

Penerapan Algoritma FP-Growth Untuk Mengetahui Variabel yang Memengaruhi Tingkat Kelulusan Mahasiswa FT UR

Nidya Nur Syafiqoh^{1a}, Feri Candra^{2b}

^{a, b}Program Studi Teknik Informatika, Fakultas Teknik, Universitas Riau
Kampus Bina Widya Jl. HR. Soebrantas Km 12,5 Simpang Baru, Panam
Pekanbaru, 28293, 0761-63272

¹nidya.nursyafiqoh5285@student.unri.ac.id, ²feri@eng.unri.ac.id

Abstract

Information is an important element in every fragment of human life that can be obtained in various possible ways. One of them is data mining. Data mining can form certain knowledge model and discover related patterns from sets of big data. Data mining application with association rules method and fp-growth algorithm can be utilized to discover variables that work on graduation rate. This study focuses on graduation, especially in the Engineering Faculty of Universitas Riau. Variables used in this study are the ways to get into university, the origin of student's high schools, the high schools' region, study majors, gender, age, time of the study, and GPA (Grade Point Average). The information displayed through the mining process is the value from support, confidence, and final of each variable tested. The result of this study is displayed on a web based and can be used by the dean and the head of study majors to cognize their student's graduation rate. Moreover, the result can be used for making a decision, such as quotation of student admission.

Keywords: data mining, association rules, fp-growth, graduation rate, web.

Abstrak

Informasi merupakan elemen penting pada setiap fragmen kehidupan manusia yang bisa didapatkan dengan berbagai cara, salah satunya adalah *data mining*. *Data mining* dapat membentuk suatu model pengetahuan dan menemukan pola-pola berkaitan dari kumpulan *big data*. Penerapan *data mining* dengan menggunakan metode *association rules* dan algoritma *fp-growth* dapat dimanfaatkan untuk mengetahui variabel-variabel yang mempengaruhi tingkat kelulusan mahasiswa. Pada penelitian ini khususnya Fakultas Teknik Universitas Riau. Variabel-variabel yang digunakan antara lain jalur masuk, asal sekolah, kota asal sekolah, program studi, jenis kelamin, usia, lama studi, dan IPK (Indeks Prestasi Kumulatif). Informasi yang ditampilkan melalui proses *mining* adalah nilai *support*, *confidence*, dan final dari setiap variabel yang diujikan. Hasil penelitian ini diimplementasikan dalam bentuk web dan dapat dimanfaatkan oleh dekan, ketua-ketua jurusan dan program studi untuk mengetahui tingkat kelulusan mahasiswanya serta sebagai penunjang pengambilan keputusan seperti penentuan kuota penerimaan mahasiswa baru.

Kata kunci: *data mining*, *association rules*, *fp-growth*, tingkat kelulusan, web.

© 2019 Jurnal CTIA

1. Pendahuluan

Informasi merupakan hasil pengolahan data yang telah diberi makna dan memiliki maksud tertentu. Namun, pengolahan data dengan metode konservatif tidak dapat menggali informasi secara efisien dan efektif dari data yang jumlahnya sangat banyak (*big data*). Dibutuhkan teknologi yang dapat menganalisis secara mendalam untuk menangani *big data* sehingga informasi-informasi baru yang tersembunyi bisa diperoleh. Salah satu teknologi tersebut adalah *data mining*.

Data mining dapat membentuk suatu model pengetahuan dan menemukan pola-pola yang saling berkaitan pada kumpulan *big data*. *Data mining* dapat dimanfaatkan pada berbagai bidang, tidak terkecuali pemanfaatan dan pengolahan data pada institusi perguruan tinggi. Setiap perguruan tinggi memiliki data mahasiswa yang setiap tahun akan terus bertambah. Data tersebut bisa dimanfaatkan untuk menemukan informasi yang sebelumnya tidak diketahui, misalnya adalah variabel-variabel yang dapat memengaruhi tingkat kelulusan mahasiswa.

Dengan mengetahui tingkat kelulusan mahasiswa, perguruan tinggi dapat memahami faktor yang memperlambat atau mempercepat kelulusan mahasiswa, serta mendapatkan informasi penunjang untuk pengambilan keputusan penentuan kuota penerimaan mahasiswa baru.

Penelitian terdahulu oleh Alfannisa Annurullah Fajrin dan Algifanri Maulana (2018) yang melakukan penerapan *data mining* untuk analisis pola pembelian konsumen dengan algoritma *fp-growth* pada data transaksi penjualan *spare part* motor CV. Tjahaja Baru menemukan hasil penjualan *spare part* motor *sport* yang paling terjual sehingga CV. Tjahaja Baru terbantu dalam aktivitas pemesanan *spare part* pada kantor pusat [1]. Kemudian penelitian oleh Heni Sulastri dan Acep Irham Gufroni (2017) yang melakukan penerapan *data mining* dalam mengelompokkan penderita *thalassaemia* pada POPTI Tasikmalaya berhasil mengelompokkan penderita *thalassaemia* menjadi tiga kelompok berdasarkan Hb level dan usia, sehingga dapat menjadi acuan pelaksanaan transfuse, menjaga kadar Hb, persediaan labu darah, pemberian obat klasi besi, dan

rekomendasi untuk operasi pengangkatan limpa [2]. Penelitian yang dilakukan oleh Susanto dkk (2015) yakni penerapan *data mining classification* untuk prediksi perilaku pola pembelian terhadap waktu transaksi menggunakan metode Naïve Bayes menghasilkan *density* tertinggi perilaku pola pembeli mencapai 16.8% pada kelas siang di hari minggu [3].

Pada penelitian yang dilakukan Rizka Nurul Arifin (2015) mengenai implementasi algoritma *fp-growth* untuk menentukan asosiasi antar produk di Nadiamart yang menghasilkan beberapa pola asosiasi yang memenuhi syarat, diantaranya jika membeli snack maka membeli susu instant. Pola yang dihasilkan dapat membantu perusahaan untuk penentuan keputusan memberikan paket diskon (*bundling*) terhadap pola pembelian item yang memiliki nilai *confidence* tinggi namun memiliki nilai *support* yang rendah [4]. Lalu penelitian berjudul aplikasi *data mining* untuk menampilkan informasi tingkat kelulusan mahasiswa pada Fakultas MIPA Universitas Diponegoro oleh Nuqson Masykur Huda (2010) berhasil berjalan dengan baik menampilkan informasi tingkat kelulusan mahasiswa Fakultas MIPA Universitas Diponegoro dan memperoleh kesimpulan semakin tinggi nilai *support* dan *confidence* maka semakin kuat nilai hubungan antar atribut [5].

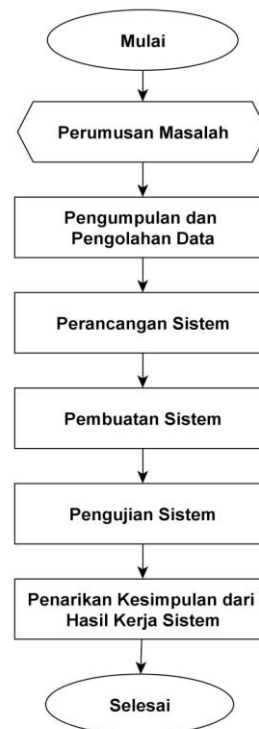
Berdasarkan latar belakang di atas, maka dilakukanlah penelitian penerapan algoritma *fp-growth* untuk mengetahui variabel yang memengaruhi tingkat kelulusan mahasiswa Fakultas Teknik Universitas Riau. Adapun batasan masalah penelitian ini sebagai berikut:

1. Data yang digunakan pada penelitian ini adalah data induk mahasiswa program sarjana reguler Fakultas Teknik Universitas Riau angkatan 2011-2014 dan data mahasiswa program sarjana Fakultas Teknik Universitas Riau yang wisuda pada bulan Februari 2016 hingga bulan Juli 2018.
2. Penelitian ini menggunakan algoritma *fp-growth* saja.
3. Implementasi berupa sistem informasi berbasis web dan bersifat *online*.
4. Sistem hanya dapat diakses oleh admin, ketua-ketua program studi dan jurusan, serta Dekan Fakultas Teknik Universitas Riau.

2. Metodologi Penelitian

2.1. Tahapan Penelitian

Penelitian ini dilakukan dengan beberapa tahapan seperti yang terlihat pada gambar berikut.



Gambar 1. Tahapan penelitian

2.1.1. Perumusan Masalah

Perumusan masalah merupakan tahap persiapan penelitian, yakni menemukan permasalahan untuk kemudian diteliti solusinya. Permasalahan utama pada penelitian ini adalah bagaimana melakukan penerapan algoritma *fp-growth* untuk mengetahui variabel yang memengaruhi tingkat kelulusan mahasiswa Fakultas Teknik Universitas Riau.

2.1.2. Pengumpulan dan Pengolahan Data

Pengumpulan data pada penelitian ini dilakukan dengan cara observasi dan pengambilan data, serta studi pustaka. Pengolahan data terbagi menjadi beberapa tahapan yaitu pembersihan data, integrasi data, transformasi data, *data mining*, evaluasi pola, dan penyajian pengetahuan.

2.1.3. Perancangan Sistem

Perancangan sistem merupakan tahap penggambaran, perencanaan, dan pembuatan sketsa atau pengaturan dari beberapa elemen yang terpisah ke dalam satu kesatuan yang utuh dan berfungsi. Perancangan sistem terdiri dari perancangan proses sistem, perancangan basis data, dan perancangan tampilan antarmuka sistem.

2.1.4. Pembuatan Sistem

Tahap pembuatan sistem ini dilakukan dengan pembuatan *coding* menggunakan bahasa PHP, HTML, CSS, dan JS sehingga rancangan tampilan dapat terimplementasikan dengan baik pada website, serta menggunakan *database* MySQL.

2.1.5. Pengujian Sistem

Tahap pengujian sistem dilakukan setelah sistem selesai dikerjakan untuk menjadi tolak ukur sistem sudah berjalan sesuai dengan tujuan yang diharapkan atau tidak. Pengujian sistem dilakukan dengan dua metode yakni pengujian *software* sistem dan pengujian manual.

2.1.6. Penarikan Kesimpulan dari Hasil Kerja Sistem

Penarikan kesimpulan dari hasil kerja sistem merupakan tahapan akhir penelitian yang didasarkan pada studi pustaka dan pembahasan permasalahan, serta hasil analisis dari penelitian.

2.2. Analisis Data

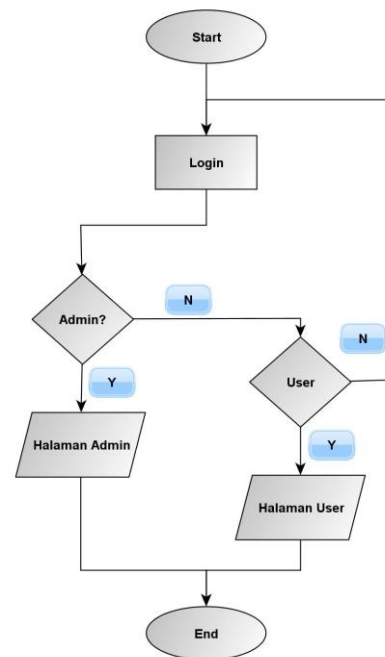
Penelitian ini bertujuan mencari nilai *support* dan *confidence* dari hubungan data kelulusan dan data induk mahasiswa. Namun tidak seluruh atribut pada data induk mahasiswa yang akan dicari hubungannya dengan data kelulusan, hanya beberapa atribut yang berguna dan sebarannya tidak terlalu acak. Data yang sebarannya terlalu acak akan membuat proses *mining* berjalan sangat lambat dan tingkat hubungannya menjadi rendah. Berikut adalah keterangan atribut yang digunakan dalam proses *mining*:

1. Hubungan tingkat kelulusan dengan jalur masuk
2. Hubungan tingkat kelulusan dengan asal sekolah
3. Hubungan tingkat kelulusan dengan kota asal sekolah
4. Hubungan tingkat kelulusan dengan program studi
5. Hubungan tingkat kelulusan dengan jenis kelamin
6. Hubungan tingkat kelulusan dengan usia

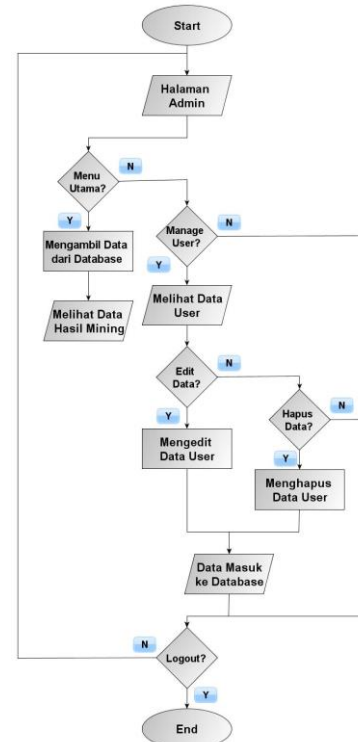
2.3. Perancangan Sistem

2.3.1. Rancangan Proses Sistem

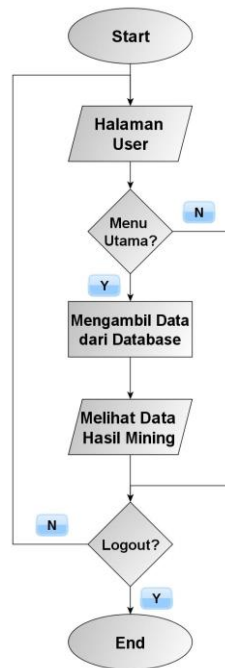
Rancangan proses sistem adalah penggambaran mengenai proses-proses yang ada pada sistem dan alur berjalannya sistem. Pada penelitian ini, rancangan proses sistem menggunakan *flowchart*. *Flowchart* berguna untuk menggambarkan dan menyederhanakan rangkaian proses atau prosedur sehingga mudah dipahami dan diatur berdasarkan urutan langkah proses.



Gambar 2. Alur *flowchart* halaman login



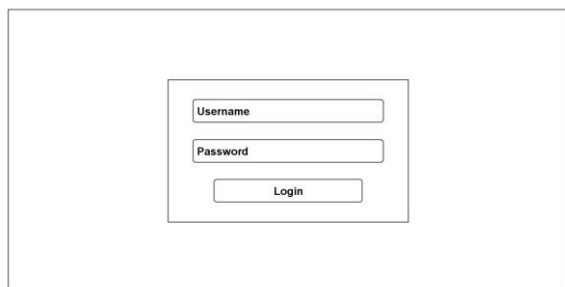
Gambar 3. Alur *flowchart* halaman admin



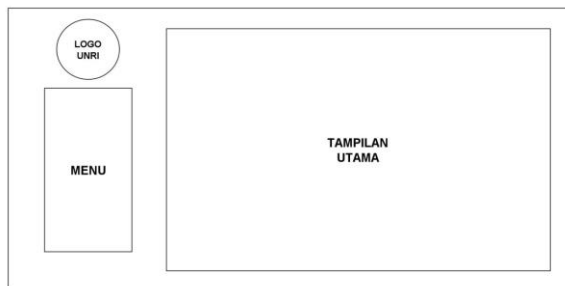
Gambar 4. Alur *flowchart* halaman user

2.3.2. Rancangan Tampilan Antarmuka

Rancangan tampilan antarmuka adalah penggambaran terhadap tampilan sistem yang meliputi halaman-halaman yang ada pada sistem.



Gambar 5. Rancangan halaman *login*



Gambar 6. Rancangan halaman utama



Gambar 7. Rancangan tampilan utama

2.4. Integrasi dan Transformasi Data

Proses integrasi data dilakukan ketika proses ETL (*extract*, *transform*, dan *load*) sehingga beberapa tabel sumber data digabungkan menjadi satu. Sedangkan transformasi data merupakan proses pengubahan dan penggabungan data ke dalam format yang sesuai agar bisa diproses pada *data mining*.

Tabel 1. Tabel Predikat Kelulusan Mahasiswa Universitas Riau Program Sarjana

Indeks Prestasi	Predikat	Keterangan
4,00	<i>Summa Cum Laude</i> (Sangat Terpuji)	1. Masa studi tidak lebih dari 8 semester efektif 2. Pernah mendapat penghargaan terbaik yang diputuskan oleh Universitas
3,50<IPK<4,00	<i>Cum Laude</i> (Dengan Pujian)	1. Tidak mempunyai nilai mata kuliah lebih rendah dari B- dan tanpa ada nilai perbaikan 2. Menyelesaikan program studinya dalam waktu tidak lebih dari 8 (delapan) semester efektif
2,75<IPK<3,50	Sangat Memuaskan	1. Tidak mempunyai nilai mata kuliah lebih rendah dari C 2. Menyelesaikan program studinya dalam waktu tidak lebih dari 10 (sepuluh) semester efektif
2,00<IPK<2,75	Memuaskan	

Berdasarkan tabel di atas dan tidak adanya wisudawan yang meraih IPK 4,00 (empat) pada data kelulusan tahun 2017-2018, maka parameter IPK dikategorikan menjadi:

1. Kategori pujian dengan rentang IPK 3,51-4,00
2. Kategori sangat memuaskan dengan rentang IPK 2,76-3,50
3. Kategori memuaskan dengan rentang IPK 2,00-2,75

Sedangkan pengkategorian data kelulusan berdasarkan lama studi yaitu:

1. Sesuai jadwal apabila lama studi adalah 4 tahun atau kurang dari 4 tahun.
2. Tidak sesuai jadwal apabila lama studi lebih dari 4 tahun.

Parameter IPK dan lama studi tersebut kemudian ditransformasikan menjadi enam kategori baru yang ditunjukkan pada tabel berikut.

Tabel 2. Tabel Transformasi Data	
Kategori	Keterangan
A1	lama studi 4 tahun atau kurang dari 4 tahun dengan rentang IPK 3,51-4,00
A2	lama studi 4 tahun atau kurang dari 4 tahun dengan rentang IPK 2,76-3,50
A3	lama studi 4 tahun atau kurang dari 4 tahun dengan rentang IPK 2,00-2,75
B1	lama studi lebih dari 4 tahun dengan rentang IPK 3,51-4,00
B2	lama studi lebih dari 4 tahun dengan rentang IPK 2,76-3,50
B3	lama studi lebih dari 4 tahun dengan rentang IPK 2,00-2,75

2.5. Penggunaan Algoritma FP-Growth

Berikut adalah contoh penggunaan algoritma *fp-growth* pada penelitian ini. Diketahui *threshold* atau batas nilai ambang = 3.

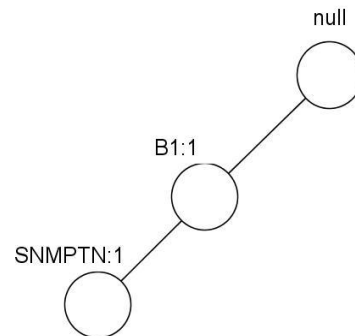
Tabel 3. Tabel Data Awal		
NIM	Kategori Kelulusan	Jalur Masuk
1207113617	B1	SNMPTN
1107114214	B2	SNMPTN
1207113557	B2	SNMPTN
1107121026	B2	PBUD
1107111936	B2	SNMPTN
1207113650	B2	SNMPTN
1107112028	B2	SNMPTN
1107152086	B2	Bina Lingkungan
1207136364	B2	Ujian Mandiri
1107114310	B1	SNMPTN
1207136334	B2	Ujian Mandiri
1107136564	B2	Ujian Mandiri
1107136525	B2	Ujian Mandiri
1207113629	B1	SNMPTN
1207154465	B2	Bina Lingkungan

Tabel 4. Tabel Item Support Count	
Item	Support Count
B1	3
B2	12
PBUD	1
SNMPTN	8
Bina Lingkungan	2
Ujian Mandiri	4

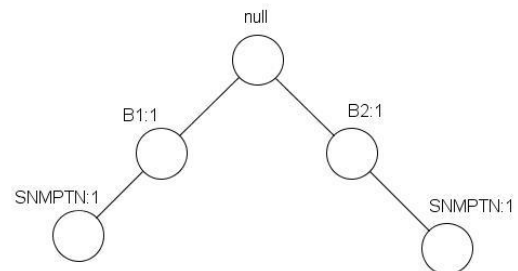
Tabel 5. Tabel Itemset Terurut	
TID	Ordered Itemset
1	{B1;SNMPTN}
2	{B2;SNMPTN}
3	{B2;SNMPTN}
4	{B2}
5	{B2;SNMPTN}
6	{B2;SNMPTN}
7	{B2;SNMPTN}
8	{B2}
9	{B2;Ujian Mandiri}

10	{B1;SNMPTN}
11	{B2;Ujian Mandiri}
12	{B2;Ujian Mandiri}
13	{B2;Ujian Mandiri}
14	{B1;SNMPTN}
15	{B2}

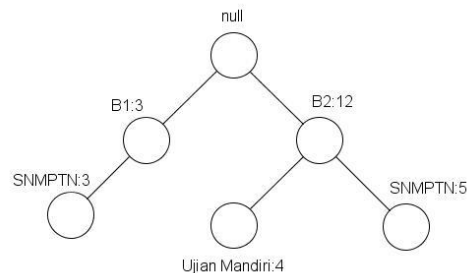
Berikutnya adalah pembentukan *fp-tree* yang ditunjukkan gambar-gambar berikut.



Gambar 8. Hasil pembentukan *fp-tree* setelah pembacaan TID 1

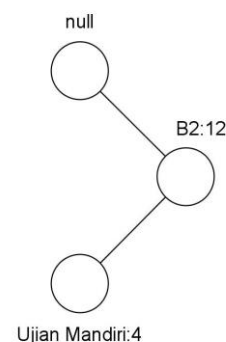


Gambar 9. Hasil pembentukan *fp-tree* setelah pembacaan TID 2

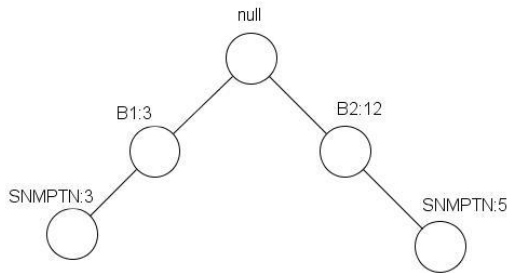


Gambar 10. Hasil pembentukan *fp-tree* setelah pembacaan TID 15

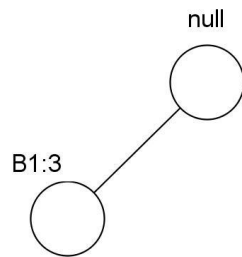
Selanjutnya adalah tahapan *conditional pattern base*.



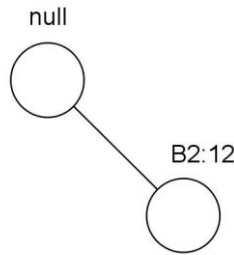
Gambar 11. Lintasan yang memiliki item Ujian Mandiri



Gambar 12. Lintasan yang memiliki item SNMPTN



Gambar 13. Lintasan yang memiliki item B1



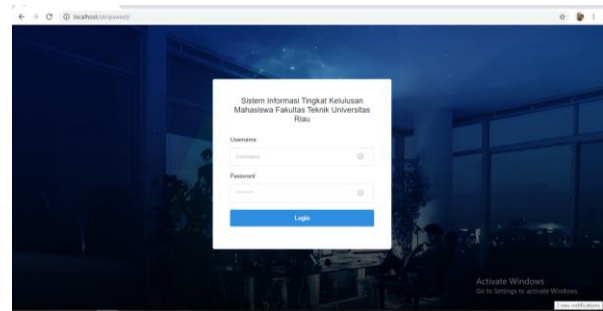
Gambar 14. Lintasan yang memiliki item B2

Suffix	Frequent Itemset
Ujian Mandiri	{B2;Ujian Mandiri}
SNMPTN	{B2;SNMPTN};{B1;SNMPTN}
B1	{B1}
B2	{B2}

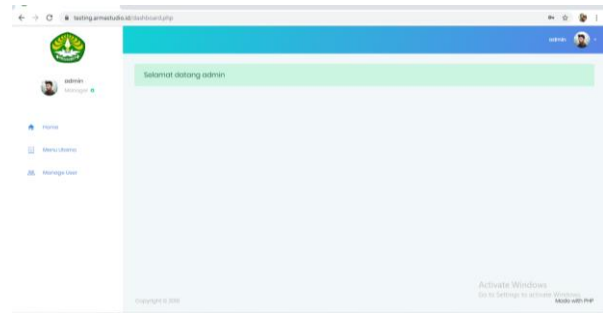
Itemsets	Support	Count	Final
B2, Ujian Mandiri	4/15 (27%)	4/12 (33%)	8,91%
B2, SNMPTN	5/15 (33%)	5/12 (42%)	13,86%
B1, SNMPTN	3/15 (20%)	3/3 (100%)	20%

3. Hasil dan Pembahasan

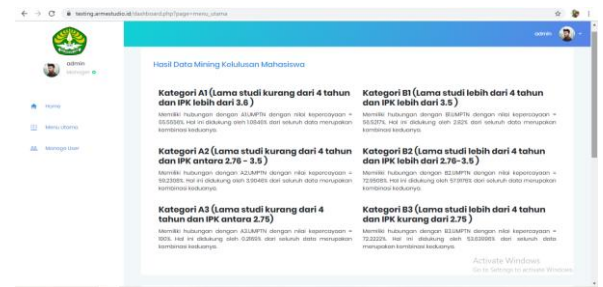
Hasil dari penelitian ini berupa implementasi sistem berbasis web, berikut adalah tampilannya.



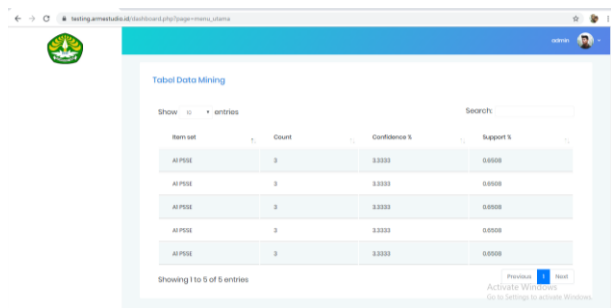
Gambar 15. Tampilan halaman login



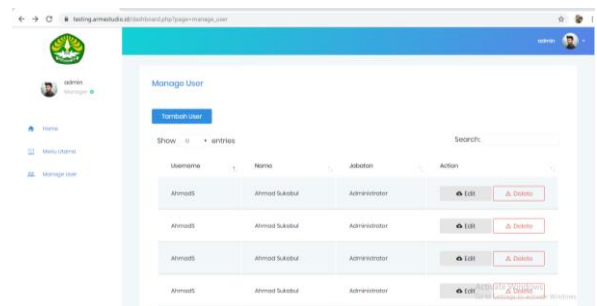
Gambar 16. Tampilan halaman awal



Gambar 17. Keterangan kategori pengelompokkan



Gambar 18. Tampilan halaman yang menyajikan hasil proses mining.



Gambar 19. Tampilan halaman *manage user*.

4. Kesimpulan

Setelah melakukan penelitian ini, kesimpulan yang didapatkan antara lain:

1. Algoritma *fp-growth* dan metode *association rules* dapat diimplementasikan dengan baik dan sesuai harapan.
2. Hasil dari pengolahan *data mining* ini dapat digunakan sebagai bahan pertimbangan dalam pengambilan keputusan, misalnya penambahan kuota pada program studi tertentu berdasarkan tingkat kelulusan prodi tersebut.

Daftar Rujukan

- [1] Fajrin, Alfannisa Annurullah & Algifanri Maulana, 2018. Penerapan *Data Mining* untuk Analisis Pola Pembelian Konsumen dengan Algoritma *FP-Growth* pada Data Transaksi Penjualan *Spare Part Motor*. *Kumpulan Jurnal Ilmu Komputer (KLIK)*, 5(1), pp.27-36.
- [2] Sulastri, Heni & Acep Irham Gufroni, 2017. Penerapan *Data Mining* dalam Pengelompokan Penderita *Thalassaemia*. *Jurnal Nasional Teknologi dan Sistem Informasi*, 3(2), pp.299-305.
- [3] Susanto, dkk, 2015. Penerapan *Data Mining Classification* untuk Prediksi Perilaku Pola Pembelian Terhadap Waktu Transaksi Menggunakan Metode *Naive Bayes*. *Konferensi Nasional Sistem & Informatika*, pp.313-318.
- [4] Arifin, Rizka Nurul, 2015. *Implementasi Algoritma Frequent Pattern Growth (FP-Growth) Menentukan Asosiasi Antar Produk (Studi Kasus Nadiamart)*. S.T. Semarang: Universitas Dian Nuswantoro.
- [5] Huda, Nuqson Masykur, 2010. *Aplikasi Data Mining untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa (Studi Kasus di Fakultas MIPA Universitas Diponegoro)*. S.Kom. Semarang: Universitas Diponegoro.