

BAB II

LANDASAN TEORI

Pada bab ini akan diuraikan teori yang menjadi landasan pembuatan tugas akhir ini, yaitu : penelitian sebelumnya, *Data Mining, Decision Tree*, fitur yang dipilih, pemilihan fitur, kondisi ibu, macam obat, algoritma FDR dan algoritma *Decision Tree ID3*.

2.1. Penelitian Sebelumnya

Penelitian sebelumnya yang menggunakan metode *decision tree C4.5* adalah Budanis Dwi Meilani Achmad dan Fauzi Slamet, lulusan universitas ITATS. Penelitian mereka berjudul klasifikasi data karyawan untuk menentukan jadwal kerja menggunakan metode *decision tree*. Atribut yang dilakukan sebagai data latih sistem adalah NP (Nomor Pegawai), nama, umur, jenis kelamin, pendidikan, agama, level, lama kerja, kesehatan dan jadwal. Atribut jadwal digunakan sebagai kelas data.

Pre processing pada penelitian tersebut menghilangkan fitur data yang tidak memiliki pola tertentu dan jumlah yang cukup besar. Setelah dilakukan *pre processing*, data dihitung dengan metode *decision tree C4.5*. Terbentuklah pohon keputusan yang akarnya adalah atribut umur dengan masing-masing node adalah jenis kelamin, pendidikan dan level. Setelah pohon keputusan terbentuk, akan diketahui rule untuk membentuk sebuah jadwal kerja yang baru. Untuk menguji apakah rule tersebut sudah sesuai, maka dilakukan pengujian dengan data-data asli.

Berdasarkan pengujian tersebut, diperoleh nilai akurasi dari rule yang telah terbentuk. Nilai akurasi inilah yang akan menentukan rule digunakan atau tidak. Rule yang digunakan untuk membuat jadwal kerja yang baru sebagai berikut :

- Jika umur = umur3 maka jadwal = B
- Jika umur = umur3 dan jenis kelamin = L maka jadwal = A
- Jika umur = umur3 dan jenis kelamin = P maka jadwal = B

- Jika umur = umur3 dan pendidikan = SMA maka jadwal = A
- Jika umur = umur3 dan pendidikan = SMA dan level = junior maka jadwal = A
- Jika umur = umur3 dan pendidikan = SMA dan level = senior maka jadwal = B

Rule tersebut berasal dari data bulan januari 2011. Tingkat akurasi pohon keputusan mencapai 87%. Kesimpulan dari penelitian tersebut, yakni : rule yang digunakan adalah rule januari 2011 karena rule tersebut memiliki nilai akurasi paling tinggi yaitu sebesar 87%, atribut hasil yang digunakan adalah atribut jadwal. Jika atribut lain yang digunakan, maka pohon keputusan tidak akan terbentuk karena kurangnya jumlah data dan data baru yang diinputkan harus memiliki atribut pada pohon. Jika ada salah satu atribut pada pohon yang tidak disertakan pada data baru, maka jadwal kerja yang baru tidak akan terbentuk. (Budanis D.M., Achmad. 2012).

2.2. Data Mining

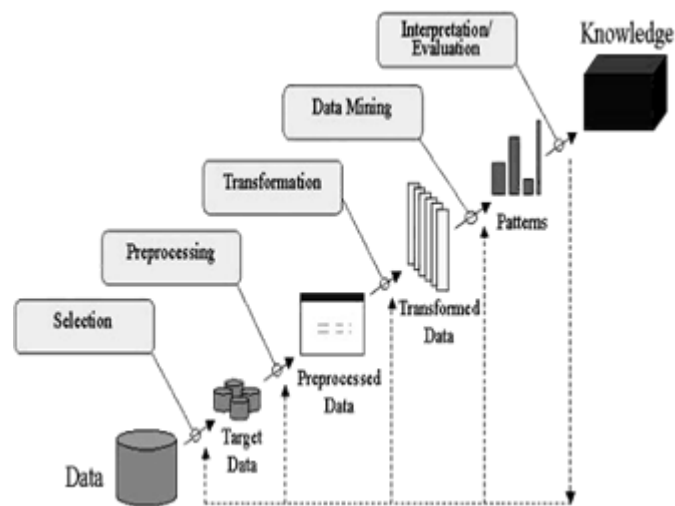
2.2.1. Pengertian Data Mining

Kehadiran *data mining* dilatar belakangi dengan problema data explosion yang dialami akhir-akhir ini dimana banyak organisasi telah mengumpulkan data sekian tahun lamanya (data pembelian, data penjualan, data nasabah, data transaksi dan sebagainya). Hampir semua data tersebut dimasukkan dengan menggunakan aplikasi komputer yang digunakan untuk menangani transaksi sehari-hari yang kebanyakan adalah OLTP (*On Line Transaction Processing*).

Data mining adalah sebuah proses untuk menemukan pola atau pengetahuan yang bermanfaat secara otomatis atau semi otomatis dari sekumpulan data dalam jumlah besar. *Data mining* hadir dianggap sebagai bagian dari *Knowledge Discovery in Database (KDD)* yaitu sebuah proses mencari pengetahuan yang bermanfaat dari data. (Santosa, Budi. 2007). KDD terdiri dari beberapa langkah yaitu :

1. Pembersihan data (membuang noise dan data yang tidak konsisten)
2. Integrasi data (penggabungan data dari beberapa sumber)

3. Seleksi data (memilih data yang relevan yang akan digunakan untuk analisa)
4. *Data mining*
5. Evaluasi pola
6. Presentasi pengetahuan dengan teknik visualisasi



Gambar 2.1. Proses Knowledge Discovery in Database

2.2.2. Metode *Data Mining*

Pada umumnya metode *data mining* dapat dikelompokkan kedalam dua kategori yaitu *deskriptif* dan *prediktif*. Metode *deskriptif* bertujuan untuk mencari pola yang dapat dimengeti oleh manusia yang menjelaskan karakteristik dari data. Metode *prediktif* menggunakan ciri-ciri tertentu dari data untuk melakukan prediksi.

Metode-metode yang ada dalam data mining adalah sebagai berikut:

1. Classification

Klasifikasi (*Classification*) merupakan proses untuk menemukan sekumpulan model yang menjelaskan dan membedakan kelas-kelas data, sehingga model tersebut dapat digunakan untuk memprediksi nilai suatu kelas yang belum diketahui pada sebuah objek. Untuk mendapatkan model, kita harus melakukan analisis terhadap data latih (*training set*). Sedangkan data uji (*test set*) digunakan untuk mengetahui tingkat akurasi dari model yang telah dihasilkan. Dalam proses klasifikasi pohon keputusan tradisional, fitur (atribut) dari tupel adalah kategorikal atau numerikal. Biasanya definisi ketepatan nilai

(point value) sudah didefinisikan di awal. Pada banyak aplikasi nyata, terkadang muncul suatu nilai yang tidak pasti. (Tsang, Smith., 2009). Klasifikasi dapat digunakan untuk memprediksi nama atau nilai kelas dari suatu objek data. Metode inilah yang digunakan dalam tugas akhir ini.

2. *Clustering*

Pengelompokkan (*Clustering*) merupakan proses untuk melakukan segmentasi. Digunakan untuk melakukan pengelompokkan secara alami terhadap atribut suatu set data. Termasuk kedalam *unsupervised task*. Contoh *clustering* seperti mengelompokkan dokumen berdasarkan topiknya.

3. *Association*

Tujuan dari metode ini yaitu untuk menghasilkan sejumlah rule yang menjelaskan sejumlah data yang terhubung kuat satu dengan yang lainnya. Sebagai contoh *association analysis* dapat digunakan untuk menentukan produk yang datang dibeli secara bersamaan oleh banyak pelanggan atau bisa juga disebut dengan *market basket analysis*.

4. *Regression*

Regression mirip dengan klasifikasi. Perbedaan utamanya adalah terletak pada atribut yang diprediksi berupa nilai yang kontinyu.

5. *Forecasting*

Prediksi (*Forecasting*) berfungsi untuk melakukan prediksi kejadian yang akan datang berdasarkan data sejarah yang ada.

6. *Sequence Analysis*

Tujuan dari metode ini adalah untuk mengenali pola dari data diskrit. Sebagai contoh adalah menemukan kelompok *gen* dengan tingkat ekspresi yang mirip.

7. *Deviation Analysis*

Tujuan dari metode ini adalah untuk menemukan penyebab perbedaan antara data yang satu dengan data yang lain dan biasa disebut sebagai *oulier detection*. Sebagai contoh adalah apakah sudah terjadi penipuan terhadap pengguna kredit dengan melihat catatan transaksi yang tersimpan dalam basis data perusahaan kartu kredit. (Santosa, Budi. 2007).

2.3. Decision Tree

2.3.1. Pengertian Decision Tree

Decision tree merupakan metode klasifikasi *data mining*. *Decision tree* dalam istilah pembelajaran merupakan sebuah struktur pohon dimana setiap *node* pohon mempresentasikan atribut yang telah diuji. Setiap cabang merupakan suatu pembagian hasil uji dan *node* daun (*leaf*) mempresentasikan kelompok kelas tertentu. (Jianwei, Han. 2001). Level *node* teratas dari sebuah *Decision Tree* adalah *node* akar (*root*) yang biasanya berupa atribut yang paling memiliki pengaruh terbesar pada suatu kelas tertentu. Pada umumnya *Decision Tree* melakukan strategi pencarian secara top-down untuk solusinya. Pada proses mengklasifikasi data yang tidak diketahui, nilai atribut akan diuji dengan cara melacak jalur dari *node* akar (*root*) sampai *node* akhir (daun) dan kemudian akan diprediksi kelas yang dimiliki oleh suatu data baru tertentu. (Santosa, Budi. 2007).

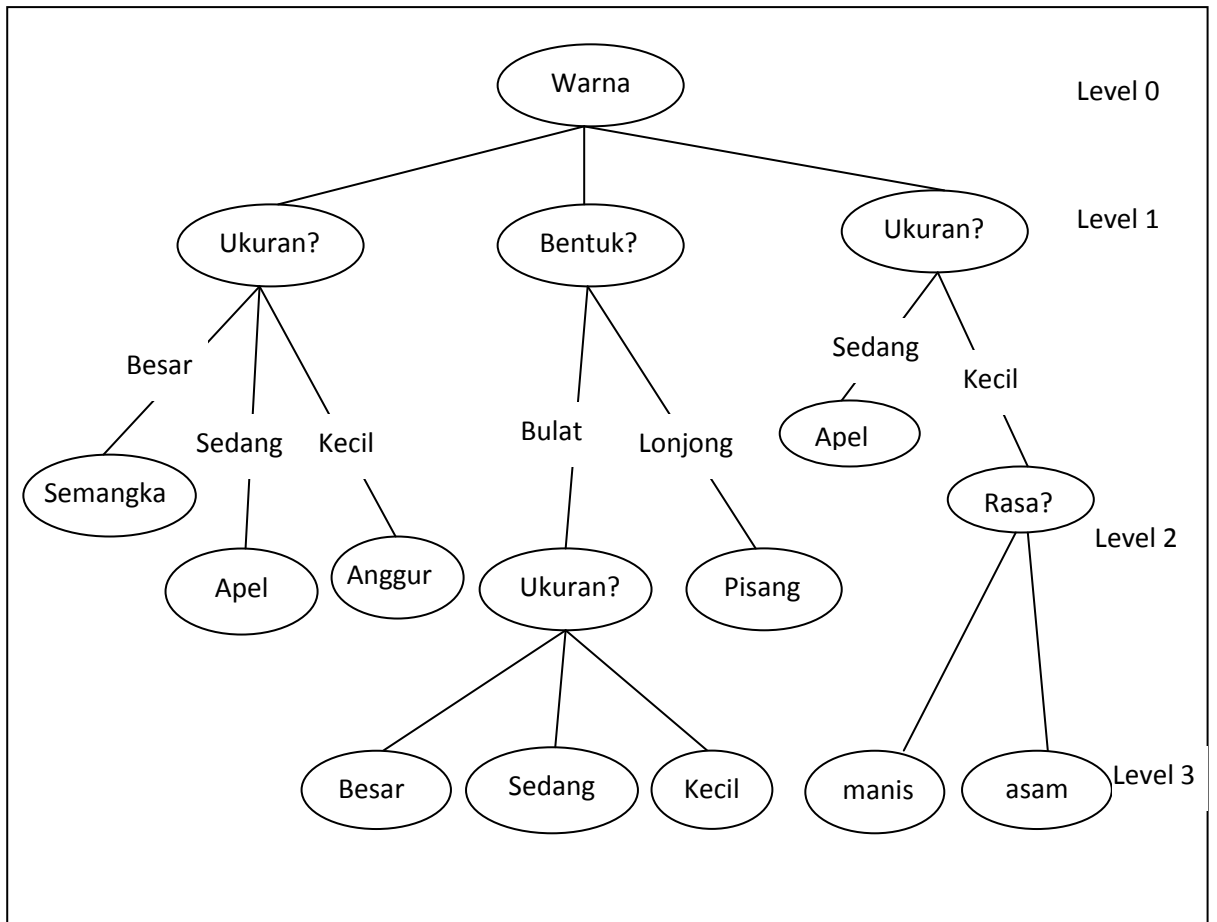
2.3.2. Ciri Kasus Untuk Diterapkan Pada Decision Tree

Decision tree sesuai digunakan untuk kasus-kasus dimana outputnya bernilai diskrit. Walaupun banyak variasi model *decision tree* dengan tingkat kemampuan dan syarat yang berbeda, pada umumnya beberapa ciri kasus berikut cocok untuk diterapkan pada *decision tree* :

1. Data / *example* dinyatakan dengan pasangan atribut dan nilainya.

Misalnya atribut satu *example* adalah temperatur dan nilainya adalah dingin. Biasanya untuk satu *example* nilai dari satu atribut tidak terlalu banyak jenisnya. Dalam contoh atribut warna ada beberapa nilai yang mungkin yaitu hijau, kuning, merah. Sedang dalam atribut temperatur, nilainya bisa dingin, sedang atau panas. Tetapi untuk beberapa kasus bisa saja nilai temperatur berupa numerik.

Berikut ini disajikan contoh *decision tree* yang bisa mempunyai lebih dari dua nilai diskrit :



Gambar 2.2 *Decision Tree*; atribut bisa mempunyai lebih dari dua nilai diskrit.

2. Label / output data biasanya bernilai diskrit

Output ini biasanya bernilai ya atau tidak, sakit atau tidak sakit, diterima atau ditolak. Dalam beberapa kasus mungkin saja outputnya tidak hanya satu kelas. Tetapi, penerapan *decision tree* lebih banyak untuk kasus binary.

3. Data mempunyai *missing value*

Misalkan untuk beberapa *example*, nilai dari suatu atributnya tidak diketahui. Dalam keadaan seperti ini *decision tree* masih mampu memberi solusi yang baik. Contoh model *decision tree* ditunjukkan pada gambar 2.2. dengan cara ini akan mudah mengelompokkan objek ke dalam beberapa kelompok. Dalam *decision tree* setiap atribut ditanyakan di simpul. Jawaban dari atribut ini dinyatakan dalam cabang sampai akhirnya ditemukan kategori atau jenis dari suatu objek di simpul terakhir. (Santosa, Budi. 2007). Untuk membuat *decision tree* kita perlu memperhatikan hal-hal sebagai berikut :

1. Atribut mana yang akan dipilih untuk pemisahan objek.
2. Urutan atribut mana yang akan dipilih terlebih dahulu.
3. Struktur *tree*.
4. Kriteria pemberhentian.
5. *Prunning*.

2.3.3. Jenis-Jenis *Decision Tree*

Beberapa model *decision tree* yang sudah dikembangkan antara lain C4.5 atau ID3 dan CART. Berikut ini akan dijelaskan model dari *decision tree* tersebut:

1. C4.5 atau ID3

Decision Tree menggunakan algoritma ID3 atau C4.5, yang diperkenalkan dan dikembangkan pertama kali oleh Quinlan yang merupakan singkatan dari *Iterative Dichotomiser 3* atau *Induction of Decision 3*. Algoritma ID3 membentuk pohon keputusan dengan metode *divide-and-conquer* data secara rekursif dari atas ke bawah. Strategi pembentukan *Decision Tree* dengan algoritma ID3 adalah:

- A. Pohon dimulai sebagai node tunggal (akar/root) yang merepresentasikan semua data.
- B. Sesudah node root dibentuk, maka data pada node akar akan diukur dengan *information gain* untuk dipilih atribut mana yang akan dijadikan atribut pembagiannya.
- C. Sebuah cabang dibentuk dari atribut yang dipilih menjadi pembagi dan data akan didistribusikan ke dalam cabang masing-masing.
- D. Algoritma ini akan terus menggunakan proses yang sama atau bersifat rekursif untuk dapat membentuk sebuah *Decision Tree*. Ketika sebuah atribut telah dipilih menjadi node pembagi atau cabang, maka atribut tersebut tidak diikuti lagi dalam penghitungan nilai *information gain*.
- E. Proses pembagian rekursif akan berhenti jika salah satu dari kondisi dibawah ini terpenuhi :
 - a. Semua data dari anak cabang telah termasuk dalam kelas yang sama.

- b. Semua atribut telah dipakai, tetapi masih tersisa data dalam kelas yang berbeda. Dalam kasus ini, diambil data yang mewakili kelas yang terbanyak untuk menjadi label kelas pada node daun. Tidak terdapat data pada anak cabang yang baru. Dalam kasus ini, node daun akan dipilih pada cabang sebelumnya dan diambil data yang mewakili kelas terbanyak untuk dijadikan label kelas.

Metode C4.5 dan ID3 memiliki perbedaan dalam nilai tiap atribut. Metode C4.5 menggunakan atribut yang bernilai kategorikal dan numerikal, sedangkan metode ID3 menggunakan atribut yang bernilai kategorikal. Metode *decision tree ID3* inilah yang digunakan dalam tugas akhir ini.

2. CART

CART adalah singkatan dari *Classification And Regression Tree*. Dalam CART ada dua langkah penting yang harus diikuti untuk mendapatkan *tree* dengan performansi yang optimal. Yang pertama adalah pemecahan objek secara berulang berdasarkan atribut tertentu. Yang kedua, *prunning* (pemangkasan) dengan menggunakan data validasi.

Misalkan kita mempunyai variabel independent $x_1, x_2, x_3, \dots, x_n$ dan variabel dependent atau output y . Pemecahan secara berulang berarti kita bagi objek ke dalam kotak-kotak berdasarkan nilai variabel x_1, x_2 atau x_r . Cara ini diulang sehingga dalam suatu kotak sebisa mungkin berisi observasi dalam kelompok atau kelas yang sama.

Langkah berikutnya sesudah dilakukan pemecahan objek atau data secara berulang adalah melakukan *prunning*. Dalam *prunning* kita ingin memangkas *tree* yang mungkin terlalu besar dan terjadi fenomena *overfitting*. *Overfitting* merupakan sebuah satu buah pengelompokan yang mungkin hanya berisi satu data yang memungkinkan data tersebut merupakan *noise* yang ada di data training dan bukan pola yang mungkin terjadi dalam data testing atau data validasi. *Prunning* terdiri dari beberapa langkah pemilihan secara berulang simpul yang akan dijadikan simpul daun. Dengan mengubah simpul menjadi

simpul daun artinya tidak akan dilakukan pemecahan lagi sesudah itu. Dengan demikian ukuran *tree* akan berkurang. (Santosa, Budi. 2007).

2.4. Pemilihan Fitur

Pemilihan fitur disebut juga pra-pemrosesan yakni pekerjaan yang dilakukan sebelum masuk ke proses inti pengenalan pola. Tujuan dari pemilihan fitur, yakni:

1. Mempercepat komputasi algoritma utama
2. Menghindari error yang besar
3. Menyeimbangkan kondisi data yang diproses

Pemilihan fitur adalah cara untuk memilih fitur yang paling penting diantara kandidat tersebut sehingga dapat mengurangi jumlahnya dan memungkinkan memberikan diskriminasi kelas dengan baik. Cara yang digunakan adalah dengan menguji fitur satu per satu dan mengombinasikan pengujian pada dua fitur atau lebih. Pemilihan fitur ini dapat dilakukan dengan metode Fisher's Discriminant Rasio (FDR). Metode FDR digunakan untuk mengukur kekuatan diskriminasi fitur individu dalam memisahkan dua kelas. (Prasetyo, Eko. 2013).

2.5. Kondisi Ibu

Ketika ibu melahirkan seorang bayi, ibu memerlukan imunisasi, imunisasi diberikan apabila ibu mengalami keadaan sebagai berikut ini :

1. Normal

Keadaan atau kondisi ibu dikatakan normal apabila ibu tidak mengalami faktor perinatal, sehingga tidak menyebabkan ibu mengalami neonatus.

2. PER

Preeklampsia adalah timbulnya hipertensi di sertai proteinuria edema akibat kehamilan setelah usia kehamilan 20 hminggu atau segera setelah persalinan. PER (Pre-eklampsia ringan) adalah timbulnya hipertensi yang disertai protein urine dan atau idema setelah 20 minggu, hipertensi mengalami kenaikan tekanan diastolik 15 mmhg atau >90 mmhg dalam 2 pengukuran berjarak 1 jam atau tekanan diastolic sampai 110 mmhg dan adanya protein urine.

3. BSC

Apabila keadaan ibu Bsc, berarti ibu pernah melakukan operasi cesaer sebelum melahirkan lagi yang akan memberikan resiko dalam melahirkan.

4. Febris

Febris (demam) yaitu meningkatnya temperature tubuh secara abnormal (Asuhan Keperawatan Anak 2001). Suhu tubuh yang melewati batas normal yaitu lebih dari 38 0C (Fadjari Dalam Nakita 2003). Febris (demam) yaitu merupakan respon yang sangat berguna dan menolong tubuh dalam memerangi infeksi (KesehatanAnak 1999). Menurut Suriadi (2001), demam adalah meningkatnya temperatur suhu tubuh secara abnormal.

5. Fase Laten

Keadaan ibu disebut fase laten, bila mana kontraksi makin teratur dan pembukaan bertambah sampai 3 cm pada ibu atau pasien.

6. Prolonged Aktif

Prolonged Aktif adalah Fase aktif yang memanjang (prolonged active phase) yakni masa melahirkan bayi melebihi 8 jam.

7. APB

Ante Partum Bleeding (APB) adalah perdarahan yang terjadi setelah kehamilan 28 minggu. Biasanya lebih banyak dan lebih berbahaya dari pada perdarahan kehamilan sebelum 28 minggu.

8. Oligohidramnion

Oligohidramnion adalah suatu keadaan dimana air ketuban kurang dari normal, yaitu kurang dari 500 cc.

9. Polihidromnion

Polihidramnion adalah kondisi di mana ibu hamil memiliki cairan ketuban terlalu banyak. Dokter dapat mengukur jumlah cairan melalui beberapa metode yang berbeda, paling sering melalui pengukuran indeks cairan amnion (AFI) atau pengukuran kantung. Jika AFI menunjukkan kadar cairan lebih dari 25 cm (atau di atas persentil ke-95), pengukuran kantung mendalam.

10. Mioma

Penyakit Mioma adalah sejenis penyakit tumor jinak terletak pada dinding rahim. Mioma uteri juga disebut Mioma yang berasal dari otot polos rahim, Myom, Tumor Otot Rahim atau Tumor fibroid. Beberapa teori menyebutkan pertumbuhan tumor ini disebabkan rangsangan hormon estrogen.

11. IUGR

IUGR adalah istilah yang digunakan untuk mendeskripsikan fetus atau neonatal yang mempunyai ukuran lebih kecil daripada normal, seringkali juga digunakan untuk mendeskripsikan intrauteri yang rusak atau terbatas.

12. Letli

Letak lintang (letli) adalah suatu keadaan dimana janin melintang (sumbu panjang janin kira-kira tegak lurus dengan sumbu panjang tubuh ibu) di dalam uterus dengan kepala pada sisi yang satu sedangkan bokong berada pada sisi yang lain. Bila sumbu panjang tersebut membentuk sudut lancip, hasilnya adalah letak lintang oblik. Letak lintang oblik biasanya hanya terjadi sementara karena kemudian akan berubah menjadi posisi longitudinal atau letak lintang saat persalinan.

13. Fetal Distress

Fetal distress pada ibu, memiliki ciri-ciri, yaitu : penurunan kemampuan membawa oksigen ibu, anemia yang signifikan, penurunan aliran darah uterin, posisi supine atau hipotensi lain dan preeklampsia

14. IUD

IUD diciptakan oleh Richter dari Polandia pada tahun 1909 dan kemudian oleh Grafenberg dari Jerman pada tahun 1929. Pada awalnya bentuk IUD seperti cincin dari logam dan dikelilingi benang sutera, kemudian sesuai dengan perkembangan zaman metode IUD dikembangkan dan disempurnakan kembali, baik dari bentuk maupun bahannya. IUD dipasang dua atau tiga hari sesudah haid atau tiga bulan sesudah melahirkan dan pemasangannya harus dilakukan oleh tenaga terlatih (MASjfulk Zuhdi, Masail Fiqiyah., hlm 72).

15. Hbs Ag

Keadaan ibu dikatakan HbsAg berarti ibu mengalami darah tinggi saat melahirkan bayi.

16. AFI

AFI singkatan dari *amniotic fluid index* adalah metode semikuantitatif untuk memperkirakan volume cairan amnion. AFI adalah jumlah dari kantung amnion vertikal maksimum dalam cm pada masing-masing empat kuadran uterus. AFI normal pada usia kehamilan lebih dari 20 minggu, yakni 5 sampai 20 cm. AFI tidak normal pada usia kehamilan lebih dari 20 minggu yakni kurang dari 5 cm.

17. Pro SC

Pro SC merupakan keadaan dimana ibu merasakan nyeri pada daerah operasi dengan riwayat ibu telah melakukan *caesar* pada kelahiran sebelumnya.

18. Kala 2 Lama

Kala 2 merupakan kala pengeluaran janin, his terkoordinir, kuat, cepat dan lebih lama. Kira-kira 2-5 menit sekali. Kepala janin telah turun masuk ke ruang panggul, sehingga terjadilah tekanan pada otot-otot dasar panggul yang secara reflektoris yang menimbulkan mase mengedan karena tekanan pada rectum, ibu seperti merasa mau buang air besar, dengan terasa tanda anus membuka pada waktu his, kepala janin mulai kelihatan. Vulva membuka dan perineum meregang. Dengan his mengedan yang terpimpin, akan lahirlah kepala dengan diikuti badan rahim. (Sinopsis Obstertri 1998:95).

19. PLR

Plasenta letak rendah (Low Lying Placenta) adalah tepi plasenta berada 3 sampai 4 cm diatas pinggir pembukaan. Pada pemeriksaan dalam tidak teraba. Dan plasenta yang implantasinya rendah tapi tidak sampai keostium uteri internum.(Sastrowinata, Sulaiman., 1983).

20. Potua

Potua merupakan keadaan dimana terjadi pendarahan saat proses persalinan yang menyebabkan gangguan plasenta previa yang letaknya abnormal yaitu

pada segmen bawah rahim sehingga menutupi sebagian atau seluruh pembukaan jalan lahir (*ostium uteri internum*).

21. Proterminasi

Proterminasi merupakan keadaan takut dan panik yang sering dialami dalam suatu kehamilan, yang disebabkan oleh janin yang terbukti memiliki defek yang berat.

22. Fase Aktif

Fase aktif merupakan frekuensi dan lama kontraksi uterus akan meningkat secara bertahap (kontraksi di anggap adekuat / memadai jika terjadi tiga kali atau lebih dalam waktu 10 menit dan berlangsung selama 40 detik atau lebih). Pembukaan yang terjadi pada ibu mencapai 4 cm hingga mencapai pembukaan lengkap atau 10 cm, akan terjadi dengan kecepatan rata-rata 1 cm per jam (nuli para atau primigravida) atau lebih dari 1 cm hingga 2 cm (multipara) dan terjadi penurunan bagian terbawah janin.

23. PEB

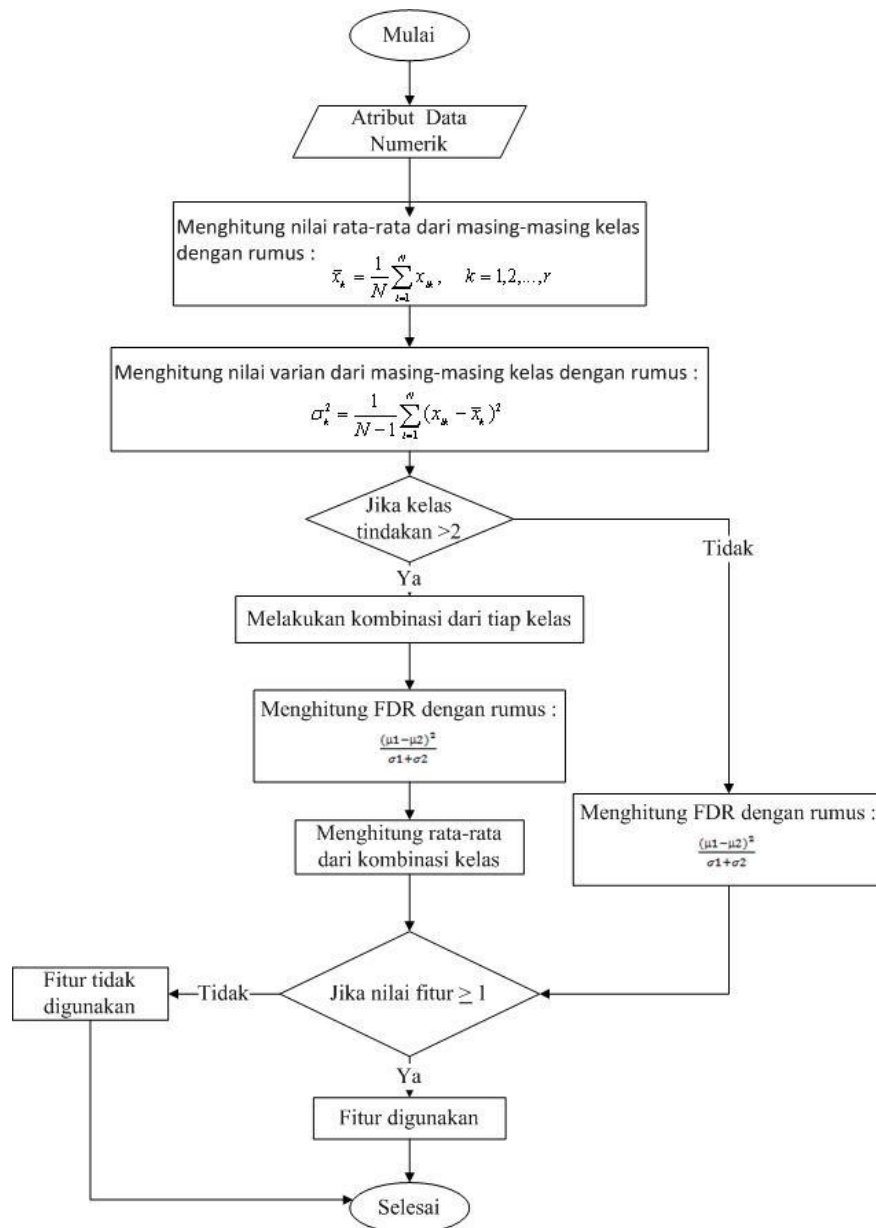
PEB atau Pre eklampsia berat merupakan suatu sindrom hipertensi yang terjadi karena kehamilan disertai protein urine, oedema dan sering kali terdapat gangguan pada sistem organ lain. (Angraeni, Dwi. 2009).

2.7. Macam Obat

Obat yang diberikan pada pasien yang melakukan persalinan meliputi pemberian multivitamin, vitamin C, maag dan antibiotik. Pemberian multivitamin dan vitamin C, contoh merk obatnya adalah vice. Apabila mengalami maag contoh merk obatnya adalah ranitidin, sedangkan pasien yang mengalami infeksi persalinan diberikan antibiotik yang macam merk obatnya adalah procefa, taxegram, mikiasin, lapixim, taxfor, cefotaxim dan vicellin. (C.H, Ennie, 2013)

2.8. Algoritma FDR

Berikut gambar mengenai algoritma preprocessing dengan metode FDR :



Gambar 2.3 Preprocessing dengan metode FDR

Penjelasan dari gambar 2.3, sebagai berikut ini :

1. Memilih atribut yang bernilai numerikal
2. Menghitung nilai rata-rata dari tiap kelas yang dimiliki atribut tersebut, dengan rumus : $\mu = \frac{\sum_{i=1}^n X_i}{n}$, dimana μ adalah rata-rata hitung sampel, n adalah ukuran sampel dan x_i adalah data ke- i .

3. Menghitung nilai varian dari tiap kelas yang dimiliki atribut tersebut, dengan rumus :

$$\sigma = \frac{1}{(n-1)} \times (\sum_{i=1}^n X_i - \mu)^2$$
, dimana σ adalah varian hitung sampel, n adalah ukuran sampel, x_i adalah data ke- i dan μ adalah rata-rata hitung sampel.

4. Jika kelas data dari atribut tersebut lebih dari 2, dilakukan kombinasi dengan membandingkan kelas dalam menghitung FDR yang selanjutnya dihitung rata-rata dari kombinasi kelas. Contohnya diketahui terdapat 5 kelas, sehingga dilakukan kombinasi sebanyak 10. Seperti contoh berikut ini:

a. $\frac{(\mu_1 - \mu_2)^2}{\sigma_1 + \sigma_2}$	f. $\frac{(\mu_2 - \mu_4)^2}{\sigma_4 + \sigma_4}$
b. $\frac{(\mu_1 - \mu_3)^2}{\sigma_1 + \sigma_3}$	g. $\frac{(\mu_2 - \mu_5)^2}{\sigma_2 + \sigma_5}$
c. $\frac{(\mu_1 - \mu_4)^2}{\sigma_1 + \sigma_4}$	h. $\frac{(\mu_3 - \mu_4)^2}{\sigma_3 + \sigma_4}$
d. $\frac{(\mu_1 - \mu_5)^2}{\sigma_1 + \sigma_5}$	i. $\frac{(\mu_3 - \mu_4)^2}{\sigma_3 + \sigma_4}$
e. $\frac{(\mu_2 - \mu_3)^2}{\sigma_2 + \sigma_3}$	j. $\frac{(\mu_4 - \mu_5)^2}{\sigma_4 + \sigma_5}$

Dari kombinasi tersebut, dihitung nilai rata-rata dari kombinasi tersebut dengan rumus : $\mu = \frac{\sum_{i=1}^n X_i}{n}$, dimana μ adalah rata-rata hitung kombinasi, n adalah ukuran kombinasi dan x_i adalah data kombinasi ke- i .

5. Jika kelas data dari atribut tersebut tidak lebih dari 2, maka melakukan perhitungan FDR.
6. Setelah dilakukan perhitungan FDR, diketahui nilai FDR dari atribut tersebut. Jika nilai FDR tidak lebih dari atau sama dengan 1, maka atribut atau fitur tersebut tidak berpengaruh. Jika nilai FDR lebih dari atau sama dengan 1, maka atribut atau fitur tersebut berpengaruh. (Prasetyo, Eko. 2013).

2.9. Algoritma Decision Tree ID3

Dalam mengetahui algoritma decision tree, diberikan permasalahan untuk memberikan penjelasan dari perhitungan tersebut. Misalkan diketahui sebuah data berjumlah 80, atribut berjumlah 4 dan output sebagai entropy global sejumlah 2

kategori, yakni ya dan tidak. Berikut ini cara penyelesaian dengan metode decision tree :

1. Menghitung entropy global

Menghitung entropy global, digunakan rumus berikut : $(y) = -\frac{y_c}{y} \log_2 \frac{y_c}{y}$. Diketahui kategori berjumlah 2 dan data berjumlah 80, maka perhitungannya : $-\frac{38}{80} \log_2 \frac{38}{80} - \frac{42}{80} \log_2 \frac{42}{80}$

2. Menghitung entropy dari setiap atribut

Menghitung entropy dari setiap atribut dengan memasangkan nilai atribut dengan entropy global, rumusnya : $-p_1 \log_2 p_1 - p_2 \log_2 p_2 \dots \dots - p_n \log_2 p_n$. Misalnya untuk menghitung sebuah atribut, dimisalkan atribut tersebut adalah kondisi yang nilainya baik, sedang dan buruk. Selanjutnya menghitung jumlah atribut kondisi dengan output.

Contohnya nilai baik dan nilai output ya berjumlah 20, sedangkan nilai baik dan nilai output tidak berjumlah 5. Nilai sedang dan nilai output ya berjumlah 20, sedangkan nilai sedang dan nilai output tidak berjumlah 15. Nilai buruk dan nilai output ya berjumlah 20, sedangkan nilai buruk dan nilai output tidak berjumlah 15.

Berikut perhitungan dari entropy dengan atribut kondisi :

- Kondisi dengan nilai baik, perhitungannya : $-\frac{20}{25} \log_2 \frac{20}{25} - \frac{5}{25} \log_2 \frac{5}{25}$
- Kondisi dengan nilai sedang, perhitungannya : $-\frac{20}{35} \log_2 \frac{20}{35} - \frac{15}{35} \log_2 \frac{15}{35}$
- Kondisi dengan nilai buruk, perhitungannya : $-\frac{20}{35} \log_2 \frac{20}{35} - \frac{15}{35} \log_2 \frac{15}{35}$

Selanjutnya menghitung entropy total atribut kondisi yakni dengan cara menghitung jumlah nilai tiap atribut dibagi dengan jumlah data, berikut ini contoh kondisi dengan nilai baik berjumlah 25 dengan penjelasan 20 dengan output ya dan 5 dengan output tidak, perhitungannya : $\frac{25}{80} \times$ nilai entropy kondisi dengan nilai baik

3. Menghitung information gain

Menghitung information gain dari hasil pengurangan entropy global dan entropy tiap atribut. Rumusnya : $(y, A) = \text{entropy global } (y) - \sum_{c \in \text{nilai } (A)} \frac{y_c}{y} \text{entropy } (y_c)$

4. Memilih information gain tertinggi untuk dijadikan root (akar).

5. Atribut yang telah terpilih menjadi akar akan diketahui kelas tindakannya apabila nilai dari atribut jumlah datanya kurang dari atau sama dengan syarat pemecahan. Misalnya syarat pemecahan data adalah data yang berjumlah 8. Apabila kelas tindakan hanya memiliki satu nilai kelas tindakan, maka dipilih kelas tindakan tersebut. Apabila kelas tindakan memiliki nilai lebih dari satu, maka dipilih kelas yang terbanyak. Apabila kelas tindakan yang nilainya lebih dari satu mempunyai jumlah kelas sama, maka dipilih salah satu. Berikut ini diberikan contoh penentuan kelas tindakan apabila terpilih atribut kondisi ibu :

Tabel 2.1 Penentuan Kelas Tindakan Dengan Satu Kelas

No	Cara Lahir	Warna Ketuban	Kondisi Ibu	Kondisi Janin	Kelas
1	SC	Keruh	AFI	Normal	Antibiotik

Dari tabel 2.1 diatas, apabila terpilih atribut kondisi ibu, maka dapat ditentukan tindakan dari atribut kondisi ibu yang bernilai AFI, tindakan medisnya adalah antibiotik.

Tabel 2.2 Penentuan Kelas Tindakan Dengan Dua Kelas Memilih Kelas Terbanyak

No	Cara Lahir	Warna Ketuban	Kondisi Ibu	Kondisi Janin	Kelas
1	Spt B	Hijau keruh	Fase laten	Normal	Antibiotik
2	Spt B	Hijau keruh	Fase laten	Normal	Antibiotik
3	Spt B	Meconial	Fase laten	Normal	Antibiotik
4	SC	Keruh	Fase laten	Normal	Antibiotik
5	SC	Jernih	Fase laten	Normal	Kosong
6	SC	Jernih	Fase laten	Normal	Kosong
7	Spt B	Meconial	Fase laten	Normal	Antibiotik
8	Spt B	Hijau keruh	Fase laten	Normal	Kosong

Dari tabel 2.2, diketahui kondisi ibu yang bernilai fase laten, jumlah datanya adalah 8 dan memiliki dua kelas tindakan, yakni antibiotik dan kosong. Karena terdapat lebih dari satu kelas, dipilih kelas dengan jumlah terbanyak. Kelas antibiotik berjumlah 5 dan kelas kosong berjumlah 3, maka kelas yang terpilih adalah antibiotik.

Tabel 2.3 Penentuan Kelas Tindakan Dengan Dua Kelas Berbeda Dengan Jumlah Kelas Sama

No	Cara Lahir	Warna Ketuban	Kondisi Ibu	Kondisi Janin	Kelas
1	SC	Keruh	Oligohidromnion	Normal	Multivitamin
2	SC	Keruh	Oligohidromnion	Normal	Antibiotik

Dari tabel 2.3, diketahui nilai kondisi ibu oligohidromnion memiliki 2 kelas yakni multivitamin dan antibiotik. Jumlah dari kelas multivitamin dan antibiotik adalah 1 data, sehingga dipilih salah satunya, misalnya dipilih multivitamin.

6. Algoritma ini akan terus menggunakan proses yang sama atau bersifat rekursif untuk dapat membentuk sebuah Decision Tree seperti langkah 1-5. Ketika sebuah atribut telah dipilih menjadi node pembagi atau cabang, maka atribut tersebut tidak diikuti lagi dalam penghitungan nilai information gain.
7. Setelah diketahui pohon keputusannya, akan terbentuk rule.
8. Rule yang terbentuk dijadikan sebagai tolak ukur untuk memberikan tindakan atau prediksi. (Santosa, Budi. 2007).