

Desain Dashboard Deskriptif dan Prediktif Analitik untuk Frequent Itemset Mining serta Cross-selling pada Produk Bakery

3 April 2023

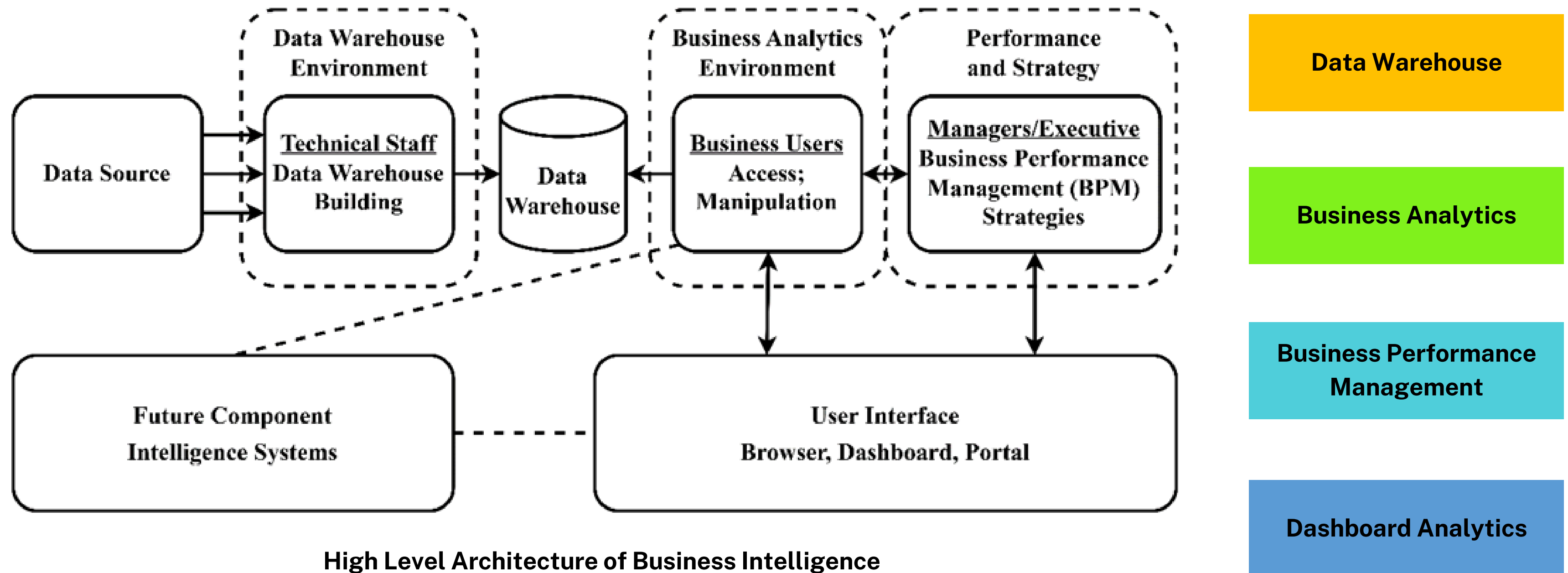
Presented by
Hary Yusuf 119190044

Content



- | | |
|-------------------------|-----------------------------|
| ➊ Latar Belakang | ➋ Informasi Dataset |
| ➌ Rumusan Masalah | ➍ Pemodelan SC 1 dan 2 |
| ➎ Tujuan Penelitian | ➏ Desain Dashboard |
| ➐ Batasan Penelitian | ➑ Hasil Ekstraksi Informasi |
| ➒ Penelitian Terdahulu | ➓ Dashboard Analitik |
| ➔ Metodologi Penelitian | ➑ Kesimpulan dan Saran |

Latar Belakang



[3] W. Eckerson, "Smart Companies in the 21st Century: The Secrets of Creating Successful Business Intelligent Solutions," Seattle, 2003.

Latar Belakang (2)

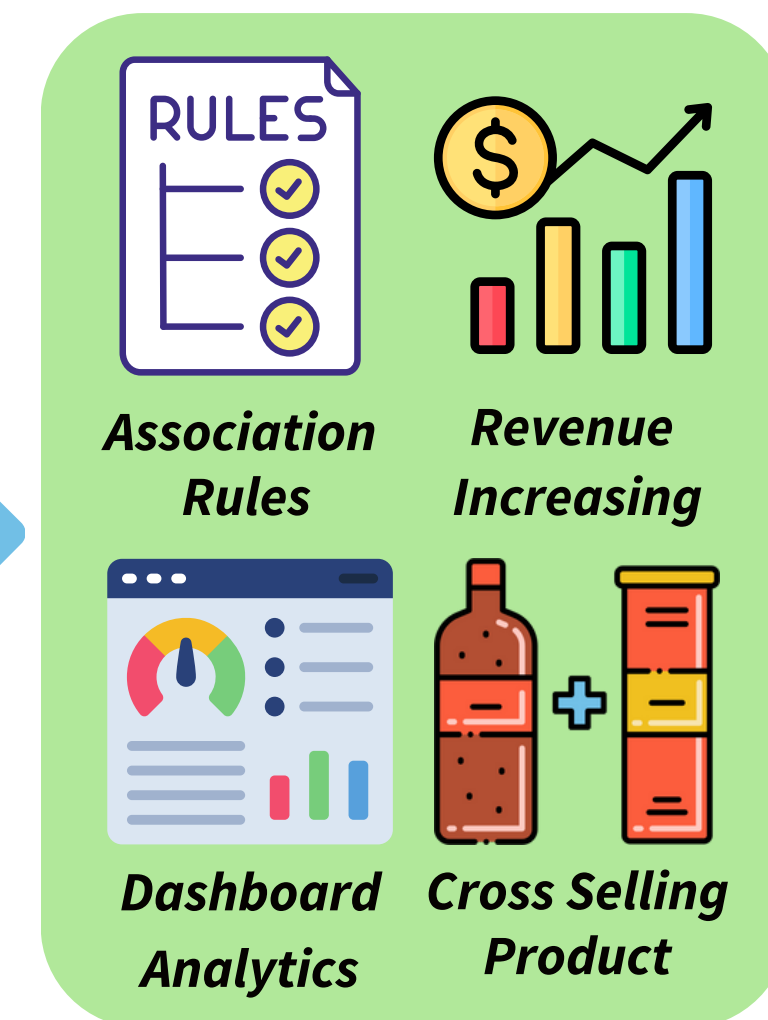
Input



Process



Output



Penelitian ini berfokus pada **pengembangan sistem pemasaran produk bakery**, dengan **memanfaatkan data transaksi** untuk **mendukung perumusan strategi pemasaran** yang disajikan dalam bentuk **dashboard analitik**

Rumusan Masalah

- 1 Perlunya **analisis entitas sistem** yang akan dikembangkan
- 2 Perlunya **ekstraksi frequent itemset** berdasarkan pola penjualan yang terjadi pada dataset produk bakery
- 3 Perlunya **penerapan aturan asosiasi** yang dihasilkan dari proses frequent itemset mining terhadap strategi pemasaran produk bakery
- 4 Perlunya **rancangan dashboard analitik** untuk mendukung kinerja pemasaran produk bakery

Tujuan Penelitian

- 1 **Menganalisis kebutuhan dan proses bisnis** dalam pengembangan sistem pemasaran produk bakery
- 2 Mengekstraksi pengetahuan (knowledge extraction) berupa **aturan asosiasi berbasis frequent itemset mining** dan **mengukur kinerja rule** tersebut
- 3 **Merancang paket cross-selling** dari hasil association rules sebagai pendukung pengambilan keputusan dalam penerapan strategi pemasaran
- 4 **Merancang dashboard deskriptif dan prediktif analitik** dari hasil knowledge extraction berbasis frequent itemset mining

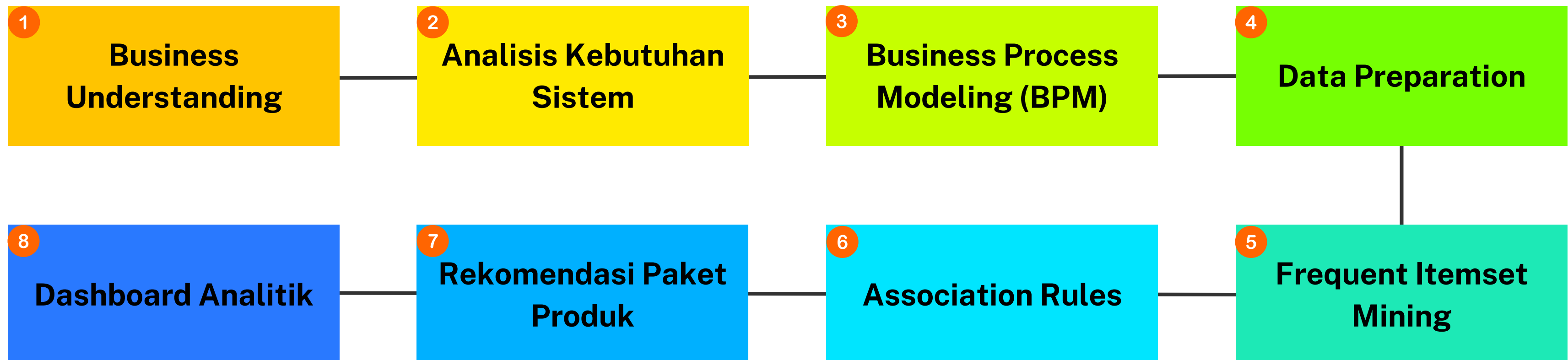
Batasan Penelitian

- 1 Penelitian menggunakan **dataset produk bakery** yang berasal dari **situs kaggle** (data source)
- 2 Penelitian yang dilakukan berfokus pada **analisis dan perhitungan data transaksi konsumen** untuk memperoleh **frequent itemset** dan **association rules**
- 3 Penelitian ini **tidak membutuhkan variabel kuantitas produk** yang dibeli per transaksi yang dilakukan konsumen **dalam memperoleh aturan asosiasi**
- 4 Hasil **visualisasi dashboard analitik** hanya menjabarkan **hasil analisis deskriptif dan prediktif** produk bakery

Penelitian Terdahulu

No	Judul	Tahun	Metode	Hasil
1	Penerapan <i>Market Basket Analysis</i> dengan Menggunakan Metode <i>Association Rule</i> Untuk Pengenalan Pola Perilaku Konsumen [6]	2022	<i>Association Rules, Algoritma Apriori</i>	Rule {Gula→Teh} dapat menjadi rekomendasi bagi Kasimura <i>Supermarket</i> karena memiliki nilai support ≥ 0.2 dan nilai confidence = 0.667
2	Analisis Pola Pembelian Konsumen pada Transaksi Penjualan Menggunakan Teknik <i>Association Rules</i> dengan Algoritma <i>Apriori</i> Studi Kasus: Ghiyas <i>Cake & Bakery</i> [10]	2022	<i>Association Rules, Algoritma Apriori</i>	Terdapat 4 <i>rules</i> dengan min_support 15% dan min_conf 70%, yaitu {soes → kue talam}, {kue talam, buka ambon → soes}, {bika ambon, soes → kue talam}, {brownies, burger → karamel}
3	Implementasi Algoritma <i>Apriori</i> pada <i>Market Basket Analysis</i> (MBA) Aksesoris Telepon Seluler [19]	2022	<i>Association Rules, Algoritma Apriori</i>	Rule {Spiral Cord, Kepala Casan Samsung→Kabel USB} memiliki nilai confidence tertinggi, yaitu 100%
4	<i>Market Basket Analysis with Equivalence Class Transformation Algorithm</i> (ECLAT) for Inventory Management Using Economic Order Quantity [17]	2022	<i>Association Rules, ECLAT Algorithm</i>	Terdapat 20 <i>rules</i> yang dihasilkan dari penggunaan algoritma ECLAT pada 212 <i>Mart</i> dengan nilai min_supp = 0.1% dan min_conf = 30%
5	<i>Analysis of Contributory Factors of Fatal Pedestrian Crashes by Mixed Logit Model and Association Rules</i> [22]	2022	<i>Association Rules</i>	Aturan asosiasi menyediakan 119 aturan valid berupa faktor penyebab kecelakaan dengan menganalisis 101,032 kasus kecelakaan pejalan kaki di Italia
6	Analisis <i>Frequent Itemset</i> dan Desain Rekomendasi Paket Produk pada Adiksi <i>Coffee</i> berbasis <i>Market Basket Analysis</i> [30]	2022	<i>Association Rules, Algoritma Apriori</i>	Terdapat 7 <i>rules</i> yang dihasilkan pada Adiksi <i>Coffee</i> berdasarkan pengelompokan data <i>all transaction</i> (semua periode waktu) dengan min_supp = 0.01, min_conf = 0.2, dan lift > 1

Metodologi Penelitian



Catatan:

- Tujuan Penelitian 1 diperoleh melalui **step 1, 2, dan 3**
- Tujuan Penelitian 2 diperoleh melalui **step 4, 5, dan 6**
- Tujuan Penelitian 3 diperoleh melalui **step 7**
- Tujuan Penelitian 4 diperoleh melalui **step 8**

Analisis Kebutuhan dan Proses Bisnis

1 Aktor yang terlibat dalam pengembangan sistem pemasaran produk bakery:

Technical Staff

Data Analyst

Manajer

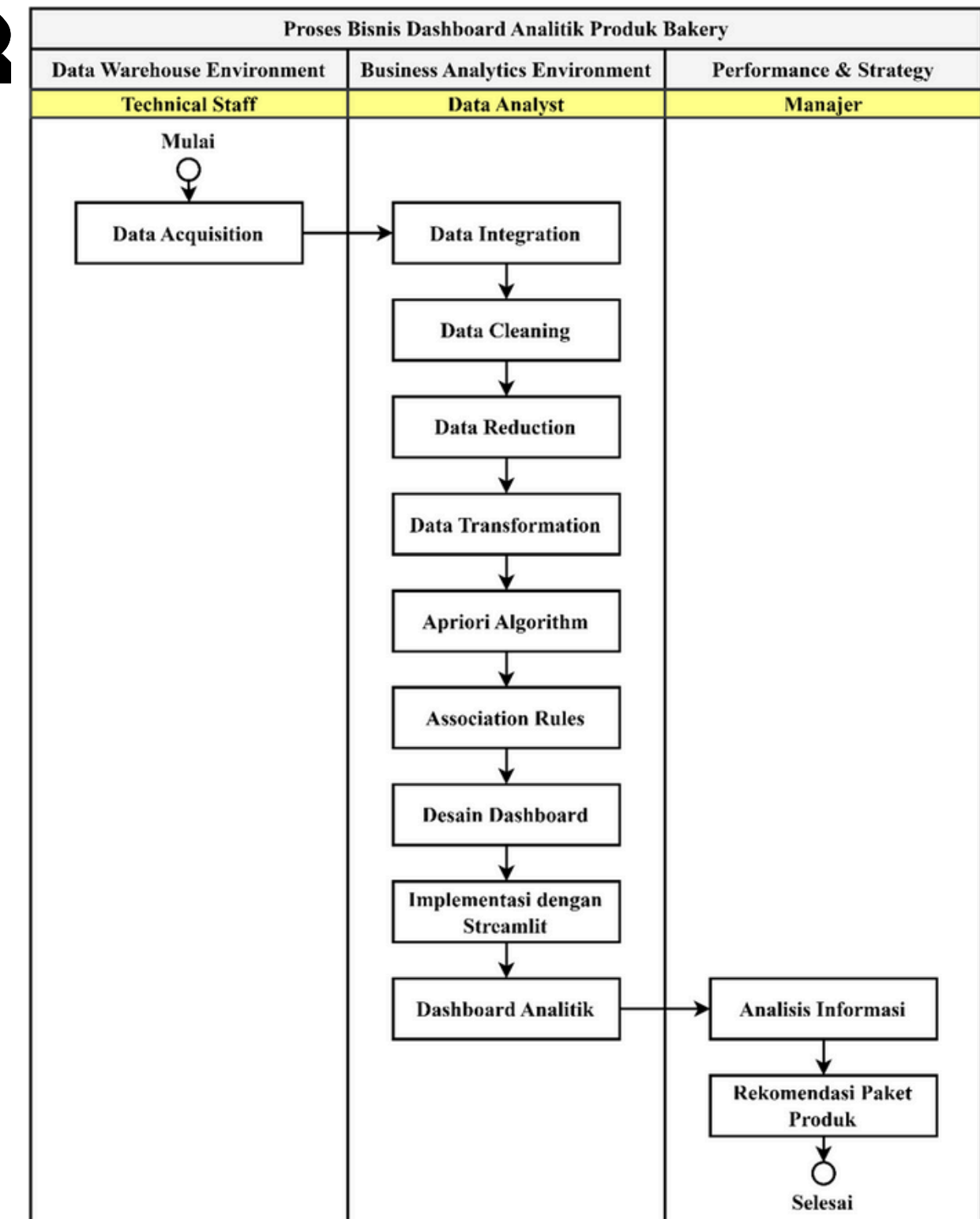
Analisis Entitas Sistem Dashboard Analitik

No	Input	Process	Output
1	Dataset produk bakery	Mengakumulasi total transaksi, total <i>item</i> yang terjual, dan <i>item</i> yang memiliki penjualan terbanyak	KPI (<i>key performance indicator</i>) berupa total transaksi, total <i>item</i> terjual, dan <i>item best seller</i>
2	Tabel <i>frequent itemset</i>	Menentukan variabel yang menjadi nilai x dan y berturut-turut, yaitu atribut <i>item</i> dan total	Total penjualan per <i>item</i> dalam bentuk <i>bar chart</i>
3	Dataset produk bakery dan dataset hasil transformasi atribut <i>Item</i>	Menampilkan dataset produk bakery dan dataset hasil transformasi	Tabel dataset penjualan dan tabel transaksi
4	Dataset <i>association rules</i>	Menentukan variabel yang menjadi nilai x, y, dan z berturut-turut, yaitu atribut <i>support</i> , <i>confidence</i> , dan <i>lift ratio</i>	Aturan asosiasi dalam bentuk <i>scatter plot</i>
5	Dataset <i>association rules</i>	Menampilkan dataset <i>association rules</i> dari hasil <i>knowledge extraction</i>	Tabel basis aturan asosiasi dan tabel aturan yang terpilih

3

terdapat 5 fitur yang menyajikan hasil analisis deskriptif (1, 2, 3) dan prediktif (4, 5) pada dataset produk bakery

2



Informasi Dataset

Tampilan dataset produk bakery

<i>Date</i>	<i>Time</i>	<i>Transaction</i>	<i>Item</i>
10/30/2016	9:58:11	1	Bread
10/30/2016	10:05:34	2	Scandinavian
10/30/2016	10:05:34	2	Scandinavian
10/30/2016	10:07:57	3	Hot chocolate
...
4/9/2017	14:32:58	9682	Coffee
4/9/2017	14:32:58	9682	Tea
4/9/2017	14:57:06	9683	Coffee
4/9/2017	14:57:06	9683	Pastry
4/9/2017	15:04:24	9684	Smoothies

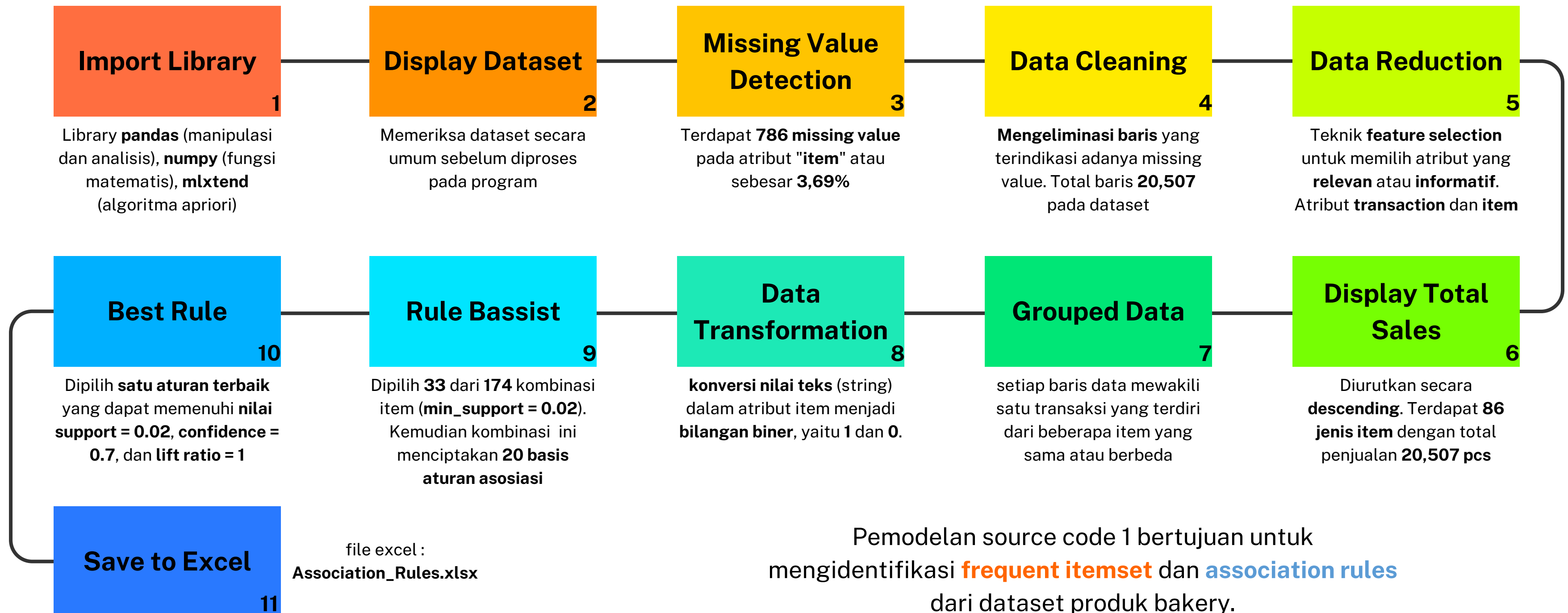
Dataset produk bakery berasal dari situs **kaggle** yang diunggah oleh **Sulman Sarwar** pada tahun **2019**.

Deskripsi Atribut

No	Atribut	Deskripsi
1	<i>Date</i>	Merupakan tanggal dari pencatatan setiap <i>item</i> yang terjual, memiliki format yang dimulai dari <i>month</i> , <i>day</i> , dan <i>year</i> (mm/dd/yyyy)
2	<i>Time</i>	Berisi waktu terjualnya setiap <i>item</i> , memiliki format AM/PM yang dimulai dari <i>hour</i> , <i>minute</i> , dan <i>second</i> (hh:mm:ss)
3	<i>Transaction</i>	Merupakan identitas nomor transaksi untuk setiap <i>item</i> yang terjual
4	<i>Item</i>	Berisi nama produk yang terjual per <i>item</i>

Memiliki **21,293 baris**, **4 kolom** (atribut), **9,684 transaksi**, dan waktu pencatatan selama **162 hari**.

Proses Pemodelan Source Code 1



Pemodelan source code 1 bertujuan untuk mengidentifikasi **frequent itemset** dan **association rules** dari dataset produk bakery.

Desain Dashboard Analitik

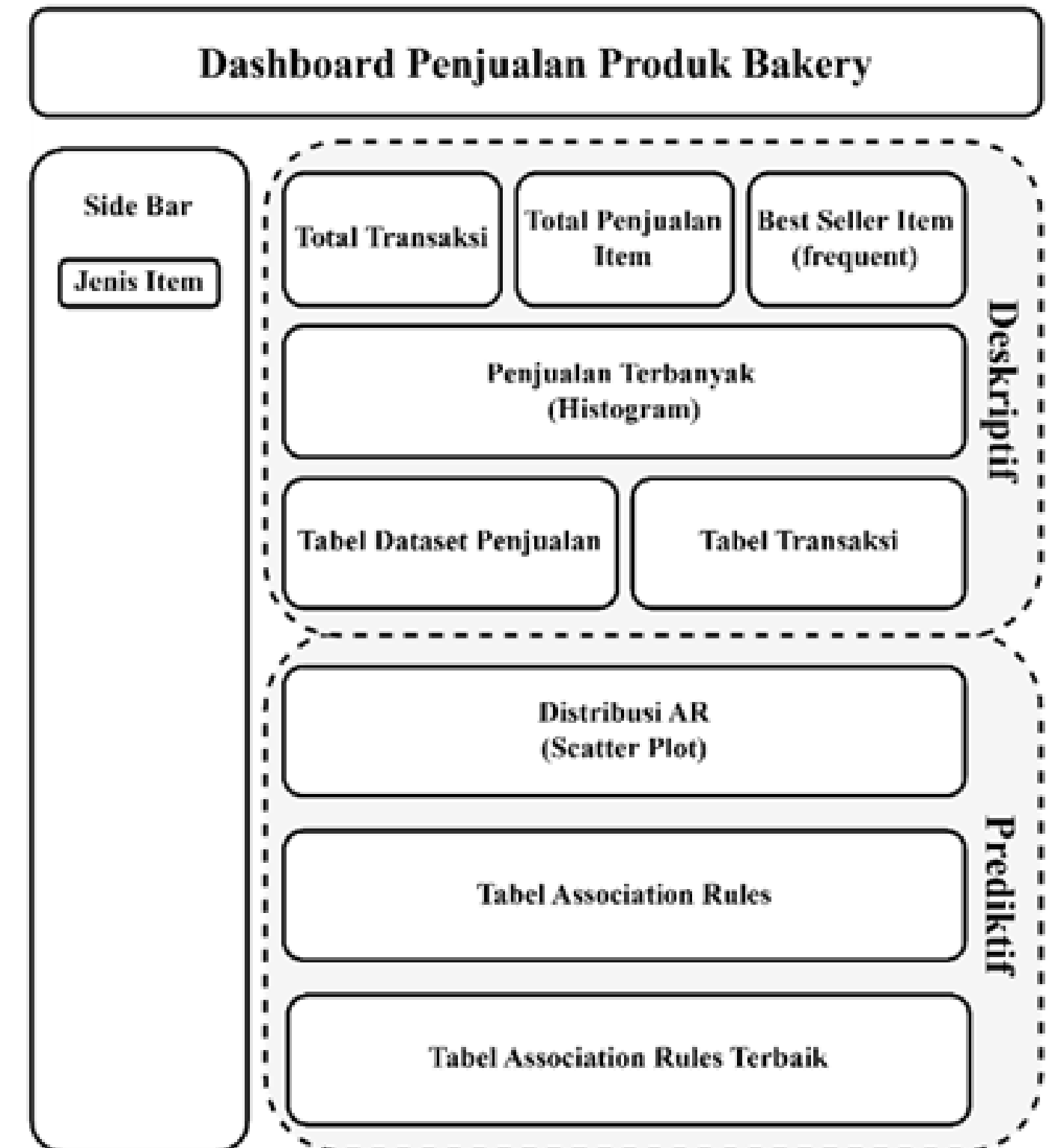
Menurut [40] terdapat elemen penting pada desain dashboard seperti **data, structure, visual representation, page layout, screen space, interaction, meta data, dan color**

Elemen tersebut **diintegrasikan** untuk menjadi satu kesatuan desain yang baik untuk **meningkatkan penggunaannya** oleh user

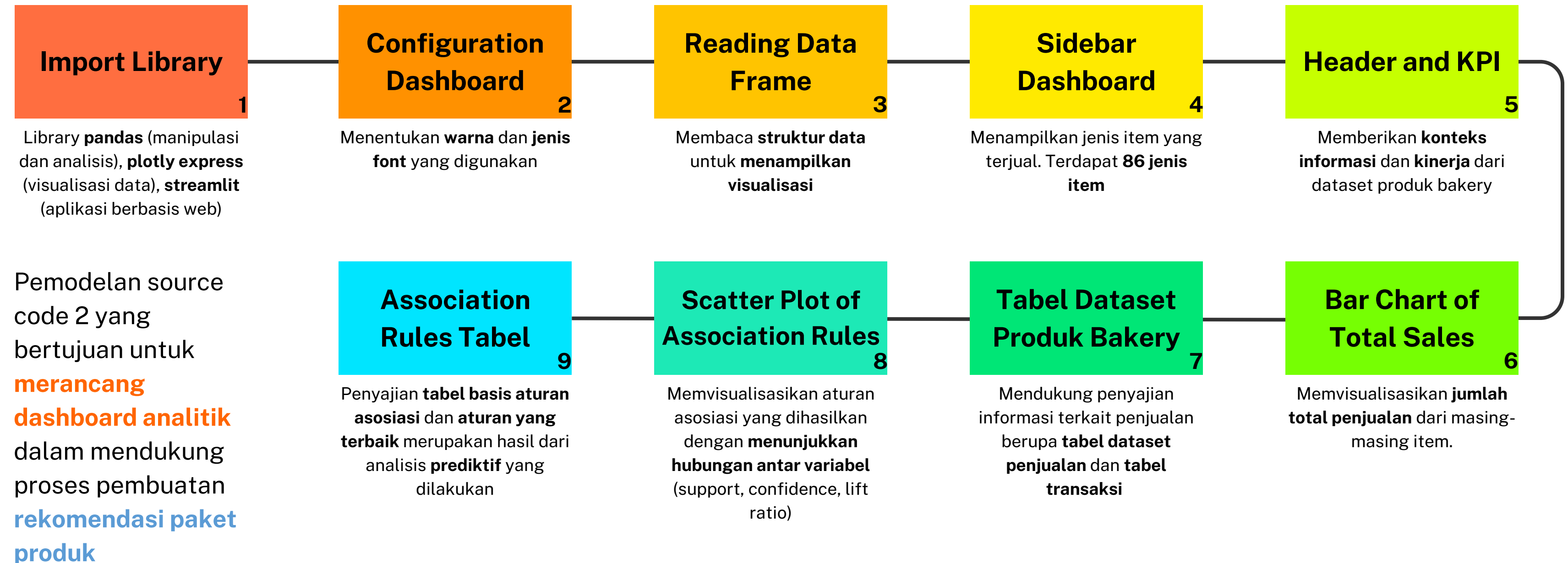
Pemilihan warna yang digunakan berasal dari **situs colorhunt.co**, yaitu memilih palet warna yang memiliki **popularitas tertinggi**

Jenis font yang digunakan adalah "**sans-serif**" karena memiliki tampilan yang jelas, mudah dibaca, modern, profesional, dan fleksibel

No	Warna	Kode Warna (HEX)	Keterangan
1		#222831	Background color
2		#393E46	Secondary background color
3		#00ADB5	Primary color
4		#EEEEEE	Text color



Proses Pemodelan Source Code 2



No	Itemsets	Support
1	{'coffee'}	0.478
2	{'bread'}	0.327
3	{'tea'}	0.143
4	{'cake'}	0.104
5	{'coffee', 'bread'}	0.090
6	{'pastry'}	0.086
7	{'sandwich'}	0.072
8	{'medialuna'}	0.062
9	{'hot chocolate'}	0.058
10	{'cake', 'coffee'}	0.055
11	{'cookies'}	0.054
12	{'tea', 'coffee'}	0.050
13	{'coffee', 'pastry'}	0.048
14	{'brownie'}	0.040
15	{'farm house'}	0.039
16	{'juice'}	0.039
17	{'muffin'}	0.038
18	{'coffee', 'sandwich'}	0.038
19	{'alfajores'}	0.036
20	{'coffee', 'medialuna'}	0.035
21	{'scone'}	0.035
22	{'soup'}	0.034
23	{'toast'}	0.034
24	{'coffee', 'hot chocolate'}	0.030
25	{'bread', 'pastry'}	0.029
26	{'scandinavian'}	0.029
27	{'cookies', 'coffee'}	0.028
28	{'tea', 'bread'}	0.028
29	{'cake', 'tea'}	0.024
30	{'coffee', 'toast'}	0.024
31	{'cake', 'bread'}	0.023
32	{'coffee', 'juice'}	0.021
33	{'truffles'}	0.020

Frequent Itemset

Frequent Itemset
dihasilkan melalui
penggunaan
algoritma apriori
dengan menetapkan
nilai minimum
support = 0.02

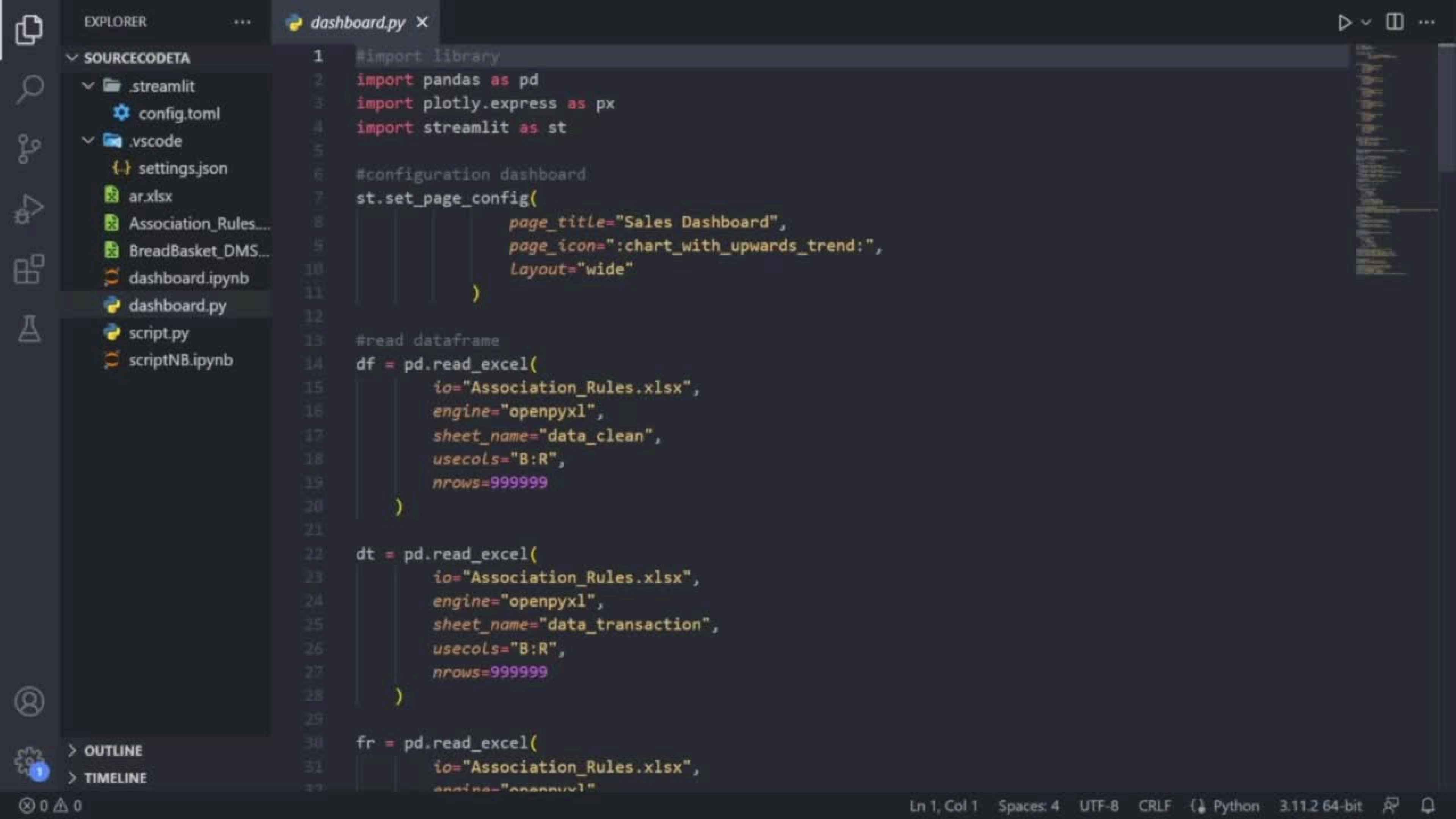
Penentuan aturan
asosiasi terbaik
dipilih dari perolehan
nilai confidence =
0.7, dan **nilai lift**
ratio = 1.

Rule {**toast** → **coffee**}
menjadi dasar
penentuan dari
rekomendasi paket
penjualan produk
bakery

Hasil Ekstraksi Informasi

No	Antecedents	Consequents	Support	Confidence	Lift Ratio
1	{'toast'}	{'coffee'}	0.024	0.704	1.472
2	{'medialuna'}	{'coffee'}	0.035	0.569	1.190
3	{'pastry'}	{'coffee'}	0.048	0.552	1.154
4	{'juice'}	{'coffee'}	0.021	0.534	1.117
5	{'sandwich'}	{'coffee'}	0.038	0.532	1.113
6	{'cake'}	{'coffee'}	0.055	0.527	1.102
7	{'cookies'}	{'coffee'}	0.028	0.518	1.084
8	{'hot chocolate'}	{'coffee'}	0.030	0.507	1.060
9	{'pastry'}	{'bread'}	0.029	0.339	1.035
10	{'cake'}	{'tea'}	0.024	0.229	1.605
11	{'tea'}	{'cake'}	0.024	0.167	1.605
12	{'coffee'}	{'cake'}	0.055	0.114	1.102
13	{'coffee'}	{'pastry'}	0.048	0.099	1.154
14	{'bread'}	{'pastry'}	0.029	0.089	1.035
15	{'coffee'}	{'sandwich'}	0.038	0.080	1.113
16	{'coffee'}	{'medialuna'}	0.035	0.074	1.190
17	{'coffee'}	{'hot chocolate'}	0.030	0.062	1.060
18	{'coffee'}	{'cookies'}	0.028	0.059	1.084
19	{'coffee'}	{'toast'}	0.024	0.049	1.472
20	{'coffee'}	{'juice'}	0.021	0.043	1.117
Keterangan					
Aturan asosiasi yang terbaik					

Association Rules



Kesimpulan

- 1 Pengembangan sistem pemasaran produk bakery terletak pada **pemanfaatan data transaksi** untuk mendukung **pengambilan keputusan** yang disajikan pada **dashboard analitik**
- 2 Dataset produk bakery hanya **menghasilkan satu aturan asosiasi terbaik** yang memenuhi ukuran kinerja rule dari total 20 aturan yang telah diciptakan, yaitu **jika membeli toast maka akan membeli coffee**
- 3 Rule {toast → coffee} menjadi **dasar penentuan** dari **rekomendasi paket penjualan produk bakery**
- 4 Dashboard analitik produk bakery menyajikan informasi terkait hasil **analisis deskriptif** dan **prediktif**.

Saran

Penelitian selanjutnya dapat menggunakan dataset yang berasal dari industri bakery yang memungkinkan untuk dilakukannya **penerapan hasil penelitian secara realistis**

Penelitian selanjutnya diharapkan mampu untuk **mengintegrasikan API** dalam **penggunaan dashboard analitik**

Sekian dan Terima Kasih

Proses Perhitungan Secara Manual

- Min Support = 30%
- Min Confidence = 70%

$$Support(A) = \frac{\text{Jumlah transaksi item A pada dataset}}{\text{Total transaksi}}$$

$$Support(A \cap B) = \frac{\text{Jumlah transaksi yang berisi item A dan B}}{\text{Total Transaksi}}$$

1

Tid	Itemset
1	coklat, stroberi, abon
2	tart, coklat
3	coklat, stroberi
4	brownies, tart, abon
5	stroberi, abon, coklat

Tabel 3.4 Iterasi 1 Contoh Perhitungan Nilai Support

2

Items	Support
coklat	0.8
stroberi	0.6
abon	0.6
brownies	0.4
tart	0.2

Tabel 3.5 Iterasi 2 Contoh Perhitungan Nilai Support

3

Items	Support
coklat, stroberi	0.6
coklat, abon	0.4
coklat, brownies	0.2
stroberi, abon	0.4
stroberi, brownies	0
abon, brownies	0.2

- Jika mencari nilai *confidence* A terhadap B:

$$Confidence(A \mid B) = \frac{Support(A \cap B)}{Support(B)} \quad (3.4)$$

- Jika mencari nilai *confidence* A, B terhadap C:

$$Confidence(A, B \mid C) = \frac{Support(A \cap B \cap C)}{Support(A \cap B)} \quad (3.5)$$

Tabel 3.7 Hasil dari Contoh Perhitungan Nilai Confidence

5

Rules	Confidence
coklat → stroberi	0.75
stroberi → coklat	1
coklat → abon	0.50
abon → coklat	0.67
stroberi → abon	0.67
abon → stroberi	0.67
coklat, stroberi → abon	0.67
coklat, abon → stroberi	1
abon, stroberi → coklat	1

$$Support(A \cap B \cap C) = \frac{\text{Jumlah transaksi yang berisi item A, B, dan C}}{\text{Total Transaksi}}$$

Tabel 3.6 Iterasi 3 Contoh Perhitungan Nilai Support

4

Items	Support
coklat, stroberi, abon	0.4

Proses Perhitungan Secara Manual (2)

- Min Support = 30%
- Min Confidence = 70%

Tabel 3.7 Hasil dari Contoh Perhitungan Nilai *Confidence*

6

Rules	Confidence
coklat → stroberi	0.75
stroberi → coklat	1
coklat → abon	0.50
abon → coklat	0.67
stroberi → abon	0.67
abon → stroberi	0.67
coklat, stroberi → abon	0.67
coklat, abon → stroberi	1
abon, stroberi → coklat	1



Tabel 3.8 Contoh Hasil Perhitungan Nilai *Lift Ratio*

7

Rules	Lift
coklat → stroberi	1.25
stroberi → coklat	1.25
coklat, abon → stroberi	1.67
abon, stroberi → coklat	1.25



Tabel 3.9 Ketentuan *Lift Ratio*

8

Nilai <i>Lift Ratio</i>	Keterangan
< 1	Korelasi Negatif
= 1	Tidak Berkorelasi
> 1	Korelasi Positif

- Jika mencari nilai *lift ratio* A terhadap B:

$$Lift\ Ratio(A \mid B) = \frac{Support(A \cap B)}{Support(A) \times Support(B)} \quad (3.6)$$

- Jika mencari nilai *lift ratio* A, B terhadap C:

$$Lift\ Ratio(A, B \mid C) = \frac{Support(A \cap B \cap C)}{Support(A \cap B) \times Support(C)} \quad (3.7)$$

Dashboard Analitik

Mengapa **Pengembangan dashboard** menggunakan Python **tidak membutuhkan** diagram Entity Relationship Diagram (ERD) dan Data Flow Diagram (DFD)?

- **Dashboard** pada umumnya digunakan untuk **visualisasi data**, bukan untuk **memodelkan struktur data atau alur data**
- **ERD dan DFD** biasanya digunakan dalam **pengembangan sistem basis data atau perangkat lunak yang lebih kompleks**, di mana pemodelan data dan analisis alur data diperlukan
- Sedangkan untuk penggunaan **python** dalam membuat dashboard, **data sudah ada dan siap untuk digunakan**

Dasar pertimbangan mengapa menggunakan library **streamlit** untuk pengembangan **dashboard analitik**?

- Kelebihan
 - **Mudah dipelajari dan digunakan**, (memiliki interface yang sederhana, sehingga proses pengembangannya menjadi lebih cepat dan efisien)
 - **Memiliki banyak komponen dan fitur**: Streamlit menyediakan banyak komponen dan fitur yang berguna untuk membuat dashboard yang interaktif dan dinamis. Contohnya adalah plot dan grafik yang bisa langsung digunakan dan disesuaikan dengan kebutuhan.
 - **Dukungan terhadap banyak jenis data**: Streamlit mendukung banyak jenis data, termasuk data tabular, data waktu, dan data geografis. Hal ini memudahkan pengembang untuk menampilkan data dengan berbagai format dan jenis.
 - **Pengembangan dashboard yang responsif**: Streamlit memungkinkan pengembangan dashboard yang responsif, yang dapat disesuaikan dengan berbagai ukuran layar dan perangkat.
- Kekurangan
 - **Tidak cocok untuk aplikasi kompleks; Tidak ada dukungan terhadap database**
 - **Terbatas pada Python; Masih dalam tahap pengembangan** (baru diperkenalkan tahun **2019**)