

Pengembangan Algoritma Support Vector Machine (SVM) untuk Mengklasifikasi Penyakit Diabetes

Samsul Arifin¹, Raul Satria², Tarwoto³

¹Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Purwokerto, Indonesia

²Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Purwokerto, Indonesia

³Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Purwokerto, Indonesia

surel: ¹21sa1066@mhs.amikompurwokerto.ac.id, ²21sa1019@mhs.amikompurwokerto.ac.id, ³tarwoto@amikompurwokerto.ac.id

Info Artikel

Sejarah artikel:

Diterima 09-05-2024

Revisi 23-05-2024

Diterima 13-08-2024

Kata kunci:

Diabetes Mellitus

Support Vector Machine

Klasifikasi

Diagnosis

Algoritma Machine Learning

ABSTRAK

Diabetes Mellitus (DM) merupakan penyakit kronis yang ditandai dengan kadar gula darah tinggi dan berpotensi menimbulkan berbagai komplikasi serius, seperti serangan jantung, stroke, gagal ginjal, amputasi kaki, dan gangguan pada sistem saraf. Deteksi dini DM menjadi kunci untuk mencegah terjadinya komplikasi-komplikasi tersebut. Penelitian ini bertujuan mengembangkan algoritma Support Vector Machine (SVM) untuk mengklasifikasikan DM berdasarkan data klinis pasien. Algoritma SVM bekerja dengan cara memisahkan data menjadi dua kelompok dengan mencari garis hyperplane yang optimal untuk mendapatkan hasil akurasi yang tinggi. Algoritma SVM dilatih dan diuji menggunakan data sekunder dari rekam medis pasien DM dan non-DM yang diperoleh dari Kaggle.com. Data ini mencakup berbagai variabel klinis yang relevan, seperti kadar glukosa darah, tekanan darah, indeks massa tubuh, dan lain-lain. Hasil dari penelitian menunjukkan bahwa algoritma SVM menghasilkan performa terbaik dengan akurasi 77,27%, dibandingkan dengan penelitian sebelumnya yang hanya mencapai akurasi 65%. Tentunya, ini menunjukkan potensi algoritma SVM sebagai alat bantu diagnosis DM yang lebih akurat dan efisien. Keberhasilan algoritma SVM dalam meningkatkan akurasi diagnosis DM sangat penting, mengingat deteksi dini dan pengelolaan yang tepat dapat mengurangi risiko komplikasi serius pada pasien. Selain itu, algoritma ini juga dapat diadaptasi dan ditingkatkan lebih lanjut untuk mencapai akurasi yang lebih tinggi, misalnya dengan menggunakan teknik pra-pemrosesan data yang lebih canggih atau dengan menggabungkan SVM dengan metode machine learning lainnya. Implementasi algoritma SVM yang lebih baik dapat membantu tenaga medis dalam pengambilan keputusan klinis yang lebih cepat dan tepat. pengembangan teknologi seperti SVM untuk deteksi dini DM dapat berdampak signifikan terhadap kesehatan masyarakat secara keseluruhan. Dengan diagnosis yang lebih cepat dan tepat, pasien dapat mendapatkan perawatan yang sesuai lebih awal, sehingga mengurangi beban penyakit dan meningkatkan kualitas hidup mereka.

Penulis yang sesuai:

Samsul Arifin

Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Purwokerto

Email: samsulaerifin@gmail.com

1. PENDAHULUAN

Di zaman sekarang banyak orang yang bekerja keras untuk memenuhi kebutuhan sehari-hari tanpa memikirkan kesehatan tubuh. Dengan hal tersebut di perburuk lagi dengan jarang berolahraga, makan dan minum instan dan mengonsumsi gula berlebihan. Pola hidup yang tidak sehat seperti ini rentan terkena penyakit Diabetes Mellitus (DM) [1]. Diabetes Mellitus (DM) merupakan penyakit kronis yang ditandai dengan kadar gula darah tinggi dalam darah dikarenakan metabolisme tubuh yang tidak berfungsi dengan baik sehingga tubuh tidak bisa menghasilkan insulin sesuai dengan kebutuhan tubuh [2].

Penyakit diabetes merupakan penyakit yang berbahaya yang menyebabkan berbagai komplikasi [3]. Komplikasi yang sering di alami penderita diabetes adalah serangan jantung, stroke, gagal ginjal, amputasi kak dan gangguan pada sistem syaraf [4]. Deteksi dini DM menjadi kunci untuk mencegah komplikasi tersebut. Apabila didiagnosa lebih awal maka penanganan bisa di lakukan lebih cepat dan komplikasi yang membahayakan bisa dihindari [5].

Dengan perkembangan teknologi sekarang yang begitu pesat diperlukannya sistem yang bisa mempercepat pendeteksian penyakit diabetes, sistem yang dapat mengetahui dengan cepat agar komplikasi bisa di hindari lebih awal, yaitu dengan menerapkan data *mining* untuk mengklarifikasi dengan memanfaatkan *machine learning* dengan algoritma *Support Vector Machine*. *machine learning* mesin yang dikembangkan agar bisa belajar dengan otomatis yang di kembangkan berdasarkan ilmu statistika, matematika, data mining [6]. *Support Vector Machine* (SVM), menawarkan solusi yang menjanjikan untuk meningkatkan akurasi dan efisiensi diagnosis DM. *Support Vector Machine* SVM adalah algoritma klasifikasi yang bekerja dengan mencari hyperplane yang memisahkan data menjadi dua kelas [7] atau menurut . Penelitian yang di lakukan oleh [8] dilakukan klarifikasi menggunakan *support vector machine* yang menghasilkan nilai rata – rata akurasi sebesar 65%

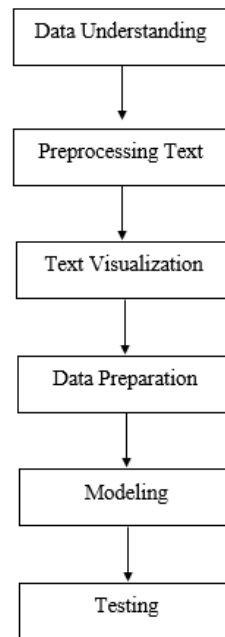
Penelitian ini bertujuan untuk mengembangkan algoritma SVM yang akurat dan efisien untuk mengklasifikasikan DM berdasarkan data klinis pasien. Algoritma SVM dilatih dan diuji menggunakan data sekunder dari rekam medis pasien DM dan non-DM di rumah sakit. Penelitian ini diharapkan dapat memberikan kontribusi untuk meningkatkan akurasi dan efisiensi diagnosis DM.

Implementasi algoritma SVM yang lebih baik tidak hanya bermanfaat bagi tenaga medis dalam pengambilan keputusan klinis yang lebih cepat dan tepat, tetapi juga dapat menjadi alat bantu yang penting dalam praktik kesehatan sehari-hari. Algoritma ini dapat menyediakan analisis data yang lebih mendalam dan memberikan rekomendasi berdasarkan pola-pola data yang mungkin tidak terlihat secara langsung oleh manusia. Dengan begitu, algoritma ini dapat berfungsi sebagai partner dalam memberikan pelayanan kesehatan yang lebih baik kepada pasien. Dalam jangka panjang, pengembangan teknologi seperti SVM untuk deteksi dini DM dapat berdampak signifikan terhadap kesehatan masyarakat secara keseluruhan. Dengan diagnosis yang lebih cepat dan tepat, pasien dapat mendapatkan perawatan yang sesuai lebih awal, sehingga mengurangi beban penyakit dan meningkatkan kualitas hidup mereka.

2. METODE

Pada penelitian ini memiliki tahapan penyusunan dalam merancang sistem mengklarifikasi penyakit diabetes. Seperti yang dapat di lihat pada gambar 1.





Gambar 1. Metode Penelitian

3. HASIL DAN PEMBAHASAN

Support Vector Machine (SVM) adalah salah satu algoritma yang sangat efektif untuk klasifikasi, terutama dalam kasus dengan dimensi fitur yang tinggi. Untuk mengembangkan algoritma SVM yang mengklasifikasikan penyakit diabetes, menggunakan dataset pasien Diabetes yang diperoleh dari sumber public. Dataset ini memiliki beberapa fitur medis pasien yang dapat digunakan untuk menentukan apakah pasien menderita diabetes atau tidak.

3.1. Dataset

Dataset yang kami peroleh mengandung informasi medis dari pasien yang memuat Fitur-fitur dalam dataset meliputi:

No	Fitur	Keterangan
1	Pregnancies	Jumlah kehamilan
2	Glucose	Kadar glukosa
3	BloodPressure	Tekanan darah diastolik (mm Hg)
4	SkinThickness	Ketebalan kulit triceps (mm)
5	Insulin	Kadar insulin serum 2 jam (mu U/ml)
6	BMI	Indeks massa tubuh (berat dalam kg/(tinggi dalam m) ²)
7	DiabetesPedigreeFunction	Fungsi silsilah diabetes (probabilitas genetik)
8	Age	Usia (tahun)
9	Outcome/ Hasil	Variabel target (0 jika tidak diabetes, 1 jika diabetes)

3.2. Metode Data Understanding

Tahap awal penelitian yang bertujuan memahami data yang akan di gunakan untuk menyelesaikan masalah. Dalam penelitian ini ada beberapa aktivitas yang pertama pengumpulan data di mana data dalam penelitian ini di ambil dari Kaggle.com, yang kedua mengidentifikasi dimana kita melihat jenis data yang kita peroleh misal tipe data dan atribut data yang kita miliki misal berapa umur tertinggi atau umur terendah.



	Kehamilan	Glukosa	Tekanan Darah	Ketebalan Kulit	Insulin	BMI	DiabetesPedigreeFunction	Umur	Hasil
0	0	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	0	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1
...
763	10	101	76	48	180	32.9	0.171	63	0
764	2	122	70	27	0	36.8	0.340	27	0
765	5	121	72	23	112	26.2	0.245	30	0
766	1	126	60	0	0	30.1	0.349	47	1
767	1	93	70	31	0	30.4	0.315	23	0

768 rows x 9 columns

Gambar 2. Jumlah Data dan Fitur

3.3. Preprocessing Text

Dalam tahap ini sangat penting untuk di lakukan, karena data yang di peroleh tidak dalam kondisi baik untuk di proses, seperti membersihkan data kosong (*missing value*) dan data berulang (*redundant*) oleh karena itu di perlukan nya tahap Preprocessing, agar data yang bermasalah tidak mengganggu dalam proses klarifikasi data.

3.3.1. Missing value

missing value merupakan permasalahan pada bagian data terdapat data yang tidak lengkap atau hilang [9]. Hal ini bisa menghambat klarifikasi karena akan menjadi tidak efisien.

3.3.2. Redundant

Data di katakan *Redundant* jika dalam suatu data terdapat lebih dari satu record yang berisi nilai yang sama [10].

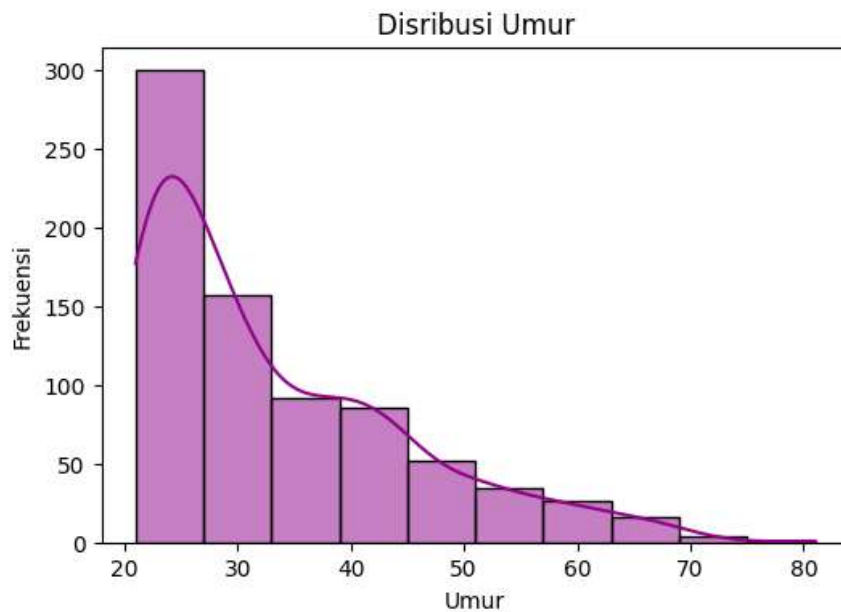
```
df.isnull().sum()
Kehamilan      0
Glukosa         0
Tekanan Darah   0
Ketebalan Kulit 0
Insulin         0
BMI            0
DiabetesPedigreeFunction 0
Umur           0
Hasil          0
dtype: int64

df.duplicated().sum()
0
```

Gambar 3. menghapus data yang duplicate atau ganda

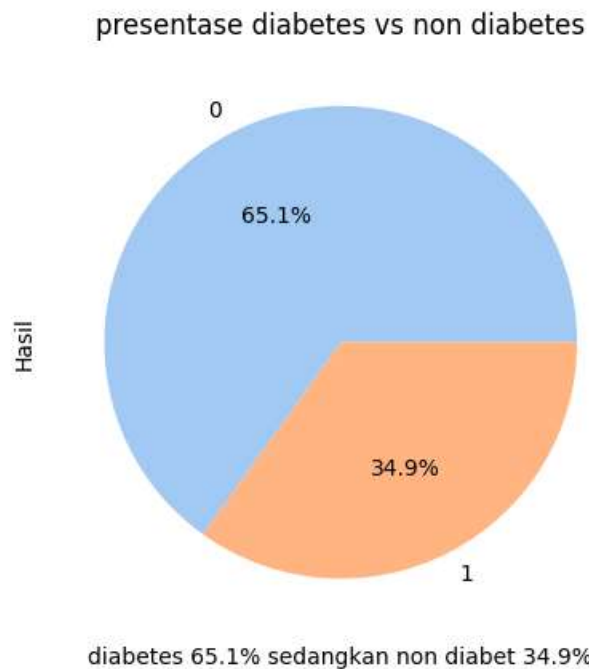
3.4. Text visualization

Dalam tahapan ini data di proyeksikan atau memvisualisasikan menjadi sebuah diagram di mana ini bertujuan agar memudahkan kita dalam memahami dan menganalisa data yang kita miliki.



Gambar 4. Distribusi Umur

Gambar di atas menunjukkan umur pasien yang lebih dari 20 tahun lebih banyak lebih dari 300 orang dan yang paling sedikit adalah pasien berumur 80 tahun.



Gambar 5. Perbandingan Presentase Diabetes dan Non diabetes

3.5. Data preparation

Tahapan preparation adalah melakukan persiapan sebelum dilakukan modeling [11]. Dimana pada tahap ini proses mengubah data mentah menjadi siap atau standarisasi untuk algoritma Algoritma Support Vector Machine (SVM).

3.6. Modeling

Pada penelitian ini di gunakan model algoritma Support Vector Machine (SVM) untuk melakukan proses klasifikasi. Klasifikasi adalah proses untuk menghasilkan fungsi menjelaskan dan membedakan konsep untuk memperkirakan kelas yang tidak di ketahui [12] Untuk tahapan ini memisahkan dua data menjadi 2 kelompok dengan mencari garis hyperplane untuk mendapatkan hasil akurasi.

	precision	recall	f1-score	support
0	0.80	0.86	0.83	99
1	0.71	0.62	0.66	55
accuracy			0.77	154
macro avg	0.76	0.74	0.74	154
weighted avg	0.77	0.77	0.77	154

Akurasi SVM: 77.27%

Best Parameters: {'C': 10, 'gamma': 0.001, 'kernel': 'rbf'}

Gambar 6. Hasil Akurasi Model SVM

Dari hasil modeling data di peroleh akurasi model yang terdiri data presisi, recall, fi score, dan support dengan akurasi model 77.27%. Akurasi sebesar 77.27% menunjukkan bahwa model dapat mengklasifikasikan 77.27 dari 100 kasus dengan benar. Meskipun ini adalah awal yang baik, ada ruang untuk perbaikan.

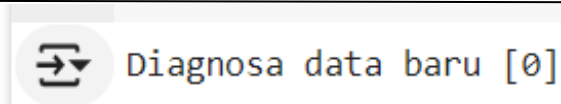
3.7. Testing

Testing tahap terakhir dalam penelitian ini di mana setelah memodelkan yaitu dengan menggunakan algoritma Algoritma Support Vector Machine (SVM), untuk testing sendiri memuat data baru dan di tentukan oleh modeling yang sudah di buat [13].

	Kehamilan	Glukosa	Tekanan Darah	Ketebalan Kulit	Insulin	BMI	DiabetesPedigreeFunction	Umur
0	1	85	66	29	0	26.6	0.351	31

Gambar 7. Data baru untuk pengujian

Setelah membuat data baru dimana membuat dua variabel yang pertama dimana kita melakukan transform terhadap data baru yang sudah di buat, lalu yang ke dua membuat hasil prediksi terhadap data pertama yang telah dibuat [14].



Gambar 8. Hasil prediksi data baru

Dari hasil diagnosis menunjukkan angka [0] berarti data tersebut di diagnosa tidak mengalami diabetes, tetapi jika hasilnya [1] maka data tersebut di diagnosa mengalami diabetes.

4. KESIMPULAN

Pengembangan algoritma Support Vector Machine (SVM) untuk mengklasifikasikan penyakit diabetes menunjukkan hasil yang cukup memuaskan dengan akurasi sebesar 77.27%. Metrik evaluasi lainnya, seperti presisi sebesar 76%, recall sebesar 74%, dan F1-score sebesar 0.74, juga menunjukkan kinerja model yang baik dalam mengidentifikasi pasien yang menderita diabetes.

Model SVM yang dikembangkan untuk klasifikasi penyakit diabetes memberikan hasil yang menjanjikan dan dapat digunakan sebagai alat bantu untuk mendeteksi diabetes. Namun, untuk mencapai kinerja yang optimal, perlu dilakukan pengembangan lebih lanjut melalui penyesuaian hyperparameter, pemilihan kernel yang tepat, dan teknik penyeimbangan data. Penggunaan SVM dalam konteks medis harus selalu disertai dengan validasi dan interpretasi hasil yang cermat oleh ahli kesehatan untuk memastikan keputusan yang diambil berdasarkan model ini adalah yang terbaik bagi pasien [15].

REFERENSI

- [1] M. D. Purbolaksono, M. Irvan Tantowi, A. Imam Hidayat, and A. Adiwijaya, "Perbandingan Support Vector Machine dan Modified Balanced Random Forest dalam Deteksi Pasien Penyakit Diabetes," *J. RESTI (Rekayasa Sist. dan Teknol. Informatika)*, vol. 5, no. 2, pp. 393–399, 2021, doi: 10.29207/resti.v5i2.3008.
- [2] I. M. Karo Karo and H. Hendriyana, "Klasifikasi Penderita Diabetes menggunakan Algoritma Machine Learning dan Z-Score," *J. Teknol. Terpadu*, vol. 8, no. 2, pp. 94–99, 2022, doi: 10.54914/jtt.v8i2.564.
- [3] Muhammad Hilmy Haidar Aly, "Klasifikasi Diabetes Menggunakan Algoritma Support Vector Machine Radial Basis Function," *J. Tek. Inform. dan Teknol. Inf.*, vol. 4, no. 1, pp. 28–38, 2024, doi: 10.55606/jutiti.v4i1.3420.
- [4] D. A. Agatsa, R. Rismala, and U. N. Wisesty, "Klasifikasi Pasien Pengidap Diabetes menggunakan Metode Support Vector Machine," *e-Proceeding Eng.*, vol. Vol.7, no. No.1, p. 2517, 2020.
- [5] J. Bonardo Junior, R. Rohmat Saedudin, and V. P. Widharta, "Perbandingan Akurasi Algoritma Decision Tree Dan Algoritma Support Vector Machine Pada Penyakit Diabetes," *e-Proceeding Eng.*, vol. 8, no. 5, pp. 9749–9756, 2021.
- [6] E. Lestari and W. I. Rahayu, "Prediksi Keganasan Kanker Payudara Dengan Pendekatan Machine Learning: Sytematic Literature Review," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 3, pp. 1967–1971, 2023.
- [7] Fatmawati and N. A. K. Rifai, "Klasifikasi Penyakit Diabetes Retinopati Menggunakan Support Vector Machine dengan Algoritma Grid Search Cross-validation," *J. Ris. Stat.*, pp. 79–86, 2023, doi: 10.29313/jrs.v3i1.1945.
- [8] C. Aldama and M. Nasir, "Klasifikasi penyakit Diabetes menggunakan metode support vector machine pada Rumah Sakit Umum Prabumulih," *J. Ilm. Betrik*, vol. 14, no. 02, pp. 376–383, 2023, [Online]. Available: <https://ejournal.pppmitpa.or.id/index.php/betrik/article/view/117>
- [9] M. R. A. Prasetya, A. M. Priyatno, and Nurhaeni, "Penanganan Imputasi Missing Values pada Data Time Series dengan Menggunakan Metode Data Mining," *J. Inf. dan Teknol.*, vol. 5, no. 2, pp. 52–62, 2023, doi: 10.37034/jidt.v5i2.324.
- [10] A. A. Sinica, "E Ć 1 , 2 , 3," vol. XX, no. X, 2018, doi: 10.16383/j.aas.2018.cxxxxxx.
- [11] A. Noviriandini, H. Hermanto, and Y. Yudhistira, "Klasifikasi Support Vector Machine Berbasis Particle Swarm Optimization Untuk Analisa Sentimen Pengguna Aplikasi Pedulilindungi," *JIKA (Jurnal Inform.)*, vol. 6, no. 1, p. 50, 2022, doi: 10.31000/jika.v6i1.5681.
- [12] B. Yogyakarta, "Naskah publikasi penelitian pemula," pp. 1–15, 2019.
- [13] N. Leli, F. Rakhmawati, and R. Widyasari, "Penerapan Metode Support Vector Machine (Svm) Untuk



- Klasifikasi Uang Kuliah Tunggal Di Universitas Islam Negeri Sumatera Utara,” *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 4, no. 2, pp. 903–915, 2023, doi: 10.46306/lb.v4i2.354.
- [14] M. Ichwan, I. A. Dewi, and Z. M. S, “Klasifikasi Support Vector Machine (SVM) Untuk Menentukan TingkatKemanisan Mangga Berdasarkan Fitur Warna,” *MIND J.*, vol. 3, no. 2, pp. 16–23, 2019, doi: 10.26760/mindjournal.v3i2.16-23.
- [15] D. Fierro, Iván; Pinto, Diego; Afanador, “No 主観的健康感を中心とした在宅高齢者における 健康関連指標に関する共分散構造分析Title,” vol. 28, no. August, pp. 1–43, 2014.