

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Tinjauan Penelitian Terdahulu**

Penelitian dibidang ini telah banyak dilakukan, pembahasan topik yang diangkat untuk dijadikan dasar serta sumber data. Hal ini bertujuan untuk mendukung penelitian yang sedang dilakukan. Selain itu untuk mengajukan gagasan atau argumen diperlukan data yang relevan dengan penelitian sehingga dapat dikatakan valid. Tujuan lain dilakukannya studi terhadap penelitian terdahulu adalah untuk menggali kekurangan dan kelebihan suatu metode yang akan digunakan. Dengan begitu peneliti dapat menghindari atau mengatasi permasalahan terkait.

Oleh sebab itu sebelum memulai penelitian perlu dilakukan eksplorasi dan studi pustaka baik melalui internet maupun jurnal – jurnal yang relevan dengan topik aturan asosiasi penjualan produk di kafe. Berikut adalah beberapa penelitian yang dijadikan tinjauan pustaka. Pada beberapa halaman selanjutnya disediakan tabel perbandingan antar-penelitian yang dapat dilihat pada Tabel 2.1

Penelitian pertama dilakukan oleh Achmad Nur Cahyono (2016) dengan judul penelitian *“Implementasi Data Mining Algoritma Apriori Pada Penjualan Sparepart Motor Di Ahas Putra Motor”*. Penelitian ini bertujuan untuk menghasilkan sebuah pemrosesan data transaksi penjualan yang akan diolah menjadi sebuah informasi

untuk meningkatkan produktifitas penjualan serta pemasaran pada Ahas Putra Motor. Penelitian ini mengimplementasikan Algoritma Apriori berbasis web pada toko sparepart Ahas Putra Motor yang memberikan kemudahan kepada administrator dalam mengelolah data penjualan dan memberikan kemudahan kepada pemilik untuk mengetahui tingkat penjualan barang.

Penelitian kedua dilakukan oleh Gilang Abi Saputro (2017) dengan judul penelitian "*Penerapan Algoritma Apriori Untuk Mencari Pola Penjualan Di Cafe*". Penelitian ini bertujuan untuk membantu pemilik kafe menemukan susunan menu *item* di kafe untuk proses pengembangan promosi dengan menggunakan algoritma apriori. Penelitian ini mengimplementasikan Algoritma Apriori pada kafe *Journey Coffee* yang menunjukan pola beli konsumen dan seberapa kuat suatu item mempengaruhi item lain.

Penelitian ketiga dilakukan oleh Lucki Darmawan (2017) dengan judul "*Penerapan Algoritma Apriori Dalam Menentukan Pola Pembelian Konsumen di Cafe Hidden Toast and Float*". Penelitian ini bertujuan untuk membantu mengetahui pola pembelian konsumen dan rekomendasi menu pada konsumen di Cafe Hidden Toast and Float. Penelitian ini mengimplementasikan dua buah sistem berbasis *website* dengan sistem admin digunakan untuk mengolah data yang ada pada *database* serta *website* pelanggan yang akan digunakan untuk pembelian secara *online*, serta penerapan algoritma apriori dengan hasil uji coba data sampel dan data

real yang menghasilkan rekomendasi kombinasi menu yang paling sering dibeli berdasarkan seluruh data transaksi.

Penelitian keempat dilakukan oleh Achmad Zaenal Hidayat (2017) dengan judul *“Penerapan Algoritma Apriori Untuk Menentukan Strategi Penjualan Pada Rumah Makan Dapoer Emak Pati”*. Penelitian ini bertujuan untuk meningkatkan penjualan pada rumah makan Dapoer Emak sehingga meminimalisir menu yang tidak laku dan bahan makanan yang basi atau busuk. Penelitian ini menerapkan perangkat lunak RapidMiner sebagai alat bantu dalam proses terciptanya aturan yang akan digunakan pada aplikasi.

Tabel 2.1 Penelitian Terdahulu

No	Peneliti	Judul Penelitian	Perumusan Masalah	Hasil/Temuan
1.	Achmad Nur Cahyono(2016)	Implementasi Data Mining Algoritma Apriori Pada Penjualan Sparepart Motor Di Ahas Putra Motor	Bagaimana cara mengatasi penumpukan data dengan menggunakan Algoritma Apriori ?	Hasil prosedur data transaksi yaitu ‘Supplier memberikan barang yang dibeli kepada petugas’, ‘Petugas mengecek dan input data barang, masuk ke database’, ‘Barang diberikan lagi kepada supplier beserta nota pembelian’, ‘Petugas membuat laporan penjualan kepada pimpinan’.



- Percobaan pertama menggunakan data transaksi Journey coffee selama periode bulan Desember 2015 (31 hari) dengan parameter nilai minimum *support* sebesar 4% dan nilai minimum *confidence* sebesar 50%.
- Percobaan kedua menggunakan data transaksi Journey coffee selama periode bulan Desember 2015 (31 hari) dengan parameter nilai minimum *support* sebesar 8% dan nilai minimum *confidence* sebesar 50%.
- Percobaan ketiga menggunakan data transaksi Journey coffee selama periode bulan Desember 2015 (31 hari) dengan

3.	Lucki Darmawan (2017)	<p>Penerapan Algoritma Apriori Dalam Menentukan Pola Pembelian Konsumen Di Cafe Hidden Toast And Float</p>	<p>• Bagaimana cara penerapan Algoritma Apriori dalam menentukan pola pembelian konsumen pada Kafe Hidden Toast And Float ?</p> <p>• Data apa sajakah yang akan diolah oleh Algoritma Apriori sehingga</p>	<p>• Hasil uji coba dari data sampel dan data real dengan menggunakan minimum support 20% dan minimum confidence 60% dari data sampel menghasilkan 4 pola kombinasi menu dengan kombinasi menu tertinggi yaitu Dark Choco Jam dan Cappucino dengan Support 37% dan Confidence 70%.</p> <p>• pada data training dan data sampel tidak ditemukan pola kombinasi asosiasi. Kemudian penulis mencoba untuk menurunkan nilai minimum support menjadi 5% dan minimum confidence 10%.</p> <p>• Pada data sampel didapatkan 58 kombinasi asosiasi dengan pola tertinggi</p>
----	-----------------------	--	--	---

4.	Achmad Zaenal Hidayat(2012)	<p>Penerapan Algoritma Apriori Untuk Menentukan Strategi Penjualan Pada Rumah Makan “Dapoer Emak” Pati</p>	<p>Bagaimana menentukan strategi penjualan dengan menggunakan Algoritma Apriori di rumah makan “Dapoer Emak”?</p>	<p>Kesimpulan yang didapatkan dari hasil penelitian ini adalah metode Association Rule dapat diimplementasikan untuk strategi penjualan pada rumah makan “dapoer emak” pati dengan menggunakan 142 data transaksi penjualan menu makanan dan di temukan ada 13 rule yang terbentuk dengan bantuan tools Rapidminer dan penyeleksian hasil. Berdasarkan pengujian sistem hasil rata – rata Confidentdiperoleh hasil minimal adalah 81,2% dan Confident terbesar 94 %.</p>
----	-----------------------------	--	---	--

## 2.2 Teori Dasar Yang Digunakan

Pada penelitian ini terdapat beberapa teori dasar yang digunakan sebagai acuan dan referensi terkait pembahasan mengenai pengolahan data transaksi penjualan produk pada kafe. Sumber ini didapat dari jurnal yang relevan dengan penelitian ini serta dari studi pustaka melalui internet.

### 2.2.1 Proses Data Mining

Proses yang umumnya dilakukan oleh data mining antara lain: deskripsi, prediksi, estimasi, klasifikasi, clustering dan asosiasi. Secara rinci proses data mining dijelaskan sebagai berikut (Larose, 2005):

#### *a. Deskripsi*

Deskripsi bertujuan untuk mengidentifikasi pola yang muncul secara berulang pada suatu data dan mengubah pola tersebut menjadi aturan dan kriteria yang dapat mudah dimengerti oleh para ahli pada domain aplikasinya. Aturan yang dihasilkan harus mudah dimengerti agar dapat dengan efektif meningkatkan tingkat pengetahuan (*knowledge*) pada sistem. Tugas deskriptif merupakan tugas data mining yang sering dibutuhkan pada teknik *postprocessing* untuk melakukan validasi dan menjelaskan hasil dari proses data mining. *Postprocessing* merupakan proses yang digunakan untuk memastikan hanya hasil yang valid dan berguna yang dapat digunakan oleh pihak yang berkepentingan.

#### *b. Prediksi*

Prediksi memiliki kemiripan dengan klasifikasi, akan tetapi data diklasifikasikan berdasarkan perilaku atau nilai yang diperkirakan pada masa yang akan datang. Contoh dari tugas prediksi misalnya untuk memprediksikan adanya

pengurangan jumlah pelanggan dalam waktu dekat dan prediksi harga saham dalam tiga bulan yang akan datang.

*c. Estimasi*

Estimasi hampir sama dengan prediksi, kecuali variabel target estimasi lebih ke arah *numerik* dari pada ke arah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Sebagai contoh, akan dilakukan estimasi tekanan darah sistolik pada pasien rumah sakit berdasarkan umur pasien, jenis kelamin, berat badan, dan level sodium darah. Hubungan antara tekanan darah sistolik dan nilai variabel prediksi dalam proses pembelajaran akan menghasilkan model estimasi.

*d. Klasifikasi*

Klasifikasi merupakan proses menemukan sebuah model atau fungsi yang mendeskripsikan dan membedakan data ke dalam kelas-kelas. Klasifikasi melibatkan proses pemeriksaan karakteristik dari objek dan memasukkan objek ke dalam salah satu kelas yang sudah didefinisikan sebelumnya.



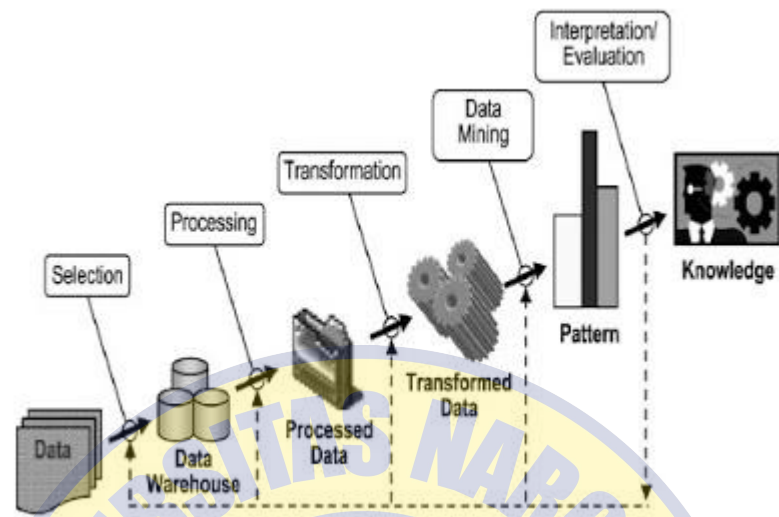
*e. Clustering*

*Clustering* merupakan pengelompokan data tanpa berdasarkan kelas data tertentu ke dalam kelas objek yang sama. Sebuah kluster adalah kumpulan *record* yang memiliki kemiripan suatu dengan yang lainnya dan memiliki ketidakmiripan dengan *record* dalam kluster lain. Tujuannya adalah untuk menghasilkan pengelompokan objek yang mirip satu sama lain dalam kelompok-kelompok. Semakin besar kemiripan objek dalam suatu cluster dan semakin besar perbedaan tiap cluster maka kualitas analisis cluster semakin baik.

*f. Asosiasi*

Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam suatu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja (*market basket analysis*). Tugas asosiasi berusaha untuk mengungkap aturan untuk mengukur hubungan antara dua atau lebih atribut.

Tahapan yang dilakukan pada proses data mining diawali dari seleksi data dari data sumber ke data target, tahap *preprocessing* untuk memperbaiki kualitas data, transformasi, data mining serta tahap interpretasi dan evaluasi yang menghasilkan *output* berupa pengetahuan baru yang diharapkan memberikan kontribusi yang lebih baik. Secara detail dijelaskan sebagai berikut (Fayyad, 1996):



Tahapan Data Mining

### 1. Data selection

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

### 2. Pre-processing / cleaning

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data.

### 3. Transformation

*Coding* adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses coding dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

### 4. Data mining

*Data mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

### 5. Interpretation / evaluation

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut interpretation. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

Data mining mempunyai fungsi yang penting untuk membantu mendapatkan informasi yang berguna serta meningkatkan pengetahuan bagi pengguna. Pada dasarnya, data mining mempunyai empat fungsi dasar yaitu:

1. Fungsi Prediksi (*prediction*). Proses untuk menemukan pola dari data dengan menggunakan beberapa variabel untuk memprediksi variabel lain yang tidak diketahui jenis atau nilainya.
2. Fungsi Deskripsi (*description*). Proses untuk menemukan suatu karakteristik penting dari data dalam suatu basis data.
3. Fungsi Klasifikasi (*classification*). Klasifikasi merupakan suatu proses untuk menemukan model atau fungsi untuk menggambarkan class atau konsep dari suatu data. Proses yang digunakan untuk mendeskripsikan data yang penting serta dapat meramalkan kecenderungan data pada masa depan.
4. Fungsi Asosiasi (*association*). Proses ini digunakan untuk menemukan suatu hubungan yang terdapat pada nilai atribut dari sekumpulan data.

### **2.2.2 Association Rule Algoritma Apriori**

*Assosiation rule* merupakan bentuk jika “kejadian sebelumnya” kemudian “konsekuensinya” (*if antecedent, then consequent*), yang diikuti dengan perhitungan aturan *support* dan *confidence*. Contoh dalam sebuah transaksi pembelian suatu produk di sebuah kafe didapat *association rule* kopi → roti bakar.

Pelanggan yang membeli kopi berkemungkinan besar pelanggan tersebut juga akan membeli roti bakar dalam waktu yang bersamaan.

Dalam menentukan suatu *association rule*, terdapat suatu *interestingness measure* (ukuran kepercayaan) yang didapat dari hasil pengolahan data dengan perhitungan tertentu yaitu:

- *Support*: suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu item/itemset dari keseluruhan transaksi. Ukuran ini menentukan apakah suatu item/itemset *confidence*-nya (misal, dari keseluruhan transaksi yang ada, seberapa besar tingkat dominasi yang menunjukkan bahwa item A dan B dibeli secara bersamaan)
- *Confidence*: suatu ukuran yang menunjukkan hubungan antara 2 item secara *conditional* (misal, seberapa sering item B jika orang membeli item A). Misalkan item A dan B dalam total transaksi, maka *support* (A) adalah jumlah transaksi yang ada A dibagi total transaksi yang disebut *support* 1-item set, dan *support* (A dan B) adalah transaksi yang ada A dan B dibagi total transaksi yang disebut *support* 2-item set. Demikian selanjutnya untuk item barang yang lebih banyak. Sedangkan *confidence* dianalisa mulai dari 2 barang karena berhubungan dengan keinginan membeli barang secara bersamaan.

Kedua ukuran ini nantinya berguna dalam menentukan suatu pola dengan membandingkan pola tersebut dengan nilai minimum kedua parameter tersebut. Bila suatu pola memenuhi kedua parameter kedua minimum parameter yang sudah ditentukan, maka pola tersebut dapat disebut sebagai *interesting rule*. Kedua ukuran ini nantinya berguna dalam menentukan *interesting association rules*, yaitu untuk dibandingkan dengan batasan (*threshold*) yang ditentukan oleh user. Batasan tersebut umumnya terdiri dari *min\_support* dan *min\_confidence*.

Algoritma apriori termasuk jenis aturan pada data mining. Menurut Dan Toomey, Apriori adalah algoritma kelas yang membantu mempelajari aturan asosiasi. Ini bekerja melawan transaksi. Algoritma mencoba untuk menemukan himpunan bagian yang umum dalam kumpulan data. Ambang batas minimum harus dipenuhi agar asosiasi dapat dikonfirmasi.

Analisis asosiasi adalah teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi item. Algoritma yang paling populer dikenal sebagai Apriori dengan paradigma *generate and test*, yaitu pembuatan kandidat kombinasi item tersebut memenuhi syarat *support minimum* (Santosa,2007).

Kombinasi item yang memenuhi syarat tersebut disebut *frequent itemset*, yang nantinya dipakai untuk membuat aturan-aturan yang memenuhi syarat *confidence minimum*. Suatu aturan asosiatif dapat diketahui dengan dua parameter, yaitu *support* dan *confidence*. *Support* (nilai penunjang) adalah presentase kombinasi

item tersebut dalam database, sedangkan *confidence* (nilai kepastian) adalah kuatnya hubungan antar item dalam aturan asosiasi (Lutfi, dkk, 2009).

Proses utama yang dilakukan dalam algoritma apriori untuk mendapat *frequent itemset* yaitu:

- *Join* (penggabungan)

Proses ini dilakukan dengan cara pengkombinasian item dengan item lainnya sehingga tidak bisa terbentuk kombinasi lagi

- *Prune* (pemangkasan)

Proses pemangkasan yaitu hasil dari item yang telah dikombinasikan kemudian dipangkas dengan menggunakan *minimum support* yang telah ditentukan.

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.

Rumus *support* adalah sebagai berikut:

$$\text{Support (A,B)} = P (A \cap B)$$

$$\text{Support (A,B)} =$$



$$\frac{\sum \text{transaksi mengandung } A \text{ dan } B}{\sum \text{total transaksi}} \times 100\%$$

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(B|A) =

$$\frac{\sum \text{transaksi mengandung } A \text{ dan } B}{\sum \text{total transaksi } A} \times 100\%$$

Cara kerja apriori :

- 1) Tentukan minimum support Iterasi 1 : hitung item-item dari *support* (transaksi yang memuat seluruh item) dengan men-*scan* database untuk 1-itemset, setelah 1-*itemset* didapatkan, dari 1-*itemset* apakah diatas *minimum support*, apabila telah memenuhi *minimum support*, 1-*itemset* tersebut akan menjadi pola *frequent* tinggi. Iterasi 2 : untuk mendapatkan 2-*itemset*, harus dilakukan kombinasi dari *k-itemset* sebelumnya, kemudian *scan* database lagi untuk hitung item-item yang



memuat *support*. itemset yang memenuhi minimum support akan dipilih sebagai pola *frequent* tinggi dari kandidat.

2) Tetapkan nilai *k-itemset* dari *support* yang telah memenuhi *minimum support* dari *k-itemset*. lakukan proses untuk iterasi selanjutnya hingga tidak ada lagi *k-itemset* yang memenuhi *minimum support*. Kelebihan dari algoritma apriori ini adalah lebih sederhana dan dapat menangani data yang besar. tentunya berpengaruh terhadap banyaknya item yang diproses serta mudah di pahami struktur kerja dan implementasinya. Apriori memiliki kelemahan karena harus melakukan scan database setiap kali iterasi, sehingga waktu yang diperlukan bertambah dengan makin banyak iterasi.

