

# PREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN ALGORITMA C4.5

## (STUDI KASUS: FAKULTAS TEKNOLOGI INFORMASI UNIVERSITAS BALE BANDUNG)

Muhamad Ghasa Saputra<sup>#1</sup>, Rustiyana, S.T, M.T <sup>\*2</sup>, Nurul Imamah, S.T., M.T <sup>#3</sup>

*Teknik Informatika, Fakultas Teknologi Informasi, Universitas Bale Bandung  
Jl.R.A.A Wiranatakusuma No. 7 Baleendah,Kabupaten Bandung, Jawa Barat 40375*

<sup>1</sup>[ajitau088@gmail.com](mailto:ajitau088@gmail.com)

<sup>2</sup>[rustiyana@unibba.ac.id](mailto:rustiyana@unibba.ac.id)

<sup>3</sup>[nurulimamah@unibba.ac.id](mailto:nurulimamah@unibba.ac.id)

**Abstract:** *In everyday life humans can produce a lot of data, one of them is in the world of education where a university can produce a lot of data about students such as student GPA data, student origin school data, alumni data, data on the number of male students and the number of female students, and data from the area of students. From the amount of data generated, the data can be utilized using data mining techniques and Algoritma C4.5. With Data Mining, each data set or data warehouse can provide important knowledge that can be used as valuable information for the benefit of the university. By using C4.5 Algorithm, the data can be processed, one of them is by developing a system that has the ability to see the pattern of student graduation rates, so that it can later become a strategy in the lecture process. In research conducted at the Information Technology Faculty of Bale Bandung University, the data to be used is Student data in the form of Year, Semester, Student's Number, Name, Study Program, GPA, and Student TOEFL Value. Where the data will be processed or extracted into a decision tree to predict student graduation rates on time and not on time. After the formation of the decision tree has been made then measuring the performance of the decision tree using a confusion matrix to improve the accuracy of the decision tree. From the measurement of the performance of the decision tree that has been made, will produce values of precision, recall, and accuracy, then each value of the confusion matrix will be averaged according to the results that have been made using a confusion matrix. Next, an application is made to predict the level of graduation of students who in the source code represent the rules in the c4.5 algorithm.*

**Keywords:** *C4.5 Algorithm, Students*

**Abstrak-** Dalam kehidupan sehari-hari manusia dapat menghasilkan banyak sekali data, salah satunya dalam dunia pendidikan dimana suatu universitas dapat menghasilkan banyak sekali data tentang mahasiswa seperti, data IPK mahasiswa, data sekolah asal mahasiswa, data alumni, data jumlah mahasiswa berjenis kelamin laki-laki dan jumlah mahasiswa berjenis kelamin perempuan, dan data asal daerah mahasiswa. Dari banyaknya data yang dihasilkan, data tersebut dapat dimanfaatkan dengan menggunakan teknik data mining dan Algoritma C4.5. Dengan Data Mining, setiap kumpulan data atau gudang data dapat memberikan pengetahuan penting yang dapat

dijadikan sebuah informasi yang berharga bagi kepentingan universitas. Dengan menggunakan Algoritma C4.5 data tersebut dapat diolah, salah satunya dengan menyusun sistem yang mempunyai kemampuan melihat pola tingkat kelulusan mahasiswa, untuk selanjutnya bisa menjadi strategi dalam proses perkuliahan. Dalam penelitian yang dilakukan di Fakultas Teknologi Informasi Universitas Bale Bandung data yang akan digunakan berupa data Mahasiswa berupa Tahun, Semester, Nim, Nama, Prodi, Nilai IPK, dan Nilai TOEFL mahasiswa. Dimana data tersebut akan diolah atau di extract untuk dijadikan suatu pohon keputusan untuk memprediksi tingkat kelulusan mahasiswa tepat waktu dan tidak tepat waktu. Setelah pembentukan pohon keputusan telah dibuat kemudian dilakukan pengukuran kinerja dari pohon keputusan menggunakan confusion matrix untuk meningkatkan akurasi dari pohon keputusan. Dari pengukuran kinerja pohon keputusan yang telah dibuat, akan menghasilkan nilai precision, recall, dan akurasi, kemudian tiap-tiap nilai dari confusion matrix akan dirata-ratakan sesuai dengan hasil yang telah dilakukan menggunakan confusion matrix. Selanjutnya dibuatkan suatu aplikasi untuk memprediksi tingkat kelulusan mahasiswa yang didalam source code nya merepresentasikan aturan-aturan dalam algoritma c4.5.

**Kata Kunci:** Algoritma C4.5, Mahasiswa

### PENDAHULUAN

Perkembangan teknologi informasi yang begitu maju saat ini, menyebabkan tingkat akurasi suatu data sangat dibutuhkan dalam kehidupan sehari-hari. Setiap informasi yang ada menjadi suatu hal penting untuk menentukan setiap keputusan dalam situasi tertentu. Hal ini menyebabkan penyediaan informasi menjadi sarana untuk dianalisa dan diringkas menjadi suatu pengetahuan dari data yang bermanfaat ketika pengambilan suatu keputusan dilakukan. (Nuqson Masyqur Huda, 2010).

Fakultas Teknologi Informasi Universitas Bale Bandung merupakan salah satu organisasi pendidikan yang bergerak di bidang pendidikan teknologi dan informasi. Informasi tingkat kelulusan mahasiswa Fakultas Teknologi Informasi Universitas Bale Bandung sangatlah penting untuk menentukan kelulusan mahasiswa sehingga dapat meningkatkan pelayanan yang dapat membuat mahasiswa nyaman dalam proses perkuliahan dan bisa lulus tepat waktu.

Penggunaan Algoritma C4.5 dapat dijadikan pertimbangan dalam mengambil keputusan lebih lanjut tentang faktor yang mempengaruhi kelulusan mahasiswa.

Dikutip dari <https://forlap.ristekdikti.go.id> jumlah mahasiswa Fakultas Teknologi Informasi Universitas Bale Bandung berjumlah 220 mahasiswa untuk Program Studi Teknik Informatika pada semester genap 2018, Sedangkan untuk Program Studi Sistem Informasi berjumlah 52 mahasiswa untuk semester genap 2018 sehingga dari keseluruhan mahasiswa Fakultas Teknologi Informasi Universitas Bale Bandung pada semester genap 2018 berjumlah 272 mahasiswa.

Pada penelitian (Mujib Ridwan, Hadi Suyono, M Sarosa, 2013) menjelaskan bahwa faktor yang paling berpengaruh dalam penentuan klasifikasi kinerja akademik mahasiswa adalah Indeks Prestasi Kumulatif (IPK), Indeks Prestasi Semester (IPS) semester 1, semester 4 dan jenis kelamin. Pada penelitian berdasarkan attribute jenis kelamin, asal sekolah SMA dan IPS semester 1 sampai dengan IPS semester 6. Algoritma C4.5 merupakan algoritma klasifikasi pohon keputusan yang banyak digunakan karena memiliki kelebihan utama dari algoritma yang lainnya. Kelebihan algoritma C4.5 dapat menghasilkan pohon keputusan yang mudah diinterpretasikan, memiliki tingkat akurasi yang dapat diterima, efisien dalam menangani attribute bertipe diskret dan numerik. Dalam mengkonstruksi pohon keputusan, algoritma C4.5 membaca sample data training dari storage dan memuatnya ke memori. Hal inilah yang menjadi salah satu kelemahan algoritma C4.5 dalam kategori 'skalabilitas' adalah algoritma ini dapat digunakan jika data training dapat disimpan secara keseluruhan dan pada waktu yang bersamaan pada memori.

Oleh karena itu penyusun pada penelitian ini mengembangkan teknik algoritma C4.5 untuk memprediksi tingkat kelulusan mahasiswa dengan menggunakan data mahasiswa Fakultas Teknologi Informasi Universitas Bale Bandung sebagai objek penelitian. Dimana pada hasil akhir penelitian akan menampilkan tahapan-tahapan dalam memanfaatkan data dan mengolah data menjadi sebuah informasi yang berguna.

## I LANDASAN TEORI

### A. Prediksi

Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi. (Herdianto, 2013)

Pengertian Prediksi sama dengan ramalan atau perkiraan. Menurut kamus besar bahasa Indonesia, prediksi adalah hasil dari kegiatan memprediksi atau meramal atau memperkirakan nilai pada masa yang akan datang dengan menggunakan data masa lalu. Prediksi menunjukkan apa yang akan terjadi pada suatu keadaan tertentu dan merupakan input bagi proses

perencanaan dan pengambilan keputusan.

Prediksi bisa berdasarkan metode ilmiah ataupun subjektif belaka. Ambil contoh, prediksi cuaca selalu berdasarkan data dan informasi terbaru yang didasarkan pengamatan termasuk oleh satelit. Begitupun prediksi gempa, gunung meletus ataupun bencana secara umum.

### B. Kelulusan

Proses kelulusan adalah kegiatan paling akhir dari manajemen peserta didik. Kelulusan adalah pernyataan dari lembaga pendidikan (sekolah) tentang telah diselesaikannya program pendidikan yang harus diikuti oleh peserta didik. Setelah peserta didik selesai mengikuti seluruh program pendidikan disuatu lembaga pendidikan dan berhasil lulus dan ujian akhir, maka kepada peserta didik tersebut diberikan surat keterangan lulus atau sertifikat.

Umumnya surat keterangan tersebut sering disebut ijazah atau surat tanda tamat belajar (STTB). Kompetensi lulusan- yaitu kemampuan minimal yang harus dicapai oleh peserta didik setelah tamat mengikuti pendidikan pada jenjang atau satuan pendidikan tertentu.

Ketika peserta didik sudah lulus, maka secara formal hubungan antara peserta didik dan lembaga telah selesai. Namun demikian, diharapkan hubungan antara para alumni dan sekolah tetap terjalin. Dari hubungan sekolah dan alumni ini, lembaga pendidikan (sekolah) bisa memanfaatkan hasil-hasilnya. Lembaga pendidikan (sekolah) bisa menjangkau berbagai informasi. Misalnya informasi tentang materi pelajaran mana yang sangat membantu untuk studi selanjutnya. Mungkin juga informasi tentang lapangan kerja yang bisa dijangkau bagi alumni lainnya.

Hubungan antara sekolah dengan para alumni dapat dipelihara lewat pertemuan-pertemuan yang diselenggarakan oleh para alumni, yang biasa disebut "reuni". Bahkan saat ini setiap lembaga pendidikan (sekolah) ada organisasi alumni, misalnya IKA (Ikatan Alumni). Prestasi yang dicapai oleh alumni dari lembaga pendidikan (sekolah) ini perlu didata atau dicatat oleh lembaga. Sebab catatan tersebut sangat berguna bagi lembaga dalam mempromosikan lembaga pendidikannya. (Tim Dosen UPI, 2012)

### C. Mahasiswa Yang Lulus

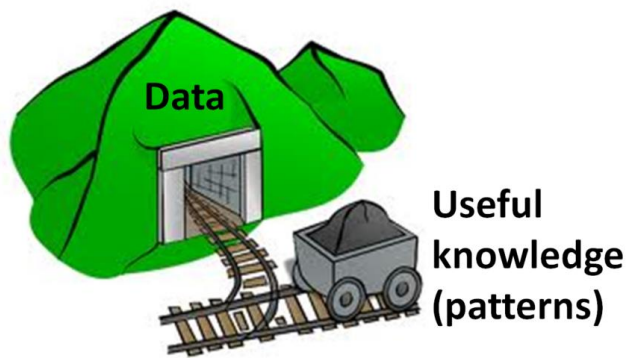
Mahasiswa yang lulus merupakan mahasiswa yang telah dinyatakan lulus program sarjana apabila memenuhi kriteria yang telah ditentukan, misalnya mahasiswa tersebut telah menyelesaikan sekurang-kurangnya sejumlah 144-148 sks yang diwajibkan, IPK minimal 2,00, Nilai mata kuliah wajib minimal C, Jumlah SKS dengan nilai D tidak lebih dari 10% jumlah SKS total, tidak ada nilai E dan lulus ujian skripsi.

### D. Data Mining

Nama data mining sebenarnya mulai dikenal sejak tahun 1990 ketika pekerjaan pemanfaatan data menjadi sesuatu yang penting dalam berbagai bidang, mulai dari bidang akademik, bisnis hingga medis (Goronescu, 2011). Data mining dapat diterapkan pada berbagai

bidang yang mempunyai sejumlah data, tetapi karena wilayah penelitian dengan sejarah yang belum lama dan belum melewati masa 'remaja', maka data mining masih diperdebatkan posisi bidang pengetahuan yang memilikinya.

Terlepas dari 'remaja-nya' data mining, ternyata data mining diproyeksikan menjadi jutaan dollar di dunia industri pada tahun 2000, sedangkan pada saat yang sama, ternyata data mining dipandang sebelah mata oleh sejumlah peneliti sebagai dirty word in statics (Goronescu, 2011). Mereka adalah orang-orang yang tidak memandang data mining sebagai sesuatu yang menarik bagi mereka pada saat itu.



Gambar 1 Data Mining

Munculnya data mining didasarkan pada jumlah data yang tersimpan dalam basis data semakin besar. Misalnya dalam sebuah supermarket, ada beberapa transaksi pelanggan yang terjadi dalam sehari dan ada berapa juta data yang sudah tersimpan dalam sebulan. Dalam perusahaan, ada berapa juta data yang sudah tersimpan dari setiap kegiatan produksi untuk setiap produk yang dibuat dalam beberapa tahun. Contoh lain, jika anda mempunyai kartu kredit, mungkin anda sering menerima surat penawaran barang atau jasa. Jika bank mempunyai 1.000.000 nasabah dan biaya pengiriman surat per nasabah adalah 500 rupiah, maka biaya yang harus dikeluarkan bank adalah 500 juta rupiah, padahal nasabah yang mungkin benar-benar hanya membeli sekitar 15%. Akibatnya ada pembuangan biaya sekitar 85% dari 500 juta atau sekitar 425 juta; sungguh sia-sia. Jika perusahaan dapat memanfaatkan data-data yang ada sehingga hanya nasabah yang berpotensi untuk membeli saja yang dikirim surat maka biaya pengiriman tersebut dapat ditekan.

Yang menjadi pertanyaan untuk data-data dalam perusahaan yang semakin menggunung jumlahnya dari waktu ke waktu adalah mau diapakan data tersebut, apakah hanya dibuat untuk laporan akhir tahun kemudian dibuang? Apakah hanya akan dikubur dalam gudang data dan tidak diapa-apakan? Tentu sayang sekali jika data-data tersebut tidak dimanfaatkan untuk kepentingan perusahaan atau instansi-instansi yang berkepentingan.

#### E. Algoritma C4.5

*Algoritma C4.5 merupakan salah satu algoritma yang digunakan untuk membentuk decision tree berdasarkan training data. Algoritma ini merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal (Wenefrida Tulit Ina, 2013). Decision tree adalah model prediksi menggunakan struktur tree atau*

*strukturberhirarki. Konsep dari decision tree adalah mengubah data menjadi decision tree dan aturan-aturan keputusan (decision rules) (eko prasetyo, 2014).*

Beberapa pengembangan yang dilakukan algoritma C4.5 adalah bisa mengatasi missing value, continue data, dan pruning. Algoritma C4.5 mempunyai input berupa training samples dan samples, training samples berupa data contoh yang akan digunakan untuk membangun sebuah tree yang telah diuji kebenarannya, sedangkan untuk samples merupakan field-field data yang nantinya akan kita gunakan sebagai parameter dalam melakukan klasifikasi data (eko prasetyo, 2014).

Algoritma C4.5 dan *decision tree* merupakan dua model yang tidak terpisahkan. Karena untuk membangun *decision tree* dibutuhkan algoritma C4.5 (Anik Andriani, 2013). Di akhir tahun 1970 hingga di awal tahun 1980, J Ross Quinlan seorang peneliti di bidang mesin pembelajaran membuat pengembangan sebuah *decision tree* yang dinamakan *ID3 (Iterative Dichotomizer)*, walaupun sebenarnya proyek ini telah dibuat sebelumnya oleh E.B. Hunt, J. Marin dan PT. Stone, kemudian Quinlan membuat algoritma dari pengembangan *ID3* yang dinamakan C4.5 yang berbasis *supervised learning* (Anik Andriani, 2013).

#### F. Aplikasi

Aplikasi adalah program siap pakai yang dapat di gunakan untuk menjalankan perintah-perintah dari pengguna aplikasi tersebut dengan tujuan mendapatkan hasil yang lebih akurat sesuai dengan pembuatan aplikasi tersebut, aplikasi adalah program komputer yang di tulis dalam suatu bahasa pemrograman dan di pergunakan untuk menyelesaikan masalah tertentu (Robi Muhamad, 2019).

#### G. WEB

Secara basic, website di gunakan untuk publikasi informasi. Adapun informasi yang akan di sediakan adalah beraneka ragam dari profil pribadi hingga copany profil. Website sering juga di sebut web, dapat diartikan suatu kumpulan halaman-halaman yang menampilkan berbagai macam informasi teks, data gambar diam ataupun bergerak, data dinamis, suara maupun gabungan dari semuanya, baik itu yang bersifat statis maupun dinamis yang dimana membentuk suatu rangkaian bangunan yang saling berkaitan dimana masing-masing dihubungkan dengan jaringan halaman atau hyperlink (Nurhadi, 2017).

Definisi lainnya dari website adalah kumpulan dari berbagai macam halaman situs, yang terangkum dari sebuah domain atau juga subdomain yang lebih tepatnya berada di dalam WWW (World Wide Web) yang tentunya terdapat di dalam internet.

#### H. My Structured Query Language (MySQL)

MySQL adalah Sebuah program database server yang mampu menerima dan mengirimkan datanya sangat cepat, multi user serta menggunakan perintah dasar SQL (Structured Query Language). MySQL

merupakan dua bentuk lisensi, yaitu FreeSoftware dan Shareware. MySQL yang biasa kita gunakan adalah MySQL FreeSoftware yang berada dibawah Lisensi GNU/GPL (General Public License), (Haris Spuro, 2012).

### I. Hypertext Preprocessor (PHP)

Bahasa pemrograman PHP merupakan bahasa pemrograman yang cukup populer dan banyak di gunakan oleh para programmer di dunia. PHP atau yang memiliki kepanjangan Hypertext Preprocessor, merupakan suatu bahasa pemrograman yang di fungsikan untuk membangun suatu website dinamis, PHP menyatu degan kode HTML. HTML digunakan seagai pembangun atau pondasi dari kerangka layout web, sedangkan PHP difungsikan sebagai prosesnya, sehingga dengan adanya PHP tersebut sebuah web akan mudah di maintenance, (Agus Saputra, 2013).

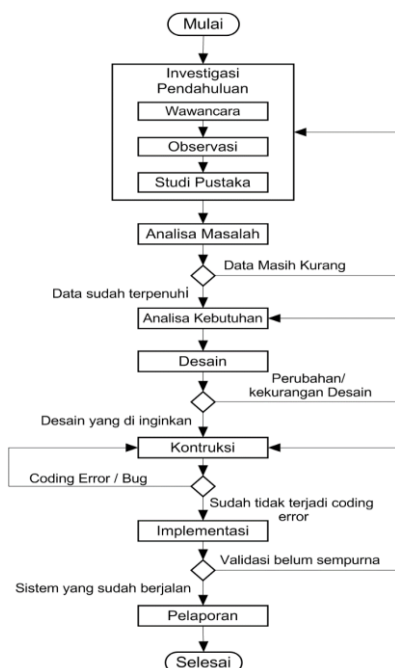
PHP berjalan pada sisi server sehingga PHP disebut juga sebagai bahasa Sever Side Scripting, artinya bahwa dalam setiap/unuk menjalankan PHP, akan mebutuhka web server untuk menjalankannya. PHP ini besifat open source sehinga dapat di pakai secara gratis, dan mampu lintas platform, yaitu dapat berjalan pada sistem operasi widow maupun linux.

### I. PEKERJAAN DAN DISKUSI HASIL

#### A. Proses pekerjaan

##### 1. Kerangka berfikir

Pada perancangan sistem Prediksi kelulusan menggunakan algoritma c4.5, sebagai pedoman penulis untuk mencapai tujuan dari penelitian yang telah ditentukan sebelumnya adalah sebagai berikut.



Gambar 2 Kerangka Berfikir

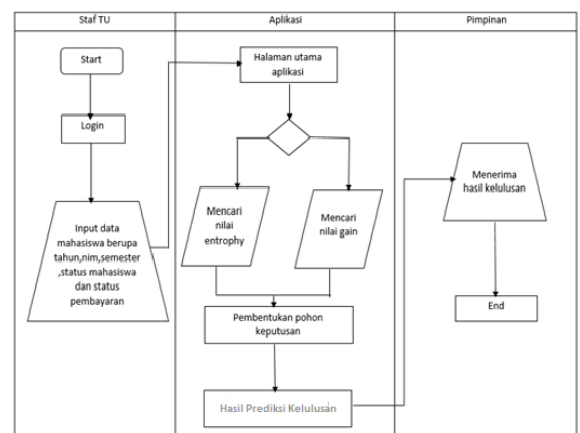
## 2. Analisis Sistem

Analisis sistem ini dilakukan untuk memberikan solusi terhadap permasalahan penilaian kinerja prediksi kelulusan yang ada di Fakultas Teknologi Informasi Universitas Bale Bandung.

### a. Analisis Sistem yang Masih Berjalan

#### A. Analisis Sistem Usulan

Analisis Usulan dilakukan untuk memberikan usulan dalam analisis sistem ini agar didapatkannya sistem yang dapat membantu terkait permasalahan yang ada. Pada Analisis Sistem Usulan ini tentunya mempunyai kelebihan yaitu dengan memakai aplikasi yang dapat memudahkan pengguna dalam melakukan perhitungan dan pendataan.



Gambar 3 Sistem yang diusulkan

## 2. Perancangan

Perancangan data mining ini penulis menggunakan algoritma C4.5 proses pada pohon keputusan adalah mengubah bentuk data (table) menjadi model pohon, mengubah model pohon menjadi rule, dan menyederhanakan rule.

Data akan dibuat pohon keputusan untuk menentukan mahasiswa yang lulus tepat waktu atau tidak dengan melihat nilai IPK mahasiswa dan nilai TOEFL mahasiswa. Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

- Pilih atribut akar.
- Buat cabang untuk tiap-tiap nilai.
- Bagi kasus dalam cabang.

Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama. Untuk memilih atribut sebagai akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung *gain*, seharusnya mencari nilai *entropy* terlebih dahulu.

Sebelum melakukan pembentukan *decision tree*, penulis menggunakan teknik data mining terlebih dahulu untuk memproses data seperti langkah-langkah sebagai berikut :



## f. Pembentukan *Decision Tree*

### a. Pembersihan Data

Data yang akan digunakan dalam proses ini adalah data mahasiswa pada tahun 2018 pada ajaran Semester 8 Genap. Data mahasiswa tersebut diambil dari Fakultas Teknologi Informasi Universitas Bale Bandung. Tidak semua data akan diambil untuk melakukan proses perhitungan, Data yang akan digunakan meliputi nilai IPK mahasiswa dan nilai TOEFL mahasiswa akan digunakan sebagai atribut dan keterangan lulus akan digunakan sebagai atribut target.

### b. Pembersihan dan pemrosesan awal data

Pembersihan dilakukan untuk membersihkan *noise* pada data. Dalam penelitian ini pembersihan pada data dilakukan menggunakan aplikasi *microsoft excel* 2010. Pada pembersihan tersebut telah diatur sedemikian rupa sehingga setiap data tentang mahasiswa, termasuk data nilai IPK mahasiswa dan nilai TOEFL mahasiswa yang disimpan tidak boleh kosong.

### c. Transformasi Pada Data

Transformasi data digunakan untuk mengolah data menjadi nilai dengan format tertentu. Dalam proses ini digunakan data yang bersifat *Diskrit*, sehingga data yang bersifat *kontinu* akan diubah menjadi data yang bersifat *diskrit*. Pengelompokan IPK akan dibagi menjadi dua kelas dimana  $IPK \geq 3.00$  akan dikelompokkan menjadi IPK tinggi sedangkan  $IPK < 3.00$  akan dikelompokkan menjadi IPK rendah. Sedangkan untuk nilai TOEFL akan dibagi menjadi dua kelas yaitu dimana nilai  $IPK \geq 400$  dikelompokkan menjadi menengah atas sedangkan  $TOEFL < 400$  akan dikelompokkan menjadi menengah bawah.

### d. Data Mining

Tahap data *mining* dimulai dengan pembagian pada data tentang mahasiswa yang memiliki atribut target lulus dan tidak lulus yang telah di *transformasikan* sebagai data latih dan data uji. Data latih akan digunakan sebagai pembentukan *decision tree* sedangkan data uji akan digunakan sebagai pengukuran kinerja pada *decision tree*. Setelah data latih telah dibagi akan digunakan dalam proses pembentukan *decision tree*.

### e. Analisis Data

Analisis data merupakan tahap untuk menganalisis dan mengidentifikasi data untuk kebutuhan perancangan sistem yang akan dibuat. Dalam hal ini, penulis menggunakan data mahasiswa Fakultas Teknologi Informasi Universitas Bale Bandung sebagai objek penelitian ini dimana data tersebut menggunakan data mahasiswa Fakultas Teknologi Informasi.

Proses klasifikasi tiap-tiap *field* pada tabel mahasiswa

#### 1. Klasifikasi total seluruh kasus

Setelah seluruh kasus telah di klasifikasikan kemudian dicari nilai *entropi* dari keseluruhan kasus tersebut dengan perhitungan sebagai berikut:

Diketahui :

Jumlah kasus (a)

$$= 56$$

Jumlah kasus mahasiswa yang tidak lulus (b) = 11

Jumlah kasus mahasiswa yang lulus (c) = 45

$$\left(-\frac{b}{a} * \log_2 \left(\frac{b}{a}\right)\right) + \left(-\frac{c}{a} * \log_2 \left(\frac{c}{a}\right)\right)$$

$$= (0.461199) + (0.253528)$$

$$Entropi \text{ total} = 0.714727$$

#### 2. Klasifikasi Nilai IPK Tinggi

Setelah seluruh kasus Status IPK tinggi telah di klasifikasikan kemudian dicari nilai *entropi* dari keseluruhan kasus mahasiswa IPK tinggi tersebut dengan perhitungan sebagai berikut:

Diketahui :

Jumlah kasus (a)

$$= 50$$

Jumlah kasus mahasiswa IPK tinggi yang tidak lulus (b)

$$= 5$$

Jumlah kasus mahasiswa IPK tinggi yang lulus (c)

$$= 45$$

$$\left(-\frac{b}{a} * \log_2 \left(\frac{b}{a}\right)\right) + \left(-\frac{c}{a} * \log_2 \left(\frac{c}{a}\right)\right)$$

$$= (0.332193) + (0.136803)$$

$$Entropi \text{ total} = 0.468996$$

Telah diketahui bahwa nilai *entropi* dari keseluruhan kasus mahasiswa yang IPK tinggi adalah 0.468996.

#### 3. Klasifikasi IPK Rendah

Setelah seluruh kasus mahasiswa IPK rendah telah di klasifikasikan kemudian dicari nilai *entropi* dari keseluruhan kasus mahasiswa IPK rendah tersebut dengan perhitungan sebagai berikut:

Diketahui :

Jumlah kasus mahasiswa IPK rendah (a)

$$= 6$$

Jumlah kasus mahasiswa IPK rendah yang tidak lulus (b) = 6

Jumlah kasus mahasiswa IPK rendah yang lulus

(c) = 0

$$\left(-\frac{b}{a} * \log_2 \left(\frac{b}{a}\right)\right) + \left(-\frac{c}{a} * \log_2 \left(\frac{c}{a}\right)\right)$$

$$= (0) + (0)$$

$$Entropi \text{ total} = 0$$

Telah diketahui bahwa nilai *entropi* dari keseluruhan kasus mahasiswa yang aktif adalah 0.

Telah diketahui bahwa nilai *entropi* dari keseluruhan kasus adalah 0.714727.

#### 4. Klasifikasi Nilai TOEFL

Setelah seluruh kasus Status nilai TOEFL menengah atas telah di klasifikasikan kemudian dicari nilai *entropi* dari keseluruhan kasus TOEFL menengah atas tersebut dengan perhitungan sebagai berikut:

Diketahui :

Jumlah kasus nilai TOEFL menengah atas (a)  
= 50

Jumlah kasus TOEFL menengah atas tidak lulus (b)  
= 5

Jumlah kasus TOEFL menengah atas yang lulus (c)  
= 45

$$\left(-\frac{b}{a} * \log_2 \left(\frac{b}{a}\right)\right) + \left(-\frac{c}{a} * \log_2 \left(\frac{c}{a}\right)\right)$$

$$= (0.332193) + (0.136803)$$

*Entropi* total = 0.468996

Telah diketahui bahwa nilai *entropi* dari keseluruhan kasus TOEFL menengah atas adalah 0.468996.

#### 5. Klasifikasi TOEFL Menengah Bawah

Setelah seluruh kasus Status TOEFL menengah bawah telah di klasifikasikan kemudian dicari nilai *entropi* dari keseluruhan kasus TOEFL menengah bawah tersebut dengan perhitungan sebagai berikut:

Diketahui :

Jumlah kasus TOEFL menengah bawah (a)  
= 6

Jumlah kasus TOEFL menengah bawah yang tidak lulus (b)  
= 6

Jumlah kasus TOEFL menengah bawah yang lulus (c)  
= 0

$$\left(-\frac{b}{a} * \log_2 \left(\frac{b}{a}\right)\right) + \left(-\frac{c}{a} * \log_2 \left(\frac{c}{a}\right)\right)$$

$$= (0) + (0)$$

*Entropi* total = 0

Telah diketahui bahwa nilai *entropi* dari keseluruhan kasus TOEFL menengah bawah adalah 0.

Setelah masing-masing atribut telah dicari nilai *entropi* nya (tabel 4.12) kemudian dicari nilai *gain* nya.

Tabel 1 Nilai Entropi

Atribut	Total	Lulus	Tidak Lulus	Entropi
Seluruh Kasus	56	45	11	0.714727
Mahasiswa IPK tinggi	50	5	45	0.468996
Mahasiswa IPK rendah	6	0	6	0
TOEFL menengah atas	50	45	5	0.468996
TOEFL menengah bawah	6	0	6	0

Setelah nilai *entropi* dari masing-masing nilai atribut telah diketahui, selanjutnya mencari nilai *gain* dari setiap atribut.

Mencari nilai *gain* Nilai IPK

Diketahui :

*Entropi* kasus keseluruhan (a) =  
0.714727

*Entropi* Kasus Nilai IPK tinggi (b) =  
0.468996

*Entropi* Kasus Nilai IPK rendah (c) = 0

Jumlah Kasus Keseluruhan (d) = 56

Jumlah Kasus Nilai IPK tinggi (e) = 50

Jumlah Status Nilai IPK rendah (f) = 6

$$\text{Gain (Total, Nilai IPK)} = a - \left(\frac{e}{d} * b\right) + \left(\frac{f}{d} * c\right)$$

$$\text{Gain (Total, Nilai IPK)} = 0.418746429 + 0$$

$$= 0.418746429$$

Telah diketahui bahwa nilai *gain* dari nilai IPK yaitu 0.418746429.

Mencari Nilai *Gain* Nilai TOEFL

Diketahui :

*Entropi* kasus keseluruhan (a) =  
0.714727

*Entropi* nilai TOEFL menengah atas (b) =  
0.468996

*Entropi* nilai TOEFL menengah bawah (c) =  
0

Jumlah Kasus Keseluruhan (d) = 56

Jumlah Kasus Nilai TOEFL menengah atas (e) = 50

Jumlah Kasus Nilai TOEFL menengah bawah (f) = 6

$$\text{Gain (Total, Nilai TOEFL)} = a - \left(\frac{e}{d} * b\right) + \left(\frac{f}{d} * c\right)$$

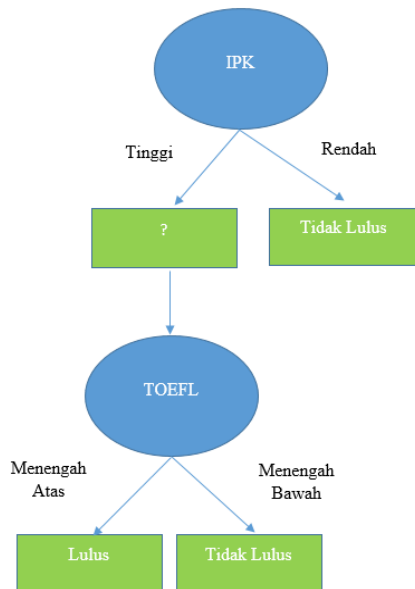
$$\text{Gain (Total, Nilai TOEFL)} = 0.418746429 + 0$$

$$= 0.418746429$$

Telah diketahui bahwa nilai *gain* dari Nilai TOEFL yaitu 0.418746429.

Setelah nilai *entropi* dan nilai *gain* dari masing-masing atribut telah diketahui, kemudian melakukan pembentukan *decision tree* dimana nilai *gain* tertinggi yaitu Nilai IPK akan dijadikan sebagai root/node nya sehingga akan menghasilkan suatu *rule* dari hasil perhitungan dan pembentukan *decision tree* dari data mahasiswa yang

digunakan. Untuk pembentukan pohon keputusan dapat dilihat pada gambar.



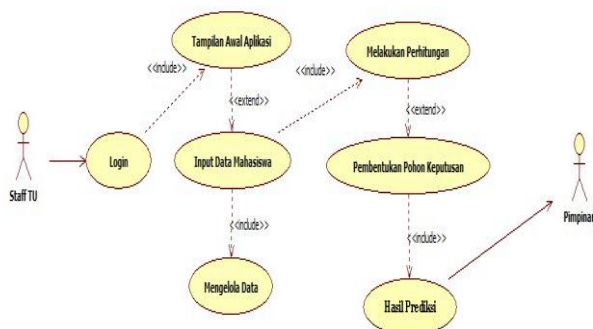
Gambar 4 Decision Tree

Pada pembentukan pohon keputusan, menghasilkan suatu *rule* seperti berikut :

if Nilai IPK = Rendah  
 then = Tidak Lulus  
 if Nilai IPK tinggi = Aktif and  
 Nilai TOEFL = Menengah atas  
 then = Lulus  
 if Nilai IPK = Tinggi and  
 Nilai TOEFL = Menengah bawah  
 then = Tidak Lulus

### 3. Use Case Diagram

Diagram Use Case di atas menunjukkan Actor yang terlibat di dalam Aplikasi perediksi kelulusan ada 2 Actor.



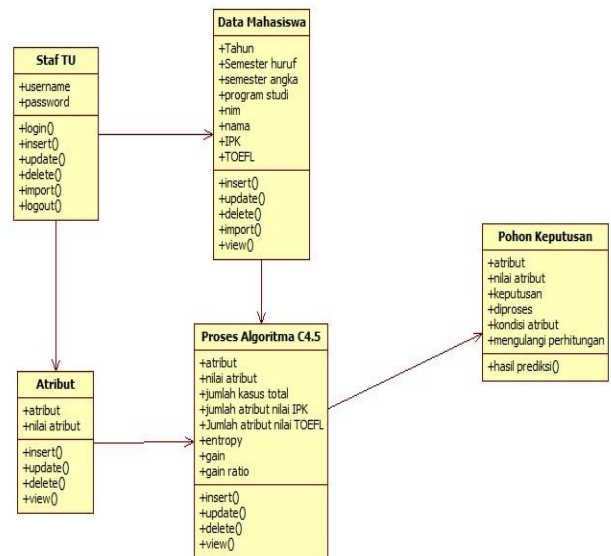
Gambar 5 Use Case Diagram

*Use case* menggambarkan kegiatan yang akan terjadi antara aktor dan aplikasi sistem. Diagram *use case* digunakan untuk memahami fungsi- apa saja yang ada dalam sebuah sistem.

### 4. Class Diagram

*Class Diagram* menggambarkan kelas-kelas dalam sebuah sistem dan hubungannya antara satu dengan

yang lain, serta dimasukan pula atribut dan operasi. *Class diagram* (diagram kelas) ini menggambarkan struktur sistem aplikasi yang diusulkan dari segi pendefinisian kelas-kelas yang akan dibuat untuk membuat aplikasi prediksi kelulusan mahasiswa.



Gambar 6 Activity Diagram

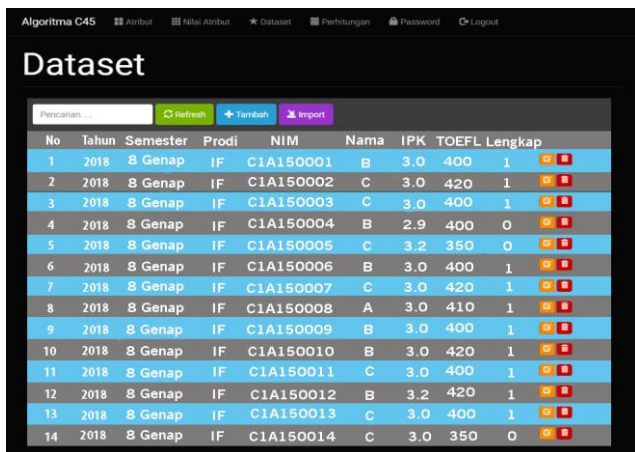
### 4. Pengujian Sistem

Berdasarkan perancangan yang dilakukan telah diketahui bahwa terdapat beberapa tampilan *form* pada aplikasi yang bertujuan untuk mempermudah *user* dalam pengoperasian-nya yang terdiri dari *Login*, halaman dataset, halaman atribut data, halaman nilai atribut data mahasiswa dan halaman perhitungan data mahasiswa. Berikut ini adalah tahapan pengujian dari berbagai *form*.

#### a. Tampilan Halaman Login

Gambar 7 Tampilan Halaman Login

## b. Tampilan Halaman Dataset Mahasiswa




No	Tahun	Semester	Prodi	NIM	Nama	IPK	TOEFL	Lengkap
1	2018	8 Genap	IF	C1A150001	B	3.0	400	1
2	2018	8 Genap	IF	C1A150002	C	3.0	420	1
3	2018	8 Genap	IF	C1A150003	C	3.0	400	1
4	2018	8 Genap	IF	C1A150004	B	2.9	400	0
5	2018	8 Genap	IF	C1A150005	C	3.2	350	0
6	2018	8 Genap	IF	C1A150006	B	3.0	400	1
7	2018	8 Genap	IF	C1A150007	C	3.0	420	1
8	2018	8 Genap	IF	C1A150008	A	3.0	410	1
9	2018	8 Genap	IF	C1A150009	B	3.0	400	1
10	2018	8 Genap	IF	C1A150010	B	3.0	420	1
11	2018	8 Genap	IF	C1A150011	C	3.0	400	1
12	2018	8 Genap	IF	C1A150012	B	3.2	420	1
13	2018	8 Genap	IF	C1A150013	C	3.0	400	1
14	2018	8 Genap	IF	C1A150014	C	3.0	350	0

Gambar 8 Tampilan Halaman Dataset Mahasiswa

Tampilan halaman dataset merupakan halaman awal aplikasi yang akan muncul setelah *user* berhasil melakukan *login* terlebih dahulu, dan sistem akan menampilkan data mahasiswa yang sebelumnya telah dimasukkan, data tersebut berupa tahun, semester, prodi (Program Studi), NIM (Nomor Induk Mahasiswa), Nilai IPK dan Nilai TOEFL.

## c. Tampilan Halaman Nilai Atribut Data Mahasiswa

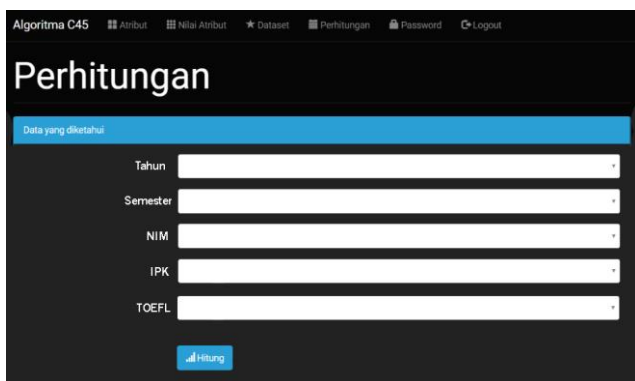


No	Nama Atribut	Nama Nilai Atribut
1	Nilai IPK	Tinggi
2	Nilai TOEFL	Rendah
3	Nilai IPK	Menengah Atas
4	Nilai TOEFL	Menengah Bawah

Gambar 9 Tampilan Halaman Nilai Atribut

Tampilan halaman nilai atribut ini merupakan atribut yang sangat berpengaruh dalam menentukan kelulusan mahasiswa Fakultas Teknologi Informasi Universitas Bale Bandung.

## d. Tampilan Halaman Perhitungan



Gambar 10 Tampilan Halaman Perhitungan

Tampilan halaman perhitungan ini merupakan tampilan untuk melakukan perhitungan data mahasiswa sehingga akan menghasilkan sebuah pohon keputusan.

## 5. Kesimpulan dan Saran

No	Tujuan	Input	Output diharapkan	Output sistem
1	Masuk kedalam menu <i>dashboard</i> aplikasi	<i>Username &amp; password</i>	Login dan aplikasi menampilkan <i>dashboard</i>	*Sukses * <i>Dashboard</i> tampil sesuai dengan <i>output</i>

Berdasarkan penelitian yang di lakukan penulis melalui beberapa tahapan yang di lakukan pada bab-bab sebelumnya maka penulis dapat menyimpulkan :

1. Dalam prediksi kelulusan mahasiswa menggunakan algoritma c4.5 pada data mahasiswa Fakultas Teknologi Informasi Universitas Bale Bandung, nilai IPK lebih dominan dalam prediksi kelulusan Mahasiswa dibandingkan dengan Nilai TOEFL.
  2. Hasil dari kinerja pohon keputusan menghasilkan nilai rata-rata *precision* 65%, *recall* 61% dan Akurasi 70%, sehingga nilai *confidence* 0,20 menghasilkan tingkat akurasi yang lebih baik dibandingkan dengan nilai *condifence* 0,15.
2. Saran

Dalam proses pembuatan Aplikasi *prediksi kelulusan mahasiswa*, penulis masih banyak sekali memiliki kekurangan dan jauh dari kata sempurna, sehingga penulis berharap untuk peneliti selanjutnya dapat mengembangkan lagi aplikasi ini, diantaranya :

1. Aplikasi ini masih memiliki banyak kekurangan dari segi *interface* dan fungsionalitasnya, maka alangkah lebih baik jika dapat di kembangkan lagi sehinga dapat lebih memudahkan user dalam pengoperasian aplikasi ini.
2. Aplikasi ini dapat ditambahkan fitur atau performansinya, selain itu dapat dikembangkan berbasis mobile dengan bahasa pemrograman lain

## 6. Daftar Pustaka

- Nuqson Masykur Huda. (2010). Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa.
- Mujib Ridwan, Hadi Suyono, M.Sarosa. (2013). PENERAPAN DATA MINING UNTUK EVALUASI KINERJA AKADEMIK MAHASISWA MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER.
- Agus Romadhona, Suprapedi, H.Himawan. (2017). Prediksi kelulusan mahasiswa tepat waktu berdasarkan usia, jenis kelamin, dan indeks prestasi menggunakan algoritma decesion tree.
- Anik Andriani. (2013). SISTEM KEPUTUSAN BERBASIS DECISION TREE DALAM PEMBERIAN BEASISWA STUDI KASUS: AMIK BSI YOGYAKARTA. *Semiar nasional*



*teknologi informasi dan komunikasi 2013*  
(SENTIKA 2013).

eko prasetyo. (2014). *DATA MINING-Mengolah data menjadi informasi menggunakan matlab*. Yogyakarta: ANDI.

Herdianto. (2013). Prediksi Kerusakan Motor Induksi Menggunakan Metode Jaringan Syaraf Tiruan Backpropagation. 8-9.

Indah Puji Astuti. (2017). Prediksi Ketepatan Waktu Kelulusan Dengan Algoritma Data Mining C4.5.