

BAB II

LANDASAN TEORI

2.1. Tinjauan Pustaka

Pada Penelitian ini referensi yang digunakan sebagai landasan teori diperoleh dari berbagai media seperti jurnal, thesis, skripsi, dan buku. Adapun penelitian yang terkait dengan penelitian ini yaitu :

- a) Penelitian yang dilakukan oleh (Dewi Kartika Pane, 2013)

Penjelasan Penelitian :

Dalam penelitian ini menerapkan *data mining* pada penjualan produk elektronik dengan algoritma *apriori* (studi kasus : kreditplus). Dengan menggunakan data penjualan laptop yang terjual berdasarkan merk diperlukan algoritma *apriori* untuk dapat mengetahuinya, dan dengan bantuan *tools tanagra*, produk dengan penjualan terbanyak dapat diketahui dengan dua tolak ukur , yaitu : *support* dan *confidence*. *Support* (nilai penunjang) adalah persentase kombinasi item tersebut dalam *database*, sedangkan *confidence* (nilai kepastian) adalah kuatnya hubungan antar-item dalam aturan asosiasi. Jadi, berdasarkan aturan asosiasi diatas mengetahui merek produk elektronik yang paling banyak terjual pada toko adalah Acer dan Toshiba, dengan diketahuinya produk yang paling banyak terjual tersebut, sehingga perusahaan dapat menyusun strategi pemasaran untuk memasarkan produk dengan merek lain dengan meneliti apa kelebihan produk yang paling banyak terjual tersebut dengan produk lainnya dan dapat menambah persediaan Acer dan Toshiba. Algoritma *Apriori* dapat membantu mengembangkan strategi pemasaran toko dengan cara memberikan saran kepada konsumen. Pengimplementasian Algoritma *apriori* pada *tanagra* dimulai dengan penginputan data penjualan perbulan yang menjadi *database* pada Ms.Excel, semakin banyak data maka pembuatan tabel tabular akan semakin sulit. Tabel tabular tersebut yang kemudian dikoneksikan ke dalam *tools tanagra*, dan

mulailah pembentukan *support* dan *confidence*, kemudian akan menghasilkan asosiasi final yang memenuhi *support* dan *confidence*.

- b) Publikasi majalah ilmiah (Kennedi Tampubolon , Hoga Saragih, dan Bobby Reza 2013)

Penjelasan Publikasi ini :

Penelitian ini membahas tentang implementasi *data mining* menggunakan algoritma *apriori* dengan studi kasus pada Apotek Kelambir – 2 Medan dari bulan Oktober – November 2013 dan yang menjadi populasi penelitian ini adalah data-data transaksi penjualan alat-alat kesehatan. Pengujian yang dilakukan dalam penelitian ini menggunakan program aplikasi Microsoft Exel 2007 dan Tanagra 1.4.48. Sistem informasi persediaan alat-alat kesehatan di Apotek Kelambir – 2 Medan sudah *terkomputerisasi* artinya semua yang terkait dengan seluruh aktifitas penjualan dan pembelian menggunakan komputer yang berbasis jaringan dengan *database* yang hanya terpusat dalam satu *server*. Sistem informasi persediaan barang yang ada di Apotek Kelambir – 2 Medan terdiri dari bagian-bagian data obat dan alat-alat kesehatan, gudang farmasi, pelayanan resep obat dan penjualan. Yang ditampilkan dalam gudang farmasi merupakan bagian persediaan, laporan pembelian, laporan penjualan sedangkan data alat-alat kesehatan keluar terdiri atas *attribute* nomor, tanggal, nomor slip, nama, jumlah barang, harga, dan total harga. Setiap transaksi penjualan minimal terdiri atas satu jenis alat-alat kesehatan. Untuk mendapatkan suatu analisis data maka data penjualan di *export* kedalam *database* Microsoft excel karena *database* Microsoft excel bersifat *spreadsheet* sehingga sangat mendukung dalam analisis data. Microsoft excel dalam analisis data digunakan untuk *database* karena Microsoft excel sangat mendukung sebagian aplikasi *data mining*, dimana aplikasi (perangkat lunak) *data mining* digunakan sebagai tempat uji coba atau implementasi. Dari proses diatas disimpulkan bahwa algoritma *apriori*

dapat menemukan kecenderungan pola *itemset* hal ini dapat dijadikan sebuah informasi yang sangat berharga dalam sistem pendukung keputusan untuk mempersiapkan stok barang yang diperlukan di waktu yang akan datang.

- c) Penelitian yang dilakukan oleh (Almon Junior Simanjuntak, 2013)

Penjelasan Penelitian :

Penelitian dengan judul “Implementasi data mining untuk pemodelan pembelian barang dengan menggunakan algoritma apriori”. Dimulai dari pengolahan data lalu mendapatkan sebuah *itemset* yang *frequent* lengkap dengan *minimum support* dan *minimum confidence*, algoritma *apriori* dapat menghasilkan sebuah komposisi *itemset* dengan frekuensi tinggi yang terjadi pada jumlah data yang ada pada transaksi penjualan. sehingga dapat diketahui kombinasi barang yang sering dibeli pelanggan secara bersamaan dengan pengolahan data transaksi penjualan.

- d) Penelitian yang dilakukan oleh (A.A Gede Bagus Mariana dan I Made Dwi Putra Asana 2013)

Penjelasan Penelitian :

Penelitian ini menjelaskan tentang analisa keranjang belanja pada perusahaan *retail* dengan mengaplikasikan *data mining* dengan menggunakan metode algoritma *apriori* untuk menemukan pola frekuensi tinggi. untuk memproses informasi selanjutnya. Pada algoritma *apriori* untuk menentukan kandidat-kandidat yang mungkin muncul dengan cara memperhatikan *minimum support*. Adapun dua proses utama yang dilakukan dalam algoritma *apriori*, yaitu : *Join* (penggabungan). Pada proses ini setiap item dikombinasikan dengan item yang lainnya sampai tidak terbentuk kombinasi lagi dan *Prune* (pemangkasan). Pada proses ini, hasil dari item yang telah dikombinasikan tadi lalu dipangkas dengan menggunakan *minimum*

support yang telah ditentukan oleh peneliti. Hasil dari penelitian ini dapat menentukan kecenderungan kebiasaan pelanggan dalam membeli barang-barang secara bersamaan sehingga dengan mengetahui informasi pola kebiasaan belanja konsumen pihak perusahaan dapat melakukan beberapa tindakan untuk meningkatkan penjualan. Contoh tindakan yang dapat dilakukan seperti promosi harga untuk barang yang banyak diminati pelanggan, mengatur letak barang yang sering dibeli secara bersamaan dengan rak yang bersebelahan, dan menjaga stok barang yang sering dibeli untuk diperbanyak dan mengurangi stok barang yang kurang laku.

- e) Penelitian yang dilakukan oleh (Wijhah Islamika 2014)

Penjelasan Penelitian :

Dalam Penelitian ini Wijhah Islamika melakukan analisis asosiasi data dengan algoritma *apriori* sebagai sistem pendukung keputusan penjualan barang yang berbasis *web* dengan menghasilkan *association rule* pola *if-then* dimana *k-itemset* digunakan untuk mengeksplorasi *itemset* dengan melakukan pendekatan *iterative* yang dikenal dengan pencarian *level-wise*.

- f) Penelitian yang dilakukan oleh (Robi Yanto, Riri Khoiriah 2015)

Penjelasan Penelitian :

Penelitian ini melakukan analisis pola pembelian obat konsumen pada sebuah Apotik dengan pengolahan data terlebih dahulu melakukan identifikasi masalah yang ada dan sering dihadapi oleh pihak Apotik, untuk kemudian mendeskripsikan masalah-masalah tersebut untuk diperoleh solusinya. Perusahaan Apotik Musi Rawas membutuhkan suatu sistem yang dapat memberikan ilmu pengetahuan yang bisa membentuk level manajerial dalam mengambil sistem pendukung keputusan khususnya dalam penyediaan obat yang efektif dan efisien hal ini yaitu obat yang sering dibeli konsumen dari

penentuan pola pembelian obat, sehingga obat yang paling sering dibeli tersebut dapat menjadi acuan untuk mengembangkan strategi dalam pola penentuan pembelian obat. Dilakukan analisis masalah menggunakan teknik *data mining* dengan algoritma *apriori*. Diketahui pola frekuensi tinggi kombinasi item yang memenuhi syarat *minimum* dari nilai *support* dalam basis data setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiasi yang memenuhi syarat *minimum* untuk *confidence* dengan menghitung *confidence* aturan asosiatif dari hasil tersebut maka sistem ini dapat membantu pola pembelian obat berdasarkan kecenderungan obat yang dibeli konsumen yang terdiri dari 2 item obat, kemudian dari hasil pengujian ini juga dapat membantu pihak karyawan dalam mengatur tata letak obat yang terdiri dari 2 item obat secara berdekatan untuk memudahkan karyawan dalam mengetahui keberadaan obat.

2.1.1.State of the art

Tabel 2.1 *State of the art*

No	Peneliti	Tahun	Judul	Metode	Hasil
1	Dewi Kartika Pane	2013	Implementasi <i>Data Mining</i> pada Penjualan Produk Elektronik dengan Algoritma <i>Apriori</i> (Studi Kasus : Kreditplus)	Algoritma <i>Apriori</i>	Berdasarkan aturan asosiasi diketahui merek produk elektronik yang paling banyak terjual pada toko adalah Acer dan Toshiba, dengan diketahuinya produk yang paling banyak terjual tersebut, sehingga perusahaan dapat menyusun strategi pemasaran untuk memasarkan produk dengan merek lain dengan meneliti apa kelebihan produk yang paling banyak terjual tersebut dengan produk

No	Peneliti	Tahun	Judul	Metode	Hasil
					lainnya dan dapat menambah persediaan Acer dan Toshiba.
2	Kennedi Tampubolon, Hoga Saragih, Bobby reza	2013	Implementasi <i>Data Mining</i> Algoritma <i>Apriori</i> pada Sistem Persediaan alat-alat kesehatan	Algoritma <i>Apriori</i>	Algoritma <i>apriori</i> sangat efektif dan dapat menemukan kecenderungan pola dalam <i>itemset</i> hal ini dapat dijadikan sebuah informasi yang sangat berharga dalam sistem pendukung keputusan untuk mempersiapkan stok barang yang diperlukan pada Apotik Kelambir 2 Medan di waktu yang akan datang.
3	Almon Junior Simanjuntak	2013	Aplikasi <i>Data Mining</i> Untuk Pemodelan Pembelian barang dengan menggunakan Algoritma <i>Apriori</i>	Algoritma <i>Apriori</i>	Algoritma <i>apriori</i> dapat menghasilkan sebuah komposisi <i>itemset</i> dengan frekuensi tinggi yang terjadi pada jumlah data yang ada pada transaksi penjualan.
4	A.A Gede Bagus Mariana, I Made Dwi Putra Asana	2013	Analisis Keranjang Belanja dengan Algoritma <i>Apriori</i> pada Perusahaan Retail	Algoritma <i>Apriori</i>	Mengetahui informasi pola kebiasaan belanja konsumen pihak perusahaan dapat melakukan beberapa tindakan untuk meningkatkan penjualan. Contoh tindakan yang dapat dilakukan seperti promosi harga untuk

No	Peneliti	Tahun	Judul	Metode	Hasil
					barang yang banyak diminati pelanggan, mengatur letak barang yang sering dibeli secara bersamaan dengan rak yang bersebelahan, dan menjaga stok barang yang sering dibeli untuk diperbanyak dan mengurangi stok barang yang kurang laku.
5	Wijhah Islamika	2014	Analisis Asosiasi Data Dengan Algoritma <i>Apriori</i> Untuk Sistem Pendukung Keputusan Penjualan Barang Berbasis <i>Web</i>	Algoritma <i>Apriori</i>	Sistem pendukung keputusan penjualan barang yang berbasis <i>web</i> dengan menghasilkan <i>association rule</i> pola <i>if-then</i> dimana <i>k-itemset</i> digunakan untuk mengeksplorasi <i>itemset</i> dengan melakukan pendekatan <i>iterative</i> yang dikenal dengan pencarian <i>level-wise</i> .
6	Robi Yanto, Riri Khoiriah	2015	Implementasi <i>Data Mining</i> dengan Metode Algoritma <i>Apriori</i> dalam Menentukan Pola Pembelian Obat	Algoritma <i>Apriori</i>	Proses penentuan pola pembelian obat dapat dilakukan dengan menerapkan data mining dengan metode algoritma <i>apriori</i> . Sehingga dapat dilakukan pengaturan tata letak obat secara berdekatan untuk memudahkan keberadaan obat.

Setelah membaca jurnal-jurnal dan penelitian yang sudah dilakukan diatas sebagai studi literature dan penelusuran ilmiah, maka peneliti semakin mengetahui banyak hal mengenai *data mining* terutama bagaimana sebuah algoritma *apriori* bisa digunakan dalam mempelajari *market basket analisis* atau perilaku belanja konsumen dengan aturan asosiasi sebuah data bisa diolah sebagai sebuah informasi sehingga dapat bermanfaat untuk masa yang akan mendatang. Penelitian yang dilakukan yaitu melakukan analisis data transaksi pada sebuah swalayan Indomaret yang bertujuan untuk mengetahui pola pembelian konsumen atau dikenal juga dengan istilah *market basket analisis*, dari pola pembelian konsumen tersebut didapatkan sebuah item barang apa yang banyak dibeli oleh konsumen. Untuk itu peneliti memilih sebuah swalayan yang sangat populer yaitu Indomaret, adapun Indomaret yang akan diteliti merupakan sebuah Indomaret yang bertempat sangat strategis di kota Semarang tepatnya di jalan Indraprasta , peneliti mengambil data sampel yang kemudian dapat diolah dan dijadikan sebagai sebuah objek penelitian.

2.2. Analisis

Menurut KBBI (kamus besar bahasa Indonesia) analisis adalah kata benda yang berarti penyelidikan terhadap suatu peristiwa untuk mengetahui keadaan yang sebenarnya. Nosi lain untuk analisis adalah penjabaran sesudah dikaji sebaik-baiknya dan pemecahan persoalan yang dimulai dengan dugaan akan kebenarannya.

2.3. Pola Pembelian Konsumen

Pola adalah bentuk atau model (atau, lebih abstrak, suatu set peraturan) yang bisa dipakai untuk membuat atau untuk menghasilkan suatu atau bagian dari sesuatu, khususnya jika sesuatu yang ditimbulkan cukup mempunyai suatu yang sejenis untuk pola dasar yang dapat ditunjukkan atau terlihat, yang mana sesuatu itu dikatakan memamerkan pola.

Definisi Pembelian adalah Suatu peristiwa atau tindakan yang dilakukan oleh dua belah pihak dengan tujuan menukarkan barang atau jasa dengan menggunakan alat transaksi yang sah dan sama-sama memiliki kesepakatan dalam transaksinya, dalam pembelian terkadang akan terjadi tawar-menawar antara pembeli dan penjual hingga mendapatkan kesepakatan harga yang kemudian akan melakukan transaksi penukaran barang atau jasa dengan alat tukar yang sah dan disepakati kedua belah pihak.

Konsumen adalah setiap orang pemakai barang dan atau jasa yang tersedia dalam masyarakat, baik bagi kepentingan diri sendiri, keluarga, maupun orang lain.

Konsumsi, dari bahasa Belanda *consumptie*, ialah suatu kegiatan yang bertujuan mengurangi atau menghabiskan daya guna suatu benda, baik berupa barang maupun jasa, untuk memenuhi kebutuhan dan kepuasan secara langsung. Konsumen adalah setiap orang pemakai barang dan atau jasa yang tersedia dalam masyarakat, baik bagi kepentingan diri sendiri, keluarga, orang lain, maupun makhluk hidup lain dan tidak untuk diperdagangkan. Jika tujuan pembelian produk tersebut untuk dijual kembali (Jawa: kulakan), maka dia disebut pengecer atau distributor. Pada masa sekarang ini bukan suatu rahasia lagi bahwa sebenarnya konsumen adalah raja sebenarnya, oleh karena itu produsen yang memiliki prinsip *holistic marketing* sudah seharusnya memperhatikan semua yang menjadi hak-hak konsumen.

2.4. Data Mining

2.4.1 Pengertian Data Mining

Data mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data.” (Larose, 2006). “*Data mining* merupakan bidang dari beberapa keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola,

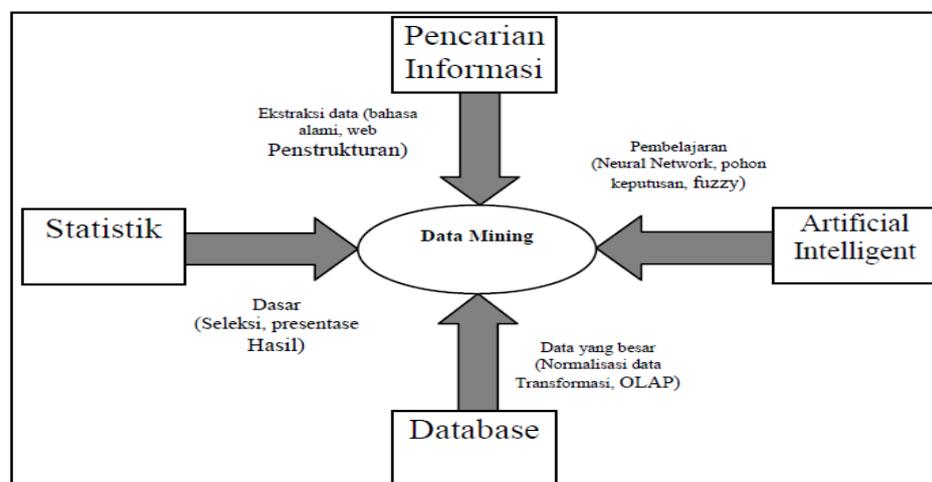
statistik, *database*, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari *database* yang besar.” (Larose, 2006).

Kemajuan luar biasa yang terus berlanjut dalam bidang *data mining* didorong oleh beberapa faktor, antara lain : (Larose, 2006)

1. Pertumbuhan yang cepat dalam kumpulan data.
2. Penyimpanan data dalam data *warehouse*, sehingga seluruh perusahaan memiliki akses ke dalam *database* yang baik.
3. Adanya peningkatan akses data melalui *navigasi web* dan *intranet*.
4. Tekanan kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi.
5. Perkembangan teknologi perangkat lunak untuk *data mining* (ketersediaan teknologi).
6. Perkembangan yang hebat dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan.

Berdasarkan definisi-definisi yang telah disampaikan, hal penting yang terkait dengan data mining adalah :

1. *Data mining* merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses berupa data yang sangat besar.
3. Tujuan *data mining* adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat.



Gambar 2.1 Bidang Ilmu *Data Mining*

Istilah *data mining* dan *Knowledge Discovery In Databases* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining. Proses KDD secara garis besar dapat dijelaskan sebagai berikut (Fayyad, 1996):

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, seperti kesalahan cetak (*tipografi*). Juga dilakukan proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.

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*

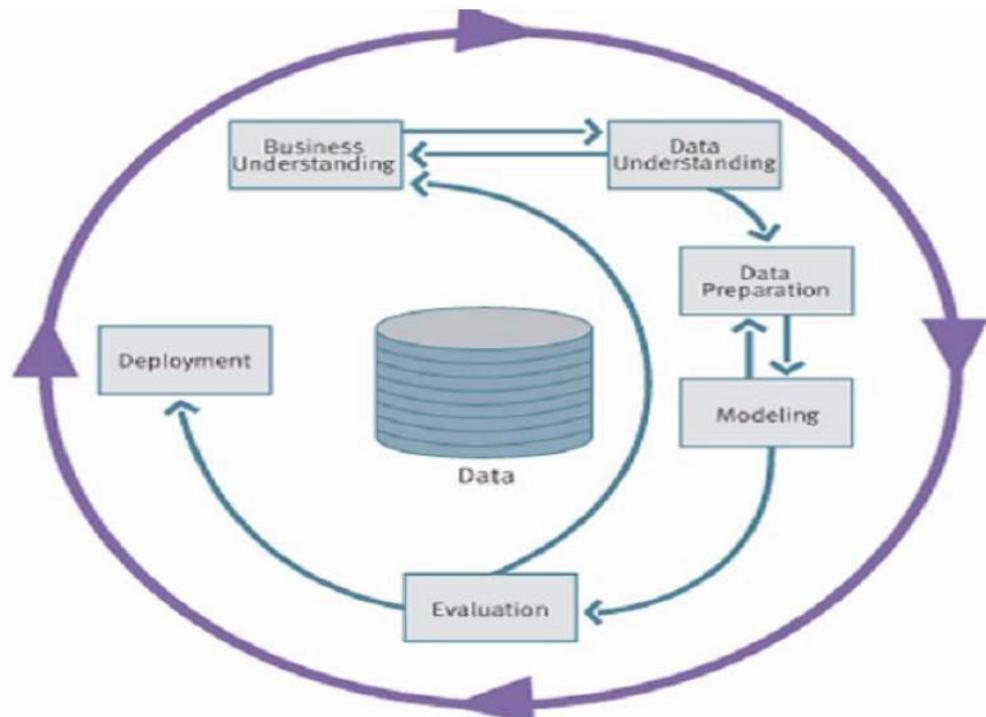
Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

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.

2.4.2 CRISP-DM

Cross-Industry Standard Process for Data Mining (CRISP-DM) yang dikembangkan tahun 1996 oleh analis dari beberapa industri menyediakan standar proses *data mining* sebagai strategi pemecahan masalah secara umum dari bisnis atau unit penelitian. Dalam CRISP-DM, sebuah proyek data mining memiliki siklus hidup yang terbagi dalam 6 (enam) fase. Keseluruhan fase berurutan yang ada tersebut bersifat adaptif dan fase berikutnya dalam urutan bergantung kepada keluaran dari fase sebelumnya. Hubungan penting antar fase digambarkan dengan panah. Sebagai contoh, jika proses berada pada fase modeling. Berdasar pada perilaku dan karakteristik model, proses mungkin harus kembali kepada fase data preparation untuk perbaikan lebih lanjut terhadap data atau berpindah maju kepada fase evaluation. Proses data mining menurut CRISP-DM dapat dilihat pada gambar 2.2.



Gambar 2.2 Data Mining menurut CRISP-DM

Dalam CRISP-DM, siklus hidup data mining yang terbagi dalam 6 (enam) fase yaitu:

1. Fase pemahaman bisnis (*Business Understanding Phase*)
 - a. Penentuan tujuan proyek dan kebutuhan secara detail dalam hidup bisnis atau unit penelitian.
 - b. Menerjemahkan tujuan dan batasan menjadi formula dari permasalahan data mining.
 - c. Menyiapkan strategi awal untuk mencapai tujuan.
2. Fase pemahaman data (*Data Understanding Phase*)
 - a. Mengumpulkan data.
 - b. Menggunakan analisis penyelidikan data untuk mengenali lebih lanjut data dan pencarian pengetahuan awal.
 - c. Mengevaluasi kualitas data.
 - d. Jika diinginkan, pilih sebagian kecil grup data yang mungkin mengandung pola dari permasalahan.

3. Fase pengolahan data (*Data Preparation Phase*)
 - a. Siapkan data awal, kumpulkan data yang akan digunakan untuk keseluruhan fase berikutnya. Fase ini merupakan pekerjaan berat yang perlu dilaksanakan secara intensif.
 - b. Pilih kasus dan variabel yang ingin dianalisis dan yang sesuai analisis yang akan dilakukan.
 - c. Lakukan perubahan pada beberapa variabel jika dibutuhkan.
 - d. Siapkan data awal sehingga siap untuk perangkat pemodelan.
4. Fase pemodelan (*Modeling Phase*)
 - a. Pilih dan aplikasikan teknik pemodelan yang sesuai.
 - b. Perlu diperhatikan bahwa beberapa teknik mungkin untuk digunakan pada permasalahan data mining yang sama.
 - c. Jika diperlukan, proses dapat kembali ke fase pengolahan data untuk menjadikan data ke dalam bentuk yang sesuai dengan spesifikasi kebutuhan teknik data mining tertentu.
5. Fase evaluasi (*Evaluation Phase*)
 - a. Pengevaluasi satu atau lebih model yang digunakan dalam fase pemodelan untuk mendapatkan kualitas dan efektivitas sebelum disebarkan untuk digunakan.
 - b. Menetapkan apakah terdapat model yang memenuhi tujuan pada fase awal.
 - c. Menentukan apakah terdapat permasalahan penting dari bisnis atau penelitian yang tidak tertangani dengan baik.
 - d. Mengambil keputusan yang berkaitan dengan penggunaan hasil dari data mining.
6. Fase penyebaran (*Deployment Phase*)
 - a. Menggunakan model yang dihasilkan. Terbentuknya model tidak menandakan telah terselesaikannya proyek.
 - b. Contoh sederhana penyebaran: pembuatan laporan.
 - c. Contoh kompleks penyebaran: penerapan proses data mining secara paralel pada departemen lain.

2.4.3 Pengelompokan Data Mining

Pengelompokan *Data Mining* berdasarkan tugas yang dapat dilakukan yaitu [9]:

1. Deskripsi

Terkadang peneliti dan analis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh, petugas pengumpulan suara mungkin tidak dapat mengumpulkan keterangan atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden. deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelesan untuk suatu pola atau kecenderungan.

2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih kearah numerik daripada ke arah kategori. Model dibangun dengan record lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Sebagai contoh, akan dilakukan estimasi tekanan darah sistolik pada pasien rumah sakit berdasarkan umur pasien, jenis kelamin, indeks berat badan, dan level sodium darah. Hubungan antara tekanan darah sistolik dan nilai variabel prediksi dalam proses pembelajaran akan menghasilkan model estimasi. Model estimasi yang dihasilkan dapat digunakan untuk kasus baru lainnya.

3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada dimasa mendatang.

Contoh prediksi dalam bisnis dan penelitian adalah:

Prediksi harga beras dalam tiga bulan yang akan datang. Prediksi persentase kenaikan kecelakaan lalu lintas tahun depan jika batas bawah dinaikan.

Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

4. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori yaitu: pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

Contoh lain klasifikasi dalam bisnis dan penelitian adalah:

Menentukan apakah suatu transaksi kartu kredit merupakan transaksi yang curang atau bukan. Memperkirakan apakah suatu pengajuan hipotek oleh nasabah merupakan suatu kredit yang baik atau buruk. mendiagnosis penyakit seorang pasien untuk mendapatkan termasuk kategori penyakit apa.

5. Pengklusteran

Pengklusteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lainnya dan tidak memiliki kemiripan dengan *record-record* dalam kluster lain.

Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklusteran. pengklusteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan record dalam suatu kelompok akan bernilai maksimal, sedangkan kemiripan dengan record dalam kelompok lain akan bernilai minimal.

Contoh pengklusteran dalam bisnis dan penelitian adalah:

Mendapatkan kelompok-kelompok konsumen untuk target pemasaran dari suatu produk sebuah perusahaan yang tidak memiliki dana pemasaran yang besar.

Untuk tujuan audit akuntansi, yaitu melakukan pemisahan terhadap perilaku finansial dalam baik maupun mencurigakan.

Melakukan pengklusteran terhadap ekspresi dari en, untuk mendapatkan kemiripan perilaku dari gen dalam jumlah besar.

6. Asosiasi

Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang pasar.

Contoh asosiasi dalam bisnis dan penelitian adalah:

- a. Meneliti jumlah pelanggan dari perusahaan telekomunikasi seluler yang diharapkan untuk memberikan respon positif terhadap penawaran upgrade layanan yang diberikan.
- b. Menemukan barang dalam supermarket yang dibeli secara bersamaan dan barang yang tidak pernah dibeli secara bersamaan.

2.5. Algoritma Apriori

apriori termasuk jenis aturan asosiasi pada *data mining*. Aturan yang menyatakan asosiasi antara beberapa atribut sering disebut *affinity analysis* atau *market basket analysis*. Analisis asosiasi atau *association rule mining* adalah teknik *data mining* untuk menemukan aturan asosiatif antara suatu kombinasi item. Contoh analisis asosiatif dari analisis pembelian di suatu pasar swalayan adalah dapat diketahuinya berapa besar kemungkinan seseorang pelanggan membeli roti bersamaan dengan susu. Dengan pengetahuan tersebut, pemilik pasar swalayan dapat mengatur penempatan barangnya atau merancang kampanye pemasaran dengan memakai kupon diskon untuk kombinasi barang tertentu. Karena analisis asosisasi menjadi terkenal karena aplikasinya untuk menganalisis isi keranjang di pasar

swalayan, analisis asosiasi juga sering disebut *market basket analysis*. (Kusrini dan Emha Taufiq Luthfi, 2009:149)

Analisis asosiasi dikenal juga sebagai salah satu teknik data mining yang menjadi dasar dari berbagai teknik data mining lainnya. Secara khusus, salah satu tahap analisis asosiasi yang menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien adalah analisis pola frekuensi tinggi (*frequent pattern mining*). Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, yaitu *support* dan *confidence*. *Support* (nilai penunjang) adalah persentase kombinasi item tersebut dalam *database*, sedangkan *confidence* (nilai kepastian) adalah kuatnya hubungan antar item dalam aturan asosiasi. Sebagai contoh, aturan asosiasi biasanya dinyatakan dalam bentuk: {roti, mentega} -> {susu} (*support* = 40%, *confidence* = 50%). Aturan tersebut berarti “50% dari transaksi di *database* yang memuat item roti dan mentega juga memuat item susu. Sedangkan 40% dari seluruh transaksi yang ada di *database* memuat ketiga item itu.” Dapat juga diartikan: “Seorang konsumen yang membeli roti dan mentega punya kemungkinan 50% untuk juga membeli susu. Aturan ini cukup signifikan karena mewakili 40% dari catatan transaksi selama ini.” Analisis asosiasi didefinisikan suatu proses untuk menemukan semua aturan asosiasi yang memenuhi syarat minimum untuk *support* (*minimum support*) dan syarat *minimum* untuk *confidence* (*minimum confidence*). (Kusrini dan Emha Taufiq Luthfi, 2009:150)

Algoritma *apriori* merupakan algoritma yang paling terkenal untuk menemukan pola frekuensi tinggi. algoritma *apriori* dibagi menjadi beberapa tahap yang disebut narasi atau pass (Efori Buulolo, 2013) yaitu:

1. Pembentukan kandidat *itemset*, kandidat *k-itemset* dibentuk dari kombinasi (k-1)-*itemset* yang didapat dari iterasi sebelumnya. Satu cara 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 menscan *database* untuk menghitung jumlah transaksi yang memuat semua item didalam kandidat *k-itemset* tersebut. Ini adalah juga ciri dari algoritma *apriori* dimana diperlukan penghitungan dengan cara 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 bagian 1.

Metode dasar analisis asosiasi terbagi menjadi dua tahap :

1. Pola frekuensi tinggi
2. Pembentukan Aturan Asosiasi

2.5.1 Pola frekuensi tinggi

Adalah tahap dimana pembentukan kombinasi antar item yang telah memenuhi syarat *minimum support* yang telah ditentukan sebelumnya. Nilai *support* pada suatu *itemset* dapat diperoleh dengan menggunakan sebuah rumus di bawah ini:

$$Support(X) = \frac{\sum \text{Transaksi yang mengandung } X}{\text{Total Transaksi}} \quad (1)$$

Rumus diatas memiliki arti untuk menentukan nilai *support* pada satu *item* jumlah transaksi yang mengandung item X dibagi dengan jumlah transaksi yang ada pada *database*.

Sedangkan pada dua *itemset* atau lebih diperoleh dengan rumus:

$$Support(X, Y) = \frac{\sum \text{Transaksi yang mengandung } X \text{ dan } Y}{\text{Total Transaksi}} \quad (2)$$

Pada rumus diatas untuk menentukan nilai *support* pada dua *itemset* atau lebih, jumlah transaksi yang mengandung item X dan Y dibagi dengan total transaksi yang terjadi pada *dataset*.

2.5.2 Pembentukan Aturan Asosiasi

Langkah yang dilakukan setelah menentukan nilai *support* pada *itemset* berfrekuensi tinggi lalu dibentuk aturan asosiasi yang menyatakan kuatnya hubungan kombinasi *itemset* pada transaksi. Untuk menentukan aturan asosiasi yang terbentuk minimal *itemset* harus memiliki dua kandidat X dan Y. Untuk menentukan aturan $X \rightarrow Y$ digunakan rumus:

$$\text{Confidence}(X \rightarrow Y) = \frac{\Sigma \text{Transaksi yang mengandung } X \text{ dan } Y}{\Sigma \text{Transaksi mengandung } X}$$

Pada rumus diatas untuk menentukan nilai *confidence* pada aturan asosiasi, jumlah transaksi yang mengandung item X dan Y dibagi dengan jumlah transaksi yang mengandung item X.

2.5.3 Join (Penggabungan)

Join merupakan proses pengkombinasian item dengan item yang lainnya sehingga tidak dapat terbentuk item lagi.

2.5.4 Prune (Pemangkasan)

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

2.6. Market Basket Analysis

Market Basket Analysis (MBA). Analisis ini merupakan salah satu metode dalam penambangan data (*data mining*) yang mempelajari tentang perilaku kebiasaan konsumen dalam membeli barang secara bersamaan dalam satu waktu. Metode analisis pola perilaku belanja MBA menggunakan bantuan algoritma *apriori*, yang merupakan algoritma MBA yang digunakan untuk menghasilkan *association rule*, dengan pola “if then ”[1]. *Market Basket Analysis* bertujuan untuk mengetahui hubungan antar produk yang dibeli oleh konsumen dalam satu kali transaksi. *Market basket analysis* merupakan salah satu bahasan dalam *data mining*. Dalam *Market basket analysis* dibutuhkan data transaksi yang cukup besar sehingga pola hubungan antar produk yang

didapat semakin valid. Pola hubungan antar produk ini berupa *Interesting Rules*. *Market basket analysis* memberikan informasi produk apa saja yang sering dibeli oleh konsumen secara bersamaan. Contohnya jika seorang konsumen membeli senter, maka ia juga akan membeli baterai. Informasi-informasi inilah yang digali dari data mining dengan *Market basket analysis*. Produk-produk yang sering dibeli secara bersamaan dapat ditempatkan secara berdekatan sehingga konsumen dapat dengan mudah menemukan apa yang dicari. Dengan demikian para konsumen akan merasa puas dan penjualan juga akan meningkat.

2.7. Swalayan Indomaret

Supermarket atau pasar swalayan adalah sebuah toko yang menjual segala kebutuhan sehari-hari. Kata yang secara harfiah yang diambil dari bahasa Inggris ini artinya adalah pasar yang besar. Barang-barang yang dijual di swalayan biasanya adalah barang-barang kebutuhan sehari-hari. Seperti bahan makanan, minuman, perlengkapan alat mandi, dan lain sebagainya.

Indomaret adalah jaringan peritel waralaba di Indonesia. Indomaret merupakan salah satu anak perusahaan Salim Group.

Indomaret merupakan jaringan minimarket yang menyediakan kebutuhan pokok dan kebutuhan sehari-hari dengan luas penjualan kurang dari 200 M2. Dikelola oleh PT. Indomarco Prismatama, cikal bakal pembukaan Indomaret di Kalimantan dan toko pertama dibuka di Ancol, Jakarta Utara, pada tahun 1988.

Tahun 1997 perusahaan mengembangkan bisnis gerai waralaba pertama di Indonesia, setelah Indomaret teruji dengan lebih dari 230 gerai. Hampir di setiap Kota di Indonesia ada cabang Indomaret karena barangnya yang lengkap, tempat mudah dijangkau, dan harganya yang murah banyak masyarakat yang berbelanja kebutuhan rumah tangga maupun sekedar beli makanan dan minuman di Indomaret.

Indomaret mudah ditemukan di daerah perumahan, gedung perkantoran dan fasilitas umum karena penempatan lokasi gerai didasarkan pada motto "mudah dan hemat".

Saat ini Indomaret berkembang sangat pesat dengan jumlah gerai mencapai lebih dari 5.000 di wilayah Jawa, Madura, Bali dan Sumatra, terdiri dari 40% gerai milik terwaralaba dan 60% gerai milik Perusahaan. Sebagian besar pasokan barang dagangan untuk seluruh gerai berasal dari 15 pusat distribusi Indomaret yang menyediakan lebih dari 4.800 jenis produk makanan dan non-makanan tersedia untuk memenuhi kebutuhan konsumen sehari-hari.(Wikipedia)

2.8. WEKA

Weka adalah aplikasi *data mining open source* berbasis Java. Aplikasi ini dikembangkan pertama kali oleh Universitas Waikato di Selandia Baru sebelum menjadi bagian dari Pentaho. Weka terdiri dari koleksi algoritma *machine learning* yang dapat digunakan untuk melakukan generalisasi atau formulasi dari sekumpulan data sampling. Weka memiliki metode dan implementasi algoritma yang cukup bervariasi, Algoritma *clustering* yang tersedia dalam Weka adalah COBWEB,EM,K-Means,Farthest First, sedangkan untuk mencari *association rule* adalah *apriori*, *Predictive Apriori*, dan *Tirtius*. Tugas pengumpulan data yang berkualitas tinggi dan pengetahuan pemodelan dan penggunaan algoritma yang tepat diperlukan untuk menjamin keakuratan formulasi yang diharapkan.

2.8.1 Fitur-fitur Weka

a. *Explorer*

Explorer adalah modul utama untuk memvisualisasikan dan *preprocessing* masukan data dan algoritma mesin menerapkan belajar untuk itu.

b. *Loading Data*

Data biasanya disimpan dalam *spreadsheet* atau *database* dan juga disebut *Dataset*. *Dataset* Setiap terdiri dari kasus, yang diwakili

oleh baris dalam *spreadsheet* atau tabel *database*. Data asli format penyimpanan dari WEKA adalah ARFF (*Attribute-Relation File Format*). Ini terdiri dari bagian *header* dan data. Bagian pertama berisi metadata yang menggambarkan kedua. Ini terdiri dari atribut semua contoh dan mereka jenis. Bagian kedua terdiri dari nilai atribut dipisahkan dengan koma. . Semuanya dapat ditemukan dalam paket `weka.core.converters`, yang dapat diperpanjang untuk lebih. Data dapat juga di-load dari database menggunakan JDBC dan dari URL.

c. *Preprocessing Data*

Setelah data *di-load*, yang terlihat di panel '*preproses*' dari *Explorer*.

Ringkasan statistik yang tersedia untuk setiap atribut dari dataset. Jika atribut adalah nominal distribusi kasus menurut atribut nilai ditampilkan. Jika atribut numerik minimum, maksimum, berarti dan deviasi standar diberikan. Selama operasi dataset pengeditan sederhana, seperti mengedit nilai tunggal untuk contoh beton dan kolom menghapus untuk semua kasus, dapat dilakukan dengan tangan. Operasi otomatis dapat dilakukan dengan filter. Biasanya format data kebutuhan untuk diubah karena berbagai alasan tergantung pada mesin belajar skema yang akan digunakan. Sebagai contoh algoritma pembelajaran mesin mungkin hanya menerima nilai-nilai numerik dari atribut, sehingga semua atribut non-numerik harus diubah agar ini algoritma yang akan digunakan. Setelah memilih filter yang tepat dapat diterapkan pada awal *dataset*. Hasil transformasi ini ditampilkan di panel '*preproses*'. Transformasi berturut-turut dapat diterapkan dalam *preprocessing* tambahan kasus diperlukan. *Dataset* berubah juga dapat disimpan sebagai file.

d. *Building classifiers*

Setelah *dataset input* diubah dalam format yang cocok untuk

mesin skema pembelajaran, dapat diberi makan ke sana. Membangun atau melatih sebuah *classifier* adalah proses menciptakan struktur fungsi atau data yang akan digunakan untuk menentukan nilai yang hilang dari atribut kelas dari unclassified baru contoh. *Classifier* beton dapat dipilih dari panel ‘Klasifikasi’ dari *Explorer*. Ada pengklasifikasi banyak tersedia. Masing-masing memiliki deskripsi tentang cara kerjanya dan referensi untuk semua parameter yang digunakannya. Sebagian besar pengklasifikasi dijelaskan secara rinci dalam buku ini tapi karena ada yang versi baru pengklasifikasi WEKA baru diimplementasikan dan dapat dipilih.

e. Association Rules

Ada beberapa asosiasi aturan algoritma diimplementasikan dalam WEKA. Mereka mencoba untuk menemukan hubungan antara atribut yang berbeda daripada mencoba untuk memprediksi nilai dari atribut class. Antarmuka untuk memilih dan mengkonfigurasi mereka adalah sama seperti untuk filter dan pengklasifikasi. Tidak ada pilihan untuk memilih tes dan set pelatihan. Hasil ditampilkan dalam panel keluaran cukup mirip ini dihasilkan setelah membangun *classifier*.

f. Clustering

Ada algoritma *clustering* sembilan diterapkan di WEKA. Mereka juga melakukan tidak mencoba untuk memprediksi nilai atribut class tapi untuk membagi pelatihan diatur ke dalam kelompok. Semua contoh dalam satu kelompok yang dekat, menurut sebuah sesuai metrik, untuk semua kasus di kelompok yang sama dan jauh dari kasus dalam kelompok lain. Antarmuka untuk memilih dan mengkonfigurasi mereka adalah sama seperti untuk filter dan pengklasifikasi. Ada pilihan untuk memilih tes dan set pelatihan. Hasil ditampilkan dalam panel keluaran cukup mirip ini dihasilkan setelah membangun *classifier*.