

BAB III

METODOLOGI PENELITIAN

3.1. Subyek dan Obyek Penelitian

Subyek penelitian yang dipilih oleh peneliti adalah perilaku pelanggan dalam membeli barang. Obyek penelitian yang menjadi fokus peneliti adalah data transaksi toko swalayan yang diperoleh secara daring.

3.2. Alat dan Bahan Penelitian

Alat dan bahan yang digunakan dalam penelitian ini antara lain:

3.2.1 Perangkat Keras

Perangkat keras yang digunakan adalah laptop yang memiliki spesifikasi sebagai berikut:

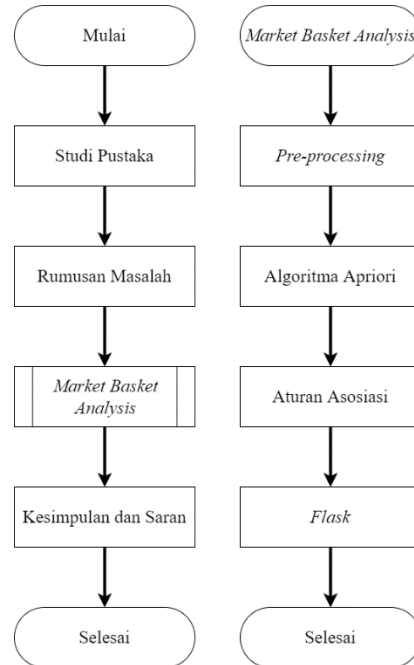
- *Processor Intel® Core™ i7-4510U CPU @ 2.0 GHz ~ 2.6 GHz*
- *RAM 4 GB*
- *Graphic card NVIDIA GeForce 940M 4 GB*
- *SSD ADATA 256 GB + HDD 1 TB*

3.2.2 Perangkat Lunak

Perangkat Lunak yang digunakan dalam penelitian ini antara lain:

- *Windows 10 Home 64-Bit*
- *Jupyter Notebook*
- *Google Colaboratory*
- *Python 3.9.0*
- *Microsoft Edge*

3.3. Diagram Alir Penelitian



Gambar 3.1. Diagram Alir Penelitian

3.3.1 Studi Pustaka

Tahap awal penelitian ini, penulis melakukan kegiatan membaca dan memahami konsep serta memahami permasalahan mengenai analisis *market basket* dengan algoritma apriori yang bersumber pada jurnal, prosiding, dan buku. Dengan melakukan studi pustaka penulis memiliki dasar dalam melakukan dan menulis penelitian.

3.3.2 Rumusan Masalah

Tahap selanjutnya adalah merumuskan masalah. Rumusan masalah memiliki peran penting sebagai kerangka pemikiran dan konsep dari topik yang diambil. Anjuran terhadap rumusan yang dipilih adalah sejalan dengan visi Institut Teknologi Telkom Purwokerto yaitu menjadi perguruan tinggi yang unggul di tingkat internasional dalam pengembangan ilmu pengetahuan berbasis teknologi dengan keunggulan pada bidang *Healthcare, Agro-industry, Tourism, dan Small-Medium Enterprise*.

3.3.3 Market Basket Analysis

Proses *market basket analysis* dimulai dengan melakukan *pre-processing* data transaksi yang telah diperoleh dari website *kaggle*. Contoh format data yang didapatkan melalui website *kaggle*, ditampilkan pada Gambar 3.2.

	Member_number	Date	itemDescription
0	1808	21-07-2015	tropical fruit
1	2552	05-01-2015	whole milk
2	2300	19-09-2015	pip fruit
3	1187	12-12-2015	other vegetables
4	3037	01-02-2015	whole milk

Gambar 3.2. Contoh Format Data Transaksi

Format data tersebut masih berupa data tabel yang memiliki komposisi 1 baris untuk memuat 1 barang, maka dari itu perlu dilakukan pengelompokan sehingga menjadi 1 baris untuk 1 nota dan beberapa barang yang dibeli secara bersamaan di dalamnya, data setelah dikelompokkan ditampilkan pada Gambar 3.3.

Member_number	Date	
1000	15-03-2015	[sausage, whole milk, semi-finished bread, yog...
	24-06-2014	[whole milk, pastry, salty snack]
	24-07-2015	[canned beer, misc. beverages]
	25-11-2015	[sausage, hygiene articles]
	27-05-2015	[soda, pickled vegetables]

Gambar 3.3. Pengelompokan Barang Berdasarkan Tanggal Pembelian

Setelah format data sesuai dengan pembelian 1 keranjang dengan beberapa barang di dalamnya, dilanjutkan dengan ekstraksi data sehingga mendapat sebuah *list* yang nantinya akan dirubah menjadi *one hot encoding table*. Penghapusan kolom yang dilakukan antara lain pada kolom nota dan tanggal sehingga tersisa daftar barang seperti berikut.

```
[['sausage', 'whole milk', 'semi-finished bread', 'yogurt'],
 ['whole milk', 'pastry', 'salty snack'],
 ['canned beer', 'misc. beverages'],
 ['sausage', 'hygiene articles'],
 ['soda', 'pickled vegetables'],
 ['frankfurter', 'curd'],
 ['sausage', 'whole milk', 'rolls/buns'],
 ['whole milk', 'soda'],
 ['beef', 'white bread'],
 ['frankfurter', 'soda', 'whipped/sour cream'],
 ['frozen vegetables', 'other vegetables'],
```

Gambar 3.4. Bentuk Data Dalam List

Setelah berhasil dikelompokkan dan dibersihkan, tahap selanjutnya adalah merubah format data menjadi *one hot encoding format* yang memudahkan menggunakan *library transaction encoder*. Hasil dari proses perubahan format ditampilkan pada Gambar 3.5.

	Instant food products	UHT- milk	abrasive cleaner	artif. sweetener	baby cosmetics	bags	baking powder
0	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False

Gambar 3.5. Data List dirubah menjadi binary table

Setelah melakukan *pre-processing* data, tahap selanjutnya dari *market basket analysis* adalah melakukan pembuatan candidate *itemset* dan melakukan perhitungan support. Hal ini dipermudah dengan menggunakan *library mlxtend* bagian *apriori*, *library* ini menerima masukan *one hot encoding format* lalu saat pemrosesan selesai akan menghasilkan *pandas dataframe* yang memiliki kolom *support* dan *itemset* yang memenuhi nilai *minimum support*. Nilai *minimum support* dapat ditentukan secara bebas, belum ada publikasi yang menjelaskan tentang penetapan nilai *minimum support*. Hasil dari fungsi *apriori* dapat dilihat pada Gambar 3.6.

	support	itemsets	length
0	0.004010	(Instant food products)	1
1	0.021386	(UHT-milk)	1
2	0.001470	(abrasive cleaner)	1
3	0.001938	(artif. sweetener)	1
4	0.008087	(baking powder)	1
...
745	0.001136	(whole milk, rolls/buns, sausage)	3
746	0.001002	(whole milk, soda, rolls/buns)	3
747	0.001337	(whole milk, yogurt, rolls/buns)	3
748	0.001069	(soda, whole milk, sausage)	3
749	0.001470	(yogurt, whole milk, sausage)	3

Gambar 3.6. Hasil perhitungan menggunakan algoritma apriori

Setelah proses *pre-processing* dan proses pembangkitan *itemset* menggunakan algoritma apriori dilakukan, tahap selanjutnya dalam melakukan *market basket analysis* adalah membuat aturan – aturan asosiasi berdasarkan *itemset* yang telah dihasilkan sebelumnya pada proses pembangkitan *itemset* dengan algoritma apriori. Sehingga proses pembangkitan aturan – aturan asosiatif menghasilkan informasi yang ditampilkan pada Gambar 3.7.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(bottled water)	(UHT-milk)	0.060683	0.021386	0.001069	0.017621	0.823954	-0.000228	0.996168
1	(UHT-milk)	(bottled water)	0.021386	0.060683	0.001069	0.050000	0.823954	-0.000228	0.988755
2	(other vegetables)	(UHT-milk)	0.122101	0.021386	0.002139	0.017515	0.818993	-0.000473	0.996060
3	(UHT-milk)	(other vegetables)	0.021386	0.122101	0.002139	0.100000	0.818993	-0.000473	0.975443
4	(rolls/buns)	(UHT-milk)	0.110005	0.021386	0.001804	0.016403	0.767013	-0.000548	0.994934

Gambar 3.7. Hasil pembentukan aturan asosiasi

Aturan – aturan tersebut adalah hasil dari *market basket analysis* yang dapat digunakan sebagai rekomendasi untuk menyusun strategi pemasaran, seperti pembentukan paket promo, pemberian diskon, dan pembenahan peletakan barang di pajangan. Setelah model dibuat, tahap selanjutnya merupakan implementasi model ke website *dengan framework flask* dan *html*.

3.3.4 Kesimpulan dan Saran

Setelah tahap *market basket analysis* selesai dilakukan, tahap selanjutnya adalah menarik kesimpulan dari penelitian yang dilakukan, mengenai apa saja yang ditemukan selama penelitian ini dilakukan. Lalu memberikan saran – saran terhadap kekurangan apa saja yang terjadi di penelitian ini, sehingga dapat menyempurnakan penelitian sejenis di masa yang akan datang.

3.4. Teknik Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah dataset publik yang disediakan oleh website kaggle melalui akun *heeral dedhia* atau melalui link <https://www.kaggle.com/datasets/heeraldedhia/groceries-dataset>

3.5. Analisis Data

Tahap analisis data merupakan tahap awal sebelum *market basket analysis*. Pada tahap ini juga dilakukan pra-pemrosesan data sehingga diperoleh data keluaran yang siap diproses di dalam algoritma apriori. Data transaksi ini peroleh dari *website www.kaggle.com* yang diunggah oleh akun *heeral dedhia*, memiliki 38.765 baris dari pembelian pelanggan pada toko swalayan. Dataset ini memiliki 3 buah kolom yaitu *Member_number*, *Date*, dan *itemDescription*, data yang diperoleh merupakan berjangka 1 tahun dimulai pada tanggal 1 Januari 2014 hingga tanggal 31 Oktober 2015.

Barang yang ada di dalam data transaksi berupa bahan sehari – hari rumah tangga, meliputi bahan makanan mentah, bahan makanan kaleng, makanan instan, pembersih serbaguna, kosmetik, dan barang – barang yang dapat ditemukan di rumah lainnya. Data transaksi ini dapat digunakan sebagai acuan atau proyeksi terhadap data realita pada penelitian *market basket analysis*. Pada website *kaggle*, terdapat 87 penelitian yang dilakukan pada data transaksi ini, mulai dari *exploratory data analysis*, *market basket analysis*, hingga visualisasi data. Data transaksi yang telah diperoleh, ditampilkan pada Gambar 3.8.

	Member_number	Date	itemDescription
0	1808	21-07-2015	tropical fruit
1	2552	05-01-2015	whole milk
2	2300	19-09-2015	pip fruit
3	1187	12-12-2015	other vegetables
4	3037	01-02-2015	whole milk
...
38760	4471	08-10-2014	sliced cheese
38761	2022	23-02-2014	candy
38762	1097	16-04-2014	cake bar
38763	1510	03-12-2014	fruit/vegetable juice
38764	1521	26-12-2014	cat food

38765 rows x 3 columns

Gambar 3.8. Kilas Dataset

Data pada Gambar 3.8. masih berbentuk 1 baris 1 barang dalam nomor anggota yang sama, maka perlu dilakukan pengelompokkan sesuai dengan nomor anggota dan tanggal pembelian. Menggunakan fungsi *groupby* yang disediakan pandas, maka pengelompokkan berdasarkan nomor anggota dan tanggal pembelian menghasilkan data terkelompok seperti gambar 3.9.

Member_number	Date	itemDescription
1000	15-03-2015	[sausage, whole milk, semi-finished bread, yog...
	24-06-2014	[whole milk, pastry, salty snack]
	24-07-2015	[canned beer, misc. beverages]
	25-11-2015	[sausage, hygiene articles]
	27-05-2015	[soda, pickled vegetables]
...
4999	24-01-2015	[tropical fruit, berries, other vegetables, yo...
	26-12-2015	[bottled water, herbs]
5000	09-03-2014	[fruit/vegetable juice, onions]
	10-02-2015	[soda, root vegetables, semi-finished bread]
	16-11-2014	[bottled beer, other vegetables]

14963 rows x 1 columns

Gambar 3.9. Hasil Pengelompokan Data Transaksi

Setelah dikelompokkan, dapat diketahui bahwa jumlah baris berkurang dari 38.765 baris menjadi 14.963 baris. Tahap selanjutnya adalah menghapus kolom *Member_number* dan *Date* yang sudah tidak terpakai, karena barang yang dibeli sudah terkelompok serta menghitung banyaknya barang yang ada dalam 1 transaksi. Tahap ini tidak memerlukan fungsi khusus, namun menggunakan perulangan pada umumnya. Hasil dari penghapusan kolom *Member_number* dan *Date* serta menghitung banyaknya barang dalam 1 transaksi dapat dilihat pada gambar 3.10.

	itemset	len_itemset
0	[sausage, whole milk, semi-finished bread, yog...	4
1	[whole milk, pastry, salty snack]	3
2	[canned beer, misc. beverages]	2
3	[sausage, hygiene articles]	2
4	[soda, pickled vegetables]	2
...
14958	[tropical fruit, berries, other vegetables, yo...	6
14959	[bottled water, herbs]	2
14960	[fruit/vegetable juice, onions]	2
14961	[soda, root vegetables, semi-finished bread]	3
14962	[bottled beer, other vegetables]	2

14963 rows x 2 columns

Gambar 3.10. Hasil Penghapusan Kolom

2	10012
3	2727
4	1273
5	338
1	205
6	179
7	113
8	96
9	19
10	1

Name: len_itemset, dtype: int64

Gambar 3.11. Jumlah barang dan Jumlah Transaksi

Setelah kolom dihapus maka kita dapat menganalisis perilaku pelanggan dengan memanfaatkan fungsi *value_counts()* pada *library*

pandas, untuk menghitung kolom “*len_itemset*”. Tahap ini menghasilkan jumlah barang yang dibeli dan berapa banyak transaksi dengan kesamaan pembelian jumlah barang, seperti ditampilkan pada gambar 3.11. Dapat diketahui bahwa mayoritas pelanggan pada toko swalayan membeli 2 barang dalam 1 waktu. Dari kesimpulan analisis perilaku pelanggan ini dapat dijadikan proyeksi bahwa mayoritas aturan asosiasi yang dihasilkan dengan algoritma apriori akan berhubungan kuat dengan 2 sampai 4 barang akibat dari munculnya frekuensi tersebut. Tahap selanjutnya adalah merubah *format dataset* yang sebelumnya berbentuk *array*, menjadi *one hot encoding format*.

Tahap ini dilakukan sebagai syarat untuk melakukan pembangkitan *itemset* yang akan dijalankan menggunakan fungsi *apriori()* yang dimiliki oleh *library mlxted*. Tahap perubahan format ini dipermudah dengan bantuan fitur *TransactionEncoder* yang dimiliki oleh *library mlxtend*, sehingga data transaksi berbentuk *one hot encoding*. Bentuk ini yang diinginkan oleh fungsi *apriori()* seperti yang ditampilkan oleh Gambar 3.12.

	Instant food products	UHT-milk	abrasive cleaner	artif. sweetener	baby cosmetics	bags	baking powder	bathroom cleaner	beef	berries	...
0	False	False	False	False	False	False	False	False	False	False	...
1	False	False	False	False	False	False	False	False	False	False	...
2	False	False	False	False	False	False	False	False	False	False	...
3	False	False	False	False	False	False	False	False	False	False	...
4	False	False	False	False	False	False	False	False	False	False	...
...
14958	False	False	False	False	False	False	False	False	False	True	...
14959	False	False	False	False	False	False	False	False	False	False	...
14960	False	False	False	False	False	False	False	False	False	False	...
14961	False	False	False	False	False	False	False	False	False	False	...
14962	False	False	False	False	False	False	False	False	False	False	...

Gambar 3.12. Hasil Proses *TransactionEncoder*

Dalam bentuk *one hot encoding format*, data transaksi yang diperoleh telah siap untuk diproses menggunakan algoritma apriori.