

APLIKASI DATA *MINING*
UNTUK MENDETEKSI PENYAKIT HEPATITIS
MENGGUNAKAN METODE
SUPPORT VECTOR MACHINE (SVM) POLYNOMIAL
(STUDI KASUS : DATA PASIEN HATI INDIA)



SKRIPSI

Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Komputer
Pada Departemen Ilmu Komputer/Informatika

Disusun Oleh :

LUSY OCTORIA SITANGGANG

24010314120022

DEPARTEMEN ILMU KOMPUTER/INFORMATIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO

2018

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini:

Nama : Lusy O. Sitanggang

NIM : 24010314120022

Judul : Aplikasi Data Mining Untuk Mendeteksi Penyakit Hepatitis Menggunakan Metode *Support Vector Machine (SVM) Polynomial* (Studi Kasus: Data Pasien Hati India).

Dengan ini saya menyatakan bahwa dalam tugas akhir/ skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang tertulis diacu dalam naskah ini dan disebutkan di dalam daftar pustaka.

Semarang, 04 September 2018



Lusy O. Sitanggang

24010314120022

HALAMAN PENGESAHAN

Judul : Aplikasi Data *Mining* Untuk Mendeteksi Penyakit Hepatitis Menggunakan Metode *Support Vector Machine (SVM) Polynomial*. (Studi Kasus: Data Pasien Hati India)

Nama : Lusy O. Sitanggang

NIM : 24010314120022

Telah diujikan pada sidang tugas akhir pada tanggal dan dinyatakan lulus pada tanggal 04 September 2018.

Semarang, 28 September 2018

Mengetahui,

Ketua Departemen Ilmu
Komputer/Informatika
FSM UNDIP



Dr. Reno Kusumaningrum, S.Si, M.Kom
NIP. 198104202005012001

Panitia Penguji Tugas Akhir

Ketua,

Drs. Eko Adi Sarwoko, M.Kom
NIP. 196511071992031003

HALAMAN PENGESAHAN

Judul : Aplikasi Data *Mining* Untuk Mendeteksi Penyakit Hepatitis menggunakan Metode *Support Vector Machine (SVM) Polynomial* (Studi Kasus : Data Pasien Hati India.

Nama : Lusy O. Sitanggang

NIM : 24010314120022

Telah diujikan pada sidang tugas akhir pada tanggal 04 September 2018

Semarang, 28 September 2018

Pembimbing



Nurdin Baktiar, S.Si, M.T

NIP. 197907202003121002

ABSTRAK

Perkembangan teknologi informasi saat ini telah merambah ke berbagai sektor termasuk sektor kesehatan. Dalam sektor kesehatan, perkembangan ilmu kedokteran mengalami kemajuan yang sangat pesat yang ditandai dengan ditemukannya penyakit-penyakit baru yang belum teridentifikasi sebelumnya. Salah satu penyakit yang berkembang saat ini yaitu penyakit pada organ hati. Salah satunya adalah penyakit Hepatitis. Diagnosa awal penyakit ini setelah memperhatikan gejala adalah melakukan tes fungsi hati yang biasa disebut LFT (*Liver Function Test*). Dengan beberapa atribut dari hasil pemeriksaan LFT tersebut akan mudah digunakan untuk menganalisis penyakit tersebut. Salah satu teknologi kecerdasan buatan yang dapat digunakan untuk menganalisis penyakit tersebut adalah *machine learning*. *Machine Learning* telah banyak digunakan dalam bidang medis yaitu untuk menganalisa dataset medis. Salah satu metode *machine learning* adalah *Support Vector Machine (SVM)*. Ciri dari metode ini adalah menemukan fungsi pemisah (*klasifier*) yang optimal yang bisa memisahkan dua set data dari dua kelas yang berbeda. Data yang digunakan pada penelitian ini didapat dari UCI (*Universitas California Invene*) *Machine Learning Repository* yang berjumlah 579 data pasien. Dalam dataset tersebut, terdapat 11 atribut yang akan digunakan untuk mendiagnosis penyakit dengan menggunakan metode *support vector machine polynomial*. Dengan menggunakan *Cross validation*, menggunakan pengujian pada 10 atribut data Pasien Hati India memiliki rata-rata nilai akurasi sebesar 87.65%.

Kata Kunci : *Support Vector Machine, Hepatitis, Machine Learning, Cross Validation, polynomial*

ABSTRACT

The development of information technology has now penetrated into various sectors including the health sector. In the health sector, the development of medical science has progressed very rapidly marked by the discovery of new diseases that have not been identified before. One of the diseases that develops now is the disease in the liver. One of them is hepatitis. The initial diagnosis of this disease after paying attention to symptoms is to perform a liver function test which is commonly called the LFT (Liver Function Test). With some attributes from the results of the LFT examination it will be easy to use to analyze the disease. One of the artificial intelligence technologies that can be used to analyze the disease is machine learning. Machine Learning has been widely used in the medical field to analyze medical datasets. One machine learning method is Support Vector Machine (SVM). The characteristic of this method is finding the optimal separator function (classifier) that can separate two data sets from two different classes. The data used in this research was obtained from UCI (University of California Invene) Machine Learning Repository, which amounted to 579 patient data. In the dataset, there are 11 attributes that will be used to diagnose the disease using the support vector machine polynomial method. By using the dataset, there are 11 attributes that will be used to diagnose the disease using the support vector machine polynomial method. By using Cross validation, using tests on 10 attributes of Indian Heart Patients has an average accuracy value of 87.65%.

Keywords : *Support Vector Machine, Hepatitis, Machine Learning, Cross Validation, polynomial*

KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Tuhan Yang Mahas Esa atas rahmat dan anugerah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Aplikasi Data Mining untuk Mendeteksi Penyakit Hepatitis menggunakan Metode *Support Vector Machine (SVM) Polynomial* (Studi Kasus: Data Pasien Hati India)”.

Dalam penyusunan laporan ini penulis mendapat banyak bimbingan, bantuan, dan dukungan dari berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih kepada:

1. Ibu Prof. Dr. Widowati, S.Si., M.Si., selaku Dekan Fakultas Sains dan Matematika Universitas Diponegoro.
2. Ibu Dr. Retno Kusumaningrum, S.Si, M.Kom, selaku Kepala Departemen Ilmu Komputer/ Informatika Fakultas Sains dan Matematika Universitas Diponegoro.
3. Bapak Helmie Arif Wibawa, S.Si, M.Cs., selaku Koordinator Skripsi Departemen Ilmu Komputer/ Informatika Fakultas Sains dan Matematika Universitas Diponegoro.
4. Bapak Nurdin Bahtiar, S.Si, M.T, selaku Dosen Pembimbing Skripsi yang telah membantu dalam membimbing dan mengarahkan penulis hingga selesainya skripsi ini.
5. Semua pihak yang telah membantu hingga selesainya skripsi ini yang tidak dapat penulis sebutkan satu persatu.

Penulis menyadari bahwa dalam laporan ini masih terdapat banyak kekurangan, baik dalam penyampaian materi maupun isi dari materi tersebut. Hal ini disebabkan oleh keterbatasan kemampuan dan pengetahuan dari penulis. Oleh karena itu, kritik dan saran yang bersifat membangun sangat penulis harapkan. Semoga skripsi ini dapat bermanfaat bagi penulis dan juga pembaca pada umumnya.

Semarang, 13 September 2018

Penulis

DAFTAR ISI

HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	Error! Bookmark not defined.
HALAMAN PENGESAHAN	ii
HALAMAN PENGESAHAN	iii
ABSTRAK	iv
ABSTRACT	v
KATA PENGANTAR.....	vi
DAFTAR ISI	vii
DAFTAR GAMBAR.....	ix
DAFTAR TABEL	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan dan Manfaat	3
1.4 Ruang Lingkup.....	3
1.5 Sistematika Penulisan	4
BAB II LANDASAN TEORI	5
2.1 Penyakit Hepatitis	5
2.2 Data Mining.....	8
2.3 Support Vector Machine (SVM).....	10
2.4 Metode Kernel SVM.....	14
2.5 Cross Validation	17
2.6 Confusion Matrix	18
2.7 Model Proses Perangkat Lunak.....	20
2.8 Pemodelan Fungsional	21
2.9 Bahasa Pemrograman PHP	23
2.10 Sistem Manajemen Basis Data MySQL.....	24
2.11 Black Box Testing.....	24
2.12 Flowchart	24

BAB III METODOLOGI PENELITIAN	28
3.1 Pengumpulan Data	28
3.2 <i>Preprocessing</i> Data	29
3.2.1 Pembersihan Data (<i>Data Cleaning</i>).....	30
3.2.2 Seleksi Data (<i>Data Selection</i>).....	30
3.2.3 Transformasi Data (<i>Data Transformation</i>)	31
3.3 Proses Klasifikasi <i>Polynomial</i>	33
3.4 Analisa Kebutuhan Aplikasi	52
3.4.1 Deskripsi Umum	52
3.4.2 Karakteristik Pengguna	52
3.4.3 Kebutuhan Fungsional	52
3.4.4 Pemodelan Fungsional	53
3.5 Desain Aplikasi	55
3.5.1 Deskripsi Aplikasi	56
3.5.2 Desain Antarmuka.....	56
3.6.3 Desain Fungsi.....	62
BAB IV IMPLEMENTASI DAN PEMBAHASAN.....	65
4.1 <i>Implementation</i>	65
4.1.1 Implementasi Basis Data.....	65
4.1.2 Implementasi Source Code	65
4.1.3 Implementasi Antarmuka.....	66
4.2 Pengujian Klasifikasi <i>Polynomial</i>	73
4.3 <i>Testing</i>	78
4.3.1 Rencana Pengujian.....	78
4.3.2 Hasil Pengujian	78
BAB V KESIMPULAN DAN SARAN	80
5.1 Kesimpulan.....	80
5.2 Saran.....	80
DAFTAR PUSTAKA.....	81
LAMPIRAN-LAMPIRAN	83

DAFTAR GAMBAR

Gambar 2.1. Tahap - Tahap <i>Data Mining</i> . (Han & Kamber, 2006)	9
Gambar 2.2 Prinsip SVM	11
Gambar 2.3 Usaha SVM dalam menemukan hyperplane terbaik secara nonlinear	11
Gambar 2.4 Ilustrasi Pembagian Dataset.....	18
Gambar 2.5 Model Waterfall.....	21
Gambar 2.6 Conceptual Flowchart.....	26
Gambar 2.7 Detail Flowchart	27
Gambar 3.1 Garis Besar Permasalahan Penelitian	28
Gambar 3.2 Data Pasien Hati India	30
Gambar 3.3 Hasil Normalisasi data Pasien Hati India	33
Gambar 3.4 DCD Aplikasi Deteksi Penyakit.....	54
Gambar 3.5 DFD Level 1 Aplikasi Deteksi Penyakit	55
Gambar 3.6 Desain Halaman Awal	57
Gambar 3.7 Desain Halaman Info	57
Gambar 3.8 Desain Halaman Bantuan	58
Gambar 3.9 Desain Halaman Pelatihan.....	59
Gambar 3.10 Desain Halaman Tambah Data Pelatihan	60
Gambar 3.11 Desain Halaman Konsultasi Hasil	60
Gambar 3.12 Desain Halaman Identifikasi Penyakit.....	61
Gambar 3.13 Desain Halaman pengujian.....	62
Gambar 3.14 Desain Halaman Hasil Pengujian	62
Gambar 3.15 Desain Fungsi Pelatihan	63
Gambar 3.16 Desain Fungsi Hasil Konsultasi.....	63
Gambar 3.17 Desain Fungsi Pengujian	64
Gambar 4.1 Implementasi Tabel Pasien.....	66
Gambar 4.2 Implementasi Antarmuka Halaman Awal	67
Gambar 4.3 Implementasi Antarmuka Halaman Info	68
Gambar 4.4 Implementasi Antarmuka Halaman Bantuan.....	68
Gambar 4.5 Implementasi Antarmuka Halaman Pelatihan	69
Gambar 4.6 Implementasi Antarmuka Halaman Tambah Data	70
Gambar 4.7 Implementasi Antarmuka Halaman Konsultasi Hasil.....	71

Gambar 4.8 Implementasi Antarmuka Halaman Identifikasi Penyakit.....	72
Gambar 4.9 Implementasi Antarmuka Halaman Pengujian	72
Gambar 4.10 Implementasi Antarmuka Halaman Identifikasi penyakit	73

DAFTAR TABEL

Tabel 2.1 Kernel yang umum dipakai dalam SVM.....	15
Tabel 2.2 Contoh Confusion Matrix dengan 2 Kelas	18
Tabel 2.3 Penomoran Level pada DFD	22
Tabel 2.4 Simbol-simbol pada DFD.....	22
Tabel 2.5 Simbol Penghubung	25
Tabel 2.6 Simbol Proses	25
Tabel 2.7 Simbol Input-Output.....	26
Tabel 3.1 Keterangan Atribut Data	29
Tabel 3.2 Tabel contoh data uji <i>Polynomial</i> 10 atribut	33
Tabel 3.3 Tabel Transpose Data <i>Polynomial</i> 10 atribut	34
Tabel 3.4 Tabel Perbandingan Data <i>Polynomial</i> 10 atribut.....	34
Tabel 3.5 Hasil Perhitungan Kernel <i>Polynomial</i> 10 atribut.....	43
Tabel 3.6 Matriks Data <i>Polynomial</i> 10 atribut	47
Tabel 3.7 Hasil Perhitungan Nilai <i>Error Polynomial</i> 10 atribut	47
Tabel 3.8 Hasil Perhitungan delta alpha <i>polynomial</i> 10 atribut	49
Tabel 3.9 Hasil perhitungan alpha baru <i>Polynomial</i> 10 atribut.....	49
Tabel 3.10 Tabel data uji <i>Polynomial</i> 10 atribut	50
Tabel 3.11 Hasil perhitungan <i>dot product</i> data uji dengan data latih 10 atribut	51
Tabel 3.12 Karakteristik Pengguna Aplikasi Deteksi Penyakit.....	52
Tabel 3.13 Kebutuhan Fungsional Aplikasi	53
Tabel 4.1 Hasil Pengujian <i>Black Box Testing</i>	78

BAB I

PENDAHULUAN

Bab ini membahas mengenai latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup serta sistematika penulisan tugas akhir mengenai Aplikasi Data Mining untuk Mendeteksi Penyakit Hepatitis menggunakan metode *Support Vector Machine (SVM) Polynomial* (Studi kasus : Data Pasien Hati India).

1.1 Latar Belakang

Pada era globalisasi saat ini, Teknologi Informasi turut berkembang sejalan dengan perkembangan peradaban manusia. Perkembangan teknologi informasi ini meliputi perkembangan infrastruktur, seperti *hardware*, *software*, teknologi penyimpanan data (*storage*), dan teknologi. Saat ini perkembangan teknologi informasi telah merambah ke berbagai sektor termasuk kesehatan.

Dalam sektor kesehatan, perkembangan ilmu kedokteran mengalami kemajuan yang sangat pesat yang ditandai dengan ditemukannya penyakit-penyakit baru yang belum teridentifikasi sebelumnya. Salah satu penyakit yang berkembang saat ini yaitu penyakit pada organ hati. Salah satunya adalah penyakit hepatitis.

Hepatitis adalah kelainan hati berupa peradangan sel hati. Penyakit ini disebabkan adanya gangguan atau kerusakan membran hati. Ada dua faktor penyebabnya yaitu faktor infeksi dan faktor non infeksi. Faktor penyebab infeksi antara lain virus hepatitis dan bakteri yang terdapat pada hati (Lika, 2007).

Peradangan ini ditandai dengan meningkatnya kadar enzim hati. Diagnosis awal penyakit ini setelah memperhatikan gejala adalah melakukan tes fungsi hati yang biasa disebut LFT (*Liver Function Test*) (Green, 2005). Dengan beberapa atribut dari hasil pemeriksaan tersebut akan mudah digunakan untuk menganalisis penyakit tersebut. Salah satu teknologi kecerdasan buatan yang dapat digunakan untuk menganalisis penyakit tersebut adalah *machine learning*.

Machine learning telah banyak digunakan dalam bidang medis untuk menganalisa dataset medis (Munawarah, 2016). Salah satu metode *machine learning*

adalah *Support Vector Machine* (SVM). *Support Vector Machine* (SVM) adalah sistem pembelajaran yang menggunakan ruang hipotesis berupa fungsi-fungsi *linier* dalam sebuah ruang fitur (*feature space*) berdimensi tinggi, dilatih dengan algoritma pembelajaran yang didasarkan pada teori optimasi dengan mengimplementasikan *learning bias* yang berasal dari teori pembelajaran statistik (Sembiring 2007). Ciri dari metode ini adalah menemukan fungsi pemisah (*klasifier*) yang optimal yang bisa memisahkan dua set data dari dua kelas yang berbeda.

Teori yang mendasari SVM sendiri sudah berkembang sejak 1960-an, tetapi baru diperkenalkan oleh Vapnik, Boser, dan Guyon pada tahun 1992 dan sejak itu SVM berkembang dengan pesat. SVM merupakan salah satu teknik yang relative baru dibandingkan dengan teknik lain, tetapi memiliki performansi yang lebih baik di berbagai bidang aplikasi seperti *bioinformatics*, pengenalan tulisan tangan, klasifikasi teks dan lain sebagainya (Ramana, 2011). SVM juga dikenal sebagai teknik pembelajaran mesin (*machine learning*) paling mutakhir setelah pembelajaran mesin sebelumnya yang dikenal sebagai *Neural Network* (NN)(Kerami, 2004). SVM berada dalam satu kelas dengan ANN dalam hal fungsi dan kondisi permasalahan yang bisa diselesaikan. Keduanya masuk dalam kelas *supervised learning*. Terbukti dalam banyak implementasi, SVM memberi hasil yang lebih baik dari ANN, terutama dalam hal solusi yang dicapai. SVM selalu mencapai solusi yang sama untuk setiap *running* (Santoso, 2007).

Dengan menggunakan metode klasifikasi *Support Vector Machine* (SVM) akan dibangun sebuah aplikasi yang dapat mendiagnosis seseorang terkena penyakit Hepatitis dengan berbasis *web*. Saat ini, penggunaan *Web* sangat mudah diakses dimana saja dan kapan pun. Sehingga, aplikasi ini dirancang untuk memberi kemudahan kepada masyarakat atau tenaga medis untuk mengetahui diagnosis penyakit Hepatitis dengan mengakses *web* ini. Aplikasi ini dibangun dengan menggunakan bahasa pemrograman PHP. Bahasa pemrograman ini sangat sederhana untuk dipahami dalam mengelola situs *web* dibandingkan dengan bahasa pemrograman lainnya.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana membangun suatu aplikasi yang dapat mendiagnosis seseorang terkena penyakit Hepatitis menggunakan metode klasifikasi *Support Vector Machine (SVM) Polynomial*?
2. Bagaimana mengetahui kinerja dari metode Klasifikasi *SVM Polynomial* pada studi kasus data Pasien Hati India ?

1.3 Tujuan dan Manfaat

Tujuan penelitian ini adalah:

1. Menghasilkan sebuah aplikasi yang dapat mendiagnosis seseorang terkena penyakit Hepatitis menggunakan metode klasifikasi *Support Vector Machine (SVM) Polynomial*.
2. Mengetahui kinerja dari metode klasifikasi *SVM Polynomial* pada studi kasus data Pasien Hati India.

Manfaat penelitian ini adalah :

1. Memberikan kemudahan kepada pengguna untuk mengetahui seseorang terdiagnosis penyakit hepatitis menggunakan metode *Support Vector Machine (SVM) Polynomial*.
2. Memberikan kemudahan kepada pengguna untuk menunjukkan kinerja dari klasifikasi metode *SVM Polynomial* pada studi kasus data Pasien Hati India.

1.4 Ruang Lingkup

Ruang lingkup aplikasi yang dapat mendiagnosa seseorang terkena penyakit Hepatitis menggunakan metode *Support Vector Machine (SVM) Polynomial* pada studi kasus data Pasien Hati India adalah sebagai berikut :

- a. Input data pelatihan dan data pengujian berupa dataset yang diambil dari *UCI Learning Dataset*
- b. Pengujian yang dilakukan menggunakan *K-Fold Cross Validation*.

- c. Pembuatan aplikasi ini menggunakan bahasa pemrograman PHP.

1.5 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam tugas akhir ini terbagi menjadi beberapa pokok bahasan, yaitu :

BAB I PENDAHULUAN

Bab ini membahas tentang latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, dan sistematika penulisan dalam penyusunan skripsi yang berjudul Aplikasi Data Mining Untuk Mendeteksi Penyakit Hepatitis Menggunakan Metode *Support Vector Machine (SVM) Polynomial* (Studi Kasus: Data Pasien Hati India).

BAB II LANDASAN TEORI

Bab ini berisi penjelasan mengenai teori-teori dalam penyusunan skripsi yang berjudul Aplikasi Data Mining Untuk Mendeteksi Penyakit Hepatitis Menggunakan Metode *Support Vector Machine (SVM) Polynomial* (Studi Kasus: Data Pasien Hati India).

BAB III METODOLOGI PENELITIAN

Bab ini membahas mengenai tahapan yang dilakukan dalam penelitian skripsi ini. Tahapan tersebut membahas metodologi penelitian untuk pengembangan aplikasi seperti gambaran umum aplikasi, pengumpulan data, preprocessing data, proses klasifikasi, pengujian klasifikasi, metode pengembangan aplikasi, fase *requirement, analysis, dan design*.

BAB IV IMPLEMENTASI DAN PEMBAHASAN

Bab ini membahas mengenai fase *implementation, dan testing* pada Aplikasi Deteksi Penyakit Hepatitis.

BAB V PENUTUP

Bab ini membahas mengenai kesimpulan dari uraian yang telah dijabarkan pada bab-bab sebelumnya dan saran untuk pengembangan peneliti lebih lanjut.