

# Implementasi Decision Tree Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu

## *Implementation of Decision Tree for Prediction of Student Graduation On Time*

Christin Nandari Dengan\*<sup>1</sup>, Kusrini<sup>2</sup>, Emha Taufiq Luthfi<sup>3</sup>

<sup>1,2,3</sup>Universitas Amikom Jl. Ring Road Utara, Condong Catur Yogyakarta, Tlp (0274) 884 201

<sup>1,2,3</sup>Magister Teknik Informatika, Universitas Amikom, Yogyakarta

e-mail: \*<sup>1</sup>[christin.dengan@students.amikom.ac.id](mailto:christin.dengan@students.amikom.ac.id), <sup>2</sup>[kusrini@amikom.ac.id](mailto:kusrini@amikom.ac.id),

<sup>3</sup>[emhataufiqluthfi@amikom.ac.id](mailto:emhataufiqluthfi@amikom.ac.id)

### **Abstrak**

Fakultas Ilmu Komputer dan Teknologi Informasi (FKTI) merupakan salah satu fakultas baru dalam lingkup Universitas Mulawarman. Mahasiswa yang diterima setiap tahun semakin meningkat, namun tidak semua mahasiswa dapat lulus tepat waktu. Dalam penelitian ini, dengan memanfaatkan data mining dapat mendukung dan membantu mahasiswa untuk memprediksi kelulusan. Adapun tujuan dari penelitian ini adalah untuk membantu mahasiswa untuk memprediksi kelulusan agar dapat meminimalisir mahasiswa yang tidak lulus tepat waktu. Dalam prediksi kelulusan mahasiswa tepat waktu menggunakan data kelulusan mahasiswa tahun 2013 dengan menggunakan atribut jenis kelamin, ipk, predikat kelulusan dan toefl. Penelitian ini, mengimplementasikan decision Tree untuk prediksi kelulusan mahasiswa yang didukung dengan simulasi menggunakan aplikasi WeKa untuk menghitung dan mendapatkan nilai akurasi. Dari hasil simulasi WeKa, maka didapatkan tingkat akurasi penggunaan decision tree terhadap prediksi kelulusan mahasiswa sebesar 60%. Dengan adanya penelitian ini, diharapkan dapat membantu pihak Program Studi Teknik Informatika Universitas Mulawarman agar membimbing mahasiswa sehingga dapat lulus tepat waktu.

**Kata kunci**— *Prediksi Kelulusan, Data Mining, Decision Tree*

### **Abstract**

The Faculty of Computer Science and Information Technology (FKTI) is one of the new faculties within the scope of Mulawarman University. Students are accepted every year, but not all students can graduate on time. In this research, utilizing data mining can support and help students to predict graduation. The purpose of this study is to help students predict graduation in order to minimize students who do not graduate on time. In predicting student graduation on time using 2013 student graduation data using the attributes of gender, GPA, graduation predicate and TOEFL. This study implemented a decision tree for student graduation predictions supported by simulations using the WeKa application to calculate and get accuracy values. From the results of WeKa simulation, the accuracy of decision tree use on student graduation predictions is 60%. With this research, it is expected to help the Mulawarman University Informatics Engineering Study Program to guide students so they can graduate on time.

**Keywords**— *Graduation Prediction, Data Mining, Decision Tree.*

### 1. PENDAHULUAN

Mahasiswa salah satu parameter penting dalam sebuah evaluasi program studi[1]. Kehadiran mahasiswa, prestasi yang dicapai mahasiswa dan profil kelulusan seharusnya mendapatkan perhatian yang serius. Mahasiswa yang diterima setiap tahun semakin banyak , namun tidak sedikit yang dapat menyelesaikan studi dengan tepat waktu[1].

Dalam mencapai sebuah kelulusan, ada tahapan atau proses yang harus dilalui oleh setiap mahasiswa seperti menyelesaikan sejumlah mata kuliah, kuliah kerja nyata, praktik kerja lapangan maupun seminar tugas akhir. Tahapan atau proses tersebut dilakukan dalam waktu yang telah ditentukan oleh pihak universitas [1].

Data tentang mahasiswa yang lulus terus bertambah di tiap tahunnya dan menumpuk seperti data yang terabaikan karena jarang digunakan[2]. Dalam system pendidikan mahasiswa merupakan aset penting dalam sebuah institusi pendidikan, untuk itu perlu diperhatikan tingkat kelulusan mahasiswa tepat waktu. Prediksi kelulusan mahasiswa dapat digunakan lebih lanjut untuk membantu universitas maupun fakultas dalam mengevaluasi dan memperbaiki system pembelajaran sehingga universitas dapat menghasilkan lulusan yang berkualitas[2]

Dari latar belakang tersebut, adapun tujuan dari penelitian ini yaitu menghasilkan penelitian yang membentuk *Decision Support System (DSS)*. Hasil dari analisa diterapkan sebagai dasar perancangan pada system pengambil keputusan untuk mengidentifikasi dan mengklasifikasi kelulusan mahasiswa tepat waktu dan tidak tepat waktu. Manfaat dari penelitian ini yaitu diharapkan dengan adanya penelitian ini dapat menjadi bahan pembelajaran bagi pihak fakultas dalam membimbing mahasiswa agar dapat lulus tepat waktu.

Data terkait mahasiswa yang lulus dapat memberikan informasi yang berguna bagi universitas maupun program studi jika dimanfaatkan secara maksimal [1]. Salah satu cara memanfaatkan data tentang mahasiswa yang lulus yaitu dengan mengolah menjadi *data mining*[2]. *Data mining* merupakan sebuah proses untuk mendapatkan informasi yang berguna dari dari gudang basis data besar[3]. Proses *data mining* dapat ditemukan pola atau aturan yang dapat digunakan untuk menghasilkan suatu informasi seperti prediksi kelulusan mahasiswa[2].

Pemilihan metode sangat berpengaruh dalam melakukan tahap penelitian. *CRISP-DM* menyediakan standar proses baku untuk *data mining* yang dapat diterapkan ke dalam strategi pemecahan masalah umu pada bisnis atau pada unit penelitian[4]. Prediksi kelulusan telah banyak diterapkan dalam beberapa penelitian, salah satunya pada jurnal “*Matching Preprocessing Methods for Improving the Prediction of Student's Graduation*”. Penelitian [5] membahas prediksi kelulusan siswa untuk membantu dalam meningkatkan kualitas system pendidikan dengan membandingkan dua metode preprocessing yaitu algoritma SMOTE dan relief. Dalam penelitian ini menggunakan 544 data set yang diperoleh dari system informasi pendaftaran pada Universitas Teknologi Bangkok. Data set di kelola dengan empat pembelajaran berbasis aturan (DT, OneR, PART dan DTNB), dengan hasil percobaan bahwa DTNB memberikan peningkatan presisi, daya ingat ukuran-f dan rata-rata g dibandingkan dengan metode lain.

*Decision tree* merupakan salah satu algoritma yang dapat digunakan dalam pembentukan pohon keputusan[6]. Dalam penelitian ini menggunakan *decision tree*, karena sesuai kebutuhan dimana hasil akhir berupa keputusan terhadap prediksi kelulusan mahasiswa. *Decision tree* juga memiliki tingkat akurasi yang lebih tinggi saat diaplikasikan untuk jumlah data yang besar dibandingkan algoritma pohon keputusan yang lain[6] dan dapat menghasilkan pohon keputusan yang mudah diinterpretasikan[7].

Dalam penelitian ini akan membahas aspek yang berpengaruh dalam prediksi kelulusan mahasiswa tepat waktu dengan mengimplementasikan *decision tree* dan dilakukan pengujian untuk mengukur tingkat akurasi menggunakan *confusion matrix* pada aplikasi WeKa.

## 2. METODE PENELITIAN

### 2.1 Prosedur Pengumpulan Data

Dataset yang digunakan diperoleh berdasarkan data primer yang didapatkan langsung dari narasumber[8]. Narasumber dalam penelitian ini adalah Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Mulawarman dan bidang kemahasiswaan dan alumni untuk mendapatkan data kelulusan dari *database*.

Tabel 1. Contoh Dataset Kelulusan Mahasiswa

| Nama                     | Jenis Kelamin | IPK  | Predikat         | TOEFL | Keterangan |
|--------------------------|---------------|------|------------------|-------|------------|
| Dwi Kinasih Widiyati     | P             | >3.5 | Cum Laude        | >400  | TW         |
| Fajar Maitari            | P             | >3.5 | Sangat Memuaskan | >400  | TTW        |
| Ketaren, Muhammad Hafizh | L             | >3.5 | Sangat Memuaskan | <400  | TTW        |
| Hadriana Agus Shifa      | L             | >3.5 | Sangat Memuaskan | >400  | TTW        |
| Abraham Ody Watulingas   | L             | >3.5 | Sangat Memuaskan | >400  | TW         |
| Azriana Sari             | P             | >3.5 | Sangat Memuaskan | >400  | TTW        |
| Randi Saputra            | P             | <3.5 | Sangat Memuaskan | >400  | TTW        |
| Danny Agniawan           | L             | <3.5 | Sangat Memuaskan | <400  | TTW        |
| Lara Silvia              | L             | >3.5 | Sangat Memuaskan | <400  | TTW        |

### 2.2 Desain Penelitian

Penulis mengaplikasikan model standarisasi data mining yaitu *CRISP-DM* (*Cross Industry Standar Process for Data Mining*), dengan beberapa tahapan sebagai berikut[4]:

#### 1) Pemahaman Bisnis

Penulis mengedepankan penentuan tujuan proyek dan kebutuhan lengkap dalam memahami proses bisnis atau tujuan bisnis mengenai penelitian tersebut. Setelah itu, penelitian tersebut mampu mengolah sebuah data yang didapatkan dari program studi teknik informatika universitas mulawarman menjadi sebuah formula atau rules.

#### 2) Pemahaman Data

Penelitian ini menggunakan sumber data primer, yaitu data-data yang didapatkan dari *database* FKTI Universitas Mulawarman dan data sekunder, yaitu data-data yang didapatkan secara tidak langsung berupa buku, jurnal dan lain sebagainya.

#### 3) Pengolahan Data

Pengolahan data berlangsung dari data mentah menjadi data bentuk hasil normalisasi. Berikut penjelasan dalam tahapan pengolahan data.

- Data cleaning and data selection*, merupakan tahapan awal dalam memproses *data mining*. Kegiatan ini bertujuan untuk membuang atribut yang tidak dibutuhkan agar lebih efisien.
- Data integration and data transformation*, merupakan tahap selanjutnya dalam proses *data mining* yaitu memindahkan *database* dan memproses transformasi data. Pada data kelulusan mahasiswa, data dari file yang berbeda seperti data kelulusan tahun 2013 disusun menjadi satu kesatuan dalam file.
- Data relection*, yaitu penggunaan data sesuai dengan atribut dan jumlah *record* yang dibutuhkan. Jadi, terdapat beberapa bagian *field* yang perlu dihilangkan karena tidak lengkapnya data atau kurang relevan dalam penelitian.

#### 4) Pemodelan

Tahap awal dilakukan pemrosesan data set seperti, konversi data dan *replacing data*, kemudian dilakukan pemodelan menggunakan algoritma C4.5.

#### 5) Evaluasi

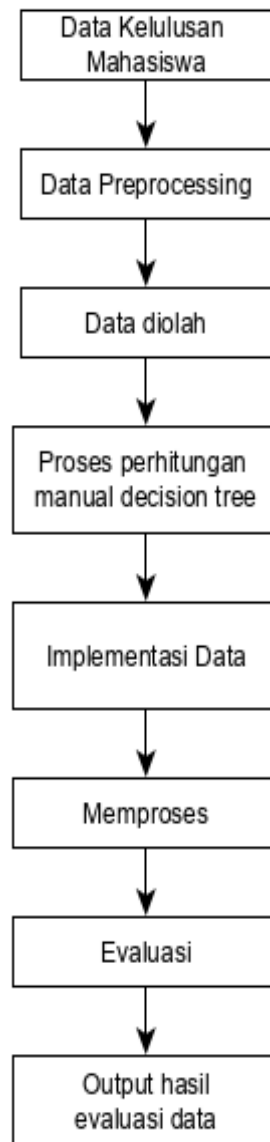
Dilakukan validasi, yaitu pengukuran akurasi dari hasil sebuah pemodelan algoritma yang digunakan dengan memanfaatkan *tools WeKa*.

#### 6) Penyebaran

Penulis menghasilkan penelitian yang membentuk *Decision Support System (DSS)*. Hasil dari analisa diterapkan sebagai dasar perancangan pada system pengambil keputusan untuk mengidentifikasi dan mengklasifikasi kelulusan mahasiswa tepat waktu dan tidak tepat waktu.

### 2.3 Metode yang diusulkan

Desain metode pada penelitian ini menggunakan *decision tree* untuk melakukan perhitungan klasifikasi dan prediksi.



Gambar 1. Metode yang diusulkan

Berdasarkan gambar 1, maka langkah-langkah metode dapat dijelaskan sebagai berikut:

- Data yang akan diimplementasikan sebagai data penelitian adalah data kelulusan pada tahun 2013.
- Data diolah dengan cara pre-processing, yaitu dengan seleksi data, transformasi data dan reduksi data. Pengolahan tersebut dilakukan supaya data terhindar dari *missing value* (data yang kurang lengkap).

- c. Tahap selanjutnya setelah dilakukan data pre-processing, maka akan menghasilkan data *training* yang siap diolah di tahap berikutnya.
- d. Pada tahap ini dilakukan perhitungan manual dengan mulai menghitung *entropy* dan *gain* berdasarkan data pre-processing.
- e. Hasil dari perhitungan tersebut diselaraskan dengan hasil yang dilakukan dalam *tools WeKa*.
- f. Dalam *tools* tersebut dilakukan proses mengolah data sesuai dengan algoritma yang digunakan yaitu algoritma C4.5.
- g. Dilakukan simulasi untuk melihat presentase akurasi data yang di dapat dari algoritma tersebut.
- h. Ouput akan memperlihatkan hasil dari pengolahan data tersebut dalam aplikasi *WeKa*.

### 3. HASIL DAN PEMBAHASAN

Untuk melakukan pengujian model tersebut diuji pada data kelulusan mahasiswa tahun 2013. Data tersebut akan dibuat pohon keputusan untuk menentukan aspek yang berpengaruh dalam prediksi kelulusan mahasiswa.

#### 3.1 Perhitungan Manual

Berikut perhitungan manual prediksi kelulusan mahasiswa tepat waktu di Program Studi Teknik Informatika Universitas Mulawarman.

##### 1. Mengambil sampel data acak

Sampel data acak digunakan sebagai data awal untuk melakukan perhitungan manual decision tree dengan menggunakan data kelulusan tahun 2013 pada Program Studi Teknik Informatika Universitas Mulawarman.

Tabel 2. Data Training Tahap 1

| Nama                     | JK | IPK   | Masa Studi | Predikat         | TOEFL | Ket |
|--------------------------|----|-------|------------|------------------|-------|-----|
| Dwi Kinasih Widiyati     | P  | > 3,5 | 4 tahun    | Cum Laude        | > 400 | TW  |
| Fajar Maitari            | P  | > 3,5 | > 4 tahun  | Sangat Memuaskan | > 400 | TTW |
| Ketaren, Muhammad Hafizh | L  | > 3,5 | > 4 tahun  | Sangat Memuaskan | < 400 | TTW |
| Hadriana Agus Shifa      | L  | > 3,5 | > 4 tahun  | Sangat Memuaskan | > 400 | TTW |
| Abraham Ody Watulingas   | L  | > 3,5 | 4 tahun    | Sangat Memuaskan | > 400 | TW  |
| Azriana Sari             | P  | > 3,5 | > 4 tahun  | Sangat Memuaskan | > 400 | TTW |
| Randi Saputra            | P  | < 3,5 | > 4 tahun  | Sangat Memuaskan | > 400 | TTW |
| Danny Agniawan           | L  | < 3,5 | > 4 tahun  | Sangat Memuaskan | < 400 | TTW |
| Lara Silvia              | L  | > 3,5 | > 4 tahun  | Sangat Memuaskan | < 400 | TTW |
| Umayatul Choerohnur      | P  | > 3,5 | 4 Tahun    | Sangat Memuaskan | > 400 | TW  |

##### 2. Menentukan entropy tahap 1

Setelah menentukan data awal, langkah selanjutnya yaitu melakukan perhitungan dengan mencari nilai entropy dari masing-masing data, menggunakan rumus entropy [9] :

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

Berdasarkan rumus (1) maka dapat dilakukan perhitungan seperti :

$$\left( - \left( \frac{26}{50} \right) * \log_2 \left( \frac{26}{50} \right) + \left( - \left( \frac{24}{50} \right) * \log_2 \left( \frac{24}{50} \right) \right) \right) = 0.998845536$$

Jenis Kelamin (P) = 0.721928095  
 Jenis Kelamin (L) = 0.918295834  
 IPK (>3.5) = 0.877962001  
 IPK (<3.5) = 0

|                    |   |             |
|--------------------|---|-------------|
| Predikat Cumlaude  | = | 0           |
| Predikat Sangat M  | = | 0.979868757 |
| Predikat Memuaskan | = | 0           |

3. Menentukan gain tahap 1

Setelah melakukan perhitungan entropy, langkah selanjutnya yaitu melakukan perhitungan nilai gain dari masing-masing atribut.

$$\text{Gain (A)} = \sum_{i=1}^k \frac{|S_i|}{|S|} \times \text{Entropi}(S_i) \dots\dots\dots(2)$$

Berdasarkan rumus gain, maka diperoleh perhitungan sebagai berikut :

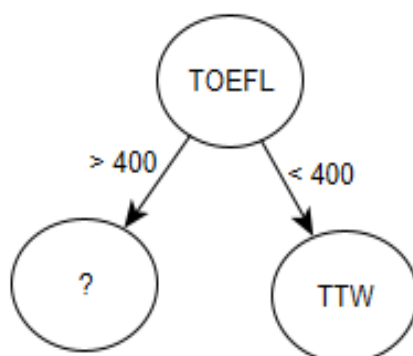
|   |   |          |
|---|---|----------|
| Gain (Jenis Kelamin)  |   |          |
| $0.99884553 - ((20/50 \cdot 0.721928095) + (30/50 \cdot 0.918295834))$        | = | 0.159097 |
| Gain IPK  |   |          |
| $0.99884553 - ((37/50 \cdot 0.877962001) + (13/50 \cdot 0))$                  | = | 0.349154 |
| Gain Predikat   |   |          |
| $0.99884553 - ((11/50 \cdot 0) + (36/50 \cdot 0.979868757) + (3/50 \cdot 0))$ | = | 0.29334  |
| Gain Toefl  |   |          |
| $0.99884553 - ((34/50 \cdot 0.787126586) + (16/50 \cdot 0))$                  | = | 0.463599 |

Setelah ditentukan data awal, langkah selanjutnya yaitu melakukan perhitungan dengan mencari nilai entropy dan gain dalam sebuah table agar lebih rinci dan jelas.

Tabel 3. Node Tree Tahap 1

| Atribut       | Nilai atribut | Total | TW | TTW | Entropy     | Gain        |
|---------------|---------------|-------|----|-----|-------------|-------------|
|               |               | 50    | 26 | 24  | 0.998845536 |             |
| Jenis Kelamin |               |       |    |     |             | 0.159096798 |
|               | P             | 20    | 16 | 4   | 0.721928095 |             |
|               | L             | 30    | 10 | 20  | 0.918295834 |             |
| IPK           |               |       |    |     |             | 0.349153655 |
|               | > 3,5         | 37    | 26 | 11  | 0.877962001 |             |
|               | < 3,5         | 13    | 0  | 13  | 0           |             |
| Predikat      |               |       |    |     |             | 0.293340031 |
|               | Cumlaude      | 11    | 11 | 0   | 0           |             |
|               | Sangat M      | 36    | 15 | 21  | 0.979868757 |             |
|               | Memuaskan     | 3     | 0  | 3   | 0           |             |
| TOEFL         |               |       |    |     |             | 0.463599457 |
|               | > 400         | 34    | 26 | 8   | 0.787126586 |             |
|               | < 400         | 16    | 0  | 8   | 0           |             |

Berdasarkan tabel 3, dapat dilihat bahwa nilai entropy dan gain sudah sesuai aturan antar 0 sampai 1. Setelah dilakukan perhitungan nilai entropy dan gain, maka didapatkan hasil gain tertinggi sebagai akar yaitu nilai gain dari TOEFL sebesar 0,463599457. Hasil pohon keputusan sebagai berikut:



Gambar 2. Pohon Keputusan Node 1

4. Mengambil Data Training Tahap 2

Setelah dilakukan tahap 1 maka, langkah selanjutnya yaitu melakukan perhitungan ulang dengan menggunakan data baru dengan atribut TOEFL dengan nilai >400 karena memiliki nilai gain terbesar pada perhitungan sebelumnya.

Tabel 4. Data Training Tahap 2

| Nama                   | JK | IPK   | Predikat         | TOEFL | Ket |
|------------------------|----|-------|------------------|-------|-----|
| Dwi Kinasih Widiyati   | P  | > 3,5 | Cum Laude        | > 400 | TW  |
| Fajar Maitari          | P  | > 3,5 | Sangat Memuaskan | > 400 | TTW |
| Hadriana Agus Shifa    | L  | > 3,5 | Sangat Memuaskan | > 400 | TTW |
| Abraham Ody Watulingas | L  | > 3,5 | Sangat Memuaskan | > 400 | TW  |
| Azriana Sari           | P  | > 3,5 | Sangat Memuaskan | > 400 | TTW |
| Randi Saputra          | P  | < 3,5 | Sangat Memuaskan | > 400 | TTW |
| Umayatul Choerohnur    | P  | > 3,5 | Sangat Memuaskan | > 400 | TW  |

5. Menentukan Entropy Tahap 2

Perhitungan entropy tahap 2 sebagai berikut :

$$\left(-\left(\frac{25}{34}\right) * \log_2\left(\frac{25}{34}\right) + \left(-\left(\frac{9}{34}\right) * \log_2\left(\frac{9}{34}\right)\right)\right) = 0.833764907$$

|                           |   |             |
|---------------------------|---|-------------|
| Jenis Kelamin (P)         | = | 0.543564443 |
| Jenis Kelamin (L)         | = | 0.981940787 |
| IPK (>3.5)                | = | 0.757878463 |
| IPK (<3.5)                | = | 0           |
| Predikat Cumlaude         | = | 0           |
| Predikat Sangat Memuaskan | = | 0.954434003 |

6. Menentukan Gain Tahap 2

Setelah melakukan perhitungan entropy, dilanjutkan dengan menghitung nilai gain dari masing-masing atribut, sebagai berikut :

$$\text{Gain Jenis Kelamin} = 0.833764907 - ((16/34 * 0.543564443) + (19/34 * 0.981940787)) = 0.029238$$

$$\text{Gain IPK} = 0.833764907 - ((32/34 * 0.757878463) + (2/34 * 0)) = 0.120468$$

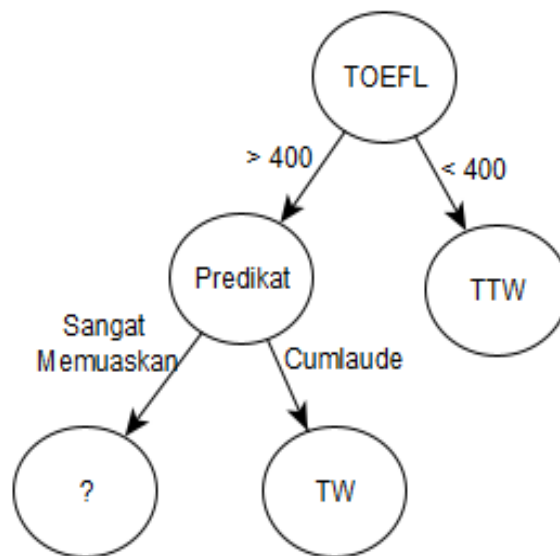
$$\text{Gain Predikat} = 0.833764907 - ((10/34 * 0) + (24/34 * 0.954434003)) = 0.160047$$

Agar lebih memudahkan untuk membaca perhitungan entropy dan gain, disajikan seperti pada tabel 5.

Tabel 5. Nilai Entropy dan Gain Tahap 2

| Atribut       | Nilai Atribut    | Total | TW | TTW | Entropy     | Gain        |
|---------------|------------------|-------|----|-----|-------------|-------------|
| TOEFL         | > 400            | 34    | 25 | 9   | 0.833764907 |             |
| Jenis Kelamin |                  |       |    |     |             | 0.029238259 |
|               | P                | 16    | 14 | 2   | 0.543564443 |             |
|               | L                | 19    | 11 | 8   | 0.981940787 |             |
| IPK           |                  |       |    |     |             | 0.120467531 |
|               | > 3,5            | 32    | 25 | 7   | 0.757878463 |             |
|               | < 3,5            | 2     | 0  | 2   | 0           |             |
| Predikat      |                  |       |    |     |             | 0.160046787 |
|               | Cumlaude         | 10    | 10 | 0   | 0           |             |
|               | Sangat Memuaskan | 24    | 15 | 9   | 0.954434003 |             |

Berdasarkan tabel 5, dapat diketahui bahwa nilai gain tertinggi ada pada atribut predikat dengan nilai 0,160046787. Maka yang akan digunakan sebagai acuan untuk perhitungan selanjutnya adalah predikat dengan nilai sangat memuaskan. Atribut predikat juga digunakan sebagai lanjutan dari pohon keputusan pada gambar 3 sebagai berikut.



Gambar 3. Pohon Keputusan Node 2

7. Menentukan Data Training Tahap 3

Setelah didapatkan pohon keputusan sampai pada tahap 2, maka kembali dilakukan perhitungan ulang untuk menentukan data baru berdasarkan hasil sebelumnya, seperti pada:

Tabel 6. Data Training Tahap 3

| Nama                   | JK | IPK   | Predikat         | TOEFL | Ket |
|------------------------|----|-------|------------------|-------|-----|
| Fajar Maitari          | P  | > 3,5 | Sangat Memuaskan | > 400 | TTW |
| Hadriana Agus Shifa    | L  | > 3,5 | Sangat Memuaskan | > 400 | TTW |
| Abraham Ody Watulingas | L  | > 3,5 | Sangat Memuaskan | > 400 | TW  |
| Azriana Sari           | P  | > 3,5 | Sangat Memuaskan | > 400 | TTW |
| Umayatul Choerohnur    | P  | > 3,5 | Sangat Memuaskan | > 400 | TW  |



8. Menentukan Nilai Entropy Tahap 3  
Perhitungan entropy dilakukan berdasarkan data training tahap 3.

$$\left(-\left(\frac{15}{24}\right) * \log_2\left(\frac{15}{24}\right) + \left(-\left(\frac{9}{24}\right) * \log_2\left(\frac{9}{24}\right)\right)\right) = 0.954434003$$

Jenis Kelamin (P) = 0.918295834  
 Jenis Kelamin (L) = 0.979868757  
 IPK (>3.5) = 0  
 IPK (<3.5) = 0

9. Menentukan Nilai Gain Tahap 3  
Setelah melakukan perhitungan entropy, dilanjutkan perhitungan gain tahap 3 seperti dibawah ini :

Gain Jenis Kelamin  
 $0.954434003 - ((12/24 * 0.918295834) + (12/24 * 0.979868757)) = 0.005351708$

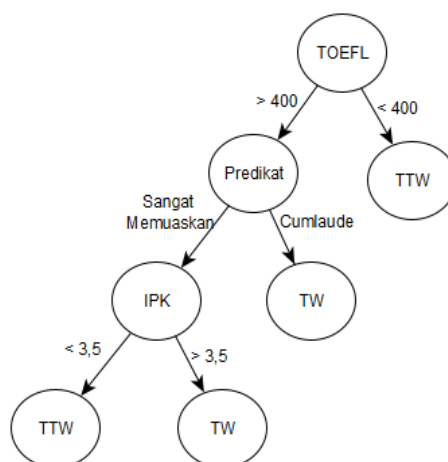
Gain IPK  
 $0.954434003 - ((22/24 * 0) + (2/24 * 0)) = 0.954434003$

Agar lebih memudahkan untuk membaca perhitungan entropy dan gain, disajikan seperti pada tabel 7.

Tabel 7. Nilai Entropy dan Gain Tahap 3

| Atribut       | Nilai atribut    | Total | TW | TTW | Entropy     | Gain        |
|---------------|------------------|-------|----|-----|-------------|-------------|
| Predikat      | sangat memuaskan | 24    | 15 | 9   | 0.954434003 |             |
| Jenis Kelamin |                  |       |    |     |             | 0.005351708 |
|               | P                | 12    | 8  | 4   | 0.918295834 |             |
|               | L                | 12    | 7  | 5   | 0.979868757 |             |
| IPK           |                  |       |    |     |             | 0.954434003 |
|               | >3,5             | 22    | 22 | 0   | 0           |             |
|               | < 3,5            | 2     | 0  | 2   | 0           |             |

Berdasarkan tabel 7, dapat diketahui bahwa nilai gain tertinggi yaitu IPK dengan nilai 0,954434003. Maka atribut IPK dapat dijadikan node selanjutnya, dimana jika IPK < 3,5 akan lulus TTW dan jika IPK > 3,5 akan lulus TW. Berikut hasil pohon keputusan yang dibentuk pada gambar 4.



Gambar 4. Pohon Keputusan

Gambar 4 merupakan hasil akhir dari pohon keputusan setelah dilakukan perhitungan algoritma C4.5. Dari gambar 4 dapat dilihat bahwa TOEFL < 400 maka lulus TTW, namun jika TOEFL > 400 maka akan masuk dalam node predikat dimana jika predikatnya *cumlaude* maka akan lulus TW, jika predikat sangat memuaskan akan masuk dalam node IPK. Jika IPK mahasiswa > 3,5 maka lulus TW, namun jika < 3,5 maka lulus TTW.

### 3.2 Pengujian Tingkat Akurasi Algoritma C4.5

Percobaan yang dilakukan pada implementasi algoritma C4.5 dalam penelitian ini menggunakan aplikasi WeKa. Data yang diolah menggunakan data tahun kelulusan 2013. Atribut yang digunakan yaitu jenis kelamin, ipk, predikat kelulusan, dan toefl. Dari hasil pengujian algoritma C4.5 pada aplikasi WeKa, didapatkan tingkat keberhasilan 100% dan error 0%, seperti pada gambar 5.

|                                  |         |     |   |
|----------------------------------|---------|-----|---|
| Correctly Classified Instances   | 5       | 100 | % |
| Incorrectly Classified Instances | 0       | 0   | % |
| Kappa statistic                  | 1       |     |   |
| Mean absolute error              | 0.1448  |     |   |
| Root mean squared error          | 0.187   |     |   |
| Relative absolute error          | 29.0893 | %   |   |
| Root relative squared error      | 37.5459 | %   |   |
| Total Number of Instances        | 5       |     |   |

Gambar 5. Tingkat Keberhasilan

Berkaitan dengan data training yaitu data kelulusan tahun 2013 yang berisi atribut jenis kelamin, ipk, predikat dan toefl, kemudian dilakukan pemodelan menggunakan *decision tree*. Selanjutnya dilakukan pengujian tingkat akurasi menggunakan *confusion matrix*. *Confusion Matrix* menggunakan tabel matriks untuk melakukan pengujian dan dataset hanya terdiri dari dua kelas yaitu positif dan negative[10]. Berdasarkan pemodelan tersebut diperoleh hasil dalam bentuk *confusion matrix* pada gambar 6.

```

=== Confusion Matrix ===
      a b  <-- classified as
      3 0 | a = TW
      0 2 | b = TTW
    
```

Gambar 6. Confussion Matrix

Didapatkan hasil dari *confusion matrix* yaitu untuk jumlah *true positif* sebanyak 3, jumlah *true negative* sebanyak 0, jumlah *false positif* 0 dan jumlah *false negative* sebanyak 2. Dari hasil tersebut dapat dianalisis bahwa nilai akurasi prediksi kelulusan mahasiswa tepat waktu menggunakan perantara aplikasi WeKA sebesar 60%. Adapun perhitungan manual menentukan akurasinya adalah sebagai berikut:

$$\begin{aligned}
 \text{Akurasi} &= \frac{((TP+TN) / (TP+TN+FP+FN)) * 100\%}{=} \\
 &= \frac{((3+0) / (3+0+0+2)) * 100\%}{=} \\
 &= \frac{(3 / 5) * 100\%}{=} \\
 &= 0.6 * 100\% \\
 &= 60\%
 \end{aligned}$$

#### 4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, maka dapat ditarik kesimpulan bahwa:

1. *Decision tree* dapat diterapkan untuk prediksi kelulusan mahasiswa tepat waktu di Program Studi Teknik Informatika Universitas Mulawarman dengan tingkat akurasi sebesar 60%.
2. Penelitian ini dapat membantu pihak program studi untuk memprediksi kelulusan mahasiswa tepat waktu.

#### 5. SARAN

Berdasarkan hasil penelitian ini, maka peneliti menyarankan beberapa hal untuk penelitian selanjutnya yaitu menggunakan lebih banyak data kelulusan agar tingkat akurasi dapat lebih tinggi dan dapat menambahkan atribut terkait kelulusan mahasiswa tepat waktu.

#### DAFTAR PUSTAKA

- [1] M. Ridwan, "Sistem Rekomendasi Proses Kelulusan Mahasiswa Berbasis Algoritma Klasifikasi C4.5," *J. Ilm. Inform.*, vol. 2, no. 1, pp. 105–111, 2017.
- [2] R. Putri and I. Waspada, "Penerapan Algoritma C4 . 5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika," *Khazanah Inform.*, vol. 4, no. 1, pp. 1–7, 2018.
- [3] P. M. Hasugian, "Penguujian Algoritma Apriori Dengan Aplikasi Weka Dalam," *J. Mantik Penusa*, vol. 1, no. 2, pp. 98–103, 2017.
- [4] A. P. Fadillah, "Penerapan Metode CRISP-DM untuk Prediksi Kelulusan Studi Mahasiswa Menempuh Mata Kuliah (Studi Kasus Universitas XYZ)," *J. Tek. Inform. dan Sist. Inf.*, vol. 1, no. 3, pp. 260–270, 2018.
- [5] P. D. Prachuabsupakij Wanthanee, "Matching Preprocessing Methods for Improving the Prediction of Student ' s Graduation Wanthanee Prachuabsupakij \*, Pafan Doungpaison," pp. 33–37, 2016.
- [6] Kusrini and E. T. Luthfi, *Algoritma Data Mining*. ANDI OFFSET, 2009.
- [7] D. H. Kamagi and S. Hansun, "Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa," *J. Ultim.*, vol. 6, no. 1, pp. 15–20, 2014.
- [8] N. Z. Rahma and A. Setyono, "Penerapan Algoritma C4.5 Dalam Memprediksi Kesiapan Siswa SMP IT PAPB Semarang Menghadapi Ujian Nasional," *Sisfotenika*, vol. 1, 2018.
- [9] R. A. Raharjo, "Kajian Komparasi Penereapan Algoritma C4.5 Neural Network dan SVM Dengan Teknik PSO Untuk Pemilihan Karyawan Teladan PT.XYZ," *J. String*, vol. 1, no. 3, pp. 345–356, 2017.
- [10] R. Geometry and G. Analysis, "Penerapan Algoritma C4.5 Pada Program Klasifikasi Mahasiswa Dropout," pp. 139–147.