

**ANALISIS ALGORITMA C4.5 UNTUK  
MENENTUKAN TINGKAT PROFIT USAHA  
PADA PT MEGACIPTA SEJATI**

**SKRIPSI**



**Oleh:  
Harina  
161510006**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS TEKNIK DAN KOMPUTER  
UNIVERSITAS PUTERA BATAM  
2021**

**ANALISIS ALGORITMA C4.5 UNTUK  
MENENTUKAN TINGKAT PROFIT USAHA  
PADA PT MEGACIPTA SEJATI**

**SKRIPSI**

**Untuk memenuhi salah satu syarat  
guna memperoleh gelar Sarjana**



**Oleh:  
Harina  
161510006**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS TEKNIK DAN KOMPUTER  
UNIVERSITAS PUTERA BATAM  
2021**

## SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini saya:

Nama : Harina  
NPM/NIP : 161510006  
Fakultas : Teknik dan Komputer  
Program Studi : Sistem Informasi

Menyatakan bahwa “**Skripsi**” yang saya buat dengan judul:

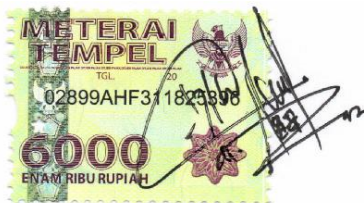
### **Analisis Algoritma C4.5 Untuk Menentukan Tingkat Profit Usaha Pada PT Megacipta Sejati**

Adalah hasil karya sendiri dan bukan “duplikasi” dari karya orang lain. Sepengetahuan saya, didalam naskah Skripsi ini tidak terdapat karya ilmiah atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis dikutip didalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata di dalam naskah Skripsi ini dapat dibuktikan terdapat unsur-unsur PLAGIASI, saya bersedia naskah Skripsi ini digugurkan dan gelar akademik yang saya peroleh dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya tanpa ada paksaan dari siapapun.

Batam, 20 Januari 2021



**Harina**  
161510006

**ANALISIS ALGORITMA C4.5 UNTUK  
MENENTUKAN TINGKAT PROFIT USAHA PADA PT  
MEGACIPTA SEJATI**

Oleh  
**Harina**  
161510006

**SKRIPSI**

Untuk memenuhi salah satu syarat  
guna memperoleh gelar Sarjana

Telah disetujui oleh Pembimbing pada tanggal  
seperti tertera di bawah ini

**Batam, 20 Januari 2021**



**Rika Harman, S.Kom., M.SI.**  
Pembimbing



---

**Universitas Putera Batam**

## ABSTRAK

Perkembangan asrama di Batam dari tahun ke tahun menunjukkan peningkatan yang signifikan. Banyak orang yang datang dari kota yang jauh dan banyak juga orang yang meninggalkan kampung halaman untuk datang dan mencari kerja di kota Batam sehingga tidak memiliki tempat tinggal membuat banyak perusahaan yang sedang mencari pekerja memberikan fasilitas tempat tinggal seperti asrama untuk para pekerja. PT Megacipta Sejati merupakan perusahaan yang bergerak dibidang sewa *dormitory* kepada perusahaan yang memberikan fasilitas tempat tinggal seperti asrama untuk para pekerjanya. Selama ini perusahaan memiliki sekumpulan data terkait pencapaian kinerja perusahaan yang telah menghabiskan banyak biaya namun data tersebut belum dapat dimanfaatkan dengan baik. Melalui penerapan teknologi data mining, sekumpulan data yang belum dimanfaatkan tersebut akan diproses dan menghasilkan new knowlage (pengetahuan baru) yang menjadi sesuatu yang berharga dan berguna bagi perusahaan terutama dalam peningkatan *profit*. Berdasarkan permasalahan tersebut, saya sebagai peneliti tertarik untuk meneliti analisa algoritma c4.5 untuk menentukan tingkat *profit* usaha pada PT Megacipta Sejati. Data Mining merupakan penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data dalam jumlah besar yang diharapkan dapat mengatasi kondisi tersebut. Data Mining sendiri memiliki beberapa teknik salah satunya klasifikasi. Teknik klasifikasi terdiri beberapa metode, dan *decision tree* adalah bagian dari metode klasifikasi. Kemudian metode *decision tree* memiliki algoritma, algoritma C4.5 adalah salah satu dari algoritma yang memiliki *decision tree*. Hasil dari penelitian ini akan menghasilkan rule-rule keputusan *profit* dan kerugian perusahaan.

**Kata Kunci:** Algoritma C4.5, *Data Mining*, Pohon Keputusan, *Profit*.

## **ABSTRACT**

*The development of boarding houses in Batam from year to year shows a significant increase. Many people come from distant cities and many people leave their hometowns to come and look for work in Batam, so they do not have a place to live. Many companies are looking for workers to provide housing facilities such as dormitories for workers. PT Megacipta Sejati is a company engaged in the rental of dormitory to companies that provide housing facilities such as dormitories for their workers. During this time the company has a set of data related to the achievement of company performance which has cost a lot but the data cannot be utilized properly. Through the application of data mining technology, a collection of untapped data will be processed and produce new knowledge that becomes valuable and useful for companies, especially in increasing profits. Based on these problems, I as a researcher are interested in researching the c4.5 algorithm analysis to determine the level of business profit at PT Megacipta Sejati. Data Mining is mining or finding new information by looking for certain patterns or rules from a large amount of data that is expected to overcome these conditions. Data Mining itself has several techniques, one of which is classification. The classification technique consists of several methods, and the decision tree is part of the classification method. Then the decision tree method has an algorithm, C4.5 algorithm is one of the algorithms that has a decision tree. The results of this study will produce rules for the company's profit and loss decisions.*

**Keywords:** Algorithm C4.5, Data Mining, Decision Tree, Profit.

## KATA PENGANTAR

Dengan mengucapkan puji dan syukur kepada Tuhan yang Maha Esa yang telah melimpahkan segala rahmat dan karunianya, sehingga penulis dapat menyelesaikan skripsi yang merupakan salah satu persyaratan untuk menyelesaikan program studi strata satu (S1) pada Program Studi Sistem Informasi Universitas Putera Batam.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Karena itu, kritik dan saran akan senantiasa penulis terima dengan senang hati.

Dengan segala keterbatasan, penulis menyadari pula bahwa skripsi ini takkan terwujud tanpa bantuan, bimbingan, dan dorongan dari berbagai pihak. Untuk itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terima kasih kepada:

1. Ibu Dr. Nur Elfi Husda, S.Kom., M.SI, selaku Rektor Universitas Putera Batam.
2. Bapak Welly Sugianto, S.T., M.M, selaku Dekan Fakultas Teknik Dan Komputer Universitas Putera Batam.
3. Bapak Muhammad Rasyid Ridho, S.Kom., M.SI., selaku Ketua Program Studi Sistem Informasi Universitas Putera Batam.
4. Bapak Rika Harman, S.Kom., M.SI., sebagai pembimbing skripsi yang telah menyediakan waktu dalam memberikan bimbingan dan petunjuk sampai selesainya skripsi ini.
5. Seluruh dosen dan staff Universitas Putera Batam.
6. Dan pihak-pihak yang telah memberikan dukungannya hingga tersusunnya penelitian ini yang tidak dapat disebutkan oleh penulis satu persatu.

Semoga Tuhan YME membalas kebaikan dan selalu mencurahkan hidayah serta taufik-Nya, Amin.

Batam, 20 Januari 2021

Penulis





---

**Universitas Putera Batam**

# DAFTAR ISI

## Halaman

HALAMAN SAMPUL	
HALAMAN JUDUL	
HALAMAN PERNYATAAN ORISINALITAS	
HALAMAN PENGESAHAN	
ABSTRAK .....	v
<i>ABSTRACT</i> .....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR .....	x
DAFTAR TABEL.....	xi
DAFTAR RUMUS .....	xii
<b>BAB I PENDAHULUAN</b>	
1.1 Latar Belakang Penelitian .....	1
1.2 Identifikasi Masalah.....	6
1.3 Pembatasan Masalah .....	6
1.4 Perumusan Masalah .....	7
1.5 Tujuan Penelitian .....	8
1.6 Manfaat Penelitian .....	8
1.6.1 Manfaat Teoritis.....	9
1.6.2 Manfaat Praktis .....	9
<b>BAB II TINJAUAN PUSTAKA</b>	
2.1 Teori.....	11
2.1.1 Teori Umum.....	11
2.1.1.1 <i>Knowledge Discovery in Database (KDD)</i> .....	11
2.1.1.2 <i>Data Mining</i> .....	14
2.1.1.3 Pengelompokan <i>Data Mining</i> .....	16
2.1.1.4 Klasifikasi .....	20
2.1.1.5 Pohon Keputusan .....	22
2.1.1.6 Algoritma C4.5 .....	25
2.1.2 Teori Khusus.....	29
2.1.2.1 Profit .....	29
2.1.2.2 <i>Waikato Environment for Knowledge Analysis (WEKA)</i> .....	30
2.2 Kerangka Berpikir.....	31
2.3 Hipotesis Penelitian .....	32
2.4 Penelitian Terdahulu .....	33
<b>BAB III METODE PENELITIAN</b>	
3.1 Desain Penelitian .....	36
3.2 Objek Penelitian.....	38

3.2.1	Lokasi Penelitian.....	38
3.2.2	Jadwal Penelitian .....	38
3.3	Populasi dan Sampel .....	39
3.3.1	Populasi.....	39
3.3.2	Sampel.....	39
3.4	Variabel Penelitian.....	40
3.5	Teknik Pengumpulan Data.....	41
3.6	Model Penelitian .....	42
<b>BAB IV HASIL DAN PEMBAHASAN</b>		
4.1	Hasil .....	43
4.2	Pembahasan.....	45
4.2.1	Melakukan Pra Proses Data Penelitian .....	46
4.2.2	Analisa Data.....	48
4.2.2.1	Pembersihan Data .....	48
4.2.2.2	Seleksi Data .....	48
4.2.2.3	Transformasi Data.....	48
4.2.3	Pohon Keputusan .....	52
4.2.4	Keputusan Hasil Pengujian .....	76
4.2.5	Pengujian Menggunakan <i>WEKA</i> .....	77
<b>BAB V SIMPULAN DAN SARAN</b>		
5.1	Simpulan .....	86
5.2	Saran .....	88
DAFTAR PUSTAKA .....		90
LAMPIRAN		

## DAFTAR GAMBAR

	<b>Halaman</b>
<b>Gambar 1.1</b> Data Penghuni.....	3
<b>Gambar 2.1</b> Proses Dalam KDD atau Knowledge Discovery in Database.....	15
<b>Gambar 2.2</b> Bidang Ilmu Data Mining.....	16
<b>Gambar 2.3</b> Pengelompokan Teknik Data Mining.....	20
<b>Gambar 2.4</b> Pengelompokan Teknik Klasifikasi.....	22
<b>Gambar 2.5</b> Pohon Keputusan Identifikasi.....	25
<b>Gambar 2.6</b> Tampilan Utama <i>WEKA</i> .....	31
<b>Gambar 2.7</b> Kerangka Berpikir.....	33
<b>Gambar 3.1</b> Desain Penelitian.....	36
<b>Gambar 4.1</b> Pohon Keputusan <i>Node 1</i> .....	61
<b>Gambar 4.2</b> Pohon Keputusan <i>Node 2</i> .....	69
<b>Gambar 4.3</b> Pohon Keputusan <i>Node 3</i> .....	76
<b>Gambar 4.4</b> Data Variabel Pada Aplikasi Microsoft Excel.....	77
<b>Gambar 4.5</b> Data Variabel Pada Aplikasi Notepad.....	78
<b>Gambar 4.6</b> Tampilan Awal Aplikasi <i>WEKA</i> .....	79
<b>Gambar 4.7</b> Tampilan Pemilihan <i>File</i> Pengujian <i>WEKA</i> .....	80
<b>Gambar 4.8</b> Tampilan Pemilihan Atribut Pengujian <i>WEKA</i> .....	80
<b>Gambar 4.9</b> Tampilan <i>Tab Menu Classify</i> .....	81
<b>Gambar 4.10</b> Tampilan Pemilihan <i>Classifier</i> .....	82
<b>Gambar 4.11</b> Tampilan Pemilihan <i>Classifer</i> .....	83
<b>Gambar 4.12</b> Tampilan Visualisasi Pohon.....	84
<b>Gambar 4.13</b> Tampilan Hasil Pengujian <i>WEKA</i> .....	84

## DAFTAR TABEL

	<b>Halaman</b>
<b>Tabel 3.1</b> Jadwal Penelitian.....	38
<b>Tabel 4.1</b> <i>Format</i> Pra Proses Data Profit Usaha.....	47
<b>Tabel 4.2</b> Biaya Tenaga Kerja.....	48
<b>Tabel 4.3</b> Biaya Operasional.....	49
<b>Tabel 4.4</b> Biaya Perlengkapan.....	49
<b>Tabel 4.5</b> Biaya Perawatan.....	49
<b>Tabel 4.6</b> Biaya Pajak.....	50
<b>Tabel 4.7</b> Biaya Legalitas.....	50
<b>Tabel 4.8</b> <i>Format</i> Data Akhir Profit Usaha.....	51
<b>Tabel 4.9</b> Hasil Perhitungan Manual <i>Node 1</i> .....	60
<b>Tabel 4.10</b> Hasil Perhitungan Manual <i>Node 2</i> .....	68
<b>Tabel 4.11</b> Hasil Perhitungan Manual <i>Node 3</i> .....	75

## DAFTAR RUMUS

	Halaman
<b>Rumus 2.1</b> Perhitungan <i>Gain</i> .....	26
<b>Rumus 2.2</b> Rumus Perhitungan <i>Entropy</i> .....	27



---

**Universitas Putera Batam**

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Perubahan demi perubahan terjadi terutama dalam dunia bisnis seiring dengan pesatnya pertumbuhan teknologi dan informasi. Teknologi yang semakin berkembang, menjadi sorotan dan memikat animo banyak orang dari berbagai tempat dan wilayah. Salah satu bentuk perkembangan teknologi ialah sebuah perangkat keras, yakni komputer. Komputer berevolusi dengan beradaptasi seiring bergantinya tahun demi tahun yang mampu mewujudkan penyampaian informasi secara ekspres sehingga menghemat waktu banyak orang guna mendapat informasi yang diperlukan. Hal ini juga dialami oleh usaha-usaha yang menggeluti dunia sewa menyewa asrama atau *dormitory*, yang mana untuk mengembangkan usaha ini membutuhkan proses memperhitungkan sebuah rencana secara akurat bahkan secara berkala dari tahun ke tahun.

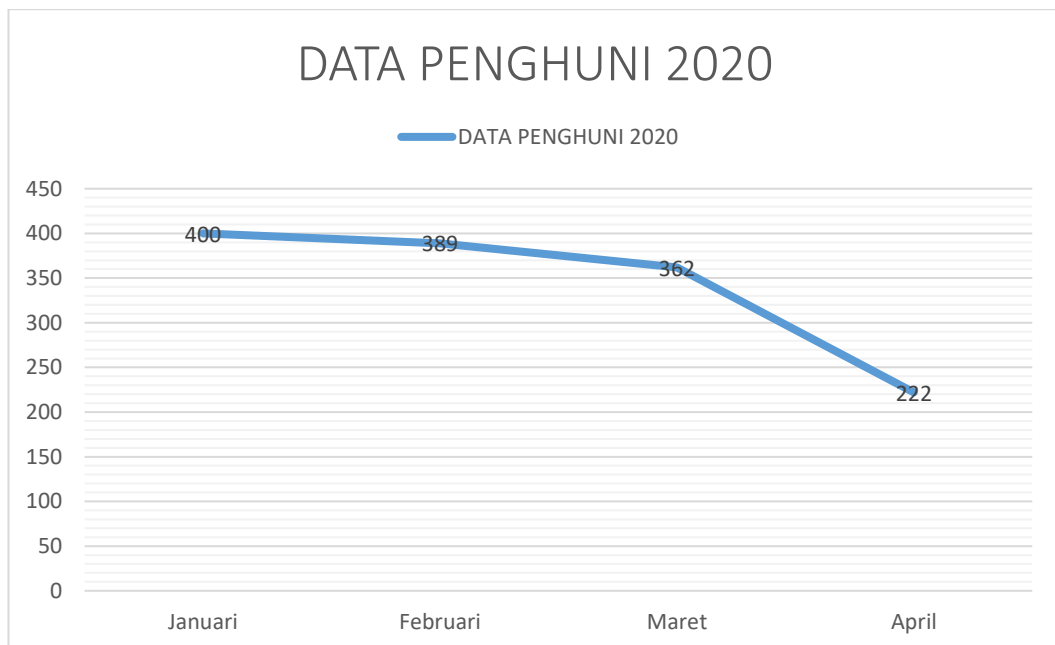
Perkembangan asrama di Batam dari waktu ke waktu memperlihatkan eskalasi yang cukup tinggi. Kota Batam sebagai kota industri terdapat banyak kawasan industri, ini menjadi perhatian banyak orang yang berada diluar kota batam untuk datang mencari pekerjaan dikota ini. Banyak orang yang datang dari kota yang jauh dan banyak juga orang yang meninggalkan kampung halaman untuk datang dan mencari kerja di kota batam sehingga tidak memiliki tempat tinggal membuat banyak perusahaan yang sedang mencari pekerja memberikan fasilitas tempat tinggal seperti asrama untuk para pekerja. Asrama dapat diartikan



sebagai sebuah gedung yang dijadikan sebagai tempat berdomisili beberapa orang dengan latar belakang yang sejenis (Departemen Pendidikan dan Kebudayaan, 1993, 53). Seperti yang terpublikasi dalam situs wikipedia.org (2009), mendefinisikan asrama sebagai sebuah gedung yang difungsikan untuk menginap personel dari suatu komunitas, lazimnya siswa dari suatu sekolah. Asrama dapat dikatakan sebagai suatu gedung yang terbagi-bagi ke dalam beberapa kamar dan dipenuhi oleh orang-orang untuk ditinggali. Biasanya orang-orang tersebut bertempat tinggal dengan durasi yang relatif lama sesuai dengan keperluannya masing-masing. Ada berbagai alasan mengapa seseorang memilih tinggal di asrama dilihat dari segi harga dan lokasi dibanding dengan memilih tempat lainnya.

Menurut *The Encyclopedia Americana*, asrama berasal dari kata *Dormitory*, ialah dari dasar kata *Dormotorius* (Latin), yang artinya *a sleeping place*, dapat didefinisikan sebagai gedung yang digunakan seluruhnya digunakan untuk keperluan edukasi dimana tata letak ruangnya terdiri atas kamar tidur dan kamar belajar. *Dormitory* dalam perusahaan atau organisasi lebih dikenal dengan sebutan mess yaitu kamar yang dibangun untuk tempat tinggal karyawan dan pegawai. *Dormitory* yang disediakan oleh entitas lazimnya mempunyai ruang kamar yang relatif banyak dengan daya tampung yang besar. *Dormitory* pada perusahaan disediakan dengan porsi tertentu yakni diberi dengan cuma-cuma atau bisa termasuk dalam kesepakatan pekerjaan yang telah ditandatangani. Ada juga *Dormitory* yang memang dengan sengaja disewakan kepada karyawan selama karyawan tersebut masih berkerja dengan perusahaan tersebut.

PT Megacipta Sejati merupakan perusahaan yang bergerak dibidang sewa *dormitory* kepada perusahaan yang memberikan fasilitas tempat tinggal seperti asrama untuk para pekerjanya sejak tahun 2012 yang beralamat di Jalan Hang Kesturi belakang Cammo Industrial Estate Blok B, Batam, Kepulauan Riau. Dalam 1 bulan PT Megacipta Sejati memiliki penghuni sekitar 300 penghuni sampai dengan 400 penghuni yang tinggal di dormitory. Demikian grafik data penghuni perusahaan PT Megacipta Sejati di Kota Batam, sebagai berikut:



**Gambar 1.1** Data Penghuni

**Sumber:** (PT Megacipta Sejati, 2020)

Berdasarkan gambar dari grafik data penghuni yang tinggal di PT Megacipta Sejati Tahun 2020, pada bulan Januari hingga Februari mengalami penurunan yang tidak cukup besar yaitu sekitar 11 penghuni. Kemudian pada bulan Maret ke April terjadi penurunan penghuni yang cukup besar yaitu sekitar

140 penghuni namun dapat disimpulkan dalam tahun 2020 sejak bulan Januari hingga bulan April PT Megacipta Sejati rata-rata mengalami penurunan dalam penghuni setiap bulan pada tahun 2020.

Kesalahan yang dilakukan oleh perusahaan ialah tidak memanfaatkan data yang berisi pembiayaan yang dimilikinya dalam rangka mencapai sasaran-sasarannya. Data yang tidak dimanfaatkan tersebut dapat diolah kembali dengan menggunakan teknologi data mining untuk menciptakan *new knowledge* (pengetahuan baru) yang lebih berfaedah sehingga data tersebut mampu dijadikan sebagai bahan acuan guna meningkatkan laba.

Melalui beberapa penjabaran permasalahan yang telah diungkapkan pada bagian sebelumnya, data tidak diperhitungkan atau dapat dikatakan tidak bernilai oleh perusahaan mampu diolah kembali menjadi data yang bernilai menggunakan penerapan teknologi data mining dengan algoritma C4.5. Data Mining diartikan sebagai penggalian atau invensi dari embaran yang sebelumnya tidak diperhitungkan menjadi diperhitungkan dengan memeriksa arketipe atau motif berasal dari generalisasi data yang berjumlah relatif besar guna memecahkan sebuah permasalahan dalam situasi tertentu. Klasifikasi menjadi satu dari banyaknya teknik yang mampu diterapkan pada perealisasi Data Mining. Teknik klasifikasi itu sendiri meliputi metode tetapi yang paling sering digunakan ialah *decision tree*. *Decision tree* merupakan metode yang mengandalkan algoritma dan yang dipakai pada penelitian ini ialah algoritma C4.5.

Pada penerapannya didunia usaha, seperti yang telah diteliti oleh (Nofitri, 2017) mengenai analisa algoritma C4.5 dalam memprediksi pencapaian profit menyatakan bahwa data mining dapat dikatakan sebagai metode penganalisan sekumpulan data guna mendapatkan sebuah corak tertentu dari data yang telah dianalisis. Data mining terjadi kedalam sebuah embaran yang membentuk corak tertentu guna dijadikan bahan pertimbangan dalam sebuah pengambilan keputusan dari proses analisa data berukuran besar. Sedangkan, pada penelitian yang pernah dilakukan oleh (Elisa, 2018) yang meneliti tentang prediksi profit pada perusahaan dengan klasifikasi algoritma C4.5 mengatakan bahwa algoritma C4.5 merupakan suatu algoritma yang menjadi unsur pembangun sebuah *decision tree* (pengambilan keputusan), memperlihatkan berbagai hal yang memberikan pengaruh pada pencapaian profit yakni pembiayaan untuk tenaga ahli, pembiayaan untuk tenaga teknik, pembiayaan untuk tenaga pendukung, pembiayaan untuk operasional kantor, pembiayaan untuk peralatan, pembiayaan untuk pelaporan dan dokumentasi, serta pembiayaan untuk mobilisasi. Dari metode yang telah digunakan yakni klasifikasi dengan algoritma C4.5 diharapkan mampu mendukung proses pemutusan sebuah hasil dalam menentukan tingkat profit usaha.

Setelah penjabaran latar belakang yang telah tersaji pada bagian sebelumnya, maka peneliti tertarik untuk mengambil sebuah judul penelitian yakni **“ANALISIS ALGORITMA C4.5 UNTUK MENENTUKAN TINGKAT PROFIT USAHA PADA PT MEGACIPTA SEJATI”**.

## 1.2 Identifikasi Masalah

Merujuk pada uraian latar belakang yang telah dijabarkan sebelumnya, peneliti dapat menetapkan pengidentifikasian masalah atas apa yang terjadi PT Megacipta Sejati, sebagai berikut:

1. Tidak adanya metode atau sistem perusahaan guna mendiagnosis kenaikan dan penurunan profit perusahaan dalam hal ini dalam pemanfaatan dari data yang telah dikumpulkan atau diterima oleh perusahaan.
2. Tidak adanya barometer atau parameter yang resmi yang dibuat oleh perusahaan dalam menentukan tingkat profit pada perusahaan, sehingga menyulitkan pihak manajemen perusahaan dalam menentukannya.

## 1.3 Pembatasan Masalah

Berdasarkan uraian masalah diatas agar penelitian tidak meluas, menjadi lebih tepat guna, lebih terorganisir dan mampu dipelajari lebih lanjut, maka peneliti melakukan pembatasan masalah yang dibatasi oleh peneliti, sebagai berikut:

1. Penelitian ini dilakukan pada PT Megacipta Sejati dengan mengelola serta menganalisa data transaksi sewa dormitory yang sedang berjalan.
2. Penelitian ini menggunakan metode Algoritma C4.5 dengan *software WEKA 3.8.4* yang akan digunakan guna mengidentifikasi dan memberikan hasil profit usaha.

3. Penelitian ini dilakukan pada PT Mega Cipta Sejati dengan mengelola data transaksi sewa dormitory selama 3 bulan terakhir yaitu Maret, April dan Mei.
4. Penelitian ini akan memanfaatkan pola yang dihasilkan dari hasil pengujian WEKA 3.8.4 dan pencarian manual dengan *Microsoft Excel* berbentuk pohon keputusan yang akan dipakai sebagai bahan pengambilan keputusan.

#### 1.4 Perumusan Masalah

Setelah dijabarkan latar belakang, pengidentifikasian masalah, dan pembatasan masalah yang telah dilakukan sebelumnya, permasalahan yang dirumuskan oleh peneliti, antara lain:

1. Bagaimanakah penerapan *Data Mining* khususnya teknik klasifikasi menggunakan Algoritma C4.5 dalam menentukan tingkat profit usaha pada PT Megacipta sejati?
2. Apakah teknik klarifikasi menggunakan *Algoritma C4.5* dapat menemukan pola dalam menentukan tingkat profit usaha?
3. Bagaimanakah cara mengimplementasikan teknik klarifikasi *Algoritma C4.5* sebagai alat untuk menemukan pola yang digunakan dalam menentukan tingkat profit usaha?
4. Bagaimanakah tindakan kedepannya dengan adanya hasil yang diperoleh dari proses penelitian yang dilakukan mengenai tingkat profit usaha sehingga kedepannya minim kesalahan?

## 1.5 Tujuan Penelitian

Perumusan masalah yang ada menimbulkan tujuan yang diharapkan terpenuhi guna menerima masukan dan embaran yang sesuai dan menggunakan alat-alat analisis data. Adapun tujuan dari penelitian ini, adalah:

1. Menerapkan *Data Mining* khususnya teknik klasifikasi menggunakan algoritma C.45 dalam menentukan kenaikan dan penurunan profit usaha pada PT Megacipta.
2. Pengujian Teknik Klasifikasi dan *Algoritma C.45* dalam wujud penerapan dari datamining yang digunakan dalam menentukan kenaikan dan penurunan profit usaha pada PT Megacipta.
3. Pengujian *Algoritma C4.5* menggunakan *software data Mining WEKA* dalam menentukan kenaikan dan penurunan profit usaha pada PT Megacipta.
4. Menguji dan menganalisa penerapan *Data Mining* teknik klasifikasi menggunakan algoritma C4.5 dalam menentukan kenaikan dan penurunan profit usaha pada PT Megacipta.

## 1.6 Manfaat Penelitian

Dihat dari penjabaran yang telah dilakukan pada bagian sebelumnya, diharapkan penelitian yang telah diteliti oleh peneliti memiliki manfaat bagi pembaca dan bahan pendukung bagi penelitian selanjutnya terutama di dunia edukasi. Berikut beberapa manfaat yang dilihat dari segi teoritis dan praktis:

### **1.6.1 Manfaat Teoritis**

Dari segi teoritis, peneliti berharap mampu memperkuat dan memperkaya referensi dan teori dan menjadi pedoman pada penelitian-penelitian selanjutnya. Manfaat teoritis yang diharapkan lahir dari setiap pengujian yang dilakukan oleh peneliti, yakni:

1. Hasil penelitian ini dapat memperkuat dan memperkaya referensi dan teori penggunaan data mining yaitu teknik klasifikasi dengan pohon keputusan menggunakan algoritma C4.5 dalam menentukan tingkat profit usaha.
2. Hasil penelitian ini dapat digunakan sebagai pedoman pada penelitian – penelitian selanjutnya yang berhubungan dengan data mining yaitu teknik klasifikasi dengan pohon keputusan menggunakan algoritma C4.5.

### **1.6.2 Manfaat Praktis**

Dari segi praktis, peneliti berharap mampu memberikan hal-hal yang praktikal. Manfaat praktis yang diharapkan lahir dari setiap pengujian yang dilakukan oleh peneliti, yakni:

1. Bagi Mahasiswa  
Mampu digunakan untuk dijadikan kepustakaan dan selanjutnya juga dijadikan acuan informasi serta mampu menambah pengetahuan mengenai data mining khususnya algoritma C.45.
2. Bagi Peneliti



Mampu meningkatkan pemahaman dan dapat memberikan pengalaman bagi peneliti sehingga mampu menjadi bekal dalam melakukan penelitian selanjutnya juga diharapkan menjadi referensi bagi peneliti yang lain.

### 3. Bagi Perusahaan

Diharapkan dapat memberi kontribusi positif kepada PT Megacipta dalam hal menentukan profit usaha, yang selama ini sangat sulit untuk ditentukan oleh pihak manajemen perusahaan.



---

**Universitas Putera Batam**

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Teori**

Peneliti menguraikan beberapa teori atau prinsip dasar yang dipakai guna mendukung pengambilan hasil, adapun teori yang dimaksud terdiri atas dua teori besar yaitu teori umum dan teori khusus, adapun uraiannya, antara lain:

##### **2.1.1 Teori Umum**

Sebuah penelitian harus didasarkan pada teori-teori yang telah dikemukakan sebelumnya, salah satunya ialah teori umum. Teori umum ini dimaksudkan untuk memperkuat isi dari teori yang ada hal ini dimaksudkan supaya penelitian yang dilakukan memperoleh hasil yang lebih baik. Adapun isi dari teori umum tersebut dapat dijelaskan sebagai berikut:

##### **2.1.1.1 *Knowledge Discovery in Database (KDD)***

*Knowledge Discovery In Database* (KDD) diartikan sebagai sistem guna menemukan informasi dari *database* yang tersedia. *Database* berisi tabel-tabel yang berkolerasi. Informasi-informasi yang telah ditemukan pada *database* tersebut dikonversikan menjadi sebuah basis pengetahuan (*knowledge base*) guna dijadikan sebagai pertimbangan dalam mengambil keputusan.

*Knowledge Discovery in Database* (KDD) dan *data mining* dikolaborasikan oleh pihak-pihak berkepentingan sesuai keperluan pada kondisi tertentu guna menjabarkan operasi penambangan informasi yang potensial tetapi

belum ditemui pada sebuah basis data yang cukup besar. Dua komponen tersebut mempunyai konsepsi yang tidak sama, namun memiliki kolerasi diantara keduanya, bahkan salah satu hierarki dari prosedur KDD ialah *data mining*. Berikut prosedur KDD yang dijabarkan seperti dibawah ini:

### 1. *Data Selection*

Tahap awal yang perlu dijalankan ialah melakukan seleksi pada tahapan *Knowledge Discovery in Database*. Setelah diseleksi, hasil yang didapatkan dilakukan penyimpanan secara terpisah dari basis data operasional yang akan dipergunakan sebagai bahan pendukung *data mining*.

### 2. *Pre-processing / Cleaning*

Tahap selanjutnya merupakan tahap yang perlu diperhatikan, yakni proses *cleaning*. Hal-hal yang harus dilakukan pada proses *cleaning* meliputi membuang data yang terdupliaksi, melakukan pemeriksaan pada data yang tidak konsisten, serta melakukan perbaikan pada data yang dianggap salah atau tidak sesuai, contohnya ialah kesalahan pencetakan. Selanjutnya, terdapat sebuah proses yang bernama *enrichment*, yakni tahapan dimana “memperkaya” data menggunakan data lain yang sejalan dengan data yang telah tersedia sebelumnya guna digunakan pada *Knowledge Discovery in Database* (KDD), semisal informasi eksternal.

### 3. *Transformation*

Pada tahapan ini, terdapat proses kunci yang biasa disebut *coding* sehingga data yang telah ada sesuai dengan keperluan pada saat pemrosesan *data mining*. Proses ini meliputi mengkategorian data sehingga membentuk pola-pola yang memudahkan pengambilan informasi berdasarkan jenis data dan tergantung apa yang dikehendaki oleh pengolah data.

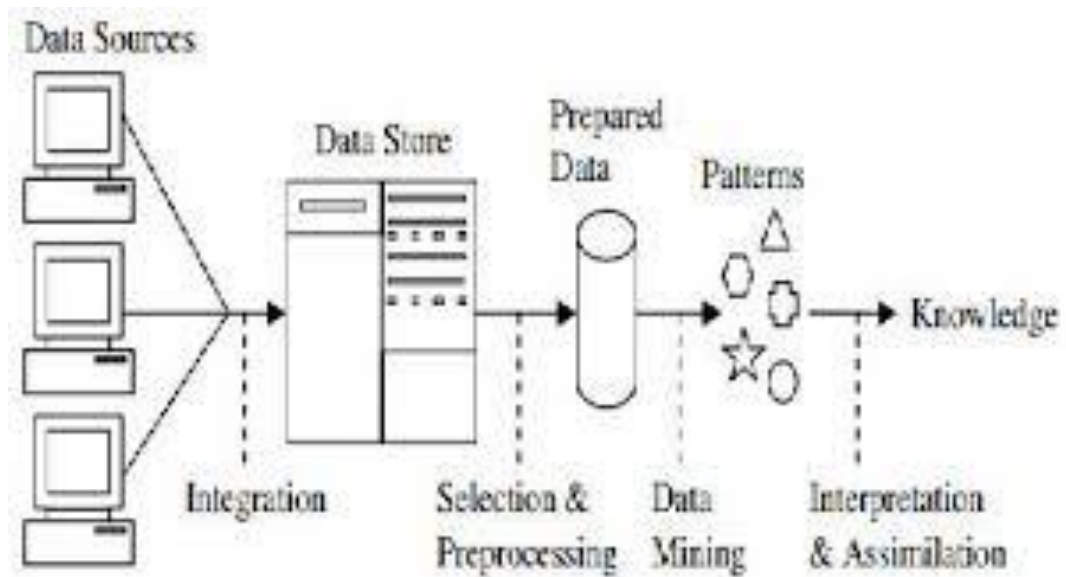
#### 4. *Data Mining*

*Data mining* dapat diartikan sebagai pencarian sebuah corak tertentu atas data yang telah dipilih sebelumnya melalui sistem atau kaidah yang ada. Terdapat varian cara atau algoritma yang dapat dipergunakan pada *Data Mining*. Cara dan algoritma yang ada dipilih sesuai dengan sasaran dan keperluan pada *Knowledge Discovery in Database* (KDD) yang akan dilakukan oleh pengolah data.

#### 5. *Interpretation / Evaluation*

Corak yang terbentuk pada tahap transformasi dan *data mining* harus menghasilkan sebuah hasil yang memudahkan pengambilan keputusan dari semua pihak yang memiliki kepentingan. Oleh karena itu, proses ini adalah proses terakhir yakni *interpretation*. Pada tahapan ini, dilakukan evaluasi pada informasi serta pola apakah data berkontradiktif dengan yang sebenarnya terjadi atau jawaban sementara yang telah dimuskani.

Pendapat dari (Mardi, 2016) proses KDD atau *Knowledge Discovery in Database* mampu dibuat skemanya sebagai berikut:



**Gambar 2.1** Proses Dalam KDD atau *Knowledge Discovery in Database*

**Sumber:** (Mardi, 2016)

### 2.1.1.2 *Data Mining*

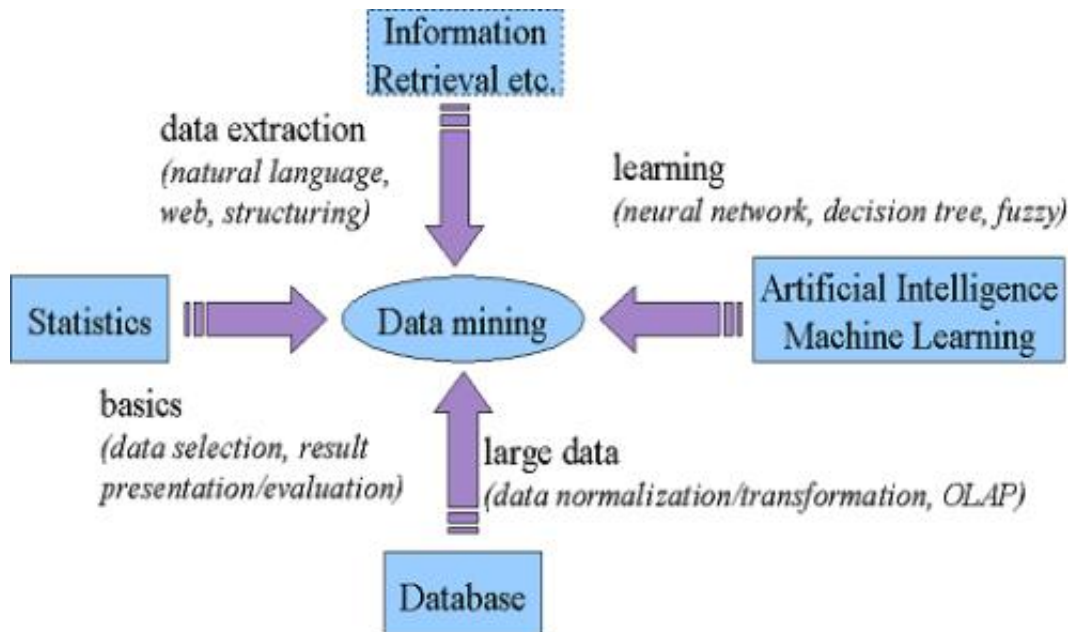
Seperti yang dikemukakan oleh Gartner Group, *data mining* dapat diartikan sebagai tahapan yang dilakukan guna mencari kolerasi yang memiliki pengertian, corak dan kualifikasi tertentu dengan pemilihan dari begitu banyak data kemudian dikumpulkan pada sebuah media penyimpanan berbasis teknologi statistik atau arimatika. *Data mining* sendiri merupakan kolaborasi dari berbagai pengetahuan yakni machine learning, pengenalan pola, statistik, database, dan visualisasi guna menangani masalah-masalah yang terjadi pada pengambilan data seharusnya mampu digunakan untuk pengambilan keputusan.

Berdasarkan pendapat dari David Hand, Heikki Mannila, dan Padhraic Smyth dari MIT, menyatakan bahwa *data mining* dapat dikatakan sebagai proses penganalisan terhadap data yang berskala besar guna mencari kolerasi antara

data tersebut dengan membuat kesimpulan atas data tersembunyi sehingga ditemukan dan mampu dipakai oleh pihak yang berkepentingan. *Data mining* biasanya terdiri dari beberapa komponen yakni statistik, aritmatika, *artificial intelligence*, dan keilmuan tentang mesin guna memilah-milah dan melakukan identifikasi data sesuai dengan keilmuan yang dibutuhkan. *Data mining* dikatakan sebagai berbagai rangkaian tahapan guna menelusuri informasi-informasi baru berbentuk keilmuan apabila tahapan dilakukan secara manual. Terdapat unsur-unsur yang berkaitan dengan *data mining*, yakni sebagai berikut:

1. *Data mining* dapat dikatakan sebagai data yang transformasikan secara otomatis.
2. Data yang dilakukan pemrosesannya biasanya berskala besar.
3. Tujuan *data mining* ialah memperoleh kolerasi dengan motif tertentu sehingga ditemukan informasi yang mampu digunakan untuk pengambilan keputusan.

*Data mining* sudah sering digunakan tetapi tidak terlalu terkenal. Kelemahan dari *data mining* ialah membutuhkan banyak keilmuan dari sektor-sektor yang cukup banyak sebagai acuan atau dasar pengolahannya serta wajib terbukti kebenarannya. Gambar dibawah ini menjelaskan bahwa *data mining* memiliki berakar dari banyaknya aspek keilmuan meliputi kecerdasan buatan (*artificial intelligent*), *machine learning*, statistik, *database*, dan juga *information retrieval*.



**Gambar 2.2** Bidang Ilmu Dari Data Mining

**Sumber:** (Mardi, 2016)

### 2.1.1.3 Pengelompokan *Data Mining*

Dilihat dari peran dan fungsinya, *Data mining* dapat dikelompokkan kedalam tugas berikut:

#### 1. *Description* (Deskripsi)

Analisis melakukan upaya-upaya guna menemukan teknik sederhana atau yang memudahkan sehingga menghasilkan penggambaran corak atau motif tertentu dan tren yang terjadi pada data yang diteliti. Contohnya ialah pada pemilihan umum dalam memilih kepala negara, yakni kemungkinan-kemungkinan yang terjadi atas fakta atau penjelasan ilmiah mengenai seperti apa pilihan pemilih. Pendeskripsian atas motif dan kecenderungan memberikan penjelasan-penjelasan tertentu mengenai hal tersebut.



## 2. *Estimation* (Estimasi)

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Sebagai contoh yaitu estimasi nilai indeks prestasi kumulatif mahasiswa program pasca sarjana dengan melihat nilai indeks prestasi mahasiswa tersebut pada saat mengikuti program sarjana.

## 3. *Prediction* (Prediksi)

*Prediction* tidak jauh berbeda dengan dugaan-dugaan dan kemungkinan yang ada, terkecuali adanya jawaban sementara atas apa yang terjadi dikemudian hari. Dapat diberikan contoh sebagai berikut:

- a. Estimasi atas harga beras dalam dua bulan kedepan .
- b. Estimasi tingkat kemiskinan dalam 5 tahun kedepan.
- c. Estimasi angka harapan hidup dalam beberapa tahun kedepan apabila ketersediaan akan pangan mengalami peningkatan.

Terdapat berbagai sistem dan cara-cara yang dimampu diterapkan terhadap menghasilkan dugaan dan kemungkinan untuk diprediksi.

## 4. *Classification* (Klasifikasi)

Pada fungsi ini, analisis ditemukan sasaran variabel berkategori. Semisal penggolongan pendapatan dikategorikan kedalam tiga tingkatan, yakni

pendapatan tinggi, pendapatan sedang, dan pendapatan rendah. Terdapat contoh lainnya, yakni:

- a. Mendiagnosis penggunaan dengan kartu debit merupakan hal yang efektif atau tidak.
- b. Mendiagnosis pengajuan hutang kepada bank dalam jumlah besar mampu memberikan keuntungan atau kerugian pada perusahaan.
- c. Mendiagnosis kadar obat yang cocok sesuai dengan penyakit yang diderita oleh pasien.

#### 5. *Clustering* (Pengkusteran)

Pengkusteran merupakan pengelompokan record, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan record-record dalam kluster lain.

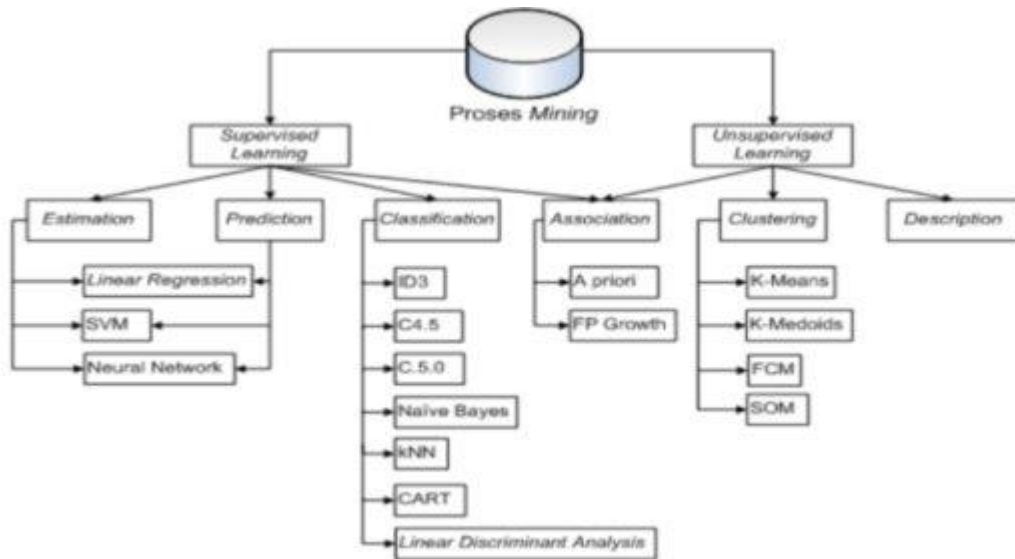
Pengkusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengkusteran. Pengkusteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma pengkusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan record dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan record dalam kelompok lain akan bernilai minimal. Contoh pengkusteran dalam bisnis dan penelitian adalah :

- a. Mengelompokkan target pasar kemudian konsumen-konsumen potensial sehingga mampu menghemat biaya pemasaran.
- b. Mengelompokkan beberapa transaksi yang tidak wajar pada kegiatan operasional perusahaan dan biasanya dilakukan oleh akuntan sebuah perusahaan.
- c. Mengelompokkan kriteria-kriteria dari gen akan penyakit bawaan orangtua yang akan diturunkan kepada anaknya.

6. *Association* (Asosiasi)

*Association* memiliki peran untuk mendapatkan instrumen-instrumen yang berasosiasi satu dengan yang lainnya pada waktu tertentu. Istilah lain yang digunakan ialah analisis keranjang belanja. Berikut contoh-contoh *association* :

- a. Melakukan penelitian dari pengguna akses internet sehingga mampu menerima tanggapan positif ketika dilakukan promosi-promosi dan peningkatan paket internet yang digunakan.
- b. Melakukan pengamatan atas barang-barang kosmetik yang memiliki ketertakaitan satu dengan yang lainnya sehingga memungkinkan barang tersebut dibeli dalam suatu waktu yang sama.



**Gambar 2.3** Pengelompokan Teknik Data Mining

**Sumber:** (Hermawan et al., 2019)

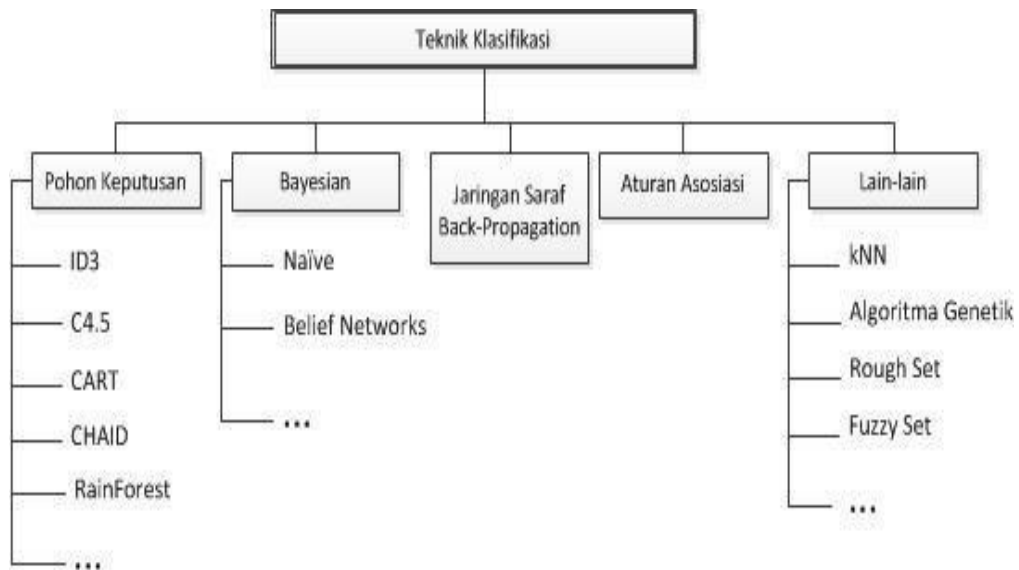
#### 2.1.1.4 Klasifikasi

Seperti yang dikemukakan oleh (Azwanti, 2018), klasifikasi data dapat dikatakan sebagai kegiatan menyatukan instrumen-instrumen yang sama pada suatu kelompok atas objek-objek yang diambil pada suatu basis data dan mengkategorikannya kedalam golongan-golongan tertentu seperti yang dikehendaki oleh analis. Klasifikasi bertujuan guna menentukan model dari training set yang memilah-milah atribut ke dalam kelas yang sesuai, kemudian atribut tersebut diklasifikasikan ke kelas yang belum ada sebelumnya. Terdapat metode-metode pada klasifikasi umum, yakni pengklasifikasi pohon keputusan, pengklasifikasi *Bayesian*, pengklasifikasi *k-nearest neighbour*, penalaran berbasis kasus, algoritma genetika dan teknik logika *fuzzy*.

Satu dari banyaknya tujuan penggunaan *data mining* memanglah klasifikasi. Pengklasifikasian dilakukan untuk kali pertama pada bidang biologi yakni mengklasifikasikan spesies makhluk hidup, pencetusnya ialah Carolus von Linne (atau dikenal dengan nama Carolus Linnaeus) yakni melakukan klasifikasi spesies hewan berdasarkan karakteristik fisiknya sehingga sampai hari ini, ia dikenal sebagai Bapak Klasifikasi.

Berikut model-model yang mendukung proses pengklasifikasian data, yakni sebagai berikut:

- a. Pohon keputusan
- b. Pengklasifikasi *bayes/naive bayes*
- c. Jaringan saraf tiruan
- d. Analisis statistik
- e. Algoritma genetik
- f. *Rough sets*
- g. Pengklasifikasi *k-nearest neighbour*
- h. Metode berbasis aturan
- i. *Memory based reasoning*
- j. *Support vector machine*



**Gambar 2.4** Pengelompokan Teknik Klasifikasi

**Sumber:** (Della, 2017)

### 2.1.1.5 Pohon Keputusan

Salah satu bentuk metode dalam pengklasifikasian data ialah metode pohon keputusan atau *decision tree*. Metode ini dapat diartikan sebagai sebuah metode yang mengubah data dilapangan dalam jumlah besar menjadi sebuah pohon keputusan yang mencerminkan sebuah pilihan-pilihan keputusan. [iluhan-piluhan keputusan tersebut memudahkan para pemegang kepentingan memahami data yang menjadi bahasa alami.

Pohon keputusan itu sendiri digunakan untuk melakukan pembagian pada banyaknya data yang ada menjadi kelompok-kelompok kecil secara terstruktur sehingga menghasilkan pilihan-pilihan guna mendapatkan sebuah kesimpulan. Dikarenakan adanya pembagian anggota-anggota kelompok maka hasil yang

ditunjukkan menjadi serupa. Pohon keputusan biasanya diterjemahkan ke dalam bentuk tabulasi dengan atribut dan record. Atribut sendiri diartikan sebagai suatu acuan sebagai satuan pengkategorian yang membentuk pohon keputusan itu sendiri. Contohnya ialah seseorang yang ingin bermain sepak bola harus memperhatikan kriteria-kriteria tertentu seperti cuaca, kondisi lapangan, dan lainnya. Terdapat atribut yang memberikan solusi dari masing-masing butir data yakni dinamakan target atribut. Nilai yang terkandung di dalam atribut disebut instance. Contohnya ialah atribut kondisi lapangan mempunyai instance yakni lapangan yang rata, berlubang, berlumpur, dan lain sebagainya. Secara garis besar, peran pohon keputusan ialah mentransformasikan data yang ada menjadi tabel, tabel disusun menjadi sebuah model pohon, model pohon yang ada diklasifikasi menjadi *rule*, dan terakhir membuat *rule* menjadi suatu bentuk yang lebih sederhana dan mudah dimengerti.

Terdapat beberapa algoritma yang biasanya membentuk sebuah pohon keputusan, yakni ID3, CART, dan C4.5. Algoritma C4.5 merupakan turunan dari algoritma ID yang dikembangkan lebih lanjut. Sebuah pohon keputusan dibentuk seperti sebuah pohon yang asli dengan bagian-bagian sebagai berikut:

1. *Root Node* : berada paling atas pada sebuah pohon keputusan.
2. *Internal Node* : seperti cabang yang bersumber pada satu input dan mengeluarkan dua output lagi.
3. *Leaf Node* : berada pada ujung pohon yang memiliki satu input dan tidak bercabang lagi

Seperti yang dijelaskan oleh Triisant, terdapat beberapa kelebihan dan kekurangan pada penggunaan pohon keputusan ini, antara lain:

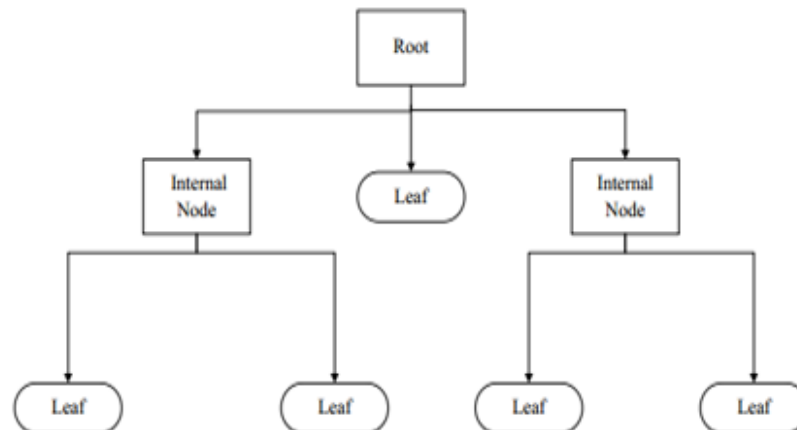
1. Kelebihan Pohon Keputusan:

- Kekompleksan area yang digunakan untuk mengambil keputusan dapat dibuat menjadi lebih sederhana.
- Menyederhanakan berbagai perhitungan data dikarenakan terjadinya eliminasi pada data yang tidak diperlukan oleh analisis.
- Proses pemilihan fitur dari *internal node* yang berbeda lebih fleksibel.
- Meminimalisir kesalahan yang mungkin memunculkan masalah baru karena pohon keputusan menyederhanakan kriteria yang ada tanpa mengurangi mutu hasil yang akan didapatkan.

2. Kekurangan Pohon Keputusan :

- Adanya kemungkinan terjadinya *overlap* yang disebabkan banyaknya hasil keputusan dan kriteria dalam jumlah yang relatif besar. Melalui kejadian ini akan memakan waktu untuk pengambilan keputusan lebih lama dan membutuhkan kapasitas penyimpanan yang semakin besar.
- Adanya jumlah *error* yang banyak pada setiap pohon keputusan.
- Adanya kesulitan dalam merancang pohon keputusan yang menunjukkan hasil maksimal.
- Mutu keputusan yang diambil sangat bergantung pada desain pohon keputusan tersebut dibuat.





**Gambar 2.5** Model Pohon Keputusan

**Sumber:** (Hamidah, 2012)

#### 2.1.1.6 Algoritma C4.5

Algoritma C4.5 diaplikasikan guna mendukung pohon keputusan. (Bening, 2014) mengemukakan sebuah pendapat bahwa algoritma C4.5 dikategorikan sebagai satu dari banyaknya *machine learning*. Mesin atau dalam hal ini komputer yang menggunakan algoritma C4.5 menunjukkan hasil kelompok data yang mampu dipelajari dengan nama *learning dataset*. Algoritma C4.5 menghasilkan sebuah kelompok data yang terbagi kedalam kelas-kelas tertentu melalui proses *test dataset*.

Hasil dari pengolahan yang telah dijabarkan pada bagian sebelumnya akan diolah kembali dengan sebutan *test dataset*. Algoritma C4.5 menjadi acuan dasar guna membuat pohon keputusan yang dinagi enjadi atribut-atribut yang berprioritas tertinggi atau dapat dikatakan mempunyai nilai *gain* tertinggi yang diambil dari nilai *entropy atribut* sebagai pedoman dari pengklasifikasian atribut yang telah ditetapkan sebelumnya. Setelah itu, dibentuklah cabang-cabang pohon

yang dilakukan perluasan sehingga menyerupai seperti pohon. Tertulis dalam kamus IGI Global (*International Publisher of Progressive Academic*), *entropy* ialah banyaknya data-data yang tidak sejalan atau relevan dengan data yang telah pada suatu himpunan data. Gain dapat didefinisikan sebagai data yang diambil yang dipengaruhi oleh adanya perubahan *entropy* pada himpunan data, dapat karena adanya observasi tertentu atau adanya keikutsertaan data tersebut pada suatu set data.

Seperti yang dikemukakan oleh (Andriani, 2012) menyatakan bahwa terdapat empat tahapan C4.5 yang membentuk suatu pohon keputusan, antara lain:

1. Mempersiapkan *data training*. *Data training* dapat diperoleh dari data-data yang sudah ada sebelumnya (*data history* atau data masa lalu) kemudian mengkategorikan data tersebut kedalam golongan-golongan yang dikehendakinya.
2. Menghitung akar dari pohon. Atribut-atribut yang dipilih oleh analis disebut sebagai akar melalui proses perhitungan *gain* yang didapat karena adanya perubahan *entropy* pada setiap atribut tersebut. Apabila *gain* menunjukkan tertinggi, maka atribut tersebut akan dijadikan akar pokok atau paling atas. Nilai *gain* berasal dari nilai *entropy*. Berikut dirumuskan proses perhitungan nilai *entropy* pada sebuah atribut:

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

**Rumus 2.1** Perhitungan *Gain*

Rumus diatas, diartikan sebagai persamaan sistematis untuk menemukan nilai *gain* yang berasal dari *entropy*, dengan keterangan sebagai berikut:

- S : himpunan kasus
- A : atribut
- N : jumlah partisi atribut A
- |Si| : jumlah kasus pada partisi ke-i
- |S| : jumlah kasus dalam S

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

### **Rumus 2.2** Perhitungan *Entropy*

Rumus diatas, diartikan sebagai persamaan sistematis untuk menemukan keandalan suatu hasil *entropy* yang telah dilakukan pada Rumus 2.1, dengan keterangan sebagai berikut:

- S: himpunan kasus
  - A : fitur
  - N : jumlah partisi S
  - $P_i$  : proporsi dari  $S_i$  terdapat S
3. Ulangi langkah ke-1 dan langkah ke-2 pada setiap atribut yang telah dipilih sampai seluruh *record* ikut berpartisipasi.
  4. Ada hal-hal yang harus diperhatikan yakni kapan partisi pohon keputusan harus dihentikan, yakni ketika:
    - a. Semua *record* dalam simpul N mendapat kelas yang sama.

- b. Tidak ada atribut di dalam *record* yang dipartisi lagi.
- c. Tidak ada *record* di dalam cabang yang kosong.

Menurut pendapat dari Marwana (2013) bahwa terdapat dua hal pokok atau mendasar dari proses menemukan hasil *entropy*, antara lain.

1. Apabila pada kolom “benar” atau “tidak” menunjukkan nilai 0 (nol) dengan demikian nilai dari *entropy*-nya berbanding lurus atau akan menunjukkan nilai 0 (nol).
2. Apabila pada kolom “benar” dan “tidak” menunjukkan nilai yang sama dengan demikian nilai dari *entropy*-nya sudah pasti menunjukkan nilai 1 (satu).

Hasil nilai *gain* dan *entropy* yang dilakukan pengujiannya bisa menggunakan alat bantu yakni *software Waikato Environment for Knowledge Analysis (WEKA)*. Seperti yang telah dilakukan oleh (Rifqo & Arzi, 2016), menemukan bahwa algoritma C4.5 mendukung pembuatan pohon keputusan yang dapat dilakukan dengan cara-cara dibawah ini:

1. Pilih atribut sebagai akar, merupakan atribut yang tidak memiliki cabang masukan dan berpengaruh paling besar pada suatu kelas tertentu.
2. Buat cabang untuk tiap-tiap nilai, adalah membuat struktur cabang dari *decision tree* yang sudah berisi nilai-nilai.
3. Bagi kasus dalam cabang, adalah melakukan pembagian atau pengelompokkan dari permasalahan yang terdapat pada struktur cabang dari *decision tree*.

4. Ulangi proses untuk setiap cabang sampai kasus pada cabang memiliki kelas yang sama, adalah melakukan proses pengulangan yang ada pada tahapan pertama hingga ketiga sampai seluruh cabang pada struktur *decision tree* memiliki kelas dan kategori yang sama.

### **2.1.2 Teori Khusus**

Dijabarkan teori khusus yang diambil oleh peneliti, dimana penjelasannya dapat diuraikan seperti dibawah ini:

#### **2.1.2.1 Profit**

Menurut (Elisa, 2018), profit atau nama lainnya ialah laba, didefinisikan oleh Martono dan Harijito sebagai selisih antara pendapatan yang menjadi penghasilan yang diterima oleh perusahaan dari satu pihak maupun beberapa pihak dengan biaya yang menjadi beban yang harus dikeluarkan guna memenuhi kewajibannya kepada pihak lain. Profit atau laba adalah peningkatan nilai modal dari peristiwa-peristiwa masa lalu operasional perusahaan yang terjadi karena adanya transaksi-transaksi keluar masuknya uang yang memberikan pengaruh pada perusahaan dalam jangka waktu tertentu. Lain halnya dengan profit bersih atau laba bersih, didefinisikan sebagai selisih lebih atau bernilai positif dari pendapatan/penghasilan dengan beban/biaya kemudian dikurangi lagi dengan bunga dan pajak.

Berdasarkan pendapat (Elisa, 2018), Sistem yang dibangun akan dipergunakan untuk Prediksi Pencapaian Profit Pada Perusahaan. Data yang semula mentah diolah menjadi berbentuk angka maupun non-angka kemudian

diklasifikasikan menjadi beberapa kelas sehingga mempermudah penganalisisan data pada tahap berikutnya. Kemudian data tersebut dibentuk berdasarkan desain pohon keputusan sebagai output. Beberapa hal untuk mengambil keputusan dalam prediksi pencapaian profit pada perusahaan yakni seperti dibawah ini:

A. Biaya Langsung Personil

1. Biaya Tenaga Ahli
2. Biaya Tenaga Teknik
3. Biaya Tenaga Pendukung

B. Biaya Langsung Non Personil

1. Biaya Operasional Kantor
2. Biaya Peralatan
3. Biaya Pelaporan & Dokumen
4. Biaya Mobilisasi

Variabel yang ditentukan untuk diambil keputuannya ialah BAIK dan TIDAK BAIK.

#### **2.1.2.2 Waikato Environment for Knowledge Analysis (WEKA)**

WEKA adalah ialah sebuah aplikasi dari *machine learning* praktis. “WEKA” berasal dari singkatan *Waikato Environment for Knowledge Analysis*, berasal dari sebuah perguruan tinggi, yakni Universitas Waikato, New Zealand digunakan pada penelitian, edukasi serta aplkasi-aplikasi lainnya. WEKA memiliki kemampuan untuk mengatasi berbagai permasalahan *data mining* di

dunia nyata, terutama pengklasifikasian dengan dasar pendekatan *machine learning*. Aplikasi ini ditulis dalam hirarki *class Java* dengan metode berorientasi objek dan dapat berjalan hampir di semua platform.

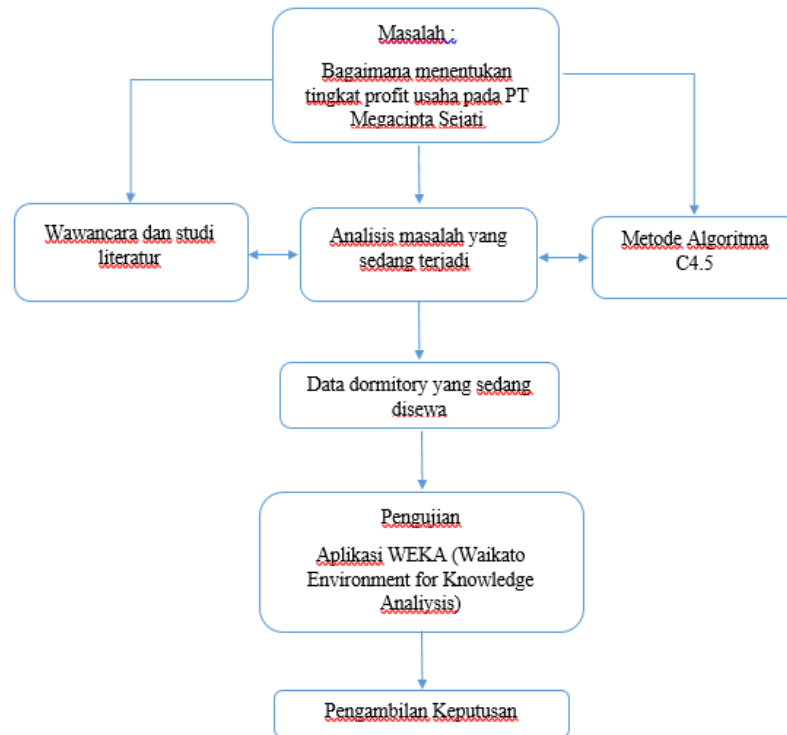


**Gambar 2.6** Tampilan Utama WEKA

**Sumber:** (Witten, 2011)

## 2.2 Kerangka Berpikir

Kerangka berpikir merupakan penyelesaian masalah dari penelitian yang ada berdasarkan langkah-langkah yang telah disusun, kerangka berpikir yang telah disusun seperti dibawah ini:



**Gambar 2.7** Kerangka Berpikir

**Sumber:** (Hasil Penelitian, 2020)

### 2.3 Hipotesis Penelitian

hipotesis merupakan pernyataan yang berisi gambaran atau prediksi adanya hubungan antara dua atau lebih variabel, yang pembuktian kebenarannya masih harus melalui beberapa pengujian. Berdasarkan uraian di atas, dirumuskan hipotesis sebagai berikut :

1. Teknik *datamining* yaitu teknik klasifikasi dengan pohon keputusan menggunakan algoritma C4.5 dapat menganalisis data profit usaha.
2. Hasil penelitian dapat membantu menentukan tingkat profit usaha pada PT Megacipta Sejati.



## 2.4 Penelitian Terdahulu

Berikut diuraikan beberapa penelitian yang telah dilakukan oleh peneliti terdahulu yang dijadikan sebagai acuan dari penelitian ini. Penelitian terdahulu bertujuan agar penelitian yang dilakukan memperoleh hasil yang optimal dan valid. Berikut adalah beberapa penelitian yang dijadikan acuan.

1. Anu And Chauhan Harvinder (2013) penelitian yang berjudul “Implementation Of Decision Tree Algoritma C4.5”. Penelitian ini menunjukkan algoritma ini lebih cocok untuk data sets yang mengandung tidak terlalu banyak data karena hasil yang lebih akurat dan pengimplementasian pohon keputusan dengan metode algoritma C4.5 sangat mudah dan cepat.
2. Jambur (2016) penelitian yang “Menganalisa jumlah pelanggan aktif dengan menerapkan teknik *datamining* metode algoritma C4.5”. Hasil yang didapat mampu membantu dalam memfilter pelanggan yang aktif dan tidak aktif pada perusahaan yang menunjukkan bahwa pelanggan dengan jumlah bayar tinggi, pembelian pertahun sedang, diskon rendah maka merupakan pelanggan aktif dan sebaliknya maka merupakan pelanggan pasif.
3. Pujiono et al., (2013) penelitian tentang “Penggunaan *WEKA* untuk menganalisis kepuasan publik dalam mewujudkan *good governance* di kota Yogyakarta”. Penelitian ini menunjukkan bahwa pelaksanaan *good governance* dapat disimpulkan akan berjalan dengan baik karena pelayanan yang diberikan Dinas Perizinan selama ini cukup baik.

4. Harman (2018) penelitian tentang “Memprediksi indikator *website* yang baik dengan menerapkan teknik *datamining* metode algoritma C4.5”. Penelitian ini menunjukkan bahwa indikator navigasi dan graphic design adalah indikator yang paling mempengaruhi indikator *website* yang baik.
5. Mardin (2017) penelitian tentang “klasifikasi menggunakan algoritma C4.5”. Penelitian ini menunjukkan bahwa algoritma C4.5 dapat menghasilkan pohon keputusan dengan metode klasifikasi yang dapat memberikan informasi yang dibutuhkan.
6. Chen, Zhang, Li, & Shahabi, (2018) penelitian yang berjudul “*Perfomance evaluation of the GIS-based datamining techniques of best-first decision tree, random forest, and naive bayes tree for landslide suscepttibility modeling*”. Penelitian ini menunjukkan bahwa teknik data mining yang paling akurat dalam memprediksi merupakan teknik *random forest model*.
7. Elisa (2018) penelitian tentang “Memprediksi profit pada perusahaan dengan menggunakan teknik *datamining* metode algoritma C4.5”. Hasil penelitian mampu menjadi sebuah referensi bagi perusahaan yang ingin meningkatkan profitnya dengan cara *saving cost*. Hasil penelitian ini merupakan biaya yang paling banyak mempengaruhi profit perusahaan adalah biaya mobilisasi dan biaya tenaga kerja.
8. Azwanti (2018) penelitian tentang “Memprediksi mahasiswa yang mengulang mata kuliah menggunakan teknik *datamining* metode algoritma C4.5. Penelitian ini dilakukan di AMIK Labuhan Batu

menyimpulkan bahwa algoritma C4.5 merupakan algoritma yang sangat membantu dalam mengklasifikasi data karena dapat memperoleh dengan jelas karakteristik data yang diklasifikasi baik dalam bentuk pohon keputusan maupun *if-then*.



---

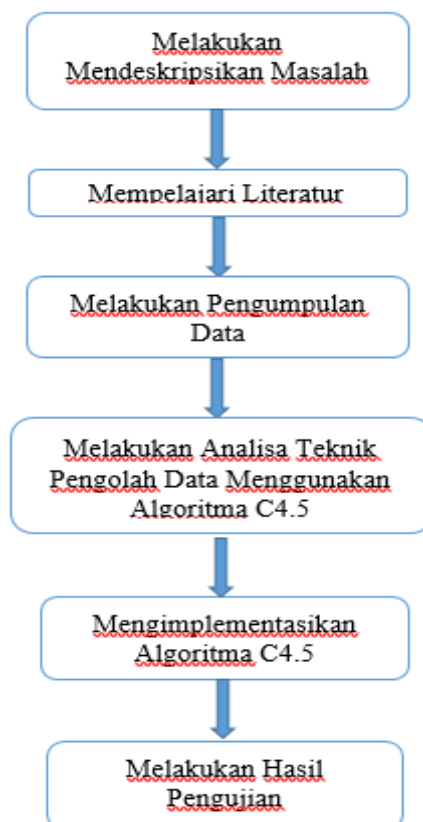
**Universitas Putera Batam**

## BAB III

### METODE PENELITIAN

#### 3.1 Desain Penelitian

Desain penelitian mencerminkan teknik untuk mengumpulkan data dan model analisa untuk menguji data. Desain penelitian itu sendiri berguna sebagai arahan guna menemukan jawaban atas permasalahan yang timbul dan sebagai batasan jika terjadi hal-hal yang menyimpang sehingga mempengaruhi hasil keputusan dari penelitian.



**Gambar 3.1** Desain Penelitian

**Sumber :** (Hasil Penelitian, 2020)

Berikut uraian mengenai langkah-langkah yang berasal dari desain penelitian diatas:

1. Mendeskripsikan Masalah, melakukan pendeskripsikan berbagai permasalahan yang akan teliti dan merupakan langkah pertama yakni menentukan, mendefinisikan, serta membatasi masalah yang akan diteliti, sehingga dapat mengetahui kenaikan dan penurunan profit usaha.
2. Mempelajari Literatur, membantu peneliti dalam proses mencari literatur serta sumber-sumber pustaka yang perlu digunakan dalam penelitian.
3. Pengumpulan Data, proses dalam mengumpulkan data yang diperlukan guna diproses dalam penelitian pengumpulan data, sehingga dapat menentukan tingkat profit usaha pada PT Megacipta sejati.
4. Melakukan Analisa Teknik Pengolah Data Menggunakan algoritma C4.5, peneliti pada tahap ini dilakukan pengolahan menggunakan algoritma C4.5 untuk menghasilkan pohon keputusan dan menghasilkan *rule* tingkat profit usaha PT Megacipta Sejati.
5. Mengimplementasikan Algoritma C4.5. Peneliti mengimplementasikan algoritma C4.5 yang dilakukan menentukan tingkat profit usaha berdasarkan data yang telah menganalisa.
6. Melakukan Hasil Pengujian, peneliti melakukan pengujian dan hasil dari pengolahan dan implementasi dengan menggunakan *Software Data Mining open source WEKA*.

### 3.2 Objek Penelitian

Objek penelitian yang diambil oleh peneliti dijabarkan mengenai keterangan lokasi penelitian dan penjelasan mengenai jadwal penelitian antara lain:

#### 3.2.1 Lokasi Penelitian

Penelitian dilaksanakan di PT Megacipta Sejati yang merupakan salah satu perusahaan yang sewa dormitory yang ada di Kota Batam dengan alamat Jalan Hang Kesturi belakang Cammo Industrial Estate Blok B, Batam, Kepulauan Riau.

#### 3.2.2 Jadwal Penelitian

Demikian penjelasan penulis mengenai rencana waktu yang akan dilakukan dalam melaksanakan kegiatan analisa mengenai kepuasan pelanggan, seperti dibawah ini :

**Tabel 3.1** Jadwal Penelitian

No.	Kegiatan	Bulan																			
		Maret 2020				April 2020				Mei 2020				Juni 2020				Juli 2020			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1.	Penentuan Judul																				
2.	Bimbingan																				
3.	Identifikasi Masalah																				
4.	Studi																				

	<b>Literatur</b>																		
<b>5.</b>	<b>Pengumpulan Data</b>																		
<b>6.</b>	<b>Pengolahan Data</b>																		
<b>7.</b>	<b>Pengujian Hasil</b>																		
<b>8.</b>	<b>Penyusunan Laporan</b>																		

Sumber: Hasil Penelitian (2020)

### 3.3 Populasi dan Sampel

Penelitian menjabarkan mengenai keterangan yang dimiliki oleh populasi dan sampel yang akan diteliti dan dilakukan pengujiannya guna menjawab hipotesis yang telah dinyatakan pada bagain sebelumnya, sebagai berikut:

#### 3.3.1 Populasi

Populasi pada penelitian ini ialah data transaksi sewa dormitory di PT Megacipta Sejati pada periode Maret 2020 sampai dengan Mei 2020.

#### 3.3.2 Sampel

Sampel yang ditentukan oleh peneliti yang berasal dari populasi menggunakan sebanyak 50 data sampel dari populasi data transaksi sewa dormitory di PT Megacipta Sejati pada periode Maret 2020 sampai dengan Mei 2020.



### 3.4 Variabel Penelitian

Berdasarkan pendapat dari (Sugiyono, 2001) variabel penelitian ialah suatu instrumen penelitian yang memiliki karakteristik tertentu. Variabel dibutuhkan guna menjadi bahan untuk dilakukan penganalisisan dan diolah menjadi sebuah keputusan, data yang diambil oleh peneliti dalam penelitian ini ialah data profit usaha yang sedang berjalan dengan variabel penilaian tingkat profit usaha yang meliputi:

1. Biaya Tenaga Kerja

Biaya yang berupa pembayaran untuk para karyawan sebagai imbal jasa dari apa yang telah diberikan oleh karyawan kepada perusahaan.

2. Biaya Operasional

Biaya yang berupa pengeluaran uang guna menjalankan aktivitas operasional perusahaan, yakni diklasifikasikan kedalam biaya penjualan dan administrasi yang digunakan untuk menambah pendapatan.

3. Biaya Perlengkapan

Biaya yang berupa pengeluaran uang untuk membeli barang-barang yang bersifat habis dipakai atau tidak bisa dipakai berulang kali.

4. Biaya Administrasi

Biaya yang berupa pengeluaran untuk membayar semua biaya bank dan pajak yang secara langsung berhubungan dengan kegiatan operasional.

5. Biaya Perawatan

Biaya yang berupa pengeluaran guna memelihara aset bahkan sistem yang ada didalamnya selama aset tersebut beroperasi.

## 6. Biaya Legalitas

Biaya yang berupa pembayaran surat izin dan dokumen yang berkaitan dengan operasional perusahaan secara langsung.

Variabel yang akan menjadi keputusan adalah profit dan defisit.

### **3.5 Teknik Pengumpulan Data**

Peneliti menentukan beberapa teknik untuk mengumpulkan data informasi yang dibutuhkan agar peneliti dapat untuk mencapai tujuan penelitian yakni:

#### 1. Teknik Studi Literatur

Dengan melakukan studi literatur ada buku-buku, jurnal, dan penelitian telah dilakukan data yang didapat dari studi literatur ini dijadikan sebagai acuan untuk membuat penelitian .

#### 2. Teknik Observasi

Teknik observasi digunakan guna mengetahui data – data dormitory mulai dari bulan Januari 2020 sampai April 2020 dari PT Megacipta Sejati.

#### 3. Teknik Dokumentasi

Dalam teknik dokumentasi penulis mengumpulkan dokumen – dokumen dari dormitory mulai dari bulan Januari 2020 – April 2020 dari PT Megacipta Sejati.

#### 4. Teknik Wawancara

Teknik wawancara dilakukan penulis proses bertanya jawab kepada pemilik PT Megacipta Sejati.

### 3.6 Model Penelitian

Model penelitian penulis menggunakan *datamining* dan algoritma C4.5 yakni data yang diklasifikasi sesuai dengan kategorinya, sebelum mendapatkan nilai *gain* dan *entropy* harus menghitung klasifikasi terlebih dahulu sesuai dengan kategorinya supaya bisa dapatkan hasil nilai *gain* dan *entropy* guna terbentuk suatu desain pohon keputusan. Dalam proses pencarian nilai akan dilakukan berulang-ulang dan terus-menerus sehingga bisa dapatkan *node* pada pohon keputusan sehingga tidak memiliki turunan sampai didapatkan pola aturan- aturan dasar keputusan sebagai bahan acuan dalam menentukan tingkat profit usaha pada PT Megacipta Sejati.