

**KLASIFIKASI KANKER PAYUDARA MENGGUNAKAN SUPPORT
VECTOR MACHINE (SVM) DAN FORWARD SELECTION**

TESIS

Oleh:

LELI ANA HARAHAHAP

NIM. 184211081



**PROGRAM STUDI MAGISTER TEKNOLOGI INFORMASI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER**

MIKROSKIL

MEDAN

2020

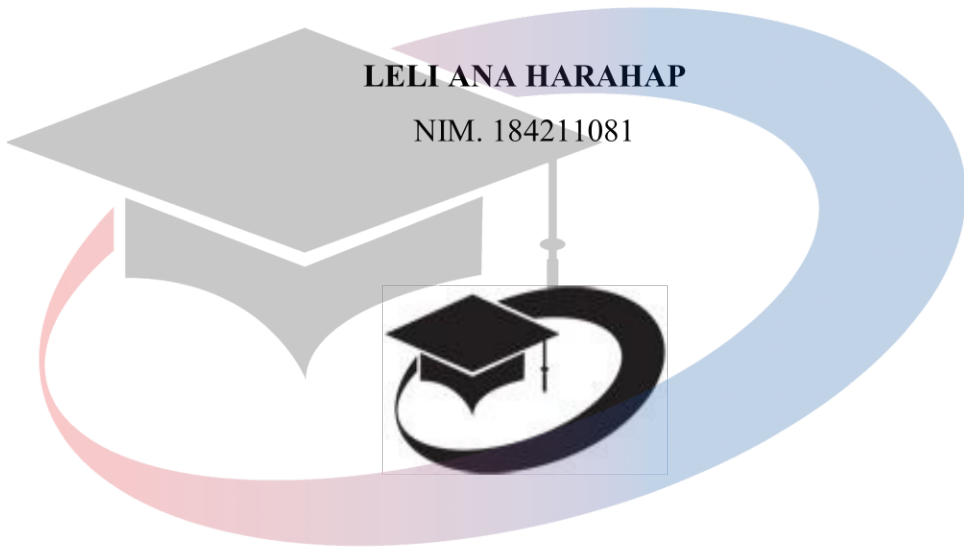
**BREAST CANCER CLASSIFICATION USING SUPPORT
VECTOR MACHINE (SVM) AND FORWARD SELECTION**

THESIS

Oleh:

LELI ANA HARAHAHAP

NIM. 184211081



**UNIVERSITAS
INFORMATION TECHNOLOGY DEPARTEMENT
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
MIKROSKIL
MEDAN
2020**

LEMBAR PENGESAHAN

KLASIFIKASI KANKER PAYUDARA MENGGUNAKAN *SUPPORT VECTOR MACHINE (SVM)* DAN *FORWARD SELECTION*

TESIS

Diajukan untuk Melengkapi Persyaratan Guna
Mendapatkan Gelar Magister
Program Studi Magister Teknologi Informasi

Oleh :

LELIANA HARAHAH

NIM. 184211081

Disetujui Oleh:

Dosen Pembimbing I.

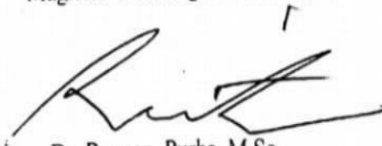


Ir. Erwin Setiawan Panjaitan.MMSI,Ph.D

Medan, 30 Januari 2020

Diketahui dan Disahkan Oleh:

Ketua Program Studi
Magister Teknologi Informasi.



Dr. Ronsen Purba, M.Sc.

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa karena atas berkat dan rahmat-Nya, penulis dapat menyelesaikan Tesis yang berjudul “Klasifikasi Kanker Payudara Menggunakan Support Vector machine dan Forward Selection”. Tesis ini dibuat untuk melengkapi persyaratan kurikulum pada Program Studi Magister Teknologi Informasi Strata Dua, STMIK Mikroskil Medan. Penulis menyampaikan banyak terima kasih kepada:

1. Bapak Ir. Erwin Setiawan Panjaitan, MMSI., Ph.D., selaku Dosen Pembimbing yang telah membimbing penulis dalam proses penyelesaian tesis ini.
2. Bapak Muhammad Fermi Pasha, B.Sc., M.Sc., Ph.D., selaku Dosen Pendamping Pembimbing yang telah membimbing penulis dalam proses penyelesaian tesis ini.
3. Bapak Dr. Pahala Sirait, S.T., M.Kom., selaku Ketua STMIK Mikroskil Medan.
4. Bapak Dr. Ronsen Purba, M.Sc., selaku Ketua Program Studi Magister Teknologi Informasi STMIK Mikroskil Medan.
5. Bapak atau Ibu Dosen STMIK Mikroskil Medan yang telah membantu proses penulisan tesis ini.
6. Anggota keluarga, teman, saudara dan semua pihak yang terus memberikan dukungan penuh kepada penulis selama proses penulisan tesis ini.

Penulis menyadari bahwa masih terdapat banyak kekurangan yang ada. Oleh sebab itu, kritik dan saran yang bersifat membangun akan sangat diterima. Akhir kata, semoga Tesis ini dapat bermanfaat bagi masyarakat. Terima kasih.

Medan, 30 Januari 2020

Penulis

LEMBAR PERNYATAAN

Saya yang membuat pernyataan ini adalah mahasiswa Jurusan/Program Studi S-2 Magister Teknologi Informasi STMIK Mikroskil Medan dengan identitas mahasiswa sebagai berikut:

Nama : Leliana Harahap
Nim : 184211081
Peminatan : Teknologi Informasi

Saya telah melaksanakan penelitian dan penulisan Tesis dengan judul “KLASIFIKASI KANKER PAYUDARA MENGGUNAKAN *SUPPORT VECTOR MACHINE* DAN *FORWARD SELECTION*”, dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan Tesis tersebut merupakan hasil karya saya sendiri (tidak meminta orang lain untuk mengerjakannya) dan semua sumber, baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Bila dikemudian hari ternyata terbukti bahwa bukan saya yang mengerjakannya, maka saya bersedia dikenakan sanksi yang telah ditetapkan oleh STMIK Mikroskil Medan, yakni pencabutan ijazah yang telah saya terima dan ijazah tersebut dinyatakan tidak sah.

Selain itu, demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada STMIK Mikroskil Medan Hak Bebas Royalti Non-eksklusif (Non-exclusive Royalty Free Right) atas Tesis saya beserta perangkat yang ada (jika diperlukan). Dengan hak ini, STMIK Mikroskil Medan berhak menyimpan, mengalihmedia/formatkan, mengelolah dalam bentuk pangkalan data (database), merawat dan mempublikasikan Tesis saya secara keseluruhan atau hanya elektronik, selama tetap mencantumkan nama saya sebagai penulis/pencipta dan pemilik hak cipta. Menyatakan juga bahwa saya akan mempertahankan hak eksklusif saya untuk menggunakan seluruh atau sebagian isi Tesis saya guna pengembangan karya di masa depan, misalnya dalam bentuk artikel, buku, ataupun perangkat lunak.

Demikian pernyataan ini saya perbuat dengan sungguh-sungguh, dalam keadaan sadar dan tanpa ada tekanan dari pihak manapun.

Medan, 30 Januari 2020

Saya yang membuat pernyataan,

Leliana Harahap

ABSTRAK

Kanker payudara mewakili sekitar 12% dari semua kasus kanker baru dan 25% dari semua kanker pada wanita. Deteksi dini dan klasifikasi kanker sangat penting untuk menyelamatkan nyawa seseorang. Salah satu penyakit mengerikan yang mempengaruhi wanita adalah kanker payudara dan itu menjadi perhatian utama di bidang medis. Kanker payudara muncul dari jaringan sel-sel payudara. Penyebab kanker payudara adalah multi faktorial dan melibatkan riwayat keluarga, obesitas, hormon, terapi radiasi, dan bahkan faktor reproduksi. Setiap tahun, satu juta wanita baru didiagnosis menderita kanker payudara, menurut laporan organisasi kesehatan dunia setengah dari mereka akan mati, karena biasanya terlambat ketika dokter mendeteksi kanker. *Machine Learning* adalah cabang ilmu pengetahuan yang mengimplementasikan algoritma matematika ke dalam pemrograman komputer untuk mengidentifikasi pola data dan meningkatkan kinerja secara iteratif. Aplikasi pembelajaran mesin telah memecahkan banyak masalah seperti, prediksi pasien kanker dan prediksi kebangkrutan perusahaan. *Forward selection* adalah jenis regresi bertahap yang dimulai dengan model kosong. Dalam seleksi *forward*, variabel pertama yang dipilih untuk entri ke dalam model yang sudah dibangun adalah yang memiliki korelasi terbesar dengan variabel dependen. Setelah variabel telah dipilih kemudian dievaluasi berdasarkan kriteria tertentu. Jika variabel yang dipilih pertama memenuhi kriteria untuk dimasukkan, maka pemilihan maju berlanjut. Prosedur berhenti, ketika tidak ada variabel lain yang tersisa yang memenuhi kriteria entri dan menambahkan variabel satu per satu. Tingkat akurasi pada Support Vector Machine dipengaruhi oleh beberapa faktor antara lain perbandingan antara jumlah data latih dan data uji yang disesuaikan dengan k-fold validation-nya. Pada perbandingan data latih dan data uji, akurasi yang dihasilkan mencapai 97,68% dengan jumlah komposisi data latih sebanyak 345 data (50%) dan data uji sebanyak 345 data (50%). Pada pengujian yang dilakukan, diperoleh akurasi *Support Vector Machine* dan *Forward Selection* sebesar 97,68%.

Kata Kunci : Kanker Payudara, Klasifikasi, Support Vector Machine, Forward Selection

ABSTRACT

Breast cancer represents about 12% of all new cancer cases and 25% of all cancers in women. Early detection and classification of cancer is very important to save a person's life. One terrible disease that affects women is breast cancer and it is a major concern in the medical field. Breast cancer arises from tissue of breast cells. The cause of breast cancer is multifactorial and involves family history, obesity, hormones, radiation therapy, and even reproductive factors. Every year, one million new women are diagnosed with breast cancer, according to a report by the World Health Organization half of them will die, because it is usually too late when doctors detect cancer. Machine Learning is a branch of science that implements mathematical algorithms into computer programming to identify data patterns and improve performance iteratively. Machine learning applications have solved many problems such as cancer patient predictions and company bankruptcy predictions. Forward selection is a type of stepwise regression that starts with an empty model. In the forward selection, the first variable selected for entry into the model that has been built is the one that has the greatest correlation with the dependent variable. After the variables have been selected then evaluated based on certain criteria. If the first selected variable meets the criteria for inclusion, the selection continues. The procedure stops, when there are no other variables left that meet the entry criteria and add variables one by one. The accuracy of the Support Vector Machine is influenced by several factors including a comparison between the amount of training data and test data adjusted for the k-fold validation. In the comparison of training data and test data, the resulting accuracy reached 97.68% with a total composition of training data of 345 data (50%) and test data of 345 data (50%). In the tests carried out, obtained an accuracy of Support Vector Machine and Forward Selection of 97.68%.

Keywords: Breast Cancer, Classification, Support Vector Machine, Forward Selection

DAFTAR ISI

Abstrak	i
Abstrack	ii
Kata Pengantar	iii
Daftar Isi.....	iv
Daftar Gambar.....	vi
Daftar Tabel	vii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Masalah Penelitian	4
1.3 Tujuan dan Manfaat Penelitian	4
1.4 Ruang Lingkup/Pembatasan Masalah.....	4
1.5 Metodologi Penelitian	5
1.6 Sistematika Penulisan	5
BAB II KAJIAN LITERATUR.....	7
2.1 Kanker Payudara.....	7
2.2 Data Mining	8
2.3 Klasifikasi	12
2.4 Support Vector Machine.....	13
2.5 Pre Pocessing	18
2.6 Normalisasi	18
2.6.1 Tahapan Normalisasi.....	19
2.6.2 Konsep Normalisasi	20
2.7 K-Fold Cross Validation.....	21
2.8 Feature Selection.....	22

BAB III METODOLOGI PENELITIAN	25
3.1 Analisis Masalah.....	25
3.2 Metode Penelitian	25
3.3 Dataset Yang digunakan	27
3.4 Kerangka Kerja	28
3.5 Pre Processing Data	30
3.6 Support Vector Machine	30
3.7 Perangkat dan Kebutuhan Dalam Penelitian	30
BAB IV HASIL DAN PEMBAHASAN	32
4.1 Hasil	32
4.1.1 Pembagian Data.....	32
4.1.2 Normalisasi Data.....	33
4.1.3 Penentuan Data Uji dan Data Latih	38
4.1.3 Hasil Pengujian dan Analisis	39
4.2 Pembahasan.....	41
4.2.1 Hasil Normalisasi.....	42
4.2.2 Hasil Klasifikasi.....	43
BAB V KESIMPULAN DAN SARAN	30
5.1 Kesimpulan	45
5.2 Saran	45
Daftar Pustaka.....	48

DAFTAR GAMBAR

Gambar 2.1 <i>Data mining</i> merupakan inti dari Proses penemuan Pengetahuan	9
Gambar 2.2 Proses data mining secara umum	12
Gambar 2.3 Memisahkan hyperplane dan hyperplane pemisah yang optimal	14
Gambar 2.4 Hyperplane pemisah optimal yang digeneralisasi.....	17
Gambar 2.5 Fungsi Φ memetakan data ke ruang vektor yang berdimensi lebih tinggi	17
Gambar 2.6 Tahapan Normalisasi.....	20
Gambar 2.7 Algoritma Forward Selection.....	24
Gambar 3.1 Metodologi Penelitian	27
Gambar 3.2 Flowchart Forward Selection	29
Gambar 3.3 Flowchart Kerangka Kerja Yang diusulkan.....	31
Gambar 4.1 Grafik hasil akurasi untuk pengujian dengan variasi nilai k-fold	46

UNIVERSITAS
MIKROSKIL

DAFTAR TABEL

Tabel 3.1 Confusion Matrix	25
Tabel 4.1 Informasi Dataset.....	34
Tabel 4.2 Data sebelum Normalisasi	35
Tabel 4.3 Data setelah Normalisasi.....	39
Tabel 4.4 Hasil akurasi untuk pengujian dengan variasi nilai k-fold.....	43
Tabel 4.4 Hasil Normalisasi Data	35



UNIVERSITAS
MIKROSKIL