

OPTIMASI PARAMETER SUPPORT VECTOR MACHINE (SVM) UNTUK PREDIKSI RISIKO KANKER PAYUDARA MENGGUNAKAN BREAST CANCER COIMBRA DATASET

Andita Titi Sulanjari, Aprilisa Arum Sari, Hanifah Permatasari

Teknik Informatika, Universitas Duta Bangsa Surakarta

Jl. Bhayangkara No.55, Tipes, Kec. Serengan, Kota Surakarta, Jawa Tengah 57154

220103174@mhs.udb.ac.id

ABSTRAK

Kanker payudara merupakan jenis kanker dengan angka kejadian tertinggi di dunia, dengan 2,3 juta kasus baru dan 685.000 kematian pada tahun 2020. Deteksi dini menjadi kunci utama dalam meningkatkan harapan hidup pasien, namun ketergantungan pada metode konvensional yang bersifat subjektif masih menyisakan celah kesalahan diagnosis yang signifikan. Penelitian ini bertujuan mengoptimalkan kinerja algoritma *Support Vector Machine* (SVM) untuk prediksi risiko kanker payudara menggunakan *Breast Cancer Coimbra Dataset* melalui *hyperparameter* tuning sistematis. Metode yang diterapkan adalah *Grid Search Cross Validation* untuk mengeksplorasi 108 kombinasi *hyperparameter* secara menyeluruh, dikombinasikan dengan *K-Fold Cross Validation* untuk memastikan stabilitas dan generalisasi model. Hasil optimasi menunjukkan bahwa konfigurasi terbaik diperoleh pada kernel RBF dengan $C = 10$ dan $\gamma = \text{scale}$. Model yang telah dioptimasi mencapai akurasi 83,33%, *precision* 81,82%, *recall* 81,82%, dan *F1-score* 81,82%, meningkat dibandingkan model sebelum tuning yang hanya memperoleh akurasi 75,00%. Validasi *10-Fold Cross Validation* mengonfirmasi stabilitas model dengan akurasi rata-rata 73,33% tanpa indikasi *overfitting*.

Kata kunci : *Breast Cancer Coimbra Dataset, Support Vector Machine (SVM), Grid Search Cross Validation, K-Fold Cross Validation*

1. PENDAHULUAN

Kanker payudara merupakan jenis kanker dengan angka kejadian tertinggi di dunia. Menurut laporan WHO, tercatat 2,3 juta kasus baru dan 685.000 kematian pada tahun 2020, dengan sekitar 7,8 juta penyintas yang masih bertahan hidup dalam lima tahun terakhir setelah terdiagnosis [1].

Penyakit ini tidak hanya mengancam perempuan dewasa, tetapi juga berisiko menyerang sejak usia pubertas, sehingga deteksi dini menjadi aspek krusial dalam upaya penanganannya. Deteksi dini terbukti mampu meningkatkan efektivitas pengobatan, menekan angka kematian, dan memperbaiki harapan hidup pasien secara signifikan [2].

Namun demikian, praktik diagnosis kanker payudara saat ini masih bertumpu pada metode pemeriksaan konvensional yang rentan terhadap penilaian subjektif tenaga medis. Kondisi ini menyisakan celah kesalahan diagnosis yang cukup signifikan, yang pada akhirnya dapat menghambat pemberian penanganan yang tepat waktu dan efektif [3].

Oleh karena itu, dibutuhkan pendekatan berbasis data yang mampu mengklasifikasikan risiko kanker secara objektif dan konsisten.

Penerapan data mining menawarkan solusi atas permasalahan tersebut dengan memungkinkan sistem melakukan klasifikasi dan prediksi risiko kanker payudara secara otomatis. Proses ini bekerja dengan mengekstraksi hubungan antar variabel klinis dari kumpulan data berskala besar untuk mengenali pola-

pola yang menjadi landasan akurasi sistem prediksi medis [4].

Salah satu algoritma yang terbukti efektif untuk keperluan ini adalah *Support Vector Machine* (SVM), yang bekerja dengan mencari *hyperplane* optimal guna memisahkan kelas data non-linear berdimensi tinggi. SVM mampu mengolah fitur kompleks secara akurat bahkan dengan jumlah data yang terbatas, asalkan menggunakan fungsi kernel yang sesuai [5].

Penelitian ini menggunakan *Breast Cancer Coimbra Dataset* yang memuat data hasil tes darah dan pengukuran antropometri tubuh sebagai pembeda antara individu sehat dan pasien kanker payudara [6].

Dataset ini dinilai tepat untuk skenario deteksi dini yang cepat, meskipun pemanfaatannya secara klinis masih terus dikembangkan. Berdasarkan urgensi tersebut, penelitian ini bertujuan mengoptimalkan kinerja algoritma SVM pada dataset tersebut melalui *hyperparameter tuning* sistematis menggunakan metode *Grid Search* [7].

Strategi penyelesaian yang ditempuh mencakup implementasi SVM dengan penyetelan *hyperparameter* otomatis, evaluasi performa model menggunakan metrik klasifikasi yang relevan, serta analisis komparatif hasil sebelum dan sesudah tuning. Hasil penelitian ini diharapkan memberikan kontribusi teoritis pada pengembangan ilmu data mining, khususnya dalam optimasi model klasifikasi untuk deteksi dini kanker payudara.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Berbagai penelitian telah dikembangkan dengan memanfaatkan teknologi *machine learning* sebagai alat bantu dalam proses skrining dan deteksi kanker payudara. Salah satu contohnya adalah penerapan algoritma *Support Vector Machine* (SVM) untuk membedakan antara tumor bersifat jinak dan ganas. Dengan mengombinasikan teknik validasi *K-Fold Cross Validation*, model yang dikembangkan berhasil memperoleh tingkat akurasi sebesar 98,21%. Lebih lanjut, pengintegrasian model ke dalam antarmuka berbasis *Flask* memungkinkan masyarakat umum untuk mengakses layanan skrining secara lebih praktis dan efisien [8].

Di samping pemilihan algoritma yang tepat, konfigurasi parameter atau yang dikenal sebagai *hyperparameter tuning* memiliki peran yang tidak kalah penting dalam menentukan kualitas performa suatu model. Dalam konteks algoritma SVM, dua parameter utama yang perlu mendapat perhatian khusus adalah parameter C sebagai pengendali regularisasi, serta nilai gamma yang mendefinisikan bentuk *decision boundary* pada fungsi kernel RBF. Penetapan nilai kedua parameter tersebut secara sembarangan tanpa melalui proses optimasi yang terstruktur berisiko menurunkan akurasi klasifikasi secara drastis dan menghasilkan prediksi yang tidak dapat diandalkan [9].

Sebuah penelitian yang mengklasifikasikan skizofrenia menggunakan sinyal EEG menunjukkan bahwa dengan menerapkan optimasi menggunakan *Grid Search Cross Validation* (GSCV), akurasi dapat ditingkatkan secara signifikan dari 90,25% menjadi 99,75% [10].

2.2. Kanker Payudara

Pada jaringan payudara, transformasi genetik seperti mutasi, delesi kromosom, amplifikasi, atau duplikasi anomali memicu timbulnya tumor maligna yang dikenal sebagai kanker payudara, di mana sel-sel mengalami proliferasi dan mitosis tanpa kendali [11].

2.3. Sistem Prediksi

Proses prediksi melibatkan upaya meramalkan variabel masa depan melalui analisis data historis kuantitatif. Dalam kerangka sistem kecerdasan buatan, tahap ini mencakup pembangunan model atas pola historis rumit guna menghasilkan estimasi yang akurat dan mendekati realitas aktual, tanpa menjanjikan kepastian absolut terhadap kejadian yang akan datang [12].

2.4. SVM

Algoritma *Support Vector Machine* (SVM) melakukan klasifikasi dengan mengidentifikasi *hyperplane* optimal yang memaksimalkan pemisahan data antar kelas secara paling tajam dan efektif [13].

2.5. K-Fold Cross Validation

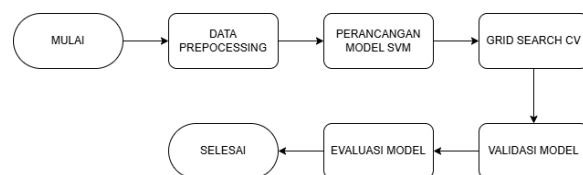
Teknik *K-Fold Cross Validation* membagi dataset secara sistematis menjadi subset pelatihan dan pengujian melalui proses iteratif, memastikan setiap elemen data bergantian berperan sebagai set uji. Pendekatan ini secara efektif mengurangi bias sampel akibat pembagian data sekali pakai, sehingga menghasilkan evaluasi model yang lebih obyektif, representatif, dan mencerminkan kondisi data riil [14].

2.6. Grid Search CV

Metode *Grid Search* mengeksplorasi beragam kombinasi hyperparameter SVM secara sistematis guna mengoptimalkan akurasi prediksi pada data tak dikenal. Praktiknya menggabungkan *K-Fold Cross-Validation* untuk menghasilkan metrik evaluasi yang andal dan imparial, sehingga menentukan serta menetapkan konfigurasi parameter superior sebagai fondasi model prediksi primer [10].

3. METODE PENELITIAN

Pendekatan eksperimental diterapkan dalam studi ini guna mengoptimalkan hyperparameter algoritma *Support Vector Machine* (SVM) menggunakan *Breast Cancer Coimbra Dataset*, dengan alur kerja yang terstruktur dalam tahapan-tahapan kunci berikut:



Gambar 1. Tahapan penelitian

Tahapan utama dalam penelitian ini mencakup:

- Pra-pemrosesan Data:** Membersihkan dan menyiapkan dataset agar siap dimanfaatkan oleh model.
- Konstruksi Model SVM:** Mengembangkan sistem klasifikasi berbasis algoritma *Support Vector Machine*.
- Pencarian Grid:** Dengan Validasi Silang Mengidentifikasi kombinasi *hyperparameter* superior.
- Pengujian Kestabilan Model:** Memverifikasi keandalan melalui teknik *K-Fold Cross Validation*.
- Penilaian Performa:** Menganalisis efektivitas model menggunakan metrik evaluasi standar.

3.1. Jenis dan Sumber Data

Penelitian ini memanfaatkan data sekunder dari *Breast Cancer Coimbra Dataset*, yang diakses melalui repositori publik Kaggle. Koleksi data ini mencakup 116 rekam medis pasien beserta variabel klinis utama, termasuk usia (*Age*), indeks massa tubuh (BMI), level glukosa, insulin, HOMA, leptin, adiponektin, *resistin*, serta MCP-1. Spesifikasi lengkap fitur tersaji dalam tabel berikut:

Tabel 1. Informasi Dataset

No	Fitur	Tipe Data	Keterangan
1	Age	Numerik	Usia pasien (tahun)
2	BMI	Numerik	Body Mass Index (indeks massa tubuh)
3	Glucose	Numerik	Kadar glukosa dalam darah
4	Insulin	Numerik	Kadar insulin dalam darah
5	HOMA	Numerik	Indeks resistensi insulin (<i>Homeostasis Model Assessment</i>)
6	Leptin	Numerik	Hormon leptin yang berhubungan dengan metabolisme lemak
7	Adiponectin	Numerik	Protein yang berperan dalam regulasi metabolisme
8	Resistin	Numerik	Hormon yang berkaitan dengan resistensi insulin
9	MCP-1	Numerik	Protein inflamasi (<i>Monocyte Chemoattractant Protein-1</i>)
10	Classification	Kategorial	Label kelas (2=sehat, 1=kanker)

3.2. Data Preprocessing

Selama pra-pemrosesan, analisis dan pembersihan data dilakukan dengan membuang fitur tak relevan, entri duplikat, serta rekam tidak lengkap untuk meningkatkan kualitas dataset [15].

Keberadaan data hilang berpotensi memicu bias serta menekan ketepatan prediksi, sehingga baris dengan nilai null dieliminasi guna memastikan kebersihan data dan hanya memanfaatkan entri valid untuk pelatihan model [12].

Proses persiapan data mencakup [10]:

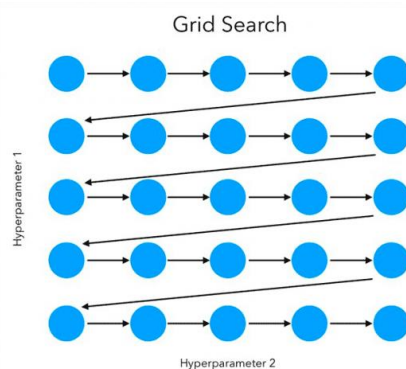
- Seleksi Atribut: Memilih variabel yang signifikan terhadap tugas klasifikasi.
- Sanitasi Data: Menangani nilai null, fitur irelevan, dan inkonsistensi.
- Konversi Data: Mengubah data kategorikal menjadi representasi numerik.

3.3. Perancangan Sistem SVM

Studi ini menerapkan algoritma Support Vector Machine (SVM) guna melakukan klasifikasi dengan mengidentifikasi *hyperplane* pemisah yang memaksimalkan margin antar kelas. Keunggulan SVM terletak pada kemampuannya mengelola data nonlinier melalui proyeksi ke ruang berdimensi tinggi, sehingga pencapaian kategorisasi tetap efisien [16].

Algoritma ini menetapkan *hyperplane* superior berdasarkan margin serta vektor pendukung, di mana fungsi pemisah optimal dipilih di antara kelas berlawanan via pencarian margin terluas untuk klasifikasi linier [13].

3.4. Grid Search Cross Validation



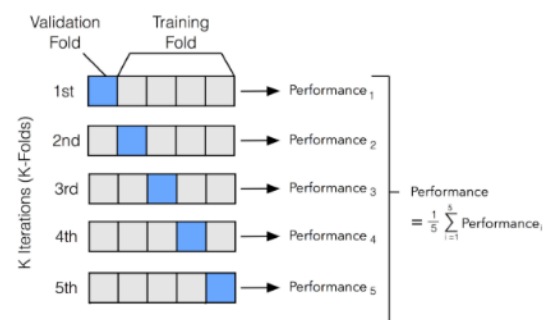
Gambar 2. Grid search cv

Tahap ini fokus pada penyetelan *hyperparameter* untuk menemukan kombinasi parameter superior yang memaksimalkan performa model. Pustaka *scikit-learn* menyediakan *GridSearchCV*, yang secara otomatis mengeksplorasi beragam kombinasi *hyperparameter* melalui validasi silang sistematis, sehingga menghasilkan akurasi optimal bagi model SVM [17].

3.5. Validasi Model

Metode *K-Fold Cross Validation* mempartisi dataset menjadi k subset (*fold*), di mana model dilatih dan diuji secara bergantian sebanyak k iterasi [8].

Rata-rata akurasi dari k model tersebut menjadi dasar evaluasi performa, sementara *grid search* memanfaatkannya untuk menyempurnakan parameter klasifier guna mencapai tingkat akurasi maksimal [10].



Gambar 3. Iterasi k-fold validation

3.6. Evaluasi Model

Performa model klasifikasi *machine learning* dinilai melalui *confusion matrix* guna menentukan algoritma superior yang meningkatkan akurasi melebihi pendekatan konvensional [15].

Penilaian ini membandingkan hasil model pasca penyetelan *hyperparameter* dengan versi tanpa tuning [16].

Sebagai instrumen analisis, *confusion matrix* mengukur efektivitas klasifikasi dalam membedakan objek antar kelas, dengan empat metrik pokok: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), serta *False Negative* (FN). Metrik-metrik ini menghasilkan skor akurasi, presisi, dan recall untuk mengevaluasi model, dihitung melalui persamaan berikut: [17]

$$Accuracy = \frac{TP+PN}{TF+FP+FN+TN} \times 100\% \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (3)$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

4. HASIL DAN PEMBAHASAN

4.1. Preprocessing Data

Breast Cancer Coimbra Dataset, yang mencakup 116 rekam pasien dengan berbagai atribut klinis serta label target klasifikasi, mengalami pra-pemrosesan. Pemeriksaan awal mengonfirmasi tidak adanya nilai hilang, sehingga seluruh data tetap utuh tanpa perlu sanitasi lanjutan. Atribut terakhir, *Classification*, mengklasifikasikan subjek ke dalam kelompok kontrol (sehat) atau pasien. Statistik deskriptif setiap fitur tersaji pada tabel berikut:

Tabel 2. Distribusi kelas

Status Label	Kondisi Diagnosa	Jumlah Sampel	Presentase
Kelas 0	Kontrol (Sehat)	52	44.8%
Kelas 1	Pasien (Kanker)	64	55.2%

Dataset selanjutnya dipartisi menjadi set pelatihan dan pengujian dengan proporsi 80:20 melalui stratified sampling guna menjaga keseimbangan kelas. Normalisasi *MinMaxScaler* (rentang 0–1) diterapkan khusus pada data latih, dengan transformasi serupa untuk data uji berdasarkan parameter tersebut, mencegah kebocoran informasi (*data leakage*).

Tabel 3. Diskripsi fitur dataset

Fitur	Mean	Std	Min	Max
Age	57.30	16.11	24.00	89.00
BMI	27.58	5.02	18.37	38.57
Glucose	97.79	22.53	60.00	201.00
Insulin	10.01	10.07	2.43	58.46
HOMA	2.69	3.64	0.47	25.05
Leptin	26.62	19.18	4.31	90.28
Adiponectin	10.18	6.84	1.65	38.04
Resistin	14.73	12.39	3.21	82.10
MCP.1	534.65	345.91	45.84	1698.40

Distribusi kelas relatif merata, tanpa isu ketidakseimbangan kelas yang mencolok. Pasca pra-pemrosesan, data telah optimal untuk tahap pelatihan dan penilaian model.

4.2. Perancangan Model SVM

Penyelarasan parameter dilakukan secara sistematis melalui mekanisme *Grid Search* yang terintegrasi dengan *5-fold Stratified Cross-Validation*. Strategi ini diterapkan untuk mengeksplorasi ruang

pencarian (*search space*) secara menyeluruh guna menemukan koordinat parameter yang menghasilkan batas keputusan (*decision boundary*) paling optimal. Fase optimasi ini melibatkan pengujian terhadap 108 kombinasi unik yang berasal dari tiga hiperparameter utama:

Tabel 4. Konfigurasi parameter

Parameter	Nilai yang Diuji	Jumlah
Complexity (C)	[0.01, 0.1, 1, 10, 100, 1000]	6
Gamma	['scale', 'auto', 0.001, 0.01, 0.1, 1]	6
Kernel	['linear', 'rbf', 'poly']	3
Total Kombinasi		108

Seluruh proses komputasi ini dijalankan dengan menetapkan *Random State* pada nilai 42 untuk menjamin keterulangan (*reproducibility*) hasil eksperimen. Penggunaan *5-Fold Stratified Cross-Validation* sebagai validasi internal dalam *Grid Search* memastikan bahwa setiap iterasi parameter dinilai secara objektif dengan tetap mempertahankan proporsi kelas pada setiap lipatan data. Melalui pendekatan sistematis ini, model SVM yang dihasilkan diharapkan memiliki ketangguhan dalam mendeteksi risiko kanker payudara dengan tingkat generalisasi yang tinggi.

4.3. Optimasi Grid Search CV

Proses pencarian parameter secara menyeluruh melalui mekanisme *Grid Search CV* telah menghasilkan pemetaan performa dari berbagai kombinasi hiperparameter. Berdasarkan data hasil iterasi, ditemukan bahwa konfigurasi model mencapai titik efektivitas tertinggi pada peringkat pertama dengan detail sebagai berikut:

Tabel 5. Tabel hasil grid search

Rank	Kernel	C	Gamma	Mean Score	Std Score
1	RBF	10.0	scale	76.20%	0.0768
2	RBF	1.0	scale	75.09%	0.0694
3	RBF	100.0	1	74.91%	0.0743
4	POLY	100.0	1	74.85%	0.0703
5	RBF	10.0	1	73.98%	0.0771
6	RBF	100.0	auto	73.92%	0.0406
6	RBF	100.0	0.1	73.92%	0.0204
8	POLY	1.0	scale	73.80%	0.0844
8	POLY	10.0	1	73.80%	0.0682
10	POLY	10.0	scale	72.63%	0.0828
11	POLY	1000.0	auto	70.64%	0.0751
12	RBF	100.0	scale	70.53%	0.0772
13	RBF	1000.0	0.1	70.47%	0.0869
14	POLY	0.1	scale	69.59%	0.0533
14	POLY	1000.0	0.1	69.59%	0.0533
14	POLY	1.0	1	69.59%	0.0533
17	LINEAR	100.0	0.1	69.42%	0.0796
17	LINEAR	100.0	0.01	69.42%	0.0796
17	LINEAR	100.0	auto	69.42%	0.0796
17	LINEAR	100.0	0.001	69.42%	0.0796

Hasil optimasi mengungkap konfigurasi unggul kernel RBF, $C = 10$, $\gamma = scale$, yang mencapai akurasi validasi silang 76,20% indikasi generalisasi model yang solid. Penggunaan kernel RBF mencerminkan sifat nonlinier relasi data, memerlukan pemetaan ke ruang dimensi tinggi untuk pemisahan efektif. Nilai $C = 10$ menjaga harmoni antara margin luas dan toleransi error klasifikasi, sedangkan $\gamma = scale$ memungkinkan adaptasi otomatis sensitivitas berdasarkan jumlah fitur, sehingga model lebih peka terhadap pola inheren data.

Kesimpulannya, *Grid Search Cross Validation* sukses menghasilkan pengaturan hyperparameter yang meningkatkan kapabilitas SVM dalam memprediksi risiko kanker payudara.

4.4. Validasi K-Fold Cross Validation

Stabilitas kinerja diuji lebih lanjut melalui validasi Setelah mendapatkan parameter optimal dan menguji performanya pada subset data pengujian, dilakukan validasi tambahan menggunakan metode 10-Fold Cross-Validation. Langkah ini krusial untuk memastikan bahwa model memiliki tingkat reliabilitas yang tinggi dan tidak bersifat bias terhadap partisi data tertentu. Hasil agregasi dari sepuluh iterasi menunjukkan stabilitas performa yang solid. Rincian capaian setiap lipatan (*fold*) beserta parameter evaluasinya disajikan pada tabel berikut:

Tabel 6. Validasi k-fold

Fold	TP	FP	FN	TN	Akurasi	Presisi	Recall	F1-Score
1	3	2	3	4	0,5833	0,6000	0,5000	0,5455
2	5	2	1	4	0,7500	0,7143	0,8333	0,7692
3	4	3	1	4	0,6667	0,5714	0,8000	0,6667
4	5	0	0	7	1,0000	1,0000	1,0000	1,0000
5	4	1	1	6	0,8333	0,8000	0,8000	0,8000
6	1	2	4	5	0,5000	0,3333	0,2000	0,2500
7	5	2	0	4	0,8182	0,7143	1,0000	0,8333
8	3	0	2	6	0,8182	1,0000	0,6000	0,7500
9	3	2	2	4	0,6364	0,6000	0,6000	0,6000
Agregasi					0,7333	0,7083	0,6933	0,6881

Dengan akurasi rata-rata 73,33% meski ada fluktuasi antar *fold*, menandakan kapabilitas generalisasi yang konsisten terhadap dataset. Konsistensi nilai akurasi di setiap fold membuktikan bahwa model SVM yang telah dioptimasi memiliki kemampuan generalisasi yang baik untuk mendeteksi risiko kanker payudara pada data yang bervariasi.

4.5. Hasil Evaluasi Model

Setelah parameter optimal ditetapkan, pengujian dilakukan pada data independen untuk mengukur peningkatan efektivitas dibandingkan dengan model SVM standar. Hasil evaluasi menunjukkan penguatan signifikan pada hampir seluruh metrik klasifikasi.

Tabel 7. Perbandingan model

Metrik Evaluasi	Sebelum Optimasi	Setelah Optimasi
Akurasi	75.00%	83,33%
Presisi	77.78%	81,82%
Recall	63.64%	81,82%
F1-Score	70.00%	81,82%

Temuan evaluasi mengindikasikan bahwa SVM raih akurasi 83,33%, *precision* 81,82%, serta *recall* 81,82%. Data tersebut mengonfirmasi bahwa penalaan parameter secara presisi berhasil meningkatkan akurasi sebesar 4,35%. Meskipun tingkat sensitivitas (*recall*) tidak mengalami perubahan, peningkatan presisi sebesar 5,49% menandakan bahwa model kini lebih selektif dan akurat dalam mendeteksi subjek yang benar-benar berisiko kanker payudara.

Tabel 8. Klasifikasi hasil setelah optimasi

	Precision	Recall	F1-Score	Support
0	0.85	0.85	0.85	13
1	0.82	0.82	0.82	11
Accuracy			0.83	24
Macro Avg	0.83	0.83	0.83	24
Weighted Avg	0.83	0.83	0.83	24

Setelah membangun model menggunakan parameter tersebut, selanjutnya melakukan evaluasi sesuai skenario pengujian yang telah dirancang. Dokumentasi hasil evaluasi mencakup perhitungan metrik performa dan visualisasi *Confusion Matrix* dalam bentuk diagram. Berikut adalah rumus perhitungan performa model setelah optimasi:

Tabel 9. Confusion matrix model terbaik

	Pred: Kontrol	Pred: Kanker
Aktual: Kontrol	11	2
Aktual: Kanker	2	9

a. Accuracy

$$\begin{aligned}
 &= \frac{TP+TN}{TP+FP+FN+TN} \times 100\% \\
 &= \frac{9+11}{9+11+2+2} \times 100\% \\
 &= \frac{20}{24} \times 100\% \\
 &= 83.33\%
 \end{aligned}
 \tag{5}$$

b. Precision

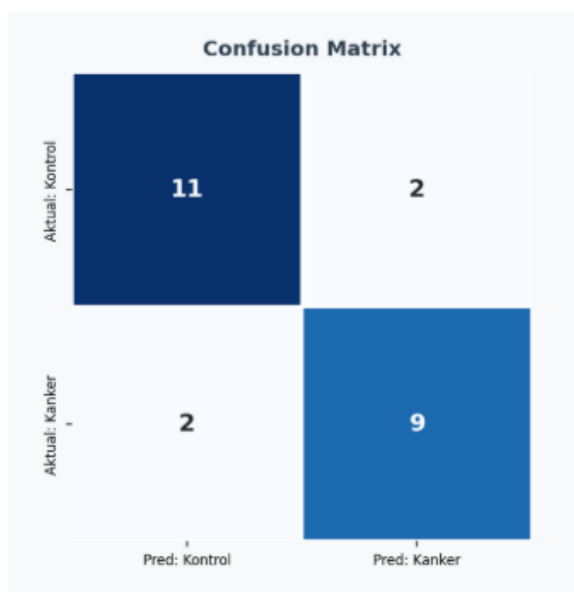
$$\begin{aligned}
 &= \frac{TP}{TP+FP} \times 100\% \\
 &= \frac{9}{9+2} \times 100\% \\
 &= \frac{9}{11} \times 100\% \\
 &= 81.82\%
 \end{aligned}
 \tag{6}$$

c. Recall

$$\begin{aligned}
 &= \frac{TP}{TP+FN} \times 100\% \\
 &= \frac{9}{9+2} \times 100\% \\
 &= \frac{9}{11} \times 100\% \\
 &= 81.82\%
 \end{aligned}
 \tag{7}$$

$$\begin{aligned}
 \text{d. } F1 &= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\
 &= \frac{2 \times 0.812 \times 0.812}{0.812 + 0.812} \\
 &= 81.82\%
 \end{aligned} \quad (8)$$

Analisis confusion matrix mengungkap bahwa model sukses mengenali 11 sampel kontrol dan 9 kasus kanker dengan tepat. Meski demikian, 2 sampel kontrol terdeteksi salah sebagai kanker (*false positive*), serta 2 kasus kanker dikira kontrol (*false negative*). Pencapaian ini mencerminkan kinerja model yang solid dan seimbang dalam mendeteksi kedua kategori, yakni kanker maupun kontrol. Secara keseluruhan, evaluasi mengonfirmasi bahwa SVM pasca-optimalisasi menampilkan performa memadai untuk memprediksi risiko kanker payudara.



Gambar 4. Confusion matrix

4.6. Pembahasan

Model *Support Vector Machine* (SVM) pasca optimasi menunjukkan kinerja solid dalam memprediksi risiko kanker payudara, mencapai akurasi 83,33% artinya sebagian besar sampel uji terklasifikasi akurat. *Precision* 81,82% mencerminkan ketepatan model dalam mengonfirmasi kasus kanker sejati, sementara *recall* serupa (81,82%) menandakan deteksi efektif terhadap mayoritas instance kanker pada data uji. Keseimbangan keduanya menghasilkan *F1-score* 81,82%, yang unggul dibandingkan konfigurasi pra-optimalisasi.

Confusion matrix mengungkap 2 *false negative* yaitu pasien kanker yang salah diprediksi sebagai kontrol konstitusi krusial di ranah medis karena berpotensi menunda intervensi diagnostik. Sebaliknya, 2 *false positive* menjaga paritas error, menunjukkan klasifikasi yang lebih netral dan adil antar kelas ketimbang model awal.

Pengujian *10-Fold Cross Validation* raih akurasi rata-rata 73,33% dengan fluktuasi *fold* yang terkendali,

mengonfirmasi stabilitas model serta minimnya *overfitting* terhadap data pelatihan. Walaupun kernel polynomial kadang unggul tipis pada iterasi tertentu, kernel RBF terpilih sebagai yang superior berkat optimasi komprehensif via *Grid Search Cross Validation*, berkat kepekaannya terhadap pola nonlinier data untuk performa holistik optimal.

Pada intinya, penelitian ini membuktikan keberhasilan *Grid Search Cross Validation* dalam menyempurnakan SVM, selaras dengan sasaran utama yakni konfigurasi *hyperparameter* superior guna tingkatkan prediksi kanker payudara.

5. KESIMPULAN DAN SARAN

Berdasarkan rangkaian pengujian dan analisis optimasi *hyperparameter Support Vector Machine* (SVM) untuk prediksi risiko kanker payudara pada *Breast Cancer Coimbra Dataset*, dapat disimpulkan bahwa metode *Grid Search Cross Validation* berhasil mengidentifikasi kombinasi parameter optimal yang menghasilkan klasifikasi lebih presisi dibandingkan model tanpa tuning. Konfigurasi *hyperparameter* terbaik diperoleh menggunakan kernel *Radial Basis Function* (RBF) dengan $C = 1$ serta $\gamma = 1$. Model akhir mencapai akurasi sebesar 83,33%, dengan nilai *precision*, *recall*, dan *F1-score* masing-masing sebesar 81,82%, yang menunjukkan keseimbangan performa yang baik antara kemampuan deteksi positif dan negatif. Selain itu, validasi menggunakan *K-Fold Cross Validation* mengonfirmasi bahwa model memiliki kemampuan generalisasi yang kuat tanpa indikasi *overfitting* terhadap data baru.

Guna kemajuan riset mendatang dan penguatan reliabilitas sistem, rekomendasi berikut layak diadopsi: Ekspansi dan Penyeimbangan Dataset: Adopsi dataset lebih luas atau metode *oversampling* untuk optimalkan *recall* serta kurangi *false negative*. Perbandingan Algoritma: Uji SVM melawan *Random Forest*, *Naive Bayes*, atau *Neural Network* guna pilih model dengan akurasi superior. Pengembangan Aplikasi: Integrasikan model ke UI/UX web atau mobile agar tenaga kesehatan dapat memanfaatkannya secara langsung sebagai alat bantu keputusan.

DAFTAR PUSTAKA

- [1] Adriana Sainafat, "Pemahaman Pemeriksaan Payudara Sendiri (SADARI) sebagai Upaya Preventif Kanker Payudara pada Remaja Putri," *INSOLOGI: Jurnal Sains dan Teknologi*, vol. 5, no. 2, hlm. 544–552, Apr 2026, doi: 10.55123/insologi.v5i2.7888.
- [2] Rosita Natania Maulani dan Zaehol Fatah, "Klasifikasi Data Kanker Payudara Menggunakan Algoritma Decision Tree Berbasis Rapidminer," *Jurnal Mahasiswa Teknik Informatika*, vol. 4, no. 2, hlm. 127–132, Okt 2025, doi: 10.35473/jamastika.v4i2.4504.
- [3] H. Setiani dan N. Trisanti, "Penerapan Metode Correlated Naive Bayes Untuk Klasifikasi Penyakit Kanker Payudara," *JOISIE (Journal Of*

- Information Systems And Informatics Engineering*), vol. 9, no. 1, hlm. 18, Feb 2025, doi: 10.35145/joisie.v9i1.4769.
- [4] N. Nurjanah, A. Nur Rani, H. Hikmayanti Handayani, A. Fitri Nur Masruriyah, U. Buana Perjuangan Karawang Kota Karawang, dan P. Korespondensi Diajukan, “Implementasi Model Klasifikasi Jenis Kanker Payudara Menggunakan Algoritma SVM dan Logistic Regression berbasis Web,” *Riset dan E-Jurnal Manajemen Informatika Komputer*, vol. 7, No.4, no. 4, hlm. 2023, Okt 2023, doi: 10.33395/remik.v7i4.12817.
- [5] Tinaliah dan T. Elizabeth, “Prediksi Jenis Kanker Payudara Menggunakan Metode Support Vector Machine Berbasis Recursive Feature Elimination,” *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 11, No.3, hlm. 1–5, Sep 2024, doi: 10.35957/jatisi.v11i3.10079.
- [6] G. Chuiko dan D. Honcharov, “Breast Cancer Dataset from Coimbra: Pre-Ratings of Its Value to Machine Learning and Diagnosis,” *Digital Technologies Research and Applications*, vol. 4, no. 2, hlm. 182–193, Sep 2025, doi: 10.54963/dtra.v4i2.1348.
- [7] S. Hurriyati, “Implementasi Metode Support Vector Machine Pada Klasifikasi Diagnosis Penyakit Kanker Payudara,” 2023. Diakses: 12 Juni 2026. [Daring]. Tersedia pada: <http://etheses.uin-malang.ac.id/id/eprint/52582>
- [8] N. Everi Rahmawati, K. Aeni, J. Raya Pagoejengan Km, dan J. Tengah, “02-Implementasi SVM untuk Klasifikasi Kanker Payudara Berbasis Website,” 2026.
- [9] Adil Setiawan, Andri Armaginda Siregar, N. Setiawan, Jalaluddin Nasution, Naufal Dhiya Putra Dalimunthe, dan Farhan Sardy Abdillah, “Optimasi Performa Model SVM dan Random Forest untuk Klasifikasi Kanker Payudara Menggunakan Penyetelan Hyperparameter,” *Jurnal Komputer Teknologi Informasi Sistem Informasi (JUKTISI)*, vol. 4, No.3, no. 3, hlm. 2141–2149, Feb 2026, doi: 10.62712/juktisi.v4i3.789.
- [10] M. Desiawan dan A. Solichin, “SVM Optimization with Grid Search Cross Validation for Improving Accuracy of Schizophrenia Classification Based on EEG Signal,” *JURNAL TEKNIK INFORMATIKA*, vol. 17, No.1, no. 1, hlm. 10–20, Mei 2024, doi: 10.15408/jti.v17i1.37422.
- [11] N. Cahyani, R. Irsyada, dan A. Y. Kartini, “Implementasi Machine Learning Model sebagai Sistem Prediksi Penyakit Breast Cancer,” *Digital Transformation Technology*, vol. 4, no. 2, hlm. 1112–1120, Jan 2025, doi: 10.47709/digitech.v4i2.5209.
- [12] R. Kamaluddin, “Optimasi Model Support Vector Regression (Svr) Dengan Gridsearchcv Untuk Prediksi Harga Saham Bank Bri,” Mar 2025. doi: 10.58486/jsr.v9i1.469.
- [13] N. T. R. Adiningrum, R. Rianti, dan C. Priyanto, “Rancang Bangun Aplikasi Prediksi Kanker Payudara Dengan Pendekatan Machine Learning,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 11, No.3, no. 3s1, hlm. 2830–7062, Sep 2023, doi: 10.23960/jitet.v11i3s1.3351.
- [14] M. Yordanus dan W. Senjaya, “Prediksi Diagnosis dan Prognosis Breast Cancer menggunakan Machine Learning,” *Jurnal STRATEGI-Jurnal Maranatha*, vol. 6, No.2, no. Vol 6 No 2 (2024): Strategi, hlm. 284–290, Nov 2024.
- [15] A. Rifa’i dan M. Ardiyanto, “Analisis Kinerja Algoritma Support Vector Machine dengan Pendekatan Oversampling pada Prediksi Kanker Payudara,” *Prosiding Seminar Nasional Amikom Surakarta*, vol. 3, hlm. 15–25, Jan 2025.
- [16] I Nyoman Adi Mahendra Putra dan Cokorda Pramatha, “Optimasi Hyperparameter Algoritma Support Vector Machine dalam Klasifikasi Penyakit β -Thalassemia,” *Jurnal Nasional Teknologi Informasi dan Aplikasinya*, vol. 3, No.2, hlm. 283–294., Feb 2025.
- [17] Z. Maisat Eka Darmawan dan A. Fauzan Dianta, “Implementasi Optimasi Hyperparameter GridSearchCV Pada Sistem Prediksi Serangan Jantung Menggunakan SVM,” *Teknologi*, vol. 13, no. 1, hlm. 8–15, Jan 2023, doi: 10.26594/teknologi.v13i1.3098