

# Optimasi Algoritma C4.5 Menggunakan Metode Forward Selection Dan Stratified Sampling Untuk Prediksi Kelayakan Mahasiswa Penerima Beasiswa

Bentar Candra P\*, Kusrini, Tonny Hidayat

Pascasarjana, Magister PJJ Informatika, Amikom Yogyakarta, Yogyakarta, Indonesia

Email: <sup>1,\*</sup>ben@students.amikom.ac.id, <sup>2</sup>kusrini@amikom.ac.id, <sup>3</sup>tonny@amikom.ac.id

Email Penulis Korespondensi: ben@students.amikom.ac.id

**Abstrak**—Setiap calon mahasiswa memiliki peluang untuk mendapatkan beasiswa di lingkungan institusi pendidikan namun seringkali tidak tepat sasaran sehingga dibutuhkan pendekatan data mining yang lebih akurasi. Namun, algoritma C4.5 memiliki kelemahan pada tingkat akurasi saat mengelola data dalam jumlah yang besar sehingga perlu di optimasi. Penelitian ini bertujuan untuk mengoptimasi algoritma C4.5 menggunakan metode stratified sampling dan forward selection dalam penentuan kelayakan penerima beasiswa. Data berasal dari calon mahasiswa di Universitas Anwar Medika dengan jumlah sampel sebanyak 263 *record* yang kemudian diolah menggunakan aplikasi RapidMiner untuk algoritma C4.5 tanpa optimasi dan algoritma C4.5 dengan optimasi metode *stratified sampling+forward selection*. Hasil penelitian menunjukkan tingkat akurasi lebih tinggi pada algoritma C4.5 dengan optimasi menggunakan metode *stratified sampling+forward selection* yakni sebesar 81,75% dibandingkan dengan tingkat akurasi pada algoritma C4.5 tanpa optimasi yakni sebesar 80,23%. Dengan demikian, kesimpulan dari penelitian ini yaitu algoritma C4.5 dengan optimasi menggunakan metode *stratified sampling* dan *forward selection* lebih efektif dan dapat mengatasi kekurangan algoritma C4.5 tanpa optimasi.

**Kata Kunci:** Data Mining; Stratified Sampling; Forward Selection; Algoritma C4.5; Beasiswa

**Abstract**—Every prospective student has the opportunity to get a scholarship within an educational institution, but it is often not on target so a more accurate data mining approach is needed. However, the C4.5 algorithm has a weakness in its level of accuracy when managing large amounts of data so it needs to be optimized. This research aims to optimize the C4.5 algorithm using stratified sampling and forward selection methods in determining the eligibility of scholarship recipients. The data came from prospective students at Anwar Medika University with a sample size of 263 records which were then processed using the RapidMiner application for the C4.5 algorithm without optimization and the C4.5 algorithm with optimization of the stratified sampling + forward selection method. The research results show a higher level of accuracy in the C4.5 algorithm with optimization using the stratified sampling + forward selection method, namely 81.75% compared to the accuracy level in the C4.5 algorithm without optimization, namely 80.23%. Thus, the conclusion of this research is that the C4.5 algorithm with optimization using stratified sampling and forward selection methods is more effective and can overcome the shortcomings of the C4.5 algorithm without optimization.

**Keywords:** Data Mining; Stratified Sampling; Forward Selection; Algorithm C4.5 Optimization; Scholarship

## 1. PENDAHULUAN

Dunia pendidikan merupakan aspek yang penting dalam kehidupan mengingat Ki Hajar Dewantara (bapak pendidikan nasional) menyebutkan bahwa pendidikan itu mengalami proses yang panjang hingga akhir hayat mengingat kehidupan adalah bagian dari proses belajar [1]. Proses itulah yang dibentuk dalam pendidikan untuk membangun suatu karakter maupun kemampuan setiap individu. Pada proses pendidikan meliputi tiga aspek yaitu kognitif, afektif, dan psikomotorik. Ketiga tersebut harus dicapai dengan baik dan sempurna untuk menciptakan suatu keberhasilan sehingga pemerintah terus mengupayakan penerapan ketiga aspek tersebut di institusi pendidikan [2]. Seperti pasal 31 ayat 1-2 yang tertuang pada UUD 1945 bahwa tiap warga negara Indonesia memiliki hak dalam memperoleh pendidikan. Berbagai program pemerintah terus di upayakan supaya warga negara Indonesia terus bisa memperoleh pendidikan salah satunya menwarkan program beasiswa bagi warga yang kurang mampu [3] terutama beasiswa untuk program perguruan tinggi. Setiap perguruan tinggi negeri maupun swasta pasti terdapat program beasiswa karena biaya perkuliahan masih terhitung tinggi. Program beasiswa yang ditawarkan terbagi menjadi dua kelompok yaitu beasiswa internal dan eksternal. Beasiswa internal merupakan beasiswa yang berasal dari institusi itu sendiri baik itu secara penuh maupun sebagian sedangkan beasiswa eksternal merupakan beasiswa yang berasal dari luar institusi yang sumber pendanaannya biasanya diadakan oleh pemerintah dan swasta [4].

Setiap warga negara Indonesia memiliki peluang untuk mendapatkan beasiswa [5] sehingga setiap institusi pendidikan banyak sekali yang menyumbangkan beasiswa yang ditujukan kepada calon mahasiswa yang ingin kuliah seperti beasiswa yang diberikan oleh lembaga pemerintah, yayasan atau perusahaan seperti Universitas Anwar Medika yang mempunyai Program Beasiswa yang disumbangkan oleh yayasan Anwar Medika. Program tersebut bermaksud untuk memberikan bantuan biaya kuliah yang diberikan kepada calon mahasiswa yang orang tuanya di rasa kurang mampu. Untuk memperoleh beasiswa tersebut, dilakukan seleksi oleh pihak universitas supaya dapat dimanfaatkan sebaik mungkin sehingga bagi yang mendapatkannya, calon mahasiswa tersebut dapat menjalankan perkuliahan dengan baik. Namun, saat dilakukan seleksi, sering menjumpai permasalahan yang dialami. Salah satunya, banyaknya penerima beasiswa yang tidak tepat sasaran [6]. Hal ini dikarenakan banyaknya calon mahasiswa yang mengajukan beasiswa tersebut dan sistem seleksi yang dilakukan oleh pihak universitas masih bersifat konvensional. Sehingga pemanfaatan beasiswa masih kurang optimal [7]. Dengan demikian, berdasarkan permasalahan tersebut dibutuhkan pendekatan data mining untuk memprediksi kelayakan penerima beasiswa. Data mining adalah proses

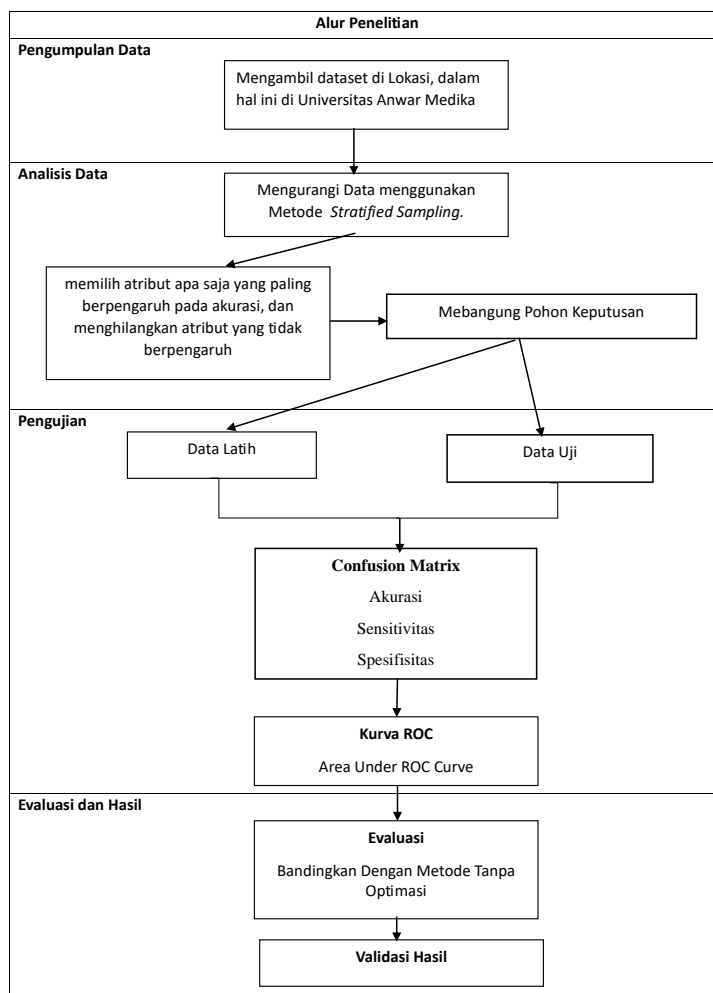
penggalan atau penggumpalan informasi pada sebuah data penting dengan kadar besar. Biasanya digunakan dalam berbagai macam bidang ilmu seperti artificial intelligence, machine learning atau pembelajaran mesin, teknik statistika maupun matematika [6]. Data Mining harus mempelajari metode untuk mencari pengetahuan atau menemukan pola dari suatu data yang besar [8].

Ada beberapa algoritma data mining yang bisa digunakan dalam prediksi kelayakan penerima beasiswa, seperti Naïve Bayes, K-Nearest Neighbor (KNN), dan C4.5 [9]. Algoritma K-Nearest Neighbor mempunyai kelebihan pada model pembelajaran yang tidak mengansumsikan apa-apa terhadap distribusi data (non parametrik), tetapi memiliki kelemahan rentan dengan atribut yang non-informatif. Naïve bayes mempunyai kelebihan pada model yang dibuat dan proses begitu cepat tetapi memiliki kelemahan hanya mengandalkan satu probabilitas (peluang) saja saat mengukur akurasi sedangkan algoritma C4.5 dapat mengatasi kerentanan atribut non-informatif dan pengukuran akurasi yang mengandalkan satu probabilitas saja dengan membuat model pohon keputusan [10]. Akan tetapi, pohon keputusan memiliki kelemahan ketika dihadapkan dengan atribut dan jumlah data yang besar [11]. Penelitian lain juga menyebutkan bahwa tingkat akurasi dengan algoritma C4.5 dalam menentukan penerimaan beasiswa masih kurang yakni hanya 57% [12]. sehingga perlu dilakukan optimasi algoritma C4.5 dengan metode Forward Selection dan Stratified Sampling karena menurut penulis dapat meminimalisir kelemahan dan membatasi besarnya jumlah data. Berdasarkan uraian diatas, penulis tertarik melakukan penelitian tentang prediksi kelayakan penerima beasiswa menggunakan algoritma C4.5 yang akan di optimasi dengan metode stratified sampling dan forward selection.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahap Penelitian

Dalam menjalankan penelitian ini, peneliti melakukan beberapa tahapan. Berikut pada gambar 1 merupakan tahapan - tahapan yang dilakukan dalam penelitian :



Gambar 1. Tahap Penelitian

Berdasarkan gambar 1, tahap penelitian ini terbagi menjadi empat bagian yaitu Pengumpulan data, analisis data, pengujian dan Evaluasi dan hasil . Berikut penjelasan tentang tahapannya :

a. Pengumpulan data

- Mengumpulkan dataset mahasiswa Universitas Anwar Medika yang akan digunakan sebagai bahan penelitian
- Analisis Data  
Menganalisa data yang terkumpul yang kemudian diseleksi menggunakan metode *stratified sampling* dengan cara memilih atribut yang paling berpengaruh dan menghilangkan atribut yang tidak berpengaruh yang selanjutnya dibuat pohon keputusan
  - Pengujian  
Hasil dari analisa data dilakukan pengujian menggunakan optimasi Algoritma C4.5 menggunakan metode *Stratied Sampling* dan *Metode Forwad Selection*
  - Evaluasi dan Hasil  
Hasil dari pengujian akan di evaluasi dengan membandingkan metode Algoritma C4.5 tanpa optimasi dan menggunakan Algoritma C4.5 dengan optimasi yang selanjutnya dilakukan tahap validasi hasil

## 2.2 Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang menjadi bagian gabungan dari algoritma decision tree[13]. Pada langkah ini dilakukan dengan perancangan dari model sistem yang menggunakan algoritma C4.5 agar dapat membentuk sebuah pohon keputusan hingga dapat menghasilkan rule prediksi penerimaan beasiswa. Algoritma C4.5 dianggap algoritma yang dapat membantu dalam melakukan sebuah klasifikasi data karena dalam karakter data yang diklasifikasikan dapat diperoleh dengan baik dan tepat, maka dapat memudahkan peneliti menggali informasi terhadap data tersebut [14]. Algoritma C4.5 yang dianggap sebagai yang sangat dalam membantu klasifikasi data karena karakter data yang akan diklasifikasikan terdapat dengan jelas, baik dalam gambaran struktur pohon keputusan (rule) ataupun ke dalam aturan if then, sehingga dapat memudahkan user dalam melakukan observasi informasi dengan data yang berkaitan. Pada penyeleksian atribut dalam penelitian ini sangat mempengaruhi dalam pengolahan algoritma C4.5 karena keputusan ini sangat bergantung pada atribut yang dipilih[15].

Dengan penjelasan di atas, maka peneliti menggunakan algoritma C4.5 untuk membantu agar lebih mudah memprediksi suatu nilai informasi, dan menurut peneliti, algoritma C4.5 merupakan pengembangan dari ID3 dengan membentuk pohon keputusan untuk memprediksi atau memperkirakan suatu peristiwa atau kejadian, untuk memperoleh informasi yang dibutuhkan. Adapun tahapan dalam Algoritma C4.5 [16] sebagai berikut:

- Menentukan nilai *gain* dengan cara memilih atribut setiap simpul pada pohon keputusan untuk menentukan variabel mana yang akan menjadi *node* pada suatu pohon keputusan. Adapun nilai *gain* dihitung dengan cara sebagai berikut:

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

Dengan  $S$  adalah himpunan kasus,  $A$  adalah atribut,  $n$  adalah jumlah partisi atribut  $A$ ,  $|S_i|$  adalah jumlah kasus pada partisi ke  $i$ , dan  $|S|$  adalah jumlah kasus dalam  $S$ .

- Menentukan nilai entropy. Nilai tersebut digunakan untuk mengetahui perbedaan keputusan pada nilai tertentu dengan rumus sebagai berikut:

$$Entropy \sum_{i=1}^n -p_i * \log_2 p_i \quad (2)$$

Dengan  $S$  adalah himpunan kasus,  $A$  adalah Fitur,  $n$  adalah jumlah partisi, dan  $P_i$  adalah proporsi dari  $S_i$  dan  $S$ .

## 2.3 Stratified Sampling

Stratified Sampling merupakan teknik pengambilan sampel secara acak dengan memperhatikan tingkatan (strata) pada populasi. Dataset dibagi menjadi beberapa bagian yang terpisah (strata), kemudian sampel diambil secara acak berdasarkan strata yang sudah dibuat. Adapun tahapan Stratified Sampling [11] sebagai berikut:

- Populasi  $N$  dibagi menjadi beberapa sub populasi yang tiap sub populasi terdiri dari elemen  $N_1, N_2, N_3, \dots, N_L$ .
- Antara sub-sub populasi tidak boleh ada tumpang tindih, sehingga  $N_1 + N_2 + N_3 + \dots + N_L = N$ .
- Mengambil sampel secara acak dari masing masing sub populasi dengan alokasi sampel yang proporsional. Sebelum pengambilan sampel, ditentukan ukuran sampel karena sampel yang diambil harus mencerminkan populasi.

## 2.4 Metode Forward Selection

Data set bisa berisi ratusan atribut, yang diantaranya mungkin tidak relevan dengan pola yang ingin dicari. Misalnya ingin mengklasifikasi status kredit nasabah yang kreditnya macet atau lancar, atribut seperti nomor telepon mungkin tidak relevan bila dibandingkan dengan atribut usia dan riwayat kredit. Mempertahankan atribut yang tidak relevan dapat memperlambat proses pencarian pola dan membuat algoritma data mining bingung sehingga menghasilkan pola yang buruk. Untuk mengatasi atribut yang tidak relevan, dapat dilakukan dengan cara dihapus [11].

Forward Selection merupakan salah satu Feature Selection yang bertujuan untuk mendapatkan kumpulan atribut yang minim dari sekumpulan atribut yang berjumlah banyak. Manfaat Feature Selection adalah membantu membuat pola yang mudah dipahami. Prosedur ini dimulai dengan himpunan kosong dari atribut yang akan dikurangi. Selanjutnya atribut tersebut diuji satu-persatu dan dipilih atribut terbaik dengan akurasi paling tinggi. Kemudian dilakukan iterasi pengujian berikutnya secara terus menerus dan berhenti jika atribut yang diuji tidak memberikan akurasi yang signifikan. Adapun tahapannya sebagai berikut [6]:

Atribut asli: {A01, A02, A03, A04, A05, A06}

Tahap pengurangan atribut:

1. {A01}
2. {A01, A04}
3. Pengurangan atribut {A01, A04, A06}

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Hasil Pengelompokan Data

Penelitian ini menggunakan metode penelitian eksperimen dan studi pustaka, Dataset yang diambil berasal dari Universitas Anwar Medika dengan jumlah 263 *record* terdiri dari 65 sampel layak dan 198 tidak layak. Dataset ini memiliki 9 atribut yang terdiri dari 8 atribut nominal dan 1 atribut numerik. Adapun karakteristik dataset pada tabel 1 sebagai berikut:

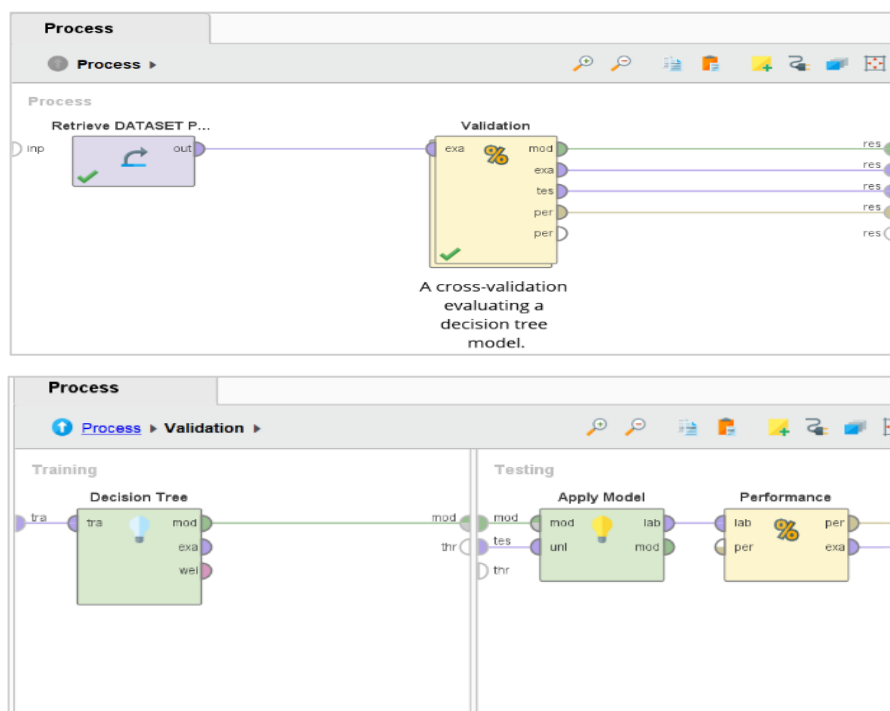
**Tabel 1.** Karakteristik Dataset di Universitas Anwar Medika

| No | Nama                   | Tipe       |
|----|------------------------|------------|
| 1  | Nama Sesuai Ijazah     | Polynomial |
| 2  | Program Studi          | Polynomial |
| 3  | Jenis Kelamin          | Binomial   |
| 4  | Alamat                 | Polynomial |
| 5  | Kecamatan              | Polynomial |
| 6  | Alat Transportasi      | Polynomial |
| 7  | Pekerjaan Ayah         | Polynomial |
| 8  | Penghasilan Ayah       | Polynomial |
| 9  | Layak atau Tidak Layak | Binomial   |

Pada tabel 1 diketahui atribut polynomial sebanyak 7 sedangkan binomial sebanyak 2 dari total 9 atribut. Adapun atribut yang termasuk tipe polynomial dalam penelitian ini yaitu nama, program studi, alamat, kecamatan, alat transportasi, pekerjaan ayah, dan penghasilan ayah sedangkan yang termasuk tipe binomial yaitu atribut jenis kelamin dan layak atau tidak layak. Untuk atribut Nama Sesuai Ijazah, Alamat, dan Kecamatan di hapus karena tidak diperlukan dalam penentuan pohon keputusan.

#### 3.2 Hasil Pengujian Algoritma C4.5 Tanpa Optimasi

Pada penelitian ini, dataset di uji dengan algoritma C4.5 tanpa optimasi menggunakan aplikasi RapidMinner yang terlihat pada gambar 1.



**Gambar 1.** Tampilan Uji Algoritma C4.5 tanpa optimasi di aplikasi RapidMinner

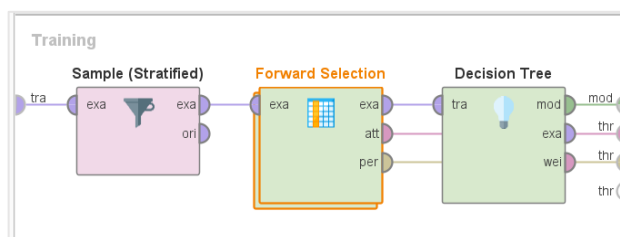
**Tabel 2.** Hasil Uji Algoritma C4.5

| Model                         | Akurasi | Sensitivitas | Spesifisitas | PPV    | NPV    | AUC    |
|-------------------------------|---------|--------------|--------------|--------|--------|--------|
| Algoritma C4.5 tanpa Optimasi | 80,23%  | 93,43%       | 40,00%       | 82,59% | 66,67% | 0,769% |

Pada tabel 2 diketahui nilai algoritma C4.5 tanpa optimasi, nilai akurasi sebesar 80,23%, sensitivitas sebesar 93,43%, spesifisitas 40,00%, PPV sebesar 82,59%, NPV sebesar 66,67%, dan AUC sebesar 0,769%. Hal ini sesuai dengan penelitian [17] bahwa algoritma C4.5 efektif dalam memprediksi kelayakan penerimaan beasiswa dengan metode manual sebelumnya. Namun penggunaan algoritma C4.5 kurang optimal jika digunakan dalam sampel yang lebih besar sehingga perlu dilakukan optimasi untuk mengetahui perubahan peningkatan akurasi [18].

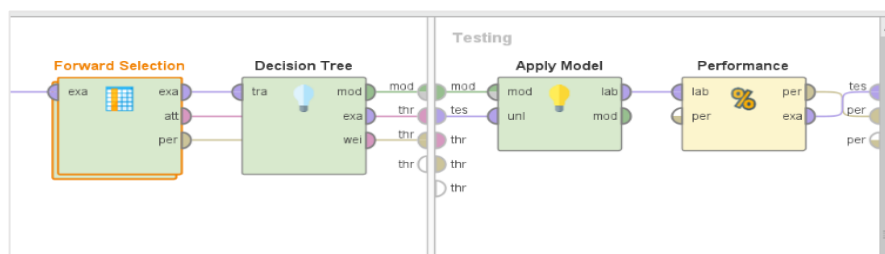
### 3.3 Hasil Pengujian Optimasi Algoritma C4.5 dengan *Stratified sampling* dan *Forward Selection*

Pada tahap ini pengujian dilakukan dengan optimasi menggunakan metode *stratified sampling* (gambar 2). Data yang sudah terkumpul diolah menggunakan *Stratified sampling* sebelum dibuat pemodelan. Sampel data yang diambil secara acak dari populasi data memperhatikan tingkat distribusi data untuk membentuk sebuah dataset baru.



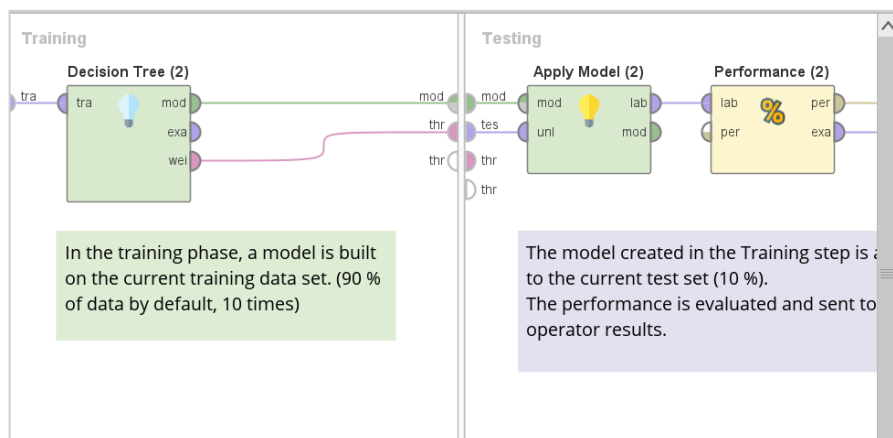
**Gambar 2.** Tampilan Metode *Stratified Sampling* pada aplikasi RapidMiner

Hasil dari *stratified sampling* terjadi pemilihan atribut terbaik dan atribut yang tidak memiliki kontribusi ke akurasi menggunakan metode *Forward Selection* (gambar 3) dimana prosesnya dilakukan secara terus-menerus. Proses akan berhenti jika atribut yang diuji sudah tidak memberikan peningkatan akurasi yang signifikan.



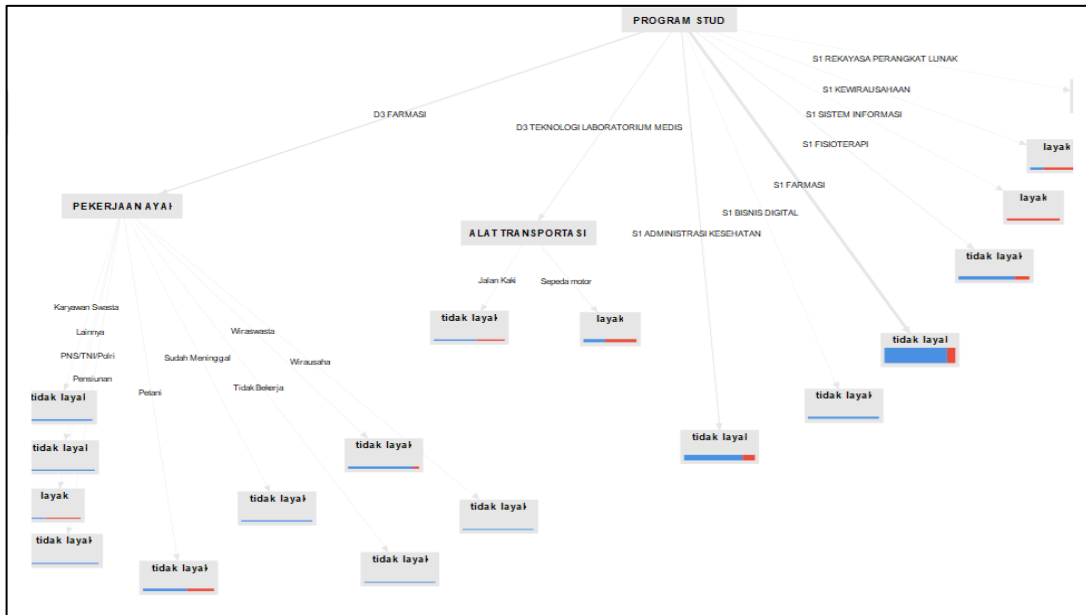
**Gambar 3.** Tampilan metode *Forward Selection* pada Aplikasi RapidMiner

*Tool* yang digunakan bagian kanan pada gambar 4 yaitu *apply model* dan *performance*. Kedua *tool* tersebut berfungsi untuk menerapkan model yang telah dilatih pada data baru dan membuat prediksi berdasarkan model tersebut. *Tools Performance* digunakan untuk memudahkan melihat nilai akurasi dan efektifitas model. Pada tahap ini, peneliti membagi data latih (data training) dan data uji (data testing) menjadi 10 subset secara acak (number of folds =10)



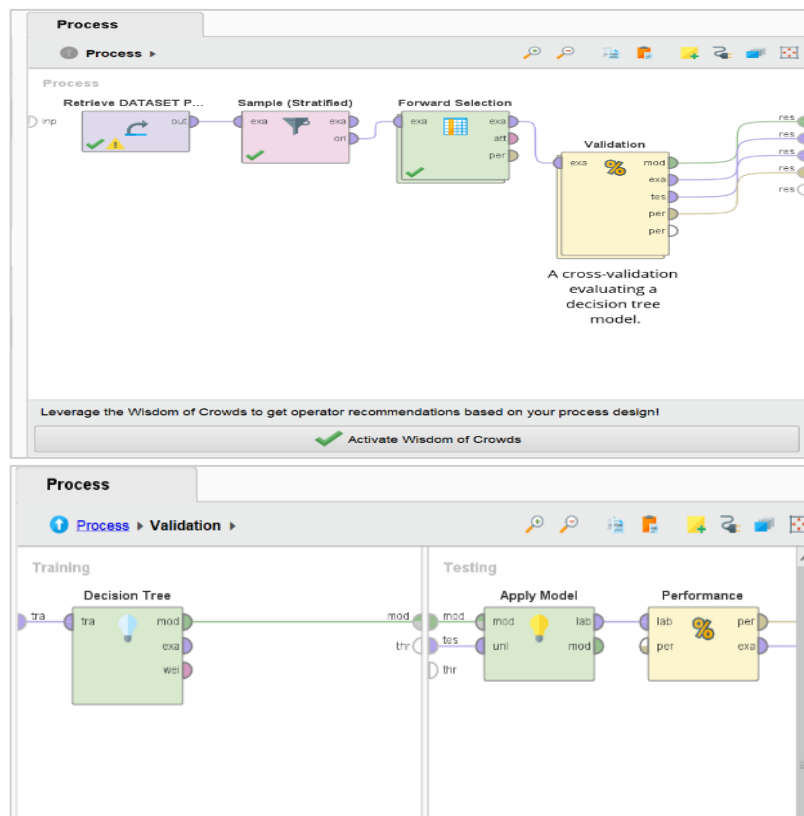
**Gambar 4.** Tampilan proses *modelling* pada Aplikasi RapidMiner

Hasil *modelling* didapatkan tampilan pohon keputusan seperti pada gambar 5. Selanjutnya hasil dari pohon keputusan dilakukan uji *cross validation* untuk mengetahui nilai akurasi dan AUC.



Gambar 5. Tampilan Pohon Keputusan

Proses *cross validation* setelah dilakukan optimasi terlihat pada gambar 6 menggunakan aplikasi RapidMiner. Bagian kanan gambar, *tool* yang digunakan yaitu *apply model* dan *performance*.



Gambar 6. Tampilan Pengujian Model setelah dilakukan optimasi dan pengukuran *performance* pada Aplikasi RapidMiner

Tabel 3. *Confusion Matrix*

|                      | Kenyataan Tidak Layak | Kenyataan Layak | Total |
|----------------------|-----------------------|-----------------|-------|
| Prediksi Tidak Layak | 187                   | 37              | 224   |
| Prediksi Layak       | 11                    | 28              | 39    |
| Total                | 198                   | 65              | 263   |

Tabel 3 merupakan *confusion matrix* (hasil klasifikasi prediksi) sebelum diketahui nilai akurasi dan AUC. Pada klasifikasi (prediksi) diketahui dari 263 sampel terdapat prediksi tidak layak sebesar 187 data dan prediksi layak sebesar 11 data sehingga kenyataan tidak layak menunjukkan jumlah data sebesar 198. Kemudian pada prediksi tidak layak pada kenyataan layak didapatkan 39 data dan prediksi layak sebesar 26 data sehingga jumlah kenyataan layak sebesar 65 data. Oleh karena itu, hasil diatas dilakukan perhitungan akurasi, sensitivitas, spesifitas, PPV, dan NPV dengan persamaan sebagai berikut:

$$\text{Akurasi} = \frac{187+28}{187+28+37+11} \times 100\% = 81,75\%$$

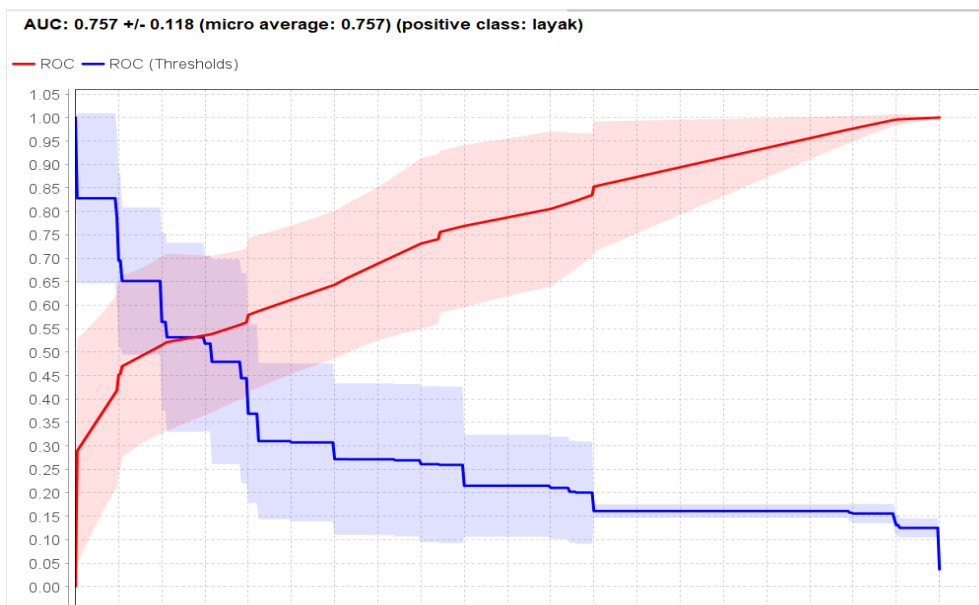
$$\text{Sensitivitas} = \frac{187}{187+11} \times 100\% = 94,44\%$$

$$\text{Spesifisitas} = \frac{28}{28+37} \times 100\% = 43,08\%$$

$$\text{PPV} = \frac{187}{187+37} \times 100\% = 83,48\%$$

$$\text{NPV} = \frac{28}{28+11} \times 100\% = 71,79\%$$

Hasil perhitungan diatas didapatkan hasil akurasi sebesar 81,75%; sensitivitas sebesar 94,44%; spesifitas sebesar 43,08%; PPV sebesar 83,48%; dan NPV sebesar 71,79%. Berdasarkan hasil perhitungan tersebut, perlu dilakukan pembuatan kurva ROC (gambar 7) untuk mengetahui nilai *Area Under ROC Curve* (AUC)[19] dan hasil menunjukkan nilai AUC sebesar 0.757.



**Gambar 7.** Tampilan Kurva ROC

### 3.4 Hasil Perbandingan Algoritma C4.5 tanpa Optimasi dan Algoritma C4.5 dengan Optimasi+ *Stratified sampling + Forward Selection*

Pada penelitian ini menunjukkan bahwa terdapat perbedaan hasil pengukuran algoritma C4.5 tanpa optimasi dengan algoritma C4.5 menggunakan optimasi metode *stratified sampling* dan *forward selection* dimana pada tabel 5 algoritma C4.5 tanpa optimasi didapatkan hasil akurasi sebesar 80,23%, sensitivitas sebesar 93,43%; spesifitas 40,00%, PPV sebesar 82,59%, NPV sebesar 66,67%, dan AUC sebesar 0,769% sedangkan algoritma C4.5 dengan optimasi menggunakan optimasi didapatkan hasil akurasi sebesar 81,75%, sensitivitas sebesar 94,44%, spesifitas sebesar 43,44%, PPV sebesar 83,08%, NPV sebesar 71,79%, dan AUC sebesar 0,757%.

**Tabel 4.** Hasil Pengukuran sebelum optimasi dan sesudah optimasi

| Model   | Akurasi | Sensitivitas | Spesifisitas | PPV    | NPV    | AUC    |
|---|---------|--------------|--------------|--------|--------|--------|
| Algoritma C4.5  | 80,23%  | 93,43%       | 40,00%       | 82,59% | 66,67% | 0,769% |
| Algoritma C4.5+Starified Sampling + Forward Selection | 81,75%  | 94,44%       | 43,08%       | 83,48% | 71,79% | 0,757% |

Hal ini sesuai dengan penelitian [11] bahwa algoritma C4.5 dengan optimasi menggunakan metode *stratified sampling* dan *forward selection* berhasil meningkatkan akurasi sebesar 1,52%. Penelitian lain [20] juga menyebutkan bahwa optimasi dengan *stratified sampling* dan *forward selection* dapat mengatasi kelemahan dari algoritma C4.5. Dengan demikian dari hasil pengukuran yang dilakukan diatas, menunjukkan bahwa nilai akurasi pada algoritma C4.5

tanpa optimasi memiliki nilai rendah dibandingkan dengan algoritma C4.5 yang sudah dioptimasi dengan metode *stratified sampling* dan *forward selection*.

#### 4. KESIMPULAN

Berdasarkan hasil dari penelitian dan pembahasan diatas, dapat disimpulkan untuk menjawab rumusan masalah dalam penelitian ini, yang pertama adalah Algoritma C4.5 telah terbukti efektif dalam memprediksi kelayakan penerima beasiswa di Universitas Anwar Medika dengan hasil nilai akurasi sebesar 80,23% sedangkan kesimpulan yang kedua setelah dilakukan optimasi pada Algoritma C4.5 dengan menggunakan Metode Stratified Sampling dan Forward Selection dengan hasil nilai akurasi sebesar 81,75%. Dan kesimpulan yang terakhir atau ketiga bahwa dengan dilakukan optimasi algoritma C4.5 menggunakan metode *stratified sampling* dan *metode forwarding selection* telah terbukti berhasil meningkatkan akurasi algoritma C4.5 sebesar 1,525% dalam memprediksi kelayakan penerima beasiswa di Universitas Anwar Medika.

#### UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada Universitas Anwar Medika yang telah mengizinkan dan bersedia dalam mendukung penelitian ini serta berbagai pihak yang telah membantu melancarkan penelitian ini.

#### REFERENCES

- [1] A. Hidayat, S. Fatimah, dan D. N. Rosidin, "Challenges and Prospects of Islamic Education Institutions and Sustainability in The Digital Era," *Nazhruna J. Pendidik. Islam*, vol. 5, no. 2, hal. 351–366, 2021.
- [2] E. N. Qorimah dan S. Sutarna, "Studi Literatur: Media Augmented Reality (AR) Terhadap Hasil Belajar Kognitif," *J. Basicedu*, vol. 6, no. 2, hal. 2055–2060, 2022.
- [3] A. Hakim, "Faktor penyebab anak putus sekolah," *J. Pendidik.*, vol. 21, no. 2, hal. 122–132, 2020.
- [4] M. Hatta, "Implementasi Metode Simple Additive Weighting Pada Sistem DSS Seleksi Penerimaan Beasiswa Perguruan Tinggi," *J. Ilm. INTECH Inf. Technol. J. UMUS*, vol. 2, no. 01, hal. 31–40, 2020.
- [5] I. Kaunang, I. Abdul, A. Maruwae, F. Bumulo, dan A. Bahsoan, "PENGARUH PEMANFAATAN BEASISWA PROGRAM INDONESIA PINTAR (PIP) TERHADAP MOTIVASI BELAJAR SISWA," *Damhil Educ. J.*, vol. 4, no. 1, hal. 29–38, 2024.
- [6] S. Yunita dan V. N. Alaeida, "Penerapan Algoritma C4. 5 Untuk Prediksi Penerimaan Beasiswa di SD 4 Pelangsan," *ICIT J.*, vol. 8, no. 2, hal. 181–193, 2022.
- [7] A. Nurjana, A. P. Windarto, dan H. Qurniawan, "IMPLEMENTASI DATA MINING DALAM MEMREDIKSI PRESTASI SISWA DENGAN ALGORITMA C4. 5," *Smart EDU Bul. Educ.*, vol. 1, no. 4, hal. 171–180, 2022.
- [8] R. Amalia, "Penerapan data mining untuk memprediksi hasil kelulusan siswa menggunakan metode naïve bayes," *J. Inform. Dan Sist. Inf.*, vol. 6, no. 1, hal. 33–42, 2020.
- [9] B. Baskoro, S. Sriyanto, dan L. S. Rini, "Prediksi Penerima Beasiswa dengan Menggunakan Teknik Data Mining di Universitas Muhammadiyah Pringsewu," in *Prosiding Seminar Nasional Darmajaya*, 2021, vol. 1, hal. 87–94.
- [10] A. Byna dan M. Basit, "Penerapan Metode Adaboost Untuk Mengoptimasi Prediksi Penyakit Stroke Dengan Algoritma Naïve Bayes," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 9, no. 3, hal. 407–411, 2020.
- [11] I. Ubaedi dan Y. M. Djaksana, "Optimasi Algoritma C4. 5 Menggunakan Metode Forward Selection Dan Stratified Sampling Untuk Prediksi Kelayakan Kredit," *JSII (Jurnal Sist. Informasi)*, vol. 9, no. 1, hal. 17–26, 2022.
- [12] A. Azahari dan N. Nursobah, "Rekomendasi Penerimaan Beasiswa Yayasan Untuk Siswa Baru SMK TI Airlangga dengan Algoritma C4. 5," *J. MEDIA Inform. BUDIDARMA*, vol. 5, no. 2, hal. 609–614, 2021.
- [13] A. Surip, M. A. Pratama, I. Ali, A. R. Dikananda, dan A. I. Purnamasari, "Penerapan Machine Learning menggunakan algoritma C4. 5 berbasis PSO dalam Menganalisa Data Siswa Putus Sekolah," *Informatics Educ. Prof. J. Informatics*, vol. 5, no. 2, hal. 147–155, 2021.
- [14] N. Nadiyah, S. Soim, dan S. Sholihin, "Implementation of Decision Tree Algorithm Machine Learning in Detecting Covid-19 Virus Patients Using Public Datasets," *Indones. J. Artif. Intell. Data Min.*, vol. 5, no. 1, hal. 37, 2022.
- [15] D. A. R. Saragih, M. Safii, dan D. Suhendro, "Penerapan Data Mining Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pelayanan Sistem Informasi di Program Studi Sistem Informasi," *J. Inf. Syst. Res.*, vol. 2, no. 2, hal. 173–177, 2021.
- [16] V. S. Ginting, K. Kusriani, dan E. Taufiq, "Implementasi algoritma c4. 5 untuk memprediksi keterlambatan pembayaran sumbangan pembangunan pendidikan sekolah menggunakan python," *Inspir. J. Teknol. Inf. dan Komun.*, vol. 10, no. 1, hal. 36–44, 2020.
- [17] Y. C. Raya dan S. Arfida, "Penerapan Algoritma Decision Tree C4. 5 Untuk Penerimaan Beasiswa Kip Bagi Mahasiswa Baru Berbasis Website: Penerapan Algoritma Decision Tree C4. 5," *Tek. J. Ilm. Bid. Ilmu Rekayasa*, vol. 18, no. 2, hal. 377–388, 2024.
- [18] I. Yulianti, R. A. Saputra, M. S. Mardiyanto, dan A. Rahmawati, "Optimasi Akurasi Algoritma C4. 5 Berbasis Particle Swarm Optimization dengan Teknik Bagging pada Prediksi Penyakit Ginjal Kronis," *Techno. Com*, vol. 19, no. 4, hal. 411–421, 2020.
- [19] Y. T. Widayati, Y. Prihati, dan S. Widjaja, "Analisis Dan Komparasi Algoritma Na Ve Bayes Dan C4. 5 Untuk Klasifikasi Loyalitas Pelanggan Mnc Play Kota Semarang," *J. Transform.*, vol. 18, no. 2, hal. 161–172, 2021.
- [20] A. Ardiyansyah, R. Saadah, L. Lisnawanty, dan D. Purwaningtiyas, "Peningkatan Akurasi Metode C4. 5 Untuk Memprediksi Kelayakan Kredit Berbasis Stratified Sampling Dan Optimize Selection," *KLIK-KUMPULAN J. ILMU Komput.*, vol. 10, no. 2, hal. 239–249, 2023.