

BAB III

METODOLOGI PENELITIAN

3.1 Metode Pengumpulan Data

Jenis data yang dikumpulkan dalam tugas akhir ini adalah data primer dan sekunder. Data primer adalah data yang dikumpulkan langsung dari objek yang diteliti yaitu data penjualan produk. Sedangkan data sekunder adalah data yang dikumpulkan melalui referensi pembacaan buku-buku yang berkaitan erat dengan kajian masalah yang diteliti, misalnya pengumpulan bahan dari skripsi, dari media internet dan jurnal serta referensi buku. Metode yang digunakan dalam proses pengumpulan data adalah sebagai berikut:

1. Observasi

Peneliti melakukan pengamatan secara langsung kelapangan untuk mempelajari permasalahan yang ada dilapangan yang erat kaitannya dengan objek yang diteliti. Pengamatan ini dilakukan peneliti untuk menambah pengetahuan mengenai topik yang diangkat penulis.

2. Wawancara

Metode yang dilakukan dengan cara melakukan diskusi serta tanya jawab dengan sumber yang dianggap memiliki pengetahuan lebih dalam, atas permasalahan yang dijadikan objek dalam penelitian. Dalam penelitian ini wawancara dilakukan dengan dengan Pimpinan PT. *Fast Food* Indonesia.

3. Studi Pustaka

Pengumpulan data dilakukan dengan mempelajari dan membaca untuk mendapatkan landasan teori atau pemikiran serta data-data empiris yang bersumber dari buku, internet, maupun penelitian penelitian sebelumnya.

3.2 Alat dan Bahan Analisis Kebutuhan Sistem

Dalam penerapan *Algoritma K-Means* membutuhkan alat dan bahan untuk menganalisa, dimana alat dan bahan ini sangat diperlukan sebagai analisis

kebutuhan sistem. Adapun alat dan bahan yang dibutuhkan dalam penelitian ini adalah sebagai berikut:

3.2.1 Alat

Alat yang akan digunakan sebagai implementasi sebagai kebutuhan non fungsional yang merupakan apa saja yang dimiliki oleh sistem agar dapat berjalan, alat yang digunakan sebagai berikut:

1. Perangkat Lunak (*Software*)

Perangkat lunak yang dibutuhkan dalam analisis *data mining* menggunakan metode *K-Means Clustering* adalah sebagai berikut:

a. Sistem Operasi

Dibutuhkannya sistem operasi berupa *windows 10* dikarenakan *support* dalam pemakaian aplikasi *Rapid Miner 9.8* pada *windows 64 bit* dan *windows 10* ini juga termasuk meminimalisir kendala yang tidak diinginkan.

b. Data Set

Microsoft excel digunakan sebagai alat perhitungan manual antar *cluster* yang dimulai dari titik awal *centroid* sampai menjadi iterasi yang menetap dan tidak berubah ubah lagi.

c. Analisis Data

Rapid Miner adalah aplikasi yang dimana dalam penerapan metode klusterisasi dapat membentuk hasil visualisasi juga perhitungan yang instan sehingga hasil manual pada *microsoft excel* dan hasil yang diterapkan dalam *rapid miner* menggunakan metode algoritma *K-Means* harus relevan, sama dan sebanding.

2. Perangkat Keras (*Hardware*)

Perangkat keras yang digunakan dalam menganalisa data adalah sebagai berikut:

a. Processor Core-i3

Dibutuhkannya *Processor Core-i3* agar penggunaan dalam pemakaian aplikasi implementasi dapat berjalan dengan lancar karena jenis processor ini mampu bekerja secara multi proses pada bagian kecil.

b. Hardisk 320 GB

Hardisk 320 GB memiliki kapasitas yang cukup besar sehingga mampu dijadikan sebagai alat acuan dalam pemakaian aplikasi yang digunakan untuk analisis data tersebut.

c. RAM 4 GB

RAM 4 GB sudah cukup untuk memenuhi standar kriteria yang digunakan dalam menyimpan data set yang cukup banyak untuk penerapan dan pemakaian aplikasi analisis data dan segala macam aplikasi perhitungan seperti *microsoft excel*.

d. Keyboard dan Mouse Standar

Keyboard dan mouse standar sudah cukup mampu sebagai alat yang digunakan untuk membantu dalam penelitian ini.

3.2.2 Bahan

Bahan yang digunakan dalam penerapan metode *clustering K-Means* yaitu dimuat dalam aplikasi *microsoft excel*, bahan yang dianalisis yaitu *data set*. Yang mana *data set* merupakan sekumpulan data tabel dengan berupa relasi antar tabel yang digunakan untuk mempresentasikan data beserta relasinya seperti *data set* yang digunakan pada penerapan *Algoritma K-means Clustering* yaitu mendeskripsikan data penjualan dimana terdapat *field* beserta keterangan yang akan dianalisis.

Tabel 3.1 Data Set

Deskripsi	Field	Keterangan
Data Penjualan	PLU	Kode Produk
	Nama Produk	Nama Produk
	<i>Eat In</i>	Jumlah Penjualan Produk yang makan di tempat
	<i>Take Away</i>	Jumlah Penjualan Produk yang dibawa keluar
	Total	Jumlah Keseluruhan Penjualan Produk

3.3 Tahapan Penelitian

Metode data mining yang digunakan dalam eksperimen ini menggunakan model *Cross-Standard Industry for Data Mining (CRISP-DM)* yang terdiri dari 6 fase (Sumathi, 2006) adalah sebagai berikut:

1. Fase Pemahaman Bisnis (*Bussiness Understanding*)

Pada tahap ini berfokus pada pemahaman mengenai tujuan dari proyek dan kebutuhan secara persepektif bisnis, kemudian mengubah hal tersebut menjadi sebuah permasalahan *data mining* dan rencana awal untuk mencapai tujuan tersebut. Kegiatan yang dilakukan antara lain: menentukan tujuan dan persyaratan dengan jelas secara keseluruhan, menerjemahkan tujuan tersebut serta menentukan pembatasan dalam perumusan masalah *data mining* dan selanjut nya mempersiapkan strategi awal untuk mencapai tujuan tersebut.

2. Fase Pemahaman Data (*Data Understanding*)

Pada tahap ini dilakukan pengumpulan terhadap data, lalu kemudian mempelajari data tersebut dengan tujuan untuk mengenal data, melakukan identifikasi dan mengetahui kualitas dari data, serta mendeteksi subset yang menarik dari data yang dapat dijadikan hipotesa bagi informasi yang tersembunyi.

3. Fase Pengolahan Data (*Data Preparation*)

Pada tahap ini dilakukan persiapan mengenai data yang akan digunakan pada tahap berikutnya. Kegiatan yang dilakukan antara lain, memilih kasus dan parameter yang akan dianalisis (*select data*), melakukan transformasi terhadap parameter tertentu (*transformation*), dan melakukan pembersihan data agar data siap untuk tahap *modeling (cleaning)*.

Data preprocessing bertujuan untuk mendapatkan data yang bersih dan siap untuk digunakan dalam penelitian. Tahapan yang dikerjakan adalah melakukan pengabaian atribut pada data mentah yang dianggap tidak relevan dengan hasil pengujian dan perubahan terhadap nilai data bahkan tipe data pada atribut *dataset* dengan tujuan untuk mempermudah pemahaman terhadap isi *record* dengan memperhatikan konsistensi data, *missing value*, dan *redundancy* pada data.

4. Fase Pemodelan (*Modelling*)

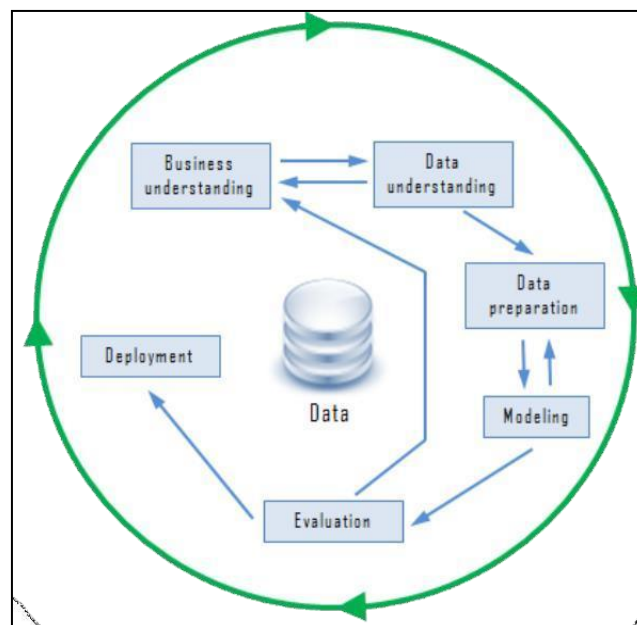
Pada tahap ini dilakukan penentuan terhadap teknik *data mining*, alat bantu *data mining*, dan algoritma *data mining* yang akan diterapkan. Lalu selanjutnya adalah melakukan penerapan teknik dan algoritma data mining tersebut kepada data dengan bantuan alat bantu. Jika diperlukan penyesuaian data terhadap teknik data mining tertentu, dapat kembali ke tahap persiapan data.

5. Fase Evaluasi (*Evaluation*)

Pada tahap ini dilakukan pengujian terhadap model-model yang dikomparasi untuk mendapatkan informasi model yang paling akurat. Evaluasi dan validasi menggunakan metode *confusion matrix* dan kurva ROC.

6. Fase Penyebaran (*Deployment*)

Setelah pembentukan model dan dilakukan analisa dan pengukuran pada tahap sebelumnya, selanjutnya pada tahap ini diterapkan model yang paling akurat dengan memakai data baru diluar *data training* dan *data testing*.



Gambar 3.1 Tahapan Metode CRISP-DM

3.4 Implementasi Metode CRISP-DM

Adapun penerapan metode CRISP-DM pada data penjualan produk PT. *Fast Food* Indonesia cabang KFC Kedaton Lampung adalah sebagai berikut:

1. Pemahaman Bisnis

Seperti yang sudah dijelaskan bahwa metode *data mining* pada aplikasi ini adalah berhubungan dengan data penjualan produk PT. *Fast Food* Indonesia. Faktor-faktor apa saja yang mempengaruhi penjualan produk sesuai data yang sudah ada. Memeriksa keterkaitan antara tiap atribut yang mempengaruhi penjualan produk yang nantinya dapat dijadikan acuan untuk memprediksi hasil penjualan produk selanjutnya sehingga memaksimalkan pemahaman data yang ada. Adapun batasan-batasan dalam metode *data mining* pada aplikasi ini adalah diantaranya sebagai berikut:

- a. Tidak ada faktor luar diluar sumber data yang ikut menjadi atribut yang mempengaruhi proses implementasi *data mining* ini. Seperti cuaca, musim, angin, dan lain-lain.
- b. Jumlah perusahaan yang sama cukup banyak dan memiliki omset penjualan yang berbeda, sehingga akan mempengaruhi penjualan pada PT. *Fast Food* Indonesia. Dengan batasan ini diharapkan dapat memaksimalkan pemrosesan data yang ada sehingga bisa mencapai mengenai potensi penjualan yang lebih tinggi.

2. Pemahaman Data

Pada tahap ini pengumpulan data terlebih dahulu dilakukan, semua data yang diperlukan nanti akan dilakukan pengolahan secara menyeluruh. Pengumpulan data berupa dengan mengambil data yang ada di PT *Fast Food* Indonesia.

Data akan diintegrasikan secara langsung, tanpa harus adanya proses *data warehouse* seperti pada tahapan data mining dan tanpa adanya migrasi data terlebih dahulu. Sistem ini tidak melakukan pemrosesan pemeriksaan data yang diinput pada aplikasi *Rapid Miner*. Format hasil proses *input* sudah sesuai dengan prosedur pemrosesan data yang akan dilakukan pada sistem ini. Analisis data yang akan dilakukan mengenai penyelidikan data penjualan beserta faktor-faktor yang mempengaruhi akan hasil penjualan. Bagaimana faktor-faktor tadi dapat dijadikan atribut acuan dalam mendapatkan informasi atau pengetahuan sesuai dengan pada tahap pemahaman bisnis sebelumnya.

Kemudian dilakukan pengevaluasian dari kualitas data yang ada, data yang banyak dan tidak beraturan dan di dalamnya bisa terjadi adanya kesalahan sehingga perlu adanya pemeriksaan dan memilah bagian atribut yang akan digunakan dalam pencarian pengetahuan, sehingga dapat dihasilkan pelaporan yang nantinya akan benar-benar digunakan.

3. Pengolahan Data

Pada fase ini akan dilakukan proses pemilihan dan pengolahan data yang nanti akan diperlukan dalam tahap permodelan sehingga nanti permodelan yang dilakukan dapat memberikan hasil maksimal sesuai dengan target yang diinginkan. Dengan menggunakan *tools Rapid Miner* untuk melakukan proses komparasi data, bagaimana data dapat diolah, dan dianalisa sehingga dapat ditentukan suatu permodelan yang sesuai dengan data sehingga dapat menghasilkan suatu hasil yang diinginkan mengenai potensi produksi yang sesuai dengan kondisi data yang ada.

Adapun data yang akan dipilah merupakan data laporan penjualan periode 1-31 Desember 2020. Berikut adalah data mentah berupa laporan penjualan PT. *Fast Food* Indonesia cabang KFC Kedaton adalah sebagai berikut:

No	Nama Produk	Eat In	Take Away	Total	Tinggi	Sedang	Rendah	Jarak	Cluster	Data
1	Ipc Chicken OR	4	35	39	50,97057975	310,615518	38,26225294	38,26225294	Rendah	8
2	7 Up Medium	3	0	3	1,414213562	360,2110492	12,56980509	1,414213562	Tinggi	34
3	K SStar HC	27	0	27	35,35533906	339,4937407	26,19160171	26,19160171	Rendah	67
4	K SStar OR	1	0	1	1,414213562	362,0276233	14,76482306	1,414213562	Tinggi	
5	K SStar Mix 1OR+1HC	2	0	2	0	361,1177093	13,6381817	0	Tinggi	
6	K SStar Triple HC	2	0	2	0	361,1177093	13,6381817	0	Tinggi	
7	Chick Cheez Burger	15	14	29	33,07567082	329,9818177	19,6468827	19,6468827	Rendah	
8	Perkedel	3	11	14	16,30950643	344,8912872	5,099019514	5,099019514	Rendah	
9	Mirinda Medium	1	0	1	1,414213562	362,0276233	14,76482306	1,414213562	Tinggi	
10	Pepsi Medium	21	12	33	38,28837944	327,6217331	25,49599757	25,49599757	Rendah	
11	Charge Take Away	0	357	357	503,4659869	169,4148155	490,7871229	161,412515	Sedang	LITERASI 1
12	English Breakfast	0	1	1	2,449489743	361,5494434	14,69693846	2,449489743	Tinggi	
13	Mocca Float	150	38	188	240,7155998	230,2911201	228,7400271	238,7400271	Rendah	
14	Ovaline	6	2	8	7,483314774	354,7196076	7,071067812	7,071067812	Rendah	
15	Hot Brulle Late	0	4	4	4,898979486	357,3765521	11,22497216	4,898979486	Tinggi	
16	Iced Tea Tarik Brulle	1	0	1	1,414213562	362,0276233	14,76482306	1,414213562	Tinggi	
17	Mango Float	77	22	99	124,5712647	273,1153602	112,5255527	112,5255527	Rendah	
18	Espresso Double Shoot	1	0	1	1,414213562	362,0276233	14,76482306	1,414213562	Tinggi	

Gambar 3.2 Laporan Penjualan Produk KFC Kedaton Lampung

4. Permodelan

Pemilihan Data merupakan dalam fase Pengolahan data dimana menyesuaikan dengan proses mining yang akan dilakukan. Proses pengklasifikasian yang dilakukan yakni melihat tinggi rendahnya hasil penjualan setiap produk yang ada.

5. Evaluasi

Dilakukan 2 proses evaluasi, yakni pengujian dan perbandingan. Dengan pengujian diambil 30 data secara acak dinilai secara manual dan dibandingkan dengan hasil permodelan yang dilakukan. Kemudian dengan perbandingan dengan menggunakan Fungsi kondisional kemudian dibandingkan dengan permodelan yang sudah dimodelkan tadi. Hasil dari evaluasi ini, permodelan yang dilakukan telah sesuai dengan tujuan pengimplementasian *data mining* pada sistem ini.

6. Penyebaran

Setelah tahap evaluasi dimana menilai secara detail hasil dari permodelan dilakukan pengimplementasian dari keseluruhan model yang telah dirancang. Selain itu juga dilakukan penyesuaian dari model dengan sistem yang akan dibangun sehingga dapat menghasilkan suatu hasil yang sesuai dengan target diawal tahap CRISP-DM ini.

3.5 Analisis Data

Pada tahap ini merupakan tahap untuk memastikan data penjualan produk yang dipilih telah layak untuk dilakukan proses pengolahan. Data yang berjenis numerikal seperti jumlah produk yang terjual harus dilakukan proses inialisasi data terlebih dahulu ke dalam bentuk nominal berikut adalah data penjualan yang digunakan untuk menganalisa penjualan menggunakan metode *K-Means Clustering*. Adapun tahapan-tahapan yang dilakukan untuk menganalisa data penjualan produk menggunakan metode *K-Means Clustering* sebagai berikut:

1. Menentukan Jumlah *Clustering*

Pada tahapan ini penulis menentukan jumlah *clustering* sebanyak 3 *clustering* yaitu *clustering* tinggi, *clustering* sedang dan *clustering* rendah.

2. Menentukan Titik Awal *Cluster/Centroid*

Pada tahapan ini penulis mengambil titik awal *cluster (centroid)* yaitu data ke-6, data ke-32 dan data ke-65 pada data penjualan produk.

3. Melakukan *Iterasi/Jarak Data Dengan Titik Awal Pusat Cluster*

Pada tahapan ini penulis melakukan iterasi data penjualan untuk menghitung jarak semua data penjualan dengan titik awal pusat *cluster* logika K-Means *Clustering* sebagai berikut:

$$D(i,j) = (x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2 \dots \text{dst.}$$

Keterangan:

$D(i,j)$ = Jarak antara data dengan titik pusat awal *cluster* tinggi

I = Data Penjualan

J = Titik pusat awal (*centroid*) *cluster* tinggi

X_{ki} = Data ke i pada atribut data ke k

X_{kj} = Titik pusat ke j pada atribut ke k

4. Menentukan Titik Awal Pusat *Cluster* Berikutnya

Pada tahapan ini dilakukan pencarian titik awal pusat *cluster* yang ke-2 menggunakan rumus sebagai berikut:

$$\text{Centroid} = \text{Jumlah Data} / \text{Jumlah Cluster}$$

setelah mendapat titik awal pusat *cluster* yang ke-2 proses berikutnya melakukan iterasi berikutnya sampai semua data tidak mengalami perpindahan *clustering*.