

BAB II LANDASAN TEORI

2.1 *Knowledge Discovery In Database*

Knowledge Discovery In Database (KDD) merupakan metode yang digunakan untuk mencari pengetahuan atau informasi yang belum diketahui dari sebuah database. *Knowledge Discovery In Database (KDD)* merupakan nama lain dari *Data Mining* walaupun sesungguhnya kedua istilah tersebut memiliki konsep yang tidak sama, namun berkaitan satu sama lain, dan salah satu dari tahapan proses keseluruhan *Knowledge Discovery In Database (KDD)* merupakan *Data Mining* yang menjadi inti dari proses KDD. Merujuk pada penelitian KDD adalah sebuah proses untuk mencari dan mengidentifikasi *pattern* dalam sebuah *database*, pada sebuah *Knowledge Discovery In Database* atau KDD memiliki beberapa tahapan diantaranya :

1. Seleksi Data

Proses Seleksi Data dilakukan dengan memilih data yang relevan dengan tugas menganalisis dari database, menciptakan himpunan data target, atau memfokuskan pada contoh data dimana *discovery* akan dilakukan dan hasil dari seleksi disimpan dalam suatu berkas terpisah dari *database* operasional.

2. Pemrosesan dan Pembersihan Data

Pada tahapan ini dilakukan menghilangkan *noise* dan data yang inkonsisten. Sebelum proses *Data Mining* dikerjakan, perlu dilakukannya proses *cleaning* pada data yang akan menjadi fokus KDD dan proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data. Dilakukan proses *enrichment*, yaitu proses “memperkaya” data yang sudah ada dengan data lain yang relevan untuk keperluan KDD.

3. Transformasi

Transformasi adalah proses transformasi pada data yang dipilih, sehingga data tersebut sesuai untuk proses *Data Mining*. Proses ini

merupakan proses kreatif dan sangat tergantung pada jenis atau *pattern* informasi yang akan dicari pada *database*.

4. *Data Mining*

Pada tahapan ini merupakan proses mencari *pattern* atau pola dan informasi dari sebuah *database* dengan menggunakan teknik atau metode. Pada proses *Data Mining* terdapat banyak teknik, metode atau algoritma yang dapat digunakan dan sangat bervariasi dan untuk menentukan pemilihan metode yang akan digunakan tergantung pada tujuan dan proses KDD secara keseluruhan.

5. Evaluasi

Tahap ini merupakan bagian dari proses KDD yang mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya. Proses ini merupakan proses penerjemahan pola-pola yang didapatkan dari *Data Mining*. Pola informasi yang dihasilkan dari proses *Data Mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan.

2.2 *Data Mining*

Data Mining merupakan sebuah inti dari proses KDD, meliputi dugaan algoritma yang mengeksplor data, membangun model dan menemukan pola yang belum diketahui. KDD bersifat otomatis, dapat didefinisikan sebagai pengorganisasian proses untuk pengidentifikasian yang benar, berguna dan penemuan pola dari kumpulan data yang besar dan kompleks *Data Mining* merupakan salah satu teknik untuk menemukan, mencari, atau menggali informasi atau pengetahuan baru dari sekumpulan data yang sangat besar, dengan integrasi atau penggabungan dengan disiplin ilmu lain seperti statistika, kecerdasan buatan, serta *machine learning*, menjadikan *Data Mining* sebagai salah satu alat bantu untuk menganalisa data yang kemudian menghasilkan informasi yang berguna. Dari penjelasan tersebut, *Data Mining* dapat diartikan sebagai sebuah proses dari sekumpulan data yang memiliki jumlah besar untuk mendapatkan data yang hasilnya tidak hanya sekedar informasi melainkan merupakan sebuah pengetahuan atau *knowledge* yang tersembunyi dari data tersebut. *Knowledge* yang diperoleh

dari hasil *mining* terhadap sekumpulan data tersebut dapat membantu dalam pengambilan sebuah keputusan.

2.3 AssociationRule

Associationrule merupakan suatu proses pada *Data Mining* untuk menentukan semua aturan asosiatif yang memenuhi syarat minimum untuk *support (minsup)* dan *confidence (minconf)* pada sebuah *database*. Kedua syarat yang digunakan untuk *interesting association rules* dibandingkan dengan batasan yang telah ditentukan dengan *minimum support* dan *minimum confidence*. *Associationrule* adalah teknik *Data Mining* untuk menemukan aturan asosiasi antara suatu kombinasi barang. Fungsi *associationrules* seringkali disebut dengan “*market basket analysis*”, yaitu Analisis dari kebiasaan membeli *customer* dengan mencari asosiasi dan korelasi antara item-item berbeda yang diletakkan *customer* dalam keranjang belanjanya. Contoh aturan asosiasi dari analisis pembelian di suatu produk pada toko sembako adalah dapat diketahuinya berapa besar kemungkinan seorang konsumen membeli beras bersamaan dengan membeli gula. Dengan pengetahuan tersebut, pemilik toko dapat menyediakan stok lebih barang tertentu yang sering dibeli oleh konsumen.

Dalam menentukan nilai *minimum support* sebuah *item* dapat menggunakan rumus persamaan seperti di bawah ini :

$$1. \text{Support (A)} = \frac{\text{Jumlah Transaksi Mengandung A}}{\text{Total Transaksi}}$$

$$2. \text{Support (A} \cap \text{B)} = \frac{\text{Jumlah Transaksi Mengandung A dan B}}{\text{Total Transaksi}}$$

Sedangkan untuk menentukan nilai *minimum confidence* sebuah *item* dapat menggunakan rumus persamaan seperti di bawah ini :

$$1. \text{Confidance (A} \rightarrow \text{B)} = \frac{\text{Jumlah Transaksi Mengandung A dan B}}{\text{Total Transaksi}}$$

Jumlah Transaksi Mengandung A

2.4 Algoritma Apriori

Algoritma Apriori merupakan salah satu algoritma yang melakukan pencarian frequent itemset dengan menggunakan teknik association rule. Untuk mencari association rule dari kumpulan data, yang harus dilakukan yaitu mencari Frequent itemset terlebih dahulu. Frequent itemset adalah kumpulan item yang sering muncul secara bersamaan. Penting tidaknya suatu asosiasi dapat diketahui dengan dua tolak ukur yaitu nilai support dan nilai confidence. Support adalah nilai penunjang atau persentase dari kombinasi sebuah item dalam database sedangkan confidence adalah nilai kepastian yaitu kuatnya hubungan antar item dalam asosiasi. Algoritma apriori adalah metode untuk menemukan pola hubungan antara satu atau lebih banyak item dalam kumpulan data. Algoritma Apriori merupakan algoritma yang paling terkenal yang sering kali digunakan peneliti untuk menemukan pola frekuensi tinggi. Pola frekuensi tinggi adalah pola item yang terdapat di dalam suatu database yang memiliki frekuensi atau support di atas ambang batas tertentu yang disebut dengan istilah minimum support. Dari penjelasan di atas dapat disimpulkan bahwa Algoritma Apriori digunakan sebagai teknik asosiasi sederhana yang menentukan pola frequent itemset yang dijalankan pada sekumpulan data. Pada intinya semua aturan asosiasi yang memenuhi syarat minimum untuk support dan syarat minimum untuk confidence

2.5 Tanagra

Tanagra merupakan salah satu software data mining yang didalamnya disediakan beberapa metode data mining mulai dari mengeksplorasi analisis data, pembelajaran statistik, pembelajaran mesin, dan data mining kebanyakan. Tanagra suatu software berbasis open source dimana semua orang dapat mengakses source code, dan menambahkan algoritma mereka sendiri, sejauh dia setuju dan menyesuaikan dengan lisensi pendistribusian software. Tanagra adalah alat pembelajaran mesin yang paling banyak digunakan oleh para peneliti untuk analisis data, Tanagra pada dasarnya digunakan dalam bentuk tekstual. Spina

adalah representasi grafis dari Tanagra. Ini sangat berguna dalam menambang data dan peneliti memanfaatkan akses terbukanya. Ini memberi peneliti platform pengembangan baru untuk menambahkan fungsionalitas baru dengan bantuan algoritma yang ada.

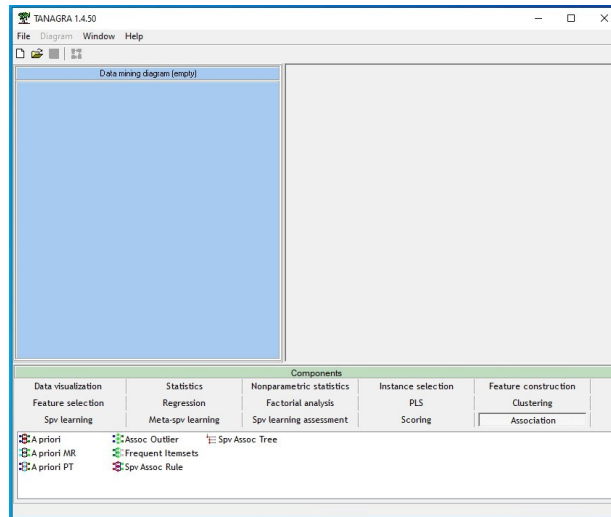
Tanagra adalah software data mining untuk keperluan akademik dan riset. Di dalamnya disediakan beberapa metoda data mining mulai dari mengeksplorasi analisis data, pembelajaran statistik, pembelajaran mesin, dan database. Proyek ini merupakan suksesor dari sebuah software lain yang bernama *SPINA* yang mengimplementasikan bermacam algoritma pembelajaran terkontrol, khususnya konstruksi pohon keputusan visual yang interaktif. Unjuk kerja Tanagra lebih baik, di mana selain memiliki beberapa pembelajaran terkontrol juga paradigma lain seperti clustering, analisis faktorial, statistik parametrik dan non parametrik, aturan asosiasi, feature selection, dan construction algorithms.

Tanagra merupakan suatu proyek open source di mana semua peneliti dapat mengakses source code nya, dan menambahkan algoritma mereka sendiri, sejauh dia setuju dan menyesuaikan dengan lisensi pendistribusian softwarena.

Ada beberapa tujuan dari proyek Tanagra ini, yaitu:

1. Memberi peneliti dan mahasiswa suatu software data mining yang mudah digunakan (*easy to use/ user friendly*), sesuai dengan aturan yang ada dari pengembangan software dalam daerah ini (terlebih lagi dalam mendisain GUI nya dan cara penggunaannya), dan memperbolehkan menganalisa data real maupun sintetis.
2. Mengusulkan pada peneliti suatu arsitektur yang memperbolehkan mereka untuk secara mudah menambah metode data mining mereka sendiri, untuk membandingkan performannya. Tanagra bertindak lebih sebagai platform percobaan dalam hal memperbolehkan mereka menuju ke pekerjaan pokok mereka, menyalurkan mereka berhadapan dengan bagian yang tak menyenangkan dalam pemograman tool seperti ini: manajemen data.
3. Tujuan terakhir yaitu, searah dengan developer pemula, bertujuan untuk menyebarkan metodologi yang mungkin untuk membangun software seperti ini. Mereka harus mengambil keuntungan dalam hal akses bebas ke source code, melihat bagaimana software seperti ini dibangun, masalah

untuk dihindari, langkah utama dari proyek, dan tools atau libraries yang mana yang harus digunakan dan untuk apa digunakan. Dalam hal ini, Tanagra dapat dianggap sebagai alat untuk mendidik untuk belajar teknik pemrograman. Terlihat screenan Tanagra pada gambar 2.1 dibawah ini



Gambar 2.1 Sistem Tanagra

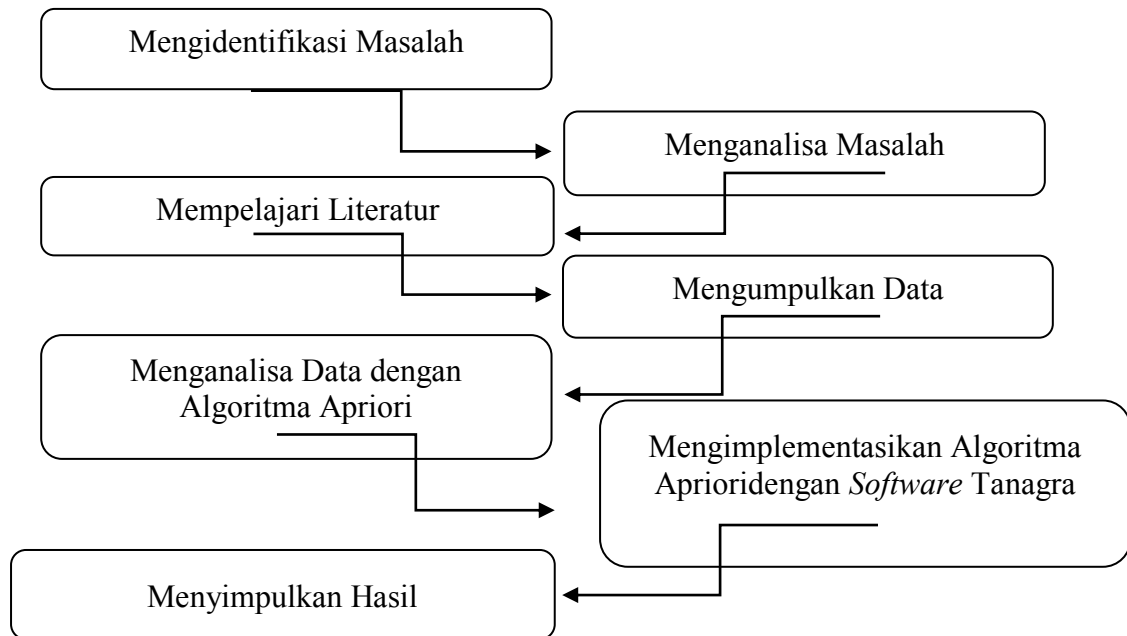
2.6 Metodologi Penelitian

Dalam melakukan penelitian agar hasilnya bisa maksimal, penulis mengikuti kaidah-kaidah (metode) yang telah ditetapkan. Metode penelitian yang akan diterapkan adalah metode penelitian ilmiah yang telah diajarkan di bangku perkuliahan sarjana. Metodologi penelitian ini memuat tentang kerangka kerja penelitian yang akan dibahas di bawah ini. Pada bab ini akan dibahas metodologi penelitian untuk mengidentifikasi permasalahan, analisa permasalahan dan pada akhirnya mencari penyelesaian masalah di dalam menganalisa data transaksi untuk mengetahui pola-pola penjualan dalam meningkatkan penjualan pada Cofe AyoS. Metodologi penelitian merupakan cara untuk mengetahui hasil dari sebuah permasalahan yang spesifik dari suatu penelitian.

2.6.1 Kerangka Kerja

Kerangka kerja dalam penelitian merupakan kumpulan konsep penelitian yang tersusun secara sistematis supaya tujuan dari penelitian tercapai dengan baik. Kerangka penelitian ini dibentuk sebelum langkah penelitian dilakukan, kerangka kerja penelitian juga merupakan konsep suatu penelitian yang menghubungkan antara visualisasi satu variabel dengan variable lainnya, sehingga penelitian yang

dilakukan dapat diterima oleh semua pihak. Dengan adanya kerangka kerja diharapkan penelitian ini dapat diselesaikan dengan baik, gambaran kerangka kerja penelitian ini dapat lihat pada gambar 2.2.



Gambar 2.2 Kerangka Kerja Penelitian

Dari **Gambar 2.2** di atas terlihat jelas bahwa penelitian ini dilakukan secara bertahap dan sistematis, berikut ini adalah penjelasan dari gambar kerangka kerja di atas:

1. Mengidentifikasi Masalah

Tahap awal dari penelitian ini adalah mengidentifikasi masalah-masalah yang ada, permasalahan-permasalahan yang teridentifikasi inilah yang akan menjadi bahan acuan untuk mengerjakan tahap berikutnya.

2. Menganalisa Masalah

Adanya penelitian karena adanya permasalahan yang harus diselesaikan, artinya analisa masalah merupakan tahapan yang harus dilewati oleh setiap peneliti, pada tahapan ini peneliti melakukan proses analisa masalah pada data penjualan Coffe Ayos. Dengan analisa ini sehingga pokok masalah yang diteliti akan tampak jelas (batasan, lingkup, latar belakang, dan

signifikansinya) setelah dilakukan analisis terhadap pokok permasalahan yang bersangkutan.

3. Mempelajari Literatur

Dengan mempelajari literatur akan meningkatkan pemahaman peneliti mengenai teori-teori yang relevan, pemahaman akan teori-teori ini diharapkan peneliti dapat menjelaskan, membedakan, memprediksi fenomena-fenomena dan gejala-gejala yang berhubungan dengan masalah penelitian. Dalam studi literatur ini sebelum mengumpulkan data yang akan di proses, terlebih dahulu mengetahui pengetahuan-pengetahuan atau *knowledge* yang akan diterapkan dalam metode algoritma apriori ini, literatur yang akan dipelajari ini bersumber dari jurnal-jurnal ilmiah yang di publikasikan di internet.

4. Mengumpulkan Data

Dalam pengumpulan data dilakukan observasi yaitu pengamatan secara langsung di tempat penelitian sehingga permasalahan yang ada dapat diketahui secara jelas. Kemudian dilakukan interview yang bertujuan untuk mendapatkan informasi atau data yang dibutuhkan

5. Menganalisa Data Menggunakan Algoritma Apriori

Pada tahap ini akan dilakukan analisa menggunakan algoritma apriori pada data transaksi penjualan untuk menentukan alternatif yang terbaik dari sekumpulan alternatif yang ada melalui suatu proses yang terstruktur dan analisa bagaimana apriori untuk mendapatkan hasil sebagai tujuan yang akan dicapai kemudian dapat dijadikan oleh pihak Coffe Ayo sebagai pengetahuan dalam meningkatkan penjualan. Proses analisa *Data Mining* dengan algoritma apriori dilakukan berdasarkan data yang telah terkumpul. Dan bagaimana mengembangkan proses analisa *Data Mining* dengan *Association Rule* dan Apriori untuk mendapatkan hubungan antar tiap produk yang ada. Proses ini dilakukan pada data penjualan untuk mengetahui pola pembelian yang paling banyak dibeli oleh konsumen.

Ada dua tahapan dalam melakukan perhitungan menggunakan algoritma apriori, yaitu yang pertama Analisa Pola Frekuensi Tinggi dan yang kedua Pembentukan Aturan Asosiasi.

- a. Analisa Pola Frekuensi Tinggi
Mengumpulkan data awal dan dikelompokan sesuai dengan transaksinya, kemudian menetapkan besarnya besaran Φ dan nilai minimum Support dan Confidence
 - b. Pembentukan Aturan Asosiasi
Setelah semua pola frekuensi tertinggi ditemukan, maka selanjutnya akan dicari aturan asosiasi yang memenuhi syarat minimum confidence yang telah ditentukan sebelumnya.
6. Mengimplementasikan Algoritma Aprioridengan *Software* Tanagra
Pada penelitian ini penulis mengimplementasikan berdasarkan hasil analisa data dengan metode algoritma apriori menggunakan alat bantu komputer dengan sistem operasi *windows* dan *software* Tanagra.
 7. Menyimpulkan Hasil
Setelah dilakukan proses analisis data penjualan produk makanan dan minuman pada Coffe Ayos menggunakan Algoritma Apriori serta mengimplementasinya dengan Aplikasi Tanagra, maka hasil yang didapat adalah pola-pola penjualan dan kemudian dijadikan acuan atau bahan evaluasi oleh pemilik usaha “owner” untuk mempersiapkan jenis-jenis makanan dan minuman mana yang perlu distok lebih banyak lagi sehingga dapat meningkatkan penjualan.