BAB II

LANDASAN TEORI

2.1 Bantuan Langsung Sementara Masyarakat (BLSM)

Bantuan Langsung Sementara Masyarakat (BLSM), merupakan bantuan tunai langsung sementara untuk membantu mempertahankan daya beli Rumah Tangga miskin dan rentan agar terlindungi dari dampak kenaikan harga akibat penyesuaian harga Bahan Bakar Minyak (BBM). BLSM disalurkan untuk membantu Rumah Tangga miskin dan rentan dalam memenuhi kebutuhan hidup Rumah Tangga, pembelian obat-obatan kesehatan, biaya pendidikan dan keperluan-keperluan lainnya [SUS05].

BLSM merupakan program bantuan tunai yang serupa dengan program yang sudah dilaksanakan pada kenaikan harga BBM sebelumnya, yaitu Subsidi Langsung Tunai (SLT) atau Bantuan Langsung Tunai (BLT) 2005/2006 dan BLT 2008/2009. BLSM bertujuan membantu mempertahankan daya beli ruta miskin dan rentan dari dampak kenaikan harga berbagai komoditas akibat penyesuaian harga BBM. Dengan demikian BLSM merupakan bantuan langsung berupa uang tunai yang dilaksanakan membantu mempertahankan daya beli Rumah Tangga miskin dan rentan agar terlindungi dari dampak kenaikan harga akibat penyesuaian harga BBM [1].

Menurut Badan Pusat Statistika terdapat 14 kriteria kemiskinan yang memperoleh bantuan langsung sementara masyarakat, yaitu [1]:

1. Luas lantai bangunan tempat tinggal kurang dari 8 meter persegi untuk masing-masing anggota keluarga
2. Jenis lantai bangunan tempat tinggal terbuat dari tanah, bambu, kayu berkualitas rendah
3. Jenis dinding bangunan tempat tinggal terbuat dari bambu, rumbia, kayu berkualitas rendah
4. Fasilitas jamban tidak ada, atau ada tetapi dimiliki secara bersama-sama dengan keluarga lain

5. Sumber air untuk minum/memasak berasal dari sumur/mata air tak terlindung, air sungai, danau, atau air hujan
6. Sumber penerangan di rumah bukan listrik
7. Bahan bakar yang digunakan memasak berasal dari kayu bakar, arang, atau minyak tanah
8. Makan dalam sehari hanya satu kali atau dua kali
9. Dalam seminggu tidak pernah mengonsumsi daging, susu, atau hanya sekali dalam seminggu
10. Dalam setahun paling tidak hanya mampu membeli pakaian baru satu stel
11. Tidak mampu membayar anggota keluarga berobat ke puskesmas atau poliklinik
12. Pekerjaan utama kepala rumah tangga adalah petani dengan luas lahan setengah hektare, buruh tani, kuli bangunan, tukang batu, tukang becak, pemulung, atau pekerja informal lainnya.
13. Pendidikan tertinggi yang ditamatkan kepala rumah tangga bersangkutan tidak lebih dari SD
14. Tidak memiliki harta dengan nilai Rp500.000,00 seperti misalnya tabungan, perhiasan emas, TV berwarna, ternak, sepeda motor (kredit atau nonkredit), kapal motor, tanah, atau barang modal lainnya.

Berdasarkan pengamatan di lapangan, hanya terdapat data kepala keluarga Kelurahan Pekauman yang intinya sudah mencakup pada kriteria yang sudah ditentukan oleh pemerintah, maka dari itu kriteria yang dijadikan pedoman pada penelitian ini adalah data kepala keluarga di Kelurahan Pekauman Kecamatan Gresik.

2.2 *Data Mining*

2.2.1 Pengertian *Data Mining*

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakit dari berbagai database besar [TUR06].

Menurut [HAN03], *data mining* adalah proses menemukan pola yang menarik dan pengetahuan dari data dalam jumlah besar. Dari beberapa pernyataan tersebut, dapat disimpulkan bahwa *data mining* merupakan proses ekstraksi informasi dari database yang berukuran besar untuk mendapatkan pengetahuan yang tersimpan dari data tersebut. Istilah *data mining* kadang disebut juga *Knowledge Discovery in Database* (KDD).

Istilah *data mining* sering dipakai, mungkin karena istilah ini lebihpendek dari *Knowledge Discovery in Database*. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Data mining dianggap hanya sebagai suatu langkah penting dalam KDD. Proses KDD secara garis besar dapat dijelaskan sebagai berikut [HAN03]:

1. Pembersihan data, untuk menghilangkan *noise* dan data yang tidak konsisten.
2. Integrasi data, di mana beberapa sumber data dapat dikombinasikan. Sebuah tren populer di industri informasi adalah untuk melakukan pembersihan dan integrasi data sebagai langkah preprocessing, dimana data yang dihasilkan akan disimpan dalam *data warehouse*.
3. Seleksi data, di mana data yang relevan dengan tugas analisis yang diambil dari database.
4. Data transformasi (dimana data diubah dan digabung ke dalam bentuk yang sesuai untuk pertambangan dengan melakukan ringkasan atau agregasi operasi) Terkadang transformasi data dilakukan sebelum proses seleksi data, khususnya dalam kasus *data warehouse*.
5. Data mining, merupakan proses esensial dimana metode cerdas diaplikasikan untuk mengekstrak data pola.
6. Evaluasi Pola, untuk mengidentifikasi pola yang benar-benar menarik yang mewakili pengetahuan.
7. Presentasi pengetahuan, dimana visualisasi dan teknik representasi pengetahuan digunakan untuk menyajikan pengetahuan hasil *data mining* kepada pengguna.

Langkah 1 sampai 4 merupakan berbagai bentuk preprocessing data, dimana data dipersiapkan untuk *data mining*. Hal ini, menunjukkan bahwa data mining sebagai salah satu langkah dalam proses KDD, karena dapat mengungkap pola-pola tersembunyi yang digunakan untuk evaluasi.

2.2.2 Tugas *Data Mining*

Secara umum, tugas *data mining* dapat diklasifikasikan kedalam 2 kategori [HAN03], yaitu :

a. Prediktif

   Tujuan dari tugas prediktif adalah untuk memprediksi nilai dari atribut tertentu berdasarkan pada nilai atribut-atribut lain. Atribut yang diprediksi umumnya dikenal sebagai target atau variable tak bebas, sedangkan atribut-atribut yang digunakan untuk membuat prediksi dikenal sebagai *explanatory* atau *variable* bebas.

b. Deskriptif

   Tujuan dari tugas deskriptif adalah untuk menurunkan pola-pola (korelasi, *trend, cluster,* teritori, dan anomali) yang meringkas hubungan yang pokok dalam data. Tugas *data mining* deskriptif sering merupakan penyelidikan dan seringkali memerlukan teknik *post-processing* untuk validasi dan penjelasan hasil.

2.2.3 Fungsi *Data Mining*

Fungsi *data mining* dan macam-macam pola yang dapat ditemukan menurut [HAN03], yaitu:

1. *Concept/Class Description*: *Characterization and Discrimination*

   Data characterization adalah ringkasan dari semua karakteristik atau fitur dari data yang telah diperoleh dari target kelas. Data yang sesuai dengan kelas yang telah ditentukan oleh pengguna biasanya dikumpulkan di dalam *database*. Misalnya, untuk mempelajari karakteristik produk perangkat lunak dimana pada tahun lalu seluruh penjualan telah meningkat sebesar 10%, data yang terkait dengan produk-produk tersebut dapat dikumpulkan dengan menjalankan sebuah *query SQL*.

Data discrimination adalah perbandingan antara fitur umum objek data target kelas dengan fitur umum objek dari satu atau satu set kelas lainnya. target diambil melalui *query database*. Misalnya, pengguna mungkin ingin membandingkan fitur umum dari produk perangkat lunak yang pada tahun lalu penjualannya meningkat sebesar 10% tetapi selama periode yang sama seluruh penjualan juga menurun setidaknya 30%.

2. *Mining Frequent Patterns, Associations, and Correlations*

Frequent Patterns adalah pola yang sering terjadi di dalam data. Ada banyak jenis dari *frequent patterns*, termasuk di dalamnya pola, sekelompok *item set*, *sub-sequence*, dan sub-struktur. Sebuah *frequent patterns* biasanya mengacu pada satu set item yang sering muncul bersama-sama dalam suatu kumpulan data transaksional, misalnya seperti susu dan roti.

Frequent patterns sering mengarah pada penemuan asosiasi yang menarik dan korelasi dalam data.*Associations Analysis* adalah pencarian aturan-aturan asosiasi yang menunjukan kondisi-kondisi nilai atribut yang sering terjadi bersama-sama dalam sekumpulan data. Analisis asosiasi sering digunakan untuk menganalisa *Market Basket Analysis* dan data transaksi.

3. Classification and Prediction

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep dengan tujuan memprediksikan kelas untuk data yang tidak diketahui kelasnya. Model yang diturunkan didasarkan pada analisis dari training data (yaitu objek data yang memiliki label kelas yang diketahui).

Model yang diturunkan dapat direpresentasikan dalam berbagai bentuk seperti *If-then* klasifikasi, *decision tree*, naïve bayes, dan sebagainya.Teknik *classification* bekerja dengan mengelompokkan data berdasarkan *data training* dan nilai atribut klasifikasi. Aturan pengelompokan tersebut akan digunakan untuk klasifikasi data baru ke dalam kelompok yang ada.

Dalam banyak kasus, pengguna ingin memprediksikan nilai-nilai data yang tidak tersedia atau hilang (bukan label dari kelas). Dalam kasus ini nilai data yang akan diprediksi merupakan data *numeric*. Disamping itu, prediksi lebih menekankan pada identifikasi *trend* dari distribusi berdasarkan data yang tersedia.

4. *Cluster Analysis*

   Cluster adalah kumpulan objek data yang mirip satu sama lain dalam kelompok yang sama dan berbeda dengan objek data di kelompok lain. Sedangkan, *Clustering* atau Analisis *Custer* adalah proses pengelompokkan satu set benda-benda fisik atau abstrak kedalam kelas objek yang sama. Tujuannya adalah untuk menghasilkan pengelompokan objek yang mirip satu sama lain dalam kelompok-kelompok. Semakin besar kemiripan objek dalam suatu *cluster* dan semakin besar perbedaan tiap *cluster* maka kualitas analisis *cluster* semakin baik.

5. *Outlier analysis*

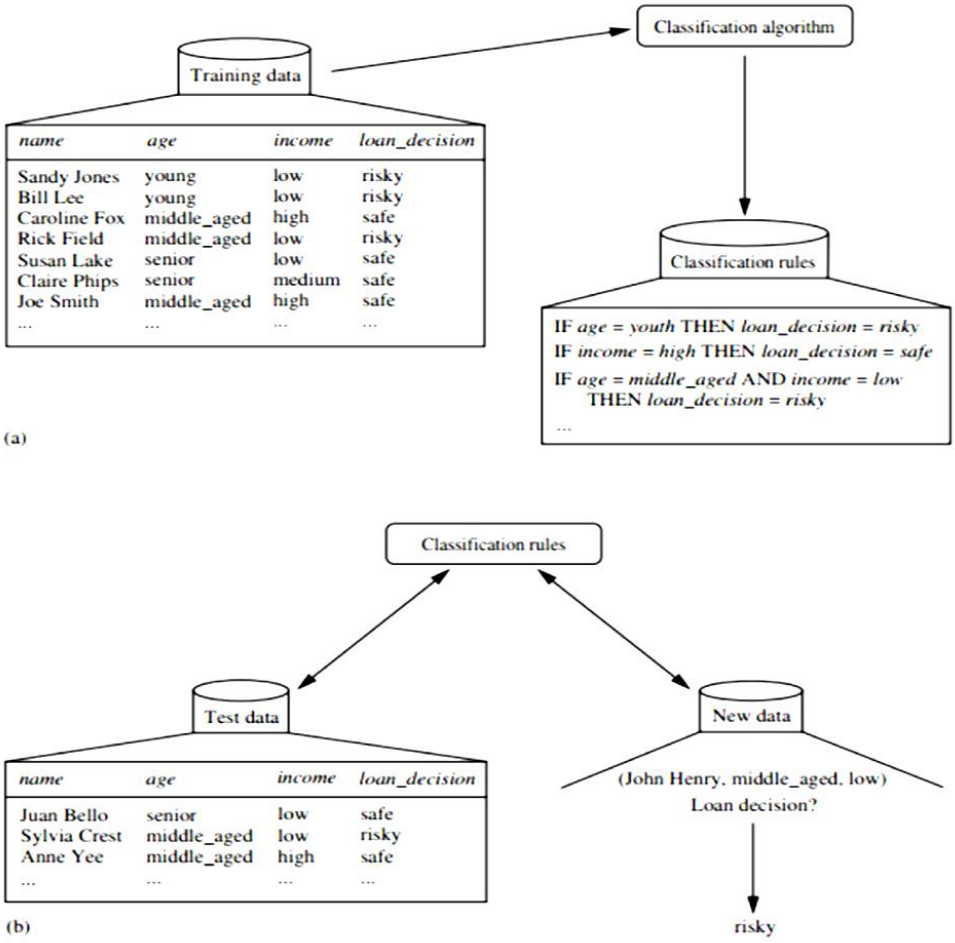
   Outlier merupakan objek data yang tidak mengikuti perilaku umum dari data. *Outlier* dianggap sebagai noise atau pengecualian. Analisis *data outlier* dapat dianggap sebagai *noise* atau pengecualian. Analisis *data outlier* dinamakan *Outlier Mining*. Teknik ini berguna dalam *fraud detection* dan *rare events analysis*.

6. *Evolution Analysis*

   Analisis evolusi data menjelaskan dan memodelkan *trend* dari objek yang memiliki perilaku yang berubah setiap waktu. Teknik ini dapat meliputi karakterisasi, diskriminasi, asosiasi, klasifikasi, atau *clustering* dari data yang berkaitan dengan waktu.

2.3 Klasifikasi

Menurut Mike Chapple, klasifikasi adalah teknik data mining yang dilakukan untuk memprediksi kelas atau properti dari setiap instance data [CHA02].



Gambar 2.1 Tahapan Klasifikasi Data Mining.

Tahapan dari klasifikasi dalam data mining menurut [HAN03]terdiri dari :

a. Pembangunan Model

   Pada tahapan ini dibuat sebuah model untuk menyelesaikan masalah klasifikasi class atau attribut dalam data. Tahap ini merupakan fase pelatihan, dimana data latih dianalisis menggunakan algoritma klasifikasi, sehingga model pembelajaran direpresentasikan dalam bentuk aturan klasifikasi.

b. Penerapan Model

Pada tahapan ini model yang sudah dibangun sebelumnya digunakan untuk menentukan attribut / class dari sebuah data baru yang attribut / classnya belum diketahui sebelumnya. Tahap ini digunakan untuk memperkirakan keakuratan aturan klasifikasi terhadap data uji. Jika model dapat diterima, maka aturan dapat diterapkan terhadap klasifikasi data baru.

2.4 Teorema Bayes

Pengklasifikasian adalah sebuah fungsi yang menugaskan data tertentu kedalam sebuah kelas. Ide dasar aturan Bayes adalah hasil dari hipotesis atau peristiwa (H) dapat diperkirakan berdasarkan pada beberapa evidence (E) yang diamati.

Hal penting dalam Bayes adalah

a. Sebuah probabilitas awal/priori H atau P(H), adalah probabilitas dari suatu hipotesis sebelum bukti diamati.
b. Sebuah probabilitas posterior H atau P(H|E), adalah probabilitas dari suatu hipotesis setelah bukti-bukti yang diamati ada.

$$P(H \mid E) = \frac{P(E \mid H) \times P(H)}{P(E)} \text{.....(2.1)}$$

Keterangan :

P(H|E) : Probabilitas posterior bersyarat (*Conditional Probability*) suatu hipotesis H terjadi jika diberikan evidence/bukti E terjadi.

P(E|H) : Probabilitas sebuah evidence E terjadi akan mempengaruhi hipotesis H.

P(H) : Probabilitas awal (priori) hipotesis H terjadi tanpa memandang evidence apapun.

P(E) : Probabilitas awal (priori) evidence E terjadi tanpa memandang hipotesi/evidence yang lain.

Contoh Teorema Bayes adalah sebagai berikut :

Peramalan cuaca untuk memperkirakan terjadinya hujan, misal ada faktor yang mempengaruhi terjadinya hujan yaitu mendung. Jika diterapkan dalam Naïve Bayes maka probabilitas terjadinya hujan jika bukti mendung sudah diamati :

$$P(Hujan \mid Mendung) = \frac{P(Mendung \mid Hujan) \times P(Hujan)}{P(Mendung)} \quad \text{.....(2.2)}$$

Keterangan :

P(Hujan|Mendung) adalah nilai probabilitas hipotesis hujan terjadi jika bukti mendung sudah diamati.

P(Mendung|Hujan) adalah probabilitas bahwa mendung yang diamati akan mempengaruhi terjadinya hujan.

P(Hujan) adalah probabilitas awal hujan tanpa memandang bukti apapun

P(Mendung) adalah probabilitas terjadinya mendung.

Teorema Bayes juga bisa menangani beberapa evidence, misalnya ada E_1, E_2, dan E_3, maka probabilitas posterior untuk hipotesis hujan:

$$P(H \mid E_1, E_2, E_3) = \frac{P(E_1, E_2, E_3 \mid H) \times P(H)}{P(E_1, E_2, E_3)} \quad \text{.....(2.3)}$$

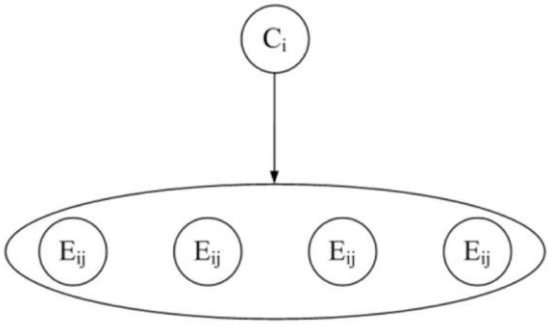
Bentuk diatas dapat diubah menjadi:

$$P(H \mid E_1, E_2, E_3) = \frac{P(E_1 \mid H) \times P(E_2 \mid H) \times P(E_3 \mid H) \times P(H)}{P(E_1, E_2, E_3)} \quad \text{.....(2.4)}$$

Untuk contoh diatas, jika ditambahkan evidence suhu udara dan angin

$$P(Hujan \mid Mendung, Suhu, Angin) = \frac{P(Mendung \mid Hujan) \times P(Suhu \mid Hujan) \times P(Angin \mid Hujan) \times P(Hujan)}{P(Mendung, Suhu, Angin)} \quad \text{.....(2.5)}$$

Pengklasifikasian menggunakan Teorema Bayes ini membutuhkan biaya komputasi yang mahal (waktu prosessor dan ukuran memory yang besar) karena kebutuhan untuk menghitung nilai probabilitas untuk tiap nilai dari perkalian kartesius untuk tiap nilai atribut dan tiap nilai kelas.



Gambar 2.2 Ilustrasi Teorema Bayes

2.5 Naïve Bayes Classifier

Klasifikasi *Naïve Bayes* adalah metode yang berdasarkan probabilitas dan Teorema Bayes dengan asumsi bahwa setiap variabel bersifat bebas (*independence*) dan mengasumsikan bahwa keberadaan sebuah fitur tidak ada kaitannya dengan keberadaan fitur yang lain. Asumsi keindependenan atribut akan menghilangkan kebutuhan banyaknya jumlah data latih dari perkalian kartesius seluruh atribut yang dibutuhkan untuk mengklasifikasikan suatu data.

Formulasi Naïve Bayes untuk klasifikasi adalah

$$P(Y|X) = \frac{P(Y)\prod_{i=1}^{q} P(X_i|Y)}{P(X)} \quad \text{.....(2.6)}$$

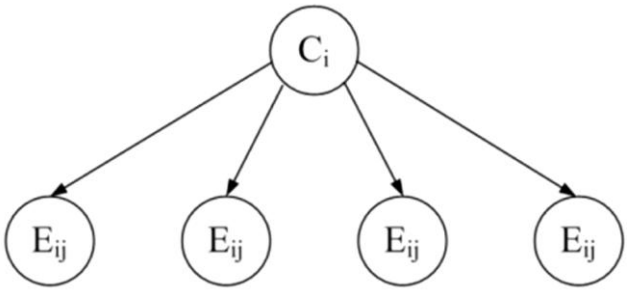
Keterangan :

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

P(Y) = Probabilitas awal kelas Y

$\prod_{i=1}^{q} P(X_i|Y)$ = Probabilitas independen kelas Y dari semua fitur dalam vektor X

Karena P (X) selalu tetap, sehingga dalam perhitungan prediksi nantinya cukup hanya dengan menghitung $P(Y)\prod_{i=1}^{q} P(X_i|Y)$.



Gambar 2.3 Ilustrasi Naive Bayes.

Umumnya, Naïve Bayes mudah dihitung untuk fitur bertipe kategoris seperti pada contoh diatas. Namun untuk tipe numerik (kontinu), ada perlakuan khusus sebelum dimasukkan dalam Naïve Bayes, yaitu :

1. Melakukan diskretisasi pada setiap fitur kontinu dan mengganti nilai fitur kontinu tersebut dengan nilai interval diskret. Pendekatan ini dilakukan dengan mentransformasi fitur kontinu ke dalam fitur ordinal.

2. Mengasumsikan bentuk tertentu dari distribusi probabilitas untuk fitur kontinu dan memperkirakan parameter distribusi dengan data pelatihan. Distribusi Gaussian biasanya dipilih untuk mempresentasikan probabilitas bersyarat dari fitur kontinu pada sebuah kelas P(Xi|Y). Untuk setiap kelas yj, probabilitas bersyarat kelas yj untuk fitur Xi adalah

$$P(X_i = x_i \mid Y = y_j) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} \exp^{-\frac{(x_i-\mu_{ij})^2}{2\sigma_{ij}^2}} \quad \text{.....(2.7)}$$

Keterangan :

μ_{ij} = mean sampel Xi ($\bar{x}$) dari semua data latih.

$2\sigma_{ij}^2$ = varian sampel (s^2) dari data latih.

2.5.1 Algoritma Klasifikasi Naïve Bayes

Algoritma Klasifikasi Naïve Bayes dihitung sesuai dengan rumus Naïve Bayes $P(Y)\prod_{i=1}^{q} P(X_i|Y)$, yang langkah-langkah perhitungannya dijelaskan sebagai berikut [PRA04] :

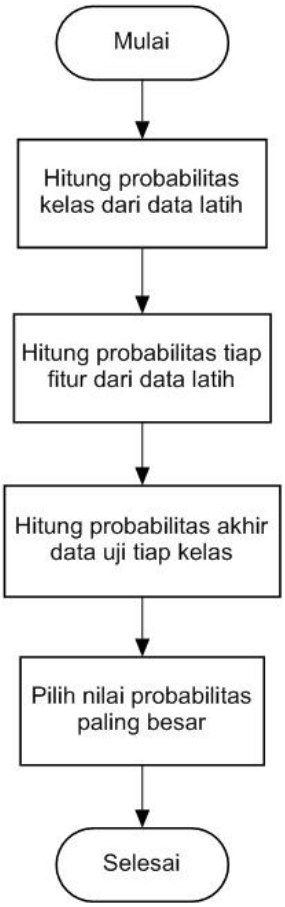
1. Menghitung nilai probabilitas kelas berdasarkan data latih $\rightarrow P(Y)$
2. Menghitung nilai probabilitas tiap fitur berdasarkan data latih$\rightarrow \prod_{i=1}^{q} P(X_i|Y)$

   Untuk fitur bertipe numerik menggunakan rumus berikut :

$$P(X_i = x_i \mid Y = y_j) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} \exp^{-\frac{(x_i-\mu_{ij})^2}{2\sigma_{ij}^2}}$$

   Fitur numerik berikut ini dihitung tiap data uji.
3. Menghitung nilai probabilitas akhir
   - Mengalikan hasil dari $P(Y)$ dan $\prod_{i=1}^{q} P(X_i|Y)$ pada masing-masing kelas dan data uji.
4. Data uji akan diklasifikasikan pada kelas dengan nilai probabilitas akhir terbesar.

Berikut flowchart perhitungan naïve bayes berdasarkan penjelasan diatas.



Gambar 2.4 *Flowchart* Naïve Bayes.

2.5.2 Karateristik Naïve Bayes

Karakteristik *Naïve Bayes* [PRA04] bekerja berdasarkan teori probabilitas yang memandang semua fitur dari data sebagai bukti dalam probabilitas. Hal ini memberikan karakteristik *Naïve Bayes* sebagai berikut :

1. Metode *Naïve Bayes* teguh (*robust*) terhadap data-data yang terisolasi yang biasanya merupakan data dengan karakteristik berbeda (*outlier*). *Naïve Bayes* juga bisa menangani nilai atribut yang salah dengan mengabaikan data latih selama proses pembangunan dan prediksi.
2. Tangguh menghadapi atribut yang tidak relevan.
3. Atribut yang mempunyai korelasi bisa mendegradasi kinerja klasifikasi *Naïve Bayes* karena asumsi independensi tersebut sudah tidak ada.

Naive Bayes memiliki beberapa keuntungan dan kekurangan yaitu sebagai berikut :

1. Keuntungan *Naive Bayes*
   - Cepat dan efisiensi ruang.
   - Kokoh terhadap atribut yang tidak relevan.
   - Hanya memerlukan sejumlah kecil data pelatihan untuk mengestimasi parameter (rata-rata dan variansi dari variabel) yang dibutuhkan untuk klasifikasi.
   - Menangani Kuantitatif dan data diskrit.
2. Kekurangan *Naive Bayes*
   - Tidak berlaku jika *Probabilitas* kondisionalnya adalah nol, apabila nol maka *Probabilitas* prediksi akan bernilai nol juga.
   - Mengasumsikan Variabel bebas.

Contoh perhitungan Naïve Bayes adalah sebagai berikut [PRA04] :

1. Jika ada sebuah data uji berupa hewan musang dengan nilai fitur: penutup kulit = rambut, melahirkan = ya, berat = 15, masuk kelas manakah untuk hewan musang tersebut ?

Tabel 2.1 Contoh Data Latih Klasifikasi Hewan

Nama hewan	Penutup kulit	Melahirkan	Berat	Kelas
Ular	Sisik	Ya	10	Reptil
Tikus	Bulu	Ya	0.8	Mamalia
Kambing	Rambut	Ya	21	Mamalia
Sapi	Rambut	Ya	120	Mamalia
Kadal	Sisik	Tidak	0.4	Reptil
Kucing	Rambut	Ya	1.5	Mamalia
Bekicot	Cangkang	Tidak	0.3	Reptil
Harimau	Rambut	Ya	43	Mamalia
Rusa	Rambut	Ya	45	Mamalia
Kura-Kura	Cangkang	Tidak	7	Reptil

Tabel 2.2 Contoh Perhitungan Nilai Probabilitas Tiap Fitur

Penutup kulit		**Melahirkan**	
Mamalia	Reptil	Mamalia	Reptil
Sisik = 0 Bulu = 1 Rambut = 5 Cangkang = 0	Sisik = 2 Bulu = 0 Rambut = 0 Cangkang = 2	Ya = 6 Tidak = 0	Ya = 1 Tidak = 3
P(Kulit = Sisik | Mamalia) = 0 P(Kulit = Bulu| Mamalia) = 1/6 P(Kulit = Rambut | Mamalia) = 5/6 P(Kulit = Cangkang | Mamalia) = 0	P(Kulit = Sisik | Reptil) = 0.5 P(Kulit = Bulu | Reptil) = 0 P(Kulit = Rambut | Reptil) = 0 P(Kulit = Cangkang | Reptil) = 0.5	P(Lahir = Ya | Mamalia) = 1 P(Lahir = Tidak | Mamalia) = 0	P(Lahir = Ya | Reptil) = 0.25 P(Lahir = Tidak | Reptil) = 0.75

Berat		**Kelas**	
Mamalia	Reptil	Mamalia	Reptil
$\bar{x}_{mamalia} = 38.55$ $s^2_{mamalia} = 1960.255$ $s_{mamalia} = 44.275$	$\bar{x}_{reptil} = 4.425$ $s^2_{reptil} = 23.6425$ $s_{reptil} = 4.8624$	Mamalia = 6 P(Mamalia) = 6/10 = 0.6	Reptil = 4 P(Reptil) = 4/10 = 0.4

Hitung nilai probabilitas untuk fitur dengan tipe numerik yaitu berat.

$$\text{P(Berat} = 15 \mid \text{Mamalia)} = \frac{1}{\sqrt{2\pi}\,44.275}\exp^{-\frac{(15-38.55)^2}{2\times 1960.255}} = 0.0078$$

$$\text{P(Berat} = 15 \mid \text{Reptil)} = \frac{1}{\sqrt{2\pi}\,4.8624}\exp^{-\frac{(15-4.425)^2}{2\times 23.6425}} = 0.0077$$

Hitung probabilitas akhir untuk setiap kelas:

P(X | Mamalia) = P(Kulit = Rambut | Mamalia) x P(Lahir = Ya | Mamalia) x P(Berat = 15 | Mamalia)

= 5/6 x 1 x 0.0078 = 0.0065

P(X | Reptil) = P(Kulit = Rambut | Reptil) x P(Lahir = Ya | Reptil) x P(Berat = 15 | Reptil)

= 0 x 0.25 x 0.0077 = 0

Nilai tersebut dimasukkan untuk mendapatkan probabilitas akhir:

- P(Mamalia | X) = α x P(Mamalia) x P(X | Mamalia)

  = α x 0.6 x 0.0065

  = 0.0039α
- P(Reptil | X) = α x P(Reptil) x P(X | Reptil) = α x 0.4 x 0 = 0

Untuk α = 1/P(X) pasti nilainya konstan sehingga tidak perlu diketahui karena terbesar dari dua kelas tersebut tidak dapat dipengaruhi P(X).

Karena nilai probabilitas akhir (posterior probability) terbesar ada di kelas **mamalia (0.0039α)**, maka data uji musang diprediksi sebagai kelas **mamalia**.

2.6 Riset – Riset Terkait

Naïve Bayes merupakan metode popular yang banyak digunakan untuk klasifikasi. Beberapa riset yang telah dilakukan berkaitan dengan kasus klasifikasi yang menggunakan metode Naïve Bayes dan penelitian tentang bantuan langsung sementara masyarakat, antara lain :

Penelitian oleh Nur Indah Sari yang berjudul "*Klasifikasi Kecendrungan Penyelesaian Studi Mahasiswa Baru Dengan Menggunakan Metode Naïve Bayes"*. Atribut yang digunakan 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, adapun data yang diambil dalam penilitian ini adalah sampel dari 40 *record* dengan kelas "Lama" dan "Tidak Lama" masing-masing berjumlah 19 untuk kelas lama dan 21 untuk kelas tidak lama yang akan dibagi menjadi data latih data uji. Dan mengunakan 30 data latih, kelas lama 19 dan kelas tidak lama 21.10 data uji kelas lama 5 dan kelas tidak lama 5. Hasil dari 10 data uji, terdapat 3 data yang hasil klasifikasi dengan mengaunkan metode naïve bayes yang tidak sama dengan hasil data yang sebenarnya. Keakurasian ketepatan hasil klasifikasi dengan naïve bayes

untuk mengetahui tingkat kecenderungan penyelesaian studi mahasiswa sistem memiliki akurasi kebenaran sistem sebesar 70% dan nilai error sebesar 30%.

Nur Rochmah Dyah, Edy Nugroho dan Eko Aribowo melakukan penelitian mengenai "*Sistem Penentuan Penerima Bantuan Langsung Tunai (BLT) dengan Metode Analitycal Hirarchy Process"*. Dalam penelitian ini dilakukan pada Badan Pusat Statistika di Banjarnegara, dengan menghasilkan sebuah sistem pendukung keputusan untuk proses penerimaan Bantuan Langsung Tunai dengan menggunakan pendekatan *top down* dengan inputan data calon penerima BLT, semua data calon, kriteria, dan nilai.

Penelitian yang berjudul *"Sistem Prediksi Prestasi (IPK) Berdasarkan Latar belakang Sekolah Asal dan Atribut Mahasiswa Ketika Masuk Kuliah Menggunakan Naïve Bayes"* oleh Meinggan Vilian Sari. Adapun Data yang diolah dalam penelitian ini adalah data mahasiswa Teknik Informatika Universitas Muhammadiyah Gresik angkatan 2010 semester 6 sejumlah 103 mahasiswa. Atribut yang digunakan adalah IPK dan dibagi menjadi 2 kelas, yaitu IPK Tinggi dan Rendah. Hasil penelitian menunjukkan bahwa akurasi tertinggi yang didapatkan pada pengujian sistem ini adalah 84.62%.

Selain itu, penelitian yang terkait mengenai klasifikasi adalah penelitian yang dilakukan oleh Ervina Rizka Anandita dengan judul "*Klasifikasi Tebu dengan Menggunakan Algoritma Naïve Bayes Classification pada Dinas Kehutanan dan Perkebunan Pati*". Data yang digunakan pada penelitian ini adalah data dari Dinas Kehutanan dan Perkebunan Kabupaten Pati, dengan atribut hasil panen berdasarkan jenis tebu yang dimiliki oleh perkebunan tebu yang meliputi jenis tebu, hasil produksi, umur panen, tinggi tanaman, diameter batang, daerah tanam, bobot batang, rendeman dan macam got yang digunakan. Berdasarkan atribut tersebut, klasifikasi tebu dibagi menjadi 2 kelas yaitu jenis tebu yang produktif atau tidak produktif dengan akurasi tertinggi sebesar 73,3%.