

Penerapan Data Mining dengan Algoritma Fp-Growth untuk Mendukung Strategi Promosi Pendidikan (Studi Kasus Kampus STMIK Triguna Dharma)

Ali Ikhwan^{*1}, Dicky Nofriansyah^{#2}, Sriani^{#3}

^{*1}Magister Ilmu Komputer Universitas Putra Indonesia “YPTK” Padang

^{#2,3}Program Studi Sistem Komputer, STMIK TrigunaDharma

Jl. Raya Lubuk Begalung, Sumbar – Indonesia

Jl. A.H. Nasution No. 73 F-Medan

Email :Aliikhwan053@gmail.com

ABSTRAK

Saat ini semakin banyaknya jumlah kampus yang semakin berkembang yang membuat para pengelola ingin mendapatkan strategi promosi yang lebih baik. Adapun cara dalam menemukan strategi promosi yang tepat akan dapat mengurangi biaya promosi dan mencapai sasaran promosi yang tepat. Salah satu cara yang dapat dilakukan untuk penentuan strategi promosi adalah dengan menggunakan teknik data mining. Adapun teknik yang digunakan dalam hal ini adalah Algoritma FP-Growth. 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. Sedangkan di dalam algoritma FP-Growth tidak dilakukan generate candidate karena FP-Growth menggunakan konsep pembangunan tree dalam pencarian frequent itemset. Penelitian dilakukan dengan mengamati beberapa variabel penelitian yang sering dipertimbangkan oleh perguruan tinggi khususnya bagian marketing dalam menentukan sasaran promosinya, Yaitu Pendidikan Terakhir, Alamat Rumah, Jurusan, Pilihan Prodi. Hasil penelitian ini adalah berupa suatu perangkat lunak dengan mengimplementasikan algoritma FP-Growth yang menggunakan konsep pembangunan FP-Tree dalam mencari Frequent Itemset.

Kata Kunci : Data Mining, Association Rules, Frequent Itemset, FP-Growth.

ABSTRAC

Nowadays more and more growing amount of campus that make the managers want to get better promotion strategy. So how to find the right promotional strategy will be able to reduce costs and achieve targeted promotional campaign right. One of proces that can be done to determine promotion strategy is to use data mining techniques . The technique used in this case is the FP-Growth Algorithm. FP-Growth algorithm is one of the alternatives that can be used to determine the set of data that appears most frequently (frequent item sets) in a set of data. FP-Growth algorithm is an extension of Apriori algorithm. While in the FP-Growth algorithm do not generate candidate because the FP-Growth using the concept of development in the search for frequent itemset tree. The research doing by observing and some research variables are often considered by colleges especially the marketing in targeting the promotion, Namely Latest Education, Home Address, Subject, Options Prodi. The results of this research are in the form of a software to implement the FP-Growth algorithm that uses the concept of development of FP-Tree in finding Frequent Itemset.

Keywords: Data Mining, Association Rules, Frequent Itemset, FP-Growth.

1. PENDAHULUAN

1.1 Latar Belakang Masalah

Persaingan di dalam dunia bisnis ini sangat ketat, khususnya dalam mempromosikan perguruan tinggi. Setiap kampus berusaha memberikan fasilitas yang terbaik. Oleh karena itu setiap kampus berusaha mencari strategi yang tepat untuk mempromosikan kampusnya. Menurut para pengembang untuk menemukan suatu strategi yang dapat menentukan strategi pemasaran dalam mempromosikan perguruan tinggi dengan memaksimalkan pelayanan kepada masyarakat. Salah satu teknik yang dipakai dalam penerapan *Data Mining* adalah di dalam bidang promosi. Bila sasaran promosi tidak ditentukan secara baik, dalam arti tidak diupayakan mencari sasaran promosi yang potensial, maka hanya akan menghabiskan banyak waktu dan biaya yang seharusnya bisa diminimalisir melalui pemilihan target promosi yang baik.

Salah satu cara yang dapat diterapkan adalah dengan menerapkan penggunaan *Data Mining*. Karena di dalam data mining sendiri terdapat cara dan teknik dalam pemenuhan kebutuhan salah satunya adalah kebutuhan informasi yang luas, dan dari informasi yang kita dapat bisa kita gunakan sebagai suatu keputusan atau menentukan sebuah kualitas dalam menentukan suatu keputusan. Adapun kumpulan dari sebuah data atau informasi memiliki banyak potensi untuk dijadikan suatu kesimpulan dalam mengambil sebuah keputusan dengan melakukan analisis dan menggali suatu informasi yang terdapat di dalam sebuah data. Maka dimungkinkan untuk dibuat strategi dalam mendukung promosi pendidikan. Dari definisi *Data Mining* yang luas terdapat banyak jenis

teknik analisa yang dapat digolongkan dalam *Data Mining*. Teknik *Data Mining* yang akan digunakan dalam penelitian ini adalah *Association Rule Mining*. *Association rule mining* adalah teknik *mining* untuk menemukan aturan asosiatif antara suatu kombinasi *item*.

Adapun dalam penelitian ini akan dibahas bagaimana cara mengimplementasikan salah satu algoritma dalam data mining, yaitu algoritma *frequent Pattern-Growth (FP-Growth)*. Algoritma ini adalah bagian dari teknik asosiasi pada data mining. Adapun *FP-Growth* sendiri adalah salah satu alternatif algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (*frequent itemset*) dalam sekumpulan data. Karakteristik algoritma *FP-Growth* adalah struktur data yang digunakan dalam *tree* yang disebut *FP-Tree*. Dengan menggunakan *FP-Tree*, algoritma *FP-Growth* dapat langsung mengekstrak *frequent itemset* dari *FP-Tree*.

1.2 Perumusan Masalah

1. Bagaimana sistem memberikan informasi kepada user dalam mempromosikan pendidikan di Perguruan Tinggi Swasta secara efektif dan efisien?
2. Bagaimana sistem ini menentukan target dalam mempromosikan pendidikan?
3. Bagaimana menerapkan algoritma *FP-Growth* di dalam promosi pendidikan sehingga dalam proses tersebut dapat menentukan strategi yang tepat ?
4. Bagaimana melakukan pengujian Metode *FP-Growth* dengan *Rapidminer* sehingga menghasilkan penyelesaian masalah dalam mencari strategi promosi yang tepat?

1.3 Batasan Masalah

penulis mencoba memberikan batasan masalah antara lain:

1. Algoritma *FP- Growth* yaitu algoritma yang digunakan di dalam penyeleksian untuk mempercepat proses penentuan *frequent itemset* sebelum *generate rull* sebagai rekomendasi keputusan.
2. Penelitian ini khusus membahas tentang strategi promosi pendidikan pada Perguruan Tinggi khususnya di STMIK TRIGUNA DHARMA .
3. Data yang digunakan sebagai sumber merupakan data yang diambil di STMIK TRIGUNA DHARMA 1 tahun terakhir, yang meliputi asal sekolah, daerah asal, jurusan dan pilihan prodi.
4. *Tools* yang digunakan di dalam pengolahan data adalah *Rapidminer 5.2*

1.4 Tujuan Penelitian

1. Untuk menganalisa permasalahan promosi pendidikan di STMIK TRIGUNA DHARMA.
2. Untuk menerapkan metode *FP- Growth* di dalam strategi promosi pendidikan di STMIK TRIGUNA DHARMA.
3. Untuk melakukan pengujian metode *FP- Growth* dengan menggunakan tools *Rapidminer* sehingga menghasilkan strategi promosi pendidikan yang efektif dan efisien.
4. Menghasilkan sebuah hubungan antara variable atau item berdasarkan implementasi dari algoritma *FP- Growth*.
5. Menguji hubungan variable atau item tersebut untuk mendapatkan pengetahuan baru dari proses ekstraksi menggunakan *Tools* yang ada.

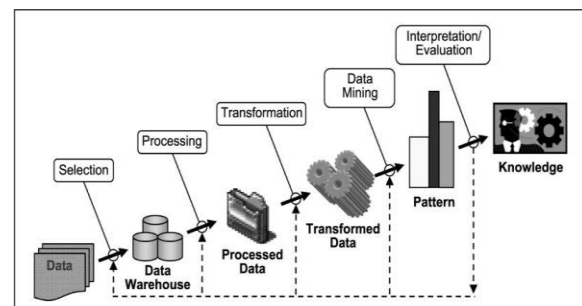
1.5 Manfaat Penelitian

1. Mengetahui algoritma *FP-Growth* dapat memberikan alternatif pengetahuan tentang menemukan strategi dalam promosi pendidikan.
2. Menemukan strategi atau sasaran yang lebih tepat dalam mempromosikan pendidikan.
3. Menjadikan hasil software ini dapat digunakan pihak kampus untuk menentukan strategi dalam promosi pendidikan.

2. LANDASAN TEORI

2.1 Knowledge Discovery In Database (KDD)

Knowledge Discovery in Database (KDD) adalah proses menentukan informasi yang berguna serta pola-pola yang ada dalam data. Informasi ini terkandung dalam basis data yang berukuran besar yang sebelumnya tidak diketahui dan potensial bermanfaat. *Data Mining* merupakan salah satu langkah dari serangkaian proses *iterative* KDD. Tahapan proses KDD dapat dilihat pada gambar 2.1.



Gambar 2.1 Tahapan dalam KDD

Tahapan proses KDD terdiri dari:

a. Data Selection

Pada proses ini dilakukan pemilihan himpunan data, menciptakan himpunan data target, atau memfokuskan pada subset

variable (sampel data) dimana penemuan (*discovery*) akan dilakukan. Hasil seleksi disimpan dalam suatu berkas yang terpisah dari basis data operasional.

b. *Pre-Processing* dan *Cleaning* Data

Pre-Processing dan *Cleaning* Data dilakukan membuang data yang tidak konsisten dan *noise*, duplikasi data, memperbaiki kesalahan data, dan bisa diperkaya dengan data *eksternal* yang relevan.

c. *Transformation*

Proses ini *mentransformasikan* atau menggabungkan data ke dalam yang lebih tepat untuk melakukan proses *mining* dengan cara melakukan peringkasan (*agregasi*).

d. *Data Mining*

Proses *Data Mining* yaitu proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik, metode atau algoritma tertentu sesuai dengan tujuan dari proses KDD secara keseluruhan.

e. *Interpretation / Evaluasi*

Proses untuk menerjemahkan pola-pola yang dihasilkan dari *Data Mining*. Mengevaluasi (menguji) apakah pola atau informasi yang ditemukan bersesuaian atau bertentangan dengan fakta atau hipotesa sebelumnya. Pengetahuan yang diperoleh dari pola-pola yang terbentuk dipresentasikan dalam bentuk visualisasi.

2.2 Data Mining

Data Mining adalah proses untuk mendapatkan informasi dengan melakukan pencarian pola dan relasi-relasi yang tersembunyi di dalam timbunan data yang

banyak. (Fadlina, 2014) *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*.

2.3 Tahapan dalam Data Mining

Menurut (Buulolo, 2013) ada empat tahap yang dilalui dalam data mining, antara lain:

1. Tahap pertama : *Precise statement of the problem* (pernyataan tepat terhadap permasalahan)

Sebelum mengakses perangkat lunak data mining, seorang analis harus memiliki kejelasan perihal 'pertanyaan apa yang akan ingin dijawabnya'. Jika tidak ada *formula* yang tepat untuk problematika yang ada maka anda hanya akan membuang-buang dan uang dalam membuat solusinya.

2. Tahap kedua : *Initial exploration*

Tahap ini dimulai dengan mempersiapkan data yang juga juga termasuk kedalam data mining "*cleaning*" (misalnya : mengidentifikasi dan menyikikan data yang dikodekan salah), transformasi data, memilih *subset record, data set*, langkah awal seleksi. Mendeskripsikan dan memvisualisasikan data adalah kunci dari tahap ini.

3. Tahap tiga : *Model building and validation*

Tahap ini melibatkan pertimbangan terhadap ragam permodelan dan memilih yang terbaik bagi performansi prediktif.

4. Tahap ke-empat : *Deployment*

Memilih aplikasi yang tepat berikut permodelan untuk membuat (*generate*) prediksi. Selanjutnya kita kan melihat rincian perihal tahapan-tahapan data mining.

2.4 Pengelompokan *Data Mining*

Data Mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu (Kusrini dan Emha Taufiq Luthfi, 2009):

1. Deskripsi

Terkadang peneliti dan analis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh, petugas pengumpulan suara mungkin tidak dapat menentukan keterangan atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden.

2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih kearah numerik daripada kearah kategori. Sebagai contoh akan dilakukan estimasi tekanan darah sistolik pada pasien rumah sakit berdasarkan umur pasien, jenis kelamin, indeks berat badan, dan level sodium darah. Hubungan antara tekanan darah sistolik dan nilai variabel prediksi dalam proses pembelajaran akan menghasilkan model estimasi. Model estimasi yang dihasilkan dapat digunakan untuk kasus baru lainnya.

3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang.

Contoh prediksi bisnis dan penelitian adalah:

a. Prediksi harga beras dalam tiga bulan yang akan datang.

b. Prediksi persentasi kenaikan kecelakaan lalu lintas tahun depan jika batas bawah kecepatan dinaikkan.

Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

4. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang dan pendapatan rendah.

Contoh lain klasifikasi dalam bisnis dan penelitian adalah:

a. Menentukan apakah suatu transaksi kartu kredit merupakan transaksi yang curang atau tidak.

b. Memperkirakan apakah suatu pengajuan hipotek oleh nasabah merupakan suatu kredit yang baik atau buruk.

c. Mendiagnosis penyakit seorang pasien untuk mendapatkan termasuk kategori penyakit apa.

5. Pengklusteran

Pengklusteran merupakan pengelompokan *record*, pengamatan atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan.

Contoh pengklusteran dalam bisnis dan penelitian adalah:

- a. Mendapatkan kelompok-kelompok konsumen untuk target pemasaran dari satu suatu produk bagi perusahaan yang tidak memiliki dana pemasaran yang besar.
- b. Untuk tujuan audit akuntansi, yaitu melakukan pemisahan terhadap perilaku *financial* dalam baik dan mencurigakan.
- c. Melakukan pengklusteran terhadap ekspresi dari *gen*, untuk mendapatkan kemiripan perilaku dari *gen* dalam jumlah besar.

6. Asosiasi

Tugas asosiasi dalam *Data Mining* adalah menemukan *attribut* yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja. Contoh asosiasi dalam bisnis dan penelitian adalah:

- a. Meneliti jumlah pelanggan dari perusahaan telekomunikasi seluler yang diharapkan untuk memberikan respon positif terhadap penawaran *upgrade* layanan yang diberikan.
- b. Menentukan barang dalam supermarket yang dibeli secara bersamaan dan yang tidak pernah dibeli secara bersamaan.

2.5 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* (Fadlina, 2014).

A. Data Preparation.

Pada banyak bidang keilmuan, terutama komputer sains, diperlukan data yang berkualitas melalui proses *preparation* dari data mentah. Dalam prakteknya, ditemukan bahwa data *cleaning* dan *preparation* membutuhkan total usaha 80% dari usaha untuk merekayasa data, sehingga menjadikan data *preparation* sebagai proses yang krusial. Pentingnya proses ini dapat dilihat dari tiga aspek, yakni:

1. Data *real world* merupakan data kotor. Data *real world* bisa mengandung data yang tidak, terdapat *noise*, tidak konsisten, yang dikarenakan:
 - a. Tidak lengkap (*incomplete*), yaitu kekurangan nilai atribut atau hanya mengandung agregat data (contoh : address = " ")
 - b. *Noise*, yaitu masih mengandung *error* dan *outliers*.
 - c. Tidak konsisten (*inconsistent*), yaitu data yang mengandung discrepansi dalam code dan nama atau singkatnya datanya tidak konsisten.
2. Sistem *mining* dengan performa tinggi membutuhkan data yang berkualitas. Data *preparation* atau *preparation* menghasilkan dataset yang lebih sedikit daripada dataset yang asli, ini bias meningkatkan efisiensi dari *Data Mining*. Langkah ini mengandung:
 - a. Memilih data yang relevan
 - b. Mengurangi data.
3. Data yang berkualitas menghasilkan pola yang berkualitas. Dengan data *preparation*, maka data yang dihasilkan adalah data yang berkualitas, yang mengarah pada pola yang berkualitas pula, dengan:

- a. Mengembalikan data yang tidak lengkap
- b. Membenarkan *error*, atau menghilangkan *outliers*
- c. Membenahi data yang bertentangan.

B. Association Rules

Association rule merupakan salah satu metode yang bertujuan mencari pola yang sering muncul di antara banyak transaksi, dimana setiap transaksi terdiri dari beberapa *item* sehingga metode ini akan mendukung sistem rekomendasi melalui penemuan pola antar *item* dalam transaksi-transaksi yang terjadi. Metodologi dasar analisis asosiasi terbagi menjadi dua tahap:

1. Analisa pola frekuensi tinggi

Tahap ini mencari kombinasi *item* yang memenuhi syarat minimum dari nilai *support* dalam *database*. Nilai *support* sebuah *item* diperoleh dengan rumus berikut:

$$\text{Support (A)} = \frac{\text{Jumlah Transaksi Mengandung A}}{\text{Total Transaksi}} \quad (1)$$

$$\text{Support (A C B)} = \frac{\text{Jumlah Transaksi Mengandung A dan B}}{\text{Total Transaksi}} \quad (2)$$

2. Pembentukan aturan asosiatif

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiatif yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan asosiatif A_B. Nilai *confidence* dari aturan A_B diperoleh dari rumus berikut:

C. Frequent Itemset

Langkah pertama pada *association rule* adalah menghasilkan semua *itemset*

yang memungkinkan dengan kemungkinan *itemset* yang muncul dengan *m-item* adalah 2^m . Karena besarnya komputasi untuk menghitung *frequent itemset*, yang membandingkan setiap kandidat *itemset* dengan setiap transaksi, maka ada beberapa pendekatan untuk mengurangi komputasi tersebut, salah satunya dengan algoritma apriori.

D. Apriori

Algoritma apriori digunakan untuk mencari *frequent itemset* yang memenuhi *minsup* kemudian mendapatkan *rule* yang memenuhi *minconf* dari *frequent itemset* tadi. Algoritma ini mengontrol berkembangnya kandidat *itemset* dari hasil *frequent itemset* dengan *support-based pruning* untuk menghilangkan *itemset* yang tidak menarik dengan menetapkan *minsup*. Prinsip dari apriori ini adalah bila *itemset* digolongkan sebagai *frequent itemset*, yang memiliki *support* lebih dari yang ditetapkan sebelumnya, maka semua subsetnya juga termasuk golongan *frequent itemset*, dan sebaliknya.

E. Rule Generation

Setelah mendapatkan *frequent itemset* menggunakan algoritma apriori, langkah selanjutnya adalah mendapatkan *rule* yang memenuhi *confidence*. Karena *rule* yang dihasilkan berasal dari *frequent itemset*, dengan kata lain, dalam menghitung *rule* menggunakan *confidence*, tidak perlu lagi menghitung *support*-nya karena semua calon *rules* yang dihasilkan telah memenuhi *minsup* sesuai yang ditentukan. Penghitungan ini juga tidak perlu melakukan perulangan *scanning* pada *database* untuk menghitung *confidence*, cukup dengan mengambil *itemset* dari hasil *support*.

2.6 Algoritma *FP-Growth*

(Sensuse, 2012) *FP-growth* adalah salah satu alternatif algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (*frequent itemset*) dalam sebuah kumpulan data. *FP-growth* menggunakan pendekatan yang berbeda dari paradigma yang digunakan pada algoritma Apriori.

(Ririanti, 2014). *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.

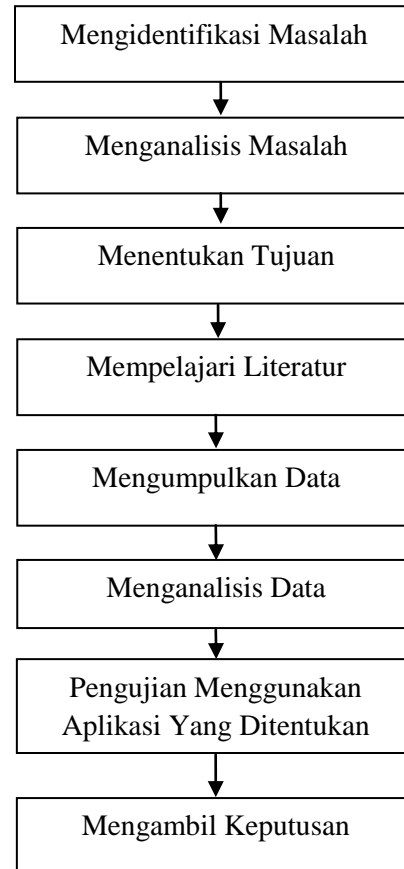
3. METODOLOGI PENELITIAN

3.1 Pendahuluan

Dalam melakukan penelitian agar hasilnya bisa maksimal, penulis mengikuti kaidah-kaidah (metode) yang telah ditetapkan. Metode penelitian yang akan diterapkan adalah metode penelitian ilmiah yang telah diajarkan di bangku perkuliahan sarjana. Metodologi penelitian ini memuat tentang kerangka kerja penelitian yang akan dibahas di bawah ini. Pada bab ini akan dibahas metodologi penelitian untuk mengidentifikasi permasalahan, analisa permasalahan dan pada akhirnya mencari penyelesaian masalah di dalam menganalisa strategi dalam mempromosikan pendidikan pada STMIK TRIGUNA DHARMA MEDAN.

3.2 Kerangka Kerja

Dalam metodologi penelitian ada urutan kerangka kerja yang harus diikuti, urutan kerangka kerja ini merupakan gambaran dari langkah-langkah yang harus dilalui agar penelitian ini bisa berjalan dengan baik. Kerangka kerja yang harus diikuti bisa dilihat pada gambar 3.1.



Gambar 3.1 Kerangka Kerja

Berdasarkan gambar 3.1 dijabarkan urutan-urutan langkah kerja sebagai berikut:

3.2.1. Mengidentifikasi Masalah

Masalah yang diidentifikasi dalam penelitian ini untuk mendapatkan strategi promosi yang efektif dan efisien dalam mendapatkan calon mahasiswa baru

ditengah ketatnya persaingan antar Perguruan Tinggi .

3.2.2. Menganalisis Masalah

Analisis masalah pada penelitian ini dilakukan dengan dua metode yaitu metode deskriptif dan metode komperatif.

a. Metode Deskriptif

Pada metode ini data yang ada dikumpulkan, disusun, dikelompokkan dan dianalisis sehingga diperoleh beberapa gambaran yang jelas pada masalah yang akan dibahas.

b. Metode Komperatif

Pada metode ini analisis dilakukan dengan cara membandingkan teori dan praktek sehingga diperoleh gambaran yang jelas tentang persamaan dan perbedaan di antara keduanya.

3.2.3. Menentukan Tujuan

Berdasarkan uraian di atas, tujuan dari penelitian ini adalah:

- a. Untuk mengklasifikasikan setiap wilayah di Sumatera Utara dalam menentukan target promosi pendidikan di STMIK TRIGUNA DHARMA MEDAN.
- b. Untuk mendapatkan atau menghasilkan suatu knowledge (pengetahuan) tentang wilayah yang akan dipilih dalam mempromosikan pendidikan.
- c. Untuk memberikan masukan kepada pihak akademik sebagai bahan pertimbangan pihak akademik di dalam menentukan wilayah yang akan dijadikan target dalam promosikan pendidikan.

Mempelajari Literatur

Literatur-literatur yang dipakai sebagai bahan referensi dalam penelitian ini adalah dari jurnal-jurnal ilmiah, modul pembelajaran dan buku tentang *Data Mining*. literatur-literatur ini akan menjadi pedoman untuk melakukan penelitian agar memudahkan proses penelitian.

3.2.4. Mengumpulkan Data

Metode pengumpulan data dilakukan dengan cara melakukan pengamatan langsung ke kampus STMIK TRIGUNA DHARMA. Selain pengamatan, juga dilakukan wawancara kepada pihak-pihak yang terkait dengan penelitian ini. Di samping itu juga, melakukan pengambilan sampel *database* nasabah untuk menunjang penelitian ini.

3.2.5. Menganalisis Data

Pada tahap ini data yang telah dikumpulkan akan dianalisis. Analisis menggunakan *Association Rule* yang dimaksud dilakukan melalui mekanisme penghitungan *support* dan *confidence* dari suatu hubungan item. Sebuah rule asosiasi dikatakan *interesting* jika nilai *support* adalah lebih besar dari *minimum support* dan juga nilai *confidence* adalah lebih besar dari *minimum confidence* .Dengan menggunakan teknik *FP-Growth* yang menghasilkan *Frequent Itemset* tanpa melakukan *candidates generation*. dengan tujuan untuk mendapatkan pengetahuan yang baru (*knowledge*) berupa informasi tentang strategi promosi yang efektif dan efisien dalam mendapatkan calon mahasiswa baru ditengah ketatnya persaingan antar Perguruan Tinggi.

3.2.6. Pengujian Menggunakan Aplikasi Yang Ditentukan

Pada tahap ini, *rule* dites kembali atau diuji lagi menggunakan sistem *data*

mining yang sudah ada. Tools yang digunakan sebagai pengujian sistem adalah *Ramidminer 5. 2*.

Mula-mula data manual "Data Mahasiswa Baru" yang sudah ada sebelumnya, di transformasikan ke dalam *Microsoft Excel*.

Sebagai tahap awal di dalam proses pengujian yaitu *mengimport* data mahasiswa baru pada kampus STMIK TRIGUNA DHARMA yang akan dijadikan sebagai data tabel transaksi *frekuensi item* di dalam *tools Ramidminer 5 2* tersebut. Kemudian tahap selanjutnya yaitu, melakukan proses *Data Input* dan *Affinity analysis* atau *market basket analysis*. Analisis asosiasi atau *association rule* adalah teknik *Data Mining* untuk menemukan aturan asosiatif antara suatu kombinasi *item*. Analisis *FP-Growth* didefinisikan suatu proses untuk menemukan semua aturan *FP-Growth* yang memenuhi syarat *minimum support* dan syarat minimum untuk *confidence*. Dari proses *Data Input* dan *Affinity* tersebut akan terdapat pola atau kriteria tertentu dalam mendapatkan strategi promosi. Hasil dari Proses *Data Input* dan *Affinity* tersebut akan diolah kembali untuk menghasilkan suatu *association rule*. Dari *association rule* tersebut akan terlihat hasilnya. Hasilnya tersebut merupakan pengetahuan baru (*knowledge*) dan juga merupakan sebagai acuan untuk menganalisa strategi dalam mempromosikan pendidikan.

3.2.7. Pengambilan Keputusan

Setelah diuji, hasil analisis antara *manual ways* (cara manual) dan pengujian memanfaatkan *tools* akan terlihat perbandingannya. Langkah berikutnya adalah penentuan atau pengambilan keputusan terhadap *knowledge* yang baru didapat yaitu berupa penentuan dalam

menentukan strategi dalam mempromosikan pendidikan.

4. ANALISA

4.1 Analisa Data

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*. Penggalan *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:

1. Tahap pembangkitan *conditional pattern base*,
 2. Tahap pembangkitan *conditional FP-Tree*, dan
 3. Tahap pencarian *frequent itemset*.
- Ketiga tahap tersebut merupakan langkah yang akan dilakukan untuk mendapat *frequent itemset*.

Input : *FP-Tree Tree*

Output : *Rt* Sekumpulan lengkap pola *frequent*

Method : *FP-Growth (Tree, null)*

Procedure : *FP-Growth (Tree, α)*

{

01: *if Tree* mengandung *single path P*;

02: *then* untuk tiap kombinasi (dinotasikan β) dari *node-node* dalam *path do*

03: bangkitkan pola β α dengan *support* dari *node-node* dalam *path do* β ;

04: *else* untuk tiap a_1 dalam *header* dari *tree do*

}

05: bangkitkan pola

06: bangun $\beta = \alpha$ dengan $support = \alpha$
 support
 07: if Tree $\beta =$

4.2 Pra Proses

Tabel 4.1 Untuk Memberi Tanda Pada Setiap Itemset

Item
SMK Negeri (A1)
SMK Swasta(A2)
SMA Negeri (A3)
SMA Swasta (A4)
Madrasah Aliyah (A5)
Medan (B1)
Aceh (B2)
Lubuk Pakam (B3)
Deli Serdang (B4)
Perbaungan (B5)
Tebing Tinggi (B6)
Riau (B7)
IPA (C1)
IPS (C2)
TKJ (C3)
MULTIMEDIA (C4)
Sistem Informasi (D1)
Sistem Komputer (D2)
Teknik Komputer (D3)
Manajemen Informatika (D4)

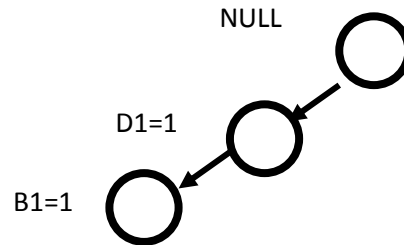
9	D1,C1,A3
10	D1,C1,A4
11	D1,C1,A4
12	A3,B1.C2
13	D1,C1
14	C1
15	D1,C1,A3
16	C1,A3
17	D1,C1,A3
18	-
19	D1,A4,C2
20	D1,A4,B1,C2
21	A3,B1,C2
22	A3,B1,C2
23	D1,B1
24	D1
25	D1,C1,A4,B1
26	D1,C1,A4,B1
27	D1,C1,A4
28	D1,C1,A3
29	C2
30	D1,A4,C2

Tabel 4.2 Frekuensi dari Setiap Transaksi

TID	ITEM
1	D1,B1
2	C1,A3,B1
3	A3,C2
4	D1,A4,B1,C2
5	D1,C1,A4
6	C1,A3
7	D1,C1,A3
8	D1

Maka langkah selanjutnya adalah membentuk pohon *FP-Tree* dengan melihat tabel 4.2 sebagai berikut:

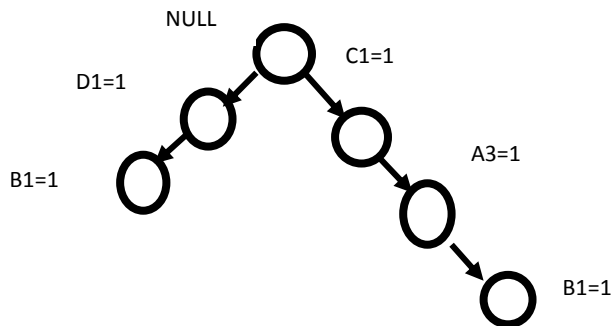
Gambar di bawah ini memberikan ilustrasi mengenai pembentukan *FP-tree* setelah pembacaan TID pada tabel 4.2



Gambar 4.1: Hasil Pembentukan FP-tree Setelah Pembacaan TID 1

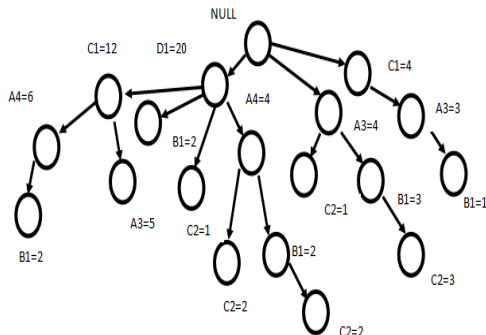
Gambar 4.1 adalah penjelasan tentang pembentukan *FP-Tree* setelah

pembacaan didapat setelah melakukan TID 1, Yaitu berisi : NULL- D1 (Sistem Informasi) =1 -(B1) Medan= 1.



Gambar 4.2 : Hasil Pembentukan FP-tree Setelah Pembacaan TID 2

Gambar 4.2 didapat setelah melakukan TID 2 ,Yaitu: NULL-C1 (IPA) =1-A3 (SMAN) =1-B1(Medan) =1.

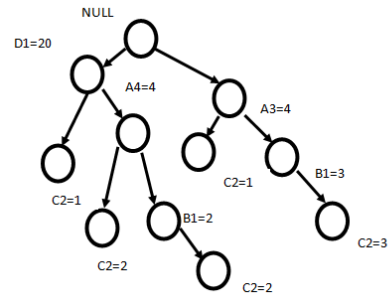


Gambar 4.3 : Hasil Pembentukan FP-tree Setelah Pembacaan TID 30

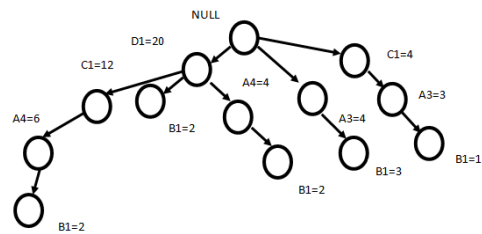
Gambar didapat setelah melakukan TID 30 yang dijumlahkan, yaitu berisi Null-Sistem Informasi (D1) = 20 -IPA (C1) =16 - SMA Negeri (A3) =12 -SMA Swasta (A4) =10-Medan (B1) = 10-IPS (C2) = 9.

Untuk menemukan *Frequent itemset* dari tabel 4.2, maka perlu ditentukan terlebih dahulu lintasan yang berakhir dengan *support count* terkecil, yaitu C2 yang diikuti dengan A4,A3, C1,D1 dan di

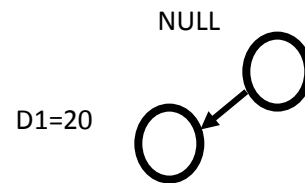
akhiri D1. Proses pembentukan masing-masing *node* dapat dilihat pada gambar 4.4.



Gambar 4.4 Lintasan yang Mengandung Simpul C2



Gambar 4.5 Lintasan yang Mengandung Simpul B1



Gambar 4.6 Lintasan yang Mengandung Simpul D1

Setelah mencari *frequent itemset* untuk beberapa akhiran *suffix* maka didapat hasil yang dirangkum dalam tabel berikut:

Tabel 4.3 Daftar Frequent Itemset Diurutkan Berdasarkan Hubungan Akhiran

Suffix	Frequent Itemset
C2	{C2},{C2,A4}{C2,A3}{C2,D1}{C2,B1,A3},{C2,B1,A4}
B1	{B1},{B1,D1},{B1,A3},{B1,A3,C1},{B1,A4,D1},{B1,A4,C1,D1}
A4	{A4},{A4,C1,D1}

A3	{A3},{A3,C1},{A3,C1,D1}
C1	{C1},{ C1,D1}
D1	{D1}

Dari *frequent itemset* yang didapat dari pembentikan *Fp-Tree* dan *FP-Growth* maka dapat dihitung nilai *Support* dan *Confidence* sebagai berikut:

$Support = \frac{Count(Sistem\ Informasi, IPA, SMA\ Swasta, Medan)}{Jumlah\ Transaksi} = \frac{1}{30}$.

$Support = \frac{Count(Sistem\ Informasi, IPA, SMA\ Negeri)}{Jumlah\ Transaksi} = \frac{1}{30}$.

$Support = \frac{Count(Sistem\ Informasi, Medan)}{Jumlah\ Transaksi} = \frac{1}{30}$.

$Support = \frac{Count(Sistem\ Informasi, SMA\ Swasta, IPS)}{Jumlah\ Transaksi} = \frac{1}{30}$.

$Support = \frac{Count(Sistem\ Informasi, SMA\ Swasta, Medan, IPS)}{Jumlah\ Transaksi} = \frac{1}{30}$.

$Support = \frac{Count(SMA\ Negeri, IPS)}{Jumlah\ Transaksi} = \frac{1}{30}$.

$Support = \frac{Count(SMA\ Negeri, Medan, IPS)}{Jumlah\ Transaksi} = \frac{1}{30}$.

$Support = \frac{Count(IPA, SMA\ Negeri, Medan)}{Jumlah\ Transaksi} = \frac{1}{30}$.

Sedangkan untuk *confidence* atau nilai kepercayaannya adalah sebagai berikut :

$Confidence = \frac{Count(SMA\ Negeri, IPS)}{Count\ Medan} = \frac{1}{9}$.

$Confidence = \frac{Count(Sistem\ Informasi, SMA\ Swasta, Medan, IPS)}{Count\ Medan} = \frac{1}{9}$.

$Confidence = \frac{Count(SMA\ Negeri, Medan, IPS)}{Count\ Medan} = \frac{1}{9}$.

$Confidence = \frac{Count(Sistem\ Informasi, SMA\ Swasta, IPS)}{Count\ Medan} = \frac{1}{9}$.

$Confidence = \frac{Count(Sistem\ Informasi, SMA\ Swasta, Medan, IPS)}{Count\ Medan} = \frac{1}{9}$.

$Confidence = \frac{Count(SMA\ Negeri, Medan, IPS)}{Count\ Medan} = \frac{1}{9}$.

$Confidence = \frac{Count(Sistem\ Informasi, IPS)}{Count\ Medan} = \frac{1}{9}$.

$Confidence = \frac{Count(Sistem\ Informasi, SMA\ Swasta, IPS)}{Count\ Medan} = \frac{1}{9}$.

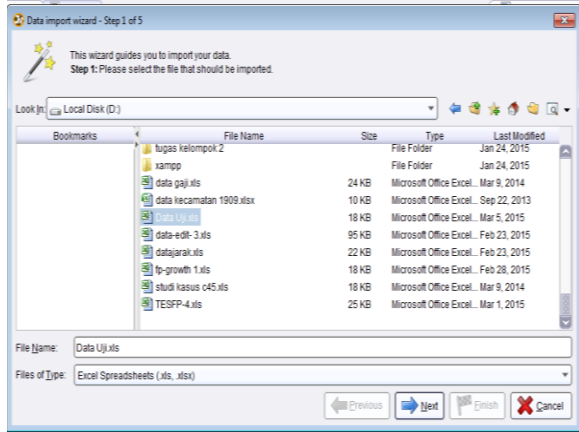
Setelah didapat nilai *support* dan *confidence* dari keseluruhan kombinasi pada data dengan perhitungan *FP-Tree* dan *FP-Growth* maka didapat nilai *support* dan *confidence* yang paling tinggi dan akurat yaitu kombinasi. (D1,C1,A3) { Sistem Informasi, IPA, SMA Negeri } yang mempunyai nilai *support* : $\frac{5}{30} = 0,16$ dan nilai *confidence* : $\frac{5}{20} = 0.25$.

5. IMPLEMENTASI SISTEM

Untuk membuktikan kebenaran hasil analisa diperlukan sebuah proses pengujian untuk menguji kebenaran dari hasil pengolahan data yang dikerjakan secara manual pada BAB IV yang telah dikerjakan sebelumnya, untuk proses pengujian tersebut kita dapat menggunakan salah satu *software* aplikasi seperti *Rapidminer*, dengan langkah-langkah sebagai berikut:

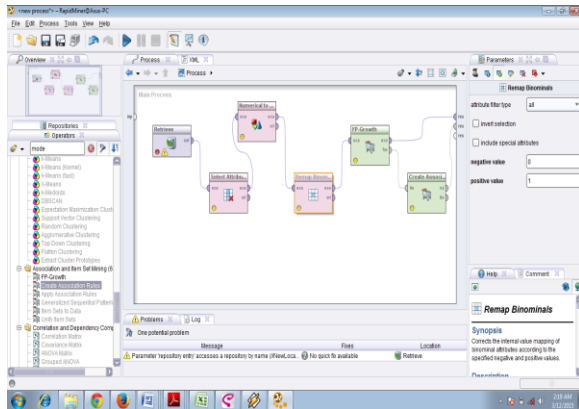
1. Dataset yang terdiri 30 record data dan berisi variabel-variabel atribut Sistem Informasi, IPA, SMA Negeri, SMA

Swasta, Medan, IPS disimpan dalam aplikasi *microsoft excel* dengan nama *file datauji.xls* dan yang akan dicoba menggunakan *software rapidminer 5.2* untuk melihat hasil sama tidaknya dengan pencarian frequent itemsetnya melalui :



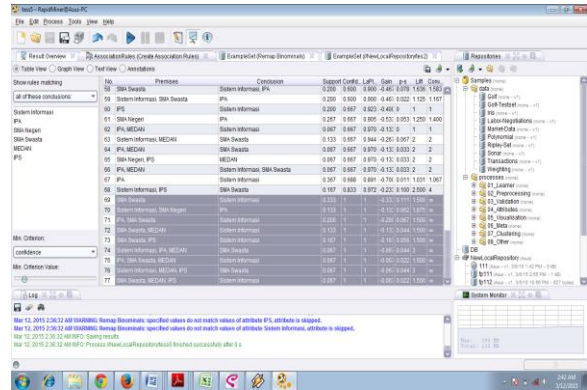
Gambar 5.1 Tampilan *Import Excel Sheet*

Gambar 5.1 adalah proses dari import data uji Xls ke dalam rapidminer



Gambar 5.2 Proses menghubungkan *FP-Growth* ke ras pertama dan *Create Association Rules* ke res kedua

Pada Gambar 5.2 dijelaskan tentang proses menghubungkan *FP-Growth* ke ras Pertama dan *Create Association Rules* ke ras kedua dan diatur pada minimum *support* dan *minimum confidencenya*.



Gambar 5.20 *Rule valid* jika nilai *Lift Ratio* > 1

Lift Ratio adalah parameter penting selain *support* dan *confidence* dalam *association rule*. *Lift Ratio* mengukur seberapa penting *rule* yang telah terbentuk berdasarkan nilai *support* dan *confidence*. *Lift Ratio* merupakan nilai yang menunjukkan kevalidan proses transaksi dan memberikan informasi apakah benar produk A dibeli bersamaan dengan produk B.

Improvement Ratio dapat dihitung dengan rumus

$$\frac{Support(A \cap B)}{SupportA * SupportB} =$$

Untuk mencari nilai *rule* yang *valid* adalah jika mempunyai nilai *Lift Ratio* > 1 dengan cara *Support Lift Ratio* = *Support* yang mengandung nilai A dan nilai B dibagi *support A* Support B* nilai *support* yang mengandung nilai A dan nilai B Adalah hasil dari minimum *support* dibagi dengan *item* (D1,C1,A3) { Sistem Informasi, IPA, SMA Negeri } Pada setiap kemunculan ditransaksi.

$$\frac{5/30}{20/30 * 16/30 * 12/30} = \frac{0,16}{0,66 * 0,53 * 0,4} = \frac{0,16}{0,1392} = 1,1494$$

Jadi dari hasil pencarian nilai *Lift Ratio* > 1 maka kita bisa menentukan sebuah *rule* yang *valid* dari sekian banyaknya *rule* yang dijalankan dengan *software Rapidminer*.

Adapun hasil dari *rule* yang paling mempengaruhi adalah:

Jika dia Asal sekolahnya SMA Swasta, Alamatnya Medan dan Asal Jurusan IPS maka dia memilih Prodi Sistem Informasi dengan tingkat kepercayaan 100% dan didukung 6% dari data keseluruhan.

Hasil dari *rule* tersebut yang akan dijadikan target dalam mempromosikan pendidikan.

6. SIMPULAN DAN SARAN

6.1 Simpulan

1. Metode dalam pencarian *Frequent Itemset* pohon keputusan menggunakan algoritma *FP-Growth* bekerja sangat baik dalam melakukan *Frequent Itemset* dengan proses pembentukan *FP-Tree* dengan menghasilkan *rule* dari data sampel mahasiswa baru.
2. Penentuan data variabel sangat menentukan tingkat akurasi *FP-Growth* yang dibuat dan besarnya presentase dalam menentukan *support* dan minimum *confidence* dipengaruhi oleh data variabel yang digunakan untuk mencari *frequent itemset* yang saling berhubungan untuk menemukan data variabel yang akan dijadikan strategi dalam promosi pendidikan.
3. Dari penelitian yang dilakukan ada beberapa atribut yang tidak digunakan dalam *rule* yang dihasilkan, sehingga

pemilihan atribut di dalam *dataset* sangat penting.

4. Algoritma *FP-Growth* dapat diterapkan untuk mendukung strategi promosi pendidikan pada Perguruan Tinggi. Adapun Informasi yang berkaitan dengan pelaksanaan promosi dapat tersedia dengan cepat, contohnya dalam pemilihan tempat yang akan dijadikan promosi pendidikan, sehingga pihak manajemen dari kampus sendiri dapat melakukan pengambilan keputusan dengan cepat.
5. Pelaksanaan promosi pendidikan pada perguruan tinggi sangat dibantu dengan adanya penerapan algoritma *FP-Growth* ini sehingga diharapkan efektifitas pelaksanaan promosi pendidikan akan dapat semakin ditingkatkan.
6. Untuk menemukan *rule* atau *knowledge* yang sangat berarti dalam menganalisa strategi dalam menentukan promosi pendidikan, dibutuhkan banyak data sampel yang digunakan untuk proses ekstraksi *rule* atau *knowledge* tersebut.

6.2 Saran

Salah satu tujuan penelitian ini adalah agar teknik dalam menentukan strategi promosi pendidikan di Kampus STMIK TRIGUNA DHARMA MEDAN dapat *diimplementasikan* atau diterapkan dengan baik, untuk itu saran yang dapat diperhatikan untuk *pengimplementasian* dari teknik ini adalah:

1. Pada penelitian ini, penulis hanya mencoba salah satu teknik dalam menemukan hubungan beberapa *item* yang digunakan untuk membentuk *FP-Tree* dalam menentukan strategi

promosi pendidikan dengan menggunakan Algoritma *FP-Growth*. Untuk menentukan strategi dalam promosi pendidikan yang baik dapat digunakan beberapa atau penggabungan beberapa teknik yang akan digabungkan dalam sistem pengambilan keputusan untuk menentukan strategi yang tepat dalam menentukan strategi promosi pendidikan atau bisa juga menggunakan algoritma *CT-Pro*. Oleh sebab itu perlunya dilakukan penelitian lebih lanjut untuk mendapatkan strategi promosi dan menentukan teori mana yang menghasilkan strategi yang paling baik.

2. Untuk mendapatkan hasil promosi yang baik sangat diperlukan sumber data yang lebih banyak dan lebih lengkap.
3. Untuk mengimplementasikan *rule* yang dihasilkan dari teknik dari algoritma *FP-Growth* dapat berkonsultasi dengan tim marketing yang ada di kampus tersebut. Sehingga dengan adanya hubungan algoritma *FP-Growth* dengan pihak manajemen dapat mengembangkan strategi dalam promosi pendidikan sehingga pihak kampus akan lebih maju lagi kedepannya.

Algoritma Association Rule Metode Apriori (Studi Kasus Di Polsekta Medan Sunggal). *Informasi dan Teknologi Ilmiah (INTI)*, Vol : III No : 1, 144-154.

RIRIANTI. 2014. Implementasi Algoritma FP-GROWTH Pada Aplikasi Prediksi Persediaan Sepeda Motor (Studi Kasus PT. Pilar Deli Labumas). *Pelita Informatika Budi Darma*, Vol : VI No: 1.

SENSUSE, G. G. D. D. I. 2012. Penerapan Metode Data Mining Market Basket Analysis Terhadap Data Penjualan Produk Buku Dengan Menggunakan Algoritma Apriori dan Frequent Pattern Growth (FP-GROWTH) : Studi Kasus Percetakan PT. GRAMEDIA. *Jurnal TELEMATIKA MKOM*, Vol.4 No.1, 118-132.

KUSRINI dan EMHA TAUFIQ LUTHFI. 2009. "Algoritma Data Mining." Jakarta : Andi. 10

7. DAFTAR PUSTAKA

BUULOLO, E. 2013. Implementasi Algoritma Apriori Pada Sistem Persediaan Obat (Studi Kasus : Apotik

Rumah Sakit Estomihi Medan). *Pelita Informatika Budi Darma*, Vol : IV No: 1, 71-83.

FADLINA. 2014. Data Mining Untuk Analisa Tingkat Kejahatan Jalanan Dengan