

BAB II

LANDASAN TEORI

2.1 Kehamilan Resiko Tinggi

Kehamilan resiko tinggi adalah kehamilan yang akan menyebabkan terjadinya bahaya dan komplikasi yang lebih besar baik terhadap ibu maupun terhadap janin yang dikandungnya selama masa kehamilan, persalinan dan nifas bila dibandingkan dengan kehamilan, persalinan dan nifas normal. Gangguan dan penyulit pada kehamilan umumnya ditemukan pada kehamilan resiko tinggi. Secara garis besar, kelangsungan suatu kehamilan sangat bergantung pada keadaan dan kesehatan ibu, keadaan plasenta dan keadaan janin.

Pemeriksaan untuk mendiagnosa jenis kelainan di masa kehamilan dibedakan menjadi dua yaitu *obstetri* dan *ginekologi*. *Obstetri* merupakan bagian Ilmu Kedokteran yang khusus mempelajari segala soal yang bersangkutan dengan lahirnya bayi. Dengan demikian, yang menjadi obyek adalah kehamilan, persalinan, nifas dan bayi yang baru dilahirkan. Berdasarkan pemeriksaan, kelainan di masa kehamilan dibagi menjadi dua yaitu kelainan yang disertai perdarahan dan kelainan yang tidak disertai perdarahan (Sastrawinata, 1984).

2.1.1 Jenis Kelainan Kehamilan

Berikut ini adalah beberapa jenis kelainan yang disertai perdarahan yang sering muncul di masa kehamilan :

1. Abortus

Berakhirnya kehamilan sebelum anak dapat hidup didunia luar disebut *Abortus*. Menurut Eastman *abortus* adalah keadaan terputusnya suatu kehamilan dimana janin atau *fetus* belum sanggup hidup sendiri diluar *uterus*. Belum sanggup diartikan apabila *fetus* itu beratnya terletak antara 400-100 gram, atau usia kehamilan kurang dari 28 minggu . Belum diketahui dengan pasti, penyebab seorang ibu hamil mengalami *abortus*. Akan tetapi ada beberapa faktor pemicu terjadinya *abortus* yaitu adanya kelainan pada janin atau rahim ibu dan daya tahan ibu

yang lemah. Abortus sendiri digolongkan menjadi tiga jenis yaitu (Mochtar, 1988) :

- *Abortus* Mengancam

Abortus mengancam biasanya terjadi di antara kehamilan trisemester pertama dan kedua awal. Tanda-tandanya; ibu mengalami kontraksi, perdarahan, dan bisa disertai keluarnya cairan. Janin bisa diselamatkan jika masih dalam kondisi baik dan *ostium* (lubang rahim) belum terbuka. Sebaliknya jika lubang rahim sudah terbuka, dokter tidak bisa berbuat banyak. Kemungkinan yang terjadi adalah *abortus* spontan.

- *Abortus* Spontan

Abortus spontan biasanya terjadi di kehamilan trimester pertama atau bahkan sebelum seorang wanita menyadari bahwa dirinya hamil. Dengan kata lain, *abortus* spontan bisa terjadi tanpa diketahui karena gejalanya mirip haid hanya lebih berat dan lebih terasa tegang. Gejala yang dialami ibu yaitu keluarnya bercak-bercak darah yang terus menerus, perdarahan disertai nyeri di bagian tengah perut dan kadang-kadang disertai sakit pinggang serta terdapat bekuan darah. Penyebab *abortus* jenis ini adalah kelainan *embrio*, janin atau plasenta kekurangan hormon, penyakit infeksi yang diderita ibu seperti gondong, cacar air dan campak atau juga reaksi *auto-immune* dimana sel-sel kekebalan ibu menyerang janin.

- *Abortus* Lanjut

Keluarnya hasil konsepsi yang disebabkan kelainan *plasenta* dan *serviks* atau ibu hamil terpapar bahan beracun seperti asap rokok, alkohol dan bahan kimia. Tanda-tandanya sama dengan abortus spontan.

2. *Plasenta Previa*

Plasenta Previa adalah plasenta yang ada di depan jalan lahir, yaitu pada segmen-bawah *uterus* sehingga dapat menutupi sebagian atau seluruh pembukaan jalan-lahir. Pada keadaan normal *plasenta* terletak dibagian

dinding depan atau dinding belakang rahim di daerah *fundus uteri*. *Plasenta* yaitu sebuah jaringan yang terbentuk di rahim saat kehamilan dan fungsinya untuk membawa makanan dan oksigen dari ibu ke janin yang sangat dibutuhkan si janin melalui tali pusar. Semua yang tidak baik akan dibuang melalui *plasenta* seperti limbah dan *CO2* namun oksigen dan sumber makanan akan terus diasup *plasenta* ke janin melalui tali pusar. *Plasenta previa* baru bisa dipastikan jika sudah memasuki usia kandungan 20 minggu. Penyebab timbulnya kasus *plasenta previa* ini sebenarnya tidak banyak diketahui namun biasanya terjadi pada para wanita yang bermasalah dengan sel telur atau lapisan rahim, memiliki riwayat *plasenta previa* atau bedah *cesar* pada kehamilan sebelumnya, gaya hidup tidak sehat dan usia ibu diatas 30 tahun (Mochtar, 1988).

Kasus *plasenta previa* terbagi menjadi 3 jenis, tergantung perletakannya :

- *Plasenta Previa Total* atau lengkap, yaitu pembukaan mulut rahim untuk jalan lahir benar-benar tertutup oleh plasenta
- *Plasenta Previa Parsial* atau sebagian, dimana hanya bagian *serviks* dalam saja yang tertutup plasenta (*internal serviks* saja)
- *Plasenta Previa Marginal*, berada pada tepi bukaan serviks, atau biasa disebut letak rendah, namun tidak menutupi *serviks*.

Gejala yang paling umum timbul adalah pendarahan pada vagina di trimester akhir kehamilan. Walaupun sedikit tetapi biasanya disertai rasa sakit dan kram. Hal ini diakibatkan *plasenta* lepas dari dinding rahim kemudian meregang. Penanganan pada kehamilan dengan *plasenta previa* adalah menghindari terjadinya resiko perdarahan. Selain itu perlu dipertimbangkan kelahiran prematur, kelahiran cacat, anemia pada janin, sehingga jalan satu-satunya bagi plasenta previa total adalah menjalani *operasi cesar* (Sastrawinata, 1984).

3. Solusio Plasenta

Solusio plasenta adalah terlepasnya *plasenta* dari tempat implantasi normalnya sebelum janin lahir, dan definisi ini hanya berlaku apabila terjadi pada kehamilan di atas 22 minggu atau berat janin di atas 500 gram. Apabila

pelepasan plasenta terjadi sebelum minggu ke-22 disebut abortus dan apabila terjadi pelepasan plasenta pada plasenta yang rendah implantasinya maka bukan disebut *solusio plasenta* tapi *plasenta previa*.

Penyebab pasti *solusio plasenta* hingga kini belum diketahui jelas, walaupun beberapa keadaan tertentu dapat menyertainya, seperti (Prawiroharjo, 2002):

- Riwayat kesehatan ibu
 - Ibu pernah mengalami *Hipertensi, Pre-Eklamsia, Eklamsia dan solusio plasenta* pada kehamilan sebelumnya
- Faktor trauma
 - Tarikan pada tali pusat yang pendek akibat pergerakan janin yang banyak atau bebas, versi luar atau tindakan pertolongan persalinan
 - Trauma langsung, seperti jatuh, kena tendang, dan lain-lain.
- Faktor usia ibu
 - Makin tua umur ibu, makin tinggi frekuensi hipertensi menahun sehingga kemungkinan ibu mengalami *solusio plasenta* semakin tinggi
- Pengaruh lain, seperti anemia, malnutrisi/defisiensi gizi, tekanan *uterus* pada *vena cava inferior* dikarenakan pembesaran ukuran *uterus* oleh adanya kehamilan

Solusio plasenta terbagi dalam 3 macam (Mochtar, 1998) :

- Solusio plasenta parsial
 - Bila hanya sebagian saja plasenta terlepas dari tempat perletakannya
- Solusio plasenta totalis (komplis)
 - Bila seluruh plasenta sudah terlepas dari seluruh tempat perletakannya
- Prolapsus Plasenta
 - Keadaan plasenta kadang-kadang turun kebawah dan dapat teraba pada pemeriksaan dalam

Cunningham dan Gasong masing-masing dalam bukunya mengklasifikasikan solusio plasenta menurut tingkat gejala klinisnya, yaitu:

- Ringan : pendarahan <100-200 cc, *uterus* tidak tegang, belum ada tanda renjatan, janin hidup, pelepasan plasenta <1/6 bagian permukaan, kadar *fibrinogen plasma* >150 mg%

- Sedang : pendarahan lebih dari 200 cc, *uterus* tegang, terdapat tanda pre rejatan, gawat janin atau janin telah mati, pelepasan plasenta 1/4 – 2/3 bagian permukaan, kadar *fibrinogen plasma* 120-150 mg%
- Berat : *uterus* tegang dan berkontraksi tetanik, terdapat tanda rejatan, janin mati, pelepasan plasenta dapat terjadi lebih 2/3 bagian atau keseluruhan

2.2 Data Mining

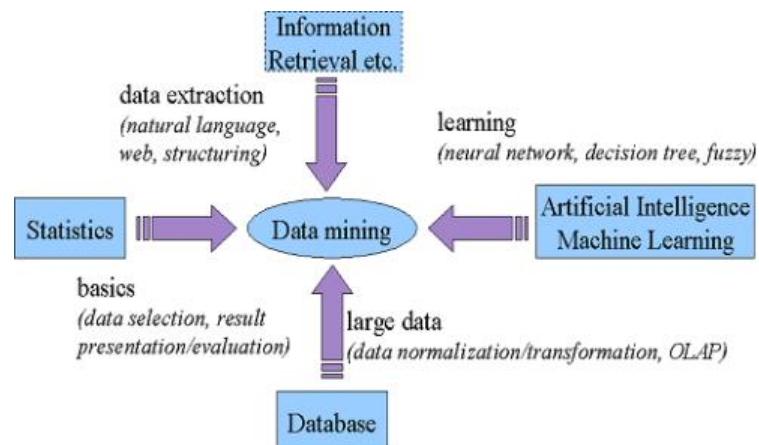
Data mining menurut Turban,dkk.(2005) adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan mesin learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait berbagai database besar. Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual.

Menurut Gartner Group data mining adalah suatu proses menemukan hubungan yang berarti, pola dan kecenderungan dengan memeriksa ke dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (larose, 2005)

Dari definisi-definisi yang telah disampaikan, hal penting yang terkait dengan data mining adalah (Kusrini dan Luthfi, 2009):

1. Data mining merupakan suatu proses otomatis terhadap data yang sudah ada
2. Data yang diproses merupakan data yang sangat besar
3. Tujuan data mining adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat

Sejarah Data mining bukanlah suatu bidang yang sama sekali baru. Salah satu kesulitan untuk mendefinisikan data mining adalah kenyataan bahwa data mining mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan terlebih dulu. Gambar 2.1 menunjukkan bahwa data mining memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, *statistic*, *database* dan juga *information retrieval* (Pramudiono, 2006).



Gambar 2.1 Bidang Ilmu Data Mining

Beberapa teknik yang sering disebut-sebut dalam literatur data mining seperti *classification*, *neural network*, *genetic algorithm* dan lain sebagainya sudah lama dikenal di dunia kecerdasan buatan. Statistik memberikan kontribusi pada data mining dengan teknik-teknik untuk menyeleksi data dan evaluasi hasil data mining selain teknik-teknik data mining seperti *clustering*. Yang membedakan persepsi terhadap data mining adalah perkembangan teknik-teknik data mining untuk aplikasi pada *database* skala besar. Sebelum populernya data mining, teknik-teknik tersebut pada umumnya diterapkan untuk data skala kecil saja. Selain itu beberapa teknik dari bidang *database* untuk transformasi data juga merupakan bagian integral dari proses data mining.

Akhir-akhir ini ada beberapa bidang ilmu seperti *information retrieval* yang juga terlibat dalam proses data mining untuk mengekstrak sumber data bagi data mining dari sumber-sumber seperti teks dan *website*. Walaupun data mining memiliki sumber dari beberapa bidang ilmu, data mining berbeda dalam beberapa aspek dibandingkan dengan bidang ilmu seperti berikut :

- Statistik : model statistik dipersiapkan oleh para ahli statistik, sedangkan data mining mengembangkan statistik untuk menangani data berjumlah besar secara otomatis
- *Expert system* (sistem cerdas) : model pada expert system dibuat berupa aturan-aturan berdasar pada pengalaman-pengalaman para ahli

- Data Warehouse (DWH) : sering terjadi kerancuan antara data mining dan data warehouse karena keduanya sering dipakai bersamaan. Pada umumnya data warehouse lebih merujuk pada tempat untuk menyimpan data yang terkonsolidasi sedangkan data mining bisa dianggap sebagai perkakas untuk menganalisa otomatis nilai dari data itu
- OLAP : seperti data warehouse, OLAP juga sering dibahas bersama data mining. Tetapi OLAP memiliki tujuan untuk memastikan hipotesa yang sudah diformulasikan terlebih dulu oleh penggunanya.

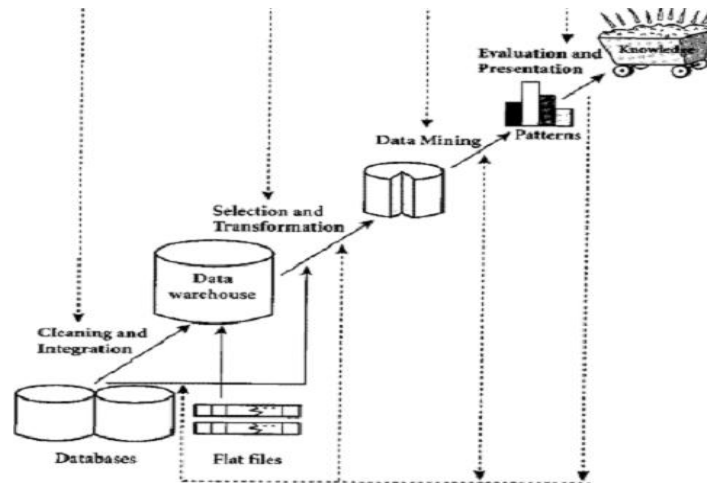
Salah satu tuntutan dari data mining ketika diterapkan pada data berskala besar adalah diperlukan metodologi sistematis tidak hanya ketika melakukan analisa saja tetapi juga ketika mempersiapkan data dan juga melakukan interpretasi dari hasilnya sehingga dapat menjadi aksi ataupun keputusan yang bermanfaat. Karenanya 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.

2.2.1 Tahapan Data Mining

Data mining merupakan proses pencarian pola-pola yang menarik dan tersembunyi (*hidden pattern*) dari suatu kumpulan data yang berukuran besar yang tersimpan dalam suatu basis data, *data warehouse*, atau tempat penyimpanan data lainnya (Tan dkk, 2006). Menurut Sumanthi dan Sivandham (2009), *data mining* juga didefinisikan sebagai bagian dari proses penggalian pengetahuan dalam *database* yang sering disebut dengan istilah *Knowledge Discovery in Database* (KDD).

KDD merupakan suatu area yang mengintegrasikan berbagai metode, yang meliputi statistik, basis data, kecerdasan buatan (*Artificial Intelligence*), *machine learning*, pengenalan pola (*Pattern Recognition*), pemodelan yang menangani ketidakpastian, visualisasi data, optimasi, Sistem Informasi Manajemen (SIM), dan sistem berbasis pengetahuan (*knowledge based-system*). Sebagai bagian dari proses yang ada di dalam KDD, maka *data mining* didahului dengan proses

pemilihan data, pembersihan data, *pre-processing*, dan transformasi data (Sumanthi dan Sivandham, 2009).



Gambar 2.2 Proses Dalam Data Mining

Sebagai suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap yang diilustrasikan pada Gambar 2.2 di atas. Tahap-tahap tersebut bersifat interaktif di mana pemakai terlibat langsung atau dengan perantaraan knowledge base. Tahapan-tahapan tersebut, diantaranya :

1. Pembersihan data (untuk membuang data yang tidak konsisten dan noise)

Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang kita miliki. Data-data yang tidak relevan itu juga lebih baik dibuang karena keberadaannya bisa mengurangi mutu atau akurasi dari hasil data mining nantinya. *Garbage in garbage out* (hanya sampah yang akan dihasilkan bila yang dimasukkan juga sampah) merupakan istilah yang sering dipakai untuk menggambarkan tahap ini. Pembersihan data juga akan mempengaruhi performansi dari sistem data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

2. Integrasi data (penggabungan data dari beberapa sumber)

Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dsb. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada. Dalam integrasi data ini juga perlu dilakukan transformasi dan pembersihan data karena seringkali data dari dua database berbeda tidak sama cara penulisannya atau bahkan data yang ada di satu database ternyata tidak ada di database lainnya.

3. Transformasi data (data diubah menjadi bentuk yang sesuai untuk *data mining*)

Beberapa teknik data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa teknik standar seperti analisis asosiasi dan klustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut *binning*. Disini juga dilakukan pemilihan data yang diperlukan oleh teknik data mining yang dipakai. Transformasi dan pemilihan data ini juga menentukan kualitas dari hasil data mining nantinya karena ada beberapa karakteristik dari teknik-teknik data mining tertentu yang tergantung pada tahapan ini.

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. Ada kalanya teknik-teknik data mining umum yang tersedia di pasar tidak mencukupi untuk melaksanakan data mining di

bidang tertentu atau untuk data tertentu. Sebagai contoh akhir-akhir ini dikembangkan berbagai teknik data mining baru untuk penerapan di bidang bioinformatika seperti analisa hasil *microarray* untuk mengidentifikasi DNA dan fungsi-fungsinya.

5. Evaluasi pola yang ditemukan (untuk menemukan yang menarik/bernilai)
 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. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti : menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba teknik data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat. Ada beberapa teknik data mining yang menghasilkan hasil analisa berjumlah besar seperti analisis asosiasi. Visualisasi hasil analisa akan sangat membantu untuk memudahkan pemahaman dari hasil data mining.
6. Presentasi pola yang ditemukan untuk menghasilkan aksi
 Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisa yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami data mining. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data mining. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil data mining.

2.2.2 Klasifikasi Dalam Data Mining

Klasifikasi pertama kali diterapkan pada bidang tanaman yang mengklasifikasikan suatu spesies tertentu, seperti yang dilakukan oleh Carolus Von Linne (atau dikenal dengan nama Carolus Linnaeus) yang pertama kali mengklasifikasikan spesies berdasarkan karakter fisik. Selanjutnya dia dikenal sebagai bapak klasifikasi. Komponen-komponen utama dari proses klasifikasi antara lain (Widodo dkk, 2013):

- Kelas, merupakan variabel tidak bebas yang merupakan label dari hasil klasifikasi. Sebagai contoh adalah kelas loyalitas pelanggan, kelas badai atau gempa bumi dan lain-lain.
- Prediktor, merupakan variabel bebas suatu model berdasarkan dari karakteristik atribut data yang diklasifikasi, misalnya merokok, minum minuman beralkohol, tekanan darah status perkawinan, dan sebagainya.
- Set data pelatihan, merupakan sekumpulan data lengkap yang berisi kelas dan prediktor untuk dilatih agar model dapat mengelompokkan kedalam kelas yang tepat. Contohnya adalah grup pasien yang telah di-*test* terhadap serangan jantung, grup pelanggan disuatu supermarket, dan sebagainya.
- Set data uji, berisi data-data baru yang akan dikelompokkan oleh model guna mengetahui akurasi dari model yang telah dibuat.

Sebagian besar istilah-istilah yang ada dalam aktifitas klasifikasi sama dengan yang digunakan dalam aplikasi data base. Namun beberapa mungkin tidak begitu dikenal, istilah-istilah tersebut antara lain (Widodo dkk, 2013) :

- Set data yang digunakan untuk proses pelatihan dikenal dengan nama-nama yang berbeda antara lain: records, tuples, vektor, instan, objek dan sampel
- Tiap set data tersebut memiliki suatu atribut.
- Pengklasifikasi (*classifier*), merupakan model matematis yang akan menentukan suatu objek masuk dalam kelas tertentu.
- Set data testing, merupakan data-data dengan sifat seperti data pelatihan untuk menguji akurasi dari model yang telah dibuat.

Dalam mesin pembelajaran (*machine learning*) dikenal istilah pembelajaran terpadu (*supervised learning*) dan pembelajaran tak terpadu (*unsupervised learning*). Pembelajaran terpadu memiliki kesamaan dengan metode prediksi yang memprediksi keluaran dari masukan tertentu. Sedangkan pembelajaran tak terpadu identik dengan metode deskriptif yang mengelompokkan dalam pola-pola tertentu

Untuk menentukan suatu model baik atau buruk, diperlukan elemen-elemen kunci antara lain (Widodo dkk, 2013) :

- Akurasi prediksi, yang menentukan seberapa akurat suatu model dalam memprediksi keluaran.
- Kecepatan, yang menunjukkan seberapa cepat suatu model dalam memproses data masukan.
- *Rosustness*, menggambarkan kemampuan suatu model melakukan prediksi yang akurat walau dalam kondisi ekstrem dan banyak gangguan yang terjadi.
- Skalabilitas, adalah kemampuan suatu model memproses data baik dalam ukuran yang lebih besar maupun data dari bidang lain yang berbeda.
- *Interpretability*, menggambarkan kemudahan suatu model untuk dipahami dan diinterpretasikan.
- Kesederhanaan, merupakan sifat yang cenderung dipilih untuk menyelesaikan suatu permasalahan.

Metode-metode/model-model telah dikembangkan oleh periset untuk menyelesaikan kasus klasifikasi. Metode-metode tersebut antara lain (Sumathi, 2006) :

- Pohon Keputusan
- Pengklasifikasi Bayes/Naive Bayes
- Jaringan Syaraf Tiruan
- Analisis Statistik
- Algoritma Genetik
- Rough Sets
- Pengklasifikasi k-Nearest Neighbour
- Metode Berbasis Aturan
- Memory Based Reasoning
- Support Vector Machine

2.3 Naive Bayes

Naive Bayes adalah metode probabilitas dan statistic (memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya). sebuah metode *information retrieval* yang menggunakan pendekatan probabilistik dalam

menginferensi, yakni berbasis pada teorema *Bayes* secara umum. Olson dan Delen (2008:102) menjelaskan *Naive Bayes* untuk setiap keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar, mengingat vektor informasi objek. Aplikasinya yang paling banyak digunakan yaitu untuk klasifikasi teks. Arti kata “*naive*” di sini adalah metode *Naive Bayes* mengasumsikan bahwa probabilitas kemunculan sebuah kata independen terhadap posisinya di dalam teks. Tujuan dari asumsi independensi posisi kata adalah untuk mengurangi kompleksitas dalam perhitungan nilai-nilai probabilitas bersyarat pada teorema *Bayes* secara umum.

Dalam sebuah aturan yang mudah, sebuah klasifikasi *Naive Bayes* diasumsikan bahwa ada atau tidaknya ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya. Untuk contohnya, buah akan dianggap sebagai sebuah apel jika berwarna merah, berbentuk bulat dan berdiameter sekitar 6 cm. Walaupun jika ciri-ciri tersebut bergantung satu sama lainnya, dalam *Bayes* hal tersebut tidak dipandang sehingga masing-masing fitur seolah tidak memiliki hubungan apapun. Berdasarkan ciri alami dari sebuah model probabilitas, klasifikasi *Naive Bayes* bisa dibuat lebih efisien dalam bentuk pembelajaran. Dalam beberapa bentuk praktiknya, parameter untuk perhitungan model *Naive Bayes* menggunakan metode *maximum likelihood*, atau kemiripan tertinggi.

Teorema *Bayes* memiliki bentuk umum sebagai berikut (Kusrini dan Luthfi, 2009):

$$P(H|X) = \frac{P(H|X)P(H)}{P(X)} \dots\dots\dots 2.1$$

Keterangan :

X = data dengan *class* yang belum diketahui

H = hipotesis data X merupakan suatu *class* spesifik

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

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

Sedangkan *Naive Bayes* dengan fitur kontinyu memiliki formula :

$$P(X|Y) = \frac{1}{\sqrt{2\pi}\sigma} \exp \frac{-(x-\mu)^2}{2\sigma^2} \dots\dots\dots 2.2$$

Keterangan :

$P(Y|X)$ = probabilitas data dengan vektor X pada kelas Y

$P(Y)$ = probabilitas awal kelas Y

$\pi_i^q = 1 p(Xi|Y)$ = probabilitas independen kelas Y dari semua fitur dalam vektor

X

μ = mean atau nilai rata-rata dari atribut dengan fitur kontinu

σ = deviasi standar

Contoh perhitungan *Naive Bayes* :

Misalnya ingin diketahui apakah suatu objek masuk dalam kategori dipilih untuk perumahan atau tidak dengan algoritma *Naive Bayes*. Untuk menetapkan suatu daerah akan dipilih sebagai lokasi untuk mendirikan perumahan, telah dihimpun 10 aturan. Ada 4 atribut yang digunakan seperti terlihat pada tabel 2.1.

Tabel 2.1. Atribut penetapan lokasi perumahan

Aturan ke-	Harga tanah (C1)	Jarak dari pusat kota (C2)	Ada angkutan umum (C3)	Dipilih untuk perumahan (C4)
1	100	2	Tidak	Ya
2	200	1	Tidak	Ya
3	500	3	Tidak	Ya
4	600	20	Tidak	Tidak
5	550	8	Tidak	Tidak
6	250	25	Ada	Tidak
7	75	15	Ada	Tidak
8	80	10	Tidak	Ya
9	700	18	Ada	Tidak
10	180	8	Ada	Ya

1. Harga tanah per meter persegi (C1),
2. Jarak daerah tersebut dari pusat kota (C2),
3. Ada atau tidaknya angkutan umum di daerah tersebut (C3), dan
4. Keputusan untuk memilih daerah tersebut sebagai lokasi perumahan (C4)

a. Mean dan deviasi standar untuk atribut harga tanah (C1)

$$\mu_{\text{tidak}} = \frac{100+200+500+80+180}{5} = 212$$

$$\mu_{\text{tidak}} = \frac{600+550+250+75+700}{5} = 435$$

$$\sigma^2_{\text{ya}} = \frac{(100-212)^2+(200-212)^2+(500-212)^2+(80-212)^2+(180-212)^2}{5-1}$$

$$= 28520,015$$

$$\sigma_{\text{ya}} = \sqrt{28520,015} = 168,8787$$

$$\sigma^2_{\text{tidak}} = \frac{(600-435)^2+(550-435)^2+(250-435)^2+(75-435)^2+(700-435)^2}{5-1} =$$

$$68624,98$$

$$\sigma_{\text{tidak}} = \sigma_{\text{tidak}} = \sqrt{68624,98} = 261,9637$$

b. Mean dan varian untuk atribut jarak dari pusat kota (C2)

$$\mu_{\text{ya}} = \frac{2+1+3+10+8}{5} = 4,8$$

$$\mu_{\text{tidak}} = \frac{20+8+25+15+18}{5} = 17,2$$

$$\sigma^2_{\text{ya}} = \frac{(2-2,8)^2+(1-2,8)^2+(3-2,8)^2+(10-2,8)^2+(8-2,8)^2}{5-1} = 15,699821$$

$$\sigma_{\text{ya}} = \sqrt{15,699821} = 3,9623$$

$$\sigma^2_{\text{tidak}} = \frac{(20-17,2)^2+(8-17,2)^2+(25-17,2)^2+(15-17,2)^2+(18-17,2)^2}{5-1}$$

$$= 39,700081$$

$$\sigma_{\text{tidak}} = \sqrt{39,700081} = 6,3008$$

Sedangkan untuk probabilitas atribut angkutan umum dan dipilih untuk perumahan terlihat pada tabel 2.2 dan tabel 2.3.

Tabel 2.2. Probabilitas kemunculan setiap nilai untuk atribut angkutan umum (C3)

Angkutan umum	Jumlah kejadian “dipilih”		probabilitas	
	ya	tidak	ya	tidak
Ada	1	3	1/5	3/5
Tidak	4	2	4/5	2/5
Jumlah	5	5	1	1

Tabel 2.3. Probabilitas kemunculan setiap nilai untuk atribut dipilih untuk perumahan (C4)

Dipilih untuk perumahan	Jumlah kejadian “dipilih”		probabilitas	
	ya	tidak	ya	tidak
Jumlah	5	5	1/2	1/2

Dari data di atas apabila diberikan $C1 = 300$, $C2 = 17$, $C3 = \text{Tidak}$, maka :

$$P_{C1=300 \text{ ya}} = \frac{1}{\sqrt{2\pi} 168,8787} \exp \frac{-(300-212)^2}{2 \times 28520,0152} = 0,0021$$

$$P_{C1=300 \text{ tidak}} = \frac{1}{\sqrt{2\pi} 261,9637} \exp \frac{-(300-435)^2}{2 \times 68624,98} = 0,001$$

$$P_{C2=17 \text{ ya}} = \frac{1}{\sqrt{2\pi} 3,9623} \exp \frac{-(17-4,8)^2}{2 \times 15,699821} = 0,0009$$

$$P_{C2=17 \text{ tidak}} = \frac{1}{\sqrt{2\pi} 6,3008} \exp \frac{-(17-17,2)^2}{2 \times 39,700081} = 0,0633$$

Sehingga :

$$\text{Nilai ya} = (0,0021) \times (0,0009) \times 4/5 \times 5/10 = 0,000000756$$

$$\text{Nilai tidak} = (0,0013) \times (0,0633) \times 2/5 \times 5/10 = 0,000016458$$

Nilai probabilitas dapat dihitung dengan melakukan normalisasi terhadap nilai ya dan tidak tersebut sehingga:

$$\text{Probabilitas ya} = \frac{0,000000756}{0,000000756 + 0,000016458} = 0,439$$

$$\text{Probabilitas tidak} = \frac{0,000016458}{0,000000756 + 0,000016458} = 0,9561$$

Dari hasil diatas, terlihat bahwa nilai probabilitas akhir terbesar ada di kelas (P|tidak) dengan hasil akhir 0,9561 sehingga dapat disimpulkan bahwa objek yang dimaksud masuk dalam klasifikasi “TIDAK” masuk dalam kategori dipilih untuk perumahan.

2.4 Penelitian Sebelumnya

Penelitian sebelumnya dilakukan Leni (2012) dengan judul Sistem Pakar Mendiagnosis Kelainan Lama Kehamilan. Sistem pakar ini dirancang untuk mendiagnosa, mengenal kelainan-kelainan yang mungkin terjadi semasa kehamilan pada ibu hamil. Disamping itu kehadiran sistem pakar yang dimaksudkan untuk mengantisipasi ketidakberadaan dokter artinya bukan menggantikan kedudukan bidan maupun dokter hanya jika dokter atau bidan tidak berada ditempat, maka sistem ini diharapkan dapat membantu masyarakat dalam memberikan informasi tentang kelainan lama kehamilan yang mungkin terjadi semasa kehamilan. penelitian ini diharapkan dapat membantu masyarakat untuk mendiagnosis permasalahan kelainan lama kehamilan. Pengembangan Program aplikasi ini menggunakan metode inferensi *forward chaining*.

Penelitian selanjutnya dilakukan Aji Darma Nugroho (2013) dengan judul Sistem Pakar Diagnosis Komplikasi Beresiko Tinggi Selama Kehamilan. Dengan menggunakan sistem pakar diharapkan Sistem pakar yang akan dibangun digunakan untuk membantu pakar dengan mengadopsi pengetahuan pakar. Sistem ini dirancang menggunakan *metode forward chaining*. *Forward Chaining* merupakan suatu penalaran yang dimulai dari fakta untuk mendapatkan kesimpulan dari fakta tersebut agar memudahkan pengguna untuk

menggunakannya. Aplikasi sistem pakar ini dipergunakan untuk mendiagnosis komplikasi beresiko tinggi pada ibu hamil yang diperoleh dari input, berupa gejala awal yang dirasakan ibu hamil. Gejala awal meliputi indikator kehamilan yaitu : perut mual, muntah-muntah dan terlambat menstruasi. 15 penyakit komplikasi beresiko tinggi pada ibu hamil yaitu : *Anemia, Intrauterine Growth Restriction, Preterm Labor, Premature Rupture of Membranes, Gestational Diabetes, Pregnancy Induced Hypertension, Placenta Previa, Hidroamnios, Rhesus, Kehamilan Post Term, Kehamilan Ganda, Kehamilan Ektopik, Keguguran, Kelahiran Mati, Pendarahan Pasca Melahirkan*. Penelitian ini menghasilkan aplikasi sistem pakar untuk diagnosis komplikasi beresiko tinggi selama kehamilan yang dapat digunakan sebagai bahan untuk berkonsultasi dengan dokter kandungan serta tingkat kepuasan konsumen 60% dari 30 sampel.

Penelitian selanjutnya dilakukan Utari Setya Hartini (2013) dengan judul perancangan dan Implementasi Sistem Pakar Deteksi Dini Gangguan Masa Kehamilan Dengan Metode *Forward Chaining*. Sistem pakar dalam penelitian ini adalah sistem yang dikembangkan dengan menerapkan metode *forward chaining*. Metode ini memberikan diagnosis penyakit masa kehamilan yang berdasarkan fakta yang ada dikumpulkan dari wawancara dengan pakar dalam sesi konsultasi. Penelitian ini bertujuan untuk mengembangkan basis pengetahuan sesuai untuk diterapkan pada masalah diagnosis pada sistem pakar. diharapkan penelitian ini dapat membantu ibu hamil dalam mendeteksi dini gangguan pada kehamilan, dan membantu calon bidan dalam studi mereka.