

Implementasi Data Mining Menggunakan Algoritma Apriori Untuk Mendapatkan Pola Kelulusan Mahasiswa

Azwar Anas, Ade Jermawinsyah Zebua

Program Studi Manajemen, STIE – Graha Karya Muara Bulian
Jl. Gajah Mada Muara Bulian-Indonesia
Email: azwarzayn@gmail.com¹, adejermawinsyahzebua9@gmail.com²

Abstract

Data transactions that occur every day will become chunks of data in an institution's database. These chunks of data have no value if mining is not carried out in obtaining interesting knowledge and information. The purpose of this study was to analyze chunks of graduation data for STIE-Graha Karya Muara Bulian students, in order to obtain patterns formed from predetermined variables. The pattern formed will later become a reference for internal institutions to see various factors that influence each other from students. The method used is the Apriori Algorithm. This is in accordance with the main function of this algorithm, which is to analyze the frequent itemset. The results showed that for the 2-itemset data the highest support value was the combination of Female Gender and Work with a value of 34%, while the highest confidence value was in the combination of a Grade Point Average above 3.5 and Work with a score of 85%. As for the 3-itemset, the highest support value is in the combination if the GPA is above 3.5, gender is female and the work status reaches 14% and confidence is 90%.

Keywords: data, mining, apriori, graduation, student.

Abstrak

Transaksi data yang terjadi setiap hari, akan menjadi kumpulan data dalam *database* suatu institusi. Kumpulan data tersebut tidak memiliki nilai apabila tidak dilakukan penambangan dalam mendapatkan pengetahuan dan informasi yang menarik. Tujuan dari penelitian ini adalah untuk menganalisa kumpulan data kelulusan mahasiswa STIE-Graha Karya Muara Bulian, agar didapatkan pola yang terbentuk dari variable yang telah ditetapkan. Pola yang terbentuk tersebut nantinya menjadi acuan bagi internal institusi untuk melihat berbagai faktor yang saling berpengaruh dari mahasiswa. Metode yang digunakan adalah Algoritma Apriori. Hal ini sesuai dengan fungsi utama algoritma ini yaitu menganalisa *frequent itemset*. Hasil penelitian menunjukkan bahwa untuk data 2-itemset nilai *support* tertinggi adalah kombinasi Jenis Kelamin Perempuan dan Kerja dengan nilai 34%, sedangkan nilai *confidence* tertinggi adalah pada kombinasi Indeks Prestasi Kumulatif di atas 3,5 dan Kerja dengan nilai 85%. Sedangkan untuk 3-itemset nilai *support* tertinggi adalah pada kombinasi jika Indeks Prestasi Kumulatif di atas 3,5 berjenis kelamin Perempuan dan status pekerjaan Kerja mencapai 14% dan *confidence* 90%.

Kata kunci: data, penambangan, apriori, kelulusan, mahasiswa.

1. Pendahuluan

Tri Dharma Perguruan Tinggi terdiri dari Pendidikan dan Pengajaran, Penelitian dan Pengabdian Kepada Masyarakat. Setiap mahasiswa yang akan menyelesaikan studinya di perguruan tinggi, wajib menyelesaikan ketiga Tri Dharma Perguruan Tinggi tersebut dengan baik. Indikator penyelesaiannya adalah dibuktikan dengan nilai yang diperoleh. Nilai-nilai yang terakumulasi dari setiap mata kuliah tersebut menjadi tolok ukur seberapa kuat pemahaman mahasiswa terhadap materi perkuliahan yang telah dijalani.

Input dari perguruan tinggi adalah mahasiswa yang mendaftar setiap tahun akademiknya. Sedangkan outputnya adalah mahasiswa yang dinyatakan menyelesaikan studinya ditandai dengan pemberian ijazah kelulusan. Selain dari nilai yang diperoleh, indikator keberhasilan mahasiswa setelah mengikuti perkuliahan adalah mampu mengaplikasikan ilmunya baik dengan menciptakan lapangan kerja maupun bekerja dengan orang lain.

Mahasiswa yang terdaftar pada perguruan tinggi tentu saja meninggalkan catatan data yang begitu besar. Semakin banyak jumlah mahasiswa, baik yang masih aktif maupun yang telah menyelesaikan studinya,

maka semakin besar pula tumpukan data yang ada pada *database* perguruan tinggi tersebut. Identitas yang tersimpan biasanya terdiri dari Nomor Induk Mahasiswa (NIM), nama, tempat dan tanggal lahir, jenis kelamin, alamat, status pernikahan dan data orang tua. Sedangkan data yang tersimpan ketika mahasiswa dinyatakan lulus adalah judul penelitian dan pembimbing.

Kumpulan data yang terakumulasi dari tahun ke tahun, tidak memberikan nilai pengetahuan bagi pemilik data maupun pembaca data tersebut. Yang tampak hanyalah bongkahan data tak berarti sama sekali. Tantangan dari tumpukan data tersebut adalah bagaimana cara memanfaatkannya untuk kemudian mendapatkan pengetahuan. Manfaat dari penambangan data adalah untuk menganalisis bongkahan data dalam suatu basis data untuk kemudian ditampilkan pengetahuan sebagai informasi menarik [1]. Tujuan dari data mining adalah mempermudah membaca informasi tersembunyi dari bongkahan data yang tersimpan dalam basis data [2].

Salah satu algoritma dalam metode penambangan data asosiasi dari sekumpulan data dalam *database* adalah algoritma apriori [3]. Parameter utama dalam pembentukan aturan asosiasi yaitu nilai *support* dan *confidence* [4]. Nilai *Support* merupakan jumlah *item* yang paling sering muncul dari jumlah total transaksi, sedangkan *confidence* adalah nilai keyakinan dari jumlah total transaksi yang muncul [5].

Dari tahun ke tahun, STIE-Graha Karya Muara Bulian meluluskan mahasiswanya. Data kelulusan tersebut tidak pernah dianalisa, padahal ada informasi berharga yang bias diambil dari sekumpulan data tersebut. Penelitian ini bertujuan untuk menghasilkan pola kelulusan mahasiswa, adapun data yang dianalisa seperti, Jenis Kelamin, Indeks Prestasi Kumulatif (IPK), lama studi, kelas (A & B) serta status Pekerjaan setelah kelulusan. Objek penelitian dilakukan di Sekolah Tinggi Ilmu Ekonomi-Graha Karya Muara Bulian, Provinsi Jambi, Indonesia. Data yang dianalisis berjumlah 525 mahasiswa yang telah menyelesaikan studinya.

Algoritma apriori telah banyak diaplikasikan oleh peneliti sebelumnya untuk melihat pola hubungan antar *item* pada transaksi [6]. Sebagaimana dilakukan oleh [7]. Hasil penelitian ini menunjukkan minat yang semakin meningkat dari para peneliti di bidang melindungi data sensitif dan pola yang ditambang dari pihak-pihak yang tidak bertanggungjawab. Kemudian [8], pada awalnya penelitian dilakukan dengan menggunakan himpunan fuzzy. Namun, karena kelemahan fuzzy, penulis lebih memilih menggunakan set neutrosodik. Hasil eksperimen menunjukkan bahwa pendekatan yang diusulkan menghasilkan peningkatan jumlah aturan. Selain itu, fungsi indeterminacy-membership telah digunakan untuk mencegah hilangnya aturan dari batasan masalah. Model yang diusulkan lebih efektif dalam memproses aturan asosiasi. Dengan membandingkannya dengan pendekatan penambangan aturan asosiasi fuzzy, pada akhirnya disimpulkan bahwa model yang diusulkan menghasilkan lebih banyak aturan asosiasi positif dan negatif, sehingga memastikan pembangunan sistem pengambilan keputusan yang nyata dan efisien. Selain itu [9], Pada penelitian ini penulis membanding beberapa algoritma yang digunakan, dan hasilnya menunjukkan algoritma apriori merupakan yang terbaik dalam teknik Market Basket Analisis, karena mudah diimplementasikan. Selain itu penulis juga menyarankan agar lingkup algoritma ini diperluas ke area lain. Lalu [10], dalam makalah ini telah dilakukan eksperimen lebih dari 57.000 eksekusi. Eksperimen dilakukan berdasarkan beberapa algoritma yang tertuang pada bagian pendahuluan makalah ini. Hasil yang diperoleh sangat menarik untuk penelitian masa depan di lapangan. Sementara itu [11] melakukan perbandingan untuk beberapa algoritma yang biasa digunakan dalam pencarian data frekuensi tinggi. Hal ini dilakukan atas dasar banyaknya algoritma yang telah diusulkan oleh peneliti sebelumnya, namun masih memiliki berbagai kekurangan dan kelebihan masing-masing. Rumus yang digunakan dalam algoritma ini [12]:

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

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

2. Metode Penelitian

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



Gambar 1. Kerangka Penelitian

Berdasarkan kerangka pada gambar 1. maka masing-masing langkahnya dapat 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 pola mahasiswa lulusan menggunakan algoritma apriori. Disamping itu juga diuraikan tentang kebutuhan data untuk menggali kaidah asosiasi. Dari hasil analisis tersebut, maka algoritma apriori hendaknya mampu menampilkan pola mahasiswa lulusan tersebut 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 mahasiswa lulusan STIE-GK Muara Bulian yang penulis amati sebanyak 500 data.
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 Weka*. 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 mahasiswa lulusan secara manual.
- b. Setelah dilakukan perhitungan manual, selanjutnya dilakukan perhitungan menggunakan *software data mining Weka* 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.

3. Hasil Penelitian

3.1. Data Mahasiswa Lulusan

Jumlah data yang penulis gunakan dalam penelitian ini sebanyak 500 mahasiswa lulusan STIE-GK Muara Bulian. Variable yang digunakan adalah Nama, Indeks Prestasi Kumulatif (IPK), Jenis Kelamin dan Status Pekerjaan. Variable tersebut dipadukan untuk dilihat aturan asosiasinya berdasarkan nilai *support* dan *confidence* yang memenuhi *threshold*.

Tabel 1. Data Mahasiswa Lulusan

MHS	IPK > 2,5	IPK > 3	IPK > 3,5	LK	PR	KERJA
1	Y	N	N	Y	N	N
2	Y	N	N	N	Y	N
3	N	Y	N	N	Y	Y
4	N	N	Y	Y	N	N
5	N	N	Y	Y	N	N
6	Y	N	N	Y	N	Y
7	Y	N	N	Y	N	N
8	Y	N	N	Y	N	N
9	N	Y	N	N	Y	Y
10	N	N	Y	N	Y	Y
500	N	N	Y	Y	N	Y
JML	159	201	140	231	269	331

Keterangan tabel:

- MHS : Mahasiswa
- IPK > 2,5 : IPK di atas 2,5
- IPK > 3 : IPK di atas 3
- IPK > 3,5 : IPK di atas 3,5
- LK : Jenis Kelamin Laki-laki
- PR : Jenis Kelamin Perempuan
- KERJA : Status Pekerjaan
- Y : Ya
- N : Tidak

Berdasarkan tabel di atas, dapat dilihat jumlah mahasiswa dengan IPK>2,5 sebanyak 159 orang, IPK>3 sebanyak 201, IPK>3,5 sebanyak 140, Jenis Kelamin Laki-laki sebanyak 231, Jenis Kelamin Perempuan sebanyak 269 dan Yang Bekerja sebanyak 331 orang.

3.2. Analisis Algoritma Apriori

Berdasarkan tabel 1, berikut disajikan nilai frekuensi data, *support* dan *confidence*.

Tabel 2. Nilai Frekuensi data, *support* dan *confidence*

No.	Variabel	Frekuensi	Support	Confidence
1	IPK>2,5	159	32%	100%
2	IPK>3	201	40%	100%
3	IPK>3,5	140	28%	100%
4	LAKI-LAKI	231	46%	100%
5	PEREMPUAN	269	54%	100%
6	KERJA	331	66%	100%

Selanjutnya adalah penghitungan nilai frekuensi, *support* dan *confidence* untuk 2-itemset sebagaimana tabel 3.

Tabel 3. Nilai Frekuensi, *support* dan *confidence* 2-itemset

No.	2-itemset	Frekuensi	Support	Confidence
1	IPK>2,5 → Laki-laki	97	19%	61%
2	IPK>2,5 → Perempuan	62	12%	39%
3	IPK>2,5 → Kerja	101	20%	64%
4	IPK>3 → Laki-laki	73	15%	36%
5	IPK>3 → Perempuan	128	26%	64%
6	IPK>3 → Kerja	111	22%	55%
7	IPK>3,5 → Laki-laki	61	12%	44%
8	IPK>3,5 → Perempuan	79	16%	56%
9	IPK>3,5 → Kerja	119	24%	85%
10	Laki-laki → Kerja	159	32%	69%
11	Perempuan → Kerja	172	34%	64%

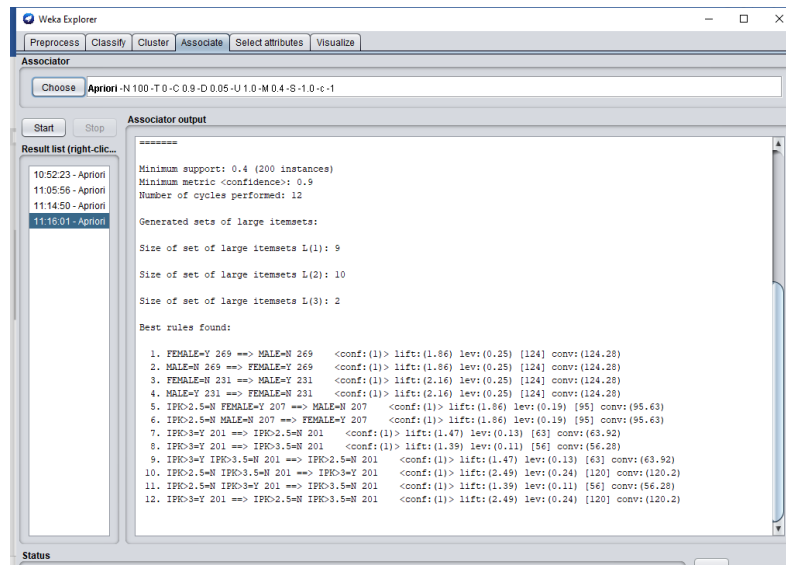
Berdasarkan tabel 3, dapat dilihat bahwa nilai *support* tertinggi adalah kombinasi Jenis Kelamin Perempuan dan Kerja dengan nilai 34%, sedangkan nilai *confidence* tertinggi adalah pada kombinasi IPK>3,5 dan Kerja dengan nilai 85%.

Tabel 4. Nilai Frekuensi, *support* dan *confidence* 3-itemset

No.	3-itemset	Frek	Support	Conf
1	IPK>2,5 → Laki-laki → Kerja	64	13%	66%
2	IPK>2,5 → Perempuan → Kerja	37	7%	60%
3	IPK>3 → Laki-laki → Kerja	47	9%	64%
4	IPK>3 → Perempuan → Kerja	64	13%	50%
5	IPK>3,5 → Laki-laki → Kerja	48	10%	79%
6	IPK>3,5 → Perempuan → Kerja	71	14%	90%

Berdasarkan tabel di atas, nilai *support* tertinggi adalah pada kombinasi IPK>3,5 → Perempuan → Kerja mencapai 14% dan *confidence* 90%.

3. Tampilan hasil Association Rule Terbaik



Gambar 4. Hasil Rule Association Terbaik

Berdasarkan gambar 4, ada 12 rule terbaik yang dihasilkan. Pola yang terbentuk pada kelulusan mahasiswa untuk kombinasi 2-itemset terbaik antara jenis kelamin dan pekerjaan adalah mahasiswa dengan jenis kelamin Perempuan dan telah bekerja dengan nilai *support* 34%. Sedangkan pola yang terbentuk kelulusan mahasiswa untuk kombinasi 3-itemset terbaik adalah jika $IPK > 3,5$ dan jenis kelamin Perempuan maka telah memiliki Pekerjaan dengan nilai *support* 14% dan *confidence* 90%.

4. Kesimpulan

Pada penelitian ini bobot masing-masing kriteria dihitung menggunakan Fuzzy AHP. Setelah itu perankingan dilakukan dengan menggunakan TOPSIS. Bobot preferensi masing-masing kriteria yang dihasilkan oleh Fuzzy AHP adalah sebesar 0.453718 untuk harga, 0.01616 untuk kriteria jumlah teknisi profesional, 0.144174 untuk kriteria metodologi pelaksanaan pekerjaan dan 0.385948 untuk kriteria pengalaman perusahaan. Perankingan yang dihasilkan TOPSIS dengan menggunakan bobot dari perhitungan Fuzzy AHP adalah PT. Nusa Network Prakarsa pada urutan pertama, PT. Mitra Integrasi Informatika pada urutan kedua dan PT. Iproteksi Telematika pada urutan ketiga. Hal ini dapat disebabkan perbedaan sedikit nilai bobot saja dapat mempengaruhi perhitungan pada TOPSIS.

Daftar Pustaka

- [1] M. Muhdi, N. S. Kastawi, And S. Widodo, “Teknik Pengambilan Keputusan Dalam Menentukan Model Manajemen Pendidikan Menengah,” *Kelola J. Manaj. Pendidik.*, Vol. 4, No. 2, P. 135, 2017, Doi: 10.24246/J.Jk.2017.V4.I2.P135-145.
- [2] K. Nisa And E. Sutinah, “Profile Matching Untuk Sistem Pendukung Keputusan Pemilihan Vendor Maintenance Server Dan Jaringan,” *J. Inform.*, Vol. 5, No. 2, Pp. 262–269, 2018, Doi: 10.31311/Ji.V5i2.3638.
- [3] M. Fajri, R. R. M. Putri, And L. Muflikhah, “Implementasi Metode Fuzzy Analytic Hierarchy Process (F-Ahp) Dalam Penentuan Peminatan Di Man 2 Kota Serang,” *J. Pengemb. Teknol. Inf. Dan Ilmu Komput.*, Vol. 2, No. 5, Pp. 2109–2117, 2018.
- [4] D. . Chang, “Application Of The Extent Analysis Method On Fuzzy Ahp,” *Eur. J. Oper. Res.* 95, Pp. 649–655, 1996.
- [5] N. C. Fitriana And B. Santosa, “Analisis Faktor-Faktor Pemilihan Suplier Material Pada Jasa Usaha Konstruksi Dengan Metode Fuzzy Ahp,” *J. Fondasi*, Vol. 9, No. 1, Pp. 1–11, 2020, Doi: 10.36055/Jft.V9i1.7440.
- [6] A. Rahmayani And M. Irawan, “Perancangan Dan Implementasi Perangkat Lunak Sistem Pendukung Keputusan Multi Kriteria Menggunakan Metode Topsis,” *J. Sains Dan Seni Its*, Vol.

- 5, No. 2, 2016, Doi: 10.12962/J23373520.V5i2.17163.
- [7] C. Surya, "Penilaian Kinerja Dosen Menggunakan Metode Topsis (Studi Kasus : Amik Mitra Gama)," *J. Resti (Rekayasa Sist. Dan Teknol. Informasi)*, Vol. 2, No. 1, Pp. 322–329, 2018, Doi: 10.29207/Resti.V2i1.119.
- [8] N. M. Rahadianto, A. A. Rumanti, And R. P. Soesanto, "Perancangan Sistem Pendukung Keputusan Penentuan Lokasi Industri Kuliner Dengan Metode Fuzzy Ahp-Topsis Design Of Decision Support System For Determining The Location Of Culinary Industry By Fuzzy Ahp-Topsis Method," Vol. 8, No. 5, Pp. 8246–8258, 2021.
- [9] S. Sutrisno, A. Susano, And A. Darmawan, "Perancangan Sistem Penentuan Objek Wisata Pantai Terbaik Di Bali Menggunakan Metode Fuzzy Ahp (F- Ahp)," *Simetris J. Tek. Mesin, Elektro Dan Ilmu Komput.*, Vol. 9, No. 1, Pp. 99–108, 2018, Doi: 10.24176/Simet.V9i1.1884.
- [10] M. A. Imawan, M. Cahyanti, M. W. Sardjono, And E. R. Swedia, "Aplikasi Perekrutan Karyawan Menggunakan Metode Topsis Berbasis Web Pada Pt. Smesco Indonesia," *Sebatik*, Vol. 23, No. 2, Pp. 343–351, 2019, Doi: 10.46984/Sebatik.V23i2.781.