

**KOMPARASI ALGORITMA KLASIFIKASI DATA MINING
MODEL C4.5 DAN NAIVE BAYES UNTUK PREDIKSI
PENYAKIT DIABETES**



TESIS

Diajukan sebagai salah satu syarat untuk memperoleh gelar

Magister Ilmu Komputer (M.Kom)

FATMAWATI

14000945

**PROGRAM PASCASARJANA MAGISTER ILMU KOMPUTER
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER**

NUSA MANDIRI

JAKARTA

2015

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini :

Nama : Fatmawati
NIM : 14000945
Program Studi : Magister Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : *Managemen Information System*

Dengan ini menyatakan bahwa tesis yang telah saya buat dengan judul: “Komparasi Algoritma Klasifikasi Data Mining Model C4.5 dan *Naive bayes* Untuk Prediksi Penyakit Diabetes” adalah hasil karya sendiri, dan semua sumber baik yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar dan tesis belum pernah diterbitkan atau dipublikasikan dimanapun dan dalam bentuk apapun.

Demikianlah surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila dikemudian hari ternyata saya memberikan keterangan palsu dan atau ada pihak lain yang mengklaim bahwa tesis yang telah saya buat adalah hasil karya milik seseorang atau badan tertentu, saya bersedia diproses baik secara pidana maupun perdata dan kelulusan saya dari Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri dicabut/dibatalkan.

Jakarta, 05 Desember 2015

Yang menyatakan,



Fatmawati

HALAMAN PENGESAHAN

Tesis ini diajukan oleh :
Nama : Fatmawati
NIM : 14000945
Program Studi : Magsiter Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : *Managemen Information System*
Judul Tesis : "Komparasi Algoritma Klasifikasi Data Mining Model C4.5 dan *Naive bayes*
Untuk Prediksi Penyakit Diabetes"

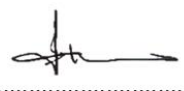
Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Magister Ilmu Komputer (M.Kom) pada Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri).

Jakarta, 05 Desember 2015
Pascasarjana Magister Ilmu Komputer
STMIK Nusa Mandiri
Direktur

Prof. Dr. Ir. R. Eko Indrajit, MSC, MBA

DEWAN PENGUJI

Penguji I : Dr. Sularso Budilaksono, M.Kom 

Penguji II : Dr. Sfenrianto, M.Kom 

Pembimbing : Dr. Windu Gata, M.Kom 

	LEMBAR KONSULTASI BIMBINGAN TESIS
	PROGRAM PASCASARJANA MAGISTER ILMU KOMPUTER

- N I M : 14000945
- Nama Mahasiswa : Fatmawati
- Konsentrasi : *Managemen Information System*
- Judul Tesis : *Komparasi Algoritma Klasifikasi Data Mining Model C4.5 dan Naive Bayes untuk Prediksi Penyakit Diabetes*

No	Tanggal Bimbingan	Pokok Bahasan	Paraf dosen Pembimbing
1.	07-11-2015	Pengajuan Judul	
2.	10-11-2015	Pengajuan Pendahuluan	
3.	14-11-2015	Revisi Pendahuluan dan Pengajuan Landasan Teori	
4.	17-11-2014	Revisi Landasan Teori dan Pengajuan Metode Penelitian	
5.	21-11-2015	Revisi Metode Penelitian dan Pengajuan Hasil serta Kesimpulan	
6.	25-11-2015	Revisi Kesimpulan	

Catatan untuk dosen pembimbing

Bimbingan Tesis

- Dimulai pada tanggal : 07 November 2015
- Diakhiri pada tanggal : 25 November 2015
- Jumlah pertemuan bimbingan : 6x Bimbingan

Disetujui oleh,
Dosen Pembimbing



[Dr. Windu Gata, M.Kom]

KATA PENGANTAR

Puji syukur alhamdulillah, penulis panjatkan kehadiran Allah, SWT, yang telah melimpahkan rahmat dan karunia-Nya, sehingga pada akhirnya penulis dapat menyelesaikan tesis ini tepat pada waktunya. Dimana tesis ini penulis sajikan dalam bentuk buku yang sederhana. Adapun judul tesis, yang penulis ambil sebagai berikut “Komparasi Algoritma Klasifikasi Data Mining Model C4.5 dan *Naive Bayes* Untuk Prediksi Penyakit Diabetes”.

Tujuan penulisan tesis ini dibuat sebagai salah satu untuk mendapatkan gelar Magister Ilmu Komputer (M.Kom) pada Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (PPs MIK STMIK Nusa Mandiri).

Tesis ini diambil berdasarkan hasil penelitian yang dilakukan dan penulis juga mencari dan menganalisa berbagai macam sumber referensi, baik dalam bentuk jurnal ilmiah, buku-buku literatur, *internet*, dan lain-lain yang terkait dengan pembahasan pada tesis ini.

Penulis menyadari bahwa tanpa bimbingan dan dukungan dari semua pihak dalam pembuatan tesis ini, maka penulis tidak dapat menyelesaikan tesis ini tepat pada waktunya. Untuk itu ijinkanlah penulis kesempatan ini untuk mengucapkan ucapan terima kasih yang sebesar-besarnya kepada :

1. Bapak Dr. Windu Gata, M.Kom selaku pembimbing tesis yang telah menyediakan waktu, pikiran dan tenaga dalam membimbing penulis dalam menyelesaikan tesis ini.
2. Orang tua dan suami tercinta yang telah memberikan dukungan material dan moral kepada penulis.
4. Seluruh staff pengajar (dosen) Program Pascasarjana STMIK Nusa Mandiri yang telah memberikan pelajaran yang berarti bagi penulis selama menempuh studi.
5. Seluruh staff dan karyawan Program Pascasarjana STMIK Nusa Mandiri yang telah melayani penulis dengan baik selama kuliah.

Serta semua pihak yang terlalu banyak untuk penulis sebutkan satu persatu sehingga terwujudnya penulisan tesis ini. Penulis menyadari bahwa penulisan tesis ini masih jauh sekali dari sempurna, untuk itu penulis mohon kritik dan saran yang

bersifat membangun demi kesempurnaan penulisan karya ilmiah yang penulis hasilkan untuk yang akan datang.

Akhir kata semoga tesis ini dapat bermanfaat bagi penulis khususnya dan bagi para pembaca yang berminat pada umumnya.

Jakarta, 05 Desember 2015

A handwritten signature in black ink, appearing to read 'Fatmawati', with a stylized flourish at the end.

Fatmawati

Penulis

**SURAT PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH
UNTUK KEPENTINGAN AKADEMIS**

Yang bertanda tangan di bawah ini, saya :

Nama : Fatmawati
NIM : 14000945
Program Studi : Magsiter Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : *Managemen Information System*
Jenis Karya : Tesis

Demi pengembangan ilmu pengetahuan, dengan ini menyetujui untuk memberikan ijin kepada pihak Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri) **Hak Bebas Royalti Non-Eksklusif (*Non-exclusive Royalti-Free Right*)** atas karya ilmiah kami yang berjudul : “Komparasi Algoritma Klasifikasi Data Mining Model C4.5 dan *Naive bayes* Untuk Prediksi Penyakit Diabetes” beserta perangkat yang diperlukan (apabila ada).

Dengan **Hak Bebas Royalti Non-Eksklusif** ini pihak STMIK Nusa Mandiri berhak menyimpan, mengalih-media atau *bentuk*-kan, mengelolanya dalam pangkalan data (*database*), mendistribusikannya dan menampilkan atau mempublikasikannya di *internet* atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari kami selama tetap mencantumkan nama kami sebagai penulis/pencipta karya ilmiah tersebut.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak STMIK Nusa Mandiri, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 05 Desember 2015

Yang menyatakan,



Fatmawati

ABSTRAK

Nama : Fatmawati
N I M : 14000945
Program Studi : Magister Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : Manajemen Informasi Sistem
Judul Tesis : Komparasi Algoritma Klasifikasi Data Mining Model C4.5
dan *Naive Bayes* untuk Prediksi Penyakit Diabetes

Penyakit diabetes merupakan salah satu penyakit yang mematikan, faktor resiko tinggi dalam keluarga yang menyebabkan diabetes dikarenakan orang gemuk yang tidak melakukan latihan fisik, dan orang-orang yang tidak memiliki gaya hidup sehat dan makanan yang berlebihan dari apa yang dibutuhkan oleh tubuh. Berdasarkan data *history* penderita diabetes dapat dibuat rekomendasi prediksi penyakit diabetes yang dapat membantu tenaga kesehatan. Klasifikasi merupakan salah satu teknik dari *data mining* yang dapat digunakan untuk membantu prediksi. Klasifikasi dapat dilakukan dengan *Decision Tree* yaitu dengan algoritma C4.5 dan *Naive Bayes*. Penelitian ini bertujuan membuat klasifikasi dan menerapkan klasifikasi *data mining*. Hasil klasifikasi data di evaluasi dengan menggunakan *Confusion Matrix* dan kurva *ROC* untuk mengetahui tingkat hasil akurasi menggunakan algoritma *Decision Tree* yaitu sebesar 73.30% dan nilai AUC dari kurva *ROC* adalah 0.733 sedangkan algoritma *Naive Bayes* sebesar 75.13% nilai AUC dari kurva *ROC* 0.810 sehingga dapat dikatakan bahwa algoritma *Naive Bayes* memiliki hasil prediksi yang baik dalam memprediksi penyakit diabetes seorang pasien.

Kata kunci: prediksi penyakit diabetes, algoritma C4.5, model *Decision Tree*, *Naive Bayes*

ABSTRACT

Nama : Fatmawati
N I M : 14000945
Program Studi : Magister Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : *Managemen Information System*
Judul Tesis : *Komparasi Algoritma Klasifikasi Data Mining Model C4.5 dan Naive Bayes untuk Prediksi Penyakit Diabetes*

Diabetes is one of the deadly disease, high risk factors in families that cause diabetes because fat people who do not do physical exercise, and those who do not have a healthy lifestyle and diet excess of what is needed by the body. Based on the history data diabetics can be made on the prediction of diabetes that can help health professionals. Classification is one of data mining techniques that can be used to help predict. Classification can be done with that Decision Tree algorithm C4.5 and Naive Bayes. This study aims to classify and apply data mining classification. Results of data classification in the evaluation using the Confusion Matrix and ROC curve to determine the level of accuracy results using algorithms Decision Tree that is equal to 73.30% and the AUC of the ROC curve was 0.733 while the algorithm Naive Bayes amounted to 75.13% AUC values of the ROC curve of 0.810, so it can be said that the algorithm Naive Bayes have the result of a good predictor in predicting diabetes patient.

Keywords: *prediction of diabetes, C4.5 algorithm, the model Decision Tree, Naive Bayes*

DAFTAR ISI

	Halaman
Cover	i
SURAT PERNYATAAN ORISINALITAS	ii
HALAMAN PENGESAHAN.....	iii
LEMBAR KONSULTASI BIMBINGAN	iv
KATA PENGANTAR	v
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS	vii
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
BAB I PENDAHULUAN	1
1.1. Latar Belakang Penulisan	1
1.2. Identifikasi dan Rumusan Masalah	2
1.2.1. Identifikasi Masalah	2
1.2.2. Rumusan Masalah.....	2
1.3. Tujuan Penelitian.....	2
1.4. Ruang Lingkup Penelitian	3
1.5. Sistematika Penulisan.....	3
BAB II LANDASAN TEORI	5
2.1. Diabetes	5
2.2. Data <i>Mining</i>	7
2.3. Klasifikasi.....	8
2.4. Algoritma C4.5.....	9
2.5. Decision <i>Tree</i>	10
2.6. Naive <i>Bayes</i>	11
2.7. Rapid Miner.....	12
2.8. Evaluasi dan Validasi	12
2.9. Confusion Matrix.....	13

2.10.	Kurva ROC	14
BAB III METODOLOGI PENELITIAN.....		15
3.1.	Desain Penelitian.....	15
3.2.	Pengumpulan Data.....	16
3.3.	Pengolahan Data Awal	16
3.4.	Metode yang di Usulkan.....	17
3.5.	Eksperimen dan Pengujian Metode	18
3.6.	Evaluasi dan Validasi Hasil.....	19
BAB IV HASIL PENELITIAN DAN PEMBAHASAN		20
4.1.	Pengumpulan Data.....	20
4.2.	Pengolahan Awal Data	22
4.3.	Pengukuran Penelitian.....	25
4.3.1.	Confucion Matrix:	25
4.3.2.	Kurva ROC (Receiver Operating Characteristic)	27
4.4.	Analisa Hasil Komparasi.....	28
BAB V PENUTUP.....		30
1.1.	KESIMPULAN	30
1.2.	SARAN.....	30
DAFTAR PUSTAKA		31
LAMPIRAN.....		33

DAFTAR TABEL

Tabel 1. Matrik Klasifikasi untuk Model 2 Class	13
Tabel 2. Tabel Atribut yang digunakan (UCI Repository)	17
Tabel 3. Spesifikasi hardware dan software.....	19
Tabel 4. Data Pasien Diabetes.....	20
Tabel 5. Atribut, Tipe, Ukuran dan Nilai Atribut	22
Tabel 6. Hasil Komparasi Algoritma C4.5 dan Naive Bayes	28

DAFTAR GAMBAR

Gambar 1. Metode yang diusulkan	18
Gambar 2. Desain Model Algoritma C4.5	22
Gambar 3. Hasil Klasifikasi Menggunakan Algoritma C4.5	23
Gambar 4. Hasil Akurasi Prediksi Penyakit Diabetes Menggunakan Algoritma C4.5	25
Gambar 5. Hasil Akurasi Prediksi Penyakit Diabetes Menggunakan Algoritma Naive Bayes	26
Gambar 6. Grafik ROC dari Model Algoritma C4.5	27
Gambar 7. Grafik ROC dari Model Naive Bayes	28
Gambar 8. Kurva ROC dengan Algoritma C4.5 dan Naive Bayes	29

BAB I

PENDAHULUAN

1.1. Latar Belakang Penulisan

Diabetes merupakan penyakit gangguan metabolik menahun akibat pankreas tidak memproduksi cukup insulin atau tubuh tidak dapat menggunakan insulin yang diproduksi secara efektif. Insulin adalah hormon yang mengatur keseimbangan kadar gula darah. Akibatnya terjadi peningkatan konsentrasi glukosa didalam darah (hiperglikemia).

Penyakit diabetes disebabkan oleh peningkatan kadar glukosa dalam darah, apabila kadar glukosa darah meningkat dalam jangka waktu yang lama maka akan menyebabkan komplikasi seperti gagal ginjal, kebutaan dan serangan jantung (Jayalskshmi & Santhakumaran, 2010).

Estimasi terakhir IDF (*International Diabetes Federation*), terdapat 382 juta orang yang hidup dengan diabetes di dunia pada tahun 2013. Dari berbagai penelitian epidemiologis di indonesia yang dilakukan oleh pusat-pusat diabetes, seekitar tahun 1980-an prevalensi diabetes melitus pada penduduk usia 15 tahun ke atas sebesar 1,5-2,3% dengan prevalensi di daerah rural/perdesaan lebih rendah dibandingkan perkotaan.

Teknik analisa konvensional secara manual yang selama ini digunakan tidak lagi efektif digunakan untuk mendiagnosa. Seiring dengan perkembangan sistem berbasis pengetahuan medis tuntutan akan adanya penggunaan sistem pengetahuan berbasis komputer sebagai teknik analisa dalam mendiagnosa penyakit menjadi semakin penting. Oleh karenanya, saat inilah waktu yang tepat untuk mengembangkan sistem pengetahuan berbasis komputer yang modern, efektif dan efisien dalam mendiagnosa penyakit (Neshat, Mehdi & Yaghobi, 2009).

Oleh karena itu, penelitian ini dilakukan untuk membantu menyelesaikan permasalahan tersebut dengan *data mining* yang berfungsi untuk memprediksi penyakit diabetes, diperlukan suatu metode atau teknik yang dapat mengolah data-data yang sudah ada. Salah satu metodenya menggunakan teknik *data mining*.

Penggunaan *data mining* dengan algoritma C4.5 dan *Naive Bayes* sebagai pilihan untuk diagnosa penyakit diabetes dapat menjadi alternatif pilihan yang tepat, tetapi sampai saat ini belum diketahui algoritma yang paling akurat dalam memprediksi penyakit diabetes.

Pada penelitian ini akan dilakukan komparasi *data mining* algoritma C4.5 dan *Naive Bayes* untuk mengetahui algoritma yang memiliki akurasi yang lebih tinggi dalam mendeteksi penyakit diabetes.

1.2. Identifikasi dan Rumusan Masalah

1.2.1. Identifikasi Masalah

Berdasarkan latar belakang permasalahan yang sudah diuraikan di atas, maka identifikasi masalah yang diangkat adalah seberapa besar akurasi dan perbandingan antara algoritma C4.5 dan *Naive Bayes* dengan menggunakan *confusion matrix* dan kurva *ROC* untuk memprediksi penyakit diabetes.

1.2.2. Rumusan Masalah

Berdasarkan latar belakang masalah, maka rumusan masalah dalam penelitian ini adalah: Penerapan algoritma apakah yang paling akurat dalam memprediksi penyakit diabetes dan bagaimana perbandingan tingkat akurasi diantara algoritma C4.5 dan *Naive Bayes*?

1.3. Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Menerapkan klasifikasi algoritma C4.5 dan *Naive Bayes*.
2. Membandingkan tingkat akurasi antara algoritma C4.5 dan *Naive Bayes*.

1.4. Ruang Lingkup Penelitian

Ruang lingkup dalam penelitian ini adalah mengenai seberapa besar tingkat akurasinya diantara model algoritma C4.5 dan *Naive Bayes* yang memiliki akurasi paling tinggi untuk memprediksi penyakit diabetes.

1.5. Sistematika Penulisan

Pembahasan dalam penyusunan tesis ini dikelompokkan dalam beberapa bab yang bertujuan agar sistematika pembahasan yang lebih teratur dan lebih jelas dalam memberikan uraian-uraian dari permasalahan yang terdapat dalam tesis ini :

BAB 1 PENDAHULUAN

Latar belakang penulisan, identifikasi masalah, rumusan masalah, tujuan penelitian, ruang lingkup penelitian dan sistematika penulisan.

BAB 2 LANDASAN TEORI

Membahas tentang teori yang melandasi penelitian dengan algoritma C4.5 dan *Naive Bayes* yang digunakan untuk melakukan perbandingan tingkat keakurasian sebagai pendukung keputusan.

BAB 3 METODE PENELITIAN

Jenis penelitian ini yaitu eksperimen. Metode yang digunakan untuk membandingkan tingkat ke akurasiannya yaitu algoritma C4.5 dan *Naive Bayes* sedangkan *tools* yang digunakan Rapid Miner. Pengujian evaluasi dan validasi untuk mengukur akurasi menggunakan *confusion matrix* dan *kurva ROC*.

BAB 4 HASIL DAN PEMBAHASAN

Membahas mengenai hasil penelitian, pengujian data dan implikasi penelitian yang berhubungan dengan masalah yang diteliti.

BAB 5 PENUTUP

Berisi tentang kesimpulan dari penelitian dan saran untuk penelitian selanjutnya

BAB II

LANDASAN TEORI

2.1. Diabetes

Diabetes adalah epidemi yang paling cepat berkembang di Barat dunia. Satu dari tiga anak-anak Amerika akan tumbuh dan terkena diabetes, 24% orang dewasa Amerika resisten insulin dan 45% orang dewasa di atas usia 60 resisten insulin (Mason, 2005). Diabetes adalah salah satu penyebab utama kematian di banyak negara dan penyebab utama kebutaan, gagal ginjal, dan *nontraumatic* amputasi (Robert, Zgonis, & Driver, 2006).

Faktor penyebab diabetes adalah (Nurrahmani, 2012):

1. Gen diabetes dalam keluarga

Gen merupakan sel pembawa sifat yang dapat diwariskan orang tua kepada keturunannya, dan diabetes merupakan penyakit yang bisa diwariskan.

2. Insulin dan gula darah

Insulin adalah karena ketidakmampuan beta sel-sel di pankreas untuk memproduksi insulin (Mason, 2005). Produksi ini disebabkan oleh tingginya kadar gula dalam darah, sehingga menyebabkan insulin diproduksi semakin tinggi.

3. Kegemukan (Obesitas)

Pada kegemukan atau obesitas sel-sel lemak yang menggemuk yang jumlahnya lebih banyak dari pada keadaan tidak gemuk, sehingga menyebabkan resistensi terhadap insulin dimana gula darah sulit masuk kedalam sel, sehingga gula darah tetap tinggi (hiperglikemi) sehingga terjadilah diabetes, khususnya terjadi pada diabetes tipe 2.

4. Asma

Penderita yang mengalami penyakit asma diharuskan untuk mengkonsumsi obat asma, sehingga memicu terjadinya diabetes, dikarenakan hormon yang digunakan pada obat asma tersebut adalah steroid yang berkerja berlawanan dengan insulin yang menaikkan gula darah.

5. KB

Pil kontrasepsi merupakan salah satu obat yang mengandung hormon steroid dengan anti insulin rendah.

Tes diagnostik diabetes tertentu dengan profil darah dapat menganalisis dasar dengan menguji glukosa, urat asam aris di bawah 5 mg/dl untuk pria dan di bawah 4 mg/dl bagi perempuan, jumlah sel darah putih harus dalam batas normal, kolesterol total aris di bawah 5 mg/dl untuk pria dan di bawah 4 mg/dl bagi perempuan, HDL (*High Density Lipoprotein* atau "lemak baik"), LDL (*Low Density Lipoprotein* atau "lemak jahat"), dan trigliserida (sejenis lemak dalam darah) harus di bawah 100x, CRP (protein yang ditemukan dalam darah) harus kurang dari 1 mg/l (Mason, 2005). Pengobatan yang sejati untuk disfungsi gula darah dalam bentuk apapun adalah dengan memilih makanan yang lebih baik, dan makan makanan alami (Mason, 2005). Ada tiga tipe utama diabetes: (Mason, 2005)

1. Tipe 1 (tergantung insulin) adalah karena ketidakmampuan beta sel-sel di pankreas untuk memproduksi insulin. Hal ini biasanya terjadi di masa kecil atau masa remaja dimana pasien-pasien harus menyuntikkan insulin, karena mereka tidak bisa memproduksinya secara alami.
2. Penderita diabetes tipe 2 (non-insulin dependent) memproduksi insulin, tetapi sel-sel hanya tidak bereaksi dengan baik, pankreas tidak hanya menghasilkan insulin tetapi biasanya overproduces, karena efektivitas sangat berkurang.
3. Jenis ketiga disebut "diabetes gestational", karena hanya mempengaruhi ibu hamil, untuk beberapa alasan wanita hamil lebih rentan terhadap diabetes dibandingkan orang lain.

Diabetes tipe 2 disebabkan oleh cara hidup dan juga dikarenakan faktor keturunan, biasanya mengenai orang dewasa (Nurrahmani, 2012). Faktor yang terkait dengan diabetes gestational adalah usia, kehamilan, berat badan, tekanan darah, riwayat keluarga diabetes serta etnis, diabetes gestational dapat meningkatkan masalah resiko bagi bayi yang belum lahir, jadi penting bahwa kadar glukosa dalam darah wanita hamil harus berada di bawah kendali (Bromuri, Schumacher, & Ruiz, 2011). Faktor resiko diabetes gestational dapat dilihat dari:

1. Berapa kali keguguran
2. Pernah melahirkan anak mati tanpa sebab
3. Pernah melahirkan bayi 4000 gram
4. Umur > 30 tahun
5. Riwayat diabetes dalam keluarga
6. Pernah terkena diabetes gestational pada kehamilan sebelumnya
7. Kegemukan
8. Berat badan ibu waktu melahirkan > 5 kg
9. Infeksi saluran kemih berulang-ulang

2.2. Data Mining

Data mining merupakan teknologi baru yang sangat berguna untuk membantu perusahaan-perusahaan menemukan informasi yang sangat penting dari gudang data mereka. Beberapa aplikasi data mining fokus pada prediksi, mereka meramalkan apa yang akan terjadi dalam situasi baru dari data yang menggambarkan apa yang terjadi di masa lalu (Witten, Frank, & Hall, 2011).

Data mining juga merupakan bagian dari *Knowledge Discovery in Database* (KDD) yang merupakan proses ekstraksi informasi yang berguna, tidak diketahui sebelumnya dan tersembunyi dari data (Bramer, 2007).

Secara garis besar *Knowledge Discovery in Database* (KDD) dapat dijelaskan sebagai berikut (Kusrini & Luthfi, 2009):

1. *Data Selection*

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD di mulai. Data hasil selksi yang akan digunakan untuk proses rpisah *data mining*, disimpan dalam suatu berkas terpisah dari basis data operasional.

2. *Pre-Processing/Cleaning*

Proses *cleaning* antara lain membuang duplikasi data, memeriksa data yang inkonsisten dan memperbaiki kesalahan pada data. Pada proses ini dilakukan juga proses *enrichment*, yaitu proses memperkaya data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD.

3. *Transformation*

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*.

4. *Data Mining*

Data Mining adalah proses mencari pola atau informasi menarikdalam data terpilih dengan menggunakan teknik atau metode tertentu.

5. *Interpretation/Evaluation*

6. Pola informasi yang dihasilkan dari proses *data mining* diterjemahkan menjadi bentuk yang lebih mudah dimengerti oleh pihak yang berkepentingan.

2.3. Klasifikasi

Klasifikasi merupakan bagian dari prediksi, dimana nilai yang diprediksi berupa label. Klasifikasi menentukan *class* atau grup untuk tiap contoh data, *input* dari model klasifikasi adalah atribut dari contoh data (*data samples*) dan *outputnya* adalah *class* dari *data samples* itu sendiri, dalam *machine learning* untuk membangun model klasifikasi digunakan metode *supervised learning* (HuiHuang, 2006). Metode *supervised learning* yaitu metode yang mencoba untuk menemukan hubungan antara atribut masukan dan atribut target, hubungan yang ditemukan diwakili dalam struktur yang disebut model.

Dalam klasifikasi kita dapat menentukan orang atau objek kedalam suatu kategori tertentu, contoh untuk masalah klasifikasi adalah menentukan apakah

seseorang pasien “mengidap” atau “tidak mengidap” penyakit tertentu. Informasi tentang pasien sebelumnya digunakan sebagai bahan untuk melatih algoritma untuk mendapatkan *rule* atau aturan.

Salah satu tujuan klasifikasi adalah untuk meningkatkan kehandalan hasil yang diperoleh dari data (Kahramanli & Allahverdi, 2008).

2.4. Algoritma C4.5

Algoritma C4.5 diperkenalkan oleh J. Ross Quinlan yang merupakan perkembangan dari algoritma ID3, algoritma tersebut digunakan untuk pohon keputusan. Pohon keputusan dianggap sebagai salah satu pendekatan yang paling populer, dalam klasifikasi pohon keputusan terdiri dari sebuah *node* yang membentuk akar, *node* akar tidak memiliki inputan. *Node* lain yang bukan sebagai akar tetapi memiliki tepat satu inputan disebut *node internal* atau *test node*, sedangkan *node* lainnya dinamakan daun. Daun mewakili nilai target yang paling tepat dari salah satu *class* (Maimon & Rokack, 2010).

Langkah-langkah membangun pohon keputusan menggunakan algoritma C4.5 adalah sebagai berikut (Kusrini & Luthfi, 2009):

1. Pilih atribut sebagai akar.

Pemilihan atribut sebagai akar berdasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung nilai *gain* tertinggi digunakan persamaan berikut:

$$Gain(S, A) = Entropy(S) - \sum_{i=0}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Keterangan:

S : himpunan kasus

A : atribut

n : jumlah partisi atribut A

|S_i| : jumlah kasus pada partisi ke-i

$|S|$: jumlah kasus dalam S

Nilai entropi dapat dihitung dengan cara berikut:

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

Dimana:

S : himpunan kasus

n : jumlah partisi S

Pi : proporsi dari S_i terhadap S

2. Buat cabang untuk tiap-tiap nilai.

3. Bagi kasus dalam cabang.

Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

2.5. Decision Tree

Pohon keputusan merupakan salah satu metode klasifikasi dan prediksi yang sangat kuat dan terkenal dalam penerapan *data mining*. Pada dasarnya *Decision Tree* mengubah data menjadi pohon keputusan (*decision tree*) dan aturan-aturan keputusan (*rule*).

Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target.

Sebuah pohon keputusan mungkin dibangun dengan saksama secara manual atau dapat tumbuh secara otomatis dengan menerapkan salah satu atau beberapa algoritma pohon keputusan untuk memodelkan himpunan data yang belum terklasifikasi. Banyak algoritma yang dipakai dalam pembentukan pohon keputusan antara lain ID3, CART dan C4.5 (Larose, 2005).

Variabel tujuan biasanya dikelompokkan dengan pasti dan model pohon keputusan lebih mengarah pada perhitungan probabilitas dari tiap-tiap record

kategori-kategori tersebut atau untuk mengklasifikasikan record dengan mengelompokkannya dalam satu kelas. Decision tree membuat set rule yang paling efisien dan kemungkinan terkecil yang membuatnya menjadi predictive model yang baik. Jika terdapat overlap diantara dua prediktor maka yang terbaik dari keduanya yang akan diambil. Pada sistem rule induction, keduanya akan diambil pada sistem ruleinductin salah satunya akan menjadi lemah atau kurang akurat.

Kelebihan-kelebihan *decision tree* adalah sebagai berikut:

1. Menyediakan visual result
2. Dibangun berdasarkan rule-rule yang dapat dimengerti dan dipahami
3. Bersifat *predictive*
4. Memungkinkan untuk melakukan prediksi
5. Menampilkan apa yang penting.

Algoritma akan mengidentifikasi atribut yang paling relevan dan akan mengidentifikasi suatu *set rule* yang akan memberikan *presentase* kemungkinan akan terjadi hal demikian dikemudian hari.

Decision tree dibuat dengan menggunakan sebuah teknik yang disebut *recursive partitioning*. Algoritma akan mendefinisikan atribut yang paling relevan dan akan men-split data yang ada berdasarkan atribut tersebut. Setiap *partition* disebut sebagai *rule*. Proses akan di ulang terus untuk setiap subgroup sampai ditemukan sebuah *good stopping point*. *Information gain measure* digunakan untuk memilih atribut mana yang akan dites pada *node tree* (Han & Kamber, 2006). Atribut dengan *information gain* yang paling tinggi akan dipilih sebagai test *attribute* untuk *current node*.

2.6. Naive Bayes

Naive Bayes merupakan metode yang tidak memiliki aturan, *naive bayes* menggunakan cabang matematika yang dikenal dengan teori probabilitas untuk mencari peluang terbesar dari kemungkinan klasifikasi, dengan cara melihat frekuensi tiap klasifikasi pada data training. *Naive Bayes* merupakan metode klasifikasi populer dan masuk dalam sepuluh algoritma terbaik dalam data mining, algoritma ini juga dikenal dengan nama *Idiot's Bayes*, *Simple Bayes* dan

Independence Bayes (Bramer, 2007). Klasifikasi *Bayes* di dasarkan pada teorema bayes, diambil dari nama seorang ahli matematika yang juga menteri Prebysterian Inggris, Thomas Bayes(1702-1761). Yaitu:

$$P(x|y) = \frac{P(y|x) P(x)}{P(y)}$$

Keterangan:

Y : data dengan kelas yang belum diketahui

X : hipotesis data y merupakan suatu kelas spesifik

$P(x|y)$: probabilitas hipotesis x berdasarkan kondisi y (posteriori probability)

$P(x)$: probabilitas hipotesis x (prior probability)

$P(y|x)$: probabilitas y berdasarkan kondisi pada hipotesis x

$p(y)$: probabilitas dari y

2.7. Rapid Miner

Rapid Miner merupakan perangkat lunak yang dibuat oleh Dr. Markus Hofmann dari Institute of Technology Blanchardstown dan Raif Klinkenberg dari rapid-i.com dengan tampilan GUI (*Graphical User Interface*) sehingga memudahkan pengguna dalam menggunakan perangkat lunak ini. Perangkat lunak ini bersifat *open source* dan dibuat dengan menggunakan bahasa java dibawah lisensi *GNU Public License* dan Rapid Miner dapat dijalankan disistem operasi manapun. Dengan menggunakan Rapid Miner, tidak dibutuhkan kemampuan koding khusus, karena semua fasilitas sudah disediakan. Rapid Miner dikhususkan untuk penggunaan data mining.

2.8. Evaluasi dan Validasi

Validasi adalah proses mengevaluasi akurasi prediksi dari sebuah model, validasi mengacu untuk mendapatkan prediksi dengan menggunakan model yang ada kemudian membandingkan hasil yang diperoleh dengan hasil yang diketahui (Gorunescu, 2011).

Mengevaluasi akurasi dari model klasifikasi sangat penting, akurasi dari sebuah model mengindikasikan kemampuan model tersebut untuk memprediksi *class* target (Vercellis, 2009).

Untuk mengevaluasi model digunakan metode *confusion matrix*, dan kurva ROC (*Receiver Operating Characteristic*).

2.9. Confusion Matrix

Confusion matrix memberikan rincian klasifikasi, kelas yang diprediksi akan ditampilkan di bagian atas matrix dan kelas yang diobservasi ditampilkan di bagian kiri (Gorunescu, 2011). Evaluasi model *confussion matrix* menggunakan tabel seperti matrix dibawah ini:

Tabel 1. Matrik Klasifikasi untuk Model 2 Class

Classifi cation	Predicted Class		
		Class=Yes	Class=No
Observ ed Class	Class =Yes	(True Positive – TP)	(False Negative – FN)
	Class =No	(False Positive – FP)	(True Negative – TN)

Sumber: Gorunescu(2011)

Akurasi dapat dihitung dengan menggunakan rumus berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

TP : Jumlah kasus positif yang diklasifikasikan sebagai positif

FP : Jumlah kasus negatif yang diklasifikasikan sebagai positif

TN : Jumlah kasus negatif yang diklasifikasikan sebagai negatif

FN : Jumlah kasus positif yang diklasifikasikan sebagai negatif

2.10. Kurva ROC

Kurva ROC banyak digunakan untuk menilai hasil prediksi, kurva ROC adalah teknik untuk memvisualisasikan, mengatur, dan memilih pengklasifikasian berdasarkan kinerja mereka (Gorunescu, 2011).

Kurva ROC adalah *tool* dua dimensi yang digunakan untuk menilai kinerja klasifikasi yang menggunakan dua *class* keputusan, masing-masing objek dipetakan ke salah satu elemen dari himpunan pasangan, positif atau negatif. Pada kurva ROC, TP *rate* diplot pada sumbu Y dan FP *rate* diplot pada sumbu X.

Untuk klasifikasi *data mining*, nilai AUC dapat dibagi menjadi beberapa kelompok (Gorunescu, 2011).

- a. 0.90-1.00 = *Excellent Classification*
- b. 0.80-0.90 = *Good Classification*
- c. 0.70-0.80 = *Fair Classification*
- d. 0.60-0.70 = *Poor Classification*
- e. 0.50-0.60 = *Failur*

The Area Under Curve (AUC) dihitung untuk mengukur perbedaan performansi metode yang digunakan. AUC dihitung menggunakan rumus (Liao & Triantaphyllou, 2007):

$$\theta^r = \frac{1}{mn} \sum_j^n = 1 \sum_i^m = 1 \psi(x_i^r, x_j^r)$$

Dimana

$$\psi(X, Y) = \begin{cases} 1 & Y < X \\ \frac{1}{2} & Y = X \\ 1 & Y > X \end{cases}$$

X= Output Positif

Y = Output Negatif

BAB III

METODOLOGI PENELITIAN

3.1.Desain Penelitian

Menurut (Dawson, 2009) ada empat metode penelitian yang umum digunakan yaitu tindakan penelitian, eksperimen, studi kasus dan survey. Dalam konteks penelitian, metode yang dilakukan mengacu kepada pemecahan masala yang meliputi mengumpulkan data, merumuskan hipotesis atau proposisi, pengujian hipotesis, menafsirkan hasil, dan kesimpulan (Berndtssom, Hansson, Olsson, & Lundell, 2008).

Dalam penelitian ini dilakukan beberapa langkah yang dilakukan dalam proses penelitian.

1. Pengumpulan data

Pada tahap ini ditentukan data yang akan diproses. Mencari data yang tersedia, memperoleh data tambahan yang dibutuhkan, mengintegrasikan semua data kedalam data set, termasuk variabel yang diperlukan dalam proses.

2. Pengolahan data awal

Ditahap ini dilakukan penyeleksian data, data dibersihkan dan ditransformasikan kebentuk yang diinginkan sehingga dapat dilakukan persiapan dalam pembuatan model.

3. Metode yang diusulkan

Pada tahap ini data dianalisis, dikelompokkan variabel mana yang berhubungan dengan satu sama lainnya. Setelah data dianalisis lalu diterapkan model-model yang sesuai dengan jenis data. Pembagian data kedalam data latihan (*training data*) dan data uji (*testing data*) juga diperlukan untuk pembuatan model.

4. Eksperimen dan pengujian metode

Pada tahap ini model yang diusulkan akan diuji untuk melihat hasil berupa *rule* yang akan dimanfaatkan dalam pengambilan keputusan.

5. Evaluasi dan validasi

Pada tahap ini dilakukan evaluasi terhadap model yang ditetapkan untuk mengetahui tingkat keakurasian model.

3.2. Pengumpulan Data

Dalam pengumpulan data terdapat sumber data, sumber data yang terhimpun langsung oleh peneliti disebut dengan sumber primer, sedangkan apabila melalui tangan kedua disebut sumber sekunder (Riduan, 2008). Data yang diperoleh adalah data sekunder karena diperoleh dari Pima Indian diabetes database dalam UCI (singkatan dari Pima Diabetes). Masalah yang harus dipecahkan di sini adalah prediksi terjadinya diabetes melitus dalam waktu 5 tahun dengan menggunakan Pima yang berisi 786 orang yang diperiksa dan sebanyak 500 pasien tidak terdeteksi terkena penyakit diabetes, sehingga 268 pasien terdeteksi penyakit diabetes. Dengan atribut dari penyakit diabetes adalah berapa kali hamil, konsentrasi glukosa, tekanan darah, ketebalan lipatan kulit, serum insulin, indeks massa tubuh, diabetes silsilah fungsi dan umur dan kelas sebagai label yang terdiri atas ya dan tidak. Data pasien penyakit diabetes bisa dilihat pada tabel 2.

3.3. Pengolahan Data Awal

Jumlah data awal yang diperoleh dari pengumpulan data yaitu sebanyak 768 data, namun tidak semua data dapat digunakan dan tidak semua atribut digunakan karena harus melalui beberapa tahap pengolahan awal data (*preparation data*).

Untuk mendapatkan data yang berkualitas, beberapa teknik yang dilakukan sebagai berikut (Vecellis, 2009):

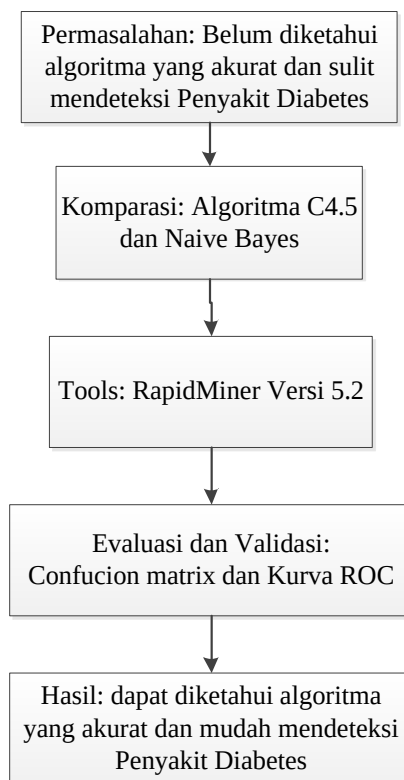
1. Data validation, untuk mengidentifikasi dan menghapus data yang ganjil (*outlier/noise*), data yang tidak konsisten, dan data yang tidak lengkap (*missing value*).
2. Data *integration and transformation*, untuk meningkatkan akurasi dan efisiensi algoritma. Data yang digunakan dalam penulisan ini bernilai kategorikal. Data ditransformasikan kedalam *software Rapidminer*. Tabel kategorikal atribut terlihat pada table 2.
3. Data *size reduction and discretization*, untuk memperoleh data set dengan jumlah atribut dan record yang lebih sedikit tetapi bersifat informative.

Tabel 2. Tabel Atribut yang digunakan (UCI Repository)

No	Atribut	Nilai
1	Berapa Kali Hamil	Berapa kali wanita hamil
2	Konsentrasi Glukosa	Konsentrasi glukosa plasma 2 jam dalam tes toleransi glukosa oral
3	Tekanan Darah	Tekanan Darah diastolik (mmHg)
4	Lipatan Kulit	Triceps ketebalan lipatan kulit (mm)
5	Serum Insulin	2-Jam serum insulin (mu U / ml)
6	Massa Tubuh	Indeks massa tubuh (berat dalam kg / (tinggi dalam m) ²)
7	Diabetes Silsilah Fungsi	Diabetes silsilah fungsi
8	Umur	Umur (tahun)

3.4. Metode yang di Usulkan

Pada tahap modeling ini dilakukan pemrosesan data traning sehingga akan membahas metode algoritma yang diuji dengan memasukan data penyakit diabetes kemudian di analisa dan dikomparasi.



Gambar 1. Metode yang diusulkan

Metode yang diusulkan dari penelitian ini, dimulai dari problem (permasalahan) analisa penyakit diabetes kemudian dibuat *approach* (model) dalam bentuk algoritma C4.5 dan *Naive Bayes* untuk memecahkan permasalahan, sedangkan *software* yang digunakan dalam penelitian ini adalah *Rapidminer* 5.2, *Microsoft Excell* 2013, pengujian evaluasi dan validasi untuk mengukur akurasi menggunakan *confusion matrix* dan *kurva ROC*, serta hasil dari penelitian didapat di antara ke dua algoritma tersebut, didapat algoritma yang terbaik dalam prediksi penyakit diabetes.

3.5.Eksperimen dan Pengujian Metode

Tahap modeling untuk menyelesaikan prediksi penyakit diabetes dengan menggunakan dua metode yaitu algoritma C4.5 dan *Naive Bayes*. Penelitian ini termasuk penelitian eksperimen, dimana penelitian ini dimulai dengan menentukan model yang digunakan, memasukan data training dan testing kedalam model dan mengujinya dengan tools rapidminer.

Dalam penelitian eksperimen digunakan spesifikasi software dan hardware sebagai alat bantu dalam penelitian pada Tabel 3:

Tabel 3. Spesifikasi hardware dan software

Software	Hardware
Sistem Operasi: Windows 8	CPU: Intel Pentium Dual Core
Data Mining: Rapid Miner versi 5.2	Memory: 2 GB
<i>Microsoft Excell</i> 2013	Hardisk: 160 GB

3.6.Evaluasi dan Validasi Hasil

Evaluasi dari penelitian ini yaitu untuk mengukur apakah model yang telah dikembangkan berhasil atau tidak, evaluasi digunakan untuk mengukur keakuratan hasil yang dicapai oleh model.

Pada penelitian ini digunakan penerapan algoritma C4.5 dengan menentukan nilai *weight* terlebih dahulu. Setelah didapatkan nilai akurasi dan AUC terbesar, nilai *weight* tersebut akan dijadikan nilai yang akan digunakan untuk mencari nilai akurasi dan AUC tertinggi. Sedangkan penerapan algoritma *Naive Bayes* beracuan pada nilai *weight* pada algoritma tersebut. Setelah ditemukan nilai akurasi yang paling ideal dari parameter tersebut langkah selanjutnya adalah menentukan nilai *weight*. Setelah ditemukan nilai akurasi yang paling ideal dari parameter tersebut langkah selanjutnya adalah menentukan *weight* sehingga terbentuk struktur algoritma yang ideal untuk pemecahan masalah tersebut.

BAB IV HASIL PENELITIAN DAN PEMBAHASAN

4.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini bersumber dari alamat web: <http://archive.ics.uci.edu/ml/>. Data merupakan hasil pemeriksaan terhadap 768 orang, 500 orang tidak terdeteksi penyakit diabetes dan 268 orang terdeteksi menderita penyakit diabetes. Pada data diabetes ini terdiri dari 9 atribut, 8 atribut predictor dan 1 atribut tujuan. Seperti terlihat pada Tabel 2:

Tabel 4. Data Pasien Diabetes

No.	Jumlah Hamil	Konsentrasi Glukosa	Tekanan Darah	Lipatan Kulit	Serum Insulin	IMB	Riwayat Diabetes	Umur	Hasil
1	6	148	72	35	0	33.6	0.627	50	1
2	1	85	66	29	0	26.6	0.351	31	0
3	8	183	64	0	0	23.3	0.672	32	1
4	1	89	66	23	94	28.1	0.167	21	0
5	0	137	40	35	168	43.1	2.288	33	1
6	5	116	74	0	0	25.6	0.201	30	0
7	3	78	50	32	88	31.0	0.248	26	1
8	10	115	0	0	0	35.3	0.134	29	0
9	2	197	70	45	543	30.5	0.158	53	1
10	8	125	96	0	0	0.0	0.232	54	1
11	4	110	92	0	0	37.6	0.191	30	0
12	10	168	74	0	0	38.0	0.537	34	1
13	10	139	80	0	0	27.1	1.441	57	0
14	1	189	60	23	846	30.1	0.398	59	1
15	5	166	72	19	175	25.8	0.587	51	1

No.	Jumlah Hamil	Konsentrasi Glukosa	Tekanan Darah	Lipatan Kulit	Serum Insulin	IMB	Riwayat Diabetes	Umur	Hasil
16	7	100	0	0	0	30.0	0.484	32	1
17	0	118	84	47	230	45.8	0.551	31	1
18	7	107	74	0	0	29.6	0.254	31	1
19	1	103	30	38	83	43.3	0.183	33	0
20	1	115	70	30	96	34.6	0.529	32	1
21	3	126	88	41	235	39.3	0.704	27	0
22	8	99	84	0	0	35.4	0.388	50	0
23	7	196	90	0	0	39.8	0.451	41	1
24	9	119	80	35	0	29.0	0.263	29	1
25	11	143	94	33	146	36.6	0.254	51	1

Sumber: <http://archive.ics.uci.edu/ml/>.

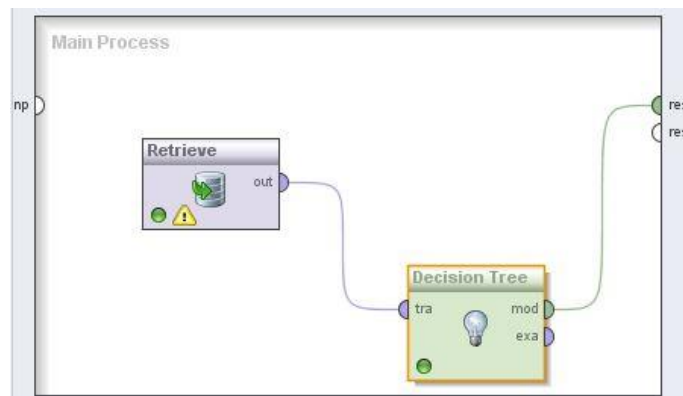
Dari tabel 2 diatas merupakan sample dari data penyakit diabetes dan data yang didapat tidak disertai keterangan yang menjelaskan maksud secara rinci mengenai maksud data, sehingga peneliti harus menganalisa dengan langkah awal melakukan pencarian informasi mengenai diabetes. Setelah melakukan pencarian tersebut, maka didapat beberapa informasi dan keterangan yang dapat membuat peneliti lebih memahami mengenai data pasien diabetes. Dari tabel 2 di atas dapat dirincikan sebagai berikut:

Tabel 5. Atribut, Tipe, Ukuran dan Nilai Atribut

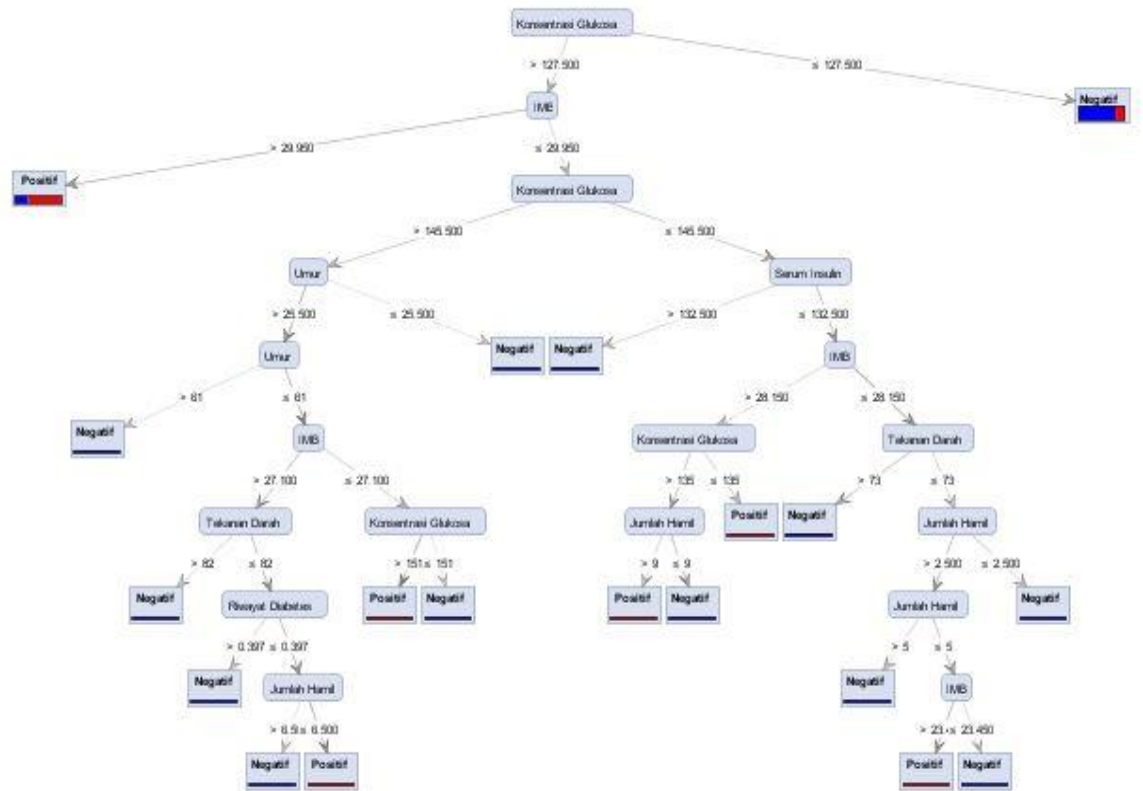
No.	Atribut	Tipe	Ukuran	Nilai Atribut
1.	Jumlah Hamil	Integer	Scale	Angka
2.	Konsentrasi Glukosa	Integer	Scale	Angka
3.	Tekanan Darah	Integer	Scale	Angka
4.	Lipatan Kulit	Integer	Scale	Angka
5.	Serum Insulin	Integer	Scale	Angka
6.	IMB	Real	Scale	Angka
7.	Riwayat Diabetes	Real	Scale	Angka
8.	Umur	Integer	Scale	Angka
9.	Hasil	String	Binominal	Positif, Negatif

4.2. Pengolahan Awal Data

Pada tahap ini menentukan data yang akan diproses. pada tabel 2 dapat dibuat suatu model pohon keputusan dengan menggunakan *Rapid Miner 5.2* dengan desain model sebagai berikut berikut:

**Gambar 2. Desain Model Algoritma C4.5**

Pada gambar 3 diatas merupakan bentuk desain model dari dataset pasien diabetes kemudian di relasikan ke algoritma C4.5 menggunakan *tools Rapid Miner* versi 5.2. Dengan desain diatas menghasilkan sebuah pohon keputusan sebagai berikut:



Gambar 3. Hasil Klasifikasi Menggunakan Algoritma C4.5

Pada Gambar 3 diatas terdapat 15 rule merupakan hasil dari klasifikasi dengan menggunakan model algoritma C4.5 dapat dijelaskan sebagai berikut:

R1: IF Konsentrasi Glukosa ≤ 127.500 THEN Hasil=Negatif

R2: IF Konsentrasi Glukosa > 127.500 ^ IMB > 29.950 THEN Hasil=Positif

R3: IF Konsentrasi Glukosa > 127.500 ^ IMB ≤ 29.950 ^ Konsentrasi Glukosa ≤ 145.500 ^ Serum Insulin > 132.500 THEN Hasil=Negatif

- R4: IF Konsentrasi Glukosa >127.500 $\wedge$ IMB ≤ 29.950 $\wedge$ Konsentrasi Glukosa ≤ 145.500 $\wedge$ Serum Insulin ≤ 132.500 $\wedge$ IMB >28.150 $\wedge$ Konsentrasi Glukosa ≤ 135 THEN Hasil=Positif
- R5: IF Konsentrasi Glukosa >127.500 $\wedge$ IMB ≤ 29.950 $\wedge$ Konsentrasi Glukosa ≤ 145.500 $\wedge$ Serum Insulin ≤ 132.500 $\wedge$ IMB >28.150 $\wedge$ Konsentrasi Glukosa >135 THEN Hasil=Negatif
- R6: IF Konsentrasi Glukosa >127.500 $\wedge$ IMB ≤ 29.950 $\wedge$ Konsentrasi Glukosa ≤ 145.500 $\wedge$ Serum Insulin ≤ 132.500 $\wedge$ IMB ≤ 28.150 $\wedge$ Tekanan Darah >73 THEN Hasil=Negatif
- R7: IF Konsentrasi Glukosa >127.500 $\wedge$ IMB ≤ 29.950 $\wedge$ Konsentrasi Glukosa ≤ 145.500 $\wedge$ Serum Insulin ≤ 132.500 $\wedge$ IMB ≤ 28.150 $\wedge$ Tekanan Darah ≤ 73 $\wedge$ Jumlah Hamil ≤ 2.500 THEN Hasil=Negatif
- R8: IF Konsentrasi Glukosa >127.500 $\wedge$ IMB ≤ 29.950 $\wedge$ Konsentrasi Glukosa ≤ 145.500 $\wedge$ Serum Insulin ≤ 132.500 $\wedge$ IMB ≤ 28.150 $\wedge$ Tekanan Darah ≤ 73 $\wedge$ Jumlah Hamil >2.500 $\wedge$ Jumlah Hamil ≤ 5 THEN Hasil=Positif
- R9: IF Konsentrasi Glukosa >127.500 $\wedge$ IMB ≤ 29.950 $\wedge$ Konsentrasi Glukosa ≤ 145.500 $\wedge$ Serum Insulin ≤ 132.500 $\wedge$ IMB ≤ 28.150 $\wedge$ Tekanan Darah ≤ 73 $\wedge$ Jumlah Hamil >2.500 $\wedge$ Jumlah Hamil >5 THEN Hasil=Negatif
- R10: IF Konsentrasi Glukosa >127.500 $\wedge$ IMB ≤ 29.950 $\wedge$ Konsentrasi Glukosa >145.500 $\wedge$ Umur ≤ 25.500 THEN Hasil=Negatif
- R11: IF Konsentrasi Glukosa >127.500 $\wedge$ IMB ≤ 29.950 $\wedge$ Konsentrasi Glukosa >145.500 $\wedge$ Umur >25.500 $\wedge$ Umur >61 THEN Hasil=Negatif
- R12: IF Konsentrasi Glukosa >127.500 $\wedge$ IMB ≤ 29.950 $\wedge$ Konsentrasi Glukosa >145.500 $\wedge$ Umur >25.500 $\wedge$ Umur ≤ 61 $\wedge$ IMB ≤ 27.100 THEN Hasil=Positif
- R13: IF Konsentrasi Glukosa >127.500 $\wedge$ IMB ≤ 29.950 $\wedge$ Konsentrasi Glukosa >145.500 $\wedge$ Umur >25.500 $\wedge$ Umur ≤ 61 $\wedge$ IMB >27.100 $\wedge$ Tekanan Darah >82 THEN Hasil=Negatif

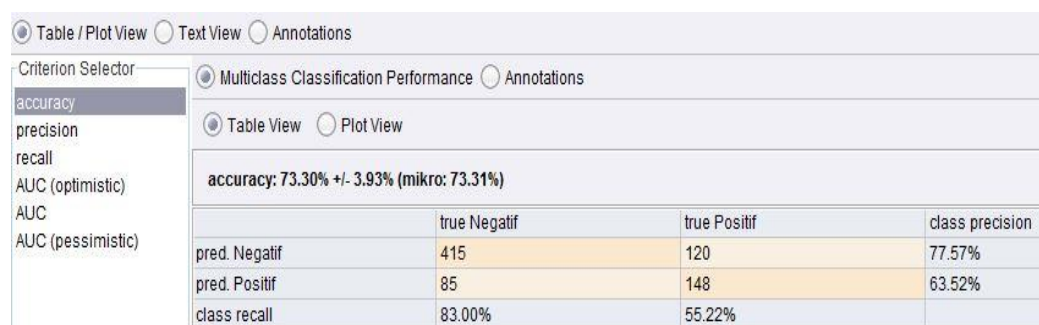
R14: IF Konsentrasi Glukosa >127.500 $\wedge$ IMB ≤ 29.950 $\wedge$ Konsentrasi Glukosa >145.500 $\wedge$ Umur >25.500 $\wedge$ Umur ≤ 61 $\wedge$ IMB >27.100 $\wedge$ Tekanan Darah ≤ 82 $\wedge$ Riwayat Diabetes >0.397 THEN Hasil=Negatif

R15: IF Konsentrasi Glukosa >127.500 $\wedge$ IMB ≤ 29.950 $\wedge$ Konsentrasi Glukosa >145.500 $\wedge$ Umur >25.500 $\wedge$ Umur ≤ 61 $\wedge$ IMB >27.100 $\wedge$ Tekanan Darah ≤ 82 $\wedge$ Riwayat Diabetes ≤ 0.397 THEN Hasil=Positif

4.3. Pengukuran Penelitian

Dari data eksperimen akan diujikan dengan menggunakan metode *10-fold cross-validation*, dimana data secara acak (*random*) akan dibagi menjadi 10 bagian. Pembagian menjadi 10 bagian merupakan metode yang paling tepat untuk mendapatkan estimasi terbaik menentukan kesalahan. Setiap bagian akan dihitung tingkat kesalahan setelah itu secara keseluruhan akan dihitung rata-ratanya. Setelah dilakukan klasifikasi model data, maka tahap selanjutnya melakukan pengujian akurasi data uji, metode yang digunakan untuk menganalisa model klasifikasi yaitu:

4.3.1. Confusion Matrix:



	true Negatif	true Positif	class precision
pred. Negatif	415	120	77.57%
pred. Positif	85	148	63.52%
class recall	83.00%	55.22%	

Gambar 4. Hasil Akurasi Prediksi Penyakit Diabetes Menggunakan Algoritma C4.5

Berdasarkan gambar 4 menunjukkan bahwa, diketahui dari 768 data pasien penyakit diabetes, ada 500 orang tidak terdeteksi diabetes tetapi pada hasil tabel *confusion matrix* diatas ada 415 pasien diprediksi negatif maka hasilnya sesuai

dengan prediksi yaitu negatif, 120 pasien diprediksi negatif tetapi hasilnya adalah positif, sedangkan 268 orang diprediksi positif tetapi pada gambar diatas menunjukkan bahwa ada 85 pasien yang diprediksi positif tetapi hasilnya adalah negatif dan 148 diprediksi positif maka hasilnya sesuai dengan prediksi yaitu positif dan tingkat akurasi dengan menggunakan algoritma C4.5 adalah 73.30%, dan dapat dihitung untuk mencari nilai *accuracy*, yaitu:

Keterangan:

TP = 148

FP = 120

TN= 415

FN = 85

$$\begin{aligned}\text{Akurasi} &= (TP+TN) / (TP+TN+FP+FN) \\ &= (148+415) / (148+415+120+85) \\ &= 0.7330 \text{ (73.30\%)}\end{aligned}$$

Sedangkan untuk algoritma Naive Bayes akan menghasilkan nilai seperti di bawah ini:

Criterion Selector	☒ Multiclass Classification Performance ☐ Annotations		
accuracy	☒ Table View ☐ Plot View		
precision	accuracy: 75.13% +/- 3.26% (mikro: 75.13%)		
recall			
AUC (optimistic)			
AUC			
AUC (pessimistic)			
	true Negatif	true Positif	class precision
pred. Negatif	425	116	78.56%
pred. Positif	75	152	66.96%
class recall	85.00%	56.72%	

Gambar 5. Hasil Akurasi Prediksi Penyakit Diabetes Menggunakan Algoritma Naive Bayes

Berdasarkan gambar 5 menunjukkan bahwa, diketahui dari 768 data pasien penyakit diabetes, ada 500 orang tidak terdeteksi diabetes tetapi pada hasil tabel *confusion matrix* diatas ada 425 pasien diprediksi negatif maka hasilnya sesuai dengan prediksi yaitu negatif, 116 pasien diprediksi negatif tetapi hasilnya adalah positif, sedangkan 268 orang diprediksi positif tetapi pada gambar diatas menunjukkan bahwa ada 75 pasien yang diprediksi positif tetapi hasilnya adalah

negatif dan 152 diprediksi positif maka hasilnya sesuai dengan prediksi yaitu positif tingkat akurasi dengan menggunakan algoritma *Naive Bayes* adalah 75,13%, dan dapat dihitung untuk mencari nilai *accuracy*, yaitu:

Keterangan:

TP = 152

FP = 116

TN = 425

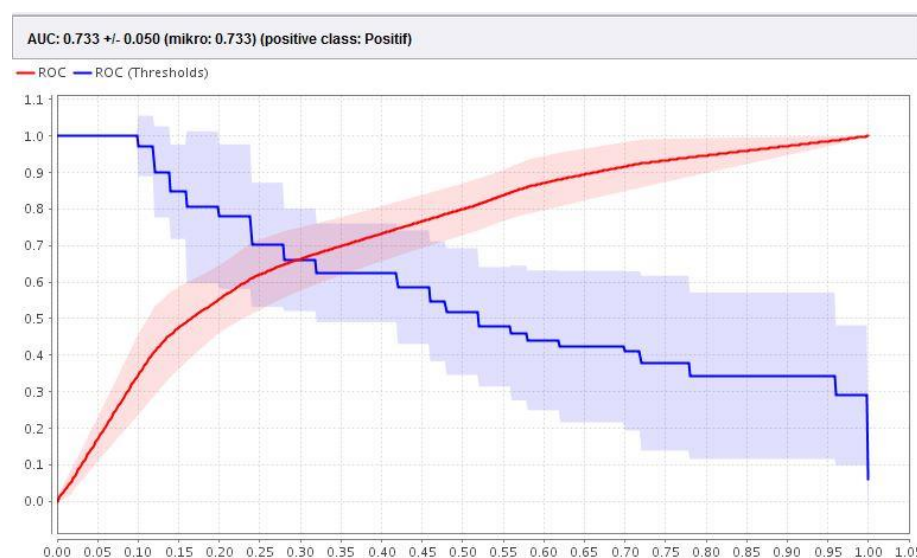
FN = 75

$$\begin{aligned}\text{Akurasi} &= (TP+TN) / (TP+TN+FP+FN) \\ &= (152+425) / (152+425+116+75) \\ &= 0.7513 \text{ (75.13\%)}\end{aligned}$$

Hasil pengujian *confusion matrix* diatas diketahui bahwa model algoritma C4.5 mempunyai akurasi 73.30% sedangkan model *Naive Bayes* memiliki akurasi 75.13%, tingkat akurasi *Naive Bayes* lebih tinggi dibandingkan dengan algoritma C4.5 sebesar 1.83%.

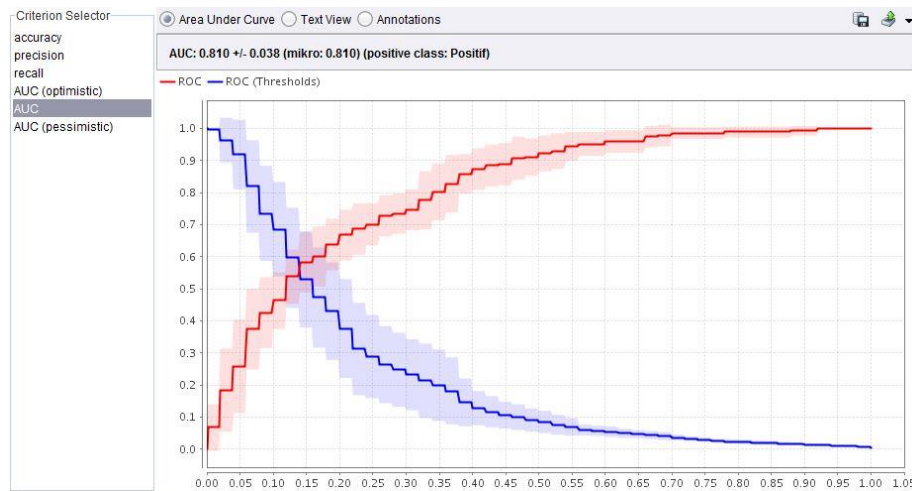
4.3.2. Kurva ROC (*Receiver Operating Characteristic*)

Data uji diatas akan dinilai hasil prediksi dengan menggunakan grafik ROC untuk algoritma C4.5, visualisasi dari grafik ROC yaitu:



Gambar 6. Grafik ROC dari Model Algoritma C4.5

Dari gambar 6 terdapat grafik ROC dengan nilai AUC (*Area Under Curve*) sebesar 0.733 dimana diagnosa hasilnya *Fair Classification*. Sedangkan visualisasi grafik ROC dengan model algoritma Naive Bayes sebagai berikut:



Gambar 7. Grafik ROC dari Model Naive Bayes

Dari gambar 7 terdapat grafik ROC dengan nilai AUC (*Area Under Curve*) sebesar 0.810 dimana diagnosa hasilnya *Good Classification*.

4.4. Analisa Hasil Komparasi

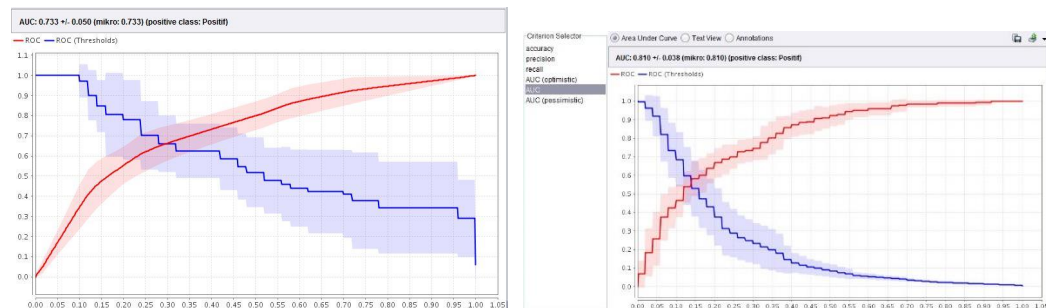
Dari hasil pengujian diatas baik evaluasi menggunakan *confusion matrix* maupun kurva *ROC* untuk model klasifikasi algoritma C4.5 dan *Naive Bayes* sebagai berikut:

Tabel 6. Hasil Komparasi Algoritma C4.5 dan Naive Bayes

	Accuracy	AUC
Decision Tree	73.30%	0.733
Naive Bayes	75.13%	0.810

Untuk evaluasi menggunakan kurva *ROC* sehingga menghasilkan nilai AUC (*Area Under Curve*) untuk model algoritma *Decision Tree* menghasilkan nilai

0.733 dengan nilai diagnosa Fair *Classification* sedangkan untuk algoritma *Naive Bayes* menghasilkan nilai 0.810 dengan nilai diagnosa *Good Classification* dan selisih nilai keduanya sebesar 1.83%. dapat dilihat pada gambar dibawah ini:



Gambar 8. Kurva ROC dengan Algoritma C4.5 dan Naive Bayes

Dengan demikian algoritma *Naive Bayes* dapat memberikan solusi untuk permasalahan dalam prediksi penyakit diabetes.

BAB V

PENUTUP

1.1. KESIMPULAN

Berdasarkan hasil pengujian dan analisis bahwa pengujian ini bertujuan untuk mengetahui diantara model algoritma C4.5 dan *Naive Bayes* yang memiliki akurasi paling tinggi untuk memprediksi penyakit diabetes. Hasil perbandingan antara C4.5 dan *Naive Bayes* diukur tingkat akurasinya menggunakan pengujian *Confusion Matrix* dan Kurva ROC. Berdasarkan hasil pengukuran tingkat akurasi kedua algoritma tersebut, diketahui bahwa nilai akurasi C4.5 adalah 73.30% dan nilai AUC adalah 0.733, sedangkan nilai akurasi *Naive Bayes* 75.13% dan nilai AUC adalah 0.810 dapat disimpulkan bahwa dengan menggunakan model *Naive Bayes* lebih tinggi tingkat akurasinya, dengan peningkatan akurasi sebesar 1.83% dan peningkatan nilai AUC sebesar 0.077 sedangkan hasil pengujian dari prediksi diabetes hasilnya termasuk *Good Clasification*.

1.2. SARAN

Beberapa saran yang dapat diberikan untuk pengembangan dan perbaikan diwaktu yang akan datang adalah:

1. Dapat melakukan komparasi lebih dari tiga metode algoritma dan dengan data yang lebih banyak lagi untuk prediksi penyakit diabetes sehingga diperoleh algoritma dengan tingkat akurasi yang lebih tinggi.
2. Dapat menggunakan metode optimasi seperti AdaBoost, dan lain-lain.

DAFTAR PUSTAKA

- [1] Berndtsson, M., Hansson, J., Olsson, B., & Lundell, B. (2008). *A Guide For Students In Computer Science And Information Systems*. London: Springer.
- [2] Bromuri, S., Schumacher, M. I., & Ruiz, J. (2011). Monitoring Gestational Diabetes Mellitus With Cognitive Agents And Agent Environments. *Acm International Conferences On Web Intelligence And Intelligent Agent Technology* , 409-414.
- [3] Bramer, M.(2007). *Principles of Data Mining* London: Springer Clark. L.A., Kochanska, G., & Ready, R. (2000). Mothers' personality and its interaction with child temperament as predictors of parenting behavior. *Journal of Personality and Social Psychology*, 79, 274-285.
- [4] Dawson, C. W. (2009). *Projects In Computing And Information System A Student's Guide*. England: Addison-Wesley.
- [5] Gorunescu, F. (2011). *Data Mining Concepts, Models and Technique*. Berlin: Springer
- [6] Han, J., & Kamber, M. (2006). *Data Mining Concept and Tehniques*. San Fransisco: Morgan Kauffman
- [7] Hui-Huang, H. (2006). *Advanced Data mining Technologies in Bioinformatics*. United States of America: Idea Group Publishing.
- [8] Jayalakshmi, T., Santhakumaran, A. (2010). *Improved Gradient Descent Back Propagation Neural Network for Diagnoses of Type II Diabetes Militus*. *Global Journal of Computer Science and Technology*. Vol.9 Issue 5.
- [9] Kahramanli, H., & Allahverdi, N. (2008). *Design of A Hybrid System for the Diabetes and Heart Diseases*. *Expert System with Application*, 82-89
- [10] Kusrini, & Luthfi, T. E.(2009). *Algoritma Data Mining*. Yogyakarta: Penerbit Andi
- [11] Larose, D. T. (2005). *Discovering knowledge in Data*. New Jersey: John Willey & Sons, Inc.
- [12] Maimon, o., & Rokach, L. (2010). *Data Mining and Knowladge Discovery Handbook Second Edition*. New York:Springer.

- [13] Mason, R. (2005). *The Natural Diabetes Cure*. Usa: 4th Printing Spring 2012.
- [14] Mehdi Neshat, and Mehdi Yaghobi. (2009, October, 20-22). “*Designing a Fuzzy Expert System of Diagnosing the Hepatitis B Intensity Rate and Comparing it with Adaptive Neural Network Fuzzy System*”. *Proceeding of the world congress on engineering and computer science* 2009, Vol II, WCECS 2009, ISBN:978-988-18210-2-7. pp 1-6, October 20-22.
- [15] Nurrahmani, U. (2012). *Stop!Diabetes Mellitus*. Yogyakarta: Familia
- [16] Riduwan. (2008). *Metode dan Teknik Menyusun Tesis*. Alfabeta. Bandung.
- [17] Robert, F. G., Zgonis, T., & Driver, V. R. (2006). *Diabetic Foot Disorders: A Clinical Practice Guideline (2006 Revision)*. *The Journal Of Foot & Ankle Surgery* , 3.
- [18] *university of California Irvine Machine learning Repository*. Dikutip 02 November 2015, dari <http://archive.ics.uci.edu/ml/machine-learning-databases/statlog/heart/hear.dat>
- [19] Vercellis, C. (2009). *Business Intelligence: Data Mining Optimization for Decision making*. United Kingdom: John Wiley & Sons.
- [20] Witten, I. H., Frank, E., & Hall, M.A.(2011) *Data Mining Practical Machine Learning Tools And Technique*. Burlington, Usa: Morgan kaufmann Publishers.

LAMPIRAN

Data Penyakit Diabetes

Jumlah Hamil	Konsentrasi Glukosa	Tekanan Darah	Lipatan Kulit	Serum Insulin	IMB	Riwayat Diabetes	Umur	Hasil
6	148	72	35	0	33.6	0.627	50	1
1	85	66	29	0	26.6	0.351	31	0
8	183	64	0	0	23.3	0.672	32	1
1	89	66	23	94	28.1	0.167	21	0
0	137	40	35	168	43.1	2.288	33	1
5	116	74	0	0	25.6	0.201	30	0
3	78	50	32	88	31.0	0.248	26	1
10	115	0	0	0	35.3	0.134	29	0
2	197	70	45	543	30.5	0.158	53	1
8	125	96	0	0	0.0	0.232	54	1
4	110	92	0	0	37.6	0.191	30	0
10	168	74	0	0	38.0	0.537	34	1
10	139	80	0	0	27.1	1.441	57	0
1	189	60	23	846	30.1	0.398	59	1
5	166	72	19	175	25.8	0.587	51	1
7	100	0	0	0	30.0	0.484	32	1
0	118	84	47	230	45.8	0.551	31	1
7	107	74	0	0	29.6	0.254	31	1
1	103	30	38	83	43.3	0.183	33	0
1	115	70	30	96	34.6	0.529	32	1
3	126	88	41	235	39.3	0.704	27	0
8	99	84	0	0	35.4	0.388	50	0
7	196	90	0	0	39.8	0.451	41	1
9	119	80	35	0	29.0	0.263	29	1
11	143	94	33	146	36.6	0.254	51	1
10	125	70	26	115	31.1	0.205	41	1
7	147	76	0	0	39.4	0.257	43	1
1	97	66	15	140	23.2	0.487	22	0
13	145	82	19	110	22.2	0.245	57	0
5	117	92	0	0	34.1	0.337	38	0
5	109	75	26	0	36.0	0.546	60	0
3	158	76	36	245	31.6	0.851	28	1
3	88	58	11	54	24.8	0.267	22	0
6	92	92	0	0	19.9	0.188	28	0
10	122	78	31	0	27.6	0.512	45	0
4	103	60	33	192	24.0	0.966	33	0

11	138	76	0	0	33.2	0.420	35	0
9	102	76	37	0	32.9	0.665	46	1
2	90	68	42	0	38.2	0.503	27	1
4	111	72	47	207	37.1	1.390	56	1
3	180	64	25	70	34.0	0.271	26	0
7	133	84	0	0	40.2	0.696	37	0
7	106	92	18	0	22.7	0.235	48	0
9	171	110	24	240	45.4	0.721	54	1
7	159	64	0	0	27.4	0.294	40	0
0	180	66	39	0	42.0	1.893	25	1
1	146	56	0	0	29.7	0.564	29	0
2	71	70	27	0	28.0	0.586	22	0
7	103	66	32	0	39.1	0.344	31	1
7	105	0	0	0	0.0	0.305	24	0
1	103	80	11	82	19.4	0.491	22	0
1	101	50	15	36	24.2	0.526	26	0
5	88	66	21	23	24.4	0.342	30	0
8	176	90	34	300	33.7	0.467	58	1
7	150	66	42	342	34.7	0.718	42	0
1	73	50	10	0	23.0	0.248	21	0
7	187	68	39	304	37.7	0.254	41	1
0	100	88	60	110	46.8	0.962	31	0
0	146	82	0	0	40.5	1.781	44	0
0	105	64	41	142	41.5	0.173	22	0
2	84	0	0	0	0.0	0.304	21	0
8	133	72	0	0	32.9	0.270	39	1
5	44	62	0	0	25.0	0.587	36	0
2	141	58	34	128	25.4	0.699	24	0
7	114	66	0	0	32.8	0.258	42	1
5	99	74	27	0	29.0	0.203	32	0
0	109	88	30	0	32.5	0.855	38	1
2	109	92	0	0	42.7	0.845	54	0
1	95	66	13	38	19.6	0.334	25	0
4	146	85	27	100	28.9	0.189	27	0
2	100	66	20	90	32.9	0.867	28	1
5	139	64	35	140	28.6	0.411	26	0
13	126	90	0	0	43.4	0.583	42	1
4	129	86	20	270	35.1	0.231	23	0
1	79	75	30	0	32.0	0.396	22	0
1	0	48	20	0	24.7	0.140	22	0
7	62	78	0	0	32.6	0.391	41	0
5	95	72	33	0	37.7	0.370	27	0

0	131	0	0	0	43.2	0.270	26	1
2	112	66	22	0	25.0	0.307	24	0
3	113	44	13	0	22.4	0.140	22	0
2	74	0	0	0	0.0	0.102	22	0
7	83	78	26	71	29.3	0.767	36	0
0	101	65	28	0	24.6	0.237	22	0
5	137	108	0	0	48.8	0.227	37	1
2	110	74	29	125	32.4	0.698	27	0
13	106	72	54	0	36.6	0.178	45	0
2	100	68	25	71	38.5	0.324	26	0
15	136	70	32	110	37.1	0.153	43	1
1	107	68	19	0	26.5	0.165	24	0
1	80	55	0	0	19.1	0.258	21	0
4	123	80	15	176	32.0	0.443	34	0
7	81	78	40	48	46.7	0.261	42	0
4	134	72	0	0	23.8	0.277	60	1
2	142	82	18	64	24.7	0.761	21	0
6	144	72	27	228	33.9	0.255	40	0
2	92	62	28	0	31.6	0.130	24	0
1	71	48	18	76	20.4	0.323	22	0
6	93	50	30	64	28.7	0.356	23	0
1	122	90	51	220	49.7	0.325	31	1
1	163	72	0	0	39.0	1.222	33	1
1	151	60	0	0	26.1	0.179	22	0
0	125	96	0	0	22.5	0.262	21	0
1	81	72	18	40	26.6	0.283	24	0
2	85	65	0	0	39.6	0.930	27	0
1	126	56	29	152	28.7	0.801	21	0
1	96	122	0	0	22.4	0.207	27	0
4	144	58	28	140	29.5	0.287	37	0
3	83	58	31	18	34.3	0.336	25	0
0	95	85	25	36	37.4	0.247	24	1
3	171	72	33	135	33.3	0.199	24	1
8	155	62	26	495	34.0	0.543	46	1
1	89	76	34	37	31.2	0.192	23	0
4	76	62	0	0	34.0	0.391	25	0
7	160	54	32	175	30.5	0.588	39	1
4	146	92	0	0	31.2	0.539	61	1
5	124	74	0	0	34.0	0.220	38	1
5	78	48	0	0	33.7	0.654	25	0
4	97	60	23	0	28.2	0.443	22	0
4	99	76	15	51	23.2	0.223	21	0

0	162	76	56	100	53.2	0.759	25	1
6	111	64	39	0	34.2	0.260	24	0
2	107	74	30	100	33.6	0.404	23	0
5	132	80	0	0	26.8	0.186	69	0
0	113	76	0	0	33.3	0.278	23	1
1	88	30	42	99	55.0	0.496	26	1
3	120	70	30	135	42.9	0.452	30	0
1	118	58	36	94	33.3	0.261	23	0
1	117	88	24	145	34.5	0.403	40	1
0	105	84	0	0	27.9	0.741	62	1
4	173	70	14	168	29.7	0.361	33	1
9	122	56	0	0	33.3	1.114	33	1
3	170	64	37	225	34.5	0.356	30	1
8	84	74	31	0	38.3	0.457	39	0
2	96	68	13	49	21.1	0.647	26	0
2	125	60	20	140	33.8	0.088	31	0
0	100	70	26	50	30.8	0.597	21	0
0	93	60	25	92	28.7	0.532	22	0
0	129	80	0	0	31.2	0.703	29	0
5	105	72	29	325	36.9	0.159	28	0
3	128	78	0	0	21.1	0.268	55	0
5	106	82	30	0	39.5	0.286	38	0
2	108	52	26	63	32.5	0.318	22	0
10	108	66	0	0	32.4	0.272	42	1
4	154	62	31	284	32.8	0.237	23	0
0	102	75	23	0	0.0	0.572	21	0
9	57	80	37	0	32.8	0.096	41	0
2	106	64	35	119	30.5	1.400	34	0
5	147	78	0	0	33.7	0.218	65	0
2	90	70	17	0	27.3	0.085	22	0
1	136	74	50	204	37.4	0.399	24	0
4	114	65	0	0	21.9	0.432	37	0
9	156	86	28	155	34.3	1.189	42	1
1	153	82	42	485	40.6	0.687	23	0
8	188	78	0	0	47.9	0.137	43	1
7	152	88	44	0	50.0	0.337	36	1
2	99	52	15	94	24.6	0.637	21	0
1	109	56	21	135	25.2	0.833	23	0
2	88	74	19	53	29.0	0.229	22	0
17	163	72	41	114	40.9	0.817	47	1
4	151	90	38	0	29.7	0.294	36	0
7	102	74	40	105	37.2	0.204	45	0

0	114	80	34	285	44.2	0.167	27	0
2	100	64	23	0	29.7	0.368	21	0
0	131	88	0	0	31.6	0.743	32	1
6	104	74	18	156	29.9	0.722	41	1
3	148	66	25	0	32.5	0.256	22	0
4	120	68	0	0	29.6	0.709	34	0
4	110	66	0	0	31.9	0.471	29	0
3	111	90	12	78	28.4	0.495	29	0
6	102	82	0	0	30.8	0.180	36	1
6	134	70	23	130	35.4	0.542	29	1
2	87	0	23	0	28.9	0.773	25	0
1	79	60	42	48	43.5	0.678	23	0
2	75	64	24	55	29.7	0.370	33	0
8	179	72	42	130	32.7	0.719	36	1
6	85	78	0	0	31.2	0.382	42	0
0	129	110	46	130	67.1	0.319	26	1
5	143	78	0	0	45.0	0.190	47	0
5	130	82	0	0	39.1	0.956	37	1
6	87	80	0	0	23.2	0.084	32	0
0	119	64	18	92	34.9	0.725	23	0
1	0	74	20	23	27.7	0.299	21	0
5	73	60	0	0	26.8	0.268	27	0
4	141	74	0	0	27.6	0.244	40	0
7	194	68	28	0	35.9	0.745	41	1
8	181	68	36	495	30.1	0.615	60	1
1	128	98	41	58	32.0	1.321	33	1
8	109	76	39	114	27.9	0.640	31	1
5	139	80	35	160	31.6	0.361	25	1
3	111	62	0	0	22.6	0.142	21	0
9	123	70	44	94	33.1	0.374	40	0
7	159	66	0	0	30.4	0.383	36	1
11	135	0	0	0	52.3	0.578	40	1
8	85	55	20	0	24.4	0.136	42	0
5	158	84	41	210	39.4	0.395	29	1
1	105	58	0	0	24.3	0.187	21	0
3	107	62	13	48	22.9	0.678	23	1
4	109	64	44	99	34.8	0.905	26	1
4	148	60	27	318	30.9	0.150	29	1
0	113	80	16	0	31.0	0.874	21	0
1	138	82	0	0	40.1	0.236	28	0
0	108	68	20	0	27.3	0.787	32	0
2	99	70	16	44	20.4	0.235	27	0

6	103	72	32	190	37.7	0.324	55	0
5	111	72	28	0	23.9	0.407	27	0
8	196	76	29	280	37.5	0.605	57	1
5	162	104	0	0	37.7	0.151	52	1
1	96	64	27	87	33.2	0.289	21	0
7	184	84	33	0	35.5	0.355	41	1
2	81	60	22	0	27.7	0.290	25	0
0	147	85	54	0	42.8	0.375	24	0
7	179	95	31	0	34.2	0.164	60	0
0	140	65	26	130	42.6	0.431	24	1
9	112	82	32	175	34.2	0.260	36	1
12	151	70	40	271	41.8	0.742	38	1
5	109	62	41	129	35.8	0.514	25	1
6	125	68	30	120	30.0	0.464	32	0
5	85	74	22	0	29.0	1.224	32	1
5	112	66	0	0	37.8	0.261	41	1
0	177	60	29	478	34.6	1.072	21	1
2	158	90	0	0	31.6	0.805	66	1
7	119	0	0	0	25.2	0.209	37	0
7	142	60	33	190	28.8	0.687	61	0
1	100	66	15	56	23.6	0.666	26	0
1	87	78	27	32	34.6	0.101	22	0
0	101	76	0	0	35.7	0.198	26	0
3	162	52	38	0	37.2	0.652	24	1
4	197	70	39	744	36.7	2.329	31	0
0	117	80	31	53	45.2	0.089	24	0
4	142	86	0	0	44.0	0.645	22	1
6	134	80	37	370	46.2	0.238	46	1
1	79	80	25	37	25.4	0.583	22	0
4	122	68	0	0	35.0	0.394	29	0
3	74	68	28	45	29.7	0.293	23	0
4	171	72	0	0	43.6	0.479	26	1
7	181	84	21	192	35.9	0.586	51	1
0	179	90	27	0	44.1	0.686	23	1
9	164	84	21	0	30.8	0.831	32	1
0	104	76	0	0	18.4	0.582	27	0
1	91	64	24	0	29.2	0.192	21	0
4	91	70	32	88	33.1	0.446	22	0
3	139	54	0	0	25.6	0.402	22	1
6	119	50	22	176	27.1	1.318	33	1
2	146	76	35	194	38.2	0.329	29	0
9	184	85	15	0	30.0	1.213	49	1

10	122	68	0	0	31.2	0.258	41	0
0	165	90	33	680	52.3	0.427	23	0
9	124	70	33	402	35.4	0.282	34	0
1	111	86	19	0	30.1	0.143	23	0
9	106	52	0	0	31.2	0.380	42	0
2	129	84	0	0	28.0	0.284	27	0
2	90	80	14	55	24.4	0.249	24	0
0	86	68	32	0	35.8	0.238	25	0
12	92	62	7	258	27.6	0.926	44	1
1	113	64	35	0	33.6	0.543	21	1
3	111	56	39	0	30.1	0.557	30	0
2	114	68	22	0	28.7	0.092	25	0
1	193	50	16	375	25.9	0.655	24	0
11	155	76	28	150	33.3	1.353	51	1
3	191	68	15	130	30.9	0.299	34	0
3	141	0	0	0	30.0	0.761	27	1
4	95	70	32	0	32.1	0.612	24	0
3	142	80	15	0	32.4	0.200	63	0
4	123	62	0	0	32.0	0.226	35	1
5	96	74	18	67	33.6	0.997	43	0
0	138	0	0	0	36.3	0.933	25	1
2	128	64	42	0	40.0	1.101	24	0
0	102	52	0	0	25.1	0.078	21	0
2	146	0	0	0	27.5	0.240	28	1
10	101	86	37	0	45.6	1.136	38	1
2	108	62	32	56	25.2	0.128	21	0
3	122	78	0	0	23.0	0.254	40	0
1	71	78	50	45	33.2	0.422	21	0
13	106	70	0	0	34.2	0.251	52	0
2	100	70	52	57	40.5	0.677	25	0
7	106	60	24	0	26.5	0.296	29	1
0	104	64	23	116	27.8	0.454	23	0
5	114	74	0	0	24.9	0.744	57	0
2	108	62	10	278	25.3	0.881	22	0
0	146	70	0	0	37.9	0.334	28	1
10	129	76	28	122	35.9	0.280	39	0
7	133	88	15	155	32.4	0.262	37	0
7	161	86	0	0	30.4	0.165	47	1
2	108	80	0	0	27.0	0.259	52	1
7	136	74	26	135	26.0	0.647	51	0
5	155	84	44	545	38.7	0.619	34	0
1	119	86	39	220	45.6	0.808	29	1

4	96	56	17	49	20.8	0.340	26	0
5	108	72	43	75	36.1	0.263	33	0
0	78	88	29	40	36.9	0.434	21	0
0	107	62	30	74	36.6	0.757	25	1
2	128	78	37	182	43.3	1.224	31	1
1	128	48	45	194	40.5	0.613	24	1
0	161	50	0	0	21.9	0.254	65	0
6	151	62	31	120	35.5	0.692	28	0
2	146	70	38	360	28.0	0.337	29	1
0	126	84	29	215	30.7	0.520	24	0
14	100	78	25	184	36.6	0.412	46	1
8	112	72	0	0	23.6	0.840	58	0
0	167	0	0	0	32.3	0.839	30	1
2	144	58	33	135	31.6	0.422	25	1
5	77	82	41	42	35.8	0.156	35	0
5	115	98	0	0	52.9	0.209	28	1
3	150	76	0	0	21.0	0.207	37	0
2	120	76	37	105	39.7	0.215	29	0
10	161	68	23	132	25.5	0.326	47	1
0	137	68	14	148	24.8	0.143	21	0
0	128	68	19	180	30.5	1.391	25	1
2	124	68	28	205	32.9	0.875	30	1
6	80	66	30	0	26.2	0.313	41	0
0	106	70	37	148	39.4	0.605	22	0
2	155	74	17	96	26.6	0.433	27	1
3	113	50	10	85	29.5	0.626	25	0
7	109	80	31	0	35.9	1.127	43	1
2	112	68	22	94	34.1	0.315	26	0
3	99	80	11	64	19.3	0.284	30	0
3	182	74	0	0	30.5	0.345	29	1
3	115	66	39	140	38.1	0.150	28	0
6	194	78	0	0	23.5	0.129	59	1
4	129	60	12	231	27.5	0.527	31	0
3	112	74	30	0	31.6	0.197	25	1
0	124	70	20	0	27.4	0.254	36	1
13	152	90	33	29	26.8	0.731	43	1
2	112	75	32	0	35.7	0.148	21	0
1	157	72	21	168	25.6	0.123	24	0
1	122	64	32	156	35.1	0.692	30	1
10	179	70	0	0	35.1	0.200	37	0
2	102	86	36	120	45.5	0.127	23	1
6	105	70	32	68	30.8	0.122	37	0

8	118	72	19	0	23.1	1.476	46	0
2	87	58	16	52	32.7	0.166	25	0
1	180	0	0	0	43.3	0.282	41	1
12	106	80	0	0	23.6	0.137	44	0
1	95	60	18	58	23.9	0.260	22	0
0	165	76	43	255	47.9	0.259	26	0
0	117	0	0	0	33.8	0.932	44	0
5	115	76	0	0	31.2	0.343	44	1
9	152	78	34	171	34.2	0.893	33	1
7	178	84	0	0	39.9	0.331	41	1
1	130	70	13	105	25.9	0.472	22	0
1	95	74	21	73	25.9	0.673	36	0
1	0	68	35	0	32.0	0.389	22	0
5	122	86	0	0	34.7	0.290	33	0
8	95	72	0	0	36.8	0.485	57	0
8	126	88	36	108	38.5	0.349	49	0
1	139	46	19	83	28.7	0.654	22	0
3	116	0	0	0	23.5	0.187	23	0
3	99	62	19	74	21.8	0.279	26	0
5	0	80	32	0	41.0	0.346	37	1
4	92	80	0	0	42.2	0.237	29	0
4	137	84	0	0	31.2	0.252	30	0
3	61	82	28	0	34.4	0.243	46	0
1	90	62	12	43	27.2	0.580	24	0
3	90	78	0	0	42.7	0.559	21	0
9	165	88	0	0	30.4	0.302	49	1
1	125	50	40	167	33.3	0.962	28	1
13	129	0	30	0	39.9	0.569	44	1
12	88	74	40	54	35.3	0.378	48	0
1	196	76	36	249	36.5	0.875	29	1
5	189	64	33	325	31.2	0.583	29	1
5	158	70	0	0	29.8	0.207	63	0
5	103	108	37	0	39.2	0.305	65	0
4	146	78	0	0	38.5	0.520	67	1
4	147	74	25	293	34.9	0.385	30	0
5	99	54	28	83	34.0	0.499	30	0
6	124	72	0	0	27.6	0.368	29	1
0	101	64	17	0	21.0	0.252	21	0
3	81	86	16	66	27.5	0.306	22	0
1	133	102	28	140	32.8	0.234	45	1
3	173	82	48	465	38.4	2.137	25	1
0	118	64	23	89	0.0	1.731	21	0

0	84	64	22	66	35.8	0.545	21	0
2	105	58	40	94	34.9	0.225	25	0
2	122	52	43	158	36.2	0.816	28	0
12	140	82	43	325	39.2	0.528	58	1
0	98	82	15	84	25.2	0.299	22	0
1	87	60	37	75	37.2	0.509	22	0
4	156	75	0	0	48.3	0.238	32	1
0	93	100	39	72	43.4	1.021	35	0
1	107	72	30	82	30.8	0.821	24	0
0	105	68	22	0	20.0	0.236	22	0
1	109	60	8	182	25.4	0.947	21	0
1	90	62	18	59	25.1	1.268	25	0
1	125	70	24	110	24.3	0.221	25	0
1	119	54	13	50	22.3	0.205	24	0
5	116	74	29	0	32.3	0.660	35	1
8	105	100	36	0	43.3	0.239	45	1
5	144	82	26	285	32.0	0.452	58	1
3	100	68	23	81	31.6	0.949	28	0
1	100	66	29	196	32.0	0.444	42	0
5	166	76	0	0	45.7	0.340	27	1
1	131	64	14	415	23.7	0.389	21	0
4	116	72	12	87	22.1	0.463	37	0
4	158	78	0	0	32.9	0.803	31	1
2	127	58	24	275	27.7	1.600	25	0
3	96	56	34	115	24.7	0.944	39	0
0	131	66	40	0	34.3	0.196	22	1
3	82	70	0	0	21.1	0.389	25	0
3	193	70	31	0	34.9	0.241	25	1
4	95	64	0	0	32.0	0.161	31	1
6	137	61	0	0	24.2	0.151	55	0
5	136	84	41	88	35.0	0.286	35	1
9	72	78	25	0	31.6	0.280	38	0
5	168	64	0	0	32.9	0.135	41	1
2	123	48	32	165	42.1	0.520	26	0
4	115	72	0	0	28.9	0.376	46	1
0	101	62	0	0	21.9	0.336	25	0
8	197	74	0	0	25.9	1.191	39	1
1	172	68	49	579	42.4	0.702	28	1
6	102	90	39	0	35.7	0.674	28	0
1	112	72	30	176	34.4	0.528	25	0
1	143	84	23	310	42.4	1.076	22	0
1	143	74	22	61	26.2	0.256	21	0

0	138	60	35	167	34.6	0.534	21	1
3	173	84	33	474	35.7	0.258	22	1
1	97	68	21	0	27.2	1.095	22	0
4	144	82	32	0	38.5	0.554	37	1
1	83	68	0	0	18.2	0.624	27	0
3	129	64	29	115	26.4	0.219	28	1
1	119	88	41	170	45.3	0.507	26	0
2	94	68	18	76	26.0	0.561	21	0
0	102	64	46	78	40.6	0.496	21	0
2	115	64	22	0	30.8	0.421	21	0
8	151	78	32	210	42.9	0.516	36	1
4	184	78	39	277	37.0	0.264	31	1
0	94	0	0	0	0.0	0.256	25	0
1	181	64	30	180	34.1	0.328	38	1
0	135	94	46	145	40.6	0.284	26	0
1	95	82	25	180	35.0	0.233	43	1
2	99	0	0	0	22.2	0.108	23	0
3	89	74	16	85	30.4	0.551	38	0
1	80	74	11	60	30.0	0.527	22	0
2	139	75	0	0	25.6	0.167	29	0
1	90	68	8	0	24.5	1.138	36	0
0	141	0	0	0	42.4	0.205	29	1
12	140	85	33	0	37.4	0.244	41	0
5	147	75	0	0	29.9	0.434	28	0
1	97	70	15	0	18.2	0.147	21	0
6	107	88	0	0	36.8	0.727	31	0
0	189	104	25	0	34.3	0.435	41	1
2	83	66	23	50	32.2	0.497	22	0
4	117	64	27	120	33.2	0.230	24	0
8	108	70	0	0	30.5	0.955	33	1
4	117	62	12	0	29.7	0.380	30	1
0	180	78	63	14	59.4	2.420	25	1
1	100	72	12	70	25.3	0.658	28	0
0	95	80	45	92	36.5	0.330	26	0
0	104	64	37	64	33.6	0.510	22	1
0	120	74	18	63	30.5	0.285	26	0
1	82	64	13	95	21.2	0.415	23	0
2	134	70	0	0	28.9	0.542	23	1
0	91	68	32	210	39.9	0.381	25	0
2	119	0	0	0	19.6	0.832	72	0
2	100	54	28	105	37.8	0.498	24	0
14	175	62	30	0	33.6	0.212	38	1

1	135	54	0	0	26.7	0.687	62	0
5	86	68	28	71	30.2	0.364	24	0
10	148	84	48	237	37.6	1.001	51	1
9	134	74	33	60	25.9	0.460	81	0
9	120	72	22	56	20.8	0.733	48	0
1	71	62	0	0	21.8	0.416	26	0
8	74	70	40	49	35.3	0.705	39	0
5	88	78	30	0	27.6	0.258	37	0
10	115	98	0	0	24.0	1.022	34	0
0	124	56	13	105	21.8	0.452	21	0
0	74	52	10	36	27.8	0.269	22	0
0	97	64	36	100	36.8	0.600	25	0
8	120	0	0	0	30.0	0.183	38	1
6	154	78	41	140	46.1	0.571	27	0
1	144	82	40	0	41.3	0.607	28	0
0	137	70	38	0	33.2	0.170	22	0
0	119	66	27	0	38.8	0.259	22	0
7	136	90	0	0	29.9	0.210	50	0
4	114	64	0	0	28.9	0.126	24	0
0	137	84	27	0	27.3	0.231	59	0
2	105	80	45	191	33.7	0.711	29	1
7	114	76	17	110	23.8	0.466	31	0
8	126	74	38	75	25.9	0.162	39	0
4	132	86	31	0	28.0	0.419	63	0
3	158	70	30	328	35.5	0.344	35	1
0	123	88	37	0	35.2	0.197	29	0
4	85	58	22	49	27.8	0.306	28	0
0	84	82	31	125	38.2	0.233	23	0
0	145	0	0	0	44.2	0.630	31	1
0	135	68	42	250	42.3	0.365	24	1
1	139	62	41	480	40.7	0.536	21	0
0	173	78	32	265	46.5	1.159	58	0
4	99	72	17	0	25.6	0.294	28	0
8	194	80	0	0	26.1	0.551	67	0
2	83	65	28	66	36.8	0.629	24	0
2	89	90	30	0	33.5	0.292	42	0
4	99	68	38	0	32.8	0.145	33	0
4	125	70	18	122	28.9	1.144	45	1
3	80	0	0	0	0.0	0.174	22	0
6	166	74	0	0	26.6	0.304	66	0
5	110	68	0	0	26.0	0.292	30	0
2	81	72	15	76	30.1	0.547	25	0

7	195	70	33	145	25.1	0.163	55	1
6	154	74	32	193	29.3	0.839	39	0
2	117	90	19	71	25.2	0.313	21	0
3	84	72	32	0	37.2	0.267	28	0
6	0	68	41	0	39.0	0.727	41	1
7	94	64	25	79	33.3	0.738	41	0
3	96	78	39	0	37.3	0.238	40	0
10	75	82	0	0	33.3	0.263	38	0
0	180	90	26	90	36.5	0.314	35	1
1	130	60	23	170	28.6	0.692	21	0
2	84	50	23	76	30.4	0.968	21	0
8	120	78	0	0	25.0	0.409	64	0
12	84	72	31	0	29.7	0.297	46	1
0	139	62	17	210	22.1	0.207	21	0
9	91	68	0	0	24.2	0.200	58	0
2	91	62	0	0	27.3	0.525	22	0
3	99	54	19	86	25.6	0.154	24	0
3	163	70	18	105	31.6	0.268	28	1
9	145	88	34	165	30.3	0.771	53	1
7	125	86	0	0	37.6	0.304	51	0
13	76	60	0	0	32.8	0.180	41	0
6	129	90	7	326	19.6	0.582	60	0
2	68	70	32	66	25.0	0.187	25	0
3	124	80	33	130	33.2	0.305	26	0
6	114	0	0	0	0.0	0.189	26	0
9	130	70	0	0	34.2	0.652	45	1
3	125	58	0	0	31.6	0.151	24	0
3	87	60	18	0	21.8	0.444	21	0
1	97	64	19	82	18.2	0.299	21	0
3	116	74	15	105	26.3	0.107	24	0
0	117	66	31	188	30.8	0.493	22	0
0	111	65	0	0	24.6	0.660	31	0
2	122	60	18	106	29.8	0.717	22	0
0	107	76	0	0	45.3	0.686	24	0
1	86	66	52	65	41.3	0.917	29	0
6	91	0	0	0	29.8	0.501	31	0
1	77	56	30	56	33.3	1.251	24	0
4	132	0	0	0	32.9	0.302	23	1
0	105	90	0	0	29.6	0.197	46	0
0	57	60	0	0	21.7	0.735	67	0
0	127	80	37	210	36.3	0.804	23	0
3	129	92	49	155	36.4	0.968	32	1

8	100	74	40	215	39.4	0.661	43	1
3	128	72	25	190	32.4	0.549	27	1
10	90	85	32	0	34.9	0.825	56	1
4	84	90	23	56	39.5	0.159	25	0
1	88	78	29	76	32.0	0.365	29	0
8	186	90	35	225	34.5	0.423	37	1
5	187	76	27	207	43.6	1.034	53	1
4	131	68	21	166	33.1	0.160	28	0
1	164	82	43	67	32.8	0.341	50	0
4	189	110	31	0	28.5	0.680	37	0
1	116	70	28	0	27.4	0.204	21	0
3	84	68	30	106	31.9	0.591	25	0
6	114	88	0	0	27.8	0.247	66	0
1	88	62	24	44	29.9	0.422	23	0
1	84	64	23	115	36.9	0.471	28	0
7	124	70	33	215	25.5	0.161	37	0
1	97	70	40	0	38.1	0.218	30	0
8	110	76	0	0	27.8	0.237	58	0
11	103	68	40	0	46.2	0.126	42	0
11	85	74	0	0	30.1	0.300	35	0
6	125	76	0	0	33.8	0.121	54	1
0	198	66	32	274	41.3	0.502	28	1
1	87	68	34	77	37.6	0.401	24	0
6	99	60	19	54	26.9	0.497	32	0
0	91	80	0	0	32.4	0.601	27	0
2	95	54	14	88	26.1	0.748	22	0
1	99	72	30	18	38.6	0.412	21	0
6	92	62	32	126	32.0	0.085	46	0
4	154	72	29	126	31.3	0.338	37	0
0	121	66	30	165	34.3	0.203	33	1
3	78	70	0	0	32.5	0.270	39	0
2	130	96	0	0	22.6	0.268	21	0
3	111	58	31	44	29.5	0.430	22	0
2	98	60	17	120	34.7	0.198	22	0
1	143	86	30	330	30.1	0.892	23	0
1	119	44	47	63	35.5	0.280	25	0
6	108	44	20	130	24.0	0.813	35	0
2	118	80	0	0	42.9	0.693	21	1
10	133	68	0	0	27.0	0.245	36	0
2	197	70	99	0	34.7	0.575	62	1
0	151	90	46	0	42.1	0.371	21	1
6	109	60	27	0	25.0	0.206	27	0

12	121	78	17	0	26.5	0.259	62	0
8	100	76	0	0	38.7	0.190	42	0
8	124	76	24	600	28.7	0.687	52	1
1	93	56	11	0	22.5	0.417	22	0
8	143	66	0	0	34.9	0.129	41	1
6	103	66	0	0	24.3	0.249	29	0
3	176	86	27	156	33.3	1.154	52	1
0	73	0	0	0	21.1	0.342	25	0
11	111	84	40	0	46.8	0.925	45	1
2	112	78	50	140	39.4	0.175	24	0
3	132	80	0	0	34.4	0.402	44	1
2	82	52	22	115	28.5	1.699	25	0
6	123	72	45	230	33.6	0.733	34	0
0	188	82	14	185	32.0	0.682	22	1
0	67	76	0	0	45.3	0.194	46	0
1	89	24	19	25	27.8	0.559	21	0
1	173	74	0	0	36.8	0.088	38	1
1	109	38	18	120	23.1	0.407	26	0
1	108	88	19	0	27.1	0.400	24	0
6	96	0	0	0	23.7	0.190	28	0
1	124	74	36	0	27.8	0.100	30	0
7	150	78	29	126	35.2	0.692	54	1
4	183	0	0	0	28.4	0.212	36	1
1	124	60	32	0	35.8	0.514	21	0
1	181	78	42	293	40.0	1.258	22	1
1	92	62	25	41	19.5	0.482	25	0
0	152	82	39	272	41.5	0.270	27	0
1	111	62	13	182	24.0	0.138	23	0
3	106	54	21	158	30.9	0.292	24	0
3	174	58	22	194	32.9	0.593	36	1
7	168	88	42	321	38.2	0.787	40	1
6	105	80	28	0	32.5	0.878	26	0
11	138	74	26	144	36.1	0.557	50	1
3	106	72	0	0	25.8	0.207	27	0
6	117	96	0	0	28.7	0.157	30	0
2	68	62	13	15	20.1	0.257	23	0
9	112	82	24	0	28.2	1.282	50	1
0	119	0	0	0	32.4	0.141	24	1
2	112	86	42	160	38.4	0.246	28	0
2	92	76	20	0	24.2	1.698	28	0
6	183	94	0	0	40.8	1.461	45	0
0	94	70	27	115	43.5	0.347	21	0

2	108	64	0	0	30.8	0.158	21	0
4	90	88	47	54	37.7	0.362	29	0
0	125	68	0	0	24.7	0.206	21	0
0	132	78	0	0	32.4	0.393	21	0
5	128	80	0	0	34.6	0.144	45	0
4	94	65	22	0	24.7	0.148	21	0
7	114	64	0	0	27.4	0.732	34	1
0	102	78	40	90	34.5	0.238	24	0
2	111	60	0	0	26.2	0.343	23	0
1	128	82	17	183	27.5	0.115	22	0
10	92	62	0	0	25.9	0.167	31	0
13	104	72	0	0	31.2	0.465	38	1
5	104	74	0	0	28.8	0.153	48	0
2	94	76	18	66	31.6	0.649	23	0
7	97	76	32	91	40.9	0.871	32	1
1	100	74	12	46	19.5	0.149	28	0
0	102	86	17	105	29.3	0.695	27	0
4	128	70	0	0	34.3	0.303	24	0
6	147	80	0	0	29.5	0.178	50	1
4	90	0	0	0	28.0	0.610	31	0
3	103	72	30	152	27.6	0.730	27	0
2	157	74	35	440	39.4	0.134	30	0
1	167	74	17	144	23.4	0.447	33	1
0	179	50	36	159	37.8	0.455	22	1
11	136	84	35	130	28.3	0.260	42	1
0	107	60	25	0	26.4	0.133	23	0
1	91	54	25	100	25.2	0.234	23	0
1	117	60	23	106	33.8	0.466	27	0
5	123	74	40	77	34.1	0.269	28	0
2	120	54	0	0	26.8	0.455	27	0
1	106	70	28	135	34.2	0.142	22	0
2	155	52	27	540	38.7	0.240	25	1
2	101	58	35	90	21.8	0.155	22	0
1	120	80	48	200	38.9	1.162	41	0
11	127	106	0	0	39.0	0.190	51	0
3	80	82	31	70	34.2	1.292	27	1
10	162	84	0	0	27.7	0.182	54	0
1	199	76	43	0	42.9	1.394	22	1
8	167	106	46	231	37.6	0.165	43	1
9	145	80	46	130	37.9	0.637	40	1
6	115	60	39	0	33.7	0.245	40	1
1	112	80	45	132	34.8	0.217	24	0

4	145	82	18	0	32.5	0.235	70	1
10	111	70	27	0	27.5	0.141	40	1
6	98	58	33	190	34.0	0.430	43	0
9	154	78	30	100	30.9	0.164	45	0
6	165	68	26	168	33.6	0.631	49	0
1	99	58	10	0	25.4	0.551	21	0
10	68	106	23	49	35.5	0.285	47	0
3	123	100	35	240	57.3	0.880	22	0
8	91	82	0	0	35.6	0.587	68	0
6	195	70	0	0	30.9	0.328	31	1
9	156	86	0	0	24.8	0.230	53	1
0	93	60	0	0	35.3	0.263	25	0
3	121	52	0	0	36.0	0.127	25	1
2	101	58	17	265	24.2	0.614	23	0
2	56	56	28	45	24.2	0.332	22	0
0	162	76	36	0	49.6	0.364	26	1
0	95	64	39	105	44.6	0.366	22	0
4	125	80	0	0	32.3	0.536	27	1
5	136	82	0	0	0.0	0.640	69	0
2	129	74	26	205	33.2	0.591	25	0
3	130	64	0	0	23.1	0.314	22	0
1	107	50	19	0	28.3	0.181	29	0
1	140	74	26	180	24.1	0.828	23	0
1	144	82	46	180	46.1	0.335	46	1
8	107	80	0	0	24.6	0.856	34	0
13	158	114	0	0	42.3	0.257	44	1
2	121	70	32	95	39.1	0.886	23	0
7	129	68	49	125	38.5	0.439	43	1
2	90	60	0	0	23.5	0.191	25	0
7	142	90	24	480	30.4	0.128	43	1
3	169	74	19	125	29.9	0.268	31	1
0	99	0	0	0	25.0	0.253	22	0
4	127	88	11	155	34.5	0.598	28	0
4	118	70	0	0	44.5	0.904	26	0
2	122	76	27	200	35.9	0.483	26	0
6	125	78	31	0	27.6	0.565	49	1
1	168	88	29	0	35.0	0.905	52	1
2	129	0	0	0	38.5	0.304	41	0
4	110	76	20	100	28.4	0.118	27	0
6	80	80	36	0	39.8	0.177	28	0
10	115	0	0	0	0.0	0.261	30	1
2	127	46	21	335	34.4	0.176	22	0

9	164	78	0	0	32.8	0.148	45	1
2	93	64	32	160	38.0	0.674	23	1
3	158	64	13	387	31.2	0.295	24	0
5	126	78	27	22	29.6	0.439	40	0
10	129	62	36	0	41.2	0.441	38	1
0	134	58	20	291	26.4	0.352	21	0
3	102	74	0	0	29.5	0.121	32	0
7	187	50	33	392	33.9	0.826	34	1
3	173	78	39	185	33.8	0.970	31	1
10	94	72	18	0	23.1	0.595	56	0
1	108	60	46	178	35.5	0.415	24	0
5	97	76	27	0	35.6	0.378	52	1
4	83	86	19	0	29.3	0.317	34	0
1	114	66	36	200	38.1	0.289	21	0
1	149	68	29	127	29.3	0.349	42	1
5	117	86	30	105	39.1	0.251	42	0
1	111	94	0	0	32.8	0.265	45	0
4	112	78	40	0	39.4	0.236	38	0
1	116	78	29	180	36.1	0.496	25	0
0	141	84	26	0	32.4	0.433	22	0
2	175	88	0	0	22.9	0.326	22	0
2	92	52	0	0	30.1	0.141	22	0
3	130	78	23	79	28.4	0.323	34	1
8	120	86	0	0	28.4	0.259	22	1
2	174	88	37	120	44.5	0.646	24	1
2	106	56	27	165	29.0	0.426	22	0
2	105	75	0	0	23.3	0.560	53	0
4	95	60	32	0	35.4	0.284	28	0
0	126	86	27	120	27.4	0.515	21	0
8	65	72	23	0	32.0	0.600	42	0
2	99	60	17	160	36.6	0.453	21	0
1	102	74	0	0	39.5	0.293	42	1
11	120	80	37	150	42.3	0.785	48	1
3	102	44	20	94	30.8	0.400	26	0
1	109	58	18	116	28.5	0.219	22	0
9	140	94	0	0	32.7	0.734	45	1
13	153	88	37	140	40.6	1.174	39	0
12	100	84	33	105	30.0	0.488	46	0
1	147	94	41	0	49.3	0.358	27	1
1	81	74	41	57	46.3	1.096	32	0
3	187	70	22	200	36.4	0.408	36	1
6	162	62	0	0	24.3	0.178	50	1

4	136	70	0	0	31.2	1.182	22	1
1	121	78	39	74	39.0	0.261	28	0
3	108	62	24	0	26.0	0.223	25	0
0	181	88	44	510	43.3	0.222	26	1
8	154	78	32	0	32.4	0.443	45	1
1	128	88	39	110	36.5	1.057	37	1
7	137	90	41	0	32.0	0.391	39	0
0	123	72	0	0	36.3	0.258	52	1
1	106	76	0	0	37.5	0.197	26	0
6	190	92	0	0	35.5	0.278	66	1
2	88	58	26	16	28.4	0.766	22	0
9	170	74	31	0	44.0	0.403	43	1
9	89	62	0	0	22.5	0.142	33	0
10	101	76	48	180	32.9	0.171	63	0
2	122	70	27	0	36.8	0.340	27	0
5	121	72	23	112	26.2	0.245	30	0
1	126	60	0	0	30.1	0.349	47	1
1	93	70	31	0	30.4	0.315	23	0