

PENERAPAN ALGORITMA KLASIFIKASI DATA MINING ID3 UNTUK MENENTUKAN PENJURUSAN SISWA SMAN 6 SEMARANG

Obbie Kristanto¹

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

¹111201005270@mh.s.dinus.ac.id

Intisari— Kurikulum yang digunakan pada Sekolah Menengah Atas / SMA saat ini adalah kurikulum 2013. Dalam perkembangan kurikulum yang masih baru ini terdapat beberapa perbedaan dengan kurikulum sebelumnya. Salah satunya yaitu perbedaan dalam menentukan penjurusan pada siswa-siswi SMA. Hal ini dikarenakan penjurusan dilakukan pada kelas X. Terlebih lagi penjurusan dilakukan dengan cara manual yang memerlukan banyak waktu dan dianggap tidak efisien. Tujuan dilakukan penelitian ini adalah untuk membantu SMAN6 Semarang dalam penjurusan siswa-siswinya dan mengetahui tingkat akurasi algoritma ID3 bila dibandingkan dengan hasil penjurusan siswa sebelumnya. Metode yang digunakan adalah *Iterative Dichotomiser 3* atau yang lebih populer dengan sebutan ID3. ID3 membentuk pohon keputusan yang nantinya dapat digunakan sebagai dasar pertimbangan dalam menentukan penjurusan di SMA, dan tools yang digunakan dalam membuat aplikasi ini adalah pemrograman Java. Hasil dari penelitian ini sendiri berupa aplikasi *data mining* untuk memprediksi penjurusan IPA atau IPS. Kesimpulan yang didapat adalah aplikasi sesuai dengan kebutuhan dan tujuan penulis, dengan menggunakan 371 dataset dan 20 data uji menghasilkan akurasi sebesar 80%.

Abstract— Indonesian high school studies use 2013 curriculum. There are a number of differences in this curriculum from the one previously in place. One of them is the method which decides the high school students' specific major. The fact that the major division is made in grade X causes it. Furthermore, majors done manually which takes time and is not considered efficient. This research is done to help students and teachers alike to face this issue better and to find out the accuracy of the algorithmic implementation of ID3 when compared to data from previous student majors. The method used in this research is the Iterative Dichotomizer 3, or better known as the ID3. ID3 makes a decision tree which will later be used as consideration to decide which major a student will be assigned to while s/he is studying in high school, and this application used Java Programming as a tools. The result of this research is a data mining application to predict whether a student should take natural science or social science. The conclusion is in accordance with the needs of the application and purpose of the author, by using 371 datasets and 20 data trainings which generate accuracy by 80%.

Keywords— Data Mining, ID3, Classification, Student Majority, High School

I. PENDAHULUAN

Dalam kurikulum yang baru penjurusan dilakukan pada saat awal masuk, yaitu pada kelas 10. Perubahan kurikulum dimaksudkan agar memungkinkan penyesuaian program pendidikan pada satuan pendidikan dengan kondisi dan kekhasan potensi yang ada di daerah peserta didik [14]. Kemungkinan yang akan terjadi jika siswa mengalami kesalahan dalam penjurusan adalah rendahnya prestasi belajar siswa atau dapat menyebabkan terjadinya ketidakcocokan dengan jurusan yang sudah dipilih oleh siswa atau siswi sebelumnya [12].

Permasalahan yang terjadi dapat bermacam-macam dari hal yang sepele sampai masalah penting yang nantinya berdampak pada pemilihan minat pada saat di perguruan tinggi. Tidak jarang juga siswa-siswi yang asal-asalan dalam menentukan jurusan yang akan mereka ambil, tidak mempunyai tujuan setelah mereka lulus sekolah menengah atas, dan lain-lain berakibat pada menurunnya prestasi belajar mereka [1]. Untuk mengatasi masalah tersebut siswa hanya

berkonsultasi secara langsung kepada guru wali kelas, BP atau dengan orangtua masing-masing yang dilakukan secara manual sehingga memakan waktu yang cukup lama.

Tetapi seiring dengan perkembangan teknologi hal itu dapat diatasi dengan teknik pengelompokan data didapat dari hasil data mining. Data mining sendiri adalah disiplin ilmu yang mempelajari metode untuk mengekstrak pengetahuan atau pola dari suatu data, sehingga data mining juga sering disebut knowledge discovery in database [13]. Data mining dapat digunakan untuk mengelompokkan data, memprediksi, mengestimasi, dan menentukan kaidah asosiasi dalam suatu data yang ada. Perlunya data mining karena adanya sejumlah besar data yang dapat digunakan untuk menghasilkan informasi dan knowledge yang berguna. Informasi dan knowledge yang didapat tersebut dapat digunakan untuk mengetahui suatu pola dalam suatu data yang banyak, terlebih lagi besarnya kebutuhan untuk mengubah data tersebut menjadi informasi yang berguna. Metode data mining bermacam-macam disesuaikan dengan kebutuhan yang ada. Untuk kasus ini penulis menggunakan algoritma klasifikasi

data mining ID3 untuk memprediksi siswa-siswi dalam menentukan pilihan jurusan yang akan mereka ambil, sehingga nantinya dapat menjadi dasar dalam pengambilan keputusan. Pemilihan algoritma ini dikarenakan algoritma ID3 cocok untuk digunakan dalam memproses data mining jika data atribut yang terdapat pada dataset berupa data kategorikal. Selain itu algoritma ID3 juga dapat men-generate pohon keputusan yang simple dan spesifik, serta eliminasi perhitungan-perhitungan yang tidak diperlukan, karena ketika menggunakan metode pohon keputusan maka sampel diuji hanya berdasarkan kriteria atau kelas tertentu. Pohon keputusan yang dihasilkan ID3 mudah dimengerti dan pembentukan pohon keputusan cepat dan pendek.

Pada penelitian ini dibuat suatu aplikasi sistem pengambilan keputusan untuk menentukan jurusan pada siswa SMA, dalam hal ini SMA yang akan menjadi objek penelitian adalah SMAN 6 Semarang yang terletak di Jl.Ronggolawe no.4 Semarang. Berdasarkan pemaparan singkat tentang beberapa hal diatas yang menjadi landasan untuk dilakukannya penelitian ini maka penulis akan membuat sebuah aplikasi yang diharapkan dapat menampilkan hasil dari data mining dengan algoritma ID3 yang berupa pohon keputusan untuk membantu dalam menentukan penjurusan di SMAN 6 Semarang.

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. Penentuan Jurusan Sekolah Menengah Atas Dengan Fuzzy C-Means.
2. Perbandingan Algoritma ID3 Dan C5.0 Dalam Identifikasi Penjurusan SMA.
3. Sistem Pendukung Keputusan Untuk Menentukan Jurusan Pada Siswa Dengan Menggunakan KNN dan SMART.
4. Rekomendasi Penjurusan di SMU YSKI dengan Algoritma K-Means.

2.2. Tinjauan Pustaka

A. Data Mining

Data mining merupakan metode untuk menemukan suatu pengetahuan dalam suatu database. Data mining adalah proses menggali dan menganalisa sejumlah data yang sangat besar untuk memperoleh sesuatu yang benar, baru, sangat bermanfaat dan akhirnya dapat menemukan suatu corak atau pola dalam data tersebut [9]. Data mining merupakan gabungan dari teori dan heuristic, berfokus pada seluruh proses penemuan knowledge dan pola dari data.

Data mining sebagai salah satu cabang ilmu yang relatif baru mempunyai potensi pengembangan yang sangat besar dan diprediksi akan menjadi salah satu yang paling

revolusioner pada dekade ini. Teknik data mining merupakan sebuah proses ekstraksi informasi untuk menggali pengetahuan (knowledge discovery) dan menemukan pola (pattern recognition) [15].

Peran utama dari data mining terbagi menjadi enam kategori, yaitu Deskripsi, Estimasi, Prediksi, Klasifikasi, Klustering, dan Asosiasi yang dimana setiap kategori memiliki algoritma masing-masing.

B. ID3

Algoritma ID3 atau Iterative Dichotomiser 3 (ID3) merupakan sebuah metode yang digunakan untuk membangkitkan pohon keputusan. Algoritma pada metode ini menggunakan konsep dari entropy informasi. Pemilihan atribut dengan menggunakan Information Gain. Pemilihan atribut pada ID3 dilakukan dengan properti statistik, yang disebut dengan information gain. Gain mengukur seberapa baik suatu atribut memisahkan training example ke dalam kelas target. Atribut dengan informasi tertinggi akan dipilih. Dengan tujuan untuk mendefinisikan gain, pertama-tama digunakanlah ide dari teori informasi yang disebut entropy. Entropy mengukur jumlah dari informasi yang ada pada atribut dengan rumus :

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

S adalah ruang (data) sample yang digunakan untuk training.

P_+ adalah jumlah yang bersolusi positif (mendukung) pada data sample untuk kriteria tertentu.

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 menyatakan suatu kelas. Semakin kecil nilai entropy maka semakin baik digunakan dalam mengekstraksi suatu kelas.

Pada algoritma ID3 pengurangan entropy disebut dengan informasi gain. Pembagian sample S terhadap atribut A dapat dihitung information gain dengan rumus [4]:

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

Dimana :

A : atribut

v : suatu nilai yang mungkin untuk atribut A

Value (A) : himpunan yang mungkin untuk atribut A

$|S_v|$: Jumlah sampel untuk nilai v

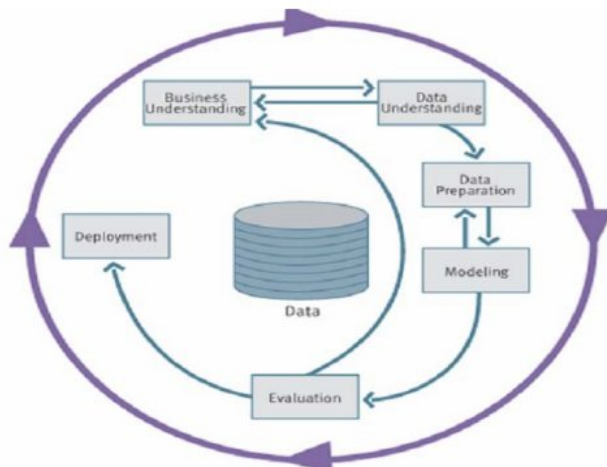
$|S|$: jumlah seluruh sampel data

Entropy(S_v) : entropy untuk sampel-sampel yang memiliki nilai v

Pilih atribut yang memiliki nilai information gain terbesar, 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.

C. CRISP-DM

Cross – Industry Standart Proses for Data Mining (CRIPS – DM) dikembangkan pada tahun 1996 oleh analis dari beberapa industri. CRIPS – DM menyediakan standart proses data mining sebagai pemecahan masalah secara umum dari bisnis atau unit penelitian. Enam fase CRISP-DM [10] :



1. Fase Pemahaman Bisnis (Business Understanding Phase)

- Penentuan tujuan proyek dan kebutuhan secara detail dalam lingkup bisnis atau unit penelitian secara keseluruhan.
- Menerjemahkan tujuan dan batasan menjadi formula dari permasalahan data mining.
- Menyiapkan strategi awal untuk mencapai tujuan.

2. Fase Pemahaman Data (Data Understanding Phase)

- Mengumpulkan data.
- Menggunakan analisis penyelidikan data untuk mengenali lebih lanjut data dan pencarian pengetahuan awal.

c. Mengevaluasi kualitas data.

d. Jika diinginkan, pilih sebagian kecil kelompok data yang mungkin mengandung pola dari permasalahan.

3. Fase Pengolahan Data (Data Preparation Phase)

- Siapkan dari data awal, kumpulan data yang akan digunakan untuk keseluruhan fase berikutnya. Fase ini merupakan pekerjaan berat yang perlu dilaksanakan secara intensif.
- Pilih kasus dan variabel yang ingin dianalisis dan yang sesuai analisis yang akan dilakukan.
- Lakukan perubahan pada beberapa variabel jika dibutuhkan.
- Siapkan data awal sehingga siap untuk perangkat pemodelan.

4. Fase Pemodelan (Modeling Phase)

- Pilih dan aplikasikan teknik pemodelan yang sesuai.
- Kalibrasi aturan model untuk mengoptimalkan hasil.
- Perlu diperhatikan bahwa beberapa teknik mungkin untuk digunakan pada permasalahan data mining yang sama.
- Jika diperlukan, proses dapat kembali ke fase pengolahan data untuk menjadikan data ke dalam bentuk yang sesuai dengan spesifikasi kebutuhan teknik data mining tertentu.

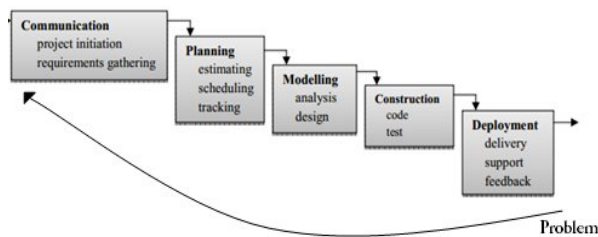
5. Fase Evaluasi (Evaluation Phase)

- Mengevaluasi satu atau lebih model yang digunakan dalam fase pemodelan untuk mendapatkan kualitas dan efektivitas sebelum disebarkan untuk digunakan.
- Menetapkan apakah terdapat model yang memenuhi tujuan pada fase awal.
- Menentukan apakah terdapat permasalahan penting dari bisnis atau penelitian yang tidak tertangani dengan baik.

- d. Mengambil keputusan berkaitan dengan penggunaan hasil dari data mining.
- 6. Fase Penyebaran (Deployment Phase)
 - a. Menggunakan model yang dihasilkan.
Terbentuknya model tidak menandakan telah terselesaikannya proyek.
 - b. Contoh sederhana penyebaran: Pembuatan laporan.
 - c. Contoh kompleks Penyebaran: Penerapan proses data mining secara paralel pada departemen lain.

D. Model Proses Waterfall

Menurut Pressman [16] model proses Waterfall, yang juga dikenal dengan classic life cycle, adalah model proses yang sistematis, pendekatan yang berurutan dalam pengembangan perangkat lunak yang diawali dengan spesifikasi kebutuhan – kebutuhan (requirement) client dan akan diawali dari kebutuhan pengguna (Requirement), berlanjut ke proses perencanaan (Planning), pemodelan (Modelling), konstruksi (Construction), dan penyebaran (Deployment) secara bertahap dan memuncak.



Gambar 2.1 Model Proses Waterfall

III. METODOLOGI PENELITIAN

Metode yang dipilih dalam pengembangan program yang dipilih adalah metode waterfall. Seperti yang telah dijelaskan pada landasan teori pada bab 2, metode ini bekerja dengan cara sistematis, yaitu pengerjaan dari suatu sistem dilakukan secara berurutan atau secara linear. Pemilihan metode ini dikarenakan mudah diterapkan dengan pembuatan aplikasi penjurusan ini. Kualitas dari sistem yang dihasilkan akan baik. Ini dikarenakan oleh pelaksanaannya secara bertahap. Sehingga tidak terfokus pada tahapan tertentu. Fase-fase dalam CRISP-DM yang digunakan peneliti dalam analisis data sangat cocok diterapkan dengan metode waterfall karena pengerjaan dilakukan secara urut dan bertahap.

Analisis Kebutuhan Pengguna

Tahapan pertama dalam pengembangan perangkat lunak menggunakan metode waterfall dimana kebutuhan perangkat lunak yang nantinya akan dibangun diidentifikasi terlebih dahulu.

Perencanaan

Berisi tentang pemahaman tentang data mining melalui CRISP-DM yang akan dijelaskan pada bagian selanjutnya dan planning untuk software yang akan dibuat.

Pemodelan

Tahap ini dilakukan sebelum melakukan coding. Tahap ini bertujuan untuk memberikan gambaran apa yang seharusnya dikerjakan dan bagaimana tampilannya. Tahap ini membantu dalam menspesifikasikan kebutuhan hardware dan sistem serta mendefinisikan arsitektur sistem secara keseluruhan.

Pembangunan

Dalam tahap ini dilakukan pembangunan perangkat lunak dengan menggunakan bahasa pemrograman Java dengan NetBeans sebagai IDE (Integrated Development Environment) dan akan dilakukan pengujian apakah perangkat lunak sudah berjalan seperti seharusnya.

Deployment

Tahap terakhir dimana perangkat lunak sudah siap digunakan dan dilakukan maintenance jika diperlukan.

IV. HASIL DAN PEMBAHASAN

4.1. Perancangan dan Analisis

A. Perancangan sistem

Berdasarkan analisis yang telah dilakukan maka dapat diketahui apa saja yang menjadi proses di dalam sistem, masukan sistem, keluaran sistem, sistem, serta antar muka sistem yang dibuat, sehingga sistem yang dibuat nantinya sesuai dengan apa yang diharapkan. Untuk lebih menjelaskan perancangan aplikasi yang dibangun, digunakan model diagram UML sebagai berikut :

4.1.1 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 aplikasi prediksi penjurusan siswa pada SMAN6 Semarang :

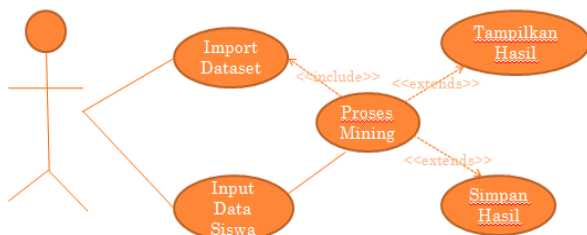
a. Identifikasi Aktor

Aktor adalah pelaku bisnis atau subjek yang menjalankan case pada software. Aktor yang terlibat dalam system ini secara khusus adalah guru BP dan Wakil Kepala Sekolah Kesiswaan SMAN6 Semarang. Namun baik pengguna secara

umum maupun siswa-siswi SMAN6 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.

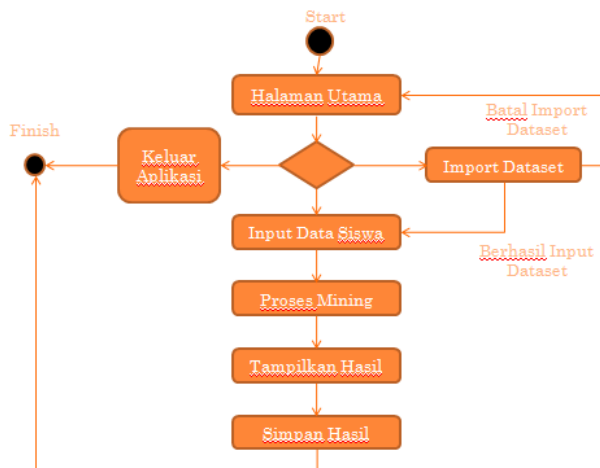
b. Diagram Model Use Case

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



Gambar 4.1 Use Case Diagram

4.1.2 Activity Diagram



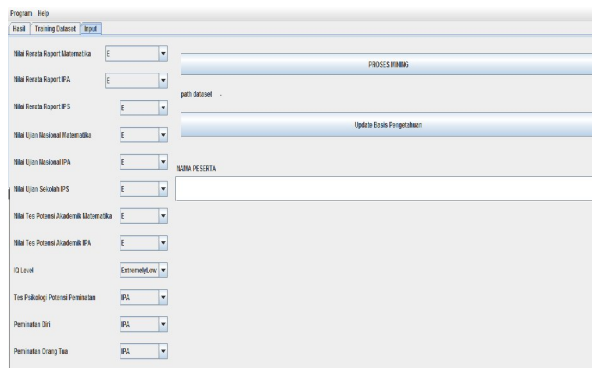
Gambar 4.2 Activity Diagram

Gambar di atas menjelaskan bagaimana pengguna melakukan proses mining. Ketika pertama kali pengguna menjalankan program, pengguna akan masuk ke dalam halaman utama. 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.

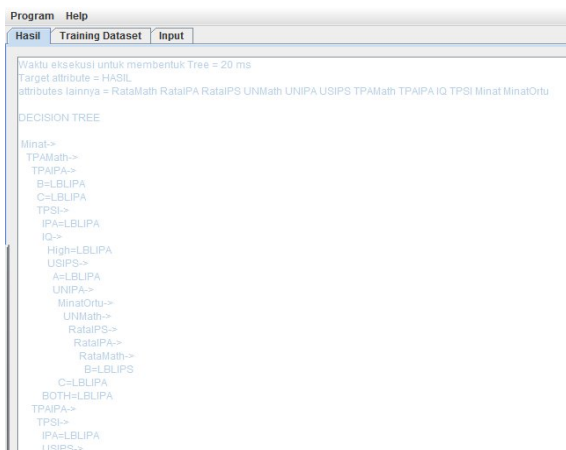
4.2 Halaman Inputan



Gambar 4.3 Halaman Input

Gambar diatas menjelaskan tentang proses pengisian form input prediksi. Pengguna mengisikan nama peserta yang akan diprediksi dengan mengisikan atribut atribut pada kolom yang tertera setelah itu memilih tombol proses mining untuk mengetahui hasil prediksi. Nilai yang diinputkan berupa nilai rata-rata matematika SMP pada masing-masing, nilai rata-rata IPA pada SMP masing-masing, nilai rata-rata IPS pada SMP masing-masing, nilai Ujian Nasional matematika, nilai Ujian Nasional IPA, nilai Ujian Sekolah IPS, nilas tes peminatan IPA, nilai tes peminatan mat, IQ, tes psikologi, peminatan anak, dan peminatan orang tua.

4.3 Halaman Hasil



Gambar 4.4 Halaman Hasil

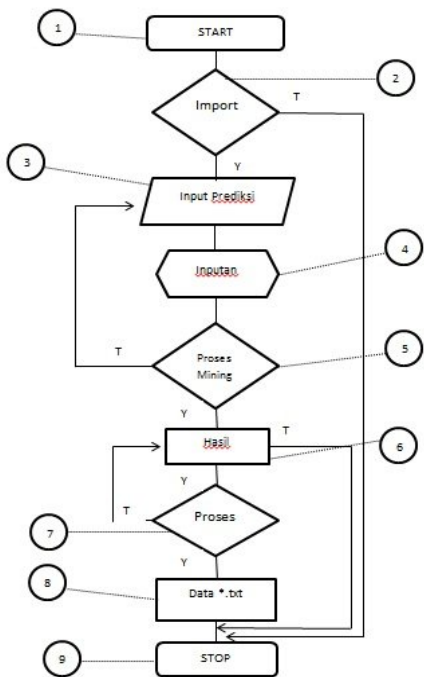
Gambar di atas menjelaskan hasil dari proses mining. Yaitu prediksi pohon tree yang terbentuk dari proses mining yang menentukan bahwa inputan yang dimasukan berdasar 12 atribut apakah target tersebut masuk IPA ataukah IPS berdasar dari pohon keputusan yang terbentuk dari dataset yang digunakan.

4.4 Pengujian dan implementasi

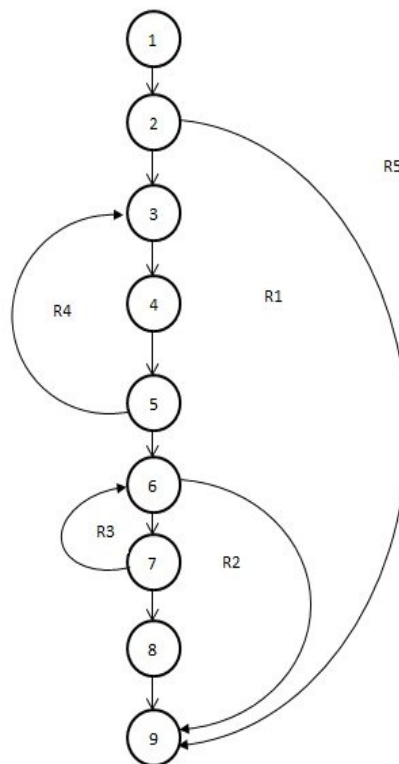
Sebelum perangkat lunak dirilis untuk pengguna hal yang perlu dilakukan adalah melakukan pengujian apakah program sudah memenuhi standar kelayakan pakai atau terdapat bug yang menyebabkan program tidak dapat berjalan dengan baik.

4.4.1 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 4.5 Bagan alir evaluasi



Gambar 4.6 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 resikonya Low (rendah).

4.4.2 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:

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

Gambar 4.7 Pengujian Black Box

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.

4.5 Pembahasan

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.

Dalam pengimplementasian algoritma ID3 dengan dataset siswa SMAN6 Semarang yang telah diolah, terlebih dahulu dilakukan proses perhitungan entropi dan information gain secara manual. Hasil ini dikarenakan agar aplikasi yang digunakan sudah mempunyai dan memenuhi algoritma ID3 yang dipakai dalam klasifikasi untuk penjurusan SMAN6 Semarang. Dari perhitungan tersebut diketahui bahwa atribut yang paling besar dalam menentukan jurusan adalah peminatan diri. Atribut ini sekaligus menjadi root (akar) dalam decision tree tersebut. Atribut ini memiliki nilai information gain yang paling besar yaitu 0.157314902. Setelah mengetahui entropi dari masing-masing nilai atribut yang ada, entropi tersebut digunakan untuk menghitung nilai information gain dari setiap atribut yang ada. Berikut table perhitungan :

	jml kasus	IPA	IPS	Entropi	Gain
total	341	271	70	0.7324	
ATAMAT					0.0049009
A	5	5	0	0	
B	336	266	70	0.7383	
C	0	0	0	0	
D	0	0	0	0	
E	0	0	0	0	
BATAIPA					0.0009737
A	1	1	0	0	
B	340	270	70	0.7335	
C	0	0	0	0	
D	0	0	0	0	
E	0	0	0	0	
BATAIPS					0.0019506
A	2	2	0	0	
B	339	269	70	0.7347	
C	0	0	0	0	
D	0	0	0	0	
E	0	0	0	0	
UNMAT					0.0324371
A	81	73	8	0.4651	
B	241	188	53	0.76	
C	18	10	8	0.9911	
D	1	0	1	0	
E	0	0	0	0	
UNIPA					0.0085377
A	15	12	3	0.7219	
B	285	231	54	0.7004	
C	40	27	13	0.9097	
D	1	1	0	0	
E	0	0	0	0	
USIPS					0.0013371
A	54	44	10	0.6913	
B	286	226	60	0.7411	
C	1	1	0	0	
D	0	0	0	0	
E	0	0	0	0	

Gambar 4.8 Perhitungan ID3

TPAMAT					0.0917806
A	0				
B	189	173	16	0.4184	
C	141	87	44	0.9541	
D	11	1	10	0.4395	
E	0	0	0	0	
TPAIPA					0.0490195
A	0	0	0	0	
B	80	74	6	0.3843	
C	225	178	47	0.7394	
D	36	19	17	0.9978	
E	0	0	0	0	
IQ					0.0054896
ExtremelyLow	0	0	0	0	
BorderLine	0	0	0	0	
Low	3	2	1	0.9183	
Average	216	167	49	0.7725	
High	104	87	17	0.6425	
Superior	17	14	3	0.6723	
Extreme	1	1	0	0	
TP					0.0159141
IPA	39	34	5	0.5525	
IPS	180	133	47	0.8284	
BOTH	122	104	18	0.6037	
PMIDIRI					0.1573149
IPA	307	260	47	0.6175	
IPS	34	1	33	0.1914	
PMORTU					0.156291
IPA	314	269	45	0.5928	
IPS	27	2	25	0.3809	

Gambar 4.9 Perhitungan ID3

Dengan hasil root yang telah diperoleh, yaitu atribut peminatan diri maka dilakukan pengujian algoritma ID3 pada aplikasi dengan cara mengambil sampel dari dataset sebanyak 20 record untuk diujikan dalam aplikasi yang telah dibuat. Label yang telah ada dihilangkan sehingga target yang akan diujikan class / labelnya tidak diketahui agar dapat mengetahui tingkat keakuratan dari aplikasi yang telah dibuat. Diambil 20 data uji dari 371 record yang ada dalam dataset peminatan siswa kelas X kurikulum 2013. Dari 20 data uji yang telah diujikan terdapat 4 data uji yang tidak cocok dengan hasil pada data yang ada pada Guru BP. 16 data uji lainnya cocok dengan hasil data yang ada pada Guru BP. Sehingga akurasi pada aplikasi penjurusan siswa SMAN6 Semarang adalah sebesar 80%. Berikut adalah tabel hasil dari data uji yang telah dibandingkan :

Hasil	RataMat	RataIPA	RataIPS	UNMat	UNIPA	USIPS	TPAMat	TPAIPA	IQ	Psikologi	Peminatan Diri	Peminatan Ortu	Hasil Mining	Ket
IPS	B	B	B	B	B	B	C	C	High	IPA	IPA	IPA	IPA	fail
IPA	B	B	B	C	C	B	D	D	Average	BOTH	IPA	IPA	IPS	fail
IPA	B	B	B	B	B	B	B	C	Low	IPS	IPA	IPA	IPA	akurat
IPS	B	B	B	B	C	B	C	C	Average	IPS	IPS	IPS	IPS	akurat
IPS	B	B	B	C	B	B	C	D	High	IPS	IPS	IPS	IPS	akurat
IPA	B	B	B	B	B	B	B	C	High	BOTH	IPA	IPA	IPA	akurat
IPA	B	B	B	B	B	B	C	C	Average	IPS	IPA	IPA	IPA	akurat
IPA	B	B	B	B	B	B	C	B	High	IPA	IPA	IPA	IPA	akurat
IPA	B	B	B	B	B	B	C	C	Average	IPS	IPA	IPA	IPA	akurat
IPS	B	B	B	B	C	B	B	C	Average	IPS	IPS	IPS	IPS	akurat
IPS	B	B	B	B	B	B	C	D	Average	IPS	IPS	IPS	IPS	akurat
IPA	B	B	B	B	B	B	C	C	Average	IPS	IPA	IPA	IPA	akurat
IPS	B	B	B	B	B	B	C	C	Average	IPS	IPA	IPA	IPS	akurat
IPS	B	B	B	B	B	B	C	C	Average	IPS	IPS	IPA	IPS	akurat
IPS	B	B	B	C	B	B	C	C	Average	IPS	IPS	IPS	IPS	akurat
IPA	B	B	B	B	B	B	C	C	Average	IPS	IPA	IPA	IPA	akurat
IPA	B	B	B	B	B	B	C	D	Average	IPS	IPA	IPA	IPS	fail
IPA	B	B	B	B	B	B	B	B	Low	IPS	IPA	IPA	IPA	akurat
IPA	B	B	B	A	B	B	C	B	Average	IPS	IPA	IPA	IPA	akurat
IPS	B	B	B	B	B	B	C	C	Average	IPS	IPA	IPA	IPA	fail

Gambar 4.10 Perbandingan Data Uji

V. PENUTUP

A. Kesimpulan

Aplikasi telah berhasil dirancang sesuai dengan kebutuhan - kebutuhan yang menjadi tujuan dari perancangan, yaitu aplikasi dapat bekerja sebagai media pembantu dalam penjurusan pada SMAN6 Semarang dan mengetahui tingkat akurasi dengan menggunakan data dari guru BP sebagai perbandingan. Dengan menggunakan 371 dataset dan 20 data uji yang diinputkan terdapat 4 kasus yang meleset dan 16 kasus berhasil, sehingga didapat akurasi sebesar 80%.

Melalui pengujian - pengujian yang sudah dilakukan setelah aplikasi dapat diimplementasikan pada Java, semua pengujian input dan output aplikasi telah sesuai dengan yang diharapkan. Aplikasi telah berhasil memproses semua input yang diberikan dan menghasilkan output yang sesuai dengan kebutuhan, sehingga aplikasi sudah layak untuk digunakan dan diterapkan sesuai dengan kebutuhan.

B. Saran

Saran yang dapat diberikan penulis dari penulisan Tugas Akhir ini adalah :

1. Diharapkan adanya penelitian dan kajian lebih lanjut mengenai pengembangan aplikasi prediksi yang menggunakan algoritma klasifikasi ID3 pada platform yang berbeda.

2. Untuk penelitian mengenai aplikasi penjurusan SMA sejenis diharapkan menggunakan algoritma klasifikasi lainnya sehingga mengetahui tingkat akurasi.
3. Diharapkan ada penelitian selanjutnya yang meneliti dan membahas kekurangan - kekurangan yang dimiliki oleh aplikasi penjurusan SMA ini agar dapat dikembangkan lebih lanjut lagi mengingat ketidaksempurnaan sistem kurikulum 2013 yang baru berjalan satu tahun.

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] Eko Sudaryanto, "Pengaruh Minat Belajar dan Penjurusan Terhadap Prestasi Belajar Siswa di SMK Katolik ST Lois Randublatung", Skripsi Fakultas Keguruan dan Ilmu Pendidikan, Universitas Muhammadiyah Surakarta, Surakarta, 2009.
- [2] Wahyudin. "Metode Iterative Dichotomizer 3 (ID3) Untuk Penerimaan Mahasiswa Baru", Program Studi Ilmu Komputer, Universitas Pendidikan Indonesia, 2009.
- [3] Bahar, "Penentuan Jurusan Sekolah Menengah Atas Dengan Algoritma Fuzzy C-Means", Tesis Magister Komputer, Universitas Dian Nuswantoro, Semarang, 2011.
- [4] Munawaroh Holisatul dkk, " Perbandingan Algoritma ID3 Dan C5.0 Dalam Identifikasi Penjurusan Siswa SMA", Jurnal Sarjana Teknik Informatika , Vol. 1, No. 1, pp. 1-12, Juni, 2013 .
- [5] Kustiyansih Yeni dkk, "Sistem Pendukung Keputusan Untuk Menentukan Jurusan Pada Siswa SMA Menggunakan Metode KNN Dan SMART", Fakultas Teknik Universitas Trunojoyo, Bangkalan, 2010.
- [6] Herawati Rosita, "Rekomendasi Penjurusan di SMU YSKI dengan Algoritma K-Means", Fakultas Ilmu Komputer Universitas Soegijapranata, Semarang, 2012.
- [7] Fathoni Muhammad dan Saniman, "Jurnal Pengantar Algoritma dan Pemrograman", *Jurnal Saintikom*, Vol.4, No.1, pp.120-133, Januari 2008.
- [8] Jain anil k., "Data clustering: 50 years beyond K-means". Elsevier, 2009.
- [9] Han, J., and Kamber, M., *Data Mining Concept and Technique*, Morgan Kaufman Publisher, San Francisco, 2006 .
- [10] larose Daniel T., "Data Mining Methods and Models", Wiley, 2006.
- [11] Kementerian Pendidikan dan Kebudayaan, "Modul Pelathian Implementasi Kurikulum 2013 untuk Guru BK / Konselor", 2013.
- [12] Departemen Pendidikan Nasional , "*Panduan Penilaian Penjurusan Kenaikan Kelas dan Pindah Sekolah*", Direktorat Pendidikan Menengah Umum, Jakarta, 2004.
- [13] Ian H. Witten, Frank Eibe, Mark A. Hall, "Data mining: Practical Machine Learning Tools and Techniques 3rd Edition", Elsevier, 2011.
- [14] Kementerian Pendidikan dan Kebudayaan, "Salinan - Permendikbud Nomor 81A Tahun 2013 Tentang Implementasi Kurikulum", 2013.

- [15] Kusri dan Emha Taufiq Luthfi, "Algoritma Data Mining", Andi, 2009.
- [16] Pressman Roger S., "Software Engineering : Practitioner's Approach 7th Edition". McGraw-Hill, Inc, New York, 2012.
- [17] <http://id.wikipedia.org/wiki/Java>, 3 Mei 2014, (Diakses pada 28 Juni 2014).