

BAB II
TINJAUAN PUSTAKA DAN LANDASAN TEORI

2.1 Tinjauan Pustaka

Dalam penyusunan penelitian ini, inspirasi dan referensi berasal dari penelitian – penelitian sebelumnya yang berkaitan dengan latar belakang masalah pada penelitian ini. Adapun penelitian yang berhubungan dengan penelitian ini dapat dilihat dalam Tabel 2.1, sebagai berikut:

Tabel 2.1 Tinjauan Pustaka

No	Peneliti	Judul Penelitian	Masalah	Metode	Hasil
1.	Teguh Iman Hermanto (2013)	<i>Data Mining</i> Menggunakan Algoritma <i>A-Priori</i> Sebagai Alat Bantu Penentuan Pola Konsumen Pada <i>E-Commerce</i> Moshi-Moshi Petshop	Diperlukannya sebuah aplikasi yang mampu memilah dan memilih data, sehingga dapat diperoleh informasi yang berguna bagi penggunaanya	Algoritma <i>A-Priori</i>	Support 10% Confidence 10%
2.	Arief Soma Darmawan (2015)	Algoritma <i>A-Priori</i> Untuk Rekomendasi Penawaran Produk Dibatik Putra Ghofur Pekalongan	Banyaknya jumlah transaksi maka perlu adanya analisis transaksi penjualan yang bisa dipergunakan sebagai promosi penawaran produk kepada pelanggan	Algoritma <i>A-Priori</i>	Support 0.07 Confidence 1
3.	Robi Yanto dan Riri Khoiriah (2015)	Implementasi <i>Data Mining</i> Dengan Metode Algoritma <i>A-Priori</i> Dalam Menentukan Pola Pembelian Obat	Sasaran pembelian obat tidak ditentukan secara baik dan tidak potensial	Algoritma <i>A-Priori</i>	Support dan confidence 0,725

No	Peneliti	Judul Penelitian	Masalah	Metode	Hasil
4.	Rizky Mei Anggraeni (2014)	Perbandingan Algoritma <i>Apriori</i> Dan Algoritma <i>Fp-Growth</i> Untuk Perekomendasi Pada Transaksi Peminjaman Buku Di Perpustakaan Universitas Dian Nuswantoro	Diperlukanya sebuah sistem perekomendasi untuk merekomendasikan buku dengan menemukan himpunan data yang paling sering muncul dalam suatu kumpulan data	Algoritma <i>Apriori</i> Dan Algoritma <i>Fp-Growth</i>	Support 30% Confidence 50%

Berdasarkan Tabel 2.1, penelitian yang terkait dengan penelitian ini adalah Teguh Iman Hermanto (2013), dengan judul penelitian *data mining* menggunakan algoritma *A-Priori* sebagai alat bantu penentuan pola konsumen pada *e-commerce* Moshi – Moshi Petshop. Menerapkan metode *association rule* sebagai salah satu fungsi *data mining* untuk menemukan pola dan aturan (*rules*) dari serangkaian himpunan data menggunakan algoritma *A-Priori*. Penerapan metode ini dilakukan dengan cara mengintegrasikan fitur analisis keranjang pasar ke dalam sebuah aplikasi berbasis *website*. Penentuan pola dan aturan pada metode *association rule* dilakukan dengan menggunakan 1000 transaksi penjualan selama enam bulan dan menggunakan minimum *support* 10% dan minimum *confidence* 10%. Hasil dari analisis ini dapat dimanfaatkan oleh manager sebagai alat untuk mendukung pengambilan keputusan bisnis pada *e-commerce* Moshi – Moshi Petshop. Sehingga transaksi penjualan dan proses bisnis menjadi lebih cepat dan praktis[8].

Penelitian yang dilakukan oleh Arief Soma Darmawan (2015), yang berjudul algoritma *A-Priori* untuk rekomendasi penawaran produk di Batik Putra Ghofur Pekalongan. Penelitian tersebut menggunakan data transaksi sebanyak 500 dengan jumlah barang yang terjual sebanyak 1370. Data tersebut diolah menggunakan metode *A-Priori* dengan menggunakan nilai *minimum support* adalah 0,07 dan dengan nilai *minimum confidence* 1. Menghasilkan informasi untuk mengambil keputusan tentang tata letak batik, serta stok batik[3].

Penelitian yang dilakukan oleh Robi Yanto dan Riri Khoiriah (2015) berjudul implementasi *data mining* dengan metode algoritma *A-Priori* dalam menentukan

pola pembelian obat. Permasalahan dalam penelitian ini adalah pihak apotik ingin mengetahui obat apa saja yang bisa dibeli oleh konsumen sehingga pihak apotik akan terus menyediakan obat tersebut sehingga dapat meningkatkan penjualan obat. Penelitian ini menggunakan metode *A-Priori* dan pengujian dilakukan berdasarkan nama obat. Metodologi pengembangan sistem yang digunakan yaitu metode *waterfall* yang terdiri dari analisis desain, pengkodean, dan pengujian. Sistem yang dibangun dengan bahasa pemrograman *Visual Basic 6.0* dan database *MySQL* menunjukkan hasil yang telah memenuhi kebutuhan dalam penentuan pola pembelian obat berdasarkan kecenderungan pembelian obat oleh pelanggan[5].

Penelitian yang dilakukan oleh Rizky Mei Anggraeni, dengan judul perbandingan algoritma *A-Priori* dan algoritma *Fp – Growth* untuk rekomendasi pada transaksi peminjaman buku di perpustakaan Universitas Dian Nuswantoro dibuat sebuah sistem di mana nantinya dapat dijadikan sistem rekomendasi menggunakan dua metode *data minin* dengan membandingkan algoritma *A-Priori* dan *Fp-Growth*. Hal tersebut bertujuan untuk menentukan buku-buku yang sering dipinjam bersamaan pada transaksi peminjaman buku sebelumnya. Dari permasalahan di atas mengenai data transaksi peminjaman buku, algoritma *A-Priori* dan *Fp-Growth* dapat digunakan dalam rekomendasi buku perpustakaan dengan teknik *data mining* menggunakan aturan asosiasi untuk menentukan buku – buku yang sering dipinjam dalam waktu bersamaan. Hasil analisis kelayakan bahwa algoritma *A-Priori* dan *Fp-Growth* membentuk aturan asosiasi dari kedua algoritma yaitu sama. Perbedaan terletak pada jumlah data yang digunakan dimana dengan menggunakan *A-Priori* jumlah data yang digunakan lebih banyak dibandingkan data menggunakan *FP - Growth*. Dari hasil kinerja kedua algoritma semakin besar *minimumsupport* nya maka *itemsets* dan *association rule* yang terbentuk tidak ada, maka dari itu dicari *minimumsupport* yang paling mendekati agar *frequentitemsets* dan *association rule* yang terbentuk sempurna atau mendekati sempurna, dan jumlah data buku yang dipinjam ditingkatkan agar pola *mining* yang terbentuk semakin bervariasi. Maka dapat disimpulkan bahwa kedua algoritma tersebut layak diimplementasikan untuk rekomendasi peminjaman buku berdasarkan aturan asosiasi[15].

Perbedaan penelitian dari tinjauan pustaka yang telah dipaparkan di atas pada penelitian ini objek penelitian adalah sebuah perusahaan deagang yang berlokasi di Purwokerto yaitu PD. Buana yang memiliki kesulitan dalam melakukan strategi penjualan. Serta penelitian ini menggunakan algoritma *A-Priori* karena berdasarkan peneliti 4, algoritma *A-Priori* lebih sederhana karena tidak melalui tahapan pencarian FP – Tree terlebih dahulu dan dapat menangani data yang besar. Sedangkan algoritma lainnya memiliki kelemahan dalam penggunaan memori saat jumlah data besar, tentunya berpengaruh terhadap banyaknya *item* yang diproses serta mudah di pahami struktur kerja dan implementasinya.

2.2 Landasan Teori

2.2.1 Strategi Pemasaran

Strategi pemasaran merupakan strategi untuk melayani pasar atau segmen pasar yang di jadikan target oleh perusahaan. Definisi strategi pemasaran adalah logika pemasaran yang digunakan oleh perusahaan dengan harapan agar unit bisnis dapat mencapai tujuan perusahaan.

Merancang strategi pemasaran yang kompetitif dimulai dengan melakukan analisis terhadap pesaing. Perusahaan membandingkan nilai dan kepuasan pelanggan dengan nilai yang diberikan oleh produk, harga, promosi, dan distribusi (*marketing mix*) terhadap pesaing dekatnya.

Strategi pemasaran didasarkan atas lima konsep strategi berikut :

- a. Segmentasi pasar, tiap pasar terdiri dari bermacam-macam pembeli yang mempunyai kebutuhan, kebiasaan membeli dan reaksi yang berbeda-beda. Perusahaan tak mungkin dapat memenuhi kebutuhan semua pembeli. Karena itu perusahaan harus mengelompok kelompokkan pasar yang bersifat heterogen ke dalam satuan–satuan pasar yang bersifat homogeny
- b. *Market positioning*, perusahaan tak mungkin dapat menguasai pasar keseluruhan. Maka prinsip strategi pemasaran kedua adalah memilih pola spesifik pemusatan pasar yang akan memberikan kesempatan maksimum kepada perusahaan untuk mendapatkan kedudukan yang kuat. Dengan kata lain perusahaan harus memilih segmen pasar yang dapat menghasilkan penjualan dan laba yang paling besar.

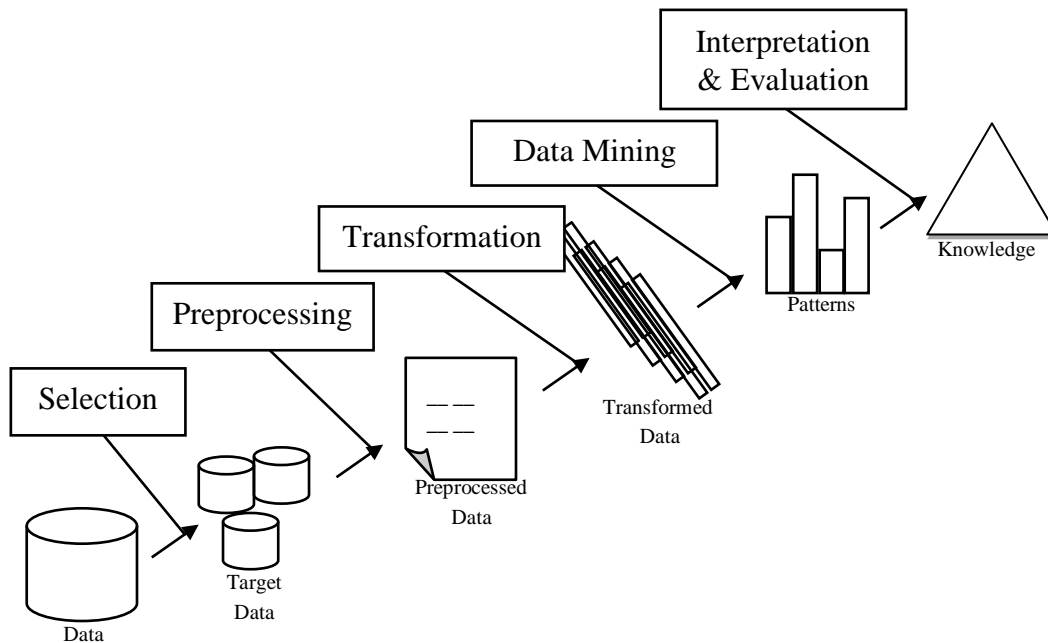
- c. *Targeting* adalah strategi memasuki segmen pasar yang dijadikan sasaran penjualan.
- d. *Marketing mix strategy*, kumpulan variable – variable yang dapat digunakan perusahaan untuk mempengaruhi tanggapan konsumen. Variabel – variable yang dapat mempengaruhi pembeli adalah variable – variable yang berhubungan dengan *product, place, promotion* dan *price* (4P).
- e. *Timing strategy*, penentuan saat yang tepat dalam memasarkan produk merupakan hal yang perlu diperhatikan. Meskipun perusahaan melihat adanya kesempatan baik. Terlebih dulu harus dilakukan persiapan baik produksi[16].

2.2.2 Data Mining

Data mining adalah melakukan ekstraksi untuk menghasilkan informasi penting yang sifatnya implisit dan sebelumnya tidak diketahui dari suatu data[10]. Peneliti lain mendefinisikan *data mining* sebagai proses dalam mempekerjakan satu atau lebih transaksi pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis[11].

Sedangkan menurut Tahyudin, penambahan data atau *data mining* merupakan serangkaian proses untuk menggali nilai tambah dari sekumpulan data berupa pengetahuan yang selama ini tersembunyi di balik data atau tidak diketahui secara manual. Proses dalam mendapatkan nilai tambah dari sekumpulan data sering juga dikenal sebagai penemuan pengetahuan dari pangkalan data (*Knowledge Discovery in Database = KDD*) yaitu tahap – tahap yang dilakukan untuk mendapatkan pengetahuan dari sekumpulan data[10].

Menurut Hermawati, proses *Knowledge Discovery in Databases* (KDD) merupakan tahapan proses dalam penggunaan *data mining*.



Gambar 2.1 Proses *Knowledge Discover in Database* (KDD)

Berdasarkan Gambar 2.1, dapat diuraikan sebagai berikut:

- a. Menggali pengetahuan awal serta apa sasaran pengguna dan memahami domain aplikasi untuk mengehui.
- b. Membuat target data-*set* yang meliputi fokus pada *subset* data dan pemilihan data.
- c. Transformasi data dan pembersihan meliputi eliminasi *derau*, *outliers*, *missing value* serta pemilihan fitur, dan reduksi dimensi.
- d. Penggunaan algoritma dari *data mining* yang terdiri atas asosiasi, sekuensial, klasifikasi, klasterisasi, dan lain – lain.
- e. Interpretasi, evaluasi, dan visualisasi pola dalam melihat apakah ada sesuatu yang baru dan menarik serta dilakukan iterasi jika diperlukan[11].

Menurut Larose, pembagian *data mining* menjadi beberapa kelompok berdasarkan tugas yang dilakukan, yaitu:[12]

1. Deskripsi

Peneliti atau analis melakukan percobaan untuk mendapatkan cara mendeskripsikan pola serta kecenderungan pada data di mana deskripsi tersebut seringkali memberi kemungkinan penjelasan terhadap suatu pola.

2. Estimasi

Estimasi hampir mirip dengan klasifikasi, hanya saja yang membedakan yaitu variabel target estimasi cenderung mengarah kepada numerik dibandingkan mengarah ke kategori.

3. Prediksi

Prediksi merupakan memperkirakan nilai masa mendatang. Teknik ini hampir sama dengan klasifikasi kecuali dalam prediksi nilai *item* dari hasil akan ada di masa mendatang.

4. Klasifikasi

Suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah diidentifikasi. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada dan telah diklasifikasi dengan menggunakan hasilnya untuk memberikan sejumlah aturan.

5. Pengklusteran

Pengklusteran yaitu mengolompokkan, mengamati, dan membangun kelas dari objek yang mempunyai kesamaan. Kluster adalah sekumpulan data yang saling mempunyai kesamaan dan mempunyai ketidaksamaan terhadap kluster yang berbeda.

6. Asosiasi

Analisis asosiasi merupakan teknik *data mining* untuk menemukan aturan asosiasi antara suatu kombinasi *item*. Analisis asosiasi menjadi terkenal karena adanya aplikasi untuk menganalisis isi keranjang belanja di pasar swalayan, analisis asosiasi juga sering disebut *market basket analysis*[4].

Andaikan sejumlah macam *item* yang tersedia dalam toko adalah $I = \{i_1, i_2, \dots, i_d\}$ sedangkan sejumlah keranjang belanja adalah $T = \{t_1, t_1, \dots, t_N\}$, setiap transaksi t_i berisi bagian (*subset*) *item* yang dipilih dari I . Dalam analisis asosiasi, *itemset* adalah kumpulan *item* di keranjang belanja dalam jumlah kosong atau lebih. Jika k -*itemset* disebut k -*itemset*, misalnya {Beras, Telur} disebut 2-*itemset*. Null (atau kosong) adalah *itemset* yang tidak berisi *item* apapun.

Lebar transaksi (*transaction width*) didefinisikan sebagai jumlah *item* yang ada dalam sebuah transaksi. Sebuah transaksi t_i dikatakan berisi *itemset* X jika X adalah *subset* t_i . Sifat penting dalam *itemset* adalah *support count*, yang

di idefinisikan sebagai jumlah transaksi yang berisi *itemset* tertentu yang secara matematis dinotasikan sebagai $\sigma(X)$. Sementara *itemset* X dapat dinyatakan sebagai berikut:

$$\Sigma(X) = |\{t_i | X \subseteq t_i, t_i \in T\}| \dots\dots\dots (2.1)$$

Lambang Σ , $|$ menyatakan jumlah *item* dalam *set*.

Aturan asosiasi adalah pernyataan implikasi bentuk $X \rightarrow Y$, dimana X dan Y adalah *itemset* yang lepas (*disjoint*) yang memenuhi persyaratan $X \cap Y = \{\}$. Kekuatan aturan asosiasi dapat diukur dengan *confidence* dan *support*. *Confidence* digunakan untuk menentukan seberapa sering *item* didalam Y muncul dalam transaksi yang berisi X sedangkan *support* digunakan untuk menentukan seberapa banyak aturan dapat diterapkan pada *set* data. Definisi formal tersebut adalah sebagai berikut:

$$Support, s(X \rightarrow Y) = \frac{\sigma(XUY)}{N} \dots\dots\dots (2.2)$$

$$Confidence, c(X \rightarrow Y) = \frac{\sigma(XUY)}{\sigma(X)} \dots\dots\dots (2.3)$$

N adalah jumlah transaksi dalam *set* data.

Support merupakan ukuran yang penting di dalam analisis asosiasi karena aturan yang sangat lemah nilai *supportnya* berarti asosiasi yang sangat jarang terjadi dalam *set* data. Jika aturan yang *supportnya* sangat rendah berarti asosiasi *supportnya* juga dipandang sebagai sesuatu yang tidak menarik dalam perspektif bisnis karena memberikan keuntungan dalam mempromosikan *item* dimana pelanggan jarang membeli barang bersama – sama.

Dalam mengukur keandalan dari inferensi yang dibuat oleh aturan digunakan *confidence*. *Confidence* memberikan cara untuk menemukan aturan asosiasi secara efisien. Untuk aturan $X \rightarrow Y$, nilai *confidence* yang tinggi menandakan banyaknya Y yang muncul dalam transaksi yang berisi X.

Penemuan aturan asosiasi (*association rule discovery*) didefinisikan sebagai berikut, diberikan sejumlah transaksi T, carilah semua aturan yang mempunyai $support \geq minsup$ dan $confidence \geq minconf$, dimana *minsup* merupakan ambang batas *support*, sedangkan *minconf* merupakan ambang batas *confidence*.

Dengan demikian metodologi dasar *association rule mining* terbagi menjadi 3 (tiga) tahap berdasarkan Tahyudin, yaitu:[10]

a. Analisa pola frekuensi tinggi

Tahap ini untuk mencari kombinasi *item* yang memenuhi syarat minimum dari *support* dalam database. Nilai dari *support* sebuah *item* diperoleh menggunakan rumus sebagai berikut:

$$Support(A) = \frac{\text{Jumlah Transaksi mengandung A}}{\text{Total Transaksi}} \dots\dots\dots (2.4)$$

Adapun nilai *support* dari dua *item* diperoleh dari rumus berikut:

$$Support(A \cap B) = \frac{\text{Jumlah Transaksi mengandung A dan B}}{\text{Total Transaksi}} \dots\dots\dots (2.5)$$

b. Pembentukan aturan asosiasi

Aturan asosiasi adalah suatu prosedur untuk mencari hubungan antar *item* dalam suatu *dataset*. Setelah semua pola frekuensi tinggi ditemukan kemudian dicari aturan asosiasi yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan asosiasi $A \rightarrow B$. Rumus yang digunakan untuk memperoleh nilai *confidence* dari aturan $A \rightarrow B$ sebagai berikut:

$$Confidence = P(B|A) = \frac{\text{Jumlah Transaksi mengandung A dan B}}{\text{Jumlah Transaksi yang mengandung A}} \dots\dots\dots (2.6)$$

c. Pengujian hasil aturan asosiasi

Hasil aturan asosiasi (*rule*) diuji berdasarkan *lift ratio* yang dihasilkan setiap *rule*. *Lift ratio* adalah parameter penting selain *support* dan *confidence* dalam *association rule*. *Lift ratio* mengukur seberapa penting *rule* yang telah terbentuk berdasarkan nilai *support* dan *confidence*. *Lift ratio* merupakan nilai yang menunjukkan keakuratan proses transaksi dan memberikan informasi apakah benar produk A dibeli bersamaan dengan produk B. Untuk menentukan nilai *rule* yang valid adalah jika memiliki nilai *lift ratio* > 1. Rumus untuk menghitung nilai *lift ratio* sebagai berikut:

$$Lift\ Ratio = \frac{Confidence}{Benchmark\ Confidence} \dots\dots\dots (2.7)$$

Benchmarkconfidence adalah perbandingan antara jumlah semua *itemconsequent* terhadap total jumlah transaksi. Rumus *benchmarkconfidence* dapat dilihat pada persamaan berikut:

$$Benchmark\ Confidence = \frac{Nc}{N} \dots\dots\dots (2.8)$$

Dimana, N_c adalah jumlah transaksi dengan *item* dalam *consequent*, dan N adalah total jumlah transaksi pada database.

2.2.3 Algoritma A-Priori

Algoritma *A-Priori* adalah suatu algoritma dasar diusulkan oleh Agrawal & Srikat pada tahun 1994 untuk penentuan frequent *itemsets* untuk menghasilkan aturan asosiasi[13]. Persoalan *association rule mining* terdiri dari dua sub persoalan:

- a. Menemukan semua kombinasi dari *item*, disebut dengan *frequent itemsets*, yang memiliki *support* yang lebih besar daripada *minimum support*.
- b. Gunakan *frequent itemsets* untuk men-generate aturan yang di kehendaki. Misal, ABCD dan AB adalah *frequent*, maka didapatkan aturan $AB \rightarrow CD$ jika *rasio* dari *support* (ABCD) terhadap *support* (AB) sedikitnya sama dengan *minimum confidence*. Aturan ini memiliki *minimum support* karena ABCD adalah *frequent*.

Algoritma *A-Priori* yang bertujuan untuk menemukan *frequent itemsets* dijalankan pada sekumpulan data. Pada *iterasi* ke k -, akan ditemukan semua *itemsets* yang memiliki k -*items*, disebut dengan k -*itemsets*. *Support* dan *confidence* adalah dua ukuran kepercayaan yang menunjukkan kepastian dan tingkat kegunaan suatu *rule* yang ditemukan. Pada umumnya *association rule* yang ditemukan dikatakan menarik apabila *rule* tersebut memenuhi baik *minimum support* maupun *minimum confidence* yang telah ditentukan.

2.2.4 RapidMiner 5 Data Mining Tool

RapidMiner yang sebelumnya dikenal sebagai *Yet Another Learning Environment* (YALE), dikembangkan pada tahun 2001 oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer. Pada tahun 2007, perubahan nama software dari YALE ke *RapidMiner*. *RapidMiner* adalah aplikasi *data mining* yang berbasis sistem *open source*, aplikasi data mining ini sangat terkemuka dan ternama. Tersedia sebagai aplikasi yang berdiri sendiri untuk analisis data dan sebagai mesin data mining untuk integrasi ke dalam produk itu sendiri[14].

Fitur yang dimiliki *RapidMiner* antara lain:

- a. Merupakan perangkat lunak berlisensi gratis (*open source*).

- b. Multi – platform karena diprogram dalam bahasa Java.
- c. Internal data berbasis XML, sehingga memudahkan pertukaran data eksperimen.
- d. Dilengkapi dengan scripting language untuk otomatisasi eksperimen.
- e. Memiliki Graphical User Interface, command line mode (batch mode), dan Java API yang dapat dipanggil dalam program lain.
- f. Dapat dikembangkan dengan mengembangkan plug – in dan extension.
- g. Fasilitas plotting untuk visualisasi data multidimensi model.