

# Prediksi Kelulusan Mahasiswa Berdasarkan Data Berkunjung dan Pinjam Buku di Perpustakaan Menggunakan Metode *C4.5*

Nurhanif<sup>a,1</sup>, Tedy Setiadi<sup>b,2</sup>

<sup>a,b</sup>Program Studi Teknik Informatika Universitas Ahmad Dahlan, Jl. Ringroad Selatan, Kragilan, Tamanan, Kec. Banguntapan, Bantul, Daerah Istimewa Yogyakarta 55191, Indonesia

<sup>1</sup>[nurhanif1500018236@webmail.uad.ac.id](mailto:nurhanif1500018236@webmail.uad.ac.id), <sup>2</sup>[tedy.setiadi@tif.uad.ac.id](mailto:tedy.setiadi@tif.uad.ac.id)

## ABSTRAK

Permasalahan yang dihadapi institusi perguruan tinggi adalah tidak tepatnya waktu kelulusan mahasiswa dan jumlah mahasiswa yang mendaftar tidak sebanding dengan jumlah kelulusan. Tidak tepatnya kelulusan mahasiswa mengakibatkan penumpukan data mahasiswa yang tidak lulus sesuai dengan masa periode kelulusannya, serta dilihat dari data transaksi perpustakaan yang banyak dan bertambah setiap harinya, misalnya data berkunjung dan peminjaman buku. Data tersebut dapat dimanfaatkan untuk mencari pengetahuan dengan mengolahnya hingga menghasilkan suatu pengetahuan penting untuk kedepannya.

Penelitian ini dilakukan untuk mencari sebuah pengetahuan menggunakan teknik *klasifikasi* dengan mengolah data untuk menemukan pola yang terjadi pada data kelulusan dan transaksi perpustakaan. Pengolahan data digunakan untuk memprediksi kelas yang belum diketahui. Teknik klasifikasi yang digunakan *decision tree* dengan penerapan *algoritma C4.5*. Atribut yang digunakan atribut bukan *class* meliputi usia, *toefl*, IPK, berkunjung, pinjam buku dan atribut *class* yaitu keterangan lulus. Dengan tahapan yang dilakukan *load data, cleaning, selection, transformasi*, pemodelan *C4.5, testing/pengujian akurasi*, hasil *rule*, prediksi, dan pengujian akurasi menggunakan *confusion matrix*.

Penelitian ini menghasilkan *rule* untuk memprediksi tepat waktu kelulusan dengan pohon keputusan. Hasil pengujian menggunakan 325 data menghasilkan 53 *rule* dengan tingkat akurasi 94%, presisi 78% dan recall 53%. Manfaat penelitian ini adalah sebagai sumber pengetahuan strategis bagi program studi untuk memprediksi masa studi dan predikat kelulusan mahasiswa.

**Kata Kunci :** Data Kelulusan, *Klasifikasi, Decision Tree, Algoritma C4.5*

## 1. Pendahuluan

Universitas merupakan suatu institusi pendidikan tinggi dan penelitian, yang memberikan gelar akademik dalam berbagai bidang. Begitupun Universitas Ahmad Dahlan, merupakan perguruan tinggi yang ada di Indonesia yang terdiri dari beberapa fakultas, terletak di Yogyakarta dan didirikan pada 18 November 1960. Salah satu permasalahan yang dihadapi institusi perguruan tinggi adalah tidak tepatnya waktu kelulusan mahasiswa. Setiap tahunnya jumlah mahasiswa yang mendaftar tidak sebanding dengan jumlah kelulusannya, dengan bertambah jumlah mahasiswa baru setiap tahunnya maka secara otomatis akan terciptanya data yang banyak dan bertambah melimpahnya data yang tersimpan dalam database.

Perpustakaan Universitas Ahmad Dahlan memiliki data transaksi kunjungan perpustakaan yang banyak dan bertambah setiap harinya. Data tersebut tersimpan didalam database, misalnya data transaksi peminjaman buku dan berkunjung kemudian data tersebut dapat dimanfaatkan untuk mencari sebuah pengetahuan. Berdasarkan hasil survey yang telah dilakukan terhadap kunjungan mahasiswa Teknik Informatika di perpustakaan, dari 20 mahasiswa yang diberikan pertanyaan, 65% mahasiswa diantaranya jarang sekali berkunjung ke perpustakaan. Hal tersebut dikarenakan berbagai alasan, yaitu karena minat baca yang rendah dan waktu luang yang terbatas. Untuk 35% mahasiswa bisa dikatakan aktif dalam berkunjung ke perpustakaan, dan diantaranya didominasi oleh mahasiswa tingkat akhir. Alasan yang yang disebutkan berbeda-beda seperti tempatnya yang nyaman untuk

belajar, banyak sumber referensi, dan tidak perlu mengeluarkan biaya. Penting dilakukannya prediksi kelulusan apakah perpustakaan mempunyai pengaruh yang sangat besar bagi keberlangsungan Mahasiswa dalam menyelesaikan program studinya, karena sebagai sumber pengetahuan strategis bagi program studi untuk memprediksi masa studi dan predikat kelulusan mahasiswa dengan menggunakan teknik data mining.

## 2. Kajian Pustaka

### 2.1. Kajian Terdahulu

Pada penelitian terdahulu yang berjudul “Penerapan algoritma C4.5 untuk klasifikasi predikat kelulusan mahasiswa Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta” (Nugroho Yusuf, 2014). Penelitian ini dilakukan bertujuan untuk menganalisa data yang bertambah banyak didalam database yang dimiliki. Sementara itu, data-data yang melimpah ini bisa dimanfaatkan sebagai sumber informasi strategis bagi program studi untuk memprediksi masa studi dan predikat kelulusan mahasiswa dengan menggunakan teknik teknik data mining. Penelitian ini menggunakan algoritma C4.5 guna mendukung proses penelitian. Penelitian diatas hampir sama kasusnya dengan yang dibuat oleh penulis, perbedaan terletak pada atribut yang digunakan yaitu memakai data transaksi pada perpustakaan seperti berkunjung dan pinjam buku di perpustakaan. Adapun keterbatasan penelitian pada penulis yaitu tidak memakai atribut nilai IPS setiap semester namun hanya memakai nilai IPK akhir dikarenakan pihak Kantor BAA tidak mengijinkan.

### 2.2. Data Mining

*Data mining* adalah proses menemukan korelasi, pola, dan tren baru yang bermakna dengan memilah-milah sejumlah besar data yang tersimpan dalam repositori, dengan menggunakan teknik pengenalan data dengan baik menggunakan teknik statistik dan teknik matematika (Larose, 2005). Beberapa metode yang sering digunakan dalam *data mining* salah satunya adalah metode algoritma C45 yaitu metode pengklasifikasian dengan menggunakan metode probabilistik dan *statistic*.

### 2.3. Tahapan Data Mining

Proses Standar Lintas-Industri untuk Penambangan Data (CRISP-DM) dikembangkan pada tahun 1996 oleh analis yang mewakili DaimlerChrysler, SPSS, dan NCR. CRISP menyediakan proses standar nonproprietary dan tersedia secara bebas untuk memasukkan data mining ke strategi pemecahan masalah umum dari bisnis atau unit penelitian (Larose, 2005). Menurut CRISP-DM, proyek penambangan data yang diberikan memiliki siklus hidup yang terdiri dari enam tahapan antara lain:

1. Tahap pemahaman bisnis

Dalam tahapan ini tentukan tujuan dan persyaratan proyek dengan jelas dalam hal unit bisnis atau penelitian secara keseluruhan, kemudian Terjemahkan sasaran dan pembatasan ini ke dalam rumusan definisi masalah penambangan data dan Siapkan strategi awal untuk mencapai tujuan-tujuan ini.

2. Tahap pemahaman data

Mengumpulkan data, mengevaluasi kualitas data dan menganalisis data eksplorasi untuk membiasakan diri dengan data dan menemukan wawasan awal.

3. Tahap persiapan data

Persiapkan dari data mentah awal, set data akhir yang akan digunakan untuk semua fase berikutnya, kemudian pilih case dan variabel yang akan dianalisis dan melakukan transformasi pada variabel tertentu sehingga siap untuk alat pemodelan.

4. Tahap pemodelan

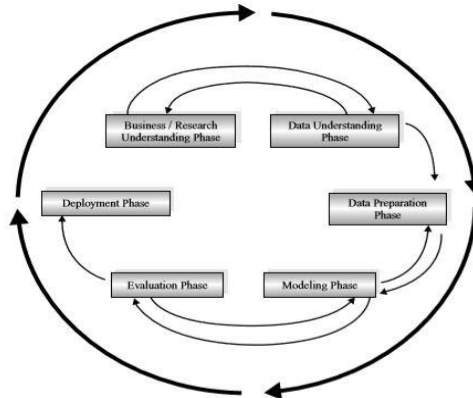
Pilih dan terapkan teknik pemodelan yang sesuai dan Kalibrasi pengaturan model untuk mengoptimalkan hasil.

5. Tahap evaluasi

Mengevaluasi satu atau lebih model yang dikirim dalam fase pemodelan untuk kualitas dan efektivitas sebelum digunakan untuk keperluan di lapangan, kemudian tentukan apakah model tersebut benar-benar mencapai tujuan yang ditetapkan untuknya pada fase pertama.

## 6. Prediksi

Memanfaatkan pemodelan yang telah dibuat. Contoh penyebaran yang lebih kompleks yaitu menerapkan proses penambangan data paralel di departemen lain.



Gambar 1. : Tahapan-tahapan *data mining* (Larose, 2005)

## 2.4. Klasifikasi

Klasifikasi merupakan model (fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui (Han & Kamber, 2012).

## 2.5. Decision Tree

*Decision tree* atau sering disebut dengan pohon keputusan merupakan pohon yang biasanya digunakan untuk mendapatkan jawaban keputusan dari suatu masalah. Dalam pohon keputusan dibagi menjadi 2 bentuk yaitu berupa pohon *biner* dan tidak *biner*. *Decision tree* banyak digunakan untuk menyelesaikan masalah seperti untuk menentukan suatu keputusan dalam berbagai bidang contohnya dalam bidang *diagnosis*, struktur data, dll. Keuntungan menggunakan pohon keputusan ini karena prosedurnya dapat diprediksi (Prasetyo, 2014).

## 2.6. Algoritma C4.5

*Algoritma C4.5* dan pohon keputusan merupakan dua model yang tak terpisahkan, karena untuk membangun sebuah pohon keputusan, dibutuhkan algoritma C4.5. Di akhir tahun 1970 hingga di awal tahun 1980-an, J. Ross Quinlan seorang peneliti di bidang pembelajaran mesin mengembangkan sebuah model pohon keputusan yang dinamakan ID3 (Iterative Dichotomiser), walaupun sebenarnya proyek ini telah dibuat sebelumnya oleh E.B. Hunt, J. Marin, dan P.T. Stone. Kemudian Quinlan membuat algoritma dari pengembangan ID3 yang dinamakan C4.5 yang berbasis supervised learning.

Ada beberapa tahapan dalam membuat sebuah pohon keputusan dalam algoritma C4.5 (Larose, 2005) yaitu :

1. Mempersiapkan data training. Data training biasanya diambil dari data histori yang pernah terjadi sebelumnya atau disebut data masa lalu dan sudah dikelompokkan dalam kelas-kelas tertentu.
2. 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:

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

Keterangan :

S : himpunan kasus  
n : jumlah partisi S  
pi : proporsi dari Si terhadap

Kemudian hitung nilai gain menggunakan rumus:

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

Keterangan:

A : Fitur  
S : Himpunan Kasus  
n : Jumlah partisi himpunan atribut A  
|Si| : Proporsi Si terhadap S

### 3. Metode Penelitian

#### 3.1. Obyek Penelitian

Dalam penelitian ini yang menjadi objek penelitian adalah data kelulusan dan data transaksi perpustakaan Universitas Ahmad Dahlan Yogyakarta, dengan mengambil data kelulusan tahun 2008 – 2014 dengan jumlah data 344 dan data transaksi perpustakaan dari tahun 2010 – 2019 dengan jumlah 2029 data.

#### 3.2. Metode Pengumpulan Data

Pengumpulan data dilakukan untuk memperoleh data atau dokumen yang dibutuhkan untuk penelitian. Adapun metode yang dilakukan yaitu melalui data *private*, wawancara dan studi pustaka. Wawancara dilakukan dengan mahasiswa yang sedang berkunjung di perpustakaan guna untuk mendapatkan informasi dan memperluas data yang diperoleh dari mahasiswanya langsung dan studi pustaka dilakukan dengan mencari serta mempelajari pustaka berupa jurnal, buku, maupun artikel mengenai prediksi masa studi mahasiswa dengan *Algoritma C4.5*.

#### 3.3. Spesifikasi Kebutuhan

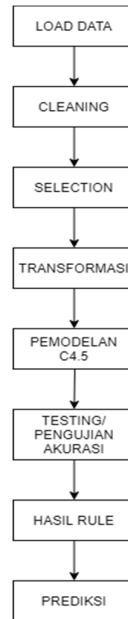
**3.3.1** *Environment* yang digunakan untuk penelitian prediksi kelulusan tepat waktu mahasiswa berdasarkan data perpustakaan Teknik informatika UAD yaitu antara lain:

1. Laptop atau computer
2. Ram 4 GB
3. Processor 7200U 3.1Ghz

**3.3.2.** *Tools* atau *Software* untuk untuk penelitian adalah:

1. *Microsoft Excel* 2013
2. *Sublime Text* 3
3. *Library Decision Tree*
4. *Xampp*
5. Browser : *Google Chrome, Mozila Firefox*
6. Sistem operasi : *Windows 10*

### 3.4. Tahapan Data Mining



Gambar 2. Tahapan Penelitian.

#### 1. Load Data

Proses load data merupakan proses awal pada tahapan penelitian ini, data yang telah dikumpulkan berupa file *excel* kemudian di *load* ke dalam program kemudian diolah.

#### 2. Cleaning

Pada proses cleaning atau pembersihan data mahasiswa yang dilakukan dengan membuang data yang tidak perlu atau tidak akan dipakai untuk prediksi. Data-data tersebut merupakan data yang sifatnya tidak konsisten seperti adanya data yang kosong.

#### 3. Selection

Tahap ini merupakan tahap untuk memutuskan atribut apa saja yang akan digunakan atau dipakai pada tahap data mining. Atribut yang akan digunakan antara lain usia saat masuk, IPK, nilai *toefl*, berkunjung, pinjam buku dan keterangan lulus tepat waktu. Tujuan dilakukan seleksi adalah untuk mendapatkan hasil yang akurat dari data yang digunakan tersebut.

#### 4. Transformasi

Pada penelitian ini transformasi data yaitu mengubah data atribut *predictor* yang berbentuk *alphabet* ke dalam bentuk numerik (angka).

#### 5. Pemodelan C4.5

Proses pada tahap ini merupakan tahap yang paling penting karena data yang sebelumnya telah ditransformasi kemudian akan dimining dengan menggunakan Algoritma C4.5. Tahap pertama dengan menentukan node akar yang akan digunakan untuk percabangan. Pemodelan tersebut mencari node atau akar sampai cabang terakhir tidak bisa dihitung lagi.

#### 6. Testing/Pengujian Akurasi

Dalam melakukan pengujian tingkat akurasi menggunakan tabel *Confusion Matrix*, ada pengukuran kinerja menggunakan *confusion matrix*, terdapat 4 (empat) istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut adalah TP, TN, FP dan FN. *Confusion matrix* dapat disajikan seperti pada Tabel 3.1.

Tabel 1. *Confusion Matrix*

Kelas	Terklasifikasi positif	Terklasifikasi negatif
-------	------------------------	------------------------

Positif	TP ( True Positif )	FN ( False Negative )
Negative	FP ( False Negative )	TN ( True Negative )

Berdasarkan nilai TN, FP, FN dan TP dapat diperoleh nilai akurasi, presisi dan recall. Nilai akurasi menggambarkan seberapa akurat sistem dapat mengklasifikasikan data secara benar. Nilai akurasi dapat diperoleh dengan Persamaan 1. Nilai presisi menggambarkan jumlah data kategori positif yang diklasifikasikan secara benar dibagi dengan total data yang diklasifikasi positif. Presisi dapat diperoleh dengan Persamaan 2. Sementara itu, recall menunjukkan berapa persen data kategori positif yang terklasifikasikan dengan benar oleh sistem. Nilai recall diperoleh dengan Persamaan 3.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \cdot 100\% \quad (1)$$

$$Presisi = \frac{TP}{FP+TP} \cdot 100\% \quad (2)$$

$$Recall = \frac{TP}{FN+TP} \cdot 100\% \quad (3)$$

## 7. Testing/Pengujian Akurasi

Pohon keputusan merupakan hasil analisis pemecahan masalah pengambilan keputusan mengenai kemungkinan atau probabilitas yang akan mempengaruhi keputusan tersebut. Proses dalam pohon keputusan yaitu mengubah bentuk data (tabel) menjadi model pohon (*tree*) kemudian mengubah model pohon tersebut menjadi aturan (*rule*).

## 8. Prediksi

Pada tahap prediksi ini merupakan hasil prediksi dari data set yang sebelumnya sudah dilakukan preprocessing dan pemodelan *C4.5* pada *training* dan *testing* data.

## 4. Hasil dan Pembahasan

### 4.1. Implementasi

#### 1. Load data

Jumlah *load data* awal Mahasiswa Prodi Teknik Informatika yang diambil dari tahun 2008 – 2014 dan data transaksi perpustakaan dari tahun 2010 – 2019 dengan menggunakan “*dataupoad.xls*” berjumlah 344 data.

#### 2. Cleaning

Pada tahap *cleaning* atau pembersihan data, data yang dihapus merupakan data yang belum lengkap atau tidak berisi, dengan data yang diperoleh sebanyak 325 data dari data awal yang berjumlah 344 data.

#### 3. Selection

Pada tahap *selection* atau seleksi data ini tidak semua atribut pada dataset awal digunakan, hanya atribut yang dianggap berpengaruh saja yang akan digunakan untuk penelitian seperti IPK, Toefl, Usia, Berkunjung, Pinjam buku. Data yang sudah diseleksi selanjut adalah tahap *transformasi*.

#### 4. Transformasi

Pada tahap transformasi yaitu data diubah dalam bentuk lain sehingga data tersebut sesuai untuk proses data mining.

Tabel 2. Transformasi

No	Data sebelum transformasi	Data setelah transformasi
1.	IPK < 2.76	Cukup
	IPK 2.76 - 3.00	Memuaskan
	IPK 3.01 - 3.50	Sangat memuaskan
	IPK 3.51 - 4.00	Dengan pujian

2.	Nilai toefl < 420	Dasar
	Nilai toefl 421 - 480	Menengah bawah
	Nilai toefl 481 - 520	Menengah atas
	Nilai toefl > 520	Mahir
3.	Berkunjung > 10	Sering
	Berkunjung 1 – 10	Jarang
4.	Pinjam buku >10	Sering
	Pinjam buku 1 – 10	Jarang
	Pinjam buku 0	Tidak pernah
5.	Atribut usia didapat dari atribut TTL dikurang Tanggal masuk.	-
6.	Masa studi 4 tahun	Tepat waktu
	Masa studi > 4 tahun	Tidak tepat waktu

### 5. Pemodelan C4.5 dan hasil rule

Pada proses pemodelan *C4.5* yaitu memasukkan data sampel disiapkan untuk proses perhitungan. Data tersebut meliputi atribut, jumlah data keseluruhan, jumlah data yang sudah terklasifikasi berdasarkan target yang ditentukan, dalam kasus ini yaitu lulus tepat waktu atau tidak, serta kolom nilai *Entropy* dan *Gain*. Tahapan selanjutnya yaitu penerapan algoritma *C4.5* yaitu menghitung nilai *Entropy* dan *Gain* pada tiap-tiap atribut untuk dijadikan bentuk *Tree*. *Tree* merupakan bentuk aturan klasifikasi yang akan diterapkan pada proses *testing*. Setelah proses perhitungan selesai maka akan menghasilkan sebuah *rule*, dapat dilihat pada gambar 4.1 berikut.

The screenshot shows a web interface for a C4.5 model. At the top, there is a navigation bar with 'Home', 'Train', and 'Predict' buttons, and a 'Logout' link. Below this, a 'Train' section indicates 'Data berhasil diproses.' and has a 'Proses' button. The main content area is titled 'Data Hasil Training' and shows a table with 10 rows of training data. Below the table, there is a 'Rekam Jejak Perhitungan' section for the C4.5 algorithm, which lists various attributes and their corresponding entropy and gain values.

No	Atribut	Parameter	Hasil	Entropi	Gain	Ortu
1	Ipk				0.0623	1
2	Ipk	Cukup	Tidak tepat waktu			1
3	Ipk	Dengan pujian		0.7794		1
4	Usia				0.0557	3
5	Usia	17		0.7025		4
6	Berkunjung				0.4401	5
7	Berkunjung	Jarang	Tidak tepat waktu	0.0		6
8	Berkunjung	Sering	Tepat waktu	0.9183		6
9	Usia	18		0.7367		4
10	Berkunluna				0.1541	9

**Rekam Jejak Perhitungan**  
**C4.5**  
 Banyak semua data: 325  
 Banyak Tepat waktu: 28. Entropi: 0.3047  
 Banyak Tidak tepat waktu: 297. Entropi: 0.1188  
 Entropi total: 0.4235

- Ipk**
  - Cukup**  
 Tidak tepat waktu: 6. Entropi: -0.0  
 Entropi: 0.0
  - Dengan pujian**  
 Tepat waktu: 20. Entropi: 0.4822  
 Tidak tepat waktu: 70. Entropi: 0.282  
 Entropi: 0.7642
  - Memuaskan**  
 Tepat waktu: 1. Entropi: 0.1598  
 Tidak tepat waktu: 30. Entropi: 0.0458  
 Entropi: 0.2056
  - Sangat memuaskan**  
 Tepat waktu: 7. Entropi: 0.1705  
 Tidak tepat waktu: 191. Entropi: 0.0501  
 Entropi: 0.2206  
 Gain: 0.0579
- Toefl**
  - Dasar**  
 Tepat waktu: 14. Entropi: 0.2636  
 Tidak tepat waktu: 192. Entropi: 0.0946  
 Entropi: 0.3582

Gambar 3. Pemodelan C4.5 dan hasil rule.



## 6. Testing atau pengujian akurasi

Pada proses *testing* ini langkah yang dilakukan yaitu memasukkan data uji atau data prediksi. Atribut yang digunakan pada proses testing ini harus sesuai dengan atribut pada proses *training*. Setiap data atribut akan dibandingkan dengan aturan yang sudah terbentuk pada perhitungan data training sebelumnya. Selanjutnya data tersebut akan diklasifikasikan berdasarkan target yang ingin diketahui yaitu data mahasiswa dan transaksi perpustakaan dengan kondisi atribut ini dapat lulus tepat waktu atau tidak. Pembagian data training sebesar 325 data dan data testing sebesar 103 data, dapat dilihat pada gambar 4.2 berikut.

**Test**

Limit

**Data Hasil Testing**

Show  entries

No	NIM	IPK	TOEFL	Usia	Berkunjung	Pinjam Buku	Keterangan Lulus	Hasil Test	Kategori Test
1	1400018053	Dengan pujian	Menengah Bawah	20	Jarang	Tidak Pernah	Tepat waktu	Tepat waktu	True positive
2	1400018068	Dengan pujian	Menengah Atas	17	Sering	Sering	Tepat waktu	Tepat waktu	True positive
3	1400018098	Dengan pujian	Menengah Bawah	18	Jarang	Jarang	Tepat waktu	Tepat waktu	True positive
4	1400018126	Dengan pujian	Dasar	19	Jarang	Tidak Pernah	Tepat waktu	Tepat waktu	True positive
5	1400018135	Dengan pujian	Menengah Bawah	18	Jarang	Tidak Pernah	Tepat waktu	Tidak tepat waktu	False negative
6	1400018146	Dengan pujian	Menengah Bawah	18	Jarang	Tidak Pernah	Tepat waktu	Tidak tepat waktu	False negative
97	12018109	Dengan pujian	Menengah Bawah	18	Jarang	Jarang	Tidak tepat waktu	Tepat waktu	False positive
98	12018118	Dengan pujian	Dasar	18	Jarang	Tidak Pernah	Tepat waktu	Tidak tepat waktu	False negative
99	12018123	Sangat Memuaskan	Dasar	18	Jarang	Jarang	Tidak tepat waktu	Tidak tepat waktu	True negative
100	12018128	Memuaskan	Dasar	17	Sering	Jarang	Tidak tepat waktu	Tidak tepat waktu	True negative

Showing 1 to 100 of 325 entries

Previous     Next

**Rekam Jejak Perhitungan**

**Confusion Matrix**

True positive: 15  
True negative: 293  
False positive: 4  
False negative: 13

Akurasi: 94.7692 %  
Presisi: 78.9474 %  
Recall: 53.5714 %

Gambar 4. Testing data

## 7. Prediksi

Pada menu prediksi menampilkan hasil prediksi 325 data yang sebelumnya sudah dilakukan proses preprocessing dan pemodelan dengan C4.5 pada *training* dan *testing*.

**Predict**

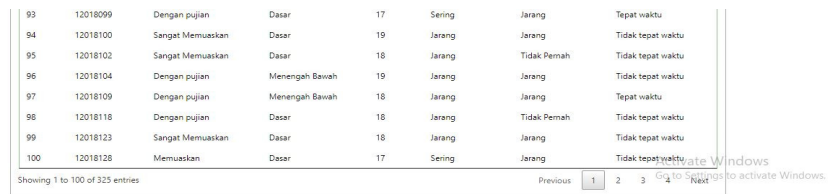
Data berhasil diproses.

**Data Hasil Predict**

Show  entries

No	NIM	IPK	TOEFL	Usia	Berkunjung	Pinjam Buku	Prediksi
1	1400018053	Dengan pujian	Menengah Bawah	20	Jarang	Tidak Pernah	Tepat waktu
2	1400018068	Dengan pujian	Menengah Atas	17	Sering	Sering	Tepat waktu
3	1400018098	Dengan pujian	Menengah Bawah	18	Jarang	Jarang	Tepat waktu
4	1400018126	Dengan pujian	Dasar	19	Jarang	Tidak Pernah	Tepat waktu
5	1400018135	Dengan pujian	Menengah Bawah	18	Jarang	Tidak Pernah	Tidak tepat waktu
6	1400018146	Dengan pujian	Menengah Bawah	18	Jarang	Tidak Pernah	Tidak tepat waktu
7	1400018147	Dengan pujian	Dasar	18	Jarang	Sering	Tepat waktu





93	12018099	Dengan pujian	Dasar	17	Sering	Jarang	Tepat waktu
94	12018100	Sangat Memuaskan	Dasar	19	Jarang	Jarang	Tidak tepat waktu
95	12018102	Sangat Memuaskan	Dasar	18	Jarang	Tidak Pernah	Tidak tepat waktu
96	12018104	Dengan pujian	Menengah Bawah	19	Jarang	Jarang	Tidak tepat waktu
97	12018109	Dengan pujian	Menengah Bawah	18	Jarang	Jarang	Tepat waktu
98	12018118	Dengan pujian	Dasar	18	Jarang	Tidak Pernah	Tidak tepat waktu
99	12018123	Sangat Memuaskan	Dasar	18	Jarang	Jarang	Tidak tepat waktu
100	12018128	Memuaskan	Dasar	17	Sering	Jarang	Tidak tepat waktu

Gambar 5. Hasil prediksi

## 5. Kesimpulan

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

1. Berdasarkan penelitian yang telah dilakukan, maka hasil dari penerapan Algoritma C4.5 dalam memprediksi kelulusan berdasarkan data transaksi kunjungan dan peminjaman buku perpustakaan menghasilkan 53 rule sebagai contoh antara lain:
  - a. Jika IPK (dengan pujian), toefl (dasar), Usia (17), berkunjung (sering), pinjam buku (sering) maka **Tidak tepat waktu**.
  - b. Jika IPK (dengan pujian), toefl (dasar), Usia (17), berkunjung (sering), pinjam buku (jarang) maka **Tepat waktu**.
  - c. Jika IPK (dengan pujian), toefl (dasar), Usia (18), berkunjung (jarang), pinjam buku (tidak pernah) maka **Tidak tepat waktu**.
  - d. Jika IPK (dengan pujian), toefl (menengah bawah), Usia (19), pinjam buku (jarang), berkunjung (jarang) ,maka **Tepat waktu**.
  - e. Jika IPK (dengan pujian), toefl (dasar), Usia (20) maka **Tidak Tepat waktu**.
2. Hasil analisis dari Prediksi kelulusan mahasiswa berdasarkan data kunjungan dan peminjaman buku di perpustakaan menggunakan algoritma C4.5 menghasilkan akurasi sebesar 94%, presisi 78% dan recall 53%.

## Ucapan Terima Kasih

Terima kasih diucapkan kepada Jurnal Sarjana Teknik Informatika (JSTIF) yang telah mengizinkan terbitnya jurnal ini pada bulan Juni 2020.

## Daftar Pustaka

- [1] Y. S. Nugroho, (2014) "Penerapan Algoritma C4.5 Untuk Klasifikasi Predikat Kelulusan Mahasiswa Fakultas Komunikasi Dan Informatika Universitas Muhammadiyah Surakarta," *Pros. Semin. Nas. Apl. Sains Teknol. 2014*, no. November, pp. 1–6.
- [2] D. H. Kamagi and S. Hansun, (2014) "Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa," *J. Ultim.*, vol. 6, no. 1, pp. 15–20.
- [3] Y. S. Nugroho, (2009) "Data Mining Menggunakan Algoritma Naïve Bayes untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro," *J. Semant. 2013*, pp. 1–11.
- [4] Veronica Sri Moertini, (2007), *Pengembangan Skalabilitas Algoritma Klasifikasi C4.5 dengan Pendekatan Konsep Operator Relasi studi kasus: pra pengolahan dan klasifikasi citra batik*. Bandung.
- [5] Hartanto Kamagi David, H. S. (2016) "Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa", *ULTIMATICS*, VI(1). doi: 10.31937/ti.v6i1.327.
- [6] Han, J., Kamber, M. (2012). *Data Mining Concepts and Techniques*, Morgan Kaufmann Publishers, USA.
- [7] Santoso, T.B, (2011) "Analisa dan Penerapan Metode C4.5 untuk Prediksi Loyalitas Pelanggan" *jurnal ilmiah Fak.teknik LIMIT'S*, vol.10.
- [8] Prasetyo, E. (2014) "*Data mining - Mengolah Data menjadi informasi menggunakan matlab*". Edited by Aldo Sahala. Yogyakarta: CV.ANDIOFFSET.
- [9] Kasus, S. et al. (2015) "Implementasi Data Mining untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5". 5. 11(2), pp. 130-138.

- [10] Larose, Daniel T. (2005). *Discovering Knowledge in Data : An Introduction to Data Mining*. John Willey & Sons, Inc.
- [11] Source: <https://uad.ac.id/id/laporan-rektor-hasil-audit-eksternal/laporan-rektor-tahun-2017>.