

**APLIKASI DATA MINING
UNTUK DIAGNOSIS PENYAKIT DIABETES
MENGUNAKAN ALGORITMA C4.5 DAN *NAÏVE BAYES*
*CLASSIFICATION***



SKRIPSI

**Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Komputer
pada Jurusan Ilmu Komputer/ Informatika**

Oleh :

MUHAMMAD RIFQI

24010310141042

**JURUSAN ILMU KOMPUTER / INFORMATIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO**

2016

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini:

Nama : Muhammad Rifqi

NIM : 24010310141042

Judul : Aplikasi Data Mining untuk Diagnosis Penyakit Diabetes Menggunakan Algoritma C4.5 dan *Naive Bayes Classification*

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

Semarang, 14 Maret 2016

Muhammad Rifqi

24010310141042

HALAMAN PENGESAHAN

Judul : Aplikasi Data Mining untuk Diagnosis Penyakit Diabetes Menggunakan
Algoritma C4.5 dan *Naive Bayes Classification*
Nama : Muhammad Rifqi
NIM : 24010310141042

Telah diujikan pada sidang tugas akhir pada tanggal 23 Februari 2016 dan dinyatakan lulus
pada tanggal 11 Maret 2016.

Semarang, 14 Maret 2016

Mengetahui,
Ketua Jurusan Ilmu Komputer/ Informatika
FSM Universitas Diponegoro,

Mengetahui,
Panitia Penguji Tugas Akhir
Ketua,

Ragil Saputra, S.Si, M.Cs
NIP. 19801021 200501 1 003

Beta Noranita, S.Si., M.Kom
NIP. 197308291998022001

HALAMAN PENGESAHAN

Judul : Aplikasi Data Mining untuk Diagnosis Penyakit Diabetes Menggunakan
Algoritma C4.5 dan *Naive Bayes Classification*
Nama : Muhammad Rifqi
NIM : 24010310141042

Telah diujikan pada sidang tugas akhir pada tanggal 23 Februari 2016.

Semarang, 14 Maret 2016
Pembimbing,

Nurdin Bahtiar, M.T.
NIP. 19790720 200312 1 002

ABSTRAK

Database rumah sakit umumnya mengandung data yang sangat banyak dengan berbagai atribut. Pemanfaatan data yang sangat banyak untuk dijadikan informasi yang relevan dari database yang besar adalah pekerjaan yang sulit. Hal tersebut memerlukan teknik-teknik tertentu sehingga penyaringan informasi dapat dilakukan dengan efektif dan efisien. Misalnya menerapkan teknik data mining yang dapat menelusuri pola data untuk tujuan analisis. Penelitian ini mempelajari bagaimana data mining dapat diterapkan untuk membantu memprediksi penyakit diabetes dari data laboratorium. Metode yang digunakan dalam penelitian ini adalah Algoritma C4.5 dan *Naive Bayes Classification*. Dalam penerapan Algoritma C4.5 memiliki tahapan-tahapan yaitu : menghitung *entropy*, *gain*, *split info* dan *ratio gain*. Langkah selanjutnya adalah membentuk *decision tree* dan menghasilkan *rule*. *Naive Bayes Classification* memiliki tahapan-tahapan yaitu menghitung probabilitas kelas dan menghitung probabilitas atribut. Pengujian dilakukan menggunakan data Rumah Sakit Pusat Pertamina dengan 3876 data dengan rincian 3.101 data pelatihan dan 775 data pengujian yang masing-masing dengan 12 variabel: umur, jenis kelamin dan hasil uji laboratorium. Hasil menunjukkan bahwa Algoritma C4.5 mempunyai akurasi prediksi maksimum untuk diabetes sebesar 75.54%, dan *Naive Bayes Classification* sebesar 75.35%.

Kata Kunci: Data mining, *Decison tree*, Algoritma C4.5, *Naive Bayes Classification*

ABSTRACT

Hospital database generally contains many data with different attributes. The utilization of many data to be the relevant information from the large databases is a difficult job. It requires certain techniques so that the information can be filtering effectively and efficiently. For example applying data mining techniques to discover patterns of data for analysis purposes. This research studied how data mining can be applied to help predict diabetes from laboratory data. The method used in this research is the algorithm C4.5 and Naive Bayes Classification. In applying of the algorithm C4.5 has stages, these are calculating the entropy, the gain, the split info and the ratio gain. The next step is creating the decision tree and generating rule. Naive Bayes Classification has stages, these are calculating the class probability and calculating the attributes of probability. The testing is using the Pertamina Central Hospital's data with 3876 data by the details 3101 training data and 775 test data, which each have 12 variables: age, sex and laboratory test results. The results showed that the algorithm C4.5 has a maximum prediction accuracy for diabetes is 75.54%, and *Naive Bayes Classification* is 75.35%.

Keywords: Data mining, *Decison tree*, Algoritma C4.5, *Naïve Bayes Classification*

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan tugas akhir yang berjudul “Aplikasi Data Mining untuk Diagnosis Penyakit Diabetes Menggunakan Algoritma C4.5 dan *Naive Bayes Classification*”.

Tugas akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana strata satu pada Jurusan Ilmu Komputer / Informatika Fakultas Sains Dan Matematika Universitas Diponegoro Semarang.

Sebagai pelaksanaan penyusunan laporan tugas akhir ini, penulis banyak mendapat bimbingan, arahan dan bantuan dari berbagai pihak. Oleh karena itu dengan segala kerendahan hati, penulis ingin mengucapkan terima kasih dengan tulus kepada :

1. Prof. Dr. Widowati, M.Si. selaku Dekan FSM UNDIP
2. Ragil Saputra, M.Cs selaku Ketua Jurusan Ilmu Komputer / Informatika
3. Helmie Arif, M.Cs. selaku Koordinator Tugas Akhir
4. Nurdin Bahtiar, S.Si., M.T. selaku dosen pembimbing
5. Herwanto dan Ema Herawati selaku orang tua penulis yang selalu memberikan dukungan dan doa
6. Semua pihak yang membantu dalam kelancaran penulisan tugas akhir ini yang tidak dapat penulis sebutkan satu persatu

Penulis menyadari bahwa dalam laporan ini masih banyak terdapat kekurangan baik dari penyampaian materi maupun isi dari materi itu sendiri. Hal ini dikarenakan keterbatasan kemampuan dan pengetahuan dari penulis. Oleh karena itu, kritik dan saran yang bersifat membangun sangat penulis harapkan.

Semoga laporan tugas akhir ini dapat bermanfaat bagi penulis dan juga pembaca pada umumnya.

Semarang, 14 Maret 2016

Penulis

DAFTAR ISI

HALAMAN PERNYATAAN KEASLIAN SKRIPSI	ii
HALAMAN PENGESAHAN	iii
HALAMAN PENGESAHAN	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan dan Manfaat	3
1.4 Batasan Masalah.....	3
1.5 Sistematika Penulisan	4
BAB II DASAR TEORI.....	5
2.1 Penyakit Diabetes	5
2.2 Data Mining.....	6
2.3 <i>Decision Tree</i>	8
2.4 <i>If-Then Rule</i>	11
2.5 Algoritma C4.5	11
2.6 <i>Naïve Bayes Classification</i>	14
2.6.1 Teorema Bayes.....	14
2.6.2 <i>Naive Bayes Classification</i>	15
2.7 Metode <i>Waterfall</i>	15
2.8 Diagram.....	17
2.8.1. <i>Entity Relationship Diagram (ERD)</i>	17
2.8.2 <i>Data Flow Diagram (DFD)</i>	18
2.9 <i>Blackbox Testing</i>	20
2.10 PHP (Hypertext Processor).....	20

2.11 DBMS MySQL.....	21
2.12 Akurasi	21
BAB III ANALISIS DAN PERANCANGAN	23
3.1 Deskripsi Umum	23
3.2 Pengumpulan Data	23
3.3 Identifikasi Kebutuhan	25
3.3.1 Identifikasi Kebutuhan Fungsional	25
3.3.2 Identifikasi Kebutuhan Non Fungsional	26
3.4 Pemodelan Sub Sistem	26
3.4.1 Sub Sistem Algoritma C4.5	26
3.4.2 Representasi Data Pelatihan	26
3.4.3 Analisis Proses Algoritma C4.5	27
3.4.4 Analisis Proses <i>Naive Bayes Classification</i>	30
3.5 Analisis Model Permasalahan	33
3.5.1 Pemodelan Data	33
3.6 Perancangan Sistem	34
3.7 Perancangan Tabel	41
3.7.1 Tabel User	41
3.7.2 Tabel Data Diabetes	41
3.7.3 Tabel Atribut.....	42
3.7.4 Tabel Frekuensi Kemunculan.....	42
3.7.5 Iterasi C4.5	43
3.7.5 Tabel Data Keputusan	43
3.7.6 Tabel Data Keputusan Kinerja.....	44
3.7.7 Tabel Data Keputusan Kinerja <i>Naive Bayes</i>	45
3.7.8 Tabel Data Keputusan <i>Naive Bayes</i>	46
3.7.9 Tabel Mining C4.5	48
3.7.10 Tabel Pohon Keputusan C45	49
3.7.11 Tabel Rule C45	49
3.7.12 Tabel Rule Penentu Keputusan.....	50
3.7.13 Tabel Frekuensi Kemunculan	50
3.8 Perancangan Antarmuka	51
3.8.1 Halaman <i>Login</i>	51

3.8.2	Halaman <i>Home</i>	51
3.8.3	Halaman Data Diabetes.....	52
3.8.4	Halaman Proses C45	53
3.8.5	Halaman Perhitungan C45	53
3.8.6	Halaman Pohon Keputusan C45.....	54
3.8.7	Halaman <i>Naive Bayes</i>	54
3.8.8	Halaman Proses Ulang	55
3.8.9	Halaman Tabel Perbandingan	55
3.8.10	Halaman Tabel Akurasi	56
3.8.11	Halaman Penentu Keputusan.....	56
3.9	Perancangan Algoritma	57
BAB IV IMPLEMENTASI DAN PENGUJIAN		60
4.1	Implementasi Tabel Data.....	60
4.2	Implementasi Antarmuka Sistem.....	67
4.2.1	Halaman Login	67
4.2.2	Halaman Data Diabetes.....	68
4.2.3	Halaman Partisi Data	69
4.2.4	Halaman Proses C45	69
4.2.5	Halaman Perhitungan C45	70
4.2.6	Halaman Pohon Keputusan C45.....	70
4.2.7	Halaman <i>Naive Bayes</i>	71
4.2.8	Halaman Proses Ulang	73
4.2.9	Halaman Tabel Perbandingan	73
4.2.10	Halaman Tabel Akurasi	74
4.2.11	Halaman Penentu Keputusan.....	74
4.3	Implementasi Algoritma	75
4.4	Pengujian Sistem.....	75
4.4.1	Lingkungan Pengujian	75
4.4.2	Rencana Pengujian.....	75
4.4.3	Pengujian Fungsi Sistem.....	76
4.4.4	Pengujian Perhitungan Sistem	79
4.4.4.3	Perhitungan Sistem.....	82
BAB V PENUTUP		87

4.2.1 Kesimpulan.....	87
4.2.2 Saran	87
DAFTAR PUSTAKA	88
LAMPIRAN – LAMPIRAN	91

DAFTAR TABEL

Tabel 2.1 Data Catatan Risiko Kredit	8
Tabel 2.2 Penjelasan Formula Teorema <i>Bayes</i>	14
Tabel 2.3 Komponen ERD	18
Tabel 2.4 Komponen DFD	19
Tabel 3.1 Nilai Referensi Data Diabetes	23
Tabel 3.2 Kategori Data Diabetes.....	24
Tabel 3.3 Kebutuhan Fungsional	25
Tabel 3.4 Kebutuhan Non Fungsional	26
Tabel 3.5 Atribut Data Laboratorium	26
Tabel 3.6 Atribut Data Pelatihan dan Pengujian	27
Tabel 3.7 Contoh Data Kategori Algoritma C4.5.....	27
Tabel 3.8 Hasil Perhitungan <i>Entropy</i> dan Gain untuk Node Akar	28
Tabel 3.9 Hasil Perhitungan Algoritma C4.5	29
Tabel 3.10 Contoh Data <i>Naive Bayes Classification</i>	31
Tabel 3.11 Pengujian.....	31
Tabel 3.12 Hasil <i>Naive Bayes Classification</i>	33
Tabel 3.13 User.....	41
Tabel 3.14 Data Diabetes	41
Tabel 3.15 Atribut.....	42
Tabel 3.16 Frekuensi Kemunculan	42
Tabel 3.17 Iterasi C4.5	43
Tabel 3.18 Data Keputusan	43
Tabel 3.19 Data Keputusan Kinerja.....	44
Tabel 3.20 Data Keputusan Kinerja <i>Naive Bayes</i>	45
Tabel 3.21 Data Keputusan <i>Naive Bayes</i>	47
Tabel 3.22 Mining C4.5	48
Tabel 3.23 Pohon Keputusan C45	49
Tabel 3.24 Rule C45	49
Tabel 3.25 Rule Penentu Keputusan	50
Tabel 3.26 Frekuensi Kemunculan	50
Tabel 4.1 Kode SQL Membuat Tabel User.....	60

Tabel 4.2 Kode SQL Membuat Tabel Diabetes	60
Tabel 4.3 Kode SQL Membuat Tabel Atribut	61
Tabel 4.4 Kode SQL Membuat Tabel Frekuensi Kemunculan	61
Tabel 4.5 Kode SQL Membuat Tabel Frekuensi Kemunculan Atribut	62
Tabel 4.6 Kode SQL Membuat Tabel Iterasi C45	62
Tabel 4.7 Kode SQL Membuat Tabel Rule Penentu Keputusan	62
Tabel 4.8 Kode SQL Membuat Tabel Rule C4.5	63
Tabel 4.9 Kode SQL Membuat Tabel Mining C4.5	63
Tabel 4.10 Kode SQL Membuat Tabel Data Keputusan Kinerja	64
Tabel 4.11 Kode SQL Membuat Tabel Data Keputusan Kinerja <i>Naive Bayes</i>	65
Tabel 4.12 Kode SQL Membuat Tabel Data Keputusan <i>Naive Bayes</i>	66
Tabel 4. 13 Kode SQL Membuat Tabel Pohon Keputusan C45	67
Tabel 4. 14 Skenario Pengujian Fungsi Sistem	76
Tabel 4.15 Hasil Pengujian Fungsi Sistem.....	77
Tabel 4.16 Data Diabetes	80
Tabel 4.17 Hasil Perhitungan	83
Tabel 4.18 Hasil Prediksi Sistem dan Data Laboratorium	85

DAFTAR GAMBAR

Gambar 2.1 Tahapan-tahapan data mining	7
Gambar 2.2 Contoh <i>Decision Tree</i>	9
Gambar 2.3 Syarat Pengujian Fitur Biner	12
Gambar 2.4 Syarat Pengujian Fitur Bertipe Nominal.....	13
Gambar 2.5 Syarat Pengujian Fitur Bertipe Ordinal.....	13
Gambar 2.6 Syarat Pengujian Fitur Bertipe Numerik.....	13
Gambar 2.7 Model <i>Waterfall</i>	16
Gambar 3.1 Contoh <i>decision tree</i> data diabetes	30
Gambar 3.2 Aplikasi Data Mining untuk Prediksi Penyakit Diabetes.....	34
Gambar 3.3 DFD Level 1 Aplikasi Data Mining Prediksi Penyakit Diabetes	35
Gambar 3.4 <i>Flowchart</i> Algoritma C4.5	35
Gambar 3.5 <i>Flowchart</i> detail perhitungan <i>entropy</i>	36
Gambar 3.6 <i>Flowchart</i> detail perhitungan <i>gain</i>	37
Gambar 3.7 <i>Flowchart</i> detail perhitungan <i>Split Info</i>	38
Gambar 3.8 <i>Flowchart Naive Bayes Classification</i>	39
Gambar 3.9 <i>Flowchart</i> detail Proses <i>Naive Bayes Classification</i>	40
Gambar 3.10 Perancangan Halaman <i>Login</i>	51
Gambar 3.11 Perancangan Halaman <i>Home</i>	52
Gambar 3. 12 Perancangan Halaman Data Diabetes	52
Gambar 3. 13 Perancangan Halaman Data Diabetes	53
Gambar 3.14 Perancangan Halman Partisi Data	53
Gambar 3.15 Perancangan Halaman Perhitungan C45	54
Gambar 3.16 Perancangan Halaman Pohon Keputusan C45	54
Gambar 3.17 Perancangan Halaman Naive Bayes	55
Gambar 3.18 Perancangan Halaman Proses Ulang	55
Gambar 3.19 Perancangan Halaman Tabel Perbandingan	56
Gambar 3.20 Perancangan Halaman Tabel Akurasi	56
Gambar 3.21 Perancangan Halaman Penentu Keputusan	57
Gambar 4.1 Implementasi Halaman <i>Login</i>	67
Gambar 4.2 Implementasi Halaman Data Diabetes.....	68
Gambar 4.3 Implementasi Halaman Data Diabetes.....	68

Gambar 4.4 Implementasi Halaman Partisi Data	69
Gambar 4.5 Implementasi Halaman Proses C45	69
Gambar 4.6 Implementasi Halaman Perhitungan C45.....	70
Gambar 4.7 Implementasi Halaman Pohon Keputusan C45	71
Gambar 4.8 Implementasi Halaman Pohon Keputusan C45	71
Gambar 4.9 Implementasi Halaman <i>Naive Bayes</i>	72
Gambar 4.10 Implementasi Halaman <i>Naive Bayes</i>	72
Gambar 4.11 Implementasi Halaman Proses Ulang	73
Gambar 4.12 Implementasi Halaman Tabel Perbandingan.....	73
Gambar 4.13 Implementasi Halaman Tabel Akurasi.....	74
Gambar 4.14 Implementasi Halaman Penentu Keputusan.....	74
Gambar 4.15 Hasil Proses Perhitungan Sistem Algoritma C4.5	82
Gambar 4.16 Hasil Proses Perhitungan <i>Naive Bayes Classification</i>	82
Gambar 4. 17 Data Laboratorium	84
Gambar 4. 18 Data Perhitungan Sistem	85

DAFTAR LAMPIRAN

Lampiran 1. Hasil Prediksi dengan Sistem dan Data Laboratorium	68
Lampiran 2. Koding <i>Naive Bayes Classification</i>	70
Lampiran 3. Koding Algoritma C4.5.....	70

BAB I

PENDAHULUAN

Bab ini membahas latar belakang, rumusan masalah, tujuan dan manfaat, batasan masalah dan sistematika penulisan dalam pembuatan tugas akhir mengenai pembangunan Aplikasi Data Mining untuk Diagnosis Penyakit Diabetes Menggunakan Algoritma C4.5 dan *Naive Bayes Classification*.

1.1 Latar Belakang Masalah

Data mining merupakan serangkaian proses untuk mendapatkan informasi berguna dari gudang basis data yang besar. Data mining juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Dalam data mining terdapat banyak teknik dalam pengerjaannya, diantaranya yaitu Algoritma C4.5, *Naive Bayes Classification*, *decision tree* dan masih banyak lainnya.

Organisasi Kesehatan Dunia (*World Health Organization/WHO*) memperkirakan bahwa 177 juta penduduk dunia mengidap penyakit *diabetes melitus* atau biasa disingkat diabetes. Jumlah ini akan meningkat hingga melebihi 300 juta pada tahun 2025. Dr Paul Zimmet, direktur dari *International Diabetes Institute (IDI)* di Victoria, Australia, meramalkan bahwa diabetes akan menjadi epidemi yang paling dahsyat dalam sejarah manusia. Tetapi data epidemiologi di negara berkembang masih belum banyak. Oleh karena itu angka *prevalensi* yang dapat ditelusuri mayoritas berasal dari negara maju.

Dengan bertambahnya angka harapan hidup di negara Indonesia, perhatian masalah kesehatan beralih dari penyakit infeksi ke penyakit degeneratif. Selain penyakit jantung koroner dan hipertensi, diabetes merupakan salah satu penyakit degeneratif yang saat ini makin bertambah jumlahnya di Indonesia.

Pada dasarnya diabetes dapat dikelompokkan menjadi dua, yaitu diabetes tipe 1 yang terjadi sejak kecil karena cacat sejak lahir dan diabetes tipe 2 yang berkembang setelah dewasa sebagai akibat gaya hidup yang salah. Diabetes tipe 2 adalah jenis diabetes yang paling banyak ditemukan. Tanpa penanganan yang efektif, jumlah

penderita diabetes akan meningkat, disebabkan oleh berbagai hal, antara lain disebabkan pola hidup yang salah. (Herwanto, 2006)

Berdasarkan hal tersebut, untuk meningkatkan upaya penurunan angka kesakitan dan prevalensi timbulnya komplikasi pada penyakit diabetes maka perlu dilakukan penelitian-penelitian yang mengarah pada pembuatan sistem yang dapat mendeteksi timbulnya penyakit diabetes sehingga dapat dilakukan upaya pencegahan serta upaya rehabilitatif bagi penderita diabetes dengan pendekatan yang menyeluruh, sehingga dampak terjadinya berbagai penyulit menahun seperti penyakit jantung koroner, penyakit pada mata, ginjal, dan syaraf dapat dikurangi.

Untuk mendiagnosis penyakit, manusia mulai mengembangkan suatu sistem untuk membantu mencari dan membuat keputusan untuk masalah-masalah dan dengan memperhitungkan berbagai macam faktor yang ada di dalam lingkup masalah tersebut dalam hal ini adalah keputusan untuk mendiagnosis penyakit diabetes. Dengan *decision tree*, manusia dapat dengan mudah mengidentifikasi dan melihat hubungan antara faktor-faktor yang mempengaruhi suatu masalah dan dapat mencari penyelesaian terbaik dengan memperhitungkan faktor-faktor tersebut. *Decision tree* ini juga dapat menganalisa nilai resiko dan nilai suatu informasi yang terdapat dalam suatu alternatif pemecahan masalah. Peranan *decision tree* sebagai alat bantu dalam mengambil keputusan (*decision support tool*) telah dikembangkan oleh manusia sejak perkembangan teori pohon yang dilandaskan pada teori graf. Kegunaan *decision tree* yang sangat banyak ini membuatnya telah dimanfaatkan oleh manusia dalam berbagai macam sistem pengambilan keputusan.

Algoritma C4.5 adalah algoritma klasifikasi data dengan teknik *decision tree* yang terkenal dan disukai karena memiliki kelebihan-kelebihan. Kelebihan ini misalnya dapat mengolah data numerik (kontinyu) dan diskret, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat diantara algoritma-algoritma yang lain. Keakuratan prediksi yaitu kemampuan model untuk dapat memprediksi label kelas terhadap data baru atau yang belum diketahui sebelumnya dengan baik. Dalam hal kecepatan atau efisiensi waktu komputasi yang diperlukan untuk membuat dan menggunakan model. (Kamagi, 2014)

Naïve Bayes Classification adalah metode pengklasifikasian paling sederhana dengan menggunakan konsep peluang, yang mana diasumsikan bahwa setiap atribut

contoh bersifat saling lepas satu sama lain berdasarkan atribut kelas. *Naïve Bayes Classification* dapat digunakan untuk klasifikasi dokumen, deteksi spam, dan masalah klasifikasi lainnya. (Zanzabila, 2015). Pada penelitian ini klasifikasi yang dimaksud adalah mengenai penyakit diabetes dan gejala-gejala yang diambil dari data diabetes tersebut.

Berdasarkan permasalahan diatas, dapat dikembangkan sebuah teknik data mining dengan memprediksi data pasien teridentifikasi penyakit diabetes dengan menggunakan metode Algoritma C4.5 dan metode *Naïve Bayes Classification* dengan harapan setelah diolah dengan teknik data mining tersebut dapat dihasilkan informasi dalam prediksi data pasien teridentifikasi penyakit diabetes.

1.2 Rumusan Masalah

Sesuai dengan latar belakang di atas, rumusan masalah yang dibahas dalam penelitian ini adalah bagaimana membangun Aplikasi Data Mining untuk Diagnosis Penyakit Diabetes Menggunakan Algoritma C4.5 dan *Naïve Bayes Classification*.

1.3 Tujuan dan Manfaat

Tujuan dan manfaat dari penelitian ini meliputi:

1. Membuat sistem antarmuka yang dapat memprediksi penyakit diabetes dengan menggunakan Algoritma C4.5 dan *Naïve Bayes Classification*.
2. Hasil dari sistem ini nantinya dapat digunakan oleh dokter maupun edukator diabetes dalam mengantisipasi peningkatan jumlah pasien diabetes.
3. Menampilkan hasil perhitungan metode Algoritma C4.5 dan *Naïve Bayes Classification* yang memiliki pendekatan terbaik untuk prediksi penyakit diabetes.

1.4 Batasan Masalah

Untuk menghindari kesalahpahaman dan meluasnya pokok bahasan, maka penulis memberikan batasan masalah sebagai berikut:

1. Penelitian ini difokuskan untuk memprediksi penyakit diabetes saja tidak untuk mencari obat diabetes dan sebagainya
2. Sumber data utama berasal dari Sistem Informasi Manajemen Rumah Sakit Pusat Pertamina
3. Pembuatan program menggunakan program PHP dan MySQL.

1.5 Sistematika Penulisan

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

BAB I PENDAHULUAN

Bab ini berisi latar belakang masalah, rumusan masalah, tujuan dan manfaat, batasan masalah, dan sistematika penulisan dalam pembuatan tugas akhir.

BAB II DASAR TEORI

Bab ini menyajikan dasar teori yang berhubungan dengan topik tugas akhir. Dasar teori digunakan dalam penyusunan tugas akhir hingga selesai terciptanya perangkat lunak tersebut dan dapat diimplementasikan.

BAB III ANALISIS DAN PERANCANGAN

Bab ini membahas tahap analisis kebutuhan dan perancangan perangkat lunak serta hasil yang didapat pada tahap ini.

BAB IV IMPLEMENTASI DAN PENGUJIAN

Bab ini membahas proses pengembangan perangkat lunak dan hasil yang didapat pada tahap implementasi. Bab ini juga berisi rincian pengujian perangkat lunak yang dibangun dengan metode *blackbox*.

BAB V PENUTUP

Bab ini berisi kesimpulan yang diambil berkaitan dengan perangkat lunak yang dikembangkan dan saran-saran untuk pengembangan perangkat lunak lebih lanjut.