

## 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 assa 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 penghasilannya 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 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 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 pengelompokan 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 Dcision Tree

### 2.5.1 Pengertian Decision Tree

*Decision tree* merupakan metode klasifikasi *data mining*. *Decision tree* dalam istilah pembelajaran merupakan sebuah struktur pohon dimana setiap *node* pohon mempresentasikan atribut yang telah diuji. Setiap cabang merupakan suatu pembagian hasil uji dan *node* daun (*leaf*) mempresentasikan kelompok kelas tertentu [JIA01]. Level *node* teratas dari sebuah *Decision Tree* adalah node akar (*root*) yang biasanya berupa atribut yang paling memiliki pengaruh terbesar pada suatu kelas tertentu. Pada umumnya *Decision Tree* melakukan strategi pencarian secara *top-down* untuk solusinya. Pada proses mengklasifikasi data yang tidak diketahui, nilai atribut akan diuji dengan cara melacak jalur dari *node* akar (*root*) sampai *node* akhir



(daun) dan kemudian akan diprediksi kelas yang dimiliki oleh suatu data baru tertentu.

### 2.5.2 Jenis – Jenis Decision Tree

Beberapa model *decision tree* yang sudah dikembangkan antara lain C4.5 atau ID3 dan CART. Berikut ini akan dijelaskan model dari decision tree tersebut :

#### 1. C4.5 atau ID3

*Decision Tree* menggunakan algoritma ID3 atau C4.5, yang diperkenalkan dan dikembangkan pertama kali oleh Quinlan yang merupakan singkatan dari *Iterative Dichotomiser 3* atau *Induction of Decision 3*. Algoritma ID3 membentuk pohon keputusan dengan metode *divide-and-conquer* data secara rekursif dari atas ke bawah. Strategi pembentukan Decision Tree dengan algoritma ID3 adalah:

- A. Pohon dimulai sebagai *node* tunggal (akar/*root*) yang merepresentasikan semua data.
- B. Sesudah *node root* dibentuk, maka data pada *node* akar akan diukur dengan *information gain* untuk dipilih atribut mana yang akan dijadikan atribut pembaginya.
- C. Sebuah cabang dibentuk dari atribut yang dipilih menjadi pembagi dan data akan didistribusikan ke dalam cabang masing-masing.
- D. Algoritma ini akan terus menggunakan proses yang sama atau bersifat rekursif untuk dapat membentuk sebuah *Decision Tree*. Ketika sebuah atribut telah dipilih menjadi *node* pembagi atau cabang, maka atribut tersebut tidak diikuti lagi dalam penghitungan nilai *information gain*.
- E. Proses pembagian rekursif akan berhenti jika salah satu dari kondisi dibawah ini terpenuhi :
  - a. Semua data dari anak cabang telah termasuk dalam kelas yang sama.

- b. Semua atribut telah dipakai, tetapi masih tersisa data dalam kelas yang berbeda. Dalam kasus ini, diambil data yang mewakili kelas yang terbanyak untuk menjadi label kelas pada *node* daun. Tidak terdapat data pada anak cabang yang baru. Dalam kasus ini, *node* daun akan dipilih pada cabang sebelumnya dan diambil data yang mewakili kelas terbanyak untuk dijadikan label kelas.

Metode C4.5 dan ID3 memiliki perbedaan dalam nilai tiap atribut. Metode C4.5 menggunakan atribut yang bernilai kategorikal dan numerikal, sedangkan metode ID3 menggunakan atribut yang bernilai kategorikal. Metode *decision tree C4.5* inilah yang digunakan dalam tugas akhir ini.

## 2. CART

CART adalah singkatan dari *Classification And Regression Tree*. Dalam CART ada dua langkah penting yang harus diikuti untuk mendapatkan *tree* dengan performansi yang optimal. Yang pertama adalah pemecahan objek secara berulang berdasarkan atribut tertentu. Yang kedua, *prunning* (pemangkasan) dengan menggunakan data validasi.

Misalkan kita mempunyai variabel independent  $x_1, x_2, x_3, \dots, x_n$  dan variabel dependent atau output  $y$ . Pemecahan secara berulang berarti kita bagi objek ke dalam kotak-kotak berdasarkan nilai variabel  $x_1, x_2$  atau  $x_r$ . Cara ini diulang sehingga dalam suatu kotak sebisa mungkin berisi observasi dalam kelompok atau kelas yang sama.

Langkah berikutnya sesudah dilakukan pemecahan objek atau data secara berulang adalah melakukan *prunning*. Dalam *prunning* kita ingin memangkas *tree* yang mungkin terlalu besar dan terjadi fenomena *overfitting*. *Overfitting* merupakan sebuah satu buah pengelompokan yang mungkin hanya berisi satu data yang memungkinkan data tersebut merupakan *noise* yang ada di data

training dan bukan pola yang mungkin terjadi dalam data testing atau data validasi. *Prunning* terdiri dari beberapa langkah pemilihan secara berulang simpul yang akan dijadikan simpul daun. Dengan mengubah simpul menjadi simpul daun artinya tidak akan dilakukan pemecahan lagi sesudah itu. Dengan demikian ukuran *tree* akan berkurang. [TUR06]

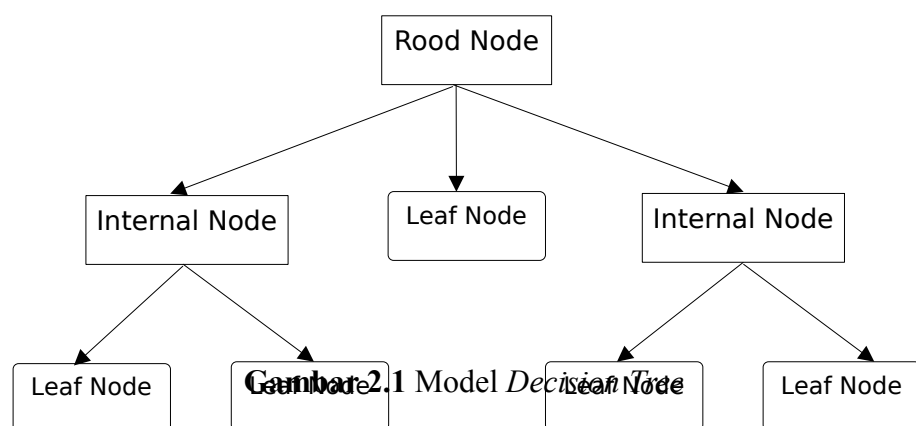
### 2.5.3 Model Decision Tree

*Decision tree* adalah *flow-chart* seperti *struktur tree*, dimana tiap *internal node* menunjukkan sebuah test pada sebuah atribut, tiap cabang menunjukkan hasil dari test, dan *leaf node* menunjukkan *class-class* atau *class distribution*.

Selain karena pembangunannya relatif cepat, hasil dari model yang dibangun mudah untuk dipahami. Pada *decision tree* terdapat 3 jenis *node*, yaitu:

- Root Node*, merupakan *node* paling atas, pada *node* ini tidak ada *input* dan bisa tidak mempunyai *output* atau mempunyai *output* lebih dari satu.
- Internal Node*, merupakan *node* percabangan, pada *node* ini hanya terdapat satu *input* dan mempunyai *output* minimal dua.
- Leaf node* atau *terminal node*, merupakan *node* akhir, pada *node* ini hanya terdapat satu *input* dan tidak mempunyai *output*.

Contoh dari model pohon keputusan yaitu seperti pada **gambar 2.1** berikut:



## 2.6 Algoritma Decision Tree C4.5

Algoritma C4.5 diperkenalkan oleh Quinlan pada tahun 1996 sebagai versi perbaikan dari ID3. Dalam ID3, induksi decision tree hanya bisa dilakukan pada fitur bertipe kategorikal (nominal atau ordinal), sedangkan tipe numerik (interval atau rasio) tidak dapat digunakan.

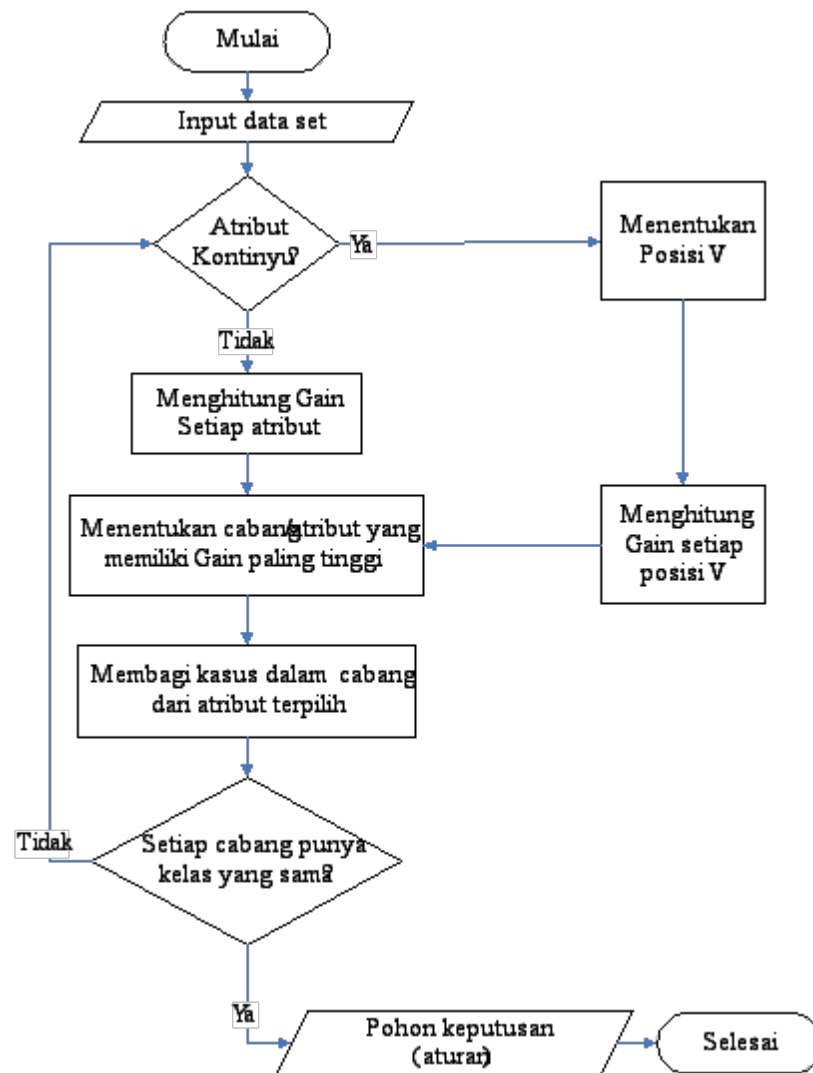
Yang menjadi hal penting dalam induksi decision tree adalah bagaimana menyatakan syarat pengujian pada node. Ada 3 kelompok penting dalam syarat pengujian node :

1. Fitur biner  
Adalah Fitur yang hanya mempunyai dua nilai berbeda. Syarat pengujian ketika fitur ini menjadi node (akar maupun interval) hanya punya dua pilihan cabang.
2. Fitur kategorikal  
Untuk fitur yang nilainya bertipe kategorikal (nominal atau ordinal) bisa mempunyai beberapa nilai berbeda. Secara umum ada 2 pemecahan yaitu pemecahan biner (*binary splitting*) dan (*multi splitting*).
3. Fitur numerik  
Untuk fitur bertipe numerik, Syarat pengujian dalam node (akar maupun internal) dinyatakan dengan pengujian perbandingan ( $A \leq V$ ) atau ( $A > V$ ) dengan hasil biner.

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

1. Pilih atribut sebagai akar.
2. Buat cabang untuk tiap-tiap nilai.
3. Bagi kasus dalam cabang.
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Berikut ini akan dijelaskan secara lebih detail algoritma C4.5 menggunakan *flowchart* yang disajikan pada **gambar 2.2**.



**Gambar 2.2** Flowchart algoritma Decision Tree C4.5

Untuk memilih atribut sebagai simpul akar (*root node*) atau simpul dalam (*internal node*), didasarkan pada nilai *information gain* tertinggi dari atribut-atribut yang ada. Sebelum perhitungan *information gain*, akan dilakukan perhitungan *entropy*. *Entropy* merupakan distribusi probabilitas dalam teori informasi dan diadopsi kedalam algoritma C4.5 untuk mengukur tingkat homogenitas distribusi kelas dari sebuah himpunan data (*data set*). Semakin tinggi tingkat *entropy* dari sebuah data maka semakin homogen distribusi kelas pada data tersebut. Perhitungan *information gain* menggunakan rumus 2.1, sedangkan *entropy* menggunakan rumus 2.2.

$$\frac{|S_i|}{|S|} * Entropy(i) \dots\dots\dots(2.1)$$

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(i)$$

dimana,

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

|Si| : Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

$$Entropy(S) = - \sum_{i=1}^n p_i * \log_2 p_i \dots\dots\dots(2.2)$$

dimana,

S : Himpunan kasus

n : Jumlah partisi S

pi : Proporsi dari Si terhadap S

Selain *Information Gain* kriteria yang lain untuk memilih atribut sebagai pemecah adalah *Rasio Gain*. Perhitungan rasio gain menggunakan rumus 2.3, sedangkan split information menggunakan rumus 2.4.

$$GainRasio(S, A) = \frac{Gain(S, A)}{SplitInformation(S, A)} \dots\dots\dots(2.3)$$

$$SplitInformation(S, A) = - \sum_{i=1}^c \frac{S_i}{S} \log_2 \frac{S_i}{S} \dots\dots\dots(2.4)$$

dimana S1 sampai Sc adalah c subset yang dihasilkan dari pemecahan S dengan menggunakan atribut A yang mempunyai sebanyak c nilai.

## 2.7 Riset – Riset Terkait

Penelitian sebelumnya yang menggunakan metode Naive Bayes dilakukan oleh Hesty Maharani Pratiwi yang berjudul *“Prediksi untuk Menentukan Kelancaran Pembayaran pada Koperasi Simpan Pinjam Menggunakan Metode Naive Bayes (Studi Kasus : Koperasi Wanita “Cempaka” Ds. Plosobuden)”*. Variabel yang digunakan sebagai adalah Jenis Usia, Status, Jumlah Anak, Pekerjaan, Gaji, Pengeluaran, dan Status Rumah. Adapun data yang diambil dalam penelitian ini adalah sampel dari 78 record dengan kelas “Lancar” dan kelas “Macet” masing-masing berjumlah 26 untuk kelas lancar dan 52 untuk kelas macet. Menghasilkan status pembayaran calon anggota yang termasuk macet atau lancar yang nantinya pihak koperasi dapat memberikan penanganan lebih awal/dini.

Penelitian selanjutnya adalah Rezki Badriza *“Implementasi Data Mining Algoritma C4.5 Untuk Memprediksi Pembayaran Pinjaman Pada Koperasi Simpan Pinjam Primkoveri Bina Bakti Pemalang”*. Primkoveri Bina Bakti Comal Pemalang Merupakan salah satu koperasi simpan pinjam yang sukses menarik banyak nasabah tiap tahunnya, terbukti dengan meningkatnya jumlah nasabah baru. Dengan adanya peningkatan ini pihak koperasi juga harus meningkatkan kualitas dan kenyamanan bagi nasabah dalam melakukan transaksi. Menghasilkan sebuah sistem yang dapat membantu pihak Koperasi simpan pinjam (Primkoveri Bina Bakti) dalam menentukan perilaku pembayaran nasabah. Variabel yang digunakan Jenis Penghasilan, Pinjaman, Status Pernikahan, dan Status Rumah.

Penelitian selanjutnya adalah Awaludin Rizky Wicaksono dan Yuniarsi Rahayu *“Klasifikasi Data Mining Untuk Menentukan Potensi Kredit Macet Pada Koperasi Simpan Pinjam Primkoveri Waleri Menggunakan Algoritma Decision Tree C4.5”*. Analisis klasifikasi prediksi nasabah yang berpotensi kredit macet yang ada di Koperasi Primkoveri masih dilakukan secara sederhana dikarenakan keterbatasan koperasi dalam melibatkan tenaga analis kredit yang membutuhkan biaya cukup mahal dan seringkali memberatkan nasabah, kesederhanaan analisis inilah yang memicu terjadinya kredit macet.

Antisipasi yang biasa dilakukan hanyalah dengan melakukan pendekatan-pendekatan personal pada semua nasabah kredit seharusnya suatu hasil analisis yang baik adalah yang mampu menghasilkan data yang akurat dan tepat dalam proses pembuatannya. Penulis memilih untuk melakukan penghitungan analisis klasifikasi prediksi nasabah yang berpotensi kredit macet dengan cara lebih efektif yang dapat digunakan untuk menggantikan cara sederhana yaitu dengan menggunakan data mining. Parameter yang digunakan untuk menentukan potensi kredit macet adalah Umur, Status, Pekerjaan, Penghasilan, Maks Kredit, Jangka Waktu Kredit dan Angsuran.