

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Pengertian Sekolah**

Sekolah adalah sistem interaksi sosial suatu organisasi keseluruhan terdiri atas interaksi pribadi terkait bersama dalam suatu hubungan organik (Wayne dalam buku Soebagio Atmodiwiro, 2000:37). Sedangkan berdasarkan undang-undang no 2 tahun 1989 sekolah adalah satuan pendidikan yang berjenjang dan berkesinambungan untuk menyelenggarakan kegiatan belajar mengajar.

Menurut Daryanto (1997:544), sekolah adalah bangunan atau lembaga untuk belajar serta tempat menerima dan memberi pelajaran.

Jadi, sekolah sebagai suatu sistem sosial dibatasi oleh sekumpulan elemen kegiatan yang berinteraksi dan membentuk suatu kesatuan sosial sekolah yang demikian bersifat aktif kreatif artinya sekolah dapat menghasilkan sesuatu yang bermanfaat bagi masyarakat dalam hal ini adalah orang-orang yang terdidik.

Dari definisi tersebut bahwa sekolah adalah suatu lembaga atau organisasi yang diberi wewenang untuk menyelenggarakan kegiatan pembelajaran. Sebagai suatu organisasi sekolah memiliki persyaratan tertentu.

Sekolah adalah suatu lembaga atau tempat untuk belajar seperti membaca, menulis dan belajar untuk berperilaku yang baik. Sekolah juga merupakan bagian integral dari suatu masyarakat yang berhadapan dengan kondisi nyata yang terdapat dalam masyarakat pada masa sekarang. Sekolah juga merupakan lingkungan kedua tempat anak-anak berlatih dan menumbuhkan kepribadiannya. (Zanti Arbi dalam buku Made Pidarta, 1997:171).

Pada tanggal 16 mei 2005 diterbitkan peraturan pemerintah (PP) nomor 19 tahun 2005 tentang standar nasional pendidikan. Dengan PP 19/2005 itu, semua sekolah di Indonesia diarahkan dapat menyelenggarakan pendidikan yang memenuhi standar nasional. pendidikan standar wajib

dilakukan oleh sekolah, delapan standar tersebut setahap demi setahap harus bisa dipenuhi oleh sekolah. Secara berkala sekolah pun diukur pelaksanaan delapan standar itu melalui akreditasi sekolah.

Berdasarkan dari beberapa teori di atas maka dapat ditarik kesimpulan bahwa sekolah adalah bagian integral dari suatu masyarakat yang berhadapan dengan kondisi nyata yang terdapat dalam masyarakat pada masa sekarang dan sekolah juga merupakan alat untuk mencapai pendidikan yang bermutu dan memenuhi standar nasional pendidikan.

## **2.2 Prestasi Siswa**

Prestasi seorang siswa itu ditentukan dengan berbagai macam proses seperti dengan belajar. Dengan belajar siswa dapat memperoleh pengetahuan secara luas, keberhasilan seorang siswa dalam belajar juga ditentukan dengan indikator yang dijadikan sebagai tolak ukur dalam menyatakan bahwa suatu proses belajar mengajar dapat dikatakan berhasil atau tidak.

### **2.2.1 Pengertian Belajar**

Belajar adalah proses usaha yang dilakukan seseorang untuk memperoleh suatu perubahan tingkah laku yang baru secara keseluruhan, sebagai hasil pengalamannya sendiri dalam interaksi dengan lingkungannya. Dan belajar merupakan suatu aktivitas mental / psikis yang berlangsung dalam interaksi yang aktif dengan lingkungan, yang menghasilkan perubahan – perubahan dalam pengetahuan, pemahaman, keterampilan dan nilai sikap.

### **2.2.2 Pengertian Prestasi Belajar**

Kemampuan intelektual siswa sangat menentukan keberhasilan siswa dalam memperoleh prestasi. Untuk mengetahui berhasil tidaknya seseorang dalam belajar maka perlu dilakukan suatu evaluasi, tujuannya untuk mengetahui prestasi yang diperoleh siswa setelah proses belajar mengajar berlangsung.

Adapun prestasi dapat diartikan hasil diperoleh karena adanya aktivitas belajar yang telah dilakukan. Namun banyak orang beranggapan bahwa yang dimaksud dengan belajar adalah mencari ilmu dan menuntut ilmu. Ada lagi yang lebih khusus mengartikan bahwa belajar menyerap pengetahuan. Belajar adalah perubahan yang terjadi dalam tingkah laku manusia. Proses tersebut tidak akan terjadi apabila tidak ada suatu yang mendorong pribadi yang bersangkutan.

### **2.2.3 Nilai Raport Siswa**

Fungsi pokok evaluasi hasil belajar siswa secara umum adalah untuk mengukur tingkat kemajuan siswa dalam belajar, untuk menyusun rencana belajar selanjutnya dan untuk memperbaiki proses pembelajaran (Muhammad Irham dan Novan A.W., 2013: 217). Laporan evaluasi hasil belajar siswa dituliskan pada sebuah dokumen yaitu rapor. Nilai rapor ditulis berdasarkan hasil belajar siswa dalam satu semester dan ditulis pada akhir semester.

Nilai Raport merupakan hasil kumpulan nilai mata pelajaran yang dimiliki setiap siswa yang berisi laporan nilai selama 1 semester. Raport diterimakan sebagai tolak ukur dan untuk mengetahui perkembangan terhadap prestasi siswa setelah mengikuti proses pembelajaran.

Melalui raport wali kelas dapat mengetahui kekuatan dan kelemahan siswa dalam kelas yang diampunya wali kelas dapat menentukan strategi dalam pengelolaan kelas yang menjadi tanggung jawabnya misalnya dengan menata strategis belajar untuk membantu siswa meningkatkan kompetensi siswa atau membantu mengatasi kesulitan belajar siswa yang lemah.

### **2.2.4 Prestasi Akademik**

Prestasi akademik merupakan hasil perubahan perilaku yang meliputi ranah kognitif, ranah afektif dan ranah psikomotor yang merupakan ukuran keberhasilan siswa. Ranah kognitif merupakan

kemampuan yang selalu dituntut kepada anak didik untuk dikuasai. Karena penguasaan kemampuan pada tingkatan ini menjadi dasar bagi penguasaan ilmu pengetahuan ( Slameto, 1995). Prestasi akademik dinyatakan sebagai pengetahuan yang dicapai atas ketrampilan yang dikembangkan dalam mata pelajaran tertentu di sekolah. Dapat disimpulkan bahwa prestasi akademik adalah besarnya penguasaan bahan pelajaran yang telah dicapai siswa yang diwujudkan berupa nilai.

### **2.2.5 Prestasi Non Akademik**

Prestasi non akademik adalah suatu prestasi yang tidak dapat diukur dan dinilai menggunakan angka, biasanya dalam hal olahraga, pramuka, PMR, atau kesenian semisal drum band, melukis, menari, dan lain – lain.

Prestasi ini yang biasa dicapai oleh siswa sewaktu mengikuti kegiatan ekstrakurikuler di Sekolah. Kegiatan ekstrakurikuler adalah berbagai kegiatan sekolah yang dilakukan dalam rangka kesempatan kepada peserta didik untuk dapat mengembangkan potensi, minat, bakat, dan hobi yang dimilikinya yang dilakukan diluar jam sekolah normal.

## **2.3 Klasifikasi**

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*, 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. *Classification* dapat direpresentasikan dalam bentuk pohon keputusan (*decision tree*). Setiap *node* dalam pohon

keputusan menyatakan suatu tes terhadap atribut *dataset*, sedangkan setiap cabang menyatakan hasil dari tes tersebut. Pohon keputusan yang terbentuk dapat diterjemahkan menjadi sekumpulan aturan dalam bentuk *IF condition THEN outcome*. (Mewati Ayub, 2007:7)

Tahapan dari klasifikasi dalam data mining terdiri dari :

1. Pembangunan model, dalam tahapan ini dibuat sebuah model untuk menyelesaikan masalah klasifikasi class atau atribut dalam data, model ini dibangun berdasarkan training set-sebuah contoh data dari permasalahan yang dihadapi, training set ini sudah mempunyai informasi yang lengkap baik atribut maupun classnya
2. Penerapan model, pada tahapan ini model yang sudah dibangun sebelumnya digunakan untuk menentukan atribut / class dari sebuah data baru yang atribut / classnya belum diketahui sebelumnya
3. Evaluasi, pada tahapan ini hasil dari penerapan model pada tahapan sebelumnya dievaluasi menggunakan parameter terukur untuk menentukan apakah model tersebut dapat diterima

## **2.4 Data Mining**

### **2.4.1 Pengertian Data Mining**

*Data mining* (Connolly dan Begg, 2010) adalah suatu proses ekstraksi atau penggalian data yang belum diketahui sebelumnya, namun dapat dipahami dan berguna dari *database* yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting.

*Data mining* biasa juga disebut dengan “Data atau *knowledge discovery*” atau menemukan pola tersembunyi pada data. *Data mining* adalah proses dari menganalisa data dari prespektif yang berbeda dan menyimpulkannya ke dalam informasi yang berguna.

*Data mining* (Han dan Kamber, 2006 : 5) didefinisikan sebagai proses mengekstrak atau menambang pengetahuan yang dibutuhkan dari sejumlah data besar.

Pada prosesnya *data mining* akan mengekstrak informasi yang berharga dengan cara menganalisis adanya pola-pola ataupun hubungan keterkaitan tertentu dari data-data yang berukuran besar. *Data mining* berkaitan dengan bidang ilmu-ilmu lain, seperti *Database System*, *Data Warehousing*, *Statistic*, *Machine Learning*, *Information Retrieval*, dan Komputasi Tingkat Tinggi. Selain itu *data mining* didukung oleh ilmu lain seperti *Neural Network*, Pengenalan Pola, *Spatial Data Analysis*, *Image Database*, *Signal Processing*.

Beberapa *survey* tentang proses pemodelan dan metodologi menyatakan bahwa, “*Data mining* digunakan sebagai penunjuk, dimana *data mining* menyajikan intisari atas sejarah, deskripsi dan sebagai standar petunjuk mengenai masa depan dari sebuah proses model *data mining*”.

Berdasarkan beberapa pengertian tersebut dapat ditarik kesimpulan bahwa *Data Mining* adalah suatu teknik menggali informasi berharga yang terpendam atau tersembunyi pada suatu koleksi data (*database*) yang sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui.

#### **2.4.2 Tugas Data Mining**

Tugas-tugas dalam *data mining* secara umum dibagi ke dalam dua kategori utama:

1. Prediktif

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

2. Deskriptif

Tujuan dari tugas deskriptif adalah untuk menurunkan pola-pola (korelasi, *trend*, *cluster*, trayektori, dan anomali) yang meringkas hubungan yang pokok dalam data. Tugas *data mining* deskriptif sering

merupakan penyelidikan dan seringkali memerlukan teknik postprocessing untuk validasi dan penjelasan hasil.

### 2.4.3 Fungsi Data Mining

*Data mining* juga memiliki beberapa fungsionalitas yaitu *Concept/Class Description: Characterization and Discrimination, Mining Frequent Patterns, Associations, and Correlations, Classification and Prediction, Cluster Analysis, Outlier analysis, dan Evolution analysis*. (Han dan Kamber, 2006. Berikut adalah penjelasan dari masing-masing fungsi diatas:

#### 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*.

Sedangkan, *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.

*Associations Analysis* adalah pencarian aturan-aturan asosiasi yang menunjukkan 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*, naive 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. *Classification* dapat direpresentasikan dalam bentuk pohon keputusan (*decision tree*). Setiap *node* dalam pohon keputusan menyatakan suatu tes terhadap atribut *dataset*, sedangkan setiap cabang menyatakan hasil dari tes tersebut. Pohon keputusan yang terbentuk dapat diterjemahkan menjadi sekumpulan aturan dalam bentuk *IF condition THEN outcome*.

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.

### 2.5 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. [PRA04]

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.

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

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 hipotesis/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 :

$$P(\text{Hujan} | \text{Mendung}) = \frac{P(\text{Mendung} | \text{Hujan}) \times P(\text{Hujan})}{P(\text{Mendung})} \dots\dots\dots(2.2)$$

Keterangan :

$P(\text{Hujan}|\text{Mendung})$  adalah nilai probabilitas hipotesis hujan terjadi jika bukti mendung sudah diamati.

$P(\text{Mendung}|\text{Hujan})$  adalah probabilitas bahwa mendung yang diamati akan mempengaruhi terjadinya hujan.

$P(\text{Hujan})$  adalah probabilitas awal hujan tanpa memandang bukti apapun

$P(\text{Mendung})$  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:

$$P(H | E_1, E_2, E_3) = \frac{P(E_1, E_2, E_3 | H) \times P(H)}{P(E_1, E_2, E_3)} \dots\dots\dots(2.3)$$

Bentuk diatas dapat diubah menjadi:

$$P(H | E_1, E_2, E_3) = \frac{P(E_1 | H) \times P(E_2 | H) \times P(E_3 | H) \times P(H)}{P(E_1, E_2, E_3)} \dots\dots\dots(2.4)$$

Untuk contoh diatas, jika ditambahkan evidence suhu udara dan angin

$$P(\text{Hujan} | \text{Mendung}, \text{Suhu}, \text{Angin}) = \frac{P(\text{Mendung} | \text{Hujan}) \times P(\text{Suhu} | \text{Hujan}) \times P(\text{Angin} | \text{Hujan}) \times P(\text{Hujan})}{P(\text{Mendung}, \text{Suhu}, \text{Angin})} \dots\dots\dots(2.5)$$

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.

## 2.6 Naive Bayes Classifier

Klasifikasi *Naive 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 Naive Bayes untuk klasifikasi adalah

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

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  $P(Y) \prod_{i=1}^q P(X_i|Y)$ .

Umumnya, Naive Bayes mudah dihitung untuk fitur bertipe kategoris. Namun untuk tipe numerik (kontinu), ada perlakuan khusus sebelum dimasukkan dalam Naive 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  $P(X_i|Y)$ . Untuk setiap kelas  $y_j$ , probabilitas bersyarat kelas  $y_j$  untuk fitur  $X_i$  adalah :

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} \exp \left[ -\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2} \right] \dots \dots \dots (2.7)$$

Keterangan :

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

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

### 2.6.1 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  $\rightarrow \prod_{i=1}^q P(X_i|Y)$

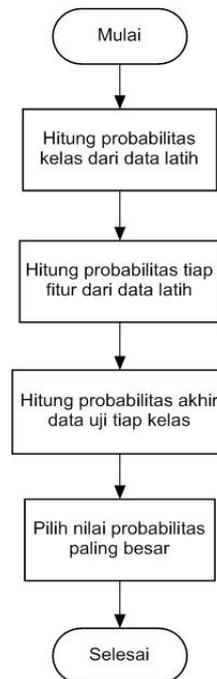
Untuk fitur bertipe numerik menggunakan rumus berikut :

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

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 2.1** Flowchart Naïve Bayes

### 2.6.2 Karakteristik Naive 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.

*Naïve Bayes* memiliki beberapa keuntungan dan kekurangan yaitu sebagai berikut :

1. Keuntungan *Naive Bayes*

1. Cepat dan efisiensi ruang.
  2. Kokoh terhadap atribut yang tidak relevan.
  3. Hanya memerlukan sejumlah kecil data pelatihan untuk mengestimasi parameter (rata-rata dan variansi dari variabel) yang dibutuhkan untuk klasifikasi.
  4. Menangani Kuantitatif dan data diskrit.
2. Kekurangan *Naive Bayes*
1. Tidak berlaku jika *Probabilitas* kondisionalnya adalah nol, apabila nol maka *Probabilitas* prediksi akan bernilai nol juga.
  2. 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
$\bar{x}_{mamalia} = 38.55$ $s_{mamalia}^2 = 1960.255$ $s_{mamalia} = 44.275$	$\bar{x}_{reptil} = 4.425$ $s_{reptil}^2 = 23.6425$ $s_{reptil} = 4.8624$	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.

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

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

Hitung probabilitas akhir untuk setiap kelas:

$$\begin{aligned} P(X | \text{Mamalia}) &= P(\text{Kulit} = \text{Rambut} | \text{Mamalia}) \times P(\text{Lahir} = \text{Ya} | \text{Mamalia}) \times \\ &\quad P(\text{Berat} = 15 | \text{Mamalia}) \\ &= 5/6 \times 1 \times 0.0078 = 0.0065 \end{aligned}$$

$$\begin{aligned}
 P(X | \text{Reptil}) &= P(\text{Kulit} = \text{Rambut} | \text{Reptil}) \times P(\text{Lahir} = \text{Ya} | \text{Reptil}) \times P(\text{Berat} = \\
 &\quad 15 | \text{Reptil}) \\
 &= 0 \times 0.25 \times 0.0077 = 0
 \end{aligned}$$

Nilai tersebut dimasukkan untuk mendapatkan probabilitas akhir:

- $P(\text{Mamalia} | X) = \alpha \times P(\text{Mamalia}) \times P(X | \text{Mamalia})$   
 $= \alpha \times 0.6 \times 0.0065$   
 $= 0.0039\alpha$
- $P(\text{Reptil} | X) = \alpha \times P(\text{Reptil}) \times P(X | \text{Reptil}) = \alpha \times 0.4 \times 0 = 0$

Untuk  $\alpha = 1/P(X)$  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 (0.0039 $\alpha$ )**, maka data uji musang diprediksi sebagai kelas **mamalia**.

## 2.7 Penelitian Terkait

*Naive Bayes* merupakan metode populer yang banyak digunakan untuk klasifikasi. Beberapa riset yang telah dilakukan berkaitan dengan kasus prediksi yang menggunakan metode *Naive Bayes*, antara lain :

Penelitian oleh Bustami yang berjudul “*Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi*”. Variabel yang digunakan sebagai data latih sistem adalah Jenis kelamin, Usia, Status, Pekerjaan, Penghasilan, Cara bayar premi, dan Masa pembayaran premi. Adapun data yang diambil dalam penelitian ini adalah sampel dari 20 *record* dengan kelas “Lancar”, “Kurang Lancar” dan “Tidak Lancar” masing-masing berjumlah 9 untuk kelas lancar, 4 untuk kelas kurang lancar dan 7 untuk kelas tidak lancar. Hasil dari klasifikasi adalah status calon nasabah tersebut masuk dalam klasifikasi “Lancar” dibandingkan dengan hasil class Lancar, Kurang Lancar dan Tidak Lancar.

Yudi Ardiyanto melakukan penelitian mengenai “*Sistem Pendukung Keputusan Kelayakan Pengajuan Kredit Sepeda Motor Menggunakan Metode Naive Bayes di Summit Otto Finance*”. Dalam penelitian sistem pendukung

keputusan ini dilakukan sebagai alat bantu bagi pihak perusahaan dalam menentukan prioritas dari pemohon kredit sebagai acuan pemberian kredit bagi perusahaan, berdasarkan criteria yang ditentukan oleh pihak Summit Otto Finance. Menghasilkan sebuah sistem pendukung keputusan untuk rekomendasi kelayakan penerimaan kredit sepeda motor dengan nilai perhitungan bayes tertinggi yang mendekati nilai kelayakan penerimaan kredit sepeda motor. Kriteria yang digunakan Jenis kelamin, Bukti kepemilikan rumah, Jenis pekerjaan, Penghasilan perbulan, Pengeluaran perbulan, Usia calon debitur dan Karakteristik calon debitur.

Selain itu, penelitian yang terkait mengenai klasifikasi adalah penelitian yang dilakukan oleh Eryina 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%.