas hingga 90,00%, menunjukkan bahwa reduksi dimensi berhasil menyederhanakan data medis berdimensi tinggi tanpa mengurangi informasi penting. Meski spesifisitas tetap rendah (50,94%), model ini lebih unggul dalam konteks medis karena deteksi dini pasien berisiko lebih diutamakan untuk mencegah komplikasi serius melalui intervensi cepat, baik berupa terapi farmakologis maupun perubahan gaya hidup.

Berbeda dengan KNN, model SVM menunjukkan performa yang lebih menekankan pada spesifisitas. Dengan akurasi 72,73% dan spesifisitas tinggi (84,00%), SVM lebih andal dalam mengenali pasien sehat sehingga mampu mengurangi kesalahan positif palsu. Namun, sensitivitas rendah (51,85%) membuat hampir separuh penderita diabetes tidak terdeteksi (*false negative*), yang dalam konteks medis lebih berbahaya karena pasien tidak segera memperoleh penanganan, berisiko mengalami komplikasi seperti kardiovaskular, nefropati, atau retinopati. Jika ditinjau dari *balanced accuracy*, PCA-KNN (70,47%) memberikan hasil yang lebih seimbang dibandingkan KNN standar maupun SVM. Oleh karena itu, PCA-KNN lebih sesuai diterapkan sebagai metode awal skrining populasi, sementara keunggulan SVM dalam spesifisitas dapat dimanfaatkan pada tahap pemeriksaan lanjutan, sehingga kombinasi keduanya berpotensi menghasilkan sistem prediksi yang lebih komprehensif dan akurat.

4. Kesimpulan

Penelitian ini berhasil menganalisis dan membandingkan performa PCA-KNN dan SVM dalam prediksi risiko diabetes dengan menyoroti keunggulan serta keterbatasan masing-masing metode. Integrasi PCA pada KNN terbukti mampu meningkatkan akurasi dan sensitivitas sehingga lebih efektif dalam mendeteksi pasien berisiko, meskipun kelemahan pada sisi spesifisitas masih berpotensi menghasilkan *false positive*. Di sisi lain, SVM menunjukkan keunggulan pada spesifisitas yang tinggi sehingga dapat mengurangi kesalahan diagnosis pada individu sehat, tetapi sensitivitas yang rendah menimbulkan risiko terlewatnya penderita diabetes (*false negative*). Kondisi ini menunjukkan bahwa tidak ada satu metode pun yang sepenuhnya optimal, melainkan masing-masing memiliki peran yang dapat saling melengkapi dalam konteks pengembangan sistem pendukung keputusan medis. Dari sudut pandang metodologis, penelitian ini masih terbatas pada penggunaan satu dataset publik dengan ukuran sampel yang relatif kecil dan tanpa eksplorasi parameter model yang lebih luas, sehingga temuan yang diperoleh belum tentu sepenuhnya merepresentasikan keragaman kondisi klinis di lapangan. Dengan mempertimbangkan hal tersebut, hasil yang diperoleh tetap memberikan gambaran penting mengenai potensi integrasi PCA-KNN pada tahap skrining awal populasi untuk deteksi dini, sementara SVM lebih sesuai digunakan sebagai metode konfirmasi pada tahap lanjutan. Pendekatan komplementer semacam ini diharapkan mampu menghasilkan sistem prediksi yang lebih akurat, seimbang, dan aplikatif dalam mendukung pengambilan keputusan medis.

Deklarasi

Kontribusi Penulis. Semua penulis berkontribusi secara bersama-sama dengan kontributor utama dalam artikel ini. Semua penulis membaca dan menyetujui versi akhir dari artikel yang diajukan.

Pernyataan Pendanaan. Tidak ada penulis yang menerima dana atau hibah dari lembaga atau badan pendanaan untuk penelitian ini.

Konflik Kepentingan. Penulis menyatakan tidak ada konflik kepentingan.

Informasi Tambahan. Tidak ada informasi tambahan dalam artikel ini.

Daftar Pustaka

- [1] American Diabetes Association, "Diagnosis and Classification of Diabetes Mellitus," *Diabetes Care*, vol. 36, suppl. 1, pp. S67–S74, Jan. 2013, doi: 10.2337/dc13-S067.
- [2] Mivtahurrahimah, and I. Sevtiyani, "Factors Contributing to Rising Diabetes Cases in Indonesia: Insights from the 2023 Indonesian Health Survey," *Jurnal Epidemiologi Kesehatan Komunitas*, vol. 10, no. 3, pp. 57–69, Jul. 2025, doi: 10.14710/jekk.v10i3.28184.
- [3] N. Suhandi, R. Gustriansyah, and A. Destria, "Klasifikasi Penyakit TBC Menggunakan Metode UMAP dan K-NN," *bit-tech*, vol. 7, no. 3, Apr. 2025, doi: 10.32877/bt.v7i3.2227.
- [4] Z. Zhang, "Introduction to Machine Learning: K-Nearest Neighbors," *Annals of Translational Medicine*, vol. 4, no. 11, pp. 1–7, Jun. 2016, doi: 10.21037/atm.2016.03.37.
- [5] A. Saifudin, E. Nirmala, and I. Kusyadi, "Penerapan Principal Component Analysis pada Model Deteksi Dini Anak Autisme," *Jurnal Teknologi Sistem Informasi dan Aplikasi*, vol. 7, no. 2, pp. 569–578, Apr. 2024, doi: 10.32493/jtsi.v7i2.38935.
- [6] I. T. Jolliffe and J. Cadima, "Principal Component Analysis: A Review and Recent Developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, Apr. 2016, doi: 10.1098/rsta.2015.0202.
- [7] F. O. Awalullaili, D. Ispriyanti, and T. Widiharih, "Klasifikasi Penyakit Hipertensi Menggunakan Metode SVM Grid Search dan SVM Genetic Algorithm (GA)," *Jurnal Gaussian*, vol. 11, no. 4, pp. 488–498, Dec. 2022, doi: 10.14710/j.gauss.11.4.488-498.
- [8] R. Guido, S. Ferrisi, D. Lofaro, and D. Conforti, "An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review," *Information*, vol. 15, no. 4, p. 235, Apr. 2024, doi: 10.3390/info15040235

-
- [9] M. I. Arsyad, A. Amran, A. Desiani, and M. J. Napitu, "Comparison of Classification Results of SVM, KNN, Decision Tree, and Ensemble Methods in Diabetes Diagnosis," *Journal of Medical Informatics Technology*, vol. 2, no. 3, pp. 76–82, Sep. 2024, doi: 10.37034/medinftech.v2i3.62.
- [10] H. S. W. Hovi, A. I. Hadiana, and F. R. Umbara, "Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM)," *Informatics and Digital Expert (INDEX)*, vol. 4, no. 1, Jul. 2022, doi: 10.36423/index.v4i1.895.
- [11] J. M. Akbar, M. Sabirin, G. S. Nugraha, and N. Alamsyah, "Penerapan Metode Principal Component Analysis (PCA) dan K-Nearest Neighbors (KNN) untuk Klasifikasi Data Kanker Paru-Paru," *Jurnal Teknologi Informasi, Komputer dan Aplikasinya (JTika)*, vol. 4, no. 2, Sep. 2022, doi: 10.29303/jtika.v4i2.232.
- [12] T. Septiana, M. A. Muda, D. Budiyanto, M. Pratama, and W. P. Jaya, "Analisis Penggunaan Support Vector Machine pada Deteksi Dini Penyakit Diabetes Melitus," *Jurnal Penelitian Inovatif (JUPIN)*, vol. 4, no. 3, pp. 1631–1640, Aug. 2024, doi: 10.54082/jupin.643.
- [13] I. N. Simbolon, "Prediksi Kualitas Air Sungai di Jakarta Menggunakan KNN yang Dioptimalisasi dengan PSO," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, Apr. 2024, doi: 10.23960/jitet.v12i2.4191.
- [14] R. Gustriansyah, J. Alie, and N. Suhandi, "A Hybrid Machine Learning Model for Market Clustering," *Engineering, Technology & Applied Science Research*, vol. 14, no. 6, pp. 18824–18828, Dec. 2024, doi: 10.48084/etasr.9259.
- [15] F. Badri and S. U. R. Sari, "Penerapan Metode Principal Component Analysis (PCA) untuk Identifikasi Faktor-Faktor yang Mempengaruhi Sikap Mahasiswa Memilih Melanjutkan Studi ke Kota Malang," *Jurnal Business Information Technology and System (BITS)*, vol. 3, no. 3, pp. 426–431 Dec. 2021, doi: 10.47065/bits.v3i3.1139.
- [16] D. H. Ramadhani, Jumadi, and G. Sandi, "Implementasi Algoritma K-Nearest Neighbors (KNN) untuk Prediksi Gizi Buruk," *SMATIKA: STIKI Informatika Jurnal*, vol. 14, no. 2, pp. 326–336, Dec. 2024, doi: 10.32664/smatika.v14i02.1360.
- [17] W. Wijiyanto, A. I. Pradana, S. Sopingi, and V. Atina, "Teknik K-Fold Cross Validation untuk Mengevaluasi Kinerja Mahasiswa," *Jurnal Algoritma*, vol. 21, no. 1, pp. 239–248, May 2024, doi: 10.33364/algoritma/v.21-1.1618.
- [18] H. Muradi, A. Saefuddin, I. M. Sumertajaya, A. M. Soleh, and D. D. Domiri, "Support Vector Regression (SVR) Method for Paddy Growth Phase Modeling Using Sentinel-1 Image Data," *Media Statistika*, vol. 16, no. 1, pp. 25–36, Jun. 2023, doi: 10.14710/medstat.16.1.25-36.