

BAB II

LANDASAN TEORI

2.1 Pengertian Koperasi

Beberapa pakar mengemukakan pengertian koperasi :

Bapak Koperasi Dr. Mohammad Hatta berpendapat Koperasi adalah bangunan organisasi sebagai badan usaha bersama berdasarkan atas kekeluargaan. Semua bertanggung jawab dan bekerja sama untuk mencapai tujuan bersama. [IHM]

Undang-undang Perkoperasian No. 17 Tahun 2012 pasal 1 ayat 1. Koperasi adalah badan hukum yang didirikan oleh orang perseorangan atau badan hukum Koperasi, dengan pemisahan kekayaan para anggotanya sebagai modal untuk menjalankan usaha, yang memenuhi aspirasi dan kebutuhan bersama dibidang ekonomi, social, dan budaya sesuai dengan nilai dan prinsip Koperasi. [UMB13]

Koperasi bertujuan menyejahterahkan anggota pada khususnya dan masyarakat pada umumnya serta ikut tatanan perekonomian nasional dalam rangka mewujudkan masyarakat yang maju, adil dan makmur berlandaskan Pancasila dan UUD 1945. Dalam Koperasi Indonesia kepribadian sebagai pencerminan kehidupan yang dipengaruhi keadaan dan lingkungan yang berdasarkan atas asa kekeluargaan bersemboyan Bhinneka Tunggal Ika. Bagi koperasi asa kekeluargaan terdapat keinsyafan, kesadaran, dan tanggung jawab terhadap kerja tanpa memikirkan kepentingan diri sendiri tapi untuk kesejahteraan bersama.

Jenis koperasi didasarkan pada kesamaan kegiatan dan kepentingan anggota. Dasar pengurus sesuai dengan undang-undang koperasi yang berlaku, terdiri atas :

- a. Koperasi Konsumen

Koperasi konsumen ialah koperasi yang anggotanya terdiri atas orang-orang yang mempunyai kepentingan langsung dalam lapangan konsumsi.

b. Koperasi Produsen

Koperasi Produsen ialah koperasi yang anggota-anggotanya terdiri atas para pengusaha, pemilik alat-alat produksi, dan karyawan yang berkepentingan, sedangkan usahanya langsung berhubungan dengan bidang industri atau kerajinan.

c. Koperasi Simpan Pinjam

d. Koperasi Pemasaran

Koperasi pemasaran ialah koperasi yang anggota-anggotanya terdiri atas orang-orang tertentu yang mempunyai kepentingan langsung dalam bidang usaha untuk pemasaran barang dan jasa.

e. Koperasi Primer

Koperasi primer adalah koperasi yang didirikan dan beranggotakan orang perseorangan.

f. Koperasi Sekunder

Koperasi sekunder adalah koperasi yang didirikan dan beranggotakan badan hukum koperasi.

2.1.1 Koperasi Simpan Pinjam

Koperasi Simpan Pinjam adalah Koperasi yang kegiatannya hanya usaha simpan pinjam yang anggotanya orang-orang yang mempunyai kepentingan langsung dalam lapangan perkreditan. Kegiatan anggota koperasi ialah menuang atau menyimpan, jumlah tabungan yang terkumpul dipinjamkan pada para anggota lainnya yang memerlukan dengan tingkat bunga yang telah diatur dalam anggaran rumah tangga koperasi.

Tujuan koperasi simpan pinjam adalah :

1. Membantu keperluan kredit para anggota yang memerlukan,
2. Mendidik para anggota supaya giat menyimpan secara teratur sehingga dapat membentuk modal,

3. Mendidik para anggota untuk hidup teratur dengan menyisihkan sebagian dari penghasilam mereka,
4. Menambah pengetahuan dan informasi tentang perkoperasian.

Salah satu koperasi yang bergerak pada simpan pinjam adalah Koperasi Wanita “Cempaka”. Koperasi yang berdiri sejak 20 November 2009 di Desa Plosobuden Kec. Deket Kab. Lamongan dengan jumlah anggota 155 orang yang terdiri dari 5 pengurus dan 3 pengawas. Pengajuan kredit pada koperasi wanita “cempaka” menggunakan sistem tanggung rentang (berkelompok).

2.1.2 Pengertian Kredit

Thomas Suyatno menyatakan bahwa istilah kredit berasal dari bahasa Yunani (*credere*) yang berarti kepercayaan (*truth atau faith*). Oleh karena itu dasar dari kredit ialah kepercayaan, seseorang atau suatu badan yang memberikan kredit (kreditor) percaya bahwa penerima kredit (debitor) pada masa yang akan datang sanggup memenuhi segala sesuatu yang telah dijanjikan. [KRE06]

Dalam Pasal 1 angka 11 Undang-undang RI No. 10 Tahun 1998 tentang Perbankan, penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan atau kesepakatan pinjam meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam untuk melunasi utangnya setelah jangka waktu tertentu dengan jumlah bunga.

Dilihat dari sudut ekonomi, kredit diartikan sebagai penundaan pembayaran. Maksudnya pengambilan atas penerimaan uang dan atau suatu barang tidak dilakukan bersamaan pada saat menerima akan tetapi pengembaliannya dilakukan pada masa tertentu yang akan datang.

Kolektibilitas adalah suatu pembayaran pokok atau bunga pinjaman oleh nasabah sebagaimana terlihat tata usaha bank berdasarkan Surat Keputusan Direksi Bank Indonesia (BI) No.

32/268/KEP/DIR tanggal 27 Pebruari 1998, maka kredit dapat dibedakan menjadi :

a. Kredit Lancar

Kredit lancar yaitu kredit yang pengambilan pokok pinjaman dan pembayaran bunga tepat waktu, perkembangan rekening baik dan tidak ada tunggakan serta sesuai dengan persyaratan kredit. Kredit lancar mempunyai kriteria sebagai berikut :

- 1) Pembayaran angsuran pokok dan bunga tepat waktu,
- 2) Memiliki mutasi rekening yang aktif,
- 3) Bagian dari kredit yang dijamin dengan uang tunai

b. Kredit Macet

Kredit macet yaitu kredit yang pengembalian pokok pinjaman dan pembayaran bunga terdapat tunggakan telah melampaui jangka waktu yang ditentukan. Kredit macet mempunyai kriteria sebagai berikut :

- 1) Terdapat tunggakan angsuran pokok yang telah melampaui jangka waktu yang ditentukan,
- 2) Kerugian operasional dituntut dengan pinjaman baru,
- 3) Jaminan tidak dapat dicairkan pada nilai wajar, baik dari segi hukum maupun dari segi kondisi pasar.

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 terakrit 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 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 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*, 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 Model Prediksi

Model prediksi berkaitan dengan pembuatan sebuah model yang dapat melakukan pemetaan dari setiap himpunan variable ke setiap targetnya, kemudian menggunakan model tersebut untuk memberikan nilai target pada himpunan baru yang didapat. Ada dua jenis model prediksi (Prasetyo, 2012):

1. Klasifikasi

Klasifikasi digunakan untuk variable target diskret, hanya beberapa jenis kemungkinan nilai target yang didapatkan dan tidak ada nilai deret waktu (*time series*) untuk mendapatkan target nilai akhir.

2. Regresi

Regresi untuk variable bersifat target kontinu, ada nilai deret waktu yang harus dihitung untuk mendapatkan nilai target akhir yang diinginkan.

2.4 Klasifikasi

Klasifikasi (*Classification*) merupakan proses untuk menemukan sekumpulan model yang menjelaskan dan membedakan 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 dihasilkan (Tang, 2005). Klasifikasi dapat digunakan untuk memprediksi nama atau nilai kelas dari suatu obyek data. Metode klasifikasi diantaranya, adalah *Artificial Neural Network* (ANN), *Support Vector Machine* (SVM), *Decision Tree*, Bayesian dan sebagainya.

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 :

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

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:

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

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

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

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 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

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

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, Naïve Bayes mudah dihitung untuk fitur bertipe kategoris. 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 $P(X_i|Y)$. Untuk setiap kelas y_j , probabilitas bersyarat kelas y_j untuk fitur X_i adalah

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

Keterangan :

μ_{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 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
$\bar{x}_{mamalia} = 38.55$ $s_{mamalia}^2 = 1960.255$ $s_{mamalia} = 44.275$	$\bar{x}_{reptil} = 4.425$ $s_{reptil}^2 = 23.6425$ $s_{reptil} = 4.8624$	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 α)**, maka data uji musang diprediksi sebagai kelas **mamalia**.

2.7 Riset – Riset 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 Naïve 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 Naïve 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%.