

***SUPPORT VECTOR MACHINE DENGAN RELIEFF, CONSISTENCY-BASED
DAN MEDICAL KNOWLEDGE-BASED FEATURE SELECTION UNTUK
KLASIFIKASI PENYAKIT JANTUNG KORONER***

SKRIPSI

**Diajukan untuk memenuhi salah satu syarat
untuk memperoleh gelar sarjana komputer**

Oleh :

AMIRUDDIN FADLLI

NIM. 2014. 69.04.0005



PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS YUDHARTA PASURUAN

2018

***SUPPORT VECTOR MACHINE DENGAN RELIEFF, CONSISTENCY-BASED DAN MEDICAL KNOWLEDGE-BASED FEATURE SELECTION
UNTUK KLASIFIKASI PENYAKIT JANTUNG KORONER***

SKRIPSI

**Diajukan untuk memenuhi salah satu syarat
untuk memperoleh gelar sarjana komputer**

Oleh :

AMIRUDDIN FADLLI

NIM. 2014. 69.04.0005



PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS YUDHARTA PASURUAN

2018

PERNYATAAN PENULIS

JUDUL : *SUPPORT VECTOR MACHINE DENGAN RELIEFF, CONSISTENCY-BASED DAN MEDICAL KNOWLEDGE-BASED FEATURE SELECTION UNTUK KLASIFIKASI PENYAKIT JANTUNG KORONER*

NAMA : AMIRUDDIN FADLLI
NIM : 2014.69.04.0005

"Saya menyatakan dan bertanggungjawab dengan sebenarnya bahwa Skripsi ini adalah hasil karya saya sendiri kecuali cuplikan dan ringkasan yang masing – masing telah saya jelaskan sumbernya. Jika pada waktu selanjutnya ada pihak lain yang mengklaim bahwa Skripsi ini sebagai karyanya, yang disertai dengan bukti – bukti yang cukup, maka saya bersedia untuk dibatalkan gelar Sarjana Komputer saya beserta segala hak dan kewajiban yang melekat pada gelar tersebut".

Pasuruan, 09 Juli 2018



Amiruddin Fadlli
Penulis

PERSETUJUAN SKRIPSI

JUDUL : *SUPPORT VECTOR MACHINE* DENGAN *RELIEFF*,
CONSISTENCY-BASED DAN *MEDICAL KNOWLEDGE-BASED*
FEATURE SELECTION UNTUK KLASIFIKASI PENYAKIT
JANTUNG KORONER

NAMA : AMIRUDDIN FADLLI
NIM : 2014.69.04.0005

Skripsi ini telah diperiksa dan disetujui
Pasuruan, 09 Juli 2018

Menyetujui,



Ketua Program Studi,

Lukman Hakim, M.Kom
NIP.Y. 06911.09.071

Pembimbing,

Muhammad Imron Rosadi, M.Kom
NIP.Y. 06902.13.121

PENGESAHAN SKRIPSI

JUDUL : *SUPPORT VECTOR MACHINE DENGAN RELIEFF, CONSISTENCY-BASED DAN MEDICAL KNOWLEDGE-BASED FEATURE SELECTION UNTUK KLASIFIKASI PENYAKIT JANTUNG KORONER*

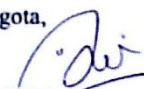
NAMA : AMIRUDDIN FADLLI
NIM : 2014.69.04.0005

Skripsi ini telah diujikan dan dipertahankan didepan Dewan Penguji pada Sidang Skripsi tanggal 11 Juli 2018. Menurut pandangan kami, Skripsi ini memadai dari segi kualitas untuk tujuan penganugerahan gelar Sarjana Komputer (S.Kom)
Pasuruan, 18 Juli 2018

Ketua Penguji,


Moch. Lutfi, M.Kom
NIP.Y. 069.16.03.004

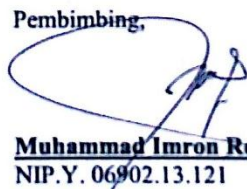
Anggota,


Arief Tri Arsanto, M.M
NIP.Y. 069.02.01.004

Dekan Fakultas Teknik,


Misbach Munir, ST., MT
NIP.Y. 069.02.01.015

Pembimbing,


Muhammad Imron Rosadi, M.Kom
NIP.Y. 069.02.13.121

ABSTRACT

Coronary heart disease is an obstruction of blood intake caused by a buildup of plaque in the coronary arteries that supply oxygen to the heart muscle. Expensive and risky coronary angiography testing to diagnose this disease leads researcher to apply Support Vector Machine for classification task and feature selection techniques such as ReliefF, Consistency-based and Medical Knowledge-based Feature Selection to overcome features in medical data that are often irrelevant and redundant. RLF and CNS reduce 13 features to 3 and 11 while increasing accuracy from 81.1111% to 84.4444%, 83.3333%. MFS reduces 13 features to 8 while decreasing accuracy from 81.1111% to 72.2222%, but combining RLF into MFS increases accuracy from 72.2222% to 83.3333%.

Keyword: Coronary heart disease, Support Vector Machine, Feature selection

Support Vector Machine Dengan ReliefF, Consistency-based Dan Medical-based Feature Selection Untuk Klasifikasi Penyakit Jantung Koroner

Amiruddin Fadlli

Program Studi Teknik Informatika, Universitas Yudharta Pasuruan

ABSTRAK

Terhambatnya asupan darah yang disebabkan oleh bertumpuknya plak di dalam arteri koroner yang mensuplai oksigen ke otot jantung disebut penyakit jantung koroner. Mahal serta beresikonya uji *coronary angiography* untuk mendiagnosis penyakit ini menuntun peneliti untuk dikembangkan sebuah metode komputerisasi yang mampu mendiagnosis keberadaan PJK.

Support Vector Machine memiliki kemampuan klasifikasi yang baik dalam menangani prediksi PJK. Namun, karna data medis seringkali mengandung sejumlah fitur yang tidak relevan dan redundan. Maka teknik seleksi fitur *ReliefF*, *Consistency-based* dan *Medical Knowledge-based Feature Selection* dapat digunakan untuk menyeleksi fitur yang relevan pada data medis.

RLF, CNS dan MFS mampu mengurangi jumlah atribut yang awalnya berjumlah 13 buah menjadi 3, 11 dan 8 buah. RLF dan CNS berhasil meningkatkan akurasi klasifikasi SVM menjadi 84.4444% dan 83.3333% dari yang pada saat menggunakan semua fitur akurasinya hanya sebesar 81.1111%. Meski MFS menurunkan akurasi dari 81.1111% menjadi 72.2222%, namun gabungan MFS dengan RLF dapat meningkatkan akurasi menjadi 83.3333% dari yang pada saat menggunakan MFS saja hanya sebesar 72.2222%.

Kata kunci: Penyakit jantung koroner, *Support Vector Machine*, *ReliefF*, *Consistency-based Feature Selection*, *Medical Knowledge-based Feature Selection*

KATA PENGANTAR

Segala puji syukur penulis panjatkan kehadirat Allah SWT yang telah memberikan rahmat, hidayah, serta inayahnya sehingga penulis dapat menyelesaikan Laporan Skripsi ini yang berjudul “*SUPPORT VECTOR MACHINE DENGAN RELIEFF, CONSISTENCY-BASED DAN MEDICAL KNOWLEDGE-BASED FEATURE SELECTION* UNTUK KLASIFIKASI PENYAKIT JANTUNG KORONER” dengan tepat waktu.

Skripsi ini merupakan salah satu syarat wajib yang harus ditempuh dalam program studi Teknik Informatika. Selain untuk menuntaskan program studi yang telah ditempuh oleh penulis, Skripsi ini juga mampu memberikan manfaat bagi penulis baik dari segi akademik maupun pengalaman yang tidak dapat penulis temukan di bangku kuliah.

Penyusunan Laporan Skripsi ini tidak dapat terselesaikan tanpa adanya bantuan dan bimbingan dari semua pihak, karenanya penulis mengucapkan beribu-ribu terima kasih kepada:

1. Kedua orang tua Ayah, Ibu dan Kakak saya yang senantiasa mendoakan dan selalu tak henti-hentinya memberi dorongan semangat dan moral atau material demi tercapinya cita-cita penulis.
2. KH. Sholeh Bahrudidn, selaku Pembina Yayasan Darut Taqwa yang selalu memberikan doa restunya.
3. Dr. Syaifullah, M.HI selaku Rektor Universitas Yudharta Pasuruan.
4. Misbachul Munir, ST, MT selaku Dekan Universitas Yudharta Pasuruan.
5. Lukman Hakim, M.Kom selaku Dosen Pembimbing serta Ketua Prodi Teknik Informatika yang telah senantiasa meluangkan tempat dan waktu untuk membimbing penulis dalam menyelesaikan Laporan Skripsi ini.
6. Teman-teman Teknik Informatika angkatan 2014, dan semua pihak yang telah membantu penulis yang tidak dapat disebutkan satu persatu atas bantuan dan dorongan semangat dalam menyelesaikan Laporan Skripsi ini.

Penulis menyadari akan segala kekurangan dari penulisan ini. Oleh karena itu, kritik dan saran yang membangun dari pembaca sangat penulis harapkan demi sempurnanya laporan ini.

Semoga laporan ini dapat dijadikan bahan referensi Skripsi di tahun-tahun berikutnya dalam merencanakan, melaksanakan, serta mengevaluasi kegiatan pembelajaran guna, meningkatkan mutu dan efektifitas pembelajaran.

Pasuruan, 09 Juli 2018

Penulis

DAFTAR ISI

	Halaman
HALAMAN SAMPUL.....	i
PERNYATAAN PENULIS	ii
PERSETUJUAN SKRIPSI.....	iii
PENGESAHAN SKRIPSI.....	iv
ABSTRACT	v
ABSTRAK	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN.....	Error! Bookmark not defined.
1.1 Latar Belakang	Error! Bookmark not defined.
1.2 Rumusan Masalah	Error! Bookmark not defined.
1.3 Tujuan Penelitian	Error! Bookmark not defined.
1.4 Manfaat Penelitian	Error! Bookmark not defined.
1.5 Batasan Masalah.....	Error! Bookmark not defined.
1.6 Sistematika Pembahasan	Error! Bookmark not defined.
BAB II TINJAUAN PUSTAKA.....	Error! Bookmark not defined.
2.1 Penelitian Terkait	Error! Bookmark not defined.
2.1.1 <i>Benchmarking Attribute Selection Techniques for Discrete Class Data Mining</i>	Error! Bookmark not defined.
2.1.2 <i>Comparative Study of Data Mining Classification Methods in Cardiovascular Disease Prediction</i>	Error! Bookmark not defined.
2.1.3 <i>Computational Intelligence for Heart Diagnosis: A Medical Knowledge Driven Approach</i>	Error! Bookmark not defined.
2.1.4 <i>Benchmarking Algoritma Pemilihan Atribut pada Klasifikasi Data Mining</i>	Error! Bookmark not defined.
2.1.5 <i>Modifikasi Seleksi Fitur Berbasis Komputer untuk Diagnosis Penyakit Jantung Koroner</i>	Error! Bookmark not defined.
2.1.6 <i>Analysis of Data Mining Techniques for Heart Disease Prediction</i>	Error! Bookmark not defined.
2.2 Landasan Teori	Error! Bookmark not defined.
2.2.1 Penyakit Jantung Koroner ...	Error! Bookmark not defined.
2.2.2 <i>Data Mining</i>	Error! Bookmark not defined.
2.2.3 Klasifikasi	Error! Bookmark not defined.
2.2.4 <i>Support Vector Machine (SVM)</i>	Error! Bookmark not defined.
2.2.5 Teknik Seleksi Fitur	Error! Bookmark not defined.
2.2.6 <i>ReliefF (RLF)</i>	Error! Bookmark not defined.
2.2.7 <i>Consistency-based Subset Evaluation (CNS)</i>	Error! Bookmark not defined.

2.2.8	<i>Medical Knowledge-based Motivated Feature Selection (MFS)</i>	Error! Bookmark not defined.
2.3	Kerangka Pemikiran	Error! Bookmark not defined.
2.4	Hipotesis Penelitian	Error! Bookmark not defined.
BAB III	METODE PENELITIAN	Error! Bookmark not defined.
3.1	Desain Penelitian	Error! Bookmark not defined.
3.2	Tahapan Penelitian	Error! Bookmark not defined.
3.2.1	Tahap <i>Business Understanding</i>	Error! Bookmark not defined.
3.2.2	Tahap <i>Data Understanding</i>	Error! Bookmark not defined.
3.2.3	Tahap <i>Data Preparation</i>	Error! Bookmark not defined.
3.2.4	Tahap <i>Modelling</i>	Error! Bookmark not defined.
3.2.5	Tahap <i>Evaluation</i>	Error! Bookmark not defined.
3.3	Spesifikasi Alat	Error! Bookmark not defined.
BAB IV	HASIL DAN PEMBAHASAN	Error! Bookmark not defined.
4.1	Preparasi Data	Error! Bookmark not defined.
4.2	Seleksi Fitur <i>ReliefF</i>	Error! Bookmark not defined.
4.3	Seleksi Fitur CNS	Error! Bookmark not defined.
4.4	Seleksi Fitur MFS	Error! Bookmark not defined.
4.5	Seleksi Fitur MFS+ <i>ReliefF</i>	Error! Bookmark not defined.
4.6	Perbandingan Nilai Evaluasi Hasil Klasifikasi	Error! Bookmark not defined.
BAB V	PENUTUP	Error! Bookmark not defined.
5.1	Kesimpulan	Error! Bookmark not defined.
5.2	Saran	Error! Bookmark not defined.
DAFTAR PUSTAKA		
LAMPIRAN		

DAFTAR TABEL

	Halaman
Tabel 2.1 Menang dan Kalah Akurasi Seleksi Fitur dengan <i>Naïve Bayes</i>	7
Tabel 2.2 Menang dan Kalah Akurasi Seleksi Fitur dengan <i>C4.5</i>	8
Tabel 2.3 Menang dan Kalah Waktu yang Dibutuhkan untuk Memilih Fitur untuk <i>Naïve Bayes</i>	8
Tabel 2.4 Menang dan Kalah Waktu yang Dibutuhkan untuk Memilih Fitur untuk <i>C.45</i>	9
Tabel 2.5 Perbandingan Model Data Mining.....	9
Tabel 2.6 Perbandingan <i>Error Rate</i> , <i>True Positive Rate</i> , dan <i>False Positive Rate</i>	10
Tabel 2.7 Perbandingan Akurasi MFS dan CFS pada Dataset H-0.....	10
Tabel 2.8 Perbandingan Akurasi MFS dan CFS pada Dataset Sick-1.....	11
Tabel 2.9 Perbandingan Akurasi MFS dan CFS pada Dataset Sick-2.....	11
Tabel 2.10 Perbandingan Akurasi MFS dan CFS pada Dataset Sick-3.....	12
Tabel 2.11 Perbandingan Akurasi MFS dan CFS pada Dataset Sick-4.....	12
Tabel 2.12 Ranking Teknik Seleksi Fitur pada Algoritma NB.....	13
Tabel 2.13 Ranking Teknik Seleksi Fitur pada Algoritma C4.5.....	14
Tabel 2.14 Perbandingan Akurasi MFS dan CFSII pada Dataset H-0.....	15
Tabel 2.15 Perbandingan Akurasi MFS dan CFSII pada Dataset Sick-1.....	15
Tabel 2.16 Perbandingan Akurasi MFS dan CFSII pada Dataset Sick-2.....	16
Tabel 2.17 Perbandingan Akurasi MFS dan CFSII pada Dataset Sick-3.....	16
Tabel 2.18 Perbandingan Akurasi MFS dan CFSII pada Dataset Sick-4.....	17
Tabel 2.19 Pengukuran Akurasi Algoritma Klasifikasi.....	17
Tabel 2.20 Pengukuran Area Under the ROC Curve (AUC).....	18
Tabel 3.1 Atribut Data Set Penyakit Jantung.....	38
Tabel 3.2 Pembagian Data Set.....	40
Tabel 4.1 Sampel Data Set.....	43
Tabel 4.2 Peringkat dan bobot atribut dengan <i>ReliefF</i>	44
Tabel 4.3 Hasil <i>evaluation SVM</i> dengan <i>ReliefF</i>	45
Tabel 4.4 Data Latih SVM- <i>ReliefF</i>	46
Tabel 4.5 Data Uji SVM- <i>ReliefF</i>	47

Tabel 4.6 <i>Confusion Matrix SVM-ReliefF</i>	47
Tabel 4.7 <i>Subset</i> keluaran dari CNS	48
Tabel 4.8 Data Latih SVM-CNS	49
Tabel 4.9 Data Uji SVM-CNS	50
Tabel 4.10 <i>Confusion Matrix SVM-CNS</i>	50
Tabel 4.11 Data Latih SVM-MFS.....	51
Tabel 4.12 Data Uji SVM-MFS	52
Tabel 4.13 <i>Confusion Matrix SVM-MFS</i>	53
Tabel 4.14 Atribut hasil seleksi MFS+RLF	54
Tabel 4.15 Data Latih SVM-MFS+RLF	54
Tabel 4.16 Data Uji SVM-MFS+RLF	55
Tabel 4.17 <i>Confusion Matrix SVM-MFS+RLF</i>	56
Tabel 4.18 Perbandingan Nilai Evaluasi	56

DAFTAR GAMBAR

	Halaman
Gambar 2.1 Blok Diagram Model Klasifikasi	21
Gambar 2.2 <i>Hyperplane</i> yang optimal dengan <i>margin</i> yang maksimal.....	23
Gambar 2.3 Komponen dari SMO	27
Gambar 2.4 Proses <i>Forward Selection Search</i>	30
Gambar 2.5 Algoritma <i>ReliefF</i>	31
Gambar 2.6 Kerangka pemikiran	35
Gambar 3.1 Pemodelan penelitian.	39
Gambar 3.2 Proses evaluasi <i>10-fold cross validation</i>	41

DAFTAR LAMPIRAN

Lampiran 1 Kartu Bimbingan

Lampiran 2 Kartu Seminar

Lampiran 3 *Curriculum Vitae*

Lampiran 4 Daftar Hadir Seminar