

BAB II

TINJAUAN PUSTAKA

2.1 Data Mining

Data ialah kumpulan informasi penting yang selama ini jarang dipergunakan. Data mining, di sisi lain, dapat didefinisikan sebagai proses penambangan atau penggalian data yang masih tersembunyi dalam jumlah besar, (Baihaqi 2017). Data mining ialah proses menemukan pola ataupun informasi penting dalam database dan melakukan operasi untuk secara otomatis menemukan informasi atau pengetahuan yang berguna dalam data besar. Metode ini terdiri dari kombinasi beberapa teknik analisis data dengan algoritma pengolahan data yang besar, (Andini et al. 2022).

Data mining adalah proses mengumpulkan dan menggunakan data sebelumnya untuk menemukan pola dan hubungan teratur data dalam jumlah besar. Data mining juga sering disebut dengan knowledge discovery atau penemuan pengetahuan dalam database (KDD), (Hutasuhut et al. 1963). Hasil dari data mining ini dapat digunakan untuk memperbaiki pengambilan keputusan di masa yang akan datang. Dalam data mining, penambangan data disimpan secara elektronik dan diproses secara otomatis, atau setidaknya disimpan di komputer. Penambangan data memecahkan masalah dengan menganalisis informasi dalam basis data.

Tahapan data mining dibagi dalam beberapa tahap, sebagai berikut:

a. Pembersihan data mining (data cleansing)

Ini adalah proses menghilangkan noise (data yang rusak), data yang tidak lengkap, atau data yang kosong dan tidak relevan.

b. Integrasi data (data integration)

Integrasi data adalah penggabungan data ke dalam satu basis data

c. Seleksi data (data selection)

Seringkali, sebagian besar data dalam database tidak terpakai, oleh karenanya diperlukan perubahan untuk data yang akan digunakan untuk analisis.

d. Data transformation

Mengkonversi data menjadi format yang cocok untuk proses data mining.

e. Proses mining

Proses penggunaan metode untuk menemukan informasi penting yang tersembunyi dalam data.

f. Persentasi pengetahuan

Terdapat visualisasi dan penjelasan tentang metode yang digunakan untuk mengumpulkan pengetahuan dari pengguna, (Andini et al. 2022), (Hariyadi 2016).

Data mining dipecah jadi beberapa kelompok berdasarkan tugas atau pekerjaan dilakukan:

- a. Deskripsi, deskripsi pola kecenderungan yang menjelaskan pola atau kecenderungan tersebut.
- b. Klasifikasi, variable kategori didefinisikan. Misalnya, klasifikasi pendapatan dibagi menjadi tiga kategori tinggi, sedang, dan kecil.
- c. Prediksi, bahwa nilai dalam prediksi memiliki hasil yang akan datang. Dalam situasi yang tepat beberapa pendekatan yang digunakan untuk penilaian dan klasifikasi juga dapat digunakan untuk peramalan.
- d. Estimasi, hampir sama dengan klasifikasi, kecuali bahwa variable target diorientasikan secara numeric daripada secara kategoris.
- e. Pengklasteran (clustering), pengklasteran adalah proses pengelompokan kumpulan data, pengamatan, atau perhatian untuk membentuk kelas objek yang memiliki karakteristik yang mirip satu sama lain dan berbeda dari record klaster lain.

- f. Asosiasi, Dalam data mining, tugas asosiasi adalah mencari fitur yang terjadi pada waktu yang bersamaan. Market basket analysis juga dikenal sebagai analisis keranjang belanja, adalah implementasi dari asosiasi. Jenis analisis ini dibahas dalam penelitian ini. , (Badrul, Studi, and Informasi 2016), (Prasetya et al. 2022), (Studi et al. 2022).

2.2 Analisis Asosiasi

Analisis asosiasi, juga dikenal sebagai aturan asosiasi mining, ialah metode data mining yang digunakan untuk menentukan aturan asosiasi antara kombinasi item, (Andini et al. 2022), (- and Tamba 2022). Ukuran daya tarik yang digunakan dalam data mining adalah:

- a. *Support*, adalah suatu ukuran yang menyatakan tingkat pengaruh suatu barang atau kelompok barang terhadap keseluruhan transaksi.
- b. *Confidance*, ialah ukuran yang secara kondisional menunjukkan hubungan antar dua objek (berdasarkan kondisi tertentu).

Ilmu tentang dasar analisis asosiasi terdiri dari dua langkah: frequent pattern atau analisis pola frekuensi tinggi dan pembentukan aturan asosiasi, (Kasus et al. 2012).

2.3 Algoritma Apriori

Algoritma apriori ialah algoritma dasar yang dikemukakan tahun 1994 oleh *Agrawal* dan *Srkant* untuk melakukan penentuan *frequent itemset* pada aturan asosiasi *Boolean*, (Purwaningtias, Defiyanti, and Apriori 2022). Dalam data mining, algoritma apriori ialah jenis aturan asosiasi. Algoritma yang melakukan analisis data dengan menggunakan aturan asosiasi untuk menentukan hubungan asosiasi antara item, (Yanto and Khoiriah, n.d.).

Algoritma apriori menggunakan teknik aturan asosiatif untuk menemukan elemen yang paling sering muncul. Untuk mencari hubungan aturan dari sekumpulan data

yang dilakukan, langkah pertama adalah mencari item yang sering muncul atau frequent itemset. Pentingnya sebuah asosiasi dapat diidentifikasi melalui support dan confidence. Support merupakan nilai pendukung sedangkan confidence ialah nilai kepastian, yaitu kekuatan hubungan antar item dalam aturan asosiasi, (Hariyadi 2016). Metode utama dalam algoritma apriori untuk mendapatkan *frequent itemset* yaitu sebagai berikut:

- a. Join (penggabungan), tindakan menggabungkan dua objek sedemikian rupa sehingga tidak kombinasi lebih lanjut yang dapat dibentuk.
- b. Prune (pemangkasan), adalah pruning yang menghasilkan produk yang digabungkan kemudian dipangkas dengan menggunakan nilai support yang telah ditentukan., (Purwaningtias, Defiyanti, and Apriori 2022).

Algoritma apriori ini dapat diaplikasikan sebagai alat bantu pendukung keputusan. Metode ini menerapkan pendekatan iterative yang diketahui dengan pencarian level-wise, di mana k-itemset diterapkan untuk mengeksplorasi atau mendapatkan itemset (k+1). Algoritma apriori dibagi menjadi beberapa langkah yang dikenal sebagai iterasi. Setiap iterasi menciptakan sampel frekuensi tinggi dan panjang yang setara, diawali dengan iterasi pertama, menghasilkan pola frekuensi tinggi dan panjang satu, (Pahlevi and Sugandi 2019). Pada tahap awal algoritma apriori, itemset paling sering muncul (frequent itemset) dilakukan dengan melihat pola frekuensi tinggi untuk menemukan kombinasi yang sesuai nilai minimum support dalam basis data. Rumus yang digunakan untuk menghitung nilai support satu item, sebagai berikut, (Hidayati, Widi Nugroho, and Nurjoko 2021), (Badaruddin and Rayendra 2022):

$$\text{Support} = \frac{\text{Jumlah transaksi}}{\text{Total transaksi}} \times 100\%$$

Untuk menghitung nilai *support* dari dua item menggunakan rumus berikut:

$$\text{Support} = \frac{\text{Jumlah transaksi A dan B}}{\text{Total transaksi}}$$

Setelah frequent itemset ditemukan, carilah aturan asosiasi yang mencukupi persyaratan *minimum confidence*. Pembentukan aturan asosiasi ketika nilai *support* sudah didapatkan dan pola kombinasi itemset telah terbentuk. Menghitung aturan asosiasi dan menghitung nilai *confidence* untuk tiap aturan asosiasi yang terbentuk. Rumus nilai *confidence* sebagai berikut:

$$\text{Confidance} = \frac{\sum \text{transaksi A dan B}}{\sum \text{transaksi A}}$$

Langkah terakhir, tahap evaluasi, menilai proses pengolahan data dengan menggunakan aturan asosiasi yang menunjukkan hubungan. Jika nilainya kurang dari yang sudah ditetapkan, maka aturan tersebut tidak akan dimunculkan karena tidak terhubung sama sekali, (Takdirillah 2020).

2.4 Analisis

Analisis adalah perincian sistem informasi yang lengkap menjadi bagian-bagiannya untuk tujuan menemukan dan memeriksa masalah, peluang, hambatan, yang telah terjadi, dan kebutuhan yang diinginkan untuk membuat rekomendasi untuk perbaikannya, (Hanik Mujiati 2013). Analisis adalah suatu proses kerja yang dimulai sebelum riset didokumentasikan melalui proses penulisan laporan. (Syahidin and Adnan 2022).

Dalam penelitian ini, aturan analisis keranjang belanja (market basket analysis). Analisis keranjang belanja adalah metode yang digunakan untuk menganalisis perilaku pembelian konsumen dan mengidentifikasi hubungan antar berbagai barang yang dimasukkan ke keranjang belanja dan dibeli pada transaksi tertentu. (Kasus et al. 2012). Dalam kasus ini, proses analisis tidak dipengaruhi oleh jumlah setiap item. Market basket analysis dilandaskan pada berbagai jenis

barang daripada banyak barang sama yang dibeli. Maksud dari analisis ini ialah untuk menemukan hubungan antara satu barang dengan yang lain, dan hasilnya dapat digunakan untuk membantu strategi penjualan di masa mendatang. (Raudhah, Muhammad, and Ramadhany 2022).

2.5 Pembelian

Pembelian ialah satu diantara fungsi utama operasi bisnis sebuah industri yang mencakup pengadaan, pemesanan, dan pengiriman barang jasa, (Heripracoyo and Persediaan 2009). Sementara pembelian juga dapat berarti pembelanjaan uang untuk barang masuk atau pembelian barang yang akan dijual, kesepakatan pembelian ini terbentuk pada penjual yang membeli barang tersebut. (Nurhayati, Josi, and Hutagalung 2018). Beralaskan penjelasan sebelumnya, dapat dirangkum bahwasanya pembelian ialah sebuah proses atau kegiatan untuk mendapatkan barang, jasa, atau bahan yang diinginkan.

2.6 Pelanggan

Orang atau entitas yang bertransaksi dengan penyedia barang, baik secara langsung maupun tidak langsung, dikenal sebagai pelanggan. Pelanggan memiliki peran penting yang dapat menguntungkan sebuah perusahaan atau organisasi. Pelanggan dapat diklasifikasikan dalam kategori berikut:

- a. Orang yang tidak tergantung pada kita, melainkan kita yang tergantung padanya ialah pelanggan.
- b. Pelanggan ialah pihak yang membawa kita ke apa yang kita inginkan.
- c. Tidak ada yang pernah menang dalam perselisihan dengan pelanggan..
- d. Target utama penjual ialah kepuasan dari pelanggan.

Oleh karena itu, pelanggan adalah individu yang memiliki pengaruh positif terhadap sebuah perusahaan, perusahaan, atau organisasi.

2.7 Aplikasi Rapidminer

Rapidminer adalah perangkat lunak pembelajaran ilmu data mining dengan sekitar seratus solusi pembelajaran untuk klasifikasi, asosiasi, dan pengkelompokan. *Rapidminer* juga dipakai untuk menganalisis dan penambangan data, dan dapat digabungkan dengan banyak bahasa pemrograman. Dibuat menggunakan Java, hingga dapat bekerja di berbagai sistem operasi. *Rapidminer* juga memiliki bentuk visualisasi atau tampilan yang ramah pengguna, membuatnya mudah digunakan. (Nasional et al. 2021).

2.8 Penelitian Terdahulu

Tinjauan pustaka yang dilakukan dalam penelitian ini berdasarkan pada penelitian sebelumnya dalam mendukung penelitian yang sedang dilakukan, tertera pada Tabel 2.1 dibawah ini:

Table 2. 1 Penelitian Terdahulu

No	Nama Peneliti	Judul Penelitian	Metode	Hasil
1.	Nurul Hidayati. Handoyo Widi Nugroho. Nurjoko, (2021).	Penerapan data mining untuk menghasilkan pola pembelian roti menggunakan algortima aprirori.	Algortima Apriori.	Data mining menggunakan algoritma apriori ternyata bisa digunakan bakal menemukan pola pembelian roti.
2.	Styawati. Andi Nurkholis. Krisma Nur Anjani (2021).	Analisa Pola Transaksi Pelanggan Menggunakan Algoritma Apriori.	Algortima Apriori.	Menerapkan Apriori pada transaksi pelanggan tuk memberikan aturan rekomendasi Toko Diengva.
3.	Multiati Badaruddin, Rayendra (2022)	Penerapan Algoritma Apriori Pada Analisa Data Penjualan E-Commerce	Algoritma Apriori.	Hasil dari pembahasan menyimpulkan bahwasanya data penjualan memuat 3 item yang dapat diprioritaskan penjualanya .

No	Nama Peneliti	Judul Penelitian	Metode	Hasil
4.	Robby Takdirillah (2020)	Penerapan Data Mining Menggunakan Algoritma Apriori Terhadap Data Transaksi Sebagai Pendukung Informasi Strategi Penjualan.	Algoritma Apriori.	Hasil pengolahan digunakan untuk menemukan aturan asosiasi data penjualan sehingga dapat dijadikan sebagai informasi pengambilan keputusan yang baru.
5.	Fitri Purwaningtias (2022).	Penggunaan Algoritma Apriori Untuk Penjualan Sparepart.	Algoritma Apriori.	Algoritma yang digunakan membantu manajemen perusahaan bagian penjualan untuk melihat sparepart apa saja yang paling banyak dibeli.
6.	Zaenal Abidin, Arsyah Kharisma, Arliyanti Nurdin (2022).	Penerapan Algoritma Apriori Pada Penjualan Suku Cadang Kendaraan Roda Dua pada Tokok Prima Motor	Algoritma Apriori.	Hasil menghasilkan aturan asosiasi yang memenuhi nilai minimum yang ditentukan agar valid.