III. LANDASAN TEORI

3.1 Pengelolaan Data Mahasiswa Baru

Saat ini di Kopertis Wilayah V Yogyakarta terdapat 46 perguruan tinggi (jenjang D-3=14 dan S-1=32) yang menyelenggarakan program studi bidang informatika dan komputer, secara nasional terdapat 1260 perguruan tinggi (jenjang D-3= 583 dan S-1=677).

Mengutip pernyataan Ashari yang bersumber dari Jawapos 17 Mei 2015, bahwa Jumlah lulusan SMA/MA tahun 2015 sebanyak 1.62 juta, yang melanjutkan kuliah 60% (975.000). Jumlah lulusan SMK sebanyak 1.17 juta, hanya 8% yang melanjutkan kuliah (94.000). Jadi, total lulusan sekolah menengah yang melanjutkan kuliah sekitar 1.1 juta orang pertahun. Jika daya tampung PTN sebesar 250.000 pertahun (rata-rata 2500 per PTN), maka potensi 850.000 calon mahasiswa akan diperebutkan oleh 3100 PTS (rata-rata 275 calon mahasiswa baru per PTS).

STMIK AKAKOM adalah perguruan tinggi yang menyelenggarakan pendidikan tinggi khusus di bidang teknologi informasi, saat ini menyelenggara-kan 5 program studi, yaitu: Teknik Informatika (S1), Sistem Informasi (S1), Manajemen Informatika (D3), Komputerisasi Akuntansi (D3), dan Teknik Komputer (D3).

Proses penerimaan mahasiswa baru di STMIK AKAKOM dilaksanakan oleh bagian admisi dan kerjasama, yang dibantu oleh *task force* yang dibentuk untuk itu, yang bertugas merumuskan strategi dan kebijakan untuk melakukan

sosialisasi dan promosi dalam bentuk pameran pendidikan dan anjangsana ke sekolah-sekolah. Strategi promosi dan sosialisasi yang dilakukan untuk keperluan jangka pendek, menengah, maupun jangka panjang, sehingga diperlukan strategi yang tepat dalam implementasinya.

Proses rekrutmen calon mahasiswa baru dilakukan melalui beberapa metode, antara lain: 1) Penerimaan calon mahasiswa melalui jalur CBT (*Computer Based Test*), calon siswa yang sudah mendaftar dapat memilih jadwal untuk mengikuti tes, calon mahasiswa yang sudah mengikuti tes pada tanggal dan jam yang telah disepakati hasilnya langsung dapat diketahui dalam waktu kurang dari 1 jam (diterima atau tidak diterima). Bagi calon mahasiswa yang diterima akan mendapatkan surat pemberitahuan diterima yang disertai persyaratan registrasi dan biaya yang harus dibayarkan pada saat registrasi. 2) Penerimaan calon mahasiswa melalui jalur prestasi akademik, apabila nilai rata-rata rapor ataui nilai UAN $\geq$ 8.00, tidak harus mengikuti seleksi, cukup menyerahkan foto-kopi rapor atau SKHUN yang telah dilegalisir sekolah. Setelah dilakukan verifikasi oleh tim, akan diterbitkan surat pemberitahuan diterima dan syarat registrasi dan biaya yang harus dibayarkan saat registrasi. 3) Penerimaan calon mahasiswa melalui jalur prestasi olahraga dan seni, dapat diterima tanpa harus mengikuti tes apabila bisa menyerahkan fotokopi sertifikat kejuaraan tingkat nasional.

Bagian admisi dan kerjasama, memiliki tanggung jawab dan kewenangan untuk mengelola data pendaftaran mahasiswa baru hingga registrasi, yang selanjutnya akan dikelola lebih lanjut oleh bagian akademik. Meningkatnya

jumlah data mahasiswa baru dari tahun ke tahun, dengan berbagai karakteristik pola data mahasiswa baru, asal jurusan di sekolah, asal SLTA, asal daerah, dan lain-lain, akan sangat bermanfaat apabila diolah lebih lanjut untuk membantu bagian admisi dan kerjasama dalam memilih strategi pemasaran dan sosialisasi yang tepat.

Terbatasnya jumlah lulusan yang masuk ke perguruan tinggi, dan banyaknya perguruan tinggi yang menyelenggarakan pendidikan di bidang informatika dan komputer, menimbulkan persaingan yang semakin ketat untuk memperoleh mahasiswa baru. Sehingga diperlukan strategi yang jitu agar perguruan tinggi dapat memperoleh target mahasiswa sesuai dengan harapan.

3.2 Data Mining

Dengan meningkatnya transaksi yang disimpan dengan sistem basis data sekarang ini, maka dibutuhkan proses untuk menangani data tersebut. Proses untuk menangani data tersebut dikenal dengan *Knowledge Discovery in Databases* (KDD). *Data Mining* adalah kegiatan untuk menemukan informasi atau pengetahuan yang berguna secara otomatis dari data yang jumlahnya besar. *Data Mining* merupakan salah satu proses dari keseluruhan proses yang ada pada *Knowledge Discovery in Databases* (KDD). KDD sendiri merupakan sekumpulan proses untuk menemukan pengetahuan yang bermanfaat dari data. KDD terdiri dari serangkaian langkah perubahan, termasuk data *preprocessing* dan juga *post processing*. Data *preprocessing* merupakan langkah untuk mengubah data mentah menjadi format yang sesuai untuk tahap analisis berikutnya. Selain itu data

preprocessing juga digunakan untuk membantu dalam pengenalan atribut dan data segmen yang relevan dengan task *data mining*. Data *preprocessing* kemungkinan akan membutuhkan waktu yang sangat lama, hal ini dikarenakan data yang mentah kemungkinan disimpan dengan format dan *database* yang berbeda. *Post processing* meliputi semua operasi yang harus dilakukan agar hasil dari *Data Mining* dapat diakses dan lebih mudah untuk diinterpretasikan oleh para analis. Teknik visualisasi juga dapat digunakan untuk mempermudah para analis untuk menggali dan memahami kegunaan dari *data mining*. Kumpulan proses dalam KDD meliputi :

a) pembersihan data (*data cleaning*),
b) integrasi data (*data integration*),
c) pemilihan data (*data selection*),
d) transformasi data (*data transformation*),
e) penambangan data (*data mining*),
f) evaluasi pola (*pattern evaluation*), dan
g) presentasi pengetahuan (*knowledge presentation*).

Berdasarkan definisi ini terlihat bahwa *data mining* hanya merupakan salah satu proses dari keseluruhan proses yang ada pada KDD, tetapi merupakan proses yang sangat penting dalam usaha menemukan pola-pola yang berguna dari sejumlah data yang besar (data tersebut bisa disimpan dalam basisdata, *data warehouse*, atau media penyimpanan informasi lainnya).

3.3 Data Mining Task

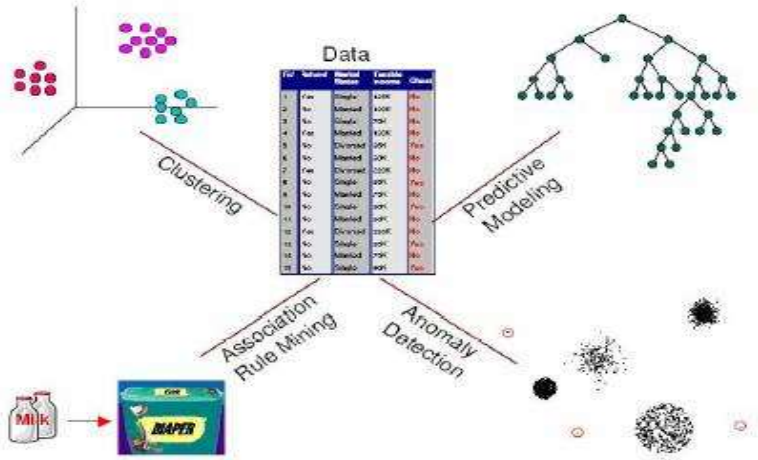
Pada umumnya, *data mining task* dibagi menjadi dua kategori yang penting, yaitu:

1) Predictive tasks

Tujuan dari *task* ini adalah untuk memprediksi nilai sebuah atribut yang penting berdasarkan nilai dari atribut yang lainnya. Atribut yang diprediksi biasanya dikenal sebagai target atau *dependent variable*, sedangkan atribut yang digunakan untuk melakukan prediksi dikenal dengan *explanatory* atau *independent variable*.

2) Descriptive task

Tujuan dari task ini adalah untuk menghasilkan pola (*correlations, trends, clusters, trajecttories dan anomalies*) yang merangkum keterhubungan dalam data.



Gambar 2.1 Data Mining Task

Dari gambar diatas , data yang ada dapat digunakan sebagai inti dari *data mining task*. *Data mining* task tersebut antara lain:

1) Predictive Modelling

Predictive modelling digunakan untuk membangun sebuah model untuk target variable sebagai fungsi dari *explanatory variable*. *Explanatory variable* dalam hal ini merupakan semua atribut yang digunakan untuk melakukan prediksi, sedangkan target variable merupakan atribut yang akan diprediksi nilainya. Predictive modeling task dibagi menjadi dua tipe yaitu : *Classification* digunakan untuk memprediksi nilai dari target variable yang *discrete* (diskret) dan *regression* digunakan untuk memprediksi nilai dari target variable yang *continu* (berkelanjutan).

2) Association Analysis

Association analysis digunakan untuk menemukan aturan assosiasi yang memperlihat-kan kondisi-kondisi nilai atribut yang sering muncul secara bersamaan dalam sebuah himpunan data.

3) *Cluster* Analysis

Tidak seperti klasifikasi yang menganalisa kelas data obyek yang mengandung label. *Clustering* menganalisa objek data tanpa memeriksa kelas label yang diketahui. Label-label kelas dilibatkan di dalam data *training*. Karena belum diketahui sebelumnya. *Cluster*ing merupakan proses pengelompokkan sekumpulan objek yang sangat mirip.

4) Anomaly Detection

Anomaly Detection merupakan metode pendeteksian suatu data dimana tujuannya adalah menemukan objek yang berbeda dari sebagian besar objek lain. Anomaly dapat di deteksi dengan menggunakan uji statistik yang menerapkan model distribusi atau probabilitas untuk data.

3.4 *Association Analysis* (*Association Rule*)

Association rule adalah salah satu teknik utama atau prosedur dalam Market Basket Analysis untuk mencari hubungan antat item dalam suatu data set dan menampilkan dalam bentuk *association rule* (Budhi dkk,2007). *Association rule* (aturan asosiatif) akan menemukan pola tertentu yang mengasosiasikan data yang satu dengan data yang lain. Untuk mencari *association rule* dari suatu kumpulan data, tahap pertama yang harus dilakukan adalah mencari *frequent itemset* terlebih dahulu. *Frequent itemset* adalah sekumpulan item yang sering muncul secara bersamaan. Setelah semua pola *frequent itemset* ditemukan, barulah mencari aturan asosiatif atau aturan keterkaitan yang memenuhi syarat yang telah ditentukan.

Diasumsikan data pada keranjang belanja pembeli di pasar swalayan, setiap transaksi pembelian dari pelanggan ada ID transaksi dan setiap transaksi terdapat sejumlah barang yang dibeli. Melalui teknik analisis asosiasi dapat diketahui pola pembelian pelanggan. Misalnya, pelanggan dari kalangan ibu rumah tangga biasanya membeli minyak, telur, gula dan beras. Namun pada saat bersamaan, jarang yang membeli baju dan buku. Sehingga, dengan mengetahui

pola pembelian dari pelanggan, manajemen pasar swalayan dapat membuat keputusan yang lebih baik, dalam menerapkan strategi penjualannya. Misalnya, kapan waktu yang tepat untuk promosi diskon barang, menentukan jumlah dan ragam barang yang harus disediakan, menentukan strategi untuk menjual barang yang kurang laku, manajemen pembelian dapat menentukan barang apa saja yang sebaiknya dibeli, dan lain-lain.

Jika diasumsikan bahwa barang yang dijual di pasar swalayan adalah semesta, maka setiap barang akan memiliki bolean variabel yang akan menunjukkan keberadaannya atau tidak barang tersebut dalam satu transaksi atau satu keranjang belanja. Pola bolean yang didapat dugunakan untuk menganalisa barang yang sering dibeli secara bersamaan. Pola tersebut dapat dirumuskan dalam sebuah *association rule*. Sebagai contoh pelanggan biasanya membeli kopi dan susu yang ditunjukkan sebagai berikut :

Kopi → susu [support =5%, confidence=65%]

Association rule diperlukan suatu variable ukuran yang ditentukan sendiri oleh user untuk menentukan batasan sejauh mana atau sebanyak apa output yang diinginkan user.

Support dan confidence adalah sebuah ukuran kepercayaan dan kegunaan suatu pola yang telah ditemukan. Nilai support 2% menunjukkan bahwa keseluruhan dari total transaksi konsumen membeli kopi dan susu secara bersamaan yaitu sebanyak 5%. Sedangkan confidence 65% yaitu menunjukkan bila konsumen membeli kopi dan pasti membeli susu sebesar 65%.

Penentuan aturan asosiasi dapat didefinisikan:

> "Diberikan sejumlah transaksi T, carilah semua aturan yang mempunyai support > minsup dan confidence > minconf, dimana minsup adalah ambang batas support, sedangkan minconf adalah ambang batas confidence".

Untuk menggali aturan asosiasi yang diinginkan, jika menggunakan pendekatan brute force, jumlah kombinasi aturan akan tumbuh secara eksponensial, namun cara tersebut sangat mahal dan lama dalam proses komputasinya. Secara spesifik, total jumlah aturan yang mungkin untuk diekstrak dari set data yang berisi d item adalah

$R = 3^d - 2^{d+1} + 1$ …… pers-1

Bila menggunakan pendekaten brute-force, misalnya kita memiliki set data yang berisi 6 item akan memerlukan komputasi support dan confidence sebanyak $= 3^6 - 2^7 + 1 = 602$ aturan. Meskipun terdapat 602 aturan, biasanya 80% dari aturan akan dibuang setelah menerapkan nilai minsup = 20% dan minconf = 50%, maka komputasi yang besar banyak yang dibuang. Sehingga, diperlukan algoritma yang lebih efisien agar tidak melakukan komputasi yang demikian besar, misalnya di bagian awal bisa memangkas aturan-aturan yang berpotensi pasti terbuang, karena adanya aturan lain yang sudah terbuang, tanpa harus menghitung *support* dan *confidence*-nya lagi.

Strategi yang biasa digunakan sejumlah algoritma dalam penggalian aturan asosiasi adalah memecah masalah ke dalam dua pekerjaan utama, yaitu:

a. *Frequent itemset generation*, tujuannya untuk mencari semua itemset yang memenuhi ambang batas *minsup*, disebut *itemset* frekuen (*frequent itemset*)
b. *Rule generation*, tujuannya untuk mengekstrak aturan dengan *confidence* tinggi dari *itemset* frekuen yang ditemukan dalam langkah sebelumnya, disebut aturan kuat (*strong rule*).

3.5 Algoritma Apriori

Algoritma apriori termasuk jenis aturan asosiasi pada data mining. Selain apriori, yang termasuk pada golongan ini adalah metode *generalized rule induction* dan *algoritma hash based*. 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 assosiatif antara suatu kombinasi item. Contoh aturan assosiatif dari analisa pembelian di suatu pasar swalayan adalah dapat diketahuinya berapa besar kemungkinan seorang 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.

Analisis asosiasi menjadi terkenal karena aplikasinya untuk menganalisa isi keranjang belanja di pasar swalayan. Analisis asosiasi juga sering disebut dengan istilah *market basket analysis*.

Analisis asosiasi dikenal juga sebagai salah satu teknik data mining yang menjadi dasar dari berbagai teknik data mining lainnya. Khususnya salah satu tahap dari analisis asosiasi yang disebut analisis pola frequensi tinggi *(frequent pattern mining)* menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien.

Penting tidaknya suatu aturan assosiatif dapat diketahui dengan dua parameter, *support* (nilai penunjang) yaitu persentase kombinasi item tersebut dalam database dan *confidence* (nilai kepastian) yaitu kuatnya hubungan antar item dalam aturan assosiatif.

Algoritma apriori merupakan salah satu algoritma yang melakukan pencarian *frequent itemset* dengan menggunakan teknik *association rule* (Erwin, 2009). Algoritma Apriori menggunakan pengetahuan frekuensi atribut yang telah diketahui sebelumnya untuk memproses informasi selanjutnya. Pada algoritma apriori penentuan kandidat yang mungkin muncul dengan cara memperhatikan minimum *support* dan minimum *confidence*. *Support* adalah nilai penunjang atau persentase kombinasi sebuah *item* dalam *database*.

Metodologi dasar analisis asosiasi terbagi menjadi dua tahap :

a. Analisa pola frekuensi tinggi

Tahap ini mencari kombinasi item yang memenuhi syarat minimum dari nilai support dalam database. Nilai support sebuah item diperoleh dengan rumus berikut:

$$Support\ (A) = \frac{\text{Jumlah transaksi mengandung A}}{\text{Total transaksi}} \times 100\% \qquad \text{..... (1)}$$

Sedangkan nilai support dari 2 item diperoleh dari rumus berikut:

$$\text{Support } (A \cap B) = \frac{\text{Jumlah Transaksi mengandung A dan B}}{\text{Total Transaksi}} \quad \ldots\ldots(2)$$

b. Pembentukan aturan assosiatif

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan assosiatif yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan assosiatif A → B. Nilai *confidence* dari aturan A → B diperoleh dari rumus berikut:

$$\textit{Confidence} = P(B \mid A) = \frac{\text{Total transaksi mengandung A dan B}}{\text{Transaksi mengandung A}} \times 100\% \quad \ldots (3)$$

Proses utama yang dilakukan dalam algoritma Apriori untuk mendapat *frequent itemset* yaitu (Erwin, 2009) :

1. *Join* (penggabungan)

   Proses ini dilakukan dengan cara pengkombinasian item dengan item yang lainnya hingga tidak dapat terbentuk kombinasi lagi.

2. *Prune* (pemangkasan)

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

Prinsip dari Algoritma Apriori antara lain :

1) Mengumpulkan item tunggal kemudian mencari item yang terbesar.
2) Menentukan *candidate pairs* kemudian dan menghitung *large pairs* dari masing-masing item.

3) Mencari *candidate triplets* dari setiap item dan seterusnya.

4) Setiap subset dari sebuah *frequent itemset* harus menjadi *frequent.*

Sebagai contoh ambil suatu data transaksi yang didapat dari penjualan sayur dengan data transaksi seperti table-3.1.

Tabel-3.1 Penjualan Item yang Dibeli

Transaksi	Item yang dibeli
1	Broccoli, Green Peppers, Corn
2	Asparagus, Squash, Corn
3	Corn, Tomatoes, Beans, Squash
4	Green Peppers, Corns, Tomatoes, Beans
5	Beans, Asparagus, Broccoli
6	Squash, Asparagus, Beans, Tomatoes
7	Tomatoes, corn
8	Broccoli, Tomatoes, Green Peppers
9	Squash, Asparagus, Beans
10	Beans, Corn
11	Green Peppers, Broccoli, Beans, Squash
12	Asparagus, Bean, Squash
13	Squash, Corn, Asparagus, Beans
14	Corn, Green Peppers, Tomatoes, Beans, Broccoli

Definisi-definisi yang terdapat pada *Association Rule*

1. I adalah himpunan yang tengah dibicarakan.

   Contoh:

   {Asparagus, Beans, …, Tomatoes}

2. D adalah himpunan seluruh transaksi yang tengah dibicarakan

   Contoh:

   {Transaksi 1, transaksi 2, …, transaksi 14}

3. Proper Subset adalah Himpunan Bagian murni

   Contoh:

Ada suatu himpunan A={a,b,c,}

Himpunan Bagian dari A adalah

Himpunan Kosong = {}

Himpunan 1 Unsur = {a},{b},{c}

Himpunan 2 Unsur = {a,b},{a,c},{b,c}

Himpunan 3 Unsur = {a,b,c,}

Proper subset-nya adalah himpunan 1 unsur dan himpunan 2 unsur

4. Item set adalah himpunan item atau item-item di I

   Item set-nya adalah {a};{b}:{c};{a,b};{a,c};{b,c}

5. K- item set adalah Item set yang terdiri dari K buah item yang ada pada I. Intinya K itu adalah jumlah unsur yang terdapat pada suatu himpunan

   Contoh:

   3-item set adalah yang bersifat 3 unsur

6. Item set frekuensi adalah jumlah transaksi di I yang mengandung jumlah item set tertentu. Intinya jumlah transaksi yang membeli suatu item set.

   Contoh:

   Digunakan tabel transaksi penjualan sayur di atas, sebagai berikut:

   - frekuensi Item set yang sekaligus membeli beans dan brocolli adalah 3
   - frekuensi item set yang membeli sekaligus membeli beans, squash dan tomatoes adalah 2

7. Frekuen item set adalah item set yang muncul sekurang-kurangnya “sekian” kali di D. Kata “sekian” biasanya di simbolkan dengan Φ. Φ merupakan batas minimum dalam suatu transaksi

Contoh:

Pertama ditentukan Φ = 3, karena jika tidak di tentukan maka frekuen item set tidak dapat di hitung.

Jika Φ = 3 untuk {Asparagus, Beans} apakah frekuen Item set?

Jika kita hitung maka jumlah transaksi yang membeli asparagus sekaligus membeli beans adalah 5.

Karena 5 >= 3 maka {Asparagus, Beans} merupakan Frekuen Item set.

8. *Fk* adalah himpunan semua frekuen item set yang terdiri dari K item.

Langkah-langkah algoritma pada Association Rule

1. Tentukan Φ
2. Tentukan semua frekuen item set
3. Untuk setiap frekuen item set dilakukan hal sbb:
   i. Ambil sebuah unsur, namakanlah s
   ii. Untuk sisanya diberi nama ss-s
   iii. Masukkan unsur-unsur yang telah di umpamakan ke dalam rule If (ss-s) then s

Mengulangi langkah ke-3 untuk dilakukan pada semua unsur.