

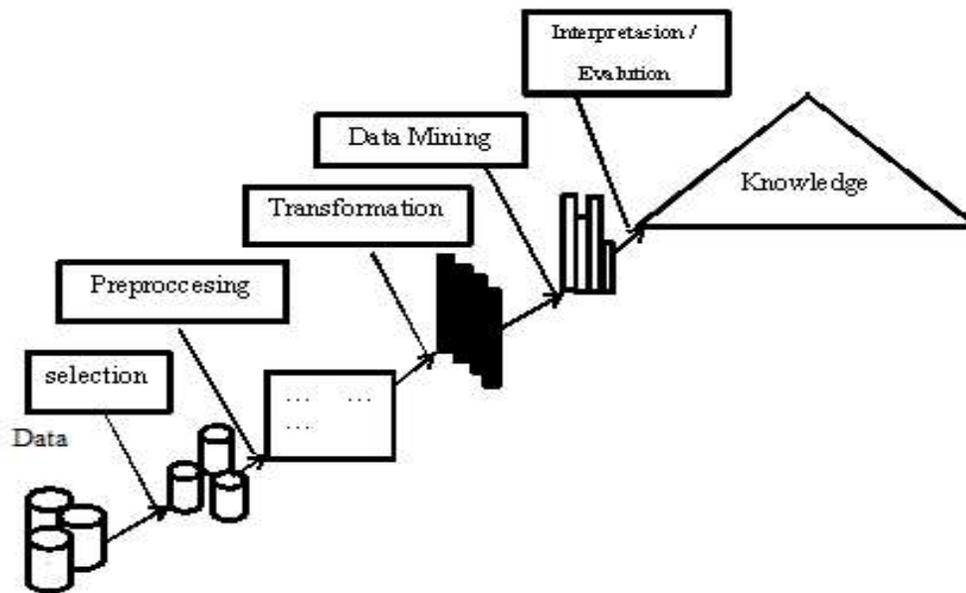
BAB II

TINJAUAN PUSTAKA

2.1 *Knowledge Discovery In Database (KDD)*

Knowledge Discovery In Database (KDD) merupakan proses pemisahan informasi tersembunyi, tersirat dan yang tidak diketahui dari sekumpulan data. Proses *knowledge discovery in database* melibatkan *output* dari proses *data mining* (bagian pemisahan dominan suatu pola data), lalu mengkonversikan *outputnya* menjadi informasi yang mudah dimengerti (Yulia & Azwanti, 2018)

Istilah *data mining* dan *knowledge discovery in database (KDD)* kerap dipakai secara bergantian dalam mengartikan cara penggalian informasi tersembunyi pada suatu basis data yang besar. Sesungguhnya sebutan kedua hal tersebut mempunyai konsep yang tidak sama, tetapi berkaitan satu dengan lainnya. *Data mining* merupakan salah satu tahap pada pada semua proses KDD . Proses KDD secara garis besar akan jelaskan seperti di bawah ini :



Gambar 2.1 Tahapan *Knowledge Discovery In Database* (KDD)

Sumber : (Retno Tri Vlandari, 2017, hal : 3)

1. *Data Selection*

Sebelum tahap pencarian informasi dalam KDD dimulai perlu dilakukan pemilihan ataupun penyeleksian data operasional dari data yang besar. Data hasil penyeleksian akan digunakan dalam proses *data mining*, disimpan dalam suatu berkas yang terpisah dari basis data yang operasional yang besar.

2. *Pre-processing/cleaning*

Proses *cleaning* pada data yang menjadi fokus KDD dilaksanakan terlebih dahulu lalu setelahnya baru proses *data mining* bisa dilakukan. Proses pembersihan tersebut meliputi menghapus data yang serupa ataupun ganda, memeriksa data yang tidak sesuai / konsekuen dan memperbaiki kesalahan pada data seperti kesalahan cetak (*tipografi*). Proses yang juga harus

dilaksanakan yaitu proses *enrichment* dimana yang berarti proses “memperkaya” data yang telah ada atau informasi lain sesuai dengan kebutuhan KDD. Contohnya data dan informasi eksternal.

3. *Transformation*

Coding ialah proses perubahan untuk data yang akan digunakan, yang akhirnya data yang dipakai konsekuen untuk proses *data mining*. Proses tersebut merupakan proses inovatif dan sangat tergantung pada pola serta jenis informasi yang diseleksi pada *database*.

4. *Data mining*

Proses *data mining* merupakan proses menemukan pola dan informasi yang menarik dari data yang dipilih bisa menggunakan teknik ataupun model tertentu. Teknik, metode dan algoritma dalam *data mining* sangat beragam. Tujuan dan proses KDD secara keseluruhan menjadi tolak ukur dalam pemilihan metode maupun algoritma yang akan digunakan secara tepat.

5. *Interpretation / Evaluation*

Tahap ini berhubungan dengan pemeriksaan apakah pola dan informasi yang dihasilkan berbeda dengan fakta atau hipotesis yang telah ada sebelumnya atau disebut juga tahap *interpretation* dan pada tahap ini harus mudah dimengerti.

6. Presentasi Pengetahuan (Knowledge)

Pada tahapan ini menggunakan metode Visualisasi

2.2 *Data Mining*

2.2.1 Pengertian *Data Mining*

Data mining yaitu teknik analisa pada data unntuk mencari informasi yang tersembunyi pada beberapa data besar yang dicatat pada saat bisnis perusahaan dijalankan. *Data mining* adalah teknik menemukan pola-pola yang tersembunyi (*hidden patern*) berupa pengetahuan (*knowledge*) yang mana sebelumnya tidak diketahui dari sekelompok data dimana data tersebut bisa terdapat dalam *database*, *data warehouse*, atau media penyimpanan informasi yang lain (Rahmawati & Merlina, 2018)

Para pakar juga mendefinisikan *data mining* yang dituangkan kedalam buku yang dikarangnya, yaitu: *Data mining* ialah suatu teknik yang digunakan dalam menelusuri nilai tambah berbentuk informasi yang tidak diketahui secara manual dari suatu basis data (Retno Tri Vulandari, 2017, hal : 2). *Data mining* adalah suatu sebutan yang dipakai untuk menjabarkan suatu penemuan pengetahuan di dalam *database*. *Data mining* adalah proses yang memakai teknik statistik, matematika, kecerdasan buatan, dan *machine learning* yang saling terkait dari beragam *database* besar (Kusrini dan Emha Taufiq Luthfi, 2009, hal : 3)

Data mining adalah metode analisa terarah pada cara penemuan pengetahuan pada basisdata atau *knowledge discovery in database* (KDD). Pengetahuan berupa data atau hubungan antar data valid (yang sebelumnya tidak diketahui) (Dr.Suyanto, 2017, hal : 1). Berlandaskan penjelasan *data mining* yang telah diuraikan sehingga bisa ditarik kesimpulan bahwasanya *data mining* itu ialah

suatu cara, teknik ataupun metode yang digunakan dalam mencari ataupun menggali data pada *database* besar dengan output informasi sebagai nilai tambah yang berguna bagi pemilik data.

2.2.2 Kegunaan *Data Mining*

Secara umum kegunaan *data mining* dibagi menjadi dua: deskriptif dan prediktif. Deskriptif berarti *data mining* dipakai untuk menemukan pola – pola yang mudah dimengerti manusia yang menjelaskan karakteristik data. Sedangkan prediktif berarti *data mining* dipakai untuk susunan model pengetahuan yang akan digunakan dalam prediksi (Dr.Suyanto,2017, hal : 3). Berdasarkan fungsionalitasnya, fungsi-fungsi *data mining* bisa dikelompokkan pada enam kelompok berikut ini :

- a) Klasifikasi (*classification*): menyamaratakan pola yang sudah diketahui untuk diaplikasikan pada data-data baru
- b) Klasterisasi (*clustering*): pengolongan data yang tidak diketahui nama kelasnya kedalam sejumlah kelompok tertentu sesuai dengan besar keserupaanya.
- c) Regresi (*regression*): fungsi yang ditemukan memodelkan data dengan kesalahan prediksi yang sangat kecil.
- d) Deteksi anomaly (*anomaly detection*): mengenali data yang tidak umum, seperti transformasi atau deviasi yang mungkin sangat penting dan perlu dilakukan investigasi selanjutnya.
- e) Pembelajaran aturan asosiasi (*association rule learning*) atau pemodelan

kebergantungan (*dependency modeling*): menemukan hubungan antar variabel.

- f) Perangkuman (*summarization*): menyediakan representasi data yang lebih simpel, berupa visualisasi dan pembuatan laporan.

2.2.3 Tahapan *Data Mining*

Data mining tidak dapat dipisahkan dari rangkaian *Knowledge in database* (KDD, hal ini karena data *mining* merupakan salah satu rangkaian dari *Knowledge in database* KDD). Serangkaian proses tersebut memiliki tahap sebagai berikut (Retno Tri Vulandari, 2017, hal : 2-3):

1. Pembersihan data (*data cleaning*)

Pembersihan diperlukan untuk membuang data yang tidak konsisten dan noise. Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performansi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

2. Integrasi data (*data integration*)

Integrasi data merupakan penggabungan data dari berbagai sumber. data yang kita peroleh biasanya tidak hanya dari *database* yang sama tapi bisa dari beberapa *database*.

3. Transformasi data (*data transformation*)

Transformasi data yaitu data diubah menjadi bentuk yang sesuai untuk untuk di-*mining*. Beberapa metode *data mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan.

4. Aplikasi Teknik *Data Mining*

Aplikasi teknik *data mining* yaitu proses ekstraksi pola dari data yang ada.

5. Evaluasi pola yang ditemukan (*Pattern Evaluation*)

Proses interpretasi pola bisa dipakai sebagai pengetahuan yang bisa dimanfaatkan sebagai pendukung pengambilan keputusan. Dalam tahap ini hasil dari teknik *data mining* berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai.

6. Presentasi pengetahuan (*Knowledge Presentasion*)

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

2.2.4 Pengelompokan *Data Mining*

Berdasarkan tugas yang dilakukan, *data mining* dibagi menjadi beberapa golongan, yaitu (Kusrini dan Emha Taufiq Luthfi, 2009, hal : 10-12):

1. Deskripsi

Adakalanya peneliti dan analis secara simpel hendak mencoba cara untuk menggambarkan pola dan kecenderungan yang ada dalam data. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan

2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih berorientasi pada numerik dari pada kearah kategori. Model estimasi yang dihasilkan dapat digunakan untuk kasus barunya.

3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan datang dimasa mendatang. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

4. Klasifikasi

Pada klasifikasi, terdapat target variabel kategori. Sebagai contoh, pengelompokan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

5. Pengklusteran

Merupakan pengelompokan *record*, observasi, atau memperhatikan dan membentuk kelas objek – objek yang menyerupai. Kluster adalah gabungan

record yang mempunyai kemiripan satu dengan yang lainya dan memiliki ketidakmiripan dengan *record – record* pada kluster lain.

6. Asosiasi

Tugas asosiasi dalam *data mining* adalah mendapatkan atribut yang tampak dalam satu waktu. Analisa keranjang belanja merupakan sebutan dalam dunia bisnis pada umumnya.

2.2.5 Teknik Pembelajaran *Data Mining*

Teknik yang dipakai pada *data mining* akrab hubunganya dengan “penemuan” (*discovery*) dan “pembelajaran” (*learning*) yang dibagi ketiga metode utama dalam pembelajaran yaitu (Retno Tri Vulandari, 2017, hal : 7-8) :

a. *Supervised Learning*

Supervised Learning ialah teknik yang sering dipakai. Teknik ini serupa dengan “*programming by example*”. Teknik ini mengaitkan *fase* pelatihan yang mana data pelatihan historis yang karakteristik-karakteristiknya bisa dipetakan ke *output* yang sudah diketahui, mengolah ke algoritma *data mining*. Proses tersebut melatih algoritma dalam mengidentifikasi variabel-variabel dan nilai-nilai kunci yang kelak akan dipakai untuk dasar pada pembuatan dugaan-dugaan saat ada data baru.

b. *Unsupervised Learning*

Teknik pembelajaran ini berlawanan arah dengan *supervised learning*. Teknik ini bergantung pada penggunaan algoritma yang mendeteksi semua pola,

seperti *associations* dan *sequences*, yang muncul dari kriteria penting yang spesifik dalam data masukan. Pendekatan ini mengarah pada pembuatan banyak aturan (*rules*) yang mengkarakterisasikan penemuan *associations*, *clusters*, dan *segments*. Beberapa aturan tersebut akan dianalisa untuk mendapatkan sesuatu yang penting.

c. *Reinforcement Learning*

Teknik pembelajaran sedikit lebih minim dipakai dibandingkan teknik pembelajaran di atas, namun mempunyai pemakaian yang terus ditingkatkan dari waktu ke waktu dan memiliki kontrol adaptif. Teknik ini hampir sama dengan kehidupan nyata yakni seperti “*on-job-training*”, dimana seorang pekerja diberikan setumpuk tugas yang membutuhkan keputusan. Pada beberapa titik waktu kelak diberikan penilaian atas *performance* pekerja tersebut kemudian pekerja diminta mengevaluasi keputusan-keputusan yang telah dibuatnya sehubungan dengan hasil *performance* pekerja tersebut. *Reinforcement learning* sangat tepat penggunaannya pada masalah yang rumit yang bersangkutan dengan waktu.

2.2.6 Penerapan *Data Mining*

Penerapan *data mining* dapat dilakukan pada (Retno Tri Vlandari, 2017, hal : 5-7):

1. Analisa pasar dan manajemen

Transaksi kartu kredit, kupon diskon, keluhan pembeli, studi tentang gaya hidup publik ditambah kartu anggota club tertentu adalah sumber data yang sering dipakai. *Data mining* dapat menyelesaikan beberapa solusi, diantaranya yaitu:

a) Menembak target pasar

Model-model konsumen dan membuat kategori pada setiap pemesan sesuai dengan kategori yang diharapkan contohnya kedukaan yang serupa, tingkat penghasilan yang sama, etiket membeli maupun karakteristik yang lain dapat dikelompokkan (*clustering*) memanfaatkan *data mining*.

b) Melihat pola beli pemakai dari waktu ke waktu

Data mining bisa dipakai guna melihat sistem pembelian dari waktu ke waktu selanjutnya. Misalnya, pada saat seseorang *merrid* bisa jadi dia selanjutnya memilih pindah dari *single account* ke *joint account*.

c) *Cross market analysis*

Data mining dapat dimanfaatkan untuk melihat hubungan antara penjualan satu produk dengan beberapa produk yang lain.

d) *Profil customer*

Data mining dapat melihat *profil customer* sehingga dapat diketahui kategori *customer* dengan pembelian produk tertentu yang sering dibeli.

e) Identifikasi kebutuhan *customer*

Dapat mengidentifikasi produk mana saja yang paling bagus untuk setiap kalangan *customer* dan faktor apa saja yang mampu menjadi daya tarik bagi pembeli.

f) Menilai *loyalitas customer*

g) Informasi *summary*

Dapat dimanfaatkan dalam menyusun laporan *summary* yang bersifat multi dimensi dan difasilitasi dengan informasi *statistic* lainnya.

2. Analisa perusahaan dan manajemen resiko

a) Perencanaan keuangan dan evaluasi aset

Data mining dapat dimanfaatkan dalam hal analisi dan prediksi *cash flow* juga untuk melakukan *contingent claim analysis* dalam mengevaluasi aset serta mampu menganalisis *trend*.

b) Perencanaan sumber daya

Dari sekumpulan ringkasan informasi juga pola pembelian dan pemasukan dari masing-masing *resource*. Maka dapat digunakan untuk *resource planning*.

c) Persaingan

Memanfaatkan *data mining* dalam hal memantau para pesaing bisnis dengan melihat *market direction* mereka. *Data mining* juga mampu mengelompokkan *customer* dan dapat memberikan variasi harga untuk setiap grup.

3. Telekomunikasi

Data mining dapat menemukan jutaan transaksi yang masuk, dan mengetahui transaksi mana sajakah yang masih harus ditangani secara manual, sasaranya yakni menambah layanan otomatis.

2.3 Metode *Data mining*

2.3.1 Algoritma C.45

Algoritma C.45 yaitu algoritma yang dipakai dalam membentuk pohon keputusan. Pohon keputusan dibentuk oleh algoritma c.45 melalui proses berikut (Kusrini dan Emha Taufiq Luthfi, 2009, hal : 13-91) :

- a. memilih atribut sebagai akar
- b. lalu pilih cabang untuk setiap nilai
- c. membagi kasus dalam cabang
- d. ulangi untuk proses pada setiap cabang hingga semua kasus pada tiap cabang hingga memiliki kelas yang sama.

2.3.2 Algoritma *Nearest Neighbor*

Pendekatan dalam mencari masalah dengan menghitung kedekatan antar kasus baru dengan kasus lama, yaitu berpusat dengan kesamaan berat dari sejumlah fitur yang sudah ada. Adapun rumus untuk melakukan kedekatan antar dua kasus adalah sebagai berikut (Kusrini dan EmhaTaufiq Luthfi, 2009, hal:93-148):

$$\text{similarity}(T,S) = \frac{\sum_{i=1}^n f_i(S_i) * w_i}{W_i}$$

Rumus 2.1 Kedekatan antar 2 kasus

Keterangan :

T : Kasus Baru

S : Kasus yang ada pada penyimpanan

n : Jumlah atribut disetiap kasus

i : Atribut individu diantara 1 s.d. n

f : Fungsi *similarity* atribut i diantara kasus T dan juga kasus S

w : Bobot yang diperkenankan pada atribut ke-i

Biasanya untuk kedekatannya berada dinilai antara 0 s.d. 1. Nilai 0 memiliki arti kedua kasus mutlak tidak mempunyai kemiripan, kebalikannya untuk nilai 1 kasus mirip dengan mutlak.

2.3.3 Algoritma *Apriori*

Salah satu jenis aturan yang ada dalam asosiasi pada *data mining* ialah algoritma *apriori* . Selain *apriori*, yang terkandung dalam kelompok ini ialah metode *Generalized Rule Induction* dan juga *Algoritma Hash Based* (Kusrini dan EmhaTaufiq Luthfi, 2009, hal : 149-176)

Aturan yang mengemukakan asosiasi diantara beberapa atribut dikenal juga dengan sebutan *affinity analysis* atau *market basket analysis*. Analisis asosiasi ataupun *association rule mining* ialah metode *data mining* dalam menemukan aturan asosiatif diantara suatu kombinasi *item*. Untuk analisis asosiasi banyak dikenal dikarenakan aplikasinya dalam analisis isi keranjang belanja di pasar swalayan, analisis asosiasi jugadikenal dengan *market basket analysis*.

Analisis asosiasi kerap kali disebut menjadi salah satu teknik *data mining* yang jadi pijakan untuk beberapa teknik *data mining* yang lain. Secara eksklusif, satu Dari banyaknya tahapan analisis asosiasi yang menjadi daya tarik perhatian beberapa peneliti dalam menciptakan algoritma yang efisien yaitu menganalisis pola frekuensi tinggi (*frequent pattern mining*). Bernilai maupun tidak bernilainya suatu aturan asosiatif akan diketahui menggunakan dua parameter, yakni satu disebut *support* dan dua disebut dengan *confidence*. *Support* (nilai penunjang) yaitu persentase kombinasi *item* pada *database*, sedang untuk *confidence* (nilai kepastian) ialah kuatnya relasi antar-*item* pada aturan asosiasi.

Analisis asosiasi ialah suatu proses dalam menemukan seluruh aturan asosiasi yang memenuhi syarat minimum untuk *support* (*minimum support*) dan syarat minimum untuk *confidence* (*minimum confidence*).

Metodologi dasar analisis asosiasi terbagi menjadi dua tahap:

1. Analisis pola frekuensi tinggi

Berfungsi menemukan kombinasi *item* yang mencukupi syarat minimum dari nilai *support* yang diberikan pada *database*. Nilai *support* sebuah *item* dihasilkan menggunakan rumus di bawah :

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

Rumus 2.2 Nilai *Support*

2. Pembentukan Aturan Asosiasi

Setelah seluruh pola frekuensi tinggi diperoleh, kemudian mencari aturan asosiasi yang mencukupi syarat minimum untuk *confidence* menggunakan cara dengan menghitung *confidence* aturan asosiatif $A \rightarrow B$.

Sementara itu, nilai *support* dari 2-item didapat dengan rumus 2.3 berikut.

$$Support = (A,B) P(A \cap B)$$

$$Support (A,B) = \frac{\sum \text{Transaksi mengandung A dan B}}{\sum \text{Transaksi}}$$

Rumus 2.3 Nilai *Support* dari 2 item

Nilai *confidence* dari aturan $A \rightarrow B$ diperoleh dari rumus berikut :

$$confidence = P(B|A) = \frac{\sum \text{Transaksi mengandung A dan B}}{\sum \text{Transaksi mengandung A}}$$

Rumus 2.4 Nilai *Confidence*

2.3.4 Fuzzy C-Means

Analisis kluster yaitu cara membagi data pada suatu himpunan kedalam sebagian himpunan yang kecocokan datanya dalam suatu himpunan lebih banyak daripada kecocokan data tersebut dengan data pada himpunan lain. Kluster secara lazim yaitu wujud kelompok bagian dari suatu kelompok data dan metode clustering dapat dikategorisasikan berdasarkan kelompok bagian yang dihasilkan: apakah *fuzzy* atau *crisp (hard)*. Terdapat banyak algoritma yang dipakai pada

clustering. *Fuzzy C-Means* (FMC) merupakan salah satunya. *Fuzzy C-Means* (FMC) ialah algoritma *clustering* data dimana semua data dapat dijadikan bagian dari suatu kluster sambil level yang diartikan sama dengan derajat keanggotaan (Kusrini dan Emha Taufiq Luthfi, 2009, hal : 177-187).

2.3.5 Bayesian Classification

Bayes classification yaitu pengelompokan statistik yang dipakai dalam memperediksi probabilitas keanggotaan suatu *class*. klasifikasi *bayes* berdasarkan pada teorema *bayes* dan terbukti mempunyai akurasi dan kecepatan tinggi dalam pengaplikasiannya pada *database* dengan data yang besar (Kusrini dan EmhaTaufiq Luthfi, 2009, hal : 189-198).

Dimana bentuk umum *teorema bayes*, dapat dilihat di bawah ini :

$$P(H|X) = \frac{P(X|H) P(H)}{P(X)}$$

Rumus 2.5 Probabilitas *Bayes*

Dalam hal ini:

X = data dengan class yang belum diketahui

H = hipotesis data X merupakan suatu *class* spesifik

P(H|X) = probabilitas hipotesis H berdasar kondisi X (*posteriori probability*)

P(H) = probabilitas hipotesis H (*prior probability*)

P(X|H) = probalitas X berdasarkan kondisi pada hipotesis H

P(X) = probabilitas dari X

2.3.6 *Backpropagation*

Backpropagation adalah algoritma yang terdapat pada pembelajaran jaringan syaraf tiruan. Proses pembelajaran ini dilaksanakan sesuai bobot- bobot jaringan syaraf tiruan dengan arah mundur berdasarkan nilai eror pada pembelajaran (Kusrini dan Emha Taufiq Luthfi, 2009, hal : 199-208).

Langkah pembelajaran dalam metode *backpropagation* adalah sebagai berikut:

1. Inisialisasi bobot jaringan secara acak (biasanya antar -1.0 hingga 1.0).
2. Pada setiap contoh data (*data training*), hitung keluaran berpedoman dari bobot jaringan saat tersebut.
3. Kemudian lakukan proses perhitungan nilai error untuk setiap keluaran (*output*) dan *hidden node* (neuron) dalam jaringan. Bobot relasi jaringan dimodifikasi.
4. Ulangi langkah 2 hingga kondisi diinginkan tercapai.

Ada beberapa cara dalam mengubah bobot-bobot jaringan saraf tiruan. Pertama, modifikasi dilakukan pada setiap akhir penghitungan setiap contoh kasus yang biasa disebut *case (online) updating*. Atau cara kedua, modifikasi bobot-bobot jaringan saraf tiruan dilakukan setelah semua contoh kasus dianalisis.

2.4 *Software* Pendukung

Software yaitu istilah khusus untuk data yang diformat, dan disimpan secara digital, termasuk program komputer, dokumentasinya, dan berbagai informasi yang bisa dibaca, dan ditulis oleh komputer. Ada beberapa *software* yang bisa

digunakan pada *data mining*, Pada penelitian ini *software* yang digunakan yaitu Tanagra versi 1.4. Tanagra adalah suatu *software* berbasis terbuka dimana setiap pengguna dapat mengakses sumber kode, dan menambahkan algoritma pengguna, selama pengguna sepakat dan sinkron terhadap lisensi pendistribusian *software* (Sikumbang, 2018).

Tanagra merupakan sebuah *software* berbasis *open source*. *Software* ini mudah dipakai selayaknya dengan aturan yang telah ada pada pengembangan *software* untuk menganalisis data *real*. Pada Tanagra tidak perlu *coding* khusus, dikarenakan semua *tools* dan *platform* sudah tersedia pada *software* tersebut.

2.5 Penelitian Terdahulu

Pada penelitian ini, peneliti menggunakan beberapa penelitian terdahulu sebagai referensi penelitian yang dikutip dari beberapa jurnal ilmiah *data mining*, yaitu:

1. Menurut penelitian (Syahdan & Sindar, 2018) dengan judul “**Data Mining Penjualan Produk Dengan Metode *Apriori* Pada Indomaret Galang Kota**”. E-ISSN: 2621-3052. Penelitian ini membahas tentang minimarket indomaret sering merasa kewalahan dengan tingkat belanjaan konsumen yang tidak beraturan. Dengan transaksi yang cukup besar maka perlu dilakukan penentuan tata letak produk makanan dan minuman. Teknik *data mining* yang dipakai yaitu *asosiatif rule* dengan algoritma *apriori* yang bertujuan mencari kombinasi dari setiap pola frekuensi transaksi. Dengan begitu akan diketahui pola frekuensi tertinggi, kemudian dicari aturan

asosiasi yang memenuhi syarat minimum untuk *confidence* aturan asosiatif $A \rightarrow B$ minimal *confidence* = 25%, nilai *confidence* dari aturan $A \rightarrow B$ dengan memakai 25 produk. Dengan setiap proses menggunakan algoritma *apriori*, maka item yang memenuhi *Support x Confidence* terbesar dan minimum *confidence* $\geq 25\%$ sehingga aturan asosiasi yang berurut yaitu Kode Produk = A, F Nama Produk = *Potabee chip bbq 35g, Plattos snack s.pg 30 g Confidence* = 33(%)

2. Menurut penelitian (Rezkiyani, 2016) dengan judul **“Implementasi data mining dengan algoritma *apriori* untuk menentukan merek sepatu yang diminati pada mahasiswa pascasarjana kelas 14.1A.01 STMK Nusa Mandiri Jakarta”**. ISBN: 978-602-72850-1-9. Penelitian ini membahas tentang cara menentukan merek sepatu yang diminati mahasiswa pascasarjana, yang mana bisa membantu staretegi penjual sepatu agar penjualan terus meningkat dengan menggunakan teknik analisis keranjang pasar. Proses pencarian asosiasi atau hubungan antar item data ini di ambil dari suatu basis data relasional. Proses tersebut menggunakan algoritma *apriori*. Dengan algoritma ini akan membantu mengetahui merek sepatu yang paling diminati dengan memanfaatkan transaksi pnjualan. Dengan nilai *support* dan *confidence* yaitu minimum *support* 20% dan minimum *confidence* 60%. Hasil yang diperoleh dari penelitian ini yaitu *Nike-Adidas* dengan *support* 25% dan *confidence* 80% dan *Converse-Adidas* dengan *support* 25% dan *confidence* 66,67%.

3. Menurut penelitian (Maulana & Fajrin, 2018) dengan judul “**Penerapan Data Mining Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma FP-Growth Pada Data Transaksi penjualan Spare Part Motor**”. ISSN: 2406-7857. CV Thaja Baru merupakan dealer yang memiliki transaksi penjualan yang besar dimana mendukung untuk mendirikan cabang baru. Dalam pemilihan lokasi cabang harus bisa menarik pola pembelian konsumen, karenanya dibutuhkan suatu analisa yang lebih dalam agar mendapatkan informasi bermanfaat dan dapat menghasilkan keuntungan. Penelitian ini menerapkan algoritma FP-Growth dalam aplikasi yang dapat mengetahui pola pembelian konsumen disetiap cabang yang memiliki karakteristik yang berbeda. Aplikasi yang dipakai dalam mengolah data pada penelitian ini yaitu *Rapidminer*. *Spare part* yang melewati minimum support dan minimum *confidence* kemudian sering terjual ialah *battery assy, screw valve adjusting, axle, cable clutch, oil seal* dan *gasket cylinder*.
4. Menurut penelitian (Rodiyansyah, 2015) dengan judul “**Algoritma apriori untuk analisa keranjang belanja pada data transaksi penjualan**”. ISSN: 2460-1861. Pada perusahaan perdagangan, sering ada transaksi dan menghasilkan kumpulan data. Tumpukan data transaksi inilah yang diolah sebagai informasi yang bermanfaat. Dengan melakukan *data mining* metode *apriori* pada data transaksi penjualan tersebut diharapkan dapat menghasilkan pengetahuan mengenai pola pembelian oleh konsumen. Pada akhirnya pola yang ditemukan ini dapat digunakan sebagai penjunjang

keputusan manajer dalam mengelola aktifitas perusahaannya. Dengan nilai *support* dan *confidence* yaitu minimum *support* 30% dan minimum *confidence* 10%.

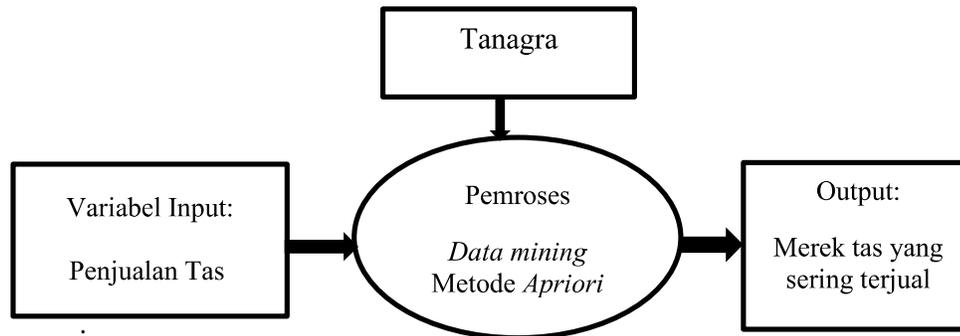
5. Menurut penelitian (Ardiansyah & Kardanawati, 2019) yang berjudul **“Implementasi Algoritma *Apriori* Untuk Menentukan *Cross Selling* Produk Pada Apotek RSUD Tugurejo Semarang”**. E-ISSN: 2528-0236. Perhitungan data transaksi yang tersimpan tiap harinya dan menyebabkan proses transaksi penjualan obat memerlukan waktu yang lama karena masih menggunakan cara yang sudah terkomputerisasi tetapi masih menggunakan cara dengan memilih daftar obat atau menentukan kode obat satu persatu berdasarkan obat yang akan dipesan. Untuk mempercepat proses penjualan tersebut dapat dilakukan dengan teknik *data mining* adalah *Association Rule* dengan memanfaatkan *cross selling* produk melalui perhitungan algoritma *Apriori*. agar proses penjualan lebih mudah dan lebih cepat sehingga dapat mempercepat proses tiap transaksi yang akan mengakibatkan berkurangnya panjang antrean para pembeli. Dengan implementasi algoritma *apriori* dengan acuan nilai minimum *support* 11% dan nilai minimum *confidence* 60% didapati pola transaksi pembelian obat dengan frekuensi tertinggi sebanyak 3 itemset yaitu jika membeli *abilify discmelt* 10mg (*aripiprazole*) *box* 10 tab dan *cath. Folley 2 way 16* (*rusch/urocare*) maka membeli *apialasit drop* dengan nilai *support* 18.6% dan nilai *confidence* 96.87% dan jika membeli *alprazolam 0.5 mg tab* dan *clonidin 0.15mg* maka membeli

nevirapine 200mg @60 (D) dengan nilai support 11.6% dan nilai confidence 90.62%.

6. Menurut penelitian (Mirmozaffari, Alinezhad, & Gilanpour, 2017) yang berjudul *"Data Mining Apriori Algorithm for Heart Disease Prediction"*. ISSN: 2349-1469. *To find the unknown trends in heart disease, Apriori algorithm in association rule are applied to a unique dataset and their accuracy are compared in two different software. To further enhance accuracy and achieve more reliable variables, the dataset is purified by Discretization unsupervised filter. Finally, better performance software for Apriori algorithm with better accuracy is introduced. The system was implemented in WEKA and MATLAB software and prediction accuracy within Apriori algorithm in 3 steps, re compared. Despite "min support", the exact number of supports are introduced. For instance, in The Second WEKA Apriori rule, only "min support" (0.6) is introduced. However, in The Second MATLAB Apriori rule, despite "min support" (0.6), the exact number of support (0.622) is evaluated.*

2.6 Kerangka Pemikiran

Untuk dapat melakukan penelitian penjualan tas ini, diperlukan kerangka pemikiran yang bisa dipakai dalam menjelaskan bagaimana cara metode ini mampu memecahkan permasalahan yang ada. Kerangka pemikiran tersebut digambarkan dengan diagram dibawah ini:



Gambar 2.2 Kerangka Pemikiran
Sumber: Data penelitian, 2019

Pada gambar 2.2 menjelaskan bahwasanya di dalam kerangka pemikiran ini mempunyai variabel input, pemroses, dan terakhir variabel *output*. Untuk variabel input ialah transaksi penjualan tas di Toko Jc Jolin *Colecction*. Kemudian variabel input diproses menggunakan teknik *data mining* menggunakan metode *apriori*, lalu diuji memakai *Software Tanagra*, agar menemukan merek tas apa saja yang paling banyak terjual. Sedangkan untuk *output* akan mendapatkan hasil *item set* yang banyak terjual dari keterkaitan *rule* atau pola.

2.7 Hipotesa

Penelitian diharapkan mampu menghasilkan sebuah keputusan dengan menggunakan teknik *data mining* metode algoritma *apriori* dimana *rule* atau pola yang terbentuk dari transaksi penjualan tas.