ANALISIS ASOSIASI PADA NILAI MAHASISWA INFORMATIKA UII MENGGUNAKAN DATA MINING DENGAN ALGORITMA APRIORI



Disusun Oleh:

N a m a : Ilma Khoiruzzidan

NIM : 18523245

PROGRAM STUDI INFORMATIKA – PROGRAM SARJANA

FAKULTAS TEKNOLOGI INDUSTRI

UNIVERSITAS ISLAM INDONESIA

2023

HALAMAN PENGESAHAN DOSEN PEMBIMBING

ANALISIS ASOSIASI PADA NILAI MAHASISWA INFORMATIKA UII MENGGUNAKAN DATA MINING DENGAN ALGORITMA APRIORI

TUGAS AKHIR

Disusun Oleh:

N a m a : Ilma Khoiruzzidan

NIM : 18523245

Yogyakarta, 26 Desember 2023

Pembimbing,

(Lizda Iswari, S.T, M.Sc)

HALAMAN PENGESAHAN DOSEN PENGUJI

ANALISIS ASOSIASI PADA NILAI MAHASISWA INFORMATIKA UII MENGGUNAKAN DATA MINING DENGAN ALGORITMA APRIORI

TUGAS AKHIR

Telah dipertahankan di depan sidang penguji sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer dari Program Studi Informatika – Program Sarjana di Fakultas Teknologi Industri Universitas Islam Indonesia

Yogyakarta, 12 Februari 2024

Tim Penguji

Lizda Iswari, S.T, M.Sc

Anggota 1

Arrie Kurniawardhani, S.Si., M.Kom.

Anggota 2

Sheila Nurul Huda, S.Kom., M.Cs.

Mengetahui,
Ketua Program Studi Informatika – Program Sarjana
Fakultas Teknologi Industri
Universitas Islam Indonesia

(Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D.)

HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan di bawah ini:

Nama : Ilma Khoiruzzidan

NIM : 18523245

Tugas akhir dengan judul:

ANALISIS ASOSIASI PADA NILAI MAHASISWA INFORMATIKA UII MENGGUNAKAN DATA MINING DENGAN ALGORITMA APRIORI

Menyatakan bahwa seluruh komponen dan isi dalam tugas akhir ini adalah hasil karya saya sendiri. Apabila di kemudian hari terbukti ada beberapa bagian dari karya ini adalah bukan hasil karya sendiri, tugas akhir yang diajukan sebagai hasil karya sendiri ini siap ditarik kembali dan siap menanggung risiko dan konsekuensi apapun.

Demikian surat pernyataan ini dibuat, semoga dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 13 Februari 2024



(Ilma Khoiruzzidan)

HALAMAN PERSEMBAHAN

Bismillahhirrahmanirrahim, puji syukur kepada Allah SWT berkat ridhonya sampai hari ini saya dapat menyusun laporan tugas akhir ini. Semoga ilmu yang selama ini saya pelajari mendapat keberkahan dan berguna baik untuk saya pribadi maupun orang-orang yang ada disekitar saya. Saya persembahkan karya ini untuk,

Kedua orang tua dan Adik-Adik saya

Yang tidak henti-hentinya mendoakan juga mempercayai saya sekaligus mendukung penuh baik secara moral juga finansial, sehingga saya dapat menyelesaikan laporan akhir ini dengan baik walaupun jauh dari kata sempurna.

Dosen Pembimbing

Ibu Lizda Iswari, S.T, M.Sc yang membimbing saya mulai dari pencarian judul, memberi ilmu tentang pengolahan data hingga penyusunan laporan akhir ini.

Dosen Pembimbing Akademik

Bapak Yudi Prayudi, S.Si., M.Kom. yang selalu mendukung kegiatan studi saya dengan baik, dalam hal penelitian maupun tidak.

Teman-teman Seperjuangan

Yang menghabiskan waktu bersama saya dalam menyusun laporan TA. Tidak hanya menyemangati tetapi memberikan dukungan penuh baik diluar konteks penelitian maupun tidak. Semoga teman-teman saya dipermudah dalam semua hal, baik dunia maupun di akhirat kelak.

Akhir kata, saya mendoakan semoga Allah mengganti kebaikan-kebaikan yang kalian berikan dengan balasan yang jauh lebih baik. Aamiiin.

HALAMAN MOTO

“Janganlah kamu bersedih, sesungguhnya Allah bersama kita”

(At-Taubah : 40)

“Never back down, never give up!!”

Nick

“You only live once, but if you do it right, once is enough.”

Mae West

“Jika ada lebih dari 5 orang yang membencimu, ada baiknya untuk kamu mengintrospeksi diri”

Dr. Tirta Mandira Hudhi

KATA PENGANTAR

Allhamdulillahirobbil'alamin, Puji dan syukur dipanjatkan sebesar-besarnya kepada Allah SWT. yang telah memberi kemudahan dan kelancaran atas segalanya sehingga penulis dapat menyelesaikan laporan tugas akhir yang berjudul "Analisis Asosiasi pada Nilai Mahasiswa Informatika UII Menggunakan Data Mining dengan Algoritma Apriori". Penulis menyadari bahwa penelitian ini belum bisa dikatakan sempurna, oleh karena itu diharapkan dari laporan akhir penelitian ini, topik yang relevan terhadap penelitian ini nantinya dapat dikembangkan lebih lanjut.

Dalam penyusunan laporan akhir ini, tentunya banyak kendala dan hambatan yang dihadapi oleh penulis, namun dari sisi ini penulis menganggapnya sebagai sebuah tantangan sehingga hal itu dapat dihadapi dengan baik. Oleh karena itu, penulis ingin menyampaikan terimakasih yang sebesar-besarnya kepada:

1. Ibu Lizda Iswari selaku dosen pembimbing saya memberi arahan untuk topik untuk laporan akhir penelitian ini sekaligus memberikan masukan, saran, dan ilmu juga bimbingan yang sangat berpengaruh dalam pembuatan laporan akhir ini.
2. Ibu dan Ayah saya tercinta, Tutik Rahmawati dan Irbabul Lubab yang memberi dukungan tiada henti dan mendoakan saya untuk menyelesaikan laporan akhir ini.
3. Teman seperjuangan saya yang tidak dapat saya sebut satu-persatu karena mereka lah yang memberi motivasi sekaligus menemani saya disaat saya mengalami kesulitan dalam menyusun laporan akhir ini.
4. Teman-teman angkatan Informatika, yang turun memberikan dukungan terhadap penelitian ini.

Sehingga dalam penyusunan laporan akhir penelitian ini, penulis berharap topik yang diangkat pada penelitian ini dapat memberi manfaat bagi siapapun yang membacanya termasuk bagi diri saya sendiri. Terima kasih.

Yogyakarta, 26 Desember 2023



(Ilma Khoiruzzidan)

SARI

Penelitian ini mengeksplorasi penerapan algoritma Apriori pada 150 nilai mata kuliah mahasiswa dengan tujuan mengidentifikasi korelasi antara nilai mata kuliah dan memberikan wawasan yang berguna baik bagi peneliti maupun institusi. Data diperoleh secara langsung dari database akademik mahasiswa. Penerapan algoritma Apriori dilakukan melalui tiga skenario dengan variasi batas nilai *minimum support* dan *confidence*. Hasil penelitian menggambarkan pola nilai mata kuliah yang cenderung muncul bersamaan. Dari ketiga skenario, kombinasi Pengembangan Gim dengan nilai A, Fundamen Informatika dengan nilai A, dan Fundamen Matematika dengan nilai A memperoleh nilai tertinggi dengan support sebesar 26% dan confidence sebesar 88.6%. Temuan ini dapat memberikan kontribusi penting dalam pemahaman faktor-faktor yang mempengaruhi kinerja akademik mahasiswa, serta memberikan landasan bagi perbaikan strategi pembelajaran di institusi pendidikan.

Kata kunci: *data mining*, analisis asosiasi, *apriori*, nilai mahasiswa.

GLOSARIUM

Apriori	algoritma dalam data mining yang digunakan untuk menemukan pola asosiasi atau hubungan antara berbagai item dalam dataset.
Asosiasi	suatu metode dalam data mining yang digunakan untuk mengidentifikasi dan menganalisis hubungan atau korelasi antara variabel atau item dalam dataset.
Confidence	nilai keyakinan yang mengindikasikan kekuatan hubungan antara dua item dalam sebuah aturan asosiasi
Support	nilai ukuran dari suatu itemset dalam dataset
Data Mining	proses ekstraksi atau penggalian pola-pola menarik atau pengetahuan yang bermanfaat dari dalam suatu dataset yang besar, kompleks, dan biasanya tidak terstruktur.
Dataset	kumpulan dari sebuah banyak data

DAFTAR ISI

HALAMAN JUDUL i
HALAMAN PENGESAHAN DOSEN PEMBIMBING ii
HALAMAN PENGESAHAN DOSEN PENGUJI iii
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR iv
HALAMAN PERSEMBAHAN v
HALAMAN MOTO vi
KATA PENGANTAR vii
SARI viii
GLOSARIUM ix
DAFTAR ISI x
DAFTAR TABEL xii
DAFTAR GAMBAR xiii
BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 3
1.3 Batasan Masalah 3
1.4 Tujuan Penelitian 3
1.5 Manfaat Penelitian 3
1.6 Sistematika Penulisan 3
BAB II LANDASAN TEORI 5
2.1 Data dan Informasi 5
2.2 Data Mining 5
2.3 Analisis Asosiasi 9
2.3.1 Parameter dalam aturan asosiasi 10
2.4 Algortima apriori 13
2.4.1 Cara kerja Algoritma Apriori 13
2.4.2 Ilustrasi penerapan algoritma apriori 16
2.4.3 Algoritma apriori dalam bentuk pseudocode 16
2.5 Python 18
2.6 Tinjauan Pustaka 19
BAB III METODOLOGI PENELITIAN 22
3.1 Subjek dan Objek penelitian 22

3.2 Sumber data penelitian 22
3.3 Metode pengumpulan data penelitian 22
3.4 Alur penelitian 23
3.5 Data *collection* 24
3.6 Data *pre-processing* 24
3.6.1 Pembentukan Dataset 24
3.6.2 Transformasi data 26
3.6.3 Penerapan hot-encoding 27
3.7 Implementasi algoritma apriori 27
3.7.1 Penetapan batas minimum support 27
3.7.2 Pencarian frequent itemsets 28
3.8 Ekstraksi aturan 28
3.9 Filtering 28
3.10 *Interpretation* 29
BAB IV HASIL DAN PEMBAHASAN 30
4.1 Dataset 30
4.2 Pencarian frequent itemsets 30
4.2.1 Implementasi algoritma apriori 30
4.2.2 Hasil pencarian frequent itemsets 31
4.3 Filtering aturan asosiasi 31
4.3.1 Pembentukkan aturan asosiasi 32
4.3.2 Filtrasi hasil pencarian aturan asosiasi pertama 33
4.3.3 Filtrasi hasil pencarian aturan asosiasi kedua 34
4.3.4 Filtrasi hasil pencarian aturan asosiasi ketiga 35
4.4 *Interpretation* 37
BAB V KESIMPULAN DAN SARAN 39
5.1 Kesimpulan 39
5.2 Saran 39
DAFTAR PUSTAKA 41
LAMPIRAN 45

DAFTAR TABEL

Tabel 2.1 Format data transaksi 9
Tabel 2.2 Format data tabular 10
Tabel 2.3 Tinjauan pustaka 19
Tabel 3.1 Dataset nilai mahasiswa 25
Tabel 3.2 Pengkodean mata kuliah 26
Tabel 4.1 Hasil pencarian frequent itemsets 31
Tabel 4.2 Hasil pembentukan aturan asosiasi 32
Tabel 4.3 Hasil filtrasi pertama 33
Tabel 4.4 Hasil filtrasi kedua 34
Tabel 4.5 Hasil Filtrasi ketiga 36

DAFTAR GAMBAR

Gambar 2.1 Proses pencarian pengetahuan dari database (KDD) 6
Gambar 2.2 Ilustrasi algoritma apriori dalam mencari *frequent itemsets* 16
Gambar 2.3 Pseudocode algoritma apriori 17
Gambar 2.4 Tampilan Google Colab sebagai IDE untuk menjalankan Python 18
Gambar 3.1 Alur penelitian 23
Gambar 3.2 Data nilai mahasiswa 25
Gambar 3.3 Hasil transpose dari salah satu data mahasiswa per-mata kuliah 25
Gambar 3.4 Format item 26
Gambar 3.5 Gambaran penerapan teknik *hot-encoding* 27
Gambar 4.1 Pencarian Frequent Itemsets 30
Gambar 4.2 Pembentukan aturan asosiasi 32
Gambar 4.3 Filtrasi pertama 33
Gambar 4.4 Filtrasi kedua 34
Gambar 4.5 Filtrasi ketiga 35

BAB I
PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi saat ini merambah ke berbagai sektor, termasuk bidang pengolahan data. Hal ini disebabkan oleh semakin banyaknya sumber data yang tersedia. Volume data yang besar memerlukan proses pengumpulan yang baik, sehingga dapat diolah menjadi informasi yang relevan bagi institiusi terkait. Menurut Puspaningrum (2020) data merupakan sumber informasi yang bentuknya masih mentah. Dengan mengolah data tesebut, kita memperoleh informasi berharga sebagai acuan dalam pengambilan keputusan.

Dalam konteks akademik, data nilai mahasiswa merupakan sumber informasi yang dapat dimanfaatkan untuk meningkatkan kualitas pembelajaran dan pengajaran. Data akademik mahasiswa mencakup informasi yang luas, mulai dari nilai mahasiswa, absensi kehadiran, masa studi, hingga biodata mahasiswa. Dalam upaya peningkatan kualitas pendidikan, evaluasi terhadap pencapaian mahasiswa menjadi hal yang sangat penting. Namun, semakin bertambahnya jumlah mahasiswa dan keberagaman mata kuliah yang ditawarkan oleh institusi, membuat proses evaluasi semakin kompleks. Oleh karena itu, penelitian penulis memfokuskan penelitian terhadap data nilai mahasiswa. Hal ini dikarenakan data nilai mahasiswa merupakan data yang paling relevan dengan pencapaian akademik mahasiswa. Nilai mata kuliah dalam sistem perguruan tinggi merupakan penilaian terhadap kegiatan dan kemajuan belajar mahasiswa dilakukan secara berkala yang diperoleh dari berbagai kegiatan, yaitu, dapat berbentuk ujian, pelaksanaan tugas, juga pengamatan dengan kriteria nilai akhir (Prodi Statistika Universitas Islam Indonesia, 2017). Nilai mahasiswa mencerminkan pemahaman dan kemampuan mahasiswa pada setiap mata kuliah yang diambilnya.

Salah satu metode efektif untuk mengoptimalkan manfaat dari data nilai mahasiswa ini adalah melalui penerapan *data mining*. *Data mining* adalah proses analisis data yang besar untuk menemukan pola dan tren yang tidak terlihat secara langsung. Dengan mengolah data nilai mahasiswa secara tepat, para mahasiswa dan tenaga pendidik dapat memanfaatkannya untuk memperoleh *insight* sebagai evaluasi dalam kegiatan ajar-mengajar (Hu, 2017). *Insight* ini dapat digunakan untuk meningkatkan kualitas pembelajaran, seperti, menentukan mata kuliah yang sulit dan perlu diberikan perhatian khusus, menentukan metode pembelajaran yang

efektif untuk setiap mata kuliah, mengetahui faktor-faktor yang mempengaruhi keberhasilan belajar mahasiswa.

Data mining adalah proses analisis data yang besar untuk menemukan pola dan tren yang tersembunyi. Pola dan tren tersebut dapat digunakan untuk mendukung pengambilan keputusan penting, termasuk dalam aktivitas dan pengembangan pendidikan (Santosa, 2007). Salah satu teknik *data mining* adalah analisis asosiasi. Analisis asosiasi bertujuan menemukan pola hubungan antar beberapa item dalam suatu dataset (Despitaria et al., 2016). Analisis asosiasi lebih dikenal sebagai *Market Based Analysis* (MBA), yaitu analisis untuk mengidentifikasi item-item yang sering dibeli secara bersamaan oleh pelanggan dalam suatu transaksi (Nurchalifatun, 2017). Analisis asosiasi merupakan teknik *data mining* yang sering digunakan untuk menganalisis transaksi belanja di *supermarket*, tetapi juga dapat digunakan untuk menganalisis data lainnya, seperti nilai mahasiswa. Berdasarkan penelitian terdahulu yang dilakukan oleh Xiaodong Wu dan Yuzhu Zeng (2019), analisis asosiasi diterapkan dengan algoritma apriori terhadap 100 nilai mahasiswa pada 34 mata kuliah. Oleh karena itu, dalam penelitian ini juga digunakan algoritma apriori untuk menemukan pola hubungan antar mata kuliah berdasarkan data nilai mahasiswa.

Pemanfaatan algoritma apriori pada penelitian ini adalah dengan mencari sebuah hubungan antara nilai mata kuliah pada satu semester dengan yang lainnya. Ketika seorang mahasiswa memperoleh nilai “A” pada mata kuliah Fundamen Matematika pada Semester ke-2 maka kita sebagai penulis dapat melihat pola apakah dia memperoleh nilai “A” juga pada mata kuliah Matematika Lanjut pada Semeseter ke-3. Hubungan-hubungan ini dibatasi dengan nilai *support* dan *confidence* pada setiap aturannya. Pengetahuan ini sangat berguna bagi para tenaga pendidik sebagai bahan evaluasi juga bahan pengambil keputusan dalam penentuan hal-hal yang terkait dengan efektivitas pembelajaran mahasiswa. Data yang digunakan pada penelitian ini adalah nilai data mahasiswa Mahasiswa UII Informatika Angkatan 2020.

Berdasarkan seluruh rincian diatas, penelitian ini akan berfokus untuk mencari korelasi antar data nilai mahasiswa pada setiap mata kuliahnya yang dihasilkan dari algoritma apriori. Harapannya hasil ini dapat memberikan insight bagi para dosen, institusi pendidikan, mahasiswa maupun peneliti yang akan melakukan penelitian serupa.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, penulis mengajukan permasalahan yaitu, bagaimana hasil analisis asosiasi terhadap nilai mata kuliah mahasiswa menggunakan algoritma apriori?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

a. Penelitian ini menggunakan data asli dari mahasiswa Informatika UII Angkatan 2020 yang diperoleh pada tanggal 09 Agustus 2023 sebanyak 150 mahasiswa.
b. Data yang digunakan adalah data nilai mahasiswa pada 18 mata kuliah wajib yang bersifat keinformatikaan dengan pengecualian mata kuliah kewarganegaraan, keaagamaan dan bahasa.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah mengidentifikasi pola keterkaitan antar mata kuliah dan menginterpretasikan aturan-aturan yang dihasilkan dari algoritma apriori terhadap nilai mata kuliah mahasiswa sebagai *insight* bagi *stakeholder* terkait.

1.5 Manfaat Penelitian

Terdapat beberapa manfaat yang diperoleh dalam penelitian ini:

a. Mengetahui bagaimana keterkaitan pola antara mata kuliah satu dengan lainnya.
b. Menjadi bahan evaluasi bagi tenaga pendidik dalam strategi pengajaran yang lebih efektif bagi mahasiswa tertentu.
c. Dapat dijadikan sebagai referensi bagi peneliti yang akan melakukan penelitian sejenis dalam topik analisis asosiasi di masa depan.

1.6 Sistematika Penulisan

Sistematika penulisan bertujuan untuk mengetahui isi dan maksud dari laporan tugas akhir tersebut, berikut sistematika penulisan yang disusun pada laporan tugas ini:

BAB I PENDAHULUAN

Bagian ini terdiri dari latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bagian ini berisi tinjauan Pustaka dan teori dasar yang diambil dari berbagai sumber yang relevan dengan penelitian ini.

BAB III METODOLOGI PENELITIAN

Bagian ini berisi tentang langkah-langkah yang dilakukan dalam penelitian mulai dari tahap awal hingga akhir.

BAB IV HASIL DAN PEMBAHASAN

Bagian ini berisi tentang pembahasan dari hasil yang telah didapat dari penerapan algoritma apriori terhadap dataset.

BAB V KESIMPULAN DAN SARAN

Bagian ini berisi mengenai kesimpulan dari seluruh hasil tahap penelitian yang telah ditulis dan juga saran bagi para peneliti yang memiliki studi kasus yang serupa dengan penelitian ini.

BAB II
LANDASAN TEORI

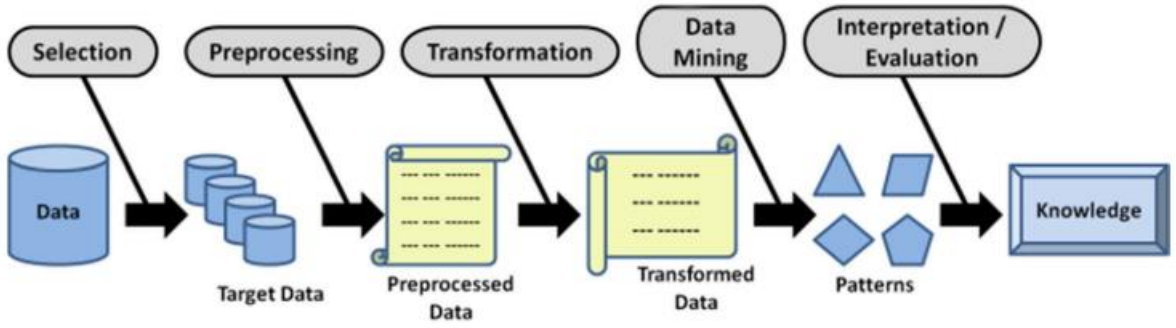
2.1 Data dan Informasi

Dalam melakukan *data mining* untuk menemukan pola dalam *dataset*, penggunaan data sangatlah penting sebagai bahan rujukan bagi algoritma itu sendiri. Menurut Lubis (2016) data adalah fakta-fakta yang menggambarkan suatu kejadian yang sebenarnya pada waktu tertentu. Ketika berususan dengan analisis asosiasi dalam *data mining*, data merupakan Kumpulan fakta-fakta yang mencerminkan berbagai kejadian atau transaksi yang terjadi. Fakta-fakta ini kemudian diorganisir dan dikelompokkan ke dalam sebuah *dataset*, yang kemudian akan menjadi landasan bagi algoritma untuk menemukan pola asosiasi di antara berbagai *item* dalam data tersebut. Hasil akhir dari pengolahan tersebut menjadi informasi baru bagi orang yang membacanya.

Informasi merupakan hal yang sangat mendasar yang diperlukan dalam pengambilan keputusan untuk mencegah bias dari pengambil keputusan dan mengurangi potensi kesalahan. Informasi juga dapat diartikan sebagai data yang telah diolah menjadi bentuk yang lebih bermakna dan berguna bagi penerima informasi tersebut. Sebagaimana yang diungkapkan oleh Anggraeni (2017), informasi didefinisikan sebagai sekumpulan data atau fakta yang diorganisasi atau diolah dengan cara tertentu sehingga memiliki arti bagi penerima. Dengan pengolahan informasi, data menjadi bahan mentah yang kemudian diolah menjadi informasi yang lebih berguna sebagai landasan untuk mengambil keputusan yang lebih baik.

2.2 Data Mining

Data mining adalah rangkaian proses yang dilakukan untuk menggali sebuah nilai berupa informasi yang belum diketahui dari suatu *database*. Informasi yang dihasilkan bisa diperoleh melalui ekstraksi pola-pola yang terbentuk dari dalam *database* itu sendiri. *Data mining* utamanya digunakan untuk mencari *insight* yang terdapat di dalam *database* maupun dataset. *Data mining* juga sering disebut sebagai *Knowledge Discovery Database* (KDD) (Vulandari, 2017) (Chee et al., 2019).



Gambar 2.1 Proses pencarian pengetahuan dari database (KDD)

Di sisi lain, menurut Syaifullah (2010) dalam penelitiannya mengatakan bahwa, *data mining* merupakan prosedur yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mencari dan mengindentifikasi informasi juga pengetahuan yang bermanfaat dari dalam suatu dataset. Terdapat ada kesamaan definisi dari kedua pendapat para ahli, yaitu, inti dari *data mining* merupakan teknik yang digunakan untuk mencari informasi yang tersembunyi menggunakan sebuah perhitungan matematika dari sebuah *dataset* atau *database.*

Pada buku yang ditulis oleh Budi Santosa (2007) terdapat 6 teknik utama dalam proses *data mining*, yaitu:

a. Deskripsi

Terkadang, untuk memahami pola dan tren dalam data bisa tugas yang merepotkan. Inilah alasan mengapa para peneliti berupaya menemukan cara-cara yang lebih mudah untuk menyajikan informasi tentang pola dan tren tersebut. Penjelasan terkait pola dan tren ini berfungsi sebagai sarana untuk menggambarkan inti dari pola atau tren tersebut dengan lebih sederhana. Teknik ini berkaitan dengan menggambarkan dan merangkum pola serta tren yang terdapat dalam data, sehingga memungkinkan untuk menyajikan informasi yang lebih mudah dipahami oleh pengguna.

b. Klasifikasi

Dalam proses klasifikasi, keluaran yang ingin diprediksi adalah data dengan karakteristik kategorikal. Model *data mining* akan melakukan analisis pada sejumlah besar data, dimana setiap data memiliki informasi mengenai keluaran yang dituju serta kumpulan input atau variabel prediksi. Dalam dunia nyata, hal ini mirip dengan bagaimana konsep manusia belajar. Misalkan kita ingin mengajari komputer untuk membedakan antara jenis-jenis bunga berdasarkan ciri-ciri tertentu. Setiap bunga adalah entitas yang memiliki ciri seperti warna, bentuk, dan ukuran kelopak dan mahkota bunga. Tujuan kita adalah mendapatkan output

berupa jenis bunga, yang juga merupakan variabel kategorikal. Model *data mining* akan menganalisis sekumpulan data yang mencakup informasi ini yang kemudian berusaha menemukan pola yang dapat membantu komputer memahami bagaimana mengklasifikasikan bunga-bunga ini secara tepat berdasarkan fitur-fitur atau ciri yang dimilikinya.

c. Estimasi

Teknik estimasi memprediksi nilai numerik berdasarkan informasi yang ada dalam *dataset*. Ini berbeda dari klasifikasi yang tujuannya adalah mengklasifikasikan data-data dengan kategori tertentu. Dalam estimasi, output dari prediksi berupa data numerik, seperti tinggi, berat dan sebagainya. Pada intinya, pada teknik estimasi menghasilkan model yang memiliki kemampuan untuk memberikan perkiraan berdasarkan data yang telah diberikan. Salah satu hasil dari estimasi contohnya adalah estimasi harga tanah dengan kombinasi faktor-faktor tertentu berdasar data historisnya.

d. Prediksi

Prediksi memiliki kemiripan dengan klasifikasi dan estimasi. Namun, perbedaannya terletak pada fokus terhadap nilai-nilai yang akan muncul di masa depan. Dalam hal ini, model *data mining* memproyeksikan nilai-nilai yang diantisipasi di masa mendatang berdasarkan informasi sebelumnya. Dengan teknik prediksi, kita dapat membuat keputusan yang cerdas dan mengelola resiko dengan lebih baik. Hasil dari prediksi contohnya adalah prediksi tren 3 bulan kedepan dalam bisnis tertentu, faktor-faktor yang mempengaruhi penjualan, dan sebagainya.

e. Clustering

Clustering adalah data teknik yang digunakan untuk mengelompokkan data kedalam sebuah kelas-kelas dengan objek yang memiliki karakteristik serupa. Dalam *clustering*, tidak ada variabel target yang ditentukan sebelumnya. Model mencari sendiri pola-pola kesamaan di antara data dan mengelompokkannya berdasarkan kemiripan karakteristik. Proses ini seringkali bersifat eksploratif, yang berarti kita mungkin tidak tahu sebelumnya berapa banyak *cluster* yang akan terbentuk atau bagaimana *cluster-cluster* ini akan terlihat. Tujuan utama dari teknik ini adalah pemahaman terhadap struktur pada data yang dapat membantu dalam pengambilan keputusan ataupun analisis lebih lanjut.

f. Asosiasi

Teknik asosiasi dalam *data mining* berfokus pada pencaian hubungan antar atribut atau item yang sering muncul bersamaan dalam data. Teknik ini memberikan pengetahuan tentang bagaimana item-item itu saling berhubungan atau saling mempengaruhi. *Market Based*

Analysis (MBA) biasanya digunakan untuk menemukan suatu aturan asosiasi antara suatu kombinasi item atau barang pada beberapa transaksi.

Data mining sering juga disebut sebagai KDD, karena prosesnya meliputi pengumpulan, pengelompokan, pemakaian data untuk menemukan suatu keteraturan, pola atau hubungan dalam dataset yang sangat besar (Santosa, 2007). Menurut Budi Santosa pada bukunya yang berjudul ”Data Mining : Teknik Pemanfaatan Data Untuk Keperluan Bisnis” secara garis besar, tahapan – tahapan dalam *data mining* dapat diuraikan sebagai berikut:

a. Data Cleaning

Data cleaning bertujuan untuk membuang data-data yang tidak cocok terhadap teknik/metode *data mining* yang akan digunakan juga melengkapi data yang hilang dari dataset/database tersebut supaya performa *data mining* dapat meningkat.

b. Data Integration

Data yang akan diproses dalam *data mining* berasal dari banyak database. Integrasi data diperlukan untuk menyatukan databse menjadi satu kesatuan database baru. Integrasi data yang hasilnya diteliti dapat meningkatkan akurasi dan kecepatan dari proses *data mining*.

c. Data Selection

Di tahap ini kita melakukan penyortiran terhadap data yang akan digunakan untuk proses *data mining*. Terkadang terdapat data yang tidak diperlukan dalam prosesnya. Pada studi kasus ini dari data nilai akademik yang diperlukan beberapa nilai mata kuliah dengan NIM yang mewakili setiap mahasiswa, tidak memerlukan nama mahasiswa pada prosesnya.

d. Data Transformation

Pada tahap transformasi data, data yang digunakan akan diubah menjadi format yang sesuai dengan metode yang digunakan pada proses *data mining*.

e. Data Mining

Tahap ini merupakan tahapan dimana kita bisa mulai melakukan penerapan metode *data mining* terhadap data yang telah diproses dan disortir sebelumnya untuk mencari pola dan informasi yang tersembunyi di dalam data tersebut.

f. Pattern Evaluation

Pola yang dihasilkan dari proses *data mining* bisa jadi berbeda dari hipotesis yang telah dibuat. Jika hal ini terjadi, maka hal yang harus kita lakukan adalah melakukan evaluasi pola terhadap hasil yang ada dan memperbaikinya. Contoh solusinya bisa dengan mengubah metode yang digunakan ataupun menerima hasil yang diberikan sistem sebagai pengetahuan baru.

g. Knowledge Presentation

Pada tahap akhir proses *data mining* ini, informasi atau pengetahuan baru yang telah dihasilkan akan disajikan terhadap stakeholder yang bersangkutan. Hal ini dikarenakan tidak setiap stakeholder memahami hasil yang ditemukan, maka yang perlu kita lakukan adalah menyajikannya dengan baik supaya dapat dipahami dengan baik juga oleh stakeholder. Salah satunya adalah memvisualisasikan data pada saat presentasi hasil proses *data mining* itu sendiri.

2.3 Analisis Asosiasi

Analisis asosiasi adalah salah satu teknik dalam *data mining* yang menjadi dasar dari salah satu teknik *data mining* lainnya untuk menemukan aturan asosiasi antar suatu kombinasi item (Dengen et al., 2019). Analisis asosiasi juga didefinisikan sebagai suatu proses untuk menemukan pola yang sering muncul, hubungan antara suatu kombinasi item dan saling membuat korelasi antara item yang telah dikelompokkan (*itemset*) atau "muncul" secara bersamaan kemudian mengambil kesimpulan berdasarkan hubungan yang terbentuk dari beberapa item juga *itemset* tadi. Dibatasi dengan aturan "*If antecedent, then consequent*" bersama dengan ukuran *support* dan *confidence* yang menjadi parameter untuk mendapatkan aturan yang optimal (Yusuf et al., 2006). *Antecedent* merupakan sebab yang menjadikan item *consequent*. Sedangkan, *Consequent* adalah sebuah akibat atau juga item yang akan dibeli setelah membeli *Antecedent* (Rahmi & Mikola, 2021).

Terdapat dua metode utama dalam merepresentasikan tipe data dalam MBA, yaitu format data transaksi dan format data tabular. Format data transaksi membutuhkan dua *field,* berupa *field* ID dan *field item*, dimana tiap *record* merepresentasikan hanya satu *item*. Berikut contoh format data transaksi sebagaimana ditampilkan pada Tabel 2.1 (Yusuf et al., 2006).

Tabel 2.1 Format data transaksi

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

Pada format data tabular, setiap catatan mewakili transaksi secara biner dengan mengandung nilai 0 atau 1 sesuai dengan keberadaan atau ketiadaan item tersebut. Format data tabular dapat diilustrasikan melalui Tabel 2.2.

Tabel 2.2 Format 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
...	...	...	...	...
n	...	...	...	...

Seperti yang terlihat pada Tabel 2.2, pada *record* terdapat ID_Transaksi 1 yang memilih item 1 dan 2, diindikasikan dengan nilai 1 pada keduanya. ID_Transaksi 3 yang tidak memilih item 1 ditunjukkan dengan nilai 0 pada kolom tersebut, tetapi pada kolom item 2 untuk ID_Transaksi 3, terdapat nilai 1 yang menunjukkan bahwa item tersebut diambil. Dengan demikian, nilai 0 atau 1 di sini mencerminkan apakah seorang pelanggan (ID_Transaksi) mengambil item tersebut atau tidak dalam transaksi belanja mereka (Yusuf et al., 2006).

2.3.1 Parameter dalam aturan asosiasi

Dalam analisis aturan asosiasi, terdapat beberapa parameter kunci yang umumnya digunakan untuk mengevaluasi dan memahami kualitas aturan yang dihasilkan yaitu, *support*, *confidence*, dan *lift*.

a. Support

Support adalah persentase dari *item* atau *itemset* tersebut muncul dalam keseluruhan *dataset*. Dalam konteks kombinasi A dan B, nilai *support* dari {A, B}mencerminkan seberapa sering transaksi dalam *dataset* memuat *item* A dan B secara bersamaan.

$$\text{Support (A)} = \frac{\text{Jumlah transaksi mengandung A}}{\text{Total transaksi}} . 100\% \qquad (2.1)$$

Persamaan (2.1) menjelaskan bahwa nilai *support* suatu *item* A dihitung dengan membagi jumlah transaksi yang mengandung A dengan total transaksi, kemudian dikalikan dengan 100 persen. Demikian pula, untuk menghitung nilai *support* dari suatu *itemset* {A, B}, dapat digunakan rumus seperti yang ditunjukkan pada Persamaan (2.2).

$$\text{Support (A, B)} = \frac{\text{Jumlah transaksi mengandung A dan B}}{\text{Total transaksi}} . 100\% \qquad (2.2)$$

Pada tahapan analisis asosiasi, terdapat istilah *minimum support*. *Minimum support* didefinisikan sebagai nilai persentase yang digunakan untuk menyaring aturan-aturan yang memiliki nilai *support* di bawah ambang batas tersebut. Penentuan nilai *minimum support* perlu disesuaikan dengan tujuan analisis dan karakteristik *dataset*. Jika nilai *minimum support* terlalu tinggi, analisis dapat kehilangan aturan-aturan yang mungkin memiliki nilai informatif. Sebaliknya, jika nilai *minimum support* terlalu rendah, dapat menyebabkan banyak aturan yang tidak relevan atau terlalu umum.

Oleh karena itu, pemilihan nilai *minimum support* memerlukan kebijaksanaan dan pemahaman yang baik terhadap konteks analisis dan karakteristik data yang sedang diteliti. Batasan yang tepat untuk *minimum support* memungkinkan analisis asosiasi memberikan aturan-aturan yang bermakna dan sesuai dengan kebutuhan penulis.

b. Confidence

Confidence adalah nilai yang mengindikasikan seberapa besar kemungkinan *item* akan muncul dalam aturan asosiasi. Sebagai contoh, Confidence (A, B) menunjukkan seberapa *item* B akan bersamaan dibeli dengan A. Rumus untuk menghitung nilai *confidence* tersebut dapat dinyatakan sebagai berikut:

$$\text{Confidence}\ (\text{A}, \text{B}) = \frac{\text{Jumlah transaksi mengandung A dan B}}{\text{Jumlah transaksi mengandung A}} \qquad (2.3)$$

Rumus umum untuk menghitung nilai *confidence* dalam bentuk persentase dapat dinyatakan dengan Persamaan (2.4).

$$\text{Confidence}\ (\text{A}, \text{B}) = \frac{\text{Support}\ (\text{A}, \text{B})}{\text{Support}\ (\text{A})} . 100\% \qquad (2.4)$$

Dengan menggunakan persamaan tersebut, kita dapat mengevaluasi seberapa kuat asosiasi antara *item* A dan B berdasarkan frekuensi kemunculan keduanya dalam transaksi. Semakin tinggi nilai *confidence*, semakin besar kemungkinan *item* B akan dibeli bersamaan dengan A dalam suatu transaksi.

Pada tahapan analisis asosiasi, penetapan nilai *minimum confidence* sangat penting karena membantu penulis untuk mengontrol kualitas dari aturan asosiasi yang dihasilkan. *Confidence* mengukur seberapa kuat hubungan antara *antecedent* dan *consequent* pada suatu

aturan. Nilai *confidence* yang tinggi menunjukkan bahwa aturan tersebut cenderung lebih kuat dan dapat diandalkan.

Penetapan nilai *minimum confidence* dilakukan oleh penulis berdasarkan batasan atau standar tertentu. Nilai ini dapat bervariasi tergantung pada tujuan penelitian, domain data, atau preferensi peneliti. Sebagai contoh, jika penelitian memiliki tujuan untuk mendapatkan aturan yang sangat kuat, nilai *minimum confidence* dapat ditetapkan lebih tinggi. Sebaliknya, jika penelitian ingin mempertahankan sejumlah besar aturan termasuk yang mungkin kurang kuat, nilai *minimum confidence* dapat ditetapkan lebih rendah.

c. Lift

Nilai *Lift* merupakan parameter yang digunakan untuk menilai kekuatan dari aturan asosisasi yang terbentuk berdasarkan nilai *support* dan *confidence*. Selain itu, nilai *lift* juga berperan sebagai indikator validitas suatu aturan asosiasi. Sebagai contoh, Lift (A, B) menunjukkan kekuatan hubungan antara *item* A dan B. Nilai *lift* juga menjadi suatu ukuran untuk mengetahui kekuatan aturan asosasi yang telah terbentuk (Ulvah, 2018). Rentang nilai lift mulai dari 0 sampai ∞ (tak hingga). Perhitungan *lift ratio* dapat dilakukan menggunakan Persamaan (2.5).

$$\text{Lift}\ (\text{A}, \text{B}) = \frac{\text{Confidence}\ (\text{A}, \text{B})}{\text{Benchmark Confidence}\ (\text{A}, \text{B})} . 100\% \qquad (2.5)$$

Nilai Lift (A, B) didapatkan dengan pembagian nilai Confidence (A, B) dan Benchmark Confidence (A, B) (Latifah et al., 2018). Selanjutnya, nilai *benchmark confidence* dapat dihitung dengan Persamaan (2.6).

$$\text{Benchmark Confidence}\ (\text{A}, \text{B}) = \frac{\text{Jumlah transaksi mengandung (B)}}{\text{Total Transaksi}} . 100\% \qquad (2.6)$$

Benchmark confidence adalah perbandingan dari jumlah item *consequent* dengan jumlah transaksi dalam dataset (Despitaria et al., 2016). Dengan menggunakan persamaan tersebut, kita dapat mengevaluasi seberapa besar peningkatan kemungkinan kemunculan *item* B ketika *item* A sudah ada dalam transaksi, dan nilai *benchmark confidence* memberikan gambaran referensi terkait kemungkinan pembelian item B secara umum.

2.4 Algortima apriori

Algoritma apriori merupakan suatu algoritma dasar yang diusulkan oleh Agrawal dan Srikant pada tahun 1994 dalam penentuan *frequent itemsets* untuk membuat aturan asosiasi (Badrul et al., 2016). Apriori adalah algoritma yang sering digunakan untuk menambang *itemset* menggunakan aturan asosiasi yang bersifat *boolean* (Chee et al., 2019). Output dari algoritma apriori adalah aturan asosiasi yang menggambarkan hubungan atau pola yang ditemukan dalam *dataset*. Aturan asosiasi ini biasanya terdiri dari dua bagian utama, yaitu *antecedent* (jika) dan *consequent* (maka). Aturan ini menyatakan bahwa jika suatu kondisi tertentu terpenuhi (*antecedent*), maka ada kecenderungan atau asosiasi untuk kondisi lain terpenuhi juga (*consequent*).

Aturan asosiasi adalah pola-pola asosiatif yang menenuhi syarat nilai minimal dari dua parameter yaitu *support* dan *confidence* (Rino Subekti, Astried Silvanie & Boy Firmansyah, 2023). Aturan asosiasi akan menyatakan asosiasi dari *item* yang ada dalam aturan tersebut. Aturan asosiasi berbentuk "*If antecedent, then consequent*", dilengkapi dengan tingkat *support* dan *confidence* aturan tersebut (Yusuf et al., 2006). Algoritma apriori adalah sebuah algoritma yang digunakan untuk menemukan *item* dan *itemset* yang relevan dengan tujuan untuk menentukan aturan asosiasi. Dalam *data mining*, algoritma apriori termasuk dalam kategori teknik analisis asosiasi yang digunakan untuk menemukan sebuah kombinasi tiap item. Dengan teknik ini, penulis dapat menghitung seberapa sering kombinasi item muncul dan kemudian menggunakan informasi ini untuk tujuan prediksi dari pola yang muncul (Agrawal & Srikant, 1995).

Output lainnya dari algoritma apriori bisa mencakup informasi mengenai nilai *support* dan *confidence* dari setiap aturan asosiasi. Nilai *support* mengukur seberapa sering aturan tersebut ditemukan dalam *dataset*, sementara *confidence* mengukur seberapa kuat keterkaitan antara *antecedent* dan *consequent* pada aturan tersebut.

2.4.1 Cara kerja Algoritma Apriori

Menurut Han (2012) dalam penelitiannya, algoritma aprori memiliki dua proses utama yang harus dilakukan, yaitu:

a. *Join* (Penggabungan)

Pada proses ini setiap item dikombinasikan dengan item yang lainnya secara terus-menerus hingga tidak dapat terbentuk kombinasi baru kembali.

b. *Prune* (Pemangkasan)

Pada proses ini hasil dari item yang telah dikombinasikan sebelumnya, akan dipangkas dengan menggunakan parameter minimum *support* yang telah ditentukan oleh penulis. Minimum *support* adalah batas ambang tertentu nilai *support* yang ditentukan oleh penulis untuk mem-filter apakah suatu *itemset* termasuk dari *frequent itemset* dengan besaran nilai *support* tertentu.

Langkah-langkah algoritma apriori

Langkah pertama algoritma apriori adalah *support* dari setiap item dihitung dengan cara melakukan *scanning* pada *database*/*dataset*. Setelah *support* dipilih sebagai *frequent itemset* dengan panjang 1 atau bisa disingkat sebagai 1-*itemset*. Kita melakukan iterasi yang menghasilkan k-*itemset* yang berarti satu *set* yang terdiri dari k *item*. (k) disini mewakili dari jumlah *item* dalam satu *set*.

Pada iterasi kedua menghasilkan 2-*itemset* yang setiap *set*-nya terdiri dari dua item. Prosesnya yaitu pertama dibuatkan kandidat 2-item dari kombinasi semua 1-*itemset*. Lalu, untuk setiap kandidat 2 *itemset* ini dihitung nilai *support*-nya dengan melakukan *scanning* pada *database/dataset*. Nilai *support* ini adalah jumlah transaksi dalam *database/dataset* yang mengandung kedua *item* dalam kandidat 2-*itemset*. Setelah nilai *support* dari setiap kandidat 2-*itemset* didapatkan, kandidat 2-*itemset* yang memenuhi syarat *minimum support* dapat ditetapkan sebagai 2-*itemset* yang juga merupakan *frequent itemset* dengan panjang 2. Hal ini akan terus berulang sampai iterasi ke-k yang dapat dibagi menjadi beberapa bagian sebagai berikut:

a. Pembentukan kandidat *itemset*

Kandidat k-*itemset* dihasilkan dengan menggabungkan (k-1) *itemset* yang diperoleh dari iterasi sebelumnya. Salah satu karakteristik algoritma apriori adalah terdapat pemangkasan kandidat k-*itemset* yang memiliki *subset* (k-1) item yang tidak termasuk dalam pola *frequent itemset* dengan panjang (k-1)

b. Perhitungan nilai *support* dari tiap kandidat k-*itemset*

Nilai *support* untuk setiap kandidat k-*itemset* dihitung dengan memeriksa seluruh *database/dataset* untuk menghitung berapa banyak transaksi yang memuat semua *item* yang ada di dalam kandidat k-*itemset* tersebut. Algoritma apriori dicirikan dengan kebutuhan untuk melakukan perhitungan dengan melakukan *scanning* pada seluru *database/dataset* sebanyak jumlah *item* yang ada pada kandidat k-*itemset* terpanjang.

c. Penetapan frequent itemsets

Frequent itemset yang berisi k-item atau k-*itemset* dipilih dari kandidat k-*itemset* yang memiliki nilai *support* lebih tinggi dari batas *minimum support* yang telah ditetapkan sebelumnya. Kemudian nilai *confidence* dari setiap kombinasi item dihitung.

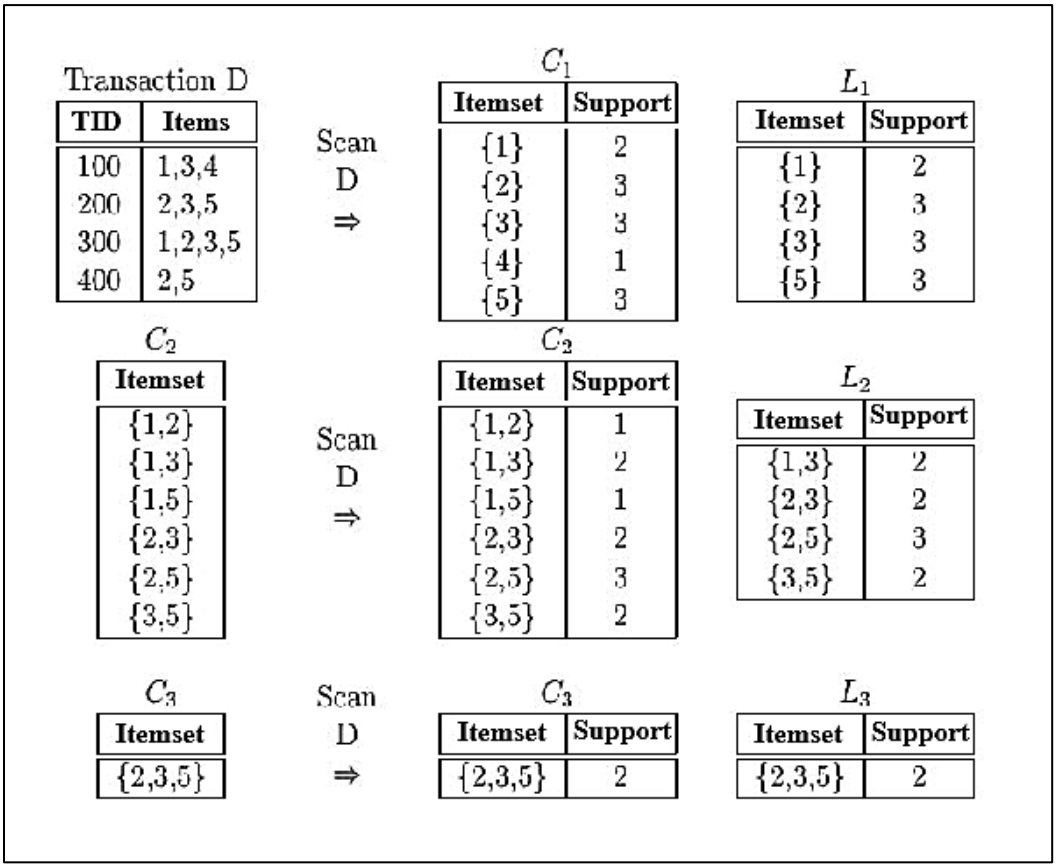
d. Akhir Iterasi

Jika tidak ditemukan *frequent itemset* baru maka seluruh proses dihentikan. Sebaliknya, jika masih ada, maka k ditambah satu dan kembali ke bagian 1.

Secara ringkas, langkah-langkah algoritma apriori adalah sebagai berikut:

1. Scanning *database*/*dataset*.
2. Menghitung nilai *support* dari masing-masing *itemset*.
3. *Pruning*, yaitu membuang *itemset* yang tidak memenuhi *minimum support*, dan melanjutkan kandidat *itemset* yang memenuhi batas *minimum support* pada *scan* berikutnya.
4. Kombinasi semua kandidat.
5. Mengulangi langkah 1 sampai 4 hingga tidak ditemukannya lagi *frequent itemset*.

2.4.2 Ilustrasi penerapan algoritma apriori



Gambar 2.2 Ilustrasi algoritma apriori dalam mencari *frequent itemsets*

Berdasarkan ilustrasi pada Gambar 2.2 yang buat oleh Agrawal dan Srikant (1994), batas *minimum support* yang digunakan dalam contoh ini adalah sebesar 50% atau setara dengan nilai 2. Pada tahap iterasi pertama, item yang memiliki *support* kurang dari nilai 2 dihapuskan dari himpunan 1-*itemset* L1. Setelah itu, kandidat 2-*itemset* C2 pada iterasi kedua terbentuk melalui permutasi dari item-item yang tersisa di L1. Setelah menghitung *support* untuk kandidat 2-*itemset* dari *database*, terbentuklah 2-*itemset* L2. Proses serupa diulangi pada iterasi selanjutnya, dan langkah-langkah tersebut akan berlanjut hingga tidak ada kandidat baru yang dapat dihasilkan pada iterasi berikutnya (Nindani, 2017)

2.4.3 Algoritma apriori dalam bentuk pseudocode

Pada Gambar 2.3 terdapat representasi visual mengenai mekanisme algoritma apriori dalam bentuk pseudo-code yang diadaptasi dari penelitian (Hui et al., 2020). *Pseudo-code* tersebut dirancang untuk menemukan *frequent itemset* dari *dataset* dengan menggunakan nilai *minimum support* yang telah ditetapkan sebelumnya. Langkah-langkah tersebut mencakup

pembentukan 1-*itemset* awal, pembentukan kandidat 2-*itemset*, dan iterasi lanjutan untuk menghasilkan *frequent itemset* dengan *support* di atas batas *minimum*.

```
Input:
- D: Dataset
- min_sup: Batas minimum support

Output:
- L: Kumpulan dari frequent itemsets

Procedure apriori_gen(Lk_1):
    Ck = []  # Inisiasi untuk kandidat itemsets

    for i = 0 to len(Lk_1) - 1:
        for j = i + 1 to len(Lk_1):
            if (Lk_1[i][1:k-2] == Lk_1[j][1:k-2]) && (Lk_1[i][k-1] < Lk_1[j][k-1]):
                c = join(Lk_1[i], Lk_1[j])  # Join step: generate candidate

                if not has_infrequent_subset(c, Lk_1):
                    Ck.append(c)  # Pruning step: Remove infrequent candidates

    return Ck

Procedure has_infrequent_subset(C, Lk_1):
    for each subset s of C:
        if s not in Lk_1:
            return true
    return false

Procedure apriori(D, min_sup):
    L = []  # Initialize an empty list for frequent itemsets
    L1 = find_frequent_1_itemsets(D)  # Find all frequent 1-itemsets
    L.append(L1)

    k = 2  # Start with frequent 1-itemsets and increment k
    while Lk_1 is not empty:
        Ck = apriori_gen(Lk_1)  # Generate candidates and prune
        for each transaction t in D:
            Ct = subset(Ck, t)  # Get a subset of t from candidate itemsets
            for each candidate C in Ct:
                C.count++

        Lk = {C in Ck | C.count >= min_sup}
        L.append(Lk)  # Store frequent k-itemsets in L
        Lk_1 = Lk
        k++

    return L
```

Gambar 2.3 Pseudocode algoritma apriori

Pseudo-code tersebut mencakup langkah-langkah rinci yang diimplementasikan oleh algoritma apriori, dimulai dari pembentukan 1-*itemset* awal hingga proses iteratif untuk menghasilkan *frequent itemset* dengan *support* di atas batas *minimum* yang ditentukan. Implementasi *pseudo-code* ini membantu pemahaman lebih mendalam terhadap proses kerja algoritma apriori dalam mengekstraksi pola-pola yang signifikan dari *dataset*.

2.5 Python

Python adalah bahasa pemrograman yang popular dalam banyak bidang, termasuk data analisis. Bahasa permrograman ini pertama kali dikembangkan oleh Guido van Rossum di Belanda pada tahun 1990. Nama Python diambil dari acara komedi favorit Guido, yaitu Monty Python's Flying Circus. Van Rossum awalnya mengembangkan Python sebagai hobi, tetapi kemudian bahasa ini menjadi populer di kalangan perusahaan dan akademisi (Cahyadi. & Purwanto. Tarjoko., 2021).

Salah satu alasan utama Python banyak digunakan adalah karena kesederhanaannya, sintax yang mudah dipahami, dan juga library yang luas (Lo et al., 2023). Library ini memungkinkan penulis untuk mengakses berbagai fungsi yang sudah ada tanpa perlu menulis kode dari awal. Python juga digunakan secara luas di perguruan tinggi dan universitas di seluruh dunia, terutama di program ilmu komputer dan teknologi informasi. Mahasiswa sering mempelajari Python untuk menyelesaikan tugas kuliah, tugas akhir, dan penelitian. Oleh karena itu, bahasa pemrograman Python akan dijadikan sebagai *tools* dalam analisis asosiasi nilai mahasiswa pada penelitian ini.



Gambar 2.4 Tampilan Google Colab sebagai IDE untuk menjalankan Python

2.6 Tinjauan Pustaka

Sejumlah penelitian sebelumnya telah mengeksplorasi analisis asosiasi terhadap nilai mahasiswa dengan memanfaatkan algoritma apriori. Pada fokus tinjauan pustaka ini, penelitian-penelitian terdahulu yang menggunakan algoritma apriori dalam konteks analisis nilai mahasiswa, meskipun pada studi kasus yang berbeda-beda, akan dianalisis. Untuk memudahkan pemahaman pembaca terhadap hasil laporan penelitian ini, kesimpulan dari tinjauan pustaka tersebut akan disajikan dalam bentuk tabel analisis, sebagaimana yang terdapat pada

Tabel 2.3.

Tabel ini akan merinci temuan-temuan yang relevan dari penelitian-penelitian terdahulu, menyajikan informasi mengenai studi kasus yang dikaji dan hasil analisis asosiasi yang ditemukan. Dengan demikian, tabel tersebut menjadi alat bantu yang efektif untuk memahami kontribusi dan temuan dari penelitian ini dalam konteks analisis asosiasi terhadap nilai mahasiswa menggunakan algoritma apriori.

Tabel 2.3 Tinjauan pustaka

No.	Judul	Studi Kasus	Hasil	Penulis
1.	Penerapan Algoritma Apriori A*ssociation Rule* untuk Analisa Nilai Mahasiswa di Universitas Gunadarma	Menemukan aturan asosiasi pada data nilai mahasiswa Universitas Gunadarma yang mengulang (mendapat nilai D dan E) pada 75 matakuliah yang telah diambil.	Berdasarkan dari aturan-aturan yang dihasilkan menunjukkan bahwa adanya hubungan antara nilai dan matakuliah yang telah diambil oleh mahasiswa. Mahasiswa yang mengulang di mata kuliah 1 cenderung mengulang juga di mata kuliah 2, hal ini mungkin dikarenakan kedua matakuliah tersebut saling berkaitan dan memiliki materi yang mirip.	(Cahyanti et al., 2017)
2.	Analisis Algoritma Asosiasi Untuk Mendapatkan Pola Pemilihan Judul Skripsi Mahasiswa STIE-GK Muara Bulian	Menemukan aturan asosiasi dari data pola pengajuan atau pengambilan judul skripsi pada mahasiswa STIE-GK Muara Bulian	Berdasarkan dari aturan-aturan yang dihasilkan menunjukkan bahwa Mahasiswa STIE-GK Muara Bulian cenderung mengajukan judul skripsi yang mereka minati dan familiar yaitu bidang Pemasaran, Keuangan dan SDM.	(Yaldi, 2018)

3.	Penerapan Algorithma Apriori Untuk Menemukan Pola Pemilihan Konsentrasi Studi Mahasiswa	Menemukan aturan asosiasi dari data nilai mahasiswa yang mengambil salah satu konsentrasi	Berdasarkan dari aturan-aturan yang dihasilkan menunjukkan bahwa terdapat mahasiswa yang tidak memenuhi kompetensi untuk memilih konsentrasi bisnis	(Jamaris et al., 2020)
4.	Application of Student achievement Analysis based on Apriori Algorithm	Menemukan aturan asosiasi dari data nilai mahasiswa pada mata kuliah wajib	Berdasarkan dari aturan-aturan yang dihasilkan menunjukkan bahwa terdapat adanya hubungan antara beberapa mata kuliah yang berbeda. Hal ini dapat digunakan sebagai evaluasi oleh perguruan tinggi untuk menyesuaikan urutan mata kuliah dan pengembangan kebijakan yang lebih efektif.	(Hui et al., 2020)
5.	Penentuan Association Rule Pada Kelulusan Mahasiswa Menggunakan Algoritma Apriori	Menemukan aturan asosasi dari data mahasiswa (IPK, masa studi, predikat kelulusan dan TOEFL)	Berdasasrkan dari aturan-aturan yang dihasilkan menunjukkan bahwa terdapat keterkaitan antara faktor-faktor yang mempengaruhi mahasiswa tepat waktu. Mahasiswa yang memiliki IPK tinggi dan nilai TOEFL tinggi memiliki kemungkinan besar untuk lulus tepat waktu dengan predikat kelulusan yang baik.	(Dengen et al., 2019)
6.	Penerapan Metode Asosiasi Data Mining Menggunakan Algoritma Apriori Untuk Mengetahui Kombinasi Antar Itemset Pada Pondok Kopi	Menemukan aturan asosiasi dari data transaksi penjualan di Pondok Kopi	Berdasarkan dari aturan-aturan yang dihasilkan menunjukkan terdapat pola pembelian konsumen di Pondok Kopi. Hal ini dapat mengembangkan manajemen inventori yang ada di Pondok Kopi dengan menyesuaikan pola pembelian konsumen.	(Nurchalifatun, 2017)
7.	Using Apriori Algorithm on Students' Performance Data for Association Rules Mining	Menemukan aturan asosiasi dari data nilai mahasiswa	Berdasarkan dari aturan yang dihasilkan dapat membuktikan bahwa metode ini dapat memberikan referensi bagi pengajar dan mahasiswa dalam penentuan mata kuliah yang tepat untuk diambilnya.	(Wu & Zeng, 2019)

Tinjauan yang diperoleh dari beberapa kajian pustaka pada

Tabel 2.3 di atas:

a. Analisis asosiasi menggunakan algoritma apriori digunakan secara luas dalam bidang pendidikan, tidak terbatas hanya dalam bidang penjualan seperti dasar contoh dari analisis asosiasi yaitu MBA (*market basket analysis*). Hal ini menunjukkan bagaimana fleksibilitas dan relevansi algoritma apriori dalam berbagai konteks
b. Penelitian analisis asosiasi pada penggunaan algoritma apriori difokuskan untuk menemukan sebuah aturan asosiasi, yang kemudian peneliti harus bisa menginterpretasikan apa yang ditemukan dari aturan-aturan asosiasi yang terbentuk. Dengan demikian, perlunya interpretasi konteks dalam penerapan algoritma apriori.
c. *Dataset* yang digunakan dalam penelitian analisis asosiasi tidak terbatas pada data kuantitatif saja, tetapi juga dapat mencakup data kualitatif. Hal ini menunjukkan bahwa algoritma apriori dapat di terapkan dalam berbagai jenis data untuk mengidentifikasi pola dan hubungan yang relevan
d. Beberapa penelitian menyoroti pentingnya penetapan parameter seperti nilai *minimum support* dan *confidence* dalam menemukan aturan asosiasi yang relevan. Penetapan parameter yang tepat dapat memengaruhi hasil analisis dan interpretasi aturan yang dihasilkan.

BAB III
METODOLOGI PENELITIAN

3.1 Subjek dan Objek penelitian

Menurut Suratnoaji (2016) subjek penelitian adalah sesuatu yang diteliti baik orang, benda ataupun lembaga/organisasi. Sederhananya subjek penelitian sesuatu hal yang akan dikenai simpulan hasil penelitian. Sedangkan objek penelitian merupakan sebuah suatu permasalahan yang ingin diteliti. Pemilihan subjek penelitian menjadi hal yang perlu diperhatikan karena subjek yang dipilih akan menjadi dasar untuk analisis dan pengambilan kesimpulan. Dalam konteks penelitian ini, subjek penelitian adalah data nilai akademik dari mahasiswa jurusan informatika UII angkatan 2020 yang status sebagai mahasiswa aktif dalam kurikulum 2020. Jumlahnya mencakup 150 mahasiswa yang dipilih sebagai sampel representatif dari populasi yang lebih besar. Subjek yang dipilih dinilai cukup relevan dengan tujuan penelitian beserta cakupan sampel yang baik diharapkan memberikan representasi yang cukup baik. Sementara itu, objek penelitian dalam penelitian ini adalah analisis asosiasi data nilai akademik mahasiswa yang ada pada subjek.

3.2 Sumber data penelitian

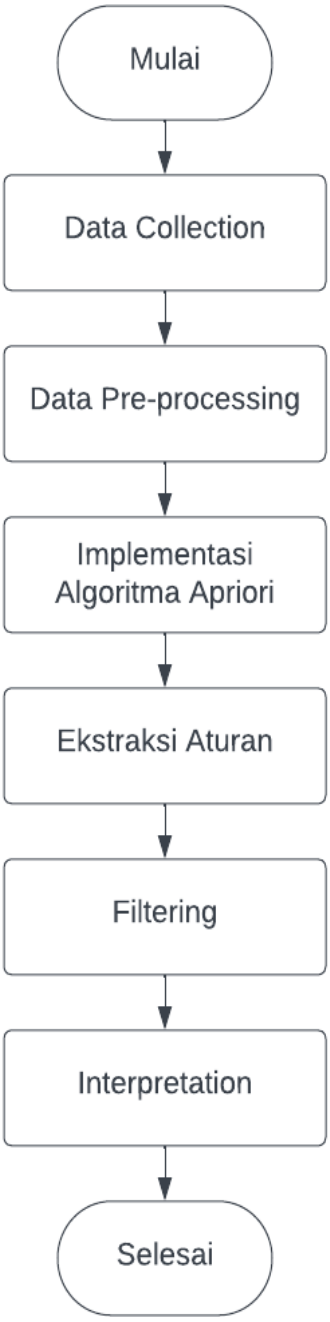
Sumber data merupakan landasan utama juga menjadi hal harus yang disiapkan pada tahap awal dari penelitian ini. Sumber data yang digunakan pada penelitian ini adalah data sekunder. Data sekunder ini diperoleh dari data akademik mahasiswa jurusan Informatika di UII. Data tersebut yang akan digunakan sebagai masukan dalam analisis asosiasi yang dilakukan dalam penelitian ini.

3.3 Metode pengumpulan data penelitian

Metode pengumpulan data pada penelitian ini dilakukan dengan metode pengambilan data secara langsung pada *database* nilai akademik mahasiswa yang diakses melalui dosen pembimbing penulis pada 09 Agustus 2023. Data tersebut akan diolah pada tahap *pre-processing* untuk membersihkan data dari *noise*, *missing value*, atau *outlier* yang mungkin ada. Kemudian, data akan disimpan dalam format .csv untuk memudahkan akses dan analisis selanjutnya. Proses pengumpulan data ini memastikan bahwa data yang digunakan dalam penelitian memiliki kualitas yang baik dan relevan dengan tujuan penelitian ini.

3.4 Alur penelitian

Alur penelitian disusun guna untuk memudahkan penulis dalam menyusun laporan penelitian. Adapun langkah-langkah penelitian ini dapat dilihat pada Gambar 3.1.



Gambar 3.1 Alur penelitian

Penelitian ini dilakukan dimulai dari data *collection*, data *pre-processing*, implementasi algoritma apriori, ekstraksi aturan, *filtering*, dan *interpretation*. Penjelasan tiap tahapan akan dilakukan pada Bab 3 ini.

3.5 Data *collection*

Data *collection* merupakan tahapan dimana penulis melakukan pengkoleksian data penelitian berupa nilai mahasiswa aktif jurusan Informatika UII angkatan 2020. Data nilai mahasiswa yang diambil sejumlah 150 mahasiswa. Data diambil secara langsung dari *database* akademik mahasiswa pada tanggal 09 Agustus 2023 melalui perantara dosen pembimbing tugas akhir. Setelah melakukan pengumpulan data, data – data nilai mahasiswa tersebut digabungkan dalam bentuk satu *dataset* nilai mahasiswa.

3.6 Data *pre-processing*

Pada tahap *pre-processing*, setiap data nilai mahasiswa dibersihkan sesuai kebutuhan penelitian. Penulis menetapkan untuk memproses data nilai mahasiswa pada 18 mata kuliah yang bersifat keinformatikaan dengan mengecualikan mata kuliah kewarganegaraan, keagamaan, dan bahasa. Hal ini dilakukan untuk meminimalisir kesalahan juga menimbang relevansi data terhadap tujuan penelitian. Proses *pre-processing* ini dilakukan di *software* Ms. Excel dan Google Sheets.

3.6.1 Pembentukan Dataset

Untuk memulai data *pre-processing* penulis memutuskan untuk melakukan pelabelan terhadap setiap data nilai yang mahasiswa yang akan diteliti. Prosesnya yaitu membuat kode mata kuliah baru yang ada pada kolom E. Kolom E ini merepresentasikan akronim dari mata kuliah digabungkan dengan semester mata kuliah tersebut diambil sebagai kode mata kuliah. Terlihat pada Gambar 3.2 pada kolom E menunjukkan gabungan dari semester mata kuliah (kolom D) dan singkatan mata kuliah (kolom F) dengan tujuan untuk memudahkan saat melakukan interpretasi data.

	A	B	C	D	E	F
1	Kode Matakuliah	Matakuliah	SMT	Nilai	MKSMT	
2	FTI000	Student Soft Skill Development	3	A	3S3D	S3D
3	SIF101	Pemikiran Desain	1	A	1PEMDES	PEMDES
4	SIF102	Fundamen Matematika	2	A-	2FUNMAT	FUNMAT
5	SIF103	Fundamen Informatika	1	A/B	1FUNDIF	FUNDIF
6	SIF104	Logika Pemrograman	1	C	1LOGPRO	LOGPRO
7	SIF201	Matematika Lanjut	3	A/B	3MATLAN	MATLAN
8	SIF202	Fundamen Pengembangan Aplikasi	2	A/B	2FPA	FPA
9	SIF203	Algoritme dan Struktur Data	1	B+	1ASD	ASD
10	SIF204	Rekayasa Perangkat Lunak	2	A/B	2RPL	RPL
11	SIF301	Sistem dan Jaringan Komputer	3	B-	3SJK	SJK
12	SIF302	Pengembangan Aplikasi Berbasis Web	3	B+	3PABW	PABW
13	SIF303	Manajemen Pengembangan Teknologi Informasi	5	A/B	5MPTI	MPTI
14	SIF401	Pengembangan Sistem Informasi	4	A/B	4PSI	PSI
15	SIF402	Sistem Cerdas dan Pendukung Keputusan	4	B	4SCPK	SCPK
16	SIF403	Fundamen Sains Data	5	B	5FSD	FSD
17	SIF404	Grafika dan Multimedia	4	A-	4GRAFMUL	GRAFMUL
18	SIF501	Bahasa Arab Alquran	3	B+	3BARAB	BARAB
19	SIF502	Pengembangan Aplikasi Bergerak	5	B+	5PAB	PAB
20	SIF503	Pengembangan Aplikasi Informatika Medis	5	B+	5PAIM	PAIM
21	SIF504	Pengembangan Gim	4	A	4GIM	GIM
22	SIF910	Deep Learning	6	A	6DL	DL
23	SIF913	Keamanan Siber dan Informasi	6	A/B	6KSI	KSI
24	SIF914	Forensika Digital	6	A-	6FORDIG	FORDIG
25	SIF915	Internet of Things	6	A	6IOT	IOT
26	UNI600	Pendidikan Agama Islam	1	B	1PAIM	PAIM
27	UNI601	Islam Ulil Albab	2	A	2ISLAMULIL	ISLAMULIL
28	UNI602	Islam Rahmatan lil 'Alamin	3	A	3ISLAMRA	ISLAMRA
29	UNI603	Pendidikan Pancasila	1	B+	1PANC	PANC

Gambar 3.2 Data nilai mahasiswa

Pada tahap ini, data diambil dari kolom D dan kolom E kemudian ditranspose ke *sheet* baru seperti yang terlihat pada Gambar 3.3 untuk melanjutkan proses selanjutnya. Proses ini akan dilakukan pada setiap data mahasiswa dari nomor induk mahasiswa (NIM) awal sampai akhir.

A	B	C	D	E	F	G	H	I	J	K
1ASD	1FUNDIF	1LOGPRO	1PAIM	1PANC	1PEMDES	2BIKI	2FPA	2FUNMAT	2ISLAMULIL	2RPL
B+	A/B	C	B	B+	A	B+	A/B	A-	A	A/B

Gambar 3.3 Hasil transpose dari salah satu data mahasiswa per-mata kuliah

Selanjutnya, data mahasiswa yang telah ditranspose digabungkan dalam satu *worksheet* untuk dijadikan sebagai satu *dataset* seperti terlihat pada Tabel 3.1. *Dataset* ini memuat seluruh data mahasiswa yang akan dianalisis, termasuk nilai untuk setiap mata kuliahnya dalam format .csv (*comma-separated values)*.

Tabel 3.1 Dataset nilai mahasiswa

NO.	1ASD	1FUNDIF	1PEMDES	2BIKI	2FPA	2FUNMAT	2RPL	3MATLAN	3PABW	4PSI	…
1	B+	A/B	A	B+	A/B	A-	A/B	A/B	B+	A/B	…
2	B+	A-	A	A	A-	A/B	A/B	A	A-	A/B	…
3	A/B	A	A/B	A/B	A-	A/B	A	A	A-	A-	…
4	A	A	A	A	A	A	A-	A	A	A	…
5	B+	A-	A	A-	A-	A	A/B	A	A/B	B+	…
6	B-	A/B	A	A	B+	A	A-	A	B+	B	…

7	A-	A-	A	A	B+	A	A/B	A	B	C	…
8	B+	A	A-	B	B+	B	A/B	A	A/B	A-	…
…	…	…	…	…	…	…	…	…	…	…	…
150	B+	A-	A-	A	B-	B+	C	A/B	C+	A/B	…

3.6.2 Transformasi data

Setelah *dataset* nilai mahasiswa terbentuk, langkah selanjutnya adalah melakukan transforimasi data. Transformasi ini bertujuan untuk mengubah data nilai per-mahasiswa menjadi bentuk format data tabular seperti pada Tabel 2.2 sesuai dengan kebutuhan dari algoritma apriori. Dalam proses transformasi ini, nilai-nilai dan kode mata kuliah akan dibentuk menjadi item tunggal (n) dalam bentuk format item.

{A.1ASD}

Gambar 3.4 Format item

Dalam format item "A.1ASD", huruf "A" sebelum titik mengindikasikan nilai yang diperoleh pada mata kuliah tersebut, sementara "1ASD" menyatakan bahwa "ASD" adalah singkatan dari nama mata kuliah "Algoritme Struktur Data", dan angka "1" mewakili semester di mana mata kuliah Algoritme Struktur Data tersebut diambil. Penyederhanaan nama mata kuliah mengacu dari referensi pada Tabel 3.2.

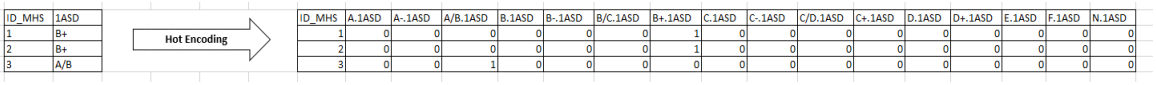
Tabel 3.2 Pengkodean mata kuliah

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

3.6.3 Penerapan hot-encoding

Dataset nilai mahasiswa yang masih dalam bentuk data transaksional tidak dapat diolah oleh algoritma apriori. Maka dilakukan transformasi data menjadi bentuk data tabular menggunakan *hot-encoding*. *Hot-encoding* adalah teknik yang digunakan untuk mengubah data kategorikal menjadi data *boolean*. *Hot-encoding* dilakukan dengan cara mendefinisikan setiap kategori sebagai varialel biner baru. Jika suatu data termasuk dalam kategori tersebut, maka variabel biner tersebut akan bernilai 1. Sebaliknya, jika tidak maka variabel biner tersebut akan bernilai 0. Proses *hot-encoding* dilakukan menggunakan fitur pivot pada *software* Google Sheets.



Gambar 3.5 Gambaran penerapan teknik *hot-encoding*

3.7 Implementasi algoritma apriori

Setelah proses *pre-processing* data, langkah berikutnya adalah menerapkan algoritma apriori dengan *tools* yang disediakan oleh library *Python*. Proses ini melibatkan serangkaian langkah yang perlu dilakukan untuk mendapatkan hasil yang akurat.

3.7.1 Penetapan batas minimum support

Proses awal dalam *data mining* untuk analisis asosiasi adalah melakukan pencarian *frequent itemsets*. *Frequent itemsets* adalah kandidat-kandidat dari item-item yang berpotensi saling memiliki pola atau ketertaitan satu dengan yang lainnya. Sebelum melakukan pencarian *frequent itemsets* peneliti harus menetapkan batas *minimum support* yang digunakan untuk memfilter seberapa besar item yang ingin dicari. Penetapan nilai *minimum support* didasarkan pada beberapa kondisi, yaitu:

a. Ukuran dataset

Jika dataset yang digunakan besar, maka batas nilai minimum support di-set lebih tinggi agar banyak item yang terfiltrasi untuk mencegah jumlah *frequent itemsets* yang mungkin terlalu banyak dan membebankan komputasi. Sebaliknya, jika dataset yang digunakan kecil,

maka nilai *minimum support* cenderung di-set rendah karena kemungkinan *frequent itemsets* yang muncul tidak terlalu banyak.

b. Tujuan analisis asosiasi

Jika menggunakan batas *minimum support* tinggi, maka banyak *itemset* yang terfiltrasi 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.

3.7.2 Pencarian frequent itemsets

Setelah menetapkan batas minimum *support*, *function* apriori pertama kali dijalankan untuk mencari *frequent itemsets,* yang akan menjadi kandidat *item* dan *itemsets* pada langkah selanjutnya dalam pembentukan aturan asosiasi (*association rules*). Hanya *itemset* yang memenuhi atau melebihi nilai *minimum support* yang dianggap sebagai *frequent itemsets*. *Frequent itemsets* yang diindentifikasi pada tahap ini menjadi dasar utama untuk proses selanjutnya, yaitu ekstraksi aturan asosiasi. Ilustrasi dari pencarian *frequent itemsets* oleh algoritma apriori terdapat pada Gambar 2.2.

3.8 Ekstraksi aturan

Setelah langkah pencarian *frequent itemsets,* tahap selanjutnya adalah ekstraksi aturan asosiasi. Dalam tahap ini, *frequent itemset* yang telah diidentifikasi akan menjadi acuan untuk menghasilkan aturan asosiasi yang relevan. Proses ekstraksi ini bertujuan untuk mengungkapkan korelasi dan pola hubungan antar-item, memungkinkan pemahaman lebih dalam pada *dataset*. Aturan asosiasi yang dihasilkan dapat memberikan wawasan tentang keterkaitan dan dependensi antar-item, sekaligus mendukung pemahaman yang lebih baik terhadap pola-pola yang ada.

3.9 Filtering

Tahap *Filtering* merupakan proses kritis dalam analisis asosiasi, di mana aturan-aturan yang dihasilkan melalui algoritma apriori disaring untuk mengekstrak hanya aturan-aturan yang memiliki nilai dibutuhkan, sesuai dengan parameter-parameter yang ditentukan sebelumnya. *Filtering* dilakukan berdasarkan metrik kualitas aturan, seperti *support*,

confidence, dan *lift*. Aturan yang tidak memenuhi nilai ambang batas yang telah ditetapkan akan dieliminasi, meninggalkan hanya aturan-aturan yang relevan dan bermakna. Tahapan ini memastikan bahwa aturan asosiasi yang diekstraksi memiliki kualitas tinggi dan dapat memberikan wawasan yang berarti terhadap hubungan antar-item dalam *dataset*.

3.10 *Interpretation*

Interpretation menjadi tahap penting setelah proses *filtering* selesai. Pada tahap ini, aturan-aturan asosiasi yang telah terpilih dianalisis secara mendalam untuk menggali makna dan relevansinya. Penggunaan wawasan yang diperoleh dari aturan-aturan ini diarahkan untuk mendukung pemahaman terhadap pola dan hubungan antar-item dalam konteks penelitian. *Interpretation* hasil ini juga memainkan peran kunci dalam merumuskan implikasi praktis dan strategi tindak lanjut, yang dapat berguna bagi *stakeholder* terkait. Dengan demikian, bab *Interpretation* menjadi bagian krusial dalam membawa penelitian ini menuju pencapaian tujuan penelitian.

BAB IV
HASIL DAN PEMBAHASAN

4.1 Dataset

Data yang dikumpulkan dari *database* akademik mahasiswa akan diubah dalam bentuk *dataset*, yang mencakup nilai setiap mahasiswa pada setiap mata kuliah yang telah diambilnya dalam format csv. Secara sekilas, struktur *dataset* dapat dilihat dalam Tabel 3.1.

4.2 Pencarian frequent itemsets

Tahapan awal penelitian, data diimpor dari penyimpanan lokal ke Google Colab untuk penerapan algoritma apriori. Data yang digunakan terdiri dari data nilai 150 mahasiswa pada 18 mata kuliah wajib, diambil dari *database* akademik mahasiswa pada bulan Agustus 2023. Implementasi algoritma apriori melibatkan pencarian *frequent itemsets* yang potensial menjadi kandidat *itemset*.

4.2.1 Implementasi algoritma apriori

Variabel frequent_itemsets akan menjalankan sebuah *function* untuk melakukan pencarian *frequent itemsets* yang ada pada *data frame*. *Data frame* adalah struktur data tabel dua dimensi yang menyimpan nilai data. “df2” adalah *data frame* yang menyimpan nilai *dataset* yang telah dipersiapkan sebelumnya. Selanjutnya, 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. Dengan nilai *support* 0,15, diharapkan mengidentifikasi lebih banyak kemungkinan *itemsets* yang muncul dibandingkan menggunakan batas nilai yang tinggi.

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

Gambar 4.1 Pencarian Frequent Itemsets

Dengan menggunakan batas nilai *support* 0,15 berarti melibatkan sebanyak mungkin *itemsets* yang dapat membantu penulis 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. Hasil pencarian *frequent itemsets* tertera pada Tabel 4.1.

4.2.2 Hasil pencarian frequent itemsets

Tabel 4.1 Hasil pencarian frequent itemsets

index	support	itemsets
0	0.153	(A/B.1ASD)
1	0.273	(B+.1ASD)
2	0.53	(A.1FUNDIF)
3	0.2	(A-.1FUNDIF)
4	0.16	(B.1LOGPRO)
...	...	...
111	0.153	(A.5MPTI, A.3MATLAN, A.3SJK, A.1FUNDIF)
112	0.173	(A.3MATLAN, A.4GIM, A.1FUNDIF, A.4GRAFMUL)
113	0.16	(A.5MPTI, A.3MATLAN, A.1FUNDIF, A.4GRAFMUL)
114	0.167	(A.3MATLAN, A.1PEMDES, A.4GIM, A.4GRAFMUL)
115	0.153	(A.5MPTI, A.3MATLAN, A.1PEMDES, A.4GRAFMUL)

Tertera pada Tabel 4.1 *frequent itemsets* yang muncul sebanyak 116 kandidat dari 1-*itemsets* hingga 4-*itemsets*. Kandidat *frequent itemsets* yang muncul mencerminkan kombinasi nilai mata kuliah yang sering muncul bersama dalam data mahasiswa. Panjang *itemset* yang mencapai 4-*itemsets* menunjukkan bahwa penelitian ini tidak hanya berfokus pada asosiasi antara dua mata kuliah, tetapi juga mencari pola yang lebih kompleks melibatkan sejumlah mata kuliah dalam satu *itemset*. 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 membuat *filtering* pada parameter utama yaitu nilai *support*, *confidence*, dan *lift*. Tujuannya adalah penulis dapat memastikan bahwa setiap temuannya terdapat korelasi antar item yang ada.

4.3 Filtering aturan asosiasi

Proses *filtering* (filtrasi) aturan asosiasi mengacu pada pembentukan aturan asosiasi dari hasil pencarian *frequent itemsets* yang telah teridentifikasi pada langkah sebelumnya dengan filtrasi 3 parameter utama yaitu, *confidence, support,* dan *lift*. *Itemsets* ini diubah menjadi aturan asosiasi yang mengandung informasi tentang korelasi antar item yang muncul dalam *dataset*. Filtrasi pertama dimulai dari pembentukan aturan asosiasi dengan batas nilai *lift* lebih dari 1.

4.3.1 Pembentukkan aturan asosiasi

```
rules = association_rules(frequent_itemsets, metric="lift", min_threshold = 1)
rules
```

Gambar 4.2 Pembentukan aturan asosiasi

Pembentukan aturan asosisasi dilakukan terhadap seluruh kandidat *frequent itemsets* yang ada pada Tabel 4.1, dengan filtrasi nilai lift *minimum* yang ditetapkan adalah 1. Rentang nilai *lift* bervariasi dari 0 hingga tak terbatas. Semakin tinggi nilai *lift* pada suatu aturan, semakin besar indikasinya akan kebermanfaatan aturan tersebut (Ulvah, 2018). Hal ini mendorong penulis untuk melakukan filtrasi awal terhadap nilai *lift* dari kandidat *frequent itemsets*, dengan tujuan untuk mendapatkan aturan yang lebih bermakna.

Tabel 4.2 Hasil pembentukan aturan asosiasi

id	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
0	(A.1PEMDES)	(B+.1ASD)	0.580000	0.273333	0.160000	0.275862	1.009251
1	(B+.1ASD)	(A.1PEMDES)	0.273333	0.580000	0.160000	0.585366	1.009251
2	(A.3MATLAN)	(B+.1ASD)	0.660000	0.273333	0.200000	0.303030	1.108647
3	(B+.1ASD)	(A.3MATLAN)	0.273333	0.660000	0.200000	0.731707	1.108647
4	(A.1PEMDES)	(A.1FUNDIF)	0.580000	0.533333	0.333333	0.574713	1.077586
...	...	...	...	...	...	...	...
401	(A.1PEMDES, A.4GRAFMUL)	(A.5MPTI, A.3MATLAN)	0.320000	0.253333	0.153333	0.479167	1.891447
402	(A.5MPTI)	(A.3MATLAN, A.1PEMDES, A.4GRAFMUL)	0.293333	0.253333	0.153333	0.522727	2.063397
403	(A.3MATLAN)	(A.5MPTI, A.1PEMDES, A.4GRAFMUL)	0.660000	0.173333	0.153333	0.232323	1.340326
404	(A.1PEMDES)	(A.5MPTI, A.3MATLAN, A.4GRAFMUL)	0.580000	0.186667	0.153333	0.264368	1.416256
405	(A.4GRAFMUL)	(A.5MPTI, A.3MATLAN, A.1PEMDES)	0.460000	0.186667	0.153333	0.333333	1.785714

Tabel 4.2 mencatat 406 baris aturan asosiasi yang berisi aturan dengan nilai *lift* lebih dari 1 dan aturan-aturan yang muncul sudah dipastikan memiliki kebermanfaatan. Kemudian,

penulis akan melakukan filtrasi dengan memfokuskan nilai *confidence* yang tinggi dan nilai *support* yang cukup dikarenakan begitu banyaknya aturan yang muncul. Oleh karena itu, dilakukan pencarian kembali aturan asosiasi dengan parameter *support* dan *confidence* yang bervariasi.

4.3.2 Filtrasi hasil pencarian aturan asosiasi pertama

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

Gambar 4.3 Filtrasi pertama

Pada Gambar 4.3 pencarian dilakukan dengan *minimum* nilai *support* 0,2 dan *minimum* nilai *confidence* 0,2 yang menghasilkan *association rules* yang terdapat pada Tabel 4.3. Batas nilai yang ditetapkan oleh penulis berdasarkan penelitian yang dilakukan oleh (Wu & Zeng, 2019). Pada penelitiannya Wu dan Zeng menetapkan batasan *minimum support* 0,06 dan *minimum confidence* 0,5. Hal ini dilakukan agar *rules-rules* memiliki potensi bermakna tidak terfilter, maka dilakukan penetapan batasan dimulai dari nilai yang kecil.

Tabel 4.3 Hasil filtrasi pertama

id	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
2	(A.3MATLAN)	(B+.1ASD)	0.660000	0.273333	0.200000	0.303030	1.108647
3	(B+.1ASD)	(A.3MATLAN)	0.273333	0.660000	0.200000	0.731707	1.108647
4	(A.1PEMDES)	(A.1FUNDIF)	0.580000	0.533333	0.333333	0.574713	1.077586
5	(A.1FUNDIF)	(A.1PEMDES)	0.533333	0.580000	0.333333	0.625000	1.077586
6	(A.2RPL)	(A.1FUNDIF)	0.286667	0.533333	0.226667	0.790698	1.482558
...	...	...	...	...	...	...	...
303	(A.1FUNDIF, A.4GIM)	(A.3MATLAN, A.1PEMDES)	0.293333	0.426667	0.200000	0.681818	1.598011
304	(A.3MATLAN)	(A.1PEMDES, A.1FUNDIF, A.4GIM)	0.660000	0.226667	0.200000	0.303030	1.336898
305	(A.1PEMDES)	(A.3MATLAN, A.1FUNDIF, A.4GIM)	0.580000	0.260000	0.200000	0.344828	1.326260
306	(A.1FUNDIF)	(A.3MATLAN, A.1PEMDES, A.4GIM)	0.533333	0.280000	0.200000	0.375000	1.339286

307	(A.4GIM)	(A.3MATLAN, A.1PEMDES, A.1FUNDIF)	0.513333	0.273333	0.200000	0.389610	1.425404

Hasilnya sebanyak 154 pada Tabel 4.3 menunjukkan bahwa *rules* yang muncul masih terlalu banyak. Secara keseluruhan, banyak *rules* yang memiliki nilai 0,2 dengan nilai *confidence* yang cukup tinggi. Contohnya, pada index ketiga terlihat *rules* " *if* (B+.ASD1) *then* (A.3MATLAN)" dengan nilai *support* 0,2 dan *confidence* 0,731707. Berdasarkan hal ini, penulis menyimpulkan bahwa meskipun tedapat *rules* dengan nilai *support* yang cukup baik dan *confidence* yang tinggi, masih terdapat potensi *rules* lain yang lebih tinggi dengan mempertahankan filter nilai support 0.2. Karena hasil yang diinginkan adalah menemukan korelasi item yang tinggi, penulis melakukan filtrasi kedua nilai *minimum confidence* dan nilai *minimum support* yang lebih tinggi.

4.3.3 Filtrasi hasil pencarian aturan asosiasi kedua

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

Gambar 4.4 Filtrasi kedua

Pada proses filtrasi kedua, penulis menetapkan parameter nilai *support* dan *confidence* di angka 0,3. Hasilnya akan disortir berdasarkan nilai *confidence* tertinggi hingga terendah karena fokus dalam penelitian ini yang ingin mencari aturan dengan korelasi antar item yang cukup kuat. Hasil filtrasi kedua dapat dilihat pada Tabel 4.4. Nilai *confidence* mengukur seberapa tinggi kemungkinan ketika suatu item muncul dengan item lainnya yang cenderung muncul secara bersamaan. Hal ini selaras dengan tujuan penulis yang ingin mencari item yang cenderung muncul bersamaan dan menunjukkan adanya korelasi pada *item-item* tersebut.

Tabel 4.4 Hasil filtrasi kedua

id	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
9	(A.1FUNDIF)	(A.3MATLAN)	0.533333	0.660000	0.420000	0.787500	1.193
55	(A.4GRAFMUL)	(A.3MATLAN)	0.460000	0.660000	0.346667	0.753623	1.141
25	(A.1PEMDES)	(A.3MATLAN)	0.580000	0.660000	0.426667	0.735632	1.114
50	(A.3SJK)	(A.3MATLAN)	0.466667	0.660000	0.340000	0.728571	1.103
29	(A.4GIM)	(A.1PEMDES)	0.513333	0.580000	0.366667	0.714286	1.231

53	(A.4GIM)	(A.3MATLAN)	0.513333	0.660000	0.366667	0.714286	1.082
31	(A.4GRAFMUL)	(A.1PEMDES)	0.460000	0.580000	0.320000	0.695652	1.199
15	(A.4GRAFMUL)	(A.1FUNDIF)	0.460000	0.533333	0.313333	0.681159	1.277
71	(A.4GRAFMUL)	(A.4GIM)	0.460000	0.513333	0.300000	0.652174	1.270
24	(A.3MATLAN)	(A.1PEMDES)	0.660000	0.580000	0.426667	0.646465	1.114
8	(A.3MATLAN)	(A.1FUNDIF)	0.660000	0.533333	0.420000	0.636364	1.193
28	(A.1PEMDES)	(A.4GIM)	0.580000	0.513333	0.366667	0.632184	1.231
5	(A.1FUNDIF)	(A.1PEMDES)	0.533333	0.580000	0.333333	0.625000	1.077
14	(A.1FUNDIF)	(A.4GRAFMUL)	0.533333	0.460000	0.313333	0.587500	1.277
70	(A.4GIM)	(A.4GRAFMUL)	0.513333	0.460000	0.300000	0.584416	1.270
4	(A.1PEMDES)	(A.1FUNDIF)	0.580000	0.533333	0.333333	0.574713	1.077
52	(A.3MATLAN)	(A.4GIM)	0.660000	0.513333	0.366667	0.555556	1.082
30	(A.1PEMDES)	(A.4GRAFMUL)	0.580000	0.460000	0.320000	0.551724	1.199
54	(A.3MATLAN)	(A.4GRAFMUL)	0.660000	0.460000	0.346667	0.525253	1.141
51	(A.3MATLAN)	(A.3SJK)	0.660000	0.466667	0.340000	0.515152	1.103
9	(A.1FUNDIF)	(A.3MATLAN)	0.533333	0.660000	0.420000	0.787500	1.193

Secara kesluruhan, terdapat 20 aturan yang melampaui batas yang diterapkan pada proses filtrasi kedua ini. Aturan yang memiliki nilai *confidence* tertinggi terletak pada indeks kesembilan yaitu, " *if* (A.1FUNDIF) *then* (A.3MATLAN) " dengan nilai *support* sebesar 0,42 dan nilai *confidence* sebesar 0,78. Aturan tersebut dapat diartikan bahwa sebanyak 42% dari kejadian dalam *dataset* melibatkan mahasiswa yang memperoleh nilai A pada mata kuliah FUNDIF di semester 1 akan memperoleh A pada MATLAN di semester 3 dengan nilai *confidence* sebesar 78%. Kejadian ini tidak terjadi secara kebetulan, sebagaimana ditunjukkan oleh nilai *lift* yang lebih dari 1, yaitu, 1.19.

Secara komprehensif, tabel ini terlihat tidak ada aturan dengan nilai *confidence* yang lebih tinggi dari 0.7 yang dengan batasan nilai *support* di atas 0,3 sampai 0,4. Oleh karena itu, dilakukan filtrasi selanjutnya dengan menetapkan batas nilai *support* sebesar 0,2 dan nilai *confidence* sebesar 0.4.

4.3.4 Filtrasi hasil pencarian aturan asosiasi ketiga

```
result = rules[(rules['support'] >= 0.2) &
               (rules['confidence'] >= 0.4)]
```

Gambar 4.5 Filtrasi ketiga

Sebanyak 126 aturan ditemukan pada filtrasi ketiga dengan batas *minimum support* sebesar 0,2 dan *minimum confidence* sebesar 0,4. Secara sekilas, banyak aturan yang berisi *item*

(A.3MATLAN) sebagai *consequents*, hal ini kemungkinan terjadi dikarenakan nilai *support* dari *item* (A.3MATLAN) yang cukup tinggi di dalam *dataset*.

Tabel 4.5 Hasil Filtrasi ketiga

id	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
128	(A.1FUNDIF, A.4GIM)	(A.3MATLAN)	0.293333	0.66000	0.26000	0.886363	1.3429
139	(A.5MPTI, A.1FUNDIF)	(A.3MATLAN)	0.233334	0.66000	0.20667	0.885714	1.3419
297	(A.1PEMDES, A.1FUNDIF, A.4GIM)	(A.3MATLAN)	0.226666	0.66000	0.2000	0.882352	1.3368
60	(A.5MPTI)	(A.3MATLAN)	0.293333	0.66000	0.25335	0.863636	1.3085
134	(A.1FUNDIF, A.4GRAFMUL)	(A.3MATLAN)	0.313335	0.66000	0.26666	0.829787	1.2572
86	(A.1PEMDES, A.1FUNDIF)	(A.3MATLAN)	0.333333	0.66000	0.27333	0.820000	1.2424
138	(A.5MPTI, A.3MATLAN)	(A.1FUNDIF)	0.253333	0.53333	0.20667	0.815789	1.5296
253	(A.3SJK, A.4GRAFMUL)	(A.3MATLAN)	0.253335	0.66000	0.20667	0.815789	1.2360
193	(A.3SJK, A.1PEMDES)	(A.3MATLAN)	0.28	0.66000	0.22666	0.809523	1.2265
…	…	…	…	…	…	…	…
7	(A.1FUNDIF)	(A.2RPL)	0.533333	0.28667	0.22666	0.425000	1.4825
100	(A.1FUNDIF)	(A.1PEMDES, A.4GIM)	0.533333	0.36664	0.22666	0.425000	1.1590
201	(A.3MATLAN)	(A.1PEMDES, A.4GIM)	0.66	0.36664	0.28000	0.424242	1.1570
232	(A.4GIM)	(A.1PEMDES, A.4GRAFMUL)	0.513333	0.32000	0.21335	0.415584	1.2987
87	(A.3MATLAN)	(A.1PEMDES, A.1FUNDIF)	0.66	0.33333	0.2733	0.414141	1.2424
106	(A.1FUNDIF)	(A.1PEMDES, A.4GRAFMUL)	0.533333	0.32000	0.22000	0.412503	1.2890
128	(A.1FUNDIF, A.4GIM)	(A.3MATLAN)	0.293333	0.66000	0.26666	0.886363	1.3429

Berdasarkan Tabel 4.5 dapat ditinjau pada aturan indeks ke-128 memiliki nilai *confidence* tertinggi dibandingkan aturan-aturan yang ada pada tabel-tabel sebelumnya. Aturan tersebut adalah " *if* (A.1FUNDIF, A.4GIM) *then* (A.MATLAN) " dengan nilai *support* sebesar 0,26 dan nilai *confidence* sebesar 0,886.

4.4 *Interpretation*

Dari hasil penerapan algoritma apriori terhadap 150 data nilai mahasiswa dengan beberapa langkah filtrasi, didapatkan beberapa aturan asosiasi data nilai mahasiswa, berikut adalah beberapa interpretasi untuk setiap aturan yang ada.

a. Pada proses filtrasi pertama terhadap aturan asosiasi yang ada pada Tabel 4.3 diambil contoh 2 aturan pada indeks ketiga dan kelima.
    1. Indeks ketiga " *if* (B+.1ASD) *then* (A.3MATLAN) " : Aturan tersebut dapat diartikan bahwa sebanyak 20% dari kejadian dalam *dataset* mahasiswa yang memperoleh nilai A pada mata kuliah PEMDES di semester 1 akan memperoleh nilai B+ pada ASD di semester 1 dengan nilai *confidence* sebesar 73%. Kejadian ini tidak terjadi secara kebetulan, sebagaimana ditunjukkan oleh nilai *lift* yang lebih dari 1, yaitu, 1,1.
    2. Indeks kelima " *if* (A.1FUNDIF) *then* (A.1PEMDES) " : Aturan tersebut dapat diartikan bahwa sebanyak 33% dari kejadian dalam *dataset* menunjukkan mahasiswa yang memperoleh nilai A pada mata kuliah FUNDIF di semester 1 cenderung memperoleh nilai A pada mata kuliah PEMDES di semester 1 dengan nilai *confidence* sebesar 62,5%. Kejadian ini tidak terjadi secara kebetulan, sebagaimana ditunjukkan oleh nilai *lift* yang lebih dari 1, yaitu, 1,07.

b. Pada proses filtrasi kedua terhadap aturan asosiasi yang ada pada Tabel 4.4 diambil contoh 3 aturan pada indeks kesembilan, ke-25, dan ke-50.
    1. Indeks kesembilan " *if* (A.1FUNDIF) *then* (A.3MATLAN) " : Aturan tersebut dapat diartikan bahwa sebanyak 42% dari kejadian dalam *dataset* menunjukkan mahasiswa yang memperoleh nilai A pada mata kuliah FUNDIF di semester 1 akan memperoleh nilai A pada mata kuliah MATLAN di semester 3 dengan nilai *confidence* sebesar 78,7%. Kejadian ini tidak terjadi secara kebetulan, sebagaimana ditunjukkan oleh nilai *lift* yang lebih dari 1, yaitu, 1,193.
    2. Indeks ke-25 " *if* (A.1PEMDES) *then* (A.3MATLAN) " : Aturan tersebut dapat diartikan bahwa sebanyak 42,67% dari kejadian dalam *dataset* menunjukkan

mahasiswa yang memperoleh nilai A pada mata kuliah PEMDES di semester 1 akan memperoleh nilai A pada mata kuliah MATLAN di semester 3 dengan nilai *confidence* sebesar 73,5%. Kejadian ini tidak terjadi secara kebetulan, sebagaimana ditunjukkan oleh nilai *lift* yang lebih dari 1, yaitu, 1,114.

3. Indeks ke-50 " *if* (A.3SJK) *then* (A.3MATLAN) " : Aturan tersebut dapat diartikan bahwa sebanyak 34% dari kejadian dalam *dataset* menunjukkan mahasiswa yang memperoleh nilai A pada mata kuliah SJK di semester 3 cenderung memperoleh nilai A pada mata kuliah MATLAN di semester 3 dengan nilai *confidence* sebesar 72,8%. Kejadian ini tidak terjadi secara kebetulan, sebagaimana ditunjukkan oleh nilai *lift* yang lebih dari 1, yaitu, 1,103.

c. Pada proses filtrasi ketiga terhadap aturan asosiasi yang ada pada Tabel 4.5 diambil contoh 3 aturan pada indeks ke-128, ketujuh, dan ke-106.
   1. Indeks ke-128 " *if* (A.1FUNDIF, A.4GIM) *then* (A.3MATLAN) " : Aturan tersebut dapat diartikan bahwa sebanyak 26% dari kejadian dalam *dataset* menunjukkan mahasiswa yang memperoleh nilai A pada mata kuliah FUNDIF di semester 1 dan nilai A pada mata kuliah GIM di semester 4 cenderung memperoleh nilai A pada mata kuliah MATLAN di semester 3 dengan nilai *confidence* sebesar 88,63%. Kejadian ini tidak terjadi secara kebetulan, sebagaimana ditunjukkan oleh nilai *lift* yang lebih dari 1, yaitu, 1,34.
   2. Indeks ketujuh " *if* (A.1FUNDIF) *then* (A.2RPL) " : Aturan tersebut dapat diartikan bahwa sebanyak 22,6% dari kejadian dalam *dataset* menunjukkan mahasiswa yang memperoleh nilai A pada mata kuliah FUNDIF di semester 1 akan memperoleh nilai A pada mata kuliah RPL di semester 2 dengan nilai *confidence* 42,5%. Kejadian ini tidak terjadi secara kebetulan, sebagaimana ditunjukkan oleh nilai *lift* yang lebih dari 1, yaitu, 1,48.
   3. Indeks ke-106 " *if* (A.1FUNDIF) *then* (A.1PEMDES, A.4GRAFMUL) " : Aturan tersebut dapat diartikan bahwa sebanyak 22% dari kejadian dalam *dataset* menunjukkan mahasiswa yang memperoleh nilai A pada mata kuliah FUNDIF di semester 1 akan memperoleh nilai A pada mata kuliah PEMDES di semester 1 dan memperoleh nilai A pada mata kuliah GRAFMUL di semester 4 dengan nilai *confidence* 41,25%. Kejadian ini tidak terjadi secara kebetulan, sebagaimana ditunjukkan oleh nilai *lift* yang lebih dari 1, yaitu, 1,28.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Analisis asosiasi menggunakan algoritma apriori dapat diterapkan untuk untuk mencari korelasi antara nilai mata kuliah mahasiswa pada data 150 mahasiswa yang mengambil mata kuliah wajib. Penerapan algoritma apriori dilakukan untuk membentuk aturan asosiasi dari *frequent itemsets* dimana aturan ini akan menjadi gambaran korelasi dari item-item yang ada pada *dataset* nilai mahasiswa. Setelah ditemukannya aturan-aturan asosiasi kemudian dilakukan filtrasi untuk memilah aturan dengan nilai *confidence*, *support* dan *lift* yang cukup baik. Dari sekian banyak aturan yang ditemukan, terdapat satu aturan yang menonjol yaitu, Indeks ke-128 " *if* (A.1FUNDIF, A.4GIM) *then* (A.3MATLAN) ". Aturan tersebut mengindikasikan bahwa sebanyak 26% dari kejadian dalam *dataset* menunjukkan mahasiswa yang memperoleh nilai A pada mata kuliah FUNDIF di semester 1 dan nilai A pada mata kuliah GIM di semester 4 cendering mendapat nilai A pada mata kuliah MATLAN di semester 3 dengan nilai *confidence* sebesar 88,63%. Temuan ini memiliki signifikansi yang tinggi, mengingat nilai *lift* yang melebihi 1, yaitu 1,34.

Melihat hal itu penting bagi mahasiswa yang ingin menekuni bidang pengembangan gim dengan memperhatikan performanya pada ketiga mata kuliah tersebut. Selain itu, penelitian ini juga menghasilkan sejumlah wawasan yang dapat menjadi pertimbangan penting dalam pengembangan sistem evaluasi akademik dan pembelajaran di institusi pendidikan. Secara keseluruhan, banyak mata kuliah MATLAN muncul di berbagai aturan, dikarenakan pada *dataset* tersebut banyak mahasiswa yang mendapatkan nilai A.

5.2 Saran

Terdapat beberapa saran yang dapat diberikan dari hasil penulisan tugas akhir ini, dalam rangka untuk pengembangan penelitian yang lebih lanjut sebagai berikut:

a. Mengembangkan model yang lebih kompleks dengan menambah variabel-variabel tambahan seperti asal sekolah mahasiswa (missal: SMA/SMK), IPK yang diperoleh, maupun lingkungan belajar.
b. Memperluas jumlah *dataset* dikarenakan semakin banyak jumlah sampel yang dimiliki semakin akurat juga hasil yang didapatkan.

c. Eksplorasi lebih lanjut terhadap variasi nilai *threshold* untuk mendapatkan aturan asosiasi yang lebih lanjut.

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.

DAFTAR PUSTAKA

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. *Proc. 20th Int. Conf. Very Large Data Bases, VLDB*, *1215*, 487–499.

Agrawal, R., & Srikant, R. (1995). Mining sequential patterns. *Proceedings of the Eleventh International Conference on Data Engineering*, 3–14.

Anggraeni, E. Y. (2017). *Pengantar sistem informasi*. Penerbit Andi.

Badrul, M., Studi, P., & Informasi, S. (2016). Algoritma Asosiasi Dengan Algoritma Apriori Untuk Analisa Data Penjualan. *Jurnal Pilar Nusa Mandiri*, *XII*(ALGORITMA ASOSIASI DENGAN ALGORITMA APRIORI UNTUK ANALISA DATA PENJUALAN), 121–129.

Cahyadi., M. D. P. A., & Purwanto. Tarjoko. (2021). Pengaruh Ketinggian Tempat Terhadap Sifat Fisiologi Dan Hasil Kopi Arabika (Coffea Arabica) Di Dataran Tinggi Desa Sarwodadi Kecamatan Pejawaran Kabupaten Banjarnegara. *Jurnal Ilmiah Media Agrosains*, *7*(1), 1–7.

Cahyanti, M., Mujahidin, M., & ... (2017). Penerapan algoritma Apriori association rule untuk analisa nilai mahasiswa di Universitas Gunadarma. *Semnasteknomedia …*, 19–24. https://ojs.amikom.ac.id/index.php/semnasteknomedia/article/view/1688%0Ahttps://ojs.amikom.ac.id/index.php/semnasteknomedia/article/viewFile/1688/1641

Chee, C. H., Jaafar, J., Aziz, I. A., Hasan, M. H., & Yeoh, W. (2019). Algorithms for frequent itemset mining: a literature review. *Artificial Intelligence Review*, *52*(4), 2603–2621. https://doi.org/10.1007/s10462-018-9629-z

Dengen, C. N., Kusrini, K., & Luthfi, E. T. (2019). Penentuan Association Rule Pada Kelulusan Mahasiswa Menggunakan Algoritma Apriori. *Jurnal Rekayasa Teknologi Informasi (JURTI)*, *3*(1), 20. https://doi.org/10.30872/jurti.v3i1.2256

Despitaria, Sujaini, H., & Tursina. (2016). Analisis Asosiasi pada Transaksi Obat Menggunakan Data Mining dengan Algoritma A Priori. *Jurnal Sistem Dan Teknologi Informasi (JUSTIN)*, *1*(1), 1–6.

Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concept and Techniques, Third Edition. *Waltham: Morgan Kaufmann Publishers*.

Hu, Z. (2017). Application and Research of Association rules and Decision tree Combination Algorithm in Student Achievement Analysis [D]. *Place of Preservation: Qingdao University of Technology*.

Hui, H., Ming-Jie, T., Qing-Tao, Z., & Xiao-Liang, Z. (2020). Application of Student achievement Analysis based on Apriori Algorithm. *Proceedings - 2020 2nd International Conference on Information Technology and Computer Application, ITCA 2020*, 19–22. https://doi.org/10.1109/ITCA52113.2020.00011

Jamaris, M., Asnal, H., & Wijaya, Y. S. (2020). Penerapan Algorithma Apriori Untuk Menemukan Pola Pemilihan Konsentrasi Studi Mahasiswa. *Al Ulum Jurnal Sains Dan Teknologi*, *5*(2), 68. https://doi.org/10.31602/ajst.v5i2.2882

Latifah, V. N., Furqon, M. T., & Santoso, N. (2018). Implementasi Algoritme Modified-Apriori Untuk Menentukan Pola Penjualan Sebagai Strategi Penempatan Barang Dan Promo. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *2*(10), 3829–3834. http://j-ptiik.ub.ac.id

Lo, R., Yunanto, A. E., Movia, R. N., Aditama, L., Wangsa, F., Lidjaja, N. A., & Ningsih, R. Y. (2023). *Penggunaan Bahasa Pemrograman Python dalam Menganalisis Hubungan Kualitas Kopi dengan Lokasi Pertanian Kopi*. *2*(2), 100–109.

Lubis, A. (2016). Basis Data Dasar. In *Deepublish*. Deepublish.

Nindani, A. A. (2017). *Analisis Association Rules Menggunakan Algoritma Apriori dan Algoritma ECLAT Pada Data Hasil Tangkapan Ikan Laut (Studi Kasus: Hasil Tangkapan Ikan Laut di …*. https://dspace.uii.ac.id/handle/123456789/27868

Nurchalifatun, F. (2017). Penerapan Metode Asosiasi Data Mining Menggunakan Algoritma Apriori Untuk Mengetahui Kombinasi Antar Itemset Pada Pondok Kopi. *Data Mining*.

Prodi Statistika Universitas Islam Indonesia, P. S. U. I. I. (2017). Buku Panduan Akademik. In *Universitas Islam Indonesia*.

Puspaningrum, A. S., Firdaus, F., Ahmad, I., & Anggono, H. (2020). Perancangan Alat Deteksi Kebocoran Gas Pada Perangkat Mobile Android Dengan Sensor Mq-2. *Jurnal Teknologi Dan Sistem Tertanam*, *1*(1), 1. https://doi.org/10.33365/jtst.v1i1.714

Rahmi, A. N., & Mikola, Y. A. (2021). Implementasi Algoritma Apriori Untuk Menentukan Pola Pembelian Pada Customer (Studi Kasus : Toko Bakoel Sembako). *Information System Journal*, *4*(1), 14–19. https://jurnal.amikom.ac.id/index.php/infos/article/view/561

Rino Subekti, Astried Silvanie, & Boy Firmansyah. (2023). Aplikasi Data Mining Untuk Menganalisis Aturan Asosiasi Tracer Study Di Ibi-Kosgoro1957. *Jurnal Disprotek*, *13*(2), 137–145. https://doi.org/10.34001/jdpt.v12i2

Santosa, B. (2007). Data mining teknik pemanfaatan data untuk keperluan bisnis. *Yogyakarta:*

Graha Ilmu, *978*(979), 756.

Suratnoaji, C. (2016). *Riset Komunikasi: Strategi praktis bagi peneliti pemula: Paradigma Penelitian dalam bidang Komunikasi*. Pusat Kajian Komunikasi Publik. Prodi Ilmu Komunikasi FISIP-UTM.

Syaifullah, M. A. (2010). Implementasi Data Mining Algoritma Apriori Pada Sistem Penjualan STMIK Amikom Yogyakarta. *Yogyakarta: Sekolah Tinggi Manajemen Informatika Dan Komputer Amikom*.

Ulvah. (2018). Implementasi Algoritma Apriori Aturan Keterkaitan Data Untuk Analisa Keranjang Belanja Sistem Persediaan Obat Pada Apotek Perdos Farma Makassar. *Jurnal INSTEK*, *3*(2), 231–240.

Vulandari, R. T. (2017). Data Mining Teori dan Aplikasi Rapidminer. *Yogyakarta: Gava Media*.

Wu, X., & Zeng, Y. (2019). *Using Apriori Algorithm on Students' Performance Data for Association Rules Mining*. *322*(Iserss), 403–406. https://doi.org/10.2991/iserss-19.2019.300

Yaldi, H. (2018). Analisis Algoritma Asosiasi Untuk Mendapatkan Pola Pemilihan Judul Skripsi Mahasiswa STIE-GK Muara Bulian. *Jurnal Ilmiah Media Sisfo*, *12*(2). http://ejournal.stikom-db.ac.id/index.php/mediasisfo/article/view/377

Yusuf, Y. W., Rian, P. F., & Gerry, T. (2006). Penerapan Data Mining Dalam Penentuan Aturan Asosiasi Antar Jenis Item. *Seminar Nasional Aplikasi Teknologi Informasi 2006 (SNATI 2006)*, *2006*(Snati), E53–E56.

LAMPIRAN

Pivot Table Data Mahasiswa

