

Buku Monograf

# ASSOCIATION RULE

Siti Aisyah  
Anita C. Sembiring  
Delima Sitanggang  
Robert



Editor  
Sasmita Z

**MONOGRAF**  
**ASSOCIATION RULE**

Penulis  
Siti Aisyah  
Anita C. Sembiring  
Delima Sitanggang  
Robert

Editor  
Sasmita Z

Penerbit  
UNPRI PRESS

Redaksi  
Jl. Sampul, Medan

Hak Cipta dilindungi Undang-Undang  
Dilarang memperbanyak karya tulis ini dalam  
bentuk dan dengan cara apapun tanpa ijin dari penerbit

## **Kata Pengantar**

Puji dan Syukur penulis panjatkan ke hadirat Allah SWT, karena atas segala berkat dan anugerah-Nya yang diberikan kepada tim penulis, sehingga penulisan buku monograf berhasil diselesaikan tepat pada waktunya. Buku monograf dengan judul *Association Rule* yang berisi tentang menentukan pola frekuensi tertinggi dengan menggunakan pola item-item dalam database yang memiliki nilai support di atas ambang batas tertentu yang disebut dengan istilah *minimum support* dan mengetahui nilai *cofindence*. Penulis menyadari masih banyak kekurangan dalam penulisan monograf ini, untuk itu saran dan kritik sangat diharapkan untuk penyempurnaan. Penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan motivasi, semangatnya dan berkontribusi dalam kemampuannya untuk dapat menerbitkan buku monograf, semoga Tuhan Yang Maha Esa membalasnya dengan balasan yang lebih baik

Penulis

Siti Aisyah, S.Kom., M.Kom

## Daftar Isi

Kata Pengantar.....	i
Daftar Isi .....	ii
BAB I .....	1
Pengertian Data Mining.....	1
1.1 Data Mining.....	1
1.2 Sejarah Singkat Data Mining .....	4
1.3 Tahapan Data Mining .....	6
1.4 Pengelompokan Data Mining .....	12
1.5 Kelebihan Dan Kekurangan Data Mining .....	14
1.5.1 Kelebihan Data Mining.....	14
1.5.2 Kekurangan Data Mining .....	16
1.6 Kelebihan Dan Kekurangan Knowledge Discovery in Databases (KDD).....	18
1.6.1 Kelebihan Knowledge Discovery in Databases (KDD).....	18
1.6.2 Kekurangan Knowledge Discovery in Databases (KDD).....	19
BAB II .....	21
ASOSIATION RULE MINING .....	21
2.1 Pengertian Asosiation Rule.....	21
2.1.1 Aturan Asosiation Rule.....	23
2.1.2 Analisa Pola Frekuensi Tinggi .....	26
2.1.3 Pembentukan Aturan Asosiatif.....	27
2.2 Kelebihan Dan Kekurangan Asosiation Rule.....	29
2.2.1 Kelebihan Asosiation Rule .....	29
2.2.2 Kekurangan Asosiation Rule .....	30
BAB III .....	32
Pengolahan Data Ekspor Terhadap Ragam Produk Kelapasawit Menggunakan Metode Association Rules .....	32
3.1 Latar Belakang .....	32

3.2	Rumusan Masalah .....	33
3.3	Batasan Masalah .....	34
3.4	Tujuan Dan Manfaat Penelitian .....	34
3.4.1	Tujuan Penelitian.....	34
3.4.2	Manfaat Penelitian .....	34
3.5	Metode Penelitian.....	35
3.5.1	Pengumpulan Data .....	35
3.5.2	Data Acquisition .....	36
3.5.3	Data Preprocessing .....	36
3.5.4	Hasil Metode Association Rules.....	37
3.5.5	Visualisasi Data.....	37
3.6	Analisis Masalah.....	38
3.7	Analisis Data.....	38
3.8	Visualisasi Data.....	45
3.8.1	Upload Dataset .....	46
3.8.2	Import Library Association Rules Dan Menampilkan Dataset ..	46
3.8.3	Visualisasi Data dengan Matplotlib .....	48
3.8.4	Visualisasi Data dengan Seaborn.....	49
3.9	Kesimpulan Dan Saran .....	51
3.9.1	Kesimpulan .....	51
3.9.2	Saran.....	51
DAFTAR PUSTAKA .....		53

## **BAB I**

### **Pengertian Data Mining**

#### **1.1 Data Mining**

Data mining diartikan sebagai suatu proses ekstraksi informasi dan pengetahuan yang berguna dari data besar atau kompleks. Dalam dunia bisnis, data mining sering digunakan untuk mengoptimalkan keputusan bisnis dan meningkatkan kinerja perusahaan. Teknologi data mining memungkinkan perusahaan untuk mengidentifikasi tren dan pola dalam data mereka, serta membantu mengidentifikasi peluang bisnis baru dan mengoptimalkan strategi pemasaran.

Dalam konteks yang lebih luas, data mining merupakan bagian dari bidang yang lebih besar yang dikenal sebagai analisis data, yang bertujuan untuk mengubah data menjadi informasi yang dapat digunakan untuk pengambilan keputusan dan pemecahan masalah.

Data mining adalah proses penggalian informasi yang terdapat dalam dataset yang besar, kompleks, dan terstruktur atau tidak terstruktur untuk menemukan pola, tren, dan informasi penting lainnya yang bermanfaat untuk pengambilan keputusan bisnis. Berikut adalah pengertian data mining menurut 10 ahli:

1. Data mining adalah proses ekstraksi pengetahuan yang tidak trivial, implisit, dan sebelumnya tidak diketahui dari data yang besar dan kompleks" (Fayyad, et al., 1996).
2. Data mining adalah proses pencarian informasi yang berguna dalam dataset besar yang mencakup model matematika, kecerdasan buatan, statistik, dan teknologi informasi" (Han dan Kamber, 2001).

3. Data mining adalah proses yang menghasilkan informasi yang tersembunyi dalam data, yang dapat digunakan untuk pengambilan keputusan bisnis" (Larose, 2005).
4. Data mining adalah proses pemrosesan data yang memungkinkan penemuan pola dan hubungan dalam data yang dapat membantu dalam pengambilan keputusan bisnis" (Kantardzic, 2011).
5. Data mining adalah suatu proses untuk mengeksplorasi dan menemukan pola yang signifikan dalam data, dengan menggunakan teknik-teknik matematika, statistika, dan kecerdasan buatan" (Witten dan Frank, 2005).
6. Data mining adalah proses untuk menemukan pengetahuan yang berharga dalam data, dengan menggunakan metode analisis statistika dan kecerdasan buatan" (Zaki dan Meira, 2014).
7. Data mining adalah suatu teknik pengolahan data untuk menemukan informasi yang tidak dapat dijangkau dengan metode konvensional dalam jumlah data yang sangat besar" (Jain dan Dubes, 1988).
8. Data mining adalah proses pencarian pola yang berharga dalam data, yang dapat digunakan untuk mendukung pengambilan keputusan bisnis" (Berry dan Linoff, 1997).
9. Data mining adalah proses penggalian informasi penting dari data dengan menggunakan teknik-teknik pengolahan data dan kecerdasan buatan" (Hastie et al., 2017).
10. Data mining adalah teknik untuk menggali pengetahuan dari data dengan menggunakan teknik-teknik pengolahan data, statistika, dan kecerdasan buatan" (Shmueli et al., 2017).

Data Mining memainkan peran penting dalam era digital saat ini dan diterapkan pada berbagai bidang seperti bisnis, kesehatan, dan sosial

untuk memahami dan memprediksi pola. Proses Data Mining melibatkan tahap-tahap seperti pengumpulan data, pre-processing, pemodelan, dan evaluasi untuk mencapai hasil yang diinginkan.

Data mining juga sering disebut sebagai *Knowledge Discovery in Database* (KDD). KDD adalah suatu proses untuk menemukan informasi dan wawasan baru dari data yang besar dan kompleks. Tujuannya adalah untuk mengubah data menjadi informasi yang berguna bagi pengguna. KDD melibatkan tahap-tahap seperti pengumpulan data, pre-processing, pemodelan, dan evaluasi untuk mencapai hasil yang diinginkan. KDD dan Data Mining sering digunakan secara bersamaan dan memiliki makna yang hampir sama.

Data Mining diterapkan dengan dukungan dari berbagai bidang ilmu seperti Neural Network, Pattern Recognition, Analisis Spasial Data, Basis Data Gambar, dan Pemrosesan Sinyal. Data Mining didefinisikan sebagai suatu proses untuk menemukan pola-pola dalam data, yang dapat dilakukan secara otomatis atau semi-otomatis. Pola-pola yang ditemukan harus memiliki makna dan memberikan manfaat, biasanya secara ekonomi, dan memerlukan data dalam jumlah besar. Data mining memiliki beberapa karakteristik, antara lain sebagai berikut:

1. **Automatisasi:** Data Mining adalah proses otomatis untuk menemukan pola-pola dalam data. Hal ini mempermudah proses menemukan pola dan mempercepat hasil analisis.
2. **Penggunaan teknik statistik dan matematika:** Data Mining menggunakan teknik statistik dan matematika untuk menganalisis

dan mengidentifikasi pola dalam data. Ini memastikan hasil analisis yang akurat dan dapat dipertanggungjawabkan.

3. Volume data besar: Data Mining membutuhkan data dalam jumlah besar. Hal ini karena pola-pola dalam data hanya dapat ditemukan dengan memproses data dalam jumlah besar.
4. Integrasi data: Data Mining membutuhkan integrasi data dari berbagai sumber untuk mencapai hasil yang diinginkan. Integrasi data memastikan bahwa data yang digunakan dalam proses Data Mining adalah data yang akurat dan relevan.
5. Pencarian pola yang bermakna: Data Mining mencari pola-pola dalam data yang memiliki makna dan memberikan manfaat bagi pengguna. Ini memastikan bahwa hasil analisis Data Mining tidak hanya memenuhi tujuan pengguna, tetapi juga memberikan manfaat bagi pengguna.
6. Ketergantungan pada algoritma: Data Mining sangat bergantung pada algoritma yang digunakan dalam proses. Algoritma yang digunakan harus memenuhi karakteristik Data Mining dan memastikan bahwa hasil analisis memenuhi tujuan pengguna.
7. Kemampuan untuk memprediksi masa depan: Data Mining memiliki kemampuan untuk memprediksi masa depan berdasarkan pola-pola dalam data. Ini membantu pengguna dalam membuat keputusan yang tepat dan memastikan bahwa hasil analisis memenuhi tujuan pengguna.

## **1.2 Sejarah Singkat Data Mining**

Data mining merupakan disiplin ilmu yang relatif baru, yang berakar dari beberapa bidang lain seperti statistika, matematika, dan ilmu komputer. Sejarah data mining dimulai pada tahun 1960-an ketika ilmuwan

mulai mengeksplorasi metode dan teknik untuk menghasilkan informasi berharga dari data yang tersedia.

Di awal perkembangan data mining, teknik-teknik yang digunakan terbatas pada analisis statistik dan eksplorasi data dasar, seperti regresi, korelasi, dan clustering. Namun, pada tahun 1980-an, dengan berkembangnya teknologi komputer dan kemajuan dalam bidang kecerdasan buatan, teknik-teknik data mining semakin berkembang dan semakin kompleks.

Pada tahun 1990-an, terjadi peningkatan dramatis dalam jumlah data yang tersedia dalam berbagai bidang seperti bisnis, keuangan, kesehatan, dan lain-lain. Kebutuhan untuk menghasilkan informasi berharga dari data semakin meningkat dan para peneliti mulai mengembangkan algoritma dan teknik data mining yang lebih canggih.

Pada tahun 2000-an, teknologi data mining semakin berkembang dan digunakan dalam berbagai aplikasi, termasuk analisis pasar, prediksi bisnis, deteksi penipuan, analisis risiko, dan lain-lain. Di masa kini, data mining telah menjadi bagian penting dari banyak bidang dan industri, dan terus berkembang dengan cepat dengan kemajuan teknologi seperti Big Data, Machine Learning, dan Artificial Intelligence.

Dalam sejarahnya, data mining telah mengalami evolusi dan perubahan yang signifikan, dari teknik analisis sederhana hingga teknik yang kompleks dan canggih. Namun, tujuannya tetap sama, yaitu untuk mengekstrak informasi berharga dari data yang tersedia dan menghasilkan pengetahuan baru yang dapat digunakan untuk meningkatkan efisiensi, produktivitas, dan pengambilan keputusan dalam berbagai bidang.

### 1.3 Tahapan Data Mining

Tahapan data mining adalah serangkaian proses sistematis dan berurutan yang harus dilakukan untuk mengidentifikasi dan mengekstrak informasi penting dari data yang besar dan kompleks. Tahapan ini bertujuan untuk memastikan bahwa data yang dianalisis sudah dalam kondisi yang tepat untuk melakukan proses data mining dan hasil dari proses ini memiliki validitas dan relevansi yang tinggi.

Data Mining dan *Knowledge Discovery in Database* (KDD) seringkali digunakan secara bersamaan dan dapat diartikan sebagai sinonim. Keduanya mengacu pada proses menemukan pola-pola dan pengetahuan baru dalam data yang besar dan kompleks. Dalam hal ini, data mining sering digunakan secara lebih luas dan mewakili proses menemukan pola dalam data dari berbagai sumber, sedangkan KDD lebih fokus pada proses menemukan pengetahuan dalam database.

Data Mining (penambangan data) adalah bagian dari proses *Knowledge Discovery in Databases* (KDD) atau Penemuan Pengetahuan dalam Basis Data. KDD adalah proses yang kompleks yang melibatkan serangkaian tahapan mulai dari pemilihan data, pre-processing, seleksi atribut, transformasi data, mining data, evaluasi model hingga interpretasi hasil untuk menghasilkan pengetahuan yang berharga bagi pengambil keputusan.

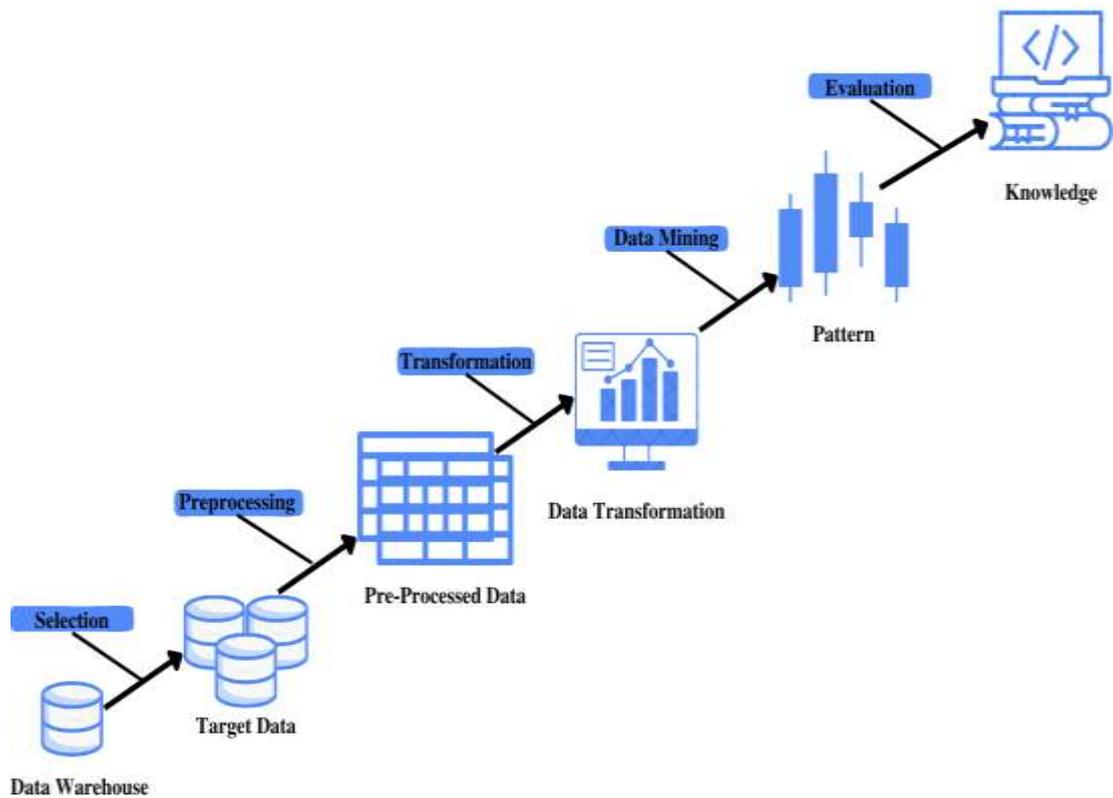
Data Mining merupakan salah satu tahap dalam proses KDD yang paling penting, karena pada tahap ini data diolah dan diproses untuk menemukan pola dan asosiasi yang tidak terlihat pada data mentah. Setelah pola dan asosiasi ditemukan, tahap berikutnya dalam proses KDD

adalah evaluasi dan interpretasi hasil Data Mining, yang kemudian mengarah ke pengambilan keputusan dan tindakan yang cerdas dan tepat.

Dalam proses KDD, Data Mining digunakan untuk mengambil informasi yang berharga dari data mentah dan menghasilkan model atau aturan yang dapat digunakan untuk memprediksi atau menjelaskan fenomena tertentu. Selain itu, Data Mining juga dapat digunakan untuk mengidentifikasi anomali atau kejadian yang tidak biasa pada data, sehingga membantu pengguna untuk mengambil tindakan yang tepat dan menghindari risiko.

Secara keseluruhan, Data Mining adalah bagian penting dalam proses KDD, karena memungkinkan pengguna untuk menemukan pola dan asosiasi yang tersembunyi dalam data mentah, yang pada gilirannya dapat digunakan untuk membuat keputusan yang lebih baik dan lebih tepat.

Ada beberapa tahapan yang umum dilakukan dalam data mining, meliputi: pengumpulan data, pembersihan data, pengintegrasian data, reduksi dimensi, pemodelan, dan interpretasi. Setiap tahap memiliki tugas dan sub-tugas masing-masing yang harus dilakukan sebelum melanjutkan ke tahap berikutnya. Tahap pertama dimulai dari sumber data dan berakhir dengan informasi yang diperoleh melalui beberapa proses yang dapat kita lihat pada gambar 1 berikut:



**Gambar 1** Tahapan Data Mining

Proses Data Mining dapat diterangkan melalui tahap-tahap sebagai berikut:

### 1. Seleksi Data

Dalam tahap ini, data yang ada dipilah dan disaring untuk memastikan bahwa data yang akan digunakan memiliki kualitas dan relevansi yang baik untuk tujuan analisis.

Beberapa faktor yang perlu dipertimbangkan dalam proses seleksi data adalah:

- a. Kualitas data: memastikan bahwa data yang akan digunakan memiliki validitas dan akurasi yang baik.
- b. Relevansi data: memastikan bahwa data yang akan digunakan relevan dengan tujuan analisis yang diinginkan.

- c. Volume data: memastikan bahwa data yang akan digunakan tidak terlalu besar atau terlalu kecil sehingga mempengaruhi hasil analisis.

Setelah data dipilah dan disaring, data yang sudah terseleksi akan digunakan dalam tahap selanjutnya dalam proses KDD seperti ekstraksi fitur, pemodelan, dan verifikasi.

## 2. Pre-Processing/Cleaning (Pemilihan Data)

Dalam tahap ini, data yang tidak relevan, tidak akurat, atau tidak lengkap dikeluarkan dari dataset. Ini membantu mengurangi noise dan memastikan bahwa analisis yang dilakukan didasarkan pada data yang valid dan akurat. Beberapa tugas yang sering dilakukan dalam tahap ini meliputi:

- a. Handling missing values: Menentukan dan mengatasi nilai-nilai yang hilang dalam dataset.
- b. Handling duplicate values: Menentukan dan mengatasi duplikat dalam dataset.
- c. Handling irrelevant data: Menentukan dan menghapus data yang tidak relevan atau tidak berguna bagi analisis.
- d. Data normalization: Memastikan bahwa data dalam dataset memiliki skala yang sama dan mudah dibandingkan.
- e. Data Transformation: Mengubah data ke dalam format yang lebih sesuai untuk analisis.

Tahap pre-processing/cleaning sangat penting karena memastikan bahwa analisis yang dilakukan didasarkan pada data yang valid dan akurat, yang akan mempengaruhi keakuratan hasil dan kesimpulan yang diambil.

### 3. Transformasi

Dalam tahap ini, data dalam dataset diterjemahkan atau diformat ulang agar lebih sesuai untuk analisis. Beberapa tugas yang sering dilakukan dalam tahap ini meliputi:

- a. *Encoding categorical data*: Mengubah data kategorik menjadi bentuk numerik seperti one-hot encoding.
- b. *Scaling data*: Memastikan bahwa data dalam dataset memiliki skala yang sama dan mudah dibandingkan.
- c. *Aggregation*: Menggabungkan data dari beberapa sumber menjadi satu dataset yang lebih besar.
- d. *Sampling*: Mengambil sampel dari dataset yang lebih besar untuk membuat analisis lebih efisien.
- e. *Dimensionality reduction*: Mengurangi jumlah dimensi data untuk membuat analisis lebih mudah.

Tahap transformasi data sangat penting karena memastikan bahwa data dalam dataset sesuai dengan format yang dibutuhkan untuk analisis, mempermudah proses analisis, dan memastikan bahwa hasil analisis yang diambil didasarkan pada data yang benar.

### 4. Data Mining

Dalam tahap ini, algoritma data mining digunakan untuk menganalisis data dan menemukan pola dan hubungan yang tidak diketahui sebelumnya. Beberapa tugas yang sering dilakukan dalam tahap ini meliputi:

- a. Association rule mining: Menemukan hubungan antara item dalam dataset.
- b. Clustering: Mengelompokkan data berdasarkan kesamaan.
- c. Classification: Menentukan kelas dari data baru berdasarkan pola yang ditemukan dalam data latih.
- d. Regression analysis: Menentukan hubungan antara variabel independen dan dependen dalam data.
- e. Anomaly detection: Mendeteksi data yang tidak sesuai dengan pola yang ditemukan dalam data.

Pada tahap data mining sangat penting karena memungkinkan analisis data yang lebih dalam dan menemukan pola dan hubungan yang tidak diketahui sebelumnya. Ini membantu dalam mengambil keputusan yang lebih baik dan mengidentifikasi peluang bisnis yang potensial.

## 5. Evaluasi

Dalam tahap ini, hasil dari analisis data yang dilakukan dalam tahap sebelumnya dievaluasi dan diterima atau ditolak. Evaluasi ini memastikan bahwa hasil analisis yang diambil didasarkan pada data yang benar dan akurat. Beberapa tugas yang sering dilakukan dalam tahap ini meliputi:

- a. Validasi model: Memastikan bahwa model data mining yang digunakan memiliki tingkat akurasi yang memuaskan.
- b. Verifikasi hasil: Memastikan bahwa hasil yang diambil sesuai dengan data dan memiliki makna praktis.
- c. Cross-validation: Memverifikasi hasil dengan menggunakan data latih dan data uji yang berbeda.

- d. Menentukan metrik evaluasi: Menentukan metrik yang akan digunakan untuk mengukur keakuratan dan kualitas hasil analisis.
- e. Visualisasi hasil: Memvisualisasikan hasil analisis untuk mempermudah pemahaman dan interpretasi.

Secara keseluruhan, tahap evaluasi sangat penting karena memastikan bahwa hasil analisis yang diambil didasarkan pada data yang benar dan akurat. Ini juga membantu dalam menentukan apakah model dan analisis yang digunakan benar-benar efektif dan sesuai untuk tujuan tertentu.

#### **1.4 Pengelompokan Data Mining**

Pengelompokan (clustering) adalah teknik data mining yang digunakan untuk mengelompokkan data berdasarkan kesamaan. Dalam pengelompokan, data dikelompokkan ke dalam beberapa kelompok (cluster) yang berbeda, dengan asumsi bahwa data yang serupa akan tergolong dalam satu kelompok yang sama.

Pengelompokan berguna untuk memahami dan mengeksplorasi data, dan juga dapat digunakan untuk aplikasi seperti segmentasi pasar, pengelompokan pelanggan, dan pengelompokan obyek geografis. Beberapa algoritma pengelompokan yang sering digunakan meliputi:

1. K-Means: Metode ini merupakan salah satu metode clustering yang paling populer dan sering digunakan. Algoritma yang menentukan k centroid untuk setiap kelompok, dan menempatkan data ke dalam kelompok yang paling dekat dengan centroid tersebut.

2. Hierarchical clustering: Algoritma yang mengelompokkan data dengan membangun hierarki dari kelompok yang semakin kecil. Metode ini membagi data menjadi kelompok-kelompok secara bertahap. Metode ini terdiri dari dua tipe, yaitu agglomerative dan divisive. Dalam tipe agglomerative, setiap objek dianggap sebagai kelompok terpisah, dan kemudian kelompok-kelompok tersebut digabungkan berdasarkan kesamaan karakteristiknya. Sedangkan dalam tipe divisive, seluruh data dianggap sebagai satu kelompok, dan kemudian data tersebut dibagi menjadi kelompok-kelompok yang lebih kecil berdasarkan perbedaan karakteristiknya.
3. Density-Based Clustering (DBSCAN): Metode ini mengelompokkan data berdasarkan kepadatan data. Data yang terletak dalam wilayah yang padat akan digolongkan ke dalam satu kelompok, sedangkan data yang terletak di wilayah yang lebih jarang akan digolongkan ke dalam kelompok yang berbeda.
4. EM (Expectation Maximization): adalah sebuah algoritma yang digunakan dalam statistik untuk menyelesaikan masalah estimasi parameter pada model yang memiliki variabel tersembunyi (hidden variable) atau model yang tidak lengkap. EM digunakan ketika kita ingin melakukan analisis terhadap data yang tidak lengkap atau data yang tidak dapat dianalisis secara langsung.
5. Grid-Based Clustering: Metode ini membagi data menjadi beberapa sel, dan kemudian kelompok-kelompok terbentuk berdasarkan sel-sel tersebut. Metode ini cocok digunakan pada data yang berdimensi tinggi. Salah satu kelemahan dari metode ini adalah jika terdapat grid yang memiliki data yang sangat banyak, maka dapat

mengakibatkan overfitting atau kelebihan penyesuaian pada grid tersebut dan membuat cluster tidak optimal.

## **1.5 Kelebihan Dan Kekurangan Data Mining**

### **1.5.1 Kelebihan Data Mining**

Berikut adalah kelebihan data mining yang dapat mempengaruhi kualitas hasil analisis data:

1. Menemukan pola dan hubungan yang tersembunyi dalam data: Data mining memungkinkan kita untuk menemukan pola dan hubungan yang tidak dapat ditemukan secara manual atau tidak terlihat dengan mudah dalam data. Ini dapat memberikan wawasan baru dan berharga dalam bisnis, sains, dan banyak bidang lainnya.
2. Meningkatkan efisiensi dan produktivitas: Data mining dapat membantu meningkatkan efisiensi dan produktivitas dengan membantu kita mengidentifikasi pola-pola dan hubungan yang berguna dalam data, yang dapat membantu kita memprediksi hasil dan membuat keputusan yang lebih baik.
3. Mendukung pengambilan keputusan yang lebih baik: Data mining dapat membantu kita membuat keputusan yang lebih baik dengan memberikan wawasan dan informasi yang lebih terperinci dan akurat. Ini dapat membantu kita mengidentifikasi peluang baru dan mengurangi risiko dalam bisnis atau organisasi.
4. Menghemat waktu dan biaya: Data mining dapat membantu menghemat waktu dan biaya dengan memungkinkan analisis data yang cepat dan efisien. Ini dapat membantu organisasi memproses data lebih cepat dan secara efektif, yang dapat menghemat biaya operasional.

5. Mendukung strategi pemasaran yang lebih baik: Data mining dapat membantu organisasi memahami perilaku pelanggan mereka dan mengidentifikasi preferensi dan kebutuhan pelanggan. Ini dapat membantu organisasi merancang strategi pemasaran yang lebih baik dan mencapai hasil yang lebih baik.
6. Meningkatkan keamanan dan pengawasan: Data mining dapat membantu organisasi meningkatkan keamanan dan pengawasan dengan memungkinkan pengidentifikasian anomali dan pola yang mencurigakan dalam data.
7. Mendukung perencanaan dan pengembangan produk: Data mining dapat membantu organisasi memahami tren dan preferensi pelanggan, yang dapat membantu dalam perencanaan dan pengembangan produk yang lebih baik dan lebih sesuai dengan kebutuhan pasar.
8. Meningkatkan efektivitas pelayanan pelanggan: Data mining dapat membantu organisasi memahami kebutuhan dan preferensi pelanggan mereka dan merespon dengan lebih baik pada permintaan pelanggan.
9. Memfasilitasi pengembangan sistem rekomendasi: Data mining dapat membantu dalam pengembangan sistem rekomendasi dengan memahami perilaku pelanggan dan memberikan rekomendasi yang lebih akurat dan terpersonalisasi.
10. Menyediakan wawasan baru dan inovasi: Data mining dapat membuka wawasan baru dan inovasi dengan mengidentifikasi pola dan hubungan yang sebelumnya tidak diketahui dalam data. Ini

dapat membantu organisasi menciptakan produk dan layanan baru dan meningkatkan kinerja dan efektivitas mereka.

### **1.5.2 Kekurangan Data Mining**

Berikut adalah kekurangan data mining yang dapat mempengaruhi kualitas hasil analisis data:

1. **Memerlukan data yang berkualitas**

Data mining memerlukan data yang berkualitas untuk menghasilkan hasil yang akurat dan bermanfaat. Kurangnya data atau data yang buruk dapat memengaruhi kualitas hasil analisis dan interpretasi.

2. **Pengolahan data yang kompleks**

Proses pengolahan data yang kompleks pada data mining dapat memakan waktu dan biaya yang signifikan. Hal ini terutama terjadi jika data yang digunakan besar dan kompleks.

3. **Ketergantungan pada algoritma**

Data mining sangat tergantung pada algoritma yang digunakan. Algoritma yang tidak cocok dengan jenis data atau tujuan analisis dapat menghasilkan hasil yang tidak akurat.

4. **Kesalahan interpretasi**

Hasil analisis data yang dihasilkan dari data mining seringkali tergantung pada interpretasi manusia. Kesalahan interpretasi dapat menyebabkan kesimpulan yang salah dan membuat keputusan yang buruk.

5. **Risiko privasi dan keamanan**

Penggunaan data mining dapat menimbulkan risiko privasi dan keamanan. Penggunaan data yang sensitif atau data pelanggan dapat menyebabkan kebocoran informasi atau penyalahgunaan data.

6. Keterbatasan dalam memprediksi masa depan

Data mining dapat membantu dalam memprediksi hasil di masa lalu dan sekarang, namun tidak selalu dapat memprediksi masa depan dengan akurat. Hal ini karena perubahan kondisi di masa depan dan variabilitas dalam data yang mungkin tidak dapat diperhitungkan.

7. Ketergantungan pada sumber data

Hasil dari data mining tergantung pada ketersediaan dan kualitas sumber data. Jika sumber data tidak lengkap atau tidak terpercaya, hasil analisis data dapat menjadi tidak akurat.

8. Memerlukan keahlian khusus

Data mining memerlukan keahlian khusus dalam analisis data dan teknik pengolahan data. Kurangnya keahlian khusus dapat mempengaruhi kualitas hasil analisis.

9. Biaya implementasi yang tinggi

Implementasi data mining memerlukan biaya yang tinggi dalam hal perangkat lunak dan hardware. Hal ini dapat memengaruhi adopsi data mining dalam organisasi kecil dan menengah.

10. Tidak dapat menggantikan keputusan manusia

Data mining dapat membantu dalam pengambilan keputusan, namun tidak dapat menggantikan keputusan manusia sepenuhnya. Keputusan manusia masih diperlukan untuk mengevaluasi dan menafsirkan hasil analisis.

## **1.6 Kelebihan Dan Kekurangan Knowledge Discovery in Databases (KDD)**

### **1.6.1 Kelebihan Knowledge Discovery in Databases (KDD)**

KDD, atau Knowledge Discovery in Databases, adalah proses mengekstraksi pengetahuan yang berguna dari data besar atau kompleks. Berikut adalah beberapa kelebihan KDD:

1. Menghasilkan pengetahuan baru

Proses KDD memungkinkan kita untuk menemukan pola atau hubungan yang tidak terduga dalam data, dan dengan demikian membantu menghasilkan pengetahuan baru yang dapat digunakan untuk mengambil keputusan yang lebih baik.

2. Meningkatkan efisiensi

KDD memungkinkan kita untuk mengidentifikasi dan menghapus data yang tidak relevan atau duplikat, sehingga meningkatkan efisiensi dalam pengolahan data.

3. Menyederhanakan data

Proses KDD dapat membantu dalam menyederhanakan data dengan mengidentifikasi fitur atau atribut yang paling penting dan relevan, sehingga memudahkan dalam analisis data.

4. Meningkatkan kecepatan pengambilan keputusan

Dengan KDD, kita dapat mengurangi waktu yang diperlukan untuk mengumpulkan dan menganalisis data, dan dengan demikian meningkatkan kecepatan pengambilan keputusan.

5. Meningkatkan kualitas hasil analisis

KDD menggunakan teknik-teknik yang canggih dalam analisis data, sehingga dapat menghasilkan hasil analisis yang lebih akurat dan dapat diandalkan.

6. Mengoptimalkan strategi bisnis

Dengan KDD, perusahaan dapat mengidentifikasi tren dan pola dalam data mereka, sehingga dapat mengoptimalkan strategi bisnis dan meningkatkan kinerja perusahaan.

7. Mendukung pengambilan keputusan yang berbasis data

KDD membantu dalam pengambilan keputusan yang lebih akurat dan obyektif, karena keputusan yang diambil didasarkan pada data dan bukan pada pendapat atau intuisi manusia.

8. Mengurangi risiko

Dengan KDD, perusahaan dapat mengidentifikasi risiko potensial dan mengambil tindakan yang tepat sebelum masalah terjadi, sehingga dapat mengurangi risiko dan kerugian yang mungkin terjadi.

Dalam keseluruhan, KDD merupakan teknologi penting yang dapat membantu kita memanfaatkan potensi informasi yang terkandung dalam data, dan dengan demikian dapat membantu meningkatkan efisiensi, mengoptimalkan keputusan, dan mengurangi risiko dalam berbagai bidang.

### **1.6.2 Kekurangan Knowledge Discovery in Databases (KDD)**

Meskipun memiliki berbagai kelebihan, proses KDD juga memiliki beberapa kekurangan yang perlu diperhatikan, antara lain:

1. Membutuhkan waktu dan sumber daya yang besar Proses KDD membutuhkan waktu dan sumber daya yang besar, karena melibatkan proses pengumpulan data, pra-pemrosesan data, pemodelan data, dan interpretasi hasil. Hal ini dapat membuat proses KDD menjadi mahal dan memakan waktu.

2. Terlalu terfokus pada data Proses KDD terkadang terlalu terfokus pada data dan mengabaikan konteks atau informasi yang tersedia di luar data. Hal ini dapat menyebabkan hasil yang dihasilkan tidak akurat atau tidak relevan dengan masalah yang ingin diselesaikan.
3. Memerlukan keahlian teknis yang tinggi Proses KDD memerlukan keahlian teknis yang tinggi, termasuk dalam penggunaan software dan algoritma analisis data. Ini dapat menjadi tantangan bagi mereka yang tidak memiliki latar belakang teknis yang memadai.
4. Data tidak selalu lengkap Data yang digunakan dalam proses KDD tidak selalu lengkap atau tidak tersedia dalam jumlah yang cukup. Hal ini dapat membatasi kemampuan proses KDD dalam menghasilkan hasil yang akurat dan berguna.
5. Tidak menggantikan keputusan manusia sepenuhnya Meskipun proses KDD dapat membantu dalam pengambilan keputusan, keputusan akhir tetap harus dibuat oleh manusia. Proses KDD hanya menyediakan informasi dan saran, dan keputusan akhir harus didasarkan pada pertimbangan etis, moral, dan bisnis yang tidak selalu dapat ditangani oleh teknologi.

Dalam keseluruhan, meskipun KDD memiliki potensi yang besar dalam memanfaatkan data dan menghasilkan pengetahuan baru, perlu diingat bahwa proses KDD juga memiliki beberapa kekurangan yang perlu diperhatikan. Oleh karena itu, penting untuk mempertimbangkan kelebihan dan kekurangan KDD secara menyeluruh sebelum memutuskan untuk menggunakannya dalam konteks tertentu.

## BAB II

### ASOSIATION RULE MINING

#### 2.1 Pengertian Asosiation Rule

Analisis asosiasi atau association rule mining adalah teknik data mining yang digunakan untuk menemukan hubungan antara item dalam data transaksi. Tujuannya adalah untuk menemukan aturan asosiasi yang menunjukkan item apa yang sering ditemukan bersama-sama dalam transaksi yang sama.

Asosiasi aturan (association rules) adalah teknik yang digunakan untuk menemukan hubungan dan pola yang tersembunyi di antara data dalam dataset. Berikut adalah definisi asosiasi aturan menurut 10 ahli:

1. Aturan asosiasi mengacu pada kemampuan menemukan korelasi atau pola dalam data. Aturan asosiasi memungkinkan kita untuk memahami hubungan antara berbagai item dalam sebuah kumpulan data" (Jiawei Han, et al., 2011).
2. Aturan asosiasi adalah teknik analisis data yang digunakan untuk menemukan hubungan atau pola antara dua atau lebih variabel atau item dalam dataset" (Dong, 1997).
3. Aturan asosiasi adalah teknik yang memungkinkan penemuan pola hubungan yang berguna dalam dataset dengan menggunakan nilai support dan confidence" (Agrawal dan Srikant, 1994).
4. Aturan asosiasi adalah teknik data mining yang digunakan untuk mengidentifikasi keterkaitan antara item atau variabel dalam suatu kumpulan data" (Berry dan Linoff, 1997).

5. Aturan asosiasi adalah teknik yang digunakan untuk mengidentifikasi keterkaitan antara item dalam kumpulan data, dan sering digunakan dalam kecerdasan bisnis untuk memprediksi perilaku pelanggan" (Witten dan Frank, 2005).
6. Aturan asosiasi adalah teknik yang memungkinkan penemuan pola hubungan dalam data yang diberikan, dengan menggunakan tingkat dukungan dan keyakinan" (Zaki dan Meira, 2014).
7. Aturan asosiasi adalah teknik analisis data yang digunakan untuk menemukan hubungan atau pola antara item dalam dataset, dan dapat digunakan dalam berbagai aplikasi seperti market basket analysis dan recommendation systems" (Kantardzic, 2011).
8. Aturan asosiasi adalah teknik yang memungkinkan penemuan hubungan atau keterkaitan antara item dalam dataset yang besar, dan dapat membantu dalam pengambilan keputusan bisnis" (Jain dan Dubes, 1988).
9. Aturan asosiasi adalah metode analisis data yang digunakan untuk mengungkapkan hubungan antara item dalam sebuah kumpulan data, dengan tujuan menemukan pola-pola tertentu yang membantu dalam pengambilan keputusan" (Shmueli et al., 2017).
10. Aturan asosiasi adalah teknik data mining yang memungkinkan identifikasi pola-pola asosiasi antara item dalam dataset, yang dapat digunakan dalam berbagai aplikasi seperti market basket analysis dan recommendation systems" (Hastie et al., 2017).

### 2.1.1 Aturan Asosiasi Rule

Aturan Asosiasi adalah salah satu teknik Data Mining yang digunakan untuk menemukan hubungan atau korelasi antara item-item yang ada dalam suatu basis data. Teknik ini sering digunakan dalam bidang pemasaran, manajemen rantai pasokan, dan analisis sosial untuk mengidentifikasi pola-pola penting dalam data.

1. Konsep Dasar Aturan Asosiasi Aturan Asosiasi digunakan untuk menemukan hubungan antara dua item atau lebih dalam basis data. Misalnya, jika seorang pelanggan membeli sebuah buku, kemungkinan besar ia juga akan membeli sebuah pena. Dalam contoh ini, "buku" dan "pena" adalah dua item yang saling terkait. Aturan Asosiasi memungkinkan kita untuk menemukan item-item yang saling terkait dan kemudian membangun aturan yang menggambarkan hubungan di antara mereka.
2. Support, Confidence, dan Lift Untuk mengevaluasi kualitas dari aturan yang dihasilkan, digunakan tiga metrik yaitu Support, Confidence, dan Lift. Support adalah proporsi dari transaksi dalam basis data yang mengandung kedua item yang saling terkait. Confidence adalah tingkat keyakinan bahwa jika suatu transaksi mengandung satu item, maka akan mengandung item lainnya juga. Lift adalah rasio antara confidence dan support dari aturan tersebut.
3. Algoritma Apriori Aturan Asosiasi biasanya ditemukan dengan menggunakan algoritma Apriori. Algoritma ini bekerja dengan cara membangun kandidat itemset berdasarkan itemset yang lebih kecil dan kemudian menghitung support dari setiap kandidat. Itemset

yang memenuhi support threshold kemudian digunakan untuk menghasilkan aturan asosiasi.

Aturan Asosiasi dapat digunakan untuk menentukan frekuensi suatu itemset dalam data. Frekuensi itemset mengacu pada jumlah kemunculan itemset tersebut dalam kumpulan data. Frekuensi ini kemudian digunakan untuk menghitung metrik seperti Support dan Confidence dalam Aturan Asosiasi. Dalam konteks Aturan Asosiasi, itemset yang sering muncul dalam kumpulan data dianggap memiliki hubungan yang kuat satu sama lain. Oleh karena itu, frekuensi suatu itemset dapat menjadi petunjuk seberapa sering itemset itu muncul bersama dengan itemset lainnya dalam data.

Aturan asosiasi menentukan frekuensi minimum dari item yang muncul bersama-sama, yang dikenal sebagai support, dan tingkat keyakinan, yang menentukan tingkat kepercayaan aturan asosiasi. Beberapa aplikasi dari analisis asosiasi meliputi:

1. Rekomendasi Produk

Menentukan item apa yang sering dibeli bersama-sama dan membuat rekomendasi produk berdasarkan aturan asosiasi. Contoh rekomendasi produk adalah sebuah toko online ingin memberikan rekomendasi produk kepada pelanggan mereka berdasarkan produk apa yang mereka beli sebelumnya. Mereka mengumpulkan data transaksi pelanggan dan menggunakan data mining untuk menemukan pola belanja pelanggan.

Hasil analisis menunjukkan bahwa pelanggan yang sering membeli produk A juga sering membeli produk B dan C. Berdasarkan informasi ini, toko online dapat membuat rekomendasi produk untuk

pelanggan yang sedang melihat produk A, seperti menawarkan produk B dan C yang sering dibeli bersama.

## 2. Market Basket Analysis

Menganalisis transaksi untuk menemukan item yang sering dibeli bersama-sama dalam sebuah toko atau pasar. Sebagai contoh sebuah toko makanan kelontong menganalisis data transaksi pelanggan mereka untuk menentukan item apa yang sering dibeli bersama-sama. Hasil analisis menunjukkan bahwa ketika seseorang membeli roti, mereka sering juga membeli mentega dan jelly. Berdasarkan informasi ini, toko dapat membuat strategi promosi yang menawarkan diskon pada item terkait, seperti menawarkan diskon pada roti jika pelanggan juga membeli mentega dan jelly. Ini dapat memotivasi pelanggan untuk membeli lebih banyak item dan memperkuat loyalitas mereka pada toko.

## 3. Fraud Detection

Menentukan aturan asosiasi yang tidak lazim untuk membantu mengidentifikasi aktivitas penipuan. Sebagai contoh Fraud Detection pada sebuah bank ingin menggunakan data mining untuk mendeteksi transaksi penipuan. Mereka mengumpulkan data transaksi kartu kredit pelanggan dan menggunakan data mining untuk menemukan pola transaksi yang tidak biasa.

Hasil analisis menunjukkan bahwa beberapa transaksi yang terlihat mencurigakan, seperti transaksi besar yang dilakukan dari lokasi geografis yang berbeda atau transaksi yang dilakukan pada waktu yang tidak biasa. Bank dapat mengambil tindakan untuk memverifikasi transaksi yang mencurigakan sebelum menyetujui pembayaran.

Analisis asosiasi sangat berguna untuk menentukan hubungan antar item dalam data transaksi dan membantu dalam aplikasi seperti rekomendasi produk, market basket analysis, dan deteksi penipuan.

### 2.1.2 Analisa Pola Frekuensi Tinggi

Analisis Pola Frekuensi Tinggi (Frequent Pattern Analysis) adalah salah satu teknik penting dalam Data Mining untuk menemukan pola-pola yang sering muncul pada data. Teknik ini digunakan untuk menemukan asosiasi antara beberapa item dalam dataset. Asosiasi didefinisikan sebagai kemunculan bersamaan beberapa item dalam dataset yang cukup sering terjadi.

Proses mencari kombinasi item yang memenuhi minimum dari nilai support yang ada dalam database dilakukan. Nilai support merupakan persentase item atau gabungan item yang terdapat dalam database. Nilai support suatu item dapat diperoleh dengan menggunakan rumus berikut:

$$\text{Support}(A) = \frac{\text{Jumlah Transaksi Untuk } A}{\text{Total Transaksi}}$$

Rumus support menjelaskan bahwa nilai support diperoleh dengan membagi jumlah transaksi yang mengandung item  $A$  dengan jumlah total transaksi. Sementara nilai support dari dua item diperoleh dari rumus berikut:

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

$$\text{Support}(A, B) = \frac{\text{Jumlah Transaksi Untuk } A}{\text{Total Transaksi}}$$

Rumus yang telah disediakan untuk mendapatkan nilai support 2-itemsets menjelaskan bahwa caranya adalah dengan membagi jumlah transaksi yang mengandung kedua item tersebut (dengan item pertama

dan item kedua yang hadir secara bersamaan) dengan jumlah total seluruh transaksi.

### **2.1.3 Pembentukan Aturan Asosiatif**

Pembentukan aturan asosiatif atau association rule mining adalah proses yang dilakukan dalam data mining untuk menemukan pola hubungan antara item atau variabel dalam dataset. Tujuannya adalah untuk menemukan hubungan antara item atau variabel dalam dataset yang muncul bersama secara teratur. Proses pembentukan aturan asosiatif dimulai dengan mengidentifikasi item atau variabel yang sering muncul bersama dalam dataset. Kemudian, dilakukan perhitungan untuk menentukan seberapa sering hubungan ini terjadi. Hal ini dilakukan dengan menghitung nilai support dan confidence pada setiap kombinasi item atau variabel.

Support adalah ukuran frekuensi kemunculan item atau variabel dalam dataset. Nilai support menunjukkan seberapa sering kombinasi item atau variabel muncul dalam dataset dibandingkan dengan total jumlah data dalam dataset. Sedangkan, confidence adalah ukuran seberapa sering dua item atau variabel muncul bersamaan dalam dataset. Nilai confidence menunjukkan seberapa sering item atau variabel kedua muncul bersamaan dengan item atau variabel pertama dalam dataset.

Setelah nilai support dan confidence telah dihitung, dilakukan proses seleksi untuk menemukan aturan asosiatif yang signifikan dan bermakna. Aturan asosiatif yang signifikan harus memiliki nilai support dan confidence yang tinggi, sehingga dapat dianggap sebagai aturan yang kuat dan dapat dipercaya. Selanjutnya, aturan asosiatif yang telah

terpilih dapat digunakan untuk membuat keputusan dan prediksi pada data baru.

Setelah semua pola yang memiliki frekuensi tinggi berhasil diidentifikasi, langkah selanjutnya adalah mencari aturan asosiasi yang memenuhi persyaratan minimum untuk nilai confidence dengan cara menghitung nilai confidence aturan asosiatif A B menggunakan rumus yang telah disediakan.

$$\text{Confidence} = P(B|A) = \frac{\sum \text{Transaksi Untuk A}}{\sum \text{Transaksi}}$$

Nilai confidence diperoleh dengan membagi jumlah transaksi yang mengandung item A dan item B (dengan item pertama dan item kedua yang hadir bersamaan) dengan jumlah transaksi yang mengandung item A (dengan item pertama atau item item di sebelah kiri), menurut rumus yang diberikan. Untuk menentukan aturan asosiasi mana yang harus dipilih, urutan harus berdasarkan nilai Support  $\times$  Confidence. Hasil terbesar dari n aturan yang dipilih akan diambil.

Analisis pola frekuensi tinggi merupakan teknik yang sangat penting dalam Data Mining karena memungkinkan pengguna untuk menemukan hubungan atau pola yang tidak terlihat pada data mentah. Dengan demikian, teknik ini dapat memberikan wawasan yang berharga bagi pemilik bisnis dalam mengambil keputusan yang cerdas dan tepat.

## **2.2 Kelebihan Dan Kekurangan Asosiation Rule**

### **2.2.1 Kelebihan Asosiation Rule**

Asosiasi rules (aturan asosiasi) adalah teknik data mining yang digunakan untuk menemukan hubungan yang kuat antara item atau variabel dalam dataset. Beberapa kelebihan dari teknik ini adalah:

1. Mengidentifikasi pola tersembunyi Asosiasi rules memungkinkan pengguna untuk mengidentifikasi pola tersembunyi dalam data, yang mungkin tidak dapat ditemukan melalui analisis manual. Ini dapat membantu pengguna memahami hubungan antara berbagai item atau variabel dalam dataset.
2. Mendukung pengambilan keputusan yang lebih baik Dengan menggunakan asosiasi rules, pengguna dapat membuat keputusan yang lebih baik berdasarkan pola hubungan yang ditemukan dalam data. Misalnya, jika sebuah toko menemukan bahwa banyak pelanggan yang membeli roti juga membeli selai kacang, toko tersebut dapat memutuskan untuk menempatkan selai kacang di dekat roti untuk mendorong pelanggan membeli keduanya.
3. Mengurangi biaya dan waktu Asosiasi rules dapat membantu organisasi mengurangi biaya dan waktu yang diperlukan untuk memperoleh informasi tentang hubungan antara item atau variabel dalam dataset. Ini karena proses asosiasi rules dapat dilakukan secara otomatis oleh algoritma, sehingga menghemat waktu dan biaya yang biasanya diperlukan untuk analisis manual.
4. Berguna untuk pemasaran dan promosi Asosiasi rules dapat membantu organisasi dalam pemasaran dan promosi, karena dapat memberikan informasi tentang item atau produk yang sering dibeli

bersama. Dengan mengetahui pola hubungan antara item atau produk, organisasi dapat mengembangkan strategi pemasaran dan promosi yang lebih efektif.

5. Dapat digunakan dalam berbagai bidang Asosiasi rules dapat digunakan dalam berbagai bidang, seperti pemasaran, penjualan, keuangan, ilmu sosial, dan lain-lain. Hal ini karena teknik ini dapat menemukan pola hubungan antara item atau variabel dalam dataset apa pun.

Dalam keseluruhan, asosiasi rules adalah teknik data mining yang berguna untuk menemukan pola hubungan antara item atau variabel dalam dataset. Beberapa kelebihan dari teknik ini termasuk mengidentifikasi pola tersembunyi, mendukung pengambilan keputusan yang lebih baik, mengurangi biaya dan waktu, berguna untuk pemasaran dan promosi, dan dapat digunakan dalam berbagai bidang.

### **2.2.2 Kekurangan Asosiation Rule**

Meskipun memiliki banyak kelebihan, teknik data mining asosiasi rules juga memiliki beberapa kekurangan, di antaranya:

1. Memerlukan data yang besar Asosiasi rules memerlukan data yang besar untuk memberikan hasil yang akurat dan bermakna. Jika dataset yang digunakan terlalu kecil, pola hubungan antara item atau variabel yang ditemukan mungkin tidak relevan atau tidak signifikan.

2. Keterbatasan dalam mengatasi variabel yang kontinu Asosiasi rules tidak efektif dalam menangani variabel kontinu, karena teknik ini hanya berlaku untuk data kategorikal atau diskrit. Sebagai hasilnya, teknik ini tidak dapat menangani jenis data yang terdapat dalam data ilmiah atau teknik yang memerlukan variabel kontinu.
3. Tidak dapat menangani data yang tidak terstruktur Asosiasi rules tidak dapat menangani data yang tidak terstruktur atau data yang tidak memiliki format tertentu, seperti data teks atau data multimedia. Hal ini karena teknik ini memerlukan data yang terstruktur agar dapat bekerja dengan baik.
4. Banyaknya asosiasi palsu Teknik asosiasi rules dapat menemukan banyak asosiasi yang mungkin hanya bersifat kebetulan atau tidak signifikan. Karena itu, diperlukan analisis yang lebih lanjut untuk menentukan apakah asosiasi yang ditemukan bermanfaat dan signifikan atau tidak.
5. Tidak dapat menangani variasi dalam pola asosiasi Asosiasi rules tidak dapat menangani variasi dalam pola asosiasi. Misalnya, jika ada pola asosiasi yang terjadi hanya pada hari-hari tertentu atau pada waktu tertentu, teknik ini tidak dapat menangkap pola tersebut.

Meskipun memiliki beberapa kekurangan, teknik asosiasi rules tetap merupakan alat yang berguna dalam data mining. Sebelum menggunakannya, perlu mempertimbangkan kelemahan dan kekuatan teknik ini dan memastikan bahwa teknik ini sesuai dengan tujuan dan jenis data yang akan dianalisis.

## BAB III

### Pengolahan Data Ekspor Terhadap Ragam Produk Kelapasawit Menggunakan Metode Association Rules

#### 3.1 Latar Belakang

Kelapa Sawit merupakan tanaman perkebunan penghasil minyak selain kelapa yang tumbuh di hutan hujan tropis. Habitat asli kelapa sawit aslinya berasal dari Afrika Barat dan Afrika Tengah (Angola dan Gambia) [1], namun benih kelapa sawit di Indonesia berawal dari 4 benih yang dibawa oleh Dr. D. T. Pryce [2], dua benih kelapa sawit berasal dari Bourbon, Mauritius dan dua benih lainnya berasal dari Hortus Botanicus, Amsterdam, Belanda pada tahun 1848 [3]. Kelapa sawit digunakan sebagai bahan baku penghasil minyak masak, minyak industri, maupun bahan bakar. Indonesia adalah penghasil minyak kelapa sawit terbesar di Indonesia. Salah satu perusahaan yang bergerak di bidang produksi minyak kelapa sawit adalah PT Permata Hijau Group. Perusahaan ini memproduksi berbagai ragam produk kelapa sawit mulai dari Palm Oil, Lauric Oil, Biodiesel, dan Fatty Acid Oleo. Melihat ragamnya produk kelapa sawit yang dipasarkan oleh PT Permata Hijau Group, maka perlu dilakukan pengambilan keputusan terhadap data produk kelapa sawit.

Data mining merupakan bidang ilmu yang melakukan proses menggali informasi dari basis data yang berjumlah besar dan diolah menjadi informasi baru yang dapat membantu dalam pengambilan keputusan [4]. Data mining menggunakan teknik statistik, matematika, kecerdasan buatan dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [5]. Association rules merupakan salah satu metode yang

bertujuan mencari pola yang sering muncul diantara banyak transaksi, dimana setiap transaksi terdiri dari beberapa item [6]. Analisis asosiasi atau association rule mining adalah teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi item. Analisis asosiasi dikenal juga sebagai salah satu teknik data mining yang menjadi dasar dari salah satu teknik data mining lainnya [7].

Penggunaan metode Association Rules sangat sesuai untuk pengklasifikasian item-item transaksi yang ada. Maka aplikasi yang akan dirancang ini diharapkan dapat membantu proses data mining untuk pengolahan data ekspor kelapa sawit [8]. Association rule merupakan salah satu metode yang bertujuan mencari pola yang sering muncul di antara banyak transaksi, dimana setiap transaksi terdiri dari beberapa item sehingga metode ini akan mendukung sistem rekomendasi melalui penemuan pola antar item dalam transaksi-transaksi yang terjadi [9]. Dari latar belakang diatas peneliti tertarik melakukan penelitian untuk mengetahui ragam produk kelapa sawit yang paling banyak diekspor, hasil ekspor kelapa sawit dalam interval waktu tertentu. Maka dari itu, judul yang diambil yaitu Pengolahan Data Ekspor Terhadap Ragam Produk Kelapa Sawit Menggunakan Metode Association Rules. Hasil dari metode Association rules yang digunakan dapat memprediksi adanya kemungkinan ragam kelapa sawit tertentu yang paling laku di pasar internasional.

### **3.2 Rumusan Masalah**

Adapun masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana hasil metode Association Rules berpengaruh terhadap data ekspor produk kelapa sawit.

2. Bagaimana memprediksi kombinasi produk yang lebih diminati oleh pasar internasional dengan menggunakan Algoritma Apriori.
3. Bagaimana melakukan penyajian data dengan visualisasi model data berupa grafik dengan menggunakan library python.

### **3.3 Batasan Masalah**

Adapun batasan masalah dalam penelitian ini adalah:

1. Penelitian ini menggunakan data ekspor olahan kelapa sawit.
2. Metode yang digunakan adalah Association Rules.
3. Data penelitian berasal dari data transaksi ekspor pada PT. Permata Hijau Group sebanyak 600 data periode bulan Januari 2022 sampai September 2022.

### **3.4 Tujuan Dan Manfaat Penelitian**

#### **3.4.1 Tujuan Penelitian**

Adapun tujuan dari penelitian:

1. Menganalisis data ekspor kelapa sawit.
2. Mengimplementasikan metode Association Rules pada data ekspor.
3. Memvisualisasikan hasil metode Association Rules dalam bentuk grafik.

#### **3.4.2 Manfaat Penelitian**

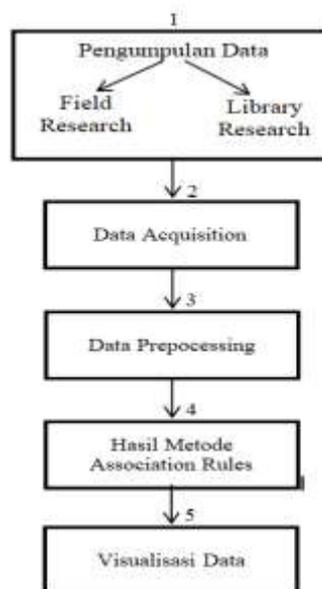
Adapun Manfaat dari penelitian:

1. Menambah keterampilan dan wawasan dari peneliti.

2. Memberikan kemudahan kepada perusahaan untuk mengolah data ekspor.
3. Memberikan informasi kepada pihak perusahaan untuk meningkatkan produksi terhadap olahan kelapa sawit tertentu yang lebih diminati pasar internasional.

### 3.5 Metode Penelitian

Penelitian ini merupakan penelitian kuantitatif, dimana data yang digunakan adalah data ekspor olahan kelapa sawit pada PT. Permata Hijau Group. Adapun Metode Penelitian ini dilakukan mengikuti tahapan-tahapan berikut ini:



**Gambar 3.1** Tahapan Alur Penelitian

#### 3.5.1 Pengumpulan Data

Dalam penelitian ini teknik pengumpulan data yang dilakukan adalah Field Research dan Library Research. Teknik Field Research bertujuan untuk mendapatkan data ekspor produk kelapa sawit yang diperoleh melalui data ekspor perusahaan, sedangkan teknik Library Research

adalah pengumpulan data yang berasal dari jurnal terdahulu yang berhubungan dengan tujuan penelitian.[10].

Tabel 3.1 Contoh Data Ekspor

No	No Transaksi	TanggalTransaksi	Negara	JenisProduk
1	IUB/22/0301/001/A	3/1/2022	Jiangmen,China	ABC
2	IU/22/0301/008/A	3/1/2022	Lianyungang, China	B
3	IM/22/0301/001/A	3/1/2022	Cat Lai, Vietnam	CDG
4	IM/22/0301/002/A	3/1/2022	Singapore	BC
5	IU/22/0401/008/A	4/1/2022	Lianyungang, China	B

### 3.5.2 Data Acquisition

Data acquisition adalah proses pengumpulan dan perekaman data dari berbagai sumber untuk tujuan analisis atau pengolahan lebih lanjut. Proses ini melibatkan pengumpulan data dari berbagai sumber, termasuk sumber internal dan eksternal, seperti sensor, perangkat pengukuran, sistem informasi, dan sumber data lainnya. Dataset ekspor produk kelapa sawit diambil melalui data perusahaan. Dataset tersebut akan diolah melalui proses pengolahan data dan digunakan untuk mengidentifikasi kombinasi produk yang diminati pasar internasional [11].

### 3.5.3 Data Preprocessing

Data preprocessing adalah proses persiapan data sebelum dilakukan analisis atau pengolahan lebih lanjut. Proses ini melibatkan serangkaian langkah untuk membersihkan, mengorganisir, dan mempersiapkan data agar dapat digunakan secara efektif dalam proses analisis data. Proses mengubah data mentah menjadi format yang mudah

dipahami. Proses ini dilakukan karena data mentah sering kali dalam bentuk format yang tidak beraturan. Tujuannya agar bisa dijadikan sumber informasi melalui sekumpulan data yang bisa diteruskan untuk diolah datanya[12].

#### **3.5.4 Hasil Metode Association Rules**

Proses menghasilkan data kombinasi produk yang memenuhi aturan asosiasi yang bisa memberikan informasi kepada perusahaan tentang kombinasi produk yang paling sering dipilih oleh pembeli [13]. Asosiasi rule atau Association Rule adalah sebuah teknik analisis data yang digunakan untuk menemukan hubungan atau pola antara item atau variabel dalam sebuah dataset. Teknik ini paling sering digunakan dalam bidang data mining dan machine learning.

#### **3.5.5 Visualisasi Data**

Visualisasi data adalah representasi visual dari data dan informasi yang dimaksudkan untuk membantu pemahaman dan interpretasi data. Tujuannya adalah untuk mengubah data yang rumit dan kompleks menjadi bentuk yang lebih mudah dipahami dan dianalisis, sehingga dapat membantu mengambil keputusan yang lebih baik dan efektif. Visualisasi data dapat dilakukan dalam berbagai bentuk, seperti grafik, diagram, peta, dan infografis. Teknik visualisasi yang digunakan tergantung pada jenis data dan tujuan dari visualisasi itu sendiri.

Visualisasi data dapat membantu kita mengungkap pola dan hubungan yang mungkin tersembunyi dalam data dan mengidentifikasi tren, anomali, dan hubungan antara berbagai variabel. Ini dapat memberikan wawasan yang berguna dalam berbagai bidang, seperti bisnis,

pemasaran, ilmu pengetahuan, teknik, kesehatan, dan banyak lagi. Akhir dari Penelitian ini adalah menggambarkan hasil dari metode Association Rules ke dalam bentuk grafik sehingga lebih mudah dipahami.

### 3.6 Analisis Masalah

Adapun analisis masalah pada penelitian kali ini adalah untuk mencari ragam olahan kelapa sawit yang paling banyak diekspor, pengujian dilakukan melalui data penjualan selama 1 bulan terakhir. Penelitian ini menggunakan metode *Association Rules* untuk menentukan nilai *Support* dan *Confidence* dari setiap ragam olahan kelapa sawit, dari hasil akhir perhitungan tersebut, data akhir akan divisualisasikan dengan menggunakan *Library Python*.

### 3.7 Analisis Data

Tabel 3.2 Sampel Tabel Transaksi Ekspor

No	No Transaksi	Tanggal Transaksi	Negara	Jenis Produk
1	IUB/22/0301/001/A	3/1/2022	Jiangmen, China	ABC
2	IU/22/0301/008/A	3/1/2022	Lianyungang, China	B
3	IM/22/0301/001/A	3/1/2022	Cat Lai, Vietnam	CDG
4	IM/22/0301/001/A	3/1/2022	Singapore	BC
5	IU/22/0401/001/A	4/1/2022	Lianyungang, China	B
6	IU/22/0401/008/A	4/1/2022	Baltimore, USA	BC
7	IU/22/0401/014/A	4/1/2022	Qingdao, China	DEF
8	IU/22/0501/005/A	5/1/2022	Qingdao, China	BCG

9	IM/22/0501/004/A	5/1/2022	Laem Chabang, Thailand	D
10	IU/22/0501/001/A	5/1/2022	Huangpu, China	E
11	IT/22/0501/002/A	5/1/2022	Tashkent, Uzbekistan	CFH
12	IUB/22/0501/001/A	5/1/2022	Qingdao, China	CDFG
13	IU/22/0601/010/A	6/1/2022	Bangkok, Thailand	BCDF
14	IU/22/0601/006/A	6/1/2022	Manzanillo, Mexico	B

Tabel 3.2 Sampel Tabel Transaksi Ekspor merupakan sampel data yang diambil dari "Tabel Transaksi Ekspor". "Tabel Transaksi Ekspor" adalah tabel transaksi ekspor yang terjadi dari bulan Januari 2022 sampai September 2022. Data yang diambil merupakan data transaksi ekspor yang akan digunakan untuk melakukan penelitian. Nama produk olahan kelapa sawit diinisialkan oleh penulis agar lebih mudah dalam proses analisis dengan metode Association rules.

Inisialnya yaitu :

A = Glycerine

B = Refined Bleached and Deodorised Palm Kernel Oil (RBDPO) C =

Stearic Acid

D = Fatty Acid E = Lauric Acid

F = Shortening 36-39

G = Hydrogenated Palm Kernel Oil (HPKO)

H = Hydrogenated Palm Kernel Stearine (HPKS)

Data transaksi ekspor produk olahan kelapa sawit dapat divisualisasikan kedalam bilangan biner seperti data yang terlihat dalam

"Tabel Reprersetasi Biner". Setiap kolom merupakan representasi olahan produk kelapa sawit dan setiap baris merupakan representasi setiap transaksi, nilai 1 menunjukkan olahan produk kelapa sawit ada dalam transaksi tersebut, atau 0 jika produk tersebut tidak terdapat dalam transaksi.

Setelah mendapatkan data berupa nilai biner, langkah selanjutnya adalah menentukan nilai support dan nilai confidence dari data transaksi ekspor produk olahan kelapa sawit. Penulis menggunakan aturan asosiasi dimana bentuk pernyataannya adalah  $X \rightarrow Y$ , dimana X dan Y adalah item set yang dipisahkan oleh kondisi if-then (hypothesis-conclusion). Hasil kekuatan metode Association Rules dapat dilihat dari nilai support dan confidence. Support digunakan untuk mengetahui seberapa banyak kombinasi yang dapat terjadi dalam suatu data, sedangkan confidence digunakan untuk mengetahui seberapa sering item set Y muncul dalam transaksi yang memiliki item set X.

Formula yang dapat digambarkan untuk menghitung nilai support dan confidence adalah sebagai berikut:

$$\text{Support} = \frac{\text{Frekuensi } (x \& y)}{N} \dots\dots(1)$$

$$\text{Confidence} = \frac{\text{Frekuensi } (x \& y)}{\text{Frekuensi } x} \dots\dots(2)$$

N = Jumlah Transaksi

Dari formula diatas dapat diimplementasikan kedalam data dengan contoh, X[B,C] dan Y [D]. Jumlah transaksi yang berisi XUY=[B,C,D] adalah 60, dan jumlah transaksi N adalah 600, sehingga support untuk nilai  $s(X \rightarrow Y) = 60/600 = 0.1$ . Nilai Confidence didapatkan dari pembagian antara pembagian antara XUY=[B,C,D] dengan X[B,C], dimana

$XUY=[B,C,D] = 60$  dan  $X[B,C] = 210$ , sehingga nilai confidence  $c(X \rightarrow Y) = 60/210 = 0.28 = 28\%$ .

Berdasarkan formula (1) dari data transaksi ekspor yang dipakai untuk mendapatkan sampel perhitungan dengan metode association rules maka diketahui nilai support minimum sebesar 0.07, sehingga nilai supcount ditunjukkan pada tabel berikut:

Tabel 3.3 Daftar *Frequent 1 Item Set*

No	Item	Support	Supcount	Supcount(%)
1	A	54	0.09	9%
2	B	336	0.56	56%
3	C	312	0.52	52%
4	D	216	0.36	36%
5	E	78	0.13	13%
6	F	78	0.13	13%
7	G	108	0.18	18%
8	H	72	0.12	12%

Tabel 3.4 Calon *frequent 2 item set*

No	Item	Support	Supcount	Supcount(%)
1	A,B	48	0,08	8%
2	A,C	54	0,09	9%
3	A,D	18	0,03	3%
4	A,E	0	0	0%
5	A,F	0	0	0%
6	A,G	0	0	0%
7	A,H	0	0	0%

8	B,C	210	0,35	35%
9	B,D	72	0,12	12%
10	B,E	0	0	0%
11	B,F	6	0,01	1%
12	B,G	60	0,1	10%
13	B,H	30	0,05	5%
14	C,D	114	0,19	19%
15	C,E	0	0	0%
16	C,F	36	0,06	6%
17	C,G	102	0,17	17%
18	C,H	60	0,1	10%
19	D,E	24	0,04	4%
20	D,F	36	0,06	6%
21	D,G	66	0,11	11%
22	D,H	18	0,03	3%
23	E,F	24	0,04	4%
24	E,G	0	0	0%
25	E,H	0	0	0%
26	F,G	6	0,01	1%
27	F,H	36	0,06	6%
28	G,H	0	0	0%

Tabel 3.5 Daftar frequent 2 item set

No	Item	Support	Supcoun t	Supcount(% )
1	A,B	48	0,08	8%
2	A,C	54	0,09	9%

3	B,C	210	0,35	35%
4	B,D	72	0,12	12%
5	B,G	60	0,1	10%
6	C,D	114	0,19	19%
7	C,G	102	0,17	17%
8	C,H	60	0,1	10%
9	D,G	66	0,11	11%

Tabel 3.6 Calon frequent 3item set

No	Item	Support	Supcount	Supcount(%)
1	A,B,C	48	0,08	8%
2	A,B,D	12	0,02	2%
3	A,B,G	0	0	0%
4	A,C,D	18	0,03	3%
5	A,C,G	0	0	0%
6	A,C,H	0	0	0%
7	A,D,G	0	0	0%
8	B,C,D	60	0,1	10%
9	B,C,G	54	0,09	9%
10	B,C,H	30	0,05	5%
11	B,D,G	24	0,04	4%
12	C,D,G	60	0,1	10%
13	C,D,H	18	0,03	3%

Tabel 3.7 Daftar frequent 3 item set

No	Item	Support	Supcount	Supcount(%)
1	A,B,C	48	0,08	8%
2	B,C,D	54	0,09	9%
3	B,C,G	54	0,09	9%
4	C,D,G	60	0,1	10%

Tabel 3.7 Daftar *frequent 4 item set*

No	Item	Support	Supcount	Supcount(%)
1	A,B,C,D	12	0,02	2%
2	A,B,C,G	0	0	0%
3	B,C,D,G	18	0,03	3%

Dari tabel dengan 4 item set tidak efisien dengan data ekspor olahan kelapa sawit, sehingga rule item produk hanya berdasarkan 3 item set saja. Dalam hal ini, penulis menetapkan nilai support minimum adalah  $\geq 0.07$ . Selanjutnya untuk mengetahui hasil data association rules adalah dengan menghitung nilai confidence.

Dengan formula 2 dan confidence yang didapatkan dengan menetapkan nilai confidence minimum  $\geq 70\%$  maka hasil terlihat pada tabel berikut.

Tabel 3.8 Daftar Association Rules

No	Frequent Item Set	Hasil Aturan Asosiasi	Support	Confidence
1	A,B	Jika beli A, maka	48	48/54   88%

		beli B			
2	A,C	Jika beli A, maka beli C	54	54/54	100%
3	G,C	Jika beli G maka beli C	102	102/108	94%
4	H,C	Jika beli H maka beli C	60	60/72	83%
5	AB,C	Jika beli AB maka beli C	48	48/48	100%
6	AC,B	Jika beli AC maka beli B	48	48/54	88%
7	A,BC	Jika beli A maka beli BC	48	48/54	88%
8	BD,C	Jika beli DB maka beli C	60	60/72	83%
9	BG,C	Jika beli GB maka beli C	54	54/60	90%
10	DG,C	Jika beli GD maka beli C	60	60/66	90%

Dari hasil rule diatas, dapat diketahui kombinasi olahan produk kelapa sawit yang paling sering diekspor berdasarkan nilai minimum support dan confidence yang ditetapkan.

### 3.8 Visualisasi Data

Visualisasi Data adalah tampilan berupa grafis atau grafik dari sebuah informasi sehingga kumpulan data menjadi lebih sederhana untuk ditampilkan. Visualisasi data dapat membantu kita mengungkap pola dan

hubungan yang mungkin tersembunyi dalam data dan mengidentifikasi tren, anomali, dan hubungan antara berbagai variabel. Ini dapat memberikan wawasan yang berguna dalam berbagai bidang, seperti bisnis, pemasaran, ilmu pengetahuan, teknik, kesehatan, dan banyak lagi.

### 3.8.1 Upload Dataset

```
[1] import pandas as pd

[2] import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

[3] from mlxtend.preprocessing import TransactionEncoder

[4] from mmap import ACCESS_DEFAULT
dataset_manual=[[ 'A', 'B', 'C'],
                 ['B'],
                 ['C', 'D', 'G'],
                 ['B', 'C'],
                 ['B'],
                 ['B', 'C'],
                 ['D', 'E', 'F'],
                 ['B', 'C', 'G'],
                 ['D'],
                 ['E'],
                 ['C', 'F', 'H'],
                 ['C', 'D', 'F', 'G'],
                 ['B', 'C', 'D', 'E'],
                 ['B']]
```

Gambar 3.2 Menginput data

Hal pertama yang harus dilakukan adalah memasukkan data kedalam sebuah variable dengan menggunakan *Google Colaboratory*.

### 3.8.2 Import Library Association Rules Dan Menampilkan Dataset

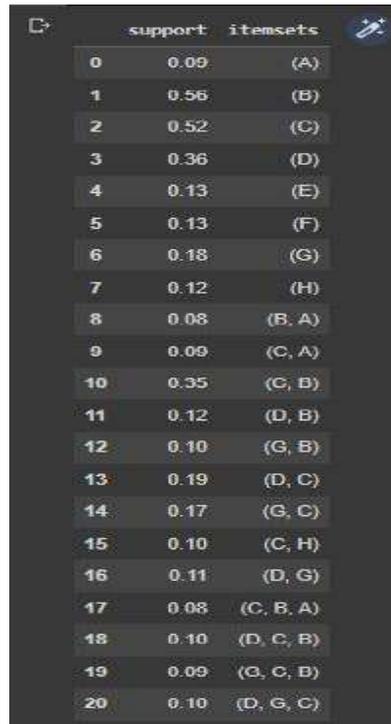
```
[5] te = TransactionEncoder()
te_ary= te.fit(dataset_manual).transform(dataset_manual)

[6] df_mlx= pd.DataFrame(te_ary, columns=te.columns_)

▶ from mlxtend import frequent_patterns
from mlxtend.frequent_patterns import apriori
frequent_itemsets = apriori(df_mlx, min_support=0.07, use_colnames=True)
frequent_itemsets
```

Gambar 3.3 Menampilkan Daftar Itemsets

Setelah melakukan input data, langkah selanjutnya adalah melakukan pemanggilan terhadap library apriori dan menentukan parameter minimum support sesuai dengan kebutuhan yaitu 0.07 atau 7%, sehingga daftar Itemsets yang memenuhi nilai minimum support =0.07 dapat dilihat pada Gambar 3.3 Daftar Itemsets.



	support	itemsets
0	0.09	(A)
1	0.56	(B)
2	0.52	(C)
3	0.36	(D)
4	0.13	(E)
5	0.13	(F)
6	0.18	(G)
7	0.12	(H)
8	0.08	(B, A)
9	0.09	(C, A)
10	0.35	(C, B)
11	0.12	(D, B)
12	0.10	(G, B)
13	0.19	(D, C)
14	0.17	(G, C)
15	0.10	(C, H)
16	0.11	(D, G)
17	0.08	(C, B, A)
18	0.10	(D, C, B)
19	0.09	(G, C, B)
20	0.10	(D, G, C)

**Gambar 3.4** Daftar Itemsets

```
[8] from mlxtend.frequent_patterns import association_rules
result_1 = association_rules(frequent_itemsets,metric="confidence",min_threshold=0.7)
result_1[['antecedents','consequents','support','confidence']]
```

**Gambar 3.5** Menampilkan Daftar Association Rules

Menentukan parameter minimum confidence sesuai kebutuhan yaitu 0.7 atau 70%, sehingga itemsets yang memenuhi nilai minimum confidence =0.7 dapat dilihat pada Gambar 3.5 Daftar Association Rules.

	antecedents	consequents	support	confidence
0	(A)	(E)	0.08	0.868889
1	(A)	(C)	0.09	1.000000
2	(G)	(C)	0.17	0.944444
3	(H)	(C)	0.10	0.833333
4	(B, A)	(C)	0.08	1.000000
5	(C, A)	(B)	0.08	0.868889
6	(A)	(B, C)	0.08	0.888889
7	(D, B)	(C)	0.10	0.833333
8	(B, G)	(C)	0.09	0.900000
9	(D, G)	(C)	0.10	0.909091

Gambar 3.6 Daftar Association Rules

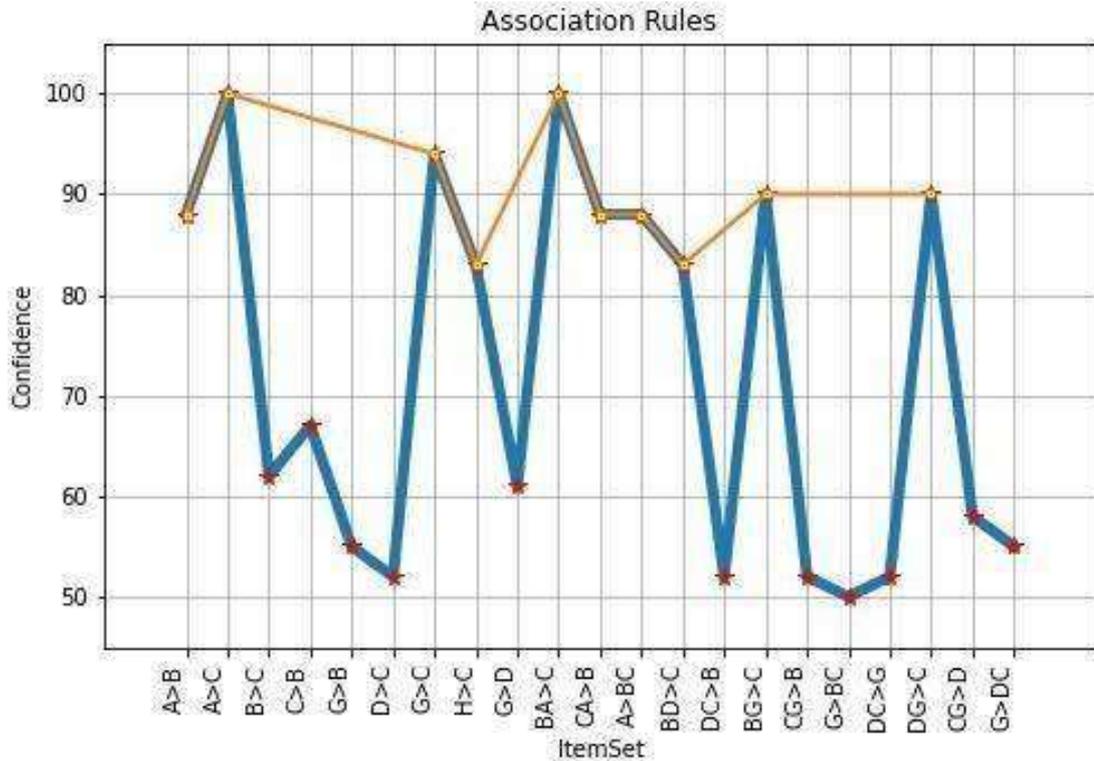
### 3.8.3 Visualisasi Data dengan Matplotlib

Visualisasi dataset Association Rules kedalam lineplot yang membandingkan antara daftar calon Association Rules dengan daftar Association Rules, data yang memenuhi aturan asosiasi ditandai dengan kedua grafik yang saling bersinggung dan titik yang memiliki dua simbol.

```
Itemsets=daftar['Association']
Confidence=daftar['confidence']
Fix_sets=fix['Association']
Fix_conf=fix['confidence']
plt.xticks(rotation = 90, ha = 'right')
plt.margins(0.1)
plt.grid(axis="x")
plt.grid(axis="y")
plt.plot(Itemsets,Confidence,linewidth=5,marker='*',markersize=8,mfc='green',mew=1,mec='red')
plt.plot(Fix_sets,Fix_conf,linewidth=2,marker='x',markersize=3,mfc='red',mew=1,mec='yellow')
plt.xlabel("ItemSet")
plt.ylabel("Confidence")
plt.title("Association Rules")
plt.ylim(45,105)
plt.rcParams["figure.figsize"] = [7,5]
plt.rcParams["figure.autolayout"] = True

y = np.linspace(-15, 15, 100)
```

Gambar 3.7 Menampilkan Grafik Lineplot



**Gambar 3.8** Grafik Association Rules Lineplot

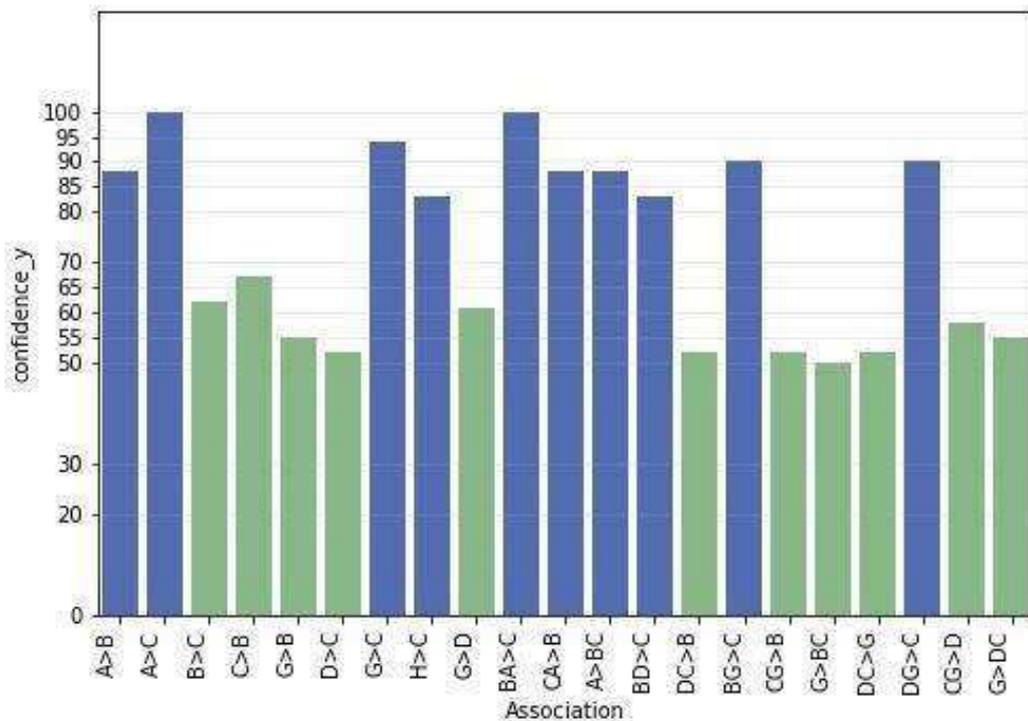
### 3.8.4 Visualisasi Data dengan Seaborn

Visualisasi dataset Association Rules kedalam barplot yang membandingkan antara daftar calon Association Rules dengan daftar Association Rules, data yang memenuhi aturan asosiasi ditandai dengan bar berwarna biru, sedangkan data yang belum memenuhi aturan asosiasi ditandai dengan bar berwarna hijau.

```
fix_zeros={"Association":daftar['Association'], "confidence":[0 for i in range (len(daftar['Association']))]}
fix_zeros_df=pd.DataFrame(fix_zeros)
fix_zeros_df
fix_merge=fix_zeros_df.merge(fix,on='Association',how='left')
fix_merge=fix_merge[['Association','confidence_y']]
fix_merge=fix_merge.fillna(0)
a = sns.barplot(data=daftar,x='Association', y='confidence',color = 'green', alpha=0.5)
b = sns.barplot(data=fix_merge,x='Association', y='confidence_y',color = 'blue', alpha=0.5)
ticks = [0,20,30,50,55,60,65,70,80,85,90,95,100]
plt.ylim(0,120)
plt.grid(axis="y",alpha=0.25)
a.set_yticks(ticks)
a.set_yticklabels(ticks)

plt.xticks(rotation = 90, ha = 'right')
plt.savefig('seaborn.png')
plt.show()
```

**Gambar 3.9** Menampilkan Grafik Barplot



**Gambar 3.10** Grafik Association Rules Barplot

Setelah melakukan visualisasi data menggunakan Library Python (Matplotlib dan Seaborn), diperoleh data bahwa produk dengan item set Glycerine→Stearic Acid dan Glycerine-RBDPO→Stearic Acid merupakan kombinasi Association Rules yang paling sering muncul yaitu 100%. Sedangkan untuk kombinasi HPKO→Stearic Acid memenuhi Association Rules dengan nilai 94%, RBDPO-HPKO→Stearic Acid dan Fatty Acid-

HPKO→Stearic Acid memenuhi Association Rules dengan nilai 90%, Glycerine→RBDPO, Glycerine→RBDPO-Stearic Acid, dan Glycerine Stearic Acid→RBDPO memenuhi Association Rules dengan nilai 88%, HPKS→Stearic Acid dan RBDPO-Fatty Acid→Stearic Acid memenuhi Association Rules dengan nilai 83%.

### **3.9 Kesimpulan Dan Saran**

#### **3.9.1 Kesimpulan**

Berdasarkan hasil yang didapatkan dari metode Association Rules pada data ekspor ragam olahan kelapa sawit, didapatkan jenis olahan kelapa sawit memenuhi aturan asosiasi yaitu produk kombinasi Glycerine-RBDPO- Stearic Acid merupakan kombinasi yang paling sering muncul yakni 100%, Glycerine-RBDPO dengan nilai asosiasi 88%, RBDPO- Stearic Acid dengan nilai asosiasi 62%, dan Stearic Acid- Fatty Acid- HPKO dengan nilai 52%. Oleh karena itu, olahan kelapa sawit Glycerine, RBDPO, Stearic Acid,HPKO, dan Fatty Acid merupakan produk yang harus ditingkatkan jumlah produksi mengingat permintaan terhadap kombinasi ragam olahan produk kelapa sawit yang lebih tinggi dibandingkan ragam produk olahan kelapa sawit lainnya .

#### **3.9.2 Saran**

Adapun saran-saran yang dapat disampaikan berdasarkan hasil pengamatan dan analisa selama penelitian yaitu:

1. Perlu penelitian lebih lanjut dengan menggunakan metode lain seperti Naïve Bayes, C4.5, dan K-Means agar memperoleh perbandingan dengan tingkat akurasi lebih tinggi.

2. Dalam menggunakan algoritma Association Rules, penulis harus menentukan nilai minimum untuk confidence dan support yang tepat, supaya hasil kombinasi yang memenuhi aturan asosiasi lebih spesifik.

## DAFTAR PUSTAKA

- [1] Lubis, R. E., & Agus Widanarko, S. P. (2011). Buku pintar kelapa sawit. AgroMedia.
- [2] Sipayung, D. I. T. (2019). Ekonomi agribisnis minyak sawit. PT Penerbit IPB Press.
- [3] Saputra, N., Abdinagoro, S. B., & Kuncoro, E. A. (2020). SUSTAINABLE GROWTH FORMULA: keterlekatan kerja dan ketangkasan belajar dari managerial resources industri minyak sawit Indonesia. SCOPINDO MEDIA PUSTAKA.
- [4] Samosir, W. A. H., & Tiur Gantini, S.T. (2022). Analisis Dataset COVID-19 menggunakan Algoritma KNN dan Random Forest. Jurnal STRATEGI-Jurnal Maranatha, 4(1), 58-69.
- [5] Emha Taufik, Luthfi, 2009, Penerapan Data Mining Algoritma Asosiasi Untuk Meningkatkan Penjualan, jurnal, Yogyakarta, STMIK AMIKOM Yogyakarta.
- [6] Nurdin, N., & Astika, D. (2019). Penerapan Data Mining Untuk Menganalisis Penjualan Barang Dengan Menggunakan Metode Apriori Pada Supermarket Sejahtera Lhokseumawe. TECHSI-Jurnal Teknik Informatika, 7(1), 132-155.
- [7] Fauzy, M., & Asror, I. (2016). Penerapan metode association rule menggunakan algoritma apriori pada simulasi prediksi hujan wilayah kota bandung. Jurnal Ilmiah Teknologi Infomasi Terapan, 2(3).
- [8] Puspasari, R., & Buluran, I. Y. (2015). Penerapan Data Mining untuk Mengolah Data Impor - Ekspor Ikan dengan Menggunakan Metode Association Rule. SEMNASTEKNOMEDIA ONLINE, 3(1), 2-1.

- [9] Hariri, F. R., & Ramadhani, R. A. (2017). Penerapan Data Mining menggunakan Association Rules untuk Mendukung Strategi Promosi Universitas Nusantara PGRI Kediri. *Prosiding SNATIKA*, 4, 138-142.
- [10] Wulandari, S., & Lubis, A. S. (2019). Analisis Perkembangan Ekspor Impor Barang Ekonomi di Provinsi Sumatera Utara. *Jurnal Administrasi Bisnis*, 8(1), 31-36.
- [11] Mahzalena, Y., & Juliansyah, H. (2019). Pengaruh inflasi, pengeluaran pemerintah dan ekspor terhadap pertumbuhan ekonomi di Indonesia. *Jurnal Ekonomi Regional Unimal*, 2(1), 37- 50.
- [12] Widians, J. A., & Rizkyani, F. N. (2020). Identifikasi Hama Kelapa Sawit menggunakan Metode Certainty Factor. *ILKOM Jurnal Ilmiah*, 12(1), 58-63.
- [13] Nugraha, M. Q., Supriadi, S., & Sabrina, T. (2015). Survei Pemetaan P-Potensial Dan PTersediaTerhadap Produksi Tanaman KelapaSawit (*Elaeis Guinensis* Jacq.) Di Perkebunan Pt. Buana Estate Kabupaten Langkat. *AGROEKOTEKNOLOGI*, 3(4).