

## **DATA MINING UNTUK MENGGALI POLA MAHASISWA BARU MENGGUNAKAN METODE *FREQUENT PATTERN GROWTH* (STUDI KASUS : INSTITUT TEKNOLOGI ADHI TAMA SURABAYA)**

Budanis Dwi Meilani, Muhammad Asadulloh

Jurusan Teknik Informatika, Institut Teknologi Adhi Tama Surabaya

E-mail : [asamilan88@gmail.com](mailto:asamilan88@gmail.com)

### **ABSTRACT**

*Nowadays state and private universities rapidly develop. Some new rules have been applied to improve education in Indonesia. One of the examples is ITATS which will perform an online student registration. Thousands of students are eager to enter state universities. Besides, the students are required by their parents to enter universities. When the students fail in the entrance selection test at state universities, then the students look for alternative universities in order to continue their education. Frequent pattern growth is one of the algorithm alternatives which can be applied for determining frequent itemset in one data group. Frequent pattern growth takes a different approach from the paradigm used in Apriori Algorithm. Frequent pattern tree is a structure of compressed data storage. Frequent Pattern Tree is set up by mapping each transaction data to any particular trajectory in frequent pattern tree. The trajectory is possibly overwritten because in every mapped transaction, there may be a transaction that has the same items. The more transaction data that have the same items, the more effective the compression process with the structure of frequent pattern tree data. The outcome of the data analysis for all data in the year of 2013 – 2014 has a minimum support of 0.01% has 163 total item, and the application is able to process the data with minimum support of 0.005%.*

**Keyword :** Student Pattern, FP-Growth, FP-Tree, minimum support

### **ABSTRAK**

Di masa sekarang ini perguruan tinggi negeri maupun swasta mengalami perkembangan yang sangat pesat. Setiap tahunnya perguruan tinggi swasta berlomba-lomba untuk mendapatkan mahasiswa baru. Banyak cara yang dilakukan baik itu dengan cara promosi di koran, ditelevisi, brosur dan masih banyak lagi. Di dalam melakukan promosi khususnya di kampus ITATS banyak terjadi kendala terutama menentukan tempat untuk melakukan promosi. Terkadang kita tidak mengetahui pasar atau seberapa banyak mahasiswa yang mendaftar dengan kriteria yang ada. Untuk itu perlu dibuatkan program aplikasi untuk menghasilkan pola mahasiswa baru dengan menggunakan algoritma *Frequent pattern growth*. Algoritma *Frequent pattern growth* adalah pola asosiasi yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (*frequent itemset*) dalam sebuah kumpulan data. *Frequent pattern growth* menggunakan pendekatan yang berbeda dari paradigma yang digunakan pada *algoritma Apriori*. *Frequent pattern tree* merupakan struktur penyimpanan data yang dimampatkan. *Frequent pattern tree* dibangun dengan memetakan setiap data transaksi ke dalam setiap lintasan tertentu dalam *frequent pattern tree*. Hasil analisa data untuk semua data pada tahun 2013 – 2014 memiliki *minimum support* sebesar 0,01 % memiliki jumlah item sebanyak 163 item, serta aplikasi ini mampu memproses data dengan *minimum support* mencapai 0,005 %.

**Kata kunci:** Pola mahasiswa, FP - Growth, FP – Tree, asosiasi rule

### **PENDAHULUAN**

Di masa sekarang ini perguruan tinggi negeri maupun swasta mengalami perkembangan yang sangat pesat. Beberapa aturan baru telah diterapkan untuk memajukan dunia pendidikan di Indonesia. Salah satu contohnya adalah SNMPTN (Seleksi Nasional Masuk Perguruan Tinggi Negeri) yang diterapkan setelah lulus dari SMA. Banyak para siswa yang berkeinginan masuk di perguruan tinggi negeri. Ketika siswa mengikuti seleksi di perguruan tinggi negeri dan hasilnya dinyatakan tidak lolos, maka siswa mencari perguruan tinggi alternatif guna melanjutkan pendidikan mereka. ITATS merupakan salah satu perguruan tinggi swasta di Surabaya. Dimana

setiap tahunnya menerima mahasiswa baru. Banyak kendala di dalam promosi untuk dapat menarik perhatian calon pendaftar. Minimnya pengetahuan tentang kriteria siswa yang berminat untuk mendaftar di ITATS adalah faktor utama menentukan lokasi promosi. Untuk itu diperlukan sistem untuk mengetahui pola pendaftar mahasiswa baru di ITATS menggunakan metode *Frequent Pattern Growth*.

## TINJAUAN PUSTAKA

*Data mining* adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam *database*. Menurut (Turban *et al*, 2005) *data mining* merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan di dalam database besar. Menurut (Larose, 2006) *data mining* adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika.

### Pola Asosiasi

Menurut (Zhao *et. al*. 2003), pola asosiasi adalah salah satu teknik dalam penggalian data yang bertujuan untuk mengekstrak korelasi yang menarik, pola – pola yang sering muncul, hubungan kumpulan *item* di dalam suatu database yang berisi *record* transaksi. Terdapat dua hal utama yang melandasi teknik ini yaitu *support* dan *confidence*. *Support* dari suatu pola asosiasi didefinisikan sebagai persentase dari *record* X U Y terhadap seluruh jumlah transaksi di dalam *database*. *Support* dapat dihitung dengan rumus sebagai berikut:

$$Support(XY) = \frac{support\ of\ XY}{total\ number\ of\ transaction} \dots\dots\dots(1)$$

*Confidence* dari suatu *associationrule* didefinisikan sebagai persentase dari jumlah transaksi yang mengandung X U Y terhadap jumlah total transaksi yang mengandung X. *Confidence* dihitung menggunakan rumus sebagai berikut:

$$Confidence(X|Y) = \frac{support(XY)}{support(X)} \dots\dots\dots(2)$$

### Algoritma Frequent Pattern Growth

*Frequent Pattern Growth* adalah salah satu alternatif algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (*frequent itemset*) dalam sebuah kumpulan data. *Frequent pattern tree* merupakan struktur penyimpanan data yang dimampatkan. *Frequent pattern tree* dibangun dengan memetakan setiap data transaksi ke dalam setiap lintasan tertentu dalam *frequent pattern tree*. Karena dalam setiap transaksi yang dipetakan, mungkin ada transaksi yang memiliki *item* yang sama, maka lintasannya memungkinkan untuk saling menimpa. Semakin banyak data transaksi yang memiliki *item* yang sama, maka proses pemampatan dengan struktur data *frequent pattern tree* semakin efektif. Adapun *frequent pattern tree* adalah sebuah pohon dengan definisi sebagai berikut:

- a. *Frequent pattern tree* dibentuk oleh sebuah akar yang diberi label *null*, sekumpulan *sub-tree* yang beranggotakan *item-item* tertentu, dan sebuah tabel *frequent header*.
- b. Setiap simpul dalam *frequent pattern tree* mengandung tiga informasi penting, yaitu label *item*, menginformasikan jenis *item* yang direpresentasikan simpul tersebut, *support count*, merepresentasikan jumlah lintasan transaksi yang melalui simpul tersebut dan *pointer* penghubung yang menghubungkan simpul-simpul dengan label *item* sama antar-lintasan, ditandai dengan garis panah putus-putus.

- c. Masukkan ke dalam tabel *frequent header* yang terdiri dari dua *field*, yaitu nama item dan penghubung *node* utama, nilai pada *node* pertama didalam *frequent pattern tree* digunakan untuk mencari nama *item*.

### **FP-Growth**

Algoritma *FP-Growth* merupakan pengembangan dari algoritma Apriori. Sehingga kekurangan dari algoritma Apriori diperbaiki oleh algoritma *FP-Growth*. *Frequent Pattern Growth (FP-Growth)* adalah salah satu alternatif algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (*frequent itemset*) dalam sebuah kumpulan data. Pada algoritma Apriori diperlukan *generate candidate* untuk mendapatkan *frequent itemsets*. Akan tetapi, di algoritma *FP-Growth generate candidate* tidak dilakukan karena *FP-Growth* menggunakan konsep pembangunan tree dalam pencarian *frequent itemsets*. Hal tersebutlah yang menyebabkan algoritma *FP-Growth* lebih cepat dari algoritma Apriori.

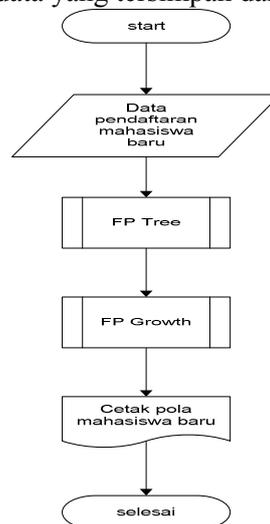
Karakteristik algoritma *FP-Growth* adalah struktur data yang digunakan adalah *tree* yang disebut dengan *FP-Tree*. Dengan menggunakan *FP-Tree*, algoritma *FP-growth* dapat langsung mengekstrak *frequent Itemset* dari *FP-Tree*.

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 sebagai :

- Tahap pembangkitan *conditional pattern base*,
- Tahap pembangkitan *conditional FP-Tree*, dan
- Tahap pencarian *frequent itemset*.

### **METODE**

Aplikasi yang akan dibuat bertujuan untuk mendapatkan pola assosiasi penerimaan mahasiswa baru ITATS berdasarkan data yang tersimpan dari tahun sebelumnya.



**Gambar 1. Flochart Utama**

### **Sumber Data**

Sumber data yang digunakan terdiri dari satu sumber data, yaitu data penerimaan mahasiswa baru. Data tersebut dibersihkan dan dipakai yang diperlukan untuk proses mining

Tabel 1. Data *Sample* Pendaftaran Mahasiswa Baru

Asal Sekolah	Kota Sekolah	Jurusan	Provinsi	Nama Sekolah
SMU	Bontang	Industri	Kaltim	SMA Pupuk Kaltim
SMK	Gresik	Informatika	Jatim	SMK Muhammadiyah 1
SMK	Surabaya	Mesin	Jatim	SMKN 2
SMU	Surabaya	Informatika	Jatim	SMKN 1
SMK	Surabaya	Geologi	Jatim	SMA Kr. Petra 5
SMK	Surabaya	Elektro	Jatim	SMK Kr. Petra
SMK	Mojokerto	Kimia	Jatim	SMK Taman Siswa
<i>Sambungan Tabel 1</i>				
SMU	Surabaya	Elektro	Jatim	SMK Kr. Petra
SMK	Surabaya	Informatika	Jatim	SMK 45
SMK	Surabaya	Elektro	Jatim	SMKN 3

### Transformasi data

Proses transformasi dilakukan untuk mempermudah dan mempercepat proses mining. Setelah semua proses transformasi data yang dilakukan, kemudian hasil transformasi dari semua data akan ditunjukkan pada tabel 2:

Tabel 2. Hasil Transformasi

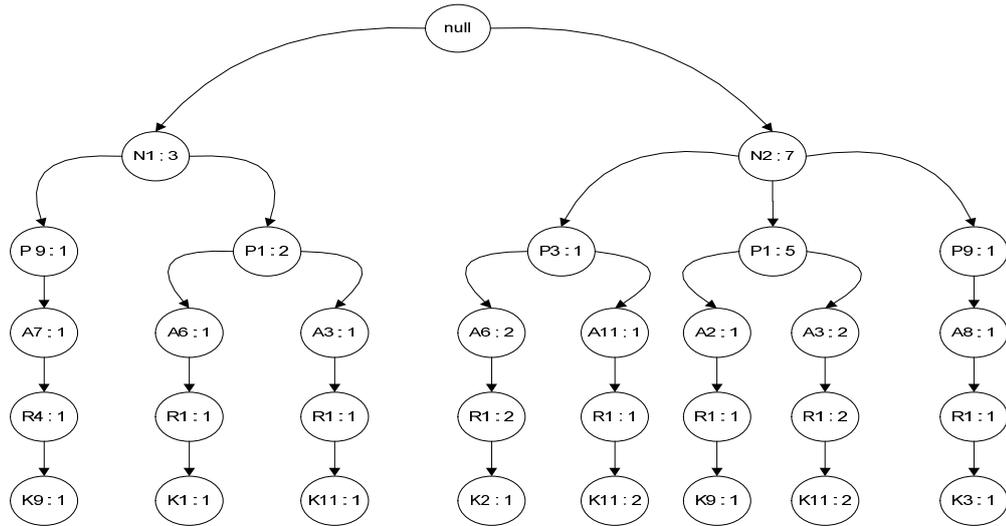
Asal Sekolah	Kota Sekolah	Jurusan	Provinsi	Nama Sekolah
N1	P9	A7	R4	K9
N2	P3	A6	R1	K11
N2	P1	A2	R1	K2
N1	P1	A6	R1	K1
N2	P1	A11	R1	K9
N2	P1	A3	R1	K11
N2	P9	A8	R1	K11
N1	P1	A3	R1	K11
N2	P1	A6	R1	K11
N2	P1	A3	R1	K3

### Proses Mining menggunakan FP-Growth

Tahapan pembentukan *FP-tree* setelah pembacaan transaksi 10 data, yang ada pada gambar 2 dibawah ini:

Proses FP-Growth dengan minimum support 20% dari 10 data transaksi menghasilkan variabel:

- N1,P1 : (SMU, Surabaya)
- N2,P1 : (SMK, Surabaya)
- N2,A3 : (SMK, Teknik Elektro)
- N2,A6 : (SMK, Teknik Informatika)
- N2,R1 : (SMK, Jawa Timur)
- N2,K11 : (SMK, SMK swasta lain)

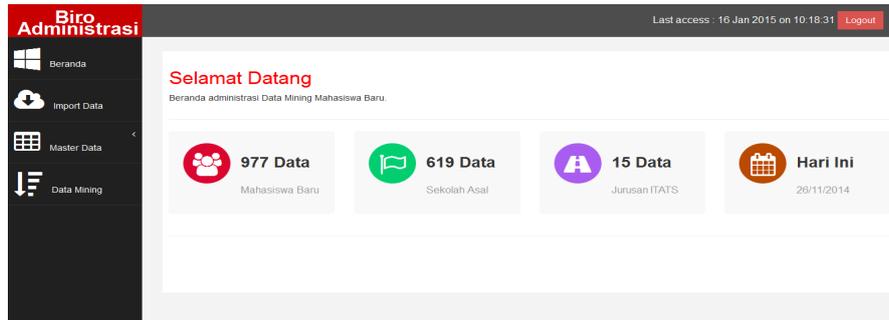


Gambar 2. Fp – Tree Pembacaan Transaksi 10 Data

## IMPLEMENTASI SISTEM

### Menu Utama

Menu utama pada program ini dapat dilihat pada gambar 3:



Gambar 3 Menu Utama

Pada gambar 3. tampilan *form* utama terdapat beberapa menu yaitu, menu utama, *import* data, master data, dan menu proses mining. Di menu master data terdapat tiga pilihan yaitu, data mahasiswa baru, data sekolah, dan data jurusan.

### Menu Master Mahasiswa Baru

Menu master data pada program ini meliputi data mahasiswa baru, sekolah, dan jurusan. Data mahasiswa baru dapat dilihat pada gambar 4 dibawah ini:

Advanced Tables

10 records per page Search:

NPM	NAMA MAHASISWA	ALAMAT MAHASISWA	TELP.	SEKOLAH	JURUSAN
01.2013.1.04585	Susilo	Tropodo Asri F-30 SDA	0898 5538243	SMK Bhina Tunas Bhakti	Sipil
01.2013.1.04586	Angga Santika Aji	Jl. Pagesangan Agung I/78 Surabaya	0857 30688308	SMKN 3 Surabaya	Sipil
01.2013.1.04587	Faris Fauzan	Jl. Griyo Benowo Indah II/K-2	0838 31430026	SMKN 5 Surabaya	Sipil
01.2013.1.04588	Nanang Sepdiantoro	Jl. Wonorejo Timur 150 Rungkut Surabaya	0852 31116060	SMKK Santo Yusuf Blitar	Sipil
01.2013.1.04589	Aliyan Azis Ra'uf	Jl. Jemursari V/31 Surabaya	0812 16361362	WL Seaton Secondary School	Sipil
01.2013.1.04590	Eko Mardi Setyo	Ds Sukorejo RT.07/02 Buduran Sidoarjo	0838 30381249	SMKN 1 Sidoarjo	Sipil
01.2013.1.04592	Ferdinandus Da Costa	Klampis Ngasem III Surabaya	0813 53350753	SMA Conis Santadia Leste	Sipil
01.2013.1.04593	Imron Hamzah	Jl. Wonocolo P Kulit 58-A Surabaya	0838 11332199	SMKN 3 Surabaya	Sipil
01.2013.1.04594	Rizki Riyanto	Jl. Kendangsari VIII/5-A Surabaya	0899 3912770	SMKN 3 Surabaya	Sipil
01.2013.1.04595	Syarifah Dhona Wardi	Kedurus IV-D Baru/91 Surabaya	0878 5114049	SMKN 5 Surabaya	Sipil

Showing 1 to 10 of 975 entries

Previous 1 2 3 4 5 ... 98 Next

Gambar 4. Menu Data Mahasiswa Baru

### Menu Sekolah

Menu sekolah pada program ini dapat dilihat pada gambar 5 dibawah ini:

View Data Sekolah

10 records per page Search:

No	Nama Sekolah	Jenis	Status	Alamat
1	SMK Bhina Tunas Bhakti	SMK	SWASTA	Semarang
2	SMAN 6 sby	SMA	NEGERI	Surabaya
3	SMK PGRI 3 Tanggul	SMK	SWASTA	Jember
4	SMA Siniolis Tidar	SMA	SWASTA	Surabaya
5	SMK Nasional Dawar Blandong	SMK	SWASTA	Lamongan
6	SMK ST Louis	SMK	SWASTA	Surabaya
7	SMA PGRI 1	SMA	SWASTA	Sidoarjo
8	SMK Kodikal Morokrembangan	SMK	SWASTA	Surabaya
9	SMK Santo Louis	SMK	SWASTA	Surabaya
10	PKBM Interaktif	Paket	SWASTA	Surabaya

Showing 1 to 10 of 619 entries

Previous 1 2 3 4 5 ... 62 Next

Gambar 5. Menu Sekolah

### Menu Proses Mining

Menu proses *mining* pada program ini dapat dilihat pada gambar 6 dibawah ini :

#### Data Mining Mahasiswa Baru

Mengetahui minat penjurusan sesuai dengan asal sekolah.

Data Sample yang dipergunakan	<input checked="" type="radio"/> Semua Data <input type="radio"/> Rentang Waktu
Minimum Support	Dalam Persen (%)
<input type="button" value="Proses Data Mining"/> <input type="button" value="Reset Data"/>	

Gambar 6. Menu Proses Mining

Gambar 7 menjelaskan bahwa sebelum diproses, kita pilih data *sample* berdasarkan rentang waktu yang telah ditentukan. Setelah kita proses dari tanggal 8 Maret 2013-8 Maret 2014, kemudian muncul 10 hasil dari *minimumsupport* yang sudah ditentukan.



Gambar 7. Proses Mining Berdasarkan Rentang Waktu PMA

**Hasil Analisa Data**

Tabel 3. Hasil Pola Mahasiswa Baru I

No.	Data Sample	Minimum Support (%)	Pola yang Dihasilkan	Jumlah Pola
1.	Semua Data Tahun 2013 – 2014	0,2	SMKN 5 Surabaya	10
			SMKN 2 Surabaya	4
			SMK PGRI 4 Surabaya	3
			SMK PGRI 1 Surabaya	3
			SMAN 1 Tual	3
			SMK Kristen Petra Sby	3
		0,4	SMKN 5 Surabaya	10
			SMKN 3 Surabaya	4
			SMK YPM 1 Taman Sidoarjo	4
		0,6	SMK Semen Gresik	5
			SMKN 5 Surabaya	6
SMKN 5 Surabaya	6			
SMKN 5 Surabaya	10			
SMKN 5 Surabaya	10			
0,8	SMKN 5 Surabaya	10		
	SMKN 5 Surabaya	10		
1,0	SMKN 5 Surabaya	10		
	SMKN 5 Surabaya	10		
1,1	-	-		

**KESIMPULAN**

Dari hasil analisa dan pembahasan yang dibuat, maka dapat disimpulkan bahwa :

1. Penentuan nilai *minimum support* akan memberikan hasil pola mahasiswa baru berbeda-beda sesuai dengan nilai *support*nya.
2. Hasil analisa data untuk tahun 2013 dengan minimum support sebesar 0,01 % memiliki jumlah *item* sebanyak 163 *item*.
3. Pada semua data pada tahun 2013 – 2014 terdapat banyak *item* dengan minimum support mencapai 1,0 %. Pada *minimum support* 0,8 % *item* SMKN 5 Surabaya berjumlah 10 *item*, sedangkan pada *minimum support* 1,0 % SMKN 5 Surabaya berjumlah 10 *item*.

**DAFTAR PUSTAKA**

Efraim, Turban. Rainer, Kelly R dan Potter, Richard. (2005). “*Introduction to Information Technology*”. 3<sup>rd</sup> Edition. USA : John Willey & Sons, Inc.

Fayyad, Usama. Piatetsky-Shapiro, G. Smyth, P dan Uthurusamy, R. (1996). *“Advances in Knowledge Discovery and Data Mining”*. Cambridge, MA : MIT Press.

J, Han. J, Pei dan Y, Yin. (2000). *“Mining Frequent Patterns Without Candidate Generation”*. In: Proceeding of The 2000 ACM-SIGMOD International Conference on Management of Data (SIGMOD'00). Dallas, TX : pp 1 - 12

Larose, Daniel T.. (2006). *“Data Mining Methods and Models”*. Hoboken New Jersey : John Willey & Sons, Inc.

Larose, Daniel T.. (2006). *“Discovering Knowledge in Data: An Introduction to Data Mining”*. USA : John Willey & Sons. Inc.

Marlinda, Linda. (2004). *“Sistem Basis Data/Linda Marlinda”*. Yogyakarta : Andi.

Yuswanto et. al. (2005). *“Mengolah Database dengan SQL Server 2000”*. Jakarta : Prestasi Pustakarya.

Zhao QianKun and Bhowmick S. Sourav (2003). *“Association Rule Mining : A Survey”*. CAIS Nanyang Technological University, Singapore.