

**KINERJA ALGORITMA *CLASSIFICATION AND  
REGRESSION TREE (CART)* DALAM  
MENGKLASIFIKASIKAN LAMA MASA STUDI MAHASISWA  
YANG MENGIKUTI ORGANISASI DI UNIVERSITAS NEGERI  
JAKARTA**





**NURUL INDAH PRABAWATI**

**5235122710**




Skripsi ini Ditulis untuk Memenuhi Sebagian Persyaratan dalam Memperoleh  
Gelar Sarjana Pendidikan

**PROGRAM STUDI PENDIDIKAN TEKNIK INFORMATIKA  
DAN KOMPUTER  
FAKULTAS TEKNIK  
UNIVERSITAS NEGERI JAKARTA  
2017**

## HALAMAN PENGESAHAN

NAMA DOSEN	TANDA TANGAN	TANGGAL
Widodo, M.Kom (Dosen Pembimbing I)		24-8-2017
M. Ficky Duskarnaen, ST., M.Sc (Dosen Pembimbing II)		25-8-2017

## PENGESAHAN PANITIA UJIAN SKRIPSI

NAMA DOSEN	TANDA TANGAN	TANGGAL
Prasetyo Wibowo Yunanto, M.Eng (Ketua Penguji)		24-8-2017
Drs. Bachren Zaini, M.Pd (Sekretaris)		24-8-2017
Vina Oktaviani, M.T (Dosen Ahli)		24-8-2017

## HALAMAN PERNYATAAN

Dengan ini saya menyatakan bahwa:

1. Karya tulis skripsi saya yang berjudul Kinerja Algoritma *Classification and Regression Tree (CART)* dalam Mengklasifikasikan Lama Masa Studi Mahasiswa yang Mengikuti Organisasi di Universitas Negeri Jakarta adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik sarjana, baik di Universitas Negeri Jakarta maupun di perguruan tinggi lain.
2. Karya tulis yang berjudul Kinerja Algoritma *Classification and Regression Tree* dalam Mengklasifikasikan Lama Masa Studi Mahasiswa yang Mengikuti Organisasi di Universitas Negeri Jakarta adalah murni gagasan, rumusan, dan penelitian saya sendiri dengan arahan dosen pembimbing.
3. Dalam karya tulis, tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka.
4. Pernyataan saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh karena karya tulis ini, serta sanksi lainnya sesuai dengan norma yang berlaku di Universitas Negeri Jakarta.

Jakarta, 14 Agustus 2017

Yang membuat pernyataan



Nurul Indah Prabawati  
5235122710

## KATA PENGANTAR

Puji dan syukur saya panjatkan kehadirat Allah SWT, shalawat serta salam kita panjatkan kejunjungan kita Nabi Muhammad SAW yang telah memberikan rahmat, karunia, dan hidayah-Nya, sehingga saya dapat menyelesaikan skripsi dengan judul “Kinerja Algoritma *Classification and Regression Trees (CART)* dalam Mengklasifikasikan Lama Masa Studi Mahasiswa yang Mengikuti Organisasi di Universitas Negeri Jakarta”, yang merupakan persyaratan untuk meraih gelar Sarjana Pendidikan Teknik Informatika dan Komputer pada Jurusan Teknik Elektro, Fakultas Teknik, Universitas Negeri Jakarta.

Skripsi tidaklah dapat terwujud dengan baik tanpa adanya bimbingan, dorongan, saran – saran, dan bantuan dari berbagai pihak. Pada kesempatan ini saya ingin menyampaikan ucapan terima kasih kepada:

1. Bapak Widodo, M.Kom selaku pembimbing I dan Bapak M. Ficky Duskarnaen, ST., M.Sc selaku pembimbing II yang telah memberikan motivasi, arahan dan kepercayaan kepada penulis dalam menyelesaikan skripsi ini.
2. Ibu Yuliatri Sastrawidjaya, M.Pd., selaku Ketua Program Studi Pendidikan Teknik Informatika dan Komputer, Jurusan Teknik Elektro, Fakultas Teknik, Universitas Negeri Jakarta.
3. Keluarga besar Bapak Wardjijoso (alm.) dan Ibu Suyati yang tiada hentinya memanjatkan doa dan memberikan semangat kepada penulis dalam menyelesaikan skripsi ini.
4. Kakak – Kakakku yang selalu memberikan dukungan baik moral maupun materi.
5. Kepala Biro Akademik, Kemahasiswaan, dan Hubungan Masyarakat Universitas Negeri Jakarta dan para staff nya yang telah membantu dan memberikan kesempatan kepada peneliti untuk melakukan penelitian.
6. Teman-teman PTIK Reguler 2012, terutama Hanif dan Vina yang selalu mengingatkan dan memberi bantuan, Nisa, Nillam, Sidik, Caca, dan Lintang yang sama-sama berjuang dan senantiasa memberikan dukungan dan semangat kepada penulis.

7. Keluarga besar Kelompok Sosial Pencinta Anak Universitas Negeri Jakarta, Khususnya bidang Sarana dan Prasarana, Bintang, Erika, dan Nurmiya yang telah berbagi suka duka dalam berorganisasi.

Saya menyadari bahwa skripsi masih jauh dari kesempurnaan, karenanya saya mohon maaf apabila terdapat kekurangan dan kesalahan baik dari isi maupun tulisan.

Jakarta, Agustus 2017

Nurul Indah Prabawati

5235122710

**KINERJA ALGORITMA *CLASSIFICATION AND REGRESSION TREE*  
(*CART*) DALAM MENGLASIFIKASIKAN LAMA MASA STUDI  
MAHASISWA YANG MENGIKUTI ORGANISASI DI UNIVERSITAS  
NEGERI JAKARTA**

**NURUL INDAH PRABAWATI**

**ABSTRAK**

Organisasi kemahasiswaan adalah fasilitas yang disediakan oleh perguruan tinggi sebagai wadah untuk mengembangkan kemampuan non akademis, minat dan bakat mahasiswa. Namun, dalam kenyataannya banyak mahasiswa yang mengikuti organisasi mengalami penurunan prestasi hingga tidak dapat lulus tepat waktu. Di Universitas Negeri Jakarta belum adanya sistem yang dapat mengklasifikasikan lama masa studi mahasiswa yang mengikuti organisasi. Sebelum membangun sistem pengambilan keputusan, diperlukan penelitian mengenai akurasi suatu algoritma agar sistem keputusan yang dibuat memiliki tingkat akurasi yang tinggi. Penelitian ini menggunakan algoritma *data mining* yaitu algoritma *Classification and Regression Tree (CART)*. *CART* merupakan metode pohon keputusan biner. *CART* dikembangkan untuk melakukan analisis klasifikasi pada peubah respon baik yang nominal, ordinal, maupun kontinu. Metode klasifikasi *CART* terdiri dari dua metode yaitu metode pohon regresi dan pohon klasifikasi. Data mahasiswa yang mengikuti organisasi yang lulus tepat waktu dan tidak lulus tepat waktu akan diolah menggunakan algoritma *CART*. Setelah diklasifikasikan data tersebut akan dihitung hasil akurasinya menggunakan *K-fold Cross Validation* dengan nilai  $K = 5$ ,  $k = 10$ , dan  $K = 20$ . Berdasarkan hasil contoh data mahasiswa yang mengikuti organisasi menunjukkan bahwa hasil perhitungan akurasi algoritma *CART* terbaik diperoleh ketika nilai  $K = 20$ . Algoritma *CART* telah mampu mengklasifikasikan lama masa studi mahasiswa yang mengikuti organisasi di Universitas Negeri Jakarta. Algoritma *CART* menghasilkan rata-rata akurasi 80%.

Kata Kunci : Kinerja, Data Mining, Organisasi, Akurasi, *CART*.

**PERFORMANCE OF CLASSIFICATION AND REGRESSION TREES  
(CART) IN CLASSIFYING THE LENGTH OF STUDY PERIOD OF  
STUDENT THAT JOINED ORGANIZATION IN JAKARTA STATE  
UNIVERSITY**

**NURUL INDAH P**

**ABSTRACT**

Student organization is a facility provided by the university as a forum to develop non-academic abilities, interests and talents of students. However, in reality many students who has joined the organization have to experience a decrease in achievement and not being able to pass on time. At the State University of Jakarta there is no system that can classify the length of study period of students who joined the organization. To build a decision-making system, research is needed on the accuracy of an algorithm in order to make the decision system has a high degree of accuracy. This research uses data mining algorithm that is Classification and Regression Tree (CART) algorithm. CART is a binary decision tree method. CART was developed to perform classification analysis on both nominal, ordinal, and continuous response variables. The CART classification method consists of two methods, namely the regression tree method and the classification tree. Student data that follow organizations that pass on time and not pass on time will be processed using CART algorithm. After being classified, the data will be calculated using K-fold Cross Validation with  $K = 5$ ,  $k = 10$ , and  $K = 20$  to. Based on the results of student data samples that joined the organization, the best CART algorithm accuracy calculation results obtained when the value of  $K = 20$ . The CART algorithm itself has been able to classify the length of study period of the students who attended the organization at the State University of Jakarta. The CART algorithm yields an average accuracy of 80%.

Keywords : Performance, Data Mining, Organization, Accuracy, CART.

## DAFTAR ISI

<b>HALAMAN JUDUL</b> .....	<b>i</b>
<b>LEMBAR PENGESAHAN</b> .....	<b>ii</b>
<b>LEMBAR PERNYATAAN</b> .....	<b>iii</b>
<b>KATA PENGANTAR</b> .....	<b>iv</b>
<b>ABSTRAK</b> .....	<b>vi</b>
<b>ABSTRACT</b> .....	<b>vii</b>
<b>DAFTAR ISI</b> .....	<b>viii</b>
<b>DAFTAR TABEL</b> .....	<b>xi</b>
<b>DAFTAR GAMBAR</b> .....	<b>xii</b>
<b>DAFTAR LAMPIRAN</b> .....	<b>xiii</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Identifikasi Masalah .....	3
1.3 Pembatasan Masalah .....	4
1.4 Perumusan Masalah.....	4
1.5 Tujuan Penelitian.....	4
1.6 Manfaat Penelitian.....	4
<b>BAB II TINJAUAN PUSTAKA</b> .....	<b>6</b>
2.1 Kerangka Teoritik .....	6
2.1.1 Pengertian Kinerja .....	6
2.1.2 Data Mining .....	6
2.1.3 Klasifikasi.....	8
2.1.4 Decision Tree.....	9
2.1.5 Classification And Regression Trees (CART) .....	11
2.1.5.1. Pengertian Classification And Regression Trees (CART) .....	11
2.1.5.2. Indeks Gini .....	13
2.1.5.3. Tahap – Tahap Pembentukan Algoritma CART .....	13
2.1.5.4. Kelebihan dan Kekurangan CART .....	16



2.1.6 K-fold Cross Validation .....	16
2.1.7 Organisasi.....	17
2.1.8 Organisasi Kemahasiswaan.....	18
2.1.9 Lama Masa Studi.....	19
2.2 Prosedur Penelitian.....	20
2.3 Metode Penelitian.....	20
2.4 Penelitian yang Relevan .....	21
<b>BAB III METODOLOGI PENELITIAN.....</b>	<b>25</b>
3.1 Tempat dan Waktu Penelitian .....	25
3.2 Alat dan Bahan Penelitian .....	25
3.2.1 Alat .....	25
3.2.2 Bahan .....	26
3.3 Diagram Alir Penelitian Penelitian .....	26
3.4 Teknik dan Prosedur Pengumpulan Data .....	28
3.5 Teknik Analisis Data .....	28
<b>BAB IV HASIL PENELITIAN DAN PEMBAHASAN.....</b>	<b>30</b>
4.1 Deskripsi Hasil Penelitian .....	30
4.1.1 Pre-processing Data.....	30
4.1.1.1 Proses Penyortiran Data.....	30
4.1.1.2 Proses Pemangkasan Data .....	31
4.1.1.3 Pemberian Label Pada Data .....	32
4.2 Analisis Data Penelitian .....	33
4.2.1 Pengolahan Data Menggunakan Algoritma CART .....	33
4.2.2 Hasil Akurasi Menggunakan K-Fold Cross Validation.....	35
4.3 Pembahasan .....	38
4.4 Aplikasi Hasil Penelitian .....	38
<b>BAB V KESIMPULAN DAN SARAN .....</b>	<b>39</b>
5.1 Kesimpulan.....	39
5.1 Saran.....	39

<b>DAFTAR PUSTAKA .....</b>	<b>40</b>
<b>LAMPIRAN .....</b>	<b>42</b>

## DAFTAR TABEL

Tabel 2.1 Predikat Program Sarjana .....	20
Tabel 3.1 Spesifikasi Perangkat Keras .....	25
Tabel 3.2 Spesifikasi Perangkat Lunak .....	26
Tabel 3.3 Ilustrasi Tahapan K-fold Cross Validation.....	29
Tabel 4.1 Potongan Data Mentah Sebelum Dipersempit .....	31
Tabel 4.2 Potongan Data Hasil Pemangkasan Data Sesuai Parameter.....	32
Tabel 4.3 Potongan Data yang Sudah diberi Label .....	32
Tabel 4.4 Perbandingan Nilai Akurasi Algoritma CART .....	38

## DAFTAR GAMBAR

Gambar 2.1 Contoh Pemisahan dari Algoritma CART .....	12
Gambar 2.2 Metode Klasifikasi .....	21
Gambar 3.1 Diagram Alir Penelitian .....	26
Gambar 4.1 Potongan Coding untuk Membaca Dataset .....	33
Gambar 4.2 Potongan Coding Pembentukan Node .....	34
Gambar 4.3 Potongan Coding Pembagian Dataset Berdasarkan Nilai K .....	34
Gambar 4.4 Hasil Akurasi Algoritma CART K(5) .....	35
Gambar 4.5 Hasil Akurasi Algoritma CART K(10)... ..	36
Gambar 4.6 Hasil Akurasi Algoritma CART K(20) .....	37

## DAFTAR LAMPIRAN

Lampiran 1 Hasil Rata - Rata Akurasi Algoritma CART .....	42
Lampiran 2 <i>Coding</i> Algoritma CART.....	43
Lampiran 3 Data Mahasiswa Angkatan 2011-2012 .....	46
Lampiran 4 Sampel Data Mentah.....	57
Lampiran 5 Surat Keterangan Penelitian Skripsi .....	63
Lampiran 6 Tentang Penulis.....	64

# **BAB I**

## **Pendahuluan**

### **1.1 Latar Belakang Masalah**

Perguruan tinggi merupakan kelanjutan pendidikan menengah yang diselenggarakan untuk mempersiapkan peserta didik untuk menjadi anggota masyarakat yang memiliki kemampuan akademis dan professional yang dapat menerapkan, mengembangkan dan menciptakan ilmu pengetahuan, teknologi dan kesenian (UU No.2 tahun 1989, Pasal 16 ayat (1)).

Tujuan dari pendidikan tinggi adalah mempersiapkan generasi penerus bangsa agar dapat memajukan bangsa menjadi lebih baik. Peserta didik di perguruan tinggi disebut mahasiswa. Kata mahasiswa sendiri diambil dari kata “maha” yang berarti sesuatu yang lebih tinggi tingkatannya sedangkan siswa adalah pelajar atau orang yang menuntut ilmu. Jadi mahasiswa dapat diartikan sebagai pelajar atau siswa yang tingkatannya lebih tinggi dibandingkan dengan siswa biasa.

Selain kemampuan akademis, mahasiswa juga dituntut untuk memiliki kemampuan-kemampuan non akademis seperti kemampuan bekerjasama, kemampuan dalam memimpin, kemampuan berpikir kritis, rasa tanggung jawab dan lain-lain. Kemampuan non akademis ini tidak dapat dimiliki apabila mahasiswa hanya mengandalkan materi yang diberikan oleh dosen. Oleh karena itu perguruan tinggi menyediakan organisasi kemahasiswaan.

Organisasi kemahasiswaan adalah fasilitas yang disediakan oleh perguruan tinggi sebagai wadah untuk mengembangkan kemampuan non akademis, minat dan bakat mahasiswa. Dengan mengikuti organisasi, mahasiswa diharapkan dapat

meningkatkan kemampuan yang dimilikinya sehingga dapat menjalankan perannya dengan lebih baik.

Di balik semua manfaat positif yang bisa didapatkan oleh mahasiswa, kegiatan kuliah sambil mengikuti organisasi merupakan hal yang sulit dilakukan. Mahasiswa harus dapat membagi waktu dan konsentrasinya serta bertanggung jawab terhadap komitmen dari kedua aktivitas tersebut.

Apabila mahasiswa tidak dapat membagi waktu dan konsentrasinya dengan baik, maka resiko mengalami penurunan prestasi akademik dapat meningkat. Tak jarang pula mahasiswa yang aktif berorganisasi mengalami kegagalan dalam matakuliah dan harus mengulang matakuliah tersebut di semester berikutnya sehingga tidak dapat lulus tepat waktu.

Untuk mengetahui lama masa studi mahasiswa yang mengikuti organisasi, data mahasiswa dapat diklasifikasikan berdasarkan waktu kelulusan. Namun mengklasifikasikan data mahasiswa cukup sulit jika dilakukan secara manual. Universitas Negeri Jakarta belum memiliki sistem yang dapat mengklasifikasikan lama masa studi mahasiswa yang mengikuti organisasi. Sebelum membangun sistem pengambilan keputusan, diperlukan penghitungan algoritmik untuk mengetahui kinerja suatu algoritma agar sistem keputusan yang dibuat memiliki tingkat akurasi yang tinggi.

Data mining merupakan serangkaian proses untuk mendapatkan informasi yang berguna dari data yang besar. Data mining juga dapat diartikan sebagai pernekekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Dalam data mining terdapat banyak jenis teknik, diantaranya yaitu algoritma CART.

*Classification and Regression Trees* (CART) dikembangkan oleh Leo Breiman, Jerome H. Friedman, Richard S. Olshen dan Charles J pada tahun 1984. CART dikembangkan untuk melakukan analisis klasifikasi pada peubah respon baik yang nominal, ordinal, maupun kontinu. Metode klasifikasi CART terdiri dari dua metode yaitu metode pohon regresi dan pohon klasifikasi. Jika variable dependen yang dimiliki bertipe kategorik maka CART akan menghasilkan pohon klasifikasi (*classification tree*). Sedangkan jika variable dependen, yang dimiliki bertipe kontinu atau numerik maka CART akan menghasilkan pohon regresi (*regression tree*).

Data mahasiswa yang mengikuti organisasi kemahasiswaan memiliki banyak atribut yang bertipe kategorik baik nominal maupun ordinal. Sehingga untuk mengolah data tersebut dapat menggunakan algoritma CART. Untuk itu peneliti ingin menguji kinerja algoritma CART dalam mengklasifikasikan lama masa studi mahasiswa yang aktif berorganisasi di Universitas Negeri Jakarta.

## **1.2 Identifikasi Masalah**

Berdasarkan latar belakang di atas, dapat diidentifikasi beberapa masalah sebagai berikut:

1. Universitas Negeri Jakarta belum memiliki aplikasi sistem pembuat keputusan yang dapat memprediksi lama masa studi mahasiswa yang mengikuti organisasi.
2. Sulitnya mengklasifikasikan data mahasiswa secara manual.
3. Diperlukan penghitungan algoritmik untuk mengetahui kinerja suatu algoritma agar sistem keputusan yang dibuat memiliki tingkat akurasi yang tinggi.



### 1.3 Pembatasan Masalah

Pembatasan masalah pada penelitian ini adalah sebagai berikut:

1. Penelitian dilakukan untuk mengetahui kinerja algoritma CART dalam mengklasifikasikan data dan tidak membangun sistem.
2. Data yang diproses adalah data mahasiswa angkatan 2011 dan 2012 yang mengikuti organisasi kemahasiswaan pada tahun 2015.

### 1.4 Rumusan Masalah

Berdasarkan latar belakang, identifikasi dan pembatasan masalah di atas maka permasalahan pada penelitian ini adalah “Bagaimana kinerja algoritma CART untuk mengklasifikasikan lama masa studi mahasiswa yang aktif berorganisasi di Universitas Negeri Jakarta?”

### 1.5 Tujuan Penelitian

Tujuan penelitian ini adalah untuk mengetahui kinerja algoritma CART dalam mengklasifikasikan lama masa studi mahasiswa yang aktif berorganisasi di Universitas Negeri Jakarta.

### 1.6 Manfaat Penelitian

Manfaat yang diharapkan dari hasil penelitian ini adalah sebagai berikut:

1. Bagi mahasiswa
  - a. Menjadi acuan untuk penelitian selanjutnya
  - b. Mengetahui bagaimana penerapan algoritma CART dalam mengklasifikasikan data.
2. Bagi Dosen
  - a. Mengetahui kinerja algoritma CART.
  - b. Sebagai bahan referensi untuk mengajar *data mining*.

### 3. Bagi Universitas

- a. Mengetahui kinerja algoritma CART dalam mengklasifikasikan lama masa studi mahasiswa yang mengikuti organisasi.
- b. Menjadi bahan pertimbangan dalam menentukan algoritma pengklasifikasian data untuk masalah lama masa studi mahasiswa.

## BAB II

### TINJAUAN PUSTAKA

#### 2.1. Kerangka Teoritik

##### 2.1.1 Pengertian Kinerja

Menurut Bernadin dan Russel, kinerja adalah suatu catatan tentang hasil yang diperoleh dari fungsi-fungsi pekerjaan atau aktivitas selama kurun waktu tertentu (Ruky, 2002:15). Menurut Irawan (2002:11), kinerja adalah hasil kerja yang bersifat konkret, dapat diamati, dan dapat diukur. Menurut Nawawi H. Hadari yang dimaksud dengan kinerja adalah hasil dari pelaksanaan suatu pekerjaan, baik yang bersifat fisik/mental maupun non fisik/non mental. Sedangkan menurut Payaman, kinerja adalah tingkat pencapaian atas pelaksanaan tugas tertentu (Simanjuntak, 2005:1).

Berdasarkan teori para ahli tersebut dapat disimpulkan bahwa kinerja adalah suatu hasil dari pelaksanaan tugas atau pekerjaan tertentu yang bersifat konkret, dapat diamati, dan dapat diukur dalam jangka waktu tertentu.

Kinerja algoritma adalah nilai yang dihasilkan sebuah algoritma dalam melaksanakan pekerjaan tertentu. Pengukuran kinerja algoritma dilakukan dengan membandingkan antara hasil prediksi algoritma dengan nilai target variabel data *testing* sebagai data sebenarnya (Faisal, 2017: 69). Secara logika sederhana, kinerja algoritma dapat disimpulkan sebagai berikut:

$$kinerja = \frac{\text{jumlah instance yang diprediksi benar}}{\text{jumlah instance}} \quad (1)$$

##### 2.1.2 Data Mining

*Data mining* sering juga disebut *knowledge discovery in database* (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk

menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. (Santosa, 2007: 10). Data mining adalah istilah yang digunakan untuk menggambarkan proses pemilahan *database* yang besar untuk mencari pola yang menarik dan belum diketahui sebelumnya (Rokach, 2007:2).

Menurut Han dan Kamber (2012:8), *data mining* adalah proses pencarian pola yang menarik dan pengetahuan dari data yang besar. Sumber data yang termasuk diantaranya *database*, *data warehouses*, *theWeb*, repositori informasi lainnya, atau data yang mengalir ke dalam sistem secara dinamis.

*Data mining* adalah suatu istilah yang digunakan untuk menemukan pengetahuan dalam *database*. *Data mining* merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan yang berguna dari *database* yang besar. Menurut Gartner Group, data mining adalah proses rekayasa pola matematis dari kumpulan data yang besar. Pola ini dapat berupa aturan, persamaan, hubungan, kecenderungan, atau model prediksi (Turban dkk., 2005: 263).

Berdasarkan dari teori para ahli tersebut dapat disimpulkan bahwa data mining adalah sebuah teknik yang digunakan untuk mencari pola dan pengetahuan dari data yang besar.

Model pembelajaran data mining dibagi menjadi dua jenis yaitu *supervised learning* dan *unsupervised learning*. *Supervised learning* digunakan untuk memprediksi suatu nilai menggunakan label kelas yang disediakan pada masing-masing *training record*. Sedangkan *unsupervised learning* digunakan untuk mencari struktur intrinsik atau relasi dalam suatu data. Pada model *unsupervised*

*learning* label kelas pada masing-masing *training record* tidak diketahui dan jumlah set kelas yang akan dipelajari mungkin tidak diketahui sebelumnya. (Han dan Kamber, 2012: 330).

*Data Mining* dibagi menjadi beberapa kelompok menurut fungsi dan tugas masing-masing, yaitu:

1. Prediksi ( *Prediction* )

Proses untuk menemukan pola dari data dengan menggunakan beberapa variabel untuk memprediksikan variabel lain yang tidak diketahui jenis atau nilainya.

2. Deskripsi ( *Description* )

Proses untuk menemukan suatu karakteristik penting dari data dalam suatu *database*.

3. Klasifikasi ( *Classification* )

Klasifikasi merupakan suatu proses untuk menemukan model atau fungsi untuk menggambarkan *class* atau konsep dari suatu data. Proses yang digunakan untuk mendeskripsikan data yang penting serta dapat meramalkan kecenderungan data pada masa depan.

4. Asosiasi ( *Association* )

Proses ini digunakan untuk menemukan suatu hubungan yang terdapat pada nilai atribut dari sekumpulan data.

### **2.1.3 Klasifikasi**

Klasifikasi adalah proses mencari sebuah model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep. (Han dan Kamber, 2012: 18). Klasifikasi data memiliki dua langkah proses yang terdiri dari *learning*

*step* di mana model klasifikasi dibangun dan *classification step* di mana model digunakan untuk memprediksi label kelas dari data yang diberikan. (Han dan Kamber, 2012: 328)

Proses klasifikasi memiliki empat komponen yaitu (Gorunescu, 2011: 15) :

1. Kelas

Kelas merupakan variabel terikat yang merepresentasikan label yang terdapat pada objek.

2. *Predictor*

*Predictor* adalah variabel bebas yang direpresentasikan oleh atribut data.

3. *Training Dataset*

*Training Dataset* adalah set data yang digunakan untuk menentukan kelas yang cocok berdasarkan *predictor*.

4. *Testing Dataset*

*Testing Dataset* adalah set data baru yang akan diklasifikasikan oleh model yang telah dibuat.

#### **2.1.4 Decision Tree**

Menurut Han, dkk (2012:330), *decision tree* adalah struktur pohon seperti diagram alir, di mana masing-masing simpul internal (simpul non daun) menunjukkan sebuah tes pada atribut, masing-masing cabang mewakili hasil pengujian, dan setiap *node* daun (atau *node* terminal) memegang label kelas. Simpul paling atas di pohon adalah simpul akar (*root*).

*Decision tree* merupakan metode klasifikasi yang paling populer digunakan. Selain karena pembangunannya relatif cepat, hasil dari model yang dibangun mudah untuk dipahami.

*Decision tree* dapat digunakan untuk menggambarkan model pengklasifikasi dan model regresi. Ketika *decision tree* digunakan untuk mengklasifikasikan data biasanya disebut pohon klasifikasi (*classification tree*). Sedangkan ketika *decision tree* digunakan untuk regresi disebut pohon regresi (*regression tree*) (Rokach, 2007:10).

Pada *decision tree* terdapat 3 jenis node, yaitu:

- a. *Root Node*, merupakan *node* awal yang tidak memiliki input dan bisa tidak mempunyai output atau mempunyai *output* lebih dari satu.
- b. *Internal Node*, merupakan *node* percabangan. *Node* ini hanya memiliki satu input dan memiliki minimal dua *output*.
- c. *Leaf node* atau *terminal node*, merupakan *node* akhir, pada *node* ini hanya terdapat satu input dan tidak mempunyai output.

Persyaratan yang harus dipenuhi dalam menerapkan *decision tree* adalah sebagai berikut:

1. Algoritma *decision tree* merepresentasikan *supervised learning*, oleh karena itu membutuhkan variabel target *preclassified*. Set *training* data harus menyediakan nilai-nilai variabel target.
2. Set *training data* harus bervariasi. *Decision tree* belajar melalui contoh sehingga jika contoh yang tersedia memiliki *subset record* yang kurang didefinisikan dengan baik maka pengklasifikasian dan prediksi untuk *subset* tersebut akan sulit.

3. Kelas atribut target harus diskrit.

Beberapa model *decision tree* yang sudah dikembangkan antara lain IDS, C4.5, dan *CART*. *CART* adalah singkatan dari *classification and regression trees* (Santosa, 2007: 95).

### **2.1.5 Classification And Regression Trees (*CART*)**

#### **2.1.5.1. Pengertian Classification And Regression Trees**

*CART* (Classification and Regression Trees) adalah salah satu metode atau algoritma dari salah satu teknik eksplorasi data *decision tree*. Metode ini dikembangkan oleh Leo Breiman, Jerome H. Fridman, Richard S. Olshen dan Charles J pada tahun 1984 dalam makalah mereka "*Classification and Regression Trees*". (Timofevv, 2005 :7)

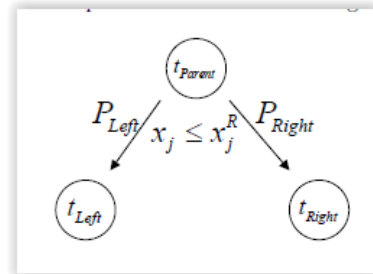
Metode klasifikasi *CART* (Classification And Regression Trees) merupakan metode nonparametrik yang berguna untuk mendapatkan suatu kelompok data yang akurat sebagai pencari dari suatu pengklasifikasian. Metode klasifikasi *CART* terdiri dari dua metode yaitu metode pohon regresi dan pohon klasifikasi. Jika variabel dependen yang dimiliki bertipe kategorik maka *CART* menghasilkan pohon klasifikasi (*classification trees*). Sedangkan jika variabel dependen yang dimiliki bertipe kontinu atau numerik maka *CART* menghasilkan pohon regresi (*regression trees*).

Algoritma *CART* mengembangkan pohon keputusan dengan memilih percabangan yang paling optimal bagi tiap simpul. Pemilihan dilakukan dengan menghitung segala kemungkinan pada tiap variabel.

Algoritma *CART* akan membentuk pohon biner, dimana setiap cabang akan dibagi dua berdasarkan kriteria *split*. Setiap simpul akar memiliki satu variabel



input ( $x$ ) dan titik *split* pada variabel tersebut dengan asumsi variabelnya numerik. *Node* daun dari pohon keputusan *CART* mengandung variabel *output* ( $y$ ) yang digunakan untuk membuat prediksi.



**Gambar 2.1 Contoh *Split* pada Algoritma CART**

Gambar 2.1 adalah contoh *split node root* menjadi cabang keputusan pada algoritma CART.

Menurut Breiman, dkk ada lima ukuran *impurity* yang sering digunakan . tiga diantaranya digunakan untuk pohon klasifikasi dan dua untuk pohon regresi (De'ath dan Fabricius, 2000:3182). Untuk pohon klasifikasi, *impurity* didefinisikan dalam proporsi,  $c$ , dari tanggapan pada masing-masing kategori. Tiga kriteria (indeks) umum adalah:

1. Indeks *Entropy*. Indeks ini identik dengan indeks keragaman Shannon-Weiner, dan membentuk kelompok dengan meminimalkan keragaman dalam kelompok.
2. Indeks Gini. Indeks ini memisahkan kategori besar menjadi grup terpisah, sedangkan indeks informasi cenderung membentuk kelompok yang terdiri lebih dari satu kategori pada awal pemecahan.
3. *Twoing Index*. Indeks ini dapat digunakan untuk lebih dari dua kategori. Indeks ini mendefinisikan dua “kategori super” di setiap perpecahan yang ditetapkan oleh indeks gini.

Sedangkan untuk pohon regresi, terdapat dua kriteria (indeks) umum yaitu:

1. Jumlah kuadrat grup rata-rata (mean). Ini setara dengan model kuadrat terkecil.
2. Jumlah penyimpangan mutlak terhadap median. Ini akan memberikan pohon yang kuat. Namun, untuk data yang didominasi dengan nilai nol (0), kriteria ini bisa jadi tidak efektif, terutama bila variabel penjelas bersifat kategoris.

### 2.1.5.2. Indeks Gini

Indeks gini adalah kriteria berbasis ketidakmurnian yang mengukur perbedaan antara distribusi probabilitas dari nilai atribut target. (Rokach, 2015; 62)

Indeks gini merupakan aturan pemecahan data yang paling banyak digunakan.

Indeks gini pada simpul  $t$  didefinisikan sebagai:

$$i(t) = \sum_{i,j} C(i|j)p(i|t)p(j|t) \quad (2)$$

### 2.1.5.3. Tahap-Tahap Pembentukan Algoritma *CART*

Algoritma *CART* melalui tiga tahapan, yaitu pembentukan pohon klasifikasi, pemangkasan pohon klasifikasi dan penentuan pohon klasifikasi optimum (Sumartini dan Purnami, 2015:2338).

1. Pembentukan pohon klasifikasi

Pada tahap ini, hal pertama yang harus dilakukan adalah menentukan variabel dan *threshold* untuk dijadikan pemilah tiap simpul. Tahapan pembentukan pohon klasifikasi terdiri dari:

- a. Penentuan Pemilah

Pada tahap ini sampel data *learning* akan dipilah untuk mencari himpunan data yang lebih homogen. Data akan dipilah menggunakan indeks gini

karena indeks gini akan selalu memisahkan kelas dengan anggota paling besar/ kelas terpenting dalam simpul terlebih dahulu. persamaan indeks gini adalah:

$$i(t) = \sum_{i,j-1} p(j|t)p(i|t), i \neq j \quad (3)$$

Dengan  $p(j|t)$  adalah proporsi kelas  $j$  pada simpul  $t$ . Pemilahan yang terpilih akan membentuk suatu himpunan kelas yang disebut dengan simpul. Simpul tersebut akan melakukan pemilahan secara rekursif sampai diperoleh *terminal nodes*. Tahap berikutnya adalah menentukan kriteria *goodness of split* untuk mengevaluasi pemilah dari pemilah  $s$  pada simpul  $t$  dengan persamaan:

$$\phi(s, t) = \Delta i(s, t) = i(t) - P_L i(t_L) - P_R i(t_R) \quad (4)$$

$t_L$ = cabang kiri dari nokhtah keputusan  $t$

$t_R$ = cabang kanan dari nokhtah keputusan  $t$

Pemilah yang menghasilkan  $\phi(s, t)$  lebih tinggi merupakan pemilah terbaik.

#### b. Penentuan Simpul Terminal

Pengembangan pohon akan berhenti ketika jumlah pengamatan pada simpul kurang dari standar yang telah ditentukan ( $N_{min}$ ) atau ketika sudah mencapai batasan tingkat kedalaman (*depth*) dalam pohon maksimal.

#### c. Penandaan Label Kelas

Label kelas pada simpul terminal ditentukan berdasarkan aturan jumlah terbanyak, yaitu jika:

$$p(j_0|t) = \max_j \frac{N_j(t)}{N(t)} \quad (5)$$

Label kelas untuk simpul terminal  $t$  adalah  $j_0$  yang memberikan nilai dugaan kesalahan pengklasifikasian pada simpul  $t$  yang paling kecil sebesar:

$$r(t) = 1 - \max_j p(j|t). \quad (6)$$

## 2. Pemangkasan pohon klasifikasi

Pemangkasan pohon dilakukan untuk mencegah *overfitting*. *Overfitting* adalah situasi di mana algoritma menghasilkan pengklasifikasian yang sangat sesuai dengan data *training* namun kehilangan kemampuan untuk menggeneralisasikan kejadian yang tidak direpresentasikan selama pelatihan (Rokach, 2007: 57). Pemangkasan pohon dapat dilakukan dengan pengukuran *cost complexity minimum* dengan rumus:

$$R_\alpha(T) = R(T) + \alpha |\tilde{T}| \quad (7)$$

$R_\alpha(t)$  merupakan kombinasi linear biaya dan kompleksitas pohon yang dibentuk dengan menambahkan *cost penalty* bagi kompleksitas terhadap biaya kesalahan klasifikasi pohon. Selanjutnya, dilakukan pencarian pohon bagian  $T(a) < T_{max}$  yang meminimumkan  $R_\alpha(t)$  yaitu:

$$R_\alpha(T(a)) = \min_{T < T_{max}} R_\alpha(t) \quad (8)$$

## 3. Penentuan pohon klasifikasi optimum

Pohon klasifikasi yang terbentuk mungkin akan memiliki tingkat kompleksitas yang tinggi. Oleh karena itu perlu dioptimalkan terlebih dahulu sebelum digunakan untuk mengklasifikasikan data baru. Optimasi pohon akan memilih ukuran pohon yang tepat dan memotong *nodes* yang tidak signifikan. Salah satu algoritma pemangkasan yang dapat digunakan adalah *cross validation*.

#### 2.1.5.4. Kelebihan dan Kekurangan *CART*

Algoritma *CART* memiliki kelebihan sebagai berikut (Timofeev, 2005: 19-24):

1. *CART* adalah metode non parametrik. Oleh karena itu *CART* tidak membutuhkan specification of any functional form.
2. *CART* tidak harus menentukan variabel di awal proses. Algoritma *CART* akan memilih sendiri variable yang akan paling signifikan dan membuang variabel yang tidak signifikan.
3. *CART* dapat dengan mudah menangani *outliers*. *Outliers* adalah data yang memiliki karakteristik unik yang terlihat berbeda jauh dari observasi-observasi lainnya dan muncul dalam bentuk nilai ekstrim baik untuk sebuah variabel tunggal atau variabel kombinasi.
4. *CART* tidak memiliki asumsi dan memiliki komputasi yang cepat.
5. *CART* fleksibel dan memiliki kemampuan untuk menyesuaikan diri sesuai kebutuhan.

Sedangkan kekurangan Algoritma *CART* adalah:

1. *CART* mungkin akan menghasilkan pohon keputusan yang tidak stabil.
2. *CART* membagi data hanya berdasarkan satu variabel.

#### 2.1.6 *K-fold Cross Validation*

*Cross Validation* adalah metode statistik untuk mengevaluasi dan membandingkan algoritma belajar dengan membagi data menjadi dua bagian yaitu data *training* dan data *test*. Bentuk sederhana dari *cross validation* adalah *k-fold cross validation* (Refaeilzadeh, dkk 2009 : 532).

*K-fold Cross Validation* adalah salah satu cara menemukan parameter terbaik dari satu model dengan cara menguji besarnya *error* pada data *test*. Dalam *cross validation* data dibagi menjadi  $k$  - subset dengan ukuran yang sama. Kemudian subset tersebut akan dibagi menjadi *data train* dan *data test* (Santosa, 2007: 220).

Tujuan dari *cross validation* adalah untuk mendefinisikan dataset untuk "menguji" model dalam tahap pelatihan (yaitu, *validasi data*), dalam rangka untuk membatasi masalah seperti terjadinya *overfitting*, memberikan wawasan tentang bagaimana model akan menggeneralisasi independen dataset dan lain-lain.

### **2.1.7 Organisasi**

Pengertian organisasi adalah perkumpulan atau wadah bagi sekelompok orang untuk bekerjasama, terkendali dan dipimpin untuk tujuan tertentu (Hidayah, 2014). Organisasi biasanya memanfaatkan suatu sumber daya tertentu misalnya lingkungan, cara atau metode, material, mesin, uang, dan beberapa sumberdaya lain dalam rangka mencapai tujuan organisasi tersebut.

Menurut para ahli terdapat beberapa pengertian organisasi sebagai berikut:

(1). Organisasi Menurut Stoner

Organisasi adalah suatu pola hubungan-hubungan yang melalui mana orang-orang di bawah pengarahan manajer mengejar tujuan bersama.

(2). Organisasi Menurut James D. Mooney

Organisasi adalah bentuk setiap perserikatan manusia untuk mencapai tujuan bersama.

(3). Organisasi Menurut Chester I. Bernard

Organisasi merupakan suatu sistem aktivitas kerja sama yang dilakukan oleh dua orang atau lebih.

Berdasarkan teori para ahli tersebut, dapat disimpulkan bahwa organisasi adalah wadah bagi sekelompok orang untuk bekerjasama dalam mencapai tujuan tertentu.

Ciri-ciri organisasi menurut Siwanto (2007: 73) yaitu :

- (1). Suatu organisasi adalah adanya sekelompok orang yang menggabungkan diri dengan suatu ikatan norma, peraturan, ketentuan dan kebijakan yang telah dirumuskan dan masing-masing pihak siap untuk menjalankannya dengan penuh tanggung jawab.
- (2). Dalam suatu organisasi yang terdiri atas sekelompok orang tersebut saling mengadakan hubungan timbal balik, saling memberi dan menerima dan juga saling bekerjasama untuk melahirkan dan merealisasikan maksud (*purpose*), sasaran (*objective*) dan tujuan (*goal*).
- (3). Dalam suatu organisasi yang terdiri atas sekelompok orang yang saling berinteraksi dan bekerjasama tersebut diarahkan pada suatu titik tertentu., yaitu tujuan bersama dan ingin direalisasikan.

### **2.1.8 Organisasi Kemahasiswaan**

Organisasi kemahasiswaan merupakan bentuk kegiatan di perguruan tinggi yang diselenggarakan dengan prinsip dari, oleh dan untuk mahasiswa (Sukirman, 2004:72, diacu dalam Pratiwi, 2017: 56). Organisasi tersebut merupakan wahana dan sarana pengembangan diri mahasiswa ke arah perluasan wawasan peningkatan ilmu dan pengetahuan, serta integritas kepribadian mahasiswa. Organisasi kemahasiswaan juga sebagai wadah pengembangan kegiatan ekstrakurikuler mahasiswa diperguruan tinggi yang meliputi pengembangan penalaran, keilmuan, minat, bakat dan kegemaran mahasiswa itu sendiri (Paryati Sudarman, 2004:34-

35). Menurut Kepmendikbud RI. No. 155/U/1998 Tentang Pedoman Umum Organisasi Kemahasiswaan di Perguruan Tinggi, organisasi kemahasiswaan intra-perguruan tinggi adalah wahana dan sarana pengembangan diri mahasiswa ke arah perluasan wawasan dan peningkatan kecendikiaan serta integritas kepribadian untuk mencapai tujuan pendidikan tinggi.

Sedangkan menurut Silvia Sukirman (2004:69), organisasi kemahasiswaan adalah kegiatan tidak wajib atau pilihan yang penting diikuti oleh setiap mahasiswa selama studinya sehingga melengkapi hasil belajar secara utuh. Berdasarkan pendapat para ahli tersebut, dapat disimpulkan bahwa organisasi kemahasiswaan adalah wadah yang disediakan oleh kampus untuk mahasiswa sebagai tempat mengembangkan keterampilan yang dimilikinya.

#### **2.1.9 Lama Masa Studi**

Mahasiswa yang telah menyelesaikan keseluruhan beban program studi yang telah ditetapkan dapat dipertimbangkan dalam penetapan kelulusan dan predikat studi akhir program. Untuk menentukan kelulusan dan predikat, fakultas menggunakan dokumen/format resmi hasil penilaian studi mahasiswa yang sudah dicek dan disahkan kebenaran isinya oleh ketua jurusan setelah diperiksa oleh tim verifikasi pada masing-masing jurusan. Kelulusan dan predikat ditetapkan berdasarkan Indeks Prestasi Kumulatif (IPK) hasil ujian semua matakuliah yang diwajibkan untuk menyelesaikan kebulatan program yang diikuti sesuai dengan ketentuan struktur program dan sebaran Sistem Kredit Semester (SKS), termasuk hasil Ujian Skripsi/Tugas Akhir. Predikat program sarjana dan diploma tercantum dalam tabel 2.1.



**Tabel 2.1 Predikat Program Sarjana**

<b>Predikat</b>	<b>IPK</b>
Terpuji	3,51 – 4,00
Sangat Memuaskan	2,76 – 3,50
Memuaskan	2,00 – 2,75

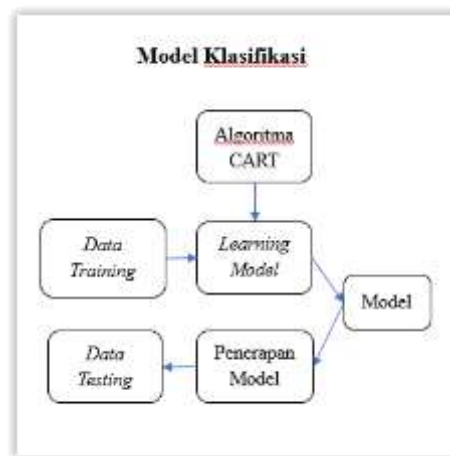
Setiap mahasiswa yang akan mengakhiri studinya harus menyelesaikan beban studi yang telah ditetapkan. Beban studi program sarjana adalah 144-146 SKS yang dijadwalkan untuk 8 (delapan) semester dan paling lama 14 (empat belas) semester. (Pedoman akademik Fakultas Teknik, 2012). Lama masa studi adalah lamanya waktu yang dibutuhkan mahasiswa dalam menyelesaikan beban studi yang telah ditentukan.

## **2.2. Prosedur Penelitian**

Prosedur yang dilakukan pada penelitian ini tersusun sebagai berikut: (1) Mengajukan surat permohonan izin mengadakan penelitian kepada Wakil Rektor III Universitas Negeri Jakarta, (2) Mengolah data mahasiswa yang mengikuti organisasi di Universitas Negeri Jakarta sehingga sesuai dengan parameter yang dibutuhkan, (3) Membuat algoritma *CART* menggunakan *Python ver 2.7.2*. (4) Menguji kinerja algoritma *CART*, (5) Analisis hasil kinerja algoritma *CART*, dan (6) Penarikan kesimpulan.

## **2.3. Metode Penelitian**

Metode yang digunakan dalam penelitian ini adalah metode klasifikasi data mining. Pada metode klasifikasi, *dataset* akan dibagi menjadi dua yaitu data *train* dan data *test*. Proses metode klasifikasi digambarkan dengan gambar 2.2 berikut:



**Gambar 2.2 Metode Klasifikasi**

Gambar 2.2 adalah proses metode klasifikasi. Pada penelitian ini peneliti akan membuat sebuah program pengklasifikasian data menggunakan algoritma *CART*. Program tersebut akan memproses data training untuk mencari pola unik pada data dan membentuk sebuah model yang akan diterapkan pada *data testing* untuk kemudian diukur nilai akurasi.

#### **2.4. Penelitian yang Relevan**

Berikut ini beberapa penelitian yang terkait, yaitu :

- (1). Prediksi Masa Studi Sarjana dengan Artificial Neural Network
  - a. Penelitian dilakukan oleh Muhamad Hanif Meinanda dkk. Mahasiswa Fakultas Teknologi Industri, Institut Teknologi Bandung.
  - b. Tujuan penelitian ini adalah untuk mengetahui variable-variable yang mempengaruhi masa studi mahasiswa.
  - c. Analisis penelitian ini menggunakan metode Artificial Neural Network
  - d. Berdasarkan penelitian ini, metode Artificial Neural Network merupakan model terbaik untuk memprediksi lama masa studi.

- (2). Penerapan Klasifikasi dengan Algoritma *CART* untuk Prediksi Kuliah Bagi Mahasiswa Baru
  - a. Penelitian ini dilakukan oleh Mardiani, mahasiswi STMIK MDP Palembang.
  - b. Tujuan penelitian ini adalah untuk mengetahui prediksi jumlah mahasiswa baru yang akan melanjutkan kuliah.
  - c. Penelitian ini menggunakan metode *CART*.
  - d. Hasil dari penelitian ini didapatkan 20 mahasiswa tidak akan melanjutkan kuliah.
- (3). Implementasi Data Mining untuk Memprediksi Kelulusan Mahasiswa Fakultas Komunikasi dan Informatika UMS Menggunakan Metode Naïve Bayes
  - a. Penelitian ini dilakukan oleh Ahmad Fikri Mauriza dan Yusuf Sulisty Nugroho, mahasiswa Fakultas Komunikasi dan Informatika, Universitas Muhammadiyah Surakarta.
  - b. Tujuan penelitian ini adalah untuk memprediksi jumlah mahasiswa Fakultas Komunikasi dan Informatika UMS yang lulus tepat waktu.
  - c. Analisis menggunakan Naïve Bayes.
  - d. Berdasarkan penelitian ini diprediksikan hanya sekitar 25,15% mahasiswa yang lulus tepat waktu sedangkan mahasiswa yang akan lulus terlambat mencapai 74,85%.
- (4). Perbandingan Performansi Algoritma C4.5 dan *CART* dalam Klasifikasi Data Nilai Mahasiswa Prodi Teknik Komputer Politeknik Negeri Padang

- a. Penelitian ini dilakukan oleh Indri Rahmayuni, dosen jurusan Teknologi Informasi, Politeknik Negeri Padang.
  - b. Tujuan penelitian ini adalah untuk mengetahui perbandingan akurasi algoritma C4.5 dan *CART*.
  - c. Analisis menggunakan algoritma C4.5 dan *CART*.
  - d. Berdasarkan penelitian ini dapat disimpulkan beberapa hal:
    - Algoritma C4.5 memberikan akurasi yang lebih baik dari pada algoritma *CART* dalam klasifikasi data nilai mahasiswa.
    - Algoritma C4.5 memberikan hasil lebih baik karena data nilai mahasiswa berupa data kelompok yang cocok dengan sifat klasifikasi algoritma C4.5.
    - Algoritma *CART* memberikan hasil dibawah C4.5 karena *CART* lebih cocok digunakan unuk data berjenis numerik.
- (5). Prediksi Penyakit Jantung Dengan Algoritma *Classification And Regression Trees*.
- a. Penelitian ini dilakukan oleh Nita Apriliani Puteri, Warih Maharani, dan Mahmud Dwi Suliiyo, mahasiswa Fakultas Teknik Informatika, Universitas Telkom.
  - b. Tujuan penelitian ini adalah:
    - Mengetahui hasil prediksi penyakit jantung dengan algoritma *Classification And Regression Trees*.
    - Mengetahui variabel utama yang berpengaruh dalam prediksi penyakit jantung.
  - c. Analisis menggunakan algoritma *Classification and Regression Trees*.

Berdasarkan penelitian ini akurasi terbesar pada data training sejumlah 800 data dengan jumlah data *training* yang bernilai positif 211 mencapai 76% sedangkan pada data *training* sejumlah 1000 data dengan jumlah data *training* yang bernilai positif 330 data yakni 77%.

## BAB III

### METODOLOGI PENELITIAN

#### 3.1. Tempat dan Waktu Penelitian

Penelitian dilakukan di program studi Pendidikan Teknik Informatika dan Komputer, Fakultas Teknik, Universitas Negeri Jakarta. Waktu penelitian dilaksanakan pada bulan Februari 2017 – Juli 2017.

#### 3.2. Alat dan Bahan Penelitian

##### 3.2.1. Alat

Dalam melakukan penelitian ini dibutuhkan dukungan perangkat keras dan perangkat lunak. Untuk pengolahan data dengan algoritma *CART* menggunakan bahasa pemrograman *Python ver 2.7.2* dan perangkat lunak *IDE* yang digunakan adalah *PyCharm Community Edition 2017.1.4* serta *Microsoft Excel* untuk mengolah data secara manual. Spesifikasi perangkat keras dan perangkat lunak yang digunakan pada penelitian ini ditunjukkan pada table 3.1 dan table 3.2 sebagai berikut:

**Tabel 3.1. Spesifikasi Perangkat Keras**

<b>Perangkat Keras</b>	
<b>Processor</b>	Intel ® Core™ i7-4500U CPU @ 1.80GHz
<b>Memori</b>	RAM 8 GB DDR 3
<b>Hard Disk Drive</b>	256 GB

**Tabel 3.2. Spesifikasi Perangkat Lunak**

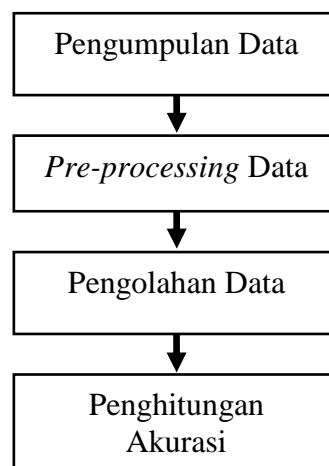
<b>Perangkat Lunak</b>	
<b>Sistem Operasi</b>	<i>Ms. Windows 8, 64 bit</i>
<b>Bahasa Pemrograman</b>	<i>Python ver 2.7.2</i>
<b>IDE</b>	<i>PyCharm Community Edition 2017.1.4</i>
<b>Pengolah Data</b>	<i>Microsoft Excel</i>

### 3.2.2. Bahan

Bahan penelitian berupa data mahasiswa angkatan 2011 dan 2012 yang mengikuti organisasi kemahasiswaan di Universitas Negeri Jakarta pada tahun 2015 sebanyak 200 data.

### 3.3. Diagram Alir Penelitian

Metode yang digunakan untuk proses analisis data dalam penelitian ini adalah sebagai berikut:

**Gambar 3.1 Diagram Alir Penelitian**

Hal pertama yang dilakukan dalam penelitian ini adalah mengumpulkan data mahasiswa angkatan 2011 dan 2012 yang mengikuti organisasi kemahasiswaan

pada tahun 2015. Untuk mendapatkan data tersebut, peneliti mengajukan surat permohonan data skripsi kepada bidang kemahasiswaan Universitas Negeri Jakarta. Data yang diperoleh sebanyak 401 data mahasiswa angkatan 2010-2015.

Setelah mendapatkan data dari bidang kemahasiswaan Universitas Negeri Jakarta, peneliti melakukan tahap selanjutnya yaitu *pre-processing data*. Pada tahap *pre-processing* ini terdapat tiga proses. Proses pertama adalah penyortiran data. Pada tahap ini data mahasiswa disortir berdasarkan tahun masuk Universitas Negeri Jakarta kemudian dipilih data mahasiswa yang masuk pada tahun 2011 dan 2012. Setelah proses penyortiran didapatkan 200 data mahasiswa.

Proses kedua adalah pemangkasan atribut data. Data yang didapatkan peneliti memiliki banyak unsur yang tidak diperlukan dalam penelitian, maka peneliti menghapus atribut yang tidak diperlukan tersebut dan mengambil data yang sesuai dengan atribut yang telah ditentukan yaitu, nama organisasi, jenis kelamin, fakultas, IPK dan kelulusan.

Proses ketiga adalah pemberian label pada masing-masing data. Data mahasiswa yang lulus tepat waktu diberi label angka 1 (satu), sedangkan data mahasiswa yang tidak lulus tepat waktu diberi label angka 0 (nol).

Setelah tahap *pre-processing data* selesai, maka tahap selanjutnya adalah pengolahan data menggunakan algoritma CART. Pada tahap ini data akan diproses sesuai dengan alur algoritma CART.

Tahap terakhir adalah penghitungan akurasi. Proses penghitungan akurasi algoritma CART ini menggunakan metode *K-fold Cross Validation*. Pada akhir penelitian akan diperoleh nilai akurasi algoritma CART dalam mengklasifikasikan data yang digunakan dalam penelitian ini.



### 3.4. Teknik dan Prosedur Pengumpulan Data

Peneliti mengumpulkan data pada penelitian ini menggunakan pendekatan data primer. Prosedur yang dilakukan oleh penulis pada penelitian ini adalah penulis menyerahkan surat penelitian skripsi kepada bidang kemahasiswaan Universitas Negeri Jakarta. Kemudian data tersebut diberikan langsung kepada peneliti dalam bentuk *soft copy* yang berisi data mahasiswa mengikuti organisasi kemahasiswaan pada tahun 2015. Jumlah data yang digunakan dalam penelitian ini adalah 200 data yang terdiri dari 159 data mahasiswa yang lulus tidak tepat waktu dan 41 data mahasiswa yang lulus tepat waktu. Parameter prediktor yang digunakan dalam penelitian ini adalah nama organisasi, jenis kelamin, fakultas, dan IPK. Sedangkan parameter yang akan dijadikan target adalah parameter kelulusan. Parameter kelulusan akan diberi label angka 1 (satu) untuk mahasiswa yang lulus tepat waktu dan nilai 0 (nol) untuk mahasiswa yang tidak lulus tepat waktu.

### 3.5. Teknik Analisis Data

*K-fold Cross Validation* adalah salah satu cara menemukan parameter terbaik dari satu model dengan cara menguji besarnya *error* pada data *test*. Dalam *cross validation* data dibagi menjadi  $k$  - subset dengan ukuran yang sama. Kemudian subset tersebut akan dibagi menjadi *data train* dan *data test*. (Santosa, 2007: 220).

Pada penelitian ini akan dilakukan *K-fold Cross Validation* sebanyak 3 kali dengan nilai  $k = 5, 10, \text{ dan } 20$ . Untuk nilai  $k = 5$  data akan dibagi menjadi 5 subset, 4 subset digunakan untuk *training* dan 1 subset digunakan sebagai data *testing*. Sedangkan untuk nilai  $k = 10$ , data akan dibagi menjadi 10 subset dengan 9 *data training* dan 1 *data test*. Begitu pula dengan nilai  $k = 20$ , data akan dibagi menjadi

20 subset dengan 19 data training dan 1 data test. Tahapan *k-fold cross validation* dengan nilai  $k = 10$  dapat diilustrasikan menjadi seperti Tabel 3.1 berikut.

**Tabel 3.3 Ilustrasi Tahapan *K-fold Cross Validation***

<b>Eksperimen</b>	<b><i>Data Train</i></b>	<b><i>Data Test</i></b>
<b>1</b>	K2,K3,K4,K5,K6,K7,K8,K9,K10	K1
<b>2</b>	K1,K3,K4,K5,K6,K7,K8,K9,K10	K2
<b>3</b>	K1,K2, K4,K5,K6,K7,K8,K9,K10	K3
<b>4</b>	K1,K2,K3, K5,K6,K7,K8,K9,K10	K4
<b>5</b>	K1,K2,K3,K4, K6,K7,K8,K9,K10	K5
<b>6</b>	K1,K2,K3,K4,K5, K7,K8,K9,K10	K6
<b>7</b>	K1,K2,K3,K4,K5,K6, K8,K9,K10	K7
<b>8</b>	K1,K2,K3,K4,K5,K6,K7, K9,K10	K8
<b>9</b>	K1,K2,K3,K4,K5,K6,K7,K8, K10	K9
<b>10</b>	K1,K2,K3,K4,K5,K6,K7,K8,K9	K10

## **BAB IV**

### **HASIL PENELITIAN DAN PEMBAHASAN**

#### **4.1. Deskripsi Hasil Penelitian**

Data yang digunakan dalam penelitian ini adalah data mahasiswa angkatan 2011 dan 2012 yang mengikuti organisasi kemahasiswaan di Universitas Negeri Jakarta pada tahun 2015. Data ini diperoleh dari Wakil Rektor III bagian Kemahasiswaan Universitas Negeri Jakarta.

##### **4.1.1. *Pre-processing Data***

Sebelum data diproses menggunakan algoritma *CART*, data yang telah dikumpulkan harus melalui tahap *pre-processing data* terlebih dahulu. *Pre-processing data* dilakukan agar data sesuai dengan atribut yang diperlukan dalam penelitian sehingga mengurangi kesalahan pemrosesan data pada saat dilaksanakannya eksperimen.

##### **4.1.1.1. Proses Penyortiran Data**

Data awal yang didapatkan oleh penulis masih berupa data mentah dari bidang Kemahasiswaan Universitas Negeri Jakarta. Oleh karena itu data tersebut harus disortir terlebih agar sesuai dengan kebutuhan penelitian. Data yang diberikan bidang kemahasiswaan Universitas Negeri Jakarta adalah data mahasiswa angkatan 2011 – 2014. Sedangkan data yang dibutuhkan pada penelitian ini adalah data mahasiswa angkatan 2011 dan 2012. Dari hasil penyortiran tersebut diperoleh 200 data mahasiswa angkatan 2011 dan 2012 dengan 159 data mahasiswa yang lulus tidak tepat waktu dan 41 data mahasiswa yang lulus tepat waktu.

#### 4.1.1.2. Proses Pemangkasan Data

Setelah melalui tahap penyortiran, tahap selanjutnya adalah proses pemangkasan data. Pada tahap penyortiran, data masih memiliki banyak unsur yang tidak diperlukan dalam penelitian, maka data perlu dipersempit dengan menghapus unsur yang tidak diperlukan tersebut.

**Tabel 4.1 Potongan Data Mentah Sebelum Dipersempit**

Nama Lengkap	ORMAWA/OPMAWA	Jabatan	Tempat Tanggal Lahir	Jenis Kelamin	Alamat	Nomor Telepon/ Hp	Golongan Darah	Tinggi Badan/ Berat Badan	E-mail	Program Studi	Fakultas	IPK	Semester	Tahun Masuk UNJ
Oka Bintoro	MENWA	Komandan Menwa	Jakarta, 19-Oktober-1993	Laki-laki	pintu air no.52 Pondokmelati Bekasi	pondokmelati bekasi	O	174cm/62kg	okabintoro@yahoo.com	Psikologi	FIP	2.81	IX	2011
Elien Maydia Sarsh	MENWA	DANP OKPAS	Pemalang, 28 Mei 1994	Perempuan	Taman Tytvan Indah Blok W1/10 Rt004 Rw012. Medan Satria Kali Baru Bekasi	81283433707	B	161cm/60kg	elienmaydiasarah@yahoo.com	Penjas	FIK	3	VII	2012
Nurul Fazriah	MENWA	Kepala Kesekretariatan	Bekasi, 03 April 1994	Perempuan	Ujung harapan, Rt. 20/08 No.22, Kel. Bahagia, Kec. Bahelan, Kab.	85710267228	O	160cm/54kg	nurul.fazriah.kuliah@gmail.com	Bimbingan dan Konseling	FIP	3.42	VII	2012
Deriana Darna Wijaya	MENWA	Wakil Komandan Pokok Pasukan	Jakarta, 15-07-1995	Laki-laki	sarwangan elok b3 no 2 r102/011 duren seribu bojongsari depok	89650210015	A	169cm/74kg	derunadarna@gmail.com	Pendidikan Khusus	FIP	3	VII	2012

Tabel 4.1 adalah potongan data mentah yang diberikan oleh bidang Kemahasiswaan UNJ kepada peneliti. Data tersebut masih mengandung banyak atribut – atribut yang tidak diperlukan seperti *timestamp*, jabatan, tempat tanggal lahir, alamat, nomor telepon, golongan darah, tinggi badan, dan email. Oleh karena itu peneliti menghapus atribut – atribut tersebut sehingga hanya menyisakan atribut – atribut yang akan digunakan yaitu, Nama Organisasi, Jenis Kelamin, Fakultas, IPK, dan Kelulusan.

**Tabel 4.2 Potongan Data Hasil Pemangkasan Data Sesuai Parameter**

Nama Lengkap	ORMAWA	nis Kelam	Fakultas	IPK
Oka Bintoro	MENWA	L	FIP	2.81
Elien Maydia Sarah	MENWA	P	FIK	3
Nurul Fazriah	MENWA	P	FIP	3.42
Deriana Darma W	MENWA	L	FIP	3
Muhammad Adam L	MENWA	L	FIP	3
Kartina Hayati	MENWA	P	FIP	3.22
Nurul Fazriah	MENWA	P	FIP	3.52
Alfian Septiandy N	MENWA	L	FIS	3.19
Shita Fitriana	MENWA	P	FIP	3.2
Muhamad Abdul R	MENWA	L	FMIPA	3.53
Rifqi Budi Wirawan	MENWA	L	FIK	2.94
Rahma Amalia	MENWA	P	FT	3.24

Tabel 4.2 adalah potongan dari hasil pemangkasan data sesuai dengan parameter yang dibutuhkan.

#### 4.1.1.3. Pemberian Label Pada Data

Setelah data dipersempit, maka selanjutnya data diberikan label untuk mempermudah proses pembelajaran data (*learning step*).

**Tabel 4.3 Potongan Data yang Sudah diberi Label**

Nama Lengkap	ORMAWA	Jenis Kelamin	Fakultas	IPK	Kelulusan
Warnadi	KSR PMI Unit UNJ	L	FT	2.96	0
Annisa'ul Chasanah Wahyu Rahayuningrum	KSR PMI Unit UNJ	P	FIP	3.4	0
Reva Siti Ardhita	KSR PMI Unit UNJ	P	FT	2.97	0
Nurmalis Dawati	KSR PMI Unit UNJ	P	FIP	3.73	1
Rendiyana Setiawati	KSR PMI Unit UNJ	P	FBS	3.22	0
Hendry	KOPMA	L	FIK	3.2	0
Rizka Ardiani Gasalba	KOPMA	P	FE	3.64	1
Jodi Sugiarto	KOPMA	L	FT	3	0
Aryono Fajar Harianto	KOPMA	L	FIK	3.75	0
Alfi Aflahal Amrul Muflih	KOPMA	L	FT	2.8	0

Tabel 4.3 adalah contoh data yang sudah diberi label untuk mempermudah algoritma dalam mengklasifikasikan data. Data mahasiswa yang lulus tepat waktu diberi label angka 1 (satu), sedangkan untuk data mahasiswa yang tidak lulus tepat waktu diberi label angka 0 (nol). Setelah itu data disimpan dengan format .csv untuk memudahkan pengolahan data menggunakan *python*.

## 4.2. Analisis Data Penelitian

Setelah data selesai melalui tahap *pre-processing*, tahap selanjutnya adalah tahap pengolahan data. Dataset yang sudah diberi label tersebut akan diolah menggunakan algoritma *CART*. Alat yang peneliti gunakan pada penelitian ini adalah *Jetbrains Pycharm Community Edition 2017.1.4* dengan dasar *Python 2.7.2*.

### 4.2.1. Pengolahan Data Menggunakan Algoritma *CART*

Pada tahap pengolahan data menggunakan algoritma *CART*, peneliti mulai melakukan *coding* algoritma menggunakan *python*. Peneliti mengambil *source code* program dari *web machinelearningmastery.com*, kemudian *source code* tersebut dimodifikasi sesuai dengan kebutuhan penelitian.

```
# CSV file
with open("dataset.csv", 'rb') as file:
    lines = reader(file)
    dataset = list(lines)
```

**Gambar 4.1 Potongan *Coding* untuk Membaca Dataset.**

Gambar 4.1 adalah *coding* untuk membaca file csv yang akan digunakan sebagai dataset. Pada *coding* di atas, program akan membuka file *dataset.csv* dan akan menyimpannya menjadi variabel *dataset* yang akan digunakan proses selanjutnya.

```

# Create child splits for a node or make terminal
def split(node, max_depth, min_size, depth):
    left, right = node['groups']
    del(node['groups'])
    # check for a no split
    if not left or not right:
        node['left'] = node['right'] = to_terminal(left + right)
        return
    # check for max depth
    if depth >= max_depth:
        node['left'], node['right'] = to_terminal(left), to_terminal(right)
        return
    # process left child
    if len(left) <= min_size:
        node['left'] = to_terminal(left)
    else:
        node['left'] = get_split(left)
        split(node['left'], max_depth, min_size, depth+1)
    # process right child
    if len(right) <= min_size:
        node['right'] = to_terminal(right)
    else:
        node['right'] = get_split(right)
        split(node['right'], max_depth, min_size, depth+1)

```

**Gambar 4.2 Potongan Coding Pembentukan Node**

Gambar 4.2 adalah potongan *coding* yang digunakan untuk pembentukan pohon klasifikasi *CART*. Fungsi *coding* tersebut adalah untuk menentukan cabang pohon keputusan.

```

# Split a dataset into k folds
def cross_validation_split(dataset, K):
    dataset_split = list()
    dataset_copy = list(dataset)
    fold_size = int(len(dataset) / K)
    for i in range(K):
        fold = list()
        while len(fold) < fold_size:
            index = randrange(len(dataset_copy))
            fold.append(dataset_copy.pop(index))
        dataset_split.append(fold)
    return dataset_split

```

**Gambar 4.3 Potongan Coding Pembagian Dataset Berdasarkan Nilai K**

Gambar 4.3 adalah *coding* yang digunakan untuk membagi dataset menjadi beberapa kelompok sesuai dengan nilai K. Fungsi *Coding* ini adalah untuk

membagi dataset menjadi beberapa kelompok dengan ukuran yang sama yang kemudian akan digunakan untuk penghitungan akurasi algoritma menggunakan *K-fold Cross Validation*.

#### 4.2.2. Hasil Akurasi Menggunakan *K-fold Cross Validation*

Setelah data diolah menggunakan algoritma *CART*, selanjutnya dilakukan penghitungan akurasi dengan menggunakan *K-fold Cross Validation*. Hasil dari penghitungan akurasi tersebut dianggap bagus atau akurat apabila hasil akurasi mendekati atau sama dengan 1, sebaliknya apabila hasil akurasi mendekati nol (0) maka algoritma tersebut dianggap tidak bagus atau tidak akurat. Pada penelitian ini akan dilakukan 3 (tiga) kali proses *K-fold Cross Validation* dengan nilai  $K = 5$ ,  $K = 10$ , dan  $K = 20$ .

Untuk mendapatkan nilai akurasi maka digunakan rumus:

$$\text{Akurasi} = \frac{\text{Jumlah Data Yang Sesuai}}{\text{Jumlah Keseluruhan Data}}$$

Untuk mendapatkan rata – rata nilai akurasi maka digunakan rumus:

$$\text{Rata – Rata Akurasi} = \frac{\text{Total Jumlah Akurasi}}{\text{Jumlah Iterasi}}$$

Pada nilai  $K = 5$ , maka dataset akan dibagi menjadi 5 kelompok dan dilakukan iterasi sebanyak 5 kali.

```
K = 5 :
Iterasi 1 : 0.6382978723404256
Iterasi 2 : 0.7446808510638298
Iterasi 3 : 0.8723404255319149
Iterasi 4 : 0.8297872340425532
Iterasi 5 : 0.9148936170212766

Rata-Rata Akurasi K(5): 0.800
```

**Gambar 4.4 Hasil Akurasi Algoritma *CART* K(5)**



Gambar 4.4 adalah hasil akurasi algoritma *CART* menggunakan *K-fold Cross Validation* dengan nilai  $K = 5$ . Berdasarkan gambar di atas, terdapat 5 hasil akurasi dari 5 iterasi yaitu, akurasi pada iterasi 1 adalah 0.6383, iterasi 2 adalah 0.7447, iterasi 3 adalah 0.8723, iterasi 4 adalah 0.8298, iterasi 5 adalah 0.9149. Rata – rata hasil akurasi algoritma *CART* dengan  $K = 5$  adalah 0.800.

Selanjutnya pada nilai  $K = 10$ , maka dataset akan dibagi menjadi 10 kelompok dan dilakukan iterasi sebanyak 10 kali.

```

K = 10 :
Iterasi 1 : 0.782608695652174
Iterasi 2 : 0.8695652173913043
Iterasi 3 : 0.9130434782608695
Iterasi 4 : 0.9565217391304348
Iterasi 5 : 0.782608695652174
Iterasi 6 : 0.8260869565217391
Iterasi 7 : 0.5217391304347826
Iterasi 8 : 0.8695652173913043
Iterasi 9 : 0.8695652173913043
Iterasi 10 : 0.782608695652174

Rata-Rata Akurasi K(10): 0.817.

```

**Gambar 4.5 Hasil Akurasi Algoritma *CART* K(10)**

Gambar 4.5 adalah hasil akurasi algoritma *CART* menggunakan *K-fold Cross Validation* dengan nilai  $K = 10$ . Berdasarkan gambar di atas, terdapat 10 hasil akurasi dari 10 iterasi yaitu, akurasi pada iterasi 1 adalah 0.7826, iterasi 2 adalah 0.8696, iterasi 3 adalah 0.9130, iterasi 4 adalah 0.9565, iterasi 5 adalah 0.7826, iterasi 6 adalah 0.8261, iterasi 7 adalah 0.5217, iterasi 8 adalah 0.8695, iterasi 9 adalah 0.8695, iterasi 10 adalah 0.7826. Rata – rata hasil akurasi algoritma *CART* dengan  $K = 10$  adalah 0.817.

Terakhir pada nilai  $K = 20$ , maka dataset akan dibagi menjadi 20 kelompok dan dilakukan iterasi sebanyak 20 kali.

```
K = 20 :  
Iterasi 1 : 0.8181818181818182  
Iterasi 2 : 0.8181818181818182  
Iterasi 3 : 0.8181818181818182  
Iterasi 4 : 0.09090909090909091  
Iterasi 5 : 0.9090909090909091  
Iterasi 6 : 0.8181818181818182  
Iterasi 7 : 0.9090909090909091  
Iterasi 8 : 1.0  
Iterasi 9 : 0.7272727272727273  
Iterasi 10 : 0.9090909090909091  
Iterasi 11: 0.8181818181818182  
Iterasi 12 : 0.9090909090909091  
Iterasi 13: 0.9090909090909091  
Iterasi 14 : 0.8181818181818182  
Iterasi 15 : 0.7272727272727273  
Iterasi 16 : 1.0  
Iterasi 17 : 0.8181818181818182  
Iterasi 18 : 0.9090909090909091  
Iterasi 19 : 0.9090909090909091  
Iterasi 20 : 0.7272727272727273  
  
Rata-Rata Akurasi K(20): 0.818
```

**Gambar 4.6 Hasil Akurasi Algoritma CART K(20)**

Gambar 4.6 adalah hasil akurasi algoritma *CART* menggunakan *K-fold Cross Validation* dengan nilai  $K = 20$ . Berdasarkan gambar di atas, terdapat 20 hasil akurasi dari 20 iterasi yaitu, akurasi pada iterasi 1 adalah 0.8182, iterasi 2 adalah 0.8182, iterasi 3 adalah 0.8182, iterasi 4 adalah 0.0909, iterasi 5 adalah 0.9091, iterasi 6 adalah 0.8182, iterasi 7 adalah 0.9091, iterasi 8 adalah 1.0, iterasi 9 adalah 0.7273, iterasi 10 adalah 0.9091, iterasi 11 adalah 0.8182, iterasi 12 adalah 0.9091, iterasi 13 adalah 0.9091, iterasi 14 adalah 0.8182, iterasi 15 adalah 0.7273, iterasi 16 adalah 1.0, iterasi 17 adalah 0.8182, iterasi 18 adalah 0.9091, iterasi 19 adalah 0.9091, iterasi 20 adalah 0.7273. Rata – rata hasil akurasi algoritma *CART* dengan  $K = 20$  adalah 0.818.

### 4.3. Pembahasan

Berdasarkan hasil penelitian di atas, peneliti akan membahas secara detail mengenai hasil akurasi dari algoritma *CART*. Kinerja algoritma dapat dilihat dari nilai rata – rata akurasi algoritma.

**Tabel 4.4 Perbandingan Nilai Akurasi Algoritma CART.**

Hasil Akurasi Algoritma CART		
<b>K = 5</b>	<b>K =10</b>	<b>K = 20</b>
0.800	0.817	0.818

Tabel 4.4 adalah perbandingan hasil rata-rata akurasi algoritma *CART* menggunakan metode *K-fold Cross Validation*. Pada nilai  $K = 5$  diperoleh nilai rata-rata akurasi sebesar 0.800. Sedangkan pada nilai  $K = 10$ , nilai rata-rata akurasi bertambah menjadi 0.817. Terakhir, ketika nilai  $K = 20$ , nilai rata-rata akurasi bertambah menjadi 0.818. Berdasarkan hasil tersebut, kinerja algoritma *CART* telah mencapai nilai akurasi 80%. Sedangkan nilai rata-rata akurasi tertinggi adalah ketika nilai  $K = 20$  yaitu 0.818.

### 4.4. Aplikasi Hasil Penelitian

Setelah penelitian dilakukan dan didapatkan hasil penelitian berupa nilai akurasi, maka algoritma *Classification and Regression Tree* dapat digunakan untuk mengklasifikasikan lama masa studi data mahasiswa yang mengikuti organisasi di Universitas Negeri Jakarta. Peneliti berharap hasil dari penelitian ini dapat berguna untuk penelitian selanjutnya. Aplikasi dari hasil penelitian ini sendiri dapat berupa sebuah program yang dapat memprediksikan lama masa studi mahasiswa yang mengikuti organisasi di Universitas Negeri Jakarta.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat diambil kesimpulan sebagai berikut:

1. Kinerja algoritma *CART* menggunakan metode uji *K-fold Cross Validation* dengan nilai  $K = 5, 10,$  dan  $20$  diperoleh nilai akurasi tertinggi pada nilai  $K = 20$  dengan nilai rata – rata akurasi  $0.818$ .
2. Algoritma *CART* telah mampu mengklasifikasikan lama masa studi mahasiswa yang mengikuti organisasi di Universitas Negeri Jakarta.

#### 5.2. Saran

Peneliti memiliki beberapa masukan untuk penelitian selanjutnya yang berhubungan dengan klasifikasi data mahasiswa yang mengikuti organisasi di Universitas Negeri Jakarta, dengan harapan hasil pengklasifikasian lebih akurat.

- Menambahkan *data train*, semakin banyak *data train* maka akan semakin akurat hasil dari klasifikasi.
- Mencoba menggunakan algoritma yang lain agar lebih akurat.
- Menambah parameter klasifikasi.

## DAFTAR PUSTAKA

- Anonim. (2010). Teori Kinerja. Tersedia: <https://teorionline.wordpress.com/2010/01/25/teori-kinerja/>. Diakses 9 Agustus 2017.
- De'ath, G. & Fabricius, K. E. (2000). Classification and regression trees: a powerful yet technique for ecological data analysis. *Ecology*, 81(11):3178-3192.
- Faisal, M. R. (2017). *Seri Belajar Data Science: Klasifikasi dengan Bahasa Pemrograman R*. [tempat tidak diketahui]: M Reza Faisal.
- [FT] Fakultas Teknik. (2012). *Pedoman Akademik*. Jakarta: Fakultas Teknik, Universitas Negeri Jakarta.
- [FT] Fakultas Teknik. (2015). *Buku Panduan Penyusunan Skripsi dan Non Skripsi*. Fakultas Teknik, Universitas Negeri Jakarta.
- Gorunescu, F. (2011). *Data Mining: Concepts, Models and Techniques*. Berlin: Springer. Tersedia: <https://books.google.com/books?isbn=3642197213>. Diakses 15 Mei 2017.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques*. Ed ke-3. Amerika Serikat: Elsevier.
- Hidayah, dkk. (2014). Organisasi Manajemen. Tersedia: <http://myrealblo.blogspot.co.id/2015/11/dasar-dasar-manajemen-organisasi.html>. Diakses 20 Juli 2017.
- Nawawi, H. H. (2005). *Manajemen Sumber Daya Manusia untuk Bisnis yang Kompetitif*. Yogyakarta: Universitas Gajah Mada.

- Pratiwi, S. S. (2017). Pengaruh keaktifan mahasiswa dalam organisasi dan motivasi belajar terhadap prestasi belajar mahasiswa Fakultas Ekonomi Universitas Negeri Yogyakarta. *Jurnal Pendidikan dan Ekonomi*, 6(1): 54-64.
- Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. *Encyclopedia of Database Systems*. 532-538.
- Rokach, L. & Maimon, O. (2015). *Data Mining with Decision Trees – Theory and Applications*. Ed ke-2 Singapura: World Scientific Publishing Co. Pte. Ltd.
- Ruky, A. S. (2006). *Sistem Manajemen Kinerja*. Jakarta: Gramedia Pustaka Utama.
- Santosa, B. (2007). *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Simanjuntak, P. J. (2005). *Manajemen dan Evaluasi Kerja*. Jakarta: Fakultas Ekonomi Universitas Indonesia.
- Siswanto. (2007). *Pengantar Manajemen*. Jakarta : PT. Bumi Aksara.
- Sumartini, S. H. & Purnami, S. H. (2015). Penggunaan metode classification and regression trees (CART) untuk klasifikasi rekurensi pasien kanker serviks di RSUD Dr. Soetomo Surabaya. *Jurnal Sains dan Seni ITS*, 4(2):2337-3520.
- Timofeev, R. (2005). *Classification and Regression Trees (CART) Theory and Applications*. [tesis]. Berlin: Humboldt University.
- Turban, E., Aronson, J. E., & Liang, T. P. (2005). *Decision Support Systems and Intelligent Systems*. Ed ke-7. New Jersey: Prentice-Hall, Inc.

## **TENTANG PENULIS**



Nama lengkap Nurul Indah Prabawati, lahir di Tangerang pada tanggal 21 September 1994. Merupakan anak ke-empat dari empat bersaudara dari pasangan Bapak Wardjjoso (alm.) dan Ibu Suyati. Penulis berkebangsaan Indonesia dan tinggal di Jalan Danau Kelapa Dua VIII no 62 RT 10 RW 05 Kelapa Dua, Tangerang.

Penulis memiliki riwayat pendidikan jenjang sekolah dasar di SDN Perumnas Bumi Kelapa Dua tahun 2000-2006, Penulis melanjutkan ke Sekolah tingkat menengah pertama di SMPN 19 Tangerang pada tahun 2006-2009 dan melanjutkan ke SMK Islamic Village Tangerang pada tahun 2009-2012. Pada tahun 2012 melanjutkan jenjang pendidikan di Universitas Negeri Jakarta, Program Studi Pendidikan Teknik Informatika dan Komputer. Penulis mengikuti program PKL (Praktek Kerja Lapangan) di Radio Republik Indonesia. Penulis juga mengikuti program PKM (Praktek Kegiatan Mengajar) di SMK Taman Siswa 2 Jakarta.