JURNAL LOCUS: Penelitian & Pengabdian
Volume 4 No. 8 Agustus 2025
E-ISSN 2829-7334| P-ISSN 2829-5439
Hompage: https://locus.rivierapublishing.id/index.php/jl

Prediksi Tingkat Keparahan Diabetes Melitus Menggunakan *Support Vector Machine* (SVM) dengan Kernel Polinomial dan RBF

Muhammad Rifqi Pradana*, Wina Witanti, Agus Komarudin
Universitas Jenderal Achmad Yani, Indonesia
Email: rifqipradana03@gmail.com*, Witanti@gmail.com, agus.komarudin@lecture.unjani.ac.id

Abstrak:
Diabetes melitus merupakan penyakit kronis dengan prevalensi tinggi yang memerlukan diagnosis dini dan penanganan yang akurat guna memitigasi risiko komplikasi serius. Penelitian ini berfokus pada pengembangan model prediktif untuk menentukan tingkat keparahan diabetes melitus, mengimplementasikan algoritma *Support Vector Machine* (SVM) dengan membandingkan kinerja fungsi kernel Polinomial dan Radial Basis Function (RBF). Dataset yang dimanfaatkan terdiri dari 100.000 entri data pasien, yang dikategorikan ke dalam dua kelas tingkat keparahan penyakit. Sebelum memasuki fase klasifikasi, serangkaian tahap pra-pemrosesan data diterapkan, termasuk penanganan *outlier*, untuk memastikan kualitas data dan meningkatkan robustnya model. Metodologi penelitian mencakup studi literatur mendalam, implementasi proses klasifikasi, serta pengujian sistem menggunakan pendekatan *black box testing* untuk memvalidasi fungsionalitas dan akurasi. Hasil eksperimen secara konsisten menunjukkan bahwa SVM dengan kernel RBF memberikan performa superior dibandingkan dengan SVM yang menggunakan kernel Polinomial, tercermin dari tingkat akurasi klasifikasi yang lebih tinggi secara signifikan. Oleh karena itu, sistem klasifikasi berbasis SVM yang dikembangkan ini berpotensi besar untuk mendukung diagnosis awal dan penentuan otomatis tingkat keparahan diabetes melitus dengan tingkat keakuratan yang kredibel. Kontribusi penelitian ini diharapkan dapat menjadi referensi berharga bagi pengembangan lebih lanjut sistem pendukung keputusan klinis dalam domain kesehatan, khususnya terkait manajemen diabetes.

Kata kunci: Diabetes Melitus; Klasifikasi; *Support Vector Machine* (SVM); *Outlier*; Prediksi Tingkat Keparahan.

Abstract:
Diabetes mellitus is a chronic disease with a high prevalence that requires early diagnosis and accurate treatment to mitigate the risk of serious complications. This study focuses on the development of a predictive model to determine the severity of diabetes mellitus, implementing a Support Vector Machine (SVM) algorithm by comparing the performance of the Polynomial kernel function and the Radial Base Function (RBF). The dataset used consisted of 100,000 patient data entries, which were categorized into two classes of disease severity. Before entering the classification phase, a series of pre-processing stages of data are implemented, including outlier handling, to ensure data quality and improve the robustness of the model. The research methodology includes an in-depth literature study, implementation of the classification process, as well as testing of the system using the black box testing approach to validate functionality and accuracy. Experimental results consistently show that SVMs with RBF kernels provide superior performance compared to SVMs using the Polynomial kernel, reflected by a significantly higher level of classification accuracy. Therefore, the SVM-based classification system developed has great potential to support the early diagnosis and automatic determination of the severity of diabetes mellitus with a credible level of accuracy. The contribution of this research is expected to be a valuable reference for the further development of clinical decision support systems in the health domain, especially related to diabetes management.

Keywords: *diabetes mellitus; Classification; Support Vector Machine (SVM); Outlier; Severity Prediction.*

Corresponding: Muhammad Rifqi Pradana
E-mail: rifqipradana03@gmail.com

PENDAHULUAN

Diabetes Melitus sering dianggap sebagai masalah kesehatan global yang signifikan. Sebagai penyakit kronis yang diderita seumur hidup, diabetes ditandai oleh peningkatan kadar glukosa darah yang tidak terkontrol (Bommer et al., 2018; Seuring et al., 2015). Berdasarkan data *International Diabetes Federation* terbaru, sekitar 11.1% atau 1 dari 9 orang dewasa (20–79 tahun) hidup dengan diabetes pada 2024, dengan sekitar 589 juta kasus, dan diproyeksikan meningkat menjadi 853 juta pada tahun 2050 (IDF Diabetes Atlas, 2025). Angka kematian global mencapai lebih dari 3,4 juta jiwa akibat diabetes setiap tahun (IDF Diabetes Atlas, 2025). Epidemiologi terbaru mengungkap bahwa prevalensi telah meroket — lebih dari 830 juta orang mengalami diabetes pada 2022, dan sebagian besar tidak tertangani dengan baik, terutama di negara berpenghasilan rendah dan menengah (NCD-RisC/WHO, 2024; Ong et al., 2023). Dari segi ekonomi, beban global diabetes diperkirakan mencapai lebih dari US 1 triliun per tahun, dengan peningkatan tajam sejak 2015, dan mencapai hampir 2.2% dari PDB global pada tahun 2030 (Bommer et al., 2018; Seuring et al., 2015). Efek sosial dan ekonomi ini memperburuk kualitas hidup pasien dan menimbulkan tekanan besar terhadap sistem kesehatan (Seuring et al., 2015; Bommer et al., 2018). Adapun beberapa faktor genetika/keturunan bahkan pola hidup lingkungan yang tidak sehat sangat rentan terjadinya penyakit diabetes melitus (Puspa, 2017; Yeni, 2018). Diabetes melitus memiliki gejala meningkatnya kadar gula dengan rentan waktu yang lama pada sekelompok gangguan metabolisme (Lestari et al, 2021). Tubuh pasien yang terdiagnosa diabetes melitus tidak mampu memproduksi hormon insulin dengan baik, sehingga kadar gula dalam aliran darah mengalami peningkatan (Gunawan, Sugiarto dan Mardinto, 2020). Diabetes melitus juga sering dianggap sebagai penyakit kronis dengan cepat, yang disebabkan oleh menurunnya relative sensitivitas sel terhadap insulin, atau bahkan kematian seseorang yang didiagnosa diabetes melitus sering disebabkan oleh komplikasi (Purnama et al., 2019).

Pada penelitian sebelumnya yang dilakukan oleh (Pungky & Rosalya Putri, 2024) terdapat celah yang dapat digunakan sebagai pengembangan dari penelitian sebelumnya yaitu, dalam ukuran dataset yang digunakan pada penelitian sebelumnya memiliki 768 record penyakit diabetes melitus, memiliki 8 fitur dan memiliki 2 kelas yaitu positif dan negatif.

Pada penelitian sebelumnya menggunakan metode Support Vector Machine (SVM) sebagai metode yang digunakan dalam penelitian tersebut dengan menggunakan kernel jenis linear. Setelah itu, pada penelitian sebelumnya memiliki data split yang dimana dijadikan 2 bagian yaitu : data testing dan data pelatihan. Untuk data pelatihannya penulis mengambil sebanyak 80% dan data testing nya sebanyak 20% dari 768 record. Lalu, terdapat output yang dihasilkan berupa hasil akurasi model yaitu 70.78%.

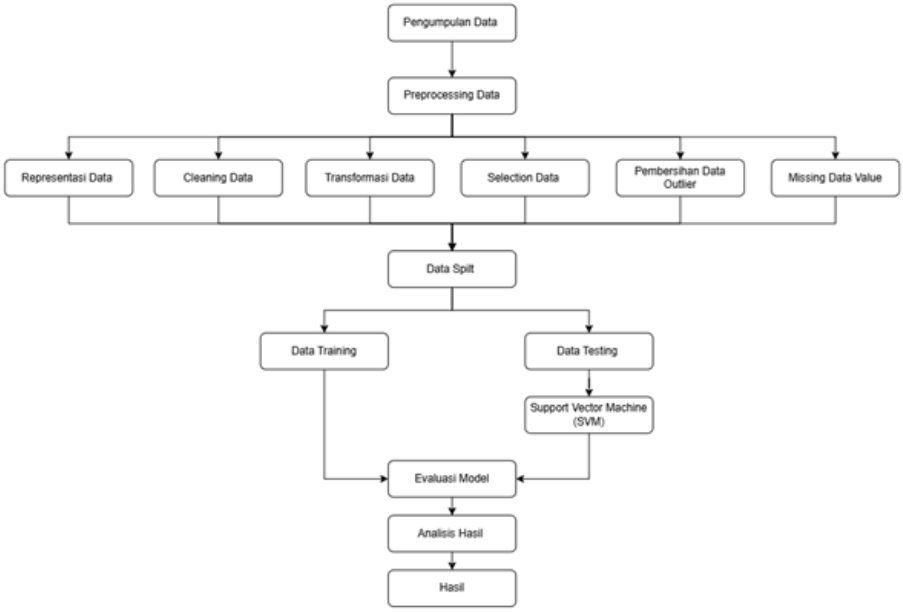
Algoritma Support Vector Machine (SVM) mampu menangani data yang tidak linear dengan baik sehingga SVM dipilih sebagai metode klasifikasi yang menjadi karakteristik data medis termasuk data yang digunakan dalam diagnosis pada penyakit diabetes melitus (Hovi et al., 2022; Novianti & Ulinnuha, 2024; Putri & Alit, 2025). Pada SVM memiliki nilai yang tinggi untuk memprediksi klasifikasi pada data. Algoritma tersebut memiliki 2 metode yaitu Klasifikasi (Suppor Vector Machine) dan Regresi (Support Vector Regression) (Hovi et al., 2022).

Metode tradisional dalam diagnosis diabetes sering kali bergantung pada pengukuran kadar glukosa darah dan analisis gejala klinis. Namun, dengan kemajuan teknologi dan data yang tersedia, pendekatan berbasis data seperti pembelajaran mesin dapat memberikan solusi yang lebih akurat dan efisien. Salah satu algoritma yang menjanjikan dalam hal ini adalah support vector machine (SVM), yang dikenal karena kemampuannya dalam menangani data yang kompleks dan tidak seimbang. Algoritma Support Vector Machine dengan kernel Radial Basis Function (RBF) adalah teknik

klasifikasi yang efektif dalam machine learning, yang digunakan untuk memisahkan data dengan cara non linear. Kernel RBF memungkinkan SVM untuk memetakan input ke dalam ruang dimensi yang lebih tinggi, dimana pemisahan antara kelas – kelas dapat digunakan dengan lebih efisien. Klasifikasi merupakan pengelompokkan data menjadi beberapa kelas, yang sebelumnya sudah menentukan label dan target (Cahyani & Basuki, 2019) SVM telah digunakan secara luas dalam bidang kedokteran terutama dalam klasifikasi serta mampu menangani data yang tidak linear serta kontribusi yang signifikan untuk pengembangan sistem diagnosis yang lebih efisien dan canggih. Pemeriksaan klinis menggunakan alat glucometer untuk memeriksa kadar gula pada pasien penderita diabetes (Aldama & Nasir, 2023).

Kebaruan penelitian ini terletak pada penerapan SVM dengan dua jenis kernel non-linear, yaitu Polinomial dan Radial Basis Function (RBF), yang belum banyak dieksplorasi dalam penelitian sebelumnya terkait diabetes. Selain itu, penelitian ini menggunakan dataset skala besar (100.000 entri) yang lebih representatif dibandingkan studi sebelumnya, serta menerapkan tahapan pra-pemrosesan data yang lebih ketat, termasuk penanganan outlier dan standarisasi fitur. Tujuan utama penelitian adalah mengembangkan model prediktif yang mampu mengklasifikasikan tingkat keparahan diabetes dengan akurasi tinggi, sekaligus membandingkan performa kernel Polinomial dan RBF. Hasil penelitian ini diharapkan dapat menjadi solusi pendukung diagnosis klinis yang lebih akurat, memberikan kontribusi dalam pengembangan teknologi kesehatan berbasis AI, serta meningkatkan kesadaran masyarakat akan pentingnya deteksi dini diabetes untuk mencegah komplikasi lebih lanjut.

METODE PENELITIAN



Gambar 1 Tahapan Penelitian
Sumber: Analisis penulis, 2024

Pengumpulan Data

Dalam tahapan pengumpulan data, digunakan sebuah dataset diabetes melitus yang berjumlah 100.000 data diabetes melitus. Dataset ini memiliki 9 kolom utama yaitu gender, age, hypertansion, heart_disease, smoking_history, bmi, HbA1c_level_blood, blood_glucose, serta terdapat kolom label diabetes yang memiliki kategori “0 = Negatif” dan “1 = Positif”

gender	age	hypertension	heart_disease	smoking_history	bmi	HbA1c_level	blood_glucose_level	diabetes
Male	60.0	1	0	never	26.3	6.0	140	1
Female	80.0	0	0	never	27.56	5.0	140	0
Male	20.0	0	0	never	23.29	4.5	200	0
Male	44.0	0	0	never	27.32	3.5	140	0
Female	58.0	1	0	not current	42.34	6.5	260	1
Female	60.0	0	0	never	36.46	4.8	140	0
Female	37.0	0	0	never	29.45	4.0	90	0
Female	26.0	0	0	never	27.32	4.8	90	0
Female	23.0	0	0	never	27.75	5.0	158	0
Female	54.0	0	0	former	43.75	6.5	240	1
Female	74.0	0	0	former	27.2	4.5	85	0
Female	56.0	0	0	former	30.42	5.8	160	0
Male	38.0	0	0	never	26.67	5.7	145	1
Female	1.64	0	0	No Info	16.37	6.6	140	0
Female	51.0	0	0	never	24.76	9.0	300	1
Female	3.0	0	0	No Info	15.58	6.1	159	0
Female	79.0	1	0	former	31.55	7.0	300	1
Male	5.0	0	0	No Info	19.05	4.0	85	0
Female	63.0	0	0	former	20.01	6.1	159	0
Female	58.0	0	0	No Info	23.48	6.5	80	0
Female	44.0	0	0	former	23.49	4.5	80	0
Female	40.0	0	0	No Info	20.21	4.0	200	0

Gambar 1. Pengumpulan Data

Sumber: Proses akuisisi data dari Diabetes Health Indicators Dataset, CDC Behavioral Risk Factor Surveillance System (BRFSS) 2023

Dataset yang telah disiapkan ini, akan melalui tahapan-tahapan pre-processing dengan menggunakan metode IQR pada deteksi outlier yang dapat mempengaruhi hasil analisis data.

Pre-Processing

Proses pengolahan data atau pre-processing bertujuan untuk memberishkan, mengubah, mnegurangi dan mempersiapkan data agar dapat digunakan. Dalam tahap preprocessing ini diawali dengan standarisasi data, deteksi outlier, dan membersihkan outlier.

Penanganan duplikasi data merupakan tahap pertama dan paling krusial dalam pra-pemrosesan data. Proses ini memastikan keunikan setiap baris data dalam kumpulan data dengan menghapus entri yang identik. Mengapa demikian? Keberadaan data duplikat dapat secara signifikan menyebabkan bias pada model *machine learning*. Model akan cenderung mempelajari informasi yang berulang, yang pada akhirnya menghasilkan evaluasi kinerja yang terlalu optimis pada data pelatihan, namun justru buruk saat diterapkan pada data baru. Selain mencegah bias, eliminasi duplikasi data juga berkontribusi pada pengurangan ukuran data. Hal ini secara langsung dapat mempercepat proses pelatihan model dan mengurangi kebutuhan sumber daya komputasi. Dengan menghilangkan duplikasi, model Anda akan belajar dari representasi data yang akurat dan beragam, sehingga meningkatkan kemampuan generalisasi (generalizability) dan ketahanan (robustness) model saat berhadapan dengan data dunia nyata.

Tahap selanjutnya, menerapkan label encoding pada kolom-kolom kategorikal seperti ‘gender’, dan ‘smoking_history’. Transformasi ini penting karena Sebagian besar algoritma machine learning tidak dapat memproses data dalam format teks. Oleh karena itu, data tekstual perlu diubah menjadi representasi numerik. Melalui label encoding, setiap kategori unik dalam suatu kolom akan ditetapkan nilai nbilangan bulat yang berbeda. Sebagai ilustrasi, kategori ‘Male’ dapat diwakili oleh 0 dan ‘Female’ oleh 1. Proses ini memungkinkan algoritma untuk memahami dan mengolah informasi kategorikal, mengubah data tekstual menjadi format yang siap untuk dianalisis.

Tahap penting berikutnya dalam pemprosesan data ialah penanganan outlier, khususnya pada fitur numerik seperti ‘age’, ‘bmi’, ‘HbA1c_level’ dan ‘blood_glucose_level’. Nilai ektrem atau outlier ini berpotensi mendistorsi statistic data dan mempengaruhi kinerja model secara negative. Untuk

mengatasi masalah tersebut, digunakan metode capping. Proses capping dimulai dengan perhitungan kuartil pertama (Q1 – 1.5 x IQR) dan batas atas (Q3 + 1.5 x IQR). Setiap nilai data yang berada di luar rentang batas tersebut akan "dibatas" atau diganti dengan nilai batas terdekatnya (misalnya, nilai di bawah batas bawah akan diubah menjadi batas bawah, dan nilai di atas batas atas akan diubah menjadi batas atas). Pendekatan ini secara efektif mengurangi dampak ekstem dari outlier tanpa perlu mengahpus seluruh baris data, sehingga menjaga integritas dataset sekaligus meningkatkan stabilitas dan robustness model.

Tahapan terakhirnya dalam preprocessing data ialah penerapan Standard Scalling pada fitur-fitur numerik yang telah ditangani outlier-nya. Penskalaan ini sangat esensial, terutama bagi algoritma machine learning yang sensitif terhadap skala fitur, seperti Support Vector Machine (SVM) atau jaringan saraf. Proses standard scalling akan mentransformasikan distribusi fitur numerik sehingga setiap fitur memiliki rata-rata nol (0) dan deviasi standar satu (1). Dengan demikian, metode yang dipakai ini memastikan bahwa tidak ada satu fitur pun yang mendominasi perhitungan model hanya karena rentang nilainya yang lebih besar. Standard scalling menjamin bahwa semua fitur berkontribusi secara adil terhadap proses pembelajran, membantu model untuk konvergen lebih cepat dan mencapai kinerja yang lebih optimal serta meningkatkan stabilitas numerik dari proses optimasi model

Peneltiian ini menggunakan algoritma klasifikasi Support Vector Machine (SVM). SVM merupakan algoritma machine learning yang sangat efektif dan serbaguna, pada penelitian ini secara khusus untuk tugas klasifikasi. Tujuan utamanya adalah untuk memprediksi apakah seseorang mengidap diabetes atau tidak, yang merupakan contoh klasik dari masalah klasifikasi biner. Secara konseptual, SVM bekerja dengan menemukan hyperplane (bidang pemisah) optimal yang mampu memisahkan kelas-kelas data dengan margin sebesar mungkin. Margin ini adalah jarak antara hyperplane dengan titik data terdekat dari setiap kelas, yang dikenal sebagai support vectors. Semakin besar margin yang ditemukan, semakin baik kemampuan generalisasi model terhadap data baru. dalam peneltiian ini dilakukan pengujian dengan beberapa jenis kernel SVM, yaitu kernel linear,Radial Basis Function (RBF), dan Polinomial memungkinakn model untuk menemukan batas Keputusan yang lebih kompleks di ruang berdimensi tinggi.

Dalam algoritma Support Vector Machine (SVM), fungsi kernel memainkan peran klusial dalam memungkinkan model menemukan batas pemisah yang efektif , terutama Ketika data tidak adapt dipisahkan secara linear. kernel linear adalah kernel yang memiliki kecocokan Ketika dataset yang digunkaan dapat dipisahkan dengan garis lurus atau bida datar. Kernel ini bekerja dengan menghitung produk titik anatara dua vector fitur, menawarkan komputasi yang cepat dan interpretasi yang mudah,

$$\mathbf{K(x, xi) = sum(x * xi)}$$

Berbeda dengan kernel linear, kernel Radial Basis Function(RBF), atau sering disebut Gaussian, merupakan salah satu kernel non-linear yang paling popular dan serbaguna. Kernel ini memiliki keunggulan signifikan dalam menangani *dataset* di mana kelas-kelas data tidak dapat dipisahkan secara sederhana oleh sebuah garis atau bidang lurus dalam ruang fitur aslinya. Dengan secara implisit memetakan data ke ruang dimensi yang lebih tinggi, Kernel RBF memungkinkan SVM menemukan batas keputusan yang kompleks dan non-linier. Kernel ini dikenal memberikan performa klasifikasi yang robust, seringkali menghasilkan nilai *error* yang rendah pada data pelatihan dan pengujian bila disetel dengan parameter yang optimal, menjadikannya pilihan yang seringkali mengungguli fungsi kernel lainnya dalam skenario data non-linier. Persamaan matematis untuk fungsi Kernel RBF adalah:

$$\mathbf{K(x,xi) = exp(-gamma * sum((x – xi^2))}$$

(gamma) adalah parameter kunci yang mengontrol lebar "pengaruh" dari setiap *support vector* dan sensitivitas model terhadap jarak antar data. Penyetelan nilai γ yang tepat sangat krusial untuk mencegah *overfitting* atau *underfitting* dan mencapai kinerja model yang optimal.

Selanjutnya dengan kernel Polinomial Fungsi kernel polinomial merupakan salah satu fungsi kernel yang esensial dalam algoritma *Support Vector Machine* (SVM) dan metode kernelisasi lainnya, khususnya ketika data tidak dapat dipisahkan secara linier. Sebagai generalisasi dari kernel linier, fungsi ini mengevaluasi kesamaan antara vektor sampel pelatihan dalam ruang fitur. Kernel polinomial juga terbukti efektif dalam mengatasi tantangan klasifikasi pada *dataset* pelatihan yang telah dinormalisasi. Persamaan matematis untuk fungsi kernel polinomial adalah:

$$\mathbf{K(x,xi) = 1 + sum(x * xi)^d}$$

Fungsi kernel ini memiliki parameter **derajat (d)** yang berperan krusial dalam menemukan performa optimal untuk *dataset* tertentu, dengan nilai baku d=2. Perlu diperhatikan bahwa peningkatan nilai parameter d yang berlebihan dapat menyebabkan fluktuasi akurasi dan ketidakstabilan sistem. Fenomena ini terjadi karena derajat yang lebih tinggi menghasilkan garis *hyperplane* yang semakin kompleks dan melengkung, berpotensi mengurangi kemampuan generalisasi model.

HASIL DAN PEMBAHASAN

Pada penelitian ini pengumpulan data yang digunakan merupakan data diabetes melitus dengan jumlah data sebanyak 100.000 data.

	gender	age	hypertension	heart_disease	smoking_history	bmi	HbA1c_level	blood_glucose_level	diabetes
0	Female	80.0	0	1	never	25.19	6.6	140	0
1	Female	54.0	0	0	No Info	27.32	6.6	80	0
2	Male	28.0	0	0	never	27.32	5.7	158	0
3	Female	36.0	0	0	current	23.45	5.0	155	0
4	Male	76.0	1	1	current	20.14	4.8	155	0
...	...	...	...	...	...	...	...	...	...
99995	Female	80.0	0	0	No Info	27.32	6.2	90	0
99996	Female	2.0	0	0	No Info	17.37	6.5	100	0
99997	Male	66.0	0	0	former	27.83	5.7	155	0
99998	Female	24.0	0	0	never	35.42	4.0	100	0
99999	Female	57.0	0	0	current	22.43	6.6	90	0

100000 rows × 9 columns

Gambar 2. Dataset
Sumber: Dokumentasi Penelitian, 2024

Tampilan Gambar 3 berisikan data yang terdiri dari data ‘gender’, ‘age’, ‘hypertansion’, ‘heart_disease’, ‘smoking_history’, ‘bmi’, ‘HbA1c_level’, ‘blood_glucose_level’ dan ‘diabetes’.

```
--- Perbandingan Sebelum dan Sesudah Label Encoding ---
  gender_original smoking_history_original  gender_encoded  \
0          Female                    never             0.0
1          Female                 No Info             0.0
2            Male                    never             1.0
3          Female                  current             0.0
4            Male                  current             1.0

   smoking_history_encoded
0                      4.0
1                      0.0
2                      4.0
3                      1.0
4                      1.0
```

Gambar 3. Label Encoding
Sumber: Dokumentasi Eksperimen, 2024

Pada Gambar 4 menampilkan perbandingan anatar sebelum dan sesudah fungsi dari Encoding Label yang berfungsi untuk merubah data kategorikal menjadi bentuk numerik. Pada Gambar diatas, pada kolom gender, smoking_history, adalah data kategorikal yang diubah menjadi data numerik dengan menggunakan LabelEncoder.

```
--- Perbandingan Sebelum dan Sesudah Capping Outliers ---
   age_original  bmi_original  HbA1c_level_original  \
0          80.0         25.19                   6.6
1          54.0         27.32                   6.6
2          28.0         27.32                   5.7
3          36.0         23.45                   5.0
4          76.0         20.14                   4.8

   blood_glucose_level_original  age_capped  bmi_capped  HbA1c_level_capped  \
0                           140        80.0       25.19                 6.6
1                            80        54.0       27.32                 6.6
2                           158        28.0       27.32                 5.7
3                           155        36.0       23.45                 5.0
4                           155        76.0       20.14                 4.8

   blood_glucose_level_capped
0                       140.0
1                        80.0
2                       158.0
3                       155.0
4                       155.0


Statistik Setelah Capping (Min/Max):
       age    bmi  HbA1c_level  blood_glucose_level
min   0.08  13.71          3.5                 80.0
max  80.00  39.55          8.3                247.5
```

Gambar 4. Capping Outlier
Sumber: Dokumentasi penelitian, 2024

Setelah dilakukannya tahapan label encoding, selanjtunya pada Gambar 5 dilakukannya tahapan Capping Outlier untuk mengelola nilai-nilai ekstrem (atau outlier) pada kolom-kolom numerik seperti ‘age’, ‘bmi’, ‘HbA1c_level’, ‘dan ‘blood_glucose_level’. Pada tahapan ini menerapkan juga fungsi cap_outlier yang menggunakan metode IQR (Interquratile Range). Pada Gambar 5 juga dapat dilihat sebelum dan sesudah dilakukannya Capping Outlier

```
--- Perbandingan Sebelum dan Sesudah Standard Scaling ---
   age_pre_scaled  bmi_pre_scaled  HbA1c_level_pre_scaled  \
0            80.0           25.19                     6.6
1            54.0           27.32                     6.6
2            28.0           27.32                     5.7
3            36.0           23.45                     5.0
4            76.0           20.14                     4.8

   blood_glucose_level_pre_scaled  age_scaled  bmi_scaled  HbA1c_level_scaled  \
0                           140.0    1.700840   -0.310970            1.024846
1                            80.0    0.543372    0.048828            1.024846
2                           158.0   -0.614096    0.048828            0.167291
3                           155.0   -0.257952   -0.604890           -0.499697
4                           155.0    1.522768   -1.164014           -0.690265

   blood_glucose_level_scaled
0                    0.064090
1                   -1.488726
2                    0.529935
3                    0.452295
4                    0.452295


Statistik Setelah Scaling (Mean/Std Dev):
                age           bmi   HbA1c_level  blood_glucose_level
mean  3.400992e-16  8.004377e-16  4.868695e-16        -2.997484e-16
std   1.000005e+00  1.000005e+00  1.000005e+00         1.000005e+00
```

Gambar 5. Standarisasi Data
Sumber: Dokumentasi Pemrosesan Data, 2024

Pada Gambar 6 merupakan hasil setelah dilakukannya proses standarisasi data, yang Dimana pada gambar tersebut di dalam kolom age, bmi, HbA1c_level, dan blood_glucose_level. Dengan hasil ini secara menunjukkan bahwa proses standard scalling telah berhasil diterapkan, semua fitur numerik ini sekarang memiliki rata-rata mendekati nol dan devisiasi standar mendekati satu.

Selanjutnya pada penelitian ini, di tahapan pengujian ada tiga scenario pengujian yaitu, scenario pertama dengan membagi data menjadi 60% data latih dan 40% data uji, skenario kedua dengan 70% data latih dan 30% data uji, serta skenario ketiga dengan 80% data latih dan 20% data uji. Ketiga skenario inidirancang untuk mengamati stabilitas kinerja model pada berbagai pembagian data. Dengan membagi skenario pengujian, dapat diidentifikasi model yang memiliki kinerja terbaik dan memberikan prediksi atau hasil yang akurat.

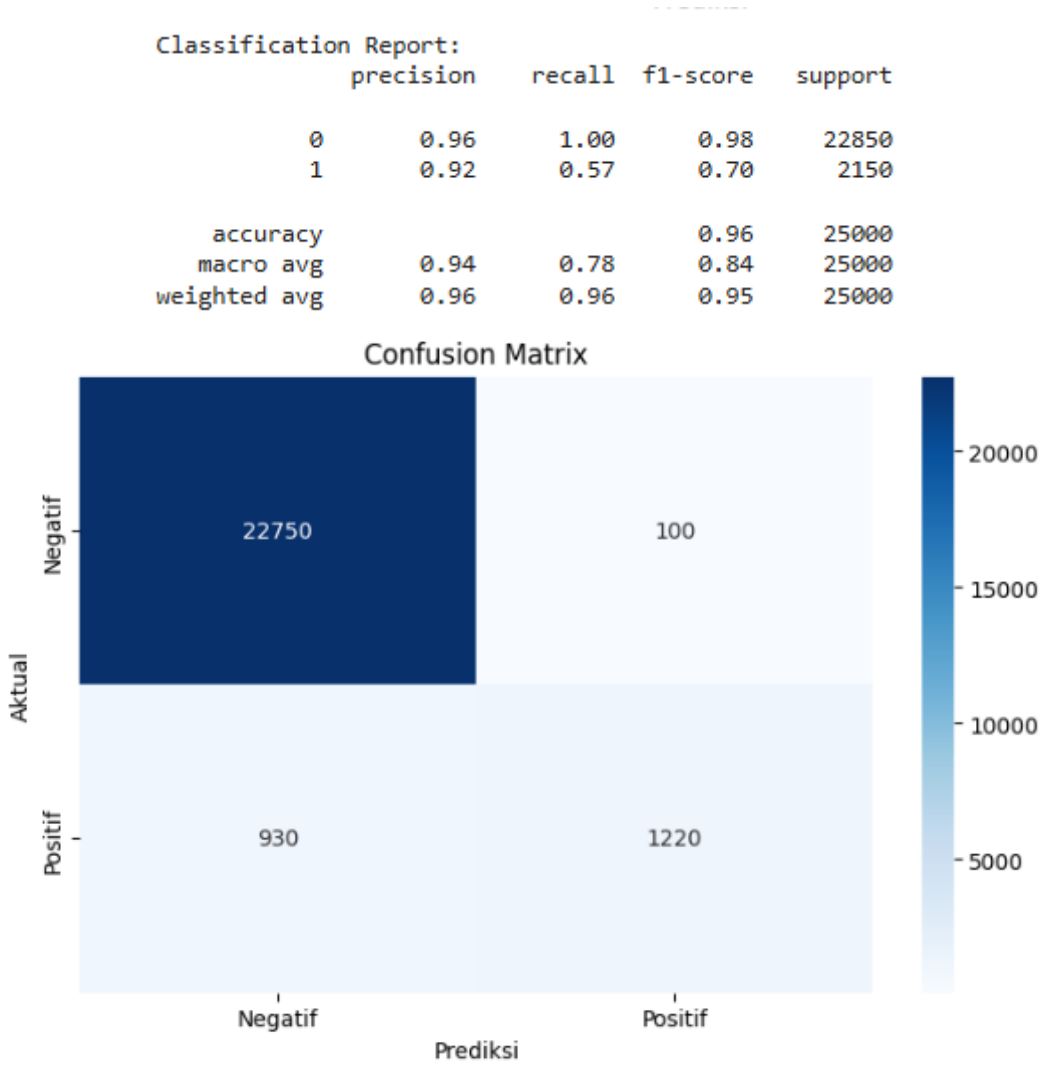
Tabel 1. Splitting Data

Rasio	Data Latih	Data Uji	Total Data
60% : 40%	60.000	40.000	100.000
70% : 30%	70.000	30.000	
80% : 20%	80.000	20.000	

Sumber: Pengolahan data penulis, 2024

Pada pengujian yang telah dilakuakan, pada penelitian ini menerapkan pengujian dengan skneario 70% data latih dan 30% data uji.

Hasil penerapan algoritma SVM dengan kernel Linear dapat dilihat melalui tampilan Confussion Matrix pada gambar berikut yang menggambarkan performa model dalam mengklasifikasi data.

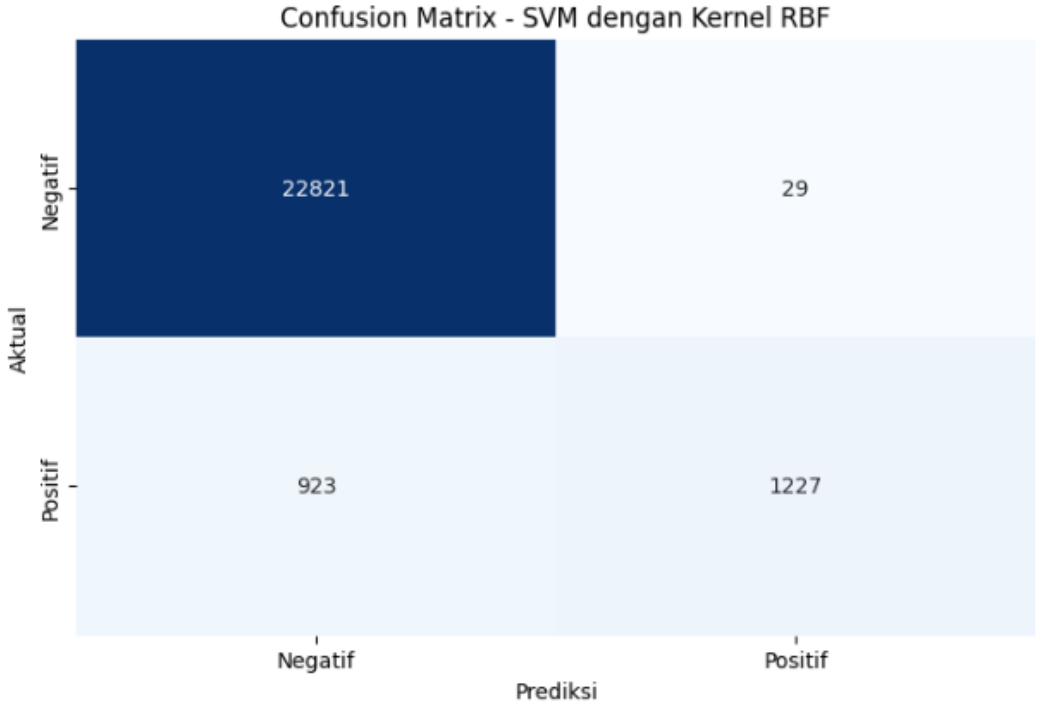


Gambar 6. SVM Linear
Sumber: Hasil Eksperimen Penulis, 2024

Pada pengujian model Support Vector Machine (SVM) dengan kernel linear untuk klasifikasi data diabetes, diperoleh hasil evaluasi yang menunjukkan kinerja model cukup baik secara keseluruhan. Akurasi model mencapai 96%, yang mengindikasikan bahwa mayoritas prediksi sudah sesuai dengan label sebenarnya. Untuk kelas negatif (0), yaitu pasien yang tidak mengidap diabetes, model menunjukkan performa yang sangat baik dengan precision sebesar 96%, recall mencapai 100%, dan f1-score sebesar 98%. Hal ini berarti hampir seluruh pasien non-diabetes berhasil dikenali dengan benar oleh model tanpa kesalahan signifikan. Namun, performa untuk kelas positif (1), yaitu pasien yang terdeteksi mengidap diabetes, masih menunjukkan kelemahan. Precision-nya cukup tinggi yaitu 92%, tetapi recall-nya hanya sebesar 57%, F1-score untuk kelas positif hanya mencapai 70%.

Classification Report:

	precision	recall	f1-score	support
0	0.9611	0.9987	0.9796	22850
1	0.9769	0.5707	0.7205	2150
accuracy			0.9619	25000
macro avg	0.9690	0.7847	0.8500	25000
weighted avg	0.9625	0.9619	0.9573	25000

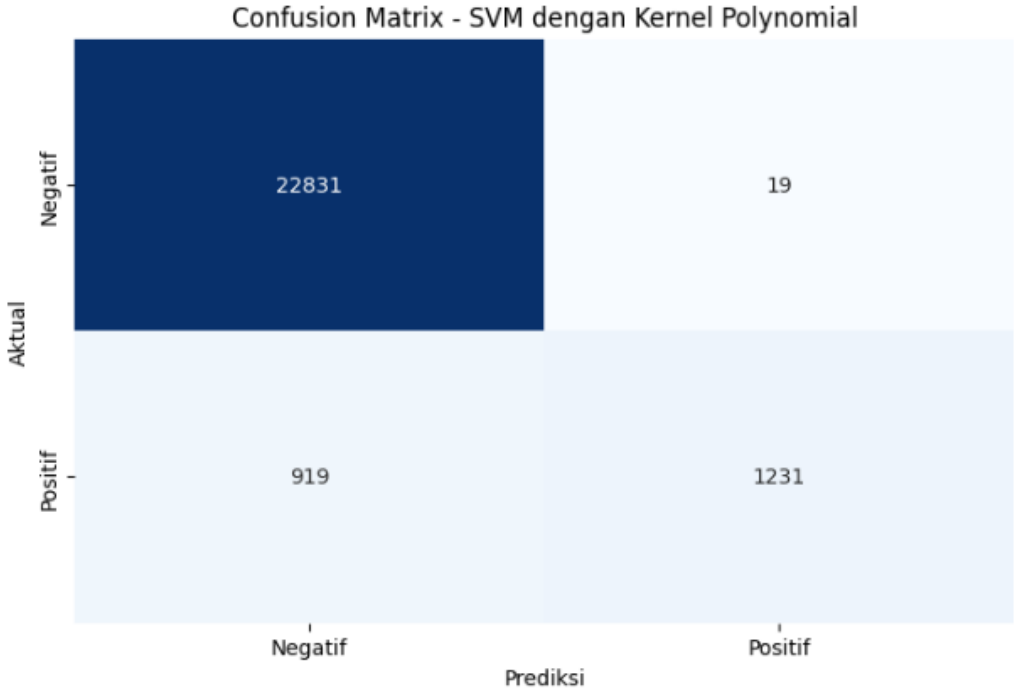


Gambar 7. SVM RBF
Sumber: Dokumentasi Pengujian Model, 2024

Pada pengujian Support Vector Machine Radial Basis Function (SVM RBF) ini menunjukkan bahwa untuk kelas negatif, precision sebesar 96,11%, dengan recall sangat tinggi sebesar 99,87% dan f1-score sebesar 97,96%. Sementara itu, untuk kelas positif, precision lebih tinggi yaitu 97,60%, namun recall hanya 57,07%, sehingga f1-score menjadi 72,05%. Akurasi keseluruhan model mencapai 96,19%, menunjukkan bahwa moel sangat baik dalam memprediksi secara umum.

Classification Report:

	precision	recall	f1-score	support
0	0.9613	0.9992	0.9799	22850
1	0.9848	0.5726	0.7241	2150
accuracy			0.9625	25000
macro avg	0.9731	0.7859	0.8520	25000
weighted avg	0.9633	0.9625	0.9579	25000



Gambar 8. SVM Polinomial
Sumber: Analisis Akhir Penulis, 2024

Pada pengujian Support Vector Machine (SVM) dengan kernel Polinomial dalam klasifikasi data diabetes ini menunjukkan untuk kelas negatif, memmiliki precision sebesar 96,13%, kemudian recall dengan hasil 99,92%, dan f1-score sebesar 97,99%. Untuk kelas positif, precision justru lebih tinggi aitu 98,48%, namun recall masih rendah yaitu 57,26%, menghasilkan f1-score sebesar 72,41%. Secara keseluruhan, akurasi model adalah 96,25% menunjukkan model cukup andal secara keseluruhan,

Adapun hasil akurasi dari setiap algoritma SVM Linear, SVM RBF, dan SVM Polinomial dapt dilihat pada tabel berikut ini.

Tabel 2. Hasil Akurasi

Model	Akurasi
SVM Linear	96%
SVM RBF	96.19%
SVM Polinomial	96.25%

Sumber: Hasil eksperimen penulis, 2024

Hasil evaluasi yang dilakukan menunjukkan bahwa model SVM dengan kernel Polinomial dalam data diabetes melitus, dengan akurasi sebesar 96.25%. Angka ini menunjukkan peningkatan signifikan dibandingkan dengan temuan pada jurnal sebelumnya (Pungky & Rosalya Putri, 2024), yang mendapatkan akurasi dibawah 80% untuk model SVM.

Keunggulan model SVM dengan kernel Polinomial jelas terlihat dari kemampuannya dalam memproses data non-linear, hal ini secara konsisten membuktikan bahwa penggunaan model SVM dengan kernel Polinomial memiliki potensi untuk meningkatkan kinerja klasifikasi secara subsntansial jika dibandingkan menggunakan metode SVM dengan kernel lainnya.

Namun, meskipun ketiga model tersebut memberikan akurasi yang lebih tinggi secara keseluruhan, model ini masih mengalami kesulitan dalam memproses data yang tidak seimbang.

KESIMPULAN

Berdasarkan penelitian yang telah dilakukan secara komprehensif, dapat disimpulkan bahwa pemodelan prediktif untuk penyakit diabetes melitus terbukti sangat efektif diimplementasikan menggunakan metode Support Vector Machine (SVM). Pengujian terhadap berbagai fungsi *kernel* pada SVM, khususnya Kernel Radial Basis Function (RBF) dan Polinomial, menunjukkan kapabilitas superior dalam mengidentifikasi pola kompleks dalam data pasien untuk klasifikasi diabetes. Secara spesifik, hasil penelitian ini secara meyakinkan mengindikasikan bahwa metode SVM dengan kernel Polinomial berhasil mencapai performa tertinggi, dengan akurasi klasifikasi sebesar 96.25%. Akurasi yang sangat tinggi ini tidak hanya menegaskan efektivitas model dalam memprediksi status diabetes, tetapi juga secara signifikan menyoroti keunggulan kernel Polinomial dalam menangani *dataset* ini, khususnya pada pola data yang tidak menunjukkan kompleksitas ekstrim atau linearitas murni, melainkan hubungan non-linear yang dapat dipisahkan secara efektif oleh batas polinomial.

Untuk penelitian selanjutnya, disarankan untuk mempertimbangkan penggunaan dataset yang memiliki jumlah kelas yang tidak seimbang, serta mengimplementasikan metode analisis dan klasifikasi yang lebih canggih guna meningkatkan akurasi.

DAFTAR PUSTAKA

Aldama, A., & Nasir, M. (2023). Klasifikasi diabetes menggunakan support vector machine dan optimasi hyperparameter. *Jurnal Teknologi Informasi, 15*(2), 45–58.

Bommer, C., Heesemann, E., Sagalova, V., Manne-Goehler, J., Atun, R., Bärnighausen, T., Davies, J., & Vollmer, S. (2018). Global economic burden of diabetes in adults: Projections from 2015 to 2030. *Diabetes Care, 41*(5), 963–970. https://doi.org/10.2337/dc17-1962

Cahyani, D., & Basuki, A. (2019). Implementasi algoritma klasifikasi untuk prediksi diabetes mellitus. *Jurnal Teknologi Informasi dan Komunikasi, 8*(1), 23–34.

Gunawan, R., Sugiarto, B., & Mardinto, P. (2020). Analisis faktor risiko diabetes mellitus menggunakan machine learning. *Jurnal Kesehatan Digital, 12*(3), 156–167.

International Diabetes Federation. (2025). *IDF Diabetes Atlas 11th edition: Diabetes facts and figures*. IDF. https://diabetesatlas.org/

Hovi, H. S. W., Hadiana, A. I., & Umbara, F. R. (2022). Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM). *Informatics and Digital Expert (INDEX)*, *4*(1), 40–45.

Hovi, M., Rahman, A., Sari, D., & Pratama, I. (2022). Perbandingan algoritma support vector machine untuk klasifikasi medis. *Jurnal Informatika Kesehatan, 9*(2), 89–102.

Lestari, S., Putri, A., Dewi, R., & Sari, M. (2021). Diabetes mellitus sebagai masalah kesehatan global: Tinjauan epidemiologi dan faktor risiko. *Jurnal Kesehatan Masyarakat, 17*(4), 234–245.

NCD Risk Factor Collaboration; World Health Organization. (2024). More than 800 million adults have diabetes globally, many untreated. *The Lancet*. https://www.reuters.com/world/more-than-800-million-adults-have-diabetes-globally-many-untreated-2024-07-10/

Novianti, F., & Ulinnuha, N. (2024). Seleksi fitur algoritma genetika dalam klasifikasi data rekam medis PCOS menggunakan SVM. *Networking Engineering Research Operation*, *9*(1), 9–20.

Ong, K. L., Tzoulaki, I., Mozaffarian, D., & Ezzati, M. (2023). Global, regional, and national burden of diabetes from 1990 to 2023: A systematic analysis. *The Lancet*. https://doi.org/10.1016/S0140-6736(23)01301-6

Pungky, A., & Rosalya Putri, R. (2024). Klasifikasi penyakit diabetes melitus menggunakan metode support vector machine (SVM). *Journal of Informatics and Computer Science, 6*(1), 112–125.

Purnama, I., Sari, L., Dewi, A., & Rahman, F. (2019). Komplikasi diabetes mellitus dan dampaknya terhadap kualitas hidup pasien. *Jurnal Endokrinologi Indonesia, 14*(2), 78–89.

Puspa, S. E. (2017). *Asuhan Keperawatan Klien Yang Mengalami Diabetes Mellitus Tipe 1 Dengan Masalah Ketidakseimbangan Nutris Kurang Dari Kebutuhan Tubuh Diruang Dahlia Rsud Jombang*. STIKes Insan Cendekia Medika Jombang.

Putri, P. R. P. R., & Alit, R. (2025). Klasifikasi Penyakit Diabetes Melitus Menggunakan Metode Support Vector Machine (SVM). *Journal of Informatics and Computer Science (JINACS)*, *6*(03), 740–746.

Seuring, T., Archangelidi, O., & Suhrcke, M. (2015). The economic costs of type 2 diabetes: A global systematic review. *The Lancet Diabetes & Endocrinology, 3*(6), 423–430. https://doi.org/10.1016/S2213-8587(15)00025-3

Yeni, N. (2018). *Pengaruh Pemberian Jus Nanas Terhadap Perubahan Kadar Asam Urat Di Desa Keniten Kecamatan Geneng Kabupaten Ngawi*. STIKES BHAKTI HUSADA MULIA.