



**PENGEMBANGAN APLIKASI DATA MINING
MENGGUNAKAN METODE INDUKSI BERORIENTASI ATRIBUT
(STUDI KASUS : DATA PPMB IPB)**

Hak Cipta dilindungi Undang-Undang.
Dilarang menyalin, memperdengarkan atau mengadaptasi tanpa izin.
a. Penggunaan hanya untuk keperluan penelitian akademik, penulis dan pengaruhnya.
b. Penggunaan tidak menyalin bagian besar tanpa izin.
c. Dilarang menggunakan dalam persyaratan teknis dan wajib kewajiban dalam bentuk sertifikat dan tesis di IPB University.

Oleh :

**ERDANI FEBRI RIYANTI
G64101034**



**DEPARTEMEN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
2005**



**PENGEMBANGAN APLIKASI DATA MINING
MENGGUNAKAN METODE INDUKSI BERORIENTASI ATRIBUT
(STUDI KASUS : DATA PPMB IPB)**

Skripsi

**Sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer
pada Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Pertanian Bogor**

Oleh :

**ERDANI FEBRI RIYANTI
G64101034**



**PROGRAM STUDI ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
2005**



ABSTRAK

ERDANI FEBRI RIYANTI. Karakterisasi dan Generalisasi Data PPMB IPB Menggunakan Induksi Berorientasi Atribut. Dibimbing oleh Imas S. Sitanggang dan Julio Adisantoso.

Metode induksi berorientasi merupakan salah satu metode dalam *data mining* untuk menggali informasi yang penting berupa karakterisasi kelas serta perbandingan antarkelas tertentu. Penelitian ini bertujuan untuk menerapkan metode induksi berorientasi tersebut pada data PPMB IPB dari tahun 2000 sampai tahun 2004 serta membuat sebuah aplikasi *data mining* dengan metode tersebut untuk mempermudah proses penggalian data.

Ide umum dari pendekatan induksi berorientasi atribut (AOI) adalah mengoleksi *task relevant data* kemudian melakukan generalisasi berdasarkan pengujian dari sejumlah nilai yang berbeda dari setiap atribut dalam kumpulan *relevant data*. Pengetahuan yang diperoleh direpresentasikan dalam bentuk aturan, tabulasi silang dan grafik.

Aplikasi yang dikembangkan dalam penelitian merujuk pada setiap tahapan yang diperlukan dalam *data mining* termasuk tahapan *data preprocessing*. Setiap langkah dalam *data preprocessing* dan *data mining* dianalisa untuk kemudian diimplementasikan menjadi sebuah aplikasi. Dengan menggunakan aplikasi tersebut, untuk karakterisasi kelas target fakultas pilihan pertama pelamar, didapatkan salah satu hasil bahwa separuh lebih pelamar tidak lolos seleksi menjadi mahasiswa IPB dan hanya 15-40% pelamar yang berhasil lolos menjadi mahasiswa IPB pada fakultas pilihan pertama yang diinginkan serta sisanya lolos seleksi bukan pada fakultas pilihan pertama. Hasil perbandingan antarkelas pulau asal pelamar dengan kelas target adalah Pulau Sumatera menunjukkan bahwa kemungkinan pelamar berasal dari Pulau Jawa lebih besar dibandingkan dengan kelas target Pulau Sumatera atau pulau lainnya. Perbandingan tersebut berdasarkan kondisi pelamar diterima pada pilihan pertamanya.

Kata kunci: *data mining*, induksi berorientasi atribut, generalisasi, karakterisasi.



Judul : Pengembangan Aplikasi *Data Mining* Menggunakan Metode Induksi Berorientasi Atribut (Studi Kasus : Data PPMB IPB)
Nama : Erdani Febri Riyanti
NRP : G 64101034

Menyetujui,

Pembimbing I

Pembimbing II

Imas S. Sitanggang, S.Si, M.Komp
NIP. 132 206 235

Ir. Julio Adisantoso, M.Komp
NIP. 131 578 807

Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Dr. Ir. Yonny Koesmaryono, MS
NIP. 131 473 999

Tanggal Lulus :



RIWAYAT HIDUP

Penulis lahir di Jakarta pada tanggal 17 Februari 1984, anak dari pasangan Maroni dan Ida Wakhidah. Penulis merupakan anak pertama dari dua bersaudara.

Tahun 2001 penulis lulus dari SMU Negeri 12 Jakarta dan pada tahun yang sama melanjutkan kuliah ke Institut Pertanian Bogor (IPB), Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam melalui jalur Undangan Seleksi Masuk IPB (USMI).

Selama mengikuti perkuliahan, penulis pernah menjadi asisten praktikum mata kuliah Rekayasa Perangkat Lunak mahasiswa D3 Informatika pada tahun ajaran 2004/2005 dan menjadi staf pengajar privat lembaga Nurul Ilmi pada tahun 2004 serta menjadi operator data *entry* Pemilu 2004. Di akhir perkuliahan, penulis berkesempatan mengikuti Praktik Lapang (PL) selama dua bulan (26 Januari-25 Maret 2005) di Departemen Pertanian, Jakarta.



PRAKATA

Puji syukur kepada Allah SWT karena atas limpahan rahmat dan karunia-Nya skripsi ini berhasil diselesaikan. Skripsi ini mengambil judul Pengembangan Aplikasi *Data Mining* Menggunakan Metode Induksi Berorientasi Atribut (Studi Kasus : Data PPMB IPB).

Penulis menyampaikan terima kasih kepada semua pihak yang telah membantu dalam penulisan karya ilmiah ini khususnya kepada Ibu Imas S Sitanggang, S.Si, M.Komp dan Bapak Ir. Julio Adisantoso M.Komp yang telah membimbing dengan penuh ketekunan dan kesabaran hingga selesainya penulisan karya ilmiah ini. Selanjutnya, penulis juga ingin mengucapkan terima kasih kepada:

- 1) Bapak Hari Agung S.Komp atas kesediaannya menjadi penguji sidang.
- 2) Bapak dan Ibu serta Helda Elisa di rumah yang senantiasa memahami kesibukan penulis serta kasih sayang, perhatian, dukungannya selama ini dan menjadi pendengar yang baik atas semua cerita dan keluh kesah selama penyusunan skripsi.
- 3) Seluruh keluarga besar Purworejo, Dian, Mba Sari, Mba Tri atas kesediaannya berbagi pengalaman, memacu semangat serta doa untuk penulis.
- 4) Hilal Ramadhan Fikri atas semangat serta keceriaannya di saat-saat tersulit bagi penulis.
- 5) Yowan Nor Herienda, S.Kom yang telah memberikan banyak waktu serta kesabarannya kepada penulis.
- 6) Anak kos Bafak 30, Nenk, Nien, Ketty, Mba Ida untuk kesediaannya mendengar rintihan hati dan keceriaan selama ini.
- 7) Teman-teman SMU, Netha, Indah, Anggit atas semangat dan doa yang tulus untuk penulis.
- 8) Rekan-rekan lab *data mining* Inu Wisnujati, Eri Herlambang, Yulia, Samuel, Hilmi atas kesediaannya berbagi ilmu kepada penulis.
- 9) Teman-teman seperjuangan Ilkom 38 untuk pengalaman dan kenangan yang tak ternilai.

Dan semua pihak yang tidak dapat disebutkan satu per satu. Semoga karya ilmiah ini dapat bermanfaat bagi semua pihak. Amin.



DAFTAR ISI

	Halaman
DAFTAR TABEL	ix
DAFTAR GAMBAR	ix
DAFTAR LAMPIRAN	x
PENDAHULUAN	
Latar Belakang	1
Tujuan.....	1
Ruang Lingkup.....	1
Manfaat Penelitian	1
TINJAUAN PUSTAKA	
<i>Knowledge Discovery in Database (KDD)</i>	2
<i>Data Preprocessing</i>	2
Integrasi Data	3
Pembersihan Data	3
Transformasi Data.....	3
<i>Data Mining</i>	4
<i>Data Mining Primitive</i>	5
<i>Data Mining Query Language (DMQL)</i>	5
Deskripsi Konsep	6
Analisis Relevansi Atribut	6
Induksi Berorientasi Atribut	7
METODE PENELITIAN	
Spesifikasi Implementasi	9
HASIL DAN PEMBAHASAN	
Gambaran Umum Aplikasi	9
Integrasi Data	9
Pembersihan Data	10
Transformasi Data.....	11
Penentuan <i>task-relevant data</i>	12
Penentuan Nilai-Nilai Tergeneralisasi Untuk Setiap Atribut Pada <i>Task-relevant Data</i>	13
Generalisasi dan Pembuangan Atribut	13
Analisis Relevansi Atribut	14
Presentasi Hasil	15
KESIMPULAN DAN SARAN	
Kesimpulan.....	16



Saran	17
DAFTAR PUSTAKA	17
LAMPIRAN	18

Hak Cipta dilindungi Undang-Undang.
Dilarang menyalin, memindai, atau mengadaptasi sebagian atau seluruh isi buku ini tanpa izin tertulis dari penerbit dan pemegang hak cipta.
a. Pengeditan buku untuk keperluan penilaian akademik, penulisan karya ilmiah, penyelesaian tugas, penulisan laporan hasil riset atau tesis di dalam institusi pendidikan
b. Pengambilan tidak menyalin bagian besar isi buku IPB University
3. Dilarang menggunakan buku ini untuk perjudian, seksual, dan politik.



DAFTAR TABEL

	Halaman
1 Ringkasan hasil analisis relevansi atribut.....	14

DAFTAR GAMBAR

	Halaman
1 Tahapan dalam KDD	2
2 <i>Form</i> integrasi tabel tanpa relasi.....	10
3 <i>Form</i> integrasi tabel dengan relasi.....	10
4 <i>Form</i> pendeksiyan data <i>null</i> dan yang mengandung <i>noise</i>	11
5 <i>Form</i> generalisasi atribut	14
6 <i>Form</i> analisis relevansi atribut.....	14



PENDAHULUAN

Latar Balakang

Masalah penumpukan data menjadi salah satu konsekuensi dari peningkatan kebutuhan manusia yang diiringi dengan perkembangan teknologi komputer yang sangat cepat. Perkembangan kemampuan media penyimpanan data serta bermunculannya teknologi perangkat lunak terbaru yang mampu mengelola data berukuran besar membuat penyimpanan beribu-ribu *record* data di dalam komputer tidak lagi menjadi sebuah masalah. Namun sering kali data yang tersimpan tersebut di kemudian hari belum dimanfaatkan secara maksimal untuk menggali pengetahuan untuk membuat suatu keputusan yang berharga karena keterbatasan kemampuan pengguna komputer untuk mengolah tumpukan data tersebut menjadi sebuah pengetahuan yang berharga.

Panitia Penerimaan Mahasiswa Baru Institut Pertanian Bogor (PPMB IPB) mengumpulkan data pelamar program sarjana setiap tahunnya meliputi data akademik pelamar, data penilaian terhadap sekolah pelamar, serta data pribadi pelamar untuk kepentingan penyeleksian penerimaan mahasiswa IPB khususnya melalui jalur USMI (Undangan Seleksi Masuk IPB). Setelah digunakan untuk kebutuhan penyeleksian calon mahasiswa, data tersebut disimpan dengan harapan dapat berguna di masa yang akan datang.

Data mining merupakan proses ekstraksi informasi atau pola yang penting dalam basis data berukuran besar (Han & Kamber 2001). Penerapan *data mining* dalam data (khususnya pelamar jalur USMI) PPMB IPB diharapkan akan mampu menggali informasi yang penting dan berharga dari tumpukan data pelamar USMI IPB tersebut sehingga dapat digunakan untuk mengambil keputusan yang berkaitan dengan seleksi penerimaan mahasiswa baru berikutnya maupun keputusan akademik lainnya. Metode *data mining* yang akan digunakan dalam penelitian ini yaitu metode induksi berorientasi atribut, diharapkan pada akhir penelitian didapatkan hasil pengetahuan berupa karakterisasi kelas serta perbandingan antara kelas target serta kelas pembanding tertentu.

Tujuan Penelitian

Tujuan dari penelitian ini adalah :

1. Untuk menerapkan metode induksi berorientasi atribut pada data pelamar USMI PPMB IPB.
2. Untuk mengembangkan aplikasi *data mining* dengan metode induksi berorientasi atribut yang dapat menghasilkan karakterisasi kelas tertentu dan membandingkan kelas-kelas tersebut berdasarkan kriteria tertentu.

Ruang Lingkup Penelitian

Ruang lingkup pada penelitian ini mencakup penerapan salah satu fungsionalitas dari *data mining* yaitu deskripsi konsep dengan metode induksi berorientasi atribut pada data akademik serta data penilaian terhadap sekolah pelamar melalui jalur USMI dari tahun 2000 sampai dengan tahun 2004.

Penelitian ini akan menghasilkan sebuah aplikasi *data mining* dengan menggunakan metode berorientasi atribut. Aplikasi menyediakan setiap langkah-langkah yang diperlukan dalam proses KDD, meliputi tahapan integrasi data, pembersihan data, pemilihan atribut, dan proses penggalian informasi penting, serta media presentasi (berupa relasi umum, tabel, dan visualisasi grafik) hasil penerapan yang akan menggambarkan karakteristik-karakteristik data dari kelas tertentu serta perbandingan antara kelas target dengan kelas pembanding tertentu. Aplikasi yang dibangun banyak menyerahkan keputusan pengolahan data kepada pengguna.

Manfaat Penelitian

Aplikasi yang dihasilkan pada penelitian ini diharapkan dapat digunakan untuk menggali informasi berupa karakterisasi kelas serta perbandingan antara kelas target dengan kelas pembanding tertentu yang diinginkan pengguna pada basis data PPMB IPB untuk tahun-tahun berikutnya.



TINJAUAN PUSTAKA

Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) adalah proses menemukan informasi yang berguna serta pola-pola yang ada dalam data (Goharian & Grossman 2003).

KDD dapat digunakan sebagai alat yang sangat berguna untuk menganalisis dan menginterpretasikan tumpukan data yang tersimpan dalam *file*, basis data maupun tempat penyimpanan lainnya, bahkan lebih lanjut KDD dapat digunakan untuk mengekstrak pengetahuan-pengetahuan yang menarik untuk membantu proses pengambilan keputusan (Zafane 1999).

Proses dalam KDD terdiri dari beberapa langkah, meliputi persiapan data, pencarian untuk pola-pola tertentu, evaluasi pengetahuan, dan beberapa perbaikan yang mencakup beberapa iterasi setelah modifikasi (Fayyad *et al.* 1996). Proses utama KDD adalah proses penggunaan metode *data mining* untuk mengekstrak (mengidentifikasi) pengetahuan apa yang diinginkan yang bergantung pada spesifikasi dari ukuran dan *threshold*, dengan menggunakan basis data *F* dengan *preprocessing* yang diperlukan, pengambilan contoh, dan transformasi dari *F* (Fayyad *et al.* 1996).

KDD merupakan sebuah proses yang terdiri dari serangkaian proses iteratif yang terurut, dan *data mining* merupakan salah satu langkah dalam proses KDD (Han & Kamber 2001). Pada Gambar 1 dapat dilihat tahapan proses KDD secara berurut.

Tahapan proses KDD menurut Han & Kamber (2001), yaitu :

1. Pembersihan data

Pembersihan terhadap data dilakukan untuk menghilangkan data yang tidak konsisten dan data yang mengandung *noise*.

2. Integrasi data

Proses integrasi data dilakukan untuk menggabungkan data dari berbagai sumber.

3. Seleksi data

Proses seleksi data mengambil data yang relevan digunakan untuk proses analisis.

4. Transformasi data

Proses ini mentransformasi atau menggabungkan data ke dalam bentuk yang tepat untuk di-*mining* dengan cara melakukan peringkasan atau operasi agregasi. Dalam beberapa kasus proses transformasi dilakukan sebelum proses seleksi terutama dalam kasus pembangunan sebuah *data warehouse*.

5. Data mining

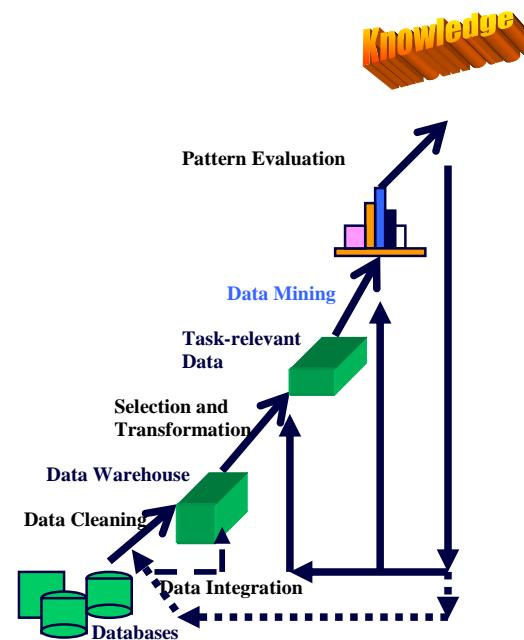
Data mining merupakan proses yang penting dimana metode-metode cerdas diaplikasikan untuk mengekstrak pola-pola dalam data.

6. Evaluasi pola

Evaluasi pola diperlukan untuk mengidentifikasi beberapa pola-pola yang menarik yang merepresentasikan pengetahuan.

7. Presentasi pengetahuan

Penggunaan visualisasi dan teknik representasi untuk menunjukkan pengetahuan hasil penggalian gunung data kepada pengguna.



Gambar 1 Tahapan dalam KDD
(Han & Kamber 2001)

Data Preprocessing

Dalam dunia nyata seringkali ditemui basis data yang kotor serta basis data yang memiliki banyak nilai-nilai yang kosong dan tidak konsisten. Untuk mendapatkan pengetahuan yang berguna dan valid dari kumpulan data



tersebut, maka sebelum memasuki tahapan ‘penggalian’ informasi, data tersebut harus melewati tahapan *preprocessing*. *Data preprocessing* meliputi tahapan pembersihan data, integrasi data, transformasi data, reduksi data, sehingga hasil akhir pengetahuan yang didapatkan memiliki kualitas yang tinggi (Han & Kamber 2001).

Pembersihan data

Pembersihan data dilakukan untuk mengisi nilai-nilai data yang hilang, mengurangi data yang kotor ketika *outlier* teridentifikasi, dan memperbaiki ketidakkonsistenan dalam data (Han & Kamber 2001).

Secara keseluruhan, permasalahan data kotor dapat diringkas menjadi tiga serta terdapat cara-cara tertentu untuk membersihkan data yang kotor tersebut (Han & Kamber 2001), yaitu :

1. Nilai yang hilang (kosong)

Untuk mengatasi basis data dengan nilai yang kosong dapat dilakukan metode penghapusan *tuple* yang memiliki nilai kosong (*null*), pengisian nilai baru secara manual, pengisian nilai konstanta umum seperti ‘tidak diketahui’ ataupun ‘∞’ (untuk metode ini sangat mudah, akan tetapi banyak informasi yang hilang), pengisian dengan nilai rata-rata dari atribut data yang kosong, pengisian dengan nilai rata-rata berdasarkan karakteristik tertentu yang sama dengan data yang kosong, serta pengisian dengan nilai yang mungkin untuk nilai-nilai yang kosong dengan metode regresi, induksi dengan pohon keputusan dan lain sebagainya.

2. Nilai yang mengandung *noise*

Nilai yang mengandung *noise* dapat diganti dengan nilai baru hasil perhitungan dengan metode *binning* (pengisian dengan nilai sekitar). Metode *binning* terdapat tiga jenis yaitu *bining* dengan nilai rataan, *bining* dengan nilai median, serta *bining* dengan nilai batasan. Pendekripsi nilai yang kotor dapat dilakukan dengan metode pengklusteran, dan pengecekan oleh manusia dan komputer, serta dengan regresi.

3. Ketidakkonsistenan data

Ketidakkonsistenan data dapat diperbaiki secara manual dengan bantuan referensi eksternal.

Integrasi Data

Untuk data yang melibatkan proses integrasi, yang menggabungkan data dari berbagai sumber ke dalam satu penyimpanan data seperti *data warehouse* terdapat beberapa hal yang harus diperhatikan seperti skema integrasi, bagaimana entitas dari dunia nyata dan dari berbagai sumber dapat disatukan. Masalah ini dikenal sebagai masalah identifikasi entitas. Sebagai contoh adalah masalah penamaan yang berbeda untuk entitas yang sama. Masalah lain yang muncul akibat pengintegrasian data yaitu *redundancy*. Ketidakkonsistensan dalam penamaan atribut juga menyebabkan terjadinya *redundancy*. Masalah *redundancy* juga terjadi pada tingkat *tuple*, dimana terdapat dua atau lebih *tuple* yang sama untuk pengisian data yang unik. Masalah lain yang muncul akibat pengintegrasian data yaitu konflik nilai data, misalnya karena perbedaan representasi sebuah nilai dalam basis data tertentu. Sebagai contoh untuk pemberian satuan berat untuk data tertentu ada yang menggunakan kilogram sedangkan yang lain menggunakan ton.

Transformasi Data

Pada tahap transformasi data, data ditransformasikan ke dalam bentuk yang tepat untuk di-mining. Menurut Han & Kamber (2001), transformasi data meliputi hal :

1. *Smoothing*

Bekerja untuk menghilangkan data yang mengandung *noise*, metode yang digunakan dapat berupa *binning*, metode *cluster*, dan regresi.

2. Agregasi

Perengkasan serta operasi agregasi diaplikasikan pada data. Sebagai contoh perhitungan penjualan per hari dapat diagregasi menjadi perhitungan penjualan per bulan.

3. Generalisasi

Generalisasi data menggantikan data yang memiliki konsep rendah dengan data yang memiliki konsep lebih tinggi melalui hirarki konsep.

4. Normalisasi

Nilai atribut diberi skala yang lebih kecil, misalnya data yang awalnya memiliki skala -1.0 sampai 1.0 diubah skalanya menjadi 0.0 sampai 1.0.



5. Pembentukan atribut

Atribut baru dibentuk dan ditambahkan ke dalam kumpulan atribut untuk membantu proses *data mining*.

Data Mining

Data mining merupakan penggunaan algoritma dalam proses KDD untuk menemukan pola yang bermanfaat (Goharian & Grossman 2003).

Beberapa teknik yang sering disebut dalam literatur *data mining* adalah deskripsi konsep, *clustering*, klasifikasi, prediksi dan *association rule*.

Untuk melakukan *data mining*, harus dikumpulkan data sebanyak mungkin dengan mentransfer data tersebut dari sekumpulan kertas ke dalam basis data yang telah ada. Disamping itu diperlukan juga restrukturisasi data yang beragam ini pada pusat informasi yang besar yaitu *data warehouse* (Connolly & Begg 2002).

Menurut Two Crows Corporation (1999), ada dua kunci sukses dalam *data mining*. Yang pertama yaitu ketepatan formulasi masalah yang akan dipecahkan. Yang kedua yaitu penggunaan data yang tepat. Setelah mendapatkan data yang akan digunakan untuk *di-mining*, ketepatan dalam mentransformasikan dan mengolahnya sangat mempengaruhi ketepatan pengetahuan hasil *data mining*.

Secara prinsip *data mining* tidak spesifik ke satu tipe media atau data. *Data mining* harus mampu diaplikasikan pada bermacam-macam tempat penyimpanan data. Tentunya algoritma dan pendekatan pasti berbeda untuk jenis data yang berbeda. Berikut tipe-tipe data yang dapat *di-mining* menurut Han & Kamber (2001) :

1. Basis data relasional

Sebuah sistem basis data terdiri dari sekumpulan koleksi dari data yang tidak berhubungan, dan sekumpulan perangkat lunak untuk mengolah serta mengakses data. Sebuah basis data relasional adalah kumpulan tabel yang diberi nama yang unik. Setiap tabel terdiri dari sekumpulan atribut (kolom atau *field*) dan menyimpan sekumpulan *tuple* (baris atau *row*). Setiap *tuple* di dalam tabel relasional merepresentasikan sebuah objek yang diidentifikasi oleh sebuah kunci yang unik, dan digambarkan oleh sekumpulan nilai

atribut. Sebuah model data seperti diagram ER (*Entity Relationship*) dibangun untuk menggambarkan hubungan antar entitas.

Data relasional dapat diakses dengan *query* basis data yang ditulis dalam bahasa *query* relasional.

2. Data warehouse

Sebuah *data warehouse* biasanya dimodelkan dengan struktur basis data multidimensional, dimana setiap dimensi berhubungan dengan sebuah atribut atau sekumpulan atribut di dalam skema, dan setiap sel menyimpan nilai dari perhitungan agregat.

3. Basis data transaksional

Secara umum, basis data transaksional terdiri dari sebuah file dimana setiap *tuple* merepresentasikan sebuah transaksi. Setiap transaksi memiliki nomor identifikasi transaksi yang unik dan serangkaian *item* yang terlibat dalam proses transaksi.

4. Sistem basis data lanjutan dan aplikasi basis data lanjutan

Untuk memenuhi kebutuhan-kebutuhan basis data yang baru seperti kebutuhan untuk menangani data spasial, data teknik (seperti bangunan, sirkuit terintegrasi, atau komponen-komponen sistem), data multimedia, data yang berelasi dengan waktu, dan WWW, muncul teknologi baru selain DMBS relasional, seperti sistem basis data berorientasi objek, sistem basis data objek relasional, sistem basis data spasial, sistem basis data untuk data temporal dan *time-series*, sistem basis data teks dan multimedia, dan lain sebagainya.

Proses *data mining* dapat dibedakan menjadi dua tujuan utama yaitu (Han & Kamber 2001):

1. Descriptive data mining

Deskripsi konsep atau *task-relevant data* dalam bentuk yang ringkas, informatif dan diskriminatif.

2. Predictive data mining

Berdasarkan analisis data dibuat model untuk kemudian dijadikan sebagai alat prediksi *trend* dan data yang tidak diketahui nilainya.

Menurut Berson *et al.* (2000), banyak organisasi menggunakan *data mining* dalam hal :

1. Menemukan pengetahuan, tujuan utama *data mining* untuk menentukan keterhubungan yang tersembunyi, pola-pola,



atau korelasi dari data yang disimpan dalam basis data perusahaan.

2. Visualisasi data, seorang analis harus mampu menarik kesimpulan dari sekumpulan data perusahaan.
3. Mengkoreksi data, dimana metode *data mining* mampu menemukan data yang tidak lengkap, mengandung kesalahan dan lain sebagainya.

Data Mining Primitive

Menurut Han & Kamber (2001) untuk membuat pengguna dapat berkomunikasi dengan sistem *data mining*, diciptakan *data mining primitive* yang dirancang untuk memfasilitasi penemuan pengetahuan secara efisien. Berikut ini *data mining primitive* menurut Han & Kamber (2001) :

1. *Task-relevant data*, digunakan untuk menspesifikasikan data yang akan *mining* karena pengguna umumnya hanya tertarik pada subset tertentu dari basis data. Dalam basis data relasional, *task-relevant data* dapat dikumpulkan melalui *query* basis data relasional seperti dengan operasi seleksi, penggabungan (*join*), dan operasi agregasi.
2. Jenis pengetahuan yang akan ditambah, meliputi deskripsi konsep (karakterisasi dan diskriminasi), asosiasi, klasifikasi, prediksi, *cluster* dan analisis evolusi.
3. Latar belakang pengetahuan yaitu hirarki konsep yang merupakan informasi tentang domain yang akan *di-mining* yang dapat bermanfaat untuk proses penemuan pengetahuan. Konsep hirarki dapat mendefinisikan serangkaian pemetaan dari konsep *berlevel* rendah menjadi *level* yang lebih tinggi.
4. Perhitungan nilai kemenarikan (*interestingness*) diperlukan karena tidak semua hasil dari *data mining* merupakan pengetahuan yang menarik pengguna sehingga diperlukan nilai tertentu yang dapat ditetapkan oleh pengguna (*threshold*) sebagai batas kemenarikan, aturan yang tidak memenuhi nilai tersebut dapat dianggap sebagai sesuatu yang tidak menarik dan dapat tidak ditampilkan sebagai pengetahuan.
5. Presentasi dan visualisasi pola yang ditemukan. Sistem *data mining* yang efektif harus mampu menampilkan pola-

pola yang ditemukan dalam bermacam-macam bentuk, seperti aturan, tabel, tabel silang, grafik, pohon keputusan, *cube*, atau visualisasi lain sehingga pengguna dari latar belakang yang berbeda dapat mengidentifikasi pola dan berinteraksi dengan sistem.

Data Mining Query Language (DMQL)

DMQL dapat membuat pengguna sistem *data mining* berinteraksi secara fleksibel karena DMQL menyediakan sebuah dasar dalam antarmuka grafik sistem *data mining* yang *user-friendly*. Setiap tujuan *data mining* pasti berbeda sehingga diperlukan perancangan DMQL yang berbeda. Bahasa dari DMQL mengadopsi sintaks SQL sehingga dapat secara mudah diintegrasikan dengan bahasa *query* relasional. Berikut beberapa DMQL menurut Han & Kamber (2001) :

1. Sintaks untuk spesifikasi *task-relevant data*.

```
<Pernyataan_Data_Mining> ::= 
  use database <nama_basis_data> | 
  use data warehouse
  <nama_data_warehouse> | { use
  hierarchy <nama_hirarki> for
  <atribut_atau_dimensi> }
  <spesifikasi_pengetahuan_yang_aka
  n_digali>
  in relevance to
  <kumpulan_atribut_atau_dimensi>
  from <relasi/cube>
  [where <kondisi>]
  [order by <grup>]
  [having <kondisi>]
  {[with
  [<nama_perhitungan_interest>]
  threshold = <nilai_threshold>
  [for <atribut>]}}
```

2. Sintaks untuk spesifikasi jenis pengetahuan yang akan digali.

Karakterisasi :

```
<Spesifikasi_pengetahuan> ::= 
  mine characteristics [as
  <nama_pola>]
  analyze <ukuran>
```

Diskriminasi :

```
<Spesifikasi_pengetahuan> ::= 
  mine comparison [as <nama_pola>]
  for <kelas_target>
  where <kondisi_target>
  { versus <kelas_pembanding_i>
  where <kondisi_pembanding_i> }
  analyze <ukuran>
```

3. Sintaks untuk spesifikasi hirarki konsep.

```
define hierarchy <nama_hirarki>
[for <atribut_atau_dimensi>]
```



```
on
<relasi_atau_cube_atau_hirarki>
>
as <deskripsi_hirarki>
[where <kondisi>]
```

4. Sintaks untuk spesifikasi perhitungan nilai kemenarikan

```
with
[<perhitungan_interestingness>]
threshold = <nilai_threshold>
```

5. Sintaks untuk visualisasi hasil

```
display as <bentuk_hasil>
```

Deskripsi Konsep

Deskripsi konsep menghasilkan sebuah deskripsi untuk karakterisasi dan perbandingan data. Karakterisasi menyediakan ringkasan yang tepat dari sekumpulan data, sedangkan perbandingan menyediakan deskripsi dari perbandingan dua atau lebih koleksi data (Han & Kamber 2001).

Deskripsi konsep erat kaitannya dengan generalisasi data. Generalisasi data adalah sebuah proses yang menentukan abstraksi sebuah himpunan yang berukuran besar dari *task-relevant data* dalam sebuah basis data dari *level konseptual* yang relatif rendah ke *level konseptual* yang lebih tinggi (Han & Kamber 2001).

Metode yang digunakan dalam deskripsi konsep menurut Han & Kamber (2001) yaitu pendekatan *data cube* dan induksi berorientasi atribut.

Induksi berorientasi atribut diperkenalkan pada tahun 1989 pada *workshop KDD* tahun 1989. Ide umum dari pendekatan induksi berorientasi atribut (AOI) adalah pertama mengoleksi *task-relevant data* menggunakan *query* basis data relasional dan kemudian melakukan generalisasi berdasarkan pengujian dari sejumlah nilai yang berbeda dari setiap atribut dalam kumpulan *task-relevant data*.

Pada dasarnya proses induksi basis data berjalan dalam generalisasi yang harus dilakukan secara hati-hati untuk mencegah *over generalisasi* (Han *et al.* 1992). Generalisasi dapat dilakukan melalui pembuangan atribut (*attribute removal*) dengan berdasarkan pada aturan; jika terdapat nilai beda yang besar untuk atribut pada *task-relevant data*, tetapi tidak terdapat operator generalisasi pada atribut tersebut (tidak ada hirarki konsep yang didefinisikan),

atau *level* konsep yang lebih tinggi dari suatu atribut dinyatakan dalam atribut lain, maka atribut tersebut dibuang (Han & Kamber 2001).

Generalisasi juga dapat dilakukan melalui generalisasi atribut. Pendekatan ini didasarkan pada aturan; jika terdapat nilai beda yang besar untuk sebuah atribut dalam *task-relevant data* dan terdapat sebuah himpunan dari operator-operator generalisasi pada atribut, maka sebuah operator generalisasi harus dipilih dan diaplikasikan terhadap atribut tersebut.

Berdasarkan pada atribut atau aplikasi yang digunakan, pengguna dapat menunjuk beberapa atribut untuk tetap pada *level abstraksi* yang rendah, sedangkan atribut yang lain digeneralisasi ke *level abstraksi* yang lebih tinggi. Proses ini diatur oleh *attribute generalization control*, jika atribut tertentu digeneralisasi terlalu tinggi, atribut tersebut akan menunjuk ke *over generalization* dan aturan yang dihasilkan dapat sangat tidak informatif. Di lain pihak, jika atribut tidak digeneralisasi ke *level konsep* yang cukup tinggi, maka dapat diperoleh *under generalization*, dimana aturan yang dihasilkan juga dapat tidak informatif.

Jika jumlah dari nilai beda tiap atribut lebih dari nilai *threshold* yang ditetapkan, maka generalisasi atribut atau pembuangan atribut harus dilakukan (Han & Kamber 2001).

Analisis Relevansi Atribut

Analisis relevansi atribut dilakukan untuk mengidentifikasi atribut yang tidak relevan atau yang kurang relevan. Selanjutnya atribut yang kurang relevan dapat dikeluarkan dari proses deskripsi konsep. Analisis relevansi atribut dilakukan dengan menghitung besaran yang digunakan untuk mengukur relevansi sebuah atribut terhadap kelas atau konsep yang diberikan. Salah satu ukuran tersebut adalah *information gain*.

Kalkulasi *information gain* dilakukan sebagai berikut (Han & Kamber 2001) :

S adalah sebuah himpunan dari *training samples*, dimana nama kelas dari setiap *sample* diketahui. Satu atribut digunakan untuk menentukan kelas dari *training sample*. Anggap bahwa terdapat *m* kelas. Misalkan *S* mengandung *s_i* *sample* dari kelas *C_i*, untuk *i* = 1, 2, ..., *m*. Sebuah *sample* sembarang anggota kelas *C_i* dengan probability *s_i/s*, dimana *s* adalah banyaknya total *sample* dalam himpunan *S*. *Expected information* diperlukan untuk mengklasifikasikan sebuah *sample* yang



diberikan. Nilai tersebut dihitung dengan formula :

$$I(s_1, s_2, \dots, s_m) = -\sum_{i=1}^m \frac{s_i}{s} \log_2 \frac{s_i}{s}$$

Sebuah atribut A dengan nilai $\{a_1, a_2, \dots, a_v\}$ dapat digunakan untuk mempartisi S ke dalam sub set $\{S_1, S_2, \dots, S_v\}$, dimana S_j mengandung sample-sample dalam S yang memiliki nilai a_j dari A. Misalkan S_j mengandung s_{ij} sample dari kelas C_i . *Expected information* yang didasarkan pada proses partisi ini oleh A dikenal sebagai *entropy* dari A. Besarannya dihitung sebagai berikut :

$$E(A) = \sum_{j=1}^v \frac{s_{1j} + \dots + s_{mj}}{s} I(s_{1j}, \dots, s_{mj})$$

Information gain diperoleh dengan proses partisi pada A didefinisikan oleh :

$$Gain(A) = I(s_1, s_2, \dots, s_m) - E(A)$$

Melalui persamaan ini dapat dihitung *information gain* untuk setiap atribut, atribut dengan *information gain* tertinggi dikatakan sebagai atribut yang paling relevan terhadap proses analisis demikian juga sebaliknya untuk atribut dengan *information gain* yang terendah dianggap sebagai atribut yang paling tidak relevan (Han & Kamber 2001)..

Selanjutnya hasil dari perhitungan *information gain* dianalisa dan dapat ditetapkan sebuah nilai *threshold* sebagai batas relevansi yang buruk, sehingga atribut yang nilai *informatin gain*-nya kurang dari nilai *threshold*, atribut tersebut dapat dibuang (Han & Kamber 2001).

Induksi Berorientasi Atribut

Di bawah ini algoritma induksi berorientasi atribut (Han & Kamber 2001) :

Input : (i) basis data relasional; (ii) query data mining; (iii) kumpulan atribut (terdiri dari atribut a_i); (iv) $Gen(a_i)$, kumpulan hierarki konsep atau operator generalisasi untuk atribut a_i ; (v) $a_gen_thres(a_i)$, threshold generalisasi untuk setiap a_i

Output : P
(Prime_generalized_relation)

Metode :

1. $W \leftarrow$ kumpulkan task-relevant data (query basis data)

2. Menentukan nilai-nilai tergeneralisasi untuk setiap atribut pada task-relevant data untuk kelas target dan kelas pembanding. Langkah ini diimplementasikan sebagai berikut :

i. Amati task-relevant data dan kumpulkan semua nilai yang berbeda untuk setiap atribut dalam task-relevant data.

ii. Untuk setiap atribut dalam task-relevant data, tentukan apakah atribut tersebut harus dibuang atau tidak. Jika tidak, hitung level konsep minimum yang diinginkan (L_i) untuk setiap atribut berdasarkan *threshold* yang diberikan. Selanjutnya tentukan pasangan-pasangan pemetaan (v, v') , dimana v adalah nilai yang berbeda dari suatu atribut dalam task-relevant data, dan v' adalah nilai tergeneralisasi yang sesuai pada level L_i .

3. Menentukan Prime_generalized_relation (P). Prime_generalized_relation, P ditentukan dengan mengganti setiap nilai v dalam initial target working class dengan nilai v' yang sesuai dengan pemetaan. Pada saat yang bersamaan, dilakukan perhitungan nilai-nilai agregat yaitu count. Langkah ini diimplementasikan dengan menggunakan dua variasi berikut :

i.Untuk setiap tuple tergeneralisasi, masukkan tuple ke dalam P yang telah diurutkan dengan menggunakan algoritma pencarian biner. Jika tuple telah ada dalam P, naikkan nilai count atau nilai agregat lainnya, selainnya masukkan tuple tersebut ke dalam P.

ii.Karena dalam banyak kasus banyaknya nilai beda setiap atribut dalam Prime_generalized_relation (P) cukup kecil, maka P dapat dikodekan sebagai array dimensi m dimana m adalah banyaknya atribut dalam P dan setiap dimensi berisi nilai atribut tergeneralisasi yang sesuai. Setiap elemen array memiliki nilai count atau nilai agregat lainnya. Penyisipan sebuah tuple tergeneralisasi dilakukan oleh ukuran agregasi dalam elemen array yang sesuai.

Setelah algortima induksi berorientasi atribut diimplementasikan, selanjutnya dapat dipresentasikan hasil dari generalisasinya. Presentasi dari hasil generalisasi dapat dilakukan berdasarkan deskripsi masing-masing atribut berdasarkan kelas target tertentu. Selain itu presentasi untuk perbandingan atribut berdasarkan kelas target dan kelas pembanding



tertentu juga dapat dilakukan dengan perhitungan yang berbeda. Untuk presentasi hasil dengan tujuan pendeskripsian atribut dilakukan dengan menghitung nilai *t-weight*, sedangkan untuk perbandingan kelas dihitung nilai *d-weight*.

T-weight merupakan sebuah ukuran nilai kemenarikan yang mendeskripsikan kekhasan dari setiap disjungsi dalam aturan atau setiap *tuple* dalam relasi tergeneralisasi yang berhubungan. Perhitungan nilai tersebut sebagai berikut (Han & Kamber 2001) :

Misal kelas target adalah kelas objek yang akan dikarakterisasi. q_a adalah *tuple* tergeneralisasi yang mendeskripsikan kelas target dan *t-weight* adalah persentase dari *tuple* dalam kelas target di *initial working relation* yang dicakup oleh q_a . Persamaannya adalah sebagai berikut :

$$\text{t-weight} = \frac{\text{count}(q_a)}{\sum_{i=1}^n \text{count}(q_i)}$$

dimana n adalah jumlah *tuple* untuk kelas target dalam relasi tergeneralisasi. q_1, \dots, q_n adalah *tuple* untuk kelas target dalam relasi tergeneralisasi, dan q_a ada dalam q_1, \dots, q_n . Nilai *t-weight* berkisar antara 0.0 sampai 1.0 atau 0% sampai 100%.

Nilai *t-weight* dapat juga direpresentasikan dalam bentuk aturan sebagai berikut :

$$\forall X, \text{kelas } _t \arg \text{et}(X) \Rightarrow \text{kondisi}_1(X) [t : w_1] \vee \dots \vee \text{kondisi}_m(X) [t : w_m]$$

Untuk perhitungan *d-weight* sebagai berikut (Han & Kamber 2001) :

Misalkan q_a merupakan *tuple* tergeneralisasi dan C_j merupakan kelas target, dimana q_a mencakup beberapa *tuple* dari kelas target dan q_a mungkin juga mencakup beberapa *tuple* dari kelas pembanding. *D-weight* untuk q_a adalah rasio perbandingan jumlah *tuple* dari *initial target class working* yang dicakup oleh q_a dengan jumlah *tuple* dalam *initial target class working* dan kelas pembanding yang dicakup oleh q_a . Persamaan untuk *d-weight* adalah sebagai berikut :

$$d\text{-weight} =$$

$$\text{count}(q_a \in C_j) / \sum_{i=1}^m \text{count}(q_a \in C_i)$$

dimana m adalah jumlah dari kelas target dan kelas pembanding, C_j berada dalam $\{C_1, \dots, C_m\}$ dan $\text{count}(q_a \in C_i)$ adalah jumlah *tuple* dari kelas C_i yang dicakup oleh q_a . *d-weight* berkisar antara 0.0 sampai 1.0 atau 0% sampai 100%. Sama seperti *t-weight*, nilai *d-weight* dapat juga direpresentasikan dalam bentuk aturan sebagai berikut :

$$\forall X, \text{kelas } _t \arg \text{et}(X) \Leftarrow \text{kondisi } (X) [d : d_weight]$$

METODE PENELITIAN

Penelitian akan dilakukan dengan metode sebagai berikut :

1. Metode analisis

Mengadakan survei dan observasi untuk mengumpulkan data yang diperlukan.

2. Metode perancangan

Mengadakan langkah-langkah proses KDD, yaitu :

a. Integrasi data, merupakan proses penggabungan data dari beberapa sumber.

b. Pembersihan data, untuk membuang data yang tidak konsisten dan data yang mengandung *noise*.

c. Transformasi data, proses transformasi data ke bentuk yang dapat di-mining.

d. Aplikasi teknik *data mining*, merupakan tahap yang paling penting karena pada tahap ini teknik *data mining* diaplikasikan terhadap data dengan langkah-langkah induksi berorientasi atribut sebagai berikut :

1. Penentuan *task-relevant data*

2. Menentukan nilai-nilai ter-generalisasi untuk setiap atribut pada *task-relevant data* untuk kelas target dan kelas pembanding.

3. Menentukan

Prime_generalized_relation P

e. Menentukan relevansi atribut.

f. Pencarian pola dalam data, pada tahap ini perhitungan terhadap nilai *t-weight* dan *d-weight* dilakukan.



- g. Presentasi pengetahuan merupakan tahap akhir dimana pada tahap ini pola yang telah ditemukan dipresentasikan ke pengguna dengan teknik visualisasi agar pengguna dapat memahaminya. Deskripsi dan perbandingan akan dipresentasikan dalam bentuk tabel, bentuk-bentuk aturan logika serta presentasi secara grafis untuk selanjutnya akan dievaluasi hasil pengetahuan yang didapatkan.

Spesifikasi Implementasi

Implementasi dirancang dan dibangun dengan perangkat keras dan perangkat lunak sebagai berikut :

Perangkat keras :

1. Prosesor Pentium 4 1.8 GHz
2. Memori 256 MB
3. Harddisk 40 GB
4. Monitor 15"
5. Alat input mouse dan keyboard

Perangkat lunak :

1. Sistem Operasi : Windows XP
2. Bahasa pemrograman : Visual Basic 6.0
3. DBMS : SQL Server 2000, Microsoft Access

HASIL DAN PEMBAHASAN

Gambaran Umum Aplikasi

Aplikasi yang dikembangkan pada penelitian ini didasarkan atas kebutuhan proses KDD termasuk salah satu proses yang terpentingnya yaitu *data mining*. Metode *data mining* yang digunakan dalam aplikasi ini yaitu metode induksi berorientasi atribut. Aplikasi dirancang atas dua tujuan utama yaitu :

1. Menyediakan sebuah fasilitas untuk mempermudah implementasi proses KDD serta algoritma induksi berorientasi atribut.
2. Menyediakan fasilitas presentasi hasil berupa tabel, aturan, serta grafik pengetahuan hasil penggalian data dengan metode induksi berorientasi atribut.

Struktur menu utama pada aplikasi ini terdiri dari menu koneksi yang digunakan untuk mengubah koneksi ke basis data serta *server* basis data yang diinginkan pengguna, menu integrasi untuk melakukan

pengintegrasian data yang akan digali pengetahuannya, menu pembersihan yang menyediakan deteksi terhadap data yang *null* dan data yang mengandung *noise* serta pilihan-pilihan untuk membersihkan data tersebut, menu generalisasi untuk melihat jumlah beda setiap atribut dan membandingkan dengan *threshold* tertentu yang ditetapkan pengguna serta menyediakan pilihan-pilihan generalisasi, penghapusan atribut, dan pembentukan atribut baru untuk atribut-atribut dengan nilai beda lebih besar dari *threshold* yang ditetapkan, menu analisis relevansi yang menyediakan fasilitas perhitungan nilai *information gain* sebagai ukuran relevansi sebuah atribut serta perhitungan nilai *t-weight* dan nilai *d-weight* sebagai ukuran kemenarikan hasil pengetahuan serta representasinya melalui aturan logika, tabel, maupun grafik, serta menu aturan untuk penginputan aturan-aturan baru pendekstrian data yang mengandung *noise*, dan menu *query* untuk melakukan *query* relasional terhadap data. Gambaran struktur menu pada aplikasi dapat dilihat pada Lampiran 1.

Integrasi Data

Untuk mempermudah proses pembersihan data terlebih dahulu dilakukan proses integrasi data. Tujuannya agar proses pembersihan data nantinya dapat sekaligus membersihkan data yang tidak konsisten akibat adanya integrasi data.

Proses integrasi data dalam penelitian ini diawali dengan konversi data nilai pelamar dari tahun 2000 sampai 2004, data nilai terpilih dari tahun 2000 sampai tahun 2004 serta data sekolah asal pelamar dari tahun ajaran 2000/2001 sampai tahun ajaran 2004/2005 yang disimpan dalam format basis data Foxpro (.dbf) menjadi format SQL Server 2000 (.mdf).

Selanjutnya, proses integrasi data dilakukan dengan dua buah modul pengintegrasian data. Modul integrasi yang pertama yaitu modul *gabung*, digunakan untuk menggabungkan tabel-tabel tanpa hubungan antar atribut tertentu dalam penggabungan, dan melakukan penggabungan penambahan baris. Setiap tabel yang akan digabungkan diidentifikasi nama atribut dan tipe datanya, untuk kemudian dibuatkan tabel baru dengan nama atribut sesuai dengan hasil identifikasi seluruh atribut pada setiap tabel yang akan digabungkan. Hasil tabel gabungan akan memiliki seluruh atribut pada seluruh tabel yang digabungkan. Jika atribut tertentu tidak dimiliki oleh salah satu tabel, maka pada hasil penggabungan, atribut tersebut akan tetap dibuat dengan nilai '*null*' untuk



seluruh *tuple* pada tabel yang tidak memiliki atribut tersebut.

Modul *gabung* ini digunakan untuk menggabungkan seluruh tabel calon dan terpilih dari tahun 2000 sampai tahun 2004. Untuk tetap menjaga informasi tahun pelamar, maka pada modul *gabung* akan dibuat satu atribut baru yaitu ‘tahun’ pada tabel hasil gabungan untuk mengidentifikasi tahun pelamar. Nilai dari atribut tahun tersebut diambil dari nama tabel yang mengandung tahun pelamar.

Pada studi kasus data PPMB, karena semua *tuple* dan atribut dalam tabel terpilih telah dicakup pada tabel calon, sehingga untuk proses selanjutnya hanya tabel calon yang diikutsertakan, sehingga pada modul integrasi yang pertama, penambahan atribut ‘status’ pada tabel hasil gabungan untuk mengidentifikasi calon yang diterima dan yang tidak diterima diperlukan. Modul *gabung* ini secara otomatis mengisi nilai atribut tersebut berdasarkan ada tidaknya nilai atribut *nrp*.

Modul integrasi yang kedua yaitu *gabung_rela*, digunakan untuk menggabungkan tabel-tabel yang membutuhkan atribut tertentu sebagai *primary key* penghubung antar tabel dan pengintegrasian dilakukan secara horizontal (penambahan atribut). Modul *gabung_rela* ini digunakan untuk menggabungkan tabel pelamar hasil gabungan dengan tabel data jumlah pelamar, jumlah pelamar yang diterima serta jumlah yang datang per sekolah (tabel PTD) dan tabel penilaian sekolah pelamar (tabel perak) dan tidak menghasilkan tabel baru. Pengguna menginputkan nama tabel hasil gabungan calon, dan modul *gabung_rela* akan menambahkan seluruh atribut pada tabel PTD dan tabel perak ke dalam tabel gabungan calon, sedangkan untuk nilai masing-masing atribut diambil dengan memperhatikan hubungan antar atribut nosla dan tahun pada tabel hasil gabungan pelamar dan atribut nosla dan T_akademik pada tabel PTD dan tabel perak sebagai *primary key*. Penggabungan tabel tanpa relasi dan dengan relasi dapat dilakukan pada *form* yang terlihat pada Gambar 2 dan Gambar 3.



Gambar 2 *Form* integrasi tabel tanpa relasi



Gambar 3 *Form* integrasi tabel dengan relasi

Setelah semua tabel diintegrasikan dengan menggunakan kedua modul integrasi ini, hasilnya didapat sebuah tabel gabungan yang memiliki 163 atribut dengan 52360 *tuple* yang secara rinci keterangan tiap atributnya dapat dilihat pada Lampiran 2.

Pembersihan Data

Sebelum proses pembersihan terhadap data, dilakukan identifikasi terhadap data yang kosong (*null*), data yang mengandung *noise*, maupun data yang tidak konsisten karena proses pengentrian data sebelumnya maupun akibat proses integrasi data.

Proses pembersihan data dilakukan dengan sebuah modul *ceknose* dalam aplikasi *data mining*. Modul *ceknose* ini mampu mengidentifikasi jumlah *tuple* bernilai *null*, *tuple* yang mengandung *noise*, *tuple* yang mengandung nilai yang tidak konsisten untuk setiap atribut, serta mampu membersihkan data yang teridentifikasi tersebut berdasarkan keputusan pengguna.

Untuk pembersihan data terhadap *tuple*-*tuple* yang bernilai *null*, modul ini menyediakan pilihan pembersihan meliputi penghapusan atribut, penghapusan *tuple* yang bernilai *null*,



pengisian secara manual, pengisian dengan nilai dari atribut lain, dan pengisian dengan nilai rataan khusus untuk atribut dengan tipe data *integer*, *double* maupun *float*. Modul *ceknoise* juga menyediakan *form* khusus untuk melakukan *query* relasional biasa untuk mempermudah pengguna melakukan pilihan pembersihan data yang lain yang tidak disediakan. Untuk atribut kategori terdapat pilihan pembersihan khusus yang direkomendasikan, yaitu pengisian nilai *null* dengan nilai rataan kategori berdasarkan nosla dan propinsi dengan terlebih dahulu mengkonversi data kategori ke angka.

Untuk pendekstrian nilai yang mengandung *noise* dilakukan berdasarkan data pada tabel aturan. Sebelum aplikasi dapat mendekripsi nilai-nilai yang mengandung *noise*, pengguna harus menginputkan ke dalam aplikasi, aturan-aturan setiap atribut dalam tabel yang akan di-*mining*. Aturan-aturan disimpan dalam tabel aturan. Setiap atribut dapat memiliki aturan berdasarkan ukuran panjang nilai yang dapat diinputkan, atau *range* nilai tertentu, atau kriteria tertentu (nilai-nilai tertentu saja yang bisa diinputkan), atau format penulisan tertentu, atau format penulisan berdasarkan atribut tertentu, atau berdasarkan operasi aritmatik tertentu dari atribut-atribut lain.

Untuk pendekstrian nilai yang mengandung *noise*, modul akan mencari berdasarkan nama atribut aturan-aturan dalam tabel aturan dan nilai-nilai yang tidak memenuhi aturan akan dianggap sebagai nilai yang mengandung *noise*. Untuk atribut-atribut yang belum diinputkan aturan ke dalam tabel aturan, maka modul akan menuliskan ‘no aturan’. Pendekstrian nilai yang mengandung *noise* juga dapat digunakan untuk mendekripsi nilai yang tidak konsisten dalam suatu atribut. Untuk pembersihan *tuple-tuple* yang terdeteksi mengandung *noise*, aplikasi menyediakan sarana untuk menginputkan ulang data yang mengandung *noise*. *Form* pendekstrian data *null* dan data yang mengandung *noise* dapat dilihat pada Gambar 4 dibawah ini.

Gambar 4 *Form* pendekstrian data *null* dan yang mengandung *noise*

Dari hasil proses pembersihan data terdapat 5033 *tuple* yang harus dibuang karena berisi data kosong (*null*) dan data yang mengandung *noise*, serta terdapat 48 atribut yang harus dibuang karena hampir seluruh nilai dalam atribut tersebut kosong (*null*) serta beberapa merupakan atribut yang jika diisi dengan nilai tertentu maka atribut tersebut akan menghasilkan *redundancy* dengan atribut lain. Dari hasil pembersihan data tersisa 47327 *tuple* yang tersisa dan 116 atribut. Hasil identifikasi data yang mengandung *noise*, *null* maupun tidak konsisten serta aksi pembersihan datanya dapat dilihat secara rinci pada Lampiran 3.

Karena dalam proses *preprocessing* melibatkan proses integrasi data dari berbagai sumber, maka perlu dianalisis tentang masalah *redundancy* data per atribut. Masalah ini dapat diatas dengan membuang salah satu atribut jika terbukti ada dua atau lebih atribut yang sama menunjuk pada satu entitas. Pengecekan *redundancy* dilakukan secara manual karena asumsi bahwa pengguna sistem sangat mengetahui tentang data.

Pada proses identifikasi atribut yang memiliki *redundancy* didapat 6 buah atribut yang saling berkorelasi yaitu atribut tahun dengan T_akademik, atribut propinsi dengan kodeprop, dan atribut nrp dengan nim, sehingga tiga atribut diantaranya harus dibuang (yaitu atribut T_akademik, kodeprop, dan nim). Secara keseluruhan hasil proses ini menghasilkan 113 atribut yang tersisa.

Transformasi Data

Proses ini dilakukan untuk mentransformasi data yang dimiliki ke dalam bentuk tertentu yang dapat di-*mining*. Proses transformasi data pada penelitian ini diperlukan untuk mengatasi banyaknya nilai beda pada satu atribut. Atribut-atribut yang berguna dalam tujuan *data mining* mungkin memiliki banyak nilai beda sehingga



nantinya akan dibuang dan tidak diikutsertakan pada proses *data mining*. Setelah dianalisis, terdapat empat buah atribut yang bisa memiliki nilai beda yang besar sehingga memungkinkan atribut tersebut akan dibuang pada tahap penghapusan dan generalisasi atribut, sedangkan atribut-atribut tersebut diperlukan untuk tujuan *data mining*. Keempat atribut tersebut ditransformasi dengan mengkonstruksi atribut-atribut baru untuk memisahkan dan menampung nilai dari masing-masing atribut sehingga jumlah beda tiap atribut akan kurang dari *threshold* yang ditetapkan. Keempat atribut tersebut yaitu kbpt, alasanol, karir, dan sumberinf.

Pada aplikasi modul untuk transformasi data digabungkan dengan modul generalisasi data. Aplikasi menyediakan dua pilihan pembentukan atribut baru, yang pertama pembentukan atribut berdasarkan jumlah digit nilai atribut, sehingga atribut baru akan dibentuk sebanyak jumlah digit nilai atribut yang diinputkan oleh pengguna, dan sistem akan mengambil satu per satu digit dalam atribut lama dan menampungnya ke dalam satu per satu atribut yang baru. Pemberian nama atribut berdasarkan atribut lama ditambah ‘X’ serta digit 1, 2, sampai panjang atribut lama. Untuk fasilitas transformasi yang ini digunakan pada pembentukan atribut baru untuk Kbpt.

Fasilitas pembentukan atribut baru yang kedua yaitu berdasarkan kemungkinan jumlah beda untuk satu atribut. Pengguna menginputkan jumlah nilai yang dapat diinputkan ke dalam atribut tersebut, lalu sistem akan membentuk atribut baru dengan nama atribut sama dengan nama atribut lama ditambah ‘X’ serta digit 1, 2, dan seterusnya yang mengidentifikasi nilai yang dapat diinputkan ke dalam atribut lama. Sistem akan mengisi atribut-atribut baru tersebut berdasarkan ada atau tidaknya nilai pada atribut lama. Jika pada atribut lama terdapat nilai tersebut, maka nilai dari atribut baru akan berisi ‘1’ dan sebaliknya jika tidak, maka nilai atribut baru akan berisi ‘0’.

Hasil dari tahap ini terdapat penambahan atribut menjadi 137 atribut.

Penentuan *Task-relevant Data*

Penentuan *task-relevant data* dilakukan untuk deskripsi kelas pelamar USMI IPB pada penelitian ini dipilih berdasarkan :

1. 8 Fakultas yang diminati oleh pelamar USMI pada pilihan pertamanya, yaitu : A (Fakultas Pertanian), B (Fakultas Kedokteran Hewan), C (Fakultas Perikanan dan Ilmu Kelautan), D (Fakultas Peternakan), E (Fakultas Kehutanan), F (Fakultas Teknologi Pertanian), G (Fakultas Matematika dan Ilmu Pengetahuan Alam), H (Fakultas Ekonomi dan Manajemen)
2. 7 pulau asal pelamar, yaitu : 1 (Sumatera), 3 (Jawa), 5 (Nusa Tenggara), 6 (Kalimantan), 7 (Sulawesi), 8 (Irian Jaya), 9 (Luar Negeri)

Penentuan *task-relevant data* dilakukan dengan *query data mining* yang kemudian ditransformasikan ke dalam bentuk *query relasional*. Berikut contoh *query data mining* dengan DMQL untuk *task-relevant data* fakultas yang dipilih pelamar pada pilihan pertama :

```
use basis data test
mine characteristics as "pilihan1"
in relevance to status, tahun,
propinsi, nosla, noinduk, nrp, nama,
jeniskel, thnmasuk, b1, b2, b3, b4,
b5, b6, b7, b8, f1, f2, f3, f4, f5,
f6, f7, f8, k1, k2, k3, k4, k5, k6,
k7, k8, m1, m2, m3, m4, m5, m6, m7,
m8, r1, r2, r3, r4, r5, r6, r7, j1,
j2, j3, j4, j5, j6, j7, Umum,
klssiswa, jumlahkls, nopindah,
tidaknaik, pilihan1,pilihan2,
pilihan3, putusani, listrik,
kelengkapan, hadir, sumberbia,
sumberbia2, biayahidup, karir2,
persepsi01, persepsi02, persepsi03,
persepsi04, persepsi05, persepsi06,
persepsi07, persepsi08, persepsi09,
persepsi10, persepsi11, persepsi12,
persepsi13, seleksi2, psdiploma,
sandiopera, kodepsbaru, kodepil,
peringkat, seleksi, namasl, usmipel,
usmitrm, usmidtg, jumlahmhs, nemipal,
nemipa2, nemmin1, nemmax1, nemratal,
NMR1, NMR2, NMR3, NMR4, NMR5,
Kategori, KbptX1, KbptX2, KbptX3,
KbptX4, KbptX5, sumberinfX1,
sumberinfX2, sumberinfX3, sumberinfX4,
sumberinfX5, sumberinfX6, sumberinfX7,
sumberinfX8, sumberinfX9, alasanolX1,
alasanpilX2, alasanpilX3, alasanpilX4,
alasanpilX5, alasanpilX6, alasanpilX7,
alasanpilX8, alasanpilX9, alasanolX10,
karirX1, karirX2, karirX3, karirX4
```

from tabel gab3

```
where pilihan1 in {"A", "B", "C", "D",
"E", "F", "G", "H"}
```

Hasil transformasi ke bahasa *query relasional* sebagai berikut :

```
use basis data test
```



```
Select status, tahun, propinsi,
nosla, noinduk, npn, nama,
jeniskel, thnmasuk, b1, b2, b3, b4,
b5, b6, b7, b8, f1, f2, f3, f4, f5,
f6, f7, f8, k1, k2, k3, k4, k5, k6,
k7, k8, m1, m2, m3, m4, m5, m6, m7,
m8, r1, r2, r3, r4, r5, r6, r7, j1,
j2, j3, j4, j5, j6, j7, Umum,
klssiswa, jumlahkls, nopindah,
tidaknaik, pilihan1,pilihan2,
pilihan3, putusanl, listrik,
kelengkapan, hadir, sumberbia,
sumberbia2, biayahidup, karir2,
persepsi01, persepsi02, persepsi03,
persepsi04, persepsi05, persepsi06,
persepsi07, persepsi08, persepsi09,
persepsi10, persepsi11, persepsi12,
persepsi13, seleksi2, psdiploma,
sandiopera, kodepsbaru, kodepil,
peringkat, seleksi, namasl,
usmipel, usmitrm, usmidtg,
jumlahmhs, nemipal, nemipa2,
nemmin1, nemmax1, nemratal, NMR1,
NMR2, NMR3, NMR4, NMR5, Kategori,
KbptX1, KbptX2, KbptX3, KbptX4,
KbptX5, sumberinfX1, sumberinfX2,
sumberinfX3, sumberinfX4,
sumberinfX5, sumberinfX6,
sumberinfX7, sumberinfX8,
sumberinfX9, alasanpilX1,
alasanpilX2, alasanpilX3,
alasanpilX4, alasanpilX5,
alasanpilX6, alasanpilX7,
alasanpilX8, alasanpilX9,
alasanpilX10, karirX1, karirX2,
karirX3, karirX4
```

From tabel gab3

```
Where pilihan1 = 'A' or pilihan1 =
'B' or pilihan1 = 'C' or pilihan1 =
'D' or pilihan1 = 'E' or
pilihan1 = 'F' or pilihan1 = 'G'
or pilihan1 = 'H'
```

Sedangkan untuk perbandingan kelas pelamar pada penelitian ini, hanya dipilih *task-relevant data* berdasarkan pulau asal pelamar. Untuk *query data mining* perbandingan dicontohkan seperti berikut ini untuk kelas perbandingan pulau asal pelamar :

```
Mine comparison as pulau
For pelamar where pulau='1'
Versus pelamar where pulau='3'
Versus pelamar where pulau='5'
Versus pelamar where pulau='6'
Versus pelamar where pulau='7'
Versus pelamar where pulau='8'
Versus pelamar where pulau='9'
Analyze count%
```

Pada aplikasi yang dibuat, penentuan *task-relevant data* dapat ditambah berdasarkan keinginan serta kebutuhan pengguna.

Penentuan Nilai-Nilai Tergeneralisasi Untuk Setiap Atribut Pada *Task-relevant Data*

Setelah *task-relevant data* ditentukan, selanjutnya pada tahap ini ditentukan banyaknya nilai beda untuk setiap atribut dalam *task-relevant data*. Penentuan banyaknya nilai beda dilakukan dengan bahasa *query* yang hasil rincinya dapat dilihat pada Lampiran 4.

Pada aplikasi, penentuan nilai beda untuk setiap atribut tergabung dengan langkah selanjutnya yaitu generalisasi dan pembuangan atribut.

Generalisasi dan Pembuangan Atribut

Setelah identifikasi jumlah nilai beda tiap-tiap atribut dalam *task-relevant data*, kemudian dilakukan analisis terhadap jumlah nilai beda tersebut. Analisis meliputi banyaknya jumlah beda per atribut, ada tidaknya operator generalisasi, dan ada tidaknya *level* yang lebih tinggi pada atribut lain. Jika jumlah beda suatu atribut melebihi nilai *threshold*, atau terdapat operator generalisasi, atau *level* yang lebih tinggi terdapat pada atribut lain, maka atribut tersebut dapat digeneralisasi ke *level* yang lebih tinggi atau atribut tersebut dapat tidak diikutsertakan pada proses selanjutnya (dibuang).

Aplikasi melakukan langkah ini pada sebuah modul *generalisasi*. Pertama, modul ini menghitung jumlah beda untuk semua atribut. Setelah itu modul akan mengecek apakah setiap atribut memiliki jumlah beda kurang dari nilai *threshold* yang sebelumnya telah ditentukan oleh pengguna. Untuk setiap atribut jika nilai bedanya kurang dari nilai *threshold*, maka modul akan memberi tahu pengguna untuk tidak membuang atribut tersebut, dan sebaliknya jika nilai beda suatu atribut tertentu melebihi nilai *threshold* maka modul akan menyarankan pengguna untuk membuang atribut tersebut.

Penanganan terhadap atribut yang memiliki jumlah beda melebihi nilai *threshold* tidak hanya terbatas pada membuang atribut. Modul *generalisasi* juga menyediakan fungsi generalisasi untuk men-generalisasi atribut-atribut dengan nilai beda yang melebihi nilai *threshold*. Generalisasi dilakukan dengan perubahan nilai-nilai pada seluruh *tuple* untuk atribut yang digeneralisasi. Generalisasi dilakukan dengan pilihan dua cara, yaitu pengambilan berapa digit tertentu pada data (berlaku untuk atribut-atribut yang memiliki operator generalisasi di dalam atribut tersebut) dan perubahan nilai atribut pada *range* tertentu.



Selain itu modul *generalisasi* ini menyediakan fungsi untuk mentransformasi data dengan mengkonstruksi atribut-atribut baru seperti yang telah dijelaskan sebelumnya. Identifikasi nilai beda untuk setiap atribut dan generalisasi serta pembuangan atributnya dilakukan pada *form* yang ditunjukkan pada Gambar 5.



Gambar 5 *Form* generalisasi atribut

Pada penelitian ini, ditetapkan nilai *threshold* sebesar 12. Pada Lampiran 5 ditunjukkan hasil lengkap generalisasi dan pembuangan atribut.

Setelah proses generalisasi atribut dan pembuangan atribut dihasilkan 95 buah atribut untuk selanjutnya diikutsertakan ke dalam analisis relevansi atribut untuk menetapkan atribut mana yang paling relevan.

Analisis Relevansi Atribut

Analisis relevansi atribut dilakukan untuk mengetahui atribut mana yang paling relevan dan atribut mana yang paling tidak relevan untuk menggambarkan hasil *data mining*. Proses analisis relevansi atribut dilakukan untuk tiap *task-relevant data*.

Aplikasi menyediakan sebuah modul *entropy* untuk menghitung nilai *expected information*, *entropy*, serta *information gain* tiap-tiap atribut berdasarkan kelas target tertentu yang dapat diubah oleh pengguna. Nilai dari *expected information*, *entropy*, serta *information gain* tersebut juga dapat diurut dari kecil ke besar berdasarkan nilai *information gain* untuk mempermudah melihat urutan relevansi tiap-tiap atribut.

Hasil ringkasan atribut yang paling relevan dan yang paling tidak relevan untuk tiap-tiap *task-relevant data* dapat dilihat pada Tabel 1. Sedangkan hasil rinci untuk

semua nilai *information gain* tiap-tiap atribut dalam masing-masing *task-relevant data* dapat dilihat pada Lampiran 6, Lampiran 7.

Tabel 1 Ringkasan hasil analisis relevansi atribut

Task relevant data	Atribut paling relevan		Atribut paling tidak relevan	
	Nama	Inf-gain	Nama	Inf-gain
Fakultas	putusan1	0.43 324	alasanpilX10	0.00 013
Pulau	kategori	0.23 102	tidak naik	0.00 008

Pada aplikasi, pengguna diharuskan menginputkan nilai *threshold* sebagai batas relevansi atribut untuk menentukan atribut-atribut yang relevan yang dapat dihitung nilai *t-weight* dan *d-weight* sebagai hasil presentasi dari *data mining*. Untuk nilai *information gain* yang kurang dari nilai *threshold*, maka atribut tersebut tidak akan diikutkan dalam presentasi hasil, dan sebaliknya atribut dengan nilai *information gain* yang lebih besar dari nilai *threshold* akan diikutsertakan. Perhitungan nilai *expected information*, *entropy*, serta *information gain* dilakukan pada *form* yang terlihat pada Gambar 6.



Gambar 6 *Form* analisis relevansi atribut

Pada penelitian dengan data PPMB ditetapkan nilai *threshold* untuk *information gain* untuk *task-relevant data* berdasarkan fakultas pilihan pertama pelamar sebesar 0.0148. Nilai *threshold* ini diambil dari rataan nilai *information gain* dari setiap atribut yang didapatkan dengan menggunakan modul *entropy*. Namun dengan tujuan untuk tetap melihat kecenderungan asal pulau pelamar, maka nilai *threshold* diturunkan sampai batas 0.0143, sehingga untuk *task-relevant data* ini terdapat 13 atribut yang dianggap relevan untuk dilihat hasil deskripsi kelasnya.



Sama dengan pada penetapan nilai *threshold* untuk batas relevansi atribut pada *task-relevant data* berdasarkan pilihan pelamar, nilai *threshold* batas relevansi atribut untuk *task-relevant data* berdasarkan pulau asal pelamar juga ditetapkan berdasarkan nilai rataan dari nilai *information gain* seluruh atribut pada *task-relevant data* pulau asal pelamar, dan didapatkan hasil sebesar 0.015. Namun, untuk tetap melihat kecenderungan fakultas pilihan pertama pelamar berdasarkan asal pulau pelamar, maka nilai *threshold* diturunkan sampai 0.01435, sehingga akan dilihat presentasi hasil 35 atribut untuk *task-relevant data* pulau asal pelamar pada langkah selanjutnya.

Presentasi Hasil

Untuk presentasi hasil pendeskripsian atribut dari kelas pelamar digunakan ukuran kemenarikan *t-weight*. Sedangkan untuk perbandingan kelas digunakan ukuran kemenarikan *d-weight*.

Pada aplikasi, presentasi hasil nilai *t-weight* dan *d-weight* setiap kelas target tertentu dapat dilihat melalui tabel, grafik dan aturan-aturan relasi.

Sebagai contoh untuk kelas target fakultas pilihan pertama pelamar untuk pendeskripsian atribut putusan1, presentasi hasil melalui aturan relasinya dapat dilihat dengan menentukan terlebih dahulu nilai *threshold* dari *t-weight* yang akan dibentuk aturan relasi. Misalnya untuk atribut putusan1 ditentukan nilai *threshold* untuk aturan relasi sebesar 16%, maka hasil presentasi dari aturan relasi didapat sebagai berikut :

$$\begin{aligned} \forall X, \text{pilihan } 1(X) = A &\Rightarrow \text{putusan } 1(X) = 0 \\ [\text{t : 74 \%}] \vee \text{putusan } 1(X) = A &[\text{t : 19 \%}] \\ \forall X, \text{pilihan } 1(X) = B &\Rightarrow \text{putusan } 1(X) = 0 \\ [\text{t : 64 \%}] \vee \text{putusan } 1(X) = B &[\text{t : 36 \%}] \\ \forall X, \text{pilihan } 1(X) = C &\Rightarrow \text{putusan } 1(X) = 0 \\ [\text{t : 72 \%}] \vee \text{putusan } 1(X) = C &[\text{t : 26 \%}] \\ \forall X, \text{pilihan } 1(X) = D &\Rightarrow \text{putusan } 1(X) = 0 \\ [\text{t : 65 \%}] \vee \text{putusan } 1(X) = D &[\text{t : 34 \%}] \\ \forall X, \text{pilihan } 1(X) = E &\Rightarrow \text{putusan } 1(X) = 0 \\ [\text{t : 71 \%}] \vee \text{putusan } 1(X) = E &[\text{t : 27 \%}] \\ \forall X, \text{pilihan } 1(X) = F &\Rightarrow \text{putusan } 1(X) = 0 \\ [\text{t : 73 \%}] \vee \text{putusan } 1(X) = F &[\text{t : 16 \%}] \end{aligned}$$

$$\begin{aligned} \forall X, \text{pilihan } 1(X) = G &\Rightarrow \text{putusan } 1(X) = 0 \\ [\text{t : 76 \%}] \vee \text{putusan } 1(X) = G &[\text{t : 16 \%}] \\ \forall X, \text{pilihan } 1(X) = H &\Rightarrow \text{putusan } 1(X) = 0 \\ [\text{t : 67 \%}] \vee \text{putusan } 1(X) = H &[\text{t : 32 \%}] \end{aligned}$$

Untuk contoh presentasi hasil melalui grafik dapat dilihat pada Lampiran 8, sedangkan untuk hasil perhitungan *t-weight* selengkapnya dapat dilihat pada Lampiran 9. Dari hasil perhitungan *t-weight* tersebut, dapat ditarik beberapa kesimpulan sebagai hasil pengetahuan dari tumpukan data yang digali bahwa sebesar 60%-70% pelamar tidak diterima pada fakultas pilihan pertama pelamar bahkan tidak diterima sama sekali menjadi mahasiswa IPB, dan hanya sebesar 15%-40% pelamar diterima pada fakultas pilihan pertamanya, sedangkan sisanya diterima pada fakultas pilihan kedua atau ketiga.

Untuk kelas target berdasarkan pulau asal pelamar, melalui hasil analisis relevansi atribut diketahui bahwa atribut kategori merupakan atribut paling relevan untuk mendeskripsikan kelas target pulau asal pelamar. Hasil selengkapnya perhitungan *t-weight* dapat dilihat pada Lampiran 10.

Dengan menetapkan nilai *threshold* sebesar 14% maka dapat dihasilkan aturan relasi sebagai berikut :

$$\begin{aligned} \forall X, \text{pulau } 1(X) = 1 &\Rightarrow \text{kategori } (X) = A \\ [\text{t : 25 \%}] \vee \text{kategori } (X) = A &[\text{t : 35 \%}] \\ \vee \text{kategori } (X) = B &[17 \%] \\ \forall X, \text{pulau } 1(X) = 3 &\Rightarrow \text{kategori } (X) = A \\ [\text{t : 36 \%}] \vee \text{kategori } (X) = A &[\text{t : 29 \%}] \\ \vee \text{kategori } (X) = A + [25 \%] \\ \forall X, \text{pulau } 1(X) = 5 &\Rightarrow \text{kategori } (X) = A \\ [\text{t : 18 \%}] \vee \text{kategori } (X) = B &[\text{t : 43 \%}] \\ \vee \text{kategori } (X) = B + [24 \%] \\ \forall X, \text{pulau } 1(X) = 6 &\Rightarrow \text{kategori } (X) = A \\ [\text{t : 16 \%}] \vee \text{kategori } (X) = B &[\text{t : 63 \%}] \\ \forall X, \text{pulau } 1(X) = 7 &\Rightarrow \text{kategori } (X) = B + \\ [\text{t : 48 \%}] \vee \text{kategori } (X) = C &[21 \%] \\ \forall X, \text{pulau } 1(X) = 8 &\Rightarrow \text{kategori } (X) = C \\ [\text{t : 27 \%}] \vee \text{kategori } (X) = C &[42 \%] \\ \forall X, \text{pulau } 1(X) = 9 &\Rightarrow \text{kategori } (X) = B \\ [\text{t : 49 \%}] \vee \text{kategori } (X) = A &[\text{t : 17 \%}] \\ \vee \text{kategori } (X) = A - [14 \%] \vee \text{kategori } (X) = B + \\ [\text{t : 14 \%}] \end{aligned}$$

Pada Lampiran 11 dapat dilihat contoh presentasi dengan grafik nilai *t-weight* dari kategori berdasarkan kelas target pulau asal pelamar. Dari hasil perhitungan *t-weight* untuk kategori, maka dapat diambil kesimpulan bahwa mayoritas sebaran kategori SMU di pulau Sumatra (1) adalah A, A-, dan B, sedangkan untuk pulau Jawa (3) adalah A, A+, dan A-. Untuk pulau Kalimantan (5) sebaran nilai kategori SMU berkisar pada nilai A, B-, dan B+. Untuk Pulau Nusa Tenggara (6) 60% SMU berkategori B+, dan 15% memiliki



kategori A-. Untuk pulau Sulawesi (7) mayoritas SMU nya memiliki kategori B+ dan C. Untuk pulau Irian Jaya nilai kategori untuk SMU di pulau tersebut berkisar pada nilai C dan C+, sedangkan terakhir untuk SMU luar negeri nilai kategori hampir merata pada nilai A, A-, B, dan B+. Kategori yang dimiliki masing-masing SMU di tiap pulau mencerminkan keberhasilan mahasiswa-mahasiswa sebelumnya yang berasal dari masing-masing pulau tersebut selama belajar di IPB, sehingga dapat disimpulkan bahwa untuk mahasiswa asal SMU dari pulau Sumatra, Jawa, Kalimantan, dan Nusa Tenggara memiliki prestasi akademik yang baik karena rata-rata kategori SMU nya berkisar B ke atas. Namun sebaliknya untuk mahasiswa asal pulau Sulawesi dan Irian Jaya, prestasi akademik mahasiswa yang berasal dari daerah tersebut kurang baik karena nilai kategori sekolah berkisar pada B dan C.

Untuk perbandingan kelas dilakukan dengan memilih kondisi tertentu yang akan dibandingkan berdasarkan kelas target dan kelas pembanding yang telah ditetapkan. Pada aplikasi, kondisi tertentu dapat diubah berdasarkan keinginan dan kebutuhan pengguna. Sebagai contoh untuk kasus data pelamar IPB dari tahun 2000 sampai dengan tahun 2004 ini akan dilihat perbandingan dari kelas target pulau asal pelamar Sumatera dibandingkan dengan kelas pembanding pulau asal pelamar Jawa, Kalimantan, Sulawesi, Nusa Tenggara, Irian Jaya, dan luar negeri. Kondisi yang dipakai sebagai contoh yaitu kombinasi antara listrik dan kode pilihan hasil keputusan seleksi IPB untuk melihat perbandingan persentase jumlah pelamar yang lolos seleksi pilihan pertama dengan mempertimbangkan kondisi listrik di rumah pelamar berdasarkan asal pulau. Hasil selengkapnya dapat dilihat pada Lampiran 12 dan untuk presentasi hasil dengan grafik dapat dilihat pada Lampiran 13, 14, 15, 16, 17, dan 18. Selain melalui tabel silang, presentasi nilai *d-weight* juga dapat dilihat dengan aturan relasi sebagai berikut (dengan penetapan nilai *threshold* sebesar 8%) :

$$\begin{aligned} \forall X, pulau(X) = 1 &\Leftarrow kodepil(X) = P1 \wedge \\ &listrik(X) = 1[d : 32\%] \\ \forall X, pulau(X) = 3 &\Leftarrow kodepil(X) = P1 \wedge \\ &listrik(X) = 1[d : 61\%] \\ \forall X, pulau(X) = 1 &\Leftarrow kodepil(X) = P1 \wedge \\ &listrik(X) = 2[d : 15\%] \end{aligned}$$

$$\begin{aligned} \forall X, pulau(X) = 3 &\Leftarrow kodepil(X) = P1 \wedge \\ &listrik(X) = 2[d : 80\%] \\ \forall X, pulau(X) = 1 &\Leftarrow kodepil(X) = P1 \wedge \\ &listrik(X) = 3[d : 14\%] \\ \forall X, pulau(X) = 3 &\Leftarrow kodepil(X) = P1 \wedge \\ &listrik(X) = 3[d : 80\%] \\ \forall X, pulau(X) = 1 &\Leftarrow kodepil(X) = P1 \wedge \\ &listrik(X) = 4[d : 12\%] \\ \forall X, pulau(X) = 3 &\Leftarrow kodepil(X) = P1 \wedge \\ &listrik(X) = 4[d : 82\%] \\ \forall X, pulau(X) = 1 &\Leftarrow kodepil(X) = P1 \wedge \\ &listrik(X) = 5[d : 8\%] \\ \forall X, pulau(X) = 3 &\Leftarrow kodepil(X) = P1 \wedge \\ &listrik(X) = 5[d : 85\%] \\ \forall X, pulau(X) = 1 &\Leftarrow kodepil(X) = P1 \wedge \\ &listrik(X) = 6[d : 14\%] \\ \forall X, pulau(X) = 3 &\Leftarrow kodepil(X) = P1 \wedge \\ &listrik(X) = 6[d : 67\%] \\ \forall X, pulau(X) = 9 &\Leftarrow kodepil(X) = P1 \wedge \\ &listrik(X) = 6[d : 13\%] \end{aligned}$$

Secara ringkas dari hasil perhitungan *d-weight*, dapat dinyatakan bahwa pada cakupan kelas target dan kelas pembanding asal pulau pelamar untuk pelamar yang diterima pada fakultas pilihan pertamanya dan untuk setiap kombinasi keadaan listrik, didapatkan pengetahuan bahwa pelamar yang memenuhi kondisi diterima pada pilihan pertama dan untuk setiap kombinasi kondisi listrik, kemungkinan pelamar tersebut berasal dari kelas pembanding Pulau Jawa lebih besar dibandingkan dengan kelas target Pulau Sumatra maupun kelas pembanding lainnya.

KESIMPULAN DAN SARAN

Kesimpulan

Aplikasi *data mining* yang dikembangkan dengan metode induksi berorientasi atribut ini mampu melakukan tahapan-tahapan KDD mulai dari tahap pengintegrasian data, pembersihan data, generalisasi, analisis relevansi atribut, serta perhitungan nilai *count*, *t-weight* dan *d-weight* sebagai hasil presentasi pengetahuan.

Aplikasi ini dibuat secara umum dan setiap keputusan analisis serta pengolahan data diserahkan kepada pengguna sehingga aplikasi ini akan mampu menerima data pelamar IPB melalui jalur USMI untuk tahun berikutnya. Pengguna dapat merubah batasan-batasan nilai tertentu serta merubah tujuan penggalian data untuk mendapatkan pengetahuan-pengetahuan lain yang diinginkan.

Hasil percobaan untuk pendeskripsiannya kelas dengan kelas target fakultas pilihan pertama pelamar menunjukkan bahwa separuh



lebih pelamar tidak lolos seleksi menjadi mahasiswa IPB dan hanya 15-40% pelamar yang berhasil lolos menjadi mahasiswa IPB pada fakultas pilihan pertama yang diinginkan serta sisanya lolos seleksi bukan pada fakultas pilihan pertama.

Sedangkan hasil percobaan untuk tujuan perbandingan kelas dengan kelas target dan kelas pembanding pulau asal pelamar menyatakan bahwa pelamar yang memenuhi kondisi diterima pada pilihan pertama dan untuk setiap kemungkinan kondisi listrik, kemungkinan pelamar tersebut berasal dari kelas pembanding Pulau Jawa lebih besar dibandingkan dengan kelas target Pulau Sumatra maupun kelas pembanding lainnya.

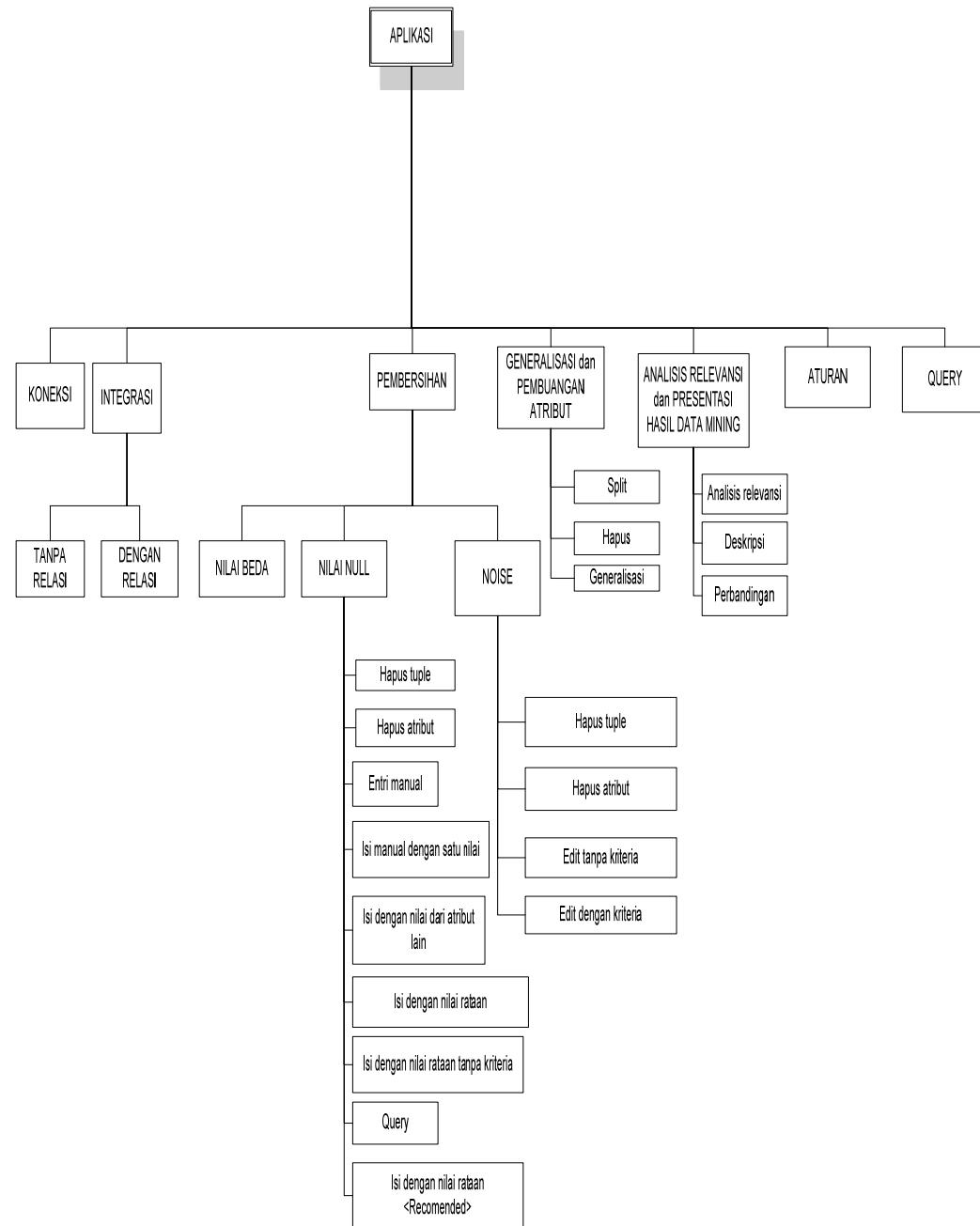
Saran

Aplikasi *data mining* yang dihasilkan pada penelitian ini masih dapat dikembangkan lebih lanjut pada :

1. Penyempurnaan modul integrasi dengan relasi yang mampu menghubungkan tabel dengan pilihan atribut yang menjadi *primary key* penghubung antar tabel dapat ditentukan oleh pengguna.
2. Pendekripsi nilai yang mengandung *noise* pada modul *ceknoise* baru terbatas pada aturan-aturan tunggal, sehingga penyempurnaan modul *ceknoise* yang dapat mendekripsi atribut-atribut dengan aturan lebih dari satu dapat dimungkinkan.
3. Untuk proses pembersihan data dapat ditambahkan metode pembersihan lain untuk data yang mengandung nilai *noise*, seperti dengan metode *binning* maupun *cluster*.
4. Penyempurnaan aplikasi dengan memperhitungkan kecepatan akses terhadap data juga dapat dilakukan.

DAFTAR PUSTAKA

- Berson, A. et al. 2000. *Building Data Mining Application for CRM*. USA : Mc Graw-Hill
- Connolly, T. & Carolyn B. 2002. *Database System A Practical Approach to Design, Implementation, and Management*. England: Addison-Wesley.
- Fayyad U. N., Piatetsky-Shapiro G., Smyth P. & Uthurusamy R. 1996. *Advances in Knowledge Discovery and Data Mining*. California : AAAI/MIT Press
- Goharian & Grossman. 2003. *Introduction to DataMining*. <http://ir.iit.edu/~nazli/cs422/CS422-Slides/DM-Introduction.pdf> [Desember 2004]
- Han, J. & Kamber, M. 2001. *Data Mining Concepts & Techniques : Slides For Text Book Chapter 1*. <http://www.cs.sfu.ca/~han/dmbook/1intro.pdf> [Desember 2004]
- Han, J. & Kamber, M. 2001. *Data Mining Concepts & Techniques*. USA : Academic Press
- Han J., Cai Y. & Cercone N. 1992. *Knowledge Discovery in Database : An attribute-Oriented Approach*. Hasil VLDB'92, Canada. Hal : 547-559
- Two Crows Corporation. 1999. *Introduction to Data Mining and Knowledge Discovery : Third Edition*. <http://www.twocrows.com/intro-dm.pdf> [Desember 2004]
- Zaine, R.O. 1999. *Principles Of Knowledge Discovery in Database Chapter1 : Introduction to Data Mining*. <http://www.cs.ualberta.ca/~zaiane/courses/comput690/slides/ch1s.pdf>
- [Desember 2004]

Lampiran 1 Struktur menu pada aplikasi *data mining*

Lampiran 2 Identifikasi atribut pada tabel gabungan seluruhnya

No	Atribut	Tipe	Panjang	Keterangan
1.	status	nvarchar	1	Atribut yang menerangkan status pelamar. 1=diterima, 0=tidak diterima.
2.	tahun	nvarchar	2	Atribut yang menyatakan tahun melamar menjadi mahasiswa IPB.
3.	propinsi	nvarchar	2	Atribut yang berisi keterangan asal propinsi pelamar. Terdiri dari 27 nilai yang berbeda untuk masing-masing propinsi yaitu 11, 12, 13, 14, 15, 16, 17, 18, 31, 32, 33, 34, 35, 51, 52, 53, 61, 62, 63, 64, 71, 72, 73, 74, 81, 82, 99.
4.	nosla	nvarchar	8	Atribut yang berisi nomor smu pelamar.
5.	noinduk	nvarchar	[3, 5]	Atribut yang berisi no induk pelamar di sekolah asal.
6.	nrp	nvarchar	9	Atribut yang berisi nrp pelamar. Hanya terisi untuk pelamar yang diterima.
7.	nrplama	nvarchar	9	Atribut yang berisi nim lama pelamar. Hanya terisi untuk pelamar yang diterima. Dan jika program studi pelamar yang diterima mengalami perubahan nim.
8.	nama	nvarchar	255	Atribut yang berisi nama pelamar.
9.	jeniskel	float	1	Atribut yang berisi kode jenis kelamin pelamar, 1 untuk laki-laki dan 0 untuk perempuan.
10.	thnmasuk	nvarchar	2	Atribut yang berisi tahun masuk sma pelamar.
11.	b1	float	3	Atribut yang berisi nilai biologi pelamar kelas 1 cawu 1.
12.	b2	float	3	Atribut yang berisi nilai biologi pelamar kelas 1 cawu 2.
13.	b3	float	3	Atribut yang berisi nilai biologi pelamar kelas 1 cawu 3.
14.	b4	float	3	Atribut yang berisi nilai biologi pelamar kelas 2 cawu 1.
15.	b5	float	3	Atribut yang berisi nilai biologi pelamar kelas 2 cawu 2.
16.	b6	float	3	Atribut yang berisi nilai biologi pelamar kelas 2 cawu 3.
17.	b7	float	3	Atribut yang berisi nilai biologi pelamar kelas 3 cawu 1.
18.	b8	float	3	Atribut yang berisi jumlah nilai biologi pelamar dari kelas 1 cawu 1 sampai dengan kelas 3 cawu 1... [b1+...+b7].
19.	f1	float	3	Atribut yang berisi nilai fisika pelamar kelas 1 cawu 1.
20.	f2	float	3	Atribut yang berisi nilai fisika pelamar kelas 1 cawu 2.
21.	f3	float	3	Atribut yang berisi nilai fisika pelamar kelas 1 cawu 3.
22.	f4	float	3	Atribut yang berisi nilai fisika pelamar kelas 2 cawu 1.
23.	f5	float	3	Atribut yang berisi nilai fisika pelamar kelas 2 cawu 2.
24.	f6	float	3	Atribut yang berisi nilai fisika pelamar kelas 2 cawu 3.
25.	f7	float	3	Atribut yang berisi nilai fisika pelamar kelas 3 cawu 1.
26.	f8	float	3	Atribut yang berisi jumlah nilai fisika pelamar dari kelas 1 cawu 1 sampai dengan kelas 3 cawu 1... [f1+...+f7].
27.	k1	float	3	Atribut yang berisi nilai kimia pelamar kelas 1 cawu 1.
28.	k2	float	3	Atribut yang berisi nilai kimia pelamar kelas 1 cawu 2.
29.	k3	float	3	Atribut yang berisi nilai kimia pelamar kelas 1 cawu 3.
30.	k4	float	3	Atribut yang berisi nilai kimia pelamar kelas 2 cawu 1.
31.	k5	float	3	Atribut yang berisi nilai kimia pelamar kelas 2 cawu 2.
32.	k6	float	3	Atribut yang berisi nilai kimia pelamar kelas 2 cawu 3.
33.	k7	float	3	Atribut yang berisi nilai kimia pelamar kelas 3 cawu 1.
34.	k8	float	3	Atribut yang berisi jumlah nilai kimia pelamar dari kelas 1 cawu 1 sampai dengan kelas 3 cawu 1... [k1+...+k7].
35.	m1	float	3	Atribut yang berisi nilai matematika pelamar kelas 1 cawu 1.
36.	m2	float	3	Atribut yang berisi nilai matematika pelamar kelas 1 cawu 2.
37.	m3	float	3	Atribut yang berisi nilai matematika pelamar kelas 1 cawu 3.
38.	m4	float	3	Atribut yang berisi nilai matematika pelamar kelas 2 cawu 1.
39.	m5	float	3	Atribut yang berisi nilai matematika pelamar kelas 2 cawu 2.
40.	m6	float	3	Atribut yang berisi nilai matematika pelamar kelas 2 cawu 3.
41.	m7	float	3	Atribut yang berisi nilai matematika pelamar kelas 3 cawu 1.



Lampiran 2 lanjutan

No	Atribut	Tipe	Panjang	Keterangan
42.	m8	float	3	Atribut yang berisi jumlah nilai matematika pelamar dari kelas 1 cawu 1 sampai dengan kelas 3 cawu 1 [$m1+...+m7$].
43.	r1	float	3	Atribut yang berisi peringkat kelas pelamar kelas 1 cawu 1.
44.	r2	float	3	Atribut yang berisi peringkat kelas pelamar kelas 1 cawu 2.
45.	r3	float	3	Atribut yang berisi peringkat kelas pelamar kelas 1 cawu 3.
46.	r4	float	3	Atribut yang berisi peringkat kelas pelamar kelas 2 cawu 1.
47.	r5	float	3	Atribut yang berisi peringkat kelas pelamar kelas 2 cawu 2.
48.	r6	float	3	Atribut yang berisi peringkat kelas pelamar kelas 2 cawu 3.
49.	r7	float	3	Atribut yang berisi peringkat kelas pelamar kelas 3 cawu 1.
50.	j1	float	3	Atribut yang berisi jumlah siswa dalam kelas pelamar saat kelas 1 cawu 1.
51.	j2	float	3	Atribut yang berisi jumlah siswa dalam kelas pelamar saat kelas 1 cawu 2.
52.	j3	float	3	Atribut yang berisi jumlah siswa dalam kelas pelamar saat kelas 1 cawu 3.
53.	j4	float	3	Atribut yang berisi jumlah siswa dalam kelas pelamar saat kelas 2 cawu 1.
54.	j5	float	3	Atribut yang berisi jumlah siswa dalam kelas pelamar saat kelas 2 cawu 2.
55.	j6	float	3	Atribut yang berisi jumlah siswa dalam kelas pelamar saat kelas 2 cawu 3.
56.	j7	float	3	Atribut yang berisi jumlah siswa dalam kelas pelamar saat kelas 3 cawu 1.
57.	ra1	float	3	Atribut yang berisi peringkat sekolah pelamar kelas 1 cawu 1.
58.	ra2	float	3	Atribut yang berisi peringkat sekolah pelamar kelas 1 cawu 2.
59.	ra3	float	3	Atribut yang berisi peringkat sekolah pelamar kelas 1 cawu 3.
60.	ra4	float	3	Atribut yang berisi peringkat sekolah pelamar kelas 2 cawu 1.
61.	ra5	float	3	Atribut yang berisi peringkat sekolah pelamar kelas 2 cawu 2.
62.	ra6	float	3	Atribut yang berisi peringkat sekolah pelamar kelas 2 cawu 3.
63.	ra7	float	3	Atribut yang berisi peringkat sekolah pelamar kelas 3 cawu 1.
64.	ja1	float	3	Atribut yang berisi jumlah siswa satu angkatan di sekolah pelamar pada kelas 1 cawu 1.
65.	ja2	float	3	Atribut yang berisi jumlah siswa satu angkatan di sekolah pelamar pada kelas 1 cawu 2
66.	ja3	float	3	Atribut yang berisi jumlah siswa satu angkatan di sekolah pelamar pada kelas 1 cawu 3.
67.	ja4	float	3	Atribut yang berisi jumlah siswa satu angkatan di sekolah pelamar pada kelas 2 cawu 1.
68.	ja5	float	3	Atribut yang berisi jumlah siswa satu angkatan di sekolah pelamar pada kelas 2 cawu 2.
69.	ja6	float	3	Atribut yang berisi jumlah siswa satu angkatan di sekolah pelamar pada kelas 2 cawu 3.
70.	ja7	float	3	Atribut yang berisi jumlah siswa satu angkatan di sekolah pelamar pada kelas 3 cawu 1.
71.	Umum	float	4	Atribut yang berisi jumlah nilai biologi, matematika, fisika, dan kimia [$b8+k8+f8+m8$].



Lampiran 2 lanjutan

No	Atribut	Tipe	Panjang	Keterangan
72.	Umum1	float	4	Atribut yang berisi jumlah nilai biologi, matematika, fisika, dan kimia dari kelas 1 cawu 1 sampai kelas 3 cawu 1 [b8+k8+f8+m8] dengan pembulatan kebawah (sampai puluhan).
73.	klssiswa	float	2	Atribut yang berisi urutan kelas pelamar di kelas 3.
74.	jumlahkls	float	2	Atribut yang berisi jumlah kelas 3 di sekolah pelamar.
75.	noklspres	nvarchar	1	Atribut yang berisi keterangan apakah urutan kelas 3 IPA didasarkan pada prestasi siswanya.
76.	kegiatan	nvarchar		Atribut yang berisi kegiatan pelamar.
77.	minat	float		Atribut yang berisi minat pelamar terhadap mata pelajaran di smu.
78.	prestasi	nvarchar		Atribut yang berisi prestasi ekstrakurikuler pelamar selama di smu.
79.	kbpt	nvarchar	5	Atribut yang berisi penilaian sekolah terhadap pelamar. Char 1 : Prestasi akademik di SMU. Char 2 : Dugaan keberhasilan di IPB. Char 3 : Kepemimpinan. Char 4 : Daya kerja sama kelompok. Char 5 : sopan santun. (C= cemerlang, B= baik, S= sedang, K=kurang)
80.	nopindah	nvarchar	8	Atribut yang berisi kode sekolah lama pelamar. Terisi hanya untuk pelamar yang pernah pindah sekolah.
81.	tidaknaik	nvarchar	1	Atribut yang berisi berapa kali pelamar pernah tidak naik kelas.
82.	pilihan1	nvarchar	3	Atribut yang berisi kode program studi pilihan pertama pelamar.
83.	pilihan2	nvarchar	3	Atribut yang berisi kode program studi pilihan kedua pelamar.
84.	pilihan3	nvarchar	3	Atribut yang berisi kode program studi pilihan ketiga pelamar. Pada tahun 2002 sampai tahun 2004 atribut ini berisi nilai null semua (karena pada tahun ini, pelamar hanya diberikan memilih dua pilihan program studi).
85.	pilihan4	nvarchar	3	Atribut yang berisi kode program studi pilihan keempat pelamar. Pada tahun 2000 sampai tahun 2004 atribut ini berisi nilai null semua (karena pada tahun ini, pelamar hanya diberikan memilih dua atau tiga pilihan program studi).
86.	pilihan5	nvarchar	3	Atribut yang berisi kode program studi pilihan kelima pelamar. Pada tahun 2000 sampai tahun 2004 atribut ini berisi nilai null semua (karena pada tahun ini, pelamar hanya diberikan memilih dua atau tiga pilihan program studi).
87.	putusan1	nvarchar	3	Atribut yang berisi kode program studi hasil keputusan pertama pelamar yang diterima. Terisi hanya untuk pelamar yang diterima.
88.	putusan2	nvarchar		Atribut yang berisi kode program studi hasil keputusan kedua pelamar yang diterima.
89.	putusan3	nvarchar		Atribut yang berisi kode program studi hasil keputusan ketiga pelamar yang diterima.
90.	listrik	float	1	Atribut yang berisi kode daya listrik yang digunakan pelamar di rumah orang tua/wali. 0=Tidak mengisi keterangan ini 1=Tidak ada 3=900 watt 4=1300 watt 2=450 watt 5=2200 watt 6=lebih dari 2200 watt



Lampiran 2 Lanjutan

No	Atribut	Tipe	Panjang	Keterangan
91.	datagab	float	1	Semua bernilai null.
92.	kelengkap a	nvarchar	2	Atribut yang berisi nama form yang tidak dibawa pelamar pada saat pendaftaran. Bernilai null jika pelamar membawa seluruh form pendaftaran yang diminta.
93.	hadir	nvarchar	1	Atribut yang berisi status kehadiran pelamar pada saat pendaftaran. Null dan Y dan 1 untuk yang hadir, dan T atau 0 untuk yang tidak hadir.
94.	sumberinf	nvarchar	9	Atribut yang berisi sumber informasi pelamar memilih program studi di IPB. 1= buku panduan IPB 2= Informasi program studi 3= mahasiswa IPB 4= Alumni IPB 5= Dosen IPB 6= Guru 7= Orangtua/Keluarga 8= Pameran pendidikan 9= Lainnya
95.	sumberinf 2	nvarchar	255	Atribut yang berisi sumber informasi memilih program studi lainnya yang disebutkan pelamar yang tidak terdapat pada pilihan.
96.	alasanpil	nvarchar	10	Atribut yang berisi alasan pelamar memilih IPB sebagai tempat untuk melanjutkan studinya. 1= Fasilitas pendidikan terbaik di Indonesia 2= Ingin jadi ilmuwan/dosen 3= Ingin mengembangkan sektor pertanian di Indonesia 4= Mudah menjadi pegawai negeri 5= Lowongan kerja di sektor swasta banyak tersedia di kota besar 6= Lulusannya masih banyak diperlukan 7= Sebagai modal berwiraswasta 8= Menunjang usaha orang tua/keluarga 9= Saran keluarga/orang tua/guru 0= lainnya
97.	alasanpil2	nvarchar	255	Atribut yang berisi alasan pilihan memilih IPB lainnya yang disebutkan pelamar yang tidak terdapat pada pilihan.
98.	sumberbia	nvarchar	1	Atribut yang berisi sumber biaya pendidikan dan biaya hidup pelamar. 1= Orang tua 2= Orang tua asuh/wali 3= Beasiswa
99.	sumberbia 2	nvarchar	255	Atribut yang berisi sumber biaya pendidikan dan hidup jika pelamar memilih beasiswa, berupa jenis/nama beasiswa.
100.	biayahidu p	nvarchar	1	Atribut yang berisi perkiraan biaya hidup bulanan yang akan diperoleh pelamar. 1=< Rp. 100.000 2= Rp. 100.000- Rp. 200.000 3= Rp. 201.000- Rp. 300.000 4= Rp. 301.000- Rp. 500.000 5=> Rp. 500.000



Lampiran 2 lanjutan

No	Atribut	Tipe	Panjang	Keterangan
101.	karir	nvarchar	4	Atribut yang berisi pilihan karir yang diinginkan oleh pelamar setelah lulus dari IPB. 1=wiraswastawan 2=pegawai negeri 3=pegawai swasta 4=lainnya
102.	karir2	nvarchar		Atribut yang berisi keterangan tambahan pilihan karir pelamar jika memilih lainnya.
103.	persepsi01	nvarchar	1	Atribut yang berisi pendapat mahasiswa tentang IPB. (- = kurang setuju, 0= sangat setuju, + = sangat setuju, ?=tidak tahu) IPB adalah lembaga pendidikan tinggi unggulan dalam ilmu pertanian di Indonesia.
104.	persepsi02	nvarchar	1	Atribut yang berisi pendapat mahasiswa tentang IPB. (- = kurang setuju, 0= sangat setuju, + = sangat setuju, ?=tidak tahu) Informasi program studi IPB memadai untuk mengisi prioritas pilihan pada lampiran 3 pada berkas pendaftaran.
105.	persepsi03	nvarchar	1	Atribut yang berisi pendapat mahasiswa tentang IPB. (- = kurang setuju, 0= sangat setuju, + = sangat setuju, ?=tidak tahu) Ada kesulitan memahami antar beberapa program studi sekurang-kurangnya menentukan prioritas pilihan.
106.	persepsi04	nvarchar	1	Atribut yang berisi pendapat mahasiswa tentang IPB. (- = kurang setuju, 0= sangat setuju, + = sangat setuju, ?=tidak tahu) IPB dan alumninya berkontribusi besar dalam pembangunan pertanian di Indonesia.
107.	persepsi05	nvarchar	1	Atribut yang berisi pendapat mahasiswa tentang IPB. (- = kurang setuju, 0= sangat setuju, + = sangat setuju, ?=tidak tahu) IPB, alumni, dan mahasiswanya berperan aktif dalam memberdayakan masyarakat pertanian pedesaan.
108.	persepsi06	nvarchar	1	Atribut yang berisi pendapat mahasiswa tentang IPB. (- = kurang setuju, 0= sangat setuju, + = sangat setuju, ?=tidak tahu) Fasilitas pendidikan (laboratorium, perpustakaan, dan kebun percobaan) sangat memadai.
109.	persepsi07	nvarchar	1	Atribut yang berisi pendapat mahasiswa tentang IPB. (- = kurang setuju, 0= sangat setuju, + = sangat setuju, ?=tidak tahu) Ada banyak dosen bereputasi nasional.
110.	persepsi08	nvarchar	1	Atribut yang berisi pendapat mahasiswa tentang IPB. (- = kurang setuju, 0= sangat setuju, + = sangat setuju, ?=tidak tahu) Kegiatan pendidikan berlangsung dengan baik.
111.	persepsi09	nvarchar	1	Atribut yang berisi pendapat mahasiswa tentang IPB. (- = kurang setuju, 0= sangat setuju, + = sangat setuju, ?=tidak tahu) Pada umumnya program sarjana dapat diselesaikan dalam 4-5 tahun.

Lampiran 2 lanjutan

No	Atribut	Tipe	Panjang	Keterangan
112.	persepsi10	nvarchar	1	Atribut yang berisi pendapat mahasiswa tentang IPB. (- = kurang setuju, 0= sangat setuju, +sangat setuju, ?=tidak tahu) Unit kegiatan kemahasiswaan (himpunan profesi, olah raga, dan kesenian) tersedia dengan memadai.
113.	persepsi11	nvarchar	1	Atribut yang berisi pendapat mahasiswa tentang IPB. (- = kurang setuju, 0= sangat setuju, +sangat setuju, ?=tidak tahu) Alumni IPB mudah memperoleh pekerjaan yang layak.
114.	persepsi12	nvarchar	1	Atribut yang berisi pendapat mahasiswa tentang IPB. (- = kurang setuju, 0= sangat setuju, +sangat setuju, ?=tidak tahu) Biaya pendidikan di IPB tergolong murah.
115.	persepsi13	nvarchar	1	Atribut yang berisi pendapat mahasiswa tentang IPB. (- = kurang setuju, 0= sangat setuju, +sangat setuju, ?=tidak tahu) Biaya hidup di Bogor tergolong murah.
116.	seleksi2	nvarchar	3	Atribut yang berisi kesediaan pelamar untuk mengikuti seleksi diploma atau seleksi masuk perguruan tinggi swasta yang bekerja sama dengan IPB jika tidak diterima. 1= psdiploma IPB 2, 3, 4, 5, 6 bervariasi artinya tiap tahun.
117.	psdiploma	nvarchar	5	Atribut yang berisi kode program diploma IPB yang diminati pelamar. Null bagi yang tidak berminat ke diploma.
118.	sandioper a	nvarchar	3	Atribut yang berisi nama singkat operator pengentri data.
119.	hitung			Semua bernilai null.
120.	kodeprop	nvarchar	2	Atribut yang berisi keterangan asal propinsi pelamar. Terdiri dari 27 nilai yang berbeda untuk masing-masing propinsi yaitu 11, 12, 13, 14, 15, 16, 17, 18, 31, 32, 33, 34, 35, 51, 52, 53, 61, 62, 63, 64, 71, 72, 73, 74, 81, 82, 99.
121.	kodepsbaru	nvarchar	3	Atribut yang berisi kode program studi baru .
122.	nim	nvarchar	9	Atribut yang berisi no induk mahasiswa.
123.	slapindah	nvarchar	255	Atribut yang berisi nama SMA tempat pelamar pindah. Terisi hanya jika pelamar pernah pindah SMA.
124.	kodepil	nvarcar	2	Atribut yang berisi kode pilihan program studi untuk pelamar yang diterima. P1, P2, atau P3 sesuai dengan pilihan pelamar.
125.	peringkat	real	4	Atribut yang berisi peringkat pelamar dalam satu sekolah asal pelamar berdasarkan kriteria dari IPB.
126.	seleksi	nvarchar	2	Atribut yang berisi keterangan apakah pelamar diikutkan dalam seleksi selanjutnya hasil dari pembahasan dalam rapat. 1=diteruskan seleksi 0=tidak diteruskan seleksi
127.	urutpang	float	8	Atribut yang berisi nomor urut panggil pelamar yang diterima saat pendaftaran ulang.
128.	tanggal	nvarchar	255	Atribut yang berisi tanggal pelamar yang diterima harus melakukan pendaftaran ulang.
129.	hari	nvarchar	255	Atribut yang berisi hari pelamar yang diterima harus melakukan pendaftaran ulang.

Lampiran 2 lanjutan

No	Atribut	Tipe	Pan-jang	Keterangan
130.	jam	nvarchar	255	Atribut yang berisi jam pelamar yang diterima harus melakukan pendaftaran ulang.
131.	perjanjian	nvarchar	255	Atribut yang berisi perjanjian tertentu yang dilakukan pada saat pelamar melakukan pendaftaran ulang.
132.	statundang	nvarchar	255	Atribut yang berisi status undangan mahasiswa. Merupakan atribut yang dirahasiakan oleh IPB
133.	akselerasi	nvarchar	255	Atribut yang berisi keterangan jika sekolah pelamar mempunyai program kelas akselerasi.
134.	koderang	nvarchar	255	Atribut yang berisi kode peringkat pelamar
135.	rataanr	nvarchar	255	Atribut yang berisi rataan rapot pelamar
136.	jumpeltdk0	nvarchar	255	Atribut yang berisi jumlah mata pelajaran yang diajarkan si sekolah pelamar.
137.	salurkans1	float	8	Semua bernilai null.
138.	akr	nvarchar	255	Atribut yang berisi akronim dari program studi
139.	verifikasi	nvarchar	255	Semua bernilai null.
140.	no	nvarchar	255	Semua bernilai null.
2.	Namasla	nvarchar		Atribut yang berisi nama sekolah pelamar.
3.	T_Akademik	nvarchar	255	Atribut yang berisi tahun akademik data sekolah tersebut.
4.	Usmipel	float	8	Atribut yang berisi jumlah pelamar usmi dari sekolah tersebut
5.	Usmitrm	float	8	Atribut yang berisi jumlah pelamar usmi yang diterima.
6.	usmidtg	float	8	atribut yang berisi jumlah pelamar usmi yang diterima yang mengambil kesempatan menjadi mahasiswa IPB .
7.	umptnpel	float	8	Atribut yang berisi jumlah pelamar umptn dari sekolah tersebut
8.	umptntrm	float	8	Atribut yang berisi jumlah pelamar umptn yang diterima
9.	umptndtg	float	8	Atribut yang berisi jumlah pelamar umptn yang diterima yang mengambil kesempatan menjadi mahasiswa IPB
10.	jumlahhmhs	float	8	Atribut yang berisi jumlah mahasiswa di IPB yang berasal dari sekolah tersebut
11.	nemipa1	float	8	Atribut yang berisi nem ipa minimal sekolah pelamar.
12.	nemipa2	float	8	Atribut yang berisi nem ipa maksimal sekolah pelamar.
13.	nemmin1	float	8	Atribut yang berisi nem minimal pelamar USMI
14.	nemmax1	float	8	Atribut yang berisi nem maksimal pelamar USMI
15.	nemrata1	float	8	Atribut yang berisi rataan dari nemmin1 dengan nemmax1
16.	nemmin2	float	8	Atribut yang berisi nem minimal pelamar UMPTN.
17.	nemmax2	float	8	Atribut yang berisi nem maksimal pelamar UMPTN.
18.	nemrata2	float	8	Atribut yang berisi rataan dari nemmin2 dengan nemmax2.
158.	NMR1	float	8	Atribut yang berisi jumlah mahasiswa dari sekolah tersebut yang memiliki IP TPB ≤ 3.5
159.	NMR2	float	8	Atribut yang berisi jumlah mahasiswa dari sekolah tersebut yang memiliki IP TPB < 3.5 dan ≥ 2.75
160.	NMR3	float	8	Atribut yang berisi jumlah mahasiswa dari sekolah tersebut yang memiliki IP TPB < 2.75
161.	NMR4	float	8	Atribut yang berisi jumlah mahasiswa dari sekolah tersebut yang memiliki IP TPB ≥ 2.1
162.	NMR5	float	8	Atribut yang berisi jumlah mahasiswa dari sekolah tersebut yang memiliki IP TPB < 2.1
163.	Kategori	nvarchar	2	Atribut yang berisi kategori penilaian IPB terhadap sekolah tersebut.

Lampiran 3 Rincian pemrosesan pembersihan data

No	Atribut	Jumlah tuple kosong	Jumlah tuple mengandung noise	Jumlah tuple tidak konsisten	Aksi
1.	status	0	0	0	
2.	tahun	0	0	0	
3.	Propinsi	18642	0	0	diisi dengan 2 nilai terdepan dari nosla
4.	nosla	0	0	0	
5.	noinduk	0	0	0	
6.	nrp	a) 38614 	b) 1 → C0440047 	c) 4936 	a) <i>update</i> = 0 b) <i>update</i> menjadi C04400047 c) <i>update</i> menjadi nrp dengan kodepsbaru
7.	nrplama	a) 51811	0	0	hapus atribut → akan jadi <i>redundancy</i> dengan nrp jika di <i>update</i>
8.	nama	0	0	0	
9.	jeniskel	0	0	0	
10.	thnmasuk	0	1 → 66	0	<i>update</i> menjadi 99
11.	b1	a) 216 	0 	0 	a) hapus tuple
12.	b2				
13.	b3				
14.	b4				
15.	b5				
16.	b6				
17.	b8				
18.	b7	48 (bukan dari tahun 2004)	0	0	hapus tuple
19.	f1	a) 1 	0 	0 	a) hapus tuple
20.	f2				
21.	f3				
22.	f4				
23.	f5				
24.	f6				
25.	f8				
26.	f7	1 (bukan dari 2004)	0	0	hapus tuple
27.	k1	0 	0 	0 	
28.	k2				
29.	k3				
30.	k4				
31.	k5				
32.	k6				
33.	k8				
34.	k7	0 (bukan dari tahun 2004)	0	0	
35.	m1	a) 4 	0 	0 	a) hapus tuple
36.	m2				
37.	m3				
38.	m4				
39.	m5				
40.	m6				
41.	m8				

Lampiran 3 lanjutan

No	Atribut	Jumlah tuple kosong	Jumlah tuple mengandung noise	Jumlah tuple tidak konsisten	Aksi
42.	m7	0	0	0	
43.	r1	13581	0	0	11 diinterpretasikan sebagai tidak mendapat peringkat (10 ke atas).
44.	r2				
45.	r3				
46.	r4				
47.	r5				
48.	r6				
49.	r7	5129 (bukan dari tahun 2004)	0	0	0 diinterpretasikan sebagai tidak mendapat peringkat.
50.	J1	a) 2178 c) 11	b) 2 tuple $j1 < r1$ dan $j1 \neq 0$	0	a) isi dengan nilai rataan berdasarkan nosla yang sama yang tidak 0 b) hapus tuple c) hapus tuple
51.	J2	a) 1916 c) 6	b) 1 tuple $j2 < r2$ dan $j2 \neq 0$	0	a) isi dengan nilai rataan berdasarkan nosla yang sama yang tidak 0 b) hapus tuple c) hapus tuple
52.	J3	a) 1838	b) 2 tuple $j3 < r3$ dan $j3 \neq 0$	0	a) isi dengan nilai rataan berdasarkan nosla yang sama yang tidak 0 b) hapus tuple
53.	J4	a) 1806 c) 4	b) 2 tuple $j4 < r4$ dan $j4 \neq 0$	0	a) isi dengan nilai rataan berdasarkan nosla yang sama yang tidak 0 b) hapus tuple c) hapus tuple
54.	J5	a) 1645 c) 2	b) 1 tuple $j5 < r5$ dan $j5 \neq 0$	0	a) isi dengan nilai rataan berdasarkan nosla yang sama yang tidak 0 b) hapus tuple c) hapus tuple
55.	J6	a) 1576 c) 1	b) 3 tuple $j6 < r6$ dan $j6 \neq 0$	0	a) isi dengan nilai rataan berdasarkan nosla yang sama yang tidak 0 b) hapus tuple c) hapus tuple
56.	J7	a) 1002 (bukan dari tahun 2004)	b) 3 tuple $j7 < r7$ dan $j7 \neq 0$		a) isi dengan nilai rataan berdasarkan nosla yang sama yang tidak 0 b) hapus tuple
57.	ra1		43751 $ra_i = r_i$		hapus atribut
58.	ra2				
59.	ra3				
60.	ra4				
61.	ra5				
62.	ra6				
63.	ra7				
64.	ja1	52360			hapus atribut
65.	ja2				
66.	ja3				
67.	ja4				
68.	ja5				
69.	ja6				
70.	ja7				
71.	Umum	0	0	0	
72.	Umum1	43332	0	0	hapus atribut

Lampiran 3 lanjutam

No	Atribut	Jumlah tuple kosong	Jumlah tuple mengandung noise	Jumlah tuple tidak konsisten	Aksi
73.	klssiswa	a) 2764	b) 6 tuple yang klssiswa>jml kelas		a) update = 0 (tidak diketahui) b) hapus tuple
74.	Jumlahkls	a) 36 b) 3	0	0	a) update dengan rataan jumlahkls berdasarkan nosla b) hapus tuple
75.	noklspres	52360	0	0	hapus atribut
76.	kegiatan	34403	0	0	hapus atribut karena data lengkapnya terdapat pada basis data biodata
77.	minat	52360	0	0	hapus atribut
78.	prestasi	52360	0	0	hapus atribut
79.	kbpt	a) 1636 b) 238 tidak 5 karakter			a) hapus tuple b) hapus tuple
80.	nopindah	a) 46272	b) 4 (berisi 1, 4) c) 1 (berisi 0)	0	a) update jadi=0 yang mengindikasikan tidak pernah pindah b) update = 1 (pernah pindah tapi tidak diketahui nosla pindahnya) c) update = 0 (tidak pernah pindah)
81.	tidaknaik	a) 47643	0	0	a) update jadi 0 (tidak pernah tidak naik)
82.	pilihan1	a) 85 f) 81	c) 77 null semua d) 8909	b)23370	a) update=0 (tidak memilih) b) update dengan kode ps baru f) hapus tuple
83.	pilihan2	a)1135	pilihan3=IPB	c)23370	a) update jadi 0 →tidak memilih c) update dengan kode ps baru
84.	pilihan3	a) 16514		e)23370	a) update jadi 0 →tidak milih d) update jadi 0 →tidak memilih e) update dengan kode ps baru
85.	pilihan4	52360	0	0	hapus atribut
86.	pilihan5	52360	0	0	hapus atribut
87.	putusan1	a) 34915 b) 496 (nrp ada tapi putusan1 tidak ada)	0	6773	a) update jadi 0 (tidak diterima) b) isi dengan 3 digit nrp
88.	putusan2	52360	0	0	hapus atribut
89.	putusan3	52360	0	0	hapus atribut
90.	listrik	341	0	0	direpresentasikan sebagai pelamar tidak mengisi keterangan listrik
91.	datagab	52360	0	0	hapus atribut
92.	kelengkapa	47465	0	0	update = 0 (lengkap)
93.	hadir	a) 46214	b) 1525	0	a) update = 1 (hadir) b) update = 0 (tidak hadir)
94.	sumberinf	a) 171	b) 2 (sumberinf kosong tapi sumberinf2 ada)	c) 27	a) update = 0 sebagai tidak mempunyai sumber informasi b) pindahkan sumberinf2 ke sumberinf c) update jadi 9 (alasan lain)
95.	sumberinf2	45330	0	0	update = 0



Lampiran 3 lanjutan

No	Atribut	Jumlah tuple kosong	Jumlah tuple mengandung noise	Jumlah tuple tidak konsisten	Aksi
96.	alasanpil	a) 243	b) 3 → ada pada kolom alasanpil	0	a) update = T (tidak tahu) b) pindahkan ke alasanpil
97.	alasanpil 2	a) 44658	b) 17 (alasanpil=0 tapi alasanpil2 kosong)	0	a) update = 0 (tidak memiliki alas an lain) b) update = 1
98.	sumberbi a	a) 105	b) 2 tuple (=4) 1 tuple (=8)	0	a) update = T (sumberbia tidak ada dalam pilihan) b) hapus tuple
99.	sumberbi a2	a) 47101	b) 8 (sumberbia tidak ada tapi sumberbia2 ada)	0	a) update = 0 (tidak ada sumberbia dari beasiswa) b) pindahkan ke sumberbia
100.	biayahid up	332	0	0	update = 0
101.	karir	a) 32861	b) 1 → 1118 c) 1 → 7127		a) update = 0 (tidak tahu) b) update = 1 c) update = 12
102.	karir2	46572	0	0	update jadi =0 (tidak ada karir lain)
103.	persepsi0 1	20586	0	0	update menjadi T (tidak mengisi)
104.	persepsi0 2				
105.	persepsi0 3				
106.	persepsi0 4				
107.	persepsi0 5				
108.	persepsi0 6				
109.	persepsi0 7				
110.	persepsi0 8				
111.	persepsi0 9				
112.	persepsi1 0				
113.	persepsi1 1				
114.	persepsi1 2				
115.	persepsi1 3				



Lampiran 3 lanjutan

No	Atribut	Jumlah tuple kosong	Jumlah tuple mengandung noise	Jumlah tuple tidak konsisten	Aksi
116.	seleksi2	a)29219	b) 4934 tanpa pilihan '1' c) 4352 (ada pilihan1 dan pilihan selain 1) d) 1 → 11 e) 5 (ada pilihan psdiploma tapi tidak mengisi 1 pada seleksi2)	0	a) update = 0 sebagai tidak mau diseleksi2 b) update = Y (bersedia ke universitas lain) c) update menjadi 1Y (psdiploma dan universitas lain) d) update = 1 e) update = 1
117.	psdiploma	a) 213 ada pilihan no 1 di seleksi 2 tapi tidak memilih psdiploma b) 34125	c) 2 → G022	0	a) update = 1 tidak punya pilihan psdiploma tapi bersedia ikut seleksi psdiploma b) update = 0 (tidak memilih psdiploma pada seleksi2) c) update → G023
118.	sandiopera	4330	0	0	update = NN (tidak ada nama)
119.	hitung	43636	0	0	hapus atribut
120.	kodeprop	34981	0	0	update sama dengan propinsi
121.	kodepsbaru	50159	0		hapus atribut
122.	nim	50159	0	0	update sama dengan nrp
123.	slapindah	52160	0	0	hapus atribut
124.	kodepil	a) 944 b) 496	0	0	a) update = P2 (karena putusan1=pilihan2) b) update = 1 (diterima tapi tidak pada ketiga pilihan)
125.	peringkat	16197	0	0	update dengan nilai umum sebagai acuan persekolah pertahun, data diurut berdasarkan umum menurun. Yang paling tinggi nilai umum diberi nilai peringkat 1, 2, dst
126.	seleksi	18285	0	0	update = 0 jika tidak punya nrp, dan 1 jika punya nrp
127.	urutpang	52360	0	0	hapus atribut
128.	tanggal	37303	0	0	hapus atribut
129.	hari	50033	0	0	hapus atribut
130.	jam	50033	0	0	hapus atribut
131.	perjanjian	52269	0	0	hapus atribut
132.	statundangan	51780	0	0	hapus atribut (rahasia ppmb)
133.	akselerasi	52360	0	0	hapus atribut



Lampiran 3 lanjutan

No	Atribut	Jumlah tuple kosong	Jumlah tuple mengandung noise	Jumlah tuple tidak konsisten	Aksi
134.	rataanr	44588	0	0	hapus atribut
135.	koderang	44301	0	0	hapus atribut
136.	jumpeltd k0	44301	0	0	hapus atribut
137.	salurkans 1	42585	0	0	hapus atribut
138.	akr	48721	0	0	hapus atribut
139.	verifikasi	43868	0	0	hapus atribut
140.	no	52360	0	0	hapus atribut
141.	Namasla	14	0	0	update = XXX
142.	T_Akade mik	14	0	0	update berdasarkan tahun
143.	Usmipel	1503	0	0	update = 0
144.	Usmitrm	1503	0	0	update = 0
145.	usmidtg	1503	0	0	update = 0
146.	umptnpel				hapus atribut karena tidak ada kaitannya dalam tujuan data mining dalam penelitian.
147.	umptntr m				
148.	umptndt g				
149.	jumlahhm hs	2761	0	0	hapus tuple
150.	nemipa1	3155	0	0	update dengan nilai rataan dari satu propinsi asal SMU pada tahun yang sama
151.	nemipa2		0	0	
152.	nemmin1	8391	0	0	update dengan nilai rataan dari satu propinsi asal SMU pada tahun yang sama
153.	nemmax 1				
154.	nemrata1				
155.	nemmin2				hapus atribut karena tidak ada kaitannya dalam tujuan data mining dalam penelitian.
156.	nemmax 2				
157.	nemrata2				
158.	NMR1	36318	0	0	update = 0
159.	NMR2	21573	0	0	update = 0
160.	NMR3	13854	0	0	update = 0
161.	NMR4	11349	0	0	update = 0
162.	NMR5	30294	0	0	update = 0
163.	Kategori	8851	0	0	update dengan nilai rataan dari satu sekolah asal SMU yang sama— sebelumnya data kategori dikonversi menjadi angka A+=12 B+=9 C+=6 D+=3 A=11 B=8 C=5 D=2 A=-10 B=-7 C=-4 D=-1

Lampiran 4 Identifikasi Jumlah beda pada setiap atribut dalam *task-relevant data*

No.	Atribut	Jumlah nilai beda	No.	Atribut	Jumlah nilai beda
1.	status	2	54.	j6	64
2.	tahun	5	55.	j7	64
3.	Propinsi	27	56.	Umum	266
4.	nosla	1529	57.	klssiswa	12
5.	noinduk	13431	58.	jumlahkls	18
6.	nrp	10328	59.	nopindah	754
7.	nama	38344	60.	tidaknaik	4
8.	jeniskel	2	61.	pilihan1	38
9.	thnmasuk	8	62.	pilihan2	40
10.	b1	10	63.	pilihan3	39
11.	b2	8	64.	putusan1	39
12.	b3	11	65.	listrik	7
13.	b4	14	66.	kelengkapa	32
14.	b5	11	67.	hadir	2
15.	b6	11	68.	sumberinf2	45331
16.	b7	11	69.	sumberbia	9
17.	b8	75	70.	sumberbia2	217
18.	f1	10	71.	biayahidup	6
19.	f2	12	72.	karir2	121
20.	f3	14	73.	alasanpil2	44659
21.	f4	12	74.	persepsi01	5
22.	f5	14	75.	persepsi02	5
23.	f6	13	76.	persepsi03	5
24.	f7	10	77.	persepsi04	5
25.	f8	73	78.	persepsi05	5
26.	k1	12	79.	persepsi06	5
27.	k2	13	80.	persepsi07	5
28.	k3	8	81.	persepsi08	5
29.	k4	14	82.	persepsi09	5
30.	k5	16	83.	persepsi10	5
31.	k6	11	84.	persepsi11	5
32.	k7	9	85.	persepsi12	5
33.	k8	76	86.	persepsi13	5
34.	m1	9	87.	seleksi2	4
35.	m2	9	88.	psdiploma	41
36.	m3	9	89.	sandiopera	16
37.	m4	10	90.	kodepil	5
38.	m5	11	91.	peringkat	172
39.	m6	12	92.	seleksi	2
40.	m7	9	93.	Namasla	2438
41.	m8	74	94.	Usmipel	86
42.	r1	47	95.	Usmitrm	51
43.	r2	46	96.	usmidtg	45
44.	r3	46	97.	jumlahmhs	50
45.	r4	50	98.	nemipa1	2487
46.	r5	46	99.	nemipa2	2482
47.	r6	47	100.	nemmin1	2036
48.	r7	47	101.	nemmax1	2105
49.	j1	64	102.	nemrata1	2534
50.	j2	62	103.	NMR1	14
51.	j3	66	104.	NMR2	28
52.	j4	65	105.	NMR3	33
53.	j5	59	106.	NMR4	45



Lampiran 4 lanjutan

No.	Atribut	Jumlah nilai beda	No.	Atribut	Jumlah nilai beda
107.	NMR5	13	122.	sumberinf X9	2
108.	Kategori	10	123.	alasanpilX1	2
109.	KbptX1	5	124.	alasanpilX2	2
110.	KbptX2	5	125.	alasanpilX3	2
111.	KbptX3	5	126.	alasanpilX4	2
112.	KbptX4	5	127.	alasanpilX5	2
113.	KbptX5	5	128.	alasanpilX6	2
114.	sumberinfX1	2	129.	alasanpilX7	2
115.	sumberinf X2	2	130.	alasanpilX8	2
116.	sumberinf X3	2	131.	alasanpilX9	2
117.	sumberinf X4	2	132.	alasanpilX10	2
118.	sumberinf X5	2	133.	karirX1	2
119.	sumberinf X6	2	134.	karirX2	2
120.	sumberinf X7	2	135.	karirX3	2
121.	sumberinf X8	2	136.	karirX4	2

Lampiran 5 Hasil rinci dari pembuangan atribut (*atribut removal*) serta generalisasi data

No.	Aksi	Atribut
1.	Tetap	status, tahun, jeniskel, thnmasuk, b1, b2, b3, b4, b5, b6, b7, f1, f2, f3, f4, f5, f6, f7, k1, k2, k3, k4, k5, k6, k7, m1, m2, m3, m4, m5, m6, m7, tidaknaik, listrik, hadir, sumberbia, biayahidup, persepsi1, persepsi2, persepsi3, persepsi4, persepsi5, persepsi6, persepsi7, persepsi8, persepsi9, persepsi10, persepsi11, persepsi12, persepsi13, seleksi2, kodepil, seleksi, Kategori, KbptX1, KbptX2, KbptX3, KbptX4, KbptX5, alasanpilX1, alasanpilX2, alasanpilX3, alasanpilX4, alasanpilX5, alasanpilX6, alasanpilX7, alasanpilX8, alasanpilX9, alasanpilX10, sumberinfX1, sumbeinfX2, sumberinfX3, sumberinfX4, sumberinfX5, sumberinfX6, sumberinfX7, sumberinfX8, sumberinfX9, karirX1, karirX2, karirX3, karirX4
2.	Dibuang	nosla, noinduk, npn, nama, b8, f8, m8, k8, j1, j2, j3, j4, j5, j6, j7, umum, klssiswa, jumlahkl, kelengkapa, sumberbia2, karir2, alasanpil2, sandiopera, kodepsbaru, peringkat, nopindah, Namasla, usmidtg, usmitrm, usmipel, jumlahmhhs, nemipa1, nemipa2, nemmin1, nemmax1, nemrata1, NMR1, NMR2, NMR3, NMR4, NMR5
3.	Generalisasi	Propinsi, r1, r2, r3, r4, r5, r6, r7, pilihan1, pilihan2, pilihan3, putusan1, psdiploma

Lampiran 6 Rincian hasil perhitungan *expected information*, *entropy*, dan nilai *information gain* untuk tiap-tiap atribut dalam *task-relevant data* berdasarkan fakultas yang diminati pelamar pada pilihan pertama

$$\text{Expected information} = I(A,B,C,D,E,F,G,H) = 2.55005$$

No.	Atribut	Entropy	Information gain
1.	PUTUSAN1	2.11681	0.43324
2.	PILIHAN2	2.42795	0.1221
3.	PSDIPLOMA	2.43134	0.11871
4.	PILIHAN3	2.51586	0.03419
5.	KODEPIL	2.51702	0.03303
6.	ALASANPILX2	2.51921	0.03084



Lampiran 6 lanjutan

No.	Atribut	<i>Entropy</i>	<i>Information gain</i>
7.	THNMASUK	2.52299	0.02706
8.	tahun	2.52327	0.02678
9.	JENISKEL	2.5246	0.02545
10.	ALASANPILX6	2.53154	0.01851
11.	M5	2.53486	0.01519
12.	PROPINSI	2.5357	0.01435
13.	KATEGORI	2.53596	0.01409
14.	M6	2.5361	0.01395
15.	K3	2.53613	0.01392
16.	K5	2.53621	0.01384
17.	R7	2.53673	0.01332
18.	K2	2.53691	0.01314
19.	K6	2.53705	0.013
20.	M4	2.53714	0.01291
21.	K4	2.53735	0.0127
22.	M7	2.53756	0.01249
23.	SELEKSI2	2.5376	0.01245
24.	M3	2.53761	0.01244
25.	K7	2.53806	0.01199
26.	KARIRX2	2.53811	0.01194
27.	ALASANPILX4	2.53828	0.01177
28.	KARIRX1	2.53873	0.01132
29.	K1	2.53914	0.01091
30.	M2	2.53948	0.01057
31.	F7	2.53951	0.01054
32.	M1	2.53954	0.01051
33.	F5	2.54024	0.00981
34.	R5	2.54033	0.00972
35.	R6	2.54075	0.0093
36.	B7	2.54087	0.00918
37.	F4	2.5413	0.00875
38.	R4	2.54134	0.00871
39.	R3	2.54162	0.00843
40.	F6	2.5417	0.00835
41.	F3	2.54185	0.0082
42.	R2	2.54258	0.00747
43.	F2	2.54259	0.00746
44.	R1	2.54291	0.00714
45.	LISTRİK	2.54292	0.00713
46.	KBPTX1	2.54312	0.00693
47.	KARIRX3	2.54314	0.00691
48.	B1	2.54349	0.00656
49.	B5	2.54359	0.00646
50.	B6	2.54386	0.00619
51.	F1	2.54398	0.00607
52.	B3	2.54431	0.00574
53.	KBPTX2	2.54445	0.0056
54.	B2	2.54456	0.00549
55.	B4	2.54463	0.00542
56.	ALASANPILX1	2.54622	0.00383
57.	BIAYAHIDUP	2.54629	0.00376
58.	status	2.54652	0.00353



Lampiran 6 lanjutan

No.	Atribut	Entropy	Information gain
59.	SELEKSI	2.54652	0.00353
60.	ALASANPILX3	2.54654	0.00351
61.	ALASANPILX7	2.54658	0.00347
62.	SUMBERINFX3	2.54711	0.00294
63.	PERSEPSI05	2.54735	0.0027
64.	PERSEPSI04	2.54745	0.0026
65.	PERSEPSI10	2.54748	0.00257
66.	PERSEPSI11	2.54752	0.00253
67.	SUMBERINFX7	2.54766	0.00239
68.	PERSEPSI03	2.5477	0.00235
69.	SUMBERINFX6	2.5477	0.00235
70.	PERSEPSI02	2.54774	0.00231
71.	PERSEPSI06	2.54803	0.00202
72.	PERSEPSI01	2.54817	0.00188
73.	ALASANPILX8	2.54818	0.00187
74.	SUMBERINFX4	2.54825	0.0018
75.	PERSEPSI07	2.54833	0.00172
76.	PERSEPSI08	2.54845	0.0016
77.	ALASANPILX9	2.54845	0.0016
78.	PERSEPSI12	2.54857	0.00148
79.	SUMBERINFX1	2.54858	0.00147
80.	SUMBERBIA	2.54872	0.00133
81.	KARIRX4	2.54875	0.0013
82.	KBPTX3	2.54885	0.0012
83.	SUMBERINFX5	2.54893	0.00112
84.	PERSEPSI13	2.54897	0.00108
85.	KBPTX4	2.54904	0.00101
86.	KBPTX5	2.54923	0.00082
87.	ALASANPILX5	2.54923	0.00082
88.	PERSEPSI09	2.54926	0.00079
89.	HADIR	2.54945	0.0006
90.	SUMBERINFX9	2.54946	0.00059
91.	SUMBERINFX2	2.54956	0.00049
92.	SUMBERINFX8	2.54983	0.00022
93.	TIDAKNAIK	2.54991	0.00014
94.	ALASANPILX10	2.54992	0.00013

Lampiran 7 Rincian hasil perhitungan *expected information*, *entropy*, dan nilai *information gain* untuk tiap-tiap atribut dalam *task-relevant data* berdasarkan pulau asal pelamar

$$\text{Expected information} = I(1,3,5,6,7,8,9) = 1.0966$$

No.	Atribut	Entropy	Information gain
1.	KATEGORI	0.86558	0.23102
2.	BIAyahidup	1.02908	0.06752
3.	B5	1.05196	0.04464
4.	B6	1.05638	0.04022
5.	F4	1.05737	0.03923
6.	F5	1.05822	0.03838
7.	F6	1.05892	0.03768
8.	B2	1.06234	0.03426
9.	F2	1.0636	0.033
10.	B3	1.06611	0.03049



Lampiran 7 lanjutan

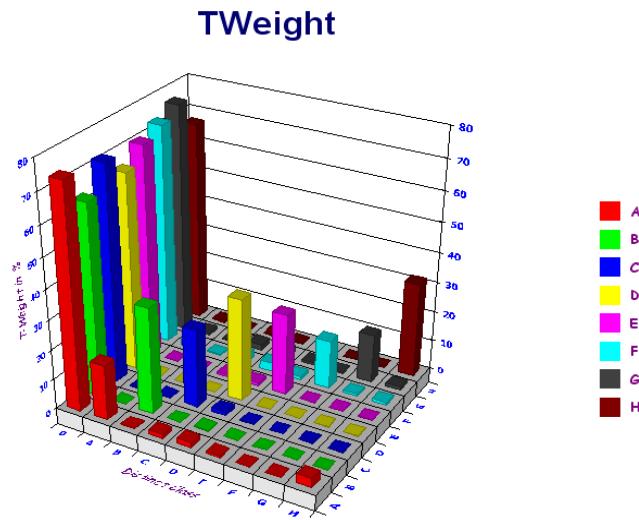
No.	Atribut	Entropy	Information gain
11.	K5	1.06683	0.02977
12.	F3	1.06712	0.02948
13.	K6	1.06774	0.02886
14.	B4	1.06776	0.02884
15.	F1	1.06786	0.02874
16.	M4	1.06889	0.02771
17.	M5	1.06919	0.02741
18.	K4	1.06952	0.02708
19.	SELEKSI2	1.07168	0.02492
20.	F7	1.07181	0.02479
21.	M6	1.07231	0.02429
22.	B1	1.07331	0.02329
23.	KBPTX5	1.07438	0.02222
24.	M2	1.0752	0.0214
25.	B7	1.0758	0.0208
26.	PERSEPSI13	1.07664	0.01996
27.	M3	1.07838	0.01822
28.	K3	1.07917	0.01743
29.	KBPTX4	1.07922	0.01738
30.	K2	1.07927	0.01733
31.	K7	1.08031	0.01629
32.	R7	1.08037	0.01623
33.	M1	1.08062	0.01598
34.	M7	1.0813	0.0153
35.	PILIHAN1	1.08225	0.01435
36.	K1	1.08319	0.01341
37.	LISTRIK	1.08361	0.01299
38.	R6	1.08389	0.01271
39.	KBPTX3	1.08476	0.01184
40.	ALASANPILX1	1.08529	0.01131
41.	PILIHAN2	1.08706	0.00954
42.	R5	1.08713	0.00947
43.	PERSEPSI11	1.0873	0.0093
44.	R4	1.08763	0.00897
45.	KBPTX2	1.08796	0.00864
46.	PUTUSAN1	1.0881	0.0085
47.	PERSEPSI08	1.08816	0.00844
48.	R3	1.08828	0.00832
49.	R1	1.08837	0.00823
50.	PERSEPSI03	1.0885	0.0081
51.	SUMBERINFX3	1.08913	0.00747
52.	KBPTX1	1.08952	0.00708
53.	R2	1.08961	0.00699
54.	PERSEPSI02	1.09007	0.00653
55.	PSDIPLOMA	1.09058	0.00602
56.	PERSEPSI01	1.0908	0.0058
57.	PERSEPSI04	1.09098	0.00562
58.	PERSEPSI05	1.09108	0.00552
59.	PERSEPSI06	1.09112	0.00548
60.	KODEPIL	1.0916	0.005
61.	PERSEPSI07	1.09216	0.00444
62.	PILIHAN3	1.09217	0.00443
63.	status	1.0924	0.0042



Lampiran 7 lanjutan

No.	Atribut	Entropy	Information gain
64.	SELEKSI	1.0924	0.0042
65.	PERSEPSI10	1.09248	0.00412
66.	ALASANPILX2	1.09301	0.00359
67.	SUMBERINFIX1	1.09329	0.00331
68.	SUMBERINFIX2	1.09379	0.00281
69.	SUMBERINFIX5	1.09403	0.00257
70.	ALASANPILX4	1.09408	0.00252
71.	ALASANPILX5	1.09415	0.00245
72.	SUMBERINFIX4	1.09422	0.00238
73.	THNMASUK	1.09447	0.00213
74.	PERSEPSI09	1.09466	0.00194
75.	PERSEPSI12	1.0947	0.0019
76.	ALASANPILX3	1.0947	0.0019
77.	ALASANPILX7	1.09498	0.00162
78.	SUMBERBIA	1.09502	0.00158
79.	KARIRX1	1.09503	0.00157
80.	tahun	1.09526	0.00134
81.	SUMBERINFIX6	1.09548	0.00112
82.	ALASANPILX8	1.09553	0.00107
83.	KARIRX2	1.0956	0.001
84.	HADIR	1.09573	0.00087
85.	SUMBERINFIX8	1.09574	0.00086
86.	JENISKEL	1.09591	0.00069
87.	SUMBERINFIX7	1.09591	0.00069
88.	KARIRX4	1.09594	0.00066
89.	SUMBERINFIX9	1.09609	0.00051
90.	ALASANPILX9	1.09613	0.00047
91.	ALASANPILX6	1.09614	0.00046
92.	KARIRX3	1.09623	0.00037
93.	ALASANPILX10	1.09649	0.00011

Lampiran 8 Hasil presentasi grafik untuk atribut putusan1 berdasarkan kelas target fakultas pilihan pertama pelamar



No.	Target Kelas		A		B		C		D		E		F		G		H	
	Atribut	Nilai beda	Count	t-weight (%)														
1.	tahun	00	4774	35.1624	574	32.1929	1420	32.9926	380	27.2597	1338	39.8808	3150	32.7818	3232	27.2146	0	0
		01	2572	18.9438	277	15.5356	579	13.4526	207	14.8494	646	19.2548	1693	17.6189	2112	17.7838	298	20.8537
		02	2315	17.0509	301	16.8817	914	21.2361	232	16.6428	603	17.9732	1732	18.0248	2194	18.4742	358	25.0525
		03	2017	14.856	312	17.4986	723	16.7983	390	27.977	416	12.3994	1594	16.5886	2317	19.5099	388	27.1519
		04	1899	13.9869	319	17.8912	668	15.5204	185	13.2712	352	10.4918	1440	14.986	2021	17.0175	385	26.9419
2.	propinsi	1	2544	18.7376	477	26.7527	967	22.4675	284	20.373	823	24.5306	1600	16.6511	1752	14.7524	215	15.0455
		3	10372	76.3939	1058	59.3382	2953	68.6106	1045	74.9641	2261	67.392	7557	78.645	9552	80.4311	1185	82.9251
		5	169	1.2448	101	5.6646	134	3.1134	20	1.4347	101	3.0104	113	1.176	238	2.004	6	0.4199
		6	172	1.2668	38	2.1312	51	1.1849	15	1.076	50	1.4903	129	1.3425	117	0.9852	2	0.14
		7	233	1.7161	66	3.7016	142	3.2993	21	1.5065	77	2.2951	155	1.6131	135	1.1367	13	0.9097
		8	80	0.5892	42	2.3556	56	1.3011	9	0.6456	41	1.2221	45	0.4683	70	0.5894	6	0.4199
		9	7	0.0516	1	0.0561	1	0.0232	0	0	2	0.0596	10	0.1041	12	0.101	2	0.14
3.	jenis kel	0	10021	73.8086	1124	63.0398	2079	48.3039	715	51.2912	1563	46.5872	5754	59.8814	6989	58.8498	985	68.9293
		1	3556	26.1914	659	36.9602	2225	51.6961	679	48.7088	1792	53.4128	3855	40.1186	4887	41.1502	444	31.0707
4.	thnmasuk	0	2014	14.8339	306	17.1621	718	16.6822	384	27.5466	416	12.3994	1586	16.5054	2298	19.3499	0	2014
		1	1890	13.9206	321	18.0034	663	15.4043	190	13.6298	345	10.2832	1433	14.9131	2023	17.0344	1	1890
		2	33	0.2431	4	0.2243	14	0.3253	1	0.0717	8	0.2385	22	0.229	29	0.2442	2	33
		3	9	0.0663	4	0.2243	4	0.0929	1	0.0717	2	0.0596	6	0.0624	9	0.0758	3	9
		96	8	0.0589	2	0.1122	0	0	0	0	0	0	2	0.0208	4	0.0337	96	8
		97	4749	34.9783	570	31.9686	1417	32.9229	380	27.2597	1338	39.8808	3148	32.761	3215	27.0714	97	4749
		98	2567	18.907	277	15.5356	575	13.3597	204	14.6341	643	19.1654	1683	17.5148	2090	17.5985	98	2567
		99	2307	16.992	299	16.7695	913	21.2128	234	16.7862	603	17.9732	1729	17.9935	2208	18.5921	99	2307
		0	10056	74.0664	1136	63.7128	3099	72.0028	912	65.4232	2369	70.611	7046	73.3271	9082	76.4736	956	66.8999
5.	putusandal	A	2515	18.524	2	0.1122	4	0.0929	0	0	5	0.149	178	1.8524	102	0.8589	6	0.4199
		B	26	0.1915	634	35.558	3	0.0697	0	0	2	0.0596	43	0.4475	37	0.3116	0	0
		C	191	1.4068	1	0.0561	1123	26.092	0	0	27	0.8048	248	2.5809	232	1.9535	4	0.2799
		D	233	1.7161	8	0.4487	57	1.3243	478	34.2898	33	0.9836	264	2.7474	239	2.0125	5	0.3499
		E	82	0.604	2	0.1122	4	0.0929	0	0	909	27.0939	112	1.1656	105	0.8841	3	0.2099
		F	16	0.1178	0	0	1	0.0232	0	0	0	0	1557	16.2036	30	0.2526	0	0
		G	32	0.2357	0	0	6	0.1394	2	0.1435	4	0.1192	85	0.8846	1918	16.1502	0	0
		H	426	3.1377	0	0	7	0.1626	2	0.1435	6	0.1788	76	0.7909	131	1.1031	455	31.8404
		0	157	1.1564	68	3.8138	34	0.79	20	1.4347	25	0.7452	243	2.5289	476	4.0081	24	1.6795
6.	pilihan2	A	3878	28.563	406	22.7706	769	17.8671	258	18.5079	630	18.7779	3119	32.4592	2423	20.4025	451	31.5605
		B	322	2.3717	0	0	104	2.4164	60	4.3042	79	2.3547	198	2.0606	235	1.9788	15	1.0497
		C	1881	13.8543	286	16.0404	1322	30.7156	309	22.1664	650	19.3741	1019	10.6046	1020	8.5888	188	13.1561
		D	785	5.7818	289	16.2086	440	10.223	365	26.1836	200	5.9613	462	4.808	352	2.964	83	5.8083
		E	1005	7.4022	204	11.4414	605	14.0567	121	8.6801	992	29.5678	709	7.3785	580	4.8838	75	5.2484
		F	2607	19.2016	180	10.0953	367	8.527	94	6.7432	339	10.1043	1465	15.2461	1575	13.262	66	4.6186
		G	2183	16.0787	300	16.8256	516	11.9888	117	8.3931	324	9.6572	2042	21.2509	4437	37.3611	193	13.5059
		H	759	5.5903	50	2.8043	147	3.4154	50	3.5868	116	3.4575	352	3.6632	778	6.551	334	23.373

Lampiran 9 Hasil perhitungan *t-weight* masing-masing atribut untuk kelas target fakultas



No.	Target Kelas		A		B		C		D		E		F		G		H	
	Atribut	Nilai beda	Count	t-weight (%)														
7.	pilihan 3	0	6480	47.7278	974	54.627	2357	54.763	818	58.6801	1435	42.772	5030	52.3468	6974	58.7235	1144	80.056
		A	1658	12.2118	130	7.2911	330	7.6673	94	6.7432	403	12.0119	1395	14.5176	1164	9.8013	86	6.0182
		B	123	0.9059	0	0	29	0.6738	23	1.6499	59	1.7586	108	1.1239	102	0.8589	5	0.3499
		C	999	7.358	106	5.945	445	10.3392	110	7.891	391	11.6542	746	7.7636	663	5.5827	26	1.8195
		D	614	4.5224	175	9.8149	198	4.6004	121	8.6801	154	4.5902	313	3.2574	304	2.5598	21	1.4696
		E	617	4.5445	94	5.272	271	6.2965	46	3.2999	399	11.8927	388	4.0379	341	2.8713	20	1.3996
		F	1103	8.124	94	5.272	215	4.9954	36	2.5825	171	5.0969	371	3.861	650	5.4732	32	2.2393
		G	1620	11.9319	192	10.7684	390	9.0613	118	8.4648	292	8.7034	1151	11.9784	1449	12.2011	65	4.5486
		H	363	2.6736	18	1.0095	69	1.6032	28	2.0086	51	1.5201	107	1.1135	229	1.9283	30	2.0994
8.	psdipлома	0	9612	70.7962	1242	69.6579	2893	67.2165	820	58.8235	2204	65.693	7511	78.1663	8587	72.3055	993	69.4892
		1	85	0.6261	5	0.2804	31	0.7203	10	0.7174	36	1.073	62	0.6452	87	0.7326	3	0.2099
		A	1936	14.2594	61	3.4212	223	5.1812	110	7.891	152	4.5306	542	5.6405	420	3.5365	178	12.4563
		B	151	1.1122	256	14.3578	37	0.8597	30	2.1521	13	0.3875	47	0.4891	88	0.741	6	0.4199
		C	223	1.6425	24	1.346	538	12.5	47	3.3716	63	1.8778	117	1.2176	148	1.2462	31	2.1693
		D	227	1.6719	41	2.2995	90	2.0911	234	16.7862	37	1.1028	116	1.2072	100	0.842	29	2.0294
		E	212	1.5615	38	2.1312	149	3.4619	40	2.8694	677	20.1788	120	1.2488	129	1.0862	25	1.7495
		F	350	2.5779	31	1.7386	87	2.0214	31	2.2238	44	1.3115	578	6.0152	179	1.5072	38	2.6592
		G	781	5.7524	85	4.7672	256	5.948	72	5.165	129	3.845	516	5.37	2138	18.0027	126	8.8174
9.	kodepi 1	0	7718	74.6711	813	61.036	2360	70.5742	561	61.4458	1804	69.1188	5526	74.6454	7083	78.1702	752	61.6899
		1	226	2.1865	2	0.1502	12	0.3589	13	1.4239	12	0.4598	102	1.3778	110	1.214	4	0.3281
		P1	1778	17.202	512	38.4384	816	24.4019	330	36.1446	633	24.2529	1133	15.3046	1280	14.1265	431	35.3568
		P2	520	5.031	5	0.3754	143	4.2763	9	0.9858	130	4.9808	552	7.4564	477	5.2643	31	2.5431
		P3	94	0.9094	0	0	13	0.3888	0	0	31	1.1877	90	1.2157	111	1.225	1	0.082
10.	kategori	A	4466	32.8939	484	27.1453	1420	32.9926	465	33.3572	990	29.5082	3273	34.0618	3877	32.6457	433	30.3009
		A-	3763	27.716	562	31.5199	1167	27.1143	462	33.142	914	27.2429	2503	26.0485	3809	32.0731	334	23.373
		A+	2921	21.5143	266	14.9187	816	18.9591	196	14.0603	644	19.1952	2443	25.4241	2483	20.9077	415	29.0413
		B	865	6.3711	108	6.0572	225	5.2277	79	5.6671	207	6.1699	487	5.0682	491	4.1344	87	6.0882
		B-	426	3.1377	80	4.4868	202	4.6933	59	4.2324	173	5.1565	244	2.5393	430	3.6207	39	2.7292
		B+	627	4.6181	155	8.6932	249	5.7853	55	3.9455	227	6.766	416	4.3293	420	3.5365	49	3.429
		C	61	0.4493	21	1.1778	49	1.1385	11	0.7891	48	1.4307	29	0.3018	59	0.4968	2	0.14
		C-	159	1.1711	30	1.6826	62	1.4405	27	1.9369	60	1.7884	69	0.7181	130	1.0946	46	3.219
		C+	37	0.2725	25	1.4021	41	0.9526	4	0.2869	22	0.6557	35	0.3642	34	0.2863	1	0.07
		D	252	1.8561	52	2.9164	73	1.6961	36	2.5825	70	2.0864	110	1.1448	143	1.2041	23	1.6095
11.	alasan pil2	0	8301	61.1402	1270	71.2283	3047	70.7946	884	63.4146	2215	66.0209	5997	62.4102	9900	83.3614	1176	82.2953
		1	5276	38.8598	513	28.7717	1257	29.2054	510	36.5854	1140	33.9791	3612	37.5898	1976	16.6386	253	17.7047
35.	alasan pil6	0	9702	71.4591	1130	63.3763	3037	70.5623	1038	74.462	2390	71.237	6396	66.5626	6343	53.4102	1109	77.6067
		1	3875	28.5409	653	36.6237	1267	29.4377	356	25.538	965	28.763	3213	33.4374	5533	46.5898	320	22.3933

Lampiran 9 lanjutan



No.	Target Kelas		A		B		C		D		E		F		G		H	
	Atribut	Nilai beda	Count	t-weight (%)	Count	t-weight (%)	Count	t-weight (%)										
36.	m5	40	3	0.0221	0	0	3	0.0697	1	0.0717	2	0.0596	5	0.052	1	0.0084	1	0.07
		50	141	1.0385	27	1.5143	45	1.0455	16	1.1478	34	1.0134	60	0.6244	94	0.7915	11	0.7698
		58	0	0	0	0	0	0	0	0	0	0	0	1	0.0084	0	0	
		60	2192	16.145	294	16.4891	790	18.355	300	21.5208	556	16.5723	1162	12.0928	1343	11.3085	266	18.6144
		70	5476	40.3329	659	36.9602	1775	41.2407	571	40.9613	1350	40.2385	3381	35.1858	4069	34.2624	647	45.2764
		74	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.07	
		80	4714	34.7205	654	36.6798	1354	31.4591	434	31.1334	1128	33.6215	3729	38.8074	4819	40.5776	434	30.3709
		86	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0.14	
		90	1038	7.6453	148	8.3006	327	7.5976	72	5.165	284	8.465	1260	13.1127	1535	12.9252	67	4.6886
		93	0	0	0	0	0	0	0	0	0	0	1	0.0104	0	0	0	
		99	13	0.0958	1	0.0561	10	0.2323	0	0	1	0.0298	11	0.1145	14	0.1179	0	0

No.	Target Kelas		1		3		5		6		7		8		9	
	Atribut	Nilai beda	Count	t-weight (%)	Count	t-weight (%)										
12.	b1	40	3	0.0346	33	0.0917	0	0	1	0.1742	1	0.1188	0	0	0	0
		50	54	0.6234	478	1.3284	9	1.0204	17	2.9617	16	1.9002	5	1.4327	0	0
		60	1101	12.7107	7087	19.6954	79	8.9569	115	20.0348	99	11.7577	58	16.6189	3	8.5714
		64	0	0	0	0	0	0	0	0	0	0	0	0	1	2.8571
		65	0	0	2	0.0056	0	0	0	0	0	0	0	0	0	0
		70	3132	36.1579	15660	43.5206	260	29.4785	214	37.2822	258	30.6413	121	34.6705	19	54.2857
		74	0	0	0	0	0	0	0	0	0	0	0	0	1	2.8571
		80	3519	40.6257	11261	31.2953	365	41.3832	188	32.7526	297	35.2732	120	34.384	9	25.7143
		90	841	9.7091	1461	4.0603	168	19.0476	39	6.7944	170	20.19	45	12.894	2	5.7143
		99	12	0.1385	1	0.0028	1	0.1134	0	0	1	0.1188	0	0	0	0
13.	b2	40	8	0.0924	9	0.025	1	0.1134	3	0.5226	0	0	0	0	0	0
		50	36	0.4156	330	0.9171	3	0.3401	4	0.6969	4	0.4751	3	0.8596	1	2.8571
		60	873	10.0785	6763	18.795	60	6.8027	110	19.1638	79	9.3824	40	11.4613	5	14.2857
		70	3079	35.5461	16171	44.9407	220	24.9433	200	34.8432	220	26.1283	111	31.8052	11	31.4286
		75	0	0	0	0	0	0	0	0	0	0	0	0	1	2.8571
		80	3695	42.6576	11136	30.9479	368	41.7234	204	35.5401	346	41.0926	138	39.5415	12	34.2857
		90	953	11.0021	1573	4.3715	229	25.9637	53	9.2334	190	22.5653	57	16.3324	5	14.2857
		99	18	0.2078	1	0.0028	1	0.1134	0	0	3	0.3563	0	0	0	0
14.	b3	40	0	0	1	0.0028	0	0	0	0	0	0	0	0	0	0
		50	20	0.2309	66	0.1834	0	0	2	0.3484	2	0.2375	1	0.2865	0	0
		60	456	5.2644	3670	10.1993	33	3.7415	67	11.6725	51	6.057	15	4.298	6	17.1429
		61	0	0	0	0	0	0	0	0	0	0	0	1	2.8571	
		70	2618	30.224	15223	42.3061	209	23.6961	194	33.7979	199	23.6342	105	30.086	12	34.2857
		75	0	0	2	0.0056	0	0	0	0	0	0	0	0	0	0
		80	4215	48.6608	14706	40.8693	373	42.2902	253	44.0767	351	41.6865	177	50.7163	11	31.4286
		83	0	0	0	0	0	0	0	0	0	0	0	1	2.8571	
		85	0	0	5	0.0139	0	0	0	0	0	0	0	0	0	0
		90	1334	15.4006	2304	6.403	267	30.2721	58	10.1045	238	28.266	51	14.6132	4	11.4286
15.	b4	99	19	0.2193	6	0.0167	0	0	0	0	1	0.1188	0	0	0	0
		40	2	0.0231	4	0.0111	0	0	0	0	1	0.1188	2	0.5731	0	0
		70 [65-74]	3658	42.2304	20745	57.6522	270	30.6122	264	45.993	240	28.5036	136	38.9685	20	57.1429
		80 [75-84]	3774	43.5696	13274	36.8896	395	44.7846	266	46.3415	344	40.8551	158	45.2722	7	20
		90 [85-94]	1228	14.1769	1959	5.4442	217	24.6032	44	7.6655	257	30.5226	53	15.1862	8	22.8571
		96	0	0	1	0.0028	0	0	0	0	0	0	0	0	0	0

Lampiran 10 Hasil perhitungan *t-weight* untuk kelas target pulau asal pelamar

No.	Target Kelas		1		3		5		6		7		8		9	
	Atribut	Nilai beda	Count	t-weight (%)												
16.	b5	40	0	0	7	0.0195	0	0	0	0	0	0	0	0	0	0
		50	18	0.2078	161	0.4474	0	0	1	0.1742	7	0.8314	4	1.1461	2	5.7143
		60	567	6.5458	4830	13.423	47	5.3288	59	10.2787	42	4.9881	17	4.8711	2	5.7143
		66	0	0	0	0	0	0	0	0	0	0	0	0	1	2.8571
		70	2749	31.7363	15907	44.207	192	21.7687	179	31.1847	145	17.2209	87	24.9284	9	25.7143
		74	0	0	0	0	0	0	0	0	0	0	0	0	1	2.8571
		79	0	0	1	0.0028	0	0	0	0	0	0	0	0	1	2.8571
		80	3931	45.3821	13053	36.2755	381	43.1973	274	47.7352	333	39.5487	189	54.1547	14	40
		85	0	0	2	0.0056	0	0	0	0	0	0	0	0	0	0
		90	1378	15.9086	2022	5.6193	253	28.6848	61	10.6272	306	36.342	52	14.8997	5	14.2857
		99	19	0.2193	0	0	9	1.0204	0	0	9	1.0689	0	0	0	0
17.	b6	50	2	0.0231	50	0.139	1	0.1134	4	0.6969	0	0	0	0	0	0
		60	355	4.0984	3144	8.7375	33	3.7415	46	8.0139	31	3.6817	20	5.7307	1	2.8571
		70	2326	26.8529	14569	40.4886	176	19.9546	208	36.2369	155	18.4086	87	24.9284	12	34.2857
		73	0	0	0	0	0	0	0	0	0	0	0	0	1	2.8571
		79	0	0	0	0	0	0	0	0	0	0	0	0	1	2.8571
		80	4233	48.8686	15393	42.7785	371	42.0635	243	42.3345	313	37.1734	164	46.9914	15	42.8571
		82	0	0	0	0	0	0	0	0	0	0	0	0	1	2.8571
		83	0	0	1	0.0028	0	0	0	0	0	0	0	0	0	0
		86	0	0	1	0.0028	0	0	0	0	0	0	0	0	0	0
		90	1721	19.8684	2821	7.8398	289	32.7664	73	12.7178	331	39.3112	78	22.3496	4	11.4286
		99	25	0.2886	4	0.0111	12	1.3605	0	0	12	1.4252	0	0	0	0
18.	b7	0	1288	14.8695	5651	15.7046	120	13.6054	62	10.8014	93	11.0451	48	13.7536	6	17.1429
		50	4	0.0462	23	0.0639	0	0	0	0	0	0	0	0	0	0
		60	228	2.6322	1036	2.8791	34	3.8549	34	5.9233	11	1.3064	22	6.3037	0	0
		70	1885	21.7617	9841	27.349	141	15.9864	179	31.1847	86	10.2138	77	22.063	7	20
		75	0	0	0	0	0	0	0	0	0	0	0	0	1	2.8571
		77	0	0	1	0.0028	0	0	0	0	0	0	0	0	0	0
		78	0	0	1	0.0028	0	0	0	0	0	0	0	0	0	0
		80	3488	40.2678	15659	43.5178	314	35.6009	229	39.8955	317	37.6485	150	42.9799	12	34.2857
		87	0	0	1	0.0028	0	0	0	0	0	0	0	0	0	0
		90	1741	20.0993	3757	10.441	267	30.2721	70	12.1951	324	38.4798	52	14.8997	9	25.7143
		99	28	0.3233	13	0.0361	6	0.6803	0	0	11	1.3064	0	0	0	0

Lampiran 10 lanjutan



No.	Target Kelas		1		3		5		6		7		8		9	
	Atribut	Nilai beda	Count	t-weight (%)												
19.	f1	30	0	0	2	0.0056	0	0	0	0	0	0	0	0	0	0
		40	19	0.2193	110	0.3057	8	0.907	0	0	3	0.3563	0	0	0	0
		50	195	2.2512	1384	3.8463	22	2.4943	20	3.4843	22	2.6128	10	2.8653	2	5.7143
		60	1811	20.9074	12441	34.5747	155	17.5737	170	29.6167	155	18.4086	103	29.5129	15	42.8571
		69	0	0	0	0	0	0	0	0	0	0	0	1	0	2.8571
		70	3654	42.1843	15401	42.8008	336	38.0952	248	43.2056	303	35.9857	149	42.6934	10	28.5714
		74	0	0	0	0	0	0	0	0	0	0	0	0	1	2.8571
		80	2561	29.5659	6077	16.8885	280	31.746	117	20.3833	251	29.81	79	22.6361	6	17.1429
		90	417	4.8141	566	1.573	80	9.0703	19	3.3101	107	12.7078	8	2.2923	0	0
		99	5	0.0577	2	0.0056	1	0.1134	0	0	1	0.1188	0	0	0	0
20.	f2	30	0	0	2	0.0056	0	0	0	0	0	0	0	0	0	0
		40	19	0.2193	110	0.3057	8	0.907	0	0	3	0.3563	0	0	0	0
		50	195	2.2512	1384	3.8463	22	2.4943	20	3.4843	22	2.6128	10	2.8653	2	5.7143
		60	1811	20.9074	12441	34.5747	155	17.5737	170	29.6167	155	18.4086	103	29.5129	15	42.8571
		69	0	0	0	0	0	0	0	0	0	0	0	1	0	2.8571
		70	3654	42.1843	15401	42.8008	336	38.0952	248	43.2056	303	35.9857	149	42.6934	10	28.5714
		74	0	0	0	0	0	0	0	0	0	0	0	0	1	2.8571
		80	2561	29.5659	6077	16.8885	280	31.746	117	20.3833	251	29.81	79	22.6361	6	17.1429
		90	417	4.8141	566	1.573	80	9.0703	19	3.3101	107	12.7078	8	2.2923	0	0
		99	5	0.0577	2	0.0056	1	0.1134	0	0	1	0.1188	0	0	0	0
21.	f3	30	0	0	2	0.0056	0	0	0	0	0	0	0	0	0	0
		40	4	0.0462	11	0.0306	0	0	0	0	0	0	0	0	0	0
		50	41	0.4733	307	0.8532	6	0.6803	6	1.0453	6	0.7126	3	0.8596	0	0
		60 [55-64]	955	11.0252	7474	20.7709	70	7.9365	96	16.7247	85	10.095	52	14.8997	15	42.8571
		70 [65-74]	3145	36.308	16228	45.0991	288	32.6531	226	39.3728	236	28.0285	138	39.5415	14	40
		80 [75-84]	3618	41.7686	10405	28.9164	366	41.4966	203	35.3659	319	37.886	134	38.3954	5	14.2857
		90 [85-94]	888	10.2517	1557	4.327	151	17.1202	41	7.1429	194	23.0404	22	6.3037	1	2.8571
		99	11	0.127	1	0.0028	1	0.1134	2	0.3484	2	0.2375	0	0	0	0

No.	Target Kelas		1		3		5		6		7		8		9	
	Atribut	Nilai beda	Count	t-weight (%)												
22.	f4	30	0	0	0	0	0	0	0	0.2375	0	0	0	0	0	0
		40	6	0.0693	28	0.0778	0	0	2	0.3484	4	0.4751	0	0	0	0
		50	59	0.6811	592	1.6452	3	0.3401	10	1.7422	10	1.1876	5	1.4327	0	0
		51	0	0	0	0	0	0	0	0	0	0	0	0	1	2.8571
		60	1122	12.9531	8314	23.1054	87	9.8639	115	20.0348	100	11.8765	62	17.765	8	22.8571
		70	3315	38.2706	16837	46.7915	270	30.6122	225	39.1986	209	24.8219	113	32.3782	16	45.7143
		75	0	0	2	0.0056	0	0	0	0	0	0	0	0	2	5.7143
		77	0	0	1	0.0028	0	0	0	0	0	0	0	0	0	0
		80	3321	38.3399	9160	25.4565	364	41.2698	194	33.7979	312	37.0546	147	42.1203	6	17.1429
		85	0	0	5	0.0139	0	0	0	0	0	0	0	0	0	0
		90	825	9.5244	1043	2.8986	156	17.6871	28	4.878	195	23.1591	22	6.3037	2	5.7143
		99	14	0.1616	1	0.0028	2	0.2268	0	0	10	1.1876	0	0	0	0
23.	f5	40	1	0.0115	23	0.0639	0	0	0	0	1	0.1188	0	0	0	0
		50	47	0.5426	341	0.9477	6	0.6803	6	1.0453	8	0.9501	8	2.2923	1	2.8571
		60	881	10.1709	6762	18.7922	62	7.0295	97	16.899	75	8.9074	42	12.0344	5	14.2857
		70 [65-74]	3154	36.4119	16626	46.2052	233	26.4172	209	36.4111	182	21.6152	135	38.6819	14	40
		80 [75-84]	3525	40.695	10791	29.9892	388	43.9909	221	38.5017	313	37.1734	130	37.2493	9	25.7143
		90 [85-94]	1038	11.9834	1439	3.9991	188	21.3152	41	7.1429	254	30.1663	34	9.7421	6	17.1429
		99	16	0.1847	1	0.0028	5	0.5669	0	0	9	1.0689	0	0	0	0
24.	f6	40	0	0	1	0.0028	2	0.2268	0	0	0	0	0	0	0	0
		50	18	0.2078	135	0.3752	1	0.1134	4	0.6969	2	0.2375	2	0.5731	0	0
		60	604	6.973	5224	14.518	68	7.7098	75	13.0662	46	5.4632	29	8.3095	5	14.2857
		70 [65-74]	2920	33.7105	16646	46.2607	243	27.551	223	38.8502	209	24.8219	120	34.384	11	31.4286
		80 [75-84]	3852	44.4701	12228	33.9827	391	44.3311	236	41.115	319	37.886	146	41.8338	17	48.5714
		90 [85-94]	1253	14.4655	1747	4.8551	177	20.068	36	6.2718	259	30.7601	52	14.8997	2	5.7143
		99	15	0.1732	2	0.0056	0	0	0	0	7	0.8314	0	0	0	0
25.	f7	0	1288	14.8695	5651	15.7046	120	13.6054	62	10.8014	93	11.0451	48	13.7536	6	17.1429
		40	0	0	2	0.0056	0	0	0	0	0	0	0	0	0	0
		50	13	0.1501	127	0.3529	0	0	0	0	0	0	0	0	0	0
		60	429	4.9527	2916	8.1038	37	4.195	55	9.5819	9	1.0689	15	4.298	1	2.8571
		70	2361	27.257	13594	37.779	215	24.3764	235	40.9408	168	19.9525	99	28.3668	8	22.8571
		77	0	0	1	0.0028	0	0	0	0	0	0	0	0	0	0
		80	3417	39.4482	11443	31.8011	331	37.5283	173	30.1394	311	36.9359	144	41.2607	16	45.7143
		90	1137	13.1263	2242	6.2307	179	20.2948	49	8.5366	247	29.3349	43	12.3209	4	11.4286
		92	0	0	1	0.0028	0	0	0	0	0	0	0	0	0	0
		99	17	0.1963	6	0.0167	0	0	0	0	14	1.6627	0	0	0	0

Lampiran 10 lanjutan

No.	Target Kelas		1		3		5		6		7		8		9	
	Atribut	Nilai beda	Count	t-weight (%)												
26.	k2	30	1	0.0115	0	0	0	0	0	0	0	0	0	0	0	0
		40	6	0.0693	14	0.0389	1	0.1134	0	0	2	0.2375	0	0	0	0
		50	70	0.8081	395	1.0977	5	0.5669	12	2.0906	15	1.7815	4	1.1461	1	2.8571
		60 [55-64]	1016	11.7294	6818	18.9478	98	11.1111	110	19.1638	83	9.8575	67	19.1977	6	17.1429
		70 [65-74]	3060	35.3267	14363	39.9161	232	26.3039	208	36.2369	247	29.3349	134	38.3954	8	22.8571
		80	3506	40.4756	12167	33.8132	341	38.6621	182	31.7073	307	36.4608	103	29.5129	18	51.4286
		90 [85-94]	995	11.487	2219	6.1668	201	22.7891	62	10.8014	183	21.734	41	11.7479	2	5.7143
		95	0	0	1	0.0028	0	0	0	0	0	0	0	0	0	0
		99	8	0.0924	6	0.0167	4	0.4535	0	0	5	0.5938	0	0	0	0
		40	1	0.0115	7	0.0195	2	0.2268	0	0	0	0	0	0	0	0
27.	k3	50	32	0.3694	158	0.4391	4	0.4535	6	1.0453	2	0.2375	11	3.1519	1	2.8571
		60	699	8.0697	5433	15.0988	90	10.2041	89	15.5052	71	8.4323	55	15.7593	6	17.1429
		70	2812	32.4636	13881	38.5766	245	27.7778	225	39.1986	236	28.0285	121	34.6705	16	45.7143
		75	0	0	3	0.0083	0	0	0	0	0	0	0	0	0	0
		80	3890	44.9088	13615	37.8373	331	37.5283	196	34.1463	316	37.5297	114	32.6648	11	31.4286
		90	1218	14.0614	2880	8.0038	205	23.2426	58	10.1045	216	25.6532	48	13.7536	1	2.8571
		99	10	0.1154	6	0.0167	5	0.5669	0	0	1	0.1188	0	0	0	0
		40	0	0	1	0.0028	0	0	0	0	0	0	0	0	0	0
28.	k4	40	7	0.0808	32	0.0889	0	0	0	0	3	0.3563	1	0.2865	0	0
		50	83	0.9582	631	1.7536	7	0.7937	13	2.2648	19	2.2565	12	3.4384	1	2.8571
		60 [55-64]	1047	12.0873	7905	21.9687	106	12.0181	106	18.4669	65	7.7197	72	20.6304	7	20
		70 [65-74]	3143	36.2849	15101	41.967	275	31.1791	234	40.7666	235	27.9097	118	33.8109	13	37.1429
		80 [75-84]	3289	37.9704	10462	29.0748	324	36.7347	182	31.7073	303	35.9857	111	31.8052	10	28.5714
		90 [85-94]	1078	12.4452	1849	5.1385	165	18.7075	39	6.7944	213	25.2969	35	10.0287	4	11.4286
		99	15	0.1732	2	0.0056	5	0.5669	0	0	4	0.4751	0	0	0	0
		40	5	0.0577	13	0.0361	2	0.2268	0	0	5	0.5938	0	0	1	2.8571
29.	k5	50	25	0.2886	375	1.0422	7	0.7937	7	1.2195	5	0.5938	6	1.7192	1	2.8571
		60	792	9.1434	6174	17.1581	115	13.0385	96	16.7247	72	8.5511	67	19.1977	4	11.4286
		70 [65-74]	2941	33.9529	15411	42.8286	250	28.3447	204	35.5401	206	24.4656	110	31.5186	13	37.1429
		80 [75-84]	3618	41.7686	11764	32.6932	309	35.034	226	39.3728	287	34.0855	121	34.6705	9	25.7143
		90 [85-94]	1267	14.6271	2245	6.2391	196	22.2222	41	7.1429	256	30.4038	45	12.894	7	20
		99	14	0.1616	1	0.0028	3	0.3401	0	0	11	1.3064	0	0	0	0

No.	Target Kelas		1		3		5		6		7		8		9	
	Atribut	Nilai beda	Count	t-weight (%)	Count	t-weight (%)	Count	t-weight (%)								
30.	k6	40	0	0	1	0.0028	0	0	0	0	0	1	0.2865	0	0	
		50	2	0.0231	93	0.2585	2	0.2268	1	0.1742	0	0	2	0.5731	0	0
		60	425	4.9065	3713	10.3188	68	7.7098	62	10.8014	25	2.9691	39	11.1748	7	20
		65	0	0	1	0.0028	0	0	0	0	0	0	0	0	0	0
		70	2645	30.5357	15089	41.9337	249	28.2313	206	35.8885	247	29.3349	117	33.5244	8	22.8571
		76	0	0	1	0.0028	0	0	0	0	0	0	0	1	2.8571	0
		77	0	0	1	0.0028	0	0	0	0	0	0	0	0	0	0
		80	4023	46.4442	14123	39.2491	310	35.1474	255	44.4251	276	32.7791	128	36.6762	14	40
		86	0	0	0	0	0	0	0	0	0	0	0	1	2.8571	0
		90	1549	17.8827	2954	8.2094	251	28.458	50	8.7108	287	34.0855	62	17.765	4	11.4286
		99	18	0.2078	7	0.0195	2	0.2268	0	0	7	0.8314	0	0	0	0
31.	k7	0	1289	14.8811	5651	15.7046	120	13.6054	62	10.8014	93	11.0451	48	13.7536	6	17.1429
		40	2	0.0231	7	0.0195	1	0.1134	0	0	0	0	0	0	0	0
		50	14	0.1616	93	0.2585	5	0.5669	0	0	0	0	2	0.5731	0	0
		60	413	4.768	2322	6.453	48	5.4422	36	6.2718	17	2.019	24	6.8768	3	8.5714
		70	2183	25.202	11896	33.0601	198	22.449	213	37.108	152	18.0523	111	31.8052	8	22.8571
		80	3323	38.363	12763	35.4695	299	33.9002	214	37.2822	302	35.867	128	36.6762	13	37.1429
		90	1427	16.4743	3242	9.0098	209	23.6961	49	8.5366	272	32.304	36	10.3152	5	14.2857
		92	0	0	1	0.0028	0	0	0	0	0	0	0	0	0	0
		99	11	0.127	8	0.0222	2	0.2268	0	0	6	0.7126	0	0	0	0
		30	4	0.0462	3	0.0083	0	0	0	0	0	0	0	0	0	0
32.	m1	40	15	0.1732	119	0.3307	4	0.4535	1	0.1742	3	0.3563	0	0	0	0
		50	185	2.1358	1226	3.4072	19	2.1542	19	3.3101	24	2.8504	9	2.5788	0	0
		56	0	0	0	0	0	0	0	0	0	0	0	1	2.8571	0
		60	1692	19.5336	10707	29.7557	160	18.1406	170	29.6167	149	17.696	79	22.6361	12	34.2857
		70	3375	38.9633	14076	39.1185	298	33.7868	227	39.547	275	32.6603	134	38.3954	8	22.8571
		80	2726	31.4708	8329	23.147	299	33.9002	133	23.1707	245	29.0974	109	31.2321	9	25.7143
		90	657	7.5849	1514	4.2075	102	11.5646	23	4.007	140	16.6271	18	5.1576	5	14.2857
		99	8	0.0924	9	0.025	0	0	1	0.1742	6	0.7126	0	0	0	0
		30	2	0.0231	1	0.0028	0	0	0	0	0	0	0	0	0	0
		40	5	0.0577	32	0.0889	1	0.1134	0	0	1	0.1188	0	0	0	0
33.	m2	50	95	1.0967	654	1.8175	7	0.7937	6	1.0453	9	1.0689	2	0.5731	1	2.8571
		55	0	0	0	0	0	0	0	0	0	0	0	1	2.8571	0
		60	1133	13.0801	8307	23.0859	111	12.585	122	21.2544	94	11.1639	69	19.7708	4	11.4286
		70	3191	36.8391	14804	41.1417	295	33.4467	226	39.3728	256	30.4038	134	38.3954	10	28.5714
		80	3317	38.2937	10319	28.6774	305	34.5805	183	31.8815	302	35.867	122	34.957	16	45.7143
		90	903	10.4248	1857	5.1608	162	18.3673	35	6.0976	176	20.9026	22	6.3037	3	8.5714
		99	16	0.1847	9	0.025	1	0.1134	2	0.3484	4	0.4751	0	0	0	0

No.	Target Kelas		1		3		5		6		7		8		9	
	Atribut	Nilai beda	Count	t-weight (%)	Count	t-weight (%)										
34.	m3	40	1	0.0115	7	0.0195	0	0	0	0	0	0	0	0	0	0
		50	31	0.3579	249	0.692	4	0.4535	5	0.8711	5	0.5938	2	0.5731	0	0
		56	0	0	0	0	0	0	0	0	0	0	0	0	1	2.8571
		60	765	8.8317	5737	15.9436	91	10.3175	79	13.7631	80	9.5012	51	14.6132	8	22.8571
		70	2814	32.4867	14416	40.0634	225	25.5102	219	38.1533	219	26.0095	116	33.2378	8	22.8571
		73	0	0	0	0	0	0	0	0	0	0	0	0	1	2.8571
		80	3813	44.0199	12515	34.7803	359	40.7029	208	36.2369	324	38.4798	140	40.1146	12	34.2857
		90	1221	14.0961	3050	8.4762	201	22.7891	63	10.9756	213	25.2969	40	11.4613	5	14.2857
		99	17	0.1963	9	0.025	2	0.2268	0	0	1	0.1188	0	0	0	0
		35.	m4	40	5	0.0577	35	0.0973	0	0	0	0	0	0	0	0
35.	m4	50	65	0.7504	460	1.2784	6	0.6803	15	2.6132	13	1.5439	9	2.5788	0	0
		60	981	11.3253	7075	19.6621	82	9.2971	85	14.8084	86	10.2138	52	14.8997	11	31.4286
		65	0	0	0	0	0	0	0	0	0	0	0	1	2.8571	
		70	2920	33.7105	15266	42.4256	247	28.0045	201	35.0174	207	24.5843	137	39.255	4	11.4286
		77	0	0	1	0.0028	0	0	0	0	0	0	0	0	0	0
		80	3625	41.8495	11021	30.6284	339	38.4354	216	37.6307	292	34.6793	124	35.5301	15	42.8571
		83	0	0	1	0.0028	0	0	0	0	0	0	0	0	0	0
		90	1050	12.1219	2116	5.8806	204	23.1293	57	9.9303	238	28.266	27	7.7364	4	11.4286
		99	16	0.1847	8	0.0222	4	0.4535	0	0	6	0.7126	0	0	0	0
		36.	m5	40	2	0.0231	14	0.0389	0	0	0	0	0	0	0	0
36.	m5	50	51	0.5888	345	0.9588	6	0.6803	11	1.9164	7	0.8314	7	2.0057	1	2.8571
		58	0	0	0	0	0	0	0	0	0	0	0	1	2.8571	
		60	784	9.051	5819	16.1715	73	8.2766	106	18.4669	62	7.3634	53	15.1862	6	17.1429
		70	2670	30.8243	14544	40.4191	225	25.5102	163	28.3972	207	24.5843	113	32.3782	6	17.1429
		74	0	0	0	0	0	0	0	0	0	0	0	1	2.8571	
		80	3801	43.8813	12460	34.6275	333	37.7551	221	38.5017	293	34.7981	146	41.8338	12	34.2857
		86	0	0	1	0.0028	0	0	0	0	0	0	0	1	2.8571	
		90	1329	15.3429	2792	7.7592	242	27.4376	72	12.5436	259	30.7601	30	8.596	7	20
		93	0	0	1	0.0028	0	0	0	0	0	0	0	0	0	0
		99	25	0.2886	7	0.0195	3	0.3401	1	0.1742	14	1.6627	0	0	0	0

Lampiran 10 lanjutan

Lampiran 10 lanjutan

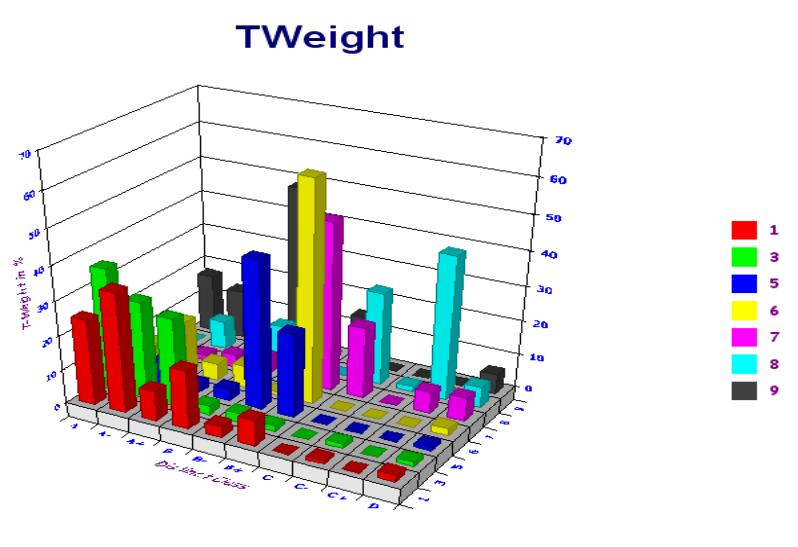
No.	Target Kelas		1		3		5		6		7		8		9	
	Atribut	Nilai beda	Count	t-weight (%)	Count	t-weight (%)										
37.	m6	40	0	0	3	0.0083	0	0	0	0	0	0	0	0	0	0
		50	9	0.1039	104	0.289	6	0.6803	0	0	1	0.1188	2	0.5731	0	0
		60	519	5.9917	4003	11.1247	63	7.1429	60	10.453	51	6.057	34	9.7421	4	11.4286
		67	0	0	0	0	0	0	0	0	0	0	0	1	2.8571	
		70	2519	29.081	14345	39.866	241	27.3243	223	38.8502	181	21.4964	133	38.1089	6	17.1429
		72	0	0	0	0	0	0	0	0	0	0	0	1	2.8571	
		75	0	0	2	0.0056	0	0	0	0	0	0	0	0	0	0
		78	0	0	0	0	0	0	0	0	0	0	0	1	2.8571	
		80	3950	45.6015	13988	38.8739	320	36.2812	204	35.5401	328	38.9549	142	40.6877	11	31.4286
		83	0	0	1	0.0028	0	0	0	0	0	0	0	0	0	0
38.	m7	90	1640	18.9333	3520	9.7824	248	28.1179	87	15.1568	265	31.4727	38	10.8883	11	31.4286
		99	25	0.2886	17	0.0472	4	0.4535	0	0	16	1.9002	0	0	0	0
		0	1288	14.8695	5651	15.7046	120	13.6054	62	10.8014	93	11.0451	48	13.7536	6	17.1429
		40	0	0	5	0.0139	0	0	0	0	0	0	0	0	0	0
		50	18	0.2078	82	0.2279	1	0.1134	1	0.1742	0	0	0	0	0	0
		60	428	4.9411	2317	6.4392	26	2.9478	51	8.885	22	2.6128	30	8.596	1	2.8571
		70	2189	25.2713	11718	32.5654	204	23.1293	194	33.7979	149	17.696	108	30.9456	8	22.8571
		78	0	0	1	0.0028	0	0	0	0	0	0	0	0	0	0
		80	3312	38.236	12677	35.2305	305	34.5805	204	35.5401	299	35.5107	126	36.1032	17	48.5714
		90	1403	16.1972	3523	9.7907	223	25.2834	62	10.8014	261	30.9976	37	10.6017	3	8.5714
39.	r7	99	24	0.2771	9	0.025	3	0.3401	0	0	18	2.1378	0	0	0	0
		0	1289	14.8811	5651	15.7046	120	13.6054	62	10.8014	93	11.0451	48	13.7536	6	17.1429
		1	870	10.0439	2694	7.4869	116	13.1519	50	8.7108	143	16.9834	39	11.1748	9	25.7143
		2	838	9.6744	2767	7.6897	98	11.1111	53	9.2334	135	16.0333	30	8.596	10	28.5714
		3	834	9.6283	2718	7.5536	93	10.5442	52	9.0592	105	12.4703	28	8.0229	6	17.1429
		4	781	9.0164	2747	7.6342	105	11.9048	43	7.4913	59	7.0071	22	6.3037	2	5.7143
		5	751	8.6701	2685	7.4619	67	7.5964	49	8.5366	59	7.0071	19	5.4441	1	2.8571
		6	665	7.6772	2450	6.8088	65	7.3696	33	5.7491	39	4.6318	25	7.1633	1	2.8571
		7	588	6.7883	2294	6.3752	41	4.6485	48	8.3624	30	3.5629	24	6.8768	0	0
		8	513	5.9224	2086	5.7972	58	6.576	38	6.6202	32	3.8005	20	5.7307	0	0
		9	419	4.8372	1697	4.7161	36	4.0816	24	4.1812	27	3.2067	12	3.4384	0	0
		10	393	4.5371	1652	4.5911	29	3.288	34	5.9233	25	2.9691	12	3.4384	0	0
40.	biayahid up	11 [11-100]	721	8.3237	6542	18.1808	54	6.1224	88	15.331	95	11.2827	70	20.0573	0	0

No.	Target Kelas		1		3		5		6		7		8		9	
	Atribut	Nilai beda	Count	t-weight (%)												
41.	pilihan1	A	2544	29.3697	10372	28.8247	169	19.161	172	29.9652	233	27.6722	80	22.9226	7	20
		B	477	5.5068	1058	2.9403	101	11.4512	38	6.6202	66	7.8385	42	12.0344	1	2.8571
		C	967	11.1637	2953	8.2067	134	15.1927	51	8.885	142	16.8646	56	16.0458	1	2.8571
		D	284	3.2787	1045	2.9041	20	2.2676	15	2.6132	21	2.4941	9	2.5788	0	0
		E	823	9.5013	2261	6.2835	101	11.4512	50	8.7108	77	9.1449	41	11.7479	2	5.7143
		F	1600	18.4715	7557	21.0016	113	12.8118	129	22.4739	155	18.4086	45	12.894	10	28.5714
		G	1752	20.2263	9552	26.5459	238	26.9841	117	20.3833	135	16.0333	70	20.0573	12	34.2857
		H	215	2.4821	1185	3.2932	6	0.6803	2	0.3484	13	1.5439	6	1.7192	2	5.7143
42.	persepsi13	-	1008	11.637	9026	25.0841	103	11.678	75	13.0662	76	9.0261	28	8.0229	6	17.1429
		?	511	5.8993	2164	6.014	41	4.6485	70	12.1951	48	5.7007	22	6.3037	2	5.7143
		+	1310	15.1235	3477	9.6629	114	12.9252	46	8.0139	119	14.133	41	11.7479	2	5.7143
		0	4443	51.293	15443	42.9175	414	46.9388	261	45.4704	453	53.8005	131	37.5358	20	57.1429
		T	1390	16.0471	5873	16.3216	210	23.8095	122	21.2544	146	17.3397	127	36.3897	5	14.2857
43.	biayahidup	1	32	0.4536	979	3.6034	16	2.3845	3	0.7143	5	0.8052	2	0.7519	0	0
		2	589	8.3499	7584	27.9142	103	15.3502	42	10	61	9.8229	34	12.782	1	5.8824
		3	1614	22.8806	9078	33.4131	236	35.1714	127	30.2381	145	23.3494	67	25.188	0	0
		4	3107	44.0459	7585	27.9178	236	35.1714	187	44.5238	234	37.6812	115	43.2331	10	58.8235
		5	1712	24.2699	1943	7.1515	80	11.9225	61	14.5238	176	28.3414	48	18.0451	6	35.2941
44.	seleksi2	0	5753	66.4165	21173	58.8417	734	83.22	462	80.4878	630	74.8219	281	80.5158	33	94.2857
		1	2051	23.6781	6690	18.5921	87	9.8639	83	14.4599	137	16.2708	56	16.0458	1	2.8571
		1Y	614	7.0884	3611	10.0353	49	5.5556	15	2.6132	36	4.2755	11	3.1519	1	2.8571
		Y	244	2.8169	4509	12.5309	12	1.3605	14	2.439	39	4.6318	1	0.2865	0	0
45.	KbptX4	B	5670	65.4583	29028	80.6714	717	81.2925	435	75.784	477	56.6508	292	83.6676	26	74.2857
		C	2852	32.9254	6371	17.7056	161	18.254	136	23.6934	345	40.9739	53	15.1862	9	25.7143
		K	2	0.0231	6	0.0167	0	0	2	0.3484	0	0	0	0	0	0
		S	138	1.5932	578	1.6063	4	0.4535	1	0.1742	20	2.3753	4	1.1461	0	0
35.	Kategori	A	2176	25.1212	13016	36.1726	163	18.4807	40	6.9686	7	0.8314	0	0	6	17.1429
		A-	3032	35.0035	10284	28.5802	62	7.0295	89	15.5052	13	1.5439	29	8.3095	5	14.2857
		A+	758	8.7509	9343	25.965	23	2.6077	29	5.0523	31	3.6817	0	0	0	0
		B	1482	17.1092	901	2.504	30	3.4014	37	6.446	44	5.2257	38	10.8883	17	48.5714
		B-	283	3.2671	921	2.5595	380	43.0839	4	0.6969	50	5.9382	15	4.298	0	0
		B+	660	7.6195	544	1.5118	215	24.3764	363	63.2404	411	48.8124	0	0	5	14.2857
		C	0	0	10	0.0278	0	0	0	0	175	20.7838	95	27.2206	0	0
		C-	88	1.0159	490	1.3618	0	0	0	0	0	0	5	1.4327	0	0
		C+	0	0	0	0	0	0	0	0	53	6.2945	146	41.8338	0	0
		D	183	2.1127	474	1.3173	9	1.0204	12	2.0906	58	6.8884	21	6.0172	2	5.7143

Lampiran 10 lanjutan



Lampiran 11 Hasil presentasi grafik untuk atribut putusan1 berdasarkan kelas target fakultas pilihan pertama pelamar



Lampiran 12 Hasil perhitungan *d-weight* untuk kelas target dan kelas pembanding berdasarkan pulau asal pelamar

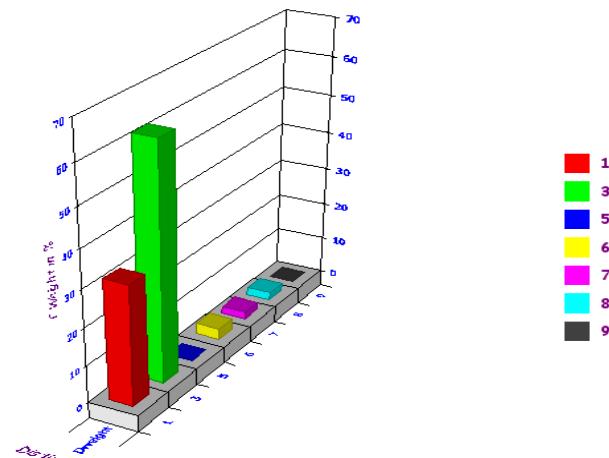
No.	Kelas Target & Pembanding		1		3		5		6	
	Kelas	Nilai beda	Count	t-weight (%)						
46.	Kodepil =P1 dan Listrik	0	12	33.3	22	61.11	0	0	2	5.56
		1	21	31.82	41	62.12	0	0	2	3.03
		2	501	15.43	2627	80.01	52	1.6	15	0.46
		3	480	13.92	2766	80.22	61	1.77	45	1.31
		4	148	12.3	985	81.88	17	1.41	18	1.5
		5	41	7.85	443	84.87	7	1.34	12	2.3
		6	24	14.37	103	61.68	4	2.4	6	3.59

No.	Kelas Target & Pembanding		7		8		9	
	Kelas	Nilai beda	Count	t-weight (%)	Count	t-weight (%)	Count	t-weight (%)
47.	Kodepil =P1 dan Listrik	0	0	0	0	0	0	0
		1	1	1.52	1	1.52	0	0
		2	41	1.26	10	0.31	1	0.03
		3	76	2.2	17	0.49	3	0.09
		4	16	1.33	19	1.58	0	0
		5	11	2.11	5	0.96	3	0.57
		6	5	2.99	4	2.4	21	12.57



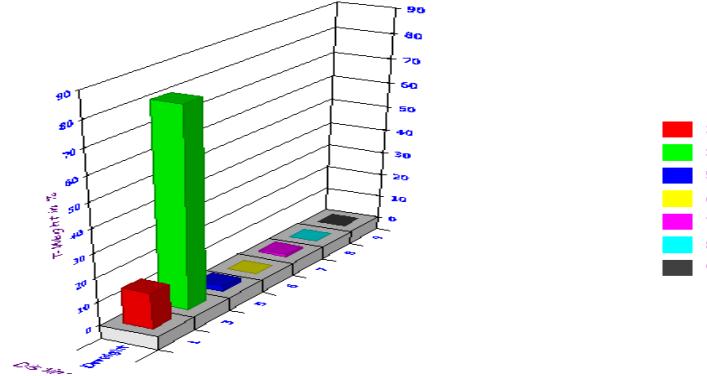
Lampiran 13 Hasil presentasi grafik perhitungan *d-weight* untuk kelas target dan kelas pembanding berdasarkan pulau asal pelamar untuk kondisi dari pelamar yang diterima pada pilihan pertama dengan kondisi listrik 1 (tidak ada)

DWeight



Lampiran 14 Hasil presentasi grafik perhitungan *d-weight* untuk kelas target dan kelas pembanding berdasarkan pulau asal pelamar untuk kondisi dari pelamar yang diterima pada pilihan pertama dengan kondisi listrik 2 (450 watt)

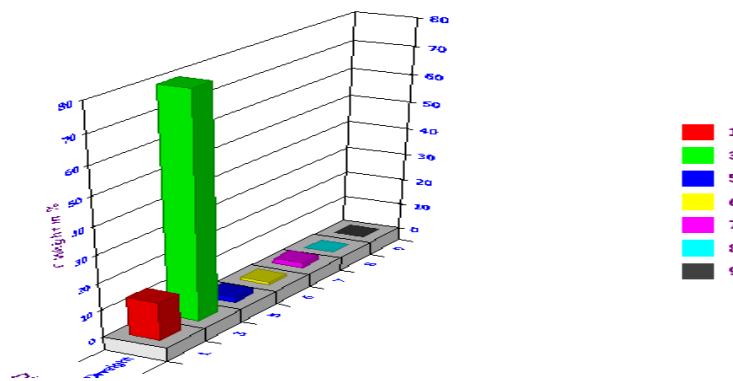
DWeight





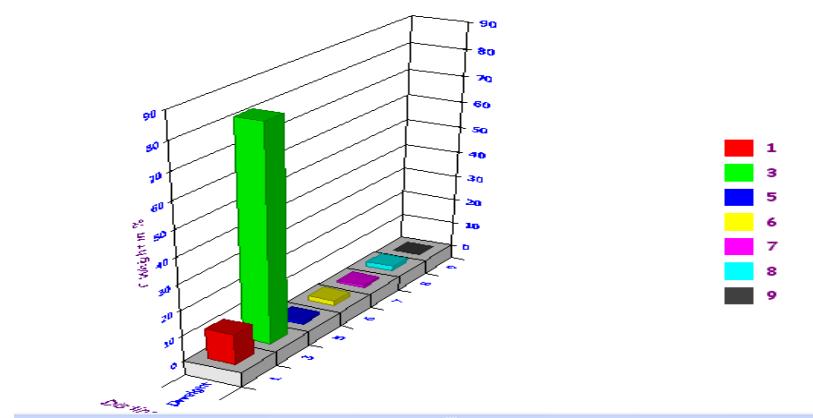
Lampiran 15 Hasil presentasi grafik perhitungan *d-weight* untuk kelas target dan kelas pembanding berdasarkan pulau asal pelamar untuk kondisi dari pelamar yang diterima pada pilihan pertama dengan kondisi listrik 3 (900 watt)

DWeight



Lampiran 16 Hasil presentasi grafik perhitungan *d-weight* untuk kelas target dan kelas pembanding berdasarkan pulau asal pelamar untuk kondisi dari pelamar yang diterima pada pilihan pertama dengan kondisi listrik 4 (1300 watt)

DWeight





Lampiran 17 Hasil presentasi grafik perhitungan *d-weight* untuk kelas target dan kelas pembanding berdasarkan pulau asal pelamar untuk kondisi dari pelamar yang diterima pada pilihan pertama dengan kondisi listrik 5 (2200 watt)



Lampiran 18 Hasil presentasi grafik perhitungan *d-weight* untuk kelas target dan kelas pembanding berdasarkan pulau asal pelamar untuk kondisi dari pelamar yang diterima pada pilihan pertama dengan kondisi listrik 6 (lebih dari 2200 watt)

