

Perbandingan Kernel Linear, Polynomial, dan RBF pada Metode Support Vector Machine untuk Klasifikasi Diabetes

Aulia Zahra Panjaitan ¹, Ella Debora Sri Karina Br. Tarigan ², Hilda Umayyah Anak Ampun ³, Peggy Mahara Dingga ⁴

^{1,2,3,4}Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Medan, 20221, Indonesia

Article Info

Article history:

Received month dd, yyyy

Revised month dd, yyyy

Accepted month dd, yyyy

Keywords:

Diabetes,
Support Vector Machine,
Kernel Polynomial,
Machine Learning.

ABSTRACT

Diabetes melitus merupakan penyakit tidak menular dengan prevalensi yang terus meningkat sehingga memerlukan metode deteksi dini yang akurat. Penelitian ini bertujuan untuk menganalisis kinerja metode *Support Vector Machine* (SVM) dengan tiga kernel, yaitu *Linear*, *Polynomial*, dan *Radial Basis Function* (RBF), dalam klasifikasi risiko diabetes. Data yang digunakan adalah dataset *Pima Indians Diabetes* yang terdiri dari 768 observasi dengan 8 variabel prediktor dan satu variabel target. Penelitian ini menggunakan pendekatan kuantitatif dengan tahapan eksplorasi data, prapemrosesan, serta evaluasi model menggunakan *10-Fold Cross Validation*. Kinerja model dievaluasi berdasarkan metrik *Accuracy*, *Sensitivity*, *Specificity*, *Kappa*, dan *ROC-AUC*. Hasil penelitian menunjukkan bahwa ketiga kernel memiliki kemampuan klasifikasi yang baik, namun kernel *Polynomial* memberikan performa terbaik dibandingkan kernel *Linear* dan *RBF*. Kernel *Polynomial* menghasilkan nilai *Accuracy* tertinggi sebesar 77,60%, serta nilai *Kappa* dan *ROC-AUC* yang lebih unggul. Selain itu, kernel *Polynomial* juga menunjukkan keseimbangan yang lebih baik dalam mengklasifikasikan pasien diabetes dan non-diabetes. Hasil ini menunjukkan bahwa pemilihan kernel pada metode SVM berpengaruh signifikan terhadap performa klasifikasi, dan kernel *Polynomial* merupakan model paling optimal pada dataset yang digunakan dalam penelitian ini.

 <https://doi.org/10.30598/stationerv4i1pp01-10>



This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution-ShareAlike 4.0 International License](#).

1. PENDAHULUAN

Diabetes melitus merupakan salah satu penyakit tidak menular yang menjadi perhatian serius di tingkat global. Kondisi ini terjadi akibat penurunan produksi hormon insulin oleh pankreas atau terjadinya resistensi insulin, sehingga glukosa dalam tubuh tidak dapat diproses secara optimal dan mengakibatkan peningkatan kadar gula darah atau hiperglikemia [1]. Komplikasi yang ditimbulkan sangat beragam, mulai dari gangguan penglihatan, penyakit jantung, stroke, hingga gagal ginjal, sehingga penanganan sejak dini menjadi sangat penting [2], [3]. Berdasarkan laporan International Diabetes Federation (IDF) tahun 2021, diperkirakan sekitar 537 juta orang di seluruh dunia hidup dengan diabetes, mewakili sekitar 10,5% dari populasi global, dan angka tersebut diproyeksikan meningkat menjadi 783 juta pada tahun 2045 [4]. Indonesia sendiri menempati peringkat kelima dengan jumlah penderita terbanyak, yakni 19,5 juta kasus pada 2021 dan diperkirakan mencapai 28,6 juta pada tahun 2045 [1].

Tingginya angka kejadian diabetes juga diperburuk oleh rendahnya kesadaran masyarakat, di mana hampir separuh dari seluruh penderita diabetes dewasa di dunia tidak menyadari kondisi medis mereka, dengan prevalensi tidak terdiagnosis tertinggi di kawasan Afrika, Pasifik Barat, dan Asia Tenggara masing-masing sebesar 53,6%, 52,8%, dan 51,3% [4]. Berbagai faktor risiko berkontribusi terhadap kejadian diabetes melitus tipe 2, meliputi usia di atas 45 tahun, obesitas, hipertensi, pola konsumsi tidak sehat, stres, serta riwayat keluarga yang terbukti memiliki hubungan bermakna dengan meningkatnya risiko penyakit ini [5]. Seseorang yang memiliki orang tua dengan riwayat diabetes memiliki risiko 3,4 hingga 3,5 kali lebih tinggi untuk terkena diabetes, dan risiko tersebut meningkat menjadi 6,1 kali apabila kedua orang tuanya menderita penyakit yang sama [1]. Kondisi ini menunjukkan betapa pentingnya upaya deteksi dini, terutama di negara-negara berkembang termasuk Indonesia.

Untuk mengatasi tantangan deteksi dini tersebut, pendekatan berbasis kecerdasan buatan dan machine learning mulai banyak dieksplorasi dalam bidang kesehatan. Pendekatan machine learning memiliki kemampuan mengenali hubungan nonlinier antar variabel dan menghasilkan prediksi yang lebih cepat serta akurat dibandingkan metode statistik konvensional [6], [7]. Berbagai algoritma seperti Naive Bayes, K-Nearest Neighbor, Decision Tree, Random Forest, dan Support Vector Machine (SVM) telah banyak diterapkan untuk klasifikasi diabetes dengan tingkat akurasi yang bervariasi [8], [9], [10]. Salah satu algoritma yang terbukti andal adalah SVM, yang bekerja dengan menemukan bidang pemisah optimal (hyperplane) untuk memisahkan data ke dalam dua kelas dengan margin yang maksimal [11], [12]. Keunggulan SVM antara lain kemampuan klasifikasi yang tinggi serta ketahanannya terhadap overfitting pada data berdimensi tinggi, sehingga sangat potensial sebagai alat bantu deteksi dini diabetes [13].

Performa SVM sangat dipengaruhi oleh fungsi kernel yang digunakan dalam proses klasifikasi. Kernel berfungsi mentransformasikan data ke dalam ruang fitur berdimensi lebih tinggi sehingga pola yang semula sulit dipisahkan menjadi dapat

terpisahkan secara linier [14]. Tiga jenis kernel yang paling umum digunakan adalah kernel linear yang efektif untuk hubungan antar fitur sederhana, kernel RBF yang unggul untuk pola nonlinier kompleks, dan kernel polynomial yang mampu memodelkan interaksi antar fitur dengan derajat tertentu [6]. Pemilihan kernel yang tepat menjadi kunci agar model SVM dapat mencapai keseimbangan antara akurasi dan kemampuan generalisasi terhadap data baru [6], [15].

Beberapa penelitian terdahulu menunjukkan hasil yang bervariasi terkait kernel terbaik dalam klasifikasi penyakit. Pada klasifikasi risiko stroke menggunakan data Indonesia Family Life Survey (IFLS), kernel linear menghasilkan akurasi tertinggi sebesar 99,0% dengan sensitivitas 97,0% dan spesifisitas 100,0%, mengungguli kernel RBF, polynomial, maupun sigmoid [14]. Pada penelitian prediksi risiko hipertensi, SVM dengan kernel linear memberikan ROC-AUC tertinggi sebesar 0,9518 meskipun kernel RBF mencatat akurasi sedikit lebih tinggi sebesar 87,38% [6]. Di sisi lain, pada klasifikasi diabetes menggunakan dataset Pima Indians, kernel RBF mampu mencapai akurasi hingga 95,45% setelah penerapan seleksi fitur [2], sementara penelitian lain menunjukkan bahwa SVM dengan kernel polynomial menghasilkan akurasi sebesar 86,48% pada dataset yang sama [11]. Perbedaan hasil ini menunjukkan bahwa tidak ada kernel yang secara universal unggul, sehingga perbandingan komprehensif antarkernel pada kasus diabetes masih perlu dikaji lebih lanjut [6], [7].

Sebagian besar penelitian terdahulu hanya berfokus pada satu jenis kernel atau satu algoritma tertentu tanpa melakukan perbandingan yang menyeluruh, serta belum menilai konsistensi performa model secara komprehensif menggunakan berbagai metrik evaluasi [6], [8], sehingga penelitian ini bertujuan untuk membandingkan kinerja *kernel Linear, Polynomial, dan Radial Basis Function (RBF)* pada metode *Support Vector Machine (SVM)* dalam mengklasifikasikan diabetes menggunakan metrik *Accuracy, Sensitivity, Specificity, Kappa*, dan *Receiver Operating Characteristic (ROC)*. Hasil penelitian diharapkan dapat menentukan kernel yang memberikan performa klasifikasi terbaik sehingga dapat digunakan sebagai referensi dalam pengembangan sistem deteksi dini diabetes."

2. METODE

2.1 Jenis dan Deskripsi Data

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen komputasional yang bertujuan untuk membandingkan kinerja beberapa model klasifikasi *Support Vector Machine (SVM)* dengan *kernel Linear, Polynomial, dan Radial Basis Function (RBF)* dalam mendeteksi risiko diabetes. Data yang digunakan merupakan data sekunder yang diperoleh dari dataset publik *Pima Indians Diabetes Dataset* yang diakses melalui sumber terbuka. Dataset ini terdiri atas 768 observasi dengan 8 variabel prediktor, yaitu *Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction*, dan *Age*, serta satu variabel target yaitu *Outcome* yang terdiri dari dua kelas, yaitu 0 (non-diabetes) dan 1 (diabetes). Seluruh proses analisis data, mulai dari *preprocessing* hingga evaluasi model dilakukan menggunakan perangkat

lunak **R Studio** dengan beberapa *package* pendukung untuk *machine learning* dan evaluasi model klasifikasi.

2.2 Tahapan Penelitian

1. Eksplorasi Data

Pada tahap ini dilakukan analisis awal untuk melihat karakteristik data seperti distribusi, nilai minimum–maksimum, serta proporsi kelas pada variabel target.

2. Pra-pemrosesan Data

Pra-pemrosesan meliputi pengecekan *missing value*, konversi variabel target menjadi faktor, serta standardisasi data untuk menghindari perbedaan skala antar variabel yang dapat memengaruhi performa SVM.

3. Pembagian Data

Data dibagi menjadi data pelatihan dan data pengujian secara implisit melalui metode *cross validation*. Evaluasi model dilakukan menggunakan metode **10-Fold Cross Validation**, yaitu membagi data menjadi 10 bagian (fold). Setiap fold secara bergantian digunakan sebagai data uji, sedangkan sisanya digunakan sebagai data latih. Proses ini diulang sebanyak 10 kali untuk mengurangi bias hasil evaluasi dan meningkatkan stabilitas model.

2.3 Metode Klasifikasi

Metode utama yang digunakan dalam penelitian ini adalah **Support Vector Machine (SVM)** dengan tiga jenis kernel, yaitu:

- **Kernel Linear**: digunakan untuk data yang bersifat linier atau hampir linier.
$$K = (X_i, X_j) = (X_i \times X_j) \quad (1)$$
- **Kernel Polynomial**: digunakan untuk menangkap hubungan nonlinier dengan derajat polinomial tertentu.
$$K = (X_i, X_j) = (X_i \times X_j + \mathbf{1})^p \quad (2)$$
- **Kernel RBF (Radial Basis Function)**: digunakan untuk menangani pola data yang kompleks dan nonlinier.
$$K = (X_i, X_j) = e^{-\gamma(X_i, X_j)^2} \quad (3)$$

Setiap kernel dievaluasi untuk melihat perbedaan performa dalam mengklasifikasikan data diabetes [16].

2.5 Metode Evaluasi

Kinerja model diukur menggunakan beberapa metrik evaluasi, yaitu:

- Accuracy (akurasi) : digunakan untuk mengukur tingkat ketepatan keseluruhan model dalam melakukan klasifikasi.
- Sensitivity (recall) : digunakan untuk melihat kemampuan model dalam mendeteksi kasus positif diabetes dengan benar.
- Specificity : digunakan untuk mengukur kemampuan model dalam mengenali kasus negatif (non-diabetes).
- Kappa : digunakan untuk menilai tingkat kesepakatan antara hasil prediksi model dan data aktual dengan mempertimbangkan kemungkinan kesalahan secara acak.
- ROC-AUC : digunakan untuk mengevaluasi kemampuan model dalam membedakan dua kelas secara keseluruhan berdasarkan kurva *Receiver Operating Characteristic*.

3. HASIL DAN PEMBAHASAN

3.1 Statistik Deskriptif

Dataset yang digunakan terdiri atas 768 observasi, 8 variabel prediktor, dan 1 variabel target (Outcome). Variabel prediktor meliputi *Pregnancies*, *Glucose*, *BloodPressure*, *SkinThickness*, *Insulin*, *BMI*, *DiabetesPedigreeFunction*, dan *Age*.

Tabel 1. Statistik Deskriptif

Variabel	Minimum	Maksimum	Rata-rata
Pregnancies	0	17	3,845
Glucose	0	199	120,9
BloodPressure	0	122	69,11
SkinThickness	0	99	20,54
Insulin	0	846	79,80
BMI	0	67,10	31,99
DiabetesPedigreeFunction	0,078	2,420	0,472
Age	21	81	33,24

Berdasarkan statistik deskriptif, rata-rata jumlah kehamilan adalah 3,845 kali, rata-rata kadar glukosa 120,9 mg/dL, tekanan darah 69,11 mmHg, ketebalan lipatan kulit 20,54 mm, kadar insulin 79,8 μ U/mL, BMI 31,99 kg/m², nilai *Diabetes Pedigree Function* 0,472, dan usia 33,24 tahun.

Variabel target *Outcome* memiliki rata-rata 0,349, yang menunjukkan bahwa sekitar 34,9% data termasuk kategori diabetes dan 65,1% termasuk kategori non-diabetes. Selain itu, beberapa variabel seperti *Glucose*, *BloodPressure*, *SkinThickness*, *Insulin*, dan *BMI* memiliki nilai minimum 0, yang mengindikasikan adanya data yang tidak terisi atau nilai yang tidak realistis secara medis. Oleh karena itu, dilakukan proses standardisasi data sebelum pemodelan menggunakan *Support Vector Machine* (SVM).

3.2 Pra-pemrosesan Data

Pada tahap pra-pemrosesan, dilakukan pengecekan missing value pada seluruh variabel menggunakan fungsi `colSums(is.na(data))`. Hasil pengecekan menunjukkan bahwa tidak terdapat missing value pada dataset, sehingga seluruh data dapat digunakan dalam proses pemodelan. Selanjutnya, variabel target *Outcome* diubah ke dalam bentuk faktor dengan dua kategori, yaitu "No" untuk pasien non-diabetes dan "Yes" untuk pasien diabetes. Transformasi ini dilakukan agar variabel target sesuai dengan kebutuhan metode klasifikasi *Support Vector Machine* (SVM) dalam proses pelatihan dan evaluasi model.

3.3 Evaluasi Menggunakan 10-Fold Cross Validation

Evaluasi model dilakukan menggunakan metode 10-Fold Cross Validation. Pada metode ini, dataset dibagi menjadi 10 bagian (*fold*) yang berukuran hampir sama. Setiap fold secara bergantian digunakan sebagai data pengujian, sedangkan sembilan fold lainnya digunakan sebagai data pelatihan. Proses ini dilakukan sebanyak 10 kali sehingga setiap data memiliki kesempatan yang sama untuk menjadi data pelatihan maupun data pengujian.

Pada penelitian ini, evaluasi model dilakukan menggunakan fungsi `trainControl()` dengan parameter `method = "cv"` dan `number = 10`. Selain itu, digunakan `classProbs = TRUE` untuk menghasilkan probabilitas kelas, `summaryFunction = twoClassSummary` untuk menghitung metrik evaluasi seperti

ROC, sensitivitas (*Sensitivity*), dan spesifisitas (*Specificity*), serta `savePredictions = "final"` untuk menyimpan hasil prediksi pada setiap fold. Penggunaan *10-Fold Cross Validation* bertujuan untuk memperoleh hasil evaluasi yang lebih stabil dan mengurangi bias yang dapat muncul akibat pembagian data pelatihan dan pengujian yang dilakukan hanya satu kali.

3.4 Hasil SVM Kernel Linear

Model SVM dengan kernel Linear diterapkan pada dataset diabetes yang terdiri atas 768 observasi dan 8 variabel prediktor. Evaluasi model dilakukan menggunakan metode *10-Fold Cross Validation* untuk memperoleh hasil yang lebih stabil dan representatif. Kinerja model diukur menggunakan Accuracy, Sensitivity, Specificity, Kappa, dan ROC.

Tabel 2. Hasil Evaluasi SVM Kernel Linear

Metrik	Nilai
Accuracy	0,755
Sensitivity	0,563
Specificity	0,858
Kappa	0,439
ROC	0,832

Berdasarkan Tabel 2, kernel Linear menghasilkan nilai *Accuracy* sebesar 77,47%, yang menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data dengan benar. Nilai *Sensitivity* sebesar 55,60% menunjukkan bahwa model mampu mendeteksi sekitar 55,60% pasien yang benar-benar menderita diabetes. Sementara itu, nilai *Specificity* sebesar 89,20% menunjukkan kemampuan yang sangat baik dalam mengenali pasien yang tidak menderita diabetes. Nilai ROC sebesar 0,829 mengindikasikan bahwa model memiliki kemampuan klasifikasi yang baik dalam membedakan kedua kelas.

3.5 Hasil SVM Kernel Polynomial

Model SVM dengan kernel Polynomial digunakan untuk mengakomodasi kemungkinan adanya hubungan nonlinier pada data. Evaluasi dilakukan menggunakan metode *10-Fold Cross Validation* dengan metrik yang sama seperti pada kernel Linear.

Tabel 3. Hasil Evaluasi SVM Kernel Polynomial

Metrik	Nilai
Accuracy	0,776
Sensitivity	0,563
Specificity	0,890
Kappa	0,479
ROC	0,838

Berdasarkan Tabel 3, kernel Polynomial menghasilkan *Accuracy* sebesar 77,60%, yang merupakan nilai tertinggi dibandingkan kernel lainnya. Nilai *Sensitivity* sebesar 56,34% menunjukkan bahwa kemampuan model dalam mendeteksi pasien diabetes sedikit lebih baik dibandingkan kernel Linear. Selain itu, nilai *Specificity* sebesar 89,00% menunjukkan bahwa model tetap mampu mengenali pasien non-diabetes dengan baik. Nilai Kappa sebesar 0,479 dan ROC sebesar 0,838 juga merupakan yang tertinggi di antara ketiga kernel, sehingga menunjukkan bahwa kernel Polynomial memiliki performa klasifikasi yang paling baik pada dataset yang digunakan.

3.6 Hasil SVM Kernel RBF

Kernel Radial Basis Function (RBF) digunakan untuk menangkap pola yang lebih kompleks dan nonlinier pada data. Seperti model sebelumnya, evaluasi dilakukan menggunakan metode 10-Fold Cross Validation.

Tabel 4. Hasil Evaluasi SVM Kernel RBF

Metrik	Nilai
Accuracy	0,755
Sensitivity	0,563
Specificity	0,858
Kappa	0,439
ROC	0,832

Berdasarkan Tabel 4, kernel RBF menghasilkan Accuracy sebesar 75,52%, yang lebih rendah dibandingkan kernel Linear dan Polynomial. Meskipun demikian, nilai Sensitivity sebesar 56,34% menunjukkan kemampuan deteksi diabetes yang sama dengan kernel Polynomial. Nilai Specificity sebesar 85,80% menunjukkan bahwa kemampuan model dalam mengenali pasien non-diabetes masih tergolong baik, meskipun lebih rendah dibandingkan dua kernel lainnya. Nilai ROC sebesar 0,832 mengindikasikan bahwa kernel RBF tetap memiliki kemampuan klasifikasi yang baik, namun performanya secara keseluruhan masih berada di bawah kernel Polynomial.

3.7 Perbandingan Ketiga Kernel

Setelah dilakukan pemodelan menggunakan kernel Linear, Polynomial, dan RBF, selanjutnya dilakukan perbandingan performa untuk menentukan kernel terbaik dalam klasifikasi diabetes. Perbandingan dilakukan berdasarkan lima metrik evaluasi, yaitu Accuracy, Sensitivity, Specificity, Kappa, dan ROC yang diperoleh melalui metode 10-Fold Cross Validation.

Tabel 5. Perbandingan Performa Kernel SVM

Kernel	Accuracy	Sensitivity	Specificity	Kappa	ROC
Linear	0,775	0,556	0,892	0,475	0,829
Polynomial	0,776	0,563	0,890	0,479	0,838
RBF	0,755	0,563	0,858	0,439	0,832

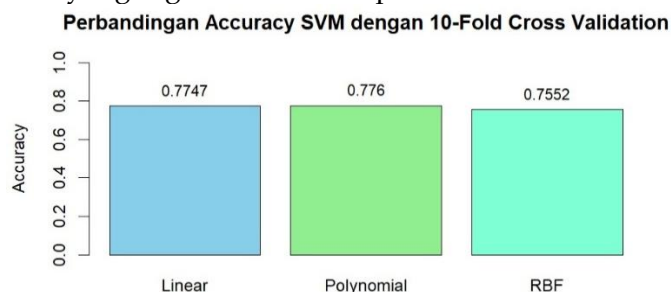
Berdasarkan Tabel 5, kernel Polynomial menunjukkan performa terbaik dengan nilai Accuracy sebesar 77,60%, Kappa sebesar 0,479, dan ROC sebesar 0,838. Nilai tersebut merupakan yang tertinggi dibandingkan kernel Linear maupun RBF, sehingga menunjukkan bahwa kernel Polynomial memiliki kemampuan klasifikasi yang lebih baik secara keseluruhan.

Ditinjau dari kemampuan mendeteksi pasien diabetes, kernel Polynomial dan RBF menghasilkan nilai Sensitivity yang sama, yaitu sebesar 56,34%, sedangkan kernel Linear menghasilkan Sensitivity sebesar 55,60%. Meskipun perbedaannya relatif kecil, hasil ini menunjukkan bahwa kernel Polynomial dan RBF sedikit lebih baik dalam mengidentifikasi pasien yang benar-benar menderita diabetes.

Sementara itu, kernel Linear menghasilkan nilai Specificity tertinggi sebesar 89,20%, yang menunjukkan bahwa kernel ini paling baik dalam mengenali pasien yang tidak menderita diabetes. Namun, karena tujuan penelitian tidak hanya mengenali

pasien non-diabetes tetapi juga mendeteksi pasien diabetes secara akurat, maka evaluasi tidak hanya berfokus pada Specificity.

Secara keseluruhan, kernel Polynomial memberikan keseimbangan yang lebih baik antara Accuracy, Sensitivity, Specificity, Kappa, dan ROC dibandingkan kedua kernel lainnya. Oleh karena itu, berdasarkan hasil evaluasi menggunakan metode 10-Fold Cross Validation, kernel Polynomial dipilih sebagai kernel terbaik untuk klasifikasi diabetes pada dataset yang digunakan dalam penelitian ini.



Gambar 1. Perbandingan Accuracy Ketiga Kernel SVM

Berdasarkan barplot pada Gambar 1 menunjukkan bahwa kernel Polynomial memiliki nilai Accuracy tertinggi, diikuti oleh kernel Linear dan kernel RBF. Meskipun selisih Accuracy antar kernel relatif kecil, kernel Polynomial tetap memberikan performa yang paling baik berdasarkan hasil evaluasi keseluruhan.

Berdasarkan Tabel 5 dan Gambar 1, kernel Polynomial menghasilkan nilai Accuracy tertinggi sebesar 77,60%, lebih tinggi dibandingkan kernel Linear sebesar 77,47% dan kernel RBF sebesar 75,52%. Hasil ini menunjukkan bahwa kernel Polynomial memiliki kemampuan klasifikasi yang lebih baik dalam mengelompokkan data pasien diabetes dan non-diabetes pada dataset yang digunakan. Selain menghasilkan Accuracy tertinggi, kernel Polynomial juga memperoleh nilai Kappa sebesar 0,479 dan ROC sebesar 0,838 yang lebih tinggi dibandingkan kernel lainnya.

Temuan ini sejalan dengan penelitian [17] yang membandingkan *kernel Linear*, *Polynomial*, RBF, dan *Sigmoid* pada klasifikasi penyakit jantung menggunakan metode SVM, di mana *kernel Polynomial* menghasilkan performa terbaik dengan akurasi sebesar 88,52%. Kesamaan hasil tersebut mengindikasikan bahwa *kernel Polynomial* memiliki kemampuan yang baik dalam memodelkan hubungan nonlinier pada data kesehatan sehingga dapat meningkatkan performa klasifikasi. Selain itu, penelitian [18] menunjukkan bahwa metode SVM mampu memberikan kinerja yang baik dalam prediksi status anemia, sehingga semakin memperkuat bahwa SVM merupakan metode yang efektif untuk permasalahan klasifikasi pada bidang kesehatan. Di sisi lain, penelitian [6] menunjukkan bahwa performa terbaik suatu kernel sangat dipengaruhi oleh karakteristik data yang digunakan, sehingga *kernel Linear* maupun RBF dapat memberikan hasil yang lebih baik pada dataset tertentu. Oleh karena itu, perbedaan hasil antarpelitian merupakan hal yang wajar karena dipengaruhi oleh jenis dataset, jumlah atribut, teknik prapemrosesan data, serta parameter model yang digunakan. Berdasarkan hasil evaluasi yang diperoleh, *kernel Polynomial* dapat ditetapkan sebagai model terbaik pada penelitian ini karena menghasilkan tingkat akurasi tertinggi dibandingkan kernel Linear dan RBF.

4. KESIMPULAN

Hasil analisis menunjukkan bahwa model *Support Vector Machine* (SVM) dengan kernel *Linear*, *Polynomial*, dan *Radial Basis Function* (RBF) mampu mengklasifikasikan data diabetes dengan kinerja yang cukup baik. Perbedaan nilai pada setiap metrik evaluasi, yaitu *Accuracy*, *Sensitivity*, *Specificity*, *Kappa*, dan ROC-AUC, mengindikasikan adanya variasi kemampuan masing-masing kernel dalam memisahkan kelas diabetes dan non-diabetes. Di antara ketiga model yang diuji, kernel *Polynomial* menghasilkan performa yang paling unggul dengan nilai *Accuracy* tertinggi serta nilai *Kappa* dan ROC-AUC yang lebih baik dibandingkan kernel *Linear* maupun RBF. Selain itu, kernel *Polynomial* juga menunjukkan keseimbangan yang lebih baik dalam mengklasifikasikan kedua kelas, sehingga memberikan hasil prediksi yang lebih stabil.

Hasil penelitian ini menunjukkan bahwa metode *Support Vector Machine* (SVM) efektif digunakan dalam permasalahan klasifikasi medis, khususnya pada data diabetes, serta pemilihan kernel yang tepat memiliki pengaruh yang signifikan terhadap kualitas hasil klasifikasi.

Informasi Pendanaan

Penulis menyatakan bahwa penelitian ini tidak menerima pendanaan dari pihak mana pun.

Pernyataan Kontribusi Penulis

Aulia Zahra Panjaitan: Konseptualisasi, pengumpulan data, mengolah data, visualisasi dan validasi.

Ella Debora Sri Karina Br. Tarigan: Penyusunan draf awal naskah, validasi, serta penulisan tinjauan.

Hilda Umayyah Anak Ampun: Konseptualisasi, pengumpulan data, metodologi, supervisi, Validasi dan penyuntingan naskah.

Peggy Mahara Dingga: Penulisan bagian tinjauan serta penelusuran, validasi, penyuntingan naskah dan pengumpulan referensi.

Seluruh penulis berkontribusi dalam diskusi hasil penelitian serta penyusunan dan penyempurnaan naskah akhir.

Pernyataan Konflik Kepentingan

Penulis menyatakan tidak memiliki konflik kepentingan, baik dalam bentuk hubungan finansial maupun hubungan pribadi yang dapat memengaruhi hasil penelitian ini.

Ketersediaan Data

Data yang digunakan dalam penelitian ini merupakan dataset publik *Pima Indians Diabetes* dari website KAGGLE. *Dataset* yang dapat diakses secara terbuka melalui repositori data <https://www.kaggle.com/code/alaabdelstar/diabetes/input>.

REFERENSI

- [1] J. P. Kamagi, B. T. Ratag, E. M. Mantjoro, and W. P. J. Kaunang, "Faktor-faktor yang berhubungan dengan diabetes melitus pada pasien," *HOLISTIK J. Kesehat.*, vol. 20, no. 1, pp. 282–291, 2026, doi: <https://doi.org/10.33024/hjk.v20i1.2457>.
- [2] N. Al Rasyid, I. Afrianty, E. Budianita, and S. K. Gusti, "Diabetes Classification using Gain Ratio Feature Selection in Support Vector Machine Method," *Bull. Informatics Data Sci.*, vol. 4, no. 1, pp. 10–21, 2025, doi: <https://doi.org/10.61944/bids.v4i1.114>.
- [3] N. D. Hendrawan, A. S. Affandi, and R. F. Fadhilrifat, "Implementasi Support Vector Machine dalam Deteksi Diabetes Melalui Indikator Kesehatan 3 rd E-proceeding SENRIABDI 2023," *Semin. Nas. Has. Ris. dan Pengabd. Kpd. Masy. Univ. Sahid Surakarta*, vol. 3, pp. 664–671, 2023, [Online]. Available: <https://jurnal.usahidsolo.ac.id/>
- [4] J. M. Hossain, M. Al-Mamun, and M. R. Islam, "Diabetes mellitus, the fastest growing global public health concern : Early detection should be focused," *Heal. Sci. Reports - WILEY*, pp. 1–5, 2024, doi: <https://doi.org/10.1002/hsr2.2004>.
- [5] R. A. S. Kabosu, A. A. Adu, and I. A. Hinga, "Faktor Risiko Kejadian Diabetes Melitus Tipe Dua di RS Bhayangkara Kota Kupang," *Timorese J. Public Heal.*, vol. 1, no. 1, pp. 11–23, 2019, [Online]. Available: <https://ojsfkmundana.science/index.php/t/notification>
- [6] Saifuddin, L. Azhari, E. Widarti, and Wartono, "Evaluasi Kinerja Kernel Linear , RBF , dan Polynomial pada Model Support Vector Machine untuk Prediksi Risiko Hipertensi," *J. Ilm. FIFO*, vol. 17, no. 2, pp. 192–201, 2025, doi: <https://doi.org/10.22441/fifo.2025.v17i2.008>.
- [7] A. M. Ridwan and G. D. Setiawan, "PERBANDINGAN BERBAGAI MODEL MACHINE LEARNING UNTUK MENDETEKSI DIABETES," *TEKNOKOM J. Teknol. dan Rekayasa Sist. Komput.*, vol. 6, no. 2, pp. 127–132, 2023, doi: <https://doi.org/10.31943/teknokom.v6i2.152>.
- [8] A. W. Mucholladin, F. A. Bachtiar, and M. T. Furqon, "Klasifikasi Penyakit Diabetes menggunakan Metode Support Vector Machine," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 2, pp. 622–633, 2021, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [9] P. R. Putri and R. Alit, "Klasifikasi Penyakit Diabetes Melitus Menggunakan Metode Support Vector Machine (SVM)," *JINACS (Journal Informatics Comput. Sci.*, vol. 06, no. 03, pp. 740–746, 2024.
- [10] E. Dritsas and M. Trigka, "Data-Driven Machine-Learning Methods for Diabetes Risk Prediction," *Sensors*, vol. 22, no. 5304, pp. 1–18, 2022, doi: <https://doi.org/10.3390/s22145304>.
- [11] D. Triyanto, "SUPPORT VECTOR MACHINE UNTUK KLASIFIKASI PENYAKIT DIABETES MELLITUS," *Media Teknol. Dan Inform.*, vol. 1, no. 3, pp. 147–152, 2024.
- [12] I. M. K. Karo and Hendriyana, "KLASIFIKASI PENDERITA DIABETES MENGGUNAKAN ALGORITMA MACHINE LEARNING DAN Z-SCORE," *J. Teknol. Terpadu*, vol. 8, no. 2, pp. 94–99, 2022, [Online]. Available: <https://journal.nurulfikri.ac.id/index.php/jtt>
- [13] T. Septiana, M. A. Muda, D. Budiyo, M. Pratama, and W. P. Jaya, "Analisis Penggunaan Support Vector Machine pada Deteksi Dini Penyakit Diabetes Melitus," *J. Penelit. Inov.*, vol. 4, no. 3, pp. 1631–1640, 2024, doi: <https://doi.org/10.54082/jupin.643>.
- [14] L. R. Safitri, N. Chamidah, T. Saifudin, and G. T. Alpandi, "Comparison of Kernel Support Vector Machine In Stroke Risk Classification (Case Study : IFLS data)," *Int. Conf. Dta Sci. Off. Stat.*, pp. 309–316, 2023.
- [15] C. Z. V. Junus, Tarno, and P. Kartikasari, "KLASIFIKASI MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DAN RANDOM FOREST UNTUK DETEKSI AWAL

RISIKO DIABETES MELITUS," J. GAUSSIAN, vol. 11, no. 3, pp. 386–396, 2023, doi: <https://doi.org/10.14710/j.gauss.11.3.386-396>.

- [16] T. M. P. Aulia, N. Arifin, and R. Mayasar, "PERBANDINGAN KERNEL SUPPORT VECTOR MACHINE (SVM) DALAM PENERAPAN ANALISIS SENTIMEN VAKSINISASI COVID-19," SINTECH J., vol. 4, no. 2, pp. 139–145, 2021, doi: <https://doi.org/10.31598>.
- [17] A. Faradisia and M. A. I. Pakereng, "Comparative Analysis of Linear , Polynomial , RBF , and Sigmoid Kernels in Support Vector Machine for Heart Disease Classification," MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 5, no. 4, pp. 1531–1537, 2025, doi: <https://doi.org/10.57152/malcom.v5i4.2321>.
- [18] T. M. Sari, L. Hakim, P. Studi, T. Informatika, and U. Y. Pasuruan, "PREDIKSI STATUS ANEMIA DENGAN PENDEKATAN PEMBELAJARAN MESIN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN SELEKSI FITUR FIREFLY ALGORITHM," J. INSTEK, vol. 10, no. 1, pp. 120–129, 2025, [Online]. Available: <https://journal.uin-alauddin.ac.id/index.php/instek>