

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1. Koperasi**

Koperasi merupakan salah satu badan usaha yang berbadan hukum dengan usaha yang beranggotakan orang-orang yang berorientasi menghasilkan nilai tambah yang dapat dimanfaatkan bagi peningkatan kesejahteraan anggotanya. Selain itu, koperasi juga Sebagai gerakan ekonomi rakyat yang berorientasi untuk menumbuhkan partisipasi masyarakat dalam upaya memperkuat struktur perekonomian nasional dengan demokrasi ekonomi yang berdasarkan atas asas kekeluargaan. Banyak jenis koperasi yang didasarkan pada kesamaan kegiatan dan kepentingan ekonomi anggotanya seperti Koperasi simpan pinjam (Koperasi jasa keuangan), Koperasi Konsumen, Koperasi Produksi, Koperasi Pemasaran dan Koperasi Serba Usaha (Puspitasari, Khautsar, Upgris, & 2018, n.d.).

#### **2.2. Kredit**

Kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan pinjam meminjam antar Bank atau perusahaan lainnya dengan pihak lain yang mewajibkan pihak peminjam melunasi hutangnya setelah jangka waktu tertentu dengan pemberian bunga (Ivandari et al., 2017).

#### **2.3. *Data mining***

Data Mining sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. Data mining juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Istilah data mining kadang disebut juga knowledge discovery.

Data Mining merupakan proses ataupun kegiatan untuk mengumpulkan data yang berukuran besar kemudian mengekstraksi data tersebut menjadi informasi–

informasi yang nantinya dapat digunakan (Mulyani, Rihadisha, Greshanayathi, Saputri, & Wulansari, 2020).

Langkah-langkah proses pelaksanaan data mining dalam tiga aktivitas adalah :

1. Eksplorasi Data, terdiri dari aktivitas pembersihan data, transformasi data, pengurangan dimensi, pemilihan ciri, dan lain-lain.
2. Membuat model dan Pengujian Validitas Model, merupakan pemilihan terhadap model-model yang sudah dikembangkan yang cocok dengan kasus yang dihadapi. Dengan kata lain, dilakukan pemilihan model secara kompetitif.
3. Penerapan model dengan data baru untuk menghasilkan perkiraan dari kasus yang ada. Tahapan ini merupakan tahapan yang menentukan apakah model yang telah dibangun dapat menjawab permasalahan yang dihadapi.

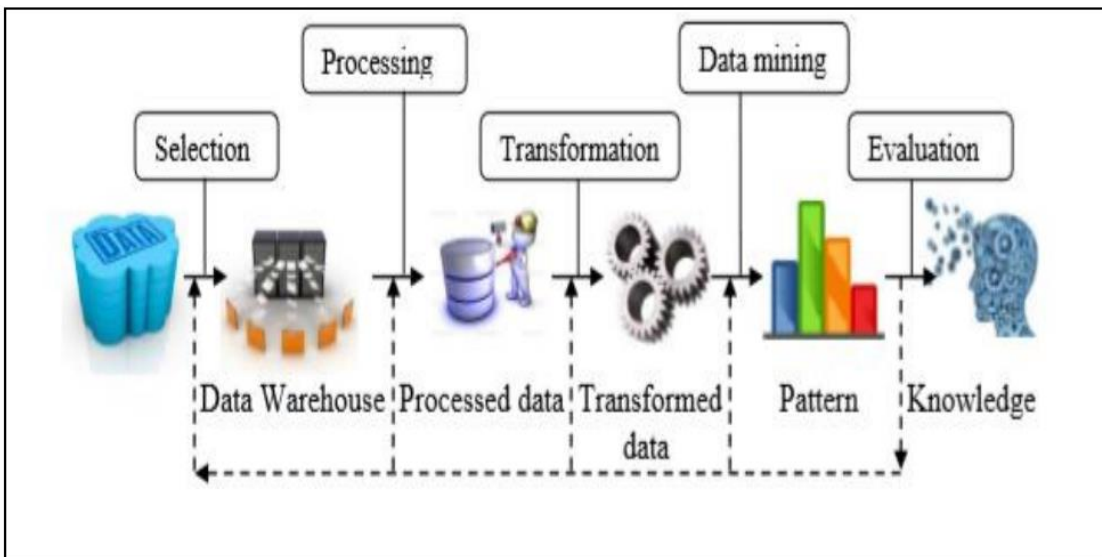
#### **2.4. Komponen-Komponen Data Mining**

Komponen-komponen utama dari proses klasifikasi adalah (Mulyani et al., 2020) :

1. Kelas, merupakan variabel tidak bebas yang merupakan label dari hasil klasifikasi.
2. Prediktor, merupakan variabel bebas suatu model berdasarkan dari karakteristik atribut data yang diklasifikasikan, misalnya merokok, minum-minuman beralkohol, tekanan darah, status perkawinan, dan sebagainya.
3. Set Data Pelatihan, merupakan sekumpulan data lengkap yang berisi kelas dan prediktor yang dilatih agar model dapat mengelompokkan ke dalam kelas yang tepat.
4. Set Data Uji, berisi data-data baru yang akan dikelompokkan oleh model guna mengetahui akurasi dari model yang telah dibuat.

## 2.5. Proses Tahapan Data Mining

Data mining merupakan salah satu dari rangkaian Knowledge Discovery in Database (KDD), KDD merupakan berhubungan dengan teknik integrasi dan penemuan ilmiah, interpretasi, dan visualisasi dari pola-pola sejumlah data. Serangkaian proses tersebut memiliki tahapan yang dapat dilihat pada Gambar 2.1 (Hadianto, Novitasari, & Rahmawati, 2019):



**Gambar 2.1** Tahapan Data Mining

Sumber : (Mohammad Rizki Ilham, 2016)

### 1. Pembersihan Data (Data Cleaning)

Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan.

### 2. Integrasi Data (Data Integration)

Integrasi data merupakan penggabungan data dari berbagai database baru.

### 3. Seleksi Data (Data Selection)

Data yang ada pada database sering kali semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database.

### 4. Transformasi Data (Data Transformasi)

Data diubah atau digabung kedalam format yang sesuai untuk diproses dalam data mining.

#### 5. Proses Mining

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data. Beberapa metode yang dapat digunakan berdasarkan pengelompokan data mining.

#### 6. Evaluasi Pola (Pattern Evaluation)

Untuk mengidentifikasi pola-pola menarik kedalam Knowledge Based yang ditentukan.

#### 7. Knowledge, yaitu sebuah hasil yang dicapai berupa pengetahuan atau sebuah informasi.

### **2.6. Fungsi Data Mining**

Menurut (Hadianto et al., 2019) fungsi data mining dapat dikelompokkan ke dalam enam kelompok yaitu:

1. Klasifikasi (classification): menggeneralisasi struktur yang diketahui untuk diaplikasikan pada data-data baru. Misalkan, klasifikasi penyakit ke dalam sejumlah jenis, klasifikasi email ke dalam spam atau bukan.
2. Klasterisasi (clustering): mengelompokkan data, yang tidak diketahui label kelasnya, ke dalam sejumlah kelompok tertentu sesuai dengan ukuran kemiripannya.
3. Regresi (regression): menemukan suatu fungsi yang memodelkan data dengan galat (kesalahan prediksi) seminimal mungkin.
4. Deteksi anomali (anomaly detection): mengidentifikasi data yang tidak umum, bisa berupa outlier, perubahan atau deviasi yang mungkin sangat penting dan perlu investigasi lebih lanjut.
5. Pembelajaran aturan asosiasi (association rule mining) atau pemodelan kebergantungan (dependency modeling): mencari relasi antar variable.
6. Perangkuman (summarization): menyediakan representasi data yang lebih sederhana, meliputi visualisasi dan pembuatan laporan.

## 2.7. Klasifikasi

Klasifikasi *adalah menentukan sebuah record data baru ke salah satu dari beberapa kategori (atau klas) yang telah didefinisikan sebelumnya.* Klasifikasi merupakan peran dalam data mining yang menggunakan metode pendekatan prediktif . (Ginting, 2019).

Secara umum proses klasifikasi dapat dilakukan dalam dua tahap, yaitu proses belajar dari data pelatihan dan klasifikasi kasus. Pada proses belajar, algoritma klasifikasi mengolah data training untuk menghasilkan sebuah model. Setelah model diuji dan dapat diterima, pada tahap klasifikasi, model tersebut digunakan untuk memprediksi kelas dari kasus baru untuk membantu proses pengambilan keputusan (Menarianti, 2015).

## 2.8. KNN (K-Nearest Neighbor)

*K-Nearest Neighbor* (K-NN) adalah suatu metode yang menggunakan algoritma *supervised* dimana hasil dari query instance yang baru diklasifikasikan berdasarkan mayoritas dari label class pada K-NN. Tujuan dari algoritma K-NN adalah mengklasifikasikan objek baru berdasarkan atribut dan training data. Algoritma K-NN bekerja berdasarkan jarak terpendek dari query instance ke training data untuk menentukan K- NN. Salah satu cara untuk menghitung jarak dekat atau jauhnya tetangga menggunakan metode *Euclidian Distance*, sering digunakan untuk menghitung jarak sehingga berfungsi menguji ukuran yang bisa digunakan sebagai interpretasi kedekatan jarak antara dua obyek (Natingga, 2017). *Euclidian Distance* dirumuskan seperti persamaan 1.

$$\sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Jika hasil nilai dari rumus di atas besar maka akan semakin jauh tingkat keserupaan antara kedua objek dan sebaliknya jika hasil nilainya semakin kecil maka akan semakin dekat tingkat keserupaan antar objek tersebut. Objek yang dimaksud adalah training data dan testing data. Dalam algoritma ini, nilai k yang terbaik itu tergantung pada jumlah data. Ukuran nilai k yang besar belum tentu menjadi nilai k yang terbaik begitupun juga sebaliknya.

Langkah-langkah untuk menghitung algoritma k-NN yaitu menentukan nilai k; menghitung kuadrat jarak euclid (*query instance*) masing-masing objek terhadap training data yang diberikan; kemudian mengurutkan objek-objek tersebut kedalam kelompok yang mempunyai jarak euclid terkecil; mengumpulkan label class Y (klasifikasi *Nearest Neighbor*); dengan menggunakan kategori *Nearest Neighbor* yang paling mayoritas maka dapat diprediksikan nilai *query instance* yang telah dihitung.

### 2.9. *Logistic Regression*

Logistic Regression adalah variasi regresi yang digunakan ketika variabel dependen bersifat biner. Model ini dapat memprediksi hasil diskrit dari satu kelompok variabel yang mungkin akan berlangsung terus menerus, kategorikal atau keduanya (Symeonidis et al., 2018). Secara garis besar algoritma logistic regression dapat dijelaskan seperti persamaan 2 berikut ini :

$$f(z) = \frac{1}{1+e^{-z}} \quad (2)$$

### 2.10. *Naive Bayes*

*Naive Bayes Classifier* adalah suatu model independen yang membahas mengenai klasifikasi sederhana berdasarkan teorema Bayes. *Naive Bayes* merupakan suatu algoritma yang dapat mengklasifikasikan suatu variable tertentu dengan menggunakan metode probabilitas dan statistik (Kurnia & Kusuma, 2020). Secara garis besar algoritma *Naive Bayes* dapat dijelaskan seperti persamaan 3 :

$$P(R|S) = \frac{P(R)P(S|R)}{P(S)} \quad (3)$$

Keterangan:

R : Data yang belum diketahui kelasnya

S : Hipotesis pada data R yang merupakan class khusus

P(R|S) : Nilai probabilitas pada hipotesis

R yang berdasarkan kondisi S

P(R) : Nilai probabilitas pada hipotesis R

$P(S|R)$  : Nilai probabilitas S yang berdasarkan dengan kondisi hipotesis R

$P(S)$  : Nilai probabilitas S

Dengan menggunakan persamaan diatas, data yang telah diperoleh dapat diproses dengan algoritma *Naive Bayes* untuk penilaian data yang akan diklasifikasikan.

### 2.11. Confusion Matrix

Untuk menguji hasil klasifikasi pada sistem yang telah dibangun, maka dibutuhkan suatu metode perhitungan evaluasi performansi yaitu dengan menghitung nilai *precision*, *recall* dan *f1-measure*. Dalam evaluasi performansi ini akan dihitung nilai akurasi dan *F1-Measure*. Akurasi adalah bagaimana suatu sistem dapat melakukan klasifikasi true pada data *true* dan *false*, sedangkan *f1-measure* untuk menilai performansi dari keseluruhan sistem dengan menghitung nilai *precision* dan *recall* (Purwanti & Handayani, 2017). Adapun perhitungan dapat dilihat pada confusion matrix :

Tabel 2.1 *Confusion Matrix*

		Nilai Aktual	
		Positive	Negative
Nilai Prediksi	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

TP (*True Positive*) merupakan prediksi positif dan nilai sebenarnya positif, TN (*True Negative*) merupakan prediksi negatif dan nilai sebenarnya negatif, FP (*False Positive*) merupakan prediksi positif dan nilai sebenarnya negatif dan FN (*False Negative*) merupakan prediksi negatif dan nilai sebenarnya positif. Perhitungan yang dihasilkan dari rumus ini terdiri dari 4 jenis keluaran, antara lain *recall*, *accuracy*, *precision*, dan *F-score*.

1. *Recall* adalah proporsi kasus positif yang diidentifikasi dengan benar, rumus *recall* dapat dilihat pada persamaan 4.

$$Recall = \frac{TP}{(TP + FN)} \quad (4)$$

2. *Accuracy* adalah perbandingan kasus yang teridentifikasi benar dengan jumlah semua kasus, rumus dari *accuracy* dapat dilihat pada persamaan 5.

$$Accuracy = \frac{(TP+TN)}{(TP + FP + FN+TN)} \quad (5)$$

3. *Precision* adalah proporsi kasus dengan hasil positif yang benar, rumus dari *Precision* dapat dilihat pada persamaan 6.

$$Precision = \frac{TP}{(TP + FP)} \quad (6)$$

4. *F-1 Score* menggambarkan perbandingan rata – rata dari *precision* dan *recall* yang dibobotkan, rumus dari *F-1 Score* dapat dilihat pada persamaan 7.

$$F - 1 \text{ score} = \frac{(2*Recall*Precision)}{(Recall+Precision)} \quad (7)$$

## 2.12 Penelitian Terdahulu

Dalam melakukan penelitian ini terdapat penelitian yang telah dilakukan sebelumnya yang menjadi bahan referensi bagi peneliti. Berikut adalah penelitian terdahulu yang digunakan sebagai referensi, dapat dilihat pada Tabel 2.2.

**Tabel 2.2** Penelitian Terdahulu

NO	Pengarang	Judul	Metode	Kesimpulan
1	Hendra Marcos, Indriana Hidayah	Implementasi Data Mining Untuk Klasifikasi Nasabah Kredit Bank "X" Menggunakan Classification Rule	Classification Rule	akurasi yang didapat dari algoritma J48 sebesar 89,18%
2	Diky Alfian Kurniawan, Yogiek Indra Kurniawan	Aplikasi Prediksi Kelayakan Calon Anggota Kredit	Naive Bayes	Aplikasi ini memiliki nilai precision sebesar



		Menggunakan Algoritma Naïve Bayes		82%, nilai accuracy sebesar 80% dan nilai recall sebesar 94%. Maka aplikasi yang telah dibangun menggunakan algoritma Naïve Bayes telah dapat digunakan.
--	--	-----------------------------------	--	--

**Tabel 2.2** Penelitian Terdahulu (lanjutan)

NO	Pengarang	Judul	Metode	Kesimpulan
3	Ridawati Manik, Pristiwanto, Kennedy Tampubolon	Prediksi Kolektibilitas Kredit Anggota Dengan Algoritma C5.0 (Studi Kasus: CU Damai Sejahtera Medan)	C5.0	Dari hasil penelitian membuktikan bahwa algoritma C5.0 membantu untuk membuat suatu keputusan dalam mengidentifikasi suatu permasalahan dan melihat hubungan antara faktor-faktor yang mempengaruhi suatu masalah. Peranan algoritma C5.0 dalam membangun pohon keputusan bisa dijadikan acuan sebagai alat bantu alternatif pemecahan masalah
4	Putri Kurnia Handayani	Model Klasifikasi Kelayakan Kredit Koperasi Karyawan Dengan Algoritma C4.5	C4.5	Decision Tree akurat untuk klasifikasi kelayakan kredit koperasi karyawan. Tingkat akurasi Decision Tree untuk data dari sebuah koperasi karyawan sebesar 92,28%.