

PENERAPAN ALGORITMA C4.5 UNTUK PREDIKSI JURUSAN SISWA SMAN 3 REMBANG

Faid Ari Prastya¹

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

¹111201105841@mhs.dinus.ac.id

Abstract— In the implementation of 2013 curriculum process majors high school students performed directly at the beginning of high school entrance or on the class X. It is different from the previous curriculum implementation in the majors where high school students conducted in class XI. In this study the authors take the object of research at National High School 3 Rembang where located at Jalan Gajah Mada No. 08 Rembang. Data mining is the process of using statistics, mathematics, artificial intelligence, and machine learning to extract and identify the information that is useful and relevant knowledge from a variety of large databases. Approach to the application of data mining with Decision Tree C4.5 algorithm will be performed to determine the specialisation that will be taken by students. The resulting accuracy of the modeling algorithm C4.5 is 74.65%. The number of true positives (tp) are 107 records, false positives (fp) are 34 records. The number of true negative (tn) are 108 records and the number of false negative (fn) are 39 records. While the Average Math Scores are being the top node of formed Decision Tree.

Keywords— Data Mining, Decision Tree C4.5, Classification, Student Majority, High School

I. PENDAHULUAN

Salah satu perubahan yang dapat dirasakan dengan penerapan kurikulum 2013 adalah proses penjurusan siswa Sekolah Menengah Atas atau SMA. Dalam penerapan kurikulum 2013 proses penjurusan siswa Sekolah Menengah Atas dilakukan langsung pada awal masuk Sekolah Menengah Atas atau pada kelas X [1]. Hal tersebut berbeda dengan penerapan kurikulum sebelumnya dimana pada proses penjurusan siswa Sekolah Menengah Atas dilakukan pada kelas XI.

Pada proses penjurusan siswa dapat memilih jurusan yang ada pada Sekolah Menengah Atas masing – masing. Ada Sekolah Menengah Atas yang membuka 3 jurusan yaitu IPA, IPS, dan Bahasa. Dan ada juga Sekolah Menengah Atas yang hanya membuka 2 jurusan yaitu IPA dan IPS. Penentuan pembukaan jurusan dalam Sekolah Menengah Atas kembali dalam kebijakan sekolah masing – masing.

Sekolah memegang peranan penting dalam mengembangkan potensi siswa sesuai keahlian atau jurusan masing – masing. Dalam hal ini bagian konseling dan kurikulum sekolah berperan penting dalam menentukan siswa kedalam jurusan – jurusan yang sesuai. Konseling sebagai bagian integral dari sistem pendidikan di sekolah memegang dan berkaitan dengan pemenuhan fungsi dan tujuan pendidikan serta peningkatan mutu pendidikan di sekolah, terlebih lagi konseling sebagai mitra kerja dalam melaksanakan tugasnya sebagai rangkaian upaya pemberi bantuan [2]. Perubahan Kurikulum dimaksudkan agar memungkinkan penyesuaian program pendidikan pada satuan

pendidikan dengan kondisi dan kekhasan potensi yang ada didaerah peserta didik [3].

Dalam penelitian ini penulis mengambil objek penelitian pada Sekolah Menengah Atas Negeri 3 Rembang yang beralamat di Jalan Gajah Mada Nomor 08 Rembang. Sekolah Menengah Atas Negeri 3 Rembang membuka 2 jurusan untuk kelangsungan proses belajar siswanya, yaitu IPA dan IPS. 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 [4].

Sementara itu, data mining adalah proses yang menggunakan statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [5]. Data mining disisi lain adalah kegiatan meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan [6].

Obbie Kristanto, dalam penelitiannya Penerapan algoritma klasifikasi data mining ID3 untuk menentukan penjurusan siswa SMAN 6 Semarang. 371 Datashet, Akurasi 80%, dan memberikan saran menggunakan Algoritma Klasifikasi Lain Untuk Pembuktian Tingkat Akurasi [1]. Holisatul Munawaroh dkk, Perbandingan algoritma ID3 dan C5.0 dalam identifikasi penjurusan siswa SMA. Kedua algoritma memiliki tingkat akurasi yang baik seiring dengan bertambahnya data testing namun hasil algoritma C5.0 lebih akurat dibanding dengan

ID3. Algoritma C5.0 tidak bisa mengklasifikasikan data yang mengandung missing value [7]. Liliانا Swastina, Penerapan algoritma C4.5 untuk penentuan jurusan mahasiswa. Algoritma C4.5 sering digunakan, Memiliki tingkat akurasi yang tinggi. Algoritma C4.5 prediksi lebih akurat dari algoritma Naive Bayes. Tingkat keakuratan 93,31% dan Akurasi Jurusan 82,64% [6]. Dan secara umum algoritma C4.5 untuk membangun keputusan meliputi pilih atribut sebagai akar, buat cabang untuk masing – masing nilai, bagi kasus dalam cabang, ulangi proses untuk masing – masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Berdasarkan pertimbangan diatas, pendekatan data mining dengan penerapan algoritma Decision Tree C4.5 akan dilakukan untuk menentukan jurusan yang akan diambil oleh siswa sesuai dengan latar belakang, minat dan kemampuannya sendiri. Dan diharapkan peluang untuk sukses dalam belajar di Sekolah Menengah Atas semakin besar..

II. STUDI PUSTAKA

2.1. Penelitian Terkait

Penulis memulai penelitian ini dengan terlebih dahulu melakukan studi kepustakaan dari penelitian-penelitian dan sumber-sumber lain. Penelitian tersebut membahas tentang topik yang terkait dengan penelitian penulis, antara lain adalah penelitian mengenai algoritma yang akan digunakan penulis.

1. Obbie Kristanto, “Penerapan Algoritma Klasifikasi Data Mining ID3 Untuk Menentukan penjurusan Siswa SMAN 6 Semarang”.
2. Liliانا Swastina, “Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa”.
3. Holisatul Munawaroh dkk, “Perbandingan Algoritma ID3 dan C5.0 Dalam Identifikasi Penjurusan Siswa SMA”.

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 [10]. Rancangan bangun dari data mining yang khas memiliki beberapa komponen utama yaitu : - Database, data warehouse, atau tempat penyimpanan informasi lainnya. - Server database atau data warehouse. - Knowledge base - Data mining engine. - Pattern evolution module - Graphical user interface[5].

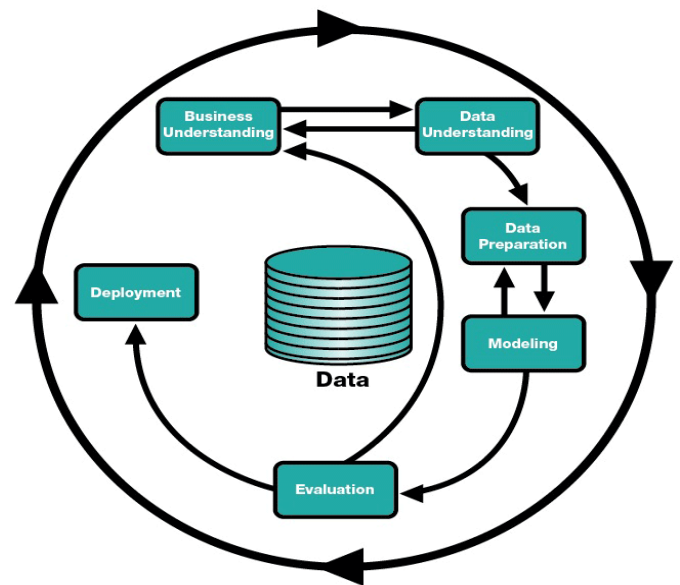
B. Tahap – tahap Data Mining

Disini akan diuraikan tahap-tahap umum dari data mining tapi perlu diingat sebelum seorang analis menerapkan tahapan-tahapan data mining tersebut, sebagai prasyarat penerapan data mining, diperlukan pemahaman terhadap data dan proses diperolehnya data tersebut. Yang lebih mendasar

lagi adalah diperlukannya pemahaman mengapa menerapkan data mining dan target yang ingin dicapai. Sehingga secara garis besar sudah ada hipotesa mengenai aksi-aksi yang dapat diterapkan dari hasilnya nanti. Pemahaman-pemahaman tersebut akan sangat membantu dalam mendesain proses data mining dan juga pemilihan teknik data mining yang akan diterapkan [11].

C. CRISP – DM

CRISP-DM (CRoss-Industry Standard Process for Data Mining) merupakan suatu konsorsium perusahaan yang didirikan oleh Komisi Eropa pada tahun 1996 dan telah ditetapkan sebagai proses standar dalam data mining yang dapat diaplikasikan di berbagai sektor industri. Berikut ini adalah gambar proses siklus hidup pengembangan dari CRISP-DM [13] :



D. Algoritma C4.5

Algoritma C4.5 merupakan salah satu algoritma machine learning. Dengan algoritma ini, mesin (komputer) akan diberikan sekelompok data untuk dipelajari yang disebut learning dataset. Kemudian hasil dari pembelajaran selanjutnya akan digunakan untuk mengolah data-data yang baru yang disebut test dataset. Karena algoritma C4.5 digunakan untuk melakukan klasifikasi, jadi hasil dari pengolahan test dataset berupa pengelompokan data ke dalam kelas-kelasnya [14].

Rumus menghitung entropy pada algoritma C4.5

$$Entropi(S) = \sum_{i=1}^k - p_i \log_2 p_i \dots\dots(1)$$

Keterangan :

S : adalah Himpunan (dataset) kasus

K : adalah banyaknya partisi S

Pi : adalah probabilitaas yang didapat dari Sum(Ya) atau Sum(Tidak) dibagi total kasus

Setelah mendapatkan entropi dari keseluruhan kasus, lakukan analisis pada setiap atribut dan nilai-nilainya dan hitung entropinya. Langkah berikutnya yaitu dengan menghitung Gain, rumus daripada Gain adalah sebagai berikut:

$$Gain(A) = Entropi(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times Entropi(S_i) \dots(2)$$

Keterangan :

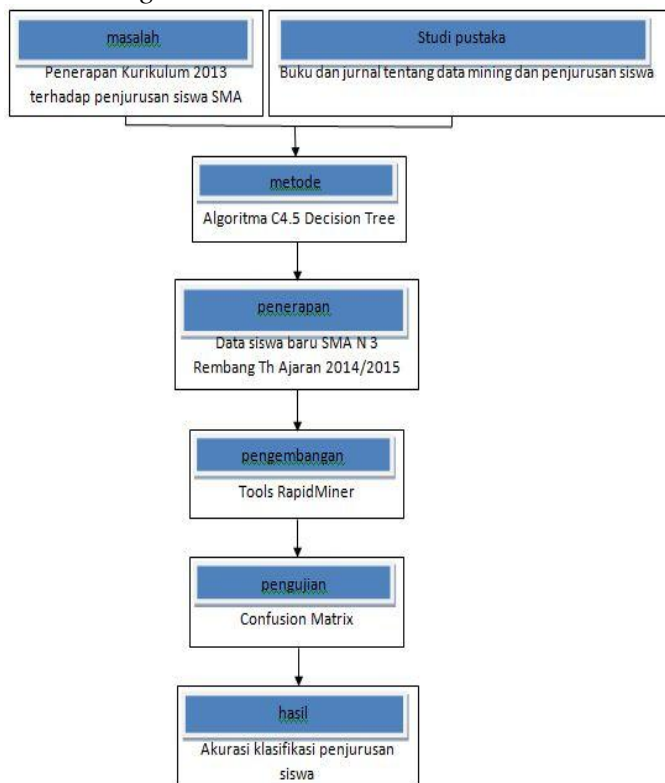
S : himpunan kasus

A : atribut

|S_i| : jumlah kasus pada partisi ke-i

|S| : jumlah kasus dalam S

2.3. Kerangka Pemikiran



III METODE PENELITIAN

3.1 Instrumen Penelitian

Berdasarkan permasalahan yang telah diuraikan pada bab sebelumnya, maka bahan dan peralatan yang diperlukan untuk penelitian ini meliputi :

A. Bahan

Dalam penelitian ini bahan yang dibutuhkan adalah Data yang diperoleh dari Sekolah Menengah Atas Negeri 3 Rembang. Data yang dikumpulkan yaitu data siswa baru tahun ajaran 2014/2015.

B. Peralatan

Peralatan dalam penelitian ini meliputi kebutuhan perangkat lunak dan kebutuhan perangkat keras.

Kebutuhan Perangkat Lunak

1. Microsoft Windows 7 sebagai OS notebook.
2. Microsoft Excel 2010 sebagai media penulisan datashet.
3. RapidMiner, framework yang akan digunakan untuk melihat hasil akurasi dari algoritma yang digunakan terhadap datashet yang sedang diteliti.
4. Matlab, framework yang akan digunakan untuk mengolah datashet dalam klasifikasi penjurusan siswa menggunakan metode data mining.

Kebutuhan perangkat keras

1. Perangkat keras notebook dengan spesifikasi :
 - a. Prosesor intel core i5 2.1 GHz
 - b. RAM 2GB
 - c. Hardisk 160GB

3.2 Pengumpulan Data

Data yang diperoleh dalam penelitian ini merupakan data primer yaitu data yang didapatkan secara langsung dari sumber data, dalam hal ini data diperoleh melalui softcopy yang tersedia dari Sekolah Menengah Atas Negeri 3 Rembang.

Selain itu dalam membantu penyusunan tugas akhir ini digunakan beberapa studi pustaka yang merupakan data sekunder yaitu:

1. Buku yang membahas data mining khususnya algoritma C4.5
2. E-book mengenai data mining dan algoritmanya.
3. Jurnal mengenai kasus klasifikasi penjurusan siswa

3.3 Desain Penelitian

Dalam penelitian ini penulis menggunakan model standarisasi data mining yaitu CRISP-DM (Cross Industry Standart Process for Data Mining), dengan langkah-langkah sebagai berikut:

Pemahaman Bisnis (*Bussiness Understanding*)

Penelitian ini mengedepankan penentuan tujuan proyek dan kebutuhan secara detail dalam lingkup bisnis atau unit penelitian secara keseluruhan, serta juga menerjemahkan tujuan dan batasan dari data yang diambil dari sekolah menengah atas menjadi formula dari permasalahan data mining mulai dari menyiapkan strategi awal hingga metode yang dibutuhkan untuk mencapai tujuan.

Pemahaman Data (*Data Understanding*)

Data yang digunakan dalam penelitian ini adalah sumber data primer. Data diperoleh dari Sekolah Menengah Atas Negeri 3 Rembang. Data yang dikumpulkan yaitu data siswa baru tahun ajaran 2014/2015.

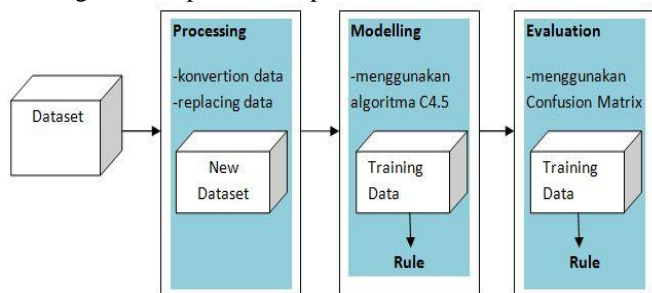
Pengolahan Data (*Data Preparation*)

Data kemudian dilakukan pemilihan atribut dan sebagian dari data dalam atribut yang ada akan dikonversikan untuk memudahkan proses data

mining, karena data akan diproses dengan tools bantu data mining.

Pemodelan (Modelling)

Metode yang akan digunakan dalam penelitian ini adalah Algoritma C4.5 untuk melakukan pengukuran akurasi dalam penelitian ini akan menggunakan tools RapidMiner. Berikut adalah gambaran pemodelan penelitian:



Validasi dan Evaluasi

Dalam tahapan ini akan dilakukan validasi serta pengukuran keakuratan hasil yang dicapai oleh model menggunakan beberapa teknik yang terdapat dalam *framework* RapidMiner yaitu Confusion Matrix untuk pengukuran tingkat akurasi model, dan *split validation* untuk validasi.

Penyebaran (Deployment)

Hasil dari penelitian ini berupa analisis yang mengarah ke *Decision Support System* (DSS) yang diharapkan dapat digunakan oleh guru atau pihak yang bersangkutan sebagai bahan pertimbangan dalam menentukan langkah penjurusan siswa dengan atribut yang telah dijabarkan, serta juga sebagai pengetahuan guru dan siswa tentang faktor – faktor yang dipertimbangkan dalam penjurusan siswa. Selain itu hasil dari analisa ini juga akan digunakan dasar perancangan sebuah sistem pengambil keputusan guna melakukan identifikasi dan klasifikasi penjurusan siswa.

IV HASIL PENELITIAN DAN PEMBAHASAN

4.1 Pengolahan Data

Data yang digunakan dalam penelitian ini adalah sumber data primer. Data diperoleh dari Sekolah Menengah Atas Negeri 3 Rembang. Data yang dikumpulkan yaitu data siswa baru tahun ajaran 2014/2015 dengan attribute Nama, Nomor Pendaftaran, Asal Sekolah, Tanggal Lahir, Jenis Kelamin, Rata Nilai Mat, Rata Nilai IPA, Nilai Ujian Mat, Nilai Ujian IPA, Piagam, Peminatan, Jurusan. Dan selanjutnya adalah pemilihan atribut data yang digunakan.

Tabel 4.1 Pemilihan Atribut

Atribut	Detail penggunaan	
Nama	√	ID
No. Pendaftaran	X	No
Asal Sekolah	X	No
Tanggal Lahir	X	No
Jenis Kelamin	X	No

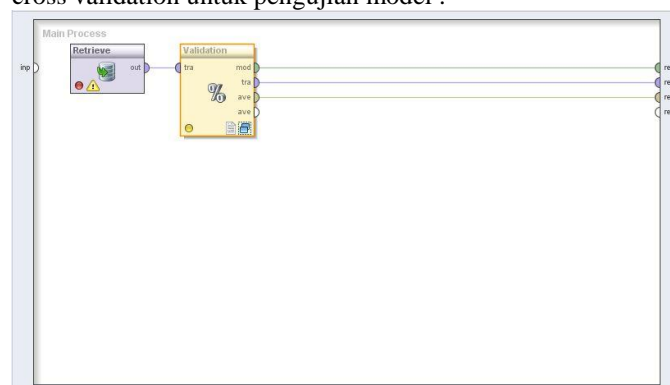
Rata Nilai Mat	√	Nilai Model
Rata Nilai IPA	√	Nilai Model
Nilai Ujian Mat	√	Nilai Model
Nilai Ujian IPA	√	Nilai Model
Piagam	√	Nilai Model
Peminatan Jurusan	√	Nilai Model
Jurusan	√	Label Target

4.2 Validasi dan Evaluasi

Pada tahap pengujian model ini data yang digunakan telah melewati tahap preprosesing. Selanjutnya data akan dilihat nilai akurasi, pohon keputusan, dan rule pada Rapid Miner. Berikut adalah gambar 4.1 tampilan proses import data :

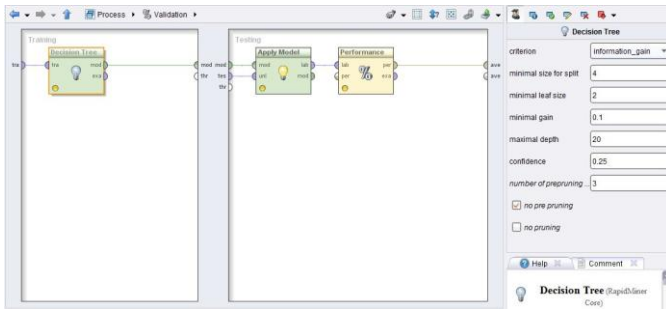
Gambar 4.1 Import Data

Dan selanjutnya adalah proses pengujian data menggunakan algoritma C4.5 untuk melihat akurasi, pohon keputusan dan rule. Berikut adalah gambar 4.2 pengujian algoritma C4.5 pada Rapid Miner, digunakan cross validation untuk pengujian model :



Gambar 4.2 Pemodelan Algoritma C4.5 oleh Cross Validation

Pada gambar 4.2 dataset penjurusan siswa baru SMAN 3 Rembang Tahun Ajaran 20 14 – 2015 dihubungkan dengan operator cross validation yang didalamnya terdapat proses seperti gambar 4.3 dibawah ini.



Gambar 4.3 Pengujian Algoritma C4.5 oleh Cross Validation

Algoritma C4.5 melakukan training terhadap data – data yang telah dibagi oleh cross validation menjadi dua kotak yaitu training dan testing. Training terdiri dari decision tree, sedangkan testing terdiri dari apply model dan performance.

Berdasarkan gambar 4.2 dan 4.3 dapat dijelaskan mengenai parameter dan operator yang digunakan pada model algoritma C4.5 sebagai berikut :

1. Retrieving Data adalah operator yang digunakan untuk mengimport dataset yang akan digunakan, pada penelitian ini data diimport dari file excel.
2. Validation operator yang digunakan dalam penelitian ini adalah split validation, validation ini hanya membagi total dari keseluruhan dataset menjadi data training dan data testing.
3. Decision tree adalah metode klasifikasi yang digunakan dalam penelitian ini.
4. Apply model adalah operator yang digunakan dalam penelitian ini yaitu algoritma C4.5.
5. Performance adalah operator yang digunakan untuk mengukur performa akurasi dari model.

Berdasarkan gambar 4.3 *Decision Tree* dapat dijelaskan mengenai parameter dan operator yang digunakan pada model algoritma C4.5 sebagai berikut :

1. Criterion, berguna memilih kriteria untuk menetapkan atribut sebagai akar dari decision tree. kriteria yang dapat dipilih, antara lain
 - a. Gain ratio merupakan varian dari information_gain. Metode ini menghasilkan information gain untuk setiap atribut yang memberikan nilai atribut yang seragam
 - b. Information gain, dengan metode ini, semua entropi dihitung. Kemudian atribut dengan entropi minimum yang dipilih untuk dilakukan perpecahan pohon (split). Metode ini memiliki bias dalam memilih atribut dengan sejumlah besar nilai.
 - c. Gini index merupakan ukuran ketidaksihan dari suatu ExampleSet. Metode ini memisahkan pada atribut yang dipilih memberikan penurunan indeks gini rata-rata yang dihasilkan subset.

d. Accuracy, metode ini memilih beberapa atribut untuk memecah pohon (split) yang memaksimalkan akurasi dari keseluruhan pohon.

2. Minimal size of split, Ukuran untuk membuat simpul-simpul pada decision tree. simpul dibagi berdasarkan ukuran yang lebih besar dari atau sama dengan parameter Minimal size of split. Ukuran simpul adalah jumlah contoh dalam subset nya.
3. Minimal leaf size, Pohon yang dihasilkan sedemikian rupa memiliki himpunan bagian simpul daun setidaknya sebanyak jumlah minimal leaf size.
4. Minimal gain merupakan nilai gain minimal yang ditentukan untuk menghasilkan simpul pohon keputusan. Gain dari sebuah node dihitung sebelum dilakukan pemecahan. Node dipecah jika gain bernilai lebih besar dari Minimal Gain yang ditentukan. Nilai minimal gain yang terlalu tinggi akan mengurangi perpecahan pohon dan menghasilkan pohon yang kecil. Sebuah nilai yang terlalu tinggi dapat mencegah pemecahan dan menghasilkan pohon dengan simpul tunggal.
5. Maximal depth, Parameter ini digunakan untuk membatasi ukuran Putusan Pohon. Proses generasi pohon tidak berlanjut ketika kedalaman pohon adalah sama dengan kedalaman maksimal. Jika nilainya diatur ke '-1', parameter kedalaman maksimal menempatkan tidak terikat pada kedalaman pohon, pohon kedalaman maksimum dihasilkan. Jika nilainya diatur ke '1' maka akan dihasilkan pohon dengan simpul tunggal.
6. Confidence, Parameter ini menentukan tingkat kepercayaan yang digunakan untuk pesimis kesalahan perhitungan pemangkasan.
7. Number of prepruning alternatives. Parameter ini menyesuaikan jumlah node alternatif mencoba untuk membelah ketika split dicegah dengan prepruning pada simpul tertentu.
 - a. No prepruning, Secara default Pohon Keputusan yang dihasilkan dengan prepruning. Menetapkan parameter ini untuk menonaktifkan benar prepruning dan memberikan pohon tanpa prepruning apapun.
 - b. No pruning Secara default Pohon Keputusan yang dihasilkan dengan pemangkasan. Menetapkan parameter ini untuk menonaktifkan benar pemangkasan dan memberikan sebuah unpruned [16].

4.3 Hasil Percobaan

Pengujian ini dilakukan untuk mengetahui kinerja dari metode algoritma C4.5 dalam melakukan klasifikasi terhadap kelas yang telah ditentukan dalam uji coba ini. Hasil akurasi confusion matrix pada percobaan ini pada gambar 4.3 sebagai berikut :

	true IPA	true IPS	class precision
pred. IPA	107	38	73.29%
pred. IPS	34	108	76.06%
class recall	75.88%	73.47%	

Gambar 4.4 Confusion Matrix

Jumlah true positif (tp) sebanyak 107 record, false positif (fp) sebanyak 34 record, jumlah true negative (tn) sebanyak 108 record dan jumlah false negative (fn) sebanyak 39 record. Akurasi yang dihasilkan dari dataset Siswa Baru SMAN 3 Rembang Tahun Ajaran 2014/2015 dengan pengujian metode algoritma C4.5 pada Rapid Miner sebesar 74,65%.

4.4 Implementasi

1. Membuka GUI dengan selanjutnya membuka *File - New - Gui*, atau mengetik guide di Command Window.
2. Merancang Desain Antar Muka.

PREDIKSI JURUSAN SISWA SMAN 3 REMBANG

Tahun Ajaran 2014/2015

Rata Nilai Mat	<input type="text"/>
Rata Nilai IPA	<input type="text"/>
Nilai Ujian Mat	<input type="text"/>
Nilai Ujian IPA	<input type="text"/>
Piagam	<input type="text"/>
Peminatan	<input type="text"/>

HASIL PREDIKSI KETERANGAN LEAF NODE

Gambar 4.5 Desain Antar Muka

3. Karena variable *Edit Text* dan *Pushbutton* akan digunakan untuk perhitungan di kode program, mengatur nama variable yang sesuai. Variabel pada Matlab diberi istilah *tag* yang tertera pada jendela inspector. Mengganti *tag* pada *edit text* menjadi RATANILAI MAT, RATANILAIIPA, NILAIUJIANMAT, NILAIUJIANIPA, PENJURUSAN, PEMINATAN, AKSIJURUSAN.

4. Menjalankan Program.

Program akan dicoba diinputkan data dari :

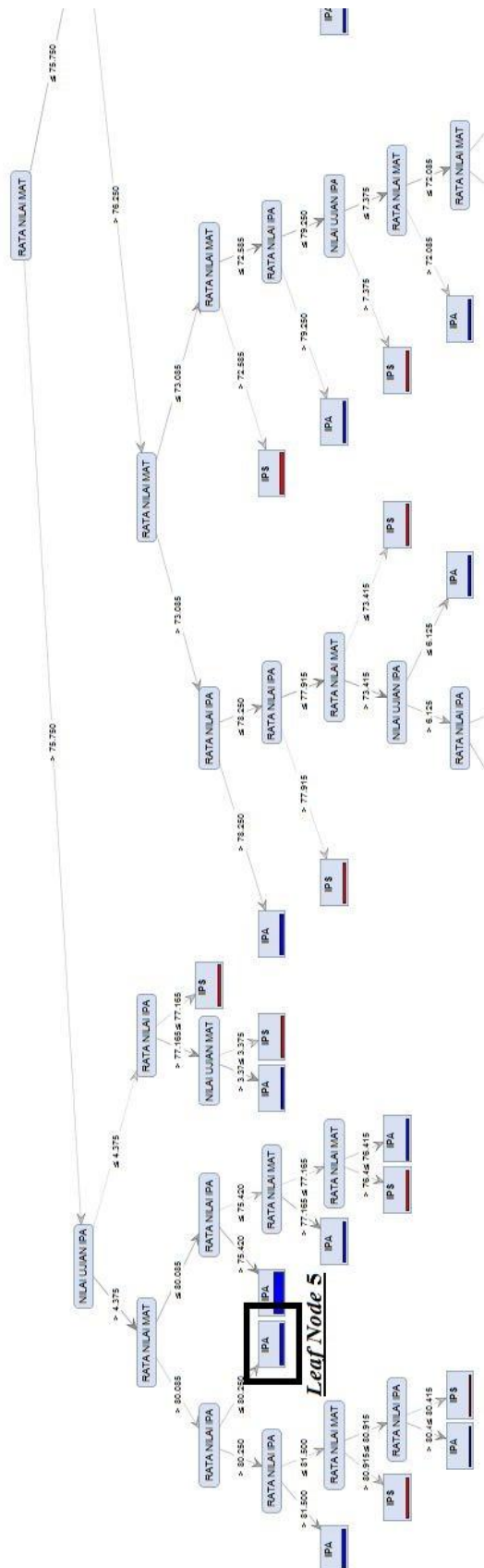
Nama : Nisya Defi Iriyanti
Rata Nilai Mat : 83.17
Rata Nilai IPA : 79.67
Nilai Ujian Mat: 7.5
Nilai Ujian IPA: 7.5

Piagam : 0
Peminatan : 1
Hasil Jurusan : "IPA"



Gambar 4.6 Keluaran Program

Hasil prediksi jurusan dari Nisya Defi Iriyanti adalah IPA dan keterangan leaf node adalah LEAF NODE KIRI KE 5. Berikut gambar 4.17 keterangan arah akar ke kiri *Leaf Node 5* :



Inputan nilai dari Nisyra Defi Iriyanti dengan Rata Nilai Mat : 83.17, Rata Nilai IPA : 79.67, Nilai Ujian Mat: 7.5, Nilai Ujian IPA: 7.5, Piagam : 0, Peminatan : 1, Hasil Jurusan : "IPA".

Dibuktikan dengan kondisi pohon keputusan yang telah terbentuk, yaitu jika RATA NILAI MAT > 75.750 dan NILAI UJIAN IPA > 4.375 dan RATA NILAI MAT > 80.085 dan RATA NILAI IPA <= 80.250 dan JURUSAN = "IPA". Dan terbukti pada arah akar ke kiri Leaf Node 5.

V KESIMPULAN DAN SARAN

a. Kesimpulan

Dari permasalahan di atas dapat disimpulkan bahwa klasifikasi data siswa baru Sekolah Menengah Atas Negeri 3 Rembang Tahun Ajaran 2014 / 2015 dapat diselesaikan menggunakan teknik data mining, yaitu dengan menggunakan algoritma C4.5. Akurasi yang dihasilkan dari pemodelan algoritma C4.5 sebesar 74,65%. Dengan jumlah true positif (tp) sebanyak 107 record, false positif (fp) sebanyak 34 record. Jumlah true negative (tn) sebanyak 108 record dan jumlah false negative (fn) sebanyak 39 record. Sedangkan yang menjadi simpul paling atas dalam pohon keputusan yang terbentuk adalah Rata Nilai Mat. Jadi klasifikasi data siswa baru Sekolah Menengah Atas Negeri 3 Rembang Tahun Ajaran 2014 / 2015 dengan algoritma C4.5 bisa menjadi pendukung keputusan yang digunakan oleh pihak Sekolah Menengah Atas Negeri 3 Rembang dalam proses penentuan jurusan siswa.

b. Saran

Agar penelitian ini terus berkembang, berikut saran – saran yang diusulkan :

1. Penelitian ini dapat dikembangkan dengan metode data mining klasifikasi lainnya untuk melakukan perbandingan.
2. Penelitian ini diharapkan dapat digunakan oleh pihak sekolah menengah atas sebagai pedoman dalam penjurusan siswa.
3. Implementasi dari penelitian ini dapat dikembangkan dengan menambah tampilan decision tree pada program.

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.

Gambar 4.7 Arah Akar Ke Kiri Leaf Node 5

REFERENSI

- [1] Kristanto Obbie, "Penerapan Algoritma Klasifikasi Data Mining ID3 Untuk Menentukan penjurusan Siswa SMAN 6 Semarang". Semarang, Indonesia, 2014.
- [2] Kementerian Pendidikan dan Kebudayaan, "Modul Pelatihan Implementasi Kurikulum 2013 untuk Guru BK atau Konselor", 2013.
- [3] Kementerian Pendidikan dan Kebudayaan, "Salinan – Permendikbud Nomor 81A Tahun 2013 Tentang Implementasi Kurikulum", 2013.
- [4] Departemen Pendidikan Nasional, "Panduan Penilaian Penjurusan Kenaikan Kelas dan Pindah Sekolah". Direktorat Pendidikan Menengah Umum, Jakarta, 2004.
- [5] Turban Efraim, Aronson Jay E, and Liang, "Decision Support Systems and Intelligent Systems", 7th ed.: Prentice Hall, Upper Saddle River, NJ, 2005.
- [6] Swastina Liliana, "Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa". Banjarmasin, Indonesia, 2013.
- [7] Munawaroh Holisatul dkk, "Perbandingan Algoritma ID3 dan C5.0 Dalam Identifikasi Penjurusan Siswa SMA". Bangkalan, Indonesia, 2013.
- [8] Bahar, "Penentuan Jurusan Sekolah Menengah Atas Dengan Algoritma Fuzzy C Means". Semarang, Indonesia, 2011.
- [9] Winanta Sendy dkk, "Implementasi Metode Bayesian Dalam Penjurusan Di SMA Bruderaan Purworejo". Purworejo, Indonesia, 2013.
- [10] Han, J., and Kamber, M., "Data Mining Concept and Technique", Morgan Kaufman Publisher, San Francisco, 2006.
- [11] Fayyad, Usama, "Advances in Knowledge Discovery and Data Mining". MIT Press, 1996.
- [12] Kusrini dan Luthfi, E. T., "Algoritma Data Mining". Andi Offset, Yogyakarta, 2009.
- [13] Larose Daniel T, Discovering knowledge in data: An Introduction to Data Mining.: Wiley Interscience, 2005.
- [14] Ilmu Biner, "Belajar Mudah Algoritma Data Mining : C4.5 ". https://www.academia.edu/6065920/Belajar_Mudah_Algoritma_Data_Mining_C4.5, (Diakses pada 2 Desember 2014).
- [15] F. Gorunescu, "Data Mining Concept, Models and Techniques". Verlag Berlin Heidelberg: Springer, 2011.
- [16] Aprilla C. Dennis dkk, "Belajar Data Mining dengan Rapid Miner". Jakarta, Indonesia, 2013.
- [17] Pujiriyanto Andry, "Cepat Mahir Matlab". Kuliah Berseri IlmuKomputer.Com, 2004.