

TUGAS AKHIR

**ANALISIS PERBANDINGAN PERFORMA
ALGORITMA *SUPPORT VECTOR MACHINE*
DENGAN *RANDOM FOREST* DALAM KLASIFIKASI
RISIKO KANKER SERVIKS**



CHANDRA AYUNDA APTA SOEMEDHY

19102301

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS INFORMATIKA
INSTITUT TEKNOLOGI TELKOM PURWOKERTO
2023**

TUGAS AKHIR

**ANALISIS PERBANDINGAN PERFORMA
ALGORITMA *SUPPORT VECTOR MACHINE*
DENGAN *RANDOM FOREST* DALAM KLASIFIKASI
RISIKO KANKER SERVIKS**

**COMPARISON ANALYSIS OF THE PERFORMANCE
OF SUPPORT VECTOR MACHINE AND RANDOM
FOREST ALGORITHM IN CERTAIN CANCER RISK
CLASSIFICATION**

Disusun Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana Komputer



CHANDRA AYUNDA APTA SOEMEDHY

19102301

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS INFORMATIKA
INSTITUT TEKNOLOGI TELKOM PURWOKERTO
2023**

LEMBAR PERSETUJUAN PEMBIMBING

ANALISIS PERBANDINGAN PERFORMA ALGORITMA *SUPPORT VECTOR MACHINE* DENGAN *RANDOM FOREST* DALAM KLASIFIKASI RISIKO KANKER SERVIKS

COMPARISON ANALYSIS OF THE PERFORMANCE OF SUPPORT VECTOR MACHINE AND RANDOM FOREST ALGORITHM IN CERTAIN CANCER RISK CLASSIFICATION

Dipersiapkan dan Disusun Oleh

Chandra Ayunda Apta Soemedhy

19102301

Fakultas Informatika

Institut Teknologi Telkom Purwokerto

Pada Tanggal: 20 Januari 2023

Pembimbing I,



(Rima Dias Ramadhani, S.Kom., M.Kom.)

NIDN. 0602039301

Pembimbing II,



(Merlinda Wibowo, S.T., M.Phil.)

NIDN. 0612059203

**ANALISIS PERBANDINGAN PERFORMA
ALGORITMA *SUPPORT VECTOR MACHINE*
DENGAN *RANDOM FOREST* DALAM KLASIFIKASI
RISIKO KANKER SERVIKS**

**COMPARISON ANALYSIS OF THE PERFORMANCE
OF SUPPORT VECTOR MACHINE AND RANDOM
FOREST ALGORITHM IN CERTAIN CANCER RISK
CLASSIFICATION**

Disusun Oleh


**CHANDRA AYUNDA APTA SOEMEDHY
19102301**

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tugas Akhir Pada 31
Januari 2023

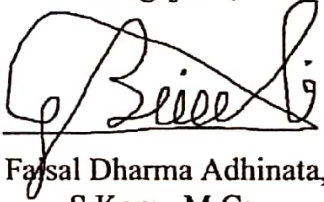
Penguji I,


Ummi Athiyah, S.Kom.,
M.Kom.
NIDN. 0621129001

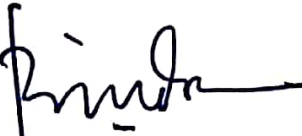
Penguji II,


Dr. Ridwan Pandiya,
S.Si., M.Sc.
NIDN. 0625088202

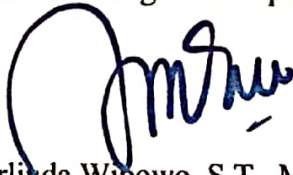
Penguji III,


Faisal Dharma Adhinata,
S.Kom., M.Cs.
NIDN. 0607079301

Pembimbing Utama,


Rima Dias Ramadhani, S.Kom., M.Kom.
NIDN. 0602039301

Pembimbing Pendamping,


Merlinda Wibowo, S.T., M.Phil.
NIDN. 0612059203

Dekan,


Auliya Burhanuddin, S.Si., M.Kom
NIDN. 0630058202

HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertandatangan di bawah ini,

Nama mahasiswa : Chandra Ayunda Apta Soemedhy
NIM : 19102301
Program Studi : S1 Teknik Informatika

Menyatakan bahwa Tugas Akhir dengan judul berikut:

ANALISIS PERBANDINGAN PERFORMA ALGORITME SUPPORT VECTOR MACHINE DAN RANDOM FOREST PADA KLASIFIKASI FAKTOR RISIKO KANKER SERVIKS

Dosen Pembimbing Utama : Rima Dias Ramadhani, S.Kom., M.Kom

Dosen Pembimbing Pendamping : Merlinda Wibowo, S.T., M.Phil

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Institut Teknologi Telkom Purwokerto maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan, dan penelitian Saya Sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing.
3. Dalam Karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggungjawab Saya, bukan tanggungjawab Institut Teknologi Telkom Purwokerto.
5. Pernyataan ini Saya buat dengan sesungguhnya, apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka Saya bersedia menerima Sanksi Akademik dengan pencabutan gelar yang sudah diperoleh serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Purwokerto, 16 Juni 2022,

Yang Menyatakan,

(Chandra Ayunda Apta Soemedhy)

KATA PENGANTAR

Segala puji dan syukur kepada Allah SWT karena telah memberikan Rahmat dan HidayahNya sehingga penulis mendapatkan kemudahan dalam menyelesaikan penulisan naskah skripsi dengan judul “Analisis Perbandingan Performa Algoritma *Support Vector Machine* dengan *Random Forest* dalam Klasifikasi Risiko Kanker Serviks” sebagai salah satu syarat menyelesaikan Program Sarjana (S1) Jurusan Teknik Informatika IT Telkom Purwokerto. Penulis menyadari bahwa pengerjaan skripsi dapat berjalan lancar tidak terlepas dari beberapa pihak yang telah memberikan dukungan, bantuan, bimbingan, dan nasehat secara moral maupun material. Penulis ucapkan terimakasih kepada:

1. Ibu Rima Dias Ramadhani, S.Kom., M.Kom, selaku pembimbing utama dan Ibu Merlinda Wibowo, S.T., M.Phil, selaku pembimbing pendamping dalam penyusunan naskah skripsi.
2. Bapak Anggi, S.T., M.Eng, selaku dosen wali yang telah membantu penulis dalam menyelesaikan studi di jurusan Teknik Informatika Institut Teknoogi Telkom Purwokerto.
3. Kedua orang tua penulis yang telah memberikan kasih sayang, doa, dan nasehat yang luar biasa dalam mendukung penulis menyelesaikan studi S1 Teknik Informatika di Institut Teknoogi Telkom Purwokerto. Tak lupa kepada adik penulis tercinta atas segala doa dan dukungan.
4. Keluarga yang telah memberikan dukungan moril maupun material.
5. Sahabat-sahabat terutama Nora Trivetisia, Dwi Puspa Martiyaningsih, Nawang Anggita Winanti, dan Tri Wulandari Utami selaku rekan penulis yang sudah menemani perkuliahan dari awal semester sampai akhir.

Dalam penyusunan skripsi ini, penulis menyadari masih banyak kekurangan, untuk itu penulis mengharapkan saran dan masukan untuk perbaikan agar skripsi ini dapat selesai dengan maksimal. Semoga skripsi ini dapat membantu dan bermanfaat baik bagi penulis maupun pembaca kedepannya.

DAFTAR ISI

HALAMAN PERSETUJUAN PEMBIMBING	iii
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	vi
DAFTAR TABEL	ix
DAFTAR GAMBAR.....	x
ABSTRAK	xiii
<i>ABSTRACT</i>	xiv
BAB I PENDAHULUAN.....	15
1.1 Latar Belakang	15
1.2 Rumusan Masalah	18
1.3 Pertanyaan Penelitian	18
1.4 Batasan Masalah.....	18
1.5 Tujuan	19
1.6 Manfaat	19
BAB II TINJAUAN PUSTAKA	20
2.1 Penelitian terdahulu	20
2.2 Dasar Teori	24
BAB III METODOLOGI PENELITIAN	38
3.1 Kerangka Berpikir	38
3.2 Subjek dan Objek Penelitian.....	38
3.3 Teknik pengambilan data	39

3.4	Alat dan bahan penelitian	39
3.5	Diagram alur penelitian	39
3.6	<i>Dataset Kanker Serviks (2017)</i>	41
3.7	<i>Preprocessing Data</i>	45
3.8	10-Fold Cross Validation	51
3.9	Alur Algoritma	51
3.10	Pelatihan Data	52
3.11	Analisis Data	52
BAB IV HASIL DAN PEMBAHASAN		59
4.1	Perhitungan Manual / Eksperimen	59
4.2	Hasil dan Pembahasan	104
4.3	Analisa Hasil	112
BAB V KESIMPULAN DAN SARAN		119
5.1	Kesimpulan	119
5.2	Saran	120
DAFTAR PUSTAKA		121

DAFTAR TABEL

Tabel 2. 1 Penelitian terkait kanker dan Algoritma klasifikasi.....	20
Tabel 2. 2 Tabel Persamaan Umum Kernel pada SVM.....	29
Tabel 3. 1 Deskripsi Atribut Dataset.....	41
Tabel 3. 2 Sampel Record Dataset Cervical Cancer (2017).....	44
Tabel 3. 3 Dataset Baru Hasil Preprocessing.....	50
Tabel 3. 4 Hasil Performa Algoritma SVM.....	50
Tabel 3. 5 Hasil Performa Algoritma SVM.....	53
Tabel 3. 6 Hasil Klasifikasi dengan Algoritma SVM.....	54
Tabel 3. 7 Pencarian Maximal Depth Terbaik.....	54
Tabel 3. 8 Hasil Performa Algoritma RF.....	55
Tabel 3. 9 Hasil Klasifikasi dengan Algoritma RF.....	55
Tabel 3. 10 Hasil Perbandingan Algoritma SVM dan RF.....	57
Tabel 4. 1 Proses Pembagian Data.....	67
Tabel 4. 2 Penggantian Nama Atribut.....	69
Tabel 4. 3 Dataset Hasil Normalisasi.....	70
Tabel 4. 4 Nilai W pada tiap atribut dan Nilai b.....	73
Tabel 4. 5 Sampel Data Uji SVM.....	74
Tabel 4. 6 Pengujian Kernel Optimal.....	80
Tabel 4. 7 Jumlah Data Training per Kategori.....	81
Tabel 4. 8 Hasil Perhitungan Entropy dan Information Gain.....	98
Tabel 4. 9 Percobaan $n_{estimators}$ Optimal.....	103
Tabel 4. 10 Perbandingan Hasil Training SVM dan RF.....	114
Tabel 4. 11 Perbandingan Hasil Testing SVM dan RF.....	115

DAFTAR GAMBAR

Gambar 2. 1 Bidang Data Mining	26
Gambar 2. 2 Arsitektur Umum Random Forest	28
Gambar 2. 3 Hyperplane Terbaik yang Memisahkan Kedua Kelas -1 Dan +1	29
Gambar 2. 4 Model K-Fold Cross Validation dengan K=10	32
Gambar 2. 5 Confusion Matrix	34
Gambar 3. 1 Kerangka Berpikir	38
Gambar 3. 2 Alur Penelitian.....	40
Gambar 3. 3 Alur Seleksi Fitur	45
Gambar 3. 4 Alur Handling Missing Value	46
Gambar 3. 5 Hasil Feature Selection Menggunakan WEKA.....	47
Gambar 3. 6 Diagram Alur Algoritma SVM	51
Gambar 3. 7 Diagram Alur Algoritma Random Forest.....	52
Gambar 3. 8 Persebaran Data Kanker Serviks	53
Gambar 3. 9 Tree pada Pengujian dengan Random Forest.....	55
Gambar 3. 10 Perbandingan Algoritma Berdasarkan Akurasi dan Waktu Komputasi	56
Gambar 3. 11 Perbandingan Algoritma Berdasarkan RMSE dan Waktu Komputasi	56
Gambar 3. 12 Pengecekan Outlier	58
Gambar 4. 1 Proses Penggantian nilai “?”	59
Gambar 4. 2 Proses Pengecekan Missing Value	60
Gambar 4. 3 Persentase Missing Value.....	60
Gambar 4. 4 Persebaran Missing Value	61
Gambar 4. 5 Penghapusan Atribut STDs: Time Since First Diagnosis dan STDs: Time Since Last Diagnosis	62
Gambar 4. 6 Korelasi Atribut terhadap Kelas	63
Gambar 4. 7 Penghapusan Atribut tidak Berkorelasi.....	64

Gambar 4. 8 Proses Mean Imputation.....	65
Gambar 4. 9 Proses Listwise Deletion	66
Gambar 4. 10 Pengecekan Ulang Dataset	66
Gambar 4. 11 Pengecekan Outlier(1).....	77
Gambar 4. 12 Pengecekan Outlier(2).....	78
Gambar 4. 13 Pengecekan Outlier(3).....	78
Gambar 4. 14 Proses Splitting Data pada Trainingset SVM.....	79
Gambar 4. 15 Proses Scalling Data SVM	79
Gambar 4. 16 Pencarian Hyperparameter Terbaik SVM.....	80
Gambar 4. 17 Proses Cross Validation SVM.....	81
Gambar 4. 18 Pohon Keputusan Perhitungan Manual RF	100
Gambar 4. 19 Distribusi Data Biopsy	101
Gambar 4. 20 Proses Splitting Data pada Trainingset RF	101
Gambar 4. 21 Proses Scalling Data RF	102
Gambar 4. 22 Pencarian Hyperparameter Terbaik RF.....	102
Gambar 4. 23 Proses Cross Validation RF.....	103
Gambar 4. 24 Proses Training SVM.....	104
Gambar 4. 25 Hasil Klasifikasi Training SVM.....	105
Gambar 4. 26 Evaluasi Training SVM.....	105
Gambar 4. 27 Hasil <i>Error</i> Training SVM.....	106
Gambar 4. 28 Proses Testing SVM.....	106
Gambar 4. 29 Hasil Klasifikasi Testing SVM	107
Gambar 4. 30 Evaluasi Testing SVM	107
Gambar 4. 31 Hasil <i>Error</i> Testing SVM.....	108
Gambar 4. 32 Proses Training RF	108
Gambar 4. 33 Hasil Klasifikasi Training RF	109
Gambar 4. 34 Evaluasi Training RF	109
Gambar 4. 35 Hasil <i>Error</i> Training RF.....	110
Gambar 4. 36 Proses Testing RF	110
Gambar 4. 37 Hasil Klasifikasi Testing RF	111
Gambar 4. 38 Evaluasi Testing RF	111

Gambar 4. 39 Hasil *Error* Testing RF 112