

Analisis Asosiasi Nilai Mahasiswa dengan Algoritma Apriori

Association Analysis of Student Grades Using the Apriori Algorithm

Ilma Khoiruzzidan¹, Lizda Iswari²

^{1,2}Jurusan Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia, Yogyakarta, Indonesia

¹18523245@students.uui.ac.id, ²lizda.iswari@uui.ac.id

Abstract

This study focuses on applying the Apriori algorithm to course grades of 150 students. The research aims to uncover correlations between course grades, providing insights for both the researcher and the institution. Data was directly extracted from the student academic database. The Apriori algorithm was implemented in three scenarios with varying minimum support and confidence thresholds. The final results depict patterns in course grades that tend to co-occur. Among the three scenarios, the combination of Pengembangan Gim with an A grade, Fundamen Informatika with an A grade, and Matematika Lanjut with an A grade yielded the highest values, with a support of 26% and confidence of 88.6%.

Keywords: data mining; association rules; apriori; student grades

Abstrak

Pada penelitian ini penulis berfokus pada menerapkan algoritma apriori terhadap nilai mata kuliah pada 150 nilai mata kuliah mahasiswa. Tujuan penelitian ini adalah untuk menemukan korelasi antara nilai mata kuliah dan memberikan wawasan bagi penulis sekaligus bermanfaat bagi institusi. Pengambilan data dilakukan secara langsung pada database akademik mahasiswa. Penerapan algoritma apriori dilakukan dalam tiga skenario dengan batas nilai minimum *support* dan *confidence* yang berbeda. Hasil akhir dari penelitian ini menggambarkan pola dari nilai mata kuliah yang cenderung muncul secara bersamaan. Dari ketiga skenario, diperoleh hasil kombinasi Pengembangan Gim dengan nilai A, Fundamen Informatika dengan nilai A dan Fundamen Matematika dengan nilai A yang memiliki nilai tertinggi dengan *support* 26% dan *confidence* 88.6%.

Kata kunci: data mining; analisis asosiasi; apriori; nilai mahasiswa

1. Pendahuluan

Perkembangan teknologi informasi saat ini menjalar kepada sektor yang lebih luas dalam fokus bidang pengolahan data. Sumber data yang begitu banyak menyebabkan kebutuhan pengolahan data menjadi sangat penting bagi sebuah institusi. Data yang begitu banyak harus dikoleksi dan dikumpulkan untuk kemudian diproses menjadi sebuah informasi yang dibutuhkan oleh institusi tersebut.

Menurut Puspaningrum [1], data adalah sumber informasi yang bentuknya masih mentah. Data-data ini berisi landasan faktual yang dapat menjelaskan bagaimana suatu hal dapat terjadi [2]. Dengan terus berkembangnya teknologi, banyak sekali data pada bidang pendidikan yang dapat dianalisis, salah satunya adalah data akademik mahasiswa. Data akademik mahasiswa mencakup informasi tentang nilai mahasiswa, presensi atau daftar hadir, masa studi, serta juga biodata mahasiswa. Pada penelitian ini, penulis memutuskan untuk memfokuskan penelitian pada data

nilai mahasiswa dikarenakan data nilai mahasiswa itu sendiri mencakup informasi mengenai pencapaian akademis masing-masing individu mahasiswa, sekaligus mencerminkan pemahaman dan kemampuan mahasiswa pada setiap mata kuliah yang diambilnya. Dengan demikian eksplorasi asosiasi dan korelasi antara mata kuliah, secara lebih khusus, lebih dimungkinkan.

Data nilai mahasiswa ini menyimpan potensi informasi yang dapat digunakan oleh institusi. Salah satu cara untuk mengolah data tersebut agar kita dapat mengambil kemanfaatannya yaitu dengan cara melakukan *data mining*. Dengan mengolah data secara tepat, para mahasiswa dan tenaga pendidik dapat memanfaatkannya untuk memperoleh *insight* sebagai evaluasi dalam kegiatan ajar-mengajar [3]. *Data mining* membantu kita menganalisis, memperkirakan tren, dan mengidentifikasi pola-pola yang tersembunyi dari data yang dianalisis dimana data tersebut sangat berguna dalam mendukung pengambilan keputusan

penting untuk menunjang aktivitas dan pengembangan pendidikan [4]. *Data mining* adalah suatu istilah yang digunakan untuk mengekstraksi informasi dari *dataset*. Salah satu teknik yang bisa digunakan untuk melakukan *data mining* yaitu analisis asosiasi.

Analisis asosiasi memiliki tujuan menemukan pola hubungan antar beberapa *item* dalam suatu *dataset* [5]. Analisis asosiasi lebih dikenal sebagai *market based analysis* yaitu analisis untuk mengidentifikasi *items* yang sering dibeli secara bersamaan oleh pelanggan dalam suatu transaksi [6].

Algoritma yang digunakan dalam analisis asosiasi untuk membentuk aturan hubungan antar *item* adalah algoritma apriori. Algoritma apriori merupakan suatu algoritma dasar yang diusulkan oleh Agrawal dan Srikant pada tahun 1994 dalam penentuan *frequent itemsets* untuk membuat aturan asosiasi yang bersifat *boolean* [7]. Algoritma apriori adalah sebuah algoritma yang digunakan untuk menemukan *item* dan *itemset* yang relevan dengan tujuan untuk menentukan aturan asosiasi.

Proses melakukan pencarian *frequent itemsets* terdapat dua langkah utama yang dilakukan oleh algoritma apriori yaitu, *join* dan *prune*.

1. Join

Dalam langkah *join*, *frequent (k-1)-itemsets* yang mengandung *k-1 item* untuk menghasilkan *k-itemsets* yang mengandung *k-item*. Contohnya, jika terdapat *frequent itemsets* {A, B} dan {A, C}, maka *k-itemsets* yang dihasilkan adalah {A, B, C}.

2. Prune

Pada langkah *prune*, algoritma apriori akan mengeliminasi *k-itemsets* yang tidak sering muncul bersama dalam *dataset* transaksi. Proses *prune* dilakukan dengan cara memeriksa apakah semua *subset* dari *k-itemsets* tersebut termasuk dalam *frequent itemsets* atau tidak. Jika tidak, maka *k-itemsets* tersebut dieliminasi [8]. Kedua langkah tersebut dilakukan secara berulang dan algoritma apriori akan menghasilkan *frequent itemsets* yang sering muncul bersama dalam *dataset* transaksi [8].

Dalam konteks *data mining*, algoritma apriori termasuk dalam kategori teknik analisis asosiasi yang digunakan untuk menemukan sebuah pola dari kombinasi tiap *item*. Dengan teknik ini, peneliti dapat menghitung seberapa sering kombinasi *item* muncul dan kemudian menggunakan informasi ini untuk tujuan prediksi [9].

Algoritma apriori telah menjadi pilihan yang efektif untuk melakukan analisis asosiasi, terutama bila terdapat beberapa hubungan *item* yang ingin dianalisa. Seperti yang diterapkan pada penelitian yang

dilakukan oleh Anis Kurniawati [10] dalam bidang promosi dan penentuan strategi pemasaran.

Penerapan analisis asosiasi pada konteks pendidikan tinggi juga telah dilakukan oleh Xiaodong Wu dan Yuzhu Zeng [8] yang menggunakan analisis asosiasi pada 100 data nilai mahasiswa dengan 34 mata kuliah yang dikonversi menjadi nilai huruf dengan tujuan untuk mengidentifikasi korelasi antar mata kuliah yang berbeda. Penting atau tidaknya suatu aturan asosiasi dapat diketahui dengan beberapa parameter yaitu *support*, *confidence* dan *lift*.

1. Support

Support adalah persentase dari jumlah kombinasi *item* dalam dataset.

$$Support(A) = \frac{Jumlah\ transaksi\ mengandung\ A}{Total\ transaksi} \cdot 100\% \quad (1)$$

Persamaan (1) merupakan rumus untuk menghitung nilai *support* dari satu item A.

$$Sup(A, B) = \frac{Jumlah\ transaksi\ mengandung\ A\ dan\ B}{Total\ transaksi} \cdot 100\% \quad (2)$$

Persamaan (2) rumus menunjukkan perhitungan untuk nilai *support* dari *itemset* {A, B}.

2. Confidence

Confidence mengindikasikan nilai dari seberapa besar kuatnya hubungan antar *item* dalam aturan asosiasi [11].

$$Conf(A, B) = \frac{Jumlah\ transaksi\ mengandung\ A\ dan\ B}{Jumlah\ transaksi\ mengandung\ A} \quad (3)$$

Dengan rumus persentasenya sebagai berikut:

$$Confidence(A \rightarrow B) = \frac{Support(A, B)}{Support(A)} \cdot 100\% \quad (4)$$

3. Lift

Lift Ratio merupakan parameter yang digunakan untuk menilai kekuatan dari aturan asosiasi yang terbentuk berdasarkan nilai *support* dan *confidence*. Pada umumnya, nilai *lift ratio* digunakan sebagai indikator untuk menentukan validitas suatu aturan asosiasi. *Antecedent* berperan sebagai penyebab terjadinya *item consequent*. Sementara itu, *Consequent* adalah hasil atau *item* yang akan dibeli setelah melakukan pembelian *Antecedent* [11]. *Lift* adalah menunjukkan nilai dari kuatnya suatu hubungan antar *item* dalam aturan asosiasi. Nilai *lift* bisa didapatkan dari rumus berikut. Rentang nilai *lift* mulai dari 0 sampai tak

hingga (∞). Nilai *Lift ratio* didapatkan dari Persamaan (5).

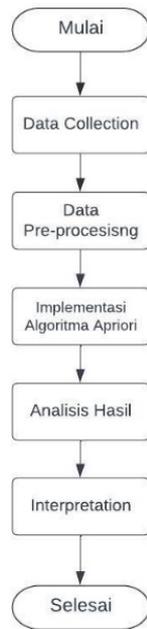
$$Lift(A, B) = \frac{Confidence(A, B)}{Benchmark\ Confidence(A, B)} \quad (5)$$

Nilai *benchmark confidence* dapat dihitung menggunakan Persamaan (6).

$$Benchmark\ Conf(A, B) = \frac{Jml\ transaksi\ mengandung\ (B)}{Total\ transaksi} \quad (6)$$

Penelitian ini berfokus untuk mencari nilai asosiasi antar nilai mata kuliah yang dihasilkan dari algoritma apriori. Hasil yang ingin dicapai adalah kemungkinan terdapat asosiasi antara mata kuliah yang dapat menjadi acuan dalam petinggi akademisi khususnya dosen dalam merancang kurikulum, mengevaluasi efektivitas pengajaran dan materi pembelajaran pada mata kuliah yang kurang berhasil, dan menjadi bahan rekomendasi kepada mahasiswa mata kuliah yang harus menjadi perhatian.

2. Metodologi Penelitian



Gambar 1. Diagram Alur Penelitian

Metode penelitian yang digunakan pada penelitian ini adalah *Knowledge Discovery in Database (KDD)*. KDD meliputi beberapa langkah yaitu *data collection*, *data pre-processing*, implementasi algoritma apriori (*data mining*), analisis hasil dan *interpretation* [12].

2.1. Data Collection

Proses pengambilan data dilakukan melalui pengambilan data secara langsung pada *database* akademik mahasiswa melalui perantara dosen pembimbing. Data yang diambil akan diseleksi dan difokuskan pada nilai mahasiswa beserta mata kuliah yang diambilnya. Data yang digunakan adalah data 150 mahasiswa pada 18 mata kuliah wajib yang

bersifat keinformatikaan dengan pengecualian mata kuliah umum seperti kewarganegaraan, agama, dan bahasa. Data-data tersebut didapatkan dari perkuliahan yang diselenggarakan oleh Jurusan Informatika, Universitas Islam Indonesia.

2.2. Data Pre-processing

Data yang telah diambil dan diseleksi kemudian dibersihkan sesuai dengan kebutuhan analisis. Dalam *pre-processing*, transformasi data dilakukan untuk mengubah data transaksional menjadi data *tabular* agar dapat diolah oleh algoritma apriori. Format data transaksional membutuhkan dua *field*, berupa *field ID* dan *field item*, dimana tiap *record* merepresentasikan hanya satu *item*. Berikut contoh format data transaksional sebagaimana ditampilkan pada Tabel 1.

Tabel 1. Data Transaksional

Field ID	Field Item
1	Item 1
1	Item 2
2	Item n
...	...

Sedangkan, pada format data *tabular*, tiap *record* merepresentasikan transaksi secara biner yang berisi nilai 0 dan 1, sebanyak itemnya. Berikut merupakan contoh format data *tabular*.

Tabel 2. Data Tabular

ID Transaksi	Item 1	Item 2	...	Item n
1	1	1	0	0
2	0	0	0	1
3	0	1	0	0
...
...

2.3. Implementasi Algoritma Apriori

Proses awal dalam data mining untuk analisis asosiasi adalah melakukan pencarian *frequent itemsets*. Pada pencarian *frequent itemsets* perlu untuk melakukan penetapan batas *minimum support*. Penetapan ini didasarkan pada beberapa kondisi, yaitu:

1. Ukuran Dataset

Jika *dataset* yang digunakan besar, maka batas nilai *minimum support* diatur lebih tinggi agar banyak *item* yang ter-filter untuk mencegah jumlah *frequent itemsets* yang mungkin terlalu banyak dan membebankan komputasi. Sebaliknya, jika *dataset* yang digunakan kecil, maka nilai *minimum support* cenderung diatur rendah karena kemungkinan *frequent itemsets* yang muncul tidak terlalu banyak.

2. Tujuan Analisis Asosiasi

Jika menggunakan batas *minimum support* tinggi, maka banyak *itemset* yang ter-filter sehingga kemungkinan kemunculan *itemset* yang umum lebih

banyak. Sebaliknya, jika menggunakan batas *minimum support* rendah, maka *itemset* yang tidak umum juga muncul. Maksud dari *itemset* yang umum disini adalah *item* yang sering muncul pada *frequent itemsets* dengan nilai *support* yang cukup tinggi.

Setelah menentukan batas-batas yang diperlukan, kemudian dilakukan pencarian *frequent itemsets*. *Frequent itemsets* akan dihasilkan dengan batas *minimum support* yang dibatasi oleh peneliti sendiri. Pada tahap selanjutnya, dilakukan penerapan algoritma apriori terhadap data yang akan menghasilkan *association rules* dari *frequent itemsets* yang diseleksi berdasarkan batas *minimum lift* sama dengan 1. Rentang nilai *lift* mulai dari 0 sampai tak hingga. Pertimbangan batas nilai *lift* disini berdasarkan teori yaitu, nilai *lift* dengan *minimum* 1 sudah mengindikasikan terdapat kemanfaatan pada rules tersebut.

2.4. Analisis Hasil

Pada tahap ini disajikan hasil yang ditemukan dari penerapan algoritma apriori yang menggambarkan bagaimana korelasi antara mata kuliah yang berbeda.

2.5. Interpretation

Pada tahap ini, peneliti akan menghubungkan hasil analisis dengan teori yang telah ada sebelumnya dan menjelaskannya secara detail.

3. Hasil dan Pembahasan

3.1. Data Selection

Hal yang dilakukan pertama kali pada proses analisis asosiasi adalah menyeleksi data dari *database* kemudian melakukan *pre-processing*. Bisa dilihat pada Tabel 3, bentuk data awal nilai mahasiswa masih berupa huruf atau kategorikal.

Dengan demikian, harus dilakukan *pre-processing*. Persiapan untuk melakukan transformasi data disini adalah membuat pengkodean nilai mata kuliah dan mendefinisikasi bobot penilaian tiap mata kuliah.

Tabel 3. Data Mahasiswa Terseleksi

No	Algoritma Struktur Data	Fundamen Informatika	Fundamen Matematika	Rekayasa Perangkat Lunak
1.	B+	A/B	A-	A/B
2.	B+	A-	A/B	A/B
3.	A/B	A	A/B	A
4.	A	A	A	A-
5.	B+	A-	A	A/B

3.2. Data Pre-processing

Setelah melakukan persiapan data, tranformasi data dilakukan untuk mengubah data kategorikal ke bentuk tabular dimana setiap baris hanya mewakili nilai *boolean* yaitu 1 dan 0. Hal ini dilakukan karena

algoritma apriori hanya bisa memproses data bentuk *boolean*. Transformasi data dilakukan menggunakan teknik *hot-encoding*.

Hot-encoding dilakukan menggunakan fitur pivot pada software Google Sheets. *Hot-encoding* adalah teknik untuk mengubah data kategorikal menjadi data numerik dengan cara membuat variabel biner baru untuk setiap kategori. variabel biner tersebut bernilai 1 jika data termasuk dalam kategori tersebut, dan bernilai 0 jika tidak. Hasil transformasi data ke bentuk data tabular tertera pada Tabel 4.

Tabel 4. Hasil Proses *Hot-encoding*

No.	A.1ASD	A-.1ASD	A/B.1ASD	B+.1ASD
1.	1	0	0	0
2.	0	0	1	0
3.	0	0	0	1
4.	1	0	0	0
5.	0	1	0	0

Setiap nama mata kuliah yang membentang pada *heading row* pada data Tabel 4 diubah dalam bentuk format *item* yang ada pada Gambar 2.

{A.1ASD}

Gambar 2. Format *Item*

Pada Gambar 2, huruf "A" sebelum titik mengindikasikan nilai yang diperoleh pada mata kuliah tersebut, sementara "1ASD" menyatakan bahwa "ASD" adalah singkatan dari nama mata kuliah "Algoritma Struktur Data", dan angka "1" mewakili semester di mana mata kuliah Algoritma Struktur Data tersebut diambil dari referensi pengkodean nama mata kuliah yang terdapat pada Tabel 5.

Tabel 5. Pengkodean Mata Kuliah

Nama Mata Kuliah	Pengkodean
Pemikiran Desain	PEMDES
Fundamen Matematika	FUNMAT
Fundamen Informatika	FUNDIF
Logika Pemrograman	LOGPRO
Matematika Lanjut	MATLAN
Fundamen Pengembangan Aplikasi	FPA
Algoritme dan Struktur Data	ASD
Rekayasa Perangkat Lunak	RPL
Sistem dan Jaringan Komputer	SJK
Pengembangan Aplikasi Berbasis Web	PABW
Manajemen Pengembangan Teknologi Informasi	MPTI
Pengembangan Sistem Informasi	PSI
Sistem Cerdas dan Pendukung Keputusan	SCPK
Fundamen Sains Data	FSD
Grafika dan Multimedia	GRAFMUL
Pengembangan Aplikasi Bergerak	PAB
Pengembangan Aplikasi Informatika Medis	PAIM
Pengembangan Gim	GIM

3.3. Data Mining

Berdasarkan data yang telah disiapkan, penerapan algoritma apriori dimulai dari meng-*import* data dari penyimpanan lokal menuju *software* Google Colab. Penelitian ini menggunakan 150 data mahasiswa pada 18 mata kuliah wajib yang diambil dari *database* akademik mahasiswa pada bulan Agustus 2023. Implementasi algoritma apriori diawali dari mencari *frequent itemsets* yang mungkin menjadi kandidat *itemset*.

```
frequent_itemsets = apriori(df2, min_support=0.15, use_colnames=True)
frequent_itemsets
```

Gambar 3. Generate Frequent Itemsets

	support	itemsets
0	0.153333	(A/B.1ASD)
1	0.273333	(B+.1ASD)
2	0.533333	(A.1FUNDIF)
3	0.200000	(A-.1FUNDIF)
4	0.160000	(B.1LOGPRO)
...
130	0.173333	(A.4GIM, A.3MATLAN, A.4GRAFMUL, A.1FUNDIF)
131	0.160000	(A.3MATLAN, A.4GRAFMUL, A.5MPTI, A.1FUNDIF)
132	0.160000	(A.2MATLAN, A.3MATLAN, A.4GIM, A.1PEMDES)
133	0.166667	(A.4GIM, A.3MATLAN, A.4GRAFMUL, A.1PEMDES)
134	0.153333	(A.3MATLAN, A.4GRAFMUL, A.5MPTI, A.1PEMDES)

135 rows x 2 columns

Gambar 4. Hasil Pencarian Frequent Itemsets

Variabel *frequent itemsets* akan menjalankan sebuah *function* untuk melakukan pencarian *frequent itemsets* yang ada pada *data frame*. Penetapan nilai *minimum support* sama dengan 0.15 diambil berdasarkan tujuan dari penelitian ini yaitu, ingin mengeksplorasi

sebanyak-banyaknya *item* yang kemungkinan memiliki pola atau hubungan satu dengan lainnya. Sehingga dengan nilai *support* 0.15, diharapkan banyak kemungkinan *itemsets* yang muncul dibandingkan menggunakan batas nilai yang lebih besar.

Dengan menggunakan batas nilai *support* 0.15 berarti melibatkan sebanyak mungkin *itemsets* dapat membantu peneliti menggali korelasi yang mungkin terlewatkan pada nilai *support* yang lebih tinggi. Dengan demikian diharapkan muncul hubungan-hubungan berpotensi memberikan informasi yang lebih mendalam tentang data yang diteliti.

Frequent itemsets yang muncul sebanyak 135 kandidat dari 1-*itemsets* hingga 4-*itemsets*. Perhitungan *frequent itemsets* hingga panjang 4-*itemsets* didasarkan pada kebutuhan untuk mengeksplorasi sejauh mungkin potensi hubungan atau pola yang ada pada data. Setelah semua *frequent itemsets* ditemukan, barulah dicari *association rules* berdasarkan nilai *support* dan nilai *confidence* yang ditentukan. Pencarian *association rules* pada penelitian ini dijalankan dalam beberapa skenario dengan batas-batas yang berbeda. Tujuannya adalah penulis dapat memastikan bahwa setiap temuannya terdapat korelasi antar *item* yang ada.

Skenario pertama dijalankan pada keseluruhan data dengan batas *minimum lift* sama dengan 1.

Pada Tabel 6 ditemukan sebanyak 478 baris *rules*. Namun, disini *association rules* yang muncul terlalu banyak dan yang dibutuhkan adalah *association rules* dengan nilai *support* dan nilai *confidence* yang tinggi agar *rules* yang diinterpretasikan valid. Maka, dilakukan lagi pencarian *association rules* dengan skenario selanjutnya dengan batas berbeda.

Tabel 6. Hasil Skenario Pertama

id	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
0	{A.1PEMDES}	{B+.1ASD}	0.58	0.273	0.16	0.275	1.00
1	{B+.1ASD}	{A.1PEMDES}	0.273	0.58	0.16	0.585	1.00
2	{A.3MATLAN}	{B+.1ASD}	0.66	0.273	0.20	0.303	1.10
3	{B+.1ASD}	{A.3MATLAN}	0.273	0.66	0.20	0.731	1.10
4	{A.1PEMDES}	{A.1FUNDIF}	0.58	0.53	0.33	0.574	1.07
...
473	{A.4GRAFMUL, A.3MATLAN}	{A.1PEMDES, A.5MPTI}	0.346	0.206	0.153	0.44	2.14
474	{A.5MPTI}	{A.4GRAFMUL, A.1PEMDES, A.3MATLAN}	0.293	0.253	0.153	0.522	2.06
475	{A.1PEMDES}	{A.4GRAFMUL, A.5MPTI, A.3MATLAN}	0.58	0.186	0.153	0.264	1.41
476	{A.3MATLAN}	{A.1PEMDES, A.4GRAFMUL, A.5MPTI}	0.66	0.173	0.153	0.23	1.30
477	{A.4GRAFMUL}	{A.1PEMDES, A.5MPTI, A.3MATLAN}	0.46	0.186	0.153	0.33	1.78

Tabel 7. Hasil Skenario Kedua

id	antecedents	consequents	antecedent support	consequent support	support	confidence
4	{'A.1FUNDIF'}	{'A.1PEMDES'}	0.53	0.58	0.33	0.625
5	{'A.1PEMDES'}	{'A.1FUNDIF'}	0.58	0.53	0.33	0.574
10	{'A.3MATLAN'}	{'A.1FUNDIF'}	0.66	0.53	0.42	0.636
11	{'A.1FUNDIF'}	{'A.3MATLAN'}	0.53	0.66	0.42	0.787
16	{'A.4GRAFMUL'}	{'A.1FUNDIF'}	0.46	0.53	0.31	0.681
17	{'A.1FUNDIF'}	{'A.4GRAFMUL'}	0.53	0.46	0.31	0.587
28	{'A.3MATLAN'}	{'A.1PEMDES'}	0.66	0.58	0.42	0.646
29	{'A.1PEMDES'}	{'A.3MATLAN'}	0.58	0.66	0.42	0.735
32	{'A.4GIM'}	{'A.1PEMDES'}	0.51	0.58	0.36	0.714
33	{'A.1PEMDES'}	{'A.4GIM'}	0.58	0.51	0.36	0.632
34	{'A.4GRAFMUL'}	{'A.1PEMDES'}	0.46	0.58	0.32	0.695

```
result = rules[(rules['support'] >= 0.3) &
              (rules['confidence'] >= 0.5)]
```

Gambar 5. Pencarian Skenario Kedua

Pada pencarian Gambar 5 dilakukan dengan *minimum support* 0.3 dan *minimum confidence* 0.5 menghasilkan beberapa *association rules* yang terdapat pada **Error! Reference source not found.**

Pada **Error! Reference source not found.** bisa diperhatikan bahwa tidak adanya *rules* yang memuat kandidat yang lebih panjang dari 1-*itemsets*. Index 11 menjadi perhatian utama pada skenario kedua ini dikarenakan pada *rules* tersebut menunjukkan nilai *support* 0.42 dan nilai *confidence* 0.78 yang cukup tinggi diantara *rules* yang lain.

```
result2 = rules[(rules['support'] >= 0.2) &
               (rules['confidence'] >= 0.8)]
```

Gambar 6. Pencarian Skenario Ketiga

Karena hasil yang diinginkan adalah menemukan

korelasi item yang tinggi, peneliti melakukan skenario ketiga dengan memprioritaskan nilai *minimum confidence* yang lebih tinggi dan menurunkan sedikit nilai *minimum support*-nya. Nilai *confidence* mengukur seberapa tinggi kemungkinan ketika suatu *item* muncul dengan *item* lainnya yang cenderung muncul secara bersamaan. Hal ini selaras dengan tujuan peneliti yang ingin mencari *item* yang cenderung muncul bersamaan dan menunjukkan adanya korelasi pada *item* tersebut.

Pada pencarian dilakukan dengan *minimum support* 0.2 dan *minimum confidence* 0.8 menghasilkan beberapa *association rules* yang terdapat pada **Error! Reference source not found.** Dari skenario ketiga terlihat banyak korelasi antara *itemset-2* dengan *item* "A.3MATLAN". Pada Index 150 terlihat *rules* tersebut memiliki nilai *support* 0.26 dan nilai *confidence* 0.886. Disini nilai *confidence* pada *rules* index 150 tertinggi dibandingkan dengan *rules* lainnya.

Tabel 8. Hasil Skenario Ketiga

id	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
150	{'A.4GIM', 'A.1FUNDIF'}	{'A.3MATLAN'}	0.29	0.66	0.26	0.886	1.34
162	{'A.5MPTI', 'A.1FUNDIF'}	{'A.3MATLAN'}	0.23	0.66	0.20	0.885	1.34
355	{'A.1PEMDES', 'A.4GIM', 'A.1FUNDIF'}	{'A.3MATLAN'}	0.22	0.66	0.20	0.882	1.33
76	{'A.5MPTI'}	{'A.3MATLAN'}	0.29	0.66	0.25	0.863	1.30
156	{'A.4GRAFMUL', 'A.1FUNDIF'}	{'A.3MATLAN'}	0.31	0.66	0.26	0.829	1.25
102	{'A.1PEMDES', 'A.1FUNDIF'}	{'A.3MATLAN'}	0.33	0.66	0.27	0.820	1.24
160	{'A.5MPTI', 'A.3MATLAN'}	{'A.1FUNDIF'}	0.25	0.53	0.21	0.815	1.52
311	{'A.4GRAFMUL', 'A.3SJK'}	{'A.3MATLAN'}	0.25	0.66	0.21	0.815	1.23
203	{'A.1PEMDES', 'A.2MATLAN'}	{'A.3MATLAN'}	0.24	0.66	0.2	0.810	1.22
233	{'A.1PEMDES', 'A.3SJK'}	{'A.3MATLAN'}	0.28	0.66	0.23	0.800	1.22

3.4. Interpretation

Dari hasil penerapan algoritma apriori sebelumnya dengan beberapa skenario didapatkan beberapa *association rules* pada data nilai mahasiswa, berikut adalah beberapa interpretasi untuk setiap skenario-nya.

Pada skenario pertama, ditemukan banyaknya *rules* yang melewati batas minimal *threshold* nilai *lift* sama dengan 1, yang artinya sebanyak 478 *rules* memiliki hubungan yang lumayan kuat antar *item*-nya.

Pada skenario kedua, pencarian *rules* dilakukan dengan menetapkan batas minimal *threshold* pada nilai *support* sebesar 0.3 dan nilai *confidence* 0.5. Hasilnya, terlihat di **Error! Reference source not found.** pada id ke-11 terdapat *rules* {A.1FUNDIF} dan {A.3MATLAN} dengan nilai *support* 0.42 dan nilai *confidence* 0.787. Dalam 42% data dari dataset, mengindikasikan bahwa kebanyakan mahasiswa yang memperoleh nilai A pada mata kuliah Fundamen Informatika di semester 1 cenderung memperoleh nilai A pada mata kuliah Matematika Lanjut di semester 3 dengan tingkat *confidence* 78.7%. Hal ini kemungkinan disebabkan karena mata kuliah Fundamen Informatika dan Matematika Lanjut saling berkaitan. Dengan membawa pengetahuan tentang materi keinformatikaan dasar pada mata kuliah Fundamen Informatika di semester awal, mahasiswa dapat menempuh mata kuliah Matematika Lanjut dengan baik di semester selanjutnya.

Pada skenario ketiga, pencarian *rules* dilanjutkan dengan dasar tujuan menemukan korelasi yang lebih tinggi sehingga peneliti memutuskan untuk menurunkan batas *threshold* pada nilai *support* sebesar 0.2 dan menaikkan nilai *confidence* sebesar 0.8. Pada **Error! Reference source not found.** hasil skenario diambil dengan mengurutkan nilai *confidence* tertinggi ke terendah. Hasilnya pada id ke-150 *rules* {‘A.4GIM’, ‘A.1FUNDIF’} dan {‘A.3MATLAN’} memperoleh nilai *support* 0.26 dan nilai *confidence* 0.886. Hasil menunjukkan bahwa 26% data dari dataset, sebagian besar mahasiswa yang memperoleh nilai A pada mata kuliah Pengembangan Gim di semester 4 dan mata kuliah Fundamen Informatika di semester 1 cenderung mempertahankan performa mereka dengan memperoleh nilai A pada mata kuliah Matematika Lanjut di semester 3 dengan tingkat *confidence* 88.7%.

Korelasi ketiga mata kuliah ini bisa diartikan sebagai mahasiswa yang telah memiliki nilai baik pada mata kuliah Pengembangan Gim yang melibatkan aspek perhitungan juga pada mata kuliah Fundamen Informatika yang mempelajari dasar-dasar pemrograman cenderung memiliki baik pada mata kuliah Matematika Lanjut. Artinya, terdapat temuan bahwa ketiga mata kuliah ini bisa menjadi perhatian bagi mahasiswa yang ingin mengasah keterampilannya

dalam mengembangkan gim untuk memperhatikan materi-materi yang relevan pada ketiga mata kuliah ini.

4. Kesimpulan

Analisis asosiasi menggunakan algoritma apriori dapat diterapkan untuk mencari korelasi antara nilai mata kuliah mahasiswa pada data 150 mahasiswa yang mengambil mata kuliah wajib. Penerapan algoritma apriori dilakukan dalam 3 skenario yang menggunakan batas nilai *support* dan *confidence* berbeda pada tiap skenario. Dari ketiga skenario ini diambil hasil terbaik yaitu, sebuah *rules* {‘A.4GIM’, ‘A.1FUNDIF’} dan {‘A.3MATLAN’} yang memperoleh nilai *support* 0.26 dan nilai *confidence* 0.886.

Interpretasi temuan ini menunjukkan bahwa adanya korelasi antara ketiga mata kuliah Pengembangan Gim, Fundamen Informatika dan Fundamen Matematika. Harapannya, temuan ini dapat memberikan *insight* bagi mahasiswa yang ingin menekuni bidang pengembangan gim dengan memperhatikan performanya pada ketiga mata kuliah tersebut. Hal ini juga bisa menjadi pertimbangan bagi mahasiswa baru dalam menentukan konsentrasi studi yang sesuai minat dan kemampuan mereka.

Reference

- [1] A. S. Puspaningrum, F. Firdaus, I. Ahmad, and H. Anggono, "Perancangan Alat Deteksi Kebocoran Gas Pada Perangkat Mobile Android Dengan Sensor Mq-2," *J. Teknol. dan Sist. Tertanam*, vol. 1, no. 1, p. 1, 2020, doi: 10.33365/jst.v1i1.714.
- [2] Y. Yuliana, P. Paradise, and K. Kusriani, "Sistem Pakar Diagnosa Penyakit Ispa Menggunakan Metode Naive Bayes Classifier Berbasis Web," *CSRID (Computer Sci. Res. Its Dev. Journal)*, vol. 10, no. 3, p. 127, 2021, doi: 10.22303/csrid.10.3.2018.127-138.
- [3] Z. Hu, "Application and Research of Association rules and Decision tree Combination Algorithm in Student Achievement Analysis [D]," *Place Preserv. Qingdao Univ. Technol.*, 2017.
- [4] B. Santosa, "Data mining teknik pemanfaatan data untuk keperluan bisnis," *Yogyakarta Graha Ilmu*, vol. 978, no. 979, p. 756, 2007.
- [5] Despitaria, H. Sujaini, and Tursina, "Analisis Asosiasi pada Transaksi Obat Menggunakan Data Mining dengan Algoritma A Priori," *J. Sist. Dan Teknol. Inf.*, vol. 1, no. 1, pp. 1–6, 2016.
- [6] F. Nurchalifatun, "Penerapan Metode Asosiasi Data Mining Menggunakan Algoritma Apriori Untuk Mengetahui Kombinasi Antar Itemset Pada Pondok Kopi," *Data Min.*, 2017.
- [7] M. Badrul, P. Studi, and S. Informasi, "Algoritma Asosiasi Dengan Algoritma Apriori Untuk Analisa Data Penjualan," *J. Pilar Nusa Mandiri*, vol. XII, no. ALGORITMA ASOSIASI DENGAN ALGORITMA APRIORI UNTUK ANALISA DATA PENJUALAN, pp. 121–129, 2016.
- [8] X. Wu and Y. Zeng, "Using Apriori Algorithm on Students' Performance Data for Association Rules Mining," vol. 322, no. Iserss, pp. 403–406, 2019, doi: 10.2991/iser-19.2019.300.
- [9] R. Agrawal and R. Srikant, "Mining sequential patterns," in *Proceedings of the eleventh international conference on data engineering*, IEEE, 1995, pp. 3–14.
- [10] A. Kurniawati, "Pemetaan Pola Hubungan Program Studi Dengan Algoritma Apriori – Studi Kasus Spmu Unnes," *Edu*

- Komputika J.*, vol. 1, no. 1, pp. 51–58, 2014.
- [11] A. N. Rahmi and Y. A. Mikola, “Implementasi Algoritma *Apriori* Untuk Menentukan Pola Pembelian Pada Customer (Studi Kasus : Toko Bakoel Sembako),” *Inf. Syst. J.*, vol. 4, no. 1, pp. 14–19, 2021, [Online]. Available: <https://jurnal.amikom.ac.id/index.php/infos/article/view/561>
- [12] N. Kumar, S. Jain, and K. Chauhan, “Knowledge discovery from data *mining* techniques,” *Int. J. Eng. Res. Technol.*, vol. 7, no. 12, pp. 1–3, 2019.