



Klasifikasi Teknologi Big Data dalam Diagnosa Pola Penyakit Diabetes dan Jantung menggunakan Algoritma *Support Vector Machine* (SVM) di Klinik Pratama Yakrija

Iik Sukira

STMIK Widuri, Jakarta, Indonesia

Email : iiksukira86@gmail.com

Article Info

Article history:

Received August 20, 2025

Revised August 25, 2025

Accepted August 27, 2025

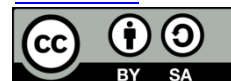
Keywords:

SVM Classification Model, Disease Pattern Diagnosis, Diabetes and Heart Disease, Model Evaluation.

ABSTRACT

Diabetes and heart disease are two common chronic conditions that require early diagnosis to prevent serious complications. At Yakrija Primary Clinic, the diagnostic process is still performed manually, often leading to inaccuracies and delays in decision-making. Therefore, an automated system is needed to improve diagnostic accuracy. This study aims to develop a classification model using the Support Vector Machine (SVM) algorithm to automatically diagnose patterns of diabetes and heart disease. The model is expected to enhance the accuracy and efficiency of the diagnostic process. This research employs the SVM method to classify patient data based on medical attributes such as Glucose, Blood Pressure, BMI, and age. The steps involved include data collection, pre-processing, normalization, training the SVM model, and evaluating model performance based on accuracy, precision, recall, and F1-Score using a Confusion Matrix. The developed SVM model showed an accuracy of 82.41% for diabetes, while for heart disease it achieved an accuracy of 95.24%. These results indicate an improvement in diagnostic accuracy compared to the manual method. Implementing the SVM model can provide significant enhancements in the accuracy and speed of diagnosing diabetes and heart disease. This model can be used as a support tool for medical decision-making, although further testing with more diverse data is necessary to improve generalization.

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



Article Info

Article history:

Received August 20, 2025

Revised August 25, 2025

Accepted August 27, 2025

Kata Kunci :

Model Klasifikasi SVM, Diagnosa Pola Penyakit, Diabetes dan Jantung, Evaluasi Model.

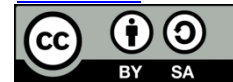
ABSTRAK

Penyakit diabetes dan jantung adalah dua kondisi kronis yang umum terjadi dan memerlukan diagnosa dini untuk mencegah komplikasi serius. Di Klinik Pratama Yakrija, proses diagnosa masih dilakukan secara manual, yang sering kali mengarah pada ketidakakuratan dan keterlambatan dalam pengambilan keputusan. Oleh karena itu, diperlukan sebuah sistem otomatis yang mampu meningkatkan akurasi diagnosa. Penelitian ini bertujuan untuk mengembangkan model klasifikasi menggunakan algoritma Support Vector Machine (SVM) untuk mendiagnosa pola penyakit diabetes dan jantung secara otomatis. Model ini diharapkan dapat meningkatkan akurasi dan efisiensi dalam proses diagnosa. Penelitian ini menggunakan metode SVM untuk mengklasifikasikan data pasien berdasarkan atribut-atribut medis seperti Glukosa, Tekanan Darah, BMI, dan usia. Langkah yang dilakukan meliputi pengumpulan data, pre-processing, normalisasi, training model SVM, dan evaluasi kinerja model berdasarkan akurasi, presisi, recall dan F1-Score menggunakan



Confusion Matrix. Model SVM yang dikembangkan menunjukkan akurasi sebesar 82.41% untuk penyakit diabetes, sementara untuk penyakit jantung mencapai akurasi 95.24%. Hasil ini menunjukkan peningkatan dalam akurasi diagnosa dibandingkan metode manual. Implementasi model SVM dapat memberikan peningkatan signifikan dalam akurasi dan kecepatan diagnosa penyakit diabetes dan jantung. Model ini dapat digunakan sebagai alat bantu untuk mendukung keputusan medis, meskipun perlu dilakukan pengujian lebih lanjut dengan data yang lebih beragam untuk meningkatkan generalisasi.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Iik Sukira

STMIK Widuri

E-mail: iiksukira86@gmail.com

PENDAHULUAN

Di zaman digital dan pertumbuhan teknologi informasi yang pesat, pemanfaatan big data semakin signifikan. Big data mengacu pada jumlah data yang besar yang berasal dari beragam sumber seperti sensor, perangkat *mobile*, jejaring sosial, dan sistem informasi lainnya. Dalam bidang kesehatan, *big data* dapat dimanfaatkan untuk menemukan pola penyakit, meramalkan risiko kesehatan, serta meningkatkan proses pengambilan keputusan medis (Muhammad Prasetyo et al., 2022).

Di Klinik Pratama Yakrija, terdapat beberapa penyakit yang sering dialami oleh pasien, termasuk diabetes. Oleh karena itu, dalam penelitian kali ini, peneliti mengadopsi model klasifikasi penyakit dengan fokus pada dua penyakit umum, yaitu: diabetes dan penyakit jantung, untuk mengklasifikasikan pola penyakit. Penyakit diabetes dan jantung adalah dua kondisi kronis yang umum terjadi dan memerlukan diagnosa dini untuk mencegah komplikasi serius. Di Klinik Pratama Yakrija, proses diagnosa masih dilakukan secara manual, yang sering kali mengarah pada ketidakakuratan dan keterlambatan dalam pengambilan keputusan. Oleh karena itu, diperlukan sebuah sistem otomatis yang mampu meningkatkan akurasi diagnosa. Namun, dengan kemajuan teknologi saat ini, proses tersebut dapat disederhanakan melalui prediksi atau deteksi penyakit menggunakan algoritma dan metode tertentu (Wafa et al., 2022).

Meramalkan dengan cepat dapat membantu mencegah komplikasi yang serius dan memungkinkan pengambilan langkah untuk sebuah keputusan yang tepat pada saat yang tepat. Penggunaan algoritma prediktif dalam bidang ilmu data telah menunjukkan kemungkinan besar untuk memdiagnosis penyakit, terutama dalam beberapa dekade terakhir ini. (Rizky et al., 2023).

Salah satu teknik dalam proses tersebut adalah penggunaan metode klasifikasi data mining, di mana data yang digunakan memiliki label atau target kelas. Dalam mendukung teknik ini, algoritma *Support Vector Machine* (SVM) digunakan, yang merupakan pendekatan terbaru untuk prediksi baik dalam klasifikasi maupun regresi (Fathurrahman, 2020).

Menurut penelitian yang dilakukan oleh Osisanwo dan rekannya, disetujui bahwa algoritma *Support Vector Machine* (SVM) sangat efektif dan akurat dalam melakukan klasifikasi. SVM adalah suatu teknik yang digunakan untuk menemukan fungsi pemisah yang optimal antara dua kelas yang berbeda. Salah satu keunggulan dari SVM adalah efisiensi



komputasinya yang tinggi karena menggunakan vektor pendukung (Martika Pasaribu et al., 2021).

Munawaroh (2014) melakukan penelitian terkait identifikasi tingkat kerusakan pada jalan rel kereta api dengan menggunakan metode *Support Vector Machine* (SVM). Klasifikasi dilakukan dalam kategori kerusakan berat, sedang, dan ringan. Proses algoritme melibatkan normalisasi data, penerapan kernel polynomial, penghitungan matriks hessian, pelatihan SVM secara berurutan, evaluasi nilai $f(x)$, dan penerapan strategi one-against-all untuk pengklasifikasian. Hasil penelitian menunjukkan tingkat akurasi sebesar 91,24% (Wijayanti et al., 2018).

Penelitian ini bertujuan untuk mengeksplorasi potensi teknologi *Big Data* dan algoritma SVM dalam meningkatkan akurasi diagnosa penyakit di klinik Pratama Yakrija. Diharapkan penelitian ini memberikan kontribusi penting dalam pengembangan sistem kesehatan yang lebih efektif dan efisien, serta meningkatkan kualitas layanan kesehatan. Penelitian ini akan berfokus pada pengembangan sistem yang dapat mengintegrasikan teknologi *Big Data* dan algoritma SVM secara efisien dalam proses diagnosa penyakit di klinik pratama yakrija.

Setelah menjelaskan sebelumnya, peneliti telah memilih judul penelitian “Klasifikasi Teknologi *Big Data* dalam Klasifikasi Pola Penyakit Menggunakan Algoritma *Support Vector Machine* (SVM) di Klinik Pratama Yakrija”.

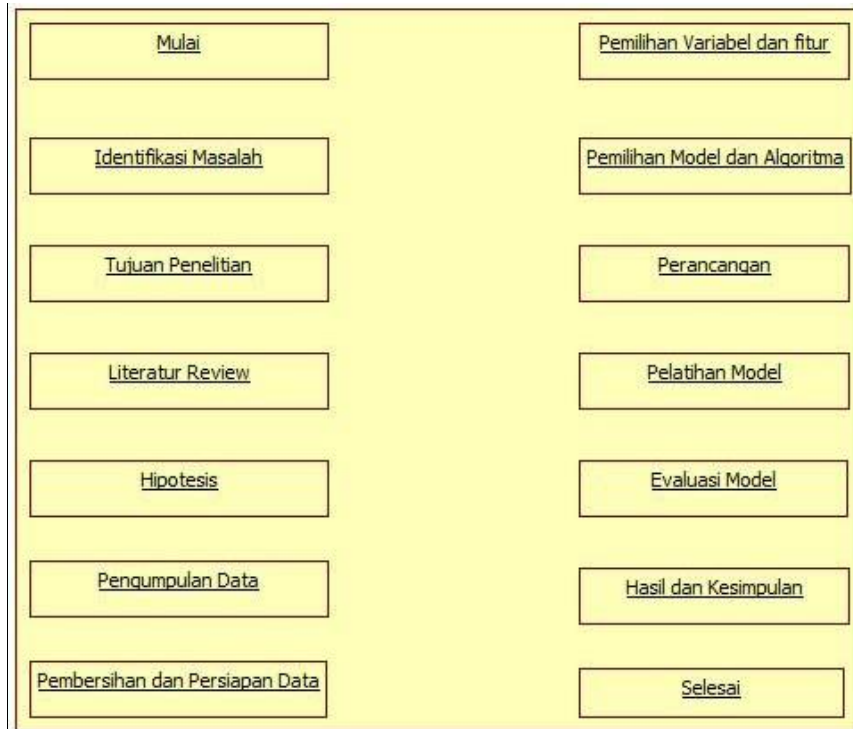
METODE PENELITIAN

Penelitian ini menggunakan algoritma *Support Vector Machine* (SVM) adalah sebuah teknik atau algoritma yang digunakan untuk melakukan klasifikasi serta prediksi. Metode ini beroperasi dengan mencari ruang pemisah yang paling baik antara kelas-kelas yang berbeda dalam *dataset* (Apriyani & Kumiaty, 2020). Konsep utama *Support Vector Machine* (SVM) adalah mencari *hyperplane* pemisah yang memiliki margin maksimal antara dua kelas, yang berarti *hyperplane* tersebut dapat memisahkan data dari kedua kelas dengan jarak terbaik di antara anggota-anggota terdekat dari kedua kelas. *Hyperplane* yang mencapai margin terbesar ini disebut *Optimal Hyperplane* (Purbolaksono et al., 2021).

Penelitian ini menggunakan metode desain eksperimental dengan pendekatan kuantitatif untuk mengklasifikasikan pola penyakit diabetes dan jantung. Sumber data utama dalam penelitian ini adalah rekam medis pasien dari Klinik Pratama Yakrija. Data yang digunakan mencakup informasi medis pasien yang relevan, seperti riwayat kesehatan, hasil pemeriksaan laboratorium, dan diagnosis sebelumnya. Teknik pengumpulan data dilakukan melalui ekstraksi data dari sistem informasi klinik yang telah terkomputerisasi.

Langkah-Langkah Penelitian

Proses ini membantu peneliti dalam merencanakan, menjalankan, dan melaporkan riset dengan cara yang terstruktur dan teratur. Berikut adalah langkah-langkah riset yang memuat langkah penelitian.



Gambar 1. Langkah-Langkah Penelitian

Model Evaluasi *Confusion Matrix*

Confusion Matrix adalah sebuah tabel yang digunakan untuk mengevaluasi seberapa baik model klasifikasi dapat memprediksi kelas atau label data secara benar atau salah. Implementasi menggunakan tools RapidMiner, lalu metode evaluasi *confusion matrix* yang digunakan meliputi pengukuran akurasi, presisi, recall, dan F1-score untuk menilai performa model dalam mengklasifikasikan pola penyakit diabetes dan jantung dalam mengevaluasi hasil kinerja model.

Tabel 1. *Confusion Matrix*

		True Values	
		True	False
Prediction	True	TP Correct result	FP Unexpected result
	False	FN Missing result	TN Connect absence result

Sumber: (Wafa et al., 2022)

Empat sel utama dalam *Confusion Matrix*:

1. *True Positives* (TP): Jumlah kasus di mana model memprediksi dengan benar bahwa pasien menderita penyakit tertentu berdasarkan data sebenarnya.
2. *False Positives* (FP): Jumlah kasus di mana model salah memprediksi bahwa pasien menderita penyakit tertentu ketika sebenarnya tidak.
3. *True Negatives* (TN): Jumlah kasus di mana model memprediksi dengan benar bahwa pasien tidak menderita penyakit tertentu berdasarkan data sebenarnya.



4. *False Negatives* (FN): Jumlah kasus di mana model salah memprediksi bahwa pasien tidak menderita penyakit tertentu ketika sebenarnya ia menderita penyakit tersebut.

Dengan demikian, untuk menghitung berbagai metrik evaluasi kinerja model, digunakan rumus-rumus berikut:

1. Akurasi (*Accuracy*): menunjukkan tingkat akurasi model yang digunakan dalam klasifikasi.

$$\left(\frac{TP + TN}{TP + TN + FP + FN} \right) \times 100\% \quad (1)$$

2. Presisi (*Precision*): Menggambarkan kesesuaian antara data yang diminta dengan hasil prediksi yang diberikan oleh model.

$$\left(\frac{TP}{TP + FP} \right) \times 100\% \quad (2)$$

3. Recall (*Sensitivity* atau *True Positive Rate*): menggambarkan keberhasilan model dalam memulihkan informasi.

$$\left(\frac{TP}{TP + FN} \right) \times 100\% \quad (3)$$

4. F-1 Score: menggambarkan perbandingan rata-rata yang dibobotkan antara precision dan recall.

$$\left(2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \right) \times 100\% \quad (4)$$

HASIL DAN PEMBAHASAN

Hasil Penelitian klasifikasi teknologi big data dalam memprediksi diagnosis pola penyakit menggunakan algoritma SVM mencakup hasil evaluasi model SVM, performa vektor dari akurasi, serta evaluasi model menggunakan *confusion matrix*."

Hasil Evaluasi Model SVM

Di bawah ini terdapat hasil evaluasi model SVM tools RapidMiner, sebagaimana dapat dilihat pada gambar.

accuracy: 82.41%			
	true Diabetes	true Non Diabetes	class precision
pred. Diabetes	25	6	80.65%
pred. Non Diabetes	13	64	83.12%
class recall	65.79%	91.43%	

Gambar 2. Akurasi Model SVM Diagnosa Diabetes

precision: 83.12% (positive class: Non Diabetes)			
	true Diabetes	true Non Diabetes	class precision
pred. Diabetes	25	6	80.65%
pred. Non Diabetes	13	64	83.12%
class recall	65.79%	91.43%	

Gambar 3 Presisi Model SVM Diagnosa Diabetes



recall: 91.43% (positive class: Non Diabetes)			
	true Diabetes	true Non Diabetes	class precision
pred. Diabetes	25	6	80.65%
pred. Non Diabetes	13	64	83.12%
class recall	65.79%	91.43%	

Gambar 4 Recall Model SVM Diagnosa Diabetes

accuracy: 95.24%			
	true Jantung	true Non Jantung	class precision
pred. Jantung	31	0	100.00%
pred. Non Jantung	2	9	81.82%
class recall	93.94%	100.00%	

Gambar 5 Akurasi Model SVM Diagnosa Jantung

precision: 81.82% (positive class: Non Jantung)			
	true Jantung	true Non Jantung	class precision
pred. Jantung	31	0	100.00%
pred. Non Jantung	2	9	81.82%
class recall	93.94%	100.00%	

Gambar 6 Presisi Model SVM Diagnosa Jantung

recall: 100.00% (positive class: Non Jantung)			
	true Jantung	true Non Jantung	class precision
pred. Jantung	31	0	100.00%
pred. Non Jantung	2	9	81.82%
class recall	93.94%	100.00%	

Gambar 7 Recall Model SVM Diagnosa Jantung

PerformanceVector

Berikut adalah hasil *PerformanceVector* yang diperoleh dari tools RapidMiner yang menunjukkan hasil persentase dan evaluasi model *confusion matrix*, seperti yang ditunjukkan pada gambar di bawah ini.

PerformanceVector Diabetes

PerformanceVector		
PerformanceVector:		
accuracy: 82.41%		
ConfusionMatrix:		
True:	Diabetes	Non Diabetes
Diabetes:	25	6
Non Diabetes:	13	64

Gambar 8. *PerformanceVector* Diabetes



Performance Vector Jantung

Performance Vector		
PerformanceVector:		
accuracy: 95.24%		
ConfusionMatrix:		
True:	Jantung	Non Jantung
Jantung:	31	0
Non Jantung:	2	9

Gambar 9. Performance Vector Jantung

Evaluasi Model Pengujian Confusion Matrix

Berikut adalah perhitungan *Confusion Matrix* yang mencakup akurasi, presisi, recall, dan F1-Score pada gambar di atas.

Confusion Matrix Diabetes

Akurasi:

$$\left(\frac{25 + 63}{25 + 63 + 6 + 13} \right) \times 100\% = \left(\frac{89}{108} \right) \times 100\% = 0.8241 = 82.41\%$$

Presisi Diabetes:

$$\left(\frac{25}{25 + 6} \right) \times 100\% = \left(\frac{25}{31} \right) \times 100\% = 0.8065 = 80.65\%$$

Presisi Non Diabetes:

$$\left(\frac{64}{64 + 13} \right) \times 100\% = \left(\frac{64}{77} \right) \times 100\% = 0.8312 = 83.12\%$$

Recall Diabetes:

$$\left(\frac{25}{25 + 13} \right) \times 100\% = \left(\frac{25}{38} \right) \times 100\% = 0.6579 = 65.79\%$$

Recall Non Diabetes:

$$\left(\frac{64}{64 + 6} \right) \times 100\% = \left(\frac{64}{70} \right) \times 100\% = 0.9143 = 91.43\%$$

F1-Score Diabetes:

$$\left(\frac{2 \cdot (0.8065 \cdot 0.6579)}{0.8065 + 0.6579} \right) \times 100\% = \left(\frac{1.0614}{1.4644} \right) \times 100\% = 0.7243 = 72.43\%$$

F1-Score Non Diabetes;

$$\left(\frac{2 \cdot (80.17 \cdot 89.42)}{80.17 + 89.42} \right) \times 100\% = \left(\frac{1.5194}{1.7455} \right) \times 100\% = 0.8705 = 87.05\%$$



Penjelasan:

1. Akurasi: Model ini memiliki tingkat akurasi yang cukup tinggi yaitu 82,41%, menunjukkan bahwa sebagian besar prediksi model adalah benar.
2. Presisi: Presisi untuk diabetes adalah 80,65%, yang berarti sebagian besar prediksi yang diklasifikasikan sebagai diabetes adalah benar-positif. Presisi untuk non-diabetes adalah 83,12%, yang menunjukkan bahwa beberapa prediksi non-diabetes mungkin salah.
3. Recall: Recall untuk penyakit Diabetes adalah 65,79%, menunjukkan bahwa model cukup mampu mengidentifikasi sebagian besar kasus positif penyakit jantung. Recall untuk non-diabetes adalah 91,43%, menunjukkan bahwa model mampu mengidentifikasi hampir semua kasus negatif penyakit Diabetes.
4. F1-Score: Nilai F1-Score yang tinggi untuk kedua kelas menunjukkan keseimbangan yang baik antara presisi dan recall.

Dengan perhitungan ini, hasilnya dapat dibandingkan dengan yang diperoleh dari alat RapidMiner untuk memastikan konsistensi dan validitas model klasifikasi yang digunakan.

Confusion Matrix Jantung

Akurasi:

$$\left(\frac{31 + 9}{31 + 9 + 0 + 2} \right) \times 100\% = \left(\frac{40}{42} \right) \times 100\% = 0.9524 = 95.24\%$$

Presisi Jantung:

$$\left(\frac{31}{31 + 0} \right) \times 100\% = \left(\frac{31}{31} \right) \times 100\% = 1.0 = 100\%$$

Presisi Non Jantung:

$$\left(\frac{9}{9 + 2} \right) \times 100\% = \left(\frac{9}{11} \right) \times 100\% = 0.8182 = 81.82\%$$

Recall Jantung:

$$\left(\frac{31}{31 + 2} \right) \times 100\% = \left(\frac{31}{33} \right) \times 100\% = 0.9394 = 93.94\%$$

Recall Non Jantung:

$$\left(\frac{9}{9 + 0} \right) \times 100\% = \left(\frac{9}{9} \right) \times 100\% = 1.0000 = 100\%$$

F1-Score Jantung:

$$\left(\frac{2 \cdot (1.0000 \cdot 0.9394)}{1.0000 + 0.9394} \right) \times 100\% = \left(\frac{1.8788}{1.9394} \right) \times 100\% = 0.9684 = 96.84\%$$

F1-Score Non Jantung:

$$\left(\frac{2 \cdot (0.8182 \cdot 1.0000)}{0.8182 + 1.0000} \right) \times 100\% = \left(\frac{1.6364}{1.8182} \right) \times 100\% = 0.9000 = 90.00\%$$



Penjelasan:

1. Akurasi: Model ini memiliki tingkat akurasi tinggi, yaitu 95,24%. Ini menunjukkan bahwa sebagian besar prediksi yang dibuat oleh model tersebut adalah benar.
2. Presisi: Presisi untuk prediksi kelas jantung mencapai 100%, yang berarti setiap prediksi yang diklasifikasikan sebagai jantung adalah benar-benar positif. Untuk prediksi kelas non-jantung, presisinya adalah 81,82%, yang mengindikasikan bahwa beberapa prediksi non-jantung mungkin salah.
3. Recall: Tingkat recall untuk jantung adalah 93,94%, yang menunjukkan bahwa model mampu mengidentifikasi sebagian besar kasus positif jantung. Recall untuk non-jantung mencapai 100%, yang berarti model ini mampu mengenali semua kasus negatif jantung.
4. F1-Score: Nilai F1-Score untuk kedua kelas adalah tinggi, menunjukkan adanya keseimbangan yang baik antara presisi dan recall.

Penilaian ini memberikan gambaran yang jelas tentang kinerja model pada data penyakit jantung berdasarkan *confusion matrix* yang disediakan.

Actual Outcome dan Predicted Outcome Diabetes

Tabel 2. *Actual Outcome dan Predicted Outcome Diabetes*

Outcome	Prediction	Confidence (Diabetes)	Confidence (Non Diabetes)
Diabetes	Diabetes	0.7433	0.2567
Non	Non	0.2249	0.7751
Diabetes	Diabetes	0.6332	0.3668
Non	Non	0.2556	0.7444
Diabetes	Diabetes	0.5386	0.4614
Diabetes	Non	0.2776	0.7224
Non	Non	0.4273	0.5727
Diabetes	Diabetes	0.8441	0.1559
Diabetes	Non	0.3797	0.6203
Diabetes	Diabetes	0.6047	0.3953
Non	Non	0.4873	0.5127
Diabetes	Diabetes	0.2819	0.7181
Non	Non	0.1253	0.8747
Diabetes	Diabetes	0.0962	0.9038
Non	Non		
Diabetes	Diabetes		



Actual Outcome (Hasil Aktual) memiliki Jumlah kasus “Diabetes”: 20 dan Jumlah kasus “Non Diabetes”: 65, lalu untuk *Predicted Outcome* (Hasil Prediksi) memiliki Jumlah kasus “Diabetes” yang diprediksi: 17 dan Jumlah kasus “Non Diabetes” yang diprediksi: 68

Actual Outcome dan Predicted Outcome Jantung

Tabel 3. *Actual Outcome dan Predicted Outcome Jantung*

Outcome	Prediction	Confidence (Jantung)	Confidence (Non Jantung)
Jantung	Jantung	0.6987	0.3013
Jantung	Jantung	0.8160	0.1840
Jantung	Jantung	0.5681	0.4319
Jantung	Jantung	0.8010	0.1990
Jantung	Jantung	0.8247	0.1753
Jantung	Jantung	0.8754	0.1246
Jantung	Jantung	0.8992	0.1008
Jantung	Jantung	0.6148	0.3852
Jantung	Jantung	0.8118	0.1882
Jantung	Jantung	0.9743	0.0257
Jantung	Jantung	0.8240	0.1760
Jantung	Non Jantung	0.4163	0.5837
Jantung	Jantung	0.9221	0.0779
Jantung	Jantung	0.7727	0.2273

Actual Outcome (Hasil Aktual) memiliki Jumlah kasus “Jantung”: 11 dan Jumlah kasus “Non Jantung”: 3, sedangkan *Predicted Outcome* (Hasil Prediksi) memiliki Jumlah kasus “Jantung” yang diprediksi: 10 dan Jumlah kasus “Non Jantung” yang diprediksi: 31

Hyperplane dan Margin

Dalam proses garis *hyperplane* dan penentuan *margin*, diperlukan nilai bias dan bobot untuk setiap atribut atau fitur yang digunakan. Berikut ini adalah nilai bias dan bobot untuk setiap atribut atau fitur yang dihasilkan oleh *tools* RapidMiner.

Hyperplane dan Margin Diabetes

```
Total number of Support Vectors: 429
Bias (offset): 0.709

w[Pregnancies] = -0.353
w[Glucose] = -0.837
w[BloodPressure] = 0.220
w[SkinThickness] = -0.038
w[Insulin] = 0.042
w[BMI] = -0.437
w[DiabetesPedigreeFunction] = -0.195
w[Age] = 0.081
```

Gambar 10. Bias dan *Weight* Diabetes

Hyperplane dan Margin Jantung

```
Total number of Support Vectors: 170
Bias (offset): -1.110

w[age] = 0.013
w[sex] = 0.233
w[cp] = -0.431
w[trestbps] = 0.077
w[chol] = 0.001
w[fbs] = 0.022
w[restecg] = -0.162
w[thalach] = -0.114
w[exang] = 0.172
w[oldpeak] = 0.345
w[slope] = -0.131
w[ca] = 0.299
w[thal] = 0.384
```

Gambar 11. Bias dan *Weight* Jantung

PEMBAHASAN

Pada tahap ini, peneliti mengevaluasi temuan-temuan dengan mengaitkannya pada tujuan penelitian dan hipotesis yang telah dirumuskan selama proses penelitian berlangsung. Berikut adalah keterangan pada masing-masing atribut data untuk penyakit diabetes dan penyakit jantung yang dijelaskan pada tabel dibawah.

1. Diabetes

Data ini adalah penderita penyakit diabetes yang terdiri dari 8 atribut, yaitu: *Pregnancies*, *Glucose*, *BloodPressure*, *SkinThickness*, *Insulin*, *BMI (Body Massa Index)*, *DiabetesPedigreeFunction*, *Age* dan 1 label, yaitu: *Outcome* atau *Output*.

Tabel 4. Atribut Data Diabetes

No	Atribut	Keterangan
1	Pregnancies	Jumlah kehamilan yang dialami oleh pasien
2	Glucose	Konsentrasi glukosa plasma dua jam setelah tes toleransi glukosa
3	BloodPressure	Tekanan darah diastolik (mm Hg)
4	SkinThickness	Ketebalan lipatan kulit triceps (mm)
5	Insulin	Konsentrasi insulin serum dua jam (mu U/ml)
6	BMI (Body Massa Index)	Indeks massa tubuh (berat dalam kg/(tinggi dalam m) ²)
7	Diabetes PedigreeFunction	Fungsi silsilah diabetes yang mengukur kemungkinan risiko diabetes berdasarkan sejarah keluarga
8	Age	Usia pasien dalam tahun
9	Outcome	Hasil diagnosa apakah pasien menderita diabetes atau tidak

Sumber: (Klinik Pratama Yakrija 2024)



2. Jantung

Data tersebut adalah penderita penyakit diabetes yang terdiri dari 13 atribut, yaitu: *Age*, *Sex*, *ChestPainType*, *RestingBloodPressure*, *Cholesterol*, *FastingBloodSugar*, *RestingECG*, *MaxHeartRate*, *ExerciseInducedAngina*, *Oldpeak*, *Slope*, *Number of Major Vessels*, *Thal* dan 1 label, yaitu: *Outcome* atau *Output*.

Tabel 5. Atribut Data Jantung

No	Atribut	Keterangan
1	Age	Usia pasien dalam tahun
2	Sex	Jenis kelamin pasien (1 = laki-laki, 0 = perempuan)
3	ChestPainType	Tipe nyeri dada (0 = Typical Angina, 1 = Atypical Angina, 2 = Non-anginal Pain, 3 = Asymptomatic)
4	RestingBlood Pressure	Tekanan darah saat istirahat (mm Hg)
5	Cholesterol	Serum kolesterol dalam mg/dl
6	FastingBloodSugar	Gula darah puasa (> 120 mg/dl) (1 = benar, 0 = salah)
7	RestingECG	Hasil elektrokardiografi istirahat (0 = Normal, 1 = ST-T wave abnormality, 2 = menunjukkan kemungkinan atau pasti hipertrofi ventrikel kiri)
8	MaxHeartRate	Detak jantung maksimal yang dicapai selama uji stres
9	ExerciseInduced Angina	Nyeri angina yang disebabkan oleh olahraga (1 = ya, 0 = tidak)
10	Oldpeak	Depresi ST yang diinduksi oleh olahraga relatif terhadap istirahat
11	Slope	Kemiringan segmen ST puncak latihan (0 = upsloping, 1 = flat, 2 = downsloping)
12	Number of Major Vessels	Jumlah pembuluh besar (0-3) yang diwarnai oleh fluoroskopi
13	Thal	Jenis thalassemia (3 = Normal, 6 = Fixed defect, 7 = Reversible defect)
14	Outcome	Hasil diagnosa apakah pasien menderita penyakit jantung atau tidak

Sumber: (Klinik Pratama Yakrija 2024)

Dataset Penyakit Pasien Diabetes

	A	B	C	D	E	F	G	H	I
1	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
2	6	148	72	35	0	34	0.627	50	Diabetes
3	1	85	66	29	0	27	0.351	31	Non Diabetes
4	8	183	64	0	0	23	0.672	32	Diabetes
5	1	89	66	23	94	28	0.167	21	Non Diabetes
6	0	137	40	35	168	43	2.288	33	Diabetes
7	5	116	74	0	0	26	0.201	30	Non Diabetes
8	3	78	50	32	88	31	0.248	26	Diabetes
9	10	115	0	0	0	35	0.134	29	Non Diabetes
10	2	197	70	45	543	31	0.158	53	Diabetes
11	8	125	96	0	0	0	0.232	54	Diabetes
12	4	110	92	0	0	38	0.191	30	Non Diabetes
13	10	168	74	0	0	38	0.537	34	Diabetes
14	10	139	80	0	0	27	1.441	57	Non Diabetes
15	1	189	60	23	846	30	0.398	59	Diabetes
16	5	166	72	19	175	26	0.587	51	Diabetes
17	7	100	0	0	0	30	0.484	32	Diabetes
18	0	118	84	47	230	46	0.551	31	Diabetes
19	7	107	74	0	0	30	0.254	31	Diabetes
20	1	103	30	38	83	43	0.183	33	Non Diabetes
21	1	115	70	30	96	35	0.529	32	Diabetes
22	3	126	88	41	235	39	0.704	27	Non Diabetes

Gambar 12. Dataset Penyakit Pasien Diabetes

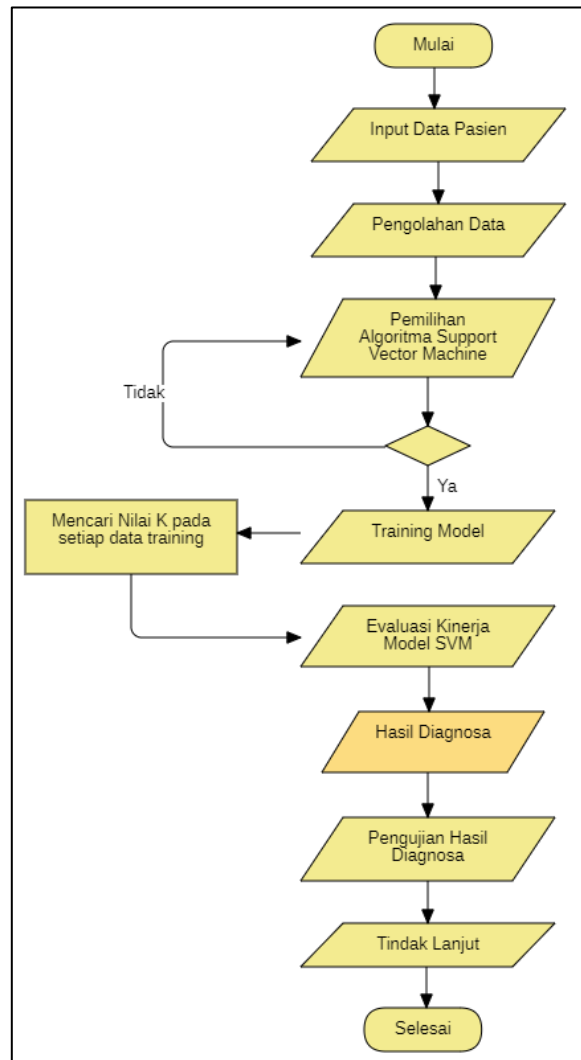
Dataset Penyakit Pasien Jantung

	C	D	E	F	G	H	I	J	K	L	M	N
1	cp	trestbps	chol	fbbs	restecg	thalach	exang	oldpeak	slope	ca	thal	Outcome
2	3	145	233	1	0	150	0	2.3	0	0	1	Jantung
3	2	130	250	0	1	187	0	3.5	0	0	2	Jantung
4	1	130	204	0	0	172	0	1.4	2	0	2	Jantung
5	1	120	236	0	1	178	0	0.8	2	0	2	Jantung
6	0	120	354	0	1	163	1	0.6	2	0	2	Jantung
7	0	140	192	0	1	148	0	0.4	1	0	1	Jantung
8	1	140	294	0	0	153	0	1.3	1	0	2	Jantung
9	1	120	263	0	1	173	0	0	2	0	3	Jantung
10	2	172	199	1	1	162	0	0.5	2	0	3	Jantung
11	2	150	168	0	1	174	0	1.6	2	0	2	Jantung
12	0	140	239	0	1	160	0	1.2	2	0	2	Jantung
13	2	130	275	0	1	139	0	0.2	2	0	2	Jantung
14	1	130	266	0	1	171	0	0.6	2	0	2	Jantung
15	3	110	211	0	0	144	1	1.8	1	0	2	Jantung
16	3	150	283	1	0	162	0	1	2	0	2	Jantung
17	2	120	219	0	1	158	0	1.6	1	0	2	Jantung
18	2	120	340	0	1	172	0	0	2	0	2	Jantung
19	3	150	226	0	1	114	0	2.6	0	0	2	Jantung
20	0	150	247	0	1	171	0	1.5	2	0	2	Jantung
21	3	140	239	0	1	151	0	1.8	2	2	2	Jantung
22	0	135	234	0	1	161	0	0.5	1	0	3	Jantung

Gambar 13. Dataset Penyakit Pasien Diabetes

Design Sistem Usulan

Design Sistem Usulan merupakan dasar yang dipergunakan untuk memulai pengembangan sistem secara praktis setelah proses perencanaan telah selesai. Tujuannya adalah untuk menjamin bahwa semua aspek teknis, arsitektur, dan alur kerja sistem telah direncanakan dengan baik sebelum memasuki tahap implementasi. Selain itu, rancangan ini dapat berfungsi sebagai pedoman untuk memonitor kemajuan sebuah proyek dan mengevaluasi sejauh mana proyek sesuai dengan yang sudah direncanakan sebelumnya.



Gambar 14. Flowchart Diagram Design Sistem Usulan

Berikut adalah Komponen Utama Sistem:

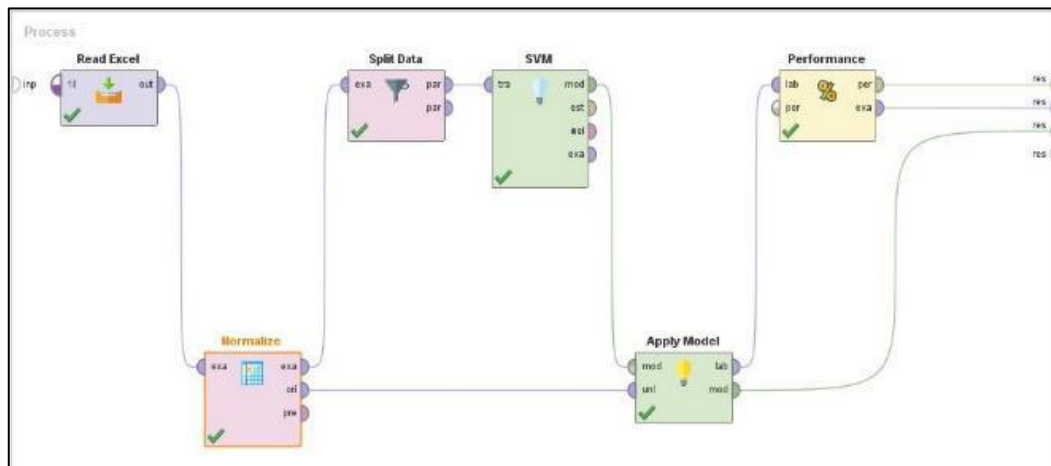
1. *Frontend Interface*: Antarmuka pengguna yang memungkinkan petugas kesehatan untuk memasukkan data pasien dan menerima hasil diagnosa.
2. *Backend Processing*: Komponen yang bertanggung jawab untuk memproses data yang dimasukkan, menerapkan algoritma SVM, dan menghasilkan hasil diagnosa.
3. Database: Tempat penyimpanan data pasien yang diperlukan untuk proses diagnosa.
4. Algoritma SVM: Algoritma machine learning yang digunakan untuk mengklasifikasikan pola penyakit berdasarkan data yang diberikan.

Dengan desain sistem yang komprehensif ini, diharapkan bahwa penelitian ini dapat meningkatkan kualitas diagnosa penyakit diabetes dan jantung di Klinik Pratama Yakrija, serta memberikan kontribusi signifikan dalam penerapan teknologi *Big Data* dan SVM di bidang medis.

Design Proses Algoritma SVM

Data yang telah dipreproses kemudian dianalisis menggunakan algoritma SVM untuk klasifikasi. SVM dipilih karena kemampuannya yang baik dalam mengklasifikasikan data yang

kompleks dan memiliki dimensi yang tinggi. Berikut gambaran design proses Algoritma SVM yang menggunakan tools RapidMiner.



Gambar 15. Design Proses Algoritma SVM

Proses pengolahan data

Proses pengolahan data menggunakan algoritma SVM pada RapidMiner melibatkan beberapa langkah utama.

1. Memuat Data: Langkah pertama adalah menggunakan fungsi *"Read Excel"* mengimpor atau memuat data yang berformat *Microsoft Excel*.
2. *Normalize*: Langkah ini melibatkan normalisasi data, meskipun rinciannya tidak dijelaskan secara lengkap.
3. Pembagian *Dataset*: Fungsi *"Split Data"* membagi *dataset* yang terdiri dari 749 data secara keseluruhan, dataset dibagi menjadi 80% data *training* dan 20% data *testing*. Pasien diabetes memiliki 537 data menjadi 429 data *training* dan 108 data *testing*, lalu untuk pasien jantung memiliki 212 data menjadi 170 data *training* dan 42 data *testing*.
4. Penggunaan Algoritma SVM: Algoritma *Support Vector Machine* (SVM) diterapkan sebagai metode klasifikasi dalam teknologi *big data* untuk memprediksi penyakit di Klinik Pratama Yakrija.
5. Penerapan Model: Fungsi *"Apply Model"* digunakan untuk melakukan prediksi atau klasifikasi pada data baru yang dimasukkan ke dalam sistem.
6. Analisis Kinerja: Fungsi *"Performance"* digunakan untuk menganalisis atau memproses data Excel yang telah diinput sehingga menghasilkan keluaran berupa *persentase* akurasi dari model yang diterapkan.

KESIMPULAN

Penelitian ini mengeksplorasi penerapan algoritma *Support Vector Machine* (SVM) untuk klasifikasi pola penyakit menunjukkan akurasi sebesar 82.41% untuk penyakit diabetes dan 95.24% untuk penyakit jantung di Klinik Yakrija Pratama dengan memanfaatkan teknologi *Big Data*. Berdasarkan hasil analisis dan pengujian, beberapa kesimpulan yang dapat ditarik adalah sebagai berikut:



1. Efektivitas Algoritma *Support Vector Machine* (SVM):
Algoritma SVM menunjukkan kemampuan yang signifikan dalam mengklasifikasikan pola penyakit diabetes dan jantung dengan tingkat akurasi, presisi, dan recall yang memuaskan. Hasil analisis menunjukkan bahwa SVM dapat diandalkan dalam proses diagnosa penyakit yang lebih akurat dan cepat dibandingkan dengan metode manual.
2. Efisiensi Proses Diagnosa:
3. Implementasi SVM mempercepat proses diagnosa dengan mengurangi waktu analisis data dari beberapa jam menjadi beberapa menit. Hal ini memungkinkan klinik untuk menangani lebih banyak pasien dan memberikan layanan yang lebih cepat dan efisien.
4. Pemanfaatan *Big Data*:
5. Teknologi *Big Data* memungkinkan pengolahan data medis yang besar dan beragam, membantu dalam pengidentifikasian pola penyakit yang lebih akurat dan pengambilan keputusan medis yang lebih tepat.

DAFTAR PUSTAKA

- Apriyani, H., & Kumiati. (2020). Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus. In *Journal of Information Technology Ampera* (Vol. 1, Issue 3). <https://journal-computing.org/index.php/journal-ita/index>
- Fathurrahman, I. (2020). Klasifikasi Penentuan Penerima Program Keluarga Harapan (PKH) Menggunakan Algoritma Support Vector Machine (Svm) Pada Kantor Dinas Sosial Lombok Timur. *Jurnal Informatika Dan Teknologi*, 3(1), 27.
- Martika Pasaribu, I., Rosyking Lumbanraja, F., Asiah Shofiana, D., & Aristoteles. (2021). *Klasifikasi Kejadian Hipertensi dengan Metode Support Vector Machine (SVM) Menggunakan Data Puskesmas di Kota Bandar Lampung* (Vol. 2, Issue 2).
- Muhammad Prasetyo, T., Amrullah, A., Syahrir, S., & Nurina Sari, B. (2022). IMPLEMENTASI ALGORITMA SVM (SUPPORT VECTOR MACHINE) DALAM KLASIFIKASI PENYAKIT PARU-PARU BERDASARKAN FITUR POLA BENTUK. *Jurnal Teknologi Informasi*, 6(1), 1–6.
- Purbolaksono, M. D., Irvan Tantowi, M., Imam Hidayat, A., & Adiwijaya, A. (2021). Perbandingan Support Vector Machine dan Modified Balanced Random Forest dalam Deteksi Pasien Penyakit Diabetes. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(2), 393–399. <https://doi.org/10.29207/resti.v5i2.3008>
- Rizky, M., Irianto, Y., & Sriyanto. (2023). Perbandingan Kinerja Algoritma Naive Bayes, Support Vector Machine dan Random forest untuk Prediksi Penyakit Ginjal Kronis Imaniar Ikko Mulya Rizky 1a* , Suhendro Yusuf Irianto 2b. Sriyanto 3c. *Institut Informatika Dan Bisnis Darmajaya*, 139–151.
- Wafa, S., Hadiana, I., & Umbara, R. (2022). *Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM) INFORMASI ARTIKEL ABSTRAK* (Vol. 4, Issue 1). <https://e-journal.unper.ac.id/index.php/informatics>



- Wijayanti, R. A., Furqon, M. T., & Adinugroho, S. (2018). *Penerapan Algoritme Support Vector Machine Terhadap Klasifikasi Tingkat Risiko Pasien Gagal Ginjal* (Vol. 2, Issue 10). <http://j-ptiik.ub.ac.id>