



Aplikasi Data Mining Untuk Menentukan Masa Studi Mahasiswa Menggunakan Metode Association Rule

Data Mining Application to Determine Student Study Period Using Association Rule Method



Alwendi^a

Andi Saputa Mandopa^b

Erwina Azizah Hasibuan^c

Article history:

Submitted: 23 August 2023

Revised: 15 September 2023

Accepted: 29 September 2023

Keywords:

Data mining applications,
Association Rules

Abstract

The development of data mining cannot be separated from technological developments, especially in determining association rules to determine the relationship between elements/attributes in the process of students receiving information at any point in time. When it comes to permissions, there are many factors that influence permissions. However, not all of these factors are considered important so research is carried out to determine the relationship between these factors. The topic research method is data collection and data analysis. By using the Association Rule technique we can obtain information from large data such as data from universities. The aim of this research is to determine the pattern of study duration for Graha Nusantara University F-KIP students. using the Association Rule Data Mining method and comparing a priori algorithms and hash-based algorithms. The data used is Graha Nusantara University F-KIP master data which is processed using the Association Rule Data Mining technique with an a priori algorithm and a hash-based algorithm with a minimum support of 1% and a minimum confidence of 1%. The results of data processing with the a priori algorithm are the same as the results of data processing with the hash-based algorithm, namely 49 2-itemset combinations.

Jurnal Pendidikan Dewantara©2023.

This is an open access article under the CC BY-NC-ND license
(<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

Corresponding author:

Alwendi

Program Studi Ilmu Komputer, Universitas Graha Nusantara, Padangsidempuan, Indonesia

Email address : alwendi60@gmail.com

^a Program Studi Ilmu Komputer, Universitas Graha Nusantara, Padangsidempuan, Indonesia

^b Program Studi Pendidikan Matematika, Universitas Graha Nusantara, Padangsidempuan, Indonesia

^c Program Studi Ilmu Komputer, Universitas Graha Nusantara, Padangsidempuan, Indonesia

1. Pendahuluan

Perkembangan data mining tidak lepas dari perkembangan teknologi informasi khususnya dalam penentuan aturan asosiasi untuk mengetahui keterkaitan factor/atribut dalam kelulusan mahasiswa tepat waktu. Dalam mencapai sebuah kelulusan, ada banyak factor yang mempengaruhi kelulusan mahasiswa. Namun tidak semua factor tersebut dianggap penting, untuk itu dilakukan penelitian untuk mengetahui keterkaitan antar factor tersebut. Adapun metode dalam penelitian ini yaitu pengumpulan data, analisa data dan alur penelitian yang menggunakan Rule Asosiasi. Data mining adalah suatu proses menemukan hubungan, pola, dan tren yang bermakna dengan memeriksa sekumpulan besar data yang disimpan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika. belajar menciptakan informasi untuk kebutuhan yang sangat penting bagi manusia, untuk mendukung semua aktivitas yang dilakukan. di luar. Terkadang kebutuhan informasi yang sangat tinggi tidak diimbangi dengan penyediaan informasi yang memadai. Oleh karena itu, informasi ini harus ditambang kembali menjadi big data. Data mining adalah metode yang umum digunakan untuk mencari informasi dalam big data. Dalam data mining terdapat beberapa teknik, termasuk aturan asosiasi. [1].

Dalam ilmu data mining, ada sebuah teknik yang dapat digunakan untuk menyelesaikan masalah ini. Teknik mining yang cocok adalah teknik asosiasi dimana teknik ini mencari kombinasi yang cocok sehingga dapat dijadikan aturan untuk menyelesaikan masalah tertentu [2]. Dengan menggunakan teknik association rule kita dapat memperoleh informasi dari data yang besar seperti data mahasiswa yang ada di perguruan tinggi. Setiap perguruan tinggi selalu berusaha meningkatkan kualitas yang dimiliki agar dapat bersaing dengan perguruan tinggi lainnya. Salah satu cara yang dilakukan perguruan tinggi untuk meningkatkan kualitas yang dimiliki yaitu melakukan evaluasi, dengan evaluasi tersebut perguruan tinggi dapat memperbaiki kekurangan-kekurangan yang sebelumnya terjadi. Salah satu bahan evaluasi adalah mengenai lama studi mahasiswa. Universitas Graha Nusantara memiliki banyak mahasiswa yang terbagi kedalam beberapa fakultas. F-KIP merupakan salah satu fakultas yang terdapat di universitas graha nusantara dan telah menghasilkan banyak lulusan dengan lama studi yang bervariasi. Oleh karena itu, dengan memanfaatkan data induk mahasiswa dan data kelulusan mahasiswa, dapat diketahui informasi pola lama studi mahasiswa melalui teknik data mining.

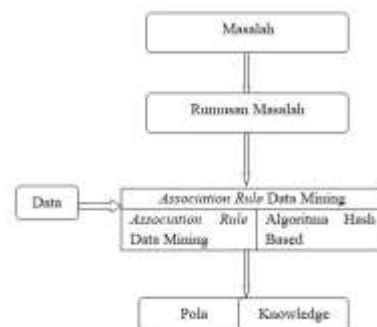
2. Metodologi Penelitian

Pada penelitian ini akan digunakan teknik associaion rule data minnig dengan dua algoritma yaitu algoritma apriori dan algoritma hash-based. Data mining merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakut dari berbagai database besar.

1. Kerangka Penelitian

Langkah-langkah pada association rule adalah sebagai berikut

1. Set - $K = 1$
2. Hitung semua itemset berukuran K
3. Hitung support untuk semua kandidat itemset-menyaring itemset berdasarkan minimum support
4. Gabungkan semua itemset yang berukuran K untuk menghasilkan calon itemset ukuran $K+1$
5. Set - $K = K+1$
6. Ulangi langkah 3 sampai 5 sampai tidak ada itemset yang lebih besar dapat dibentuk.
7. Menghasilkan set akhir itemset dengan menciptakan persatuan semua itemset berukuran K.



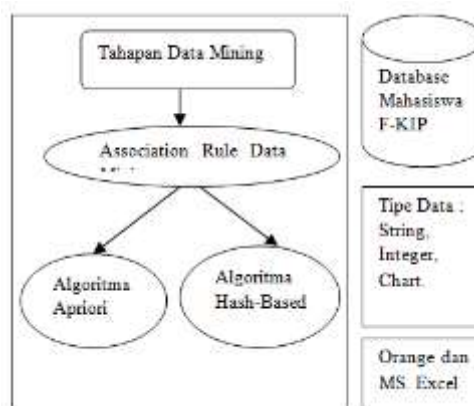
Gambar 1. Kerangka Penelitian

Berdasarkan masalah yang ada, rumusan masalah yang akan di bahas pada penelitian ini yaitu menentukan pola lama studi mahasiswa F-KIP Universitas Graha Nusantara, serta melihat perbandingan 2 algoritma yang digunakan yaitu algoritma apriori dan algoritma hash based. Penelitian ini menggunakan teknik associaion rule data minnig untuk mengelola data dengan menggunakan dua algoritma yaitu algoritma apriori dan algoritma hash-based. Hasil yang akan diperoleh dari pengolahan data ini adalah pola lama studi.

Preprocessing Data Mining dapat meningkatkan kualitas data yang di olah melalui tahap-tahap data cleaning, data integration, data selection, dan data transformation. Hal tersebut dilakukan agar data yang diolah lebih berkualitas artinya data-data tersebut bersifat objektif, representatif, memiliki sampling error yang kecil, terbaharui dan relevan. Persiapan tersebut antara lain:

1. Data Cleaning
Data Cleaning merupakan proses untuk dapat mengatasi nilai yang hilang, noise dan data yang tidak konsisten.
2. Data Integration
Data Integration merupakan proses menggabungkan data dari banyak database. Setelah dilakukan data authentication terdapat data terpisah yaitu data tanggal lahir sehingga didapatkan umur penderita, maka dilakukan proses cleaning kedua dengan mengintegrasikan data mahasiswa F-KIP Universitas Graha Nusantara.
3. Data Selection
Data Selection merupakan proses meminimalkan jumlah data yang digunakan untuk proses mining dengan tetap merepresentasikan data aslinya. Data selection dapat berupa sampling, denoising, dan feature extraction.
4. Data Transformation
Data transformation dilakukan untuk mengubah bentuk dan format data. Hal ini tentunya sangat membantu memudahkan pengguna dalam proses mining ataupun memahami hasil yang didapat Dalam proses data transformation bisa dilakukan dengan centering, normalization, dan scaling.
5. Proses mining
Proses mining merupakan suatu proses yang menggunakan dua algoritma yaitu algoritma apriori dan algoritma hash-based.
6. Evaluasi pola (pattern evaluation)
7. Digunakan untuk mengidentifikasi pola-pola menarik ke dalam knowledge based yang ditemukan.
8. Presentasi pengetahuan (knowledge presentation)
Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

2. Desain Arsitektur Metode Penelitian



Gambar 2. Proses Pengolahan Data Dalam Data Mining

Database mahasiswa F-KIP Universitas Graha Nusantara dibagi menjadi beberapa kategori antara lain , IPK dibagi menjadi $IPK > 3,5$, $3 \leq IPK \leq 3,5$, $2,5 \leq IPK < 3$, $IPK < 2,5$ lama lulus dibagi menjadi lama lulus ≤ 4 tahun, $4 \text{ tahun} < \text{lama lulus} \leq 5 \text{ tahun}$, lama lulus $> 5 \text{ tahun}$, jalur masuk dibagi menjadi jalur masuk SPMB, jalur masuk SUMIKOLA, jalur masuk T2, jalur masuk KEMITRAAN, jenis kelamin dibagi menjadi jenis kelamin laki-laki, jenis

kelamin perempuan, jurusan dibagi menjadi jurusan biologi, jurusan matematika, jurusan kimia, jurusan fisika, jurusan farmasi, asal sekolah dibagi menjadi asal sekolah kabupaten, asal sekolah kecamatan. Pada tahap analisis data, data mahasiswa F-KIP Universitas Graha Nusantara yang didapatkan akan diolah dengan software Orange dan Microsoft Office Excel 2007 menggunakan data mining.

3. HASIL DAN PEMBAHASAN

Algoritma Apriori

Algoritma apriori adalah satu algoritma dasar yang diusulkan oleh Agrawal dan Srikan pada tahun 1994 untuk menemukan frequent itemsets pada aturan asosiasi Boolean. Ide utama pada algoritma apriori adalah : pertama, mencari frequent itemset (himpunan item-item yang memenuhi minimum support.) dari basis data transaksi, kedua – menghilangkan itemset dengan frekuensi yang rendah berdasarkan level minimum support yang telah ditentukan sebelumnya. Selanjutnya membangun aturan asosiasi dari itemset yang memenuhi nilai minimum confidence dalam basis data [4]. Algoritma apriori dibagi menjadi beberapa tahapan yang disebut iterasi. Pada iterasi pertama dihasilkan pola frekuensi tinggi dengan panjang 1-itemset. Pada iterasi pertama pula, nilai support dari setiap item dihitung dengan melihat database yang ada. Setelah didapat nilai support, item yang memiliki support di atas minimum support (minsup) dipilih sebagai acuan untuk iterasi selanjutnya. Iterasi kedua menghasilkan 2-itemset yang tiap setnya memiliki dua item. Tiap kandidat 2-itemset dihitung nilai supportnya dengan melihat database sebelumnya. Support disini dimaksudkan untuk jumlah transaksi dalam database yang mengandung kedua item dalam kandidat 2-itemset. Sama seperti 1-itemset, setelah didapat nilai supportnya, maka kandidat 2-itemset yang memenuhi minsup dapat ditetapkan sebagai 2-itemset yang merupakan pola frekuensi tinggi. Data yang diperoleh dari sistem informasi mahasiswa F-KIP Universitas Graha Nusantara berupa data sekunder berupa nama, jalur masuk Universitas Sam Ratulangi, IPK, lama studi, jenis kelamin, dan jurusan serta wawancara (data primer) berupa asal sekolah diberikan simbol sebagai berikut :

Tabel 1. Analisa data berdasarkan asal sekolah

Simbol	Keterangan	Simbol	Keterangan
IPK	IPK	K	Jalur Masuk Kemitraan
A	$IPK > 3,5$	J.K	Jenis Kelamin
B	$3 \leq IPK \leq 3,5$	L	Jenis kelamin laki-laki
C	$2,5 \leq IPK < 3$	P	Jenis kelamin perempuan
D	$IPK < 2,5$	Jurusan	Jurusan
LL	Lama Lulus	Q	Jurusan biologi
E	Lama lulus ≤ 4 tahun	R	Jurusan Matematika
F	$4 \text{ tahun} < \text{lama lulus} \leq 5 \text{ tahun}$	S	Jurusan Kimia
G	Lama lulus > 5 tahun	T	Jurusan Fisika
JM	Jalur masuk	U	Jurusan Farmasi
H	Jalur masuk SPMB	Sekolah	Asal Sekolah
I	Jalur masuk Sumikola	X	Asal sekolah Kabupaten
J	Jalur masuk T2	Y	Asal sekolah Kecamatan

Dengan menggunakan software Orange data tersebut diolah dan menghasilkan kombinasi itemset sampai kombinasi 6-itemset. Berikut merupakan beberapa kombinasi itemset seperti pada tabel 2 dan tabel 3.

Tabel 2. Beberapa Kombinasi 2- Itemset

2-Itemset	Support	Confidence
J.K=P -> LL=F	25,2%	36,9%
J.K=L -> LL=G	13,8%	43,5%
SEKOLAH=X -> LL=G	21,8%	38,3%
SEKOLAH=Y -> LL=G	16%	37,2%
J. M=J -> LL=G	33%	51,2%
J. M=I -> LL=F	4,2%	53,2%
J. M=H -> LL=F	3,2%	42,2%
J. M=K -> LL=E	18,3%	90,9%

JURUSAN=S -> LL=G	14,3%	44,8%
JURUSAN=Q -> LL=G	13,3%	43,2%
JURUSAN=R -> LL=G	7,5%	38,5%
JURUSAN=U -> LL=F	5%	52,6%
JURUSAN=T -> LL=E	3,8%	46,9%
IPK=C -> LL=G	18%	48,2%
IPK=B -> LL=F	17,5%	43,4%
IPK=D -> LL=G	6,5%	65%
IPK=A -> LL=E	5%	40,5%

Pembentukan Aturan Asosiasi

Untuk pembentukan aturan asosiasi, frequent itemset yang telah didapat akan dibentuk menjadi pola “jika A, maka B ($A \Rightarrow B$)” dan akan dicari nilai confidence masing- masing pola tersebut dengan rumus :

Support dan Confidence dari kombinasi 2-itemset diatas diperoleh menggunakan rumus seperti berikut :

$$\text{Support } (X \cap G) = \frac{\text{Jumlah lulusan mengandung } X \text{ dan } G}{\text{Total Kelulusan}} = \frac{131}{600} = 0,218 = 21,8 \%$$

$$\text{Confidence} = P(G|X) = \frac{\text{Jumlah kelulusan mengandung } X \text{ dan } G}{\text{Jumlah Kelulusan Mengandung } X} = \frac{131}{342} = 0,383 = 38,3 \%$$

Pada rumus yang sama digunakan untuk mencari nilai Support dan confidence dari kombinasi 2- itemset lainnya.

1. Kombinasi 2-itemset berdasarkan asal sekolah
 - a. Jika X (asal sekolah kabupaten) maka G (lama lulus > 5 tahun) dengan nilai support 21,8% dan confidence 38,3% artinya sebanyak 21,8% lulusan yang memiliki asal sekolah kabupaten menempuh studi selama lebih dari 5 tahun dengan nilai confidence 38,3%.
 - b. Jika Y (asal sekolah kecamatan) maka G (lama lulus > 5 tahun) dengan nilai support 16% dan confidence 37,2% artinya sebanyak 16% lulusan yang memiliki asal sekolah kecamatanmenempuh studi selama lebih dari 5 tahun dengan nilaiconfidence 37,2%.
 - c. Berdasarkan pola diatas dapat dilihat bahwa alumni F-KIP Universitas Graha Nusantara yang berasal dari kabupaten dan kecamatan cenderung memiliki lama studi yang sama yaitu lebih dari 5 tahun.
2. Kombinasi 2-itemset berdasarkan Jenis Kelamin
 - a. Jika P (jenis kelamin perempuan) maka F (4 tahun < lama lulus ≤ 5 tahun) dengan nilai support25,2% dan confidence 36,9% artinya sebanyak 25,2% lulusan yang berjenis kelamin perempuan menempuh studi selama lebih dari 4 tahun dengan nilaiconfidence 36,9%.
 - b. Jika L (jenis kelamin laki-laki) maka G (lama lulus > 5 tahun) dengan nilai support 13,8% dan confidence 43,5% artinya sebanyak 13,8% lulusan yang berjenis kelamin laki-laki menempuh studi selama lebih dari 5 tahun dengan nilaiconfidence 43,5%
 - c. Berdasarkan pola diatas dapat dilihat bahwa alumni F-KIP Universitas Graha Nusantara yang berjenis kelamin perempuan cenderung lebih cepat lulus di dibandingkan dengan alumni F-KIP Universitas Graha Nusantara yang berjenis kelamin laki-laki.

KESIMPULAN

Berdasarkan hasil penelitian didapatkan kesimpulan dari Penggunaan Association Rule Data Mining Untuk Menentukan pola lama Studi Mahasiswa.

1. Dengan menggunakan association rule data mining dihasilkan pola lama studi mahasiswa F-KIP Universitas Graha Nusantara. Beberapa pola lama studi mahasiswa pada F-KIP Universitas Graha Nusantara antara lain :
 - a. Kombinasi 2-itemset antara lain sebanyak 18,3% mahasiswa yang masuk F-KIP Universitas Graha Nusantara melalui jalur masuk kemitraan menempuh studi F-KIP Universitas Graha Nusantara selama kurang dari 4 tahun dengan nilai confidence sebesar 90,9%.

- b. Kombinasi 2-itemset antara lain sebanyak 7,5% lulusan yang berasal dari jurusan matematika menempuh studi selama lebih dari 5 tahun dengan nilai confidence 38,5%.
 - c. Kombinasi 2-itemset antara lain sebanyak 5% lulusan yang memiliki IPK lebih dari 3,5 menempuh studi selama kurang dari 4 tahun dengan nilai confidence 40,5%.
 - d. Kombinasi 3-itemset antara lain sebanyak 1,2% lulusan yang berasal dari jurusan matematika dan memiliki IPK lebih dari 3,5 maka menempuh studi selama kurang dari 4 tahun dengan nilai confidence 58,3%.
 - e. Kombinasi 3-itemset antara lain sebanyak 5,2% lulusan yang berasal dari jurusan matematika dan masuk melalui jalur masuk kemitraan maka menempuh studi selama kurang dari 4 tahun dengan nilai confidence 96,9%.
2. Dengan Pembangkitan 2-itemset pada algoritma hash-based menggunakan tabel hash sedangkan pada algoritma apriori menggunakan minimum support. Namun jumlah 2-itemset yang dihasilkan tidak memiliki perbedaan yaitu sebanyak 49 kombinasi 2-itemset.

DAFTAR PUSTAKA

- [1] E. Aribowo dan O. SAD, "Analisis Perbandingan Algoritma Apriori Dan Algoritma Hash Based Pada Market Basket Analysis Di Apotek Universitas Ahmad Dahlan," *Jurnal Sarjana Teknik Informatika volume 3 Nomor 1, Februari 2015*.
- [2] D. Fitriati, "Implementasi Data Mining untuk Menentukan Kombinasi Media Promosi Barang Berdasarkan Perilaku Pembelian Pelanggan Menggunakan Algoritma Apriori," *Annual Research Seminar volume 2 Nomor 1, Desember 2016*.
- [3] W. Weku, " Implementasi Data Mining Untuk Menentukan Studi Mahasiswa Menggunakan Algoritma Apriori," *Jurnal Matematika Dan Aplikasi Volume 3 Nomor 1 Tahun 2014*.
- [4] D. Listriani dan H.Setyaningrum, " Penerapan Metode Asosiasi Menggunakan Algoritma Apriori Pada Aplikasi Analisa Pola Belanja Konsumen," *Jurnal Teknik Informatika Volume 9 Nomor. 2, Oktober 2016*.
- [5] T. Pradana, " Penggalan Kaidah Multilevel Association Rule Dari Data Mart Swalayan ASGAP," *Jurnal SPIRIT Volume. 7 Nomor. 2 Nopember 2015*.
- [6] Alwendi, A., & Aldo, D. (2020). Sistem Pendukung Keputusan Pemilihan Toko Handphone Terbaik Di Kota Padangsidempuan Menggunakan Metode Oreste. *JURSIMA (Jurnal Sistem Informasi dan Manajemen)*, 8(1), 10-17.
- [7] Alwendi, A., & Samosir, K. (2022). Perancangan Aplikasi Sistem Pakar dalam Mendiagnosa Penyakit Infertilitas pada Pria Menggunakan Metode Certainty Factor Berbasis Web. *J. Inform. dan Rekayasa Perangkat Lunak*, 4(1), 24.
- [8] Dengen, C. N., Kusriani, K., & Luthfi, E. T. (2019). Penentuan Association rule pada kelulusan mahasiswa menggunakan algoritma apriori. *J. Rekayasa Teknol. Inf*, 3(1), 20-29.