BAB 3
METODE PENELITIAN

3.1 Metode Pengumpulan Data

Untuk memperoleh data-data penelitian, penulis menggunakan beberapa metode pengumpulan data yang dilakukan untuk menunjang penelitian tugas akhir ini. Berikut beberapa metode yang dilakukan :

1. Observasi

   Dalam proses observasi, peneliti melakukan pengamatan langsung terhadap InshofMart. Tentang bagaimana sistem informasi yang ada saat ini sedang berjalan, dapat digunakan untuk apa saja data-data yang tersimpan dalam *database* sistem tersebut, intinya bagaimana InshofMart memberdayakan sistem informasi mereka dan data-data yang dimiliki. Selain mengamati sistemnya, peneliti juga mengamati keadaan InshofMart secara fisik yang mempunyai hubungan dengan masalah yang diangkat dalam penelitian ini. Yaitu gudang penyimpanan barang yang hubungannya dengan pengendalian *stock* barang, dan tatanan rak-rak barang yang hubungannya dengan asosiasi antar *item (market basket analysis).*
2. Studi Kepustakaan

   Untuk menunjang penelitian ini agar dapat dilakukan dengan baik dan setiap materi yang dicantumkan dapat dipertanggungjawabkan keabsahannya dan sitasinya, maka dilakukan studi pustaka dengan mempelajari buku-buku *literature*, laporan-laporan yang berkaitan dengan penelitian dan juga hasil penelitian terdahulu atau yang pernah dilakukan oleh peneliti lain yang ada hubungannya dengan masalah yang penulis teliti saat ini.

Dalam melaksanakan penelitian, diperlukan adanya data-data untuk mendukung penelitian agar sesuai dengan objek yang diteliti. Ada beberapa jenis dan sumber data-data yang perlu diperhatikan dalam memperolehnya, diantaranya adalah :

1. Jenis Data

   Adapun jenis data-data yang penulis gunakan di dalam melakukan penelitian tugas akhir ini adalah sebagai berikut :

   a. Data Kuantitatif

      Merupakan jenis data yang dapat dihitung dan menunjukkan jumlah sesuatu. Data kuantitatif yang digunakan dalam penelitian ini misalnya data jumlah sisa *item* produk, jumlah produk yang terjual, jumlah berapa kali transaksi terjadi, dan sebagainya. Untuk data transaksi yang akan digunakan pada proses *market basket analysis* adalah yang dikumpulkan InshofMart mulai tanggal 01 November 2016 s/d 31 Maret 2017. Sedangkan untuk *clustering* merupakan data pembelian ke *distributor* yang dikumpulkan sekitar tanggal 01 November 2016 s/d 31 Maret 2017. Karena tidak semua produk dibeli tepat tanggal 01 November 2016, sehingga digunakan data yang terjadi pada satu kali pembelian tanggal sebelumnya. Namun data dalam transaksi yang digunakan adalah jumlah sisa produk tepat pada tanggal 01 November 2016, sehingga tidak semua jumlah yang dibeli di kejadian tanggal sebelumnya digunakan. Dan untuk data nama-nama produk yang ada dalam *clustering* juga diambil dari jenis produk yang dibeli oleh pelanggan saja, atau produk yang terdapat di tabel transaksi selama 01 November 2016 s/d 31 Maret 2017 tersebut. Sehingga untuk jenis produk yang tidak terdapat dalam tabel transaksi di kisaran periode tersebut, dihapus atau tidak disertakan dalam himpunan data-data untuk di dalam melakukan proses *clustering* nantinya.

      Jumlah data yang terdapat dalam tabel transaksi selama 01 November 2016 s/d 31 Maret 2017, berjumlah 3.623 *record* data transaksi. Sedangkan untuk data jenis-jenis produk yang dibeli pada *distributor* dan terdapat dalam tabel transaksi 01 November 2016 s/d 31 Maret 2017 berjumlah 1.005 *item*. Untuk data transaksi yang diminta dari InshofMart terdiri dari 4 atribut, yaitu nomor, tanggal, nama produk, dan jumlah. Sedangkan untuk data *stock item* yang diminta terdiri dari 2 atribut, yaitu kategori dan nama produk. Berikut ini adalah merupakan bentuk tabel yang akan digunakan dalam proses penelitian, atribut-atribut dalam tabel *stock* dan transaksi :

Tabel 3.1 Sampel Data Pada Tabel *Stock*

Kategori	Nama Produk
Alat-alat Tulis (AT)	Joyko Correction Tape CF-S201-PT
	Joyko Correction Tape JK-101
	Joyko Isi Pensil PL-05 (2B)
Canned & Lauk Pauk (CN)	Pronas Chicken Sausage (325 gr)
	ABC Sardines Sauce Chilli (155 gr)
	ABC Sardines Sauce Tomato (155 gr)
Coklat (CK)	Silver Queen Almond Milk Chocolate (33 gr)
	Silver Queen Cashew Milk Chocolate (33 gr)
	Delfi Fruit & Nut (30 gr)
Kosmetik Rambut (KR)	Gatsby Watergloss Hard (150 gr)
	Gatsby Watergloss Hyper Solid (75 gr)
	Gatsby Styling Pomade Supreme Grease (80 gr)

Tabel 3.2 Sampel Data Pada Tabel Transaksi

No	Tanggal	Nama Produk	Jumlah
1	01/11/2016	Baby Happy Pants XL 1 Pcs	3
		Mitu Baby Wet Tissue Pink 10s	1
		My Baby Powder Soft & Gentle (150 gr)	1
2	01/11/2016	Aqua (1500 ml)	1
		Snack Kerupuk Tengiri Original	1
		Snack Ring Balado	1
3	01/11/2016	Aqua (1500 ml)	1
		Indofood Saus Pedas Sachet Isi 6 (9 gr)	1
		Okey Sosis Ayam / 1 Potong	3
		Sarimi isi 2 Koya jeruk Nipis (112 gr)	2
		So Good Bakso Kuah Ayam	1
5	01/11/2016	Campina Concerto Sundae Blueberry	2
		Bayclin Botol (100 ml)	1
		Milkuat Rasa Strawberry	2
		Nuvo Bar Soap Family Biru (80 gr)	2
		So Klin Smart Color Deterjen (800 gr)	1

2. Sumber Data
   a. Data Sekunder

      Merupakan data yang didapatkan secara tidak langsung untuk menambah pengetahuan, antara lain data dari literatur, jurnal, atau data pendukung lain yang bukan dari sumber utama yang berhubungan dengan topik penelitian. Misal dalam penelitian ini penulis menggunakan data *stock* produk dan data transaksi yang didapatkan dari InshofMart, bukan memperolehnya langsung dari kegiatan jual beli yang setiap harinya dilakukan antara penjual dengan pembeli dan *distributor*.

3.2 Metode Analisis

3.2.1 CRISP-DM (*Cross Industry Standard Process For Data Mining*)

Ada 6 tahap dalam fase CRISP-DM, namun pada penelitian ini hanya sampai fase ke-4 dikarenakan tahap evaluasi dan *deployment* perlu penelitian lebih lanjut dan dukungan dari pihak luar. Berikut 4 fase yang digunakan dalam penelitian :

1. *Business Understanding*

   Dinamakan juga sebagai pemahaman penelitian. Tahap ini sangat diperlukan kapasitas pemahaman yang bagus tentang substansi dari kegiatan *datamining* yang akan dilakukan, atau merupakan kebutuhan dari sudut pandang bisnis. *Business understanding* terdiri dari beberapa sub tahapan, yang diantaranya yaitu: menentukan tujuan atau sasaran bisnis, memahami proses bisnis yang sedang berjalan, dan menentukan tujuan dari *data mining*.
2. *Data Understanding*

   Merupakan fase di mana data-data mentah atau data awal dikumpulkan, lalu memahami data-data tersebut untuk dapat mengenal data apa saja yang akan digunakan dalam proses *data mining*. Fase ini mencoba mengidentifikasikan masalah apa saja yang berkaitan dengan kualitas data, mendeteksi *subset* yang menarik dari data untuk membuat hipotesa awal.
3. *Data Preparation*

   Disebut juga sebagai persiapan data, yang mana tahapan ini lah yang dapat dikatakan paling rumit atau banyak membutuhkan waktu karena ada banyak

sub-sub fase yang memang perlu dilakukan seperti *cleaning* data, *integrating*, *selecting*, bahkan pembuatan atribut baru jika dibutuhkan, dan sebagainya.

4. *Modeling*

   Merupakan fase penentuan tehnik atau algoritma *data mining* yang akan digunakan, *tools* yang dapat digunakan untuk mendukung algoritma, dan juga parameter dengan nilai yang optimal.

3.2.2 Metode *Clustering* Menggunakan K-*Means*

Untuk menemukan pola-pola data dari proses *clustering* berbasis k-*means* dan mendukung tahap pemodelan pada CRISP-DM, berikut langkah-langkah yang perlu dilakukan [6], [7], [8] :

1. Menentukan dahulu berapa jumlah *cluster* yang akan dibentuk (k-*cluster*).
2. Menentukan titik pusat *cluster (centroid)* secara acak.
3. Menghitung jarak antara setiap data ke titik pusat terdekat menggunakan rumus *Euclidean Distance*. Berikut ini merupakan persamaannya yaitu :

   $$Dn\ (x,y) = \sqrt{\sum_{k=1}^{p}\left(X_{xk} - Y_{yk}\right)^2} \qquad (3.1)$$

   Dengan :

   Dn (x,y) = jarak objek antara objek i dan j

   P = dimensi data

   X_{xk} = koordinat dari objek x pada dimensi k

   Y_{yk} = koordinat dari objek y pada dimensi k

4. Mengalokasikan masing-masing data ke titik pusat terdekat, sehingga dalam tahap ini setiap data sudah mulai bergerak ke *cluster*-nya masing-masing.
5. Menghitung titik pusat baru menggunakan jumlah keseluruhan per *cluster.*

   Untuk menghitung titik pusat *cluster* yang baru ditentukan berdasarkan rata-rata dari data-data yang ada di titik pusat yang sama. Berikut persamaannya :

   $$C_k = \left(\frac{1}{n_k}\right)\sum d_i \qquad (3.2)$$

Dengan :

C_k = titik pusat baru

n_k = jumlah dokumen dalam *cluster* k

d_i = dokumen dalam *cluster* k

6. Kembali ke langkah 3, jika masih ada data yang berpindah *cluster* atau apabila perubahan nilai *centroid* ada yang di atas nilai *threshold* atau jika perubahan nilai *objective function* yang digunakan di atas nilai *threshold* yang ditentukan.

3.2.3 Metode *Market Basket Analysis* Menggunakan *Fuzzy C-Covering*

Fuzzy c-covering adalah salah satu metode pengklasifikasian *item-item* dari suatu himpunan yang universal ke dalam partisi-partisi berbentuk *fuzzy sets*. Untuk menemukan pola data dari proses *market basket analysis* berbasis *fuzzy c-covering* berikut langkah-langkah yang perlu dilakukan [4], [5]:

1. Menentukan *max_item_threshold* sesuai yang diperlukan. *Max_item_threshold* merupakan pembatas untuk menyaring transaksi berdasar pada jumlah *item-item* yang ada dalam transaksi tersebut.
2. Mencari *record* data yang berada di tabel transaksi yang sesuai dengan batasan *max_item_threshold* dan kemudian disimpan ke dalam QT.
3. *Set* k = 1 (k merupakan variabel yang menentukan berapa jumlah kombinasi).
4. Menentukan *min_support* ke-k yang merupakan *threshold* untuk kombinasi k-*item* terhadap besaran dominasinya dari semua transaksi yang terjadi.
5. Mencari *support* yang terjadi di tiap kombinasi k-*item* yang mungkin terdapat di dalam transaksi tersebut.
6. Melakukan penyaringan pada kombinasi *item* yang terdapat di dalam transaksi yang tidak memenuhi : *support* (u) ≥ *min_support* ke-k.
7. *Set* k = k + 1, di mana jika k > ith, maka melanjutkan langkah 9 (sembilan).
8. Mencari kombinasi k-*item* yang mungkin terdapat di tiap kombinasi (k-1)-*item* yang mencapai *minimum_support* yang sudah ditentukan, dengan langkah :

a. Untuk memperoleh kombinasi k-*item*, u, harus ada kombinasi (k-1)-*item*, u', di mana u' $\subset$ u. Misalkan untuk mendapat u = $\{I_{01}, I_{02}, I_{03}, I_{04}\}$ harus ada u' = $\{I_{01}, I_{02}, I_{03}\}$, $\{I_{01}, I_{02}, I_{04}\}$, $\{I_{01}, I_{03}, I_{04}\}$, $\{I_{02}, I_{03}, I_{04}\}$.

b. Jika tidak ada kombinasi k-*item* lain yang memenuhi *min_support* yang telah ditentukan, dapat ke langkah 9, selain itu ulangi langkah 4 sampai 7.

9. Mendefinisikan setiap *item* yang didapat dari langkah sebelumnya menjadi *fuzzy set* atau bisa juga disebut sebagai *item fuzzy set* terhadap transaksi QT.
10. Mencari *candidate rules* dengan menghitung *confidence* per kombinasi k-*item* yang mencapai *min_support* ke-k (k ≥ 2) dari *item fuzzy set* yang sudah didapat pada langkah 9 (sembilan).