

RANCANG BANGUN APLIKASI *DATA MINING* ANALISIS TINGKAT KELULUSAN MENGGUNAKAN ALGORITMA *FP-GROWTH* (Studi Kasus Di Politeknik Negeri Malang)

Naufal Farras Hilmy¹, Banni Satria Andoko²

Program Studi Teknik Informatika, Jurusan Teknologi Informasi, Politeknik Negeri Malang
JL. Soekarno-Hatta No. 9 Malang 65141, Indonesia

¹ naufal_farras.hilmy@gmail.com, ² ando@polinema.ac.id

Abstrak

Kondisi kompetisi dan tuntutan oleh Direktorat Jenderal Pendidikan (Dikti) membuat perguruan tinggi harus memiliki keunggulan kompetitif dan kualitas agar dapat tetap bersaing dengan perguruan-perguruan tinggi lainnya. Teknologi informasi merupakan salah satu sumber daya yang dapat meningkatkan keunggulan bersaing dan dapat digunakan untuk mengolah informasi, menyebarkan informasi, serta pengambilan keputusan strategis. Tersedianya suatu informasi tidak terlepas dari melimpahnya data, dari pengolahan dan penggalian data tersebut informasi dapat diperoleh. Teknologi yang berkaitan dengan mengolah dan penggalian data menjadi informasi yang berguna adalah *data mining*. Teknologi *data mining* memiliki berbagai macam algoritma dan teknik untuk penyelesaian masalah yang berhubungan dengan data salah satunya adalah algoritma *fp-growth* dan teknik *association rules*. Dari uraian diatas dapat dibuat sebuah aplikasi untuk menganalisis tingkat kelulusan melalui teknik *data mining* menggunakan algoritma *fp-growth*. Informasi yang ditampilkan menggunakan teknik *association rules* yaitu menampilkan nilai *support* dan *confidence* dari masing-masing proses *mining*. Terbukti dari hasil pengujian yang dilakukan, algoritma *fp-growth* dan teknik *association rules* berjalan sesuai harapan dan dapat diimplementasikan pada aplikasi *data mining* ini.

Kata kunci: Aplikasi, *Data mining*, *Association rules*, algoritma *fp-growth*, tingkat kelulusan

1 Pendahuluan

Perguruan tinggi saat ini dihadapkan pada kondisi kompetisi yang ketat dengan berbagai tuntutan kualitas penyelenggaraan oleh Direktorat Jenderal Pendidikan (Dikti). Kondisi ini membuat perguruan tinggi harus memiliki keunggulan kompetitif dan kualitas pelayanan pelanggan agar dapat tetap bersaing dengan perguruan-perguruan tinggi lainnya. Sistem informasi adalah merupakan salah satu sumber daya yang dapat meningkatkan keunggulan bersaing. Sistem informasi dapat digunakan untuk mendapatkan, mengolah, dan menyebarkan informasi untuk menunjang kegiatan sehari-hari serta dapat digunakan untuk pengambilan keputusan strategis.

Tersedianya suatu informasi tidak terlepas dari melimpahnya data, dari pengolahan dan penggalian data tersebut informasi dapat diperoleh. Salah satu teknologi yang berkaitan dengan mengolah data menjadi informasi yang berguna adalah *data mining*. *Data mining* sendiri merupakan istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. Teknologi *data mining* memiliki berbagai macam algoritma dan teknik untuk penyelesaian masalah yang berhubungan dengan data.

Salah satu teknik pada *data mining* adalah teknik asosiasi. Teknik asosiasi adalah teknik *data mining* untuk menemukan aturan asosiatif antara

suatu kombinasi item. Teknik ini merupakan salah satu yang paling menarik dalam penggalian data. Dikatakan menarik karena teknik ini akan memberikan gambaran mengenai sejumlah sifat atau atribut tertentu yang sering muncul bersamaan dalam sebuah kumpulan data.

Frequent Pattern-Growth (FP-Growth) merupakan salah satu alternatif algoritma teknik asosiasi dalam *data mining*. *FP-Growth* dapat digunakan untuk menentukan himpunan data yang paling sering muncul dalam sebuah kumpulan data.

Dari uraian di atas, maka dapat dibuat sebuah aplikasi untuk menganalisis tingkat kelulusan melalui teknik *data mining* menggunakan algoritma *FP-Growth*. Diharapkan aplikasi ini dapat membantu dalam pengambilan keputusan oleh pihak Politeknik Negeri Malang untuk dapat mengetahui sejauh mana tingkat keberhasilan jalur masuk, tingkat keberhasilan mahasiswa dengan asal sekolah tertentu, tingkat keberhasilan jurusan atau program studi, tingkat keberhasilan berdasarkan asal kota, dan tingkat kesesuaian pekerjaan yang didapat oleh mahasiswa dengan jurusan yang diambil dengan menggunakan tolak ukur dari tingkat kelulusan.

2 Landasan Teori

2.1 Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data mining* adalah

proses yang menggunakan teknik statistic, matematika dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Turban, dkk. 2005).

Data mining bukanlah suatu bidang yang sama sekali baru. Salah satu kesulitan untuk mendefinisikan data mining adalah kenyataan bahwa data mining mewarisi banyak aspek teknik dari sebidang-bidang ilmu yang sudah mapan terlebih dahulu. Gambar 2.1 menunjukkan bahwa data mining memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (artificial Intelligent), mechine learning, statistic, database, dan juga information retrieval (Pramudiono, 2006)

2.2 Association Rules

Association rules (aturan asosiasi) atau analisis asosiasi dikenal juga sebagai salah satu metode data mining yang menjadi dasar dari berbagai metode data mining lainnya. Khususnya salah satu tahap dari analisis asosiasi yang disebut analisis pola frekuensi tinggi (frequent pattern mining) menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien. Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, support (nilai penunjang) yaitu prosentase kombinasi item tersebut dalam database dan confidence (nilai kepastian) yaitu kuatnya hubungan antar item dalam aturan asosiatif. Analisis asosiasi didefinisikan suatu proses untuk menemukan semua aturan asosiatif yang memenuhi syarat minimum untuk support (minimum support) dan syarat minimum untuk confidence (minimum confidence) (Pramudiono, 2007).

Untuk mendapatkan nilai support sebuah item A dapat diperoleh dari rumus berikut:

$$S = \frac{\sum(\text{Transaksi item A})}{\sum(\text{Transaksi})} \quad (1)$$

Sementara itu, untuk mencari nilai support dari 2-item dapat diperoleh dari rumus berikut:

$$S(A, B) = \frac{\sum(\text{Transaksi item A dan B})}{\sum(\text{Transaksi})} \quad (2)$$

Sedangkan untuk mendapatkan nilai confidence dapat diperoleh dari rumus berikut

$$C(A, B) = \frac{\sum(\text{Transaksi item A dan B})}{\sum(\text{Transaksi A})} \quad (3)$$

2.3 FP-Grwoth

Algoritma FP-Growth adalah salah satu alternative algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul. Karakteristik algoritma FP-Grwoth adalah struktur data yang digunakan adalah tree yang disebut

FP-Tree. Dengan menggunakan FP-Tree, algoritma FP-Growth dapat langsung mengekstrak frequent itemset dari FP-Tree.

Penggalian itemset yang frequent dengan menggunakan algoritma FP-Growth akan dilakukan dengan cara membangkitkan struktur data tree atau disebut dengan FP-Tree. Metode FP-Growth dapat dibagi menjadi 3 tahapan utama yaitu: (Jiawei Han, Kamber M, 2006)

1. Tahap pembangkitan conditional pattern base
2. Tahap pembangkitan conditional FP-tree
3. Tahap pencarian frequent itemset

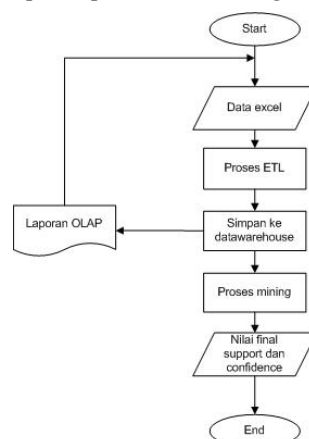
3 Perancangan Sistem

3.1 Deskripsi Sistem

Aplikasi data mining yang dibangun adalah aplikasi berbasis desktop dengan menggunakan database lokal. Rencana pengguna dari aplikasi ini adalah bagian akademik Politeknik Negeri Malang, sehingga dari aplikasi ini diharapkan dapat memberikan informasi terkait dengan tingkat kelulusan mahasiswa diantaranya adalah dapat mengetahui sejauh mana tingkat keberhasilan proses masuk mahasiswa, tingkat keberhasilan mahasiswa dengan sekolah tertentu, tingkat keberhasilan jurusan atau program studi, dan tingkat keberhasilan berdasarkan asal kota. Tingkat keberhasilan tersebut diukur dengan IPK (Indeks Prestasi Kumulatif) yang didapat mahasiswa. Selain itu dari aplikasi ini diharapkan juga dapat memberikan informasi tentang alumni yaitu tingkat kesesuaian pekerjaan yang didapat alumni dengan jurusan yang diambil.

3.2 Flowchart Sistem

Berikut adalah flowchart sistem secara keseluruhan pada aplikasi data mining ini.



Gambar 3.1 Flowchart sistem

3.3 Penggunaan FP-growth

Berikut adalah salah satu contoh penggunaan algoritma fp-growth pada aplikasi data mining ini, yaitu untuk mengetahui tingkat kelulusan berdasarkan proses masuk. Diketahui threshold atau batas nilai ambang adalah = 2.

Tabel 3.1 Data transaksi

NIM	IPK	PROSES MASUK
1141180001	A1	PSB/PMDK
1141180002	A1	PSB/PMDK
1141180003	A2	PSB/PMDK
1141180004	A1	PSB/BIDIK MISI
1141180005	A3	BARU UMPOL D4
1141180006	A2	BARU UMPOL D4
1141180007	A2	PSB/BIDIK MISI
1141180008	A1	PSB/BIDIK MISI
1141180009	A1	BARU UMPOL D3
1141180010	A2	BARU UMPOL D4
1141180011	A1	PSB/BIDIK MISI

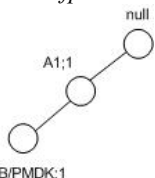
Tabel 3.2 Frekuensi kemunculan item

ITEM	FREKUENSI
A1	6
A2	4
PSB/BIDIK MISI	4
PSB/PMDK	3
BARU UMPOL D4	3

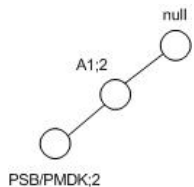
Tabel 3.3 Data transaksi terorder

TID	ITEMSET
1141180001	{A1; PSB/PMDK}
1141180002	{A1; PSB/PMDK}
1141180003	{A2; PSB/PMDK}
1141180004	{A1; PSB/BIDIK MISI}
1141180005	{BARU UMPOL D4}
1141180006	{A2; BARU UMPOL D4}
1141180007	{A2; PSB/BIDIK MISI}
1141180008	{A1; PSB/BIDIK MISI}
1141180009	{A1}
1141180010	{A2; BARU UMPOL D4}
1141180011	{A1; PSB/BIDIK MISI}

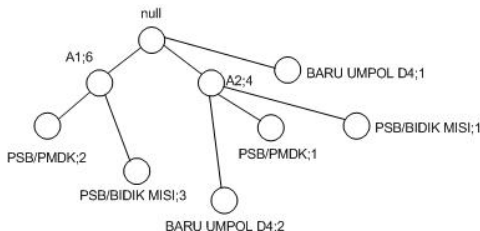
Setelah didapatkan data transaksi yang sesuai dengan nilai batas ambang atau nilai *threshold*, selanjutnya adalah proses *fp-tree*.



Gambar 3.2 tree setelah tid 1

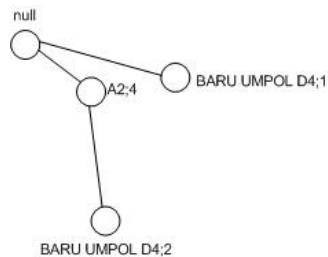


Gambar 3.3 tree setelah tid 2

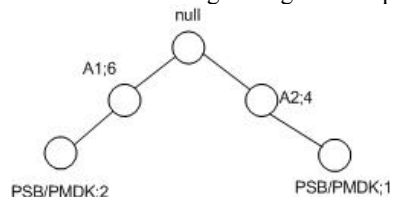


Gambar 3.4 tree setelah tid ke 11

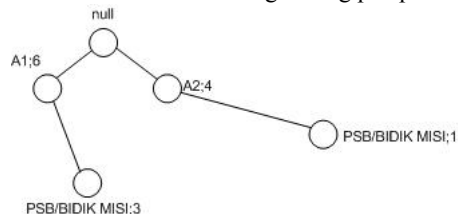
Selanjutnya langkah algoritma *fp-growth*. Berikut adalah tahap *conditional pattern base*.



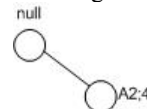
Gambar 3.5 lintasan mengandung baru umpol D4



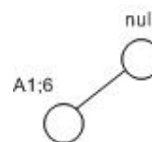
Gambar 3.6 lintasan mengandung psb/pmdk



Gambar 3.7 lintasan mengandung psb/bidik misi

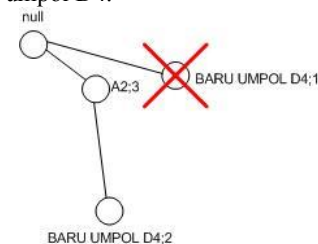


Gambar 3.8 lintasan mengandung A2



Gambar 3.9 lintasan mengandung A1

Setelah tahap *conditional pattern base*, selanjutnya adalah tahap *conditional fp-tree*. Pada tahap ini nilai item pada *tree* yang kurang dari batas nilai ambang atau nilai *threshold* akan dihapus. Berikut salah satu contoh *conditional fp-tree* pada item baru umpol D4.



Gambar 3.10 Conditional fp-tree baru umpol D4

Didapat *frequent item set* pada item baru umpol D4 adalah {A2; Baru Umpol d4}. Berikut hasil *frequent item set* secara keseluruhan.

Tabel 3.4 Frequent itemset

SUFFIX	FREQUENT ITEMSET
BARU UMPOLD4	{A2;BARU UMPOLD4}
PSB/PMDK	{A1;PSB/PMDK}
PSB/BIDIK MISI	{A1;PSB/BIDIK MISI }
A2	{A2}
A1	{A1}

Setelah didapat *frequent itemset* atau kombinasi item yang paling sering muncul. Selanjutnya adalah menghitung nilai *support* dan *confidence*.

Tabel 3.5 Hasil perhitungan

Frequent Item	Nilai Perhitungan	Hasil
{A1;PSB/BIDIK MISI}	Support	27.27 %
	Confidence	50.00 %
	Final	13.64 %
{A1;PSB/PMDK}	Support	18.18 %
	Confidence	33.33 %
	Final	6.06 %
{A2;BARU UMPOLD4}	Support	18.18 %
	Confidence	50.00 %
	Final	9.09 %

3.4 Perancangan Struktur table

Proses *mining* pada aplikasi ini adalah menggunakan *store procedure* pada *sql server* sedangkan implementasi dari algoritma *fp-growth* adalah menggunakan struktur *temporary* tabel. Alasan dari menggunakan *temporary* tabel adalah karena proses-proses *mining* tidak perlu disimpan permanen didalam *database* sehingga membuat *database* menjadi berat, jadi setiap eksekusi *mining* akan membuat *temporary* tabel. Berikut adalah struktur *temporary* tabel pada algoritma *fp-growth*.

Nama Field	Tipe data	Keterangan
tid	int	
item	varchar(1000)	

Gambar 3.10 Struktur tabel tabel_trans_fp

Nama Field	Tipe data	Keterangan
item	varchar(100)	
support	int	

Gambar 3.11 Struktur tabel tabel_total_count

Nama Field	Tipe data	Keterangan
tid	int	
item	varchar(100)	

Gambar 3.12 Struktur tabel tabel_sorted_trans

Nama Field	Tipe data	Keterangan
item	varchar(100)	
cnt	int	
path	varchar(1000)	

Gambar 3.13 Struktur tabel tabel_before_tree

Nama Field	Tipe data	Keterangan
item	varchar(100)	
cnt	int	
path	varchar(1000)	

Gambar 3.14 Struktur tabel tabel_fp_tree

Nama Field	Tipe data	Keterangan
tid	varchar(50)	
id	int	
item	varchar(100)	
cnt	int	

Gambar 3.15 Struktur tabel tabel_pattern_base

Nama Field	Tipe data	Keterangan
prefix	varchar(50)	
item	varchar(100)	
cnt	int	

Gambar 3.16 Struktur tabel tabel_cond_pattern

Nama Field	Tipe data	Keterangan
item	varchar(50)	
fp	varchar(1000)	
cnt	Int	

Gambar 3.17 Struktur tabel tabel_freq_pattern

Nama Field	Tipe data	Keterangan
item	varchar(50)	
fp	varchar(1000)	
cnt	Int	

Gambar 3.18 Struktur tabel tabel_temp_fp

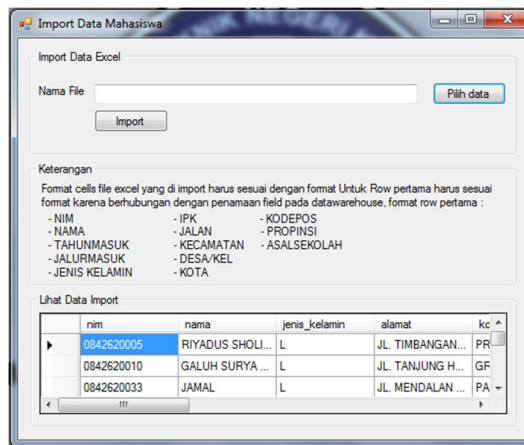
4 Implementasi

4.1 Implementasi Program

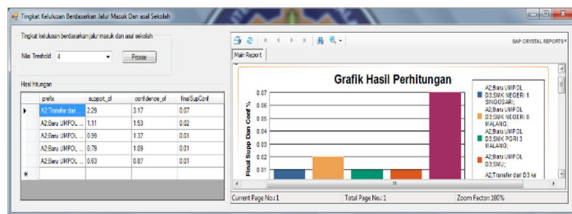
Implementasi program adalah tampilan hasil dari perancangan dan analisis sistem yang telah dibuat, berikut adalah tampilan hasil dari perancangan dan analisis sistem program.



Gambar 4.1 Borang utama



Gambar 4.2 Borang import data



Gambar 4.3 Salah satu borang perhitungan

5 Pengujian dan Pembahasan

Pengujian pada sistem ini meliputi beberapa jenis pengujian, yaitu pengujian fungsional, pengujian metode.

5.1 Pengujian Fungsional

Pengujian fungsional dilakukan untuk mengetahui apakah fitur-fitur dari aplikasi *mining* ini berjalan dengan semestinya atau tidak. Teknik yang digunakan dalam pengujian fungsional ini adalah teknik *blackbox*.

Berdasarkan hasil pengujian fungsional dapat ditarik kesimpulan bahwa aplikasi *mining* ini telah berjalan sesuai dengan harapan dan dapat menampilkan hasil perhitungan dari masing-masing proses *mining*.

5.2 Pengujian Metode

Selain melakukan pengujian fungsional pada aplikasi dilakukan juga pengujian metode dan algoritma pada aplikasi, yaitu metode *association rule* dan algoritma *fp-growth*. Pengujian dilakukan dengan cara percobaan secara manual dan percobaan terkomputerisasi pada aplikasi. Jika hasil perhitungan sesuai atau sama, baik secara manual dan terkomputerisasi maka dapat disimpulkan bahwa metode dan algoritma pada aplikasi *mining* ini berjalan sesuai dengan harapan. Berikut tabel hasil pengujian metode.

Tabel 5.1 hasil pengujian metode

Frequent Item	Nilai Perhitungan	Percobaan Manual	Percobaan Aplikasi
{A1;PSB/BIDIK MISI}	Support	27.27 %	27.27 %
	Confidence	50.00 %	50.00 %
	Final	13.64 %	13.64 %
{A1;PSB/PMDK}	Support	18.18 %	18.18 %
	Confidence	33.33 %	33.33 %
	Final	6.06 %	6.06 %
{A2;BARU UMPOLD4}	Support	18.18 %	18.18 %
	Confidence	50.00 %	50.00 %
	Final	9.09 %	9.09 %

Berdasarkan tabel hasil percobaan diatas dapat ditarik kesimpulan bahwa metode *association rule* dan algoritma *fp-growth* berjalan sesuai harapan dan berhasil diimplementasikan pada aplikasi *data mining* ini, karena perhitungan yang dihasilkan dari percobaan manual dan percobaan aplikasi cocok 100%.

6 Kesimpulan dan Saran

6.1 Kesimpulan

Setelah melakukan analisis, perancangan, implementasi beserta pengujian yang dilakukan, maka dapat ditarik beberapa kesimpulan, yaitu :

1. Algoritma *fp-growth* dapat diimplementasikan pada aplikasi *mining* ini, dari hasil analisis, perancangan, implementasi dan pengujian terbukti algoritma *fp-growth* dapat menampilkan informasi tingkat kelulusan berdasarkan asal sekolah, program studi, asal kota, proses masuk, serta dapat menampilkan informasi tingkat kesesuaian pekerjaan yang didapat alumni berdasarkan program studi.
2. Hasil dari aplikasi *mining* ini dapat digunakan sebagai pertimbangan dalam pengambilan sebuah keputusan berdasar pada tingkat kelulusan. Contoh keputusan penambahan kuota pada jalur masuk tertentu dilihat dari tingkat kelulusan.

6.2 Saran

Berdasarkan penelitian yang dilakukan, berikut saran untuk pengembangan aplikasi *data mining* ini lebih lanjut, yaitu aplikasi diubah menjadi *web based application*, karena penyajian laporan OLAP lebih dinamis dan lebih menarik dibandingkan *desktop application*.

Daftar Pustaka:

Alves, Ronnie. 2007. *Programming Relational Databases For Itemset Mining Over Large Transaction Tables*. Portugal : University of Minho.

Erwin. 2009. Analisis Market Basket Dengan Algoritma Apriori dan FP-Growth. Palembang: Universitas Sriwijaya.

Fathimah Fatihatul, Atje Setiawan, Rudi Rosadi. 2011. Asosiasi Data Mining Menggunakan Algoritma FP-Growth Untuk Market Basket Analysis. Jatinagor: Universitas Padjadjaran.

Hermawati, Fajar Astuti. 2013. *Data Mining*. Yogyakarta: Penerbit Andi.

Huda, Nuqson Masykur. 2010 *Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa (Studi Kasus di Fakultas MIPA Universitas Diponegoro)* Semarang : Universitas Diponegoro

Kusrini, dan Emha Taufiq L. 2009. *Algoritma Data Mining*. Yogyakarta: Penerbit Andi.

Ririanti. 2014. Implementasi Algoritma *FP-Growth* Pada Aplikasi Prediksi Persediaan Sepeda Motor (Studi Kasus PT.Pilar Deli Labumas). Medan: STMIK Budi Darma Medan.

Samuel, David. 2007. Penerapan Struktur FP-Tree dan Algoritma FP-Growth dalam Optomasi Penentuan Frequent Itemset. Bandung : Institut Teknologi Bandung.

Susanto, Sani, dan Dedy Suryadi. 2010. *Pengantar Data Mining*. Yogyakarta: Penerbit Andi.