

PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA FP-GROWTH UNTUK PENENTUAN STRATEGI PROMOSI FAKULTAS ILMU KOMPUTER UNIVERSITAS KUNINGAN

Heru Budianto¹⁾, Jejen Riana²⁾

¹⁾ Sistem Informasi Universitas Kuningan

²⁾ Sistem Informasi Universitas Kuningan

Jl Cut Nyak Dien No 36 A Cijoho Kabupaten Kuningan

Email : heru.budianto@uniku.ac.id¹

Abstrak

Dewasa ini, persaingan antar kampus meningkat dalam rangka menjaring sebanyak mungkin mahasiswa baru. Hal tersebut berimbas pada pengelolaan di tingkat fakultas. Dana dan tuntutan efisiensi yang terbatas menyebabkan pengelola di tingkat fakultas berusaha untuk menemukan strategi yang tepat untuk mencapai target mahasiswa yang telah ditentukan. Strategi yang tepat mampu meminimalisir penggunaan dana dan mencapai target yang telah ditetapkan. Penelitian ini bertujuan untuk melakukan analisis dengan teknik data mining dengan bantuan algoritma FP-Growth. Algoritma FP-Growth yang merupakan perkembangan dari algoritma Apriori. FP-Growth adalah salah satu alternatif algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (frequent item set) dalam sekumpulan data. Penelitian dilakukan dengan mengamati beberapa variabel penelitian yang sering dipertimbangkan oleh fakultas dalam menentukan sasaran promosinya, Yaitu Pendidikan Terakhir, Jurusan, Pilihan Prodi. Hasil penelitian ini adalah berupa sebuah rekomendasi penentuan strategi dengan menggunakan algoritma FP-Growth yang menggunakan konsep pembangunan FP-Tree dalam mencari Frequent Itemset
Kata Kunci : Data Mining, Association Rules, FP-Growth, FKOM, Universitas Kuningan

Abstract

Nowadays, competition between campuses is increasing in order to attract as many new students as possible. This has an impact on management at the faculty level. Limited funding and efficiency demands cause managers at the faculty level to try to find the right strategy to achieve the student targets that have been determined. The right strategy is able to minimize the use of funds and achieve the targets set. This study aims to analyze data mining techniques with the help of the FP-Growth algorithm. FP-Growth algorithm which is a development of the Apriori algorithm. FP-Growth is one alternative algorithm that can be used to determine the set of data that most often appears (frequent item sets) in a data set. The study was conducted by observing several research variables that are often considered by faculties in determining the promotion objectives, namely the Latest Education, Departments, Study Program Choices. The results of this study are in the form of a strategy determination recommendation using the FP-Growth algorithm that uses the concept of FP-Tree development in searching for Frequent Itemset

Keyword: Data Mining, Association Rules, FP-Growth, FKOM, Kuningan University

1. Pendahuluan

Saat ini banyak bermunculan Universitas, fakultas, serta program studi baru menimbulkan persaingan yang sangat ketat. Setiap Universitas sampai tingkat program studi berupaya untuk memberikan peningkatan fasilitas yang terbaik dalam upaya mempromosikan baik universitas, fakultas maupun program studi.

Fakultas Ilmu Komputer Universitas Kuningan memiliki 4 program studi yaitu Teknik Informatika S1, Sistem Informasi S1, Desain Komunikasi Visual S1 dan Manajemen Informatika D3 berupaya untuk meningkatkan jumlah mahasiswa barunya maupun mempertahankan jumlah dalam rangka mempertahankan nilai akreditasi, terutama salah satu program studi yang jarang diminati sehingga jumlah mahasiswa tidak sesuai dengan target. Hal tersebut mengakibatkan biaya yang dikeluarkan untuk program studi tersebut menjadi tidak seimbang.

Untuk itu diperlukan sebuah kajian guna peningkatan efisiensi dan efektifitas untuk menetapkan sasaran promosi yang tepat. Dengan adanya sasaran dan metode yang tepat maka biaya dapat diminimalisir. Peningkatan efisiensi dapat dilakukan dengan pengolahan data-data yang telah ada untuk mengetahui informasi baru berupa pola tersembunyi yang dapat digunakan untuk keperluan promosi. Pengolahan data tersebut disebut sebagai data mining. [1] Setiawan 2016

Association rule dipergunakan dalam penelitian ini dengan bantuan algoritma *FP-Growth*. Hasil dari pengolahan data dengan teknik tersebut

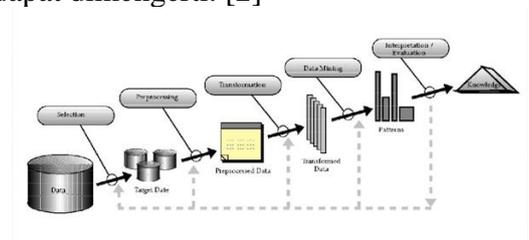
adalah pola-pola yang memiliki hubungan antara faktor didalamnya.

Penelitian ini bertujuan untuk mempercepat penyeleksian dalam proses penentuan frequent itemset sehingga dapat membantu pihak FKOM untuk merancang ulang strategi promosi yang akan dan telah dilakukan.

2. Landasan Teori

a. Knowledge Discovery In Database (KDD)

Knowledge Discovery in Databases berkaitan dengan proses penemuan pengetahuan yang diterapkan pada database. Hal ini juga didefinisikan sebagai proses *non-trivial* untuk identifikasi data yang valid, baru, berpotensi bermanfaat, dan akhirnya memiliki pola yang dapat dimengerti. [2]



Gambar 1 Tahapan KKD

Proses KDD meliputi beberapa tahapan seperti gambar 1, yaitu :

1) Data selection

Tahapan pemilihan dan pengambilan data yang akan dijadikan sebagai data penelitian dan akan disimpan pada database yang terpisah.

2) Data Cleaning

Dalam proses pengolahan data, dihindarkan data yang tidak konsisten dan cenderung menimbulkan

anomali sehingga perlu dilakukan pembersihan data.

3) Data Integration

Proses penggabungan dilakukan untuk mendapatkan data yang tepat dan seringkali diperlukan peringkasan.

4) Data Transformation

Proses perubahan menjadi sebuah bentuk yang lebih tepat dan sesuai dengan proses yang akan dilakukan.

b. Data mining

Data Mining adalah proses untuk mendapatkan informasi dengan melakukan pencarian pola dan relasi-relasi yang tersembunyi di dalam timbunan data yang banyak [3]. Data Mining atau sering disebut sebagai *knowledge discovery in database* (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar. Keluaran Data Mining ini bisa dipakai untuk membantu pengambilan keputusan di masa depan. Pengembangan KDD ini menyebabkan penggunaan *pattern recognition* semakin berkurang karena telah menjadi bagian Data Mining.

Beberapa survey tentang proses pemodelan dan metodologi menyatakan bahwa, "Data mining digunakan sebagai penunjuk, dimana data mining menyajikan intisari atas sejarah, deskripsi dan sebagai standar petunjuk mengenai masa depan dari sebuah proses model data mining" [4]

Karakteristik data mining sebagai berikut:

- a. Data mining berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
- b. Data mining biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dapat dipercaya.
- c. Data mining berguna untuk membuat keputusan kritis.

c. Tahapan Asosiasi

Analisis asosiasi dikenal juga sebagai salah satu teknik Data Mining yang menjadi dasar dari berbagai teknik 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 persentase kombinasi item tersebut dalam database dan *confidence* [3].

Dasar analisis asosiasi terbagi menjadi 2 tahap yaitu analisa frekuensi tinggi dan pembentukan aturan asosiatif.

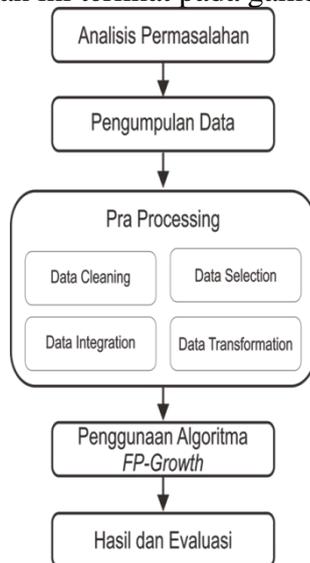
d. Algoritma FP-Growth

FP-Growth menggunakan pendekatan yang berbeda dari paradigma yang digunakan pada algoritma Apriori. [5]. FP-Growth adalah salah satu alternatif algoritma

yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (frequent item set) dalam sekumpulan data. Algoritma FP-Growth merupakan pengembangan dari algoritma Apriori. FP-growth adalah metode yang sering itemset pertambahan tanpa Generasi calon. Ini membangun sebuah struktur data yang sangat padat (FP-tree) untuk kompres database transaksi asli. Algoritma FP-Growth menggunakan struktur data tree yang disebut dengan FP-Tree.

3. Metodologi

Dalam melakukan sebuah penelitian diperlukan kaidah-kaidah atau metode yang akan menjadi petunjuk setiap langkah yang akan dikerjakan dalam penelitian ini. Adapun metode yang diterapkan dalam penelitian ini terlihat pada gambar 2



Gambar. 2 Metode Penelitian

Berdasarkan gambar 2 dapat dijelaskan tahapan yang akan dilakukan dalam penelitian ini yaitu :

1. Analisis data

Langkah-langkah dalam proses analisis meliputi pengidentifikasian masalah, analisis, dan penentuan tujuan. Permasalahan yang akan diidentifikasi dalam penelitian ini berfokus pada bagaimana mencari strategi promosi yang efektif dan efisien yang akan dipergunakan oleh Fakultas Ilmu Komputer dalam mendapatkan calon mahasiswa baru.

Adapun tujuan dari penelitian ini adalah :

- a. Melakukan pengelompokan mengenai sasaran dan target promosi Fakultas Ilmu Komputer Universitas Kuningan
 - b. Mendapatkan pengetahuan tentang pemilihan objek dan sasaran yang akan dilakukan promosi
 - c. Memberikan masukan kepada pihak fakultas sebagai bahan pertimbangan target promosi penerimaan mahasiswa baru
2. Pengumpulan data
Metode pengumpulan data dilakukan dengan cara :
 - a. Observasi atau pengamatan langsung di Fakultas Ilmu Komputer Universitas Kuningan. Pengamatan dilakukan langsung dalam proses promosi yang dilakukan selama tahun 2017 ke sekolah-sekolah baik yang ada di wilayah kuningan atau di luar daerah kuningan.
 - b. Form isian.

Sebagian besar data yang diambil diperoleh dari basis data yang dimiliki

oleh Pusat Informasi Universitas Kuningan dalam Sistem Informasi Akademik pada tahun 2016. Data yang diambil hanya data mahasiswa Fakultas Ilmu Komputer. Untuk memperoleh data lengkap diperlukan form tambahan untuk melengkapi basis data yang ada.

3. Pre-Processing

a. Data Cleaning

Bertujuan untuk menghilangkan data-data yang memiliki duplikasi, data yang kosong, dan tidak konsisten dari data yang telah terkumpul. Hasil dari data cleaning didapatkan seperti pada tabel 1

Tabel 1 hasil *data cleaning*

b. Data Integration

Pada tahap ini dilakukan proses penggabungan dengan data lainnya yang digunakan sebagai pelengkap, yaitu data yang berasal dari form isian. Sehingga data yang akan diproses menjadi seperti tabel 2

Tabel 2 Data Integration

Jurusan	Asal SMA	PRODI
IPA	a SMA	Sistem
	t NEGERI	Informasi
IPS	a SMA	Sistem
	NEGERI	Informasi
TKJ	s SMK	Sistem
	e SWASTA	Informasi
IPS	l SMA	Sistem
	e NEGERI	Informasi
MM	c SMK	Sistem
	t NEGERI	Informasi
TKRO	i SMK	Sistem
	o SWASTA	Informasi
IPA	n SMA	Sistem
	D NEGERI	Informasi
IPS	a SMA	Sistem
	t NEGERI	Informasi
TKJ	a SMK	Sistem
	SWASTA	Informasi
IPS	y SMA	Sistem
	a NEGERI	Informasi

Jurusan	Asal SMA	dst	PRODI
IPA	SMA		Sistem
	NEGERI		Informasi
IPS	SMA		Sistem
	NEGERI		Informasi
TKJ	SMK		Sistem
	SWASTA		Informasi
IPS	SMA		Sistem
	NEGERI		Informasi
MM	SMK		Sistem
	NEGERI		Informasi
TKRO	SMK		Sistem
	SWASTA		Informasi
IPA	SMA		Sistem
	NEGERI		Informasi
IPS	SMA		Sistem
	NEGERI		Informasi
TKJ	SMK		Sistem
	SWASTA		Informasi
IPS	SMA		Sistem
	NEGERI		Informasi

dst
 ng didapat kemudian

diseleksi untuk mendapatkan data-data yang relevan. Adapun tabel 2 merupakan data hasil selection dari hasil pemilihan dan penghilangan beberapa atribut yang tidak diperlukan.

d. Data transformation

Dalam proses perhitungan beberapa metode memerlukan transformasi data menjadi bentuk lain untuk lebih memudahkan. Data hasil selection kemudian ditransformasikan dengan menetapkan terlebih dahulu kode. Hasil dari pengkodean didapatkan data seperti pada tabel 3
 Tabel 3 hasil pengkodean.

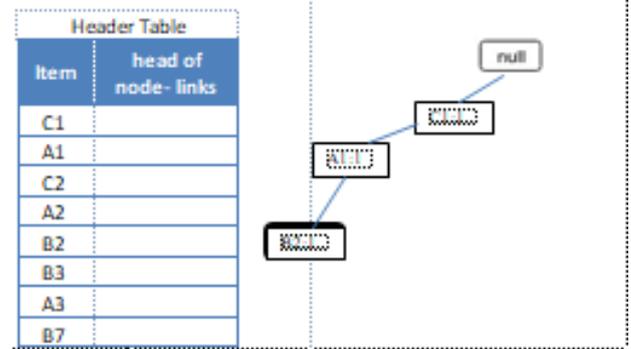
No.	Item
1	A1,B2,C1
2	A1,B3,C1
3	A3,B7,C1
4	A1,B3,C1
5	A2,B4,C1
6	A3,B9,C1
7	A1,B2,C1
8	A1,B3,C1
	A3,B7,C1
10	A1,B3,C1
	dst.

4. Hasil dan Pembahasan

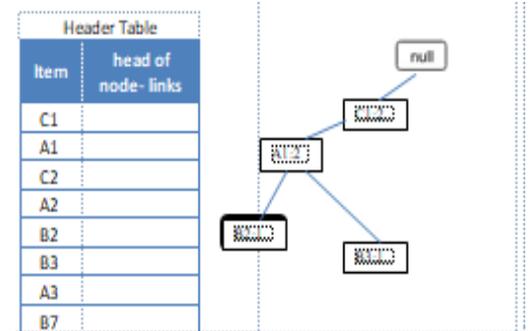
1. FP-Tree Construction

Pembentukan *FP-tree* dilakukan dengan cara pembacaan setiap data yang kemudian direpresentasikan

kedalam bentuk jalur pada *FP-tree*

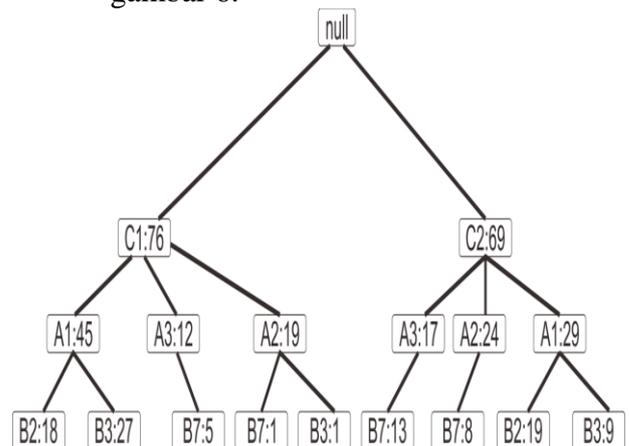


Gambar 3 FP-tree setelah pembacaan data 1



Gambar 5. FP-tree setelah pembacaan data ke 2

Setelah semua data direpresentasikan pada *FP-tree* maka akan didapatkan *FP-tree* yang telah lengkap seperti pada gambar 6.



Gambar 6. *FP-tree* setelah pembacaan data ke 147

2. Frequent Itemset generation

Pada tahap ini dilakukan 3 tahapan yaitu conditional

pattern base, conditional FP-tree dan terakhir frequent itemset. Didapatkan hasil seperti tabel 4.

item	conditional Pattern	Frequent Items
B7	{A2,C2:8},{A2,C1:1},{A3,C1:5},{A3,C2:13}	{B7,A3,C2}, {B7,A3},{B7,C2}
A3	{C1:12},{C2:17}	{A3,C2},{C2}
B3	{A1,C1:27},{A2,C1:1},{A1,C2:9}	{B3,A1,C1},{B3,A1}, {B3,A1,C2},{B2,A1},
B2	{A1,C1:18},{A1,C2:9}	{B2,A1,C2},{B2,A1},
A2	{C1:19},{C2:24}	{A2,C2},{C2}
C2	#	
A1	{C1:45},{C2:29}	{A1,C1},{C1}
C1	#	

Dari setiap kemungkinan diatas akan dilakukan perhitungan confidence dan expected confidence untuk melihat rekomendasi yang paling baik. Hanya 3 nilai dengan 3 variabel yang akan dilakukan perhitungan.

3. Confidence dan expected confidence

Nilai confidence dan expected confidence dilakukan dengan menggunakan rumus 1 dan 2.

Confidence

$$= \frac{\text{Jumlah transaksi anteseden dan konsekuen}}{\text{jumlah transaksi anteseden}} \dots 1$$

Expected Confidence

$$= \frac{\text{Jumlah transaksi yang konsekuen}}{\text{jumlah transaksi dalam database}} \dots 2$$

Hasil perhitungan confidence dan expected confidence dapat dilihat pada tabel 5.

Tabel 5. Expected confidence dan Confidence

item	expected confidence	confidence 2
{B7,A3,C2}	8,84	76,47
{B3,A1,C1}	18,37	60,00
{B2,A1,C2}	12,93	65,52

Sedangkan nilai dari lift ratio dari ketiga rekomendasi diatas dapat dilihat pada tabel 6.

item	Lift Ratio
{B7,A3,C2}	8,65
{B3,A1,C1}	3,27
{B2,A1,C2}	5,07

4. Hasil

Berdasarkan dari tabel diatas didapatkan rekomendasi terbaik yaitu promosi terbaik yaitu :

- Siswa jurusan TKJ yang berasal dari SMK swasta adalah promosi mengenai Pogram Studi Teknik Informatika dengan nilai confidence 76,47 %.
- Siswa jurusan IPS yang berasal dari SMA Negeri promosi yang harus dilakukan adalah tentang program studi Sistem Informasi dengan nilai confidence 60 %
- Siswa jurusan IPA yang berasal dari SMA Negeri promosi yang harus dilakukan adalah tentang program studi Teknik Informatika dengan nilai confidence 59,21%

Bila dilihat dari nilai lift ratio memberikan gambaran bahwa dari ketiga data tersebut mempunyai hubungan asosiasi yang kuat antar itemnya dan memiliki nilai kebermanfaatan yang tinggi karena memiliki nilai lebih dari 1.

Hasil rekomendasi ini bisa dipergunakan untuk penentuan strategi promsosi oleh FKOM UNIKU>

5. Kesimpulan

Kesimpulan dari penelitian ini adalah :

- a. Algoritma FP-Growth dapat digunakan untuk penentuan strategi promosi di FKOM UNIKU dengan memberikan rekomendasi terbaik promosi
- b. Rekomendasi yang dihasilkan yaitu untuk siswa asal jurusan TKJ yang berasal dari SMK swasta dan Jurusan IPA dari SMA Negeri, promosi yang harus dilakukan tentang Program Studi Teknik Informatika, sedangkan siswa yang berasal dari jurusan IPS dari SMA Negeri cenderung harus diberikan promosi mengenai program studi Sistem Informasi.

- [3] Mariscal, G., Marba, O., & Fernandez, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. Cambridge: Cambridge University Press.
- [4] RIRIANTI. 2014. Implementasi Algoritma FPGROWTH Pada Aplikasi Prediksi Persediaan Sepeda Motor (Studi Kasus PT. Pilar Deli Labumas). Pelita Informatika Budi Darma, Vol : VI No: 1

Daftar Pustaka

- [1] Setiawan, Rony. 2016. Penerapan Data mining menggunakan algoritma K-means Clustering untuk menentukan strategi promosi mahasiswa baru. Jurnal Lentera ICT Vol.3 No. 1 Mei 2016/ISSN 2338-3143
- [2] Kurgan, L.A. and Musilek, P. (2006) A Survey of Knowledge Discovery and Data Mining Process Models. The Knowledge Engineering Review, 21, 1-24.FADLINA. 2014. Data Mining Untuk Analisa Tingkat Kejahatan Jalanan Dengan Algoritma Association Rule Metode Apriori (Studi Kasus Di Polsekta Medan Sunggal). Informasi dan Teknologi Ilmiah (INTI), Vol : III No : 1, 144-154.