

Aplikasi Data Mining Menggunakan Algoritma ID3 Untuk Mengklasifikasi Kelulusan Mahasiswa Pada Universitas Dian Nuswantoro Semarang

Danny Himawan¹

^{1,3} Jurusan Teknik Informatika, FASILKOM UDINUS
Jln. Nakula 1 No 5-11 Semarang 50131 INDONESIA

¹111201005219@mhs.dinus.ac.id

Intisari — Seorang mahasiswa dinyatakan lulus tepat waktu jika dapat menyelesaikan pendidikan selama minimal 8 semester atau 4 tahun. Untuk mewujudkan hal tersebut, perkembangan nilai Indeks Prestasi sangat membantu dalam proses penyelesaian skripsi suatu mahasiswa. Dengan memantau hasil belajar di universitas berupa nilai IP tiap semester, seorang mahasiswa dapat dinyatakan lulus tepat waktu atau tidak. Pada penelitian ini dibuat aplikasi untuk Mengklasifikasi kelulusan mahasiswa menggunakan teknik data mining dengan algoritma ID3, hal tersebut bertujuan untuk melihat hasil perkembangan mahasiswa apakah dapat lulus tepat waktu atau tidak.

Kata kunci — Data Mining, Klasifikasi, ID3, Prediksi kelulusan mahasiswa UDINUS SEMARANG.

Abstract— One student graduated on time if it can complete education for a minimum of 8 semesters or 4 years. To achieve this goal, the development of value GPA is very helpful in the process of completion of a student thesis. By monitoring the learning outcomes in the form of the value of IP university each semester, a student can be passed on time or tidak. Pada this study made an application for graduation Classifying students using data mining techniques with ID3 algorithm, that aims to see whether the results of the development of students can graduate on time or not.

Keywords— Data Mining, Classification, ID3, Prediction Graduation UDINUS SEMARANG

I. PENDAHULUAN

Mahasiswa merupakan aspek penting dalam evaluasi keberhasilan penyelenggaraan program studi pada suatu perguruan tinggi. Dalam masa pembelajarannya, mahasiswa berkembang untuk meningkatkan pola pikir kecerdasan sebagai langkah awal untuk memasuki dalam persaingan di dunia kerja dan lingkungan masyarakat kelak [1]. Kualitas dari sebuah perguruan tinggi selain dapat dilihat dari rata-rata lama lulusannya mendapatkan pekerjaan juga dapat dilihat dari rata-rata masa atau lama studi dari mahasiswanya. Berdasar pada peraturan yang disampaikan dalam buku II standard dan prosedur tentang akreditasi institusi perguruan tinggi oleh BAN-PT (Badan Akreditasi Nasional Perguruan Tinggi) pada tahun 2011 menyatakan bahwa salah satu aspek penilaian akreditasi adalah mahasiswa dan lulusan. Khusus mengenai evaluasi standar mahasiswa dan lulusan, komponen yang dinilai adalah system rekrutmen mahasiswa baru, dan lulusan (rata-rata studi dan IPK) [1]. Jadi, tingkat kelulusan dan jumlah mahasiswa akan berpengaruh dalam proses akreditasi yang dilakukan oleh pemerintah. Salah satu permasalahan utama dari institusi perguruan tinggi adalah untuk meningkatkan kualitas pendidikan untuk mahasiswa dan untuk meningkatkan kualitas dari keputusan manajerial institusi. Sehingga tingkat kelulusan dan

jumlah mahasiswa akan berpengaruh dalam proses akreditasi yang dilakukan pemerintah. Salah satu cara untuk mencapai kualitas level mutu tertinggi dari sistem perguruan tinggi adalah dengan menggali pengetahuan dari data bidang

pendidikan sebagai atribut pembelajaran utama yang mempengaruhi pencapaian mahasiswa [2].

Untuk Mengklasifikasi data dilakukan dengan berbagai cara, salah satunya dapat pula dilakukan dengan menggunakan teknik datamining. Teknik data mining sebagai sebuah teknik analisis berbasis database dan statistic, sangatlah tepat untuk menggali berbagai potensi informasi yang terkandung dalam sebuah database.

Pendekatan data mining berkembang untuk mengatasi berbagai permasalahan menyangkut tentang pengolahan data. Beberapa peneliti menggunakan teknik data mining untuk menyelesaikan permasalahan prediksi kelulusan mahasiswa [3].

Data mining adalah suatu cara yang bertujuan dalam penemuan pola secara otomatis atau semi otomatis dari data yang sudah ada di dalam database atau sumber data lain yang dimanfaatkan untuk menyelesaikan suatu masalah melalui berbagai aturan proses [4].

II. STUDI PUSTAKA

2.1. Penelitian Terkait

Ada beberapa referensi yang diambil penulis sebagai bahan pertimbangan untuk penelitian yang dilakukan, referensi tersebut diambil dari beberapa penulisan yang dilakukan sebelumnya yang membahas permasalahan yang hampir sama, antara lain :

1. Analisa dan Komparasi Metode Klasifikasi Data Mining untuk Prediksi Kelulusan Mahasiswa. (Hilda Amalia, 2012)
2. Data Mining Classification untuk Prediksi Lama Masa Studi Mahasiswa Berdasarkan Jalur Penerimaan Dengan Metode Naïve Bayes (Jonh Fredrik Ulysses, 2012)
3. Graduation Prediction of Gunadarma Student Using Naïve Bayes and Decision Tree (Suhartina & Ernastuti, 2010)

2.2. Tinjauan Pustaka

A. Kelulusan Mahasiswa

Mahasiswa merupakan elite masyarakat yang mempunyai ciri intelektualitas yang lebih kompleks dibandingkan kelompok seusia mereka yang bukan mahasiswa, ataupun kelompok usia dibawah dan diatas mereka. Ciri intelektualitas tersebut adalah kemampuan mahasiswa menghadapi, memahami dan mencari cara pemecahan masalah secara lebih sistematis [1].

Kelulusan mahasiswa merupakan hal yang penting diperhatikan, karena penurunan jumlah kelulusan akan memberikan dampak negative kepada jumlah kelulusan dan akan menghilangkan jumlah pendapatan institusi dan juga akan berpengaruh pada penilaian pemerintah dalam bentuk status akreditasi institusi [8].

B. Data Mining

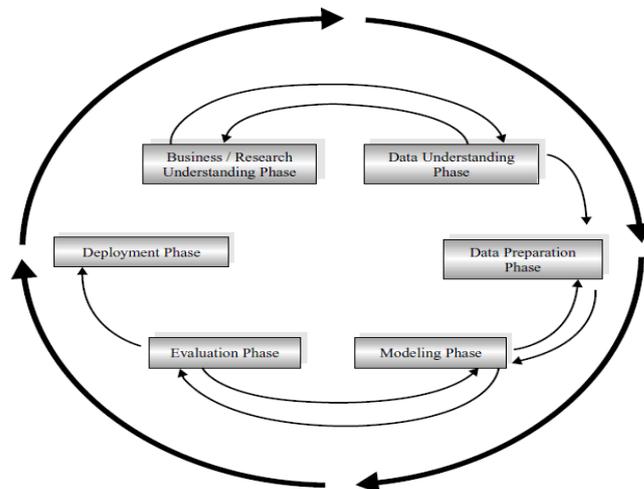
Data Mining merupakan teknologi baru yang sangat berguna untuk membantu perusahaan-perusahaan menemukan informasi yang sangat penting dari gudang data mereka. Beberapa aplikasi data mining focus pada prediksi, mereka meramalkan apa yang akan terjadi dalam situasi baru dari data yang menggambarkan apa yang terjadi di masa lalu [4]. Data mining adalah proses menggali informasi atau pola dalam data berukuran besar untuk menemukan pola dari suatu data untuk keperluan tertentu. Di dalam data mining, data disimpan secara elektronik dan pencarian data dalam data mining secara otomatis, atau dengan kata lain dijalankan secara komputerisasi. Masalah-masalah yang ada dalam data mining dipecahkan dengan menganalisis data yang sudah ada dalam database. Di dalam data mining, terdapat empat tahapan utama dalam pemrosesannya, yaitu :

1. Input (Data).
2. Metode (Algoritma Data Mining).
3. Output (Pola atau Model atau Knowledge).
4. Evaluation (Akurasi, AUC, RMSE, dan lain sebagainya).

Data mining merupakan gabungan dari teori dan heuristic, berfokus pada seluruh proses penemuan knowledge dan pola dari data. Peran utama dari data mining terbagi menjadi enam kategori, yaitu Deskripsi, Estimasi, Prediksi, Klasifikasi, Clustering, dan Asosiasi yang dimana setiap kategori memiliki algoritma masing-masing [10].

C. CRIPS-DM

Cross – Industry Standart Proses for Data Mining (CRIPS – DM) dikembangkan pada tahun 1996 oleh analis dari beberapa industry. CRIPS – DM menyediakan standart proses data mining sebagai pemecahan masalah secara umum dari bisnis atau unit penelitian [10]. CRIPS – DM memiliki siklus hidup yang terbagi dalam enam fase, yaitu :



Gambar 2.1 Siklus CRIPS – DM

1. Bussines Understanding

Merupakan tahap awal yang memahami tujuan dan kebutuhan dari sudut pandang bisnis, kemudian menterjemahkan pengetahuan ini ke dalam pendefinisian masalah dalam data mining . Selanjutnya akan ditentukan rencana dan strategi untuk mencapai tujuan tersebut.

2. Data Understanding

Dalam tahap ini dilakukan pengumpulan data, mengenali lebih lanjut data yang akan digunakan.

3. Data Preparation

Dalam tahap ini meliputi semua kegiatan untuk membangun dataset akhir (data yang akan diproses pada tahap permodelan/modeling) dari data mentah. Tahap ini dapat diulang beberapa kali. Pada tahap ini juga mencakup pemilihan tabel, record, dan atribut-atribut data, termasuk proses pembersihan dan transformasi data untuk kemudian dijadikan masukan dalam tahap permodelan (modeling).

4. Modeling

Dalam tahap ini akan dilakukan pemilihan dan penerapan berbagai teknik pemodelan dan beberapa parameternya akan disesuaikan untuk mendapatkan nilai yang optimal. Secara khusus, ada beberapa teknik berbeda yang dapat diterapkan untuk masalah data mining yang sama. Di pihak lain ada teknik pemodelan yang membutuhkan format data khusus. Sehingga pada tahap ini masih memungkinkan kembali ke tahap sebelumnya.

5. Evaluation

Pada tahap ini, model sudah terbentuk dan diharapkan memiliki kualitas baik jika dilihat dari sudut pandang analisa data. Pada tahap ini akan dilakukan evaluasi terhadap keefektifan dan kualitas model sebelum digunakan dan menentukan apakah model dapat mencapai tujuan yang ditetapkan pada fase awal (*Business Understanding*). Kunci dari tahap ini adalah menentukan apakah ada masalah bisnis yang belum dipertimbangkan. Di akhir dari tahap ini harus ditentukan penggunaan hasil proses *data mining*.

6. Deployment

Pada tahap ini, pengetahuan atau informasi yang telah diperoleh akan diatur dan dipresentasikan dalam bentuk khusus sehingga dapat digunakan oleh pengguna. Tahap *deployment* dapat berupa pembuatan laporan sederhana atau mengimplementasikan proses *data mining* yang berulang dalam perusahaan. Dalam banyak kasus, tahap *deployment* melibatkan konsumen, di samping analisis data, karena sangat penting bagi konsumen untuk memahami tindakan apa yang harus dilakukan untuk menggunakan model yang telah dibuat.

D. Algoritma ID3

Algoritma ID3 atau *Iterative Dichotomizer 3* (ID3) merupakan sebuah metode yang digunakan untuk membuat pohon keputusan. Input dari algoritma ini adalah sebuah database dengan beberapa variabel yang juga dikenal dengan atribut [14]. Setiap masukan dalam database menyajikan sebuah objek dari domain yang disebut dengan variabel bebas. Sebuah atribut didesain untuk mengklasifikasikan objek yang disebut dengan variabel tidak bebas.

Proses klasifikasi dilakukan dari node paling atas yaitu akar pohon (*roof*). Dilanjutkan ke bawah melalui cabang-cabang sampai dihasilkan node daun (*leafes*) dimana node daun ini menunjukkan hasil akhir klasifikasi. Sebuah objek yang diklasifikasi dalam pohon harus dites nilai entropinya. Entropi adalah ukuran dari teori informasi yang dapat mengetahui karakteristik *impurity* dan *homogeneity* dari kumpulan data. Dari nilai entropi tersebut kemudian dihitung nilai *information gain* (IG) masing-masing atribut *independent* terhadap atribut *dependent*-nya.

$$Entropy(S) = -P_+ \log_2 P_+ - P_- \log_2 P_-$$

Dimana :

1. S adalah ruang (data) sample yang digunakan untuk training.
2. P+ adalah jumlah yang bersolusi positif (mendukung) pada data sample untuk kriteria tertentu.
3. P- adalah jumlah yang bersolusi negative (tidak mendukung) pada data sample untuk kriteria tertentu. Dari rumus *entropy* diatas dapat disimpulkan bahwa definisi *entropy* (S) adalah jumlah bit yang diperkirakan

dibutuhkan untuk dapat mengekstrak suatu kelas (+ atau-) dari sejumlah data acak pada suatu ruang sampel S. Entropy bisa dikatakan sebagai kebutuhan bit untuk meyakinkan suatu kelas. Semakin kecil nilai entropy maka semakin baik digunakan dalam mengekstraksi suatu kelas.

Pajang kode untuk menyatakan informasi secara optimal adalah $-\log_2 p$ bits untuk *messages* yang mempunyai probabilitas p. Sehingga jumlah bit yang diperkirakan untuk mengekstraksi S ke dalam kelas adalah : $-p \log_2 p + -p \log_2 p$.

IG merupakan nilai rata-rata entropi pada semua atribut. Secara ringkas, langkah kerja algoritma ID3 dapat digambarkan sebagai berikut [7]:

1. Penghitungan *information gain* dari setiap atribut menggunakan

$$Gain(S, A) = Entropy(S) - \sum_{v \in \text{nilai}(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Dimana

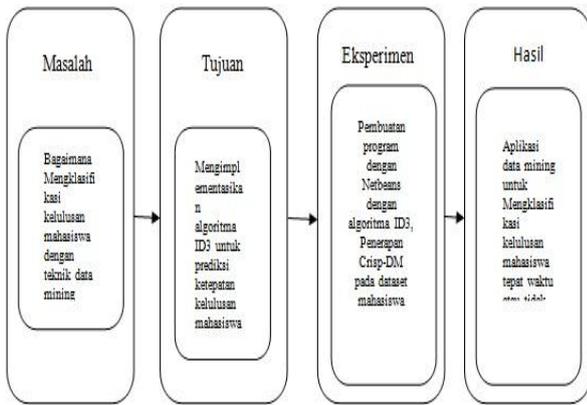
- A : atribut
- V : suatu nilai yang mungkin untuk atribut A
- Value (A) : himpunan yang mungkin untuk atribut A
- |Sv| : Jumlah sampel untuk nilai v
- |S| : jumlah seluruh sampel data
- Entropy(Sv) : entropy untuk sampel-sampel yang memiliki nilai v

2. Pemilihan atribut yang memiliki nilai *information gain* terbesar,

3. Pembentukan simpul yang berisi atribut tersebut, Ulangi proses perhitungan *information gain* akan terus dilaksanakan sampai semua data telah masuk dalam kelas yang sama. Atribut yang telah dipilih tidak diikuti lagi dalam perhitungan nilai *information gain*.

2.3. Kerangka Pemikiran

Masalah yang ditemui pada penelitian ini adalah bagaimana cara memprediksi ketepatan kelulusan mahasiswa dengan menggunakan teknik *data mining*. Dimana faktor nilai IPK sampai dengan semester IV dan faktor data induk mahasiswa yang berupa jenis kelamin, kota tempat tinggal, status bekerja atau tidak, dan status pernikahan akan menjadi variable prediktornya. Metode yang akan digunakan adalah klasifikasi dengan algoritma ID3.



Gambar 2. 2 Kerangka Pemikiran

III. METODOLOGI PENELITIAN

3.1. Instrumen Penelitian

Penelitian ini dilakukan berdasarkan permasalahan yang telah di uraikan pada bab sebelumnya. Adapun metode penelitian dalam penelitian ini adalah :

A. Pengumpulan Data

Data pada penelitian ini diambil dari PSI UDINUS Semarang yang beralamat di Jalan Nakula I No. 5-11, Semarang Provinsi Jawa Tengah pada bulan Maret 2014. Berupa data mahasiswa Fakultas Ilmu Komputer UDINUS angkatan tahun 2008, 2009 dan 2010.

B. Metode pengembangan perangkat lunak

1. Perancangan dan analisis

Dalam penelitian ini, digunakan *Cross Industry Standard Processing-Data Mining* (CRISP-DM) sebagai metode analisis data dan pemodelan dengan tahapan fase sebagai berikut :

- Fase Pemahaman Bisnis (*Business Understanding Phase*)
- Fase Pemahaman Data (*Data Understanding Phase*)
- Fase Pengolahan Data (*Data Preparation Phase*)
- Fase Pemodelan (*Modeling Phase*)
- Fase Evaluasi (*Evaluation Phase*)
- Fase Penyebaran (*Deployment Phase*)

2. Pembuatan Program

Pembuatan program adalah tahap coding software analisis data mining menggunakan bahasa pemrograman JAVA, library SPMF untuk menangani tugas data mining dan Netbeans IDE sebagai perangkat pembangunan. Beberapa kegiatan yang dilakukan pada pembuatan program meliputi:

- Desain alur program.
- Desain User Interface (GUI).

- Penyisipan algoritma ID3 pada program.
- Penyisipan dataset training
- Pembuatan fitur simpan hasil mining ke dalam bentuk *.txt sehingga dapat disimpan / dicetak hasil miningnya.
- Pembuatan fitur input untuk menginputkan atribut untuk proses mining.
- Pembuatan fitur bantuan untuk memandu user awam menggunakan program.

3. Pengujian software

Tahap pengujian adalah tahap membenahan fungsi program, pencarian dan koreksi kesalahan program baik kesalahan sintaks (Syntax Error) maupun kesalahan saat program dijalankan. Dalam pembangunan software ini, menggunakan metode pengujian White Box dan Black Box.

4. Evaluasi Kualitas Software

Adalah tahap penilaian secara objektif yang dilakukan dengan penyerahan program dan pengisian kuisioner seputar performa program prediksi data mining pada mahasiswa FIK UDINUS yang dipilih secara acak dan tidak mengetahui tentang data mining. Untuk poin-poin yang akan diuji pada tahap evaluasi kualitas software adalah :

- Probabilitas, pengujian kualitas apakah software cukup portable dalam hal instalasi.
- Kemudahan, pengujian kualitas apakah software cukup mudah untuk dioperasikan untuk pengguna awam.
- Desain, pengujian kualitas apakah desain User Interface mudah dipahami atau tidak.
- Kegunaan, pengujian kualitas seberapa besar manfaat yang nantinya dapat diperoleh dengan adanya modul data mining.

IV. HASIL DAN PEMBAHASAN

4.1. Perancangan dan Analisis

A. Perancangan sistem

Teknik pemecahan masalah pada analisis system yang menggabungkan kembali komponen system menjadi system yang utuh. Ada tiga point utama yang akan dibahas pada tahap perancangan :

4.1.1 User Communication

Komunikasi yang terjadi antara pengguna dengan program dengan respon yang diberikan oleh system. Yaitu analisis kebutuhan user, use case diagram, dan diagram aktifitas.

4.1.1.1 Analisis Kebutuhan User

Analisis kebutuhan bertujuan mendefinisikan kebutuhan yang diperlukan dalam pengembangan perangkat lunak. Suatu system yang akan dikembangkan harus memenuhi kebutuhan fungsional

sehingga program dapat berjalan dengan optimal sehingga mendukung fungsi utama dalam program agar berjalan sesuai yang diharapkan.

1. Kebutuhan Fungsional
 - a. Sistem harus menyediakan fungsi untuk mengimpor dataset sebagai data training.
 - b. Sistem harus menyediakan fungsi untuk mining sebagai layanan utama dalam program ini.
 - c. Sistem harus menyediakan fungsi untuk menyimpan hasil mining sebagai output dalam bentuk file teks.
2. Kebutuhan Non-Fungsional
 - a. Sistem menyediakan layanan bantuan agar pengguna lebih mudah dalam mengoperasikan program

4.1.1.2 Use Case Diagram

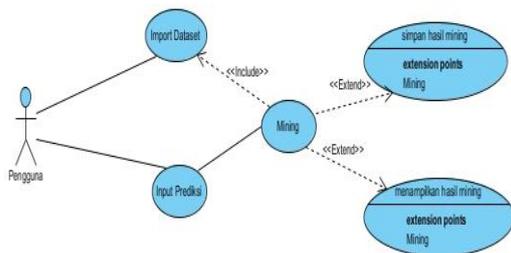
Use case diagram yang akan disusun pada perangkat lunak ini memerlukan actor sebagai pelaku yang menjalankan *case-case* yang ada. Berikut adalah langkah membuat diagram use case dalam software prediksi kelulusan mahasiswa pada Universitas Dian Nuswantoro Semarang:

1. Identifikasi Aktor

Aktor adalah pelaku bisnis atau subjek yang menjalankan case pada software. Aktor yang terlibat dalam system ini secara khusus adalah petugas akademik fakultas yang memantau perkembangan nilai mahasiswa. Namun baik pengguna secara umum maupun petugas akademik UDINUS Semarang memiliki case yang sama, hal ini dikarenakan software yang dibuat tidak memerlukan otentikasi sebagai lini keamanan sehingga software ini dapat digunakan oleh semua pengguna.

2. Diagram Model Use Case

Adalah penjabaran secara grafis kegiatan yang dilakukan oleh pengguna selama berada didalam system atau software

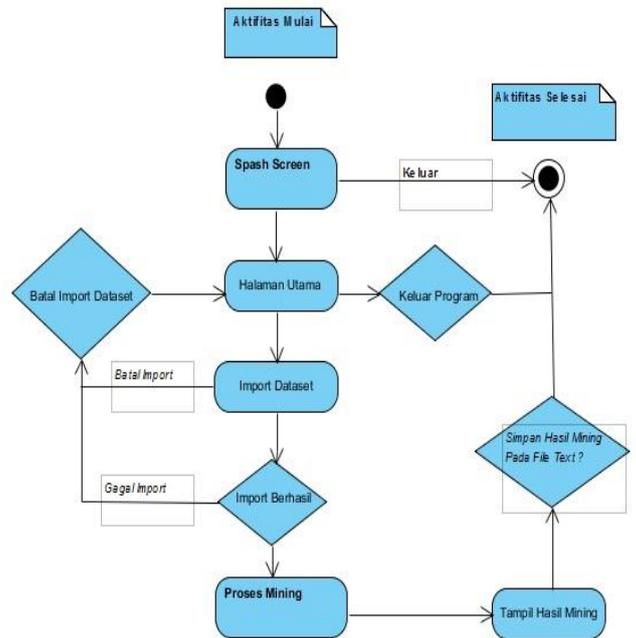


Gambar 4.1 Use Case Diagram

Diagram *use case* di atas menggambarkan bahwa pengguna dapat melakukan impor dataset sebagai inputan proses mining, proses mining, menampilkan hasil mining, dan menyimpan hasil proses mining. Untuk melakukan mining,

pengguna diharuskan mengimpor dataset sebagai inputan untuk proses mining, setelah pengguna berhasil melakukan proses mining, hasil proses yang terbentuk dapat disimpan pada file teks.

4.1.1.3 Activity Diagram



Gambar 4.2 Activity Diagram

Gambar di atas menjelaskan bagaimana pengguna melakukan proses mining. Ketika pertama kali pengguna menjalankan program, pengguna akan masuk ke dalam halaman *spashscreen*. Pengguna dapat memilih apakah masuk halaman utama atau keluar dari program. Aktifitas akan berhenti jika pengguna memilih keluar dari program. Jika pengguna memilih masuk maka pengguna akan masuk ke dalam halaman utama.

Untuk melakukan proses mining pengguna diharuskan mengimpor dataset training sebagai data training pada proses mining. Pengguna dapat membatalkan impor dataset dan kembali pada halaman utama. Jika proses impor dataset gagal maka proses impor akan dibatalkan dan kembali pada halaman utama.

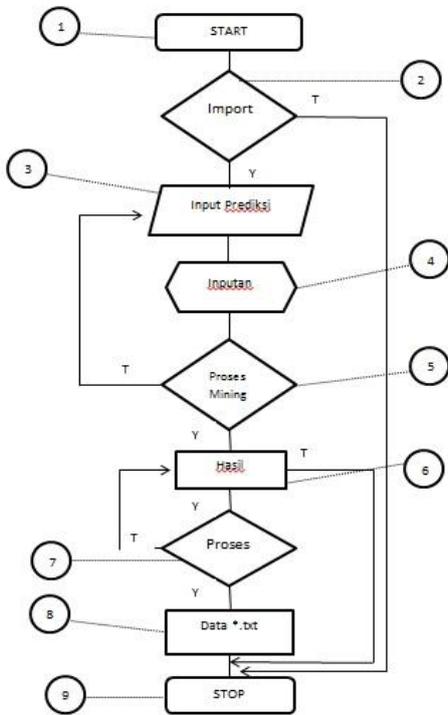
Setelah dataset diimpor oleh pengguna dan berhasil, maka proses mining dapat dilakukan. Setelah proses impor berhasil pengguna masuk ke dalam form inputan berupa kategori kategori yang akan di prediksikan. Setelah pengisian inputan maka proses mining dilakukan. Setelah proses mining berhasil program akan menampilkan hasil mining berupa statistic mining yang dapat disimpan pada bentuk file teks dan aktifitas selesai.

4.3. Pengujian dan implementasi

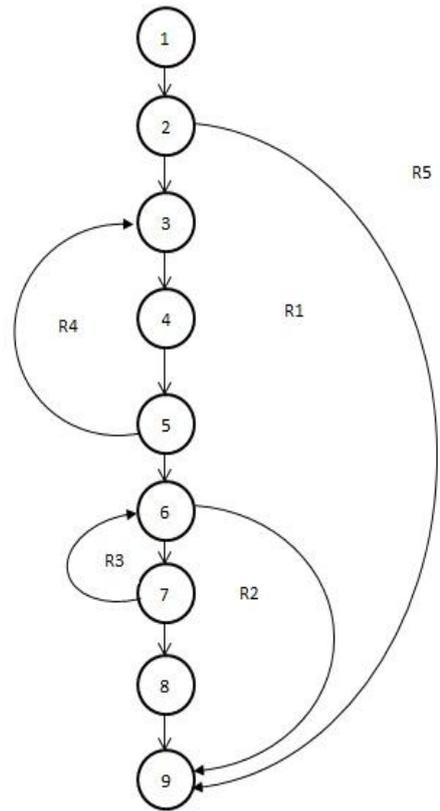
Sebelum perangkat lunak dirilis untuk pengguna hal yang perlu dilakukan adalah melakukan pengujian apakah program sudah memenuhi standar kelayakan pakai atau terdapat bug yang menyebabkan program tidak dapat berjalan dengan baik.

4.3.1 Pengujian Perangkat Lunak Dengan Metode White-Box Testing

Pengujian white box bertujuan untuk memastikan struktur semua statemen pada program telah dieksekusi paling tidak satu kali pengujian dan tidak dijumpai eror message. Pengujian ini menggunakan basis path yang memungkinkan pengukuran kompleksitas logis dari desain procedural sebagai pedoman penetapan basis set pada tiap eksekusi. Dalam penelitian ini test care dilakukan hanya satu kali untuk memastikan program sudah dapat berjalan dengan prosedur perancangan. Adapun contoh pengujian yang diambil dalam penelitian ini adalah proses output.



Gambar 4.10 Bagan alir evaluasi



Gambar 4.11 Flowgraph Evaluasi

Dari gambar 4.11 dapat diketahui :

Edge (E)	= 12
Region (R)	= 5
Predikat Node	= 5
Node (N)	= 9

a. Menghitung Nilai Cyclomatic Complexity (CC)

Cyclomatic Complexity digunakan untuk mencari jumlah path dalam satu flowgraph. Cyclomatic Complexity $V(G)$ untuk grafik alir dihitung dengan rumus:

$$V(G) = E - N + 2$$

Dimana :

E = jumlah edge pada grafik alir
N = jumlah node pada grafik alir

maka

$$V(G) = 12 - 9 + 2 = 5$$

Jadi Cyclomatic Complexity untuk flowgraph evaluasi adalah 5. Berdasarkan tabel hubungan antara Cyclomatic Complexity dan Resiko menurut McCabe, menunjukkan bahwa nilai CC 5 – 10 masuk dalam type of procedure a well structured and stable procedure (strukturnya baik dan procedure stabil) serta risikonya Low (rendah).

b. Menentukan basis path

Basis set yang dihasilkan dari jalur independent secara linier adalah jalur sebagai berikut :

Tabel 4.1 Pengujian basis path

No	Path	Input	Output	Ket.
1	1-2-3-4-5-6-7-8-9	- Klik tombol impor pada menu program - Pilih dataset yang akan diimpor - Inputkan yang akan diprediksi - Klik proses mining - Klik tombol simpan file pada menu program - Keluar Program	-Tampil halaman impor -Tampil dataset training -Tampil inputan prediksi -Tampil simpan hasil	Ok
2	1-2-9	- Klik tombol impor - Batal Impor - Keluar Program	- Tampil halaman impor - Klik keluar tombol - Keluar Program	Ok
3	1-2-3-4-5-3-6-7-8-9	- Klik tombol impor pada menu program - Pilih dataset yang akan diimpor - Inputkan yang akan diprediksi - Salah input, kembali ke inputan - Inputkan kembali - Klik proses mining - Klik tombol simpan file pada menu program - Keluar Program	- Tampil halaman impor - Tampil dataset training - Tampil inputan prediksi - Kembali ke inputan prediksi - Tampil simpan hasil	Ok
4	1-2-3-4-5-6-9	- Klik tombol impor pada menu program - Pilih dataset yang akan diimpor - Inputkan yang akan diprediksi - Klik proses mining - Tampil hasil mining - Keluar Program	- Tampil halaman impor - Tampil dataset training - Tampil input prediksi - Tampil hasil prediksi	Ok
5	1-2-3-4-5-6-7-6-7-8-9	- Klik tombol impor pada menu program - Pilih dataset yang akan diimpor - Inputkan yang akan diprediksi - Klik proses mining - Inputan salah kembali input - Proses mining - Simpan hasil mining - Keluar Program	- Tampil halaman impor - Tampil dataset training - Tampil input prediksi - Keluar pesan salah input - Kembali ke menu input - Tampil input prediksi - Tampil hasil mining - Tampil simpan mining	Ok

Pada saat aplikasi dijalankan, makan terlihat bahwa semua basis set tereksekusi dengan baik dan simpul telah

dieksekusi satu kali. Berdasarkan ketentuan tersebut dari segi kelayakan software, system ini telah memenuhi syarat.

4.3.2 Pengujian Perangkat Lunak Dengan Black-Box Testing

Pengujian black box dilakukan untuk memastikan bahwa suatu event atau masukan akan menjalankan proses yang tepat dan menghasilkan output sesuai dengan rancangan. Untuk contoh pengujian terhadap beberapa proses memberikan hasil sebagai berikut:

Tabel 4.2 Hasil Pengujian Black Box Terhadap Beberapa Proses

Input / Event	Fungsi	Hasil yang Diharapkan	Hasil Uji
Klik Menu Utama / Program	Menampilkan menu program	Halaman menu program tampil	Sesuai
Klik Menu Impor Dataset	Menampilkan menu impor dataset	Halaman menu impor tampil	Sesuai
Klik Menu Save file	Menampilkan menu safe file	Halaman safe file tampil	Sesuai
Klik Menu Reset	Mereset semua inputan dan dataset	Dataset dan inputan tereset	Sesuai
Klik Menu Keluar Program	Keluar Program	Keluar Program	Sesuai
Klik tombol Help	Menampilkan menu help	Halaman menu help tampil	Sesuai
Klik tombol Proses mining	Memproses hasil mining	Hasil mining tampil di halaman hasil	Sesuai
Klik Tombol Update basis pengetahuan	Menambah data pada dataset training	Halaman penambahan data tampil	Sesuai
Klik Menu Hasil	Menampilkan data saat proses mining terlaksana	Halaman hasil tampil	Sesuai
Klik Menu Training Dataset	Menampilkan dataset saat setelah impor dataset	Halaman Training dataset tampil	Sesuai

Dari hasil pengujian dapat disimpulkan untuk uji black box yang meliputi uji input proses dan output dengan acuan rancangan perangkat lunak telah terpenuhi dengan hasil sesuai dengan rancangan. Uji juga dilakukan pada program utama dan program pendukung lainnya.

4.4 Hasil Penelitian

Pada pembahasan, penulis akan mengevaluasi hasil prediksi mining dengan data real yang sudah terjadi, sehingga terlihat seberapa akurat proses mining pada program aplikasi data mining ini, berikut hasil penelitiannya

4.4.1 Evaluasi Hasil Data Mining Menggunakan 100 Dataset

Hasil penelitian menggunakan data mahasiswa Teknik Informatika UDINUS angkatan tahun 2009 menghasilkan hasil sebagai berikut :

Tabel 4.3 Hasil Uji 1

No.	Hasil Asli	Hasil Prediksi
1	Tidak Tepat	Tidak Tepat
2	Tepat	Tepat
3	Tidak Tepat	Tidak Tepat
4	Tidak Tepat	Tepat
5	Tepat	Tepat
6	Tepat	Tepat
7	Tidak Tepat	Tidak Tepat
8	Tepat	Tidak Tepat
9	Tepat	Tepat
10	Tidak Tepat	Tidak Tepat

Dari tabel diatas, dihasilkan prediksi menggunakan 100 dataset dengan 10 sampel prediksi adalah 2 salah dan 8 sampel benar.

4.4.2 Evaluasi Hasil Data Mining Menggunakan 1000 Dataset

Hasil penelitian menggunakan data mahasiswa Teknik Informatika UDINUS angkatan tahun 2009 menghasilkan hasil sebagai berikut :

Tabel 4.4 Hasil Uji 2

No.	Hasil Asli	Hasil Prediksi
1	Tidak Tepat	Tepat
2	Tidak Tepat	Tidak Tepat
3	Tidak Tepat	Tidak Tepat
4	Tidak Tepat	Tidak Tepat
5	Tidak Tepat	Tidak Tepat
6	Tidak Tepat	Tidak Tepat
7	Tidak Tepat	Tidak Tepat
8	Tidak Tepat	Tidak Tepat
9	Tepat	Tidak Tepat
10	Tidak Tepat	Tidak Tepat

Dari tabel diatas, dihasilkan prediksi menggunakan 1000 dataset dengan 10 sampel prediksi adalah 2 salah dan 8 sampel benar.

4.4.3 Evaluasi Hasil Data Mining Menggunakan 1469Dataset

Hasil penelitian menggunakan data mahasiswa Teknik Informatika UDINUS angkatan tahun 2009 menghasilkan hasil sebagai berikut :

No.	Hasil Asli	Hasil Prediksi
1	Tidak Tepat	Tidak Tepat
2	Tidak Tepat	Tidak Tepat
3	Tepat	Tepat
4	Tidak Tepat	Tidak Tepat
5	Tidak Tepat	Tidak Tepat
6	Tidak Tepat	Tidak Tepat
7	Tepat	Tepat
8	Tidak Tepat	Tidak Tepat
9	Tepat	Tepat
10	Tidak Tepat	Tidak Tepat

Dari tabel diatas, dihasilkan prediksi menggunakan 100 dataset dengan 10 sampel hasilnya adalah benar semua.

V. PENUTUP

A. Kesimpulan

Dari hasil penelitian yang telah dilakukan didapat kan bahwa pada percobaan dengan 100 dataset dengan 10 sampel uji hasil yang diharapkan tidak mencapai yang diinginkan karena dengan 10 sampel uji 8 sampel tepat dan 2 sampel tidak tepat. Sama hal nya dengan 1000 dataset dengan 10 sampel uji hasil yang diharapkan tidak mencapai yang diinginkan karena dengan 10 sampel uji 8 sampel tepat dan 2 sampel tidak tepat. Sedangkan pada percobaan dengan seluruh dataset / 1469 data dengan 10 sampel uji, semua sampel uji tepat. Sehingga dapat disimpulkan bahwa semakin sedikit data training yang digunakan maka hasil klasifikasinya menunjukkan ketidakakuratan yang tinggi. Sebaliknya jika data training yang digunakan semakin banyak hasilnya akan semakin akurat. Software klasifikasi yang telah disusun dapat digunakan untuk menentukan klasifikasi potensi ketepatan kelulusan pada mahasiswa UDINUS Semarang.

B. Saran

Penelitian ini memiliki banyak kekurangan dan masih dapat dikembangkan lagi pada penelitian berikutnya. Beberapa saran untuk pengembangan penelitian selanjutnya adalah

1. Dataset sebagai dataset training ditingkatkan agar pola mining yang terbentuk semakin bervariasi.
2. Variabel prediksi ditambahkan agar hasil prediksi mendekati sempurna.
3. Penelitian ini menggunakan dataset dari UDINUS SEMARANG, untuk penelitian selanjutnya dapat menggunakan dataset yang lainnya, dari dataset private atau dataset public.
4. Program pada penelitian ini masih berupa model prototype sehingga masih dapat dikembangkan lagi menggunakan tools atau bahasa pemrograman yang lainnya.
5. Hasil proses mining yang terbentuk dapat dikembangkan menjadi basis pengetahuan untuk system pendukung keputusan maupun system rekomendasi yang akan dilakukan oleh mahasiswa bila hasil yang didapat tidak sesuai dengan yang diinginkan.
6. Penelitian selanjutnya dapat dikembangkan pada jenis data yang serupa dengan metode yang lebih baik seperti c50 atau j48.

UCAPAN TERIMA KASIH

Penulis mengucapkan teima kasih kepada Universitas Dian Nuswantoro, Rektor UDINUS, Dekan Fakultas Ilmu Komputer, Kaprodi Teknik Informatika-S1, Dosen pembimbing, Dosen-dosen pengampu kuliah di Fakultas Ilmu Komputer , serta teman-teman dan sahabat yang selama ini

telah mendampingi penulis selama kuliah di Universitas Dian Nuswantoro.

REFERENSI

- [1] Muarif, Khoirul. (2013).Komparasi Pemodelan Data Menggunakan C4.5 Dan C4.5 Berbasis Particle Swarm Optimization Untuk Memprediksi Kelulusan Mahasiswa. Semarang :Universitas Dian Nuswantoro.
- [2]Abu Tair Mohammed M., El-Halees Alaa M. Mining, Educational Data to Improve Students' Performance: A Case Study, 2012, International Journal of Information and Communication Technology Research, Volume 2 No. 2, pp 140-146.
- [3]Suhartina & Ernastuti.(2010).Graduation Prediction of Gunadarma Student Using Naive Bayes and Decision Tree.Jakarta:Universitas Gunadarma.
- [4]Witten, H.I.,Eibe,F.,& Hall,A.M.(2010).Data Mining Machine Learning Tools and Techniques.Burlington:Morgan Kaufmann Publisher
- [5]Vercellis, Carlo (2009). Business Intelligent: Data Mining and Optimization for Decision Making. Southern Gate. Chichester, West Sussex: John Willey & Sons,Ltd.
- [6]Gorunescu, Florin. (2011). Data Mining: Concepts and Techniques. Verlag berlin Heidelberg: Springer.
- [7] Defiyanti Sofi.(2010). Perbandingan Kinerja Algoritma ID3 Dan C4.5 Dalam Klasifikasi Spam-Mail. Bekasi:Universitas Gunadarma.
- [8]Amalia, Hilda. (2012).”Komparasi Metode Klasifikasi Data Mining Untuk Prediksi Ketepatan Kelulusan Mahasiswa. Jakarta : STIMIK Nusa Mandiri.
- [9]Quadri, M. N., & Kalyankar, N. V. (2010). Drop Out Feature of Student Data for Academic Performance Using Decision Tree techniques. Global Journal of Computer Science and Technology, 2-4.
- [10]Larose, D. T. (2005). Discovering Knowledge in Data. New Jersey: John Willey & Sons, Inc.
- [11]Han, J., & Kamber, M. (2007). Data Mining Concepts and Techniques. San Fransisco: Mofgan Kaufan Publisher.
- [12]Wu, X., & Kumar, V. (2009). The Top Ten Algorithms in Data Mining. Taylor & Francis Group, LLC.
- [13]Kusrini, & Luthfi, E. T. (2009). Algoritma Data Mining. Yogyakarta: Andi Publishing.
- [14]Giovani Ronny, Mudjihartono Paulus, Pranowo. (2011). Sistem Pendukung Keputusan Prediksi Kecepatan Studi Mahasiswa Menggunakan Metode ID3. Yogyakarta: Universitas Atma Jaya.
- [15]P.J.Deitel, Java(tm) for Programmers: Deitel(r) Developer Series. Prentice Hall, 2009.
- [16] I. Sommerville, Software Engineering, 9th ed. Boston: Education, Inc, 2011.
- [17] J. M. Kerr and R. Hunter, Inside RAD: How to Build a Fully Functional System in 90 Days or Less. McGraw-Hill., 1993.