

BAB II

LANDASAN TEORI

2.1 Data Mining

Data mining sudah dikenal sejak tahun 1990-an, ketika adanya suatu pekerjaan yang memanfaatkan data menjadi suatu hal yang lebih penting dalam berbagai bidang, seperti *marketing* dan bisnis, sains dan teknologi, serta seni dan hiburan [14]. *Data mining* adalah metode dengan mengekstraksi informasi dalam jumlah besar, dengan cara ini membantu perusahaan fokus pada informasi penting di *data warehouse* [15]. *Data mining* menggunakan analisis matematika dalam menemukan pola kecenderungan dari suatu data, yang pada umumnya pola ini sukar ditemukan dengan eksplorasi data secara biasa. Hal ini disebabkan karena terlalu rumitnya hubungan antardata atau disebabkan besarnya data [16].

2.1.1 Metode Data Mining

Dalam [17], *Data Mining* dibagi menjadi beberapa kelompok, yaitu:

1. Deskripsi

Deskripsi adalah suatu cara untuk menggambarkan pola serta kecenderungan yang ada di dalam data. Deskripsi dari pola kecenderungan ini sering menghasilkan kemungkinan penjelasan terkait suatu pola atau kecenderungan.

2. Estimasi

Estimasi adalah suatu metode berupa memperkirakan nilai dari suatu populasi dengan menggunakan nilai dari sampel. Hampir sama dengan klasifikasi, namun estimasi lebih kearah numerik daripada kearah kategori.

3. Prediksi

Prediksi merupakan hasil dari kegiatan dalam memperkirakan apa yang akan terjadi di masa depan dari data yang sudah ada. Beberapa metode, teknik, serta algoritma yang digunakan dalam klasifikasi dan estimasi dapat

pula digunakan pada prediksi, bedanya adalah data yang digunakan merupakan data time series.

4. Klasifikasi

Klasifikasi merupakan metode untuk menyusun data secara sistematis. Dengan kata lain yaitu pengelompokan data berdasarkan suatu acuan atau ketentuan tertentu. Di dalam kelompok klasifikasi terdapat target berupa variabel kategorikal.

5. Pengklasteran

Pengklasteran adalah metode dalam menganalisa data yang berupa pengelompokan record, pengamatan atau membentuk kelas objek-objek yang memiliki suatu kemiripan. Klustering tidak memiliki target seperti klasifikasi, namun algoritma pengklasteran dapat melakukan suatu pembagian pada keseluruhan data sehingga menjadi kelompok-kelompok yang memiliki kemiripan, dimana kemiripan *record* dalam satu kelompok akan bernilai maksimal, namun *record* akan bernilai minimal untuk kemiripan dengan kelompok lainnya.

6. Asosiasi

Data mining bertugas untuk menemukan atribut-atribut yang muncul bersamaan dalam suatu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja atau market basket analysis. Metode asosiasi menggunakan pola “*If antecedent, then consequent,*” yang menggunakan pengukuran *support* dan nilai *confidence* untuk membentuk suatu aturan.

2.2 Association Rule Mining

Association rule mining merupakan suatu teknik *data mining* yang digunakan untuk menemukan aturan-aturan asosiasi dengan mengungkapkan *rules* untuk mengukur hubungan antara dua atau lebih atribut yang sering muncul secara bersamaan dalam sebuah himpunan data [18]. Banyak dimanfaatkan dalam dunia bisnis, teknik ini biasa disebut dengan analisis keranjang belanja, sebagai contohnya yaitu digunakan untuk menentukan barang-barang yang banyak dibeli

secara bersamaan ataupun tidak pernah dibeli secara bersamaan dalam *supermarket* [19].

Dalam [20] pada penelitiannya menyatakan bahwa analisis pola frekuensi tinggi (*frequent pattern mining*) menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien yang memiliki parameter penting, yaitu *support* (nilai penunjang) yang merupakan presentasi kombinasi *item* dalam *database*, dan *confidence* (nilai kepastian) yang merupakan kuatnya hubungan antar *item* dalam aturan asosiasi. Pemilihan aturan asosiasi tidak hanya melihat *rule* yang memiliki nilai *confidence* paling tinggi, namun menghitung *lift ratio* merupakan langkah untuk mengetahui kekuatan aturan asosiasi yang telah terbentuk [21].

2.2.1 Metodologi Dasar Analisis Asosiasi

Dalam [22], metodologi untuk melakukan analisis asosiasi terbagi kedalam dua tahapan, yaitu:

a. Analisa Pola Frekuensi Tinggi

Tahap ini berfungsi untuk mencari pola kombinasi *item* yang memenuhi syarat *minimum* dari nilai *support*. Misalnya, dari keseluruhan *dataset* yang ada, berapa besar tingkat dominasi yang menunjukkan kemunculan bersama dalam satu waktu antara *item* A dan *item* B. Nilai *support* sebuah *item* diperoleh dengan menggunakan rumus persamaan 2.1.

$$Support(A) = \frac{\text{Jumlah Transaksi Mengandung A}}{\text{Total Transaksi}} \times 100\% \quad (2.1)$$

Sedangkan untuk nilai *support* dari 2-*item* diperoleh dari rumus persamaan 2.2.

$$Support(A,B) = \frac{\text{Jumlah Transaksi Mengandung A dan B}}{\text{Total Transaksi}} \times 100\% \quad (2.2)$$

b. Pembentukan Pola Aturan Asosiasi

Setelah semua pola *frequent itemset* ditemukan, selanjutnya dicari aturan asosiasi yang memenuhi syarat *minimum confidence* dengan menghitung nilai *confidence* aturan asosiasi $A \rightarrow B$ yang diperoleh dari rumus 2.3.

$$Confidence = P(B|A) = \frac{\text{Total Transaksi } A \rightarrow B}{\text{Total Transaksi } A} \times 100\% \quad (2.3)$$

Nilai Kepastian atau *confidence* $A \rightarrow B$ diperoleh dengan membagi jumlah Transaksi yang mengandung A dan B dengan Total Transaksi yang di kalikan dengan 100%.

Lift ratio adalah suatu ukuran untuk mengetahui kekuatan aturan asosiasi yang telah terbentuk. Nilai *lift ratio* biasanya digunakan sebagai penentu valid tidaknya suatu aturan asosiasi yang telah didapatkan. Berikut adalah tiga kemungkinan yang akan dihasilkan saat menghitung *lift ratio*, yaitu [23]:

1. Jika $lift < 1$, maka kemunculan *item* A berkorelasi negatif terhadap kemunculan *item* B . Dengan artian jika membeli barang A maka tidak membeli barang B .
2. Jika $lift > 1$, maka *item* A dan *item* B berkorelasi positif terhadap kemunculan *item* B . Artinya jika membeli *item* A maka membeli *item* B .
3. Jika $lift = 1$, maka *item* A dan B adalah independen dan memiliki korelasi diantara kedua *item* atau tidak dapat ditarik kesimpulan. Artinya jika membeli *item* A maka belum tentu membeli *item* B .

Semakin tinggi nilai *lift* maka semakin kuat asosiasi, atau dapat dikatakan asosiasi yang menarik adalah asosiasi dengan aturan yang kuat serta memiliki nilai $lift > 1$ [24].

$$Lift (A \Rightarrow B) = \frac{Confidence A \Rightarrow B}{P(B)} = \frac{P(A \cup B)}{P(A) P(B)} \quad (2.4)$$

2.2.2 Aturan Asosiasi

Proses aturan asosiasi terdiri dari beberapa tahap. Berikut adalah tahapan pada aturan asosiasi [7]:

1. Pertama yaitu mencari kandidat *1-itemset* dengan cara sistem melakukan *scan* pada *database* dan kemudian menghitung nilai

support. Nilai *support* yang telah didapatkan kemudian dibandingkan dengan *minimum support* yang telah ditentukan. *Itemset* yang termasuk dalam *large itemset* yaitu jika nilai yang dihasilkan lebih besar atau sama dengan *minimum support*.

2. *Itemset* yang tidak termasuk dalam *large itemset* tidak akan diikuti pada iterasi selanjutnya atau dapat dihilangkan saja.
3. Hasil *large itemset* pada iterasi pertama (L1) kemudian diteruskan untuk membentuk kandidat *itemset* kedua (L2). Pada iterasi selanjutnya sistem akan menggunakan hasil *large itemset* pada iterasi sebelumnya (Lk-1) untuk membentuk kandidat *itemset* berikutnya (Lk). Sistem akan menggabungkan (*join*) hasil Lk-1 dengan Lk-1 untuk mendapatkan Lk. Sama seperti sebelumnya, iterasi yang tidak termasuk dalam *large itemset* tidak akan diikuti pada iterasi selanjutnya atau dapat dihilangkan saja.
4. Selanjutnya operasi *join* dilakukan, pasangan *itemset* baru dari hasil proses *join* tersebut dihitunglah nilai *support*.
5. Proses *join* dan *prune* untuk pembentukan kandidat akan terus dilakukan hingga himpunan kandidat *itemset* sudah tidak ada lagi kandidat yang akan dibentuk.
6. Hasil *frequent itemset* yang telah didapatkan kemudian dibentuk dalam aturan asosiasi yang memenuhi nilai *support* dan *confidence* yang telah ditetapkan.
7. Pada pembentukan aturan asosiasi ini, nilai yang sama dianggap sebagai satu nilai.
8. Hasil *association rule* yang terbentuk harus memenuhi nilai *minimum* yang telah ditentukan.

2.3 Algoritma APRIORI

Algoritma apriori adalah algoritma yang terdapat dalam *data mining* dengan menemukan pola data atau pola kemunculan (frekuensi) data dengan teknik aturan asosiasi [25]. Untuk menentukan kandidat-kandidat yang mungkin muncul pada

algoritma Apriori yaitu dengan cara memperhatikan *minimum support* pada kombinasi *itemset* [25]. Berikut adalah langkah-langkah pada proses algoritma apriori [6] :

1. Tahap pertama yaitu pembentukan kandidat *itemset*, dimana kandidat *k-itemset* terbentuk dari kombinasi $(k-1)$ -*itemset* yang didapatkan dari *iterasi* yang sebelumnya telah dicari. Untuk kandidat *k-itemset* yang subsetnya berisi $k-1$ *item* akan dipangkas karena tidak termasuk dalam pola frekuensi tinggi dengan panjang $k-1$.
2. Tahap kedua yaitu menggunakan perhitungan *support* dari setiap kandidat *k-itemset* yang sebelumnya sudah didapatkan. Cara menentukan nilai *support* tiap kandidat *k-itemset* yaitu dengan melakukan *scan database* untuk menghitung jumlah transaksi yang memuat semua *item*.
3. Setelah itu bandingkan antara nilai *support* dengan *minimum support* yang sebelumnya telah ditentukan, apabila nilai *support* lebih besar atau nilainya sama dengan *minimum support*, *itemset* terhitung dalam *large-itemset* 1 (L_1). *Itemset* yang tidak terhitung dalam *large-itemset* tidak dipakai untuk melakukan iterasi berikutnya. Untuk mencari kombinasi selanjutnya dilakukan hal yang sama. Apabila tidak terdapat lagi pola frekuensi tinggi yang baru maka proses berlanjut ke pembentukan pola aturan asosiasi.
4. Langkah selanjutnya yaitu untuk semua *large-itemset* yang terbentuk atau memenuhi nilai *minimum support* akan dibentuk aturan asosiasi setelah itu dicari juga nilai *confidence*. Nantinya seluruh aturan yang terbentuk jika nilai *confidence* kurang dari nilai *minimum confidence* yang ditetapkan, maka aturan tersebut tidak akan dipakai atau tidak termasuk dalam aturan asosiasi yang dipakai.

Berikut contoh penggunaan algoritma apriori beserta urutan proses pencarian *k-itemset*:

Tabel 2.1 Data Transaksi

No_ID	Item
1	HVS, Pena, Buku tulis
2	Pensil, HVS, Penghapus
3	HVS, Pena, HVS
4	Penghapus, Pensil, Buku tulis
5	Pena, Buku tulis, HVS

Data yang terdapat pada Tabel 2.1 merupakan data transaksi pembelian. Terdapat 5 transaksi pembelian dengan *item* yang berbeda-beda. Kemudian ditentukan *minimum support* = 30%.

Tabel 2. 2 Proses Mencari *Frequent 1-Itemset* Apriori

<i>Items</i>	<i>Count</i>	Jumlah Transaksi	<i>Support</i>
HVS	4	5	$4/5 = 0,8$
Pena	3	5	$3/5 = 0,6$
Buku tulis	3	5	$3/5 = 0,6$
Pensil	2	5	$2/5 = 0,4$
Penghapus	2	5	$2/5 = 0,4$

Setelah dihitung *minimum support* tiap *item*, semua *itemset* memiliki *support* melebihi syarat *minimum support* sehingga dapat digunakan untuk pencarian selanjutnya yaitu penyilangan pada 2-*itemset*

Tabel 2. 3 Proses Mencari *Frequent 2-Itemset* Apriori

<i>Pattern</i>	<i>Count</i>	Jumlah Transaksi	<i>Support</i>
HVS, Pena	3	5	$3/5 = 0,6$
HVS, Buku tulis	2	5	$2/5 = 0,4$
HVS, Pensil	1	5	$1/5 = 0,2$
HVS, Penghapus	1	5	$1/5 = 0,2$
Pena, Buku tulis	2	5	$2/5 = 0,4$
Pena, Pensil	0	5	$0/5 = 0$
Pena, Penghapus	0	5	$0/5 = 0$
Buku tulis, Pensil	1	5	$1/5 = 0,2$
Buku tulis, Penghapus	1	5	$1/5 = 0,2$
Pensil, Penghapus	1	5	$1/5 = 0,2$

Selanjutnya untuk mencari kombinasi 2-*itemset*, semua *itemset* memiliki *support* melebihi syarat *minimum support* sehingga dapat digunakan untuk pencarian selanjutnya yaitu penyilangan pada 2-*itemset*.

Tabel 2. 4 Hasil *Frequent 3-Itemset* Apriori

<i>Pattern</i>	<i>Count</i>	Jumlah Transaksi	<i>Support</i>
HVS, Pena, Buku tulis	2	5	$2/5 = 0,4$

Dari hasil pencarian kombinasi 3-*itemset* terdapat item { HVS, Pena, Buku tulis } dengan *support* 0,4 atau 40%. Selanjutnya yaitu mencari aturan asosiasi dengan *minimum confidence* = 70%.

Tabel 2. 5 Aturan Asosiasi Apriori

Rules	Rumus	Confidence
HVS => Pena	$\frac{\text{Supp (HVS, Pena)}}{\text{Supp (HVS)}}$	0,75
Pena => HVS	$\frac{\text{Supp (Pena, HVS)}}{\text{Supp (Pena)}}$	1
HVS => Buku tulis	$\frac{\text{Supp (HVS, Buku tulis)}}{\text{Supp (HVS)}}$	0,5
Buku tulis => HVS	$\frac{\text{Supp (Buku tulis, HVS)}}{\text{Supp (Buku tulis)}}$	0,67
Pena => Buku tulis	$\frac{\text{Supp (Pena, Buku tulis)}}{\text{Supp (Pena)}}$	0,67
Buku tulis => Pena	$\frac{\text{Supp (Buku tulis, Pena)}}{\text{Supp (Buku tulis)}}$	0,67
HVS, Pena => Buku tulis	$\frac{\text{Supp (HVS, Pena, Buku tulis)}}{\text{Supp (HVS, Pena)}}$	0,67
HVS, Buku tulis => Pena	$\frac{\text{Supp (HVS, Buku tulis, Pena)}}{\text{Supp (HVS, Buku tulis)}}$	1
Pena, Buku tulis => HVS	$\frac{\text{Supp (Pena, Buku tulis, HVS)}}{\text{Supp (Pena, Buku tulis)}}$	1

Setelah pencarian aturan asosiasi, didapatkan empat aturan yang memenuhi syarat *minimum confidence*. Untuk aturan yang dapat memenuhi syarat *minimum confidence*, dilanjutkan dengan mencari *lift ratio* sebagai berikut:

Tabel 2. 6 Hasil Proses Mencari *Lift Ratio* Apriori

Rules	Rumus	Lift
HVS => Pena	$\frac{\text{Supp (HVS, Pena)}}{\text{Supp (HVS) x Supp (Pena)}}$	1,25
Pena => HVS	$\frac{\text{Supp (Pena, HVS)}}{\text{Supp (Pena) x Supp (HVS)}}$	1,25
HVS, Buku tulis => Pena	$\frac{\text{Supp (HVS, Buku tulis, Pena)}}{\text{Supp (HVS, Buku tulis) x Supp (Pena)}}$	1,67
Pena, Buku tulis => HVS	$\frac{\text{Supp (Pena, Buku tulis, HVS)}}{\text{Supp (Pena, Buku tulis) x Supp (HVS)}}$	1,25

Setelah dicari perhitungan untuk *lift ratio* tiap aturan yang didapat, dihasilkan semua aturan mempunyai nilai *lift ratio* lebih dari 1 yang berarti memiliki korelasi positif, dan aturan yang memiliki nilai *lift ratio* tertinggi didapatkan sebesar 1,67 yaitu “Jika semakin banyak HVS dan Buku tulis dibeli, maka jumlah Pena yang dibeli semakin banyak”.

2.4 Algoritma *Equivalence Class Transformation* (ECLAT)

Algoritma *Equivalence Class Transformation* atau yang disingkat ECLAT merupakan algoritma yang proses pencariannya secara *depth-first search* pada *database*. Proses pembentukan *itemset* dimulai dengan mengubah bentuk dari transaksi yang berbentuk horizontal menjadi transaksi berbentuk vertikal dengan penggabungan *TID List* yang memiliki *item* yang sama pada transaksi. [26].

Berikut adalah proses pembentukan *itemset* pada algoritma ECLAT [27]:

1. Pertama yaitu mempresentasikan data dimulai dengan mengubah bentuk dari data transaksi berbentuk horizontal menjadi transaksi berbentuk vertikal dengan penggabungan *TID List* yang memiliki *item* yang sama pada transaksi.
2. Kemudian proses penyilangan pada *TID List*, yaitu dilakukan penyilangan *subset item X* dan *subset item Y* untuk mencari penyilangan pada 2 *itemset*. Proses penyilangan dilakukan terus menerus hingga tidak ada lagi *itemset* yang tersisa.
3. Proses selanjutnya yaitu menentukan nilai *minimum support* dari setiap *k-itemset*. *Itemset* yang memiliki nilai kurang dari syarat *minimum support* tersebut kemudian dilakukan pemangkasan.
4. Selanjutnya yaitu mencari aturan asosiasi yang memenuhi syarat *minimum* dengan menghitung *confidence* aturan asosiasi pada semua pola *frequent itemset*. Kemudian *rules* yang telah didapatkan lalu dilakukan penerjemahan ke dalam informasi yang dapat dimengerti oleh pengguna.

Berikut contoh penggunaan algoritma ECLAT beserta urutan proses pencarian *k-itemset*:

Tabel 2. 7 Data Transaksi Horizontal

No_ID	Item
1	HVS, Pena, Buku tulis
2	Pensil, HVS, Penghapus
3	HVS, Pena, HVS
4	Penghapus, Pensil, Buku tulis
5	Pena, Buku tulis, HVS

Data transaksi yang terdapat dalam transaksi pada Tabel 2.7 harus diubah menjadi bentuk vertikal terlebih dahulu karena data masih berbentuk horizontal. Tabel 2.8. Sebagai contoh Transaksi_ID 1 berisikan data HVS, Pena, dan Buku tulis. Kemudian dipecah membentuk vertikal, sehingga HVS dan Transaksi_ID yaitu 1,2,3,5. Berikut tampilan data yang sudah diubah menjadi bentuk vertikal, *minimum support* = 30% atau 0.3 digunakan untuk meneliti data tersebut.

Tabel 2. 8 Data Transaksi Vertikal

<i>Item</i>	<i>Transaksi_ID</i>	<i>Support</i>
HVS	1,2,3,5	4/5 = 0.8
Pena	1,3,5	3/5 = 0.6
Buku tulis	1,4,5	3/5 = 0.6
Pensil	2,4	2/5 = 0.4
Penghapus	2,4	2/5 = 0.4

Dilihat dari Tabel 2.8, semua *itemset* memiliki *support* melebihi syarat *minimum support* sehingga dapat digunakan untuk pencarian selanjutnya yaitu penyilangan pada *2-itemset*.

Tabel 2. 9 Proses Mencari *2-Itemset* ECLAT

<i>2-Itemset</i>	<i>Transaksi_ID</i>
HVS, Pena	1,3,5
HVS, Buku tulis	1,5
HVS, Pensil	2
HVS, Penghapus	2
Pena, Buku tulis	1,5
Pena, Pensil	0
Pena, Penghapus	0
Buku tulis, Pensil	4
Buku tulis, Penghapus	4
Pensil, Penghapus	2

Pada Tabel 2.9, terdapat beberapa *itemset* yang tidak memenuhi syarat *minimum support*, sehingga *itemset* tersebut tidak dapat lanjut ke tahap pencarian *k-itemset* berikutnya. Berikut adalah tabel *frequent 2-itemset*:

Tabel 2. 10 *Frequent 2-Itemset* ECLAT

<i>2-Itemset</i>	<i>Transaksi_ID</i>	<i>Support</i>
HVS, Pena	1,3,5	3/5 = 0.6
HVS, Buku tulis	1,5	2/5 = 0.4
Pena, Buku tulis	1,5	2/5 = 0.4

Selanjutnya dilakukan tahap penyilangan *3-itemset*. Tabel 2.11 adalah hasil penyilangan yang didapatkan dengan memenuhi syarat *minimum support*.

Tabel 2. 11 Proses mencari 3-Itemset ECLAT

3-Itemset	Transaksi_ID	Support
HVS, Pena, Buku tulis	2	$2/5 = 0.4$

Dari hasil penyilangan pada tabel 2.11, *itemset* yang didapatkan memenuhi syarat memiliki nilai *support* 0.4, maka *frequent 3-itemset* tersebut dapat memenuhi syarat *minimum support*. Selanjutnya yaitu mencari aturan asosiasi dengan *minimum confidence* = 70%.

Tabel 2. 12 Mencari Aturan Asosiasi ECLAT

Rules	Rumus	Confidence
HVS => Pena	$\frac{\text{Supp (HVS, Pena)}}{\text{Supp (HVS)}}$	0,75
Pena => HVS	$\frac{\text{Supp (Pena, HVS)}}{\text{Supp (Pena)}}$	1
HVS => Buku tulis	$\frac{\text{Supp (HVS, Buku tulis)}}{\text{Supp (HVS)}}$	0,5
Buku tulis => HVS	$\frac{\text{Supp (Buku tulis, HVS)}}{\text{Supp (Buku tulis)}}$	0,67
Pena => Buku tulis	$\frac{\text{Supp (Pena, Buku tulis)}}{\text{Supp (Pena)}}$	0,67
Buku tulis => Pena	$\frac{\text{Supp (Buku tulis, Pena)}}{\text{Supp (Buku tulis)}}$	0,67
HVS, Pena => Buku tulis	$\frac{\text{Supp (HVS, Pena, Buku tulis)}}{\text{Supp (HVS, Pena)}}$	0,67
HVS, Buku tulis => Pena	$\frac{\text{Supp (HVS, Buku tulis, Pena)}}{\text{Supp (HVS, Buku tulis)}}$	1
Pena, Buku tulis => HVS	$\frac{\text{Supp (Pena, Buku tulis, HVS)}}{\text{Supp (Pena, Buku tulis)}}$	1

Setelah pencarian aturan asosiasi, didapatkan empat aturan yang memenuhi syarat *minimum confidence*. Untuk aturan yang dapat memenuhi syarat *minimum support* maka dilanjutkan dengan mencari *lift ratio* sebagai berikut:

Tabel 2. 13 Mencari Lift Ratio Rules ECLAT

Rules	Rumus	Lift
HVS => Pena	$\frac{\text{Supp (HVS, Pena)}}{\text{Supp (HVS) x Supp (Pena)}}$	1,25
Pena => HVS	$\frac{\text{Supp (Pena, HVS)}}{\text{Supp (Pena) x Supp (HVS)}}$	1,25
HVS, Buku tulis => Pena	$\frac{\text{Supp (HVS, Buku tulis, Pena)}}{\text{Supp (HVS, Buku tulis) x Supp (Pena)}}$	1,67
Pena, Buku tulis => HVS	$\frac{\text{Supp (Pena, Buku tulis, HVS)}}{\text{Supp (Pena, Buku tulis) x Supp (HVS)}}$	1,25

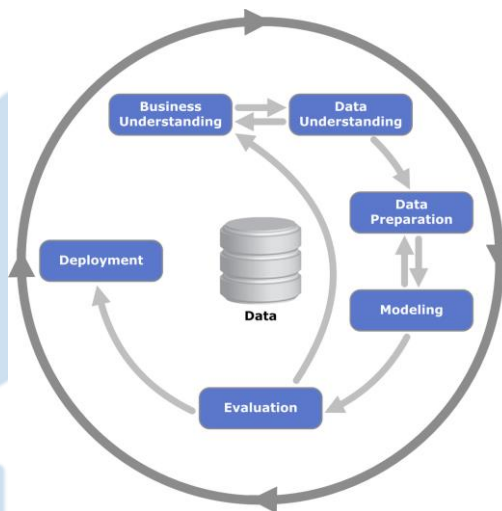
Setelah dicari perhitungan untuk *lift ratio* tiap aturan yang didapat, dihasilkan semua aturan mempunyai nilai *lift ratio* lebih dari 1 yang berarti memiliki korelasi positif, dan aturan yang memiliki nilai *lift ratio* tertinggi didapatkan sebesar 1,67 yaitu “Jika semakin banyak HVS dan Buku tulis dibeli, maka jumlah Pena yang dibeli semakin banyak”.

2.5 CRISP-DM

2.5.1 Pengertian CRISP-DM

Cross Industry Standard Process for Data Mining (CRISP-DM) merupakan standarisasi proses *data mining* yang dikembangkan pada tahun 1996 dengan tujuan untuk melakukan proses analisis strategi pemecahan masalah dari penelitian ataupun bisnis [28]. Dalam [29], dijelaskan bahwa CRISP-DM terdiri dari enam fase pengembangan yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation* dan *deployment*.

2.5.2 Fase CRISP-DM



Gambar 2. 1 Alur CRISP-DM

Berikut merupakan *life cycle* dari CRISP-DM yang biasa digunakan dalam implementasi *data mining* [29]:

1. *Business Understanding Phase*

Fase ini mendefinisikan tujuan organisasi yang akan dicapai kemudian mengkonversi tujuan menjadi formula *data mining* dan menerapkan strategi untuk mencapai tujuan tersebut.

2. *Data Understanding Phase*

Aktivitas fase ini yaitu mengumpulkan data-data dari aktivitas organisasi dalam hal ini adalah data transaksi penjualan buku. Kumpulkan data yang akan dibutuhkan dan memastikan datanya dapat dianalisa.

3. *Data Preparation Phase*

Menyiapkan data, menganalisa hingga membersihkan data seperti redundansi, data kosong, menghapus kolom dan baris yang tidak diperlukan agar data yang di masukkan ke tahap *modeling* sesuai yang dibutuhkan.

4. *Modeling Phase*

Memilih algoritma yang akan diterapkan terhadap data dan mengimplementasikan algoritmanya. Jika dibutuhkan, dapat melakukan eksperimen terhadap beberapa model algoritma sampai hasilnya sesuai yang diharapkan. Pilih salah satu metode yang akan digunakan dan sesuai dengan *business understanding phase*.

5. *Evaluation Phase*

Dari *modeling phase*, ada hasil yang didapat. Hasil dari model algoritma yang telah dipilih dievaluasi apakah sudah sesuai dengan tujuan bisnis pada fase 1 atau belum mencapai tujuan. Tahap ini bertujuan untuk menentukan apakah ada masalah bisnis yang masih belum dipertimbangkan. Penggunaan hasil proses *data mining* harus ditentukan di akhir dari tahap ini.

6. *Deployment Phase*

Dari hasil yang sudah diperoleh pada *evaluation phase*, algoritma yang sudah diterapkan dapat diimplementasikan pada sebuah laporan atau aplikasi pendukung keputusan sumber terbuka. Selain analisis data, tahap *deployment* melibatkan konsumen juga dilakukan. Hal ini dikarenakan sangat penting bagi konsumen untuk memahami dari model yang telah dibuat, tindakan apa yang harus dilakukan.

2.6 RStudio



Gambar 2. 2 Logo RStudio

Aplikasi R pada dasarnya berbasis teks atau *command line* sehingga *user* harus mengetikkan perintah-perintah tertentu dan juga harus hapal dengan penggunaan perintah-perintahnya. R dirancang untuk mengerjakan berbagai hal yang berhubungan dengan penggunaan teknik *data mining* baik statistika maupun *machine learning* [30]. R menyediakan berbagai macam teknik statistika antara lain yaitu pemodelan linier, pemodelan nonlinier, analisis deret waktu, dan lainnya.

RStudio merupakan *integrated development environment* (IDE) dari R, yaitu mengintegrasikan fasilitas-fasilitas yang diperlukan pada R [30]. RStudio mempunyai tampilan yang menarik dan dirancang untuk mempermudah proses *coding* dan *project management*. Secara umum tampilan Rstudio dibagi menjadi 4 bagian, yaitu *Script editor*, *Console editor*, *Workspace* dan *history*, dan *File*, *plot*, *packages*, serta *help*.

2.7 Penelitian Terdahulu

Tabel 2. 14 Penelitian Terdahulu

No	Judul / Jurnal	Penulis	Algoritma	Hasil
1.	Penerapan Data Mining Penjualan Sepatu Menggunakan Metode Algoritma Apriori / Jurnal Teknik Komputer Vol 4, No. 1, Februari 2018 [31]	Erma Delima Sikumbang	Apriori	Akumulasi dari transaksi diambil dari 3 <i>item</i> yang paling laris terjual disetiap bulannya sehingga didapatkan hasil 12 transaksi. Dengan menggunakan minimum <i>support</i> = 30% dan minimum <i>confidence</i> = 60% didapatkan 6 aturan.
2.	Analisis Pola Pembelian Konsumen pada Transaksi Penjualan Menggunakan Algoritma Apriori /	Irsyad Djamiludin, Agus Nursikuwagus	Apriori	Dengan menggunakan data sebanyak 282 data transaksi yang kemudian dilakukan pembersihan dengan menyisihkan 245 data untuk diuji. Didapatkan hasil dengan

No	Judul / Jurnal	Penulis	Algoritma	Hasil
	Jurnal SIMETRIS, Vol 8 No 2 November 2017 [32]			minimum <i>support</i> = 3% dan minimum <i>confidence</i> = 70% sebanyak 4 aturan pada <i>frequent 2 itemset</i> , dan sebanyak 3 aturan pada <i>frequent 3 itemset</i> .
3.	Implementasi Algoritma Apriori untuk Rekomendasi Produk pada Toko Online / Citec Journal, Vol. 7, No. 1, Januari 2020 [33]	Esha Alma'arif, Ema Utami, Ferry Wahyu Wibowo	Apriori	Dengan menggunakan data sebanyak 20 data transaksi, didapatkan hasil sebanyak 2 aturan yang menggunakan nilai minimum <i>support</i> = 40% dan minimum <i>confidence</i> = 70%.
4.	Analisis Faktor Risiko Kematian dengan Penyakit Komorbid COVID-19 menggunakan Algoritma ECLAT / Jurnal Informasi dan Teknologi, Vol. 4, No.1, Hal: 52-57, 2022 [34]	Sukma Evadini	ECLAT	Penelitian dilakukan dengan menggunakan nilai minimum <i>support</i> = 20% didapatkan hasil 8 aturan dengan <i>frequent 2 itemset</i> , 4 aturan dengan <i>frequent 3 itemset</i> , dan dengan nilai minimum <i>confidence</i> = 80%, terdapat 2 aturan dengan <i>frequent 4 itemset</i> .
5.	Analisa Keterkaitan <i>Risk Factor Stroke</i> dengan Jenis Stroke yang Diderita Menggunakan Algoritma ECLAT / Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI) 9, Fakultas Sains dan Teknologi, UIN Sultan Syarif Kasim Riau, Pekanbaru, 18-19 Mei 2017 [35]	Rio Fernando, Lia Anggraini, Alwis Nazir	ECLAT	Data <i>sampling</i> yang digunakan yaitu sebesar 630 data dengan maximum <i>length</i> = 4 dan minimum <i>support</i> = 10% didapatkan hasil sebanyak 109 aturan.
6	Implementasi Algoritma Apriori untuk Rekomendasi Kombinasi Produk Penjualan/ ULTIMATICS, Vol. XII, No. 1, 17 Juni 2020 [36]	Andre Setiawan, Farica Perdana Putri	Apriori	Data yang digunakan yaitu dari PT Grup Gaga Indonesia selama 2 tahun dengan total data sebanyak 529 data transaksi. Penelitian menggunakan algoritma Apriori dengan minimum <i>support</i> 4% dan minimum <i>confidence</i> 3% dihasilkan sebanyak 4 aturan asosiasi, 2 diantaranya memiliki nilai <i>lift ratio</i> > 1 dan 2 aturan memiliki nilai <i>lift ratio</i> < 1.

Dari lima penelitian terdahulu pada Tabel 2.14, dapat disimpulkan bahwa model dengan algoritma Apriori dan algoritma ECLAT terbukti dapat menghasilkan aturan asosiasi berdasarkan syarat minimum *support* dan minimum *confidence*. Penelitian ini akan berfokus dengan penggunaan algoritma Apriori dan algoritma ECLAT dalam proses menghasilkan aturan asosiasi.

Berikut adalah beberapa perbandingan penelitian terdahulu dengan penelitian ini:

1. Pada tabel penelitian terdahulu, dari penelitian 1 sampai dengan 4 hanya menggunakan *confidence* sebagai kualitas mencari aturan asosiasi, penelitian ke-5 hanya menggunakan *support* saja, dan pada penelitian ke-6 sudah menggunakan *support*, *confidence*, dan *lift ratio*. Pada penelitian skripsi ini kualitas aturan asosiasi yang didapatkan yaitu berdasarkan *support*, *confidence*, dan *lift ratio*.
2. Penggunaan data pada kelima penelitian tersebut dibawah 1000 transaksi. Sedangkan pada penelitian skripsi ini, data yang digunakan yaitu sebanyak 35.565 data transaksi penjualan.
3. Pada penelitian terdahulu, tiap penelitian hanya menggunakan satu algoritma saja, sedangkan pada penelitian skripsi ini dilakukan dengan menggunakan algoritma Apriori dan algoritma ECLAT untuk didapatkan hasil perbandingan dari kinerja tiap algoritma tersebut. Perbandingan diukur berdasarkan waktu eksekusi tiap algoritma.
4. Pada penelitian terdahulu, tidak ada satupun penelitian yang melakukan evaluasi dengan meminta pendapat pemilik data terkait hasil penelitian yang didapatkan. Sedangkan pada penelitian skripsi ini, dilakukan evaluasi dengan mengajukan pendapat pemilik data terhadap hasil penelitian aturan asosiasi yang didapatkan.