

BAB II

TINJAUAN PUSTAKA

2.1 Pencapaian Terdahulu

Penelitian sebelumnya digunakan peneliti sebagai referensi dalam melaksanakan penelitian ini. Metode *Apriori* banyak digunakan pada penelitian terdahulu dalam menentukan pola hubungan antar atribut dalam dataset. Dimana *support* atau penunjang dan *confidence* atau kepastian menghasilkan sebuah *rule* atau aturan yang dicek validitas menggunakan *lift ratio*. Kurniawan, A., et al. (2021). Penelitian yang dipublikasikan di *eProceedings of Engineering* (8(5)) tentang penerapan Data Mining menggunakan algoritma Apriori di Restoran Pagi Sore menunjukkan bahwa aplikasi ini terbukti sangat bermanfaat bagi perusahaan, berfungsi sebagai sarana yang efektif dalam menemukan produk penjualan yang paling laris.

Selain itu, hasil dari penelitian Agustiani, N., et al. (2020). Implementasi Data Mining dengan metode Apriori dalam analisis penjualan di Alfamart. Jurnal SENARIS, 2. Penelitian ini menghasilkan analisis dan perhitungan data penjualan dengan rentang waktu 1 hingga 21 Januari 2020, dengan total 21 transaksi yang menghasilkan 142 aturan asosiasi. Produk yang paling diminati oleh konsumen memiliki minimum confidence di atas 50%.

Penelitian lainnya yaitu oleh W. (2022) yang dipublikasikan dalam *INSOLOGI: Jurnal Sains dan Teknologi* (1(4)) menyimpulkan bahwa konsumen sering membeli Produk A dan Produk B dalam transaksi yang sama.

Berdasarkan beberapa kajian literatur penelitian sebelumnya yang telah diuraikan, penulis memilih menggunakan data mining dengan metode *apriori* karena metode yang efektif dan sederhana dengan mengolah data transaksi support dan confidence dalam menentukan rule atau aturan yang valid.

2.2 Tinjauan Teoritis

Dalam tinjauan teoritis, peneliti akan memaparkan penjelasan dan menguraikan beberapa hal yang terkait dalam penelitian ini yang bertujuan untuk memberikan persepsi yang sama, yaitu :

2.2.1 Data

Data adalah sekumpulan objek yang memiliki atribut-atribut tertentu. Atribut ini menggambarkan sifat atau karakteristik dari objek yang bersangkutan dan sering disebut sebagai fitur, variabel, field, atau karakteristik.

Data keranjang belanja (*Market basket data*) atau sering disebut sebagai data transaksi ini adalah setiap record dari set data yang mengandung beberapa item. Setiap transaksi dapat memiliki jumlah item yang berbeda dalam penyebutan data keranjang belanja dikarenakan setiap isi keranjang belanja terdapat jenis dan jumlah barang yang berbeda antar pembeli. Contoh data transaksu dapat dilihat pada Tabel 2.1. TID merupakan transaksi ID dan item set adalah nama item yang dibeli.

Tabel 2. 1 *Contoh Data Transaksi*

TID	ITEM SET
1	Bread, Milk, Diaper
2	Beer, Bread
3	Beer, Milk, Diaper
4	Coke, Bread, Milk
5	Coke, Bread, Diaper

Kualitas Data sangatlah diperlukan, apabila kualitas data tidak bagus maka hasil yang diperoleh akan tidak optimal. kualitas data dapat ditinjau dari aspek berikut:

1. Kesalahan Pengukuran (*Measurement Error*)

Kesalahan dalam pengukuran dapat timbul karena adanya perbedaan antara nilai yang tercatat dan nilai yang benar, yang dihasilkan dari proses pengukuran. berikut dipaparkan beberapa jenis kesalahan pengukuran.

- a. *Noise*, merupakan perubahan nilai asli dari sebuah data. Misalnya adalah distorsi pada video seseorang ketika berkomunikasi melalui *video call* disebabkan jaringan yang buruk.
- b. Bias merujuk pada perbedaan pengukuran antara nilai yang dihitung dengan selisih antara mean dan nilai yang telah diketahui.
- c. *Precision* merupakan tingkat keseragaman hasil ukur berulang (untuk kuantitas yang sama), yang dapat diukur dengan standar deviasi.
- d. *Accuracy* yaitu tingkat kedekatan hasil pengukuran yang mencerminkan nilai yang sebenarnya dari kuantitas yang diukur.

2. Kesalahan Pengumpulan (*Collection Error*)

Pengumpulan data yang salah dapat terjadi karena hilangnya data objek atau nilai atribut, atau karena lingkup objek data yang tidak tepat. Berikut adalah rincian data yang termasuk dalam kesalahan pengumpulan.

- a. Data outlier merujuk pada data objek yang menunjukkan perbedaan yang signifikan dari mayoritas objek data lainnya dalam suatu kumpulan data.
- b. *Missing values*, atau nilai kosong, bisa terjadi karena berbagai penyebab
 - 1) Informasi sulit diperoleh (misal informasi usia dan berat badan seseorang, akan terjadi penolakan untuk memberikan informasi tersebut).
 - 2) Atribut yang mungkin tidak relevan untuk semua kasus (seperti atribut pendapatan tahunan yang tidak berlaku pada anak-anak). Untuk mengatasi masalah ini, tahap-tahap yang dapat dilakukan adalah:
 - 1) Hilangkan objek atau atribut data;
 - 2) Perkirakan nilai yang kosong;
 - 3) Abaikan nilai yang kosong selama analisis.
- c. Data Duplikat, merupakan kumpulan data dapat mencakup objek data yang sama. Objek yang sama berupa data duplikat, biasanya terjadi saat menggabungkan data dari sumber yang bervariasi. Maka yang harus dilakukan ialah pembersihan data (*data cleaning*) terlebih dahulu.

2.2.2 *Data Mining*

Suntoro (2019:2) mengemukakan bahwa “Data mining adalah proses untuk mendapatkan informasi yang berguna dari basis data yang besar dan perlu diekstraksi agar menjadi informasi baru dan dapat membantu dalam pengambilan keputusan”.

Berdasarkan Werdiningsih (2020:17), “Data mining merupakan bidang ilmu yang digunakan untuk menangani masalah pengambilan informasi dari database yang besar dengan menggabungkan teknik dari statistic, pembelajaran mesin, visualisasi data, pengenalan pola, dan database”.

Data mining, menurut Han dan Kamber (2006), merupakan tahapan yang diterapkan untuk mengekstrak pengetahuan dari data besar. Proses ini menganalisis pola-pola atau hubungan yang ada pada data yang berjumlah besar untuk mengidentifikasi informasi yang bernilai. Dalam data mining, terdapat hubungan dengan berbagai disiplin ilmu lain, seperti *Statistic Database Warehousing*, *Komputasi Tingkat Tinggi*, *Machine Learning*, *Information Retrieval*, dan *Data System*. Selain itu, bidang seperti *Pengenalan Pola*, *Signal Processing*, *Spatial Data Analysis*, *Neural Network*, *Image Database*. turut memberikan dukungan.

Berbagai *survey* mengenai proses pemodelan dan metodologi mengemukakan bahwa, “*Data mining* digunakan sebagai penunjuk, dimana *data mining* menyajikan intisari atas sejarah, deskripsi dan sebagai standar petunjuk mengenai masa depan dari sebuah proses model data mining Karakteristik *data mining* sebagai berikut:

- a. *Data mining* berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
- b. *Data mining* biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dapat dipercaya.
- c. *Data mining* berguna untuk membuat keputusan kritis.

Berdasarkan beberapa pengertian tersebut dapat ditarik kesimpulan bahwa *Data Mining* adalah suatu teknik menggali informasi berharga yang terpendam atau tersembunyi pada suatu koleksi data (*database*) yang sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui.

Menurut Hoffer, Presscott, dan McFadden (2007), tujuan data mining adalah:

1. Explanatory, untuk menjelaskan beberapa kondisi penelitian, seperti mengapa penjualan truk pick up meningkat di colorado.
2. Confirmatory, untuk mempertegas hipotesis, seperti halnya 2 kali pendapatan keluarga lebih suka di pakai untuk membeli peralatan keluarga, di bandingkan dengan satu kali pendapatan keluarga.

3. Exploratory, menganalisis data untuk hubungan yang baru yang tidak di harapkan, seperti halnya pola apa yang cocok untuk kasus penggelapan kartu kredit.

Menurut Matthew North (2012:5), tahapan dalam proses data mining terbagi dalam beberapa langkah yang disebut dengan CRISP-DM (Cross-Industry Standard Process for Data Mining) yaitu antara lain:

- a. Business Understanding (Pemahaman Bisnis): Tahap pemahaman sistem yang berjalan dan kebutuhan apa yang dibutuhkan dalam menyelesaikan masalah yang timbul didalamnya.
- b. Data Understanding (Pemahaman Data): Tahap pemahaman dan pengumpulan data yang dibutuhkan sebelum dilakukan persiapan untuk analisa.
- c. Data Preparation (Persiapan Data): Tahap persiapan dan seleksi data yang telah dikumpulkan dan diubah menjadi bentuk yang dapat diolah dalam model yang ditentukan selanjutnya.
- d. Modeling (Pemodelan): Proses analisa dan pemodelan data yang telah disiapkan dimana dalam ini dilakukan penerapan atau penghitungan berdasarkan algoritma atau metode yang ditentukan untuk mendapatkan hasil yang diinginkan sesuai dengan kebutuhan pengguna dan melakukan representasi pemecahan masalah.
- e. Evaluation (Evaluasi): Melakukan analisa dan evaluasi dari hasil model yang telah dibuat apakah sudah sesuai standar dan telah memecahkan masalah atau memenuhi kebutuhan pengguna.
- f. Deployment (Penerapan): Tahap penerapan hasil dari model yang telah dievaluasi dan dianalisa untuk kemudian dijadikan bentuk yang dapat diolah kembali.

2.2.3 Algoritma Apriori

Algoritma apriori adalah salah satu algoritma yang melakukan pencarian *frequent itemset* dengan menggunakan teknik *association rule*. Algoritma apriori menggunakan pengetahuan frekuensi atribut yang telah diketahui sebelumnya untuk memproses informasi selanjutnya. Pada algoritma apriori menentukan kandidat yang mungkin muncul dengan cara memperhatikan minimum *support* dan minimum

confidence. *Support* adalah nilai pengunjung atau persentase kombinasi sebuah *item* dalam *database* (Erwin,2009). Rumus *support* adalah sebagai berikut :

$$Support (A) = \frac{\sum Transaksi\ mengandung\ A}{\sum Transaksi} \times 100\%$$

Selanjutnya rumus *support* untuk mencari nilai dari dua item sebagai berikut :

$$Support (A, B) = \frac{\sum Transaksi\ mengandung\ A, B}{\sum Transaksi} \times 100\%$$

Sedangkan *confidence* adalah nilai kepercayaan yaitu kuatnya hubungan antar item dalam sebuah *Apriori*. *Confidence* dapat dicari setelah pola frekuensi munculnya sebuah item ditemukan. Berikut rumus *confidence* :

$$Confidence\ P (A|B) = \frac{\sum Transaksi\ mengandung\ A\ dan\ B}{\sum Transaksi\ A} \times 100\%$$

Selanjutnya untuk pengujian penelitian ini digunakan metode *lift ratio* yakni untuk mengukur seberapa akurat rule yang telah dibentuk berdasarkan nilai *support* dan *confidence*. Hasil perhitungan *lift ratio* mencerminkan validitas transaksi yang menunjukkan seberapa sering produk A dan produk B dibeli secara bersamaan. Berikut rumus *lift ratio* :

$$Lift\ Ratio = \frac{Support\ (A \cap B)}{Support\ A \times Support\ B} \times 100\%$$

Secara garis besar cara kerja algoritma apriori adalah:

1. Pembentukan kandidat *itemset*, kandidat kitemset dibentuk dari kombinasi (k-1)-*itemset* yang didapat dari iterasi sebelumnya. Satu ciri dari algoritma *Apriori* adalah adanya pemangkasan kandidat *k-itemset* yang subset-nya yang berisi k-1 item tidak termasuk dalam pola frekuensi tinggi dengan panjang k-1.
2. Penghitungan *support* dari tiap kandidat *k-itemset*. *Support* dari tiap kandidat *k-itemset* didapat dengan men-scan database untuk menghitung jumlah transaksi yang memuat semua item di dalam kandidat *k-itemset* tsb.

Ini adalah juga ciri dari algoritme apriori dimana diperlukan penghitungan dengan scan seluruh database sebanyak *k-itemset* terpanjang.

3. Tetapkan pola frekuensi tinggi. Pola frekuensi tinggi yang memuat *k-item* atau *k-itemset* ditetapkan dari kandidat *k-itemset* yang supportnya lebih besar dari *minimum support*.
4. Bila tidak didapat pola frekuensi tinggi baru maka seluruh proses dihentikan. Bila tidak, maka *k* ditambah satu dan kembali ke bagian 1.

2.2.4 Statistik

Menurut Sudjana, S. (n.d.). Metode Statistika. Statistik ialah suatu pengetahuan yang berhubungan dengan metode pengumpulan data, pengolahan data, analisis data serta menyimpulkan data berdasarkan kumpulan dan analisis sebuah data.

Statistik ialah sebuah cabang ilmu yang berkaitan dengan pengumpulan, analisis, interpretasi dan penyajian data. statistik dibagi menjadi 2 kategori utama, statistik deskriptif dan statistik inferensial.

- a. Statistik deskriptif yang berfokus pada penggambaran dan ringkasan data.
- b. Statistik inferensial biasa digunakan untuk menarik kesimpulan tentang populasi berdasarkan sampel.

Statistik berperan penting dalam berbagai bidang, diantaranya ialah teknik, ekonomi, kesehatan dan sebagainya. Serta dapat membantu dalam mengambil keputusan berbasis data dan analisis yang lebih baik. Terdapat metode statistik yang sering digunakan yaitu seperti analisis regresi, uji hipotesis serta analisis varians. Metode ini biasa digunakan untuk menganalisis hubungan antar variabel dan menguji asumsi.

2.2.5 PHP

PHP adalah singkatan dari *Hypertext Preprocessor*, yang bahasa *scripting*-nya adalah *server-side* dan dimana pemrosesan kode program berlangsung di server dan hasilnya akan ditampilkan di *browser*. Situs *website* resmi PHP, yaitu <http://www.php.net/>. PHP adalah bahasa pemrograman yang sangat populer dikalangan *developer* untuk membuat *website* dinamis. PHP atau *Hypertext Preprocessor* adalah bahasa skrip sisi *server* yang secara khusus dibuat untuk membangun dan membuat situs web (Yudhanto & Prasetyo, 2019). PHP berjalan

pada dokumen *HTML (Hypertext Markup Language)* untuk menghasilkan konten halaman *web* sesuai dengan permintaan. *PHP* adalah perangkat lunak *open source* dan juga merupakan perangkat lunak lintas platform, sehingga bekerja dengan baik pada sistem operasi *Windows, Mac OS, dan Linux*.

2.2.6 *Laravel*

Laravel adalah sebuah kerangka kerja web open-source berbasis *PHP* yang sangat populer untuk mengembangkan aplikasi web. *Laravel* dirancang untuk memudahkan proses pengembangan dengan menyediakan berbagai alat dan fitur yang beragam. Fitur utama *Laravel*:

1. *Eloquent ORM*:

Implementasi *ActiveRecord* yang canggih untuk bekerja dengan *database*. *Eloquent* memungkinkan pengembang untuk berinteraksi dengan objek *database* dan hubungan dengan sintaks yang ekspresif.

2. *Artisan CLI*:

Antarmuka baris perintah *Laravel* yang mencakup sejumlah perintah yang membantu untuk pengembangan aplikasi. *Artisan* dapat mengotomatiskan tugas-tugas berulang seperti migrasi *database*, *seeding*, dan *scaffolding*.

3. *Blade Templating Engine*:

Mesin *templating* yang kuat dan ringan yang disediakan oleh *Laravel*. *Blade* memungkinkan penggunaan kode *PHP* biasa dalam tampilan dan juga menyediakan pintasan yang nyaman untuk tugas-tugas umum.

4. *Routing*:

Laravel menawarkan cara yang sederhana dan ekspresif untuk mendefinisikan rute. Ini mendukung routing *RESTful*, memungkinkan pembuatan *API RESTful* dengan mudah.

5. *Middleware*:

Middleware menyediakan mekanisme yang nyaman untuk memfilter permintaan *HTTP* yang masuk ke aplikasi. Ini sering digunakan untuk tugas-tugas seperti *otentikasi*.

6. *Otentikasi dan Otorisasi*:

Laravel memudahkan implementasi *otentikasi*. Ini juga menyediakan cara sederhana untuk mengelola logika otorisasi dan mengontrol akses ke berbagai bagian dari aplikasi.

7. *Testing:*

Laravel dibangun dengan pemikiran pengujian. Ini menyediakan dukungan *PHP* Unit secara bawaan dan mencakup metode pembantu yang nyaman untuk menguji aplikasi.

8. Penjadwalan Tugas:

Fitur penjadwalan tugas memungkinkan pengembang untuk menjadwalkan tugas-tugas berulang seperti mengirim email atau pembersihan database menggunakan sintaks yang lancar dan ekspresif.

9. *Queues:*

Layanan antrian *Laravel* menyediakan API terpadu di berbagai backend antrian yang berbeda. Antrian memungkinkan pengembang untuk menunda pemrosesan tugas yang memakan waktu.

