

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Objek penelitian ini berfokus pada melakukan analisis asosiasi antar produk dengan menggunakan algoritma Apriori, dan dari hasil analisis asosiasi tersebut kemudian akan diprediksi banyaknya penjualan produk Chewee dengan menggunakan algoritma XGBoost untuk meningkatkan efektivitas dalam menghasilkan prediksi yang paling akurat dan andal. Penelitian ini dilakukan, untuk melihat pola pembelian konsumen terhadap produk-produk Chewee yang sering mereka beli secara bersamaan, serta untuk memprediksikan banyak penjualan produk-produk tersebut untuk manajemen stok produk Chewee. Adapun dataset yang digunakan berasal dari data penjualan Chewee Dengan rentang waktu dari 1 Januari 2023 sampai 31 Desember 2024, dengan detail kolom yaitu ID Orderan, Tanggal Order, Nama Produk, Harga Produk per-satuan, *Quantity*, Total Harga, *Store*, Kategori, dan Merk. Namun hanya beberapa data saja yang digunakan dalam penelitian ini yaitu, ID Produk dan Merk.

1	ID Orderan	Tanggal Order	Nama Produk	Harga Produk	Quantity	Total Harga	Store	Kategori	Merk
2	#41939	12/31/2024	Prama Snack 70 gr Dog Snack	23,000.00	15	345,000.00	Denzel Store	Dog Food	Prama Snack
3	#41938	12/31/2024	BAVARO DOG FOOD ALL LIFE STAGE SUPER PREMIUM 18KG	535,000.00	1	535,000.00	Hanabi Petshop	Dog Food	Bavaro Dog
4	#41937	12/30/2024	Nature Bridge Gastrointestinal Vet Cat Food Freshpack 1kg	148,500.00	1	148,500.00	Denzel Store	Cat Food	Nature Bridge
5	#41937	12/30/2024	CHUMMY PET Beef/Lamb/Fish Dog Food Kaleng 375 Gram	14,000.00	1	14,000.00	Jdpeach Store	Dog Food	Chummy Pet
6	#41937	12/30/2024	Nature Bridge Gastrointestinal Vet Cat Food Freshpack 1kg	148,500.00	1	148,500.00	Denzel Store	Cat Food	Nature Bridge
7	#41936	12/30/2024	Shampoo Animal n Co Anjing & Kucing 1 Liter	67,000.00	6	402,000.00	Denzel Store	Cat & Dog Shampoo	Shampoo Animal
8	#41935	12/30/2024	Whiskas Pouch 80 gr 1 dus	120,000.00	2	240,000.00	Hanabi Petshop	Cat Food	Whiskas
9	#41935	12/30/2024	Whiskas Makanan Kucing Kering 1.2 kg Adult	57,000.00	2	114,000.00	Hanabi Petshop	Cat Food	Whiskas
10	#41932	12/30/2024	Royal Canin Adult Pomeranian 1,5 kg	187,000.00	1	187,000.00	Pom Pom Petshop	Cat Food	Royal Canin
11	#41931	12/29/2024	PASIR KUCING PAWSITIVE 25 LITER / 20 KG	115,000.00	1	115,000.00	Olga Cat Mansion	Cat Litter Sand	Sand Pawsitive
12	#41855	12/29/2024	BAVARO DOG FOOD ALL LIFE STAGE SUPER PREMIUM 18KG	535,000.00	1	535,000.00	Hanabi Petshop	Dog Food	Bavaro Dog
13	#41853	12/27/2024	Natures Protection Adult Veal 400 GR	51,300.00	10	513,000.00	Hanabi Petshop	Dog Food	Natures Protection
14	#41853	12/27/2024	Moochie Chicken Mousse With Goat Milk Dog 70gr Makanan Anjing Basah Import - Goatmilk	12,000.00	24	288,000.00	Jdpeach Store	Dog Food	Moochie
15	#41852	12/27/2024	Royal Canin Vet Gastrointestinal Dog Canned 400gr	61,000.00	3	183,000.00	Jdpeach Store	Dog Food	Royal Canin
16	#41850	12/27/2024	Happy Cat Profile Adult	612,500.00	1	612,500.00	VJ Pet House	Cat Food	Happy Cat

Gambar 3.1 Data Penjualan Chewee

3.2 Metode Penelitian

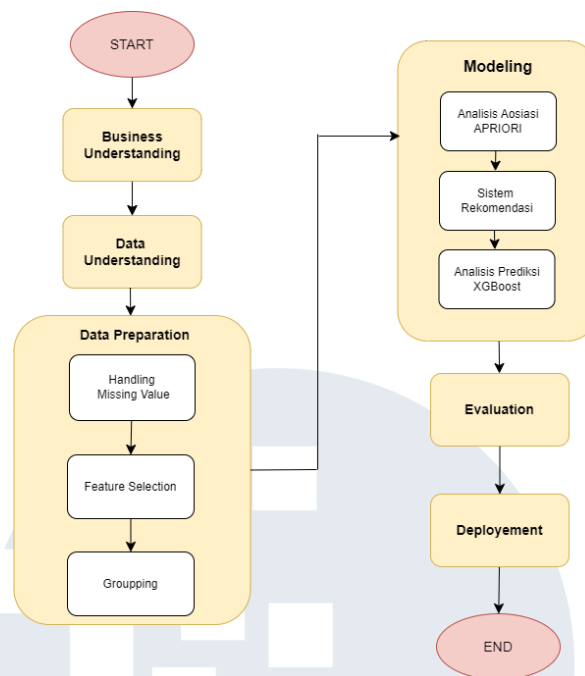
3.2.1 Alur Penelitian

Penelitian ini menggunakan pendekatan *Hybrid Model* yang menggabungkan algoritma Apriori dan XGBoost untuk menganalisis asosiasi antar produk Chewee dan memprediksi penjualannya. Alur penelitian dimulai dengan tahap pengumpulan informasi, di mana informasi penjualan dan informasi produk Chewee dikumpulkan dari sumber yang relevan. Setelah informasi terkumpul, tahap pre-processing dilakukan untuk membersihkan informasi, menangani nilai yang hilang, menghapus informasi duplikat, dan melakukan normalisasi pada fitur-fitur yang digunakan.

Selanjutnya, dilakukan analisis asosiasi menggunakan algoritma Apriori untuk mengidentifikasi pola-pola pembelian produk yang sering muncul dalam transaksi konsumen. Hasil dari analisis asosiasi ini akan digunakan untuk memahami hubungan antar produk yang dibeli bersama.

Pada tahap berikutnya, prediksi penjualan dilakukan menggunakan algoritma XGBoost untuk membangun demonstrasi prediktif berdasarkan pola yang ditemukan pada tahap analisis asosiasi serta faktor-faktor lain yang mempengaruhi penjualan. Evaluasi *show* dilakukan untuk mengukur akurasi prediksi menggunakan metrik seperti Akurasi, Accuracy, dan Review, serta melakukan tuning *demonstrate* untuk mendapatkan hasil yang ideal.

Tahap terakhir adalah analisis hasil yang mencakup interpretasi dari asosiasi yang ditemukan serta prediksi penjualan yang telah dibuat, yang kemudian digunakan untuk memberikan rekomendasi kepada pihak manajemen produk Chewee dalam meningkatkan strategi pemasaran dan penjualan.



Gambar 3.1 Proses Alur Penelitian

3.2.2 Metode Penelitian

Terdapat berbagai macam metode penelitian untuk mendukung proses penelitian, khususnya dalam konteks *data mining*. Beberapa metode penelitian yang biasanya digunakan dalam proses *data mining* yaitu, KDD, CRISP-DM, dan SEMMA. Berikut perbedaan dari ketiga model tersebut :

Tabel 3.1 Perbandingan Metode Penelitian

Aspek	KDD	CRISP-DM	SEMMA
Ruang Lingkup	Ilmiah; pipeline discovery data saja	Industri; integrasi bisnis hingga deployment	Teknis; model-centric dalam SAS Enterprise Miner
Iterasi	Iterative, namun garis bedar linear	Iterative dan feksibel antar fase	Seq-to-seq, tapi bisa diulang; kurang fleksibel dibanding CRISP-DM
Business Context	Tidak ada	Fase khusus Business Understanding	Implicit saja; tidak ada fase bisnis eksplisit
Tool-agnostic	Ada	Ada	Tidak—terintegrasi SAS Enterprise Miner
Fokus utama	Discovery pola & evaluasi ilmiah	Keseimbangan bisnis-teknis ; penuh end-to-end	Pengembangan dan evaluasi model statistik/ML

Dari ketiga perbandingan tersebut, metode penelitian yang diterapkan pada penelitian ini pun menggunakan *Cross-Industry Standard Process for Data Mining* (CRISP-DM). CRISP-DM merupakan kerangka kerja yang banyak digunakan untuk proses *data mining*, pendekatan terstruktur untuk perencanaan, pelaksanaan, sampai evaluasi proyek *data mining*. Berikut merupakan alur penelitian yang diterapkan dengan menggunakan metode CRISP-DM.

1. Business Understanding

Tahap pertama yang perlu dilakukan adalah *business understanding*. *Business understanding* merupakan tahapan untuk memahami tujuan bisnis/perusahaan, masalah yang perlu diselesaikan, dan bagaimana bentuk pembelajaran model dapat membantu perusahaan tersebut. Berdasarkan permasalahan penelitian ini, melihat Chewee yang sedang melakukan transformasi bisnis sehingga perlu dibantu dalam proses strategi pemasaran baru salah satunya dengan menerapkan teknik *bundling* tersebut melalui analisis pola kombinasi antara merek produk di Chewee

2. Data Understanding

Data understanding merupakan tahap dalam melakukan sesi pemahaman dan melakukan pemilihan data/seleksi data, yang diperlukan dalam proses data mining. Data tersebut akan diambil dari website Chewee khusus admin, dan hanya berfokus pada data penjualan produk Chewee 2024. Data penjualan Chewee mencakup data ID order, tanggal order, nama produk, harga produk per-satuan, *quantity*, total harga, kategori, dan merek dengan detail sebagai berikut

Tabel 3.2 Detail Deskripsi Data Penjualan Chewee

Data	Deskripsi
ID Orderan	Berisi kode unik setiap pesanan, untuk mengidentifikasi setiap transaksi.
Tanggal Order	Menunjukkan tanggal pembelian produk.
Nama Produk	Merupakan nama lengkap dari produk yang dibeli konsumen.

Harga Produk	Menunjukkan harga satuan per produk dalam mata uang Rp.
<i>Quantity</i>	Menunjukkan banyaknya jumlah unik produk yang dibeli oleh konsumen.
Total Harga	Total biaya produk dari harga per satuan produk dan <i>quantity</i> , yang dibeli oleh konsumen.
Kategori	Pengelompokkan kategori produk menjadi kategori spesifik seperti <i>dog food</i> , <i>cat food</i> , dll.
Merk	Menunjukkan merek atau nama brand dari produk yang dibeli konsumen.

3. Data Preparation

Tahap ketiga dari CRISP-DM yaitu *data preparation*, yang merupakan tahapan untuk mempersiapkan data seperti membersihkan data dari informasi yang acak, memiliki kesalahan, nilai yang tidak tersedia dalam dataset, mentransformasikan atau merubah data menjadi bentuk data yang sesuai dengan format pengolahan sampai pada data yang tidak relevan. Setelah tahap *preparation*, akan bersamaan juga melakukan pembersihan data dengan menghilangkan data yang memiliki informasi tidak relevan/nilai yang tidak tersedia. Hal tersebut dilakukan untuk meningkatkan kualitas data, supaya lebih siap untuk digunakan dalam analisis maupun model pembelajaran mesin. Untuk mempersiapkan data akan menggunakan Python dengan menerapkan :

- *isnull* = untuk mengetahui jumlah *missing value* pada data.
- *dropna* = menghapus kolom yang tidak akan digunakan dalam analisis ini.
- *groupby* = melakukan pengelompokkan data berdasarkan ID Orderan.

4. Modeling

Modeling merupakan tahapan untuk menerapkan algoritma maupun metode yang menjadi fokus pada penelitian ini, untuk mengidentifikasi pola data, pengetahuan, maupun informasi penting dari data tersebut. Pada penelitian ini menerapkan 2 algoritma untuk melakukan analisis asosiasi dan prediksi. Untuk analisis asosiasi antara produk menggunakan algoritma Apriori, dan analisis prediksi banyaknya pembelian produk Chewee menggunakan algoritma XGBoost. Serta menyertakan juga sistem rekomendasi, untuk melihat rekomendasi kombinasi merek produk tersebut.

5. Evaluation

Tahap kelima pada model CRISP-DM adalah *evaluation*. *Evaluation* merupakan tahapan untuk mengetahui dan mengidentifikasi hasil dari *data mining* sebelumnya, apakah telah menjawab tujuan penelitian yang telah ditetapkan di awal. Evaluasi tersebut dilakukan dengan melihat hasil analisis apriori dan analisis prediksi, yang kemudian akan disimpulkan algoritma mana yang terbaik dalam melakukan analisis asosiasi antar produk Cheweez dan analisis prediksi banyaknya penjualan produk Chewee. Pada penelitian ini akan menggunakan teknik evaluasi *cross-validation*.

6. Deployment

Tahap terakhir setelah *evaluation* adalah *deployment*. *Deployment* merupakan tahapan melakukan implementasi model untuk menjadi *insight* kepada perusahaan. Pada penelitian ini, *deploy* yang dibangun membuat suatu website yang berisi hasil dari analisis asosiasi, sistem rekomendasi, dan analisis prediksi dari kombinasi antara merek produk Chewee.

3.3 Teknik Pengumpulan Data

Data yang digunakan dalam penelitian ini menggunakan data penjualan produk di Chewee, yang diambil melalui administrator website Chewee tersebut dengan rentang waktu 1 Januari hingga 31 Desember 2024. Data yang diambil berisikan 8 atribut yang dipilih seperti ID Orderan, Tanggal Order, Nama Produk, Harga Produk per Satuan, *Quantity*, Total Harga, dan Kategori, Merek.

Rentang waktu tersebut dipilih karena, dapat memengaruhi perubahan tren, pola, atau dinamika yang terkait dengan topik penelitian dan memberikan gambaran yang lebih komprehensif tentang peristiwa atau interaksi yang terjadi dalam rentang waktu tertentu. Memilih rentang waktu tertentu juga memungkinkan peneliti untuk mendapatkan data yang relevan dan terbaru terkait dengan topik penelitian

3.4 Variabel Penelitian

3.4.1 Variabel Independen

Variabel independen merupakan variabel yang mempengaruhi atau menjadi penyebab perubahan pada variabel lain. Dalam bahasa Indonesia, sering disebut sebagai variabel bebas. Variabel ini berfungsi sebagai stimulus atau prediktor dalam penelitian. Pada penelitian ini, variabel independen yang digunakan khususnya pada tahap prediksi berfokus pada nilai *antecedent support*, *consequent support*, *confidence*, *lift*.

3.4.2 Variabel Dependen

Variabel dependen merupakan variabel yang dipengaruhi atau menjadi akibat dari variabel independen. Dalam bahasa Indonesia, sering disebut sebagai variabel terikat. Variabel ini merupakan hasil atau output yang diukur dalam penelitian untuk melihat dampak dari variabel independen. Pada penelitian ini variabel dependen yang digunakan khususnya pada tahap prediksi berfokus pada nilai *support*.