

Analisis Algoritma Asosiasi Untuk Mendapatkan Pola Pemilihan Judul Skripsi Mahasiswa STIE-GK Muara Bulian

Azwar Anas¹, Hendriyaldi²

STIE – Graha Karya Muara Bulian/Program Studi Manajemen
Jl. Gajah Mada Muara Bulian-Jambi 36613
E-mail: azwarzayn@gmail.com¹, hendriyaldi@yahoo.com²

Abstract

In this study, the authors use an association algorithm to get the pattern of choosing thesis title students STIE-GK Muara Bulian. The task of data mining is to generate all the rules of association on a transactional table, which has a support value of more than the minimum support. The rules should also have greater support from confidence. While to test the algorithm, the author uses data mining software Orange. The end result of this research is the pattern of title selection that most often appear simultaneously. Based on the discussion in this study, the value of support and confidence with reference 2-itemset which has the highest value is a combination of Marketing and Finance with 14% support and 18% confidence. As for the 3-itemset reference obtained 4 association rules that meet the minimum requirements of support and minimum confidence. It is found that all the combinations of Marketing, Finance and Human Resources titles have the highest score of 12% support and 85% confidence.

Keywords: algorithm, apriori, data, mining, title, thesis.

Abstrak

Dalam penelitian ini, penulis menggunakan algoritma asosiasi untuk mendapatkan pola pemilihan judul skripsi mahasiswa STIE-GK Muara Bulian. Tugas dari data mining adalah untuk menghasilkan semua kaidah asosiasi pada suatu tabel transaksional, yang mempunyai nilai support lebih dari minimum support. Kaidah tersebut juga harus mempunyai support yang lebih besar dari confidence. Sedangkan untuk melakukan uji coba algoritmanya, penulis menggunakan software data mining Orange. Hasil akhir dari penelitian ini adalah gambaran pola pemilihan judul yang paling sering muncul secara bersamaan. Berdasarkan pembahasan dalam penelitian ini, nilai support dan confidence dengan acuan 2-itemset yang memiliki nilai tertinggi adalah kombinasi antara Pemasaran dan Keuangan dengan support 14% dan confidence 18%. Sedangkan untuk acuan 3-itemset diperoleh 4 aturan asosiasi yang memenuhi syarat minimum support dan minimum confidence. Diperoleh bahwa semua kombinasi judul Pemasaran, Keuangan dan SDM memiliki nilai tertinggi yaitu support 12% dan confidence 85%.

Kata kunci: algoritma, apriori, data, mining, judul, skripsi.

© 2018 Jurnal MEDIASISFO.

1. Pendahuluan

Tugas pokok perguruan tinggi adalah Pendidikan dan pengajaran, Penelitian dan Pengabdian kepada Masyarakat yang tertuang dalam Tri Darma Perguruan Tinggi. Begitu pula mahasiswa, salah satu tugas pokoknya adalah melaksanakan penelitian. Dalam tahapan penelitian langkah pertama adalah pengajuan judul penelitian. Banyaknya jumlah mahasiswa yang mengajukan judul dari tahun ke tahun, maka akan banyak pula arsip atau daftar judul yang tersimpan.

Permasalahan yang ada yaitu menumpuknya jumlah judul skripsi yang telah diajukan oleh mahasiswa STIE-GK Muara Bulian. Jumlah judul skripsi yang begitu banyak justru bisa menjadi masalah bila perguruan tinggi tidak bisa dengan bijak memanfaatkannya. Data yang menumpuk hanya akan menjadi sampah apabila tidak digunakan dengan baik. Padahal data yang menumpuk merupakan salah satu aset

untuk diambil manfaatnya. Selama ini belum pernah dilakukan analisis terhadap tumpukan judul skripsi mahasiswa tersebut untuk diambil sebuah keputusan. Sehingga tidak ada perkembangan yang berarti dari pola pemilihan judul skripsi dari tahun ke tahun atau cenderung pada judul yang sama. Untuk dapat mengambil manfaat dari data masa lalu yang menumpuk tersebut, perlu dilakukan penambangan data atau istilah lainnya adalah data mining. Dalam penelitian ini, peneliti akan memanfaatkan algoritma apriori untuk mengelompokkan data judul skripsi berdasarkan kecenderungannya yang muncul bersamaan dalam suatu pengajuan.

Dalam penelitian ini, peneliti akan menggunakan algoritma apriori untuk menganalisis pola pemilihan judul skripsi dan menggunakan software Orange untuk melakukan uji coba terhadap hasil analisis tersebut.

Algoritma apriori termasuk jenis aturan asosiasi pada data mining. Selain apriori, yang termasuk dalam golongan ini adalah metode Generalized Rule Induction dan Algoritma Hash Based. Aturan yang menyatakan asosiasi antara beberapa atribut sering disebut affinity analysis atau market basket analysis.

Tugas dari data mining adalah untuk menghasilkan semua kaidah asosiasi pada suatu tabel transaksional, yang mempunyai nilai support lebih dari minimum support. Kaidah tersebut juga harus mempunyai support yang lebih besar dari confidence.

2. Tinjauan Pustaka

2.1. Landasan Teori

Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual [1]. Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [2].

Data mining dibagi menjadi beberapa teknik berdasarkan tugas yang dapat dilakukan, yaitu [3]:

- a. Deskripsi
Terkadang analis/ peneliti ingin mendeskripsikan pola dan trend yang tersimpan dalam data.
- b. Estimasi
Estimasi mirip dengan klasifikasi, kecuali variabel tujuan yang lebih ke arah numeric daripada kategori. Misalnya, akan dilakukan estimasi tekanan darah *systolic* dari pasien rumah sakit berdasarkan umur pasien, jenis kelamin, indeks berat badan, dan level sodium darah.
- c. Prediksi
Prediksi memiliki kemiripan dengan estimasi dan klasifikasi. Hanya saja, prediksi hasilnya menunjukkan sesuatu yang belum terjadi (mungkin terjadi di masa depan).
- d. Klasifikasi
Dalam klasifikasi variabel, tujuan bersifat kategorik. Misalnya, kita akan mengklasifikasikan pendapatan dalam 3 kelas, yaitu mendapatkan tinggi, pendapatan sedang, dan pendapatan rendah.
- e. *Clustering*
Clustering lebih ke arah pengelompokan *record*, pengamatan, atau kasus dalam kelas yang memiliki kemiripan. Sebuah *cluster* adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lain dan memiliki ketidakmiripan dengan *record-record* dalam *cluster* yang lain.
- f. Asosiasi
Mengidentifikasi hubungan antara berbagai peristiwa yang terjadi pada satu waktu. Pendekatan asosiasi tersebut menekankan sebuah kelas masalah yang dicirikan dengan analisis keranjang pasar.

Algoritma-algoritma *data mining* kebanyakan berasal atau merupakan bentuk pengembangan dari algoritma-algoritma dari bidang ilmu *machine learning*, statistika, intelegensia buatan dan jaringan syaraf tiruan. Karena tidak dirancang untuk menangani data dalam ukuran yang sangat besar, sedangkan *data*

mining dimaksudkan untuk menangani data dalam ukuran yang demikian, maka salah satu arah penelitian di bidang *data mining* adalah mengembangkan algoritma-algoritma tersebut agar dapat menangani data yang berukuran sangat besar.

Selain itu, integrasi teknik-teknik *data mining* ke dalam DBMS, khususnya *Object-Relation DBMS* (ORDBMS) yang merupakan teknologi DBMS terbaru, juga masih merupakan bidang penelitian yang aktif. Tujuan utama dari integrasi ini adalah untuk “melebur” algoritma *data mining* agar menjadi fungsi internal ORDBMS yang berkualitas, sehingga pengguna dapat menggunakannya sesuai dengan kebutuhan. Karena DBMS merupakan teknologi yang sudah matang, digunakan secara luas, dapat mengelola data dalam ukuran yang sangat besar, memfasilitasi kueri tabel yang mudah dengan *Structured Query Language* (SQL) dan memiliki fitur yang mengakomodasi kebutuhan pengguna ORDBMS, maka pengembangan dan peleburan algoritma *data mining* ke dalam ORDBMS yang baik adalah dengan memanfaatkan semua ini.

Banyak penerapan yang dapat dilakukan oleh *data mining*. Apalagi ditunjang kekayaan dan keanekaragaman bidang ilmu (*artificial intelligence*, *database*, statistik, pemodelan matematika, pengolahan citra, dan sebagainya) membuat penerapan *data mining* menjadi makin luas. Bidang penerapan *data mining* dapat dilakukan diantaranya adalah :

1. Analisa Pasar dan Manajemen
 - Untuk analisa pasar, banyak sumber data yang dapat digunakan seperti transaksi kartu kredit, kartu anggota klub tertentu, kupon diskon, keluhan pembeli, ditambah dengan studi tentang gaya hidup publik. Beberapa kondisi yang bisa diselesaikan dengan *data mining* diantaranya :
 - a. Menembak target pasar
 - b. Melihat pola beli pemakai dari waktu ke waktu
 - c. *Cross Market Analysis*
 - d. Profil *customer*
 - e. Identifikasi kebutuhan *customer*
 - f. Menilai loyalitas *customer*
 - g. Informasi *summary*
2. Analisa Perusahaan dan Manajemen Resiko
 - a. Perencanaan keuangan dan evaluasi aset
Data mining dapat membantu melakukan analisis dan prediksi *cash flow* serta melakukan *contingent claim analysis* untuk mengevaluasi aset. Selain itu juga dapat menggunakannya untuk analisis tren.
 - b. Perencanaan sumber daya (*Resource Planning*)
Dengan melihat informasi ringkas (*summary*) serta pola pembelajaran dan pemasukan dari masing-masing *resource*, dan dapat memanfaatkannya untuk melakukan *resource planning*.
 - c. Persaingan (*Competition*)
 - I. Sekarang ini banyak perusahaan yang berupaya untuk dapat melakukan *competitive intelligence*. *Data mining* dapat membantu memonitor pesaing-pesaing dan melihat *market direction*.
 - II. Dapat melakukan pengelompokan *customer* dan memberikan variasi harga/ layanan/ bonus untuk masing-masing grup.
 - III. Menyusun strategi penetapan harga di pasar yang sangat kompetitif. Hal ini diterapkan oleh perusahaan minyak REPSOL di Spanyol dalam menetapkan harga jual gas di pasaran.
3. Telekomunikasi
Sebuah perusahaan telekomunikasi menerapkan *data mining* untuk melihat dari jutaan transaksi yang masuk, transaksi mana sajakah yang masih harus ditangani secara manual (dilayani oleh orang). Tujuannya tidak lain adalah untuk menambah layanan otomatis khusus untuk transaksi-transaksi yang masih dilayani secara manual. Dengan demikian jumlah operator penerima transaksi manual tetap bisa ditekan minimal.
4. Keuangan
Financial Crimes Enforcement Network di Amerika Serikat baru-baru ini menggunakan *data mining* untuk menambang trilyunan dari berbagai subjek seperti properti, rekening bank dan transaksi

- keuangan lainnya untuk mendeteksi transaksi-transaksi keuangan yang mencurigakan (seperti *money laundry*), yang susah dilakukan jika menggunakan analisa standar.
5. Asuransi
Australian Health Insurance Commision menggunakan *data mining* untuk mengidentifikasi layanan kesehatan yang sebenarnya tidak perlu tetapi tetap dilakukan oleh peserta asuransi. Hasilnya, berhasil menghemat satu juta dolar per tahunnya. Tentu saja ini tidak hanya bisa diterapkan untuk asuransi kesehatan, tetapi juga untuk berbagai jenis asuransi lainnya.
 6. Olah Raga
IBM Advanced Scout menggunakan *data mining* untuk menganalisis statistik permainan NBA (jumlah *shots blocked*, *assist* dan *fouls*) dalam rangka mencapai keunggulan bersaing (*competitive advantage*) untuk tim New York Knicks dan Miami Heat.
 7. Astronomi
Jet Propulsion Laboratory (JPL) di Pasadena, Clifornia dan *Palomar Observatory* berhasil menemukan 22 quasar dengan bantuan *data mining*. Hal ini merupakan salah satu kesuksesan penerapan *data mining* di bidang astronomi dan ilmu ruang angkasa.
 8. *Internet Web Surf-Aid*
IBM Surf-Aid menggunakan algoritma *data mining* untuk mendata akses halaman web khususnya yang berkaitan dengan pemasaran guna melihat perilaku dan minat *customer* serta melihat keefektifan pemasaran melalui web.

Data mining merupakan salah satu aktivitas di bidang perangkat lunak yang dapat memberikan ROI (*return on investment*) yang tinggi. Namun demikian *data mining* tetaplah hanya alat bantu yang dapat membantu manusia untuk melihat pola, menganalisis *trend* dalam rangka mempercepat pembuatan keputusan.

Association rule merupakan salah satu metode yang bertujuan mencari pola yang sering muncul di antara banyak transaksi, dimana setiap transaksi terdiri dari beberapa *item* [4]. Contoh dari aturan asosiatif dari analisa pembelian di suatu pasar swalayan adalah dapat diketahuinya berapa besar kemungkinan seorang pelanggan membeli roti bersama dengan susu. Dengan pengetahuan tersebut, pemilik pasar swalayan dapat mengatur penempatan barangnya atau merancang kampanye pemasaran dengan memakai kupon diskon untuk kombinasi barang tertentu. Karena analisis asosiasi menjadi terkenal karena aplikasinya untuk menganalisa isi keranjang belanja di pasar swalayan, analisis asosiasi juga sering disebut dengan istilah *market basket analysis*.

Association rule merupakan teknik *data mining* untuk mencari pola hubungan dalam data atau basis data. Yang paling populer adalah *market basket analysis* (MBA). Teknik MBA ini digunakan untuk mencari pelanggan yang membeli barang pada took X, akan membeli barang Y dan seterusnya [5].

Tabel 1. Contoh Transaksi untuk Association Rule Analysis

Trans	Itemset
1	A, B, E
2	B, D
3	B, C
4	A, B, D
5	A, C
6	B, C
7	A, C
8	A, B, C, E
9	A, B, C

Dalam *association analysis* *itemset* merupakan kumpulan nol atau lebih item. Pada contoh tabel di atas diumpamakan *itemset* adalah barang yang dibeli pada took tersebut, trans adalah kode transaksi. Pada tabel tersebut dapat dilihat pada transaksi 1 yang dibeli adalah barang A, B dan E. Transaksi 2 yang dibeli adalah B dan D, demikian seterusnya. *Association analysis* akan melakukan analisis hubungan dengan aturan tertentu sering disebut sebagai *association rule*.

Sebagai contoh tabel tersebut, pembeli yang biasanya membeli barang A dan B akan membeli barang C. aturan ini dapat ditulis $\{A, B\} \rightarrow \{C\}$. Untuk memperoleh kuantifikasinya terminologi *support* dan *confidence* perlu diperhatikan. *Support* dan *confidence* dapat dihitung sebagai berikut :

$$\text{Support, } s(X \rightarrow Y) = \frac{\sigma(XUY)}{N} \quad (1)$$

$$\text{Confidence, } c(X \rightarrow Y) = \frac{\sigma(XUY)}{\sigma_X} \quad (2)$$

Di mana X dan Y sebagai *itemset* dan σ adalah *support* dari *itemset*.

Dengan rumus tersebut maka *support* yang didapat dari hubungan $s(A, B \rightarrow C)$ adalah : $\frac{2}{9} = 0, 22$. Sedangkan nilai *confidence* yang diperoleh adalah $c(A, B \rightarrow C)$ adalah : $\frac{2}{4} = 0, 5$. Nilai 2 pada *support* diperoleh dari *itemset* A, B, C berjumlah 2 yaitu pada trans 8 dan 9, jumlah keseluruhan trans adalah 9. Sedangkan 2 pada *confidence* sama seperti pada *support*, 4 diperoleh dari jumlah A, B yaitu pada trans 1, 4, 8 dan 9. Dari perolehan tersebut, kita dapat membuat aturan persentasi yang diinginkan untuk menentukan kepastian asosiasinya. Paling baik adalah apabila nilai *support* maupun *confidence* tinggi.

Apriori adalah suatu algoritma untuk melakukan pencarian *frequent itemset* untuk mendapatkan *association rules*. Sesuai dengan namanya, algoritma ini menggunakan *prior knowledge* mengenai *frequent itemset properties* yang telah diketahui sebelumnya, untuk memproses informasi selanjutnya. Apriori menggunakan pendekatan secara *iterative* yang disebut juga sebagai *level-wise search* di mana *k-itemset* digunakan untuk mencari $(k+1)$ -*itemset*.

Algoritma apriori menggunakan *knowledge* mengenai *frequent itemset* yang telah diketahui sebelumnya, untuk memproses informasi selanjutnya. Pola frekuensi tinggi adalah pola-pola item dalam *database* yang memiliki frekuensi atau *support* di atas ambang batas tertentu yang disebut istilah *minimum support*. Pola frekuensi tinggi ini digunakan untuk menyusun aturan asosiatif dan juga beberapa teknik *data mining* lainnya.

Prinsip dari algoritma apriori adalah :

1. Kumpulkan jumlah item tunggal, dapatkan item besar
2. Dapatkan *candidate pairs*, hitung \rightarrow *large pairs* dari item-item
3. Dapatkan *candidate triplets*, hitung \rightarrow *large triplets* dari item-item dan seterusnya
4. Sebagai petunjuk : setiap *subset* dari sebuah *frequent itemset* harus menjadi *frequent*

Dua proses utama dalam algoritma apriori merupakan langkah yang akan dilakukan untuk mendapat *frequent itemset*. Walaupun algoritma apriori mudah untuk dipahami dan diimplementasikan dibandingkan algoritma yang lainnya yang memang diterapkan untuk proses *association rule*, akan tetapi algoritma apriori juga memiliki kekurangan yaitu untuk melakukan pencarian *frequent itemset*, algoritma apriori harus melakukan *scanning database* berulang kali untuk setiap kombinasi item. Hal tersebut menyebabkan banyaknya waktu yang dibutuhkan untuk melakukan *scanning database*. Selain itu dibutuhkan *generate candidate* yang besar untuk mendapatkan kombinasi item dari *database*.

2.2. Penelitian Terdahulu

Penerapan *algoritma apriori* telah banyak digunakan sebelumnya untuk mendapatkan informasi berharga dari sejumlah frekuensi data. Berikut beberapa penerapan algoritma apriori yang telah dilakukan :

1. Penerapan Data Mining Algoritma Asosiasi untuk Meningkatkan Penjualan [6].
2. Analisis *Market Basket* dengan Algoritma Apriori dan FP-Growth [7].
3. Implementasi *Association Rule* terhadap Penyusunan *Layout* Makanan dan Penentuan Paket Makanan Hemat di RM Roso Echo dengan Algoritma Apriori [8].
4. Implementasi Data Mining pada Penjualan Tiket Pesawat Menggunakan Algoritma Apriori [9].
5. Implementasi Data Mining dengan Metode Algoritma Apriori dalam Menentukan Pola Pembelian Obat [10].

3. Metode Penelitian

Dalam melakukan sebuah penelitian, tentu harus berdasarkan metode penelitian yang benar sehingga dapat mempermudah jalannya penelitian tersebut. Metode penelitian merupakan kerangka kerja dalam melakukan penelitian. Dengan mengikuti kerangka kerja tersebut maka penelitian yang dilakukan akan berjalan dengan sistematis dan memberikan hasil yang baik.

Pada bagian ini akan diuraikan kerangka penelitian, kerangka ini merupakan langkah-langkah yang akan dilakukan dalam penyelesaian masalah yang akan dibahas. Langkahnya diuraikan seperti berikut ini :

1. Mendefinisikan Ruang Lingkup Masalah
Ruang lingkup masalah yang akan diteliti harus ditentukan terlebih dahulu, karena tanpa mampu menentukan serta mendefinisikan batasan masalah yang akan diteliti, maka tidak akan pernah didapat suatu solusi yang terbaik dari masalah tersebut. Jadi langkah pertama ini adalah langkah awal yang terpenting dalam penulisan ini.
2. Analisis Masalah
Langkah analisis masalah adalah untuk dapat memahami masalah yang telah ditentukan ruang lingkup atau batasannya. Dengan menganalisa masalah yang telah ditentukan tersebut, maka diharapkan masalah dapat dipahami dengan baik. Pada bagian ini digambarkan proses untuk mendapatkan perilaku konsumen dengan menggunakan algoritma apriori. Disamping itu juga diuraikan tentang kebutuhan data untuk menggali kaidah asosiasi judul skripsi. Dari hasil analisis tersebut, maka algoritma apriori hendaknya mampu mendapatkan pola pemilihan judul skripsi dengan baik.
3. Penentuan Tujuan
Berdasarkan pemahaman dari masalah, maka ditentukan tujuan yang akan dicapai dari penulisan ini. Pada tujuan ini ditentukan target yang dicapai, terutama yang dapat mengatasi masalah-masalah yang ada.
4. Mempelajari Literatur
Untuk mencapai tujuan, maka dipelajari berbagai literatur yang relevan dengan masalah yang diteliti. Kemudian literatur-literatur yang dipelajari tersebut diseleksi untuk dapat ditentukan literatur-literatur mana yang akan digunakan dalam penelitian.
Literatur diambil dari internet, yang berupa artikel dan jurnal ilmiah tentang *data mining*, *association rule* dan algoritma apriori, serta bahan bacaan lain yang mendukung penelitian.
5. Pengumpulan Data dan Informasi
Dalam pengumpulan data dilakukan observasi yaitu pengamatan secara langsung di tempat penelitian sehingga permasalahan yang ada dapat diketahui dengan jelas. Kemudian dilakukan *interview* yang bertujuan untuk mendapatkan informasi atau data yang dibutuhkan. Selain itu juga dilakukan studi kepustakaan yaitu dengan membaca buku-buku yang menunjang dalam melakukan analisa terhadap data dan informasi yang didapat.
Adapun data-data yang diperlukan dalam penelitian ini adalah data daftar judul skripsi mahasiswa STIE-GK Muara Bulian.
6. Analisis Teknik yang digunakan
Bagian ini bertujuan untuk menganalisis dan memahami teknik yang akan digunakan dalam pengolahan data yang telah diperoleh dari tempat penelitian, terutama pada proses, yang mana teknik yang digunakan untuk pengolahan data dengan algoritma apriori.
7. Implementasi
Setelah dilakukan analisis teknik yang digunakan, maka dilakukan implementasi terhadap *software* dalam hal ini adalah *software data mining Orange*. Tentu saja penerapan analisis pada *software* tersebut menggunakan algoritma apriori. Jika penerapan sistem sudah berjalan dengan lancar dan akurat, maka sistem dapat diimplementasikan sesuai dengan kebutuhan.
8. Pengujian
Pengujian dilakukan untuk mendapatkan hasil yang sesuai dengan melakukan perbandingan antara perhitungan manual dan komputerisasi, berikut ini langkah-langkah pengujian yang akan dilakukan :
 - a. Melakukan pengujian algoritma apriori terhadap data pengajuan judul skripsi secara manual.
 - b. Setelah dilakukan perhitungan manual, selanjutnya dilakukan perhitungan menggunakan *software data mining Orange* dengan memasukkan data-data dan diproses menggunakan aturan asosiasi.

- c. Tahap terakhir adalah melakukan perbandingan antara perhitungan manual dengan komputerisasi, jika hasil yang dicapai sama atau hampir sama, maka data yang diproses dan teknik yang digunakan telah benar.

4. Hasil dan Pembahasan

4.1. Analisis Asosiasi dengan Algoritma Apriori

Dalam penelitian ini, penulis hanya melakukan analisis terhadap 50 mahasiswa yang mengajukan judul skripsi. Dari 50 data yang telah diamati, penulis hanya mengamati mahasiswa yang mengajukan maksimal 3 judul.

Analisis asosiasi atau *association rule mining* adalah teknik *data mining* untuk menentukan aturan asosiatif antara aturan kombinasi item. Contoh dari aturan asosiatif dari analisis pola pengajuan judul skripsi adalah mengetahui besarnya kemungkinan seorang mahasiswa untuk mengajukan judul Keuangan bersamaan dengan Pemasaran. Dengan pengetahuan tersebut, perguruan tinggi dapat mengelompokkan judul skripsi yang diajukan oleh mahasiswa.

Dalam menentukan pola pengajuan judul skripsi mahasiswa, akan digunakan algoritma apriori untuk melakukan analisisnya. Data-data judul yang telah dikumpulkan akan diolah menggunakan prinsip kerja algoritma apriori yang mampu memecahkan data bersifat *frequent itemset* atau mencari pola berulang.

Aturan asosiasi biasanya dinyatakan dalam bentuk :

{SDM, Keuangan} → {Pemasaran} {*support* = 40%, *confidence* = 50% }

Artinya 50% dari transaksi pengajuan SDM dan Keuangan juga memuat Pemasaran. Sementara 40% dari seluruh pengajuan judul yang ada memuat ketiga *item* tersebut.

4.2. Menentukan Variabel

Data yang akan diolah dalam menentukan pola pengajuan judul ini adalah data judul skripsi mahasiswa STIE-GK. Adapun variabel-variabel data pengajuan judul yang akan diolah adalah:

1. Mahasiswa
2. Judul 1
3. Judul 2
4. Judul 3

Dari variabel-variabel di atas, setiap mahasiswa akan dibagi kedalam beberapa kelompok berdasarkan jenis judul yang sama, hal ini dilakukan agar proses analisa data dapat dilakukan dengan mudah menggunakan algoritma apriori.

4.3. Analisis Frekuensi Tinggi

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

$$\text{Support}(A) = \frac{\text{Jumlah transaksi mengandung } A}{\text{Jumlah transaksi}} \quad (1)$$

Sedangkan nilai *support* dari 2 item diperoleh dari rumus berikut :

$$\text{Support}(A,B) = P(A \cap B)$$

$$\text{Support}(A,B) = \frac{\sum \text{Jumlah transaksi mengandung } A \text{ dan } B}{\sum \text{Jumlah transaksi}} \quad (2)$$

4.4. Melakukan Proses

Berdasarkan variabel-variabel data judul di atas, maka format data dapat dilihat pada tabel berikut ini :

Tabel 2. Data Judul Skripsi

Mhs	Judul 1	Judul 2	Judul 3	Mhs	Judul 1	Judul 2	Judul 3
1	Keuangan	Pemasaran	SDM	26	Pemasaran	SDM	Keuangan
2	SDM			27	Keuangan	Pemasaran	
3	Pemasaran	Keuangan		28	SDM	Keuangan	Pemasaran
4	Pemasaran	SDM	Keuangan	29	Pemasaran	Keuangan	SDM
5	Keuangan	Pemasaran		30	SDM	Pemasaran	
6	SDM	Keuangan	Pemasaran	31	Keuangan	SDM	
7	Pemasaran	Keuangan	SDM	32	Pemasaran	SDM	Keuangan
8	SDM	Pemasaran		33	Pemasaran	SDM	Keuangan
9	Keuangan	SDM		34	SDM	Keuangan	Pemasaran
10	Pemasaran	SDM	Keuangan	35	Pemasaran	Keuangan	SDM
11	Pemasaran	SDM	Keuangan	36	SDM	Pemasaran	
12	Keuangan	Pemasaran		37	Keuangan	SDM	
13	SDM	Keuangan	Pemasaran	38	Pemasaran	SDM	Keuangan
14	Pemasaran	Keuangan	SDM	39	Pemasaran	SDM	Keuangan
15	SDM	Pemasaran		40	Keuangan	Pemasaran	SDM
16	Keuangan	Pemasaran		41	SDM		
17	SDM	Keuangan	Pemasaran	42	Pemasaran	Keuangan	
18	Pemasaran	Keuangan	SDM	43	Pemasaran	SDM	Keuangan
19	SDM	Pemasaran		44	Keuangan	Pemasaran	
20	Keuangan	SDM		45	Keuangan	SDM	
21	Pemasaran	SDM	Keuangan	46	Pemasaran	SDM	Keuangan
22	Pemasaran	SDM	Keuangan	47	Pemasaran	SDM	Keuangan
23	Keuangan	Pemasaran	SDM	48	Keuangan	Pemasaran	SDM
24	SDM			49	SDM		
25	Pemasaran	Keuangan		50	Pemasaran	Keuangan	

Setelah data judul diperoleh, selanjutnya adalah mengelompokkan mahasiswa berdasarkan judul skripsi yang diajukan sebagaimana tabel berikut.

Tabel 3. Data Judul Skripsi per Mahasiswa

Mhs	Judul	Mhs	Judul
1	Keuangan, Pemasaran, SDM	23	Keuangan, Pemasaran, SDM
2	SDM	24	SDM
3	Pemasaran, Keuangan	25	Pemasaran, Keuangan
4	Pemasaran, Keuangan, SDM	26	Pemasaran, SDM, Keuangan
5	Keuangan, Pemasaran	27	Keuangan, Pemasaran
6	SDM, Keuangan, Pemasaran	28	SDM, Keuangan, Pemasaran
7	Pemasaran, Keuangan, SDM	29	Pemasaran, Keuangan, SDM
8	SDM, Pemasaran	30	SDM, Pemasaran
9	Keuangan, SDM	31	Keuangan, SDM
10	Pemasaran, SDM, Keuangan	32	Pemasaran, SDM, Keuangan
11	Pemasaran, SDM, Keuangan	33	Pemasaran, SDM, Keuangan
12	Keuangan, Pemasaran	34	SDM, Keuangan, Pemasaran
13	SDM, Keuangan, Pemasaran	35	Pemasaran, Keuangan, SDM
14	Pemasaran, Keuangan, SDM	36	SDM, Pemasaran
15	SDM, Pemasaran	37	Keuangan, SDM
16	Keuangan, Pemasaran	38	Pemasaran, SDM, Keuangan
17	SDM, Keuangan, Pemasaran	39	Pemasaran, SDM, Keuangan
18	Pemasaran, Keuangan, SDM	40	Keuangan, Pemasaran, SDM
19	SDM, Pemasaran	41	SDM
20	Keuangan, SDM	42	Pemasaran, Keuangan
21	Pemasaran, SDM, Keuangan	43	Pemasaran, SDM, Keuangan
22	Pemasaran, SDM, Keuangan	44	Keuangan, Pemasaran

45	Keuangan, SDM	48	Keuangan, Pemasaran, SDM
46	Pemasaran, SDM, Keuangan	49	SDM
47	Pemasaran, SDM, Keuangan	50	Pemasaran, Keuangan

Data transaksional di atas lalu direpresentasikan dalam bentuk seperti terlihat pada tabel berikut ini :

Tabel 4. Representasi Data Judul Skripsi

Mhs	Judul	Mhs	Judul	Mhs	Judul
1	Keuangan	17	Pemasaran	33	SDM
1	Pemasaran	18	Pemasaran	34	SDM
1	SDM	18	Keuangan	34	Keuangan
2	SDM	18	SDM	34	Pemasaran
3	Pemasaran	19	SDM	35	Pemasaran
3	Keuangan	19	Pemasaran	35	Keuangan
4	Pemasaran	20	Keuangan	36	Pemasaran
4	Keuangan	20	SDM	37	Keuangan
4	SDM	21	Pemasaran	37	SDM
5	Keuangan	21	SDM	38	Pemasaran
5	Pemasaran	21	Keuangan	38	SDM
6	SDM	22	Pemasaran	38	Keuangan
6	Keuangan	22	SDM	39	Pemasaran
6	Pemasaran	22	Keuangan	39	SDM
7	Pemasaran	23	Keuangan	39	Keuangan
7	Keuangan	23	Pemasaran	40	Keuangan
7	SDM	23	SDM	40	Pemasaran
8	SDM	24	SDM	40	SDM
8	Pemasaran	25	Pemasaran	41	SDM
9	Keuangan	25	Keuangan	42	Pemasaran
9	SDM	26	Pemasaran	42	Keuangan
10	Pemasaran	26	SDM	43	Pemasaran
10	SDM	26	Keuangan	43	SDM
10	Keuangan	27	Keuangan	43	Keuangan
11	Pemasaran	27	Pemasaran	44	Keuangan
11	SDM	28	SDM	44	Pemasaran
11	Keuangan	28	Keuangan	45	Keuangan
12	Keuangan	28	Pemasaran	45	SDM
12	Pemasaran	29	Pemasaran	46	Pemasaran
13	SDM	29	SDM	46	SDM
13	Keuangan	29	Keuangan	46	Keuangan
13	Pemasaran	30	SDM	47	Pemasaran
14	Pemasaran	30	Pemasaran	47	SDM
14	Keuangan	31	Keuangan	47	Keuangan
14	SDM	31	SDM	48	Keuangan
15	SDM	32	Pemasaran	48	Pemasaran
15	Pemasaran	32	SDM	48	SDM
16	Keuangan	32	Keuangan	49	SDM
16	Pemasaran	33	Pemasaran	50	Pemasaran
17	SDM	33	Keuangan	50	Keuangan
17	Keuangan				

Pada pengujian ini, penulis menetapkan nilai *minimum support* sebesar 5% dan *minimum confidence* sebesar 10%. Sehingga data yang akan diproses oleh *software Orange* hanyalah *rule-rule* yang memenuhi kriteria di atas. Tabel berikut ini menunjukkan calon *2-itemset* dari data judul skripsi yang memenuhi ketentuan *minimum support* dan *minimum confidence* berdasarkan pada tabel 5.

Tabel 5. Calon 2-itemset

Kombinasi	Jumlah
Pemasaran, Keuangan	7
SDM, Pemasaran	4
Keuangan, SDM	5

Dari tabel 5 di atas, dapat disimpulkan jika ditetapkan nilai dari *threshold* di mana $\Phi = 2$, maka : $F_2 = \{(Pemasaran, Keuangan), (SDM, Pemasaran), (Keuangan, SDM)\}$.

Tabel 6. Calon 3-itemset

Kombinasi	Jumlah
Keuangan, Pemasaran, SDM	3
Pemasaran, Keuangan, SDM	6
SDM, Keuangan, Pemasaran	5
Pemasaran, SDM, Keuangan	10

Kombinasi dari *itemset* dalam F_2 dapat kita gabungkan menjadi calon *3-itemset*. *Itemset-itemset* dari F_2 yang dapat digabungkan adalah *itemset-itemset* yang memiliki kesamaan dalam $k-1$ *item* pertama. Calon *3-itemset* yang dapat dibentuk dari F_2 tampak pada tabel 7. Dengan demikian $F_3 = \{(Keuangan, Pemasaran, SDM), (Pemasaran, Keuangan, SDM), (SDM, Keuangan, Pemasaran), (Pemasaran, SDM, Keuangan)\}$.

4.5. Pembentukan Aturan Asosiasi

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiasi yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan asosiasi $A \rightarrow B$. Nilai *confidence* dari aturan $A \rightarrow B$ diperoleh dari rumus berikut.

$$Confidence = P(B | A) \quad (3)$$

Jika mengajukan judul Keuangan dan Pemasaran, maka juga mengajukan judul SDM. Untuk mendapatkan nilai *support* dari aturan ini maka masukkan nilai total total judul yang mengandung Keuangan, Pemasaran dan SDM lalu dibagi dengan total pengajuan judul. Rujukan dari rumus ini adalah pada tabel 6 dengan *3-itemset* dan tabel 5 dengan *2-itemset*. Sementara untuk mendapatkan nilai *confidence* dari aturan ini adalah dengan menghitung total pengajuan judul mengandung Keuangan, Pemasaran dan SDM lalu dibagi dengan total pengajuan judul yang mengandung Keuangan dan Pemasaran.

Aturan asosiasi yang terbentuk berdasarkan *minimum support* 5% dan *minimum confidence* 10% dapat dilihat pada tabel 7 berikut.

Tabel 7. Aturan Asosiasi 2-itemset

Aturan	Support	Confidence
Jika mengajukan judul Pemasaran, maka mengajukan judul Keuangan.	14%	18%
Jika mengajukan judul SDM, maka mengajukan judul Pemasaran.	8%	9%
Jika mengajukan judul Keuangan, maka mengajukan judul SDM.	1%	12%

Dari tabel 7 di atas dapat disimpulkan nilai *support* dan *confidence* dengan acuan *2-itemset* yang memiliki nilai tertinggi adalah kombinasi antara Pemasaran dan Keuangan dengan *support* 14% dan *confidence* 18%.

Sementara aturan asosiasi dengan *3-itemset* yang memenuhi kriteria *minimum support* dan *minimum confidence* dapat dilihat pada tabel 8 berikut.

Tabel 8. Aturan Asosiasi 3-itemset

Aturan	Sup	Conf
Jika mengajukan judul Keuangan dan Pemasaran, maka mengajukan judul SDM.	6%	42%

Jika mengajukan judul Pemasaran dan Keuangan, maka mengajukan judul SDM	12%	85%
Jika mengajukan judul SDM dan Keuangan, maka mengajukan judul Pemasaran	1%	1%
Jika mengajukan judul Pemasaran dan SDM, maka mengajukan judul Keuangan	2%	25%

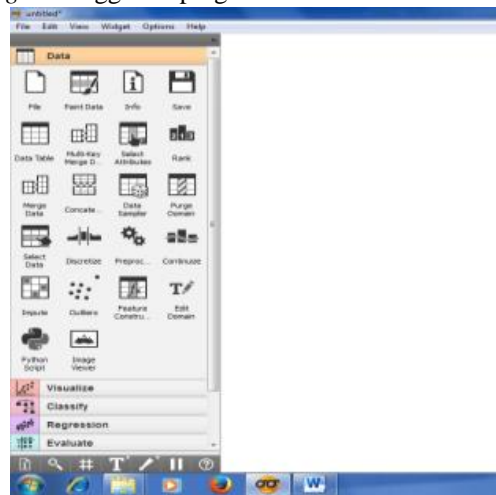
Dari tabel 8 di atas diperoleh 4 aturan asosiasi yang memenuhi syarat *minimum support* dan *minimum confidence*. Dapat pula disimpulkan bahwa semua kombinasi judul Pemasaran, Keuangan dan SDM memiliki nilai tertinggi yaitu *support* 12% dan *confidence* 85%.

4.6. Hasil Pengujian pada Software

Pengujian terhadap hasil analisa, sangat penting dilakukan untuk menentukan dan memastikan apakah hasil analisa tersebut benar atau tidak. Software yang penulis gunakan dalam pengujian ini adalah *Orange*.

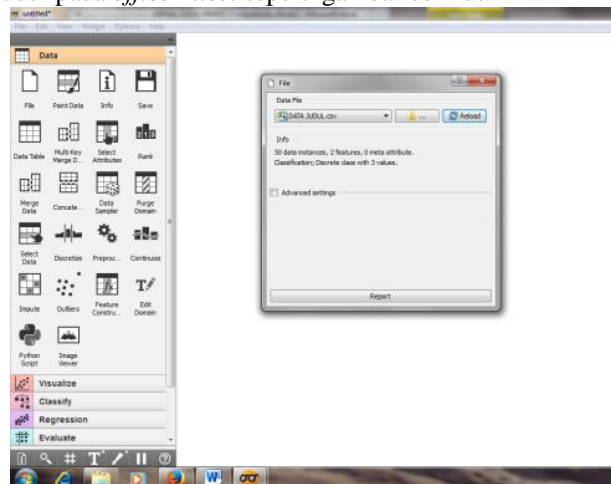
Tahapan dalam pengujian pada *Orange* adalah sebagai berikut:

1. Aktifkan *software Orange* sehingga tampil gambar berikut.



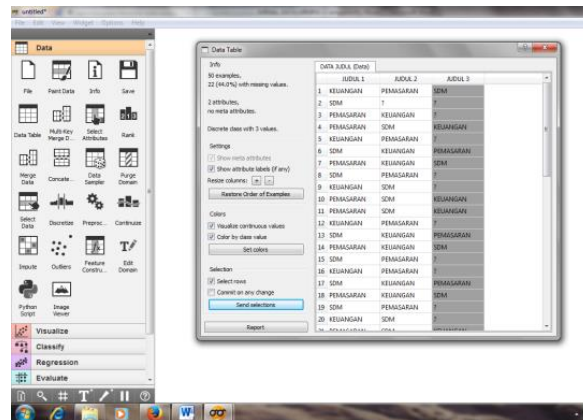
Gambar 1. Area Kerja pada Orange

2. Setelah jendela *Orange* terbuka, klik menu *File*, klik kanan lalu pilih *Open* untuk membuka file daftar pengajuan judul pada *office Excel* seperti gambar berikut.



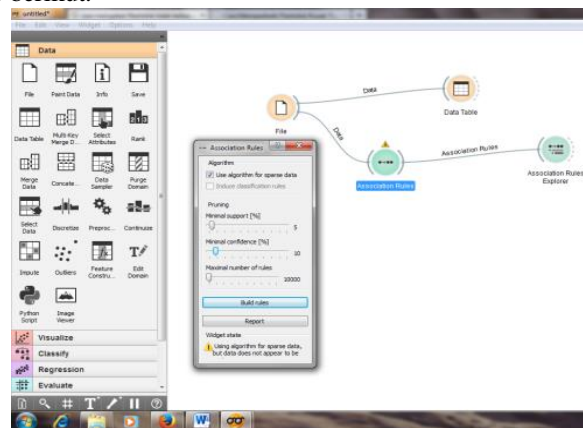
Gambar 2. Proses Pemanggilan Data

3. Langkah berikutnya adalah melakukan pengecekan apakah data terbaca pada *Orange*, caranya dengan mengklik menu *Data Table* tepat di bawah menu *File*, hubungkan menu *File* dan *Data Table* lalu klik kanan dan pilih *Open* pada *Data Table* untuk melihat tabel judul skripsi seperti gambar berikut.



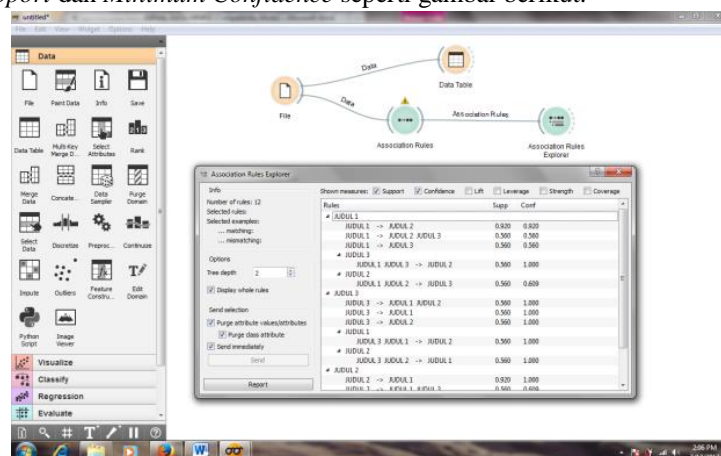
Gambar 3. Daftar Judul Skripsi pada Orange

- Langkah berikutnya adalah memasukkan menu *Association Rules*, hubungkan menu *File* dengan *Association Rules*, klik kanan *Open*, lalu atur ukuran *Minimum Support 5%* & *Minimum Confidence 10%* seperti gambar berikut.



Gambar 4. Pengaturan Support dan Confidence

- Selanjutnya masukkan menu *Association Rules Explorer*, hubungkan menu *Association Rules* dengan *Association Rules Explorer*, klik kanan lalu pilih *Open* untuk melihat *Rules* yang memenuhi *Minimum Support* dan *Minimum Confidence* seperti gambar berikut.



Gambar 5. Hasil Akhir Rule Asosiasi

5. Kesimpulan

5.1 Simpulan

Dari uraian di atas dapat ditarik kesimpulan :

1. Nilai *support* dan *confidence* dengan acuan *2-itemset* yang memiliki nilai tertinggi adalah kombinasi antara judul Pemasaran dan Keuangan dengan *support* 14% dan *confidence* 18%.
2. Diperoleh 4 aturan asosiasi yang memenuhi syarat *minimum support* dan *minimum confidence*. Kombinasi judul Pemasaran, Keuangan dan SDM memiliki nilai tertinggi yaitu *support* 12% dan *confidence* 85%.

5.2 Saran

Berdasarkan pembahasan di atas, saran penulis sebagai berikut:

1. Judul yang telah banyak diajukan oleh mahasiswa, agar dibatasi penerimaannya.
2. Judul yang kurang diminati oleh mahasiswa, agar diberikan pemahaman kepada mahasiswa.
3. Lakukan analisis secara berkelanjutan terhadap judul-judul yang telah diajukan oleh mahasiswa.

6. Daftar Rujukan

- [1] Wirdah Choiriah. (2016). Penggunaan Algoritma Apriori Data Mining Untuk Mengetahui Tingkatkesetiaan Konsumen (Brand Loyalty) Terhadap Merek Kendaraan Bermotor (Studi Kasus Dealer Honda Rumbai), *X(X)*, 44–52.
- [2] Pane, D. K. (2013). Implementasi Data Mining Pada Penjualan Produk Elektronik Dengan Algoritma Apriori (Studi Kasus : Kreditplus). *Pelita Informatika Budi Darma, volume : 1*, 25–29. <https://doi.org/2301-9425>
- [3] Tampubolon, K., Saragih, H., Reza, B., Epicentrum, K., Asosiasi, A., & Apriori, A. (2013). Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-Alat Kesehatan. *Informasi Dan Teknologi Ilmiah*, 93–106. <https://doi.org/2339-210X>
- [4] Wandu, N., Hendrawan, R. A., & Mukhlason, A. (2012). Pengembangan Sistem Rekomendasi Penelusuran Buku dengan Penggalan Association Rule Menggunakan Algoritma Apriori. *Jurnal Teknik ITS*, 1, 1–5.
- [5] Widodo. (2008). Prediksi Mata Kuliah Pilihan dengan Aturan Asosiasi. *Science*, 2008, 21–23.
- [6] Emha Taufiq Luthfi. (2009). Penerapan Data Mining Algoritma Asosiasi Untuk Meningkatkan Penjualan. *Seminar*, 10(1), 1–21.
- [7] Erwin (Jurusan Teknik Informatika, Fakultas Ilmu Komputer, U. S. A. (2009). Analisis Market Basket Dengan Algoritma. *Jurnal Generic*, 4, 26–30.
- [8] Widiati, E., & Evita Dewi, K. (2014). Implementasi Association Rule Terhadap Penyusunan Layout Makanan Dan Penentuan Paket Makanan Hemat Di Rm Roso Echo Dengan Algoritma Apriori. *Jurnal Ilmiah Komputer Dan Informatika (KOMPUTA)*, 96(2), 2089–9033.
- [9] Siregar, S. R. (2014). Implementasi Data Mining Pada Penjualan Tiket Pesawat Menggunakan Algoritma Apriori (Studi Kasus : Jumbo Travel Medan), *152*, 152–156.
- [10] Yanto, R., & Khoiriah, R. (2015). Implementasi Data Mining dengan Metode Algoritma Apriori dalam Menentukan Pola Pembelian Obat, 2, 102–113. <https://doi.org/2354-5771>