

IMPLEMENTASI ALGORITMA C.45 UNTUK MENENTUKAN JURUSAN CALON SISWA SEKOLAH MENENGAH KEJURUAN (SMK)

IMPLEMENTATION OF ALGORITHM C.45 TO DETERMINE THE MAJOR OF PROSPECTIVE STUDENTS VOCATIONAL HIGH SCHOOL (VHS)

Yessy Yee Nur Ariyanti Sekar Puspita Dewi¹⁾, Arief Hermawan²⁾

^{1,2}Prodi Informatika, Fakultas Sains dan Teknologi, Universitas Teknologi Yogyakarta
Jl. Siliwangi Jl. Ring Road Utara, Jombor Lor, Sendangadi, Kec. Mlati, Kabupaten
Sleman, Daerah Istimewa Yogyakarta

E-mail: yynaspd@gmail.com, ariefdb@uty.ac.id

ABSTRAK

Setiap tahunnya, PPDB (Penerimaan Peserta Didik Baru) mengalami kendala yang sama dimana calon siswa kesulitan untuk memilih jurusan yang akan mereka pilih dibangku SMK. Banyak dari mereka yang salah mengambil jurusan karena disuruh orang tua atau ikut-ikutan tanpa mengetahui potensi diri mereka sehingga hasil yang diperoleh relatif tidak maksimal dan membuat mereka sulit mendapatkan peluang kerja sesuai skill mereka. Dari permasalahan tersebut, penelitian ini dibuat dengan tujuan agar calon siswa mengetahui jurusan apa yang relevan berdasarkan potensi dirinya. Tes Kepribadian RIASEC yang terdiri dari kepribadian *Realistik*, *Investigatif*, *Artistik*, *Social*, *Enterprising*, dan *Conventional* berperan dalam memprediksi, menguatkan, dan meyakinkan calon siswa dalam memilih jurusan. Validitas Tes RIASEC ini dikatakan dinyatakan valid dan dapat digunakan namun harus memperhatikan dan mencocokkan kondisi lingkungan Penelitian ini menggunakan perhitungan dengan Algoritma C.45. Hasil yang diperoleh yaitu klasifikasi 4 jurusan yaitu Tata Boga, Teknik Permesinan, Teknik Instalasi Listrik, dan Multimedia yang menghasilkan akurasi data uji sebesar 75%.

Kata kunci : C.45, Jurusan SMK, Klasifikasi, Tes Riasec.

ABSTRACT

Every year, PPDB (New Student Acceptance) experiences the same problem where prospective students find it difficult to choose the major they will choose at SMK. Many of them took the wrong major because their parents told them to or went along without knowing their potential, so the results they obtained were relatively suboptimal and made it difficult for them to get job opportunities according to their skills. From these problems, this research was made with the aim that prospective students know what majors are relevant based on their potential. The RIASEC Personality Test which consists of *Realistic*, *Investigative*, *Artistic*, *Social*, *Enterprising*, and *Conventional* personalities plays a role in predicting, strengthening, and convincing prospective students in choosing majors. The validity of the RIASEC test is stated to be valid and can be used but must pay attention to and match environmental conditions. This research uses calculations with Algorithm C.45. The results obtained were the classification of 4 majors namely *Catering*, *Mechanical Engineering*, *Electrical Installation Engineering*, and *Multimedia* which resulted in a test data accuracy of 75%.

Keywords: C.45, SMK Major, Classification, Riasec.

PENDAHULUAN

Berdasarkan kurikulum yang berlaku di Indonesia, calon siswa

SMA/SMK yang akan mendaftar sekolah akan ditanyai mengenai jurusan yang akan diambil. Saat ini penerimaan calon siswa dilaksanakan dengan sistem Pelaksanaan

Penerimaan Peserta Didik Baru (PPDB). Pada PPDB SMK calon peserta diwajibkan untuk memilih 3 jurusan dengan 2 pilihan SMK. Hasil seleksi dapat dilihat pada website PPDB dan diupdate setiap hari.

Pada saat ini pemilihan jurusan ke Sekolah Menengah Kejuruan betul-betul banyak peminatnya sampai melebihi kuota yang ditentukan. Karena SMK dirancang untuk menyiapkan peserta didik atau lulusan yang siap memasuki dunia kerja dan mampu mengoptimalkan sikap profesional di bidang pekerjaannya [1]. Pemilihan jurusan merupakan permasalahan yang banyak dialami oleh calon siswa karena banyak aspek yang perlu dipertimbangkan diantaranya dari segi nilai, potensi, dan minat calon siswa SMK sehingga banyak calon siswa yang menunda pendaftaran karena memikirkan jurusan yang akan diambil. Salah satu penyebab kebingungan calon siswa dalam memilih jurusan adalah belum mengenali dengan baik minat dan bakat dirinya. Ketika individu telah mengetahui dan mengenali minat dirinya, maka akan mempermudah dirinya untuk menentukan karier melalui pemilihan jurusan yang tepat [2].

Holland mengklasifikasikan minat dan lingkungan ke dalam 6 tipe, yaitu: *Realistic, Investigative, Artistic, Social, Enterprising dan Conventional*, yang dikenal dengan RIASEC [3]. Validitas Teori Holland ini dikatakan dinyatakan valid dan dapat digunakan namun harus memperhatikan dan mencocokkan kondisi lingkungan. [4]. Ke enam tipe tersebut memiliki karakteristik masing masing. Tipe *Realistik* merupakan tipe orang yang memiliki pola pikir yang *logic* dan praktis. Tipe ini cenderung menyukai pekerjaan lapangan. Tipe *Investigatif* merupakan tipe orang yang memiliki rasa ingin tau yang cukup tinggi, pengamat yang cerdik dan solutif. Tipe ini cenderung melakukan banyak eksperimen. Tipe *Artistik* merupakan tipe orang yang cenderung kreatif, inovatif, dan mampu berekspeksi mewujudkan ide ide menggunakan imajinasinya. Tipe *Sosial* merupakan tipe orang yang menyukai interaksi dengan

banyak orang, cenderung banyak bicara untuk saling bertukar pikiran dengan orang lain. Tipe *Enterprising* merupakan tipe orang yang cenderung berani mengambil resiko, mampu mempengaruhi banyak orang dan memiliki tingkat kepercayaan diri yang tinggi. Sedangkan tipe *Conventional* merupakan tipe orang yang cenderung sistematis dalam bekerja, melakukan hal dengan berhati hati agar perfeksionis. Konsep ini dapat diterapkan untuk menentukan jurusan yang sesuai dengan minat masing-masing calon siswa.

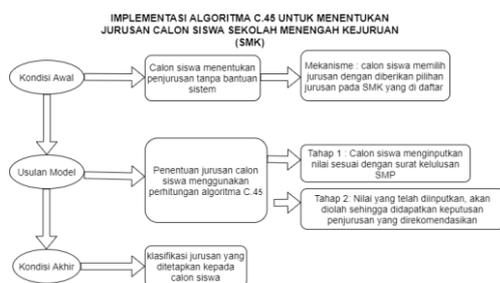
Penelitian sebelumnya yang berjudul “Penerapan Algoritma C.45 Untuk Prediksi Minat Penjurusan Siswa di SMKN N 1 Kademangan” [5] menjadi pandangan dan referesi dalam penelitian ini. Penelitian tersebut menggunakan 5 atribut sebagai input, kemudian diproses dengan algoritma C.45 dan menghasilkan 4 jurusan. Oleh karena itu, dari permasalahan tersebut maka dibutuhkan sistem pendukung keputusan yang dapat menentukan jurusan calon siswa agar lebih terarah dan efisien waktu. Sistem dibuat menggunakan algoritma C.45 yang berguna untuk mengeksplorasi data, menemukan relasi tersembunyi antara variabel input dengan sebuah variabel sasaran[6]. Penerapan algoritma C4.5 dalam melakukan prediksi sudah memiliki akurasi yang cukup baik dalam melakukan penyeleksian terhadap data dengan prosedur yang terstruktur dengan baik melalui beberapa tahapan diantaranya proses klasifikasi dengan data mining[7]. Penerapan algoritma dengan menerapkan teknik data mining metode decision tree algoritma C4.5 dapat memberikan informasi berupa klasifikasi[8]. Algoritma C4.5 terdiri dari beberapa variabel dan atribut yang dipergunakan untuk mendapatkan informasi atau pengetahuan tentang suatu aturan yang sangat mudah dipahami dengan penggambaran sebuah pohon keputusan [9][10]. Decision tree yang dihasilkan melalui perhitungan algoritma C4.5 sebagai penarik sebuah keputusan menghasilkan sebuah informasi yang dipergunakan untuk node dalam akar

pohon sehingga terbentuk sebuah pohon keputusan dengan menggunakan perhitungan atribut[11]. Algoritma C4.5 dengan decision tree memiliki struktur kerja mulai dari root menuju paling daun, jika diuji dengan data uji[12][13]. Proses pencarian nilai pada pohon keputusan yang dihasilkan melalui perhitungan dengan menggunakan nilai gain dan perhitungan entropy akurat [14]. Kelebihan dari algoritma C4.5 dapat menangani atribut continue dan diskrit beserta pengolahan training data dengan missing value untuk memperbaiki informasi pada gain menggunakan gain ratio [15].

Perbedaan penelitian ini dan penelitian sebelumnya terletak pada data yang diinput dan hasil klasifikasi. Pada sistem ini input data menggunakan nilai siswa dan hasil tes RIASEC untuk memperkuat hasil klasifikasi. Sistem dibuat untuk calon siswa SMK dengan 4 tujuan jurusan yaitu Tata Boga, Teknik Instalasi Listrik, Teknik Permesinan, dan Multimedia. Variable yang digunakan dalam proses prediksi yaitu Nilai Bahasa Indonesia, Nilai Bahasa Inggris, Nilai Matematika, Nilai PPKn, dan tes bakat minat RIASEC.

METODE

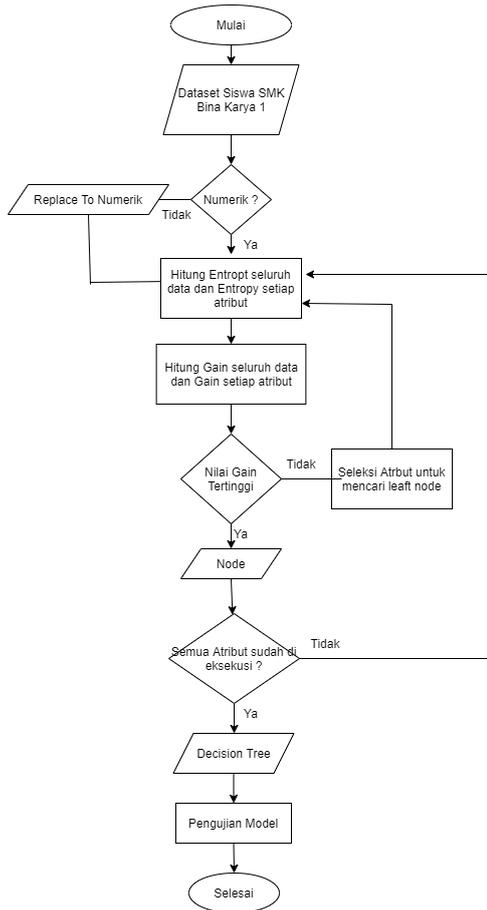
Penelitian ini untuk menganalisis minat penjurusan calon siswa SMK. Kerangka Penelitian dilakukan untuk memudahkan pengguna dalam mengidentifikasi masalah. Kerangka penelitian dalam Menentukan penjurusan SMK menggunakan C.45 sebagai berikut



Gambar 1. Tahapan Penelitian

Berdasarkan Gambar 1, Tahapan penelitian dilakukan sebagai berikut :

1) Kondisi Awal. Pada kondisi awal dalam menentukan penjurusan calon siswa yaitu calon siswa menentukan penjurusan tanpa bantuan sistem. Mereka hanya mengandalkan informasi yang didapat bahkan ada yang memilih jurusan dengan ikut teman mereka. Mekanisme yang masih berjalan saat ini adalah calon siswa memilih jurusan dengan diberikan banyaknya pilihan jurusan yang ada pada SMK yang terdaftar, hal tersebut memicu calon siswa memilih jurusan dengan asal asalan. **2) Usulan Model.** Usulan model yang diusulkan yaitu calon siswa akan dibantu sistem dalam menentukan penjurusan calon siswa. Metode sistem tersebut menggunakan Algoritma C.45. Tahap pertama usulan model adalah calon siswa menginputkan nilai yang sesuai dengan surat kelulusan SMP masing-masing. Tahap kedua yaitu nilai yang sudah diinputkan akan diproses dengan perhitungan algoritma C.45 sehingga didapatkan keputusan jurusan calon siswa. Flowchart perhitungan algoritma C.45 dapat dilihat pada gambar 2 **3) Kondisi Akhir.** Kondisi akhir yaitu berupa keputusan jurusan yang ditetapkan calon siswa. Hasil yang akan muncul adalah peminatan yang relevan dengan nilai yang diinput. Dengan hasil tersebut calon siswa akan lebih terarah menentukan penjurusan saat PPDB berlangsung sehingga lebih efisien.



Gambar 2 Flowchart Algoritma C.45

Berdasarkan Gambar 2, proses dalam melakukan klasifikasi jurusan dimulai dengan melakukan persiapan dataset, dimana data yang harus disiapkan diantaranya pemilihan atribut untuk diinputkan ke perhitungan nilai *entropy* dan *gain*.

Dalam proses pembuatan algoritma C.45, Ada beberapa tahapan dalam membuat sebuah pohon keputusan dalam algoritma C4.5[11] yaitu:

1. Mempersiapkan data training. Data training biasanya diambil dari data histori yang pernah terjadi sebelumnya atau disebut data masa lalu dan sudah dikelompokkan dalam kelas – kelas tertentu
2. Menghitung akar dari pohon. Akar akan diambil dari atribut yang akan 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*. Untuk menghitung nilai *entropy* digunakan rumus

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

Keterangan :

S : Himpunan kasus

n : Jumlah partisi *S*

p_i : Jumlah sampel pada kelas *i*

Setelah mengetahui nilai *entropy*, untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut atribut yang ada. Untuk menghitung gain digunakan rumus seperti tertera dalam persamaan berikut:

$$Gain(S, A) =$$

$$Entropy(S) - \sum_{i=1}^n \frac{S_i}{S} \times Entropy(S_i) \quad (2)$$

Keterangan :

Entropy(S) : Nilai *Entropy* seluruh himpunan Kasus

A : Atribut

n : Jumlah partisi dalam atribut *A*

S_i : Jumlah kasus pada partisi ke *i*

S : Jumlah Himpunan Kasus

3. Ulangi Langkah ke 2 hingga semua record terpartisi
4. Proses partisi pohon keputusan akan berhenti saat :
 - a) Semua record dalam simpul *N* mendapat kelas yang sama.
 - b) Tidak ada atribut didalam *record* yang dipartisi lagi.
 - c) Tidak ada record didalam cabang yang kosong

Setiap atribut memiliki variabel. Pada *node* akar, Hitung Kembali *entropy* dan *gain* setiap variabel, temukan *gain* tertinggi, hingga semua variable dieksekusi dan menemukan *node* daun. *Node* daun merupakan hasil klasifikasi.

HASIL DAN PEMBAHASAN

HASIL

Berdasarkan Gambar 2 Flowchart C.45, penelitian klasifikasi jurusan dimulai dari persiapan dataset. *Dataset* diambil dari hasil pengambilan data dan

observasi pada SMK Bina Karya 1 Karanganyar di Kabupaten Kebumen. Data sudah dimasukkan ke dalam file csv. Terdapat 6 atribut dalam file csv yang sudah di siapkan terlihat pada tabel 1

Tabel 1. Atribut

Simbol	Nama Atribut
BI	Nilai Bahasa Indonesia
MTK	Nilai Matematika
B.Ing	Nilai Bahasa Inggris
PKN	Nilai PPKn
TK	Tes kepribadian RIASEC
Jurusan	Klasifikasi jurusan

Berdasarkan tabel 1, 6 atribut telah ditetapkan berdasarkan pengambilan data dan observasi. Atribut atribut tersebut diantara nya Nilai Bahasa Indonesia yang disajikan dalam bentuk angka, Nilai Matematika yang disajikan dalam bentuk angka, Nilai Bahasa Inggris yang disajikan dalam bentuk angka, Nilai PPKn yang disajikan dalam bentuk angka, Tes Kepribadian yang disajikan dalam bentuk polinomial, dan Jurusan yang disajikan dalam polinomial.

Terdapat 39 record data. Pada atribut jurusan, terdapat 4 jurusan sebagai batasan yaitu Tata Boga (TBO), Teknik Instalasi Listrik(TIL), Teknik Permesinan (TP), dan Multimedia(MM) yang dapat dilihat pada tabel 2

Tabel 2. Penelitian

BI	MTK	BIn g	PK N	TK	Jurus an
81	86	82	81	Artistik	TBO
80	82	83	80	Artistik	TBO
89	87	83	85	Artistik	TBO
77	86	80	79	Investigati f	TBO
79	82	83	82	Artistik	TBO
87	87	85	83	Sosial	TBO
81	88	85	84	Artistik	TBO
82	88	83	83	Conventio nal	TBO
82	88	84	84	Artistik	TBO
81	82	81	82	Investigati f	TIL
82	84	89	84	Investigati f	TIL
77	82	84	83	Conventio nal	TIL
83	88	84	85	Conventio nal	TIL
82	86	82	84	Realistik	TIL

Berdasarkan tabel 2, hasil ringkasan dapat dilihat pada tabel 3

Tabel 3. Ringkasan Dataset

Jml Kasus	TBO	TIL	TP	MM
39	9	10	10	10

Perhitungan C.45 menggunakan *Tools excel* agar mengurangi resiko salah hitung. Proses perhitungan C.45 sebagai berikut:

1) Menghitung Nilai *Entropy*

Perhitungan entropy dibagi menjadi 2 yaitu *entropy* (S) dan *entropy* (Si). Hasil *Entropy* (S) dapat dilihat pada tabel 4

Tabel 4. Hasil *Entropy*(S)

Jml Kasus	TBO	TIL	TP	MM	Entropy
39	9	10	10	10	1,99855

Selanjutnya, menentukan *entropy* (Si) yang merupakan nilai *entropy* tiap varibel dari masing masing atribut. Setiap atribut akan diklasifikasi kan menjadi 2 varibel, Jika nilai lebih dari 83 maka variabel unggul, Jika kurang dari 83 kaka variabel baik. Hasil *entropy* (Si) dapat dilihat pada Tabel 5

Tabel 5. Hasil *Entropy* Si

Atribut	Jml Kasus	TB O	TI L	T P	M M	Entro py
BI						
Baik	22	9	9	3	1	1.6497
Unggul	17	0	1	7	9	1.2533
MTK						
Baik	12	2	4	1	5	1.7841
Unggul	27	7	6	9	5	1.9659
BIng						
Baik	22	6	7	3	6	1.9400
Unggul	17	3	7	7	8	1.9015
PPkn						
Baik	14	6	3	3	2	1.8773
Unggul	25	3	7	7	8	1.9215
TM						
R	9	0	2	7	0	0.7642
I	6	1	5	0	0	0.6500
A	12	6	0	1	5	1.3250
S	2	1	0	1	0	1
E	5	0	0	1	4	0.7219
C	5	1	3	0	1	1.3709

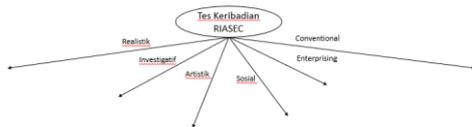
Setelah semua *entropy* tiap variable dalam atribut ditemukan, maka tentukan nilai *gain information* untuk menentukan

node akar. Hasil perhitungan *gain* dapat dilihat pada Tabel 6

Tabel 6. Hasil Gain Information

Atribut	Jml Kasus	T B O	T B I L	T P M y	MEntrop My	Gain Information
BI						0.5216
Baik	22	9	9	3	1	1.6497
Unggul	17	0	1	7	9	1.2533
MTK						0.0885
Baik	12	2	4	1	5	1.7841
Unggul	27	7	6	9	5	1.9659
BIng						0.0752
Baik	22	6	7	3	6	1.9400
Unggul	17	3	7	7	8	1.9015
PPkn						0.0928
Baik	14	6	3	3	2	1.8773
Unggul	25	3	7	7	8	1.9215
TM						0.9948
R	9	0	2	7	0	0.7642
I	6	1	5	0	0	0.6500
A	12	6	0	1	5	1.3250
S	2	1	0	1	0	1
E	5	0	0	1	4	0.7219
C	5	1	3	0	1	1.3709

Dari Tabel 6, Nilai *Gain* tertinggi adalah atribut Tes *riasec* yang memiliki nilai gain 0,99489. Maka node akar pada pohon keputusan penelitian ini adalah Tes Kepribadian. Pohon keputusan sementara dapat dilihat pada gambar 3.



Gambar 3. Hasil Pohon Keputusan Sementara

Setelah menentukan *node* akar, maka dari tiap variabel *node* akar tersebut, hitung kembali nilai *entropy* dan *gain* untuk variable selain tes kepribadian. Pertama, menentukan *entropy* dan *gain* dari variable *realistic*. Hasil perhitungan *entropy* dan *gain realistic* terlihat pada tabel 7

Tabel 7. Hasil Entropy dan Gain Realistik

Atribut	Jml Kasus	T B O	T B I L	T P M y	MEntrop My	Gain Information
Realistik						
Total	9	0	2	7	0	0.7642
BI						0.4581
Baik	3	0	2	1	0	0.9182
Unggul	6	0	0	6	0	0
MTK						0.0817

Atribut	Jml Kasus	T B O	T B I L	T P M y	MEntrop My	Gain Information
Baik	2	0	1	1	0	1
Unggul	7	0	1	6	0	0.5916
BIng						0.0247
Baik	3	0	1	2	0	0.9182
Unggul	6	0	1	5	0	0.6500
PPkn						0.0817
Baik	2	0	1	1	0	1
Unggul	7	0	1	6	0	0.5916

Berdasarkan Tabel 7, Nilai Bahasa Indonesia mendapatkan hasil tertinggi yaitu 0,45811, Oleh karena itu *node* anak selanjutnya setelah tes kepribadian adalah Nilai Bahasa Indonesia. Jika ditinjau dari detail nilai Bahasa Indonesia terdapat 2 label yang berpengaruh dengan nilai BI yaitu TIL dan TP. Dalam atribut BI variabel unggul di isi oleh jurusan TP, Oleh karena itu TP merupakan *node* daun, dan variable baik di isi oleh TIL karena hanya ada 2 jurusan yang terisi oleh tes kepribadian *riasec*

Hasil pohon keputusan sementara dapat dilihat pada gambar 4



Gambar 4. Hasil Pohon Keputusan Sementara

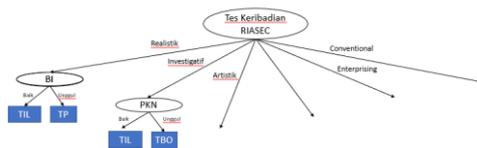
Terlihat pada gambar 4, variabel *realistic* sudah terurai sampai ke node daun, maka ulangi langkah yang sama yaitu hitung *entropy* dan *gain* hingga semua variabel pada atribut tes kepribadian tereksekusi semua. Untuk hasil perhitungan *entropy* dan *gain* variabel *investigative* dapat dilihat pada Tabel 8.

Tabel 8. Hasil *entropy* dan *Gain Investigative*

Atribut	Jml Kasus	T B O	T B I L	T P M y	MEntrop My	Gain Information
Investigatif						
Total	6	1	5	0	0	0.7642
BI						0.109
Baik	4	1	3	0	0	0.8113
Unggul	2	0	2	0	0	0
MTK						0.109
Baik	2	0	2	0	0	0
Unggul	4	1	3	0	0	0.8113
BIng						0.0817
Baik	4	1	3	0	0	0.8113

Atribut	Jml Kasus	T B O	T I L	T P	MEntropy	Gain Informasi
Unggul	2	0	2	0	0	0
PPkn						
Baik	2	1	1	0	0	1
Unggul	4	0	4	0	0	0

Dari Tabel 8, terdapat 2 jurusan yang berpengaruh pada tes kepribadian *investigatif* yaitu TBO dan TIL. Hasil *gain* tertinggi adalah Nilai PKN. Pada nilai PKN karakteristik Unggul, TIL memiliki pengaruh penuh. Oleh karena itu nilai MTK dengan karakteristik Baik berpengaruh pada jurusan TBO. Hasil pohon sementara dapat dilihat pada Gambar 5.



Gambar 5. Hasil Pohon Keputusan Sementara

Perhitungan selanjutnya adalah perhitungan variabel *artistic*. Perhitungan *entropy* dan *gain* variabel *artistic* dapat dilihat pada Tabel 9

Tabel 9. Hasil *Entropy* dan *gain* Artistik

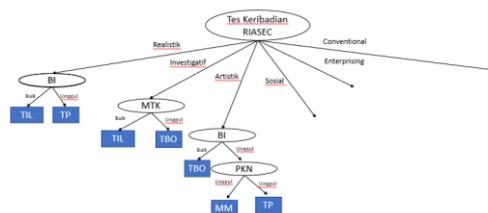
Atribut	Jml Kasus	T B O	T I L	T P	MEntropy	Gain Informasi
Artistik						
Total	12	6	0	1	5	1.3125
BI						0.6549
Baik	5	5	0	0	0	0
Unggul	7	1	0	1	5	1.1488
MTK						0.2213
Baik	4	2	0	0	2	0.5
Unggul	8	4	0	1	3	1.4056
BIng						0.1791
Baik	4	4	0	0	0	0
Unggul	8	2	0	1	5	1.2988
PPkn						0.1791
Baik	5	3	0	1	1	1.371
Unggul	7	4	0	0	4	0.8052

Dari Tabel 9 terlihat *gain* tertinggi adalah Nilai Bahasa Indonesia yaitu 0,6549. Pada nilai Bahasa Indonesia dengan variabel Baik, TBO berpengaruh penuh dalam tes kepribadian *artistic*. Pada variabel Baik terdapat 2 jurusan yang berpengaruh yaitu MM dan TP. Maka perhitungan *entropy* dan *gain* *artistic* berdasarkan nilai Bahasa Indonesia Baik dapat dilihat pada Tabel 10

Tabel 10. Hasil *Entropy* dan *Gain* Artistik berdasarkan BI

Atribut	Jml Kasus	T B O	T I L	T P	MEntropy	Gain Informasi
Artistik						
Total	7	1	0	1	5	1.1488
MTK						0.7538
Baik	2	0	0	0	2	0
Unggul	5	1	0	1	3	1.371
BIng						1
Baik	1	1	0	0	0	0
Unggul	6	0	0	1	5	0.65
PPkn						
Baik	2	0	0	1	1	1
Unggul	5	1	0	0	4	0.7219

Untuk lebih jelasnya, pohon keputusan sementara dapat dilihat pada Gambar 6



Gambar 6. Hasil Pohon Keputusan Sementara

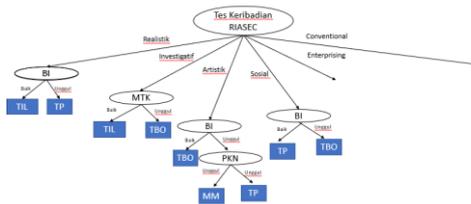
Selanjutnya adalah menentukan *entropy* dan *gain* variabel sosial. Hasil perhitungan variabel sosial terlihat pada Tabel 11.

Tabel 11. Hasil *Entropy* dan *Gain* Sosial

Atribut	Jml Kasus	T B O	T I L	T P	MEntropy	Gain Informasi
Sosial						
Total	2	1	0	1	0	1
BI						1
Baik	1	0	0	1	0	0
Unggul	1	1	0	0	0	0
MTK						0
Baik	0	0	0	0	0	0
Unggul	2	1	0	1	0	1
BIng						1
Baik	1	0	0	1	0	0
Unggul	1	1	0	0	0	0
PPkn						0
Baik	2	1	0	1	0	1
Unggul	0	0	0	0	0	0

Pada variabel Sosial, terdapat 2 jurusan yang berpengaruh yaitu TBO dan TP. Atribut BI dan Bing memperoleh hasil *gain* yang sama yaitu 1, maka pilih salah satunya. Pada BI maupun Bing, Nilai dengan karakteristik Baik sama-sama berpengaruh pada jurusan TP sedangkan

variabel unggul TBO. Agar lebih jelas, hasil pohon sementara terlihat pada Gambar 7



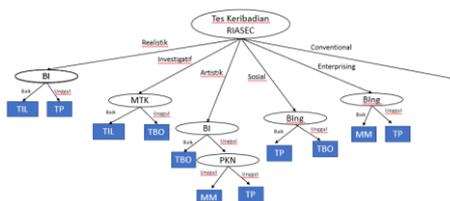
Gambar 7. Hasil Pohon Keputusan Sementara

Pada gambar 7, *node* daun pada variabel Sosial sudah terurai. Langkah selanjutnya adalah menentukan *entropy* dan *gain* variabel *Enterprising*. Hasil perhitungan *entropy* dan *gain* variabel *enterprising* terlihat pada Tabel 12

Tabel 12. Hasil *Entropy* dan *Gain Enterprising*

Atribut	Jml Kasus	T B O	T I L	T P	M Entropy	Gain Information
Enterprising						
Total	5	0	0	1	4	0.7219
BI						0.3219
Baik	2	0	0	1	1	
Unggul	3	0	0	3	0	
MTK						0.171
Baik	2	0	0	0	2	
Unggul	3	0	0	1	2	0.9183
BIing						0.7219
Baik	4	0	0	0	4	
Unggul	1	0	0	1	0	
PPkn						0.0729
Baik	1	0	0	0	1	
Unggul	4	0	0	1	3	0.8113

Pada Tabel 12, Jurusan MM dan TP berpengaruh pada tes kepribadian *enterprising*. Untuk hasil pohon keputusan sementara dapat dilihat pada gambar 8.



Gambar 8. Hasil Pohon Keputusan Sementara

Selanjutnya adalah menentukan *entropy* dan *gain* pada variabel *conventional*. Hasil perhitungan terlihat pada Gambar 17

Tabel 13. Hasil *Entropy* dan *Gain Conventional*

Atribut	Jml Kasus	T B O	T I L	T P	M Entropy	Gain Information
Conventional						
Enterprising						
Total	5	1	2	0	1	1.3171
BI						0.722
Baik	4	1	3	0	0	0.811
Unggul	1	0	0	0	1	0
MTK						0.42
Baik	2	0	1	0	0	1
Unggul	3	1	2	0	0	0.918
BIing						0.42
Baik	3	1	1	0	0	0.585
Unggul	2	0	2	0	0	0
PPkn						0.42
Baik	2	1	1	0	0	1
Unggul	3	0	2	0	0	0.918

Berdasarkan Tabel 13, nilai *gain* tertinggi adalah atribut BI. Pada atribut BI memiliki 2 variabel yaitu unggul dan baik. Pada variabel unggul jurusan MM berpengaruh penuh, sedangkan pada variabel baik terdapat 2 jurusan yang berpengaruh. Maka dari itu harus dilakukan perhitungan *entropy* dan *gain* untuk variabel *conventional* berdasarkan nilai BI Baik. Hasil perhitungan terlihat pada Tabel 14.

Tabel 14. Hasil *Entropy* dan *Gain Conventional* berdasarkan BI

Atribut	Jml Kasus	T B O	T I L	T P	M Entropy	Gain Information
Conventional						
Enterprising						
Total	4	1	3	0	0	0.811
MTK						1.054
Baik	1	0	1	0	0	0
Unggul	3	1	2	0	0	0.5283
BIing						0.971
Baik	3	1	1	0	0	1
Unggul	2	0	2	0	0	0
PPkn						0.971
Baik	2	1	1	0	0	1
Unggul	3	0	2	0	0	0

Semua variabel sudah tereksekusi, maka dari itu sudah didapatkan sebuah pohon keputusan. Pohon keputusan utuh dapat dilihat pada Gambar 9

- Siswa Kelas X MAN 2 Tangerang “Teori Kepribadian Karir John L.Holland”. *Jurnal Ilmiah Bimbingan Konseling Undiksha*, Vol 11, 2020
- [4] S.D. Anderson, “Validitas dan Reabilitas Instrumen Teori Pilihan Karir Holland di Indonesia”. *Bulletin of Conseling and Psichitherypy* vol. 2, no.2, pp. 68–73, 2020
- [5] Devinda Yulia, Kusuma A.P, D.F.H. Permadi. “Penerapan Algoritma C.45 Untuk Prediksi Minat penjurusan Siswa Di SMKN 1 Kademangan”. *Jurnal mahasiswa Teknik informatika* vol. 6, no.2, pp. 683–900, 2022.
- [6] Vol 3, P.B.N Setio, D.R.S Saputro, Wanarno Bowo. Pp 67-71, 2020.
- [7] E. Hasmin and S. Aisa, “Penerapan Algoritma C4.5 Untuk Penentuan Penerima Beasiswa Mahasiswa,” *CogITo Smart J.*, vol. 5, no. 2, p. 308, 2019, doi: 10.31154/cogito.v5i2.219.308-320.
- [8] E. P. Cynthia and E. Ismanto, “Metode Decision Tree Algoritma C.45 Dalam Mengklasifikasi Data Penjualan Bisnis Gerai Makanan Cepat Saji,” *Jurnal Ris. Sist. Inf. dan Tek. Inform.*, vol. 3, no. July, p. 1, 2018.
- [9] S. Wahyuni, “Implementation of Data Mining to Analyze Drug Cases Using C4.5 Decision Tree,” *J. Phys. Conf. Ser.*, vol. 970, no. 1, 2018
- [10] D. Kurniawan, A. Anggrawan, and H. Hairani, “Graduation Prediction System On Students Using C4.5 Algorithm,” *MATRIK Jurnal Manajemen Teknik Informatika dan Rekayasa Komputer.*, vol. 19, no. 2, pp. 358–365, 2020.
- [11] A. Yuliana and D. B. Pratomo, “Memprediksi Kepuasan Mahasiswa Terhadap Kinerja Dosen Politeknik TEDC Bandung,” *Semnasinotek 2017*, pp. 377–384, 2017.
- [12] E. Darmawan, “C4.5 Algorithm Application for Prediction of Self Candidate New Students in Higher Education,” *J. Online Inform.*, vol. 3, no. 1, p. 22, 2018.
- [13] N. Anwar, A. Pranolo, and R. Kurnaiwan, “Grouping the community health center patients based on the disease characteristics using C4.5 decision tree,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 403, no. 1, pp. 1–6, 2018.
- [14] E. P. Cynthia and E. Ismanto, “Metode Decision Tree Algoritma C.45 Dalam Mengklasifikasi Data Penjualan,” *Jurnal Riset Sistem Informasi Dan Teknik Informatika.*, vol. 3, 1–13, 2018.
- [15] Yadi, “Implementasi Algoritma C.45 Clasfifikasi Calon Penerima Beasiswa”, *Jurnal Simantec*, Vol 11. No 1, 2022.
- [16] Sumpena, Akbar Yuma, Nirat. “Penerimaan Calon Siswa Baru dan Penentuan Penjurusan dengan Algoritma C.45 SMK Plus PGRI Cibinong, *Jurnal CKI on Spot*. Vol 11, no 2, 181-191, 2018