

***IMPROVEMENT MODEL PREDIKSI KANKER PAYUDARA
MENGUNAKAN SUPPORT VECTOR MACHINE DENGAN
KERNEL LINEAR, RBF, HYPERPARAMETER***



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

VERONICA LANDREA DARNELLA OSWARI

00000028353

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2022

***IMPROVEMENT MODEL PREDIKSI KANKER PAYUDARA
MENGUNAKAN SUPPORT VECTOR MACHINE DENGAN
KERNEL LINEAR, RBF, HYPERPARAMETER***



Diajukan sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana Komputer (S. Kom.)

VERONICA LANDREA DARNELLA OSWARI

0000028353

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA**

**UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2022

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Veronica Landrea Darnella Oswari
Nomor Induk Mahasiswa : 00000028353
Program studi : Sistem Informasi

Skripsi dengan judul:

“Improvement Model Prediksi Kanker Payudara menggunakan Support Vector Machine dengan Kernel Linear, RBF, Hyperparameter” merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/ penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas akhir yang telah saya tempuh.

Tangerang, 14 Juni 2022



(Veronica Landrea Darnella Oswari)

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

HALAMAN PENGESAHAN

Skripsi dengan judul
*Improvement Model Prediksi Kanker Payudara menggunakan Support Vector
Machine dengan Kernel Linear, RBF, Hyperparameter*

Oleh

Nama : Veronica Landrea Darnella Oswari

NIM : 00000028353

Program Studi : Sistem Informasi

Fakultas : Teknik dan Informatika

Telah diujikan pada hari Kamis, 23 Juni 2022

Pukul 17.00 s.d 19.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut.

Ketua Sidang



Monika Evelin Johan, S.Kom, M.MSI.
071281

Penguji



Friska Natalia, Ph. D
026068

Pembimbing

6-7-2022



Ir. Raymond Sunardi Octama, M.C.I.S.
328046803/023897

Ketua Program Studi



06/07/2022

Ririn Ikana Desanti, S.Kom., M.Kom.

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Multimedia Nusantara, saya yang bertandatangan di bawah ini:

Nama : Veronica Landrea Darnella Oswari
NIM : 00000028353
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika
JenisKarya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Multimedia Nusantara Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul:

“Improvement Model Prediksi Kanker Payudara menggunakan Support Vector Machine dengan Kernel Linear, RBF, Hyperparameter “

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini, Universitas Multimedia Nusantara berhak menyimpan, mengalihmediakan/mengalihformatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 14 Juni 2022

Yang menyatakan,



Veronica Landrea Darnella Oswari

KATA PENGANTAR

Puji Syukur atas selesainya penulisan skripsi ini dengan judul: “*Improvement Model Prediksi Kanker Payudara menggunakan Support Vector Machine dengan Kernel Linear, RBF, Hyperparameter*” dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Jurusan Sistem Informasi Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan tugas akhir ini, sangatlah sulit bagi saya untuk menyelesaikan tugas akhir ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Universitas Multimedia Nusantara.
3. Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Universitas Multimedia Nusantara.
4. Ir. Raymond Sunardi Oetama, M.C.I.S., sebagai Pembimbing pertama yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya tesis ini.
5. Orang Tua dan keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tesis ini.
6. Teman-teman baik teman sekolah maupun teman kuliah yang telah memberikan support, dukungan, dan bantuan dalam proses pengerjaan.

Semoga karya ilmiah ini bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 14 Juni 2022



Veronica Landrea Darnella Oswari

IMPROVEMENT MODEL PREDIKSI KANKER PAYUDARA MENGUNAKAN SUPPORT VECTOR MACHINE DENGAN KERNEL LINEAR, RBF, HYPERPARAMETER

Veronica Landrea Darnella Oswari

ABSTRAK

Kanker payudara paling umum menyerang wanita dan menjadi persoalan besar di dunia kesehatan. Pertambahan kasusnya merupakan yang tertinggi di dunia dan kasus kematiannya terbanyak ketiga di dunia. Upaya deteksi dini perlu dilakukan agar penderita segera di tangani dan dilakukan perawatan sehingga dapat meminimalisir kasus kematian karena terlambatnya deteksi dan penanganan.

Salah satu solusi yang dilakukan yaitu melalui *data mining* dengan mengkombinasikan algoritma SVM dengan *Kernel Linear*, *Kernel RBF*, dan *hyperparameter*. *Kernel Linear* dan *Kernel RBF* merupakan *kernel* yang biasa dipakai pada SVM. Pemilihan *hyperparameter* dilakukan karena masih jarang ditemui penelitian kanker payudara yang menggunakan *hyperparameter*. Selain itu dari penelitian terdahulu di bidang kesehatan, *hyperparameter* sukses meningkatkan akurasi. Penelitian ini akan menggunakan *framework* CRISP-DM dengan *programming Python* dan akan menggunakan bantuan *tools Jupyter Notebook*. Penelitian ini bertujuan untuk melakukan prediksi terhadap penyakit kanker payudara.

Dari penelitian ini, diperoleh nilai akurasi SVM menggunakan *hyperparameter* dengan diketahui *Kernel* terbaik yaitu RBF dengan akurasi 98.38%, *f1-score* 98.38%, *precision* 100%, dan *recall* 96.82%. Maka diketahui *hyperparameter* merupakan metode yang cocok untuk penelitian ini. Penelitian diharapkan dapat membantu bidang kesehatan dan dapat menambah informasi pengetahuan mengenai implementasi *data mining* terlebih dalam algoritma SVM *Kernel Linear*, *Kernel RBF*, dan *hyperparameter*.

Kata kunci: *Hyperparameter*, Kanker Payudara, *Kernel Linear*, *Kernel RBF*, *Support Vector Machine*

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

***IMPROVEMENT MODEL OF BREAST CANCER PREDICTION
USING SUPPORT VECTOR MACHINE WITH LINEAR
KERNEL, RBF, HYPERPARAMETER***

Veronica Landrea Darnella Oswari

ABSTRACT

Breast cancer is the most common in women and is a big problem in the world of health. The increase in cases is the highest in the world and the number of deaths in the third in the world. Early detection efforts need to be carried out so that patients are immediately treated so that they can minimize cases of death due to delays in detection and treatment.

One solution that is done is through data mining by combining the SVM algorithm with Linear Kernel, RBF Kernel, and hyperparameters. Linear Kernel and RBF Kernel are kernels commonly used in SVM. The selection of hyperparameters was made because breast cancer research is still rarely found using hyperparameters. In addition, from previous research in the health sector, hyperparameters have succeeded in increasing accuracy. This research will use the CRISP-DM framework with Python programming and will use Jupyter Notebook tools. This study aims to predict breast cancer.

From this study, the SVM accuracy value was obtained using hyperparameters with the best-known kernel, namely RBF with 98.38% accuracy, 98.38% f1-score, 100% precision, and 96.82% recall. It is known that hyperparameter is a suitable method for this research. The research is expected to help the health sector and can add information to knowledge about the implementation of data mining, especially in the SVM Kernel Linear algorithm, Kernel RBF, and hyperparameters.

Keywords: *Breast Cancer, Hyperparameter, Kernel Linear, Kernel RBF, Support Vector Machine*

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS.....	v
KATA PENGANTAR.....	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
DAFTAR RUMUS	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	6
1.3 Batasan Masalah.....	7
1.4 Tujuan Penelitian dan Manfaat Penelitian.....	7
1.4.1 Tujuan Penelitian	7
1.4.2 Manfaat Penelitian	7
1.5 Sistematika Penulisan.....	8
BAB II LANDASAN TEORI	10
2.1 Penyakit Kanker Payudara	10
2.2 CRISP-DM	11
2.3 <i>Support Vector Machine</i>	12
2.4 <i>Hyperparameter</i>	14
2.5 <i>Label encoding</i>	14
2.6 <i>Confusion matrix</i>	15
2.7 <i>Classification report</i>	15
2.7.1 <i>Accuracy</i>	16

2.7.2	<i>Precision</i>	16
2.7.3	<i>Recall</i>	16
2.7.4	<i>F1-score</i>	17
2.8	<i>Tools</i>	17
2.9	Penelitian Terdahulu.....	17
BAB III METODOLOGI PENELITIAN		23
3.1	Gambaran Umum Objek Penelitian	23
3.2	Metode Penelitian.....	23
3.2.1	<i>Business Understanding</i>	24
3.2.2	<i>Data Understanding</i>	24
3.2.3	<i>Data Preparation</i>	25
3.2.4	<i>Modelling</i>	26
3.2.5	<i>Evaluation</i>	28
3.2.6	<i>Deployment</i>	28
3.3	Variabel Penelitian	28
3.3.1	Variabel Independen	29
3.3.2	Variabel Dependen.....	29
3.4	Teknik Pengumpulan Data	29
3.5	Teknik Pengambilan Sampel.....	30
3.6	Teknik Analisis Data	30
BAB IV ANALISIS DAN HASIL PENELITIAN		31
4.1	Tahapan <i>Business Understanding</i>	31
4.2	Tahapan <i>Data Understanding</i>	31
4.3	Tahapan <i>Data Preparation</i>	34
4.3.1	<i>Drop Unused Columns</i>	34
4.3.2	<i>Encoding Categorical Data M (Malignant) and B (Benign)</i>	35
4.3.3	<i>Normalize the Data</i>	37
4.3.4	<i>Split Data</i>	39
4.4	Tahapan Modelling.....	40
4.4.1	SVM with <i>Kernel Linear</i>	40
4.4.2	SVM with <i>Kernel RBF</i>	44
4.4.3	SVM with <i>Hyperparameter</i>	47

4.4.4	SVM <i>Linear</i> and RBF with <i>C</i> and <i>Gamma</i>	50
4.5	Tahapan <i>Evaluation</i>	56
4.5.1	Evaluasi Kinerja <i>Support Vector Machine Kernel Linear</i>	56
4.5.2	Evaluasi Kinerja <i>Support Vector Machine Kernel RBF</i>	59
4.5.3	Evaluasi Kinerja <i>Support Vector Machine Hyperparameter</i>	62
4.5.4	Evaluasi Komparasi Kinerja <i>Support Vector Machine</i> dengan <i>Kernel Linear</i> , <i>Kernel RBF</i> , dan <i>Hyperparameter</i>	64
4.5.5	Hasil Diskusi	67
BAB V	KESIMPULAN DAN SARAN	70
5.1	Kesimpulan	70
5.2	Saran	70
	DAFTAR PUSTAKA	72
	LAMPIRAN	78

UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR TABEL

Tabel 2.1 Kernel yang dapat digunakan untuk mempersiapkan Model SVM [24]	14
Tabel 2.2 Penelitian Terdahulu yang Menggunakan CRISP-DM.....	17
Tabel 2.3 Penelitian Terdahulu yang menjadi Acuan	19
Tabel 4.1 Spesifikasi Hardware dan Software	40
Tabel 4.2 Komparasi Pemodelan SVM dengan Kernel Linear, RBF, Hyperparameter.....	64
Tabel 4.3 Hasil Implementasi dengan Nilai $C = 10$ dan $\text{Gamma} = 0.1$	66
Tabel 4.4 Perbandingan Akurasi terhadap Penelitian Terdahulu.....	68

UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR GAMBAR

Gambar 1.1 Trend Kanker Payudara Hingga 1999-2018 [2].....	1
Gambar 1.2 Chart Jenis Kanker yang Menyerang Wanita 2020 [3]	2
Gambar 1.3 Data Kasus Baru dan Kematian Penyakit Kanker tahun 2020 [3]	3
Gambar 2.1 Fase CRISP-DM [20]	11
Gambar 2.2 <i>Hyperlane</i> pada SVM [23]	13
Gambar 2.3 <i>Confusion matrix</i> [28]	15
Gambar 3.1 Tahapan CRISP-DM [20]	24
Gambar 4.1 Proses melakukan <i>Retrieve</i> Data	32
Gambar 4.2 Proses Cek <i>Benign</i>	32
Gambar 4.3 Proses Pengecekan Data NULL	33
Gambar 4.4 Proses Pengecekan Tipe Data	34
Gambar 4.5 Melakukan <i>Drop</i> pada <i>Attribute</i> yang Tidak Digunakan	35
Gambar 4.6 <i>Syntax</i> Untuk Membuat <i>Barplot Malignant</i> dan <i>Benign</i>	35
Gambar 4.7 Hasil <i>Malignant</i> dan <i>Benign</i>	36
Gambar 4.8 <i>Syntax</i> untuk Menampilkan Value dari <i>Attribute</i> Diagnosis.....	36
Gambar 4.9 Rincian Jumlah <i>Attribute</i> Diagnosis.....	36
Gambar 4.10 <i>Syntax</i> dan Hasil dari Proses <i>Label encoding</i>	37
Gambar 4.11 <i>Syntax</i> untuk Mendefinisikan Independen Variabel.....	37
Gambar 4.12 Hasil dari Independen Variabel.....	38
Gambar 4.13 <i>Syntax</i> Melakukan Normalisasi	39
Gambar 4.14 Hasil dari Normalisasi Data	39
Gambar 4.15 Proses Melakukan <i>Split</i> Data.....	39
Gambar 4.16 <i>Library</i> Untuk Modelling menggunakan Support Vector Machine	41
Gambar 4.17 <i>Syntax</i> untuk SVM <i>Kernel Linear</i>	41
Gambar 4.18 Hasil Model SVM <i>Kernel Linear</i>	41
Gambar 4.19 <i>Syntax</i> untuk Membuat <i>Confusion matrix</i>	42
Gambar 4.20 Hasil <i>Confusion matrix</i>	43
Gambar 4.21 Hasil dari <i>Classification report</i>	44
Gambar 4.22 <i>Syntax</i> Membangun SVM <i>Kernel RBF</i>	44
Gambar 4.23 Hasil <i>Accuracy</i> dengan <i>Kernel RBF</i>	45
Gambar 4.24 <i>Syntax</i> Membuat <i>Confusion matrix Kernel RBF</i>	45
Gambar 4.25 Hasil <i>Confusion matrix Kernel RBF</i>	46
Gambar 4.26 Hasil <i>Classification report Kernel RBF</i>	47
Gambar 4.27 <i>Library</i> SVM <i>Hyperparameter</i>	47
Gambar 4.28 <i>Syntax</i> dan Output <i>Hyperparameter</i>	48
Gambar 4.29 Hasil Akurasi Menggunakan <i>Hyperparameter</i>	48
Gambar 4.30 <i>Syntax Confusion matrix Hyperparameter</i>	49
Gambar 4.31 Hasil <i>Confusion matrix Hyperparameter</i>	49
Gambar 4.32 <i>Syntax</i> dan Hasil <i>Classification report Hyperparameter</i>	50
Gambar 4.33 Hasil Akurasi <i>Kernel Linear</i> dengan $C = 1$, $\Gamma = 10$	51

Gambar 4.34 Syntax untuk <i>Confusion matrix Kernel Linear</i> dengan $C = 1$, $\Gamma = 10$	51
Gambar 4.35 <i>Confusion matrix</i> untuk <i>Kernel Linear</i> dengan $C = 1$, $\Gamma = 10$	52
Gambar 4.36 Syntax <i>Classification report Kernel Linear</i> dengan $C = 1$, $\Gamma = 10$	52
Gambar 4.37 Hasil <i>Classification report Kernel Linear</i> dengan $C = 1$, $\Gamma = 10$	53
Gambar 4.38 Hasil Akurasi <i>Kernel RBF</i> dengan $C = 1$, $\Gamma = 10$	53
Gambar 4.39 Syntax <i>Confusion matrix Kernel RBF</i> dengan $C = 1$, $\Gamma = 10$	54
Gambar 4.40 Hasil <i>Confusion matrix Kernel RBF</i> dengan $C = 1$, $\Gamma = 10$..	54
Gambar 4.41 Syntax membuat <i>Classification Report Kernel RBF</i> dengan $C = 1$, $\Gamma = 10$	55
Gambar 4.42 Hasil <i>Classification report Kernel RBF</i> dengan $C = 1$, $\Gamma = 10$	55
Gambar 4.43 Akurasi SVM <i>Kernel Linear</i>	56
Gambar 4.44 <i>Confusion matrix SVM Kernel Linear</i>	57
Gambar 4.45 <i>Classification report Kernel Linear</i>	58
Gambar 4.46 Akurasi SVM dengan RBF	59
Gambar 4.47 <i>Confusion matrix SVM Kernel RBF</i>	60
Gambar 4.48 <i>Classification report SVM Kernel RBF</i>	61
Gambar 4.49 Hasil Evaluasi SVM <i>Hyperparameter</i>	62
Gambar 4.50 <i>Confusion matrix SVM Hyperparameter</i>	63
Gambar 4.51 <i>Classification report SVM Hyperparameter</i>	64



DAFTAR RUMUS

Rumus 2.1 Persamaan <i>Training Data</i> [23]	13
Rumus 2.2 Persamaan <i>Hyperlane</i> [23].....	13
Rumus 2.3 Persamaan <i>Margin 1</i> [23]	13
Rumus 2.4 Persamaan <i>Margin 2</i> [23]	13
Rumus 2.5 Rumus Mencari <i>Accuracy</i> [28].....	16
Rumus 2.6 Rumus Mencari <i>Precision</i> [28].....	16
Rumus 2.7 Rumus Mencari <i>Recall</i> [28].....	16
Rumus 2.8 Rumus Mencari <i>f1-score</i> [28].....	17

UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA