

# Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Studi di SMK Karya Guna 1 Bekasi

Novi Wulandari

Juusan Sistem Informasi STMIK Pranata Indonesia, Email: noph21@yahoo.com

Muhamad Saeful Malik

Juusan Sistem Informasi STMIK Pranata Indonesia, Email: saefulmalik17@gmail.com

## Abstract

SMK Karya Guna 1 Bekasi annually accepts new students. In many cases, it is found that the selection of majors is not in accordance with their abilities, interests, personality and talents. This results in a reluctance to learn and a decrease in quality and academic achievement because students feel wrong in choosing a major. A classification method was tested with the application of a decision tree model using the C4.5 Algorithm. The criteria that will be assessed for this analysis are NUN (National Examination Score), Mathematics, Physics Scores, English Values, Computer Values and Student Interests. After doing the testing, the results showed that in predicting the level of accuracy using the decision tree algorithm C4.5, the highest accuracy level was 81.47%. While the kappa value of 0.771 can be classified into the value of the level of reliability including satisfactory.

**Kata Kunci:** Analysis, Determination of Study Majors, C4.5 Algorithm, Rapid Miner

## Abstrak

SMK Karya Guna 1 Bekasi setiap tahunnya melakukan penerimaan siswa baru. Banyak kasus dijumpai bahwa pemilihan jurusan yang tidak sesuai dengan kemampuan, minat, kepribadian dan bakatnya. Hal ini mengakibatkan keengganan belajar dan menurunnya kualitas serta prestasi akademik karena siswa merasa salah dalam memilih jurusan. Dilakukan pengujian sebuah metode klasifikasi dengan penerapan model pohon keputusan yaitu dengan menggunakan Algoritma C4.5. Kriteria yang menjadi penilaian untuk analisis ini adalah NUN (Nilai Ujian Nasional), Matematika, Nilai Fisika, Nilai Bahasa Inggris, Nilai Komputer dan Minat Siswa. Setelah dilakukan pengujian diperoleh hasil bahwa dalam melakukan prediksi tingkat ketepatan dengan menggunakan decision tree algoritma C4.5 didapatkan tingkat akurasi tertinggi sebesar 81,47 %. Sedangkan nilai kappa 0,771 dapat digolongkan kedalam nilai tingkat reliabilitas termasuk memuaskan.

**Kata Kunci:** Analisis, Penentuan Jurusan studi, Algoritma C4.5, Rapid Miner

## I. PENDAHULUAN

Pada SMK Karya Guna 1 Bekasi setiap tahunnya melakukan penerimaan siswa baru. Terdapat beberapa jurusan di antaranya: Teknik Instalasi Listrik, Teknik Audio Visual, Teknik Kendaraan Ringan dan Multi Media. Penentuan jurusan dilakukan oleh siswa tanpa campur tangan dari pihak sekolah. Keputusan para siswa terkadang dipengaruhi oleh pendapat orang tua, teman atau figur-figur yang diidolakan. Dengan hanya mendasarkan pendapat tersebut dan tanpa menelaah kemampuannya, seorang siswa bisa membuat keputusan yang sangat bertolak belakang dengan kemampuan, minat, bakat dan kepribadiannya. Hal ini menyebabkan timbulnya keengganan belajar dan menurunnya kualitas serta prestasi akademik karena siswa merasa salah dalam

memilih jurusan. Diperlukan suatu analisis yang dapat membantu pihak sekolah untuk penentuan jurusan dalam bidang studi serta dapat membantu para siswa yang akan masuk sekolah pada SMK Karya Guna 1 Bekasi dalam memilih bidang studi yang sesuai dengan kemampuan, minat, bakat dan kepribadiannya.

Kriteria yang menjadi penilaian untuk analisis ini adalah NUN (Nilai Ujian Nasional), Matematika, Nilai Fisika, Nilai Bahasa Inggris, Nilai Komputer dan Minat Siswa. Banyak metode yang dapat digunakan untuk analisis ini yang memiliki beberapa kriteria yang dijadikan pertimbangan untuk pengambilan keputusan. Algoritma klasifikasi data mining dapat diterapkan untuk menentukan jurusan

dalam bidang studi yang akan diambil oleh siswa (Rahayu, 2013) (Kamagi, dkk, 2013).

Algoritma C4.5 (*decision tree*) metode tidak memerlukan proses pengelolaan terlebih dahulu dan dapat menyelesaikan dengan sederhana kasus-kasus yang memiliki dimensi yang besar. Metode ini juga mempunyai akurasi sangat baik asalakan datanya yang akan dijadikan patokan merupakan data yang akurat. Metode ini banyak diterapkan dalam berbagai bidang antara lain: kesehatan dan pengobatan, pendidikan, finansial, produksi, astronomi, hingga biologi molekuler.

## II.METODE PENELITIAN

### Data Mining

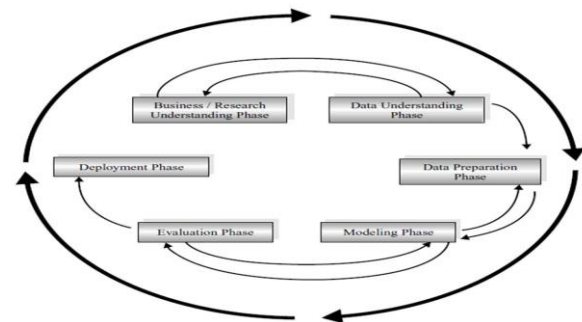
*Data mining* adalah analisa terhadap data (biasanya data yang berukuran besar) untuk menemukan hubungan yang jelas serta menyimpulkan yang belum diketahui sebelumnya dengan cara terkini dan berguna bagi pemilik data tersebut (Widodo, dkk. 2013). *Data mining* sering dimanfaatkan untuk mendeteksi kejadian-kejadian ganjil seperti penyakit tertentu, transaksi mencurigakan, mendeteksi telepon yang dimaksudkan untuk penipuan, penyalahgunaan kartu kredit, bahkan pasar swalayan saat ini sudah menggunakan teknik data mining untuk mengatur peletakan stand penjualan agar mempermudah pembeli dalam pergerakannya didalamnya.

Aplikasi yang menggunakan *data mining* bermaksud menyelesaikan permasalahan dengan membangun model berdasarkan data yang sudah digali untuk diterapkan terhadap data yang lain (Widodo, dkk, 2013).

Secara umum ada dua jenis topologi aplikasi *data mining*. Metode prediksi, yang bermaksud memprediksi nilai yang akan datang berdasarkan data-data yang telah ada variabelnya seperti klasifikasi, regresi, detikasi anomali dan lain lain. Yang kedua adalah metode deskriptif, yang bermaksud membantu user agar mudah melihat pola-pola yang berasal dari data yang ada.

Sebagai satu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap. Tahap-tahap tersebut bersifat interaktif dimana pemakai terlibat langsung atau dengan perantara *knowledge base* (Larose,

2011). Tahapan skema data mining ditunjukkan pada gambar 1.



**Gambar 1.** Tahapan Data Mining

#### *Business Understanding*

*Business Understanding* atau pemahaman domain (penelitian), pada fase ini dibutuhkan pemahaman tentang substansi dari kegiatan data mining yang akan dilakukan, kebutuhan dari perspektif bisnis. Kegiatannya menentukan sasaran atau tujuan bisnis, memahami situasi bisnis, menentukan tujuan data mining dan membuat perencanaan strategi serta jadwal penelitian.

#### *Data Understanding*

*Data Understanding* atau pemahaman data adalah fase mengumpulkan data awal, mempelajari data untuk bisa mengenalkan data yang akan dipakai. Fase ini mencoba mengidentifikasi masalah yang berkaitan dengan kualitas data, mendeteksi subset yang menarik dari data untuk membuat hipotesa awal.

#### *Data Preparation*

*Data Preparation* atau persiapan data, fase ini sering disebut sebagai fase yang padat karya. Aktivitas yang dilakukan antara lain memilih table dan field yang akan di transformasikan kedalam database baru untuk bahan *data mining* (set data mentah).

#### *Modelling*

fase ini menentukan teknik *data mining* yang digunakan, menentukan *tools data mining*, teknik *data mining*, algoritma *data mining*, menentukan parameter dengan nilai yang optimal.

#### *Evaluation*

*Evaluation* adalah fase interpretasi terhadap hasil data mining yang ditujukan dalam proses pemodelan pada fase sebelumnya. Evaluasi dilakukan secara

mendalam dengan tujuan menyesuaikan model yang didapat agar sesuai dengan sasaran yang ingin dicapai.

### Algoritma C4.5

Algoritma C4.5 merupakan struktur pohon dimana terdapat simpul yang mendiskripsikan atribut-atribut, setiap cabang menggambarkan atribut yang diuji, dan setiap daun menggambarkan kelas (Widodo, dkk, 2013).

Pohon keputusan (*decision tree*) merupakan salah satu teknik terkenal dalam *data mining* dan merupakan salah satu metode yang populer dalam menentukan keputusan suatu kasus” (Widodo, dkk, 2013). Hal ini karena metode ini tidak memerlukan proses pengelolaan pengetahuan terlebih dahulu dan dapat menyelesaikan dengan sederhana kasus-kasus yang memiliki dimensi yang besar. Akurasi sangat baik asalkan data yang akan dijadikan patokan merupakan data yang akurat. Metode ini banyak diterapkan dalam berbagai bidang antara lain: kesehatan dan pengobatan, finansial, produksi, astronomi, hingga biologi molekuler (Aprilla 2013). Tugas yang paling umum yang diserahkan kepada pohon keputusan adalah klasifikasi. Dari set *database* kita bisa mengetahui apakah suatu nasabah yang baik atau tidak riwayatnya, seseorang berpeluang terkena suatu penyakit tertentu berdasarkan riwayat dan lain-lain.

Pohon keputusan merupakan teknik yang paling efisien. Teknik regresi sangat banyak, tetapi paling terkenal adalah algoritma yang diperkenalkan oleh Prof. Brieman dengan istilah *The Classification and Regression Tree* (CART).

Masalah pertama pada pembuatan pohon keputusan adalah variabel manakah yang menjadi akar dari pohon tersebut. Akar disini adalah pemisah pertama dari pohon keputusan. Dikenal istilah *Bayesian Score* yang menilai suatu variabel, atau dalam pohon keputusan terkenal dengan sebutan *entropy*.

Algoritma pohon keputusan yang terkenal adalah C4.5. pada akhir tahun 1970 sampai awal 1980 J. Rose Quinlan, seorang peneliti di bidang *machine learning*, membuat sebuah algoritma *decision tree* yang dikenalkan dengan ID3 (*Iterative dichotomiser*). Quinian kemudian membuat algoritma C4.5 (sering disebut dengan pohon keputusan) yang merupakan

pengembangan dari algoritma ID3. Algoritma ini memiliki kelebihan, yaitu mudah dimengerti, fleksibel, dan menarik karena dapat divisualisasikan dalam bentuk gambar (pohon keputusan) (Gorunescu, 2011). Algoritma C4.5 secara rekursif mengunjungi setiap simpul keputusan, memilih pembagian yang optimal, sampai tidak bisa dibagi lagi. Algoritma C4.5 menggunakan konsep *information gain* atau *entropy reduction* untuk memilih pembagian yang optimal. Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5.

Mempersiapkan *data training*.

*Data training* diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu.

Menentukan akar dari pohon.

Akar diambil dari atribut yang terpilih, dengan cara menghitung nilai *gain* dari masing-masing atribut, nilai *gain* yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut, hitung dahulu nilai *entropy* dengan menggunakan persamaan 1:

$$Entropy(S) = \sum_{i=1}^n - p_i \cdot \log_2 p_i \quad (1)$$

S = himpunan kasus

n = jumlah partisi S

$p_i$  = Proporsi  $S_i$  terhadap S

Kemudian *gain* dihitung dengan menggunakan persamaan 2:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

S = himpunan kasus

A = fitur

n = Jumlah partisi atribut A

$|S_i|$  = proporsi  $S_i$  terhadap kasus S

$|S|$  = jumlah kasus dalam S

Ulangi langkah perhitungan *entropy* hingga semua *record* terpartisi.

Proses partisi pohon keputusan akan berhenti saat semua record dalam simpul N mendapat kelas yang sama, tidak ada atribut di dalam *record* yang terpartisi lagi, tidak ada *record* di dalam cabang yang kosong.

### Confusion Matrix

Metode evaluasi yang akan digunakan adalah pengukuran *confusion matrix* untuk mendapatkan nilai AUC.

Metode ini menggunakan tabel matrix, apabila dataset hanya terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lain negatif. Evaluasi dengan *confusion matrix* menghasilkan nilai *accuracy*, *precision*, dan *recall* (Kuswandi, dkk, 2011). *Accuracy* dalam klasifikasi adalah presentase ketepatan *record* data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi. Sedangkan *precision* atau *confidence* adalah proporsi kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya. *Recall* atau *sensitivity* adalah proporsi kasus positif yang sebenarnya yang diprediksi positif secara benar.

*True positives* adalah jumlah *record* positif yang diklasifikasikan sebagai positif, *false positive* adalah jumlah *record* negatif yang diklasifikasikan sebagai positif, *false negative* adalah *record* positif yang diklasifikasikan sebagai negatif, *true negative* adalah *record* negatif yang diklasifikasikan sebagai negatif.

**Tabel 1.** Konsep Tabel *Confusion Matrix*

| Correct<br>classification | Classified As          |                        |
|---------------------------|------------------------|------------------------|
|                           | +                      | -                      |
| +                         | <i>True Positives</i>  | <i>False Negatives</i> |
| -                         | <i>False Positives</i> | <i>True Negatives</i>  |

Kurva ROC menunjukkan akurasi dan perbandingan klasifikasi secara visual. ROC mengekspresikan *confusion matrix*. ROC adalah grafik dua dimensi dengan *false positives* sebagai garis horisontal dan *true positive* sebagai garis vertikal.

Hasil perhitungan divisualisasikan dengan kurva ROC (*Receiver Operating Characteristic*) (Gorunescu, 2011). ROC memiliki tingkat nilai diagnosa yaitu:

1. Akurasi bernilai 0,90 – 1,00 = *excellent classification*
2. Akurasi bernilai 0,80 – 0,90 = *good classification*
3. Akurasi bernilai 0,70 – 0,80 = *fair classification*
4. Akurasi bernilai 0,60 – 0,70 = *poor classification*

5. Akurasi bernilai 0,50 – 0,60 = *failure*

### Persiapan Data

Pada penelitian ini akan dilakukan di SMK Karya Guna 1 Bekasi yang akan difokuskan untuk mencari analisa tentang penentuan jurusan studi, sesuai dengan kriteria yang dibutuhkan oleh SMK Karya Guna 1 Bekasi. Hasil akhir penilaiannya adalah Multi Media, Teknik Instalasi Listrik, Teknik Audio Visual, Teknik Kendaraan Ringan. Dan pembuatan Analisis menggunakan algoritma C4.5. Agar penyeleksian dapat diketahui dengan cepat dan akurat, data yang akan digunakan dalam penelitian berupa data siswa masuk tahun periode Juni 2015 – Juli 2016.

Data yang didapat dari SMK Karya Guna 1 Bekasi adalah data siswa masuk pada tahun 2015-2016, jumlah data sebanyak 254 data. Variabel yang menjadi dasar penentuan penjurusan, dan variabel tujuan yaitu variabel yang dijadikan sebagai penjurusan. Variabelnya yaitu nilai tes Bahasa Inggris, Komputer, Matematika, Fisika, Minat siswa, dan NEM. Sedangkan variabel tujuannya adalah hasil jurusan diantaranya MM (Multi Media), TAV (Teknik Audio dan Visual), TTIL (Teknik Instalasi Listrik), TKR (Teknik Kendaraan Ringan). Contoh data siswa yang akan diolah dapat dilihat pada tabel 1.

**Tabel 2.** Data Siswa Kelas A, B, C, D SMK KG 1 Bekasi

| NO  | NAMA                       | L/P | ASAL SEKOLAH          | KOMP | ING | MTK | FISIKA | MINAT | NEM |
|-----|----------------------------|-----|-----------------------|------|-----|-----|--------|-------|-----|
| 1   | SASYA NINDA SYAHIRA        | P   | SMPN 6 TAMBUN SELATAN | 84   | 56  | 82  | 74     | MM    | 24  |
| 2   | JUANTO LATIEF              | L   | SMPN 18 BEKASI        | 46   | 52  | 82  | 92     | TTIL  | 23  |
| 3   | AHMAD DZULHAM NUR AMDANI   | L   | SMPN 1 BABELAN        | 46   | 48  | 82  | 68     | TTIL  | 19  |
| 4   | AHMAD FAUZI                | L   | SMPN 6 TAMBUN SELATAN | 74   | 60  | 82  | 56     | TAV   | 18  |
| 5   | FAKHRI SUPRIYATNA          | L   | MTS. AL-KHAIRIYAH     | 78   | 55  | 94  | 14     | TAV   | 16  |
| 6   | MUHAMMAD SYAFUDIN          | L   | SMPN 3 TAMBUN UTARA   | 44   | 56  | 86  | 64     | TKR   | 21  |
| 7   | JEHEM HENSEN               | L   | SMPN 6 TAMBUN SELATAN | 56   | 54  | 82  | 74     | TKR   | 15  |
| 8   | ALDI RAMADHANI IKSAN       | L   | SMPN 6                | 52   | 56  | 82  | 80     | TKR   | 17  |
| 9   | DEVINA NADITYA PUTRI       | P   | SMPN 3 TAMBUN UTARA   | 72   | 52  | 82  | 94     | MM    | 16  |
| 10  | RAMADHAN PUTRA             | L   | SMPN 6                | 48   | 56  | 94  | 78     | MM    | 33  |
| 11  | FERRIYANTO SETIAWAN        | L   | SMPN 6                | 86   | 44  | 86  | 10     | MM    | 30  |
| 12  | TRISNA HARYADI             | L   | SMPN 6 TAMBUN SELATAN | 80   | 44  | 82  | 52     | MM    | 19  |
| 13  | NAFIS HARDI PUTRA          | L   | SMPN 4 JATIMULYA      | 44   | 42  | 90  | 82     | TKR   | 18  |
| 14  | AINA FIRDHA ARINI          | P   | SMPN 3 TAMBUN UTARA   | 82   | 56  | 90  | 72     | MM    | 31  |
| 15  | RAFIFAH BUNGA PRIANA       | P   | SMPN 3 TAMBUN UTARA   | 68   | 44  | 82  | 70     | TAV   | 17  |
| 16  | MUHAMMAD IQBAL DWI SAPUTRO | L   | SMPN 8 TAMBUN SELATAN | 88   | 60  | 84  | 48     | TKR   | 25  |
| 17  | REYVALDI DEAR NAVORA       | L   | SMPN 3 JATIWANGI      | 64   | 58  | 84  | 60     | TKR   | 28  |
| 253 | ZIKI DAMA PERDANA          | L   | MTSN 1 BEKASI         | 66   | 78  | 74  | 76     | TKR   | 32  |
| 254 | IMAM AL FARIZI             | L   | SMPN 18 BEKASI        | 62   | 90  | 78  | 20     | TAV   | 26  |

data akan di transformasikan agar mempermudah dalam pembacaan proses *data mining*. data yang digunakan sejumlah 254 data, yang kemudian dibagi menjadi dua bagian yaitu sebagai data *training* dan data *testing*. Persentase untuk pembagian data yang digunakan yaitu 70 : 30, yang menghasilkan 178 data sebagai data training dan 76 data untuk data testing.

perlu dilakukan tahap persiapan (*preprocessing*), dengan tujuan untuk mencari *field* data yang kosong (*missing value*) dan data dengan format yang tidak sesuai (*noise*), agar akurasi dari hasil pengolahan data dapat meningkat

## Pemodelan

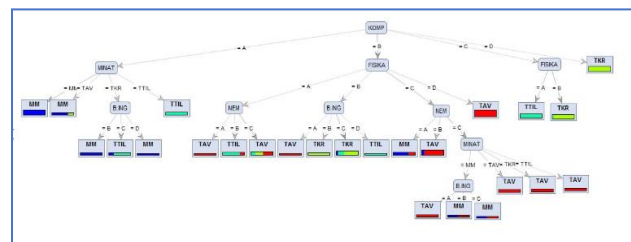
Teknik *data mining* yang akan digunakan adalah klasifikasi dan algoritma yang akan digunakan adalah C4.5. Langkah berikutnya mengimplementasikan data training yang telah dipersiapkan kedalam algoritma C4.5 untuk pembuatan pohon keputusan.

Akar diambil dari atribut yang terpilih pada proses data preparation, dengan cara menghitung nilai *gain* dari masing-masing atribut, nilai *gain* yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut, terlebih dahulu hitung nilai *entropy*. Hasil Perhitungan entropy untuk *node* 1 dapat dilihat pada tabel 2:

**Tabel 3.** Data Perhitungan *Node* 1

| NO | Nama       | Atribut | Jumlah | MM | TAV | TKR | TTL | Entropy | Gain  |
|----|------------|---------|--------|----|-----|-----|-----|---------|-------|
|    | TOTAL      |         | 178    | 35 | 69  | 42  | 32  | 1,925   |       |
| 1  | Komputer   | A       | 34     | 24 | 0   | 1   | 9   | 1,011   | 0,745 |
|    |            | B       | 104    | 11 | 69  | 13  | 11  | 1,451   |       |
|    |            | C       | 25     | 0  | 0   | 13  | 12  | 0,997   |       |
|    |            | D       | 15     | 0  | 0   | 15  | 0   | 0       |       |
| 2  | B. Inggris | A       | 33     | 1  | 26  | 5   | 1   | 0,991   | 0,182 |
|    |            | B       | 49     | 7  | 23  | 14  | 5   | 1,765   |       |
|    |            | C       | 78     | 20 | 17  | 22  | 19  | 1,991   |       |
|    |            | D       | 18     | 7  | 3   | 1   | 7   | 2,011   |       |
| 3  | MTK        | A       | 34     | 1  | 0   | 26  | 7   | 0,913   | 0,224 |
|    |            | B       | 90     | 19 | 37  | 16  | 18  | 2,301   |       |
|    |            | C       | 36     | 6  | 23  | 0   | 7   | 1,301   |       |
|    |            | D       | 18     | 9  | 9   | 0   | 0   | 1,301   |       |
| 4  | FISIKA     | A       | 35     | 0  | 7   | 4   | 24  | 1,194   | 0,065 |
|    |            | B       | 58     | 10 | 2   | 38  | 8   | 1,497   |       |
|    |            | C       | 57     | 20 | 37  | 0   | 0   | 0,933   |       |
|    |            | D       | 28     | 5  | 23  | 0   | 0   | 0,675   |       |
| 5  | MINAT      | MM      | 59     | 24 | 20  | 11  | 4   | 1,769   | 0,275 |
|    |            | TAV     | 32     | 4  | 22  | 6   | 0   | 1,073   |       |
|    |            | TKR     | 60     | 7  | 19  | 19  | 15  | 1,911   |       |
|    |            | TTL     | 27     | 0  | 8   | 6   | 13  | 1,508   |       |
| 6  | NEM        | A       | 28     | 15 | 5   | 6   | 2   | 1,672   | 0,188 |
|    |            | B       | 65     | 16 | 21  | 9   | 19  | 1,935   |       |
|    |            | C       | 85     | 4  | 43  | 27  | 11  | 1,610   |       |
|    |            |         |        |    |     |     |     |         |       |

Atribut Tanggung Jawab mempunyai nilai *gain* terbesar dan akan dijadikan *node* (akar) pertama pada pohon keputusan kelayakan kualitas kerja karyawan. Pehitungan *entropy* untuk mendapatkan *gain* dilakukan selama beberapa kali sehingga didapatkan suatu pohon keputusan yang terlihat seperti pada gambar 2.



**Gambar 2.** Pohon Keputusan Siswa SMK KG 1 Bekasi

## III. HASIL DAN PEMBAHASAN

### Evaluasi

Tingkat akurasi model yang telah dibuat sebesar 81,47 %. Perhitungan *confusion matrix* menunjukkan hasil prediksi MM, TTIL, TKR dan TAV yang dihasilkan dari model yang dibuat. Untuk prediksi MM yang true MM berjumlah sebanyak 26 data, true TTIL sebanyak 4 data, true TKR sebanyak 1 data, true TAV sebanyak 4 data, yang dimaksud disini adalah 26 data diprediksikan hasilnya MM oleh model yang memang benar hasilnya MM. Sedangkan 9 data di prediksikan hasilnya MM yang ternyata hasilnya adalah TTIL, TKR dan TAV. kesimpulannya untuk prediksi MM, akurasi model mencapai 74,29 %, Prediksi TTIL mencapai 85,19 %, Prediksi TKR 80,00%, sedangkan prediksi TAV mencapai 84,51%. Sedangkan nilai *kappa* dari proses validasi yang dihasilkan adalah 0,771.

### Program Aplikasi

Model klasifikasi yang telah didapatkan dapat di implementasikan kedalam sebuah program yang berupa sistem penunjang keputusan penentuan jurusan pada SMK Karya Guna 1 Bekasi. Tampilan aplikasi penjurusan dari hasil implementasi algoritma C4.5 dapat dilihat pada gambar 3.



**Gambar 3.** Aplikasi Analisis Penjurusan SMK KG 1 Bekasi

#### IV. KESIMPULAN

1. Berdasarkan hasil penilaian untuk nilai akurasi klasifikasi algoritma C4.5 senilai 81.47 % dan nilai  $kappa$  sebesar 0.771 dengan kategori nilai ISTIMEWA. Pohon keputusan tersebut akan menghasilkan *rules* sebagai dasar klasifikasi penentuan jurusan studi di SMK karya Guna 1 Bekasi.
2. *Rules* yang didapatkan dari pohon keputusan, diimplementasikan ke dalam sebuah sistem pendukung keputusan, yang dapat membantu untuk mempermudah pengambilan keputusan terhadap penentuan jurusan studi bagi calon siswa di SMK Karya Guna 1 Bekasi.

#### V.REFERENSI

- Aprilla, Dennis, dkk. (2013) *Belajar Data Mining dengan RapidMiner*. 8-11,-27 Jakarta.
- Gorunescu, F.(2011). *Data Mining Concept, Model, and Techniques*
- Kamagi.D. H, dkk. (2014). *Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Kelulusan Mahasiswa*, Vol : VI. 15-20, .Universitas Multimedia: Tangerang.
- Kuswandi, dkk (2011). *Delta Delapan Langkah dan Tujuh Alat Statistik untuk Peningkatan Mutu Berbasis Komputer*, Jakarta, PT.Alexputindo.
- Larose, D. (2011). *Discovering Knowledge in Data. An Introduction to Data Mining*
- Rahayu.E.R(2013). *Algoritma C4.5 Untuk Penjurusan Siswa Sma Negeri 3 Pati*. Surakarta: Universitas Dian Nuswantoro
- Widodo Prabowo Pudjo, dkk. (2013). *Penerapan Data Mining dengan Matlab*. Bandung: Rekayasa Sains