

**PENERAPAN DECISION TREE C4.5 SEBAGAI
SELEKSI FITUR DAN SUPPORT VECTOR
MACHINE (SVM) UNTUK DIAGNOSA KANKER
PAYUDARA**

TESIS



**Oleh
PAKARTI RISWANTO
NPM. 1621211054**

**MAGISTER TEKNIK INFORMATIKA
PROGRAM PASCASARJANA
INFORMATICS & BUSINESS INSTITUTE DARMAJAYA
BANDAR LAMPUNG
2019**

**PENERAPAN DECISION TREE C4.5 SEBAGAI
SELEKSI FITUR DAN SUPPORT VECTOR
MACHINE (SVM) UNTUK DIAGNOSA KANKER
PAYUDARA**

TESIS

Diajukan Sebagai Salah Satu Syarat untuk Mencapai Gelar
MAGISTER TEKNIK INFORMATIKA
Pada Program Studi Teknik Informatika
IIB Darmajaya Bandar Lampung



Oleh
PAKARTI RISWANTO
NPM. 1621211054

**MAGISTER TEKNIK INFORMATIKA
PROGRAM PASCASARJANA
INFORMATICS & BUSINESS INSTITUTE DARMAJAYA
BANDAR LAMPUNG
2019**

PERNYATAAN KEASLIAN LAPORAN TESIS

Saya yang bertandatangan dibawah ini adalah :

Nama : PAKARTI RISWANTO
NPM : 1621211054
Program Studi : Magister Teknik Informatika
Judul Tesis : PENERAPAN DECISION TREE C4.5 SEBAGAI SELEKSI FITUR DAN SUPPORT VECTOR MACHINE (SVM) UNTUK DIAGNOSA KANKER PAYUDARA

Dengan ini menyatakan bahwa :

1. Tesis ini adalah asli dan bukan plagiat sebagai atau keseluruhan, serta belum pernah diajukan untuk mendapatkan gelar akademik dimanapun.
2. Karya tulis ini adalah murni gagasan, rumusan, dan penelitian saya sendiri melalui proses pembimbingan dan pengujian.
3. Semua kutipan, petikan dari karya orang lain saya tuliskan namanya dengan jelas serta saya cantumkan dalam daftar pustaka.
4. Pernyataan ini saya buat dengan sebenar-benarnya, jika terdapat penyimpangan di kemudian hari, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh, serta sanksi lainnya sesuai dengan norma dan aturan yang berlaku.

Bandar lampung, 27 Maret 2019



Pakarti Riswanto
NPM. 1621211054

LEMBAR PERSETUJUAN TESIS

Judul Tesis : **PENERAPAN DECISION TREE C4.5
SEBAGAI SELEKSI FITUR DAN
SUPPORT VECTOR MACHINE (SVM)
UNTUK DIAGNOSA KANKER PAYUDARA**

Nama Mahasiswa : **Pakarti Riswanto**

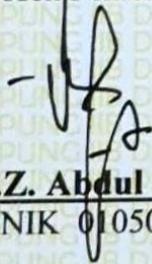
Nomor Pokok Mahasiswa : 1621211054

Program Studi : **Magister Teknik Informatika**

Telah diperiksa dan disetujui untuk diajukan dalam ujian sidang tesis guna memperoleh gelar Magister Teknik Informatika pada program pascasarjana IIB Darmajaya

Bandar Lampung, 27 Maret 2019

Menyetujui,
Dosen Pembimbing



Dr. R.Z. Abdul Aziz, M.T
NIK 01050904

Mengetahui,
Kep. Prodi Magister Teknik Informatika



Abdul Darmawan, S.T., M.T.I
NIK 01150305

LEMBAR PENGESAHAN TESIS

Pada hari Rabu, tanggal 27 Maret 2019, telah diselenggarakan ujian sidang Tesis dengan judul **PENERAPAN DECISION TREE C4.5 SEBAGAI SELEKSI FITUR DAN SUPPORT VECTOR MACHINE (SVM) UNTUK DIAGNOSA KANKER PAYUDARA**. Untuk memenuhi persyaratan akademik guna memperoleh gelar Magister Teknik Informatika pada Program Pascasarjana Informatics and Business Institute Darmajaya, bagi mahasiswa:

Nama Mahasiswa : **Pakarti Riswanto**

Nomor Pokok Mahasiswa: 1621211054

Program Studi : Magister Teknik Informatika

Telah dinyatakan Lulus oleh Dewan Penguji yang terdiri dari :

Nama

Tim Penguji

Tandatangan

1. **Dr. R.Z. Abdul Aziz, M.T**

Pembimbing

2. **Dr. Suhendro Yusuf Irianto**

Ketua Sidang

3. **Dr. Joko Triloka**

Anggota

Dekan Pascasarjana
IIB Darmajaya,

Sriyanto, S.Kom., M.M, Ph.D

NIK 00210800

DAFTAR RIWAYAT HIDUP

Data Pribadi

Nama : Pakarti Riswanto
Tempat, Tanggal Lahir : Candimas, 13 Januari 1989
Jenis Kelamin : Laki-Laki
Agama : Islam
Kewarganegaraan : Indonesia
Alamat : Jalan Way Tabuhan Rt 02/Rw .01, Desa
Kalibening Raya, Kecamatan Abung Selatan,
Kabupaten Lampung Utara.
E-mail : tutiriswanto@gmail.com

Riwayat Pendidikan

- SD Negeri 6 Candimas Lulus Tahun 2001
- SMP Negeri 1 Abung Selatan Lulus Tahun 2004
- SMK Negeri 1 Kotabumi Lulus Tahun 2007
- D3 AMIK DCC Bandarlampung Lulus Tahun 2009
- S1 STMIK Tunas Bangsa Jurusan Sistem Informasi Lulus Tahun 2014

Pengalaman Pekerjaan

- Kolektor di PT. Indomobil Finance Indonesia (2010 - 2011)
- Admin Kantor di PT. Arta Boga Cemerlang (2011 – 2013)
- Akuntan di PT. Andini Agro Loka (2013 - 2016)
- Dosen STMIK DCC Kotabumi (2016 – sekarang)
- Guru SMK Negeri 1 Kotabumi (2016 – sekarang)
- Marbot di Masjid Al – Musywaroh Candimas (2016 – sekarang)

MOTTO DAN PERSEMBAHAN

MOTTO

"Keteladanan Itu Lebih Baik daripada Seribu Nasihat "

(Pakarti Riswanto)

PERSEMBAHAN

Alhamdulillah, atas rahmat dan hidayah-Nya, saya dapat menyelesaikan Tesis ini dengan baik. Karya sederhana ini penulis persembahkan untuk:

1. Bapak dan Ibu tercinta, Bapak Sugito Niti Wijoyo (Alm) dan Ibu Parmini yang selalu memberikan do'a dan dukungan, memberi motivasi dan semangat serta memberikan kasih sayang yang besar yang tak mungkin bisa terbalas dengan apapun.
2. Tuti Ningsih, S.Pd Istri tercinta dan Pranaja Ahsan Riswanto Anaku tercinta yang selalu memberikan semangat dan dukungan.
3. Teman Seperjuangan MTI Angkatan 14 B, Mukrim, Sanusi, Rustam, Dewi Sinaga, Andi Kasyadi, Eka Supriyanto, Aditya.
4. Kakaku Mukarson Aribowo, S.Sos, Adiku – adiku Triowali Rosandi, S.Kom,.M.TI dan Rahmad Aan Setian serta mertuaku Bapak Tusam dan Ibu Umiyati. atas dukunganya selama perkuliahan.

**PENERAPAN DECISION TREE C4.5 SEBAGAI SELEKSI FITUR
DAN SUPPORT VECTOR MACHINE (SVM) UNTUK
DIAGNOSA KANKER PAYUDARA**

PAKARTI RISWANTO

ABSTRACT

In the field of medicine, the use of data mining has a quite important and evolutionary role that can change the perspective of doctors, practitioners and health researchers in the process of detecting breast cancer in a patient. There are 2 classification applications in it, namely the process of diagnosing (diagnosing) cancer cells that distinguishes between tumors (benign cancer) or malignant cancer and prognosis (prognosis) to determine the possibility of reappearance of cancer cells in patients who have been operated on in the future. Data mining aims to describe new findings in the dataset and explain a process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and related knowledge from the database.

Classification with data mining can be done using several methods, namely Decision Tree, K-Nearest Neighbor, Naive Bayes, ID3, CART, Linear Discriminant Analysis, etc., which certainly have advantages and disadvantages of each. But in this study, the author focuses on the classification of data mining using the Support Vector Mechine and Deccision Tree algorithms.

This study will analyze the Breast Cancer Wisconsin Original data set obtained from the UCI Machine Learning Repository (repository of research data) to classify breast cancer malignancies. This time the author correlates between the Decision Tree classifier algorithm which has good ability to process large databases as a feature selection, then with a proper and relevant SVM Method used in analyzing and diagnosing breast breast cancer patients because it has accurate results for existing problems and several bases .

Keywords— *Data Mining, diagnosis, Decision Tree, SVM Method*

PENERAPAN DECISION TREE C4.5 SEBAGAI SELEKSI FITUR DAN SUPPORT VECTOR MACHINE (SVM) UNTUK DIAGNOSA KANKER PAYUDARA

PAKARTI RISWANTO

ABSTRAK

Dalam bidang kedokteran, penggunaan data mining mempunyai peranan yang cukup penting dan evolusioner yang dapat mengubah cara pandang para dokter, praktisi dan peneliti kesehatan dalam melakukan proses deteksi penyakit kanker payudara pada seorang pasien. Terdapat 2 aplikasi klasifikasi di dalamnya, yaitu proses diagnose (diagnosis) sel kanker yang membedakan antara tumor (kanker jinak) atau kanker ganas dan proses prognosa (prognosis) untuk mengetahui kemungkinan munculnya kembali sel kanker pada pasien yang telah dioperasi di masa yang akan datang. data mining bertujuan untuk menguraikan temuan baru di dalam dataset dan menjelaskan suatu proses yang menggunakan teknik statistik, matematis, artificial intelligence, dan machine learning untuk melakukan ekstrak dan identifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari database tersebut.

Klasifikasi dengan data mining dapat dilakukan menggunakan beberapa metode yaitu Decision Tree, K-Nearest Neighbor, Naive Bayes, ID3, CART, Linear Discriminant Analytc dan lain sebagainya yang tentunya memiliki kelebihan dan kekurangan masing - masing. Namun pada penelitian kali ini, penulis berfokus pada klasifikasi data mining dengan menggunakan algoritma Support Vector Mechine dan Deccision Tree.

Penelitian ini akan menganalisis Breast Cancer Winconsin Original data set yang diperoleh dari UCI Machine Learning Repository (repositori data penelitian) untuk melakukan klasifikasi keganasan kanker payudara. Kali ini penulis mengkombinasikan antara algoritma Decision Tree classifier yang memiliki kemampuan baik untuk mengolah database yang besar sebagai feature selection kemudian dengan Metode SVM yang layak dan relevan digunakan dalam menganalisis dan mengdiaknosa Pasien Kanker Payudara payudara karena memiliki hasil yang akurat atas permasalahan yang ada dan beberapa landasan.

Kata Kunci— Data Mining, diagnosis, Decision Tree Metode SVM

DAFTAR ISI

HALAMAN JUDUL LUAR	i
HALAMAN JUDUL DALAM.....	ii
HALAMAN PERNYATAAN KEASLIAN LAPORAN TESIS	iii
HALAMAN LEMBAR PERSETUJUAN TESIS	iv
HALAMAN LEMBAR PENGESAHAN SIDANG TESIS	v
HALAMAN MOTTO DAN PERSEMBAHAN	vi
KATA PENGANTAR	vii
ABSTRACT	ix
ABSTRAK.....	x
DAFTAR RIWAYAT HIDUP	xi
DAFTAR ISI	xii
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xv

BAB I PENDAHULUAN

1.1 Latar Belakang.....	1
1.2 Perumusan Masalah.....	3
1.3 Batasan Penelitian	3
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penulisan.....	5

BAB II DAFTAR PUSTAKA

2.1 Data Mining.....	6
2.2 Support Vector Machine (SVM)	11
2.3 Algoritma C4.5	13
2.4 Kanker Payudara	15

2.5 Karsinoma Papiler Invasif	15
2.6 Karsinoma Tubuler	16
2.7 Karsinoma Adenokistik	16
2.8 Karsinoma Apokrin	16
2.9 Klasifikasi Kanker Payudara Menurut UICC	16
2.10 Definisi Confusion Matrix	19
2.11 Penelitian Terkait	22

BAB III METODOLOGI PENELITIAN

3.1 Tempat dan Waktu Penelitian	26
3.2 Alat dan Bahan	26
3.3 Metode Pengumpulan Data	26
3.4 Data Preprocessing	26
3.5 Penerapan Data Mining	27

BAB IV HASIL DAN PEMBAHASAN

4.1 Evaluasi dan Validasi	40
4.2 Data Preprocessing	41
4.3 Penerapan Data Mining	41
4.4 Penerapan C4.5 untuk seleksi fitur	44
4.5 Penerapan metode seleksi fitur pada klasifikasi SVM	52

BAB V KESIMPULAN DAN SARAN

5.1 Kesimpulan	56
5.2 Saran	56

DAFTAR PUSTAKA

DAFTAR RIWAYAT HIDUP

LAMPIRAN

DAFTAR GAMBAR

Gambar 2.1 Tahapan Data Mining	8
Gambar 2.2 Proses Klasifikasi	11
Gambar 2.3 Rumus mencari Entropy Total.....	14
Gambar 2.4 Rumus mencari Entropy Partisi	14
Gambar 2.5 Model Classifier pada Klasifikasi.....	20
Gambar 2.6 Model Classifier pada Klasifikasi	21
Gambar 3.1 Alur Penelitian	25
Gambar 3.2 Tabel Alur Perhitungan SVM.....	36
Gambar 4.3 Block Processing Algoritma SVM	33
Gambar 4.4 Analisis Hasil Akurasi SVM ratio 50%, 50%	42
Gambar 4.5 Analisis Hasil akurasi SVM ratio 70%, 30%	43
Gambar 4.6 Analisis Hasil akurasi SVM ratio 80%, 20%	43
Gambar 4.7 Model Pohon Keputusan pada Algoritma C4.5.....	45
Gambar 4.8 Gambar Root Not/Akar	46
Gambar 4.9 Gambar Note 1.....	47
Gambar 4.10 Gambar Note 2	49
Gambar 4.11 Gambar Tect View pada Pohon Keputusan.....	51
Gambar 4.12 Pohon Keputusan sebagai Seleksi Fitur.....	52
Gambar 4.13 Block Algoritma SVM.....	53
Gambar 4.14 Hasil Akurasi 1	53
Gambar 4.15 Hasil Akurasi 2	54
Gambar 4.16 Hasil Akurasi 3	54

DAFTAR TABEL

Tabel 2. Penelitian Terkait	22
Tabel 3.1 Perhitungan Algoritma C4.5.....	32
Tabel 3.2 Perhitungan Manual Algoritma	38
Tabel 4.1 Hasil Pengujian.....	55

KATA PENGANTAR

Dengan memanjatkan puji dan syukur ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya pada penulis, akhirnya penulis dapat menyelesaikan penyusunan tesis yang berjudul PENERAPAN DECISION TREE C4.5 SEBAGAI SELEKSI FITUR DAN SUPPORT VECTOR MACHINE (SVM) UNTUK DIAGNOSA KANKER PAYUDARA. Tesis ini ditulis dalam rangka memenuhi salah satu persyaratan memperoleh gelar Magister Teknik Informatika di Program Pascasarjana IBI Darmajaya Lampung.

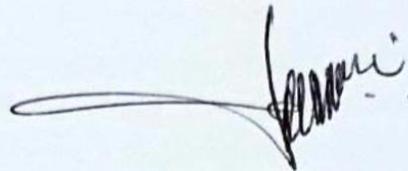
Penulis menyadari bahwa tesis ini dapat diselesaikan berkat dukungan dan bantuan dari berbagai pihak. Oleh karena itu, penulis berterimakasih kepada semua pihak yang secara langsung dan tidak langsung memberikan kontribusi dalam penyelesaian tesis ini. Secara khusus pada kesempatan ini penulis menyampaikan terimakasih kepada:

1. Bapak Ir. Firmansyah YA, M.B.A., M.Sc. selaku Rektor IBI Darmajaya.
2. Bapak Dr. R.Z Abdul Aziz, MT, selaku Direktur Pascasarjana IBI Darmajaya.
3. Bapak Sriyanto, S.Kom., M.M., Ph.D selaku Dekan Pascasarjana IBI Darmajaya sekaligus pembimbing yang telah membimbing dan mengarahkan penulis selama penyusunan tesis ini dari awal hingga selesai.
4. Bapak Abdi Darmawan, S.T., M.T.I, selaku ketua program studi Magister Teknik Informatika IBI Darmajaya.
5. Kedua orang tua tercinta. Bapak Sugito Niti Wijoyo (Alm) dan Ibu Parmini yang telah mendidik dan membesarkan penulis, mendoakan dan memberikan cinta serta kasih sayang yang tak terhingga.
6. Para dosen Pascasarjana IBI Darmajaya Lampung yang telah membimbing dengan ilmu-ilmu yang bermanfaat dan memotivasi penulis dalam penyelesaian tesis ini.

7. Kakaku Mukarson Aribowo, S.Sos, Adiku – adiku Triowali Rosandi, S.Kom,.M.TI dan Rahmad Aan Setiawan serta mertuaku Bapak Tusam dan Ibu Umiyati. atas dukunganya selama perkuliahan.

Penulis berharap tesis ini mudah-mudahan dapat memberikan sumbangsih bagi pendidikan yang selalu menghadapi tantangan seiring dengan tuntutan zaman.

Bandar Lampung, 27 Maret 2019

A handwritten signature in black ink, appearing to read 'Pakarti Riswanto', with a long horizontal flourish extending to the left.

Pakarti Riswanto

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penyakit Kanker merupakan suatu penyakit yang paling ditakuti oleh masyarakat karena proses penyembuhannya yang sulit, efek yang ditimbulkan dan memerlukan biaya yang cukup besar untuk pengobatan dan juga perawatan. Banyak masyarakat beranggapan bahwa kanker sama seperti tumor, padahal kenyataannya tumor yang muncul tidak berarti kanker. Munculnya benjolan yang aneh dari segi bentuk maupun letak tumbuhnya perlu dicurigai, karena itu bisa dipastikan adalah tumor. Tumor jinak dan tumor ganas adalah dua jenis tumor yang berbeda. Semua jenis kanker ganas bisa diklasifikasikan sebagai kanker. Kanker dapat menimpa setiap orang di dunia ini, karena di dalam tubuh setiap manusia memiliki sel kanker yang ada sejak lahir. Namun tidak selalu bahwa setiap orang positif mengidap penyakit kanker, ada beberapa faktor yang menyebabkan meningkatnya resiko kanker pada diri seseorang seperti terpaparnya bahan kimia dari asap rokok atau dari industri (dalam kurun yang waktu lama), penyinaran yang berlebihan (sinar radio aktif, sinar ultraviolet, sinar X dan radiasi), adanya jenis virus yang berhubungan erat dengan perubahan sel normal menjadi sel kanker, hormon dan makanan. Zat yang dihasilkan kelenjar tubuh yang memiliki fungsi untuk mengatur kegiatan organ tubuh dari selaput tertentu biasa disebut dengan faktor hormon. Kenyataannya dari sejumlah penelitian dapat diketahui bahwa pemberian hormon tertentu secara berlebihan tidak baik untuk tubuh dan dapat menyebabkan meningkatnya beberapa jenis resiko penyakit kanker seperti Kanker Panyudara dan beberapa kanker lainnya.

Pada tahun 2012 kasus kanker payudara di Indonesia mencapai kurang lebih 40 kasus untuk setiap 100.000 penduduk dan kejadian itu meningkat pada setiap tahunnya (Menurut Data WHO). Data terbaru yang didapatkan oleh Riset Kesehatan Dasar 2013, kasus kematian pasien penyakit kanker payudara meningkat menjadi kasus kematian tertinggi dengan angka 21,5 pada setiap

100.000 penduduk. Dibandingkan dengan Malaysia, di Indonesia penderita kanker payudara lebih dominan diderita oleh wanita berusia muda dan pada tahap yang lebih lanjut. Angka kematian yang terus meningkat dari kasus Kanker Payudara membutuhkan perhatian khusus untuk upaya pencegahan dini dan penanggulangan dengan diagnosa gejala awal kanker payudara.

Diagnosis dini penyakit kanker payudara atau breast cancer dapat dilakukan dengan metode data mining. Dimana proses data mining bertujuan untuk menguraikan temuan baru di dalam dataset dan menjelaskan suatu proses yang menggunakan teknik statistik, matematis, artificial intelligence, dan machine learning untuk melakukan ekstrak dan identifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari database tersebut. Klasifikasi dengan data mining dapat dilakukan menggunakan beberapa metode yaitu Decision Tree, K-Nearest Neighbor, Naive Bayes, ID3, CART, Linear Discriminant Analytic dan lain sebagainya yang tentunya memiliki kelebihan dan kekurangan masing - masing. Namun pada penelitian kali ini, penulis berfokus pada klasifikasi data mining dengan menggunakan algoritma Support Vector Machine dan Decision Tree sebagai seleksi fitur.

Berdasar dari penelitian terdahulu terkait dengan klasifikasi menggunakan algoritma Support Vector Machine yang diterapkan pada data kesehatan seperti diagnosis gejala penyakit hasil keakuratannya bisa dibilang sangat baik. Pada kasus Breast Cancer Wisconsin atau kanker payudara yang pernah dituangkan dalam jurnal hasil akurasi klasifikasi dengan algoritma Support Vector Machine yaitu untuk partisi data training dan testing 50:50 sebesar 92,48%, partisi data training dan testing 70:30 sebesar 88,89%, dan untuk partisi data training dan testing 80:20 sebesar 94,34% dengan rata-rata ketepatan klasifikasi sebesar 91,9%. Dari kedua hasil tersebut maka penulis ingin melakukan peningkatan akurasi pada klasifikasi keganasan kanker payudara dari data Breast Cancer Wisconsin Original dengan menambahkan seleksi fitur yaitu dengan algoritma Decision Tree C4.5. Perpaduan dua metode yang digunakan akan membantu dalam peningkatan hasil akurasi.

Feature Selection atau seleksi fitur adalah salah satu teknik penting yang sering digunakan dalam preprocessing data pada data mining untuk mempercepat serta memaksimalkan kerja suatu algoritma (dalam penelitian ini algoritma yang dimaksud adalah algoritma Decision Tree classifier. Teknik seleksi fitur adalah memilih subset dari fitur dengan jumlah besar yaitu mengarah pada penghilangan fitur - fitur yang tidak memiliki pengaruh untuk melakukan klasifikasi. Seleksi fitur juga dapat meningkatkan akurasi dari sebuah penelitian.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan di atas, didapatkan suatu perumusan masalah sebagai berikut :

1. Bagaimana cara Decision Tree C4.5 bekerja sebagai seleksi fitur terhadap diagnosa kanker payudara?
2. Bagaimana Algoritma Decision Tree C4.5 sebagai seleksi fitur untuk meningkatkan kinerja Support Vector Machine terhadap diagnosa kanker payudara ?

1.3 Batasan Penelitian

Batasan masalah perlu disertakan dalam penelitian ini sebagai acuan tidak terjadi perluasan masalah yang dibahas dan supaya penelitian ini berfokus pada pokok pembahasan yaitu sebagai berikut :

- a. Data diagnosa didapat dari University California Irvine (UCI) Machine Learning Repository yaitu Breast Cancer Wisconsin Original Data Set yang berjumlah 683 record.
- b. Penerapan algoritma Decision Tree C4.5 sebagai fitur seleksi terhadap dataset diagnosa kanker payudara.
- c. Penerapan peningkatan akurasi dengan algoritma Support Vector Machine terhadap diagnosa kanker payudara.

1.4 Tujuan Penelitian

Penelitian dengan penerapan Support Vector Mechine dan Metode Deccision Tree C4.5 tentang Kanker Payudara ini sebagai berikut :

- a. Meningkatkan akurasi algoritma Support Vector Mechine setelah dilakukan seleksi fitur terhadap dataset kanker payudara.
- b. Membuktikan apakah dengan melakukan seleksi fitur akan menghasilkan tingkat akurasi yang baik.

1.5 Manfaat Penelitian

- a. Manfaat bagi Masyarakat

Sistem seleksi fitur dengan penerapan dengan algoritma Deccision Tree C4.5 akan diperuntukan kepada pengguna data sebagai rekomendasi terhadap dataset analisa kanker payudara.

- b. Manfaat bagi Mahasiswa

1. Sebagai acuan serta pembuka wawasan untuk mahasiswa mengenai metode yang tepat bagi permasalahan diagnosis keganasan kanker payudara yang banya muncul di sekitar kita.
2. Dapat menerapkan ilmu yang telah diajarkan selama masa perkuliahan khususnya dibidang data mining.
3. Memberikan referensi untuk penulisan penelitian dengan menggunakan algoritma yang sama.
4. Menjadi tolak ukur mahasiswa untuk penelitian selanjutnya agar lebih baik lagi dari penelitian sebelumnya.

- c. Manfaat Bagi Akademis Universitas

1. Dapat mengembangkan penelitian dari mahasiswa sehingga bisa meningkatkan minat mahasiswa untuk melakukan penelitian atau pengembangan penelitian yang sudah ada.

2. Dapat menjadi nilai tambah bagi Universitas dalam meningkatkan kualitas Univeristas. Menambah kuantitas dalam jumlah penelitian yang telah dilakukan mahasiswa sehingga dapat berdampak pada akreditasi perguruan tinggi.

1.6 Sistematika Penulisan

Laporan penelitian tesis ini terdiri dari lima bab dengan sistematika penulisan sebagai berikut:

BAB I PENDAHULUAN

Dalam pendahuluan tercantum antara lain latar belakang, ruang lingkup, rumusan masalah, tujuan penelitian, manfaat peneliti andan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini memuat tentang teori-teori yang mendukung penelitian yang akan dilakukan oleh penulis/peneliti. Penelitian yang menggunakan analisis statistik, bab ini memuat kerangka pikir dan hipotesis (bila diperlukan).

BAB III METODOLOGI PENELITIAN

Dalam bab ini berisi objek penelitian, alat dan bahan, metode pengumpulan data, prosedur penelitian, pengukuran variabel dan metode analisis (metode-metode pendekatan penyelesaian permasalahan yang dipakai dan metode analisis data).

BAB IV HASIL DAN PEMBAHASAN

Dalam bab ini disajikan hasil, implementasi, analisis dan pembahasan penelitian. Hasil dan implementasi dapat berupa gambar alat/program dan aplikasinya. Untuk penelitian lapangan hasil dapat berupa data (kualitatif maupun kuantitatif). Analisis dan pembahasan berupa hasil pengolahan data.

BAB V SIMPULAN DAN SARAN

Dalam bab ini disajikan simpulan dan saran dari hasil pembahasan.

BAB II

TINJAUAN PUSTAKA

2.1. Data Mining

2.1.1 Pengertian Data Mining

Data Mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data. (Larose, 2005 dalam Kusriani, 2009: 4)

Data mining juga disebut sebagai serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. (Pramudiono, 2006 dalam Kusriani, 2009: 3)

2.1.2 Pengelompokan Data Mining

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu (Larose, 2005 dalam Kusriani, 2009: 4)

1. Deskripsi

Terkadang peneliti dan analis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih kearah numerik daripada kearah kategori. Model dibangun menggunakan record lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.

3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi. Kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang.

4. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, sedang dan rendah.

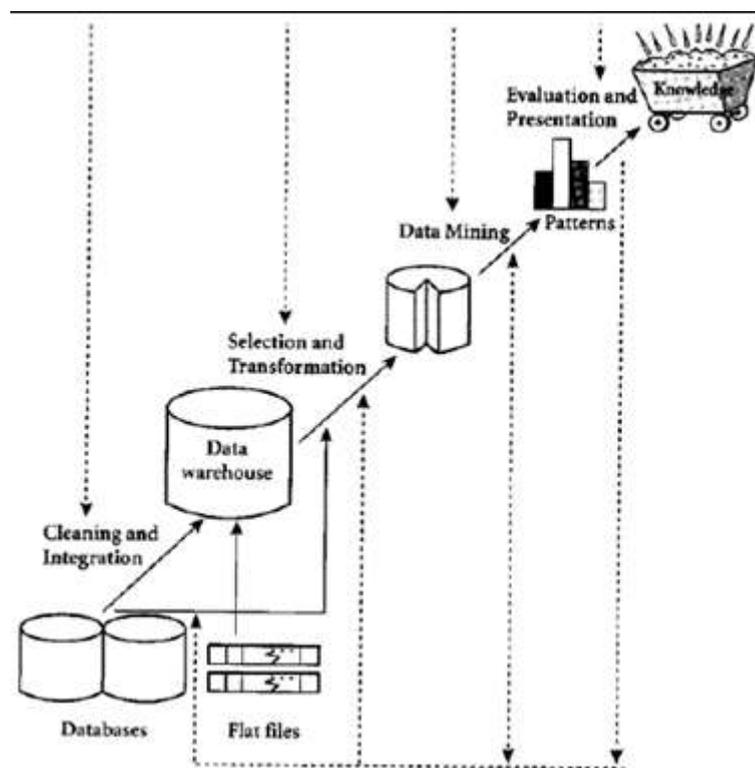
5. Pengklusteran

Pengklusteran merupakan pengelompokkan record, pengamatan atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan record-record dalam kluster lain. Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklusteran. Pengklusteran tidak mencoba melakukan klasifikasi, estimasi atau memprediksi nilai dari variabel target. Akan tetapi, algoritma pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan record dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan record dalam kelompok lain akan bernilai minimal.

6. Asosiasi

Tugas asosiasi adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

2.1.3 Tahap-tahap Data Mining



Gambar 2.1. Tahapan Data Mining

1. Pembersihan data (*Cleaning data*)

Untuk menghilangkan data yang tidak diperlukan, data yang diperoleh dari tahap pengambilan dataset akan disaring untuk menghasilkan data yang benar-benar dibutuhkan. umumnya data tersebut memiliki nilai yang tidak sempurna seperti data yang hilang. Selain itu, ada juga atribut-atribut data yang tidak sesuai dengan pemrosesan data mining yang akan digunakan. Data-data yang tidak relevan itu juga lebih baik dibuang karena keberadaannya bisa mengurangi mutu atau akurasi dari hasil data mining nantinya. Pembersihan data juga akan mempengaruhi performansi dari sistem data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

2. Integrasi data

Data yang akan digunakan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lain-lain. Pada tahap ini hal yang perlu dilakukan untuk lebih detail dan cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan keputusan pada akhirnya. Dalam integrasi data ini juga perlu dilakukan transformasi dan pembersihan data karena seringkali data dari dua database berbeda tidak sama cara penulisannya atau bahkan data yang ada di satu database ternyata tidak ada di database lainnya.

3. Seleksi Data

Data diseleksi untuk menentukan variabel apa saja yang akan diambil agar tidak terjadi kesamaan dan perulangan yang tidak diperlukan dalam pengolahan teknik data mining. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus *market basket analysis*, tidak perlu mengambil namapelanggan, cukup dengan id pelanggan saja.

4. Transformasi data

Pengubahan data menjadi format ekstensi yang sesuai untuk pengolahan dalam *data mining*. Beberapa metode *data mining* membutuhkan format data yang khusus sebelum bisa diproses dalam teknik data mining. Misalnya sebagian metode standar seperti analisis asosiasi dan klustering hanya bisa menerima input data kategorial. Karenanya data berupa angka numerik yang berlanjut perlu dibagi menjadi beberapa interval.

5. Proses *mining*,

Untuk memproses teknik utama saat metode diterapkan agar menemukan pengetahuan berharga, data yang terkumpulkan sesuai prosedur harus diterapkan pada proses mining setelah data melalui tahap transformasi.

6. Evaluasi pola

Tahap ini yaitu mengidentifikasi pola-pola menarik kedalam *knowledge based* yang diidentifikasi. Dalam tahap ini hasil dari teknik *data mining* berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah kajian yang ada sudah memenuhi target yang diinginkan. Jika ternyata hasil yang diperoleh tidak sesuai kajian ada beberapa alternatif dengan mencoba metode *data mining* lain agar lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

7. Presentasi pengetahuan

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami data mining. Karenanya presentasi hasil *datamining*.

2.1.4. Konsep Klasifikasi

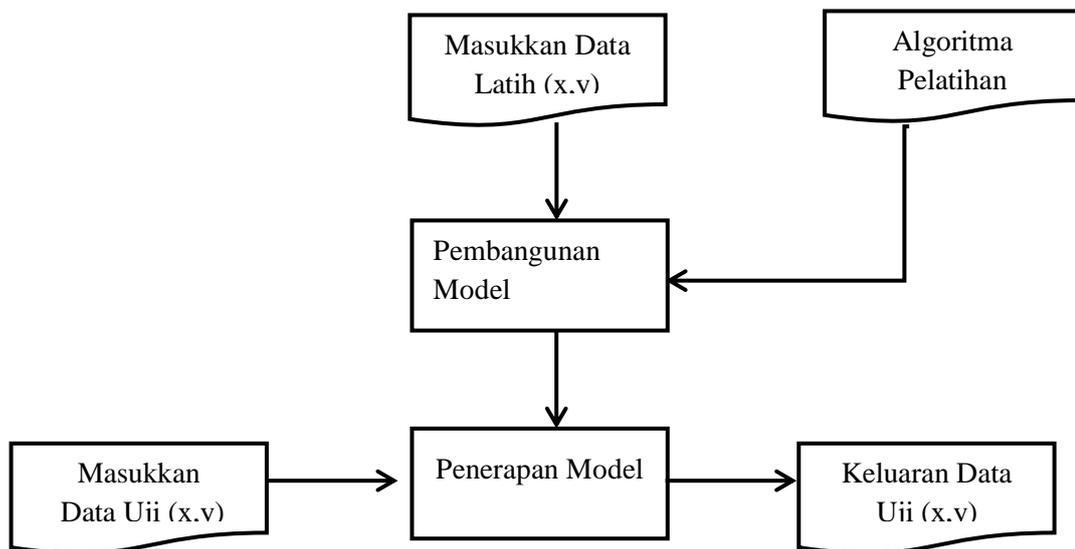
Terdapat dua pekerjaan utama dalam klasifikasi, yaitu:

- (1) pembangunan model sebagai prototype untuk disimpan sebagai memori dan
- (2) penggunaan model tersebut untuk pengenalan / klasifikasi/ prediksi pada suatu objek data lain untuk membedakan objek data tersebut dengan data yang telah disimpan.

Klasifikasi juga didefinisikan sebagai pekerjaan yang melakukan pelatihan / pembelajaran terhadap fungsi target f yang memetakan setiap set atribut (fitur) x ke satu dari sejumlah label kelas y yang tersedia. Pekerjaan tersebut akan menghasilkan suatu model yang kemudian disimpan sebagai memori.

Klasifikasi adalah metode data mining yang dapat digunakan untuk proses pencarian sekumpulan model (fungsi) yang dapat menjelaskan dan membedakan kelas-kelas data atau konsep, yang tujuannya supaya model tersebut dapat digunakan memprediksi objek kelas yang labelnya tidak diketahui atau dapat

memprediksi kecenderungan data-data yang muncul di masa depan. Metode klasifikasi juga bertujuan untuk melakukan pemetaan data ke dalam kelas yang sudah didefinisikan sebelumnya berdasarkan pada nilai atribut data (Han dan Kamber, 2006). Proses klasifikasi tersebut seperti terlihat pada gambar 2.1.



Gambar 2.2 Proses Klasifikasi

2.2 Support Vector Machine (SVM)

Support vector machine (SVM) pertama kali dikenalkan oleh Vapnik pada tahun 1992 pada saat dipresentasikan di Annual Workshop on Computational Learning Theory. Prinsip dasar SVM adalah linier classifier, yaitu kasus klasifikasi yang secara linier dapat dipisahkan. Misalkan diberikan himpunan $X = \{x_1, x_2, x_3, \dots\}$ dinyatakan sebagai kelas positif jika $f_x \geq 0$ dan yang lainnya termasuk ke dalam kelas negatif. SVM melakukan klasifikasi himpunan vektor training berupa set data berpasangan dari dua kelas,

$$(X_i, y_i), X_i \in R^n, y_i \in \{1, -1\}, i = 1, \dots, n,$$

Pemisahan hyperplane dengan bentuk canonical mengikuti constraint atau batasan berikut,

$$y_i [(\mathbf{w}^T \mathbf{x}_i) + b] \geq 1, i = 1, 2, \dots, n$$

Hyperplane yang optimal diperoleh dengan meterbesarkan atau meminimumkan $\varphi(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2$ kemudian permasalahan optimasi ini dapat diselesaikan dengan menggunakan Fungsi Lagrange berikut.

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i [y_i (\mathbf{w}^T \mathbf{x}_i + b) - 1]$$

dimana α_i adalah pengganda fungsi Lagrange. Persamaan (9) merupakan primal space sehingga perlu ditransformasi menjadi dual space agar lebih mudah dan efisien untuk diselesaikan. Sehingga dual problemnya menjadi seperti berikut.

$$\hat{\alpha} = \arg \min_{\alpha} \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (\mathbf{x}_i^T \mathbf{x}_j) - \sum_{i=1}^n \alpha_i \frac{2}{\|\mathbf{w}\|}$$

dengan batasan

$$\alpha_i \geq 0, i = 1, \dots, n \text{ dan } \sum_{i=1}^n \alpha_i y_i = 0$$

Pada kasus non-separabel beberapa data mungkin tidak bisa dikelompokkan secara benar atau terjadi misclassification. Sehingga fungsi obyektif maupun kendala dimodifikasi dengan mengikutsertakan variabel slack $\xi > 0$. Formulasinya menjadi sebagai berikut :

$$\hat{\alpha} = \arg \min_{\alpha} \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i=1}^n \alpha_i$$

Pada kasus separabel dan kasus non-separabel perbedaan keduanya hanya terletak dengan adanya penambahan kendala

Pada kasus separabel dan kasus non-separabel perbedaan keduanya hanya terletak dengan adanya penambahan kendala $0 \leq \alpha_i \leq C$ Pada masalah non-separabel.

Pada kasus non-linier optimasi persamaan (10) menjadi sebagai berikut :

$$\phi(\mathbf{w}, \xi) \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^{\lambda} \xi_i$$

Dengan kendala

$$y_i [(\mathbf{w}^T \mathbf{x}_i) + b] + \xi_i \geq 1, \quad i = 1, 2, \dots, n$$

Dengan batasan

$$0 \leq \alpha_i \leq C, \quad i = 1, \dots, n \text{ dan } \sum_{i=1}^n \alpha_i y_i = 0$$

$K(\mathbf{x}_i, \mathbf{x}_j)$ adalah fungsi kernel yang digunakan untuk mengatasi data non-linier. Berdasarkan langkah langkah yang telah dijelaskan dalam kasus linier, diperoleh fungsi sebagai berikut

$$\begin{aligned} f(\mathbf{x}) &= \text{sign} \left(\sum_{i=1}^n y_i \hat{\alpha}_i \left(\phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \right) + \hat{b} \right) \\ &= \text{sign} \sum_{i=1}^n y_i \hat{\alpha}_i \left(K(\mathbf{x}_f, \mathbf{x}_f) + \hat{b} \right) \end{aligned}$$

dengan fungsi sign semua nilai $f(\mathbf{x}) < 0$ diberi label -1 dan nilai $f(\mathbf{x}) > 0$ diberi label +1.

Fungsi kernel yang biasanya digunakan dalam literatur SVM.

2.3 Algoritma C4.5

1. Kernel Linier : $(\mathbf{x}^T \mathbf{x})$
2. Kernel Polinomial : $(\mathbf{x}^T \mathbf{x} + 1)^p$
3. Kernel RBF : $\exp \left(\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2} \right)$

Secara umum Algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

- a. Pilih atribut sebagai akar
- b. Buat cabang untuk masing-masing nilai
- c. Bagi kasus dalam cabang
- d. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus seperti yang tertera berikut:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Gambar 2.3 Rumus mencari Entropy Total

Keterangan:

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

|Si| : Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

Sebelum mendapatkan nilai Gain adalah dengan mencari nilai Entropi. Entropi digunakan untuk menentukan seberapa informatif sebuah masukan atribut untuk menghasilkan sebuah atribut. Rumus dasar dari Entropi adalah sebagai berikut:

$$Entropy(S) = \sum_{i=1}^n - p_i * \log \log_2 p_i$$

Gambar 2.4. Rumus mencari Entropy Partisi

Keterangan:

S : Himpunan Kasus

n : Jumlah partisi S

pi: Proporsi dari Si terhadap S

1.4 Kanker Payudara

1.4.1 Etiologi Kanker Payudara

Ada 3 pengaruh penting pada kanker payudara:

a. Faktor genetik

Faktor ini berpengaruh dalam peningkatan terjadinya kanker payudara. Pada percobaan tikus dengan galur sensitif kanker, melalui persilangan genetik didapatkan tikus yang terkena kanker. Ada faktor turunan pada suatu keluarga yaitu di lokus kecil kromosom 17q21 pada kanker payudara yang tumbuh di usia muda.

b. Hormon

Kelebihan estrogen endogen atau ketidakseimbangan hormon terlihat sangat jelas pada kanker payudara. Banyak faktor resiko yang dapat disebutkan seperti masa reproduksi yang lama, nulipara dan usia tua yang memili anak pertama. Wanita post menopause dengan tumor ovarium fungsional dapat terkena tumor payudara karena kelebihan hormon estrogen. Suatu penelitian menyebutkan bahwa kelebihan jumlah estrogen di air seni, frekuensi ovulasi dan usia saat menstruasi dihubungkan dengan meningkatnya resiko terkena kanker payudara. Epitel payudara normal memiliki reseptor estrogen dan progesteron. Kedua reseptor ditemukan pada sebagian besar kanker payudara. Berbagai bentuk growth promoters (transforming growth factor- alpha/ epitelial growth factor, PDGF) disekresi oleh sel kanker payudara manusia. 13 dari susunan jaringan yang tidak teratur berisi sel tumor tanpa diferensiasi, sebagian besar sel berbentuk signet-ring.

2.4.2 Karsinoma meduler

Sel berukuran besar berbentuk polygonal/lonjong dengan batas sitoplasma tidak jelas. Diferensiasi dari jenis ini buruk, tetapi memiliki prognosis lebih baik daripada karsinoma duktus infiltratif. Biasanya terdapat infiltrasi limfosit yang nyata dalam jumlah sedang diantara sel kanker, terutama dibagian tepi jaringan kanker.

1.5 Karsinoma papiler invasif

Komponen invasif dari jenis karsinoma ini berbentuk papiler.

1.6 Karsinoma tubuler

Pada karsinoma tubuler, bentuk sel t dan tersusun secara tubuler selapis, dikelilingi oleh stroma fibrous. Jenis ini merupakan karsinoma dengan diferensiasi tinggi.

1.7 Karsinoma adenokistik

Jenis ini merupakan karsinoma invasive dengan karakteristik sel yang berbentuk kribriiformis. Sangat jarang ditemukan pada payudara.

1.8 Karsinoma apokrin

Karsinoma ini didominasi dengan sel yang memiliki sitoplasma eosinofilik, sehingga menyerupai sel apokrin yang mengalami metaplasia. Bentuk karsinoma apokrin dapat ditemukan juga pada jenis karsinoma payudara yang lain.

1.9 Klasifikasi kanker payudara menurut UICC

(International Union Against Cancer) Stadium klinis kanker payudara yang banyak digunakan adalah klasifikasi kanker payudara menurut International Union Against Cancer (UICC) yang berdasarkan besar tumor, kelenjar aksila dan metastasis yang disebut dengan TNM. Berdasarkan gambaran gejala klinik, Klasifikasi TNM menurut International Union Against Cancer (UICC) adalah :

T = Tumor Primer

Tx = Tumor primer tak dapat diperiksa

T0 = Tidak terdapat tumor primer

Tis = Karsinoma in situ

Tis (DCIS) Ductal carcinoma in situ

Tis (LCIS) Lobular carcinoma in situ

Tis (Paget) Paget disease

T1 = Ukuran tumor 2 cm atau kurang

T1a = Ukuran tumor lebih dari 0,1 cm dan tidak lebih dari 0,5 cm

- T1b = Ukuran tumor lebih dari 0,5 cm dan tidak lebih dari 1 cm
- T1c = Ukuran tumor lebih dari 1 cm dan tidak lebih dari 2 cm
- T2 = Ukuran tumor lebih dari 2 cm dan tidak lebih dari 5 cm
- T3 = Ukuran tumor lebih dari 5 cm
- T4 = Semua ukuran tumor dengan ekstensi ke dinding dada atau kulit.
- T4a = Ekstensi ke dinding dada.
- T4b = Edem (termasuk peau d'orange), atau ulserasi kulit payudara, atau satelit nodul pada payudara ipsilateral.
- T4c = T4a dan T4b
- T4d = Inflammatory carcinoma
- N = Limfonodi Regional
- Nx = Limfonodi Regional tak dapat diperiksa
- N0 = Tak ada metastasis di Limfonodi Regional
- N1 = Metastasis di Limfonodi aksila ipsilateral mobile
- N2 = Metastasis di Limfonodi aksila ipsilateral fixed
- N2a = Metastasis di Limfonodi aksila ipsilateral fixed antar limfonodi atau fixed ke struktur jaringan sekitarnya
- N2b = Metastasis di Limfonodi mamaria interna
- N3a = Metastasis di Limfonodi infraklavikuler ipsilateral
- N3b = Metastasis di Limfonodi mamaria interna dan aksila ipsilateral
- N3c = Metastasis di Limfonodi supraklavikuler
- M = Metastasis jauh
- Mx = Metastasis jauh tak dapat diperiksa
- M0 = Tak ada Metastasis jauh
- M1 = Metastasis Jauh

2.9.1 Stadium Kanker Payudara

- Stadium 0 = Tis N0 M0
- Stadium I = T1 N0 M0
- Stadium IIA = T0 N1 M0
T1 N1 M0

	T2 N0 M0
Stadium IIB	= T2 N1 M0
	T3 N0 M0
Stadium III A	= T0 N2 M0
	T1 N2 M0
	T2 N2 M0
	T3 N1,N2 M0
Stadium III B	= T4 N0,N1,N2 M0
Stadium III C	= Setiap T N3 M0
Stadium IV	= Setiap T Setiap N M1

2.9.2 Modalitas terapi kanker payudara.

Modalitas terapi kanker payudara yang standart dan sering dipergunakan bisa berupa:

1. Pembedahan
2. Kemoterapi
3. Hormonal terapi
4. Monoklonal antibody (gen terapi)
5. Immunoterapi

Modalitas terapi kanker payudara merupakan multi modalitas dengan tujuan sebagai berikut :

1. Kuratif , terapi ini untuk “penyembuhan” dapat dinilai dengan lamanya bebas tumor dan lamanya bertahan hidup (long overall survival)
2. Paliatif , terapi ini bertujuan untuk meningkatkan dan mengembalikan kualitas hidup yang dapat dinilai secara fungsional, kosmetik, dan penampilan sosial.
3. Preventif , terapi ini bertujuan untuk mencegah terjadinya kanker payudara.

Adapun pembagian kanker payudara yang ditinjau dari modalitas terapi yang dipergunakan, dibagi menjadi beberapa kategori :

1. Stadium 0 atau lesi in situ non metastasis yaitu ductal carcinoma in situ (DCIS). dan lobular carcinoma in situ (LCIS).

2. Stadium dini invasif (stadium 1 dan beberapa stadium II)
3. Stadium intermediate operabel (stadium II dan IIIA)
4. Stadium in operabel atau stadium lokal (stadium IIIA – IIIC)
5. Stadium lanjut (stadium IV)

Modalitas terapi yang dapat dipergunakan pada kanker payudara adalah :

1. Modalitas terapi pembedahan, dapat berupa :
 - Breast Conservation Surgery (BCS)
 - Simple Mastectomy (SM)
 - Radical Mastectomy (RM)
 - Modified Radical Mastectomy (MRM).

Pembedahan merupakan suatu upaya terapi yang bersifat lokalregional.

2. Radioterapi, merupakan upaya terapi yang bersifat lokoregional.
3. Kemoterapi, bersifat sistemik
4. Hormonal terapi, bersifat sistemik
5. Molekular targeting terapi (biologi terapi).

Penggunaan modalitas terapi ini tergantung pada tujuan pengobatan dan stadium kanker payudara.

2.10 Definisi Confusion Matrix

Confusion matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Rumus ini melakukan perhitungan dengan 4 keluaran, yaitu: recall, precision, accuracy dan error rate.

- Recall adalah proporsi kasus positif yang diidentifikasi dengan benar. Rumus dari recall = $d/(c+d)$
- Precision adalah proporsi kasus dengan hasil positif yang benar. Rumus dari Precision = $d/(b+d)$
- Accuracy adalah perbandingan kasus yang diidentifikasi benar dengan jumlah semua kasus. Rumus dari accuracy = $(a+c)/(a+b+c+d)$

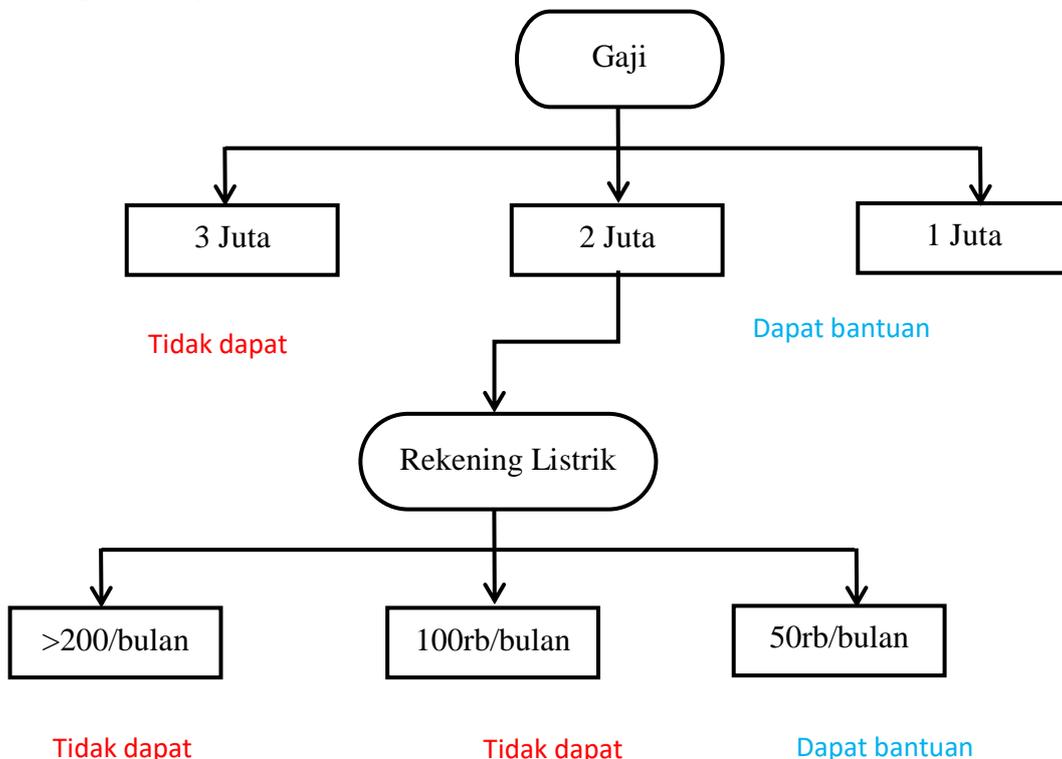
- Error Rate adalah kasus yang diidentifikasi salah dengan sejumlah semua kasus

$$\text{Rumus dari Error Rate} = (b+c)/(a+b+c+d)$$

Keterangan:

- jika hasil prediksi negatif dan nilai sebenarnya negatif.
- jika hasil prediksi positif sedangkan nilai sebenarnya negatif.
- jika hasil prediksi negatif sedangkan nilai sebenarnya positif.
- jika hasil prediksi positif dan nilai sebenarnya positif.

Contoh perhitungan confusion matrix adalah sebagai berikut, akan saya beri contoh sederhana yaitu pengambilan keputusan untuk mendapatkan bantuan beasiswa. Saya menggunakan dua atribut, yaitu rekening listrik dan gaji. Ini adalah pohon keputusannya.



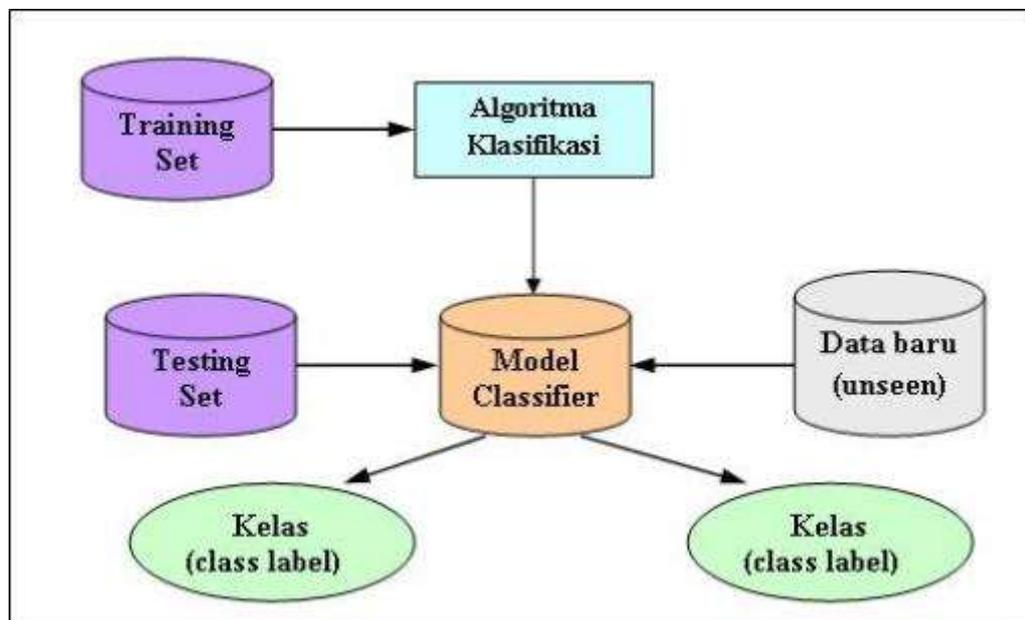
Gambar 2.5 Model Classifier Pada Klasifikasi

2.11 Training set, Testing set, dan Validation set

Ketiga istilah ini sangat berkaitan dengan klasifikasi data. Sebuah model classifier pada klasifikasi data dibentuk berdasarkan data yang sudah ada, dan kemudian

model tersebut digunakan untuk klasifikasi dan prediksi data baru yang belum pernah ada.

Pada umumnya, proses dapat digambarkan seperti ini:



Gambar 2.6 Model Classifier Pada Klasifikasi

Data umumnya dibagi menjadi training set dan testing set. Training set digunakan oleh algoritma klasifikasi (misalnya: decision tree, bayesian, neural network, SVM) untuk membentuk sebuah model classifier. Model ini merupakan representasi pengetahuan yang akan digunakan untuk prediksi kelas data baru yang belum pernah ada. Testing set digunakan untuk mengukur sejauh mana classifier berhasil melakukan klasifikasi dengan benar. Karena itu, data yang ada pada testing set seharusnya tidak boleh ada pada training set sehingga dapat diketahui apakah model classifier sudah “pintar” dalam melakukan klasifikasi. Lain lagi halnya dengan validation set. Umumnya beberapa algoritma klasifikasi memerlukan beberapa parameter. Misalnya: jumlah hidden layer dan learning rate pada neural network; parameter kernel pada SVM. Biasanya sebagian dari training set diambil untuk validation set. Validation set ini digunakan untuk mencari parameter yang paling baik untuk sebuah algoritma klasifikasi.

Memisahkan data menjadi training dan testing set dimaksudkan agar model yang diperoleh nantinya memiliki kemampuan generalisasi yang baik dalam melakukan klasifikasi data. Tidak jarang sebuah model klasifikasi dapat melakukan klasifikasi data dengan sangat baik pada training set tetapi sangat buruk dalam melakukan klasifikasi data yang baru dan belum pernah ada. Hal ini dinamakan overfitting. Model tersebut masih belum baik.

2.11 Penelitian Terkait

Tabel 2.1 Penelitian Terkait

NO.	Nama Peneliti dan Tahun	Masalah	Metode	Hasil
1.	Fourina Ayu Novianti dan Santi Wulan Purnami, September 2012	Penelitian ini menganalisis 17ogist-faktor apa saja yang mempengaruhi kanker payudara menggunakan metode regresi 17ogistic dan support vector machine (SVM).	Supprot Vector Mechine dan Regresi Logistic	Pengklasifikasian menggunakan regresi 17ogistic biner menghasilkan ketepatan klasifikasi sebesar 88,72% Sedangkan Pengklasifikasian Menggunakan SVM menghasilkan ketepatan klasifikasi sebesar 94,34%.
2	Laily Hermawanti, Maret 2012	Menerapkan algoritma klasifikasi C4.5 untuk peningkatan akurasi dalam mendiagnosis penyakit kanker payudara.	Decision Tree C4.5	Nilai akurasi untuk algoritma klasifikasi C4.5 senilai 94.56 % dan nilai AUC untuk algoritma C4.5 senilai 0.941.
3	Fairudz Shahura, Oni Soesanto, Fatma Indriani,	bagaimana performansi metode Radial Basis Probabilistic Neural Network untuk	Radial Basis Probabilistic Neural Network	Hasil klasifikasi dengan menggunakan metode Radial Basis Probabilistic Neural Network untuk kanker

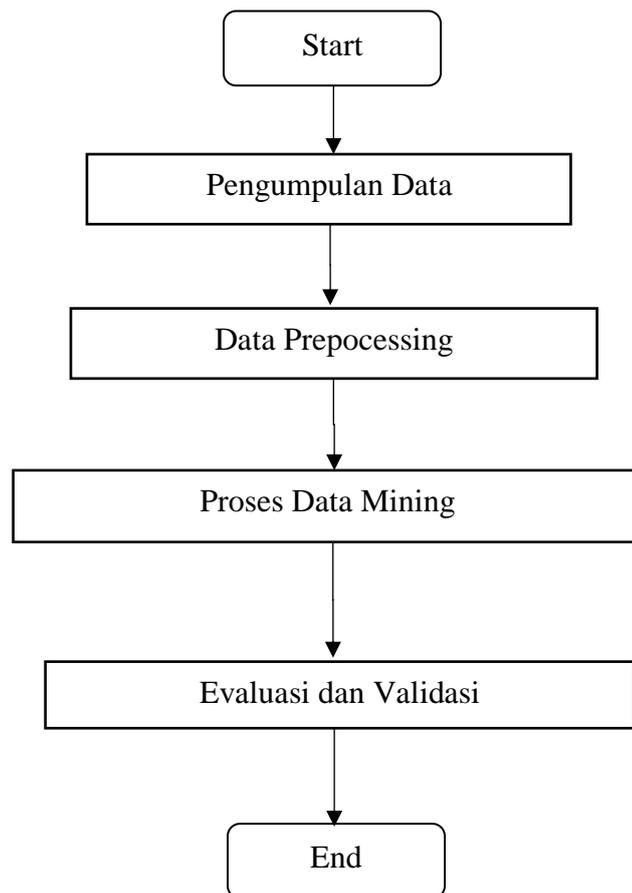
	September 2016	mengklasifikasikan kanker payudara.		payudara dari sistem yang dibuat sebesar 91.87%, 93.19%, 92.31%, dan 92.09%, dan untuk data testing berturut –turut sebesar 89.47%, 90.35%, 89.47%, dan 89.47%.
4	Al Riza Khadafy, Desember 2015	Penerapan algoritma pengklasifikasi Naïve Bayes dapat mengurangi data noise pada dataset berukuran besar dan memiliki banyak kelas atau multi kelas sehingga akurasi klasifikasi algoritma DT dapat meningkat.	Decision Tree dan Naive Bayes.	Penerapan algoritma pengklasifikasi NB dapat mengurangi data noise pada dataset berukuran besar dan memiliki banyak kelas atau multi kelas sehingga akurasi klasifikasi algoritma DT dapat meningkat
5.	Cheng-Lung Huang, Hung-Chang Liao, Mu-Chen Chen, 2008	Pasien yang terkena Virus DNA – HSV-1, EBV, CMV, PV, dan HHV-8 – didiagnosis terjangkit tumor payudara	Support Vector Machine (SVM)	Penelitian ini menunjukkan bahwa pengklasifikasi berbasis SVM untuk fibroadenoma atau klasifikasi diagnosis kanker payudara model memuaskan baik dalam akurasi klasifikasi dan dalam menemukan fitur-fitur penting untuk membedakan antara Fibroadenoma atau kanker payudara.

6.	Ol Mangasarian, W. Nick Street dan William H. Wolberg , 1994	Penelitian ini dilakukan agar mendapatkan hasil yang optimal dalam pendiagnosaan kanker payudara guna pengambilan keputusan dokter bedah operasi dalam menangan pasien sebelum penanganan lebih lanjut.	Multisurface (MSM)	Penelitian ini menunjukkan bahwa mendapatkan hasil yang akurat dengan metode yang diterapkan dengan menghindari kesalahan – kesalahan terhadap hasil keputusan dokter bedah operasi dalam menangani pasien sebelum penanganan lebih lanjut.
7	Murat Karabatak M. Cevdet Ince, 2008	Dalam penelitian ini bermaksud untuk membantu dokter dalam membuat keputusan diagnostik.	JST (Apriori	sistem diagnosis otomatis untuk mendeteksi kanker payudara berdasarkan aturan asosiasi (AR) dan jaringan saraf (NN). Dalam penelitian ini, AR digunakan untuk mengurangi dimensi data kanker payudara. NN digunakan untuk klasifikasi cerdas. Tingkat klasifikasi yang benar dari sistem yang diusulkan adalah 95,6%.

BAB III

METODE PENELITIAN

Penelitian ini akan dilaksanakan berdasarkan alur penelitian seperti yang ditunjukkan pada Gambar 3.1.



Gambar 3.1 Alur Penelitian

3.1. Tempat dan Waktu Penelitian

3.2. Alat dan Bahan

Adapun alat dan bahan yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Hardware

Kebutuhan perangkat keras (hardware) yang digunakan : Laptop Axioo Neon
Processor : Intel® Core™ 2 Duo @2.10 GHz. RAM : 2.00 GB. harddisk dengan kapasitas 80 GB.

2. Software

Kebutuhan perangkat lunak (software) yang digunakan :

- a. Rapid Miner 9.0
- b. Microsoft Excel 2010,
- c. Microsoft Windows 8.1

1.3 Metode Pengumpulan Data

Dalam penyusunan penelitian ini metode yang digunakan dalam proses pengumpulan data adalah Studi pustaka. Studi Pustaka dilakukan dengan cara menghimpun informasi yang relevan dengan topik atau masalah yang akan atau sedang diteliti. Informasi diperoleh dari buku-buku ilmiah, laporan penelitian, karangan-karangan ilmiah, tesis dan disertasi, peraturan-peraturan, ketetapan-ketetapan, ensiklopedia, dan sumber-sumber tertulis baik tercetak maupun elektronik.

3.4 Data Preprocessing

Untuk menghilangkan data yang tidak diperlukan, data yang diperoleh dari tahap pengambilan dataset akan disaring untuk menghasilkan data yang benar-benar dibutuhkan. umumnya data tersebut memiliki nilai yang tidak sempurna seperti data yang hilang. Selain itu, ada juga atribut-atribut data yang tidak sesuai dengan pemrosesan data mining yang akan digunakan. Data-data yang tidak relevan itu

juga lebih baik dibuang karena keberadaannya bisa mengurangi mutu atau akurasi dari hasil data mining nantinya.

Nilai – nilai atribut tak benar mungkin karena :

- Kegagalan instrumen pengumpulan data
- Problem pemasukan data
- Problem transmisi data
- Keterbatasan teknologi
- Ketak- konsistenan dalam konversi penamaan

Problem data lainnya yang memerlukan pembersihan data.

- Duplikasi record
- Data tak dianggap
- Data tidak konsisten

3.5 Penerapan Data Mining

3.5.1 Penerapan Algoritma Decision Tree (C4.5)

Pada pohon keputusan ini bisa memberikan keuntungan berwujud visualisasi dari pemecahan masalah yang diolah menggunakan teknik data mining yang membuat protocol dari prediksinya dapat diamati, maka dari itu konsep ini termasuk fleksibel dan atraktif. Pohon keputusan ini sendiri juga sudah banyak digunakan pada berbagai bidang ilmu pengetahuan, salah satunya yaitu bidang kesehatan untuk diagnosa penyakit pasien, ilmu computer pada struktur data, psikologi untuk teori pengambilan keputusan, dan lain-lain.

Dalam pohon keputusan sangat berhubungan dengan algoritma C4.5, karena dasar algoritma C4.5. Algoritma data mining C4.5 merupakan salah satu algoritma yang digunakan untuk melakukan klasifikasi atau segmentasi atau pengelompokan yang bersifat prediktif. Cabang-cabang pohon keputusan merupakan pertanyaan klasifikasi dan daun-daunnya merupakan kelas-kelas atau segmen-segmennya.

Rumus menghitung entropy pada algoritma C4.5

$$Entropi (S) = -\sum_{k=1}^n p_k \log_2 p_k$$

Keterangan :

S : Himpunan (dataset) kasus

k : Banyaknya partisi S

P_i : Probabilitaas yang didapat dari Sum (Ya) atau Sum (Tidak) dibagi total kasus

Setelah mendapatkan entropi dari keseluruhan kasus, lakukan analisis pada setiap atribut dan nilai-nilainya dan hitung entropinya. Langkah berikutnya yaitu dengan menghitung Gain, rumus daripada Gain adalah sebagai berikut:

$$Gain(A) = Entropi(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times Entropi(S_i)$$

Keterangan :

S : himpunan kasus

A : atribut

n : jumlah partisi atribut A

$|S_i|$: jumlah kasus pada partisi ke- i

$|S|$: jumlah kasus dalam S

Dibawah ini merupakan contoh perhitungan manual dari penerapan algoritma C4.5 untuk Fitur Seleksi Pada Algoritma C4.5 harus menentukan pohon keputusan yang kemudian akan menjadi akar untuk menghasilkan seleksi atribut mana yang paling berpengaruh dan tidak banyak pengaruh terhadap kelas.

- Entropy Total

Jumlah kasus (S) = 683

Jumlah Maligant (S_1) = 239

Jumlah Benignant (S_2) = 444

$$\begin{aligned} \text{Entropi(Total)} &= \left(-\frac{S_1}{S} * \log_2 \frac{S_1}{S}\right) + \left(-\frac{S_2}{S} \log_2 \frac{S_2}{S}\right) \\ &= \left(-\frac{239}{683} * \log_2 \frac{239}{683}\right) + \left(-\frac{444}{683} * \log_2 \frac{444}{683}\right) \end{aligned}$$

$$\log_2 \left(\frac{239}{683}\right) = \log_2(239) - \log_2(683)$$

$$\begin{aligned}
&= 1 - \frac{\ln(683)}{\ln(239)} \\
&= 1 - \frac{6.52649}{5.47646} \\
&= 1 - 1,19173 \\
&= - \mathbf{0,119173}
\end{aligned}$$

$$\begin{aligned}
\text{Log}_2(444/683) &= \text{Log}_2(444) - \text{Log}_2(683) \\
&= 1 - \frac{\ln(683)}{\ln(444)} \\
&= 1 - \frac{6.52649}{6.09582} \\
&= 1 - 1,07065 \\
&= - \mathbf{0,07065}
\end{aligned}$$

$$\begin{aligned}
\text{Entropi(Total)} &= (- \frac{239}{683} * - \mathbf{0,119173}) + (- \frac{444}{683} * - \mathbf{0,07065}) \\
&= \mathbf{0,06709} + \mathbf{0,04592} \\
&= \mathbf{0,11302}
\end{aligned}$$

Berikut diatas adalah nilai entropy total yang dihasilkan.

- Keseragaman Bentuk Sel >> 1

$$\text{Jumlah kasus (S)} = 372$$

$$\text{Jumlah Maligant (S1)} = 3$$

$$\text{Jumlah Benignant (S2)} = 369$$

$$\begin{aligned}
\text{Entropi(Total)} &= (- \frac{3}{372} * \text{Log}_2 \frac{3}{372}) + (- \frac{369}{372} * \text{Log}_2 \frac{369}{372}) \\
&= 0.035383946 + 0.001358854 \\
&= \mathbf{0.03674}
\end{aligned}$$

- Keseragaman Bentuk Sel >> 2

$$\text{Jumlah kasus (S)} = 41$$

$$\text{Jumlah Maligant (S1)} = 4$$

$$\text{Jumlah Benignant (S2)} = 37$$

$$\text{Entropi(Total)} = (- \frac{4}{41} * \text{Log}_2 \frac{4}{41}) + (- \frac{37}{41} * \text{Log}_2 \frac{37}{41})$$

$$= 0.163783025 + 0.025655281$$

$$= \mathbf{0.18944}$$

- Keseragaman Bentuk Sel >> 3

$$\text{Jumlah kasus (S)} = 39$$

$$\text{Jumlah Maligant (S1)} = 12$$

$$\text{Jumlah Benignant (S2)} = 27$$

$$\text{Entropi(Total)} = (-12/39 * \text{Log}_2 12/39) + (-27/39 * \text{Log}_2 27/39)$$

$$= 0.145946358 + 0.077242504$$

$$= \mathbf{0.22319}$$

- Keseragaman Bentuk Sel >> 5

$$\text{Jumlah kasus (S)} = 45$$

$$\text{Jumlah Maligant (S1)} = 45$$

$$\text{Jumlah Benignant (S2)} = 0$$

$$\text{Entropi(Total)} = (-45/45 * \text{Log}_2 45/45) + (-0/45 * \text{Log}_2 0/45)$$

$$= 0 + 0$$

$$= \mathbf{0}$$

- Keseragaman Bentuk Sel >> 6

$$\text{Jumlah kasus (S)} = 18$$

$$\text{Jumlah Maligant (S1)} = 18$$

$$\text{Jumlah Benignant (S2)} = 0$$

$$\text{Entropi(Total)} = (-18/18 * \text{Log}_2 18/18) + (-0/18 * \text{Log}_2 0/18)$$

$$= 0 + 0$$

$$= \mathbf{0}$$

- Keseragaman Bentuk Sel >> 7

$$\text{Jumlah kasus (S)} = 23$$

$$\text{Jumlah Maligant (S1)} = 22$$

$$\text{Jumlah Benignant (S2)} = 1$$

$$\text{Entropi(Total)} = (-22/23 * \text{Log}_2 22/23) + (-1/23 * \text{Log}_2 1/23)$$

$$= 0.013755579 - 0.043478261$$

$$= \mathbf{-0.0297}$$

- Keseragaman Bentuk Sel >> 8

$$\text{Jumlah kasus (S)} = 41$$

$$\text{Jumlah Maligant (S1)} = 40$$

$$\text{Jumlah Benignant (S2)} = 1$$

$$\text{Entropi(Total)} = (-40/41 * \text{Log}_2 40/41) + (-1/41 * \text{Log}_2 1/41)$$

$$= 0.006530534 - -0.024390244$$

$$= \mathbf{-0.0179}$$

- Keseragaman Bentuk Sel >> 9

$$\text{Jumlah kasus (S)} = 15$$

$$\text{Jumlah Maligant (S1)} = 14$$

$$\text{Jumlah Benignant (S2)} = 1$$

$$\text{Entropi(Total)} = (-14/15 * \text{Log}_2 14/15) + (-1/15 * \text{Log}_2 1/15)$$

$$= 0.024400132 - 0.066666667$$

$$= \mathbf{-0.0423}$$

- Keseragaman Bentuk Sel >> 10

$$\text{Jumlah kasus (S)} = 69$$

$$\text{Jumlah Maligant (S1)} = 69$$

$$\text{Jumlah Benignant (S2)} = 0$$

$$\text{Entropi(Total)} = (-69/69 * \text{Log}_2 69/69) + (-0/69 * \text{Log}_2 0/69)$$

$$= 0 + 0$$

$$= \mathbf{0}$$

Berikut diatas adalah nilai entropi dari partisi 1 sampai 10 dalam Atribut Keseragaman Bentuk Sel, Lakukan perhitungan yang sama terhadap atribut yang lain. Langkah selanjutnya ketika seluruh masing – masing partisi mendapatkan hasil entropinya, maka lakukan perhitungan nilai gain dari atribut dataset kanker payudara.

Perhitungan Gain atribut Keseragaman Bentuk Sel.

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

$$\begin{aligned} \text{Gain Total, Atribut Keseragaman Bentuk Sel} &= 0.11302 - ((\frac{372}{683} * \mathbf{0.03674}) + \\ &(\frac{41}{683} * \mathbf{0.18944}) + (\frac{39}{683} * \mathbf{0.22319}) + (\frac{20}{683} * \mathbf{0.2996}) + (\frac{45}{683} * 0) + (\frac{18}{683} \\ &* 0) + (\frac{23}{683} * \mathbf{-0.0297}) + (\frac{41}{683} * \mathbf{-0.0179}) + (\frac{15}{683} * \mathbf{-0.0423}) + (\frac{69}{683} * \mathbf{0})) \\ &= \mathbf{0.063120888} \end{aligned}$$

Lakukan hal yang sama terhadap perhitungan untuk mendapatkan hasil gain disetiap atributnya. Berikut adalah nilai gain yang diperoleh dari masing – masing atribut dalam bentuk table dibawah ini :

Tabel 3.1 Perhitungan Algoritma C4.5

Node	Keterangan / Nama Atribut	Partisi	Jumlah Kasus	Class		Entropy	Gain
				Maligant	Benignant		
1	Total		683	239	444	0.11302	
	Mitosis						-0.005459734
		1	139	3	136	0.0797	
		2	50	4	46	0.16579	
		3	104	12	92	0.12426	
		4	79	12	67	0.14843	
		5	128	45	83	0.16011	
		6	33	18	15	0.24673	
		7	23	22	1	0.01376	
		8	44	40	4	0.18074	
		9	14	14	0	0	
		10	69	69	0	0	
	Keseragaman Bentuk Sel						0.063120888
		1	372	3	369	0.03674	
		2	41	4	37	0.18944	
		3	39	12	27	0.22319	
		4	20	12	8	0.2996	
		5	45	45	0	0	
		6	18	18	0	0	
		7	23	22	1	-0.0297	
		8	41	40	1	-0.0179	
		9	15	14	1	-0.0423	
		10	69	69	0	0	
	Ketebalan Gumpalan						0.011540197

Tabel 3.1 Perhitungan Algoritma C4.5 (lanjutan 1)

		1	346	2	344	0.04396	
		2	58	7	51	0.15991	
		3	53	23	30	0.21025	
		4	43	31	12	0.21203	
		5	32	30	2	0.26779	
		6	29	27	2	0.28625	
		7	30	28	2	0.27978	
		8	27	26	1	-0.0259	
		9	7	7	0	0	
		10	58	58	0	0	
	Marginal Adhesion						0.010298391
		1	393	30	363	0.07018	
		2	58	21	37	0.20023	
		3	58	27	31	0.2055	
		4	33	28	5	0.21949	
		5	23	19	4	0.27304	
		6	21	18	3	0.29875	
		7	13	13	0	0	
		8	25	25	0	0	
		9	4	4	0	0	
		10	55	54	1	-0.0137	
	Keseragaman Ukuran Sel						0.018892574
		1	44	1	43	-0.0168	
		2	376	21	355	0.06217	
		3	71	43	28	0.19087	
		4	48	41	7	0.18054	
		5	39	34	5	0.19755	
		6	40	39	1	-0.0183	
		7	11	9	2	0.52189	
		8	21	19	2	0.35383	
		9	2	2	0	0	
		10	31	30	1	-0.0229	
	Ukuran Asli Nuclei						0.015588343
		1	402	15	387	0.05145	
		2	30	9	21	0.24639	
		3	28	14	14	0.26265	
		4	19	13	6	0.30438	
		5	30	20	10	0.24927	
		6	4	4	0	0	
		7	8	7	1	-0.065	
		8	21	19	2	0.35383	
		9	9	9	0	0	
		10	132	129	3	0.08291	

	Ukuran Sel Epitel Tunggal						0.008395413
		1	150	2	148	0.0857	
		2	160	7	153	0.07886	

Tabel 3.1 Perhitungan Algoritma C4.5 (lanjutan 2)

		3	161	36	125	0.13416	
		4	39	32	7	0.20527	
		5	34	30	4	0.21409	
		6	9	8	1	-0.0608	
		7	71	65	6	0.1359	
		8	28	28	0	0	
		9	11	11	0	0	
		10	20	20	0	0	
	Kromatin						0.003349992
		1	432	41	391	0.0753	
		2	36	6	30	0.21134	
		3	42	31	11	0.21161	
		4	18	17	1	-0.0365	
		5	19	17	2	0.37701	
		6	22	18	4	0.28039	
		7	16	14	2	0.41927	
		8	23	20	3	0.2824	
		9	15	15	0	0	
		10	60	60	0	0	
	Keadaan Nucleoli Normal						0.005917772
		1	563	132	431	0.10336	
		2	35	27	8	0.22297	
		3	33	31	2	0.26222	
		4	12	12	0	0	
		5	6	5	1	-0.0723	
		6	3	3	0	0	
		7	9	8	1	-0.0608	
		8	8	7	1	-0.065	
		9	0	0	0	0	
		10	14	14	0	0	

Dari hasil pada Tabel 3.1 dapat diketahui bahwa atribut dengan Gain tertinggi adalah Keseragaman bentuk Sel yaitu sebesar 0.05. Dengan demikian keseragaman bentuk sel dapat menjadi node akar dan masih terbagi lagi menurut perangkingan masing – masing atribut dibawahnya. Kemudian Ada 10 partisi atau nilai atribut dari keseragaman bentuk sel yaitu Nilai 1 sampai dengan 10. Dari sepuluh nilai atribut tersebut, nilai atribut masih perlu dilakukan perhitungan lagi.

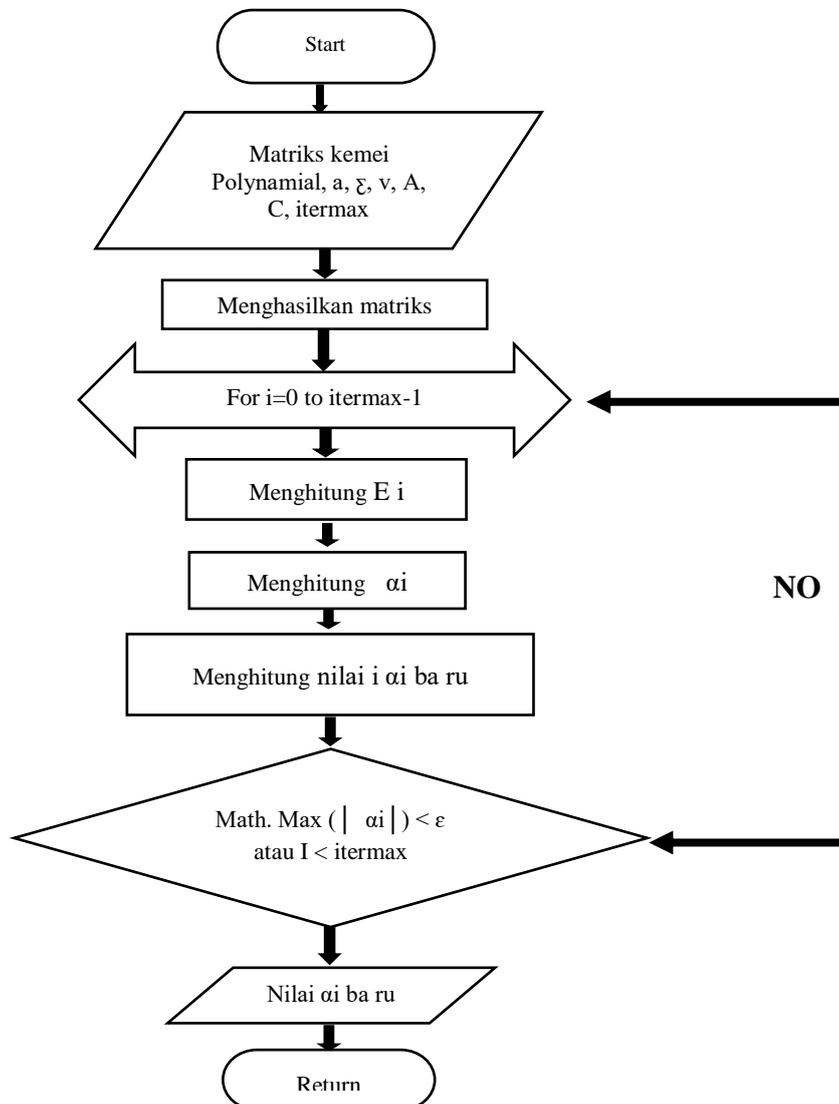
Dari perhitungan diatas didapat perankingan gain dari mulai yang terbesar sampai dengan yang terkecil. Berdasarkan perhitungan gain terbesar dapat dinyatakan sebagai akar, kemudian gain yang terkecil merupakan atribut yang kurang berpengaruh terhadap hasil. Dengan demikian atribut tersebut dapat direlease/dihapus dan kemudian dapat dilakukan optimalkan akurasi dengan menggunakan algoritma yang berbeda, yaitu SVM.

3.5.2 Proses Perhitungan Algoritma Support Vector Machine

Proses Support Vector Machine Pada langkah awal dalam proses SVM akan dilakukan input data dari database. Setelah itu data akan dihitung menggunakan kernel polynomial untuk diteruskan ke proses training dengan menggunakan proses Sequential Training SVM. Output yang akan dihasilkan adalah berupa hasil klasifikasi.

A. Sequential Training SVM

Proses Sequential Training SVM adalah proses yang digunakan untuk mencari nilai hyperplane terbaik. Proses ini dimulai dari inialisasi nilai konstanta, perhitungan matriks hessian, perhitungan E_i , $\delta\alpha_i$ hingga menghitung nilai α_i yang baru. Hasil akhir dari proses training ini adalah berupa nilai α_i yang akan digunakan untuk proses pengujian. Gambar 1 akan memperlihatkan diagram alir proses SVM.



Gambar 3.2 Tabel Alur Perhitungan SVM

B Perhitungan Manual Algoritma SVM

Pada proses selanjutnya adalah proses transformasi data dengan menggunakan SVM. Berikut adalah persamaan yang dipakai dalam algoritma yang dilakukan secara manual untuk mencari transformasi data:

- Untuk Mencari Transformasi data

$$X_n = \frac{0.8*(X-a)}{b-a} + 0.1$$

(1)

Di mana :

- X_n = Nilai ke - n
- A = Nilai angka terendah
- B = Nilai angka tertinggi
- 0.8 dan 0.1 = Ketetapan

- Untuk Menghitung Hasil Prediksi

$$f(\phi(x)) = \text{sign}(w \cdot \phi(x) + b) = \quad \text{Dimana:}$$

$$\text{Sign}\left(\sum_{i=1}^n a_i y_i \phi(x_i)^T \cdot \phi(x) + b\right)$$

- w = Nilai weight
- x = Nilai variable input
- b = Nilai bias

Pada penelitian ini diketahui bahwa jumlah kasus adalah sebanyak 683 record dari jumlah tersebut terdapat 9 atribut dan 1 Kelas yaitu Benignant atau kanker jinak dan Malignant yaitu kanker. Dari study kasus tersebut SVM akan mengklasifikasikan tingkat keakurasian data tersebut, berikut pembahasannya :

Dimana :

0.8 : Ketetapan rumus

- 0.1 : Ketetapan rumus
 10 : Jumlah kasus pada data pertama
 3 : Jumlah Kasus terendah
 69 : Jumlah Kasus tertinggi
 136 : Jumlah Nilai pada data perta
 139 : Jumlah Nilai tertinggi
 0 : Jumlah Nilai terendah

$$X1 = \frac{0.8(136-3)}{10-3} + 0.1 = \frac{5.6}{7} + 0.1$$

$$= 0.7057 + 0.1 = 0.8057$$

$$X2 = \frac{0.8(136-4)}{139-14} + 0.1 = \frac{108}{135} + 0.1$$

$$= 0.753 + 0.1 = 0.853$$

Sehingga diperoleh transformasi data x1 (Partisi atribut) = 0.9 dan transformasi data x2 (nilai) = 0.9. Lakukan hal yang sama untuk seluruh data.

Tabel 3.2 Perhitungan Manual Algoritma

X1	X2	Label
0.08057554	0.85329	-1
0.043089431	0.87398	-1
0.030232558	0.88837	-1
0.046961326	0.95967	-1
0.452380952	0.19492	-1
0.1375	0.92701	-1
0.002702703	0.86216	-1
0.06112532	0.94536	-1

0.319025522	0.34397	-1
-------------	---------	----

Jika hasil prediksi pertama adalah negatif maka label data latih (-1), jika hasil prediksi sementara positif maka label data latih (+1).

Pada penelitian ini penulis mempunyai α ta. Setelah menghitung nilai matriks kernel k, maka didapatkan nilai matriks K dengan ukuran $N \times N$ $K(x,xi)=\infty (xi)$

0.08057554	0.85329
0.043089431	0.87398
0.030232558	0.88837
0.046961326	0.95967
0.452380952	0.19492
0.137500000	0.92701
0.002702703	0.86216
0.06112532	0.94536
0.319025522	0.34397

Karena pada fitur label memiliki 10 data maka, untuk alfa juga memiliki 10 data yang sama sehingga setiap elemenmatrik kernel $K(N*N)$ yang dikali dengan $Y_i Y_j$.

Maksudnya untuk mendapatkan nilai alpha kita harus menjumlahkan semua nilai pada matriks kernel k dengan berpedoman pada label data latih pada label data latihnya - 1 maka nilai matriks juga (-) dan jika pada label data latihnya (+) maka nilai matriksnya juga (+) persamaan ini untuk masing - masing kolom matriks. Maka nilai alpha dengan persamaan adalah :

$$A_1=A_2=A_3=A_4=A_5=A_6=A_7=A_8=A_9=A_{10} = 10/6.83074$$

$$= 1.4639$$

Setelah didapatkan nilai alpha, langkah selanjutnya adalah mencari nilai weight (w) dan bias (b) dengan persamaan. Artinya nilai alpha dikali dengan nilai matriks per kolom. Nilai alpha +/- masih berpedoman pada label data latih.

0.83529
0.87398
0.88837
0.95967
0.19492
0.927010

Setelah didapatkan nilai w, selanjutnya pilih salah satu Support Vector dari kelas "+ 1" dan "-1" dari matriks mempunyai nilai (+) dan (-).

16

0.86216

BAB IV

HASIL DAN PEMBAHASAN

Didalam bab ini penulis akan membahas mengenai data yang akan digunakan dalam penelitian, data tersebut akan dihitung menggunakan feature selection dengan menggunakan algoritma C4.5 yang kemudian akan diuji menggunakan Cross Validation menggunakan algoritma Support Vector Machine (SVM). Sebagai pengimprove akurasi. Pada penelitian ini, data yang digunakan adalah Data diagnosa didapat dari Universitas California Irvine (UCI) Machine Learning Repository yaitu Breast Cancer Winconsin Original Data Set yang berjumlah 683 record.

4. 1.Evaluasi dan Validasi

Tahap pengujian data yang dilakukan menggunakan RapidMiner 7. dengan tujuan untuk melihat nilai akurasi, pohon keputusan, dan rule sebagai seleksi fitur. Pada model klasifikasi dapat diketahui hasil evaluasi berdasarkan pada banyaknya dataset record yang diklasifikasi secara benar atau tidak benar pada model klasifikasi tersebut. Dari 683 record akan dilakukan pengujian sebanyak 3 kali. Pembagian pengujian dengan data training dan testing yang berbeda. Data training 50% 50%, Data training 70% dan data testing 30% dan Data training 80% dan data testing 20%. Data training digunakan untuk membentuk model, sedangkan data testing digunakan untuk menguji ketepatan klasifikasi dari model yang telah dibentuk. Berikut merupakan tampilan proses import Data diagnosa didapat dari Universitas California Irvine (UCI) Machine Learning Repository yaitu Breast Cancer Winconsin Original Data Set yang berjumlah 683 record. Dalam bentuk file Microsoft Excel dalam rapidminer 5.3:

4.2 Data preprocessing

Berikut adalah tahapan data preprocessing pada data Mining :

1. Pembersihan data (*Cleaning data*)

Pembersihan data akan mempengaruhi performansi dari sistem data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya. Dalam study kasus kanker payudara pada dataset public repository breast cancer terdapat 16 record error dari 699 record menjadi 683 record.

2. Seleksi Data

Data diseleksi untuk menentukan variabel apa saja yang akan diambil agar tidak terjadi kesamaan dan perulangan yang tidak diperlukan dalam pengolahan teknik data mining. Dalam penanganan kasus noisy yang terjadi pada dataset public repository breast cancer, dalam penseleksian data dapat dilakukan dengan cara mengimport data ke dalam program rapidminer, data tersebut tidak akan dapat diolah karena masih terdapat beberapa error data. Program akan secara otomatis memberi tahu bahwa dataset perlu dilakukan preprocessing. Setelah data error tersebut ditemukan kita dapat menseleksi melakukan filter untuk menghapus data tersebut. Sehingga data tersebut dapat diproses sebagaimana mestinya.

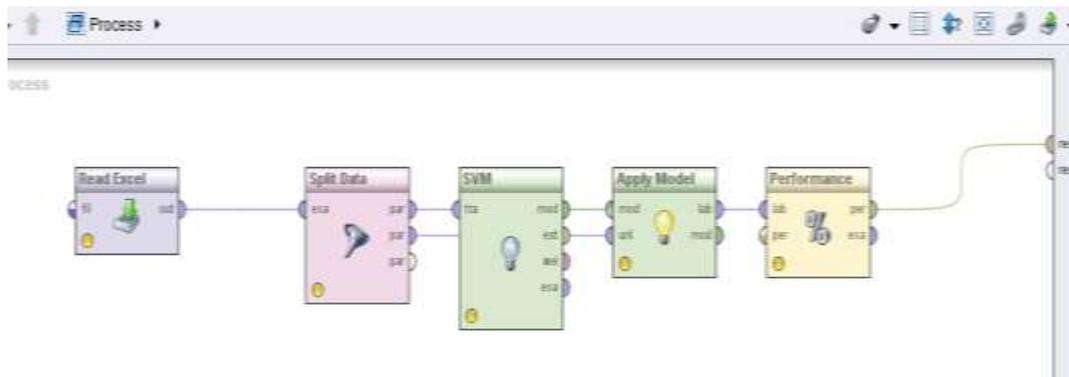
4.3 Penerapan Data Mining

4.3.1. Penerapan SVM untuk akurasi kanker payudara

Pengujian yang pertama data set dari jumlah data 683 record dengan 9 atribut dan 1 sebagai Class.

4.3.1.1 Block Process penerapan SVM

Berikut merupakan Block Process algoritma SVM untuk mengetahui hasil pengujian algoritma dengan dataset kanker payudara.



Gambar 4.3 Block Procressing Algoritma SVM

Split Data digunakan sebagai ratio pembandingan antara data training dan data testing. Data training digunakan untuk membentuk model, sedangkan data testing digunakan untuk menguji ketepatan klasifikasi dari model yang telah dibentuk.

4.3.1.2 Hasil Pengujian

Berikut merupakan hasil Pengujian berdasarkan jumlah data yang diuji dari sumber data training dan data testing. Hasil dari data keseluruhan 683 record dari dataset kanker payudara. Dengan 9 Atribut dan 1 atribut sebagai Class dapat dilihat melalui gambar berikut dibawah ini :

A. Accuracy 1

Berikut adalah Hasil dari Accuray sebelum dilakukan seleksi fitur dengan perbandingan Ratio sebesar 50%, 50% adalah sebagai berikut :

accuracy: 97.65%			
	true B	true M	class precision
pred B	218	2	99.09%
pred M	6	115	95.04%
class recall	97.32%	99.23%	

Gambar 4.4 Analisis Hasil akurasi SVM ratio 50 %, 50 %

Klasifikasi SVM pada penelitian ini menggunakan aplikasi program rapidminer dengan Data training dan testing sebesar 50:50. Ketepatan klasifikasi terbesar yang dihasilkan oleh metode SVM dari partisi data Training dan testing 50:50 yaitu sebesar 97,65 %.

B. Accuracy 2

Berikut adalah Hasil dari Accuray sebelum dilakukan seleksi fitur dengan perbandingan Ratio sebesar 70 %, 30 % adalah sebagai berikut :

Table / Plot View / Text View / Annotations			
Criterion Selector			
	<input checked="" type="radio"/> Multiclass Classification Performance <input type="radio"/> Annotations		
	<input checked="" type="radio"/> Table View <input type="radio"/> Plot View		
	accuracy: 97.56%		
	true B	true M	class precision
pred. B	132	2	98.51%
pred. M	3	68	95.77%
class recall	97.79%	97.14%	

Gambar 4.5 Analisis Hasil akurasi SVM ratio 70%, 30%

Klasifikasi SVM pada penelitian ini menggunakan aplikasi program rapidminer dengan Data training dan testing sebesar 70:30. Ketepatan klasifikasi terbesar yang dihasilkan oleh metode SVM dari partisi data Training dan testing 70:30 yaitu sebesar 97,56 %.

C. Accuracy 3

Berikut adalah Hasil dari Accuray sebelum dilakukan seleksi fitur dengan perbandingan Ratio sebesar 80%, 20% adalah sebagai berikut :

Table View / Plot View			
	accuracy: 96.35%		
	true B	true M	class precision
pred. B	92	3	96.94%
pred. M	2	40	95.24%
class recall	97.87%	99.02%	

Gambar 4.6 Analisis Hasil akurasi SVM 80%,20%

Klasifikasi SVM pada penelitian ini menggunakan aplikasi program rapidminer dengan Data training dan testing sebesar 80:20. Ketepatan klasifikasi terbesar yang

dihasilkan oleh metode SVM dari partisi data Training dan testing 80:20 yaitu sebesar 96,35 %.

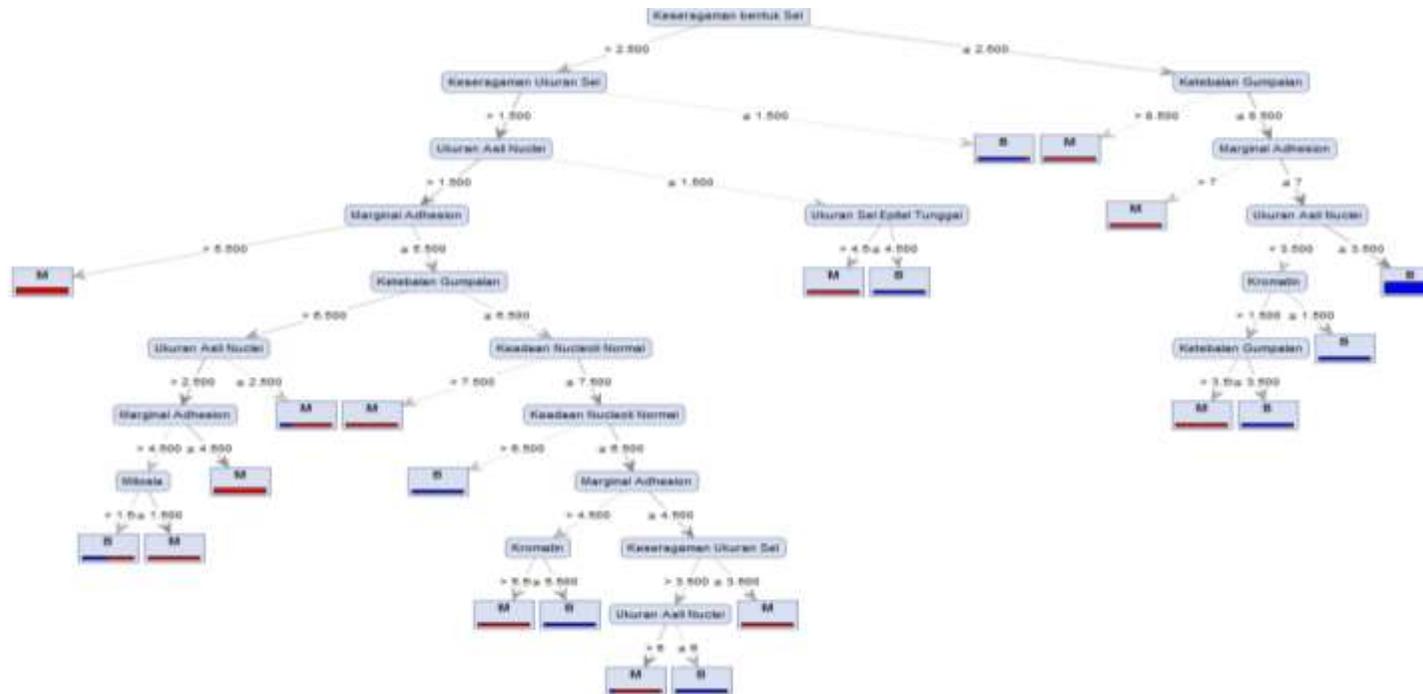
4.4 Penerapan C4.5 untuk seleksi fitur

Pada proses selanjutnya adalah proses pengujian data dengan menggunakan C4.5 untuk melihat pohon keputusan yang dihasilkan, rule yang terbentuk dan nilai akurasi dari algoritma tersebut dan Class digunakan sebagai label/ sasaran atribut. Berikut adalah pengujian algoritma pada RapidMiner 5.3, dengan model pengujian menggunakan Split Data. Dari model algoritma tersebut maka setelah dilakukan Running Program akan terlihat pohon keputusan yang dihasilkan, rule yang terbentuk fitur seleksi dari algoritma tersebut. Berikut adalah Gambar pohon keputusan pada aplikasi RapidMiner 5.3

4.2.2.1 Hasil

A. Pohon Keputusan

Berikut adalah tampilan gambar pohon keputusan setelah dilakukan running program :

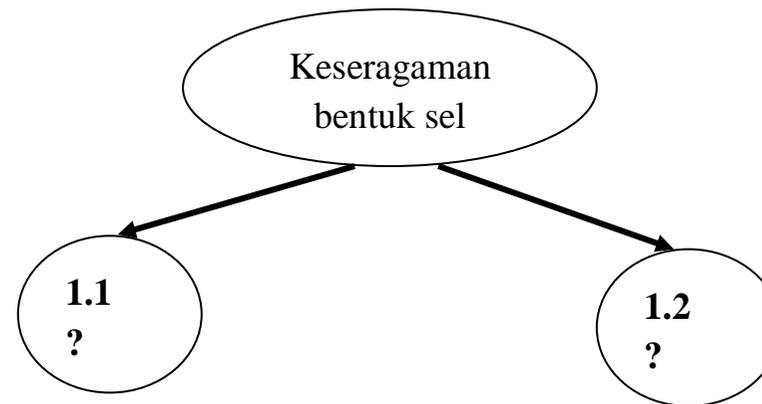


Gambar 4. 7 Model Pohon Keputusan pada Algoritma C4.5

Pohon keputusan terbentuk karena perangkingan nilai gain yang tertinggi hingga yang terendah. Nilai gain yang paling tinggi akan menjadi root note atas simpul akar dan akan memiliki percabangan dibawahnya sampai terbentuknya kelas disemua atributnya.

A. Keseragaman Bentuk Sel sebagai Root Note/ Akar

Berikut adalah Keseragaman Bentuk Sel sebagai Root Note/ Akar :



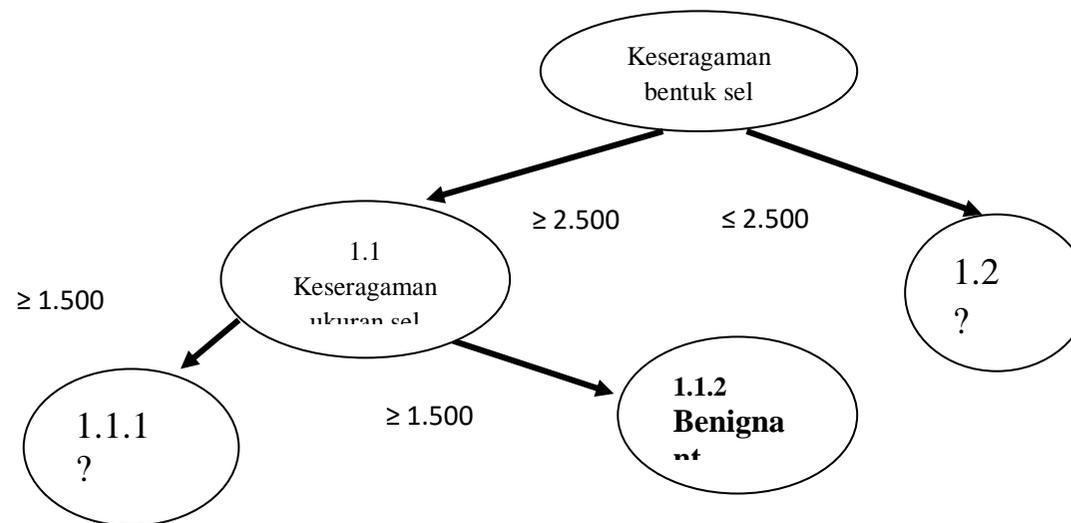
4.8 Gambar root not/ akar

Gambar diatas merupakan pohon keputusan yang dihasilkan dari perhitungan entropy dan gain untuk seluruh atribut. Terlihat bahwa atribut keseragaman bentuk sel menjadi simpul akar karena mempunyai nilai gain tertinggi. Dari simpul akar splittingnya menjadi dua simpul sesuai dengan nilai yang dimilikinya. Dalam atribut tersebut belum dapat menghasilkan class malignant atau benignant. Itu artinya, percabangan yang dilakukan masih ragu – ragu dikarenakan partisi yang dimiliki atribut keseragaman bentuk sel belum

menunjukkan class. Dalam kondisi seperti ini yang harus dilakukan adalah mencari atribut lain sebagai root not / akar selanjutnya sampai seluruh atribut mendapatkan hasil class masing - masing.

A. Keseragaman Ukuran Sel sebagai Note 1

Berikut adalah Keseragaman ukuran Sel sebagai Note 1 :

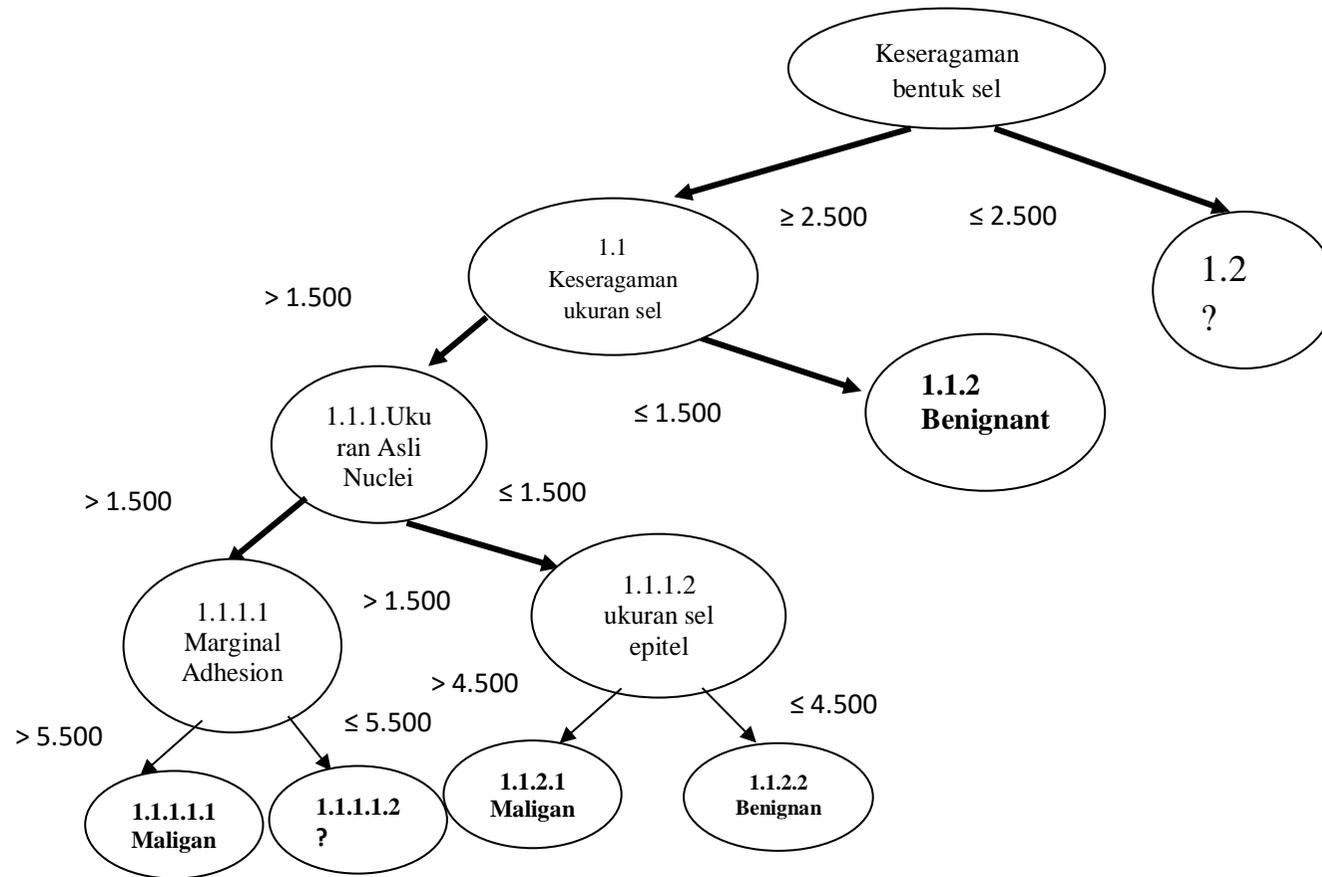


4.9 Gambar Note 1

Terlihat bahwa atribut keseragaman ukuran sel menjadi cabang dari simpul akar atribut keseragaman bentuk sel karena memiliki gain lebih tinggi dari atribut lainnya setelah atribut keseragaman bentuk sel. Langkah selanjutnya kita tentukan akar dari simpul sebelah kiri terlebih dahulu dengan range/ batas ≥ 2.500 nilai. Setelah atribut dari nilai ≥ 2.500 data terbentuk semua kelasnya baru kita akan tentukan ranting selanjutnya pada posisi sebelah kanan. Kemudian dari atribut Keseragaman ukuran sel memiliki 2 ranting cabang yaitu dengan penomoran 1.1.1 disebelah kiri dan 1.1.2 disebelah kanan. Untuk ranting simpul no. 1.1.2 atribut dapat menghasilkan kelas yaitu benignant. Jadi, untuk atribut no. 1.1.2 tidak dibuatkan percabangan kembali dikarenakan kelas telah terbentuk. Berbeda dengan no. 1.1.1 dikarenakan kelas belum dihasilkan maka kita harus mencari nilai atribut dengan gain yang tertinggi selanjutnya.

C. Ukuran asli nuclei sebagai Note 2

Berikut adalah Ukuran Asli Nuclei sebagai Note 2 :



Gambar 4.10 Note 2

Selanjutnya adalah ukuran asli nuclei menjadi atribut yang dipilih menjadi simpul berikutnya dari atribut tersebut menghasilkan dua simpul dimana simpul sebelah kiri dengan no. 1.1.1.1 Marginal Adhesion menghasilkan simpul sebelah kiri dengan kelas malignant sementara simpul sebelah kanan belum mendapatkan hasil kelas sehingga harus mencari atribut terdekat dengan gain yang tertinggi, sementara untuk simpul bagian kanan dengan no. 1.1.1.2 ukuran sel epitel tunggal telah menghasilkan dua atribut yaitu simpul sebelah kiri dengan kelas malignant dan simpul sebelah kanan benignant. Untuk perhitungan selanjutnya dapat dibuat pohon keputusan sampai semua atribut menghasilkan kelas yaitu malignant dan benignant. Berikut adalah struktur yang terbentuk oleh pohon keputusan dari mulai gain yang tertinggi sampai dengan yang terendah dengan melihat text view pada pohon keputusan dataset kanker payudara :

Kemudian atribut kromatin dan mitosis memiliki nilai gain terendah dari atribut yang lain, sehingga atribut tersebut tidak memiliki pengaruh besar terhadap hasil. Dalam penelitian ini atribut tersebut dihilangkan, kemudian dataset yang baru setelah dilakukanya seleksi fitur maka akan diuji kembali guna mengimprove akurasi dengan menggunakan algoritma yang berbeda yaitu Support Vector Mechine.

Selanjutnya adalah ukuran asli nuclei menjadi atribut yang dipilih menjadi simpul berikutnya dari atribut tersebut menghasilkan dua simpul dimana simpul sebelah kiri dengan no. 1.1.1.1 Marginal Adhesion menghasilkan simpul sebelah kiri dengan kelas malignant sementara simpul sebelah kanan belum mendapatkan hasil kelas sehingga harus mencari atribut terdekat dengan gain yang tertinggi, sementara untuk simpul bagian kanan dengan no. 1.1.1.2 ukuran sel epitel tunggal telah menghasilkan dua atribut yaitu simpul sebelah kiri dengan kelas malignant dan simpul sebelah kanan benignant. Untuk perhitungan selanjutnya dapat dibuat pohon keputusan sampai semua atribut menghasilkan kelas yaitu malignant dan benignant. Berikut adalah struktur yang terbentuk oleh pohon keputusan dari mulai gain yang tertinggi sampai dengan yang terendah dengan melihat text view pada pohon keputusan dataset kanker payudara :

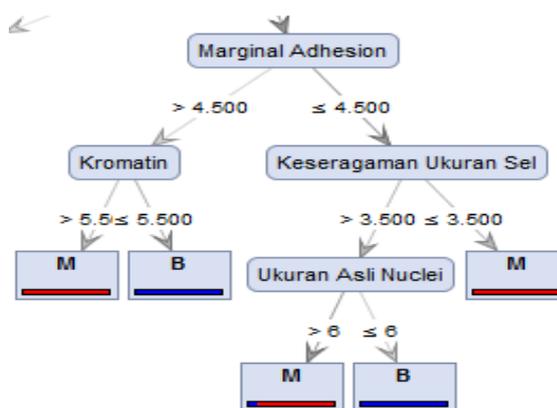
```

○ Graph View ● Text View ○ Annotations
Keseiragaman bentuk Sel > 2.500
|
|--- Keseiragaman Ukuran Sel > 1.500
|   |
|   |--- Ukuran Asli Nuclei > 1.500
|   |   |
|   |   |--- Marginal Adhesion > 5.500; M (B=0, M=79)
|   |   |--- Marginal Adhesion ≤ 5.500
|   |   |   |
|   |   |   |--- Ketebalan Gumpalan > 6.500
|   |   |   |   |
|   |   |   |   |--- Ukuran Asli Nuclei > 2.500
|   |   |   |   |   |
|   |   |   |   |   |--- Marginal Adhesion > 4.500
|   |   |   |   |   |   |
|   |   |   |   |   |   |--- Mitosis > 1.500; B (B=1, M=1)
|   |   |   |   |   |   |--- Mitosis ≤ 1.500; M (B=0, M=6)
|   |   |   |   |   |   |--- Marginal Adhesion ≤ 4.500; M (B=0, M=58)
|   |   |   |   |   |   |--- Ukuran Asli Nuclei ≤ 2.500; M (B=1, M=3)
|   |   |   |   |   |--- Ketebalan Gumpalan ≤ 6.500
|   |   |   |   |   |   |
|   |   |   |   |   |   |--- Keadaan Nucleoli Normal > 7.500; M (B=0, M=7)
|   |   |   |   |   |   |--- Keadaan Nucleoli Normal ≤ 7.500
|   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |--- Keadaan Nucleoli Normal > 6.500; B (B=2, M=0)
|   |   |   |   |   |   |   |--- Keadaan Nucleoli Normal ≤ 6.500
|   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |--- Marginal Adhesion > 4.500
|   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |--- Kromatin > 5.500; M (B=0, M=2)
|   |   |   |   |   |   |   |   |   |--- Kromatin ≤ 5.500; B (B=3, M=0)
|   |   |   |   |   |   |   |   |--- Marginal Adhesion ≤ 4.500
|   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |--- Keseiragaman Ukuran Sel > 3.500
|   |   |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |--- Ukuran Asli Nuclei > 6; M (B=1, M=7)
|   |   |   |   |   |   |   |   |   |   |--- Ukuran Asli Nuclei ≤ 6; B (B=3, M=0)
|   |   |   |   |   |   |   |   |   |--- Keseiragaman Ukuran Sel ≤ 3.500; M (B=0, M=11)
|   |   |   |   |   |   |   |--- Ukuran Asli Nuclei ≤ 1.500
|   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |--- Ukuran Sel Epitel Tunggal > 4.500; M (B=0, M=11)
|   |   |   |   |   |   |   |   |--- Ukuran Sel Epitel Tunggal ≤ 4.500; B (B=10, M=0)
|   |   |   |   |   |   |--- Keseiragaman Ukuran Sel ≤ 1.500; B (B=13, M=1)
|   |--- Ketebalan bentuk Sel ≤ 2.500
|   |   |
|   |   |--- Ketebalan Gumpalan > 8.500; M (B=0, M=4)
|   |   |--- Ketebalan Gumpalan ≤ 8.500
|   |   |   |
|   |   |   |--- Marginal Adhesion > 7; M (B=0, M=2)
|   |   |   |--- Marginal Adhesion ≤ 7
|   |   |   |   |
|   |   |   |   |--- Ukuran Asli Nuclei > 3.500
|   |   |   |   |   |
|   |   |   |   |   |--- Kromatin > 1.500
|   |   |   |   |   |   |
|   |   |   |   |   |   |--- Ketebalan Gumpalan > 3.500; M (B=0, M=2)
|   |   |   |   |   |   |--- Ketebalan Gumpalan ≤ 3.500; B (B=3, M=0)
|   |   |   |   |   |--- Kromatin ≤ 1.500; B (B=5, M=0)
|   |   |   |--- Ukuran Asli Nuclei ≤ 3.500; B (B=242, M=0)

```

4.11 Gambar Text View pada Pohon Keputusan

Dari gambar tersebut terlihat jelas nilai perankingan dari gain tertinggi hingga terendah sampai hingga akhirnya atribut kromatin dan ukuran asli nuclei memiliki nilai gain terendah dari atribut yang lain, sehingga atribut tersebut tidak memiliki pengaruh besar terhadap hasil. Berikut adalah hasil pohon keputusan untuk atribut yang akan dihilangkan karena memiliki nilai gain terendah dibandingkan dengan atribut yang lain.



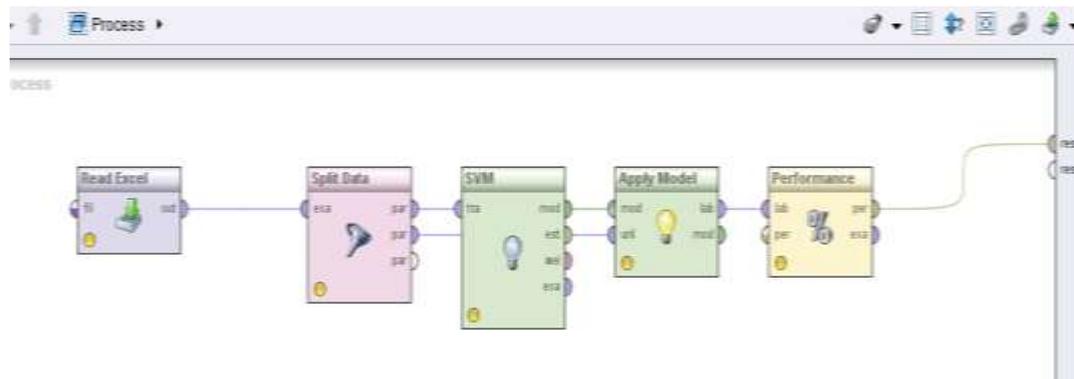
Gambar 4.12 Pohon keputusan sebagai seleksi fitur

Dalam penelitian ini atribut tersebut dihilangkan, kemudian dataset yang baru setelah dilakukanya seleksi fitur maka akan diuji kembali guna mengimprove akurasi dengan menggunakan algoritma yang berbeda yaitu Support Vector Mechine.

4.5 Penerapan metode seleksi fitur pada klasifikasi SVM

4.5.1 Blok Process algoritma SVM

Pengujian yang Kedua dataset dari jumlah data 683 record dengan 7 Atribut dan 1 sebagai Class. Berikut ini merupakan model algoritma dari klasifikasi menggunakan RapidMiner 5.3 setelah dilakukan Feature Selection oleh Deccision Tree dengan Model yang sama dengan menggunakan algoritma SVM sebagai berikut:



Gambar 4.13 Block Algoritma SVM

4.5.2 Hasil pengujian

Proses data mining dilakukan dengan bantuan perangkat lunak datamining, yaitu RapidMiner. Metode yang digunakan yaitu feature selection dengan menggunakan Algoritma C4.5. Pengujian algoritma untuk mengimprove akurasi menggunakan algoritma SVM. Dari data tersebut tersebut dilihat hasil dari akurasi dengan pembagian ratio sebesar 70 : 30 , 50 : 50 serta 80:20. Hasil dari data keseluruhan 683 record dari dataset kanker payudara. Dengan 7 Atribut dan 1 atribut sebagai Class dapat dilihat melalui gambar berikut ini :

A. Accuracy 1

Berikut adalah Hasil dari Accuracy setelah dilakukan seleksi fitur :

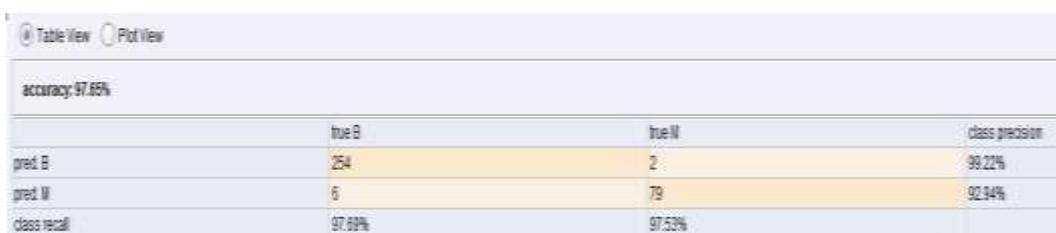
	true B	true M	class precision
pred. B	158	0	100.00%
pred. M	2	45	95.74%
class recall	98.75%	100.00%	

Gambar 4.14 Hasil Akurasi 1

Klasifikasi SVM pada penelitian ini menggunakan aplikasi program rapidminer dengan Data training dan testing sebesar 70:30. Ketepatan klasifikasi terbesar yang dihasilkan oleh metode SVM dari partisi data Training dan testing 70:30 yaitu sebesar 99,02 %. Itu artinya tingkat keakurasian meningkat dan dengan adanya seleksi fitur menggunakan algoritma Decision Tree maka dapat diimprove tingkat keakurasiannya meskipun dengan metode yang sama yaitu Support Vector Mechine (SVM).

B. Accuracy 2

Berikut adalah Hasil dari Accuracy setelah dilakukan seleksi fitur :



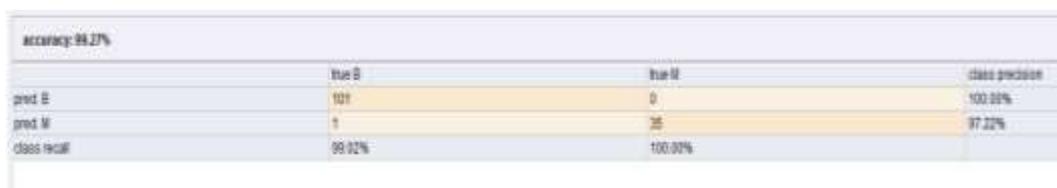
accuracy: 97.65%			
	true B	true M	class precision
pred B	254	2	99.22%
pred M	6	79	92.94%
class recall	97.69%	97.53%	

Gambar 4.11 Hasil Akurasi 2

Hasil dari Precision setelah dilakukan seleksi fitur menunjukkan angka 97,65 % dengan pembagian Data training dan testing sebesar 50:50.

C. Accuracy 3

Berikut adalah Hasil dari Accuracy setelah dilakukan seleksi fitur :



accuracy: 99.27%			
	true B	true M	class precision
pred B	101	0	100.00%
pred M	1	35	97.22%
class recall	99.02%	100.00%	

Gambar 4.12 Analisis Akurasi 3

Hasil dari Precision setelah dilakukan seleksi fitur menunjukkan angka 99,27 % dengan pembagian Data training dan testing sebesar 80:20.

4.2.4 Hasil Pengujian

Berikut adalah Hasil dari pengujian sebelum dan setelah dilakukan seleksi fitur :

Tabel 4.1 Hasil Pengujian

No.	Algoritma	Accuracy 50 : 50	Accuracy 70 : 30	Accuracy 80 : 20	Rata - rata
1	Support Vector Mechine	97.56 %	97.14 %	95.77 %	97,56 %
2	Decision Tree C4.5 + SVM	99.02 %	100 %	95.74 %	98, 65

Dari hasil pengujian pada penelitian ini, dapat dilihat bahwa performansi akurasi SVM meningkat ketika dilakukannya fitur seleksi oleh algoritma decision tree C4.5 pada dataset diagnose Kanker Payudara dari nilai akurasi 97.56 % menjadi 99.02 % kemudian dari hasil recall pun meningkat dari 97.14 % menjadi 100 %, tetapi untuk nilai Precision menurun dari 95.77 % menjadi 95.74 % hal ini disebabkan karena ada pengaruh terhadap class maligant pada saat memanggil data ternyata ada 2 data yang terindikasi classnya benignant dari 45 data yang dimiliki oleh malignant. Sementara Nilai AUC sebelum dilakukan seleksi fitur menunjukkan angka 0.996 setelah digunakan seleksi fitur hasilnya pun meningkat menjadi 1.000 itu artinya bisa diklasifikasikan sebagai excellent classification.

BAB V

SIMPULAN DAN SARAN

5.1 Simpulan

Berdasarkan hasil penelitian yang telah dilakukan, maka dapat disimpulkan sebagai berikut :

5.1.1. Akurasi SVM

Akurasi SVM meningkat ketika dilakukannya seleksi fitur oleh algoritma decision tree C4.5 pada dataset diagnosa Kanker Payudara dari nilai akurasi 97.56 % menjadi 99.02 %.

5.1.2. Recall algoritma Deccision Tree C4.5

Hasil Recall meningkat ketika dilakukannya seleksi fitur oleh algoritma decision tree C4.5 pada dataset diagnosa Kanker Payudara dari nilai 97.14 menjadi 100 %, kemudian untuk nilai AUC sebelum dilakukan seleksi fitur menunjukkan angka 0.996 setelah digunakan seleksi fitur hasilnya pun meningkat menjadi 1.000 itu artinya bisa diklasifikasikan sebagai excellent classification.

5.1.3. Precision Algoritma C4.5

Precision Algoritma C4.5 menurun dari 95.77 % menjadi 95.74 % hal ini disebabkan karena ada pengaruh terhadap class maligant pada saat memanggil data ternyata ada 2 data yang terindikasi classnya benignant dari 45 data yang dimiliki oleh malignant.

5.2 Saran

Dataset public masih terdapat eror sehingga diharapkan nantinya akan diperoleh analisis yang lebih tepat. Selain itu untuk metode Support Vector Machine dalam pengimprove akurasi dapat mengubah parameter sampling linier disesuaikan dengan dataset. Parameter SVM sebaiknya tidak menggunakan trial and error agar

efisien dan menghasilkan akurasi yang optimum. Namun apabila data missing value tersebut tidak dapat dihindarkan maka untuk penelitian selanjutnya dapat dilakukan pengembangan metode SVM untuk data missing value dan penentuan parameter SVM tanpa trial and error yang diharapkan nantinya akan memberikan akurasi yang lebih tinggi.

DAFTAR PUSTAKA

Agresti, A. (2002). *Categorical Data Analysis, Second Edition*. John Willey & Sons, New York.

Andi Publishing. D. T. Larose, (2005). *Algoritma Data Mining.: Discovering Knowledge in Data: An Introduction to Data Mining*. United States of America: John Wiley & Sons, Inc, Yogyakarta.

Daniel T. Larose, (2005). *Discovering Knowledge in Data: an Introduction to Data Mining*, *John Wiley & Sons*,

Dinas Kesehatan Nasional.(2007). Data penderita kanker payudara di Indonesia. Diakses pada tanggal 31 januari 2011 dari [<http://www.depkes.go.id/index.php/berita/press-release/1060-jika-tidak-dikendalikan-26-juta-orang-di-dunia-menderita-kanker-.html>]

Ellis, E.O., Schnitt, S.J., S.-Garau, X., Bussolati, G., Tavassaoli, F.A., Eusebi, (2003). *Pathology and Genetic of Tumours of The Breast and Female Genital Organs / WHO Classification of Tumours*. Washington: IARC Press; 2.P.10,34-6.

Ethem Alpaydin, (2014). *Introduction to Machine Learning*, 3rd ed., *MIT Press*,

Florin Gorunescu, (2011). *Data Mining: Concepts, Models and Techniques*, *Springer*,

F. Gorunescu, (2011), *Data Mining Concept Model Technique*. Romania: Springer.

Gunn, Steve. (1998). *Support Vector Machine for Clasification and Regression*. Southampton: University of Southaton.

Hosmer, D. W., dan Lemeshow, S. (2000). *Applied Logistic Regression.:* John Wiley & Sons, Inc. New York

Ian H. Witten, Frank Eibe, Mark A. (2011). *Hall, Data mining: Practical Machine Learning Tools and Techniques 3rd Edition*, *Elsevier*,

J. Han and M. Kamber, (2006). *Data Mining Concept dan Techniques*, 2nd ed. United States of America: Diane Cerra.

Jiawei Han and Micheline Kamber, (2012). *Data Mining: Concepts and Techniques Third Edition*, *Elsevier*.

Kardinah (2002). *Penatalaksanaan Kanker Payudara Terkini oleh Penanggulangan & Pelayanan Kanker Payudara Terpadu Paripurna R.S. Kanker Dharmais*. Jakarta: Pustaka Populer Obor.

Keles, A., Keles, A., dan Yavuz, U. (2011). *Expert System Based On Neuro-Fuzzy Rules For Diagnosis Breast Cancer*.

Kusrini, & Luthfi, E. T (2006). *E. Technical and P. Series, Guidelines for management of breast cancer*. World Health Organization.

Laily Hermawanti, "Penerapan Algoritma Klasifikasi C4.5 Untuk Diagnosis Penyakit Kanker Payudara," *JURNAL SAINS DAN SENI ITS Vol. 7 No. 2, Hal 57-64 Maret 2012 Vol. 7 No. 2, Hal 57-64*

Larose, D., (2006), *Data Mining Method And Model*, Canada, Inc. Hoboken, New Jersey.

Markus Hofmann and Ralf Klinkenberg, (2014) *RapidMiner: Data Mining Use Cases and Business Analytics Applications*, *CRC Press Taylor & Francis Group*,

Nugroho, A.S., Witarto, A.B., Handoko, D., (2003). *Support Vector Machine – Teori dan Aplikasi dalam Bioinformatika*. Diakses pada tanggal 9 Maret 2012 dari <http://www.ilmukomputer.com>

Purnami, S. W., dan Embong, A. (2008). *Smooth Support Vector Machine For Breast Cancer Classification*.

Santosa, B. (2006). *Data Mining: Teknik Pemanfaatan Data Untuk Keperluan Bisnis.*: Graha Ilmu. Yogyakarta

The 4th IMT-GT 2008 Conference on Mathematics, Statistics, and Their Applications (ICMSA08), Banda Aceh, Indonesia.

Wang, D., Shi, L., dan Heng, P. A. (2009). *Automatic Detectiom of Breast Cancer in Mammogrmas using Support Vector Machines*. *Neurocomputing* 72.3296-3302.

WHO. (2005). Data penderita kanker payudara di dunia. Dikases pada tanggal 3 Februari 2012 dari [<http://www.who.int/cancer/detection/braestcancer/en/index1.html>].

LAMPIRAN ORIGINAL DATASET BREAST CANCER MACHINE LEARNING
REPOSITORY

Ketebalan Gumpalan	Keseragaman Ukuran Sel	Keseragaman bentuk Sel	Marginal Adhesion	Ukuran Sel Epitel Tunggal	Kromatin	Keadaan Nucleoli Normal	Mitosis	Class
5	1	1	1	2	3	1	1	B
5	4	4	5	7	3	2	1	B
3	1	1	1	2	3	1	1	B
6	8	8	1	3	3	7	1	B
4	1	1	3	2	3	1	1	B
8	10	10	8	7	9	7	1	M
1	1	1	1	2	3	1	1	B
2	1	2	1	2	3	1	1	B
2	1	1	1	2	1	1	5	B
4	2	1	1	2	2	1	1	B
1	1	1	1	1	3	1	1	B
2	1	1	1	2	2	1	1	B
5	3	3	3	2	4	4	1	M
1	1	1	1	2	3	1	1	B
8	7	5	10	7	5	5	4	M
7	4	6	4	6	4	3	1	M
4	1	1	1	2	2	1	1	B
4	1	1	1	2	3	1	1	B
10	7	7	6	4	4	1	2	M
6	1	1	1	2	3	1	1	B
7	3	2	10	5	5	4	4	M
10	5	5	3	6	7	10	1	M
3	1	1	1	2	2	1	1	B
1	1	1	1	2	3	1	1	B
5	2	3	4	2	3	6	1	M
3	2	1	1	1	2	1	1	B
5	1	1	1	2	2	1	1	B
2	1	1	1	2	2	1	1	B
1	1	3	1	2	1	1	1	B
3	1	1	1	1	2	1	1	B
2	1	1	1	2	3	1	1	B
10	7	7	3	8	7	4	3	M
2	1	1	2	2	3	1	1	B
3	1	2	1	2	2	1	1	B
2	1	1	1	2	2	1	1	B
10	10	10	8	6	8	9	1	M

6	2	1	1	1	7	1	1	B
5	4	4	9	2	5	6	1	M
2	5	3	3	6	7	5	1	M
10	4	3	1	3	6	5	2	M
6	10	10	2	8	7	3	3	M
5	6	5	6	10	3	1	1	M
10	10	10	4	8	8	10	1	M
1	1	1	1	2	2	1	2	B
3	7	7	4	4	4	8	1	M
1	1	1	1	2	2	1	1	B
4	1	1	3	2	3	1	1	B
7	8	7	2	4	3	8	2	M
9	5	8	1	2	2	1	5	M
5	3	3	4	2	3	4	1	M
10	3	6	2	3	4	10	2	M
5	5	5	8	10	7	3	7	M
10	5	5	6	8	7	1	1	M
10	6	6	3	4	3	6	1	M
8	10	10	1	3	3	9	1	M
8	2	4	1	5	5	4	4	M
5	2	3	1	6	5	1	1	M
9	5	5	2	2	5	1	1	M
5	3	5	5	3	4	10	1	M
1	1	1	1	2	2	1	1	B
9	10	10	1	10	3	3	1	M
6	3	4	1	5	3	9	1	M
1	1	1	1	2	2	1	1	B
10	4	2	1	3	4	3	10	M
4	1	1	1	2	3	1	1	B
5	3	4	1	8	4	9	1	M
8	3	8	3	4	8	9	8	M
1	1	1	1	2	3	2	1	B
5	1	3	1	2	2	1	1	B
6	10	2	8	10	7	8	10	M
1	3	3	2	2	7	2	1	B
9	4	5	10	6	4	8	1	M
10	6	4	1	3	3	2	3	M
1	1	2	1	2	4	2	1	B

1	1	4	1	2	2	1	1	B
5	3	1	2	2	2	1	1	B
3	1	1	1	2	3	1	1	B
2	1	1	1	3	2	1	1	B
2	2	2	1	1	7	1	1	B
4	1	1	2	2	2	1	1	B
5	2	1	1	2	3	1	1	B
3	1	1	1	2	7	1	1	B
3	5	7	8	8	7	10	7	M
5	10	6	1	10	4	10	10	M
3	3	6	4	5	4	4	1	M
3	6	6	6	5	6	8	3	M
4	1	1	1	2	3	1	1	B
2	1	1	2	3	2	1	1	B
1	1	1	1	2	3	1	1	B
3	1	1	2	2	1	1	1	B
4	1	1	1	2	3	1	1	B
1	1	1	1	2	2	1	1	B
2	1	1	1	2	3	1	1	B
1	1	1	1	2	3	1	1	B
2	1	1	2	2	1	1	1	B
5	1	1	1	2	3	1	1	B
9	6	9	2	10	2	9	10	M
7	5	6	10	5	7	9	4	M
10	3	5	1	10	3	10	2	M
2	3	4	4	2	2	5	1	M
4	1	2	1	2	3	1	1	B
8	2	3	1	6	7	1	1	M
10	10	10	10	10	8	8	8	M
7	3	4	4	3	3	2	7	M
10	10	10	8	2	4	1	1	M
1	6	8	10	8	5	7	1	M
1	1	1	1	2	2	3	1	B
6	5	4	4	3	7	8	3	M
1	3	1	2	2	5	3	2	B
8	6	4	3	5	3	1	1	M
10	3	3	10	2	7	3	3	M
10	10	10	3	10	8	1	1	M

3	3	2	1	2	3	1	1	B
1	1	1	1	2	1	1	1	B
8	3	3	1	2	3	2	1	B
4	5	5	10	4	7	5	8	M
1	1	1	1	4	1	1	1	B
3	2	1	1	2	3	1	1	B
1	1	2	2	2	3	1	1	B
4	2	1	1	2	3	1	1	B
10	10	10	2	10	5	3	3	M
5	3	5	1	8	5	3	1	M
5	4	6	7	9	8	10	1	M
1	1	1	1	2	2	1	1	B
7	5	3	7	4	7	5	5	M
3	1	1	1	2	3	1	1	B
8	3	5	4	5	1	6	2	M
1	1	1	1	10	1	1	1	B
5	1	3	1	2	2	1	1	B
2	1	1	1	2	3	1	1	B
5	10	8	10	8	3	6	3	M
3	1	1	1	2	2	2	1	B
3	1	1	1	3	2	1	1	B
5	1	1	1	2	3	3	1	B
4	1	1	1	2	2	1	1	B
3	1	1	1	2	1	1	1	B
4	1	2	1	2	2	1	1	B
3	1	1	1	2	1	1	1	B
2	1	1	1	2	1	1	1	B
9	5	5	4	4	4	3	3	M
1	1	1	1	2	1	1	1	B
2	1	1	1	2	2	1	1	B
3	4	5	2	6	4	1	1	M
1	1	1	1	3	2	1	1	B
3	1	1	3	8	5	8	1	B
8	8	7	4	10	7	8	7	M
1	1	1	1	1	3	1	1	B
7	2	4	1	6	5	4	3	M
10	10	8	6	4	8	10	1	M
4	1	1	1	2	1	1	1	B

1	1	1	1	2	1	1	1	B
5	5	5	6	3	3	1	1	M
1	2	2	1	2	2	1	1	B
2	1	1	1	2	3	1	1	B
9	9	10	3	6	7	10	6	M
10	7	7	4	5	5	7	2	M
4	1	1	1	2	3	2	1	B
3	1	1	1	2	3	1	1	B
1	1	1	2	1	1	1	7	B
4	1	1	1	2	3	2	1	B
5	6	7	8	8	3	10	3	M
10	8	10	10	6	3	1	10	M
3	1	1	1	2	3	1	1	B
1	1	1	2	1	1	1	1	B
3	1	1	1	2	1	1	1	B
1	1	1	1	2	3	1	1	B
1	1	1	1	2	2	1	1	B
6	10	10	10	8	10	10	7	M
8	6	5	4	3	6	1	1	M
5	8	7	7	10	5	7	1	M
2	1	1	1	2	3	1	1	B
5	10	10	3	8	5	10	3	M
4	1	1	1	2	3	1	1	B
5	3	3	3	6	3	1	1	M
1	1	1	1	1	3	1	1	B
1	1	1	1	2	1	1	1	B
6	1	1	1	2	3	1	1	B
5	8	8	8	5	7	8	1	M
8	7	6	4	4	5	1	1	M
2	1	1	1	1	3	1	1	B
1	5	8	6	5	7	10	1	M
10	5	6	10	6	7	7	10	M
5	8	4	10	5	9	10	1	M
1	2	3	1	2	3	1	1	B
10	10	10	8	6	7	10	1	M
7	5	10	10	10	4	10	3	M
5	1	1	1	2	2	1	1	B
1	1	1	1	2	3	1	1	B

3	1	1	1	2	3	1	1	B
4	1	1	1	2	3	1	1	B
8	4	4	5	4	7	8	2	B
5	1	1	4	2	3	1	1	B
1	1	1	1	2	1	1	1	B
3	1	1	1	2	2	1	1	B
9	7	7	5	5	7	8	3	M
10	8	8	4	10	8	1	1	M
1	1	1	1	2	3	1	1	B
5	1	1	1	2	3	1	1	B
1	1	1	1	2	3	1	1	B
5	10	10	9	6	7	10	5	M
10	10	9	3	7	3	5	1	M
1	1	1	1	1	3	1	1	B
1	1	1	1	1	3	1	1	B
5	1	1	1	1	3	1	1	B
8	10	10	10	5	8	10	6	M
8	10	8	8	4	7	7	1	M
1	1	1	1	2	3	1	1	B
10	10	10	10	7	7	10	4	M
10	10	10	10	3	10	6	1	M
8	7	8	7	5	5	10	2	M
1	1	1	1	2	2	1	1	B
1	1	1	1	2	3	1	1	B
6	10	7	7	6	8	10	2	M
6	1	3	1	2	3	1	1	B
1	1	1	2	2	3	1	1	B
10	6	4	3	10	9	10	1	M
4	1	1	3	1	2	1	1	M
7	5	6	3	3	7	4	1	M
10	5	5	6	3	7	9	2	M
1	1	1	1	2	2	1	1	B
10	5	7	4	4	8	9	1	M
8	9	9	5	3	7	7	1	M
1	1	1	1	1	3	1	1	B
10	10	10	3	10	9	10	1	M
7	4	7	4	3	7	6	1	M
6	8	7	5	6	8	9	2	M

8	4	6	3	3	4	3	1	B
10	4	5	5	5	4	1	1	M
3	3	2	1	3	3	6	1	B
10	8	8	2	8	4	8	10	M
9	8	8	5	6	4	10	4	M
8	10	10	8	6	3	10	10	M
10	4	3	2	3	5	3	2	M
5	1	3	3	2	2	3	1	B
3	1	1	3	1	3	1	1	B
2	1	1	1	2	3	1	1	B
1	1	1	1	2	5	1	1	B
1	1	1	1	2	3	1	1	B
5	1	1	2	2	3	1	1	B
8	10	10	8	5	7	8	1	M
8	4	4	1	2	3	3	1	M
4	1	1	1	2	3	6	1	B
1	2	2	1	2	1	1	1	B
10	4	4	10	2	5	3	3	M
6	3	3	5	3	3	5	3	B
6	10	10	2	8	7	3	3	M
9	10	10	1	10	3	3	1	M
5	6	6	2	4	3	6	1	M
3	1	1	1	2	1	1	1	B
3	1	1	1	2	2	1	1	B
3	1	1	1	2	3	1	1	B
5	7	7	1	5	3	4	1	B
10	5	8	10	3	5	1	3	M
5	10	10	6	10	10	6	5	M
8	8	9	4	5	7	8	1	M
10	4	4	10	6	5	5	1	M
7	9	4	10	10	5	3	3	M
5	1	4	1	2	3	2	1	B
10	10	6	3	3	4	3	2	M
3	3	5	2	3	7	1	1	M
10	8	8	2	3	8	7	8	M
1	1	1	1	2	3	1	1	B
8	4	7	1	3	3	9	2	M
5	1	1	1	2	3	1	1	B

3	3	5	2	3	7	1	1	M
7	2	4	1	3	3	3	1	M
3	1	1	1	2	3	2	1	B
3	1	1	1	2	2	1	1	B
1	1	1	1	2	2	1	1	B
1	1	1	1	2	3	1	1	B
10	5	7	3	3	3	3	8	M
3	1	1	1	2	3	1	1	B
2	1	1	2	2	3	1	1	B
1	4	3	10	4	5	6	1	M
10	4	6	1	2	5	3	1	M
7	4	5	10	2	3	8	2	M
8	10	10	10	8	10	7	3	M
10	10	10	10	10	4	10	10	M
3	1	1	1	3	2	1	1	B
6	1	3	1	4	5	10	1	M
5	6	6	8	6	4	10	4	M
1	1	1	1	2	1	1	1	B
1	1	1	1	2	3	1	1	B
10	4	4	6	2	2	3	1	M
5	5	7	8	6	7	4	1	M
5	3	4	3	4	4	7	1	B
8	2	1	1	5	1	1	1	B
9	1	2	6	4	7	7	2	M
8	4	10	5	4	7	10	1	M
1	1	1	1	2	3	1	1	B
10	10	10	7	9	7	10	10	M
1	1	1	1	2	3	1	1	B
8	3	4	9	3	3	3	1	M
10	8	4	4	4	3	10	4	M
1	1	1	1	2	3	1	1	B
1	1	1	1	2	3	1	1	B
7	8	7	6	4	8	8	4	M
3	1	1	1	2	5	1	1	B
2	1	1	1	3	2	1	1	B
1	1	1	1	2	1	1	1	B
8	6	4	10	10	3	5	1	M
1	1	1	1	2	1	1	1	B

1	1	1	1	1	2	1	1	B
5	5	5	2	5	4	3	1	M
6	8	7	8	6	8	9	1	M
1	1	1	1	5	3	1	1	B
4	4	4	4	6	7	3	1	B
7	6	3	2	5	7	4	6	M
3	1	1	1	2	3	1	1	B
5	4	6	10	2	4	1	1	M
1	1	1	1	2	3	1	1	B
3	2	2	1	2	2	3	1	B
10	1	1	1	2	5	4	1	M
1	1	1	1	2	2	1	1	B
8	10	3	2	6	3	10	1	M
10	4	6	4	5	7	1	1	M
10	4	7	2	2	6	1	1	M
5	1	1	1	2	3	1	2	B
5	2	2	2	2	2	2	1	B
5	4	6	6	4	4	3	1	M
8	6	7	3	3	3	4	2	M
1	1	1	1	2	1	1	1	B
6	5	5	8	4	3	4	1	M
1	1	1	1	2	3	1	1	B
1	1	1	1	1	2	1	1	B
8	5	5	5	2	4	3	1	M
10	3	3	1	2	7	6	1	M
1	1	1	1	2	3	1	1	B
2	1	1	1	2	1	1	1	B
1	1	1	1	2	1	1	1	B
7	6	4	8	10	9	5	3	M
1	1	1	1	2	1	1	1	B
5	2	2	2	3	1	3	1	B
1	1	1	1	1	1	3	1	B
3	4	4	10	5	3	3	1	M
4	2	3	5	3	7	6	1	M
5	1	1	3	2	1	1	1	B
2	1	1	1	2	3	1	1	B
3	4	5	3	7	4	6	1	B
2	7	10	10	7	4	9	4	M

1	1	1	1	2	2	1	1	B
4	1	1	1	3	2	2	1	B
5	3	3	1	3	3	3	3	M
8	10	10	7	10	7	3	8	M
8	10	5	3	8	4	10	3	M
10	3	5	4	3	3	5	3	M
6	10	10	10	10	8	10	10	M
3	10	3	10	6	5	1	4	M
3	2	2	1	4	2	1	1	B
4	4	4	2	2	2	1	1	B
2	1	1	1	2	3	1	1	B
2	1	1	1	2	2	1	1	B
6	10	10	10	8	7	10	7	M
5	8	8	10	5	8	10	3	M
1	1	3	1	2	1	1	1	B
1	1	3	1	1	2	1	1	B
4	3	2	1	3	2	1	1	B
1	1	3	1	2	1	1	1	B
4	1	2	1	2	2	1	1	B
5	1	1	2	2	2	1	1	B
3	1	2	1	2	2	1	1	B
1	1	1	1	2	1	1	1	B
1	1	1	1	2	2	1	1	B
1	1	1	1	1	2	1	1	B
3	1	1	4	3	2	2	1	B
5	3	4	1	4	3	1	1	B
1	1	1	1	2	1	1	1	B
10	6	3	6	4	7	8	4	M
3	2	2	2	2	3	2	1	B
2	1	1	1	2	1	1	1	B
2	1	1	1	2	1	1	1	B
3	3	2	2	3	1	2	3	B
7	6	6	3	2	7	1	1	M
5	3	3	2	3	3	1	1	B
2	1	1	1	2	2	2	1	B
5	1	1	1	3	2	2	1	B
1	1	1	2	2	2	1	1	B
10	8	7	4	8	3	7	9	M

3	1	1	1	2	2	1	1	B
1	1	1	1	1	1	1	1	B
1	2	3	1	2	2	1	1	B
3	1	1	1	2	2	1	1	B
3	1	1	1	2	3	1	1	B
4	1	1	1	2	1	1	1	B
3	2	1	1	2	2	2	1	B
1	2	3	1	2	1	1	1	B
3	10	8	7	6	9	3	8	M
3	1	1	1	2	1	1	1	B
5	3	3	1	2	2	1	1	B
3	1	1	1	2	1	1	1	B
1	2	1	3	2	1	2	1	B
1	1	1	1	2	2	1	1	B
4	2	2	1	2	2	1	1	B
1	1	1	1	2	2	1	1	B
2	3	2	2	2	3	1	1	B
3	1	2	1	2	2	1	1	B
1	1	1	1	2	2	1	1	B
10	10	10	6	8	8	5	1	M
5	1	2	1	2	3	1	1	B
8	5	6	2	3	6	6	1	M
3	3	2	6	3	3	5	1	B
8	7	8	5	10	7	2	1	M
1	1	1	1	2	2	1	1	B
5	2	2	2	2	3	2	2	B
2	3	1	1	5	1	1	1	B
3	2	2	3	2	3	1	1	B
10	10	10	7	10	8	2	1	M
4	3	3	1	2	3	3	1	B
5	1	3	1	2	2	1	1	B
3	1	1	1	2	1	1	1	B
9	10	10	10	10	10	10	1	M
5	3	6	1	2	1	1	1	B
8	7	8	2	4	5	10	1	M
1	1	1	1	2	2	1	1	B
2	1	1	1	2	2	1	1	B
1	3	1	1	2	2	2	1	B

5	1	1	3	4	3	2	1	B
5	1	1	1	2	2	2	1	B
3	2	2	3	2	1	1	1	B
6	9	7	5	5	4	2	1	B
10	8	10	1	3	5	1	1	M
10	10	10	1	6	2	8	1	M
4	1	1	1	2	1	1	1	B
4	1	3	3	2	1	1	1	B
5	1	1	1	2	1	1	1	B
10	4	3	10	4	10	1	1	M
5	2	2	4	2	1	1	1	B
1	1	1	3	2	1	1	1	B
1	1	1	1	2	1	1	1	B
5	1	1	6	3	2	1	1	B
2	1	1	1	2	1	1	1	B
1	1	1	1	2	1	1	1	B
5	1	1	1	2	1	1	1	B
1	1	1	1	1	1	1	1	B
5	7	9	8	6	8	10	1	M
4	1	1	3	1	2	1	1	B
5	1	1	1	2	1	1	1	B
3	1	1	3	2	1	1	1	B
4	5	5	8	6	10	7	1	M
2	3	1	1	3	1	1	1	B
10	2	2	1	2	1	1	2	M
10	6	5	8	5	8	6	1	M
8	8	9	6	6	10	10	1	M
5	1	2	1	2	1	1	1	B
5	1	3	1	2	1	1	1	B
5	1	1	3	2	1	1	1	B
3	1	1	1	2	1	1	1	B
6	1	1	3	2	1	1	1	B
4	1	1	1	2	1	2	1	B
4	1	1	1	2	1	1	1	B
10	9	8	7	6	7	10	3	M
10	6	6	2	4	9	7	1	M
6	6	6	5	4	7	6	2	M
4	1	1	1	2	1	1	1	B

1	1	2	1	2	2	1	1	B
3	1	1	1	1	2	1	1	B
6	1	1	3	2	1	1	1	B
6	1	1	1	1	1	1	1	B
4	1	1	1	2	1	1	1	B
5	1	1	1	2	1	1	1	B
3	1	1	1	2	1	1	1	B
4	1	2	1	2	1	1	1	B
4	1	1	1	2	1	1	1	B
5	2	1	1	2	1	1	1	B
4	8	7	10	4	7	5	1	M
5	1	1	1	1	1	1	1	B
5	3	2	4	2	1	1	1	B
9	10	10	10	10	10	10	10	M
8	7	8	5	5	9	10	1	M
5	1	2	1	2	1	1	1	B
1	1	1	3	1	1	1	1	B
3	1	1	1	1	2	1	1	B
10	10	10	10	6	8	1	5	M
3	6	4	10	3	3	4	1	M
6	3	2	1	3	4	1	1	M
1	1	1	1	2	1	1	1	B
5	8	9	4	3	7	1	1	M
4	1	1	1	1	2	1	1	B
5	10	10	10	6	6	5	2	M
5	1	2	10	4	2	1	1	B
3	1	1	1	1	2	1	1	B
1	1	1	1	1	1	1	1	B
4	2	1	1	2	1	1	1	B
4	1	1	1	2	2	1	1	B
4	1	1	1	2	2	1	1	B
6	1	1	1	2	3	1	1	B
4	1	1	1	2	2	1	1	B
4	1	1	2	2	2	1	1	B
4	1	1	1	2	3	1	1	B
1	1	1	1	2	1	1	1	B
3	3	1	1	2	1	1	1	B
8	10	10	10	7	4	8	7	M

1	1	1	1	2	1	1	1	B
5	1	1	1	2	1	1	1	B
2	1	1	1	2	1	1	1	B
1	1	1	1	2	1	1	1	B
5	1	1	1	2	2	1	1	B
5	1	1	1	2	1	1	1	B
3	1	1	1	1	2	1	1	B
6	6	7	10	3	8	10	2	M
4	10	4	7	3	9	10	1	M
1	1	1	1	1	1	1	1	B
1	1	1	1	1	2	1	1	B
3	1	2	2	2	1	1	1	B
4	7	8	3	4	9	1	1	M
1	1	1	1	3	1	1	1	B
4	1	1	1	3	1	1	1	B
10	4	5	4	3	7	3	1	M
7	5	6	10	4	5	3	1	M
3	1	1	1	2	2	1	1	B
3	1	1	2	2	1	1	1	B
4	1	1	1	2	1	1	1	B
4	1	1	1	2	3	1	1	B
6	1	3	2	2	1	1	1	B
4	1	1	1	1	2	1	1	B
7	4	4	3	4	6	9	1	M
4	2	2	1	2	2	1	1	B
1	1	1	1	1	3	1	1	B
3	1	1	1	2	2	1	1	B
2	1	1	1	2	2	1	1	B
1	1	3	2	2	3	1	1	B
5	1	1	1	2	3	1	1	B
5	1	2	1	2	3	1	1	B
4	1	1	1	2	2	1	1	B
6	1	1	1	2	2	1	1	B
5	1	1	1	2	2	1	1	B
3	1	1	1	2	1	1	1	B
5	3	1	1	2	1	1	1	B
4	1	1	1	2	2	1	1	B
2	1	3	2	2	2	1	1	B

5	1	1	1	2	2	1	1	B
6	10	10	10	4	7	10	1	M
2	1	1	1	1	1	1	1	B
3	1	1	1	1	1	1	1	B
7	8	3	7	4	7	8	2	M
3	1	1	1	2	2	1	1	B
1	1	1	1	2	3	1	1	B
3	2	2	2	2	4	2	1	B
4	4	2	1	2	2	1	2	B
3	1	1	1	2	1	1	1	B
4	3	1	1	2	4	8	1	B
5	2	2	2	1	2	1	1	B
5	1	1	3	2	1	1	1	B
2	1	1	1	2	2	1	1	B
5	1	1	1	2	2	1	1	B
5	1	1	1	2	3	1	1	B
5	1	1	1	2	3	1	1	B
1	1	1	1	2	3	1	1	B
3	1	1	1	2	2	1	1	B
4	1	1	1	2	3	2	1	B
5	7	10	10	5	10	10	1	M
3	1	2	1	2	3	1	1	B
4	1	1	1	2	2	1	1	B
8	4	4	1	6	2	5	2	M
10	10	8	10	6	10	3	1	M
8	10	4	4	8	8	2	1	M
7	6	10	5	3	9	10	2	M
3	1	1	1	2	2	1	1	B
1	1	1	1	2	2	1	1	B
10	9	7	3	4	7	7	1	M
5	1	2	1	2	3	1	1	B
5	1	1	1	2	2	1	1	B
1	1	1	1	2	2	1	1	B
1	1	1	1	2	2	1	1	B
1	1	1	1	2	3	1	1	B
5	1	2	1	2	2	1	1	B
5	7	10	6	5	7	5	1	M
6	10	5	5	4	6	10	1	M

3	1	1	1	2	1	1	1	B
5	1	1	6	3	1	1	1	B
1	1	1	1	2	1	1	1	B
8	10	10	10	6	10	10	1	M
5	1	1	1	2	2	2	1	B
9	8	8	9	6	4	1	1	M
5	1	1	1	2	1	1	1	B
4	10	8	5	4	10	1	1	M
2	5	7	6	4	7	6	1	M
10	3	4	5	3	4	1	1	M
5	1	2	1	2	1	1	1	B
4	8	6	3	4	7	1	1	M
5	1	1	1	2	2	1	1	B
4	1	2	1	2	2	1	1	B
5	1	3	1	2	3	1	1	B
3	1	1	1	2	2	1	1	B
5	2	4	1	1	1	1	1	B
3	1	1	1	2	2	1	1	B
1	1	1	1	1	2	1	1	B
4	1	1	1	2	2	1	1	B
5	4	6	8	4	8	10	1	M
5	3	2	8	5	8	1	2	M
10	5	10	3	5	7	8	3	M
4	1	1	2	2	1	1	1	B
1	1	1	1	2	1	1	1	B
5	10	10	10	10	10	1	1	M
5	1	1	1	2	1	1	1	B
10	4	3	10	3	7	1	2	M
5	10	10	10	5	8	5	1	M
8	10	10	10	6	10	10	10	M
2	3	1	1	2	2	1	1	B
2	1	1	1	1	2	1	1	B
4	1	3	1	2	2	1	1	B
3	1	1	1	2	2	1	1	B
4	1	1	1	2	2	1	1	B
5	1	1	1	2	2	1	1	B
3	1	1	1	2	2	1	1	B
6	3	3	3	3	6	1	1	B

7	1	2	3	2	2	1	1	B
1	1	1	1	2	1	1	1	B
5	1	1	2	1	2	1	1	B
3	1	3	1	3	1	1	1	B
4	6	6	5	7	7	7	3	M
2	1	1	1	2	1	1	1	B
2	1	1	1	2	1	1	1	B
4	1	1	1	2	1	1	1	B
6	2	3	1	2	1	1	1	B
5	1	1	1	2	2	1	1	B
1	1	1	1	2	1	1	1	B
8	7	4	4	5	5	10	1	M
3	1	1	1	2	1	1	1	B
3	1	4	1	2	1	1	1	B
10	10	7	8	7	10	10	3	M
4	2	4	3	2	2	1	1	B
4	1	1	1	2	1	1	1	B
5	1	1	3	2	1	1	1	B
4	1	1	3	2	1	1	1	B
3	1	1	1	2	2	1	1	B
3	1	1	1	2	2	1	1	B
1	1	1	1	2	1	1	1	B
2	1	1	1	2	1	1	1	B
3	1	1	1	2	2	1	1	B
1	2	2	1	2	1	1	1	B
1	1	1	3	2	1	1	1	B
5	10	10	10	10	10	10	10	M
3	1	1	1	2	2	1	1	B
3	1	1	2	3	1	1	1	B
1	2	1	3	2	2	1	1	B
5	1	1	1	2	2	2	1	B
4	1	1	1	2	2	1	1	B
3	1	1	1	2	3	1	1	B
3	1	1	1	2	2	1	1	B
5	1	1	1	2	2	1	1	B
5	4	5	1	8	3	6	1	B
7	8	8	7	3	7	2	3	M
1	1	1	1	2	1	1	1	B

1	1	1	1	2	2	1	1	B	
4	1	1	1	2	3	1	1	B	
1	1	3	1	2	2	1	1	B	
1	1	3	1	2	2	1	1	B	
3	1	1	3	2	2	1	1	B	
1	1	1	1	2	1	1	1	B	
5	2	2	2	2	1	1	2	B	
3	1	1	1	2	3	1	1	B	
5	7	4	1	6	7	10	3	M	
5	10	10	8	5	7	10	1	M	
3	10	7	8	5	7	4	1	M	
3	2	1	2	2	3	1	1	B	
2	1	1	1	2	3	1	1	B	
5	3	2	1	3	1	1	1	B	
1	1	1	1	2	2	1	1	B	
4	1	4	1	2	1	1	1	B	
1	1	2	1	2	2	1	1	B	
5	1	1	1	2	1	1	1	B	
1	1	1	1	2	1	1	1	B	
2	1	1	1	2	1	1	1	B	
10	10	10	10	5	10	10	7	M	
5	10	10	10	4	5	6	3	M	
5	1	1	1	2	3	2	1	B	
1	1	1	1	2	1	1	1	B	
1	1	1	1	2	1	1	1	B	
1	1	1	1	2	1	1	1	B	
1	1	1	1	2	1	1	1	B	
3	1	1	1	2	2	3	1	B	
4	1	1	1	2	1	1	1	B	
1	1	1	1	2	1	1	8	B	
1	1	1	3	2	1	1	1	B	
5	10	10	5	4	4	4	1	M	
3	1	1	1	2	1	1	1	B	
3	1	1	1	2	2	1	2	B	
3	1	1	1	3	1	1	1	B	
2	1	1	1	2	1	1	1	B	
5	10	10	3	7	8	10	2	M	
4	8	6	4	4	3	10	6	1	M

4	8	8	5	4	10	4	1	M
---	---	---	---	---	----	---	---	---