

CV. RUDANG MAYANG



**Buku Hasil Penelitian**

# Penerapan Algoritma C4.5

Untuk Mengukur Tingkat  
Kepuasan Mahasiswa

**Fristi Riandari**

**Agustina Simangunsong**



# PENERAPAN ALGORITMA C4.5 UNTUK MENGUKUR TINGKAT KEPUASAN MAHASISWA

Mengukur kepuasan mahasiswa sangatlah penting untuk diperhatikan mengingat tingginya tingkat persaingan dalam dunia pendidikan seiring berkembangnya pengetahuan dan teknologi. Dalam mengukur kepuasan mahasiswa sangatlah penting diperhatikan apakah pelayanan yang diharapkan oleh mahasiswa sesuai dengan yang diterima. Mengukur kepuasan mahasiswa akan sangat membantu suatu Perguruan Tinggi dalam peningkatan kualitas pelayanan yang nantinya juga akan berdampak pada peningkatan jumlah mahasiswa. Sistem yang sangat dibutuhkan adalah Penerapan Data Mining Dalam Mengukur Kepuasan Mahasiswa. Dimana mahasiswa akan menjadi objek yang memberikan penilaian/opini terhadap variable-variabel yang memiliki karakteristik yaitu *Tangible*, *Reability*, *Responsivnes*, *Assurance*, *Empathy*. Sistem ini dibangun dengan algoritma C4.5 dan Pengujian sistem dengan bantuan aplikasi *RapidMiner*.



**Fristi Riandari**, Lahir di Sei Gelugur, 24 Februari 1993, menempuh Pendidikan SD di SD Negeri 101828 (2005), SMP Negeri 3 Pancur Batu (2008), SMK YAPIM (2011), Sarjana Komputer di STMIK Pelita Nusantara Medan (2015) dan Menempuh Magister Ilmu Komputer di Universitas Putra Indonesia "YPTK" Padang (2017). Saat ini Penulis Aktif sebagai Dosen Profesional di STMIK Pelita Nusantara Program Studi Teknik Informatika, selain Profesi Dosen, Penulis juga aktif menulis Buku dan Meneliti di bidang Computer Science.



**Agustina Simangunsong**, Sarjana Komputer di STMIK Pelita Nusantara Medan (2015) dan Menempuh Magister Ilmu Komputer di Universitas Putra Indonesia "YPTK" Padang (2017). Saat ini Penulis Aktif sebagai Dosen Profesional di STMIK Pelita Nusantara Program Studi Teknik Informatika, selain Profesi Dosen, Penulis juga aktif menulis Buku dan Meneliti di bidang Computer Science.



CV. RUDANG MAYANG  
Info@iocscience.org  
0852 9758 3991

ISBN 978-623-90597-2-9



# **Penerapan Algoritma C4.5 Untuk Mengukur Tingkat Kepuasan Mahasiswa**

**Fristi Riandari  
Agustina Simangunsong**

**CV.Rudang Mayang**

# **Penerapan Algoritma C4.5 Untuk Mengukur Tingkat Kepuasan Mahasiswa**

**Penulis:**

Fristi Riandari  
Agustina Simangunsong

**ISBN :** 978-623-90597-2-9

**Editor:**

Hengki Tamando Sihotang

**Penyunting:**

Hengki Tamando Sihotang  
Paska Marto Hasugian

**Desain Sampul dan Tata Letak:**

Tambun

**Penerbit:**

**CV.Rudang Mayang**

**Redaksi:**

Perumahan Romeby Lestari BlokC No 14C  
HP: 0852 9758 3991

**Distributor:**

**CV.Rudang Mayang**

Perumahan Romeby Lestari BlokC No 14C  
HP: 0852 9758 3991

Hak Cipta dilindungi Undang-undang

Dilarang memperbanyak karya tulis ini dalam bentuk apapun dan dengan cara  
Apapun tanpa ijin tertulis dari penerbit



# **Penerapan Algoritma C4.5 Untuk Mengukur Tingkat Kepuasan Mahasiswa**

Copyright © CV. Rudang Mayang

Penulis: Fristi Riandari, Agustina Simangunsong

Editor: Hengki Tamando Sihotang

Penyunting: Hengki Tamando Sihotang, Paska Marto Hasugian

Desain Sampul dan Tata Letak: Tambun

Penerbit:

CV.Rudang Mayang

Web : [www.iocscience.org](http://www.iocscience.org)

Email: [info@iocscience.org](mailto:info@iocscience.org)

Fristi Riandari, Agustina Simangunsong  
Penerapan Algoritma C4.5 Untuk Mengukur Tingkat Kepuasan  
Mahasiswa

CV. Rudang Mayang, 2019

ix. 76;16 x 23 cm

ISBN: 978-623-90597-2-9

Cetakan:

Redaksi:

Perumahan Romeby Lestari BlokC No 14C

HP: 0852 9758 3991

Distributor:

CV.Rudang Mayang

Perumahan Romeby Lestari BlokC No 14C

HP: 0852 9758 3991

Hak Cipta dilindungi Undang-undang

Dilarang memperbanyak karya tulis ini dalam bentuk apapun dan dengan cara  
Apapun tanpa ijin tertulis dari penerbit

## KATA PENGANTAR

Dengan mengucapkan syukur pada Tuhan Yang Maha Esa judul “*Penerapan Algoritma C4.5 Untuk Mengukur Tingkat Kepuasan Mahasiswa*” ini dapat diselesaikan pada waktunya. Sebagai hasil penelitian, monograf ini membahas data *mining* algoritma C4.5. Kegiatan penelitian dilaksanakan untuk memenuhi kewajiban seorang dosen selain kegiatan mengajar. Pelaksanaan Tri Dharma pada perguruan Tinggi ini didukung penuh Yayasan Demokrat Cemerlang dan LPPM STMIK Pelita Nusantara. Pada kesempatan ini Peneliti mengucapkan terima kasih atas dukungan seluruh pihak antara lain :

1. Dirlitjen Penguatan Riset dan Pengembangan Kementerian Riset, Teknologi dan Pendidikan Tinggi, selaku pemberi dana selama melaksanakan penelitian ini.
2. Lembaga Layanan Pendidikan Tinggi Wilayah I Sumatera Utara.
3. Bapak Drs. John F. Marpaung, M.A., selaku Pembina Yayasan Pendidikan Demokrat Cemerlang.
4. Ibu Ritha Z. Tarigan, S.E., M.M., selaku Ketua Yayasan Pendidikan Demokrat Cemerlang.
5. Bapak Hengki Tamando Sihotang, S.Kom., M.Kom., selaku Ketua STMIK Pelita Nusantara Medan.
6. Ibu Fristi Riandari, S.Kom., M.Kom., selaku Wakil Ketua I Bidang Akademik STMIK Pelita Nusantara Medan.
7. Ibu Sulindawaty, S.Kom., M.Kom., selaku Ketua Program Studi Teknik Informatika STMIK Pelita Nusantara Medan.
8. Bapak Fricles A. Sianturi, S.Kom., M.Kom., selaku Ketua LPPM STMIK Pelita Nusantara Medan.
9. Seluruh dosen dan mahasiswa STMIK Pelita Nusantara Medan.

Akhir kata untuk perbaikan hasil penelitian ini, diharapkan pembaca memberikan kritik saran yang membangun.

Penulis

## DAFTAR ISI

|                                                                            |           |
|----------------------------------------------------------------------------|-----------|
| Daftar isi .....                                                           | i         |
| Kata Pengantar .....                                                       | iii       |
| Abtrak .....                                                               | v         |
| <b>I. Pendahuluan .....</b>                                                | <b>1</b>  |
| <b>II. Tinjauan Pustaka .....</b>                                          | <b>5</b>  |
| 2.1 Knowledge Discovery in Database (KDD) .....                            | 5         |
| 2.2 Data Mining .....                                                      | 9         |
| 2.2.1 Karakteristik Data Minging .....                                     | 11        |
| 2.2.2 Pengeolahan data Mining .....                                        | 12        |
| 2.2.3 Tugas dan Fungsi Data Mining.....                                    | 16        |
| 2.2.4 Langkah-langkah Data Mining .....                                    | 17        |
| 2.3 Klasifikasi .....                                                      | 18        |
| 2.4 Decision Tree .....                                                    | 20        |
| 2.5 Algoritman C4.5.....                                                   | 22        |
| <b>III. Metode Penelitian .....</b>                                        | <b>27</b> |
| 3.1 Tahapan Penelitian .....                                               | 27        |
| 3.2 Langka kerja Penelitian .....                                          | 27        |
| <b>IV Hasil dan Luaran yang Dicapai .....</b>                              | <b>31</b> |
| 4.1 Analisa Masalah.....                                                   | 31        |
| 4.2 Analisa data Mining untuk mengukur tingkat<br>Kepuasan Mahasiswa ..... | 32        |
| 4.3 Data Selection .....                                                   | 33        |
| 4.4 Data Cleaning.....                                                     | 34        |
| 4.5 Data Transformation .....                                              | 34        |
| 4.6 Penetapan Metode C4.5.....                                             | 38        |
| 4.6.1 Proses Algoritma C4.5.....                                           | 39        |
| 4.6.2 Klasifikasi Data Algoritma C4.5 .....                                | 39        |
| <b>V Kesimpulan dan saran .....</b>                                        | <b>65</b> |
| 5.1 Kesimpulan .....                                                       | 65        |
| 5.2 Saran .....                                                            | 65        |
| <b>DAFTAR PUSTAKA.....</b>                                                 | <b>67</b> |

# **PENERAPAN ALGORITMA C4.5 UNTUK MENGUKUR TINGKAT KEPUASAN MAHASISWA**

## **Abstrak**

Mengukur kepuasan mahasiswa sangatlah penting untuk diperhatikan mengingat tingginya tingkat persaingan dalam dunia pendidikan seiring berkembangnya pengetahuan dan teknologi. Dalam mengukur kepuasan mahasiswa sangatlah penting diperhatikan apakah pelayanan yang diharapkan oleh mahasiswa sesuai dengan yang diterima. Mengukur kepuasan mahasiswa akan sangat membantu suatu Perguruan Tinggi dalam peningkatan kualitas pelayanan yang nantinya juga akan berdampak pada peningkatan jumlah mahasiswa. Sistem yang sangat dibutuhkan adalah Penerapan Data Mining Dalam Mengukur Kepuasan Mahasiswa. Dimana mahasiswa akan menjadi objek yang memberikan penilaian/opini terhadap variable-variabel yang memiliki karakteristik yaitu Tangiable, Reability, Responsivnes, Assurance, Empathy. Sistem ini dibangun dengan algoritma C4.5 dan Pengujian sistem dengan bantuan aplikasi *RapidMiner*.



**Kata Kunci:** *kepuasan mahasiswa, decision tree, data mining, C4.5.*

## **I. Pendahuluan**

Perkembangan pengetahuan dan teknologi mengakibatkan tingginya tingkat persaingan dalam dunia usaha dan salah satunya adalah dalam dunia pendidikan, yang membuat setiap instansi pendidikan berlomba-lomba dalam memberikan kualitas pelayanan terhadap mahasiswa agar instansi tersebut dapat terus berdiri dan dapat berkembang.

Kualitas pelayanan merupakan ukuran seberapa baik tingkat layanan yang diberikan sesuai harapan pelanggan (Khoirista A., Yulianto E. and Mawardi M. K., 2015).

STMIK Pelita Nusantara merupakan sebuah yayasan pendidikan yang selalu mengikuti kemajuan pendidikan, perkembangan pengetahuan dan teknologi yang dapat meningkatkan kesadaran mahasiswa terhadap kualitas pelayanan yang diberikan suatu instansi. Kualitas pelayanan dalam dunia pendidikan yang diberikan suatu instansi menjadi sangat penting karena dapat membantu instansi dalam memberikan kualitas pelayanan yang sesuai dengan yang diharapkan mahasiswa, yang nantinya dimasa yang akan datang dapat membantu memberikan solusi jenis pelayanan apa yang paling berpengaruh terhadap kepuasan mahasiswa sehingga

pelayanan tersebut dapat ditingkatkan. Kualitas pelayanan yang sesuai dengan yang diharapkan mahasiswa juga akan menjadikan hal tersebut sebagai bahan promosi bagi masyarakat luas yang dapat meningkatkan jumlah mahasiswa.

*Data mining* merupakan proses perulangan dan memerlukan interaksi manusia dalam prosesnya untuk menemukan pola atau model baru yang dapat digeneralisasi untuk masa yang akan datang, dan bermanfaat jika digunakan untuk melakukan suatu tindakan (Azmi Z. and Dahria M., 2013). *Data mining*, sering juga disebut *knowledge discovery in database* (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data dan historis untuk menemukan keteraturan, pola atau hubungan dalam set data yang berukuran besar (Santosoe H., Hariyadi I P. and Prayitno., 2016). Ada beberapa metode yang dapat digunakan dalam *data mining*, yaitu *Association Rule*, *Apriori*, *Roughset*, *K-Means*, Algoritma C4.5 dan yang lain-lain.

Algoritma C4.5 merupakan salah satu solusi pemecahan kasus yang sering digunakan untuk membuat pohon keputusan dalam pemecahan masalah pada teknik klasifikasi yang memiliki karakteristik yaitu dengan

proses penentuan nilai *entropy* dan nilai *gain*. Sebuah penelitian menjelaskan bahwa Algoritma C4.5 merupakan teknik klasifikasi dengan menggunakan *entropy* dan keuntungan informasi sebagai pemisah pada pohon keputusan (Florence A.M and Savithri.R., 2013). Sebuah penelitian melakukan perbandingan antara algoritma C4.5 dan algoritma CART dalam mengklasifikasi nilai mahasiswa menjelaskan bahwa algoritma C4.5 memiliki nilai akurasi lebih tinggi yaitu 85.61% sedangkan algoritma CART memiliki nilai akurasi 84.95% (Rahmayuni I., 2014). Dalam penelitian lain, data mining di implementasikan untuk mengukur masa studi mahasiswa dan didapatkan hasil pengujian yang menjelaskan bahwa tingkat kesalahan dalam mengukur masa studi mahasiswa hanya 5% (Haryati S., Sudarsono A., Suryana E., 2015).

Berdasarkan masalah yang dihadapi dan hasil dari penelitian-penelitian sebelumnya yang telah penulis jelaskan, penulis tertarik untuk melakukan sebuah penelitian untuk mengukur kepuasan mahasiswa dengan menggunakan algoritma C4.5. Adapun judul penelitiannya

yaitu “**Implementasi Algoritma C4.5 Untuk Mengukur Tingkat Kepuasan Mahasiswa**”.

### **Rumusan Masalah**

1. Bagaimana menganalisa algoritma C4.5 untuk mengukur tingkat kepuasan mahasiswa?



## **II. Tinjauan Pustaka**

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

*Data mining* disebut juga *Knowledge Discovery in Database (KDD)* didefinisikan sebagai ekstraksi informasi potensial implisit dan tidak dikenal dari sekumpulan data. Proses *Knowledge Discovery in Database* melibatkan hasil proses *data mining* (proses pengekstrak kecenderungan suatu pola data), kemudian mengubah hasilnya secara akurat menjadi informasi yang mudah dipahami (Tampubolon K., Saragih H. and Reza B., 2013).

Ada beberapa macam pendekatan yang berbeda yang diklasifikasikan sebagai teknik pencarian informasi/pengetahuan dalam KDD. Ada pendekatan kuantitatif, seperti pendekatan probabilistik seperti logika induktif, pencarian pola, dan analisis pohon keputusan. Pendekatan yang lain meliputi deviasi, analisis kecenderungan, algoritma genetik, jaringan saraf tiruan, dan pendekatan campuran dua atau lebih dari beberapa pendekatan yang ada. Pada dasarnya ada enam elemen yang paling esensial dalam teknik pencarian informasi/pengetahuan dalam KDD yaitu:

1. Mengerjakan sejumlah besar data.
2. Diperlukan efisiensi berkaitan dengan volume data.

3. Mengutamakan ketetapan/keakuratan.
4. Mengutamakan ketetapan/keakuratan.
5. Membutuhkan pemakaian bahasa tingkat tinggi.
6. Menggunakan beberapa bentuk dari pembelajaran otomatis.
7. Menghasilkan hasil yang menarik.

Perusahaan memerlukan kecerdasan bisnis untuk mengembangkan proses bisnis, memonitor waktu, biaya kualitas, dan pengendalian. Gartner Group mendefenisikan bahwa kecerdasan bisnis atau *business intelligence* (BI) sebagai berikut. “*Business intelligence is the process of transforming data into information and through discovery transforming that information into knowledge*”. Dalam definisi ini disampaikan bahwa kecerdasan bisnis merupakan proses pengubahan data menjadi informasi. Dari kumpulan informasi yang ada akan diambil polanya menjadi pengetahuan. Tujuan kecerdasan bisnis adalah untuk mengubah data yang sangat banyak menjadi nilai bisnis melalui laporan analistik. (Tampubolon K. et al, 2013). Proses *Knowledge Discovery in Database* (KDD) secara garis besar dapat dijelaskan sebagai berikut:

1. Data Selection

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses *datamining* disimpan dalam suatu berkas, terpisah dari basis data operasional.

## 2. *Pre- processing / Cleaning*

Sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses pembersihan pada data yang menjadi fokus KDD. Proses pembersihan mencakup antara lain membuang *duplikasi data*, memeriksa data yang *inkonsisten*, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (*tipografi*).

## 3. Transformation

*Coding* adalah transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

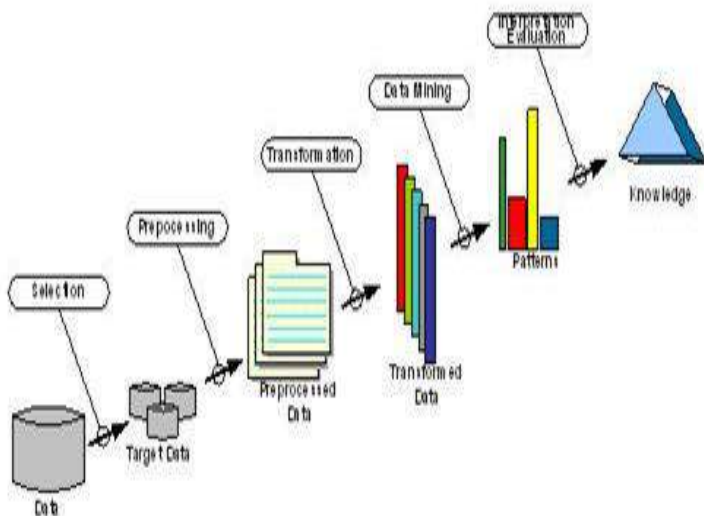
## 4. *Data mining*

*Data mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat

bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

#### 5. *Interpretation / Evaluation*

Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya (Nasari F. dan Darma S., 2015).



**Gambar 2.1 Knowledge Database Discovery (KDD)**

## 2.2 Data Mining

*Data mining* adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran *computer (machine learning)* untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. Definisi lain diantaranya adalah pembelajaran berbasis induksi (*induction-based learning*) adalah proses pembentukan definisi-definisi konsep umum yang dilakukan dengan cara mengobservasi contoh-contoh spesifik dari konsep-konsep yang akan dipelajari. *Knowledge Discovery in Databases* (KDD) adalah penerapan metode saintifik pada data mining. Dalam konteks ini data mining merupakan satu langkah dari proses KDD (Eska J., 2016).

Selain beberapa penjelasan diatas, ada juga yang menjelaskan bahwa “*Data mining* merupakan proses ekstraksi data menjadi informasi yang sebelumnya belum tersampaikan, dengan teknik yang tepat proses *data mining* akan memberikan hasil yang optimal (Abdillah G., Putra F. A. and Renaldi F., 2016)”.

*Data mining* juga didefinisikan sebagai sebuah proses untuk menemukan hubungan, pola dan tren baru yang bermakna dengan menyaring data yang sangat besar, yang tersimpan dalam penyimpanan, menggunakan teknik pengenalan pola seperti teknik Statistik dan Matematika.

*Data mining* bukanlah suatu bidang yang sama sekali baru. Salah satu kesulitan untuk mendefinisikan *data mining* adalah kenyataan bahwa *data mining* mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan terlebih dahulu.

Berawal dari beberapa disiplin ilmu, *data mining* bertujuan untuk memperbaiki teknik tradisional sehingga bisa menangani:

1. Jumlah data yang sangat besar
2. Dimensi data yang tinggi
3. Data yang heterogen dan berbeda sifat

Sedangkan menurut para ahli, *data mining* merupakan sebuah analisa dari observasi data dalam jumlah besar untuk menemukan hubungan yang tidak diketahui sebelumnya dan dua metode baru untuk meringkas data agar mudah dipahami serta kegunaannya untuk pemilihan data (Kamagi D. H dan Seng Hansun, 2014).

Maka dapat disimpulkan bahwa, Sementara menurut Turban, E. dkk 2005, *data mining* merupakan sebuah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai

database besar (Hendrian S, 2018). Secara garis besar data mining dapat dikelompokkan menjadi 2 kategori utama, yaitu: a) *Descriptive mining*, yaitu proses untuk menemukan karakteristik penting dari data dalam suatu basis data. Teknik data mining yang termasuk dalam descriptive mining adalah clustering, association, dan sequential mining. b) *Predictive*, yaitu proses untuk menemukan pola dari data dengan menggunakan beberapa variabel lain di masa depan. Salah satu teknik yang terdapat dalam predictive mining adalah klasifikasi (Andri, Kunang Y. N. and Murniati S., 2013). Berdasarkan definisi-defenisi yang telah disampaikan, hal penting yang terkait dengan *data mining* adalah:

1. *Data mining* merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses berupa data yang sangat besar.
3. Tujuan *data mining* adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat.

### **2.2.1 Karakteristik *Data Mining***

Data mining juga memiliki karakteristik eperti yang dijelaskan pada penelitian (Santose H., Haryadi I. P. and Prayitno 2016), yaitu:

1. *Data mining* berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
2. *Data mining* biasa menggunakan data yang sangat besar.
3. *Data mining* berguna untuk membuat keputusan yang kritis, terutama dalam strategi.

### **2.2.2 Pengolahan Data Mining**

Data Mining dibagi menjadi beberapa metode berdasarkan tugas yang dapat dilakukan, yaitu:

#### **1. Asosiasi (*Association*)**

Digunakan untuk mengenali kelakuan dari kejadian - kejadian khusus atau proses dimana hubungan asosiasi muncul pada setiap kejadian. Salah satu contohnya *Market Basket Analysis*, yaitu salah satu metode asosiasi yang menganalisa kemungkinan pelanggan untuk membeli beberapa item secara bersamaan. Penting tidaknya suatu aturan asosiasi dapat diketahui dengan dua parameter yaitu:

##### **a. *Support***

Suatu ukuran yang menunjukkan seberapa besar tingkat dominasi *item* / *itemset* dari keseluruhan transaksi. Ukuran ini menentukan apakah suatu *item* / *itemset* layak untuk dicari *confidence* tersebut



(contoh, dari keseluruhan transaksi yang ada, seberapa besar tingkat dominasi yang menunjukkan *item* A dan B dibeli bersamaan).

b. *Confidence*

Suatau ukuran yang menunjukkan hubungan antar dua *item* secara *conditional* (contoh, seberapa sering item B dibeli jika orang membeli item A).Nofriansyah D. et al, 2016 dalam (Kusrini dan Lutfi 2009:150)

2. Pengklusteran (*clustering*)

Pengklusteran 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. Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklusteran. Pengklusteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (*homogeny*), yang mana kemiripan

dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan *record* dalam kelompok lain akan bernilai minimal.

Contoh pengklusteran dalam bisnis dan penelitian adalah:

- a. Mendapatkan kelompok-kelompok konsumen untuk target pemasaran dari satu suatu produk bagi perusahaan yang tidak memiliki dana pemasaran yang besar.
- b. Untuk tujuan audit akuntansi, yaitu melakukan pemisahan terhadap *ial* dalam baik dan mencurigakan.
- c. Melakukan pengklusteran terhadap ekspresi dari *gen*, untuk mendapatkan kemiripan perilaku dari *gen* dalam jumlah besar. Nofriansyah D. et al, 2016 dalam (Kusrini dan Lutfi 2009:12)

### 3. Prediksi (*prediction*)

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang.

Contoh prediksi bisnis dan penelitian adalah:

- a. Prediksi harga beras dalam tiga bulan yang akan datang.

- b. Prediksi persentasi kenaikan kecelakaan lalu lintas tahun depan jika batas bawah kecepatan dinaikkan.

Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi. Nofriansyah D. et al, 2016 dalam (Kusrini dan Lutfi 2009:11).

#### 4. Estimasi (*estimation*)

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih kearah numerik dari pada kearah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Sebagai contoh akan dilakukan estimasi tekanan darah sistolik pada pasien rumah sakit berdasarkan umur pasien, jenis kelamin, indeks berat badan, dan level sodium darah. Hubungan antara tekanan darah sistolik dan nilai variabel prediksi dalam proses pembelajaran akan menghasilkan model estimasi. Model estimasi yang dihasilkan dapat digunakan untuk kasus baru lainnya. Contoh lain yaitu estimasi nilai indeks prestasi kumulatif mahasiswa program pasca sarjana dengan

melihat nilai indeks prestasi mahasiswa tersebut pada saat mengikuti program sarjana. Nofriansyah D. et al, 2016 dalam (Kusrini dan Lutfi 2009:10)

#### 5. Klasifikasi (*classification*)

Klasifikasi adalah fungsi pembelajaran yang memetakan (mengklasifikasi) sebuah unsur (*item*) data ke dalam salah satu dari beberapa kelas yang sudah didefinisikan. Contoh lain klasifikasi dalam bisnis dan penelitian adalah:

- a. Menentukan apakah suatu transaksi kartu kredit merupakan transaksi yang curang atau tidak.
  - b. Memperkirakan apakah suatu pengajuan hipotek oleh nasabah merupakan suatu kredit yang baik atau buruk.
  - c. Mendiagnosis penyakit seorang pasien untuk mendapatkan termasuk kategori penyakit apa.
- Nofriansyah D. et al, 2016 dalam (Kusrini dan Lutfi 2009:11)

### 2.2.3 Tugas dan Fungsi *Data Mining*

Data mining menganalisis data menggunakan tool untuk menemukan pola serta aturan dalam himpunan data. Perangkat lunak bertugas untuk menemukan pola dengan cara mengidentifikasi aturan dan fitur pada data. Tool data

mining diharapkan mampu mengenal pola ini dalam data dengan input minimal dari user (Tampubolon K. et al, 2013).

#### **2.2.4 Langkah-langkah *Data Mining***

Pada penelitian Tampubolon K. et al, 2013 dalam (Feen Lee & Juan Santana, 2010:37-40) “ada empat tahap yang dilalui dalam Data Mining antara lain:

1. Tahap pertama: *Precise statement of the problem* (mendefinisikan permasalahan yang ingin diketahui). Misalnya ingin mengetahui apakah seorang *customer* berpotensi memiliki kredit macet, atau mengidentifikasi seorang *customer* apakah akan pindah ke kompetitor bisnis kita, dan lain sebagainya. Setelah menemukan pertanyaan bisnis yang perlu dijawab oleh data mining, selanjutnya tentukan tipe tugas untuk menjawab pertanyaan bisnis tersebut. Tugas dasar yang menjadi dasar algoritma data mining adalah klasifikasi, regresi, segmentasi, asosiasi dan sequence analisis.
2. Tahap kedua: *Initial Exploration* (Mempersiapkan data yang menjadi sumber untuk data mining termaksud data “*cleaning*” untuk mempelajari polanya). Setelah menemukan definisi masalah, langkah berikutnya adalah mencari data yang mendukung definisi masalah. Menentukan porsi data yang digunakan men-

training data mining berdasarkan algoritma data mining yang telah dibuat. Setelah persiapan data selesai dilakukan, langkah berikutnya adalah memberikan sebagian dataa kedalam algoritma data mining.

3. Tahap ketiga: *Model building and validation*. Validasi apakah data mining memberikan prediksi yang akurat. Setelah training data selesai dilakukan, data mining tersebut perlu di-“uji” atau di-validasi keakuratannya terhadap data *testing*.
4. Tahap ke-empat: *Deployment*. Tahap ini memilih aplikasi yang tepat terhadap data mining untuk membuat prediksi.”

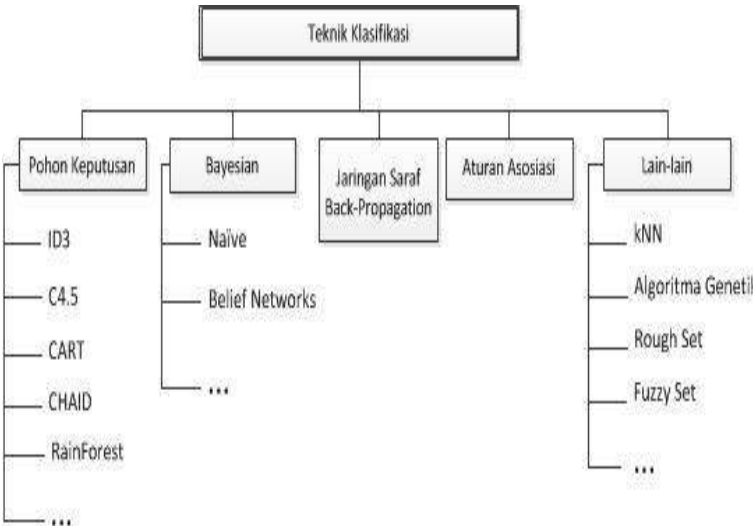
## **2.3 Klasifikasi**

Klasifikasi adalah proses menemukan sekumpulan model yang menggambarkan serta membedakan kelas-kelas data. Tujuan dari klasifikasi adalah agar model yang dihasilkan dapat digunakan untuk memprediksi kelas dari suatu data yang tidak mempunyai label kelas. Jika diberikan sekumpulan data yang terdiri dari beberapa fitur dan kelas, maka klasifikasi adalah menemukan model dari kelas tersebut sebagai fungsi dari fitur-fitur yang lain (Sukardi, Syukur A. and Supriyanto C., 2014).

Ada juga yang menjelaskan bahwa, Klasifikasi data merupakan suatu proses menemukan properti-properti yang sama pada sebuah himpunan obyek di dalam sebuah basis data dan mengklasifikasikannya ke dalam kelas-kelas yang berbeda-beda menurut model klasifikasi yang ditetapkan. Tujuan dari klasifikasi adalah untuk menemukan model dari *training set* yang membedakan atribut ke dalam kategori atau kelas yang sesuai, model tersebut kemudian digunakan untuk mengklasifikasikan atribut yang kelasnya belum diketahui sebelumnya. Teknik klasifikasi terbagi menjadi beberapa teknik yang salah satunya adalah Pohon Keputusan (Lorena S., Zarman W. and Hamidah I., 2014).

Penelitian lain menyebutkan bahwa teknik klasifikasi adalah pendekatan sistematis untuk membangun model klasifikasi dari kumpulan data masukan. Misalnya, teknik pohon keputusan, *Bayesian* (*Naive Bayesian* dan *Bayesian Belief Networks*), Jaringan Saraf Tiruan (*Backpropagation*), teknik yang berbasis konsep dari penambangan aturan-aturan asosiasi, dan teknik lain (*K-Nearest Neighbor*, algoritma genetik, teknik dengan pendekatan himpunan *rough* dan *fuzzy*). Klasifikasi merupakan teknik mengklasifikasikan data. Perbedaannya dengan metode *clustering* terletak

pada data, dimana pada *clustering* variable dependen tidak ada, sedangkan pada *classification* diharuskan ada variable dependen (Jamhur A. I., 2016).



**Gambar 2.2 Pengelompokkan Teknik Klasifikasi**

**2.4    *Decision Tree***

*Decision tree* adalah pohon keputusan yang dimana setiap cabangnya menunjukkan pilihan diantara sejumlah alternatif pilihan yang ada, dan setiap daun akan menunjukkan keputusan yang dipilih. *Decision tree* biasa digunakan untuk mendapatkan sebuah informasi untuk tujuan pengambilan sebuah keputusan.

Ada yang mendefinisikan bahwa pohon keputusan adalah model prediksi menggunakan struktur pohon atau



struktur berhirarki. Selain karena pembangunannya relatif cepat, hasil dari model yang dibangun juga mudah untuk dipahami, sehingga *Decision Tree* ini merupakan metode klasifikasi yang paling populer digunakan. *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* (Lorena S. et al, 2014), dan sebuah penelitian lainnya menjelaskan sebagai berikut “A decision tree is a flowchart-like tree structure, where each internal node (non leaf node) denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (or terminal node) holds a class label (Agrawal G. L. and Prof. Gupta H., 2013)”. Dalam definisi ini disebutkan bahwa Pohon keputusan adalah pohon *flowchart* atau diagram alir dimana masing-masing node menunjukkan atribut yang telah diuji, masing-masing cabang mewakili hasil pembagian yang telah diuji dan setiap node daun (node internal) mewakili kelompok kelas tertentu.

Manfaat dari pohon keputusan dapat dilihat sebagai berikut:

1. Mem-*break down* proses pengambilan keputusan yang kompleks menjadi lebih simpel sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan.
2. Pohon Keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel *input* dengan sebuah variabel target.
3. Pohon keputusan memadukan antara eksplorasi data dan pemodelan, sehingga sangat bagus sebagai langkah awal dalam proses pemodelan bahkan ketika dijadikan sebagai model akhir dari beberapa teknik lain.

## 2.5 Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang digunakan untuk membangun sebuah pohon keputusan (*decision tree*) dari data. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 yang juga merupakan algoritma untuk membangun sebuah pohon keputusan. Algoritma C4.5 secara rekursif mengunjungi tiap simpul keputusan, memilih percabangan optimal, sampai tidak ada cabang lagi yang mungkin dihasilkan (Rahmayuni I, 2014).

Algoritma C4.5 atau pohon keputusan merupakan metode yang umum digunakan untuk melakukan klasifikasi pada data mining. Seperti yang telah dijelaskan sebelumnya, klasifikasi merupakan suatu teknik menemukan kumpulan pola atau fungsi yang mendeskripsikan serta memisahkan kelas data yang satu dengan yang lainnya untuk menyatakan objek tersebut masuk pada kategori tertentu dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Metode ini populer karena mampu melakukan klasifikasi sekaligus menunjukkan hubungan antar atribut. Banyak algoritma yang dapat digunakan untuk membangun suatu *decision tree*, salah satunya ialah algoritma C4.5. Algoritma C4.5 dapat menangani data numerik dan diskret. Algoritma C4.5 menggunakan rasio perolehan (*gain ratio*). Sebelum menghitung rasio perolehan, perlu dilakukan perhitungan nilai informasi dalam satuan bits dari suatu kumpulan objek, yaitu dengan menggunakan konsep entropi (Supriyanti W., Kusrini and Amborowati A., 2016).

Pengertian lainnya yaitu Algoritma C4.5 merupakan generasi baru dari algoritma ID3 yang dikembangkan oleh J. Ross Quinlan pada tahun 1983. Untuk membuat sebuah pohon keputusan, algoritma ini dimulai dengan

memasukkan training *samples* ke dalam simpul akar pada pohon keputusan. *Training samples* adalah sampel yang digunakan untuk membangun model *classifier* dalam hal ini pohon keputusan. Kemudian sebuah atribut dipilih untuk mempartisi sampel ini. Untuk tiap nilai yang dimiliki atribut ini, sebuah cabang dibentuk. Setelah cabang terbentuk maka *subset* dari himpunan data yang atributnya memiliki nilai yang bersesuaian dengan cabang tersebut dimasukkan ke dalam simpul yang baru.

Ada juga yang menyebutkan bahwa algoritma C4.5 mengkonstruksi pohon keputusan dari data pelatihan, yang berupa kasus-kasus atau *record-record (tupel)* dalam *database* (Putra D. W. T., 2016).

#### 1. Konsep Entropi

Entropi (S) yaitu merupakan jumlah bit yang diperkirakan dibutuhkan untuk dapat mengekstrak suatu kelas (+ atau -) dari sejumlah data acak pada ruang sampel S. Entropi dapat dikatakan sebagai kebutuhan bit untuk menyatakan suatu kelas. Semakin kecil nilai entropy maka akan semakin entropy digunakan dalam mengekstrak suatu kelas. Entropy digunakan untuk mengukur ketidakaslisan S. Besarnya Entropi pada ruang sampel S didefinisikan dengan :

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

Dimana :  $\sum_{i=1}^n$

$S$  : himpunan kasus

$A$  : fitur

$n$  : jumlah partisi  $S$

$p_i$  : proporsi dari  $S_i$  terhadap  $S$

## 2. Konsep Gain

Gain ( $S, A$ ) merupakan Perolehan informasi dari atribut  $A$  relatif terhadap output data  $S$ . Perolehan informasi didapat dari output data atau variabel dependent  $S$  yang dikelompokkan berdasarkan atribut  $A$ , dinotasikan dengan gain ( $S, A$ ).

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i) \quad (2)$$

Dengan :

$S$  : Himpunan kasus

$A$  : Atribut

$n$  : Jumlah partisi atribut  $A$

$|S_i|$  : Jumlah kasus pada partisi ke  $i$

$|S|$  : Jumlah kasus dalam  $S$

Ada beberapa tahapan dalam membuat sebuah pohon keputusan dalam algoritma *C4.5*, yaitu :

1. Mempersiapkan data *training*. Data *training* biasanya diambil dari data *histori* yang pernah

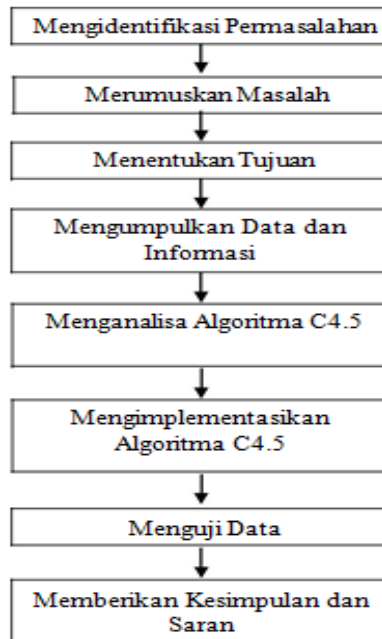
terjadisebelumnya atau disebut data masa lalu dan sudah dikelompokkan dalam kelas-kelas tertentu.

Menghitung akar dari pohon. Akar akan diambil dari atribut yang akan terpilih, dengan cara menghitung nilai *gain* dari masing-masing atribut, nilai *gain* yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut, hitung dahulu nilai *entropy*. Untuk menghitung nilai *entropy* digunakan rumus (1).

### III. Metode Penelitian

#### 3.1. Tahapan Penelitian

Dalam penyusunan penelitian ini, kerangka berpikir yang digunakan dapat dilihat pada gambar berikut:



Gambar 3.1 Tahapan Penelitian

#### 3.2. Langkah Kerja Penelitian

Berikut adalah langkah-langkah yang dilakukan peneliti dalam melakukan penelitian ini yang disertai dengan penjelasan:

1. Melakukan identifikasi masalah. Hal ini perlu dilakukan karena dengan mengidentifikasi masalah

terlebih dahulu peneliti akan lebih mudah dalam menemukan solusi yang paling tepat dari permasalahan yang akan dibahas.

2. Perumusan masalah dilakukan agar dapat memahami masalah-masalah yang telah ditentukan identifikasinya.
3. Selanjutnya, setelah masalah dapat dipahami, maka peneliti akan menentukan apa yang menjadi tujuan.
4. Mengumpulkan data dan informasi, tujuan dari tahapan ini adalah untuk mendapatkan informasi tentang kriteria-kriteria yang akan dinilai oleh mahasiswa dalam mengukur kepuasan mahasiswa. Pengumpulan data dan informasi dilakukan dengan beberapa cara, seperti mencari dan mempelajari literatur-literatur dan jurnal, melakukan observasi dengan melakukan pengamatan langsung ke STMIK Pelita Nusantara Medan untuk mendapatkan data yang dibutuhkan yang berkaitan dengan permasalahan yang sedang dibahas pada penelitian ini dan membagikan kuesioner kepada mahasiswa dengan seperangkat pertanyaan untuk dijawab dimana tingkat mahasiswa merasa puas dan tingkat tidak puas untuk mendapat gambaran dan penjelasan secara lebih terperinci,



sehingga data yang diinginkan dapat tercapai yaitu data yang lengkap dan akurat.

5. Proses analisa algoritma C4.5 merupakan tahap dimana setelah masalah yang akan dianalisa ditemukan, selanjutnya data yang ada akan diolah. Perancangan Algoritma C4.5 akan dilakukan setelah semua data yang dibutuhkan sudah lengkap. Pemrosesan data dilakukan sesuai dengan tahapan KDD, yaitu:
  - a. Selection
  - b. Preprocessing/Cleaning
  - c. Transformation
  - d. Data Mining
  - e. *Interpretation/Evaluation* (Nasari F. dan Darma S., 2015).
6. Implementasi, tujuan dari tahapan ini yaitu untuk mendapatkan hasil dari prediksi kepuasan mahasiswa terhadap kualitas pelayanan STMIK Pelita Nusantara. Untuk mendapatkan hasil yang diinginkan maka data akan dianalisa dan kemudian akan dilakukan implementasi dari metode tersebut.
7. Menguji data, cara kerja sistem dalam mengolah data dapat diketahui dengan melakukan pengujian data. Tahap pengujian dilakukan dengan menggunakan

*software* yang sudah ditentukan sesuai dengan langkah-langkah yang ada dalam *software* tersebut.

8. Kesimpulan dan saran, pada tahapan ini penulis akan menarik kesimpulan dan memberikan saran dari tahapan analisa yang sudah dilakukan dalam mengukur kepuasan mahasiswa terhadap kualitas pelayanan berdasarkan hasil kuisioner.

## **IV. Hasil Dan Luaran Yang Dicapai**

### **4.1 Analisa Masalah**

Rendahnya tingkat kenaikan jumlah mahasiswa membuat institusi kesulitan menentukan tindakan apa yang harus dilakukan dalam meningkatkan jumlah mahasiswa agar dapat meningkat pesat, mengingat sudah banyak tindakan seperti promosi diluar kampus yang dilaksanakan. Mengingat hal tersebut institusi dirasa perlu untuk melakukan suatu tindakan seperti mengukur tingkat kepuasan karena sampai saat ini kegiatan tersebut belum pernah dilakukan. Tujuan dari upaya mengukur tingkat kepuasan mahasiswa ini adalah untuk melihat apakah pelayanan yang di berikan oleh pihak institusi sudah sesuai dengan yang diharapkan mahasiswa atau tidak karena melihat masalah sebelumnya bahwa pelayanan terhadap mahasiswalah yang dianggap sebagai media promosi paling baik dan agar institusi juga dapat mengetahui pelayanan dalam bidang apa yang paling berpengaruh terhadap kepuasan mahasiswa sehingga pelayanan tersebut dapat ditingkatkan.

## **4.2 Analisa Data Mining Untuk Mengukur Tingkat Kepuasan Mahasiswa**

Penelitian ini dilakukan di STMIK Pelita Nusantara Medan, pengumpulan data dilakukan dengan cara observasi dan penyebaran angket kepada mahasiswa. Data yang diperoleh berupa data kualitatif dari mahasiswa tahun ajaran 2018/2019, jumlah mahasiswa yang diambil sebagai sampel dalam mengisi angket yaitu 104 orang dan sampel yang digunakan adalah sebesar 34 orang. Dari data-data yang telah dikumpulkan maka didapatkan variabel yang menjadi alat ukur dalam menentukan tingkat kepuasan mahasiswa.

Penelitian ini akan memberikan informasi kriteria-kriteria yang mendasari tingkat kepuasan mahasiswa terhadap fasilitas maupun pelayanan yang diberikan oleh pihak kampus, yang menjadi *input* pada penelitian ini adalah hasil angket yang sudah ditentukan oleh pihak STMIK Pelita Nusantara Medan dan telah disebarkan pada tahun 2019 dan *output* yang dihasilkan adalah syarat – syarat mahasiswa tersebut merasa puas terhadap fasilitas dan pelayanan yang diberikan.

*Output* atau hasil pada penelitian ini dibagi menjadi dua kategori yaitu PUAS dan TIDAK PUAS. Berdasarkan *output* yang dihasilkan maka digunakan data mining

dengan teknik klasifikasi karena pada teknik ini terdapat variabel kategori yang akan menghasilkan informasi:

1. *Data Selection* yaitu menggabungkan data dari beberapa sumber.
2. *Data Cleaning* yaitu untuk membuang data yang tidak konsisten dan *noise*.
3. *Data Transformation* yaitu mentransformasikan data menjadi bentuk yang sesuai untuk di *mining*.
4. *Data Mining* yaitu proses esensial di mana metode yang intelegen digunakan untuk mengekstrak pola data, kemudian *Pattern evolution* untuk mengidentifikasi pola yang benar-benar menarik yang mewakili pengetahuan berdasarkan atas beberapa tindakan yang menarik
5. *Graphical User Interface (GUI)* yaitu untuk *End User*.

#### **4.3 Data Selection**

Seluruh data dari angket yang sudah disebarkan dikumpulkan akan digabungkan menjadi satu informasi lengkap untuk dapat di proses ke tahap selanjutnya. Jumlah keseluruhan data yaitu 104 orang.

#### **4.4 Data Cleaning**

Setelah hasil seleksi didapatkan selanjutnya data hasil seleksi akan diproses ke tahap *datacleaning* untuk membuang data yang tidak konsisten dan *noise*. Pada tahap ini data hasil angket mahasiswa yang diperoleh menjadi 91 orang dari data yang sebelumnya berjumlah 104 orang.

#### **4.5 Data Transformasi**

Selanjutnya setelah data sudah di *cleaning* maka data tersebut akan diubah atau digabungkan kedalam bentuk penilaian yang sesuai untuk diproses kedalam data mining.

Berdasarkan angket yang telah diselesaikan didapatkan 5 (lima) buah kriteria yang diambil dalam mengukur tingkat kepuasan mahasiswa yaitu Tangible, Reability, Responsivnes, Assurance dan Empathy yang terdiri dari beberapa sub – sub kriteria di antaranya :

1. *Tangibles* (bukti langsung)

*Tangibles* (bukti langsung), yakni fasilitas fisik, perlengkapan, pegawai dan sarana komunikasi ini dianggap lebih penting dan paling berpengaruh dari kelima kriteria yang ada dalam memberikan kepuasan terhadap mahasiswa. Kriteria ini dibagi menjadi 3 (tiga) penilaian dan jumlah pernyataan yang dinilai

pada kriteria ini yaitu 4 (empat) pernyataan. Klasifikasi dapat dilihat pada tabel 4.1.

**Tabel 4.1 Kriteria Penilaian *Tangibles***

| NO | <i>Tangibles</i> |        |
|----|------------------|--------|
| 1  | 3.51 - 5.00      | Tinggi |
| 2  | 2.00 - 3.50      | Cukup  |
| 3  | $\leq 1.99$      | Rendah |

## 2. *Reliability* (keandalan)

*Reliability* (keandalan), yakni kemampuan memberikan jasa dengan segera dan memuaskan. *Reliability* (keandalan) dianggap penting dalam mengukur tingkat kepuasan mahasiswa dan merupakan kriteria kedua dalam mengukur tingkat kepuasan mahasiswa. Kriteria dibagi menjadi 3 penilaian dan jumlah pernyataan yang diberikan adalah 4 (empat), klasifikasi dapat dilihat pada tabel 4.2.

**Tabel 4.2 Kriteria Penilaian *Reability***

| NO | <i>Reability</i> |        |
|----|------------------|--------|
| 1  | 3.51 - 5.00      | Tinggi |
| 2  | 2.00 - 3.50      | Cukup  |
| 3  | $\leq 1.99$      | Rendah |

### 3. *Responsiveness* (daya tanggap)

*Responsiveness* (daya tanggap), yakni kemampuan untuk memberikan jasa dengan tanggap.

*Responsiveness* Pendapatan kriteria ketiga yang digunakan dalam mengukur tingkat kepuasan mahasiswa. Kriteria ini dibagi menjadi 3 (tiga) penilaian dan jumlah pernyataan yang dinilai pada kriteria ini yaitu 4 (empat) pernyataan, klasifikasi dapat dilihat pada tabel 4.3.

**Tabel 4.3 Kriteria Penilaian *Responsiveness***

| NO | <i>Responsiveness</i> |        |
|----|-----------------------|--------|
| 1  | 3.51 - 5.00           | Tinggi |
| 2  | 2.00 - 3.50           | Cukup  |
| 3  | $\leq 1.99$           | Rendah |

### 4. *Assurance* (jaminan)

*Assurance* (jaminan), yakni kemampuan, kesopanan, dan sifat dapat dipercaya yang dimiliki oleh para staf, bebas dari bahaya, resiko dan keragu – ragan. *Assurance* merupakan kriteria keempat dalam mengukur tingkat kepuasan mahasiswa. Pernyataan



yang diberikan pada kriteria ini adalah sebanyak 3 (tiga), pernyataan dinilai sesuai dengan yang sudah ditentukan, klasifikasi dapat dilihat pada tabel 4.4.

**Tabel 4.4 Kriteria Assurance**

| NO | <i>Assurance</i> |        |
|----|------------------|--------|
| 1  | 3.51 - 5.00      | Tinggi |
| 2  | 2.00 - 3.50      | Cukup  |
| 3  | $\leq 1.99$      | Rendah |

#### 5. *Empathy*

*Empathy* menjadi kriteria kelima atau kriteria terakhir yang ditentukandalam membantu mengukur tingkat kepuasan mahasiswa. Kriteria dibagi menjadi 3 penilaian dan jumlah pernyataan yang diberikan adalah 4 (empat), klasifikasi dapat dilihat pada tabel 4.5.

**Tabel 4.5 Kriteria Penilaian *Empathy***

| NO | <i>Empathy</i> |        |
|----|----------------|--------|
| 1  | 3.51 - 5.00    | Tinggi |
| 2  | 2.00 - 3.50    | Cukup  |
| 3  | $\leq 1.99$    | Rendah |

Setelah semua kriteria sudah diklasifikasikan selanjutnya data – data hasil angket yang sudah di isi oleh

mahasiswa ditransformasikan sesuai dengan hasil klasifikasi dari setiap kriteria dan dari 91 data hasil cleaning yang kemudian di transformasikan maka didapatkan 34 aturan yang berbeda yang dapat dilihat pada table 6.6 selanjutnya.

#### **4.6 Penerapan Metode C4.5**

Perancangan sistem yang akan digunakan dalam menentukan syarat – syarat dalam mengukur tingkat kepuasan mahasiswa. Data awal yang akan dibagi perkelas berupa *numeric* ataupun *nonnumeric* untuk memudahkan analisa selanjutnya. Setelah semua data yang akan diproses dibagi perkelas, selanjutnya akan dilakukan proses klasifikasi dengan yang membentuk sebuah pohon keputusan sebagai *outputnya*.

Proses dalam pengambilan keputusan untuk menentukan syarat – syarat dalam mengukur tingkat kepuasan mahasiswa adalah sebagai berikut:

1. *Tangibles*
2. *Reability*
3. *Responsiveness*
4. *Assurance*
5. *Empathy*

Kategori yang akan menjadi keputusan adalah Puas dan Tidak Puas.

#### 4.6.1 Proses Algoritma C4.5

Seperti yang sudah dijelaskan diatas jumlah sampel yang digunakan sebanyak 34 data dari hasil angket yang sudah disebarkan tahun 2019 dan variabel – variabel yang sudah ditentukan oleh kampus sebanyak 5 variabel.

Dari *attribute* yang sudah dikelompokkan atau diklasifikasi didapatkan *format* data akhir. Misalnya data pada tabel 4.3 *Tangibles* yang penilaiannya dibagi menjadi 3 yaitu Tinggi, Cukup dan Rendah begitu juga dengan *attribute* lainnya.

#### 4.6.2 Klasifikasi Data Algoritma C4.5

Dari data hasil angket mahasiswa yang sudah didapatkan, maka akan dilakukan klasifikasi data algoritma C4.5 dengan membuat pohon keputusan. Kasus tersebut kemudian akan dibuat pohon keputusan untuk mengidentifikasi syarat – syarat dalam mengukur tingkat kepuasan mahasiswa berdasarkan hasil angket mahasiswa di STMIK Pelita Nusantara Medan yang telah di sebarakan dengan melihat data hasil angket dalam mengukur tingkat kepuasan mahasiswa (*Tangibles, Reability, Responsiveness, Assurance, Empathy*). Untuk memilih atribut sebagai akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung *gain*

digunakan rumus (2), sedangkan untuk menghitung nilai *entropy* dapat dilihat pada rumus (1), pada bab 2.

Dalam pembuatan pohon keputusan, yang harus dilakukan adalah menghitung jumlah kasus, jumlah kasus untuk keputusan “Puas”(S1), jumlah kasus untuk keputusan “Tidak Puas” (S2) dan kasus yang dibagi berdasarkan atribut *Tangibles*, *Reability*, *Responsiveness*, *Assurance*, *Empathy*, kemudian dilakukan perhitungan *gain* untuk setiap atribut. Adapun langkah-langkah pembuatan pohon keputusan yaitu:

Menentukan atribut sebagai akar dan menghitung nilai informasi *gain* atribut. Untuk memilih atribut sebagai akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada. Dibutuhkan nilai *Entropy* untuk menentukan *gain* tertinggi.

Menghitung Nilai Entropy tiap-tiap atribut :

**Entropy (Total)** dengan rumus sebagai berikut :

$$Entropy (total) = - \sum_{i=1}^n -P_i \times \log_2 P_i$$

$$Entropy(total)$$

$$= \left( -\frac{23}{34} * \log_2 \left( \frac{23}{34} \right) \right) + \left( -\frac{11}{34} * \log_2 \left( \frac{11}{34} \right) \right) = 0.9082$$

*Entropy* (total) adalah menghitung nilai total keputusan diterima (23) dan tidak diterima (11), sedangkan 34 adalah jumlah keseluruhan kasus.

#### 1. *Tangibles*

Untuk menghitung *entropyTangibles* dilihat dari tabel 4.9, nilai *Tangibles* diterima yang memiliki nilai Tinggi berjumlah 11 kasus, *Tangibles* dengan nilai cukup sebanyak 14 kasus dan *Tangibles* dengan nilai rendah adalah sebanyak 9 kasus, berikut adalah *entropy* masing-masing kasus :

*Entropy(B)*

$$= \left( -\frac{12}{14} * \log_2 \left( \frac{12}{14} \right) \right) + \left( -\frac{2}{14} * \log_2 \left( \frac{2}{14} \right) \right) = 0.994$$

*Entropy(C)*

$$= \left( -\frac{11}{14} * \log_2 \left( \frac{11}{14} \right) \right) + \left( -\frac{3}{14} * \log_2 \left( \frac{3}{14} \right) \right) = 0.7496$$

*Entropy(R)*

$$= \left( -\frac{6}{9} * \log_2 \left( \frac{6}{9} \right) \right) \\ + \left( -\frac{3}{9} * \log_2 \left( \frac{3}{9} \right) \right) = 0.9183$$

Untuk kasus pada variabel lainnya lakukan hal yang sama.

## 2. *Reability*

*Entropy(B)*

$$= \left( -\frac{12}{14} * \log_2 \left( \frac{12}{14} \right) \right) \\ + \left( -\frac{2}{14} * \log_2 \left( \frac{2}{14} \right) \right) = 0.5917$$

*Entropy(C)*

$$= \left( -\frac{9}{14} * \log_2 \left( \frac{9}{14} \right) \right) \\ + \left( -\frac{5}{14} * \log_2 \left( \frac{5}{14} \right) \right) = 0.9403$$

*Entropy(R)*

$$= \left( -\frac{2}{6} * \log_2 \left( \frac{2}{6} \right) \right) \\ + \left( -\frac{4}{6} * \log_2 \left( \frac{4}{6} \right) \right) = 0.9183$$

## 3. *Responsiveness*

*Entropy(T)*

$$= \left( -\frac{6}{6} * \log_2 \left( \frac{6}{6} \right) \right) \\ + \left( -\frac{0}{6} * \log_2 \left( \frac{0}{6} \right) \right) = 0$$

*Entropy(C)*

$$= \left( -\frac{15}{17} * \log_2 \left( \frac{15}{17} \right) \right) \\ + \left( -\frac{2}{17} * \log_2 \left( \frac{2}{17} \right) \right) = 0.5226$$

*Entropy(R)*

$$= \left( -\frac{9}{11} * \log_2 \left( \frac{9}{11} \right) \right) \\ + \left( -\frac{2}{11} * \log_2 \left( \frac{2}{11} \right) \right) = 0.971$$

#### 4. Assurance

*Entropy(T)*

$$= \left( -\frac{7}{8} * \log_2 \left( \frac{7}{8} \right) \right) \\ + \left( -\frac{1}{8} * \log_2 \left( \frac{1}{8} \right) \right) = 0.5436$$

*Entropy(C)*

$$= \left( -\frac{14}{21} * \log_2 \left( \frac{14}{21} \right) \right) \\ + \left( -\frac{7}{21} * \log_2 \left( \frac{7}{21} \right) \right) = 0.9183$$

*Entropy(R)*

$$= \left( -\frac{2}{5} * \log_2 \left( \frac{2}{5} \right) \right) \\ + \left( -\frac{3}{5} * \log_2 \left( \frac{3}{5} \right) \right) = 0.971$$

5. *Empathy*

*Entropy(T)*

$$= \left( -\frac{13}{14} * \log_2 \left( \frac{13}{14} \right) \right) \\ + \left( -\frac{1}{14} * \log_2 \left( \frac{1}{14} \right) \right) = 0.3712$$

*Entropy(C)*

$$= \left( -\frac{7}{12} * \log_2 \left( \frac{7}{12} \right) \right) \\ + \left( -\frac{5}{12} * \log_2 \left( \frac{5}{12} \right) \right) = 0.9799$$

$$Entropy(R) = \left( -\frac{3}{8} * \log_2 \left( \frac{3}{8} \right) \right) + \left( -\frac{5}{8} * \right. \\ \left. \log_2 \left( \frac{5}{8} \right) \right) = 0.9544$$

Menghitung nilai gain tiap-tiap atribut:



1. *Gain (Total, Tengibles)*

$$= Entropy(S)$$

$$\begin{aligned} & - \sum_{i=1}^n \frac{|Tengibles_i|}{|Total|} * Entropy(Tengibles_i) \\ & = 0.9082 - \left( \left( \frac{11}{34} * 0.994 \right) + \left( \frac{14}{34} * 0.7496 \right) \right. \\ & \quad \left. + \left( \frac{9}{34} * 0.9183 \right) \right) = 0.0348 \end{aligned}$$

2. *Gain (Total, Reability)*

$$\begin{aligned} & = Entropy(S) - \sum_{i=1}^n \frac{|Reability_i|}{|Total|} * Entropy(Reability_i) \\ & = 0.9082 - \left( \left( \frac{14}{34} * 0.5917 \right) + \left( \frac{14}{34} * 0.9403 \right) \right. \\ & \quad \left. + \left( \frac{6}{34} * 0.9183 \right) \right) = 0.1153 \end{aligned}$$

3. *Gain(Total, Responsiveness)*

$$= Entropy(S)$$

$$\begin{aligned} & - \sum_{i=1}^n \frac{|Responsiveness_i|}{|Total|} * Entropy(Responsiveness_i) \\ & = 0.9082 - \left( \left( \frac{6}{34} * 0 \right) + \left( \frac{17}{34} * 0.5226 \right) \right. \\ & \quad \left. + \left( \frac{11}{34} * 0.971 \right) \right) = 0.3328 \end{aligned}$$

$$4. \text{Gain}(\text{Total}, \text{Assurance})$$

$$= \text{Entropy}(S)$$

$$- \sum_{i=1}^n \frac{|\text{Assurance}_i|}{|\text{Total}|} * \text{Entropy}(\text{Assurance}_i)$$

$$= 0.9082 - \left( \left( \frac{8}{34} * 0.5436 \right) + \left( \frac{21}{34} * 0.9183 \right) + \left( \frac{5}{34} * 0.971 \right) \right) = 0.0703$$

$$5. \text{Gain}(\text{Total}, \text{Empathy})$$

$$= \text{Entropy}(S)$$

$$- \sum_{i=1}^n \frac{|\text{Empathy}_i|}{|\text{Total}|} * \text{Entropy}(\text{Empathy}_i)$$

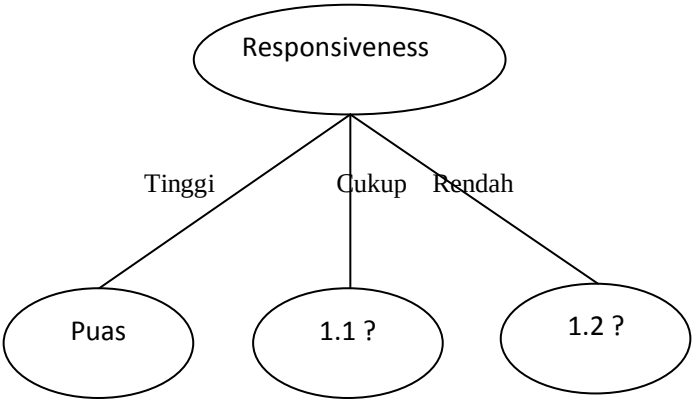
$$= 0.9082 - \left( \left( \frac{14}{34} * 0.3712 \right) + \left( \frac{12}{34} * 0.9799 \right) + \left( \frac{8}{38} * 0.9544 \right) \right) = 0.1849$$

Setelah semua nilai *entropy* dan *gain* dari masing – masing atribut dihitung, kemudian hasil dari perhitungan tersebut dimasukkan ke dalam tabel 4.6.

**Tabel 4.6 Perhitungan Node 1**

| Node | Jumlah Kasus(S) |        | Puas (S1) | Tidak Puas (S2) | Entropy | Info Gain |        |
|------|-----------------|--------|-----------|-----------------|---------|-----------|--------|
| 1    | Total           |        | 34        | 23              | 11      | 0.9082    |        |
|      |                 |        |           |                 |         |           |        |
|      | Tangibles       |        |           |                 |         |           | 0.0348 |
|      |                 | Tinggi | 11        | 6               | 5       | 0.994     |        |
|      |                 | Cukup  | 14        | 11              | 3       | 0.7496    |        |
|      |                 | Rendah | 9         | 6               | 3       | 0.9183    |        |
|      |                 |        |           |                 |         |           |        |
|      | Reability       |        |           |                 |         |           | 0.1153 |
|      |                 | Tinggi | 14        | 12              | 2       | 0.5917    |        |
|      |                 | Cukup  | 14        | 9               | 5       | 0.9403    |        |
|      |                 | Rendah | 6         | 2               | 4       | 0.9183    |        |
|      |                 |        |           |                 |         |           |        |
|      | Responsiveness  |        |           |                 |         |           | 0.3328 |
|      |                 | Tinggi | 6         | 6               | 0       | 0         |        |
|      |                 | Cukup  | 17        | 15              | 2       | 0.5226    |        |
|      |                 | Rendah | 11        | 2               | 9       | 0.971     |        |
|      |                 |        |           |                 |         |           |        |
|      | Assurance       |        |           |                 |         |           | 0.0703 |
|      |                 | Tinggi | 8         | 7               | 1       | 0.5436    |        |
|      |                 | Cukup  | 21        | 14              | 7       | 0.9183    |        |
|      |                 | Rendah | 5         | 2               | 3       | 0.971     |        |
|      |                 |        |           |                 |         |           |        |
|      | Empathy         |        |           |                 |         |           | 0.1849 |
|      |                 | Tinggi | 14        | 13              | 1       | 0.3712    |        |
|      |                 | Cukup  | 12        | 7               | 5       | 0.9799    |        |
|      |                 | Rendah | 8         | 3               | 5       | 0.9544    |        |

Dari perhitungan pada tabel 4.6 diatas dapat diketahui bahwa atribut dengan *gain* tertinggi adalah *Responsiveness* dengan nilai *gain* 0.3328 maka attribut tersebut dijadikan sebagai *node* akar atau *node* 1, dimana nilai atribut cukup dan nilai atribut rendah masih harus dilakukan perhitungan kembali, pohon keputusan dari yang dihasilkan dari *node* 1 dapat dilihat pada gambar 4.3 dibawah ini.



**Gambar 4.2 Pohon Keputusan Hasil Perhitungan Node 1**

Langkah selanjutnya dilakukan penyelesaian untuk menghitung *Node* 1.1 dan *Node* 1.2 sebagai akar untuk nilai cukup dan nilai rendah dari atribut *Empathy*, sama dengan cara yang sebelumnya di atas dengan menghitung nilai *entropy* dari atribut yang tersisa yaitu *Tangibles*, *Reability*, *Assurance* dan *Empathy* dan Status Kerja

setelah *entropy* dihitung, kemudian menghitung *gain* untuk tiap-tiap atribut.

*Entropy* (NilaiResponsiveness Cukup), dengan rumus sebagai berikut :

$$\begin{aligned} \text{Entropy (total)} &= - \sum_{i=1}^n -P_i \times \log_2 P_i \\ &= \left( -\frac{15}{17} * \log_2 \left( \frac{15}{17} \right) \right) + \left( -\frac{2}{17} * \log_2 \left( \frac{2}{17} \right) \right) = 0.5226 \end{aligned}$$

#### 1. AtributTangibles

*Entropy(T)*

$$\begin{aligned} &= \left( -\frac{3}{5} * \log_2 \left( \frac{3}{5} \right) \right) \\ &+ \left( -\frac{2}{5} * \log_2 \left( \frac{2}{5} \right) \right) = 0.957 \end{aligned}$$

*Entropy(C)*

$$\begin{aligned} &= \left( -\frac{7}{7} * \log_2 \left( \frac{7}{7} \right) \right) \\ &+ \left( -\frac{0}{7} * \log_2 \left( \frac{0}{7} \right) \right) = 0 \end{aligned}$$

*Entropy(R)*

$$= \left( -\frac{5}{5} * \log_2 \left( \frac{5}{5} \right) \right) \\ + \left( -\frac{0}{5} * \log_2 \left( \frac{0}{5} \right) \right) = 0$$

2. *Atribut Reability*

*Entropy (T)*

$$= \left( -\frac{6}{6} * \log_2 \left( \frac{6}{6} \right) \right) \\ + \left( -\frac{0}{6} * \log_2 \left( \frac{0}{6} \right) \right) = 0$$

*Entropy (C)*

$$= \left( -\frac{4}{5} * \log_2 \left( \frac{4}{5} \right) \right) \\ + \left( -\frac{1}{5} * \log_2 \left( \frac{1}{5} \right) \right) = 0.72192$$

*Entropy (R)*

$$= \left( -\frac{1}{2} * \log_2 \left( \frac{1}{2} \right) \right) \\ + \left( -\frac{1}{2} * \log_2 \left( \frac{1}{2} \right) \right) = 1$$

3. *Atribut Assurance*

*Entropy(T)*

$$= \left( -\frac{9}{9} * \log_2 \left( \frac{9}{9} \right) \right) \\ + \left( -\frac{0}{9} * \log_2 \left( \frac{0}{9} \right) \right) = 0$$

*Entropy(C)*

$$= \left( -\frac{6}{6} * \log_2 \left( \frac{6}{6} \right) \right) \\ + \left( -\frac{0}{6} * \log_2 \left( \frac{0}{6} \right) \right) = 0$$

*Entropy(R)*

$$= \left( -\frac{0}{2} * \log_2 \left( \frac{0}{2} \right) \right) \\ + \left( -\frac{2}{2} * \log_2 \left( \frac{2}{2} \right) \right) = 0$$

#### 4. Atribut *Empathy*

*Entropy(T)*

$$= \left( -\frac{7}{8} * \log_2 \left( \frac{7}{8} \right) \right) \\ + \left( -\frac{1}{8} * \log_2 \left( \frac{1}{8} \right) \right) = 0.5436$$

*Entropy(R)*

$$= \left( -\frac{6}{7} * \log_2 \left( \frac{6}{7} \right) \right) \\ + \left( -\frac{1}{7} * \log_2 \left( \frac{1}{7} \right) \right) = 0.5917$$

*Entropy(C)*

$$= \left( -\frac{2}{2} * \log_2 \left( \frac{2}{2} \right) \right) \\ + \left( -\frac{0}{2} * \log_2 \left( \frac{0}{2} \right) \right) = 0$$

Menghitung nilai *gain* tiap-tiap atribut :

1. *Gain (Total, Tangibles)*

$$= Entropy(S) \\ - \sum_{i=1}^n \frac{|Tes Tangibles_i|}{|Total|} * Entropy(Tangibles_i) \\ = 0.5226 \\ - \left( \left( \frac{5}{17} * 0.971 \right) \right. \\ \left. + \left( \frac{7}{17} * 0.91829 \right) + \left( \frac{5}{17} * 0 \right) \right) \\ = 0.237$$

2. *Gain (Total, Reability)*



$$\begin{aligned}
&= Entropy(S) \\
&- \sum_{i=1}^n \frac{|Reability_i|}{|Total|} * Entropy(Reability_i) \\
&= 0.5226 - \left( \left( \frac{9}{17} * 0 \right) + \left( \frac{6}{17} * 0 \right) + \left( \frac{2}{17} * 0 \right) \right) \\
&= 0.5226
\end{aligned}$$

### 3. Gain (TotalAssurance)

$$\begin{aligned}
&= Entropy(S) \\
&- \sum_{i=1}^n \frac{|Responsiveness_i|}{|Total|} * Entropy(Responsiveness_i) \\
&= 0.5226 \\
&- \left( \left( \frac{4}{17} * 0.8113 \right) \right. \\
&\quad \left. + \left( \frac{11}{17} * 0.4395 \right) + \left( \frac{2}{17} * 0 \right) \right) \\
&= 0.0473
\end{aligned}$$

### 4. Gain (Total, Empathy)

$$\begin{aligned}
&= Entropy(S) \\
&- \sum_{i=1}^n \frac{|Assurance_i|}{|Total|} * Entropy(Assurance_i)
\end{aligned}$$

$$= 0.61938$$

$$\begin{aligned}
 & - \left( \left( \frac{8}{17} * 0.5436 \right) \right. \\
 & \left. + \left( \frac{7}{17} * 0.5917 \right) + \left( \frac{2}{17} * 0 \right) \right) \\
 & = 0.0231
 \end{aligned}$$

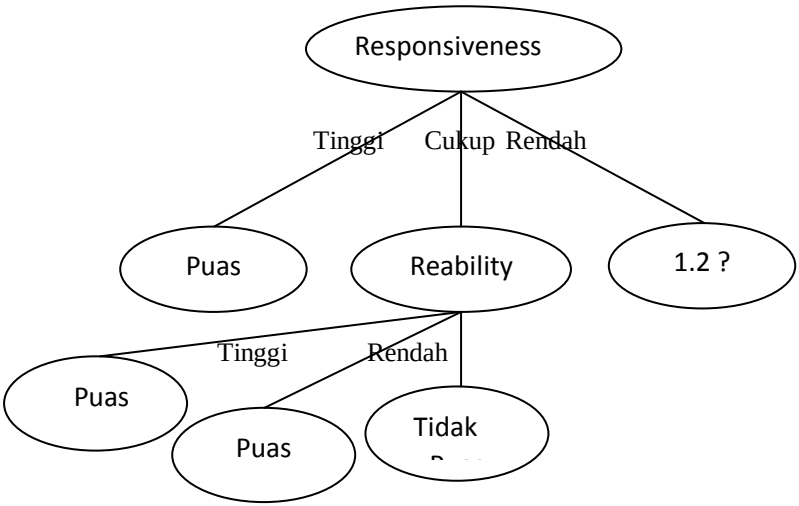
Selanjutnya setelah nilai *entropy* dan *gain* selesai dihitung, maka hasil dari perhitungan tersebut dimasukkan ke dalam tabel 4.11.

**Tabel 4.7 Perhitungan Node 1.1**

| Node | Jumlah Kasus(S) |        | Puas (S1) | Tidak Puas (S2) | Entropy | Info Gain |        |
|------|-----------------|--------|-----------|-----------------|---------|-----------|--------|
| 1    | Total           |        | 17        | 15              | 2       | 0.5226    |        |
|      |                 |        |           |                 |         |           |        |
|      | Tangibles       |        |           |                 |         |           | 0.237  |
|      |                 | Tinggi | 5         | 3               | 2       | 0.971     |        |
|      |                 | Cukup  | 7         | 7               | 0       | 0         |        |
|      |                 | Rendah | 5         | 5               | 0       | 0         |        |
|      |                 |        |           |                 |         |           |        |
|      | Reability       |        |           |                 |         |           | 0.5226 |
|      |                 | Tinggi | 9         | 9               | 0       | 0         |        |
|      |                 | Cukup  | 6         | 6               | 0       | 0         |        |
|      |                 | Rendah | 2         | 0               | 2       | 0         |        |
|      |                 |        |           |                 |         |           |        |
|      | Assurance       |        |           |                 |         |           | 0.0473 |
|      |                 | Tinggi | 4         | 3               | 1       | 0.8113    |        |
|      |                 | Cukup  | 11        | 10              | 1       | 0.4395    |        |
|      |                 | Rendah | 2         | 2               | 0       | 0         |        |
|      |                 |        |           |                 |         |           |        |
|      | Empathy         |        |           |                 |         |           | 0.0231 |
|      |                 | Tinggi | 8         | 7               | 1       | 0.5436    |        |
|      |                 | Cukup  | 7         | 6               | 1       | 0.5917    |        |
|      |                 | Rendah | 2         | 2               | 0       | 0         |        |

Dari hasil tabel 4.7 diatas maka dapat diketahui bahwa atribut *gain* tertinggi yang dijadikan sebagai *Node* 1.1 adalah atribut *Reability*, dengan nilai *gain* 0.5226. Dengan demikian maka dapat diartikan bahwa

Assurance dapat menjadi *node* akar 1.1. Pohon keputusan yang terbentuk dapat dilihat pada gambar 4.4.



**Gambar 4.3 Pohon Keputusan Hasil Perhitungan Node 1.1**

Setelah *Node* 1.1 selesai dihitung maka langkah selanjutnya adalah menghitung nilai rendah atribut *Reability* yang merupakan *Node* 1.2., sama dengan yang sebelumnya yaitu, menghitung *Entropy* dan *Gain* dari masing – masing atribut yang tersisa yaitu, *Tangibles*, *Reability*, *Assurance* dan *Empathy*.

*Entropy* (Nilai *Responsiveness* Rendah), dengan rumus sebagai berikut :

$$Entropy (total) = - \sum_{i=1}^n -P_i \times \log_2 P_i$$

$$= \left( -\frac{2}{11} * \log_2 \left( \frac{2}{11} \right) \right) + \left( -\frac{9}{11} * \log_2 \left( \frac{9}{11} \right) \right) = 0.4174$$

1. *AtributTangibles*

*Entropy(T)*

$$= \left( -\frac{0}{3} * \log_2 \left( \frac{0}{3} \right) \right) \\ + \left( -\frac{3}{3} * \log_2 \left( \frac{3}{3} \right) \right) = 0$$

*Entropy(C)*

$$= \left( -\frac{2}{5} * \log_2 \left( \frac{2}{5} \right) \right) \\ + \left( -\frac{3}{5} * \log_2 \left( \frac{3}{5} \right) \right) = 0.971$$

*Entropy(R)*

$$= \left( -\frac{0}{3} * \log_2 \left( \frac{0}{3} \right) \right) \\ + \left( -\frac{3}{3} * \log_2 \left( \frac{3}{3} \right) \right) = 0$$

2. *Atribut Reability*

*Entropy (T)*

$$= \left( -\frac{0}{2} * \log_2 \left( \frac{0}{2} \right) \right) \\ + \left( -\frac{2}{2} * \log_2 \left( \frac{2}{2} \right) \right) = 0$$

*Entropy (C)*

$$= \left( -\frac{2}{7} * \log_2 \left( \frac{2}{7} \right) \right) \\ + \left( -\frac{5}{7} * \log_2 \left( \frac{5}{7} \right) \right) = 0.8631$$

*Entropy (R)*

$$= \left( -\frac{0}{2} * \log_2 \left( \frac{0}{2} \right) \right) \\ + \left( -\frac{2}{2} * \log_2 \left( \frac{2}{2} \right) \right) = 0$$

### 3. Atribut Assurance

*Entropy(T)*

$$= \left( -\frac{1}{1} * \log_2 \left( \frac{1}{1} \right) \right) \\ + \left( -\frac{0}{1} * \log_2 \left( \frac{0}{1} \right) \right) = 0$$

*Entropy(C)*

$$\begin{aligned} &= \left( -\frac{1}{7} * \log_2 \left( \frac{1}{7} \right) \right) \\ &+ \left( -\frac{6}{7} * \log_2 \left( \frac{6}{7} \right) \right) = 0.5917 \end{aligned}$$

*Entropy(R)*

$$\begin{aligned} &= \left( -\frac{0}{3} * \log_2 \left( \frac{0}{3} \right) \right) \\ &+ \left( -\frac{3}{3} * \log_2 \left( \frac{3}{3} \right) \right) = 0 \end{aligned}$$

4. Atribut *Empathy*

*Entropy(T)*

$$\begin{aligned} &= \left( -\frac{2}{2} * \log_2 \left( \frac{2}{2} \right) \right) \\ &+ \left( -\frac{0}{2} * \log_2 \left( \frac{0}{2} \right) \right) = 0 \end{aligned}$$

*Entropy(R)*

$$\begin{aligned} &= \left( -\frac{0}{4} * \log_2 \left( \frac{0}{4} \right) \right) \\ &+ \left( -\frac{4}{4} * \log_2 \left( \frac{4}{4} \right) \right) = 0 \end{aligned}$$

*Entropy(C)*

$$= \left( -\frac{0}{5} * \log_2 \left( \frac{0}{5} \right) \right) + \left( -\frac{5}{5} * \log_2 \left( \frac{5}{5} \right) \right) = 0$$

Menghitung nilai *gain* tiap-tiap atribut :

1. *Gain (Total, Tangibles)*

$= Entropy(S)$

$$- \sum_{i=1}^n \frac{|Tes Tangibles_i|}{|Total|} * Entropy(Tangibles_i)$$

$= 0.4174$

$$- \left( \left( \frac{3}{11} * 0 \right) + \left( \frac{5}{11} * 0.971 \right) + \left( \frac{3}{11} * 0 \right) \right) = -0.024$$

2. *Gain (Total, Reability)*

$= Entropy(S)$

$$- \sum_{i=1}^n \frac{|Reability_i|}{|Total|} * Entropy(Reability_i)$$



$$= 0.4174$$

$$- \left( \left( \frac{2}{11} * 0 \right) + \left( \frac{7}{11} * 0.8631 \right) + \left( \frac{2}{11} * 0 \right) \right) = -0.1319$$

### 3. *Gain* (TotalAssurance)

$$= Entropy(S)$$

$$- \sum_{i=1}^n \frac{|Responsiveness_i|}{|Total|} * Entropy(Responsiveness_i)$$

$$= 0.4174$$

$$- \left( \left( \frac{1}{11} * 0 \right) + \left( \frac{7}{11} * 0.5917 \right) + \left( \frac{3}{11} * 0 \right) \right) = 0.0409$$

### 4. *Gain* (Total, Empathy)

$$= Entropy(S)$$

$$- \sum_{i=1}^n \frac{|Assurance_i|}{|Total|} * Entropy(Assurance_i)$$

$$= 0.4174$$

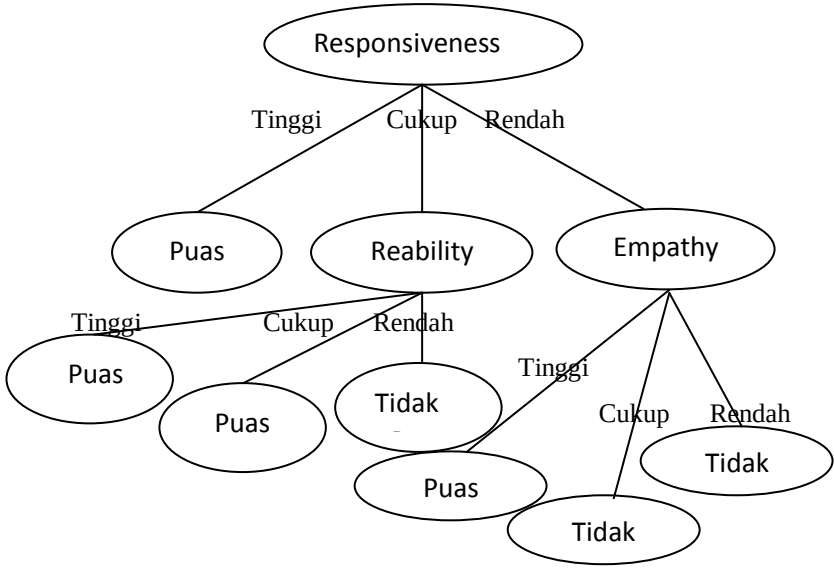
$$- \left( \left( \frac{2}{11} * 0 \right) + \left( \frac{4}{11} * 0 \right) + \left( \frac{5}{11} * 0 \right) \right) = 0.4171$$

Setelah nilai *entropy* dan *gain* dihitung, kemudian hasil dari perhitungan tersebut dimasukkan ke dalam tabel 4.11.

**Tabel 4.8 Perhitungan Node 1.2**

| Node | Jumlah Kasus(S) |        | Puas (S1) | Tidak Puas (S2) | Entropy | Info Gain |
|------|-----------------|--------|-----------|-----------------|---------|-----------|
| 1    | Total           | 11     | 2         | 9               | 0.4174  |           |
|      |                 |        |           |                 |         |           |
|      | Tangibles       |        |           |                 |         | -0.024    |
|      |                 | Tinggi | 3         | 0               | 3       | 0         |
|      |                 | Cukup  | 5         | 2               | 3       | 0.971     |
|      |                 | Rendah | 3         | 0               | 3       | 0         |
|      |                 |        |           |                 |         |           |
|      | Reability       |        |           |                 |         | -0.1319   |
|      |                 | Tinggi | 2         | 0               | 2       | 0         |
|      |                 | Cukup  | 7         | 2               | 5       | 0.8631    |
|      |                 | Rendah | 2         | 0               | 2       | 0         |
|      |                 |        |           |                 |         |           |
|      | Assurance       |        |           |                 |         | 0.0409    |
|      |                 | Tinggi | 1         | 1               | 0       | 0         |
|      |                 | Cukup  | 7         | 1               | 6       | 0.5917    |
|      |                 | Rendah | 3         | 0               | 3       | 0         |
|      |                 |        |           |                 |         |           |
|      | Empathy         |        |           |                 |         | 0.4174    |
|      |                 | Tinggi | 2         | 2               | 0       | 0         |
|      |                 | Cukup  | 4         | 0               | 4       | 0         |
|      |                 | Rendah | 5         | 0               | 5       | 0         |

Dari hasil tabel 4.8 diatas maka dapat diketahui bahwa atribut *gain* tertinggi yang dijadikan sebagai *Node* 1.2 adalah atribut *Empathy*, dengan nilai *gain* 0.4174. Dengan demikian maka dapat diartikan bahwa *Empathy* dapat menjadi *node* akar 1.1. Pohon keputusan yang terbentuk dapat dilihat pada gambar 4.4.



**Gambar 4.4 Pohon Keputusan Hasil Perhitungan Node 1.2**

Adapun aturan atau *rule* yang terbentuk berdasarkan pohon keputusan terakhir seperti pada gambar 4.5 diatas adalah sebagai berikut :

1. IF *Responsiveness* = Baik THEN Keputusan = Puas

2. IF *Responsiveness* = Cukup AND *Reability* = Tinggi  
THEN Keputusan = Puas
3. IF *Responsiveness* = Cukup AND *Reability* = Cukup  
THEN Keputusan = Puas
4. IF *Responsiveness* = Cukup AND *Reability* = Rendah  
THEN Keputusan = Tidak Puas
5. IF *Responsiveness* = Rendah AND *Empathy* = Tinggi  
THEN Keputusan = Puas
6. IF *Responsiveness* = Rendah AND *Empathy* = Cukup  
THEN Keputusan = Tidak Puas
7. IF *Responsiveness* = Rendah AND *Empathy* = Rendah  
THEN Keputusan = Tidak Puas

## **BAB VKESIMPULAN DAN SARAN**

### **5.1. Kesimpulan**

Kesimpulan yang dapat diambil dalam penelitian ini yaitu:

1. Berdasarkan hasil penelitian yang telah dilakukan, didapatkan informasi bahwa variable yang paling berpengaruh terhadap kepuasan mahasiswa adalah *Responsiveness*, *Reability* dan *Empathy*.
2. *Data mining* dengan algoritma C4.5 model *Decision Tree* dapat mengklasifikasi variable – variable dalam mengukur tingkat kepuasan mahasiswa sebagai dasar dalam membuat sebuah keputusan dalam menentukan tindakan apa yang harus dilakukan untuk dapat meningkatkan kepuasan mahasiswa.
3. Hasil implementasi di uji menggunakan salah satu aplikasi *Data mining* algoritma C4.5 yaitu *RapidMiner Studio* dan didapatkan hasil klasifikasi yang sama.

### **5.2. Saran**

Beberapa saran dari hasil penelitian ini yaitu:

1. STMIK Pelita Nusantara Medan dapat menambahkan beberapa kriteria yang berpengaruh dalam mengukur tingkat kepuasan mahasiswa.
2. STMIK Pelita Nusantara Medan dapat menerapkan sistem yang telah dibuat sebagai sebuah sistem

informasi pendukung keputusan dalam meningkatkan pelayanan – pelayanan kepada mahasiswa.

3. Riset selanjutnya agar dapat menambahkan beberapa variabel dalam melakukan klasifikasi dan menerapkan metode yang dapat mengukur akurasi hasil klasifikasi.

## DAFTAR PUSTAKA

- [1] A. I. Jamhur, "Penerapan Data Mining Untuk Menganalisa Jumlah Pelanggan Aktif Dengan Menggunakan Algoritma C4.5," *Maj. Ilm.*, vol. Vol. 23, no. No.2, pp. 12–20, 2016.
- [2] A. Khoirista, "Pengaruh Kualitas Pelayanan Terhadap Kepuasan Pelanggan (Survey Pada Pelanggan Fedex Express Surabaya)," *J. Adm. Bisnis S1 Univ. Brawijaya*, vol. 25, no. 2, pp. 1–7, 2015.
- [3] A. M. Florence.T; and Ms.Savithri.R, "International Journal of Emerging Technologies in Computational and Applied Sciences ( IJETCAS ) Talent Knowledge Acquisition Using C4 . 5 Classification Algorithm," *Int. J. Emerg. Technol. Comput. Appl. Sci.*, vol. 4, no. 4, pp. 406–410, 2013.
- [4] Andri, Y. N. Kunang, and S. Murniati, "Implementasi Teknik Data Mining Untuk Memprediksi Tingkat Kelulusan Mahasiswa Pada Universitas Bina Darma Palembang," *Semin. Nas. Inform. 2013 (semnasIF 2013)*, no. A, pp. 56–63, 2013.
- [5] D. Dicky Nofriansyah, "Penerapan Data Mining dengan Algoritma Naive Bayes Clasifier untuk Mengetahui Minat Beli Pelanggan terhadap Kartu Internet XL ( Studi Kasus di," *Saintikom*, vol. 15, no. 1978–6603, pp. 81–92, 2016.
- [6] D. H. K. S. Hansun, "Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa," *ULTIMATICS*, vol. VI, no. 1, pp. 15–20, 2014.
- [7] D. W. T. Putra, "Algoritma C4.5 untuk Menentukan Tingkat Kelayakan Motor Bekas yang

- Akan Dijual,” *J. TEKNOIF*, vol. 4, no. 1, pp. 16–22, 2016.
- [8] E. Khoirista, Afrinda, Yulianto and M. K. Mawardi, “( Survey Pada Pelanggan Fedex Express Surabaya ),” *J. Adm. Bisnisadministrasi Bisnis*, vol. 25, no. 2, pp. 1–7, 2015.
  - [9] E. S. Siska Haryati, Aji Sudarsono, “Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu),” *J. Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.
  - [10] F. Nasari and S. Darma, “Penerapan K-Means Clustering Pada Data Penerimaan Mahasiswa Baru,” *Semin. Nas. Teknol. Inf. dan Multimed. 2015*, pp. 73–78, 2015.
  - [11] G. Abdillah *et al.*, “Penerapan Data Mining Pemakaian Air Pelanggan Untuk Menentukan Klasifikasi Potensi Pemakaian Air Pelanggan Baru Di Pdam Tirta Raharja Menggunakan Algoritma K-Means,” *Sentika 2016*, vol. 2016, no. Sentika, pp. 18–19, 2016.
  - [12] G. L. A. P. H. Gupta, “Optimization of C4.5 Decision Tree Algorithm for Data Mining Application,” *OInternational J. Emerg. Technol. Adv. Eng.*, vol. 3, no. 3, pp. 341–345, 2013.
  - [13] H. Santoso, I. P. Hariyadi, and Prayitno, “Data Mining Analisa Pola Pembelian Produk Dengan Menggunakan Metode Algoritma Apriori,” *Tek. Inform. ISSN 2302-3805*, no. 1, pp. 19–24, 2016.
  - [14] I. Rahmayuni, “Perbandingan performansi algoritma c4.5 dan cart dalam klasifikasi data nilai mahasiswa prodi teknik komputer politeknik negeri padang,” *Teknoif*, vol. 2, no. 1, pp. 40–46, 2014.
  - [15] J. Eska, “Penerapan Data Mining Untuk Prekdiksi Penjualan Wallpaper Menggunakan Algoritma



- C4.5 STMIK Royal Ksieran,” *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 2, pp. 9–13, 2016.
- [16] K. Tampubolon, H. Saragih, B. Reza, K. Epicentrum, A. Asosiasi, and A. Apriori, “Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-Alat Kesehatan,” pp. 93–106, 2013.
- [17] Selvia Lorena Br Ginting; Wendi Zarman; Ida Hamidah, “Analisi Dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Memprediksi Masa Studi Mahasiswa Berdasarkan Data Nilai Akademik” *Pros. Semin. Nas. Apl. Sains Teknol. 2014 Yogyakarta*, no. November, pp. 263–272, 2014.
- [18] S. Hendrian, “Algoritma Klasifikasi Data Mining Untuk Memprediksi,” vol. 11, no. 3, pp. 266–274, 2018.
- [19] W. Supriyanti, Kusrini, and A. Amborowati, “Perbandingan Kinerja Algoritma c4.5 Dan Naive Bayes Untuk Ketepatan Pemilihan Konsentrasi Mahasiswa,” *J. Inf. Politek. Indonusa*, vol. 1, no. 3, pp. 61–67, 2016.
- [20] Z. Azmi and M. Dahria, “Decision Tree Berbasis Algoritma Untuk,” *Saintikom*, vol. 12, pp. 157–164, 2013.