

BAB 2 TINJAUAN PUSTAKA

2.1 Penelitian Terkait

Beberapa jurnal penelitian terkait yang diperoleh tentang pemanfaatan teknik *Data Mining* dengan metode *Association Rule* menggunakan Algoritma *Apriori* untuk dapat mengolah data dan menemukan *pattern* atau pola penjualan adalah sebagai berikut :

1. Penelitian oleh Mochammad Yusuf Pratama judul penelitiannya adalah “Penerapan Data Mining Association Rule Menggunakan Algoritma Apriori Untuk Meningkatkan Penjualan dan Memberikan Rekomendasi Pemasaran Produk Speedy” yang diterbitkan pada tahun 2014[3].

Objek studi kasus yang menjadi topik utama penelitian berikut ini adalah data penjualan *speedy* pada PT. Telkom Cabang Wonogiri.

Penelitian yang dibahas :

Bagaimana penerapan algoritma *apriori* dengan aturan *association rule* pada data penjualan produk *speedy* dapat meningkatkan kualitas pemasaran produk *speedy* di daerah Wonogiri. Tujuan dari penelitian ini adalah melakukan analisa terhadap *association rule* dengan menggunakan algoritma *apriori* berdasarkan jumlah kemunculan *itemset* dalam data penjualan *speedy* agar dapat menentukan pola strategi pemasaran sehingga dapat diperoleh kepuasan pada konsumen. Penelitian ini menggunakan nilai asumsi dengan *minimum support* 50% dan *minimum confidence* 80% dengan frekuensi maksimal 2 *itemset*.

2. Penelitian oleh Shona Chayy Bilqisth judul penelitiannya adalah “Analisis Pola Pembelian Konsumen Dengan Algoritma Apriori Pada Indomaret Indraprasta Semarang” yang diterbitkan pada tahun 2016[4].

Objek studi kasus yang menjadi topik utama penelitian berikut ini adalah data transaksi penjualan Indomaret Cabang Indraprasta Semarang.

Penelitian yang dibahas :

Pada penelitian ini dilakukan analisis pola pembelian konsumen penggunaan *Market Basket Analysis* dengan algoritma *apriori* yang bertujuan untuk mengetahui pola keterkaitan barang yang sering dibeli pelanggan secara bersamaan, penelitian ini bertujuan untuk meningkatkan tingkat kepuasan konsumen, penelitian dilakukan dengan menggunakan data transaksi penjualan barang bulan September dan Oktober tahun 2015 pada Indomaret daerah Indraprasta Semarang. Pembuktian data dilakukan dengan aplikasi Weka versi 3.7.13.

3. Penelitian oleh Wiwit Agus Triyanto judul penelitiannya adalah “Association Rule Mining untuk Penentuan Rekomendasi Promosi Produk” yang diterbitkan pada tahun 2014[5].

Objek studi kasus yang menjadi topik utama penelitian berikut ini adalah data transaksi penjualan barang supermarket.

Penelitian yang dibahas :

Dengan menggunakan *Association Rule* dapat digunakan sebagai salah satu strategi pemasaran produk kepada konsumen, dengan cara ini dapat mengetahui barang-barang apa saja yang dibeli pelanggan secara bersama-sama sehingga pedagang pun dapat merekomendasikan barang-barang yang mereka jual, teknik pengumpulan data menggunakan analisis keranjang pasar atau *market basket*. Nilai asumsi yang digunakan adalah *minimum support* 40% dan *minimum confidence* 80%.

Dari penelitian yang telah dibahas dapat disimpulkan :

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti dan Tahun	Masalah Terkait	Metode yang Digunakan	Hasil Penelitian
1.	Mochammad Yusuf Pratama, 2014	Strategi Pemasaran produk <i>Speedy</i> belum efisien ,beberapa daerah di kecamatan yang dapat menjadi potensi penjualan produk <i>speedy</i> masih belum belum terjangkau.	<i>Data Mining</i> dengan <i>Association Rule</i> dengan metode Algoritma <i>Apriori</i>	Hasil yang didapat dari proses <i>data mining</i> dapat dijadikan sebagai rekomendasi pemasaran produk yang lebih akurat dan efisien untuk menentukan pola strategi pemasaran agar penjualan produk <i>speedy</i> lebih meningkat.

No	Nama Peneliti dan Tahun	Masalah Terkait	Metode yang Digunakan	Hasil Penelitian
2.	Shona Chayy Bilqisth, 2016	Kurangunya stok barang yang memiliki potensi penjualan tinggi hal ini terjadi dikarenakan tidak ada pemanfaatan data transaksi yang masih kurang dan hanya bertugas sebagai rekaman.	<i>Market Basket Analysis</i> dengan metode Algoritma <i>Apriori</i>	berdasarkan proses <i>data mining</i> didapatkan hasil yaitu jumlah barang yang paling banyak dibeli secara bersamaan adalah mie instant dan air mineral, dari hasil tersebut didapatkan beberapa rekomendasi yaitu konsumen dapat memperbanyak stok barang menyusun tata letak barang secara berdekatan.

No	Nama Peneliti dan Tahun	Masalah Terkait	Metode yang Digunakan	Hasil Penelitian
3	Wiwit Agus Triyanto, 2014	Kurangnya dukungan dalam pembelian barang secara bersamaan sehingga tidak didapatkan hasil yang mendukung dalam pemilihan keputusan.	<i>Market Basket Analysis</i> dengan menggunakan metode Algoritma <i>Apriori</i>	Hasilnya adalah 3 rule ketika menggunakan <i>minimum support</i> 40% dan <i>minimum confidence</i> 80% yang dapat digunakan untuk membantu menentukan rekomendasi promosi suatu produk secara lebih tepat.

Relevansi dari ketiga jurnal tersebut dengan penelitian ini adalah :

1. Ketiga jurnal berikut memiliki keterkaitan didalam inti masalah yang sama yaitu kurangnya penggunaan data arsip yang memiliki potensi dalam meningkatkan efisiensi dalam pemasaran dan penjualan produk.
2. Ketiga jurnal tersebut menggunakan metode yang sama yaitu menggunakan *data mining* dengan metode *association rule*.

2.2 Data Mining

Beberapa Definisi dari *Data Mining* adalah :

1. Didalam *Data Mining* kata "*Mining*" berasal dari bahasa Inggris yang memiliki arti "*mine*". Dalam bahasa Inggris "*mine*" memiliki arti tambang sedangkan

“*mining*” artinya adalah menambang sumber daya yang berada didalam tanah, maka *data mining* merupakan sebuah penggalian makna yang tersembunyi didalam kumpulan data yang besar[6].

2. *Data mining* merupakan suatu proses dalam mencari pola, kecenderungan dan hubungan yang memiliki arti, dengan memprosesnya dalam sekumpulan data dalam skala besar dan tersimpan dengan memanfaatkan teknik pengenalan pola teknik statistik dan matematika[2].

Intinya dapat disimpulkan bahwa definisi dari *data mining* adalah proses pengambilan data atau informasi yang kemudian dapat diproses menjadi suatu pengetahuan yang bisa digunakan dan dapat diteliti lagi menemukan pengetahuan atau informasi yang dibutuhkan.

2.3 Tahapan KDD

KDD merupakan singkatan dari *Knowledge Discovery in Database* dapat dijabarkan diantaranya [7] :

1. *Data Selection* :

Data Selection atau Seleksi data yang dilakukan didalam sekelompok data sebelum tahap pencarian informasi dalam KDD dimulai.

2. *Preprocessing / Cleaning* :

Sebelum dilakukannya proses didalam *data mining*, tahap ini diwajibkan untuk melakukan proses pembersihan dengan tujuan menghapus data yang duplikat, mengecek yang tidak konsisten, dan memperbaiki *error* yang ada didalam data, seperti contoh *error* didalam tata huruf. Dalam tahap ini dilakukan juga proses memperkaya data dengan data atau informasi yang dibutuhkan dalam KDD, contohnya seperti data atau informasi eksternal.

3. *Transformation* :

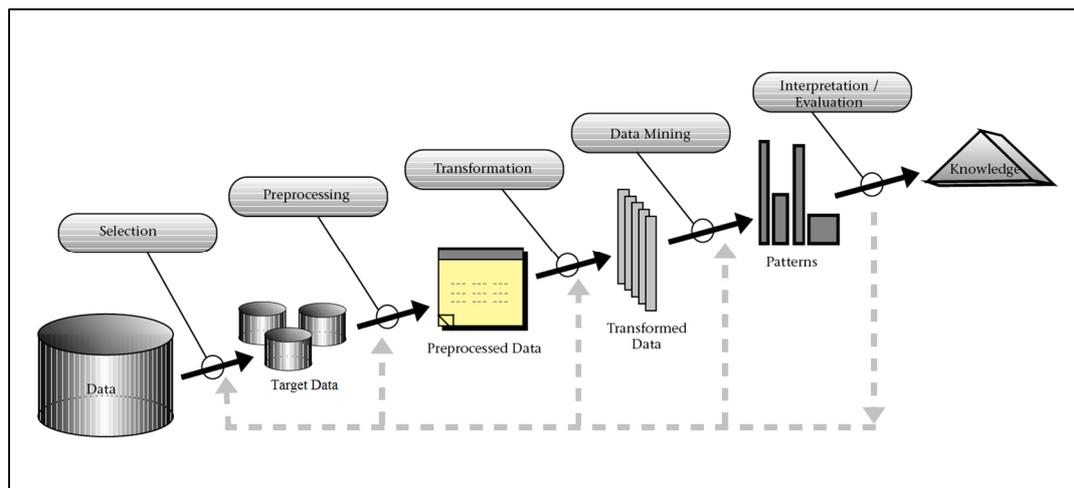
Tahap ini merupakan tahap dimana akan dilakukan transformasi data yang telah diseleksi, didalam tahap ini proses yang kreatif sangat bergantung pada pencarian pola atau jenis informasi didalam data, sehingga data yang telah mengalami tahap transformasi tersebut sesuai saat digunakan dalam proses *data mining*.

4. *Data mining* :

Salah satu tahap didalam proses *data mining* yaitu menemukan pola atau informasi yang didalam data yang telah terseleksi. Ada banyak variasi didalam metode, teknik, dan algoritma yang terdapat pada *data mining*. Kesesuaian metode atau algoritma sangat bergantung pada proses dan tujuan yang terdapat pada KDD.

5. *Interpretation / Evaluation* :

Merupakan tahap dilakukannya penerjemahan pola yang dilakukan didalam *data mining* tahap dilakukan dengan tujuan agar pola informasi yang dihasilkan nantinya dapat dimengerti oleh pihak yang berkepentingan. Didalam tahap ini juga dilakukan investigasi terhadap informasi dan pola yang ditemukan bertentangan dengan fakta atau tidak.



Gambar 2.1 Tahapan Data Mining

2.4 Teknik Data Mining

1. *Clustering*

Clustering atau sering disebut dengan Klasterisasi adalah salah satu teknik membagi *data-set* menjadi beberapa *sub-set* atau kelompok yang sedemikian rupa, sehingga elemen yang ada didalam kelompok tertentu memiliki *set property* yang dibagi bersama, dengan tingkat kesamaan yang tinggi dalam satu

kelompok dan tingkat kesamaan antar kelompok yang rendah. Teknik berikut memiliki sebuta lain yaitu *unsupervised learning*[6].

2. *Regretion*

Metode yang digunakan untuk memprediksi nilai dalam suatu *variable* berkelanjutan yang diberikan dengan dasar nilai dari *variable* lain, dengan asumsi dari model ketergantungan linier atau non-linier. Teknik tersebut sering digunakan dalam statistika dan bidang jaringan *neural network*[6].

3. *Classification*

Classification atau klasifikasi adalah teknik yang digunakan untuk menentukan suatu *record* didalam sekelompok data baru menuju kesalah satu dari beberapa kategori yang telah disebutkan. Teknik ini sering disebut juga dengan *supervised learning*[6].

4. *Association Rule*

Teknik *association rule* atau aturan asosiatif merupakan teknik *data mining* dalam mencari *rule* diantara kombinasi suatu *item*. Dalam teknik tersebut terbentuk sebuah implikasi dimana $A \rightarrow B$ merupakan sebuah *Itemset*. Contoh : $\{\text{Pena, Tinta}\} \rightarrow \{\text{Buku}\}$. Didalam *association rule* terdapat dua parameter yaitu *support* dan *confidence*. *Support* merupakan nilai penunjang yang ditunjukan dengan tingkat presentase *itemset* dari seluruh transaksi sedangkan *confidence* adalah nilai kepastian yang diperlihatkan dengan seberapa kuat relasi antar *itemset* dalam *association rule*[8]. Rumus dari *Support* dan *confidence* dapat diuraikan sebagai berikut :

$$\text{Support } A \cup B = \frac{A \cup B}{N} \quad (2.1)$$

$$\text{Confidence } A \cup B = \frac{A \cup B}{A} \quad (2.2)$$

2.5 Metode Algoritma *Apriori*

Metode algoritma *apriori* merupakan bagian dari metode *association rule* didalam *data mining*. Selain *apriori* ada 2 metode lain yang termasuk didalam golongan *association rule* yaitu metode algoritma *hash based* dan *generalized rule*[2].

Metode algoritma *apriori* adalah metode yang digunakan dalam mencari pola *high frequency*. Pola *high frequency* adalah sekumpulan dari pola *item* didalam database yang memiliki *frequency* atau *support* diatas batas yang telah ditentukan sering juga disebut dengan *minimum support* dan *threshold*. Pola *high frequency* tersebut digunakan untuk menentukan *association rule*[2].

Algoritma *apriori* terbagi menjadi beberapa proses *iteration*. Tiap *iteration* menciptakan beberapa pola *high frequency* yang memiliki panjang sama dimulai dari pertama dan akan membentuk pola *high frequency* dengan panjang satu. pada *iteration* pertama, *support* yang berasal dari setiap *item* akan dikalkulasikan dengan cara mengecek database. Setelah *support* dari setiap *item* ditemukan, *item* yang mempunyai *support* diatas *minimum support* akan ditunjuk menjadi pola *high frequency* dengan panjang satu atau dapat disingkat sebagai “1-*itemset*”. Sedangkan *k-itemset* memiliki arti satu *set* yang terisi didalam *k-item*.

Iteration kedua menciptakan 2-*itemset* pada setiap setnya mempunyai 2 *item*. Pertama dibuat kandidat 2-*itemset* ini dikalkulasikan *supportnya* dengan mengecek database. *Support* yang disebut didalam *iteration* kedua ini berarti jumlah transaksi yang berada didalam database terdapat kedua *item* didalam kandidat 2-*itemset*, setelah *support* akan diperoleh calon 2-*itemset*, calon 2-*itemset* yang telah mencukupi kriteria *minimum support* bisa diputuskan sebagai 2-*itemset* yang juga menjadi pola *high frequency* yang memiliki panjang dua.

Berikutnya *iteration* akan dibagi menjadi beberapa bagian :

1. Pembuatan Calon *itemset*

Dengan mengkombinasikan antara (k-1)-*itemset* yang diperoleh dari *iteration* terdahulu akan terbentuk *k-itemset*. Ciri dari metode algoritma *apriori* adalah

memiliki pemotongan calon *k-itemset* yang *subset*-nya berisi *k-q item* tidak masuk hitungan dalam pola *high frequency* yang memiliki panjang *k-1*.

2. Kalkulasi *support* calon *k-itemset*

Dengan melakukan cek database untuk mengkalkulasikan jumlah setiap transaksi yang memiliki *item* didalam calon *k-itemset* akan diperoleh *support* dari setiap calon *k-itemset*. Ini merupakan salah satu ciri dari metode algoritma *apriori* dimana kalkulasi *database* dilakukan dengan cara mengecek seluruh *database* dengan sejumlah *k-itemset* terpanjang.

3. Menetapkan *high frequency*

High frequency didalam *k-itemset* diputuskan melalui calon *k-itemset* dengan *support* lebih besar dari *minimum support*. Apabila tidak ada pola *high frequency* yang baru maka keseluruhan proses yang terjadi akan diberhentikan.

Jika tidak maka *k* akan ditambahkan satu dan kembali lagi menjadi bagian satu. Analisis *association* terkenal karena aplikasinya banyak digunakan untuk *market basket analysis* atau analisa keranjang pasar. Metodologi dasar dari analisis berikut terbagi menjadi dua tahap yaitu :

1. Analisis Pola *Highest Frequency*

Tahap ini akan ditelusuri kombinasi pada suatu *item* yang telah memenuhi kriteria *minimum* dari nilai *support* yang terdapat didalam *database*. Nilai *support* dalam sebuah *item* dapat dirumuskan sebagai berikut :

$$Support (A) = \frac{Total\ Transaksi\ MA}{\epsilon\ Transaksi} \quad (2.3)$$

Sementara itu, jika nilai *support* dari 2 *item* dirumuskan sebagai berikut.

$$Support (A \cup B) = \frac{Total\ Transaksi\ A\ dan\ B}{\epsilon\ Transaksi} \quad (2.4)$$

2. Pembentukan *Association Rule*

Jika semua pola *high frequency* telah ditemukan, selanjutnya dilakukan pencarian *association rule* yang telah memenuhi syarat *minimum* untuk *confidence* dengan mengkalkulasikan $A \rightarrow B$.

Nilai *confidence* pada aturan $A \rightarrow B$ dirumuskan sebagai berikut :

$$\text{Confidence}P(B|A) = \frac{\text{Total Transaksi A dan B}}{\epsilon A} \quad (2.5)$$

3. Lift Ratio

Lift digunakan untuk mengukur seberapa penting rule yang telah terbentuk berdasarkan nilai *support* dan *confidence*. *Lift ratio* memberikan nilai yang dapat menunjukkan validitas suatu proses transaksi dan memberikan informasi apakah kejadian dimana produk A dibeli bersamaan dengan produk B benar[9]. *Lift ratio* dapat dirumuskan sebagai berikut :

$$\text{Lift (A,B)} = \frac{P(AUB)}{P(A) \times P(B)} \quad (2.6)$$

2.6 Rapid Miner

RapidMiner adalah *software* yang bersifat *opensource*. RapidMiner merupakan sebuah terobosan dalam analisis yang dilakukan *prediction analysis*, *text mining* dan *data mining*. Dalam memberikan informasi kepada pengguna RapidMiner menggunakan teknik deskriptif dan prediksi sehingga nantinya akan didapat memberikan informasi berupa pengambilan keputusan yang paling efektif. RapidMiner kurang lebih memiliki 500 operator, termasuk didalamnya operator yang digunakan untuk *output*, *input*, *visualization*, dan *data preprocessing*. Software ini bersifat *stand alone* atau berdiri sendiri untuk analisis data dan mesin *data mining* yang dapat melakukan proses integrasi terhadap produknya sendiri. RapidMiner memiliki kemampuan untuk bekerja secara universal pada semua sistem operasi kemampuan ini dapat diraih karena basis dari RapidMiner adalah *java*[10].

2.7 NetBeans

NetBeans adalah sebuah aplikasi IDE yang digunakan oleh *software developer* komputer dalam menulis, meng-*compile*, mencari kesalahan, dan menjalankan

program. Netbeans mempunyai sekumpulan *software module* yang digunakan sebagai sarana untuk membuat suatu aplikasi. *Module* merupakan sebuah arsip *java* atau *java archive* yang berisikan *class Java* berfungsi sebagai sarana interaksi dengan Netbeans *Open API*[11].

Pembuatan program java dapat menggunakan *text editor* layaknya notepad++ dan aplikasi *text editor* lainnya. Dalam meng-*compile* dan menjalankan program dapat dilakukan melalui cmd atau *Command Prompt*. Penggunaan aplikasi netbeans akan memudahkan para *software developer* dalam melakukan coding program, *compile*, dan menjalankan program[11].

2.8 MySQL

MySQL merupakan *software* basis data berkemampuan *multi-thread* dan *multi-user* dengan *user* mencapai 6 juta yang tersebar diberbagai penjuru dunia. MySQL AB menciptakan MySQL menjadi sebuah *software* gratis yang berlisensi GNU *General Public Licence* (GPL), tetapi perusahaan tersebut menawarkan juga *software* ini dengan lisensi komersial bagi beberapa *user* yang merasa tidak sesuai dengan *software* berlisensi GPL[11].

Tidak seperti proyek Apache, yang dimana pengembangan *software* dan hak cipta kode sumber dimiliki oleh pembuatnya masing-masing. MySQL merupakan *software* yang berada dibawah naungan perusahaan komersial Swedia MySQL AB. Hampir dari semua hak cipta *source code* didalam MySQL dimiliki oleh perusahaan MySQL AB. MySQL AB didirikan oleh Michael “Monty” Widenius, David Axmark, dan Allan Larsson. [11].

2.9 SPMF

SPMF adalah sebuah aplikasi perangkat lunak *open source* yang berfungsi untuk melakukan pengolahan *data mining* yang menyediakan *library data mining* dan ditulis dalam pemrograman java. SPMF itu sendiri didistribusikan dibawah lisensi dari GPL v3 yang telah mengimplementasikan 115 algoritma untuk *association rule mining* dan *sequential pattern mining*[12].

2.10 Penggunaan Data Mining Dalam Perusahaan

Bisnis didalam perusahaan tidak hanya berjalan pada ruang tertutup. Oleh karena itu keberhasilan yang terdapat dalam menjalankan bisnis tidak hanya bergantung pada bagaimana cara menjalankan bisnis, akan tetapi ada pada bagaimana proses menjalankan bisnis tersebut bisa berbeda dari lainnya namun dapat memberi keuntungan yang lebih pada perusahaan tersebut. Kunci dalam membuat perbedaan tersebut ada pada penggunaan data yang tersimpan pada sistem yang digunakan untuk bisnis[13].

Data semakin lama akan semakin menumpuk seiring dengan terus bergeraknya aktivitas proses bisnis didalam perusahaan. Data ini terdiri atas sumber informasi yang berharga. Untuk mendapatkan informasi yang terkandung dalam data tersebut data sebelumnya harus diolah terlebih dahulu[13].

Setelah data diolah hasil yang didapat dari pengolahan data tersebut adalah kemampuan untuk memprediksi fakta yang terdapat pada lingkungan bisnis. Hal ini membuat perusahaan mampu melakukan tindakan proaktif dibandingkan hanya reaktif. Data memberikan pengetahuan tak ternilai tentang masa depan. Data dapat membantu perusahaan dalam mengoperasikan bisnis secara optimal, yaitu dengan cara melakukan pengembangan terhadap *market share*, meningkatkan *customer share*, dan membangun loyalitas pelanggan[13].

Dalam persaingan di dunia bisnis, khususnya industry otomotif, menuntut para manajer untuk menemukan suatu strategi jitu yang dapat meningkatkan penjualan kendaraan roda dua. Salah satu cara mengatasinya adalah dengan bagaimana cara memprediksi perilaku pelanggan dalam membeli kendaraan roda dua. Untuk dapat melakukan hal tersebut dibutuhkan sebuah data, data penjualan merupakan salah satu jenis data yang sangat cocok untuk diolah dalam meningkatkan pemasaran.

Permasalahan utamanya adalah bagaimana cara mendapatkan pengetahuan ini dari data yang dimiliki oleh perusahaan. Kuncinya yaitu *data mining*, yang menjadi teknik paling populer untuk membangun sistem pendukung keputusan[13]

Data mining memiliki banyak keuntungan yaitu dapat membantu perusahaan dalam mempercepat proses pengambilan keputusan, selain itu juga terdapat beberapa keuntungan yaitu perusahaan dapat mengelola informasi yang terkandung di dalam data transaksi menjadi sebuah pengetahuan (*knowledge*) yang baru. Lewat pengetahuan yang didapat, perusahaan dapat meningkatkan pendapatannya dan mengurangi biaya, dan pada akhirnya di masa yang akan datang perusahaan dapat lebih kompetitif [14].

Data mining tidak hanya sekedar publisitas yang berlebihan, tetapi *data mining* dapat menjadi sebuah langkah logis didalam pengumpulan pengetahuan yang merupakan model perusahaan paling strategis. Dalam melakukan langkah ini tidak berarti perusahaan harus melakukan langkah sebelumnya, tetapi dengan memperhatikan langkah-langkah tersebut akan membantu dalam persiapan kedepan. *Data mining* tidak akan mengubah sistem yang sudah ada menjadi *obsolete*, tetapi justru akan dapat memperbesar sistem dengan memperkaya data yang digunakan[13].

