Penerapan Algoritma C4.5 Untuk Klasifikasi Kelayakan Pemilihan Program Studi di Universitas Halu Oleo

Ilmi Faizan1\*

1\*Program Studi Teknik Informatika, Fakultas Teknik,

Universitas Halu Oleo Kendari, Indonesia

Email: [ilmifaizan1112@gmail.com](mailto:ilmifaizan1112@gmail.com)

# Abstrak

*Pada perguruan tinggi terdapat banyak sekali konsentrasi jurusan yang dapat dipilih bagi seseorang yang akan menempuh pendidikan sampai tingkatan ini. Faktor penentu keberhasilan dalam menjalani pendidikan dapat dilihat dari pemilihan program studi. Pemilihan program studi yang tidak sesuai akan mempengaruhi kualitas atau output dari perguruan tinggi dan individu itu sendiri, sehingga sangat penting bagi seseorang belajar sesuai dengan bidang kemampuannya. Untuk itu dilakukan identifikasi tingkat kelayakan pemilihan prodi seseorang yang nantinya dapat dijadikan sebagai bahan pertimbangan ketika menjalani program studi yang dipilih. Penelitian ini menggunakan algoritma C4.5 dalam melakukan klasifikasi. Adapun data sampel yang digunakan sebagai bahan penelitian adalah responden mahasiswa/i Universitas Halu Oleo. Hasil akhir penelitian berupa layak atau tidaknya seseorang dalam memilih program studi. Berdasarkan pengujian yang telah dilakukan, menunjukkan bahwa algoritma C4.5 dapat melakukan klasifikasi terhadap kelayakan pemilihan program studi dimana didapatkan hasil akurasi 85% yang artinya akurasi data sudah baik.*

**Kata kunci:** Klasifikasi*,* Algoritma *C4.5,* Program Studi

***Abstract***

*In college there are many concentrations of majors that can be chosen for someone who will study up to this level. The determining factor of success in undergoing education can be seen from the selection of study programs. The selection of an inappropriate study program will affect the quality or output of the college and the individual himself, so it is very important for someone to study according to his field of ability. For this reason, it is necessary to identify the feasibility level of selecting one's study program which can later be used as material for consideration when undergoing the chosen study program. This study uses the C4.5 algorithm in classifying. The sample data used as research material are student respondents at Halu Oleo University. The final result of the research is whether or not a person is eligible in choosing a study program. Based on the tests that have been carried out, it shows that the C4.5 algorithm can classify the feasibility of choosing a study program where the results are 85% accuracy, which means the data accuracy is good.*

**Keywords**: *Classification, Algorithm C4.5, Study Program*

# PENDAHULUAN

Perguruan tinggi adalah salah satu wadah di bidang pendidikan untuk menghasilkan generasi terdidik dan profesinal. Pada perguruan tinggi terdapat banyak sekali konsentrasi jurusan yang dapat dipilih bagi seseorang yang akan menempuh pendidikan sampai tingkatan ini. Faktor penentu keberhasilan dalam menjalani pendidikan dapat dilihat dari pemilihan program studi, apakah pilihan tersebut relevan atau sesuai dengan minat dan bakat. Pemilihan program studi yang tidak sesuai tentu akan mempengaruhi kualitas atau *output* dari perguruan tinggi dan individu itu sendiri, sehingga sangat penting bagi seseorang belajar sesuai dengan bidang kemampuannya. Pemilihan program studi yang tepat membuat proses perkuliahan menjadi efektif. Tetapi sebaliknya pemilihan program studi yang tidak tepat dapat menyebabkan proses perkuliahan menjadi tidak efektif, bahkan bisa jadi seseorang tidak dapat melanjutkan perkuliahan karena merasa bahwa jurusan atau program studi yang dijalani tidak selaras dengan kemampuan ataupun minat.

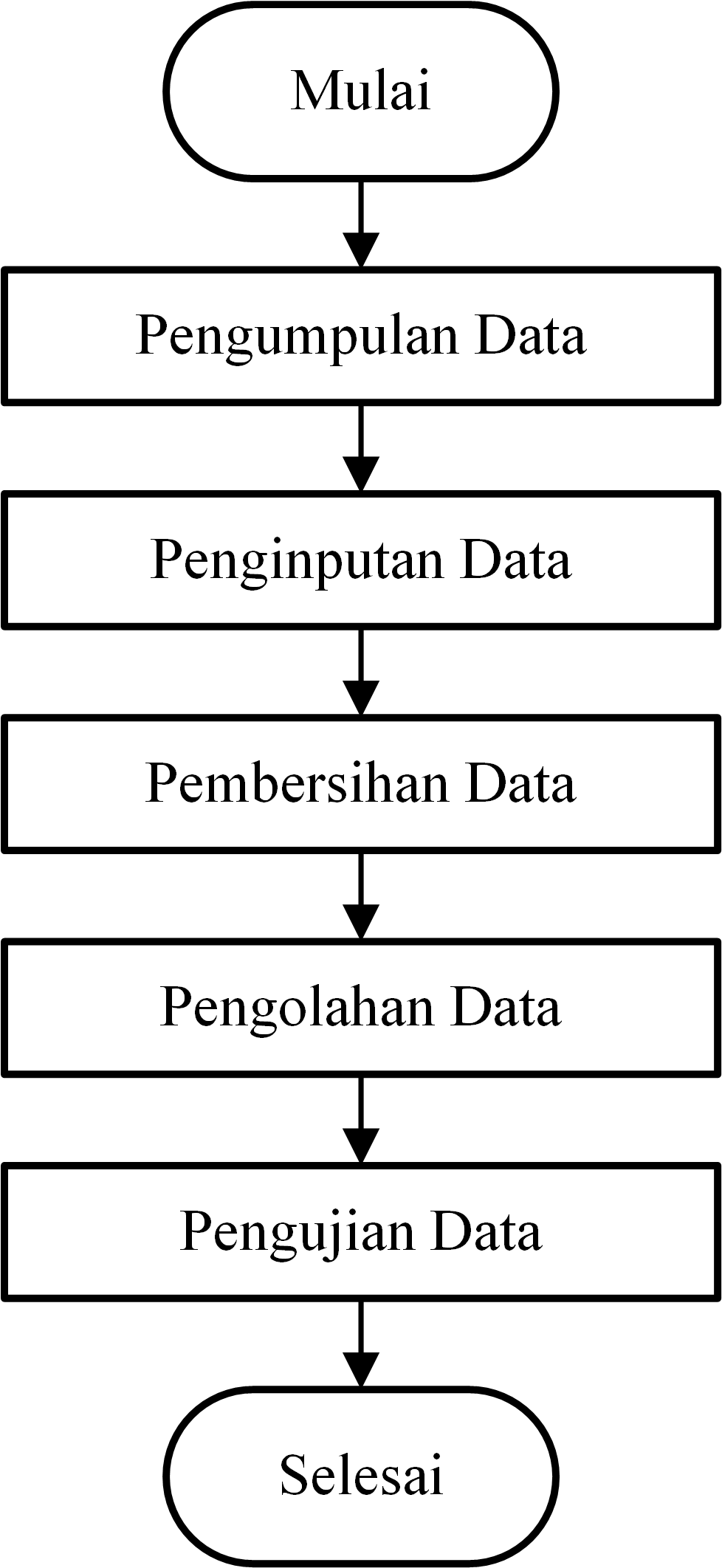
*Data mining* adalah proses ekstraksi data dengan mengumpulkan serta mengolah informasi penting pada data. Pengekstrasian data dapat dilakukan dengan bantuan perangkat lunak *(software)* atau melalui pendekatan matematika maupun statistika. Pembagian kelompok pembahasan *data mining* salah satunya yaitu metode klasifikasi *(Classifier method).* Algoritma *decision tree (C4.5)* merupakan metode pengklasifikasian dengan menggunakan model *tree* atau struktur pohon dalam setiap pengambilan keputusan.

Penelitian ini bertujuan untuk mengidentifikasi tingkat kelayakan pemilihan prodi seseorang yang nantinya dijadikan sebagai bahan pertimbangan ketika menjalani program studi yang dipilih. Hasil akhir penelitian berupa layak atau tidaknya seseorang dalam memilih program studi. Dari pemaparan masalah diatas, terdapat salah satu solusi yang dapat terapkan terhadap permasalahan tersebut yakni melakukan penelitian menggunakan pendekatan *data mining* dimana bertujuan untuk membantu mengklasifikasikan kelayakan pemilihan program studi seseorang dengan menerapkan konsep *decisin tree C4.5.*

# METODE PENELITIAN

## Tahap Penelitian

Model *decision tree* *(C4.5)* merupakan model yang digunakan untuk mengklasifikasikan data penelitian. Rancangan dimulai dari tahap pengumpulan data, penginputan data, pembersihan data, pengolahan data, dan pengujian data. Berikut rancangan penelitian dapat dilihat pada gambar dibawah ini.



Gambar 1. Tahapan Penelitian

1. **Pengumpulan Data**

Metode kuesioner adalah teknik pengambilan yang dipakai dalam penelitian ini. Metode kuesioner berupa pemberian pertanyaan tertulis kepada responden. Adapun target atau objek yang dijadikan penelitian adalah mahasiswa/i Universitas Halu Oleo mulai dari angkatan 2019 sampai angkatan 2021.

1. **Penginputan Data**

Setelah pengumpulan data selesai, selanjutnya data dimasukkan ke Microsoft Excel agar dapat diproses oleh komputer. Data-data telah di-*input-*kankemudian disimpan ke dalam format .csv atau .xlsx

1. **Pembersihan Data**

Pembersihan data atau *data cleaning* bertujuan membuat data menjadi lebih muda dibaca dan memastikan tidak adanya duplikasi pada *dataset*. Pembersihan data dilakukan dengan cara memperbaiki kesalahan struktur pada data, menghilangkan data yang *duplicate* atau tidak lengkap, dan validasi data.

1. **Pengolahan Data**

Tahap berikutnya adalah melakukan pengolahan data menggunakan *software* khusus untuk mengklasifikasikan *dataset* yang ada.

1. **Pengujian Data**

Pengujian data pada penelitian ini yaitu menggunakan metode *split validation.* Dimana membagi *dataset* yang ada secara acak, yaitu *data testing* / data uji dan *data trainig /* data latih. *Data training* digunakan untuk melatih kinerja atau performa dari sebuah algoritma. *Data testing* adalah *dataset* yang telah dibagi untuk menguji tingkat keakuratan dari hasil klasifikasi.

## *Orange Data Mining*

*Orange Data Mining* merupakan sebuah *software* pengolahan data secara visual yang dikembangkan oleh University of Ljubljana. Orange yang dapat digunakan untuk membedakan berbagai data dengan tujuan menghasilkan data yang tepat dan akurat.

## *Classification*

*Classification* atau klasifikasi merupakan pemisahan data ke dalam beberapa kelompok berbeda berdasarkan ciri atau *feature* yang sebelumnya sudah ditentukan. Klasifikasi merupakan bagian dari model *supervised learning* karena memerlukan data latih *(data training)* untuk mengklasifikasikan data ke dalam kelas-kelas tertentu.

## *Decision Tree*

*Decision tree* adalah teknik pemodelan prediktif yang digunakan dalam pengklasifikasian, pengelompokkan, dan juga prediksi. *Decision tree* menggunakan teknik “*conquer and divide*” untuk membagi ruang eksplorasi menjadi beberapa subset. Sebuah *decision tree* atau pohon keputusan diberikan label pertanyaan setelah itu setiap node yang mewakili akan memberikan jawaban untuk pertanyaan-pertanyaan yang terkait.

Secara umum, tahapan dalam membuat struktur *tree* terdiri dari beberapa tahap sebagai berikut.

1. Menentukan *feature* / attribut sebagai *root node.*
2. Membuat cabang untuk tiap-tiap node.
3. Membagi kasus dalam beberapa cabang.
4. Mengulangi proses untuk setiap cabang sampai menghasilkan *class* yang sama.



Keterangan:

S : himpunan kasus

n : jumlah partisi

pi : proporsi Si terhadap S

Untuk menghitung nilai gain, menggunakan rumus sebagai berikut:



Keterangan:

S : himpunan kasus

A : attribut

n : jumlah partisi attribut A

|Si| : jumlah kasus pada partisi ke – i

|S| : jumlah kasus dalam s

# HASIL DAN PEMBAHASAN

Penelitian ini menyajikan informasi mengenai pengolahan data, dibagi menjadi 3 bagian yaitu *dataset, data training,* dan *data testing* menggunakan algoritma *C4.5*.

1. ***Feature* Penelitian**

Adapun *feature* atau attribut yang digunakan dapat terlihat seperti tabel 1 dibawah ini.

Tabel 1. Attribut Penelitian

|  |  |  |  |
| --- | --- | --- | --- |
| No | Attribut | Tipe | Deskripsi |
| 1 | Nama | *Text* | Nama lengkap |
| 2 | Jenis Kelamin | *Categorical* | Jenis kelamin |
| 3 | Jurusan | *Text* | Jurusan mahasiswa di perguruan tinggi |
| 4 | Fakultas | *Text* | Fakultas di perguruan tinggi |
| 5 | Angkatan | *Categorical* | Angkatan mahasiswa |
| 6 | Jurusan di SMA / MA / SMK | *Categorical* | Jurusan pada saat di SMK / MA / SMK |
| 7 | Pemilihan Prodi | *Categorical* | Pemilihan program studi |
| 8 | *Basic* Prodi | *Categorical* | *Basic* mengenai program studi |
| 9 | Sesuai Minat | *Categorical* | Program studi sesuai minat |
| 10 | Urutan Pilihan Prodi | *Categorical* | Urutan pemilihan program studi |

1. ***Data Cleaning***

Pada tabel dibawah, *dataset* belum dilakukan proses *cleaning* data. Datayang terkumpul diperoleh dari hasil pengisian kuesioner di perguruan tinggi Universitas Halu Oleo berjumlah 100 record. *Dataset* sebelum p roses *cleaning* dapat dilihat pada tabel dibawah ini.

Tabel 2. *Dataset*



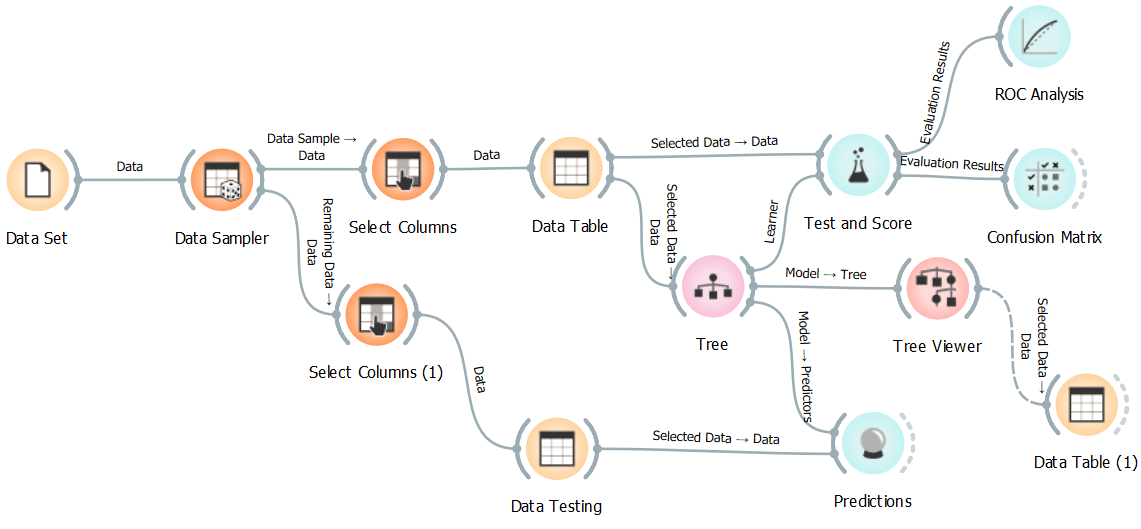
Sebelum melakukan pengujian pada data, beberapa variabel atau attribut dipilih dan pembersihan *record-record* pun juga dilakukan. Attribut pada kolom ke-3 perlu dilakukan kategorisasi. Kategorisai tersebut dibagi menjadi berdasarkan kerelevanan antara jurusan di SMA dan diperkulihan ditentukan sehingga tampilan *dataset* menjadi seperti berikut.

Tabel 3. *Final Dataset*



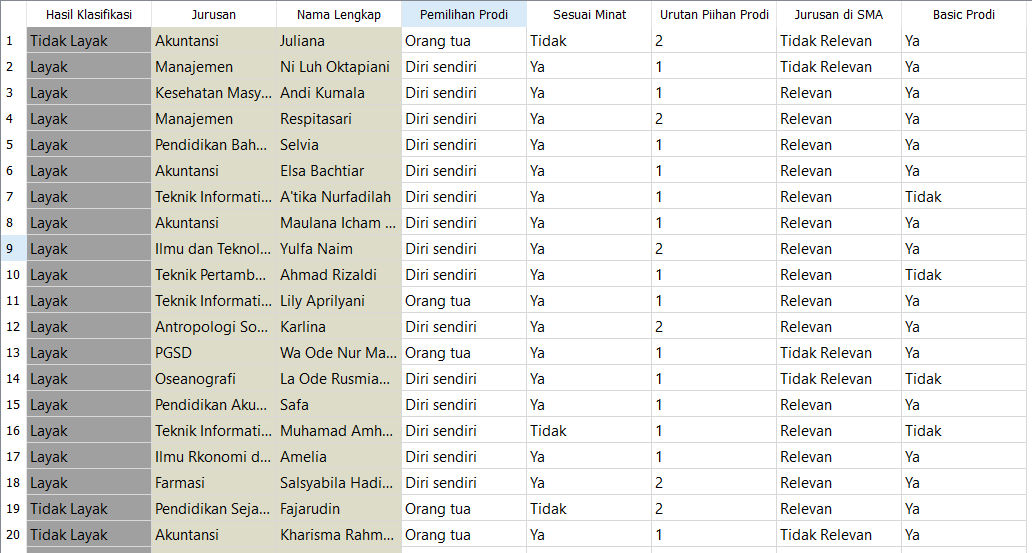
1. **Pengolahan Data Orange**

Dalam proses pengolahan data di orange, beberapa *feature* yang tidak berpengaruh terhadap klasifikasi tidak diikutsertakan seperti, attribut jenis kelamin, fakultas, dan angkatan. Tampilan desain akhir setelah dilakukan pengolahan data di orange sebagai berikut.



Gambar 2. Tampilan Desain Data di Orange

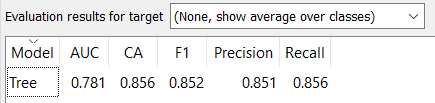
Dapat dilihat pada gambar diatas, *dataset* dibagi menjadi dua bagian menggunakan fitur *data* *sampler* di orange. Jumlah *data* *training* yaitu 90% dari total keseluruhan *dataset* sedangkan *data testing* berjumlah 10% dari total *dataset* yang ada, sehingga jumlah *data testing* yaitu 90 data dan jumlah *data testing* yaitu 10 data. Berikut tampilan dari tabel data latih dan tabel data uji*.*



Gambar 3. Data Latih

 Gambar 4. Data Uji

Klasifikasi data ujidan data latih dilakukan denganmemasukkan 7 buah attribut sebagai *feature* yaitu attribut tipe teks terdiri dari jurusan dan nama lengkap, attribut tipe numerik terdiri dari urutan pemilihan prodi, attribut tipe *categorical* terdiri dari pemilihan prodi, jurusan di SMA, sesuai bakat, urutan prodi, dan *basic* prodi, serta 1 attribut digunakan sebagai attribut *target.* Hasil *test and score* dapat dilihat pada gambar 4 dan hasil pengujian data dapat dilihat pada gambar 5.

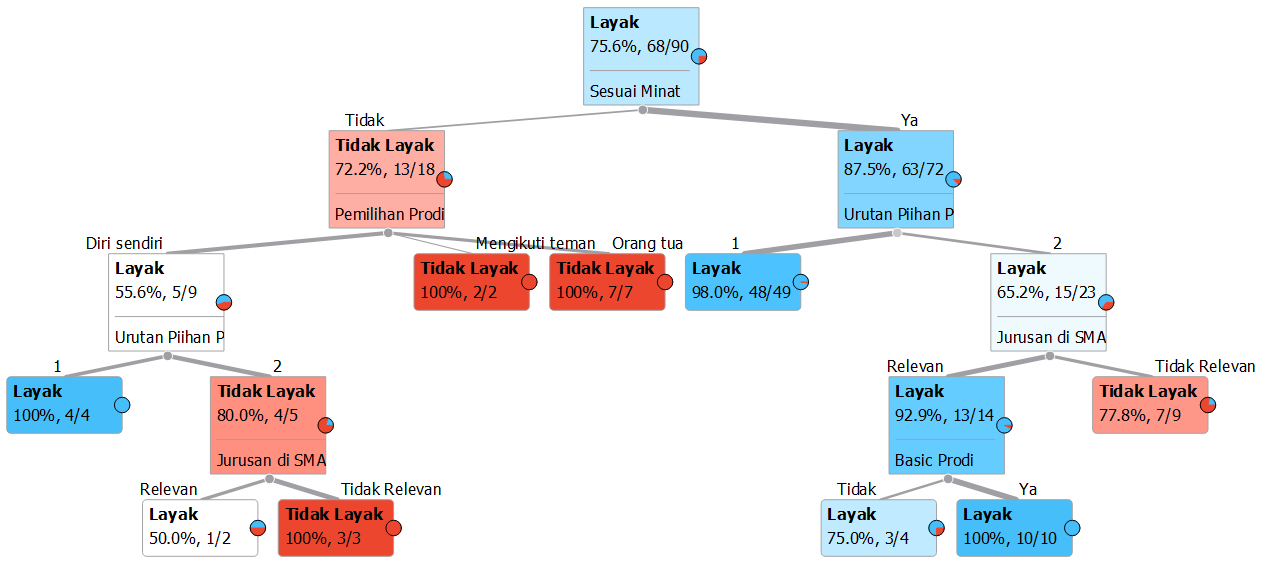


Gambar 5. Hasil *Test and Score*



Gambar 6. Hasil Pengujian

Dari beberapa data yang telah dilakukan pengujian, hasil yang diperoleh menggunakan algoritma C4.5, yaitu 0.856 untuk *accuracy*, 0.851 untuk *precision*, dan 0.856 untuk *recall*. Adapun struktur pohon yang terbentuk dari hasil klasifikasi sebagai berikut.



Gambar 7. Bentuk Pohon Keputusan

# KESIMPULAN

Berdasarkan pengujian yang telah dilakukan menunjukkan bahwa algoritma C4.5 dapat melakukan klasifikasi terhadap kelayakan pemilihan program studi dimana didapatkan hasil akurasi 85% yang artinya akurasi data sudah baik. Variabel yang menjadi faktor penentu kelayakan pemilihan prodi yaitu variabel kesesuaian minat prodi yang dipilih.

# DAFTAR PUSTAKA

1. Fadma Ristianti, D. (2019). Komparasi Algoritma Klasifikasi pada Data Mining. PROCEEDINGS OF THE 1 St STEEEM, 1(1), 148–156.
2. Hidayanti, I., Kurniawan, T. B., & Afriyudi, A. (2020). Perbandingan Dan Analisis Metode Klasifikasi Untuk Menentukan Konsentrasi Jurusan. Jurnal Ilmiah Informatika Global, 0 10 20 30 40 50 60 70 80 Decision Tree Naïve Bayes k-NN Chart Title Accuracy AUC
3. Marlina, D., & Bakri, M. (2021). Penerapan Data Mining Untuk Memprediksi Transaksi Nasabah Dengan Algoritma C4.5. Jurnal Teknologi Dan Sistem Informasi (JTSI), 2(1), 23–28.
4. Mafakhir, A. Z., & Solichin, A. (2020). Penerapan Metode Naïve Bayes Classifier Untuk Penjurusan Siswa Pada Madrasah Aliyah Al-Falah Jakarta. Fountain of Informatics Journal, 5(1), 21. [https://doi.org/10.21111/fij.v5i1.400 7](https://doi.org/10.21111/fij.v5i1.400%207)
5. Suryadi, A., & Harahap, E. (2018). Sistem Rekomendasi Penerimaan Mahasiswa Baru Menggunakan Naive Bayes Classifier Di Institut Pendidikan Indonesia. Joutica, 3(2), 171. <https://doi.org/10.30736/jti.v3i2.231>
6. Kusrini dan Luthfi, E.T., Algoritma Data Mining, Penerbit Andi, Yogyakarta, 2009.
7. Supartha, I Kadek D.G., Dewi, I Gusti Ayu P.E.P., Sistem Pendukung Keputusan Penentuan Jurusan Pada SMK Kertha Wisata Denpasar Menggunakan Fuzzy SAW, Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI), ISSN 2089 – 8673 No. 2 Vol. 3, Juli 2014.
8. Hermawati, A.F. 2013. Data Mining. Surabaya: Universitas 17 Agustus 1945 & ANDI.
9. Kamagi, D. H., & Hansun, S. (2014). Implementasi Data Mining dengan Algoritma C4 . 5 untuk Memprediksi Tingkat Kelulusan Mahasiswa, VI(1), 15–20.
10. Ariestya, W. W., & Gunadarma, U. (2017). DECISION TREE LEARNING UNTUK PENENTUAN JALUR KELULUSAN Decision Tree Learning Untuk Penentuan Jalur Kelulusan Mahasiswa, (May 2016). https://doi.org/10.22441/fifo.v8i1.1304