

BAB II

LANDASAN TEORI

2.1 DATA MINING

Data mining adalah proses mengekstraksi pola-pola yang menarik (tidak remeh-temeh, implisit, belum diketahui sebelumnya, dan berpotensi untuk bermanfaat) dari data yang berukuran besar berdasarkan (JK06). Sedangkan menurut Sid Adelman (2000, p145) data mining adalah proses pencarian pola data yang tidak diketahui atau tidak diperkirakan sebelumnya.

Menurut *Gartner Group* data mining didefinisikan sebagai suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yg tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika.

(*Pramudiono, 2006*) "Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual."

"Data mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk pengenalan permasalahan pengambilan informasi dari database yang besar." (*Larose, 2005*)

2.2 Teknik Data Mining

Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu Kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Perlu diingat bahwa katamining sendiri berarti usaha untuk mendapatkan sedikit data berharga dari sejumlah besar datadasar. Karena itu data mining

sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (artificial intelligent), machine learning, statistik dan basis data. Beberapa teknik yang sering disebut-sebut dalam literatur data mining antara lain yaitu association rule mining, clustering, klasifikasi, neural network, genetic algorithm dan lain-lain. Model maupun hasil analisisnya, salah satunya dengan kemampuan pembelajaran yang dimiliki beberapa teknik data mining seperti klasifikasi.

a. Classification

Suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Aturan-aturan tersebut digunakan pada data-data baru untuk diklasifikasi. Teknik ini menggunakan supervised induction, yang memanfaatkan kumpulan pengujian dari record yang terklasifikasi untuk menentukan kelas-kelas tambahan. Salah satu contoh yang mudah dan populer adalah dengan Decision tree yaitu salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi. Prediksi menggunakan struktur pohon atau struktur berhierarki. Decision tree adalah struktur flowchart yang menyerupai tree (pohon), dimana setiap simpul internal menandakan suatu tes pada atribut, setiap cabang merepresentasikan hasil tes, dan simpul daun merepresentasikan kelas atau distribusi kelas. Alur pada decision tree ditelusuri dari simpul akar ke simpul daun yang memegang prediksi kelas untuk contoh tersebut. Decision tree mudah untuk dikonversi ke aturan klasifikasi (*classification rules*).

B. Association

Digunakan untuk mengenali kelakuan dari kejadian-kejadian khusus atau proses dimana link asosiasi muncul pada setiap kejadian. Contoh dari aturan asosiatif dari analisa pembelian di suatu pasar swalayan adalah bisa diketahui 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. Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, support yaitu prosentasi kombinasi atribut tersebut dalam basis data dan confidence yaitu kuatnya hubungan antar atribut dalam aturan asosiatif. Motivasi awal pencarian association rule berasal dari keinginan untuk menganalisa data transaksi, ditinjau dari perilaku customer. Association rule ini menjelaskan seberapa sering suatu aktifitas muncul. Sebagai contoh, association rule “beer =>diaper (80%)” menunjukkan bahwa empat dari lima customer yang membeli beer juga membeli diaper. Dalam suatu association rule $X \Rightarrow Y$, X disebut dengan antecedent dan Y disebut dengan consequent. Rule.

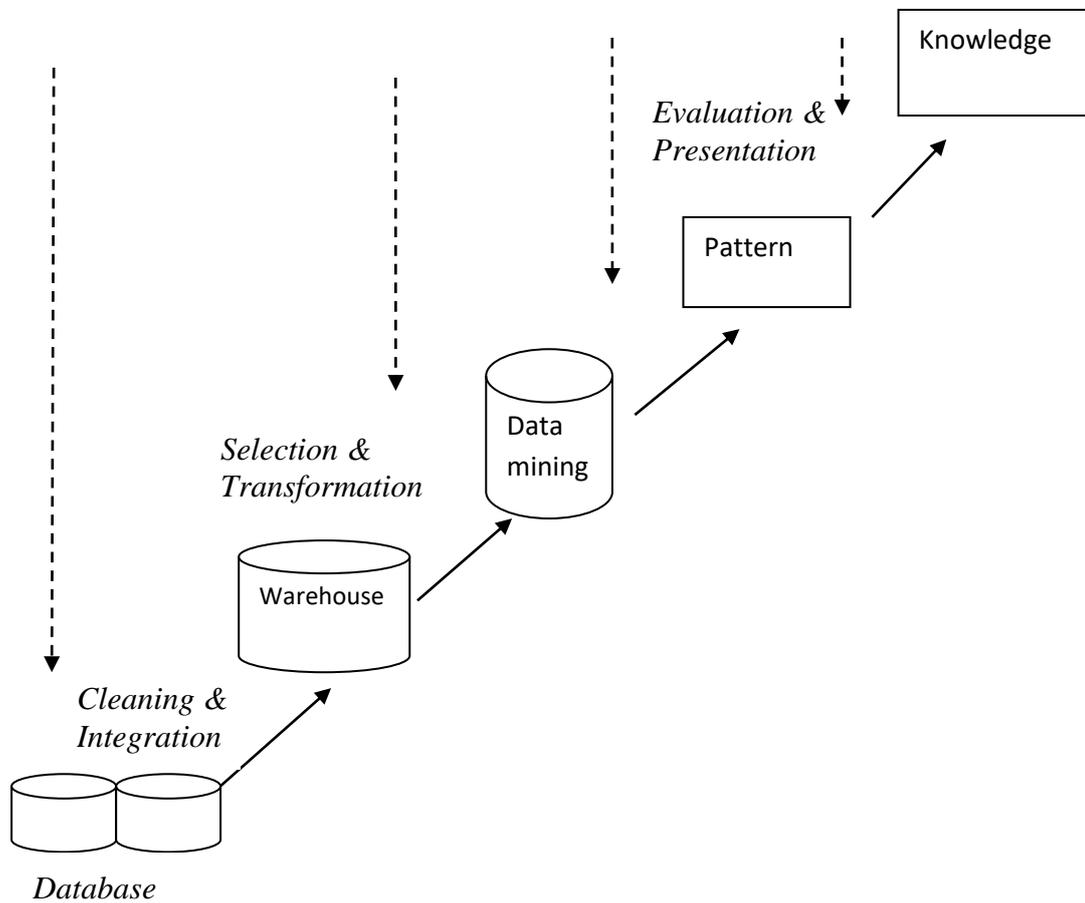
C. Clustering

Digunakan untuk menganalisis pengelompokan berbeda terhadap data, mirip dengan klasifikasi, namun pengelompokan belum didefinisikan sebelum dijalankannya tool data mining. Biasanya menggunakan metode neural network atau statistik. Clustering membagi item menjadi kelompok-kelompok berdasarkan yang ditemukan tool data mining. Prinsip dari clustering adalah memaksimalkan kesamaan antar anggota satu kelas dan meminimumkan kesamaan antar cluster. Clustering dapat dilakukan pada data yang memiliki beberapa atribut yang dipetakan sebagai ruang multidimensi. Dinyatakan dengan bidang dua dimensi, dari pelanggan suatu toko dapat dikelompokkan menjadi beberapa cluster dengan pusat cluster ditunjukkan oleh tanda positif(+). Banyak algoritma clustering memerlukan fungsi jarak untuk mengukur kemiripan antar data, diperlukan juga metoda untuk normalisasi bermacam atribut yang dimiliki data.

2.3 Tahapan Data Mining

Salah satu tuntutan dari data mining ketika diterapkan pada data berskala besar adalah diperlukan metodologi sistematis tidak hanya ketika melakukan analisa saja tetapi juga ketika mempersiapkan data dan juga melakukan interpretasi dari hasilnya sehingga dapat menjadi aksi ataupun keputusan yang bermanfaat. Data mining seharusnya dipahami sebagai suatu proses, yang memiliki tahapan-tahapan tertentu dan juga ada umpan balik dari setiap tahapan ke tahapan

sebelumnya. Pada umumnya proses data mining berjalan interaktif karena tidak jarang hasil data mining pada awalnya tidak sesuai dengan harapan analisnya sehingga perlu dilakukan desain ulang prosesnya.



(Gambar 2.3 Tahap-Tahap Data Mining)

Sebagai suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap yang diilustrasikan di Gambar 1. Tahap-tahap tersebut bersifat interaktif di mana pemakai terlibat langsung atau dengan perantara knowledge base.

1. Pembersihan data

Digunakan untuk membuang data yang tidak konsisten dan noise.

2. Integrasi Data

Data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Hasil integrasi data sering diwujudkan dalam sebuah datawarehouse karena dengan data warehouse, data dikonsolidasikan dengan struktur khusus yang efisien. Selain itu data warehouse juga memungkinkan tipe analisa seperti OLAP.

3. Transformasi data

Transformasi dan pemilihan data ini untuk menentukan kualitas dari hasil data mining, sehingga data diubah menjadi bentuk sesuai untuk di-Mining.

4. Aplikasi Teknik Data Mining

Aplikasi teknik data mining sendiri hanya merupakan salah satu bagian dari proses data mining. Ada beberapa teknik data mining yang sudah umum dipakai.

5. Evaluasi pola yang ditemukan

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

Presentasi pola yang ditemukan untuk menghasilkan aksi tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisa yang didapat. Sesuai yang tercantum dalam buku "Advances

in Knowledge Discovery Dan Data mining” terdapat definisi sebagai berikut: Knowledge discovery (data mining) in databases (KDD) adalah keseluruhan proses non-trivial untuk mencari dan mengidentifikasi pola (pattern) dalam data, dimana pola yang ditemukan bersifat sah (valid), baru (novel), dapat bermanfaat (potentially usefull), dapat dimengerti (ultimately understandable).

2.3.1 Data Selection

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 berbedakan tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining. Proses KDD secara garis besar dapat dijelaskan sebagai beriku: Pemilihan (seleksi) data dan sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisahdari basis data operasional.

2.3.2 Pre-processing/ Cleaning

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi focus 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 enrichment, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.

2.3.3 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

2.3.4 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 secara keseluruhan.

2.3.5 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 dengan interpretation. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya. Proses KDD secara garis besar memang terdiri dari 5 tahap seperti yang telah dijelaskan sebelumnya. Akan tetapi, dalam proses KDD yang sesungguhnya, dapat saja terjadi iterasi atau pengulangan pada tahap-tahap tertentu. Pada setiap tahap dalam proses KDD, seorang analis dapat saja kembali ke tahap sebelumnya. Sebagai contoh, pada saat coding atau data mining, analis menyadari proses cleaning belum dilakukan dengan sempurna, atau mungkin saja analis menemukan data atau informasi baru untuk “memperkaya” data yang sudah ada.

2.4 Arsitektur data mining

Data mining merupakan proses pencarian pengetahuan yang menarik dari data berukuran besar yang disimpan dalam basis data, data warehouse atau tempat penyimpanan informasi lainnya. Dengan demikian arsitektur sistem data mining memiliki komponen-komponen utama yaitu:

1. Basis data, data warehouse atau tempat penyimpanan informasi lainnya.

2. Basis data dan data warehouse server. Komponen ini bertanggung jawab dalam pengambilan relevant data, berdasarkan permintaan pengguna.
3. Basis pengetahuan, Komponen ini merupakan domain knowledge yang digunakan untuk memandu pencarian atau mengevaluasi pola-pola yang dihasilkan. Pengetahuan tersebut meliputi hirarki konsep yang digunakan untuk mengorganisasikan atribut atau nilai atribut ke dalam level abstraksi yang berbeda. Pengetahuan tersebut juga dapat berupa kepercayaan pengguna (*user belief*), yang dapat digunakan untuk menentukan kemenarikan pola yang diperoleh. Contoh lain dari domain knowledge adalah threshold dan metadata yang menjelaskan data dari berbagai sumber yang heterogen.
4. Data mining *engine*, Bagian ini merupakan komponen penting dalam arsitektur sistem data mining. Komponen ini terdiri modul-modul fungsional data mining seperti karakterisasi, asosiasi, klasifikasi, dan analisis cluster.
5. Modul evaluasi pola. Komponen ini menggunakan ukuran-ukuran kemenarikan dan berinteraksi dengan modul data mining dalam pencarian pola-pola menarik. Modul evaluasi pola dapat menggunakan threshold kemenarikan untuk mem-filter pola-pola yang diperoleh.
6. Antarmuka pengguna grafis. Modul ini berkomunikasi dengan pengguna dan sistem data mining. Melalui modul ini, pengguna berinteraksi dengan sistem dengan menentukan kueri atau task data mining. Antarmuka juga menyediakan informasi untuk memfokuskan pencarian dan melakukan eksplorasi data mining berdasarkan hasil data mining antara. Komponen ini juga memungkinkan pengguna untuk mencari (browse) basis data dan skema data warehouse atau struktur data, evaluasi pola yang diperoleh dan visualisasi pola dalam berbagai bentuk. Data mining dapat diaplikasikan pada berbagai jenis penyimpanan data seperti basis data relational, data warehouse, transactional database, objectoriented
7. and object-relational databases, spatial databases, time-series data and temporal data, text databases and multimedia databases, heterogeneous and legacy databases dan WWW.

7.1 Basis data Relasional, Basis data relasional merupakan koleksi dari table. Setiap table berisi atribut (field) dan biasanya menyimpan sejumlah besar tuple (record). Setiap tuple dalam table relasional merepresentasikan sebuah objek yang diidentifikasi oleh kunci unik dan dideskripsikan oleh sekumpulan nilai atribut. Data relasional dapat diakses oleh kueri basis data yang ditulis dalam bahasa kueri relasional seperti SQL atau dengan bantuan antarmuka pengguna grafis.

7.2 Data warehouse, Data warehouse merupakan tempat penyimpanan informasi yang dikumpulkan dari berbagai sumber, disimpan dalam skema yang dipersatukan (unified schema) dan biasanya bertempat pada tempat penyimpanan tunggal. Data warehouse dikonstruksi melalui sebuah proses data cleaning, data transformation, data integration, data loading dan periodic data refreshing. Untuk memfasilitasi proses pembuatan keputusan, data dalam data warehouse diorganisasikan ke dalam subjek utama seperti customer, item, supplier atau aktivitas. Data disimpan untuk menyediakan informasi dari perspektif sejarah (seperti 5-10 tahun yang lalu) dan biasanya data tersebut diringkas (summarized). Sebagai contoh, daripada menyimpan data rinci dari transaksi penjualan, data warehouse dapat menyimpan ringkasan dari transaksi per tipe item untuk setiap toko atau diringkas dalam level yang lebih tinggi seperti daerah pemasaran.

Data warehouse biasanya dimodelkan oleh struktur basis data multidimensional, dimana setiap dimensi berkaitan dengan sebuah atribut atau sekumpulan atribut dalam skema, dan setiap sel menyimpan nilai dari ukuran agregasi seperti count dan sales_amount. Struktur fisik dari data warehouse dapat berupa penyimpanan basis data relasional atau sebuah kubus data multidimensional. Selain data warehouse, terdapat istilah penyimpanan data yang lain yaitu data mart. Sebuah data warehouse mengumpulkan informasi mengenai subjek-subjek yang menjangkau seluruh organisasi, dengan demikian cakupannya enterprise-wide. Sedangkan data mart merupakan sub bagian dari data warehouse. Fokus data mart adalah pada subjek yang dipilih dan dengan demikian cakupannya adalah department-wide.

7.3 Basis data Transaksional Secara umum, basis data transaksional terdiri dari sebuah file dimana setiap record merepresentasikan transaksi. Sebuah transaksi biasanya meliputi bilangan identitas transaksi yang unik (*trans_id*), dan sebuah daftar dari item yang membuat transaksi (seperti item yang dibeli dalam sebuah *took*). Basis data transaksi dapat memiliki tabel tambahan, yang mengandung informasi lain berkaitan dengan penjualan seperti tanggal transaksi, customer ID number, ID number dari sales person dan dari kantor cabang (*branch*) dimana penjualan terjadi.

2.5 Pengertian Association Rules

Association rules (aturan asosiasi) atau *affinity analysis* (analisis afinitas) berkenaan dengan studi tentang “apa bersama apa”.

Aturan asosiasi ingin memberikan informasi tersebut dalam bentuk hubungan “if-then” atau “jika-maka”. Aturan ini dihitung dari data yang sifatnya probabilistik (Santoso, 2007).

Tahap dari analisis asosiasi yang disebut analisis pola frekuensi tinggi (*frequent pattern mining*) menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien. Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, *support* (nilai penunjang) yaitu prosentase kombinasi item tersebut dalam *database* dan *confidence* (nilai kepastian) yaitu kuatnya hubungan antar item dalam aturan asosiatif. Analisis asosiasi didefinisikan suatu proses untuk menemukan semua aturan asosiatif yang memenuhi syarat minimum untuk *support* (*minimum support*) dan syarat minimum untuk *confidence* (*minimum confidence*) (Pramudiono, 2007). Analisis pola frekuensi tinggi menggunakan satu item dan secara rekursif mengembangkan *frequent itemset* dengan dua item, tiga item dan seterusnya hingga *frequent itemset* dengan semua ukuran. Untuk mengembangkan *frequent set* dengan dua item, dapat menggunakan *frequent set item*. Alasannya adalah bila set satu item tidak melebihi *support minimum*, maka sembarang ukuran itemset yang lebih besar tidak akan melebihi *support minimum* tersebut. (Ridwan yusuf, 2012)

Secara umum, mengembangkan set dengan fc-item menggunakan frequent set dengan $k-1$ item yang dikembangkan dalam langkah sebelumnya. Setiap langkah memerlukan sekali pemeriksaan ke seluruh isi *database*. Dalam asosiasi terdapat istilah *antecedent* dan *consequent*, *antecedent* untuk mewakili bagian “jika” dan *consequent* untuk mewakili bagian “maka”. Dalam analisis ini, *antecedent* dan *consequent* adalah sekelompok item yang tidak punya hubungan secara bersama (Santosa, 2007).

Dari jumlah besar aturan yang mungkin dikembangkan, perlu memiliki aturan-aturan yang cukup kuat tingkat ketergantungan antar item dalam *antecedent* dan *consequent*. Untuk mengukur kekuatan aturan asosiasi ini, digunakan ukuran *support* dan *confidence*. *Support* adalah rasio antara jumlah transaksi yang memuat *antecedent* dan *consequent* dengan jumlah transaksi. *Confidence* adalah rasio antara jumlah transaksi yang meliputi semua item dalam *antecedent* dan *consequent* dengan jumlah transaksi yang meliputi semua item dalam *antecedent*.

1. Langkah pertama algoritma adalah: *Support* dari setiap item dihitung dengan men-scan *database*. Setelah *support* dari setiap item didapat, item yang memiliki *support* lebih besar dari *minimum support* dipilih sebagai pola frekuensi tinggi dengan panjang 1 atau sering disingkat 1-itemset. Singkatan k-itemset berarti satu set yang terdiri dari k item.
2. Iterasi kedua menghasilkan 2-itemset yang tiap set-nya memiliki dua item. Pertama dibuat kandidat 2-itemset dari kombinasi semua 1-itemset. Lalu untuk tiap kandidat 2-itemset ini dihitung *support*-nya dengan men-scan *database*. *Support* artinya jumlah transaksi dalam *database* yang mengandung kedua item dalam kandidat 2-itemset. Setelah *support* dari semua kandidat 2-itemset didapatkan, kandidat 2-itemset yang memenuhi syarat *minimum support* dapat ditetapkan sebagai 2-itemset yang juga merupakan pola frekuensi tinggi dengan panjang 2. (Pramudiono, 2007)
3. Untuk selanjutnya iterasi iterasi ke-k dapat dibagi lagi menjadi beberapa bagian.

Pembentukan kandidat itemset

Kandidat k-itemset dibentuk dari kombinasi (k-1)-itemset yang didapat dari iterasi sebelumnya. Pemangkasan kandidat k-itemset yang subset-nya yang berisi k-1 item tidak termasuk dalam pola frekuensi tinggi dengan panjang k-1. (Santosa, 2007)

2. Penghitungan support dari tiap kandidat k-itemset

Support dari tiap kandidat k-itemset didapat dengan men-scan *database* untuk menghitung jumlah transaksi yang memuat semua item di dalam kandidat k-itemset tersebut. Penghitungan dengan scan seluruh *database* sebanyak k-itemset terpanjang.

$$\text{Support (A)} = \frac{\text{Jumlah Transaksi mengandung A}}{\text{Total Transaksi}} \dots\dots(2.1)$$

3. Tetapkan pola frekuensi tinggi

Pola frekuensi tinggi yang memuat k item atau k-itemset ditetapkan dari kandidat k-itemset yang support-nya lebih besar dari *minimum support*. Kemudian dihitung *confidence* masing-masing kombinasi item. Iterasi berhenti ketika semua item telah dihitung sampai tidak ada kombinasi item lagi.

$$\text{Confidence} = P(B|A) = \frac{\text{Jumlah Transaksi mengandung A dan B}}{\text{Jumlah Transaksi mengandung A}} \dots(3.1)$$

2.6 Algoritma C4.5

Algoritma C4.5 merupakan kelompok algoritma Decision Tree. Algoritma ini mempunyai input berupa training samples dan samples. Training samples berupa data contoh yang akan digunakan untuk membangun sebuah tree yang telah diuji kebenarannya. Sedangkan samples merupakan field-field data yang nantinya akan digunakan sebagai parameter dalam melakukan klasifikasi data (Sunjana, 2010). Algoritma C 4.5 adalah salah satu metode untuk membuat decision tree berdasarkan training data yang telah disediakan. Algoritma C 4.5 dibuat oleh Ross

Quinlan yang merupakan pengembangan dari ID3 yang juga dibuat oleh Quinlan (Quinlan, 1993). Beberapa pengembangan yang dilakukan pada C4.5 antara lain adalah bisa mengatasi missing value, bisa mengatasi continue data, dan pruning.

2.7 Tanagra

Tanagra adalah software data mining gratis untuk tujuan akademik dan penelitian. Tanagra mengusulkan beberapa metode data mining dari eksplorasi analisis data, pembelajaran statistik, mesin pembelajaran dan area database. Algoritma untuk ekstraksi association rules pada awalnya dikembangkan untuk menemukan kaitan yang logis antara variabel dengan status yang sama. Asosiasi prediktif aturan pencarian asosiasi antara item yang mencirikan atribut yang saling tergantung. Pada dasarnya, algoritma ini tidak benar-benar diubah. Eksplorasi adalah hanya terbatas pada itemset yang termasuk variabel dependent. Perhitungan waktu kemudian dikurangi. Dua komponen Tanagra yang didedikasikan untuk tugas ini adalah ASSOC SPV dan SPV RULE TREE ASSOC yang berada dalam tabulasi ASOSIASI. Dibandingkan dengan pendekatan konvensional, komponen Tanagra memiliki spesifisitas tambahan dimana kita bias menentukan class value ("variabel dependent = value") yang kita inginkan untuk memprediksi. Keuntungannya adalah bahwa kita dapat mengatur parameter algoritma, secara langsung dalam kaitannya dengan karakteristik data. Hal ini penting misalnya ketika probabilitas sebelumnya dari nilai-nilai variabel dependent sangat berbeda (Ricco, 2005).

2.8 Pengertian PHP

Dikutip dari sumber wikipedia menjelaskan bahwa pengertian PHP adalah bahasa pemrograman script server-side yang didesain untuk pengembangan web. Sumber ini juga menjelaskan bahwa PHP juga bisa digunakan sebagai bahasa pemrograman umum. Pada awalnya PHP di kembangkan pada tahun 1995 oleh Rasmus Lerdorf, dan sekarang dikelola oleh The PHP Group, situs resmi PHP beralamat di <https://www.php.net>. Kemudian bahwa pengertian PHP disebut sebagai bahasa pemrograman server side karena PHP diproses pada komputer server yang selanjutnya ditampilkan pada halaman web. Berbeda dengan

JavaScript sebagai bahasa pemrograman client-side yang melakukan proses pada web browser atau client. Penjelasan pengertian php dan fungsinya sampai di sini apakah sudah dapat dipahami? karena apa itu php akan kita bahas sedikit lebih jauh lagi. Menurut para ahli pengertian PHP: Hypertext Preprocessor adalah bahasa skrip yang dapat ditanamkan atau disisipkan ke dalam HTML sehingga PHP banyak dipakai untuk memprogram situs web dinamis. Pada awalnya PHP merupakan singkatan dari Personal Home Page sebelum menjadi PHP: Hypertext Preprocessor. Oleh karena itu awalnya PHP hanya digunakan untuk membuat website pribadi. Kemudian karena perkembangannya, PHP akhirnya menjadi bahasa pemrograman web yang powerful dan sangat dinamis, sehingga tidak hanya digunakan untuk membuat halaman web sederhana, tetapi juga website populer yang digunakan oleh jutaan orang seperti wikipedia dan CMS wordpress. Terakhir menguak apa itu php adalah pada lisensinya. PHP dapat digunakan secara gratis dan bersifat Open Source. PHP dirilis dalam lisensi PHP License, berbeda dengan lisensi GNU General Public License (GPL) yang biasa digunakan untuk proyek Open Source.

2.9 Penelitian Terkait Terdahulu

No	Judul Artikel Ilmiah	Penulis	Tahun	Kesimpulan
1	Penerapan Association Rule Menggunakan Spv Assoc Rule Untuk Menentukan Faktor Penghambat Keberhasilan Siswa Sekolah Dasar Dalam Belajar	1. Ridwan Yusuf	2012	Penggalian association rule mempunyai peranan penting dalam proses pengambilan keputusan. Tahapan besar dari penggalian association rule adalah mengidentifikasi <i>frequent</i> itemset dan membentuk kaidah asosiasi dari itemset tersebut. association rule digunakan untuk menggambarkan hubungan antar item pada tabel data transaksional ataupun data relasional. Tapi semakin berkembangnya teknologi komputer di dunia industri,

				<p>semakin pesat pula perkembangan ukuran data yang dihasilkan. Dan pada data yang besar tersebut, proses pencarian <i>frequent itemset</i> sangatlah sulit.</p> <p>Penelitian ini bertujuan untuk menentukan faktor penghambat siswa sekolah dasar dalam belajar. Penentuan tersebut dengan melihat keterkaitan data hasil UASBN, status sekolah dan latar belakang serta kondisi siswa SD di Kota Metro dengan mengimplementasikan metode association rule menggunakan SPV Assoc Rule. Metode association rule dalam penelitian ini menggunakan minimum support 33% dan minimum confidence 75%.</p> <p>Pada tahap pengujian dalam penelitian dilakukan menggunakan hasil frequent 1-itemset, frequent 2-itemset dan frequent 3-itemset. Pengujian lanjutan tidak menggunakan frequent 1-itemset karena memperhatikan antecedent \neq consequent. Dalam pengujian fokus kepada itemset yang berhubungan dengan dengan hasil uasbn yaitu dengan rule if x then y dimana y adalah consequent. Hasil pengujian diperoleh 3 buah rule terbaik dengan consequent hasil UASBN yang Rendah.</p>
--	--	--	--	--

2	Implementasi Data Mining Untuk Memprediksima sa Studi Mahasiswa Menggunakan Algoritma C4.5(Studi Kasus: Universitas Dehasen Bengkulu)	1.Siska Haryati 2. Aji Sudarsono 3.Eko Suryono	Jurnal Media Infotama Vol. 11 No. 2, September 2015	Tujuan dari penelitian ini adalah dengan menggunakan pohon keputusan berbasis algoritma C4.5 dan dimplementasikan ke suatu aplikasi yaitu RapidMiner diharapkan dapat meningkatkan keakuratan analisa masa studi mahasiswa. Pada penelitian dibahas dengan memantau hasil belajar di universitas berupa nilai IPK dan Jumlah SKS yang belum akurat untuk menentukan seorang mahasiswa lulus tepat waktu atau tidak. Pada Penelitian ini untuk mengklasifikasikan kelulusan mahasiswa digunakan teknik data mining dengan algoritma C4.5 dan diImplementasikan ke Rapid Miner, hal tersebut bertujuan untuk melihat hasil perkembangan mahasiswa apakah dapat lulus tepat waktu atau tidak. Dari hasil penelitian terbukti bahwa algoritma C4.5 lebih akurat dibandingkan analisa yang dilakukan oleh analis mahasiswa. Hal ini dibuktikan dengan hasil evaluasi penelitian bahwa algoritma C4.5 mampu menganalisa tingkat ketepatan waktu mahasiswa menyelesaikan masa studinya. Kata Kunci: Data mining, Algoritma C4.5
3	Penentuan Tingkat Kelulusan Tepat Waktu Mahasiswa	1.Hermansyah 2.Vincent 3.Ika Novita	Jurnal Teknologi Informasi, Volume 13 Nomor 1,	Timely graduatuion rates in college could not be consideres easy and trivial. Many cases found that the share of the number of

	<p>Stmik Subang Menggunakan Algoritma C4.5</p>		<p>Januari 2017, ISSN 1907-3380</p>	<p>students who did not get in and who have completed their studies so that the build up of high numbers of students in every period. I need to know the factors cause students not graduating on time. Classification data mining techniques can be used to predict student graduation rates. The algorithm used is algorithmic C 4.5 with data as much as 200 students study computer engineering programs STMIK Subang. The result of the classification process is evaluated by using the confusion matrix, ROC Curve, Recall. Based on experimental results and evaluation is done then it can be inferred that the algorithm C 4.5 accurately applied to determine the level of students graduation. After testing the prediction accuracy resulting from trials reached 95,00% of the classification result generate information in the from of graph in the form of the curve results from the decision tree that is useful for institutions of higher education in taking policy. Keywords : Classification, C 4.5 Algorithm, Graduation Rate</p>
--	--	--	---	--