

PREDIKSI POLA KELULUSAN MAHASISWA MENGGUNAKAN TEKNIK DATA MINING CLASSIFICATION EMERGING PATTERN

Ida Farida¹, Spits Warnars Harco Leslie Hendric²

Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana¹

Program Doktor Ilmu Komputer, Universitas Bina Nusantara²

Jl. Raya Meruya Selatan, Kembangan, Jakarta, 11650

E-mail : dae.farida@mercubuana.ac.id¹, shendric@binus.edu

ABSTRACT

Mercu Buana University is one of the private universities in Indonesia, especially in DKI Jakarta, which has a large number of students and a number of graduations. However, the University management has difficulty predicting a pattern and graduation rate from existing student data in each academic year. Most researchers use data mining techniques to find a regularity of patterns or relationships set on large data. In this paper, to predict patterns and analyze student graduation rates researchers use data mining by focusing on the classification process using emerging pattern algorithms on the timeliness of student studies. In this study the data used came from combined data between student master data and graduation data. The results of testing the data carried out by researchers in the data mining application produce graduation patterns with various variations according to the learning attributes used, namely gender, class, study program, lecture system and student GPA.

By using the testing result of this study, it is expected that the data result can help the management of the University as a basis for analysis in planning the teaching and learning process strategy to increase the graduation rate on time and as a support for the management of Mercu Buana University

Keywords: *data mining, emerging pattern, KDD (Knowledge Discovery in Database), graduation data, classification*

ABSTRAK

Universitas Mercu Buana merupakan salah satu perguruan tinggi swasta di Indonesia khususnya di wilayah DKI Jakarta yang memiliki jumlah mahasiswa dan jumlah kelulusan yang cukup banyak. Akan tetapi pihak manajemen Universitas kesulitan untuk memprediksi suatu pola dan tingkat kelulusan dari data mahasiswa yang telah ada pada setiap tahun akademik. Sebagian besar para peneliti yang menggunakan teknik data mining untuk menemukan sebuah keteraturan pola atau hubungan set pada data yang berukuran besar. Pada tulisan ini, untuk memprediksi pola dan menganalisa tingkat kelulusan mahasiswa peneliti menggunakan data mining dengan menitik beratkan pada proses klasifikasi menggunakan algoritma emerging pattern terhadap ketepatan waktu studi mahasiswa. Dalam penelitian ini data yang digunakan berasal dari data gabungan antara data induk mahasiswa dan data kelulusan. Hasil pengujian data yang dilakukan oleh peneliti pada aplikasi data mining ini menghasilkan pola kelulusan dengan berbagai variasi sesuai dengan atribut learning yang digunakan yaitu jenis kelamin, angkatan, program studi, system kuliah dan IPK mahasiswa.

Dengan menggunakan Dari hasil pengujian penelitian ini maka diharapkan data yang dihasilkan dapat membantu pihak manajemen Universitas sebagai dasar analisis dalam merencanakan strategi proses belajar mengajar untuk meningkatkan tingkat kelulusan yang tepat waktu serta sebagai pendukung manajemen Universitas Mercu Buana dalam proses pengambilan keputusan

Kata Kunci: *data mining, emerging pattern, KDD (Knowledge Discovery in Database), data kelulusan, classification*

1. PENDAHULUAN

Angka kelulusan menjadi salah satu indikator atau tolak ukur tingkat keberhasilan perguruan tinggi dalam melaksanakan proses kegiatan belajar mengajar (KBM). Angka kelulusan tinggi bisa pula dianggap sebuah prestasi sehingga kebanggaan bagi perguruan tinggi yang bersangkutan. Tidak sampai disitu, angka kelulusan yang tinggi bisa juga menjadi bahan promosi untuk menarik minat calon mahasiswa baru. Tidak mengherankan angka kelulusan ini menjadi begitu penting dan berharga bagi perguruan tinggi, terlebih apabila diumumkan sebagai perguruan tinggi dengan tingkat kelulusan atau nilai tertinggi maka euforia akan nampak jelas sekali. Perguruan tinggi yang dapat menghasilkan kelulusan 100% maka perguruan tinggi tersebut bisa dikatakan perguruan tinggi yang sukses atau berhasil.

Perguruan Tinggi saat ini dituntut untuk memiliki keunggulan bersaing dengan memanfaatkan sumber daya yang dimiliki. Selain sumber daya sarana, prasarana dan manusia, sistem informasi adalah salah satu sumber daya yang dapat digunakan untuk meningkatkan keunggulan bersaing. Pemanfaatan data yang ada di dalam sistem informasi untuk menunjang kegiatan pengambilan keputusan, tidak cukup hanya mengandalkan data operasional saja, diperlukan suatu analisis data untuk menggali potensi-potensi informasi yang ada. Para pengambil keputusan berusaha untuk memanfaatkan gudang data yang sudah dimiliki untuk menggali informasi yang berguna membantu mengambil keputusan, hal ini mendorong munculnya cabang ilmu baru untuk mengatasi masalah penggalian informasi atau pola yang penting atau menarik dari data dalam jumlah besar, yang disebut dengan *data mining*.

Penggunaan teknik *data mining* diharapkan dapat memberikan pengetahuan-pengetahuan yang sebelumnya tersembunyi di dalam gudang data sehingga menjadi informasi yang berharga[3]. Oleh karena itu, dengan memanfaatkan data induk mahasiswa dan data kelulusan mahasiswa, dapat diketahui informasi pola kelulusan mahasiswa melalui teknik *data mining*.

Dari uraian tersebut diatas, maka penelitian akan menggunakan metode *data mining* dengan menggunakan teknik *emerging pattern* untuk mencari pola kelulusan mahasiswa Universitas Mercu Buana, dengan judul: "Prediksi Pola Kelulusan Mahasiswa Menggunakan Teknik *Data mining Classification Emerging pattern*; Studi Kasus Universitas Mercu Buana". Hasil yang diharapkan dari penelitian ini yaitu membantu mempermudah manajemen dalam mengolah data kelulusan serta manajemen dalam memprediksikan dan menganalisis kelulusan mahasiswa sehingga dapat digunakan untuk membantu proses pengambilan keputusan dalam mengambil kebijakan bagi mahasiswa yang diprediksikan tidak dapat lulus tepat waktu sebagai tindakan preventif terhadap hal tersebut.

Penelitian untuk melihat hasil tingkat kelulusan mahasiswa suatu universitas atau instansi sudah pernah dilakukan di beberapa penelitian sebelumnya, antara lain "Aplikasi *Data mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa (Studi Kasus Di Fakultas MIPA (Universitas Diponegoro))*" oleh Nuqson Masykur Huda. Pada penelitian tersebut Nuqson melakukan penelitian menggunakan algoritma Apriori dan KDD sebagai metode *data mining*-nya. Selanjutnya Alfa Shaleh melakukan penelitian dengan judul "*Penerapan Data mining Dengan Metode Klasifikasi Naïve Bayes Untuk Memprediksikan Kelulusan Mahasiswa Dalam Mengikuti English Proficiency Test*". Dalam penelitian tersebut menggunakan Naïve Bayes dengan hasil sebagai berikut Dengan metode Naïve Bayes berhasil mengklasifikasikan 49 data dari 50 data yang diuji.

2. METODOLOGI PENELITIAN

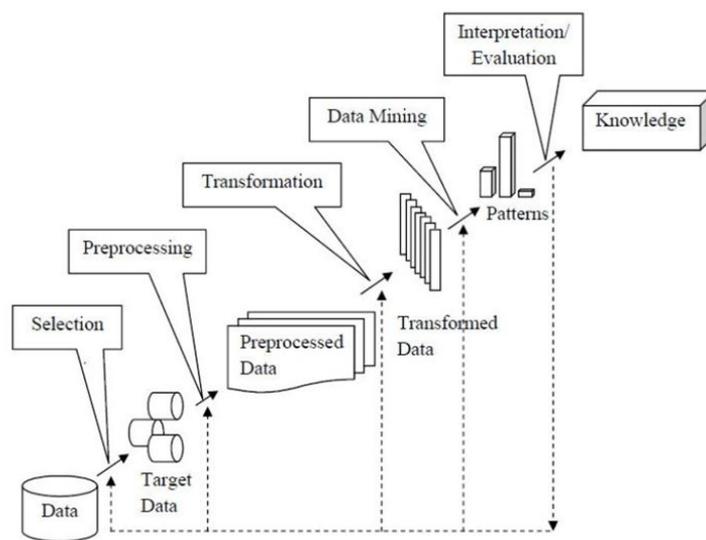
Pada penelitian ini, adapun untuk pengumpulan datanya, peneliti menggunakan metode studi pustaka dan tinjauan lapangan secara langsung pada pihak-pihak terkait. Implementasi hasil analisis dan perancangan menggunakan pemrograman PHP.

Pada tahap pemilihan sampel dilakukan dengan cara memilih *key person* yang mengetahui informasi yang dibutuhkan perusahaan dan teknis pengolahan data dan ketersediaan data pada universitas. *Key person* tersebut adalah unit Pusat Karir dan Alumni dan staff Biro Administrasi Akademik. Pengumpulan data dengan metode studi pustaka peneliti melakukan perbandingan dan

pencarian informasi dari beberapa sumber informasi dan referensi karya ilmiah yang berkaitan dengan *data mining* dan *emerging pattern*. Selain itu untuk melihat proses bisnis yang terjadi untuk keperluan penelitian ini, peneliti melakukan observasi langsung dan wawancara dengan unit pusat karir dan *Alumni* dan Biro Administrasi Akademik sebagai penyedia dan pengolah data kelulusan mahasiswa.

Sebagai instrumentasi penunjang, peneliti melakukan identifikasi kebutuhan pengguna dengan melakukan teknik FGD (*Focus Group Discussion*). Untuk mengetahui spesifikasi kebutuhan pengguna dalam pembuatan aplikasi yang dikembangkan, maka penulis mengumpulkan kebutuhan pengguna berdasarkan masalah yang ada dan diskusi

Metodologi *data mining* yang peneliti gunakan pada penelitian ini adalah KDD (*Knowledge Discovery in Database*) yang memiliki 9 tahap. Alasan peneliti menggunakan metodologi ini dibandingkan yang lain adalah pada tahap-tahap pada KDD dianggap paling mendekati dengan kebutuhan fungsional dan non fungsional dari aplikasi yang ingin peneliti kembangkan.



Gambar 1. Fase Metodologi KDD

Tahap pertama dari metodologi KDD adalah Mengembangkan pemahaman mengenai aplikasi domain dan mengidentifikasi tujuan dari proses KDD dari sudut pandang user. Dalam hal ini peneliti melihat bahwa data kelulusan yang ada pada saat ini belum efektif dan efisien karena data yang ditampilkan hanya berupa data mentah yang belum diolah yang bersumber dari Sistem Infomasi Akademik (SIA) sehingga pihak manajemen Universitas kesulitan dalam mengolah data kelulusan tersebut sehingga diperlukanlah suatu teknik akan pengolahan dan pencarian pola yang ada pada data kelulusan untuk membantu manajemen didalam pengambilan keputusan. Tahap kedua, Membuat sebuah data set sasaran, memilih data set dan berfokus pada variable atau sampel data yang proses penemuannya akan dilakukan. Pada tahap ini data set yang peneliti gunakan adalah data induk mahasiswa yang diambil dalam sampel adalah data mahasiswa angkatan 2009, 2010, dan 2011. Hal ini didasarkan pada kebutuhan data yang akan di hubungkan dengan data kelulusan, dengan asumsi bahwa mahasiswa angkatan 2009 -2011 akan lulus dari rentang waktu tahun 2013-2015. Sedangkan data kelulusan yang diambil adalah data kelulusan dari tahun 2013 sampai tahun 2015. Data yang diambil hanya dari mahasiswa sarjana (S1) program Reguler1 dan Reguler2 (KY).

Tahap ketiga selanjutnya adalah pembersihan data dan pre-processing terhadap data set yang dipilih. Pada tahapan ini sebelum data ditransfer maka data akan dipilah-pilah, dibersihkan terlebih dahulu dari data-data yang tidak relevan dan tidak dapat digunakan karena mengganggu performance dari operational database. dilakukan transformasi data dengan tujuan memodifikasi sumber data format yang berbeda dalam hal jenis dan nilai, yang bertujuan untuk menyederhanakan misalnya menyederhanakan data yang memiliki jarak yang panjang.

Pada tahap keempat yaitu menemukan fitur-fitur yang berguna untuk merepresentasikan data. Tahap ini mempunyai tujuan menggeneralisasi data yang ada untuk menghilangkan *noise* yang sangat banyak yang berada pada data level bawah sehingga setelah dinaikkan levelnya akan menjadi lebih jelas representasi data yang digunakan. Fitur-fitur yang didapatkan antara lain: Jenis Kelamin, Angkatan, Program Studi, Sistem Kuliah, dan IPK.

Tahap selanjutnya yaitu tahap kelima yaitu mencocokkan tujuan dari proses Knowledge Discovery in Database (KDD). Tahap keenam dalam proses KDD adalah mengeksplorasi analisis dan model dan menyeleksi hipotesis. Pada tahap ini peneliti mulai menggunakan algoritma klasifikasi *emerging pattern* untuk menganalisis pola dari data set digunakan dan sebagai solusi dari hipotesis permasalahan yang sudah dipaparkan sebelumnya.

Emerging pattern (EP) adalah sebuah penemuan pengetahuan dari penangkapan tren basis data yang muncul ketika diterapkan dalam *database* atau menangkap perbedaan antara data kelas ketika diterapkan pada data set dan kelas[1]. Selain itu, EP menangkap perubahan yang signifikan dan perbedaan antara data set didefinisikan sebagai item set yang mendukung (frekuensi) meningkat secara signifikan dari satu data set yang lain. Perubahan item set dari satu ke yang lain (dukungan dua rasio) disebut *growth rate* (tingkat pertumbuhan). Selanjutnya, EP menggunakan *threshold* (batasan) untuk mengurangi kandidat pattern yang akan muncul, maka dapat dikatakan EP adalah item set yang *growth rate* nya lebih besar daripada *threshold* yang diberikan.

Secara keseluruhan algoritma EP dapat dibagi menjadi proses *discriminating* antara dua data set dan klasifikasi lebih dari dua data set.

Berikut persamaan dalam pencarian nilai support dan *growth rate* pada *emerging pattern*:

$$\text{supp}_D(X) = \frac{\text{count}_D(X)}{|D|} \tag{1}$$

Dimana :

$\text{supp}_D(X)$ = support dalam data set D mengandung item set X

$\text{count}_D(X)$ = jumlah data dalam data set D yang mengandung item set X dimana $\text{count}_D(X) = t \in D$ and $X \subseteq t$, dimana t adalah bagian dari D

$|D|$ = total data dalam data set D

D = Data set

X = item set atau pattern

Diasumsikan data yang diberikan data set D1 dan D2 lalu *growth rate* item set X dari data set D1 dan D2 dinotasikan dalam persamaan 2 sebagai berikut:

$$\text{GrowthRate}_{D1 \rightarrow D2}(X) = \frac{\text{supp}_{D2}(X)}{\text{supp}_{D1}(X)} \tag{2}$$

Dimana:

$\text{supp}_{D1}(X)$ = support pada data set D1 mengandung item set X (persamaan 1)

$\text{supp}_{D2}(X)$ = support pada data set D2 mengandung item set X (persamaan 1)

Untuk EP yang saling berhubungan dengan dua data set, data set D1 disebut background data set atau bias disebut juga EP negative class dan data set D2 disebut data set target atau disebut positive class.

Jumping *Emerging pattern* (JEP) yang *growth rate* ∞ adalah support EP dengan pembagian sama dengan 0.

Growth rate pada EP dapat dilihat pada persamaan 3 terkait dengan prediksi *confidence* yang dapat digunakan oleh EP.

$$\frac{\text{GrowthRate}(X)}{\text{GrowthRate}(X)+1} = \frac{\text{supp}_{D2}(X)}{\text{supp}_{D2}(X)+\text{supp}_{D1}(X)} \tag{3}$$

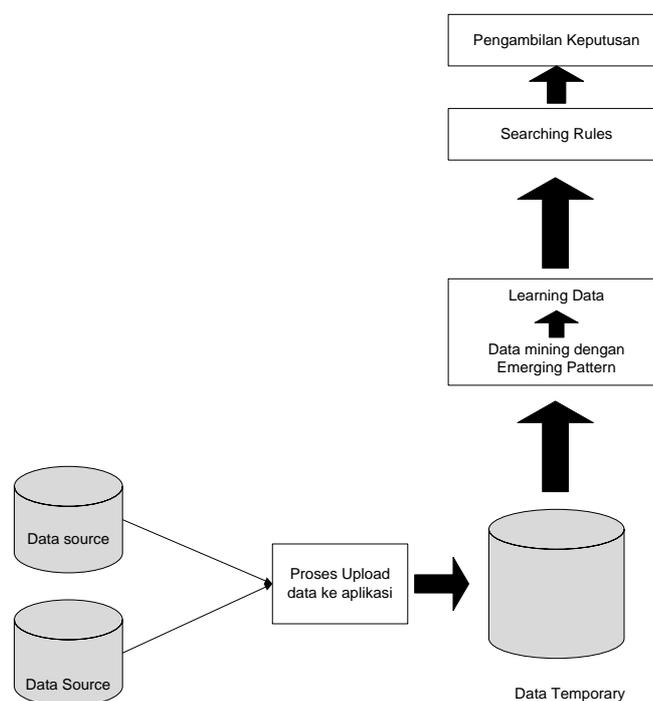
EP dengan *growth rate* yang besar dapat membedakan karakteristik 2 data set dan dapat membangun pengklasifikasian yang sangat menarik. Dengan demikian, EP dengan *growth rate* yang sangat besar akan sering muncul dalam satu kelas tetapi dalam kelas lain akan jarang muncul. Sementara itu, EP yang mempunyai support 1% sampai dengan 20% dapat memberikan pengetahuan baru yang sangat berguna dan penelitian dari para ahli, bahkan aplikasi dapat memahami dengan baik.

Setelah menentukan Algoritma yang akan digunakan, tahap selanjutnya dalam KDD yaitu tahap ketujuh adalah *Melakukan data mining* dengan mencari pola-pola yang menarik dalam bentuk representasional tertentu. Pola yang didapatkan akan membantu manajemen dalam menghasilkan informasi. Misalnya pihak Manajemen akan mendapatkan informasi data kelulusan mahasiswa yang berkaitan dengan ketepatan lulus berdasarkan program studi dari mahasiswa atau manajemen dapat melakukan analisa tentang factor-faktor yang mempengaruhi kelulusan mahasiswa.

Setelah penentuan pola didapatkan, tahap berikutnya yang merupakan tahap kedelapan dari proses KDD adalah menafsirkan pola dapat melibatkan visualisasi dari pola. Peneliti pada penelitian ini membuat suatu aplikasi untuk Implementasi *data mining* kelulusan mahasiswa ini akan dibuat dengan menggunakan bahasa pemrograman server PHP. Tahap terakhir dalam metodologi KDD ini adalah menemukan suatu pengetahuan baru, pada tahap akhir ini metodologi KDD ini bertindak terhadap pengetahuan yang berhasil ditemukan dengan menggunakan pengetahuan tersebut secara langsung, menggabungkan pengetahuan tersebut ke dalam sistem yang lain untuk langkah lebih lanjut, atau hanya mendokumentasikannya saja dan melaporkan kepada pihak yang berkepentingan menggabungkan pengetahuan tersebut ke dalam system yang lain untuk langkah lebih lanjut, atau hanya mendokumentasikannya saja dan melaporkan kepada pihak yang berkepentingan

3. HASIL DAN PEMBAHASAN

Dalam pembangunan penelitian ini, sumber data diperoleh dari dua tabel *database* terpisah yang tidak saling terkait satu sama lain, maka didapatkan arsitektur pengembangan *data mining* yang digunakan selama proses penelitian ini yang dijelaskan oleh gambar dibawah ini:



Gambar 2. Arsitektur pengembangan *data mining*

Berikut penjelasan dari gambar 2 diatas:

1. *Data Source*
Data source yang peneliti gunakan berasal dari data induk mahasiswa dan data kelulusan yang ada di Universitas Mercu Buana untuk jenjang strata-1 (S1) baik sistem kuliah Reguler1 (RE) dan Reguler2 (KY).
2. Proses Upload ke dalam aplikasi
 Proses Upload ini digunakan dengan mengambil data dari hasil *query* dari *database source* yang sudah diproses dan terbentuk satu table data gabungan yang akan digunakan untuk proses *data mining* pada penelitian ini.
3. *Data Temporary*
 Pada data temporary ini digunakan untuk menampung data hasil dari proses query yang akan digunakan dalam proses *data mining*.
4. *Data mining* dengan *Emerging pattern*
 Pembuatan model *data mining* yang dilakukan dengan menggunakan teknik *emerging pattern*.
5. *Searching Rules*
 Proses pencarian dilakukan dengan memilah data dari keseluruhan data yang telah disiapkan sebelumnya. Pencarian pola yang berupa *rules* yang digunakan pada *data mining* sehingga menghasilkan suatu informasi atau pengetahuan yang berguna dan dapat digunakan.
6. Pengambilan Keputusan
 Hasil dari proses *data mining* yang telah dilakukan diproses untuk dapat digunakan oleh pihak manajemen Universitas yang dapat dijadikan sebagai informasi dalam mengambil suatu keputusan dan penentuan kebijakan manajemen.

Dari paparan gambar diatas, telah dijelaskan bahwa dalam penelitian ini data yang digunakan berasal dari data gabungan antara data induk mahasiswa dan data kelulusan, karena pada penelitian ini peneliti mencari hubungan antara beberapa atribut dari kedua data itu.

Data input yang digunakan penelitian ini diperoleh dari Bagian Biro Administrasi Pembelajaran dan Akreditasi (BAPA) Universitas Mercu Buana. Data yang diambil meliputi mahasiswa Strata-1 Program Kelas Reguler-1 dan Program Reguler-2 Universitas Mercu Buana.

Tabel 1 Tabel Data Mahasiswa Angkatan 2009-2011

ANGT.	JML MHS	STATUS MAHASISWA							
		AKTIF	DO	CUTI	NON AKTIF	MENGUN DURKAN DIRI	PINDAH	WAFAT	LULUS
2009	2657	70	350	3	148	59	218	3	1806
2010	2439	202	444	4	184	59	65	2	1479
2011	2874	1038	493	28	164	64	50	0	1037
Total data Kelulusan Angk.2009-2011									4322

Sedangkan Atribut yang didapatkan dari *database* gabungan tersebut disesuaikan dengan data agar sesuai dengan data yang dibutuhkan. Data yang digunakan antara lain:

1. Jenis Kelamin : Digunakan dalam proses *data mining* untuk melihat hubungan jenis kelamin dengan jumlah kelulusan mahasiswa
2. Angkatan : digunakan untuk melihat hubungan antara angkatan dari mahasiswa dengan kelulusan pada proses mining.
3. Program Studi : digunakan dalam proses *data mining* untuk melihat tingkat kelulusan berdasarkan program studi
4. Sistem Kuliah : digunakan untuk melihat hubungan antara pemilihan system kuliah dengan data kelulusan.
5. IPK : digunakan sebagai ukuran tingkat keberhasilan kelulusan mahasiswa

Tabel 2. Data yang digunakan dalam data yang digunakan *data mining*:

Data yang diambil	Sumber Tabel
Jenis Kelamin	Db.mhs
Angkatan	Db.perwalian
Program Studi	Db.mhs
Sistem Kuliah	Db.mhs
IPK	Db.perwalian

Pencarian pola yang ingin peneliti lakukan dapat dilihat dan dikategorikan dengan melihat masa studi dan nilai IPK (Indeks Prestasi Kumulatif) mahasiswa. Dua parameter tersebut data diubah menjadi tipe data yang memudahkan untuk dilakukan proses *data mining* dengan teknik *emerging patterns*.

IPK (Indeks Prestasi Kumulatif) digunakan sebagai dasar penentuan predikat kelulusan program sarjana adalah sebagai berikut:

Tabel 3. Tabel Predikat Kelulusan

Indeks Prestasi	Predikat
2,00 – 2,75	Memuaskan
2,76 – 3,50	Sangat Memuaskan
3.51 – 4,00	Dengan Pujian (Cumlaude)

Predikat kelulusan dengan pujian atau Cum Laude ditentukan dengan memperhatikan masa studi maksimum, yaitu tidak lebih dari 10 (sepuluh) semester. Apabila masa studi lebih dari 10 (sepuluh) semester, maka predikat kelulusan menjadi sangat memuaskan.

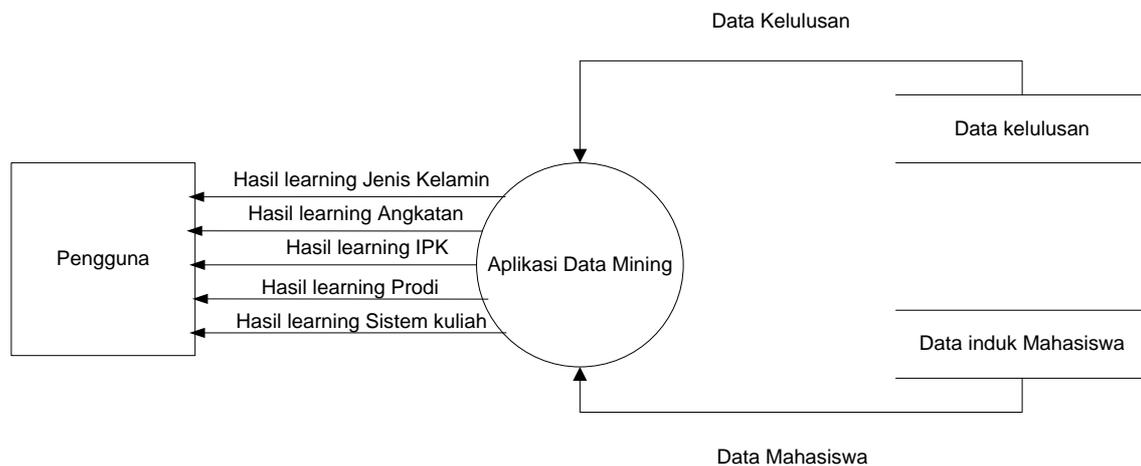
Sedangkan untuk parameter kelulusan yang kedua yaitu masa studi, peneliti membuat 2 (tiga) kategori dan Predikat kelulusan dengan pujian atau Cum Laude ditentukan dengan memperhatikan masa studi maksimum, yaitu tidak lebih dari 10 (sepuluh) semester. Apabila masa studi lebih dari 10 (sepuluh) semester, maka predikat kelulusan menjadi sangat memuaskan.

Sedangkan untuk parameter kelulusan yang kedua yaitu masa studi, peneliti membuat 2 (tiga) kategori antara lain:

1. Tepat waktu, bila lama studi kurang atau 4 tahun
2. Tidak tepat waktu, bila lama studi lebih dari 4

Untuk pemodelan sistem, pada penelitian ini Pemodelan fungsi digambarkan dengan DCD (*Data Context Diagram*), DFD (*Data Flow Diagram*) dan kamus data (*Data Dictionary*).

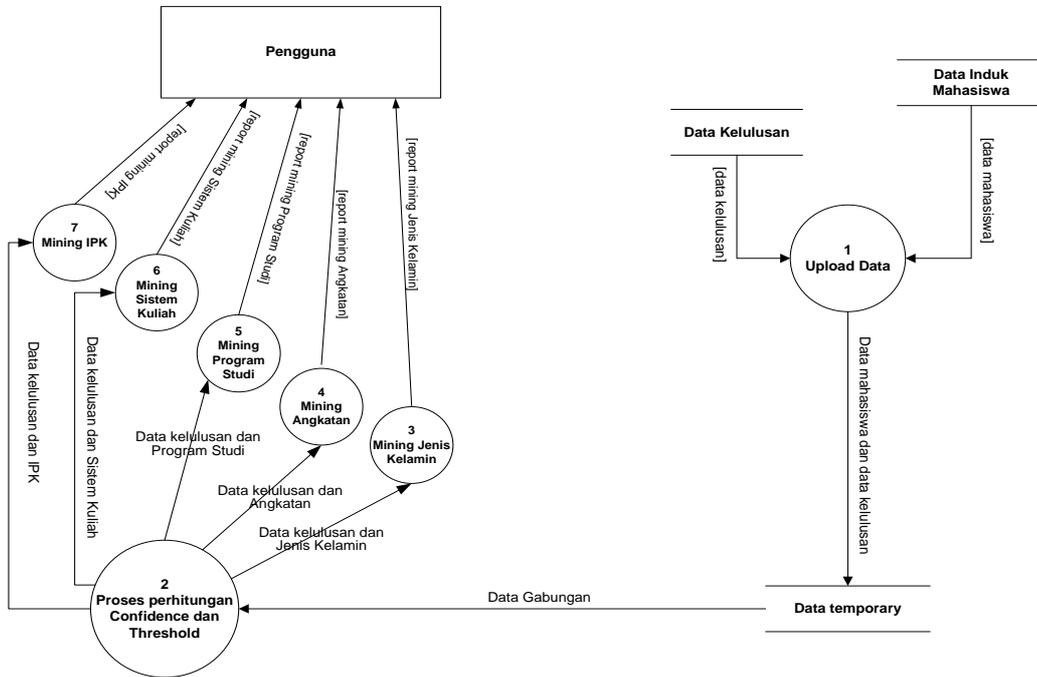
DFD / DFD Level-0



Gambar 3. Gambar DFD Level-0

Gambar 3 merupakan DCD / DFD level-0 pada Aplikasi *Data mining* yang terdiri dari 2 input dan 5 output. *External entity* berupa pengguna atau user dan dua *database* yaitu *database Kelulusan* dan *database Induk Mahasiswa*.

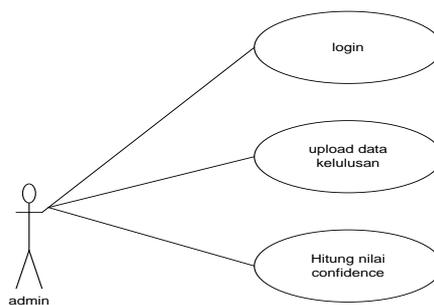
1. DFD Level-1



Gambar 4. Gambar DFD Level-1

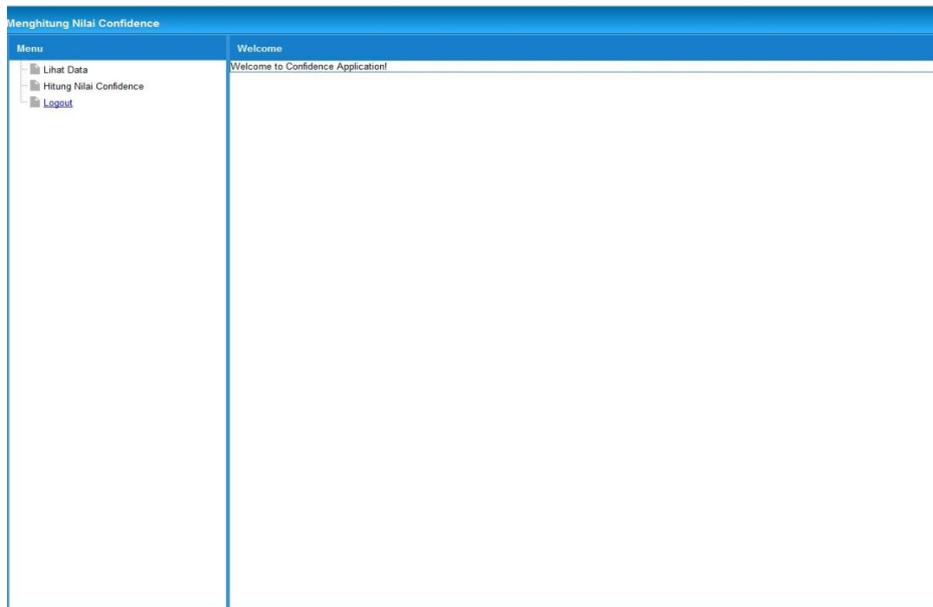
Berikut proses-proses yang terdapat pada Aplikasi *Data mining* :

1. Upload Data
Proses upload data adalah proses *load* data dari data set yang telah disiapkan dalam bentuk Ms.Excel ke dalam aplikasi *data mining*. Semua data akan dimasukkan tanpa ada penyaringan.
2. Proses perhitungan *confidence* dan *threshold*
3. Proses mining Jenis Kelamin merupakan proses mining untuk mengetahui hubungan tingkat kelulusan dengan jenis kelamin dari mahasiswa.
4. Proses mining Angkatan merupakan proses mining untuk mengetahui hubungan tingkat kelulusan dengan angkatan masuk mahasiswa
5. Proses mining Program studi merupakan proses mining untuk mengetahui hubungan tingkat kelulusan dengan program studi yang dipilih oleh mahasiswa
6. Proses mining Sistem Kuliah merupakan proses mining untuk mengetahui hubungan tingkat kelulusan dengan sistem kuliah dari mahasiswa
7. Proses mining IPK merupakan proses mining untuk mengetahui hubungan tingkat kelulusan dengan IPK akhir mahasiswa.



Gambar 5. Gambar use case diagram untuk mencari pola kelulusan mahasiswa

Pada tahap awal, peneliti harus lebih dahulu memasukan data yang akan digunakan pada proses mining ke dalam aplikasi tersebut untuk kemudian dilakukan proses mining dimana program tersebut menganalisis data yang dimasukan tadi dengan cara menghitung nilai *support*, *grow rate* dan *confidence* dari nilai masing-masing atribut yang kita pilih.



Gambar 6. Gambar Form Utama aplikasi *data mining*

Implementasi Metodologi KDD (*Knowledge Discovery in Database*)

a. Mengembangkan pemahaman mengenai aplikasi domain dan mengidentifikasi tujuan dari proses KDD dari sudut pandang user

Data kelulusan yang ada saat ini hanya dapat ditampilkan berupa data mentah yang belum diolah/diproses yang bersumber dari Sistem Informasi Akademik (SIA) sehingga pihak manajemen Universitas kesulitan dalam mengolah data kelulusan tersebut, maka dari itu perlu digunakannya *data mining* dengan menggunakan teknik *emerging pattern* akan mencari pola yang ada pada data kelulusan yang membantu manajemen didalam pengambilan keputusan.

b. Membuat sebuah data set sasaran, memilih data set, pembersihan data dan pre-processing dan mereresentasikan data

Pada tahap kedua ini, peneliti menentukan data set yang akan digunakan dalam proses aplikasi *data mining*, kemudian data set tersebut dilakukan *cleaning* (pembersihan) data untuk mempercepat proses *data mining* tersebut. Data set yang digunakan harus sesuai dengan data yang ingin digunakan dalam proses *data mining*, data tersebut antara lain: data jenis kelamin, angkatan, program studi, sistem kuliah dan IPK.

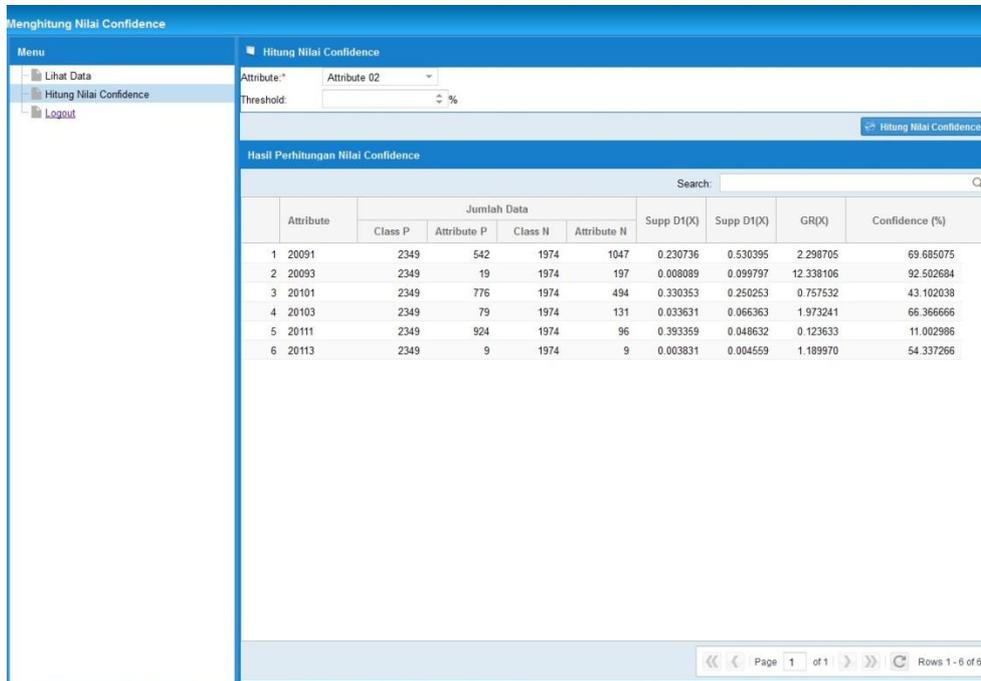
Pada penelitian ini, data yang digunakan pada saat proses *data mining* menggunakan *file based* data dan untuk menunjang kecepatan proses maka data yang digunakan pada saat pemrosesan *data mining* peneliti hanya menggunakan satu *FACT* Table seperti yang terlihat pada gambar

FACT Table
Jenis Kelamin
Angkatan
Program Studi
Sistem Kuliah
IPK

Gambar 7. Gambar *fact* Tabel yang digunakan dalam proses mining

- c. **Mencocokkan tujuan dari proses KDD, Mengeksplorasi analisis dan model dan menyeleksi hipotesis, Melakukan data mining dengan mencari pola-pola yang menarik, Menafsirkan pola dapat melibatkan visualisasi dari pola**

Hasil dari data yang diproses dalam data mining pada aplikasi tersebut akan dipisahkan sesuai dengan atribut yang dipilih saat melakukan proses data mining. Misalnya saat kita memilih atribut Angkatan untuk diproses maka hasil yang keluar adalah prosentase pola kelulusan dari masing-masing angkatan.



Gambar 8. Gambar Form Utama aplikasi data mining

- d. **Acting the discovered knowledge**

Dalam form ini terdapat hasil dari proses mining berupa nilai masing-masing kategori dari atribut yang dipilih yang mempunyai nilai confidence tertinggi. Selain itu terdapat tabel item set dengan atribut, jumlah data, support $D_1(X)$, support $D_2(X)$, $GR(X)$ (growth rate) dan confidence dari item set atribut tersebut.

Setelah didapatkan hasil data mining yang telah dijalankan maka data tersebut dapat dibaca dan dimengerti oleh pengguna untuk keperluan yang diinginkan sesuai keinginan pengguna tersebut.

- e. **Hasil Pengujian Data**

Data yang digunakan dalam pemrosesan data mining pada penelitian ini sebanyak 4321 jumlah data kelulusan mahasiswa.

Tabel 4. Tabel Hasil pencarian manual atribut jenis kelamin

Itemset	Jumlah subdata Tepat waktu	Jumlah Kemunculan Tepat Waktu	Jumlah subdata Tidak Tepat waktu	Jumlah Kemunculan Tidak Tepat Waktu
L	2348	1035	1973	1206
P	2348	1313	1973	767

Untuk perhitungan pola kelulusan menggunakan teknik *emerging pattern* maka dapat langsung kita lakukan dengan memasukan data item set dari tabel diatas.

Maka dapat dihitung nilai supportnya sebagai berikut :

$$X = \{(\text{jenis kelamin}, L)\}$$

$$\text{Support } D_1(X) = \frac{\text{count}_{D_1}(X)}{|D_1|} = 1035/2348 = 0.44$$

$$\text{Support } D_2(X) = \frac{\text{count}_{D_2}(X)}{|D_2|} = 1206/1973 = 0.61$$

Dilihat dari hasil kedua support D1 dan D2, maka didapatkan bahwa untuk support $D_1(X)$ disebut dengan EP negatif karena memiliki nilai *support* yang lebih kecil sedangkan untuk support $D_2(X)$ disebut EP positif karena nilainya lebih besar. Untuk menghitung nilai Growth rate dari item set diatas, maka

$$\text{Growth rate} = \frac{\text{supp}_{D_2}(X)}{\text{supp}_{D_1}(X)} = (1035/2348)/(1206/1973) = 0.61/0.44 = 1.38$$

$$\text{Confidence} = \frac{\text{GrowthRate}(X)}{\text{GrowthRate}(X)+1} = \frac{\text{supp}_{D_2}(X)}{\text{supp}_{D_2}(X)+\text{supp}_{D_1}(X)}$$

$$\frac{\text{GrowthRate}(X)}{\text{GrowthRate}(X) + 1} = \frac{0.61}{0.61 + 0.44} = 0.58$$

Hasil *confidence* adalah 0.58 atau bisa dikatakan subdata set yang digunakan menarik dalam *emerging pattern* sebesar 58% pada kelas tidak tepat waktu.

	Attribute	Jumlah Data				Supp D1(X)	Supp D2(X)	GR(X)	Confidence (%)
		Class TW	Freq TW	Class TTW	Freq TTW				
1	L	2348	1035	1973	1206	0.440801	0.611252	1.386685	58.100889
2	P	2348	1313	1973	767	0.559199	0.388748	1.438462	58.990542

Output: {itemset (attribute01.L) is an EP of class TW with support 1035/2348 (frequency = 0.4408068143101), in class TTW support 1206/1973 (frequency = 0.6112519006589) and GR = 1.3866854712532 and confidence 58.100888783015%
 {itemset (attribute01.P) is an EP of class TW with support 1313/2348 (frequency = 0.55919931856899), in class TTW support 767/1973 (frequency = 0.3887480993411) and GR = 1.4384618716253 and confidence 58.990541880671%

Gambar 9. Hasil Perhitungan Nilai Confidence dari atribut Jenis Kelamin

Tabel 5. Tabel Hasil atribut jenis kelamin dengan aplikasi

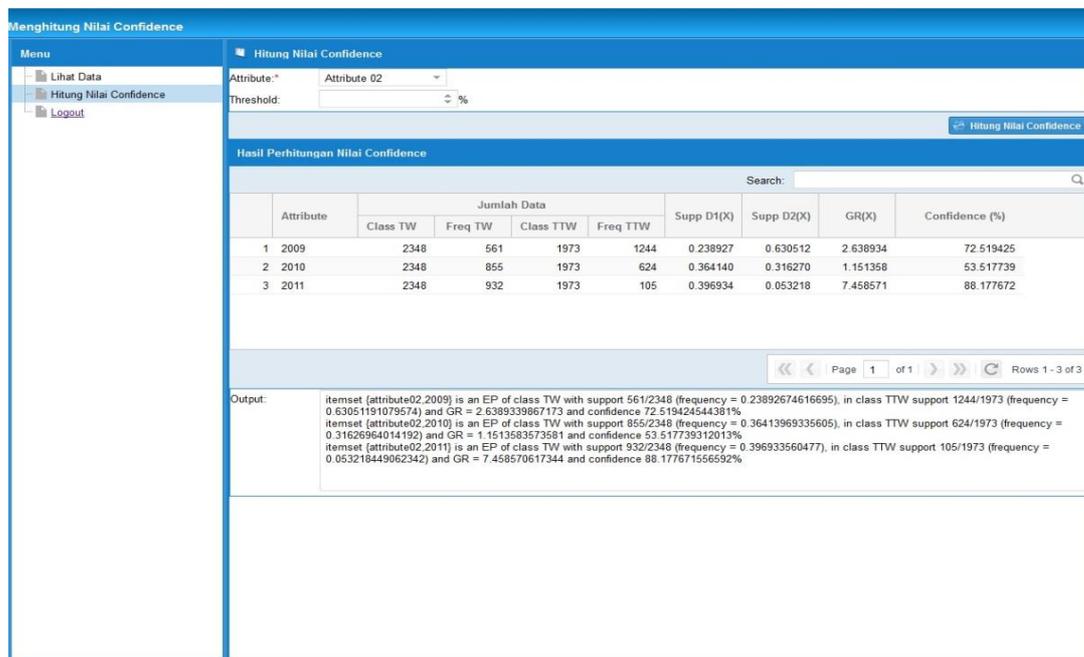
Itemset	Jumlah data				Support D ₁ (X)	Support D ₂ (X)	GR (X)	Confidence(%)
	Class TP	Freq TP	Class TTW	Freq TTW				
L	2348	1035	1973	1206	0,440801	0,611252	1,386685	58,100889
P	2348	1313	1973	767	0,559199	0,388748	1,438462	58,990542

Dari hasil proses mining diatas didapatkan bahwa mahasiswa dengan jenis kelamin Laki-laki memiliki prosentase tinggi dalam class Tidak tepat waktu lebih dominan 1.3 kali lipat dibandingkan dengan mahasiswa wanita.

Prosentase confidence pada item set laki-laki 58.09% dengan tingkat kelulusan tidak tepat waktu memiliki *frequency* 61% sedangkan tingkat *frequency* pada kelas tepat waktu hanya 44% dari total kelulusan tepat waktu.

Sedangkan untuk item set perempuan prosentase nilai confidence 58,9% dengan *frequency* 55,9% pada kelas tepat waktu dan untuk kelas tidak tepat waktu nilai *frequency* 38,85%.

Dengan melihat hasil tersebut maka dapat disimpulkan bahwa mahasiswa perempuan memiliki tingkat kelulusan dengan tepat waktu lebih dominan dibandingkan laki-laki yaitu sebanyak 1.4 kali lipat.



Gambar 10. Hasil Perhitungan Nilai Confidence dari atribut Angkatan

Pada perhitungan mining atribut angkatan terdapat 3 item set yaitu 2009,2010 dan 2011. *Threshold* yang digunakan adalah 0 untuk melihat keseluruhan pola kelulusan dari setiap angkatan serta nilai confidence yang didapatkan.

Tabel 6. Tabel hasil mining atribut angkatan dengan aplikasi

Itemset	Jumlah data				Support D ₁ (X)	Support D ₂ (X)	GR (X)	Confidence(%)
	Class TP	Freq TP	Class TTW	Freq TTW				
2009	2348	561	1973	1244	0.238927	0.630512	2.638934	72.519425
2010	2348	855	1973	624	0.364140	0.316270	1.151358	53.517739
2011	2348	932	1973	105	0.396934	0.053218	7.458571	88.177672

Dari hasil proses mining diatas, maka didapatkan bahwa angkatan yang memiliki prosentase kelulusan tepat waktu tertinggi adalah angkatan 2011. Sedangkan untuk angkatan yang memiliki kelulusan tidak tepat waktu tertinggi adalah mahasiswa angkatan 2009.

Dari tabel diatas, dapat dilihat untuk angkatan 2009, nilai support untuk mahasiswa yang memiliki kelulusan tepat waktu (TW) memiliki *frequency* 23%, sedangkan untuk mahasiswa yang memiliki kelulusan tidak tepat waktu *fequency*nya 63% dengan nilai *growth rate* 2.6 dan nilai *confidence* sebesar 72.5%. Hasil ini memperlihatkan bahwa pada angkatan 2009 ini memiliki rata-rata 2.6 kali lebih tinggi tingkat mahasiswa yang kelulusannya masuk ke dalam tidak tepat waktu (TTW) dibandingkan dengan angkatan 2010 dan 2011.

Attribute	Jumlah Data				Supp D1(X)	Supp D2(X)	GR(X)	Confidence (%)
	Class TW	Freq TW	Class TTW	Freq TTW				
1 Akuntansi	2348	531	1973	303	0.226150	0.153573	1.472587	59.556525
2 Arsitektur	2348	49	1973	52	0.020869	0.026356	1.262927	55.809446
3 Desain Inte...	2348	31	1973	8	0.013203	0.004055	3.256122	76.504434
4 Desain Pro...	2348	182	1973	114	0.077513	0.057780	1.341515	57.292607
5 Hubungan ...	2348	173	1973	149	0.073680	0.075520	1.024970	50.616553
6 Informatika	2348	99	1973	157	0.042164	0.079574	1.887276	65.365281
7 Komunika...	2348	53	1973	56	0.022572	0.028383	1.257428	55.701801
8 Manajemen	2348	386	1973	213	0.164395	0.107957	1.522778	60.361163
9 Penyiaran	2348	239	1973	256	0.101789	0.129752	1.274715	56.038447
10 Periklanan ...	2348	59	1973	150	0.025128	0.076026	3.025591	75.168928
11 Psikologi	2348	84	1973	74	0.035775	0.037506	1.048391	51.181204
12 Sistem Info...	2348	144	1973	175	0.061329	0.088697	1.446261	59.121281
13 Teknik Elek...	2348	66	1973	69	0.028109	0.034972	1.244160	55.439893
14 Teknik Indu...	2348	149	1973	46	0.063458	0.023315	2.721808	73.131336
15 Teknik Mesin	2348	53	1973	94	0.022572	0.047643	2.110683	67.852718
16 Teknik Sipil	2348	50	1973	57	0.021295	0.028890	1.356675	57.567337

Gambar 11. Hasil Perhitungan Nilai Confidence dari atribut Program studi

Pada perhitungan mining atribut Program stuitemdi terdapat 16 item set yang terdiri dari seluruh program studi program strata-1 pada universitas Mercu Buana. Berikut tabel hasil proses mining dengan atribut angkatan. *Threshold* dikosongkan untuk melihat perbedaan hasil dari pola kelulusan dari seluruh item set tersebut.

Tabel 7. Tabel hasil mining atribut program studi

Itemset	Jumlah data				Support D ₁ (X)	Support D ₂ (X)	GR (X)	Confidence(%)
	Class TP	Freq TP	Class TTW	Freq TTW				
Akuntansi	2348	531	1973	303	0.226150	0.153573	1.472587	59.556525
Arsitektur	2348	49	1973	52	0.020869	0.026356	1.262927	55.809446
Desain Interior	2348	31	1973	8	0.013203	0.004055	3.256122	76.504434
Desain Produk	2348	182	1973	114	0.077513	0.057780	1.341515	57.292607
Hubungan Masyarakat	2348	173	1973	149	0.073680	0.075520	1.024970	50.616553
Informatika	2348	99	1973	157	0.042164	0.079574	1.887276	65.365281
Komunikasi Visual	2348	53	1973	56	0.022572	0.028383	1.257428	55.701801

Manajemen	2348	386	1973	213	0.164395	0.107957	1.522778	60.361163
Penyiaran	2348	239	1973	256	0.101789	0.129752	1.274715	56.038447
Periklanan dan Komunikasi Pemasaran	2348	59	1973	150	0.025128	0.076026	3.025591	75.158928
Psikologi	2348	84	1973	74	0.035775	0.037506	1.048391	51.181204
Sistem Informasi	2348	144	1973	175	0.061329	0.088697	1.446261	59.121281
Teknik Elektro	2348	66	1973	69	0.028109	0.034972	1.244160	55.439893
Teknik Industri	2348	149	1973	46	0.063458	0.023315	2.721808	73.131336
Teknik Mesin	2348	53	1973	94	0.022572	0.047643	2.110683	67.852718
Teknik Sipil	2348	50	1973	57	0.021295	0.028890	1.356675	57.567337

Dari hasil proses mining dengan teknik *emerging pattern*, didapatkan pola kelulusan seperti gambar 11. Untuk atribut program studi hasil mining *emerging pattern* yang memiliki hasil yang paling menarik dapat dilihat pada program studi Desain Interior, program studi Periklanan & Komunikasi Pemasaran dan program studi Teknik Industri.

Untuk program studi Desain Interior pola kelulusan mahasiswa didominasi oleh mahasiswa yang memiliki waktu kelulusan tepat waktu dengan nilai confidence 76% untuk kelas tepat waktu dan growth rate 3.2 kali lipat lebih tinggi dibandingkan kelulusan tidak tepat waktu. Pada kelas tepat waktu (TW) *frequency* yang didapatkan 1.3% dan pada kelas tidak tepat waktu *frequency* 0.4%.

Hasil proses mining dari Program studi Periklanan & Komunikasi Pemasaran merupakan kebalikan dari program studi desain interior, dimana mahasiswa yang memiliki kelulusan tepat waktu jauh lebih sedikit dibandingkan mahasiswa yang memiliki kelulusan tidak tepat waktu. Pada kelas tepat waktu *frequency* pada program studi ini 2.5% dan pada kelas tidak tepat waktu 7%, untuk growth ratenya 3.02 dengan nilai confidence 75% untuk mahasiswa dengan kelulusan tidak tepat waktu.

Pada program studi Teknik Industri juga memiliki pola kelulusan yang menarik yaitu dengan nilai confidence 73 % pada kelas kelulusan tepat waktu. Dengan growth rate 2.7, yang artinya mahasiswa yang lulus tepat waktu lebih tinggi 2.7 kali lipat dibandingkan dengan mahasiswa yang lulus tidak tepat waktu.

Attribute	Jumlah Data				Supp D1(X)	Supp D2(X)	GR(X)	Confidence (%)
	Class TW	Freq TW	Class TTW	Freq TTW				
1 KY	2348	432	1973	850	0.183986	0.430816	2.341565	70.073901
2 RE	2348	1916	1973	1123	0.816014	0.569184	1.433655	58.909546

Output: itemset (attribute04 KY) is an EP of class TW with support 432/2348 (frequency = 0.1839863713799), in class TTW support 850/1973 (frequency = 0.43081621896) and GR = 2.3415648288938 and confidence 70.073900965403%
 itemset (attribute04 RE) is an EP of class TW with support 1916/2348 (frequency = 0.8160136286201), in class TTW support 1123/1973 (frequency = 0.56918398376104) and GR = 1.4336552887511 and confidence 58.909546285284%

Gambar 12. Hasil Perhitungan Nilai Confidence dari atribut sistem kuliah

Data yang digunakan dalam pemrosesan *data mining* diatas terdapat sebanyak 4321 jumlah data kelulusan mahasiswa. *Threshold* yang digunakan adalah 0 karena pada saat proses *data mining* dijalankan pada atribut jenis kelamin ini hanya terdiri dari 2 item set yaitu KY (Reguler2/kelas karyawan) dan item set RE (Reguler1), jadi *threshold* dikosongkan untuk melihat perbedaan hasil dari pola kelulusan dari kedua item set tersebut.

Tabel 8. Tabel hasil mining atribut sistem Kuliah menggunakan aplikasi

Itemset	Jumlah data				Support D ₁ (X)	Support D ₂ (X)	GR (X)	Confidence(%)
	Class TP	Freq TP	Class TTW	Freq TTW				
KY	2349	432	1974	850	0.183986	0.430816	2.341565	70.073901
RE	2349	1916	1974	1123	0.816014	0.569184	1.433655	58.909546

Pada proses mining system kuliah diatas, bisa dilihat bahwa pola kelulusan pada system kuliah KY lebih menarik dibandingkan dengan pola kelulusan pada system kuliah RE,hal ini dibuktikan dengan prosentase nilai confidence yang besar yaitu senilai 70% untuk system kuliah KY pada kelas kelulusan tidak Tepat waktu (TTW). Sedangkan pada system kuliah RE nilai confidence hanya 58% untuk kelulusan tepat waktu.

Dengan melihat hasil dari proses mining diatas, dapat disimpulkan bahwa mahasiswa yang memilih system kuliah KY/Reguler2 memiliki tingkat kelulusan tepat waktu yang lebih sedikit dibandingkan dengan dengan mahasiswa RE/reguler1.

Hal ini mungkin disebabkan karena pada mahasiswa system kuliah KY merupakan mahasiswa yang sebagian dari mereka melaksanakan kuliah sambil bekerja, sehingga waktu yang dibutuhkan untuk menyelesaikan kuliah tidak seoptimal mahasiswa pada system kuliah RE atau reguler1.

Proses mining yang terakhir adalah proses mining berdasarkan atribut IPK, terdiri dari 3 item set yaitu ipk 2.00-2.75, ipk 2.76-3.50 dan ipk 3.51-4.0.

Attribute	Jumlah Data				Supp D ₁ (X)	Supp D ₂ (X)	GR(X)	Confidence (%)
	Class TW	Freq TW	Class TTW	Freq TTW				
1 2.00-2.75	2348	6	1973	12	0.002555	0.006082	2.380132	70.415355
2 2.76-3.50	2348	964	1973	1281	0.410562	0.649265	1.581405	61.261406
3 3.51-4.0	2348	1378	1973	680	0.586882	0.344653	1.702822	63.001636

Output:
 itemset (attribute05,2.00-2.75) is an EP of class TW with support 6/2348 (frequency = 0.0025553662691652), in class TTW support 12/1973 (frequency = 0.0060821084642676) and GR = 2.3801317790167 and confidence 70.415354625881%
 itemset (attribute05,2.76-3.50) is an EP of class TW with support 964/2348 (frequency = 0.41056218057922), in class TTW support 1281/1973 (frequency = 0.64926507856057) and GR = 1.5814049638759 and confidence 61.261405652832%
 itemset (attribute05,3.51-4.0) is an EP of class TW with support 1378/2348 (frequency = 0.58688245315162), in class TTW support 680/1973 (frequency = 0.34465281297516) and GR = 1.7028221765708 and confidence 63.001635524955%

Gambar 13. Hasil Perhitungan Nilai Confidence dari atribut IPK

Tabel 9. Tabel hasil mining atribut IPK

Itemset	Jumlah data				Support D ₁ (X)	Support D ₂ (X)	GR (X)	Confidence (%)
	Class TP	Freq TP	Class TTW	Freq TTW				
2.00-2.75	2348	6	1973	12	0.002555	0.006082	2.380132	70.415355
2.76-3.50	2348	964	1973	1281	0.410562	0.649265	1.581405	61.261406
3.51-4.0	2348	1378	1973	680	0.586882	0.344653	1.702822	63.001636

Pada proses mining *threshold* yang kita gunakan 0, karena kita akan membandingkan keseluruhan prosentase nilai confidence dari setiap item set sehingga tidak dibutuhkan *threshold* (batasan).

Pada tabel 13 dapat dilihat pola kelulusan dari masing-masing item set, pola kelulusan yang paling menarik didapatkan pada item set IPK mahasiswa 2.00-2.75, dimana nilai confidence-nya sebesar 70% untuk kelas tidak tepat waktu (TTW), hasil ini menunjukkan bahwa untuk mahasiswa yang memiliki rentang ipk 2.00-2.75 didominasi oleh mahasiswa yang masa studinya lebih dari 4 tahun.

Pada item set 2.76-3.50, pola kelulusan dengan menggunakan teknik *emerging pattern* didominasi oleh mahasiswa yang termasuk kedalam kelas tidak tepat waktu (TTW) dengan *frequency* 64% dan nilai confidence 61% untuk kelas tersebut.

Berbeda dengan hasil kedua item set diatas, untuk item set ipk 3.51-4.0 ini pola kelulusan didominasi oleh mahasiswa yang termasuk kedalam kelas tepat waktu (TW) dengan *frequency* 58% sedangkan *frequency* pada kelas tidak tepat waktu 34% dengan nilai confidence 63% pada kelas tepat waktu dan nilai growth rate sebesar 1.7.

4. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan mengenai Prediksi Pola Kelulusan Mahasiswa Menggunakan Teknik *Data mining Classification Emerging pattern* Studi Kasus Universitas Mercu Buana diatas, dapat ditarik beberapa kesimpulan antara lain:

1. Dari hasil mining dengan atribut jenis kelamin maka hasil yang didapatkan bahwa pola kelulusan mahasiswa yang memiliki prosentase tepat waktu didominasi oleh mahasiswa dengan jenis kelamin Perempuan dibandingkan dengan mahasiswa laki-laki. Untuk hasil dari mining menggunakan atribut angkatan didapatkan bahwa mahasiswa pada angkatan 2009 memiliki pola kelulusan yang lebih menarik dengan prosentase dan nilai confidence sebesar 72.5% pada kelas tidak tepat waktu. Nilai ini lebih besar dibandingkan dengan pola kelulusan mahasiswa tahun angkatan 2010 dan 2011.
2. Dari proses *data mining* diatas juga didapatkan bahwa mahasiswa Kelas Reguler (Reguler1) terdapat pattern kelulusan tepat waktu dengan prosesntase lebih tinggi dibandingkan dengan mahasiswa dengan system kuliah Reguler2 atau kelas karyawan. Hal ini bisa dijadikan acuan analisis dari pihak universitas dalam mengambil tindakan strategis dalam menaikan pattern kelulusan tepat waktu pada mahasiswa dengan system kuliah Reguler2. Dari hasil tersebut diketahui bahwa *high level management* bisa menentukan rencana yang diterapkan untuk meningkatkan kelulusan mahasiswa dengan waktu yang sesuai dengan menganalisa data yang telah ada sebelumnya.
3. Untuk mendapatkan hasil yang baik percobaan terhadap suatu variable dilakukan berkali-kali dalam menentukan *threshold* atau batasan yang dapat menghasilkan data yang menarik karena nilai dari suatu *threshold* akan berbeda tergantung dari data yang akan diolah. Pada suatu data *threshold* tertentu bisa digunakan dan didapatkan data yang menarik namun pada suatu data nilai *threshold* tersebut tidak dapat digunakan.

Selain kesimpulan yang peneliti paparkan diatas, ada beberapa hal dan saran yang mungkin dapat digunakan untuk sebagai acuan untuk penelitian selanjutnya, yaitu:

1. Menggunakan data kelulusan mahasiswa untuk pencarian pattern menarik lainnya seperti *similar pattern*, *frequent pattern*, *association rules* dan pattern lainnya yang dapat digunakan sebagai penunjang dalam pihak manajemen mengambil suatu keputusan.
2. Menggunakan lebih dari dua atribut perbandingan dalam melakukan proses mining pada penelitian diatas, agar pola kelulusan mahasiswa terlihat lebih menarik dan hasil yang didapatkan lebih tajam dari sisi analisis sehingga memudahkan pihak manajemen Universitas mengambil langkah strategis dalam membuat suatu keputusan.
3. Menggunakan data kelulusan mahasiswa untuk dijalankan dengan algoritma *data mining* yang lain, guna mencari perbedaan dan persamaan dengan teknik *data mining emerging pattern*.
4. Pembuatan aplikasi *data mining* ini dapat dipadukan dengan aplikasi OLTP, sementara itu aplikasi ini menggunakan tipe *generic software* sederhana dimana user terbatas untuk pengembangan kebutuhan bisnis proses. Oleh karena itu untuk pengembangan selanjutnya Universitas Mercu Buana dapat meningkatkan dan mengembangkan tipe *customized software* yang dikembangkan sesuai kebutuhan bisnis proses yang ada pada di Universitas Mercu Buana.
5. Untuk penelitian selanjutnya, sebaiknya data yang diberikan untuk penelitian yang serupa, menggunakan data set yang lebih banyak lagi agar mengoptimalkan kinerja algoritma yang digunakan.

DAFTAR PUSTAKA

- [1] KDNuggets. http://www.kdnuggets.com/data_mining_course/x1-intro-to-data-mining-notes.html. Retrieved 30 August 2012
- [2] Larose, 2005. "Discovering Knowledge in Data: An Introduction to *Data mining*": John Willey & Sons, Inc
- [3] Santosa, Budi, 2007, "*Data mining* Teknik Pemanfaatan Data untuk Keperluan Bisnis", Graha Ilmu, Yogyakarta
- [3] Nuqson, Masykur Huda, Program Studi, Teknik Informatika, Fakultas Matematika dan Pengetahuan Alam, and Universitas Diponegoro. 2010. Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa (Studi Kasus Di Fakultas Mipa (Universitas Diponegoro).
- [4] David Hartanto Kamagi, Seng Hansun. 2014. Program Studi, Teknik Informatika, Universitas Multimedia Nusantara. "Implementasi *Data mining* dengan Algoritma C4.5 untuk memprediksi Tingkat Kelulusan Mahasiswa."
- [5] Kusriani, dan Emha Taufik Luthfi. 2009. Anatan.L, Ellitan.L. 2009. Algoritma *Data mining*. Yogyakarta , Penerbit Andi.