

**ANALISA POLA ASOSIASI JALUR MASUK TERHADAP KELULUSAN
MAHASISWA DENGAN MENGGUNAKAN METODE *FOLD-GROWTH*
(STUDI KASUS FAKULTAS SAINS DAN TEKNOLOGI)**

Mohamad Irfan

Jurusan Teknik Informatika, Fakultas Sains dan Teknologi
Universitas Islam Negeri Sunan Gunung Djati Bandung
irfan.bahaf@uinsgd.ac.id

Abstrak

Teknologi komunikasi dan informasi telah berkembang seiring dengan tuntutan untuk mendapatkan informasi yang lebih spesifik dan terukur. Data yang diolah dijadikan pendukung untuk mengambil keputusan dengan memanfaatkan basisdata yang dimiliki untuk menentukan pola asosiasi jalur masuk calon mahasiswa terhadap kelulusan. Pola asosiasi antara data mahasiswa dan data kelulusan mahasiswa. Pola asosiasi antara dua jenis data tersebut dapat menunjukkan apakah ada keterkaitan antara variabel data mahasiswa yang terdiri dari jalur masuk dan Jurusan, dengan tingkat kelulusan mahasiswa yang bisa dilihat dari lama studi dan indeks prestasi kumulatif (IPK). Penelitian mencari pola asosiasi antara jalur masuk dengan kelulusan mahasiswa menggunakan metode FOLD-Growth. Dalam metode FOLD-Growth tahapan pertama yang dilakukan adalah mencari *frequent itemset*, pemangkasan item-item yang tidak *frequent*, pembangunan FP-Tree menggunakan transaksi yang telah dipangkas dan penggalian *itemset frequent*. Pencarian pola antara jalur masuk dengan kelulusan mahasiswa disertai nilai *support* dan nilai *confident* menghasilkan jalur masuk melalui tes tulis maka lama lulusnya lebih dari 4 tahun dengan IPK berada diantara 3.00 sampai 3.49 masuk dalam kategori B2 di jurusan Teknik Informatika, dengan nilai *support* 28,73. Jalur masuk tes PPA maka lulusnya lebih dari 4 tahun dengan IPK berada diantara 3,00 sampai 3,49 atau kategori B2 dengan nilai *support* 12,80.

Kata Kunci: *Asosiasi Rule, Itemset, FOLD-Growth, Mahasiswa, Data Mining*

1. Pendahuluan

1.1 Latar Belakang Masalah

Saat ini, sudah banyak instansi pemerintahan maupun swasta yang sudah

memakai sistem informasi untuk kegiatan sehari-hari. Sistem informasi dapat digunakan untuk mendapatkan, mengolah dan menyebarkan informasi untuk menunjang kegiatan, memberikan

informasi dan sebagai sarana pengambil keputusan dengan memanfaatkan *database* yang dimiliki. *Database* merupakan sistem penyimpanan data memakai komputer dimana *database* tersebut merupakan komponen utama dalam sistem informasi, karena merupakan dasar dalam menyediakan informasi [5]. Dimana banyak berbagai bidang menggunakan *database*. Bahkan saat ini hampir semua perusahaan, perkantoran, supermarket termasuk Perguruan Tinggi menggunakan *database*.

Belum diketahuinya pola asosiasi antara data mahasiswa dan data kelulusan mahasiswa di Universitas Islam Negeri Sunan Gunung Djati Bandung. Pola asosiasi antara dua jenis data tersebut dapat menunjukkan apakah ada keterkaitan antara variabel data mahasiswa yang terdiri dari jalur masuk dan program studi, dengan tingkat kelulusan mahasiswa yang bisa dilihat dari lama studi dan IPK. Pencarian data dari pola tersebut sangat erat kaitannya dengan penggunaan data

mining, penggunaan data *mining* diharapkan dapat memberikan pengetahuan-pengetahuan yang sebelumnya tersimpan di dalam *database* sehingga menjadi informasi yang berharga. Ada sejumlah algoritma untuk menyelesaikan *association rule*, diantaranya algoritma *Fold-Growth*, yang masuk dalam algoritma asosiasi [4].

Penelitian ini dibuat dengan tujuan untuk mendapatkan informasi yang berguna tentang hubungan antara tingkat kelulusan mahasiswa dengan jalur masuk perguruan tinggi menggunakan teknik data *mining* menggunakan metode *Fold-Growth*. Informasi yang disajikan untuk mengukur tingkat kelulusan dan hubungannya dengan jalur masuk mahasiswa. Informasi yang ditampilkan berupa nilai *support* (nilai penunjang) *confident* (nilai kepastian) dan hubungan antara tingkat kelulusan dengan proses masuk mahasiswa.

1.2 Identifikasi Masalah

Berdasarkan latar belakang masalah diatas pokok permasalahan yang dicari solusinya adalah:

1. Bagaimana membuat aplikasi untuk menghasilkan informasi yang dapat digunakan untuk mengukur tingkat kelulusan dan hubungannya dengan jalur masuk mahasiswa.
2. Informasi yang ditampilkan berupa nilai *support* (nilai penunjang) *confident* (nilai kepastian) dan hubungan antara tingkat kelulusan dengan proses masuk mahasiswa.

1.3 Batasan Masalah

Adapun batasan masalah dalam aplikasi data *mining* untuk menganalisa hubungan antara jalur masuk dengan tingkat kelulusan mahasiswa menggunakan metode *Fold-Growth* ini adalah:

1. Informasi yang ditampilkan berupa nilai *support* (nilai penunjang) dan *confident* (nilai kepastian) dan

hubungan antara tingkat kelulusan dengan proses masuk mahasiswa.

2. Pembahasan dilakukan pada Program Sarjana (SI) di Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Gunung Djati Bandung.
3. Bahasa pemrograman yang digunakan yaitu bahasa Java.
4. Data yang dijadikan penelitian ini adalah data mahasiswa Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Gunung Djati Bandung dari tahun 2006 sampai 2009 dan data mahasiswa yang sudah menempuh pendidikan dari tahun 2006 sampai 2009 di Fakultas Sains dan Teknologi.
5. Input dari aplikasi ini adalah memilih data mahasiswa sesuai jurusan, memilih nilai *support* dan memilih nilai *Confidence*.
6. Output dari aplikasi ini adalah hasil dari perhitungan menggunakan metode *Fold-Growth* berupa kombinasi *itemset*, nilai *support* dan nilai *confidence*.

1.4 Tujuan

Penelitian ini dibuat dengan tujuan, menghasilkan aplikasi untuk mendapatkan informasi yang berguna tentang hubungan antara tingkat kelulusan mahasiswa dengan jalur masuk menggunakan teknik data *mining* menggunakan metode *Fold-Growth* di Fakultas Sains dan teknologi.

2. Dasar Teori

2.1 Data Mining

Data *mining* merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data.

2.2 Teknik Data Mining

Beberapa teknik dan sifat data *mining* adalah sebagai berikut:

a. Classification [Predictive]

Klasifikasi adalah merupakan sebuah *record* data baru ke salah satu dari

beberapa kategori (atau *class*) yang telah didefinisikan sebelumnya.

b. Clustering [Descriptive]

Mempartisi data-set menjadi beberapa *sub-set* atau kelompok sedemikian rupa sehingga elemen-elemen dari suatu kelompok tertentu memiliki *set property* yang dishare bersama, dengan tingkat similaritas yang tinggi dalam satu kelompok dan tingkat similaritas antar kelompok yang rendah. Disebut juga dengan '*unsupervised learning*'

c. Assosiation Rule Discovery [Descriptive]

Mendeteksi kumpulan atribut-atribut yang muncul bersamaan (*co-accur*) dalam frekuensi yang sering, dengan membentuk sejumlah kaidah dari kumpulan-kumpulan tersebut. Contohnya: 90% orang yang belanja di suatu supermarket yang membeli roti juga membeli selai, dan 60% dai semua orang yang berbelanja membeli keduanya.

d. *Sequential Pattern Discovery*
[Descriptive]

Mencari sejumlah event yang secara umum terjadi bersama-sama. Contohnya, dalam satu set urutan DNA, ACGTC diikuti oleh GTCA setelah suatu celah selebar 9 dengan probabilitas sebesar 30%.

Jika diberikan sekumpulan obyek, dengan masing-masing obyek dihubungkan dengan waktu kejadiannya maka dapatkan pola yang memprediksi ketergantungan sekuensial (*sequential dependencies*) yang kuat diantara kejadian-kejadian yang berbeda.

$$(A B) (C) \rightarrow (D E)$$

Pola-pola sekuensial pertama, pada dasarnya dibentuk dengan cara mencari semua kemungkinan pola yang ada. Nilai-nilai kejadian dalam pola diatur berdasarkan urutan waktu kejadian.

e. *Regression* [Predictive]

Memprediksi nilai dari suatu variable kontinyu yang diberikan berdasarkan nilai dari variable yang lain, dengan

mengasumsikan sebuah model ketergantungan linier atau nonlinier. Teknik ini banyak dipelajari dalam statistika, bidang jaringan saraf tiruan (*neuron network*).

2.3 Analisa Asosiasi

a. *Frequent itemset*

Semua *record* dalam satu kelompok mempunyai idtrans yang sama. Mereka bersama-sama menggambarkan suatu transaksi pelanggan yang melibatkan pembelian satu item atau lebih. Suatu transaksi terjadi pada suatu tanggal tertentu dan nama dari tiap item yang dibeli dicatat bersama dengan jumlahnya. Pembuatan tabel '*denormalized*' untuk mempermudah data *mining* biasanya dilakukan pada tahan data *cleaning* dari proses KDD.

Lebar transaksi didefinisikan sebagai jumlah item yang terdapat dalam sebuah transaksi. Suatu transaksi t_j dikatakan berisi sebuah *itemset* X jika X merupakan subset dari t_j . Contohnya transaksi

pertama pada tabel diatas berisi *itemset* {Pena, Tinta} dan bukan {Pena, Air}.

Support count(σ) merupakan jumlah transaksi yang berisi suatu *itemset* tertentu atau dengan kata lain merupakan frekuensi kejadian dari suatu *itemset*. *Support* dari suatu *itemset* adalah perbandingan dari transaksi dalam basisdata yang berisi semua item dalam *itemset*.

Biasanya jumlah himpunan item yang sering dibeli bersamaan relative kecil, khususnya saat ukuran *itemset* meningkat. Semua *itemset* yang *support*-nya lebih tinggi dari nilai minimum yang ditetapkan user yang disebut dengan *minsup*; disebut *frequentitemset*.

b. Association Rule

Association rule merupakan sebuah ekspresi implikasi yang berbentuk $X \rightarrow Y$, dimana X dan Y merupakan *disjoint itemset* ($X \cap Y = \emptyset$). Contoh :{Pena, Tinta} \otimes {jus}. Dalam *association rule*, kita dapat menghitung *support* dan *confidence*. *Confidence* menyatakan seberapa sering item-item-item dalam Y

muncul dalam transaksi yang berisi X . secara formal dapat dinyatakan dengan persamaan berikut ini:

$$s(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N}$$

$$c(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)}$$

dimana s adalah *support* dan c adalah *confidence*.

2.4 Association Rule Mining

Jika terdapat sebuah himpunan transaksi T , maka tujuan dari *association rule mining* adalah untuk menemukan semua aturan yang mempunyai *support* $\geq \text{minsup}$ dan *confidence* $\geq \text{minconf}$. Pendekatan *brute-force* untuk *association rule mining* menggunakan pendekatan dengan menghitung *support* dan *confidence* dari semua kemungkinan rule. Pendekatan *Brute-force* terdiri dari langkah-langkah berikut:

- a. Daftar semua kemungkinan *association rule*
- b. Hitung *support* dan *confidence* untuk setiap rule

- c. Pangkas rule yang tidak memenuhi $minsup$ dan $minconf$ thresholds.

Secara umum, pendekatan ini sangat mahal dalam waktu komputasinya. Jumlah kemungkinan rule dari sebuah tabel transaksi yang terdiri d item adalah sebesar:

$$R = 3^d - 2^{d+1}$$

Misalkan digunakan data pada tabel pembelian diatas dengan jumlah item $d=5$, maka jumlah kemungkinan rule adalah sebesar:

$$R = 3^5 - 2^{6+1} = 243 - 64 + 1 = 180$$

Semua rule diatas merupakan partisi biner dari *itemset* yang sama. *Rules* yang berasal dari *itemset* yang sama mempunyai *support* yang identik tetapi mempunyai *confidence* yang berbeda, sehingga kita dapat memasang ulang untuk keperluan *support* dan *confidence*.

2.5 Fold-Growth

Fold-Growth merupakan salah satu metode penggalian pola asosiasi menggunakan struktur data *SOTriellT*

(*support Ordered-Trie Itemset*) dalam penggalian *itemset* yang frequent. *SOTriellT* adalah sebuah struktur data yang melakukan ekstraksi 1-*itemset* dan 2-*itemset* dari semua transaksi dalam basis data.

Dengan menggunakan basis data transaksi yang terdiri dari kode transaksi, dan kode dari barang yang di beli, algoritma ini akan diproses untuk menghasilkan pola asosiasi. pada penelitian yang dilakukan oleh Woon et al. (2004), penggabungan dua buah algoritma yaitu *FP-GROWTH* dengan *FOLDARM* pada metode asosiasi menghasilkan sebuah algoritma baru (*FOLD-GROWTH*) yang memiliki kinerja yang lebih baik. Algoritma ini menggunakan struktur data *SOTriellT*, *SOTriellT* atau *Sub Ordered Trie Itemset* merupakan suatu *Tree* yang digabungkan dengan yang utuh. Kinerja dari algoritma tersebut menghasilkan proses yang lebih fleksibel dan efisien jika dibandingkan dengan algoritma-algoritma

aturan asosiasi lainnya seperti apriori dan *FP-Growth*.

2.6 *FP-Tree*

FP-Tree merupakan struktur penyimpanan data yang dimampatkan. *FP-Tree* dibangun dengan memetakan setiap data transaksi ke dalam setiap lintasan tertentu dalam *FR-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 *FP-Tree* semakin efektif. Kelebihan dari *FP-Tree* adalah hanya memerlukan dua kali pemindaian data transaksi yang terbukti sangat efisien. Misal $I = \{a_1, a_2, \dots, a_n\}$ adalah kumpulan dari item. Dan basis data transaksi $DB = \{T_1, T_2, \dots, T_n\}$, dimana T_i ($i \in [1..n]$) adalah sekumpulan transaksi yang mengandung item di I . Sedangkan *support* adalah perhitungan (*counter*) frekuensi kemunculan transaksi

yang mengandung suatu pola dikatakan sering muncul (*frequent pattern*) apabila *support* dari pola tersebut tidak kurang dari suatu konstanta ϵ (batas ambang *minimum support*). Permasalahan mencari pola *frequent* dengan batas ambang *minimum support count* ϵ , inilah yang dicoba untuk dipecahkan oleh *FOLD-Growth* dengan bantuan struktur *FP-Tree*.

3. Data

Data merupakan keterangan-keterangan atau fakta-fakta yang dikumpulkan dari suatu populasi atau bagian populasi untuk menjelaskan karakteristik populasi tersebut [5].

3.1 Data induk Mahasiswa

Data induk mahasiswa adalah data mahasiswa yang di data ketika mahasiswa pertama masuk ke Universitas Islam Negeri Sunan Gunung Djati Bandung setelah melakukan registrasi ulang. Data yang dicatat adalah identitas pribadi mahasiswa dan identitas sekolah asal mahasiswa. Proses pendaftaran dilakukan

di tingkat Universitas, setelah di rekap lalu disebarakan ke Fakultas dan Jurusan. Atribut yang terdaftar pada data induk Mahasiswa.

Tabel 1 Atribut Mahasiswa

Atribut	Keterangan
No_Daftar	No daftar adalah, no pendaftar pada saat siswa tersebut daftar ke Universitas Islam Negeri Sunan Gunung Djati Bandung
Angkatan	Angkat adalah keterangan untuk tahun pada saat mahasiswa pertama masuk
NIM	NIM (Nomor Induk Mahasiswa) adalah nomor Identitas mahasiswa yang diberikan oleh pihak Universitas
Nama lulus	Nama mahasiswa yang lulus seleksi masuk
Kota Lahir	Kota dimana tempat mahasiswa yang bersangkutan lahir
Tanggal Lahir	Tanggal dimana mahasiswa yang bersangkutan lahir
Alamat	Alamat tempat mahasiswa tinggal
Atribut	Keterangan
Desa	Desa dimana mahasiswa yang bersangkutan tinggal
RT	RT (Rukun Tetangga) tempat dimana mahasiswa yang bersangkutan bertempat tinggal
Kecamatan	Kecamatan tempat dimana mahasiswa yang bersangkutan tinggal
Kota	Kota dimana Tempat mahasiswa yang bersangkutan lahir
Fakultas	Fakultas tempat dimana mahasiswa tersebut diterima

Kode Prodi	Nomor unik prodi dimana tempat mahasiswa bersangkutan diterima
Prodi	Prodi dimana tempat mahasiswa tersebut diterima
Kelas	Kelas tempat dimana mahasiswa tersebut diterima
Jenis Kelamin	Jenis kelamin mahasiswa yang bersangkutan
Kode Pos	Kode Pos tempat dimana mahasiswa bertempat tinggal
No Telp	No Telp mahasiswa yang bersangkutan
Sekolah	Sekolah asal tempat mahasiswa sebelum melanjutkan ke jenjang Universitas
Nama Sekolah	Nama sekolah tempat mahasiswa yang bersangkutan sekolah sebelum melanjutkan ke jenjang Universitas
Rumpun	Asal jurusan mahasiswa tersebut pada saat di sekolah menengah atas
Warga Negara	Kewarganagaraan mahasiswa yang bersangkutan
Agama	Agama yang dianut oleh mahasiswa yang bersangkutan
Penghasilan Orang Tua	Penghasilan orang tua mahasiswa yang bersangkutan
Atribut	Keterangan
Nama Orang Tua	Nama orang tua mahasiswa yang bersangkutan
Pekerjaan Orang Tua	Pekerjaan orang tua mahasiswa yang bersangkutan
Tes	Jalur masuk tes mahasiswa yang bersangkutan pada saat pertama masuk ke Universitas. Pada Universitas Islam Negeri Sunan Gunung Djati Bandung terdapat beberapa jalur masuk diantaranya, PPA, Jalur Tulis, dan juga

	SNMPTN
Pendidikan Orang Tua	Pendidikan terakhir orang tua Mahasiswa yang bersangkutan
Status 1	Status sekolah terakhir mahasiswa yang bersangkutan
Propinsi	Propinsi dimana mahasiswa yang bersangkutan tinggal

Tabel 2 Atribut Data Kelulusan

Atribut	Keterangan
No pokok Ijazah	Nomor unik yang terdapat di ijazah pada saat mahasiswa yang bersangkutan telah menyelesaikan pendidikannya di Universitas
Tanggal sidang	Tanggapan pada saat mahasiswa tersebut sidang tugas akhir dan dinyatakan lulus
Nama Mahasiswa	Nama Mahasiswa yang bersangkutan
Tempat, tgl Lahir	Tempat dan tanggal lahir mahasiswa yang bersangkutan
Alamat	Alamat tempat mahasiswa tinggal
Yudicium	Indeks prestasi yang di capai mahasiswa yang bersangkutan selama berkuliah di universitas
Atribut	Keterangan
Judul Skripsi	Judul tugas dari mahasiswa yang bersangkutan
Pembimbing	Orang/dosen yang membimbing tugas akhir mahasiswa yang bersangkutan
Prodi	Jurusan dari mahasiswa yang bersangkutan

3.2 Data Kelulusan

Data kelulusan adalah data mahasiswa yang telah dinyatakan lulus dan sudah menyelesaikan pendidikan di Fakultas Sains dan teknologi. Dalam data kelulusan terdapat atribut data mahasiswa dan kelengkapan kelulusan diantaranya nomor pokok Ijazah, Tanggal sidang, Nama mahasiswa, Yudisium, judul skripsi, Program Studi, pembimbing, alamat serta tempat tanggal lahir. IPK dikategorikan berdasarkan prediksi kelulusan yang sudah ditetapkan dalam Universitas Islam Negeri Sunan Gunung Djati Bandung sebagai berikut [8]:

Tabel 3 Indeks Prestasi Kumulatif

No	Indek Prestasi	Predikat Kelulusan
1	3.50-4.00	Cumlaude
2	3.00-3.49	Amat Baik
3	2.50-2.99	Baik
4	2.00-2.49	Cukup
5	0.00-1.99	Tidak Lulus

3.3 Proses Mining

Data *mining* adalah proses yang memperkerjakan satu atau lebih teknik pembelajaran komputer untuk

menganalisis dan mengekstraksi pengetahuan secara otomatis [2]. Data mining berisi pencarian trend atau pola yang diinginkan dalam *database* besar untuk membantu pengam-bilan keputusan di waktu yang akan datang. Pola-pola ini dikenal oleh perangkat tertentu yang dapat memberikan suatu analisa data yang berguna dan berwawasan yang kemudian dapat dipelajari dengan lebih teliti, yang mungkin saja meng-gunakan perangkat pendukung keputusan yang lainnya.

4. Metode

4.1 Assosiasi Rule

Association Rule merupakan sebuah ekspresi implikasi yang berbentuk $X \rightarrow Y$, dimana X dan Y merupakan disjoint *itemset*. Dalam assosiasi rule kita dapat menghitung *support* dan *confidence* [1][3]. *Confidence* menyatakan seberapa sering item-item dalam Y muncul dalam transaksi yang berisi X . secara formal dapat dinyatakan dengan persamaan berikut ini:

$$s(X) = \frac{\text{Jumlah transaksi yang mengandung } X}{\text{Total Transaksi}} \quad (2.1)$$

$$s(X.Y) = \frac{\text{Jumlah transaksi yang mengandung } X \text{ dan } Y}{\text{Total Transaksi}} \quad (2.2)$$

Confidence

$$(X \rightarrow Y) = \frac{\text{Jumlah transaksi yang mengandung } X \text{ dan } Y}{\text{Jumlah Transaksi yang mengandung } X} \quad (2.3)$$

Algoritma FOLD-Growth

Algoritma *FOLD-Growth* merupakan hasil gabungan dari algoritma *FOLDARM* dan *FP-Growth*. Pada algoritma ini diharapkan dapat menggabungkan keuntungan dan kebaikan dari 2 (dua) algoritma antara *FP-Growth* dengan *FOLD-ARM*. Algoritma *FOLDARM* yang memiliki kinerja cepat dalam pengukuran *itemset frequent* maksimum(K_{max}) adalah kecil atau $K \leq 10$, sedangkan algoritma *FP-Growth* memiliki kinerja yang cepat pada saat $K_{max} > 10$ [1]. Dalam

algoritma ini terdapat pembagian tahap pengerjaan sebanyak 4 (empat) tahapan utama [7].

Tahapan Utama dalam algoritma *FOLD-Growth* yaitu:

- a. Penggalan L1 (*large 1 – itemset*) dan L2 (*large 2 – Itemset*) dengan menggunakan *SOTrieIT*. Pada tahapan ini, dilakukan pembacaan basis data sebanyak satu kali untuk membaca transaksi-transaksi yang ada dalam basis data. Untuk setiap transaksi akan dibangkitkan semua kemungkinan 1-*itemset* dan 2-*itemset* yang kemudian dicatat dalam *SOTrieIT*.
- b. Pemangkasan item-item yang tidak *frequent*. Dalam tahapan ini, akan dilakukan pemangkasan pada setiap transaksi yang ada dalam basis data dengan menggunakan 1-*itemset* dan 2-*itemset*. Untuk setiap transaksi T, pada *itemset* Lk yang terdapat dalam transaksi tersebut dimana panjang k lebih dari 2, akan dilakukan pengecekan dengan menggunakan 1-

itemset dan 2-*itemset*. Sehingga, untuk item-item yang tidak sesuai akan dilakukan pemangkasan.

- c. Membangun *FP-Tree* menggunakan transaksi – transaksi yang telah dipangkas
- d. Penggalan *itemset frequent*

5. Pencarian Pola Asosiasi

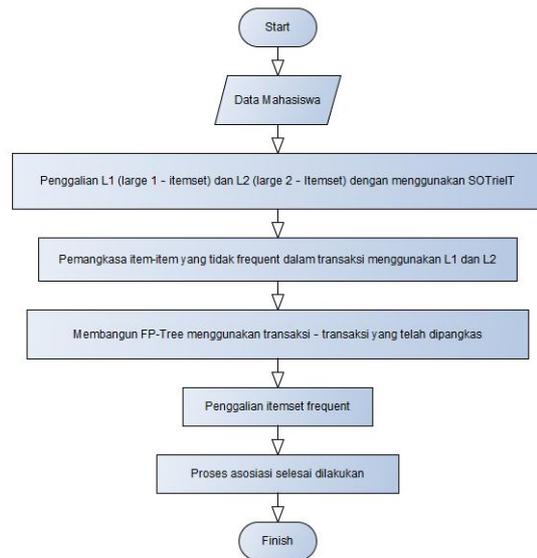
Dalam penelitian ini, variabel yang digunakan yaitu jalur masuk, program studi, lama studi dan IPK. Dari penggabungan data tersebut akan terbentuk suatu pola yang bisa memberikan informasi yang berkaitan antara jalur masuk dengan kelulusan mahasiswa. Pada data yang ada, akan ditransformasi agar mudah dibaca. Transformasi data yang dilakukan pada penelitian ini adalah data tingkat kelulusan mahasiswa. Dari data kelulusan mahasiswa dapat dilihat dari lama studi dengan IPK. Dari dua data tersebut dapat di kategorikan sesuai tingkat IPK yang sudah ditentukan oleh Universitas. Dari

data tersebut dapat dikategorikan menjadi 8 kategori, yaitu.

Tabel 4 Kategori Tingkat kelulusan Mahasiswa

Kategori	Keterangan
A1	Lama studi 4 tahun atau kurang dari 4 tahun dan IPK 3,50 – 4,00
A2	Lama studi 4 tahun atau kurang dari 4 tahun dan IPK 3,00 – 3,49
A3	Lama studi 4 tahun atau kurang dari 4 tahun dan IPK 2,50 – 2,99
A4	Lama studi 4 tahun atau kurang dari 4 tahun dan IPK 2,00 – 2,49
B1	Lama studi lebih dari 4 tahun dan IPK 3,50 – 4,00
Kategori	Keterangan
B2	Lama studi lebih dari 4 tahun dan IPK 3,00 – 3,49
B3	Lama studi lebih dari 4 tahun dan IPK 2,50 – 2,99
B4	Lama studi lebih dari 4 tahun dan IPK 2,00 – 2,49

Dalam proses pencarian pola ini ada beberapa langkah yang harus dijalani secara bertahap guna mendapatkan informasi yang baru dari dalam *database*. Pada proses ini dimulai dengan pengumpulan data. Preprocessing, data *mining*, proses FOLD-Growth



Gambar 1 Diagram alir prose FOLD-Growth

Untuk melakukan analisa terhadap data mahasiswa dengan kelulusan mahasiswa. Perlu melakukan beberapa proses yang meliputi:

- Sistem menerima inputan berupa minimum confident, minimum *support*, data mahasiswa yang telah disimpan dalam database. Kemudian sistem akan melakukan proses association rule yang menghasilkan aturan asosiasi pada data inputan dan selanjutnya melakukan pengujian rule
- Proses yang ada dalam algoritma fold-growth ini terdiri dari 4 tahapan, yaitu penggalian 1-2 *itemset* dengan

- menggunakan struktur data SOTrieIT, pemangkasan *item* yang tidak *frequent*, pembangunan FP *Tree*, dan penggalian *itemset frequent*.
- c. Proses pertama pada algoritma FOLD-Growth adalah penggalian 1-2 *itemset* dengan SOTrieIT yang dilakukan dengan membaca database sebanyak satu kali untuk membaca data-data mahasiswa dan kelulusan. Untuk setiap data akan dibangkitkan kemungkinan 1 *itemset* dan 2 *itemset* yang kemudian dicatat dalam struktur data SOTrieIT.
- d. Proses selanjutnya adalah pemangkasan *item* yang tidak *frequent* yaitu apabila nilai *support* dan nilai *count* nya kurang dari batasan minimum *support* yang sudah ditentukan. Jika minimum *support* dan minimum *confidence* tidak sesuai maka akan dipangkas.
- e. Setelah dilakukan pemangkasan *item* yang tidak *frequent*, maka *item-item* yang diproses tersebut akan diurutkan berdasarkan nilai *support* dan nilai *confidence* yang paling besar dan yang telah dipangkas.
- f. Kemudian tahapan selanjutnya adalah pembanguna FP *Tree* dengan menggunakan *item* yang telah dipangkas selanjutnya berdasarkan urutan nilai *support* dan nilai *confidence*.
- g. Tahap pada proses selanjutnya adalah tahap penggalian *itemset frequent*.
- h. Tahapan selanjutnya akan dilakukan conditional FP *Tree*, yaitu *support count* dari setiap *item* dijumlahkan. Hanya *item* dengan nilai *support* lebih besar atau sama dengan nilai minimum *support* yang akan dibangkitkan.
- i. Tahapan terakhir dalam sistem adalah pengujian terhadap rule yang dihasilkan, yaitu pengujian *lift ratio*. Pada pengujian *lift ratio* dilakukan perbandingan antara *confidence rule* yang terbentuk dengan nilai *beckmark confidence*.

j. Kemudian sistem akan menampilkan rule yang dihasilkan, nilai councidence, nilai *support*, *lift ratio* yang telah dip roses diatas.

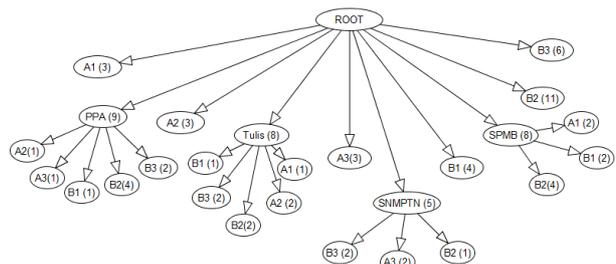
Pada perhitungan manual, dilakukan dengan sample 30 data yang sudah melalui proses *prepro-cessing* dengan *minimum support* 20% dan *minimum confident* 50%. Sample data terdapat pada tabel 3.2

Tabel 5 Sample data mahasiswa

NIM	Jalur	Kategori
206700110	PPA	B2
206700072	PPA	B2
206700076	Tulis	B1
206700103	SPMB	A1
206700105	SPMB	B1
206700112	SNMPTN	B3
206700180	Tulis	B3
206700026	Tulis	A2
206700034	PPA	A2
206700044	PPA	B2
206700046	SPMB	B2
206700071	SPMB	B2
206700091	Tulis	A1
206700094	SNMPTN	A3
206700098	PPA	A3
206700109	PPA	B3
206700141	SNMPTN	B3
206700142	SNMPTN	B2
206700143	SPMB	B2
206700146	SPMB	B1
206700169	PPA	B1
206700181	PPA	B2
206700194	Tulis	B2

206700050	Tulis	B3
207429727	PPA	B3
207429751	Tulis	A2
206700073	SNMPTN	A3
206700102	SPMB	A1
206700144	SPMB	B2
206700163	Tulis	B2

Selanjutnya adalah tahap penggalian 1-2 *itemset* dengan menggunakan struktur data *SOTrieIT*. Proses penggalian 1-2 *itemset* dengan menggunakan struktur data *SOTrieIT* dapat dilihat pada gambar.



Gambar 2 Struktur data *SOTrieIT*

penggalian 1-2 *itemset*

Pada gambar 3.3. data yang sudah di *SOTrieIT* penggalian 1-2 *itemset* terdapat beberapa jalur masuk diantaranya PPA sebanyak 9 dengan kategori kelulusan yang berada di PPA ada A2 sebanyak 1, A3 sebanyak 1, B1 sebanyak 1, B2 sebanyak 4 dan B3 sebanyak 2. Tulis

sebanyak 8 dengan kategori kelulusan yang berada di Tulis yaitu B1 sebanyak 1, B3 sebanyak 2, B2 sebanyak 2, A2 sebanyak 2, A1 sebanyak 1. SNMPTN sebanyak 5 dengan kategori kelulusan yang berada di SNMPTN yaitu B3 sebanyak 2, A3 sebanyak 2, B2 sebanyak 1. Sedangkan SPMB sebanyak 8 dengan kategori kelulusan yang berada di jalur SPMB yaitu B2 sebanyak 4, B1 sebanyak 2, dan A1 sebanyak 2.

Setelah membangun SOTrieIT yang utuh. Dilakukan penghitungan *support count* pada *1-itemset* dan *2-itemset*, pada tabel 3.13 untuk hasil penghitungan *support 1-itemset* dan untuk tabel 3.4 penghitungan *Support 2-itemset*.

Tabel 6 Perhitungan *Support 1-itemset*

<i>Itemset</i>	<i>Count</i>	<i>Support</i>
PPA	9	0.45
TULIS	8	0.40
SNMPTN	5	0.25
SPMB	8	0.40
A1	3	0.15
A2	3	0.15
A3	3	0.15
B1	4	0.20
B2	11	0.55

B3	6	0.30
----	---	------

Tabel 7 Perhitungan *Support 2-itemset*

<i>Itemset</i>	<i>Count</i>	<i>Support</i>
PPA,A2	1	0.05
PPA,B2	4	0.2
PPA,A3	1	0.05
PPA,B1	1	0.05
PPA,B3	2	0.1
TULIS,A1	1	0.05
TULIS,A2	2	0.1
TULIS,B1	1	0.05
TULIS,B2	2	0.1
TULIS,B3	2	0.1
SNMPTN,A3	2	0.1
SNMPTN,B2	1	0.05
SNMPTN,B3	2	0.1
SPMB,A1	2	0.1
SPMB,B1	2	0.1
SPMB,B2	4	0.2

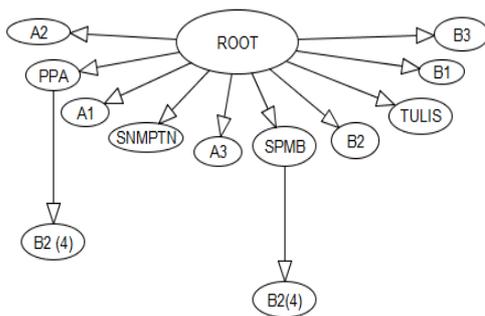
Pada tabel 3.4 terdapat penggabungan *2-itemset* dan perhitungan nilai *count* dengan nilai *support*.

Tahap selanjutnya adalah pemangkasan item-item yang tidak *frequent*. *Minimum support* yang telah ditentukan adalah 20% atau 0.2, jadi apabila nilai *support countnya* kurang dari batas *minimum support* yang telah ditentukan maka dilakukan pemangkasan.

Tabel 8 perhitungan *support 2 itemset*

<i>Itemset</i>	<i>Count</i>	<i>Support</i>
PPA,B2	4	0.2
SPMB,B2	4	0.2

Tahapan selanjutnya adalah membangun *FP-Tree* menggunakan transaksi-transaksi data yang telah di pangkas



Gambar 3 *FP-Tree*

Tahap selanjutnya adalah perhitungan nilai *confidence* pada *frequent itemset*. Pada tahapan ini, dicari nilai *confidence* pada *frequent itemset* yang telah terbentuk.

Tabel 9 *Confidence Frequent itemset*

<i>Frequent Itemset</i>	<i>Support (A,B)</i>	<i>Support (A)</i>	<i>Confidence</i>
PPA →B2	0,2	0,45	0.44
SPMB →B2	0,2	0,40	0.50

Berdasarkan minimum *confidence* yang telah ditentukan yaitu sebesar 50% maka apabila nilai *confidence* dari *frequent itemset* \geq *minimum confidence* maka akan dibangkitkan menjadi *rule*. *Rule* yang terbentuk dapat dilihat pada tabel 4.6

Tabel 10 *Rule* yang dihasilkan

<i>Frequent Itemset</i>	<i>Keterangan</i>
SPMB →B2	Jika Jalur Masuknya SPMB maka Kategori kelulusannya B2

Dari hasil perhitungan manual dengan sampel 30 data, dihasilkan SPMB→B2 dengan nilai *confident* 0,50.

6. Kesimpulan

1. Jika jalur masuknya melalui tes TULIS maka lama lulusnya lebih dari 4 tahun dan IPK berada diantara 3.00 sampai 3.49 atau masuk dalam kategori B2 untuk jurusan Manajemen pendidikan Islam.
2. Jika jalur masuknya melalui tes PPA maka lama lulusnya lebih dari 4 tahun

- dengan IPK berada diantara 3.00 sampai 3.49 atau masuk dalam kategori B2 untuk jurusan pendidikan Agama Islam, pendidikan fisika, matematika, biologi, kimia, Agroteknologi.
3. Jika jalur masuknya melalui tes TULIS maka lama lulusnya lebih dari 4 tahun dan IPK berada diantara 3.00 sampai 3.49 atau masuk dalam kategori B2 untuk jurusan Pendidikan Agama Islam, matematika, fisika, kimia, Teknik Informatika, Agroteknologi, Teknik Elektro.
 4. Jika jalur masuknya tes TULIS maka lama lulusnya kurang dari atau pas 4 tahun dan IPK berada diantara 3.00 sampai 3.49 atau masuk kategori A2 untuk jurusan Pendidikan Bahasa Arab, kimia.
 5. Jika jalur masuknya tes PPA maka lama lulusnya kurang dari atau pas 4 tahun dan IPK berada diantara 3.00 sampai 3.49 atau masuk kategori A2 untuk jurusan Pendidikan Bahasa Arab, Bahasa Inggris,
 6. Jika jalur masuknya melalui tes SNMPTN maka lama lulusnya lebih dari 4 tahun dan IPK berada diantara 3.00 sampai 3.49 atau masuk dalam kategori B2 untuk jurusan matematika, Biologi, Teknik elektro.
 7. Jika jalur masuknya melalui tes SPMB maka lama lulusnya lebih dari 4 tahun dan IPK berada diantara 3.00 sampai 3.49 atau masuk dalam kategori B2 untuk jurusan matematika, biologi.
 8. Jika jalur masuknya melalui tes SNMPTN maka lama lulusnya kurang dari atau pas 4 tahun dan IPK berada diantara 3.00 sampai 3.49 atau masuk dalam kategori A2 untuk jurusan Fisika, kimia.

DAFTAR PUSTAKA

- [1] Hamnon, Marissa. 2012. *Penerapan Metode Association Rule Dengan Algoritma FP-Growth Pada Data Transaksi Penjualan Buku*. Semarang: Universitas Brawijaya.
- [2] Hermawati, A. F. 2009. *Data Mining*. Surabaya: Penerbit Andi.

-
- [3] Slimani, Thabet. 2013. Eficient Analysis of Patten Association Rule Mining. LARODEC Lab.
- [4] Ma'ruf, Amrul, F .2013. *Aplikasi Data Mining Untuk Mengetahui Hubungan Proses Masuk Dengan Tingkat Kelulusan Mahasiswa*. Yogyakarta: AMIKOM.
- [5] Mata, T. Ramon, A. 2002. *Dasar-dasar Database Relational*. Jakarta: Penerbit Erlangga
- [6] Prasetyo, Eko. 2012. *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*. Gresik: Penerbit Andi
- [7] Prastow, Deni. 2008. *Penggunaan Struktur Data SoTrieIT untuk Pemangkasan Transaksi dengan Algoritma Data Mining Fold-Growth*. Institut Pertanian Bogor(IPB). Bogor.
- [8] UIN Bandung. 2014. *Pedoman Akademik UIN Sunan Gunung Djati Bandung*. CV. Insan Mandiri.