

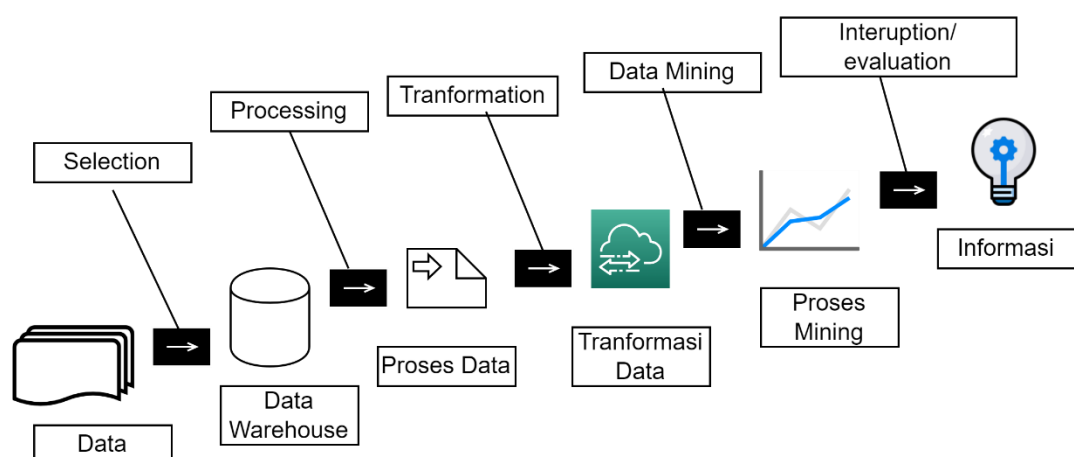
BAB 2

TINJAUAN PUSTAKA

2.1 Data Mining

Data Mining adalah teknik penggalian informasi dari volume data yang sangat besar dengan mencari pola dan hubungan yang tersembunyi di dalam jumlah data yang besar. Output data mining ini dapat digunakan untuk membantu dalam membuat keputusan berdasarkan kumpulan data besar. Pattern recognition telah menurun penggunaannya sebagai akibat munculnya Data Mining atau KDD, karena telah menjadi komponen Data Mining. Teknik pemrosesan data mining sering kali dalam statistic, matematika, kecerdasan buatan, dan machine learning (Sholikin 2019)

Berikut adalah gambar Proses data mining :



Gambar 2.1 Data Mining

Tahapan-tahapan proses KDD sebagai berikut;

1. Pembersih data (data cleaning) : Pembersihan data mencakup aktivitas seperti menghilangkan kebisingan dan data yang salah atau tidak relevan.. Data dari database dengan atribut dan item yang tidak lengkap, serta data yang hilang, berisi data yang salah atau hanya data yang salah saja. Juga disarankan agar data yang tidak berguna tersebut dihapus. Kecanggihan dalam temuan data dan teknik penambangan data sangat tergantung pada proses pembersihan data, karena kekurangan dalam data yang diproses mengurangi jumlah dan kerumitannya berkurang.
2. Integrasi data (data integration) yaitu proses pengumpulan data ke dalam database baru. Integrasi data harus dilakukan dengan teliti untuk menghindari kesalahan yang dapat membahayakan hasil pengambilan keputusan.
3. Seleksi data (data selection) Pemilihan data yang sesuai dalam proses analisis database, langkah ini dilakukan karena beberapa data dalam database seringkali tidak diperlukan.
4. Transformasi data (data transformation) yaitu data diubah menjadi format yang sesuai untuk diproses. Salah satu metode standar, misalnya, analisis asosiasi dan klustering, hanya dapat menerima input data kategoris. Akibatnya, perlu dilakukan pembagian data berupa bilangan kontinu ke dalam suatu interval.
5. Proses mining merupakan metode prosedur utama dalam penelitian ini, menggunakan pendekatan yang bertujuan untuk menemukan pengetahuan penting dan tersembunyi.

6. Evaluasi pola (pattern evaluation) Temuan fase ini dapat diidentifikasi dalam bentuk pola yang berbeda dalam bentuk analisis berbasis pengetahuan dan prediksi dan model evaluasi untuk menentukan apakah asumsi benar-benar direalisasikan atau sebaliknya.
7. Presentasi pengetahuan (knowledge presentation) Metode yang dilakukan untuk memperoleh pengetahuan melalui penyajiannya secara visualisasi. Tahap terakhir adalah proses transformasi hasil analisis menjadi keputusan dan tindakan. Setelahnya diharapkan penyajian kedalam bahasa pengetahuan dapat dipahami manusia umum sebagai salah satu tahapan yang diperlukan. Presentasi 6 juga diharapkan dapat membantu dalam menampilkan hasil data mining dalam presentasi ini.

2.2 Association Rules

Association rule adalah metode data mining yang berupaya untuk mencari sekumpulan items yang sering muncul bersamaan. Aturan asosiasi ini biasanya disebut Market Basket Analysis (Analisa Keranjang Pasar)(Chen et al. 2005). Dari keranjang belanja para pengunjung supermarket dapat dipelajari, barang-barang mana yang sering dibeli bersama dan mana yang tidak. Association rule biasanya ditulis dalam bentuk pernyataan IF-THEN, yang menggabungkan beberapa items menjadi satu(Lestari and Zakiyyah 2020). Aturan asosiasi juga merupakan metode pengembangan data yang mengidentifikasi semua aturan asosiasi dalam database yang memenuhi standar dukungan minimal (minsup) dan kepercayaan (minconf). Kedua kondisi ini, yaitu minsup dan minconf, akan dimanfaatkan untuk

menerapkan aturan asosiasi yang lebih menarik daripada batasan yang telah ditentukan. Berdasarkan definisi ini, pencarian pola aturan asosiasi menggunakan dua ukuran dukungan dan kepercayaan diri dengan nilai mulai dari 0% hingga 100%.

Berikut penjelasan mengenai support dan confidence (Yoo and Kim 2023).

Dalam menentukan suatu association rule, terdapat ukuran yang menentukan bahwa informasi atau knowledge dianggap menarik (interestingness measure).

Ukuran ini berasal dari hasil pengolahan data menggunakan rumus tertentu.

Berikut rumus support dan confidence sebagai berikut:

- SUPPORT

Support adalah nilai yang menunjukkan seberapa sering item atau kumpulan item muncul dalam keseluruhan transaksi.

Nilai Support dari 1 item diperoleh dengan menggunakan rumus berikut;

$$\text{Support (A)} : \frac{\text{Jumlah Transaksi Mengandung A}}{\text{Total Transaksi}} \times 100\%$$

Sedangkan nilai support dari 2 item diperoleh dari rumus berikut :

$$\text{Support (A} \cap \text{B)} : \frac{\sum \text{Jumlah Transaksi Mengandung A dan B}}{\sum \text{Total Transaksi}} \times 100\%$$

- CONFIDENCE

Confidence adalah nilai menunjukkan hubungan antara dua objek, khususnya seberapa sering item B muncul dalam transaksi termasuk item A. (Lestari and Zakiyyah 2020) (Hossain et al. 2017)

Nilai pada Confidence diperoleh dengan menggunakan rumus berikut;

$$\text{Confidence} : \frac{\sum \text{Jumlah Transaksi Mengandung A dan B}}{\sum \text{Jumlah Transaksi Mengandung A}} \times 100\%$$

2.3 Frequent Pattern Growth (FP-Growth)

Frequent Pattern Growth yaitu algoritma yang sering digunakan untuk menemukan hubungan asosiasi dalam data yang terdiri dari beberapa kelompok data dengan memeriksa pola yang sering muncul secara bersamaan (Lestari and Zakiyyah 2020). Algoritma Frequent Pattern Growth (FP-Growth) dikembangkan untuk menutupi kekurangan dari algoritma apriori. Proses pemindahan database algoritma apriori membutuhkan waktu yang lama dan dapat berdampak pada efisiensi waktu dalam proses data mining. Algoritma FP-Growth dianggap lebih efisien karena tidak memerlukan candidate item sets. FP-Growth membaca setiap transaksi dalam kumpulan data satu per satu dan memetakan setiap transaksi ke kelompok FP-Tree yang dibuat.

FP-Growth menggunakan struktur pohon data yang disebut dengan FP-Tree. Algoritma FP growth merupakan perkembangan dari algoritma apriori yang menggunakan prinsip divide-and-conquer dengan mengkompresi database dan menyajikan kumpulan itemset ke dalam bentuk FP-Tree yang menyimpan informasi asosiasi antar itemset (Lestari 2015). Setelah itu, database terkompresi dipisahkan menjadi berbagai set database kondisional yang berbeda, dengan masing-masing item set saling terkait dengan satu frequent item atau “pattern fragment” dan setiap database dianalisis secara terpisah. Pada setiap “pattern fragment”, hanya dataset yang berasosiasi saja yang di perlu di tambang. Adapun Langkah – Langkah dari Algoritma FP-Growth adalah sebagai berikut :

1. Menyiapkan Dataset.
2. Mencari Frequent Itemset (Item yang sering muncul).

3. Kumpulan dataset diurutkan berdasarkan priority.
4. Membuat FP-Tree berdasarkan item yang sudah diurutkan.
5. Pembangkitan Conditional Pattern Base.
6. Pembangkitan Conditional FP-Tree.
7. Pembangkitan Frequently Pattern.
8. Mencari Support.
9. Mencari Confidence.

Langkah pertama adalah menyiapkan dataset

Tabel 2.1 Data Transaksi

Id	Items
1	{ M, O, N, K, E, Y }
2	{ D, O, N, K, E, Y }
3	{ M, A, K, E }
4	{ M, U, C, K, Y }
5	{ C, O, K, I, E }

Mencari Item yang sering muncul dalam transaksi yang kemudian data transaksi tersebut diurutkan dari nilai frekuensi yang sering muncul.

Tabel 2.2 Frekuensi yang sering muncul

Item	Supcount	Frekuensi Item Set
K	5	100%

E	4	80%
M	3	60%
O	3	60%
Y	3	60%
N	2	40%
C	2	40%
D	1	20%
A	1	20%
U	1	20%
I	1	20%

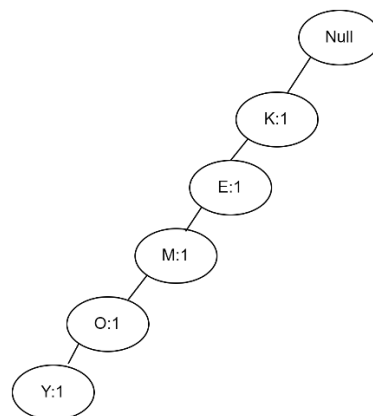
Setelah didapatkan item sesuai data yang dibutuhkan, lalu peneliti menentukan nilai support sebesar 60% { K, E, M, O, Y } yang hanya digunakan selanjutnya. Setelah menentukan batas nilai support selanjutnya membuat FP-Tree berdasarkan item set yang sudah diurutkan.

Tabel 2.3 Membuat Fp-Tree berdasarkan Item set

TID	Item
1	{ K, E, M, O, Y }
2	{ K, E, O, Y }
3	{ K, E, M }

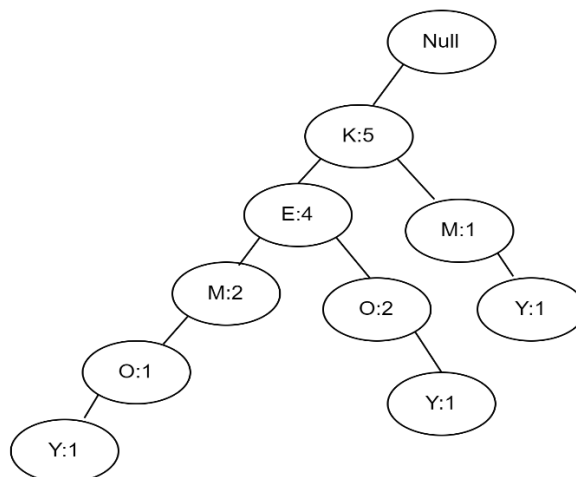
4	{ K, M, Y }
5	{ K, E, O }

langkah selanjutnya adalah membentuk pohon tree dalam proses algoritma fp-growth. Pembuatan Fp-Tree diawali dengan TID 1 yaitu { K, E, M, O, Y }.



Gambar 2.2 TID 1 FP-Tree

Setelah pembuatan TID 1, maka membuat TID 2,3,4,5 seperti gambar dibawah ini.



Gambar 2.3 TID 2,3,4,5 dst fp-tree

Setelah itu membangkitkan Conditional Pattern Base, seperti gambar dibawah ini;

Tabel 2.4 Pembangkitan Conditional Pattern Base dan Cond Fp-Tree

Item	Conditional Pattern Base	Cond FP-Tree
Y	{ K,E,M,O:1 } { K,E,O:1 } { K,M:1 }	{ K:3 }
O	{ K,E,M:1 } { K,E:2 }	{ KE:3 }
M	{ K,E:2 } { K:1 }	{ K:3 }
E	{ K:4 }	{ K:4 }
K	-	-

Perhitungan Frequent Pattern Generated

Tabel 2.5 Pembangkitan Frequent Pattern

Item	Supcount
Y	{ K, Y : 3 }
O	{ K, O : 3 } { E, O:3 }
M	{ K, M : 3 }
E	{ K, E: 4 }
K	-

Tabel 2.6 Rules yang terbentuk Support dan Confidance

Kombinasi 2 item	Frekuensi item set	Min. Support	Min. Confidance
K,Y	3	60%	60%

K, O	3	60%	60%
E, O	3	60%	60%
K, M	3	60%	60%
K, E	4	80%	80%

Hasil Asosiasi

Dari pengerjaan diatas didapatkan hasil asosiasi fp-growth dari min support dan min confidence sebagai berikut;

1. Jika konsumen membeli barang K juga membeli Y dengan nilai support 60% dan nilai confidence 60%.
2. Jika membeli barang K juga membeli O dengan nilai support 60% dan nilai confidence 60% dan seterusnya.

2.4 RapidMiner

Rapid Miner adalah perangkat lunak yang dibuat oleh Dr. Markus Hofmann dari Institut Teknologi Blanchardstown dan Ralf Klinkenberg dari rapidi.com yang berbasis GUI (Graphical User Interface) untuk memudahkan pengguna menggunakannya. RapidMiner adalah program gratis dan bersifat terbuka (open source). RapidMiner adalah solusi untuk analisis data mining, text mining dan analisis prediksi. RapidMiner menggunakan sejumlah algoritma deskriptif dan prediktif untuk menawarkan wawasan kepada pengguna yang memungkinkan mereka membuat keputusan terbaik(Wardani and Kristiana 2020)(Aprilla Dennis

2013). Rapid Miner mencakup berbagai pendekatan seperti klasifikasi, pengelompokan, asosiasi, dan lain-lain.

Menggunakan RapidMiner, tidak dibutuhkan keahlian koding khusus, karena semua fitur telah disediakan. RapidMiner khusus untuk penggunaan data mining. Model yang disediakan juga cukup banyak dan lengkap, seperti Model Bayesian, Modelling, Tree Induction, Neural Network dan lain-lain. Banyak metode yang disediakan oleh RapidMiner mulai dari kllasifikasi, klustering, asosiassi dan lain-lain.(Sudarsono et al. 2021).

Beberapa fitur dari Rapid Miner meliputi :

1. Berbagai algoritma data mining yang tersedia, seperti Decision Tree dan Self-Organization set.
2. Bentuk grafis yang lebih kompleks, seperti Diagram Histogram yang tumpang tindih, bagan pohon, dan plot sebar 3D.
3. Banyak variasi pluggin, seperti plugin teks untuk analisis teks.
4. Menerapkan teknik penambangan data dan pembelajaran mesin seperti ETL (ekstraksi, transformasi, dan pemuatan), prapemrosesan data, visualisasi, pemodelan, dan penilaian.
5. Proses data mining yang terdiri dari operator-operator yang nestable, dideskripsikan dengan XML, dan dibuat dengan GUI. Mengintegrasikan proyek data mining WEKA dan Statistika R(Ardiansyah and Walim 2018).