

Penerapan Feature Selection Pada Algoritma Decision Tree Untuk Menentukan Pola Rekomendasi Dini Konseling

Oman Somantri^{1,*}, Wildani Eko Nugroho², Abdul Rohma Supriyono¹

¹Program Studi Rekayasa Keamanan Siber, Politeknik Negeri Cilacap, Cilacap, Indonesia

²Program Studi Teknik Komputer, Politeknik Harapan Bersama, Tegal, Indonesia

Email: ^{1,*}oman_mantri@yahoo.com, ²wild4ani@gmail.com, ³a.rohman.sy@pnc.ac.id

Email Penulis Korespondensi: sandinoberutu@ukrimuniversity.ac.id

Submitted: 06/12/2022; Accepted: 20/12/2022; Published: 31/12/2022

Abstrak-Deteksi dini dalam memberikan rekomendasi untuk dilakukannya sebuah konseling bagi siswa sekolah sangat penting karena dengan cara ini maka sedini mungkin dapat menilai potensi, kepercayaan dan attitude siswa tersebut. Permasalahan yang muncul pada kasus ini adalah bagaimana mendeteksi dini seorang siswa agar dapat diketahui memerlukan bantuan konseling atau tidak dan dapat diketahui sejak dini untuk meminimalisir resiko kondisi psikis yang lebih jauh lagi. Pada artikel ini diusulkan sebuah model data mining dengan menggunakan decision tree untuk klasifikasi rekomendasi konseling bagi siswa. Selain itu, untuk dapat meningkatkan performance akurasi yang dihasilkan maka diusulkan metode feature selection dengan menggunakan forward selection dan algoritma genetika. Tahapan penelitian dilakukan dengan proses preprocessing data, penerapan algoritma, validasi data, dan optimasi model. Hasil eksperimen menunjukkan untuk tingkat akurasi terbaik dengan menggunakan model decision tree adalah sebesar 95,64%, meningkat menjadi 96,91%. setelah optimasi dengan menggunakan algoritma genetika.

Kata Kunci: Konseling; Decision Tree; Feature Selection; Forward Selection; Algoritma Genetika.

Abstract-Early detection in providing recommendations for student counseling is very important, therefore you can assess the student's potential, beliefs, and attitude as early as possible. The problem that arises in this case is how to detect a student early so that he or she needs counseling assistance or not so that it can be identified early to minimize the risk of further psychological conditions. This article proposes a data mining model using a decision tree to classify counseling recommendations for students. In addition, to improve the resulting accuracy performance, a feature selection method is proposed using forward selection and genetic algorithms. The stages of the research were carried out by pre-processing the data, implementing algorithms, validating data, and optimizing the model. The experimental results show that the best level of accuracy using the decision tree model is 95.64%. It increases to 96.91% after optimization using the genetic algorithm.

Keywords: Counseling; Decision Tree; Feature Selection; Forward Selection; Algoritma Genetika.

1. PENDAHULUAN

Perilaku seseorang sangat dipengaruhi oleh beberapa faktor yang dapat menjadikan orang tersebut dapat berubah, seperti masalah keadaan ekonomi, masalah kehidupan keluarga, kesehatan, masalah pribadi, dan lain-lain. Perubahan yang terjadi pada diri seseorang sangatlah berpengaruh terhadap kehidupan dirinya maupun keadaan disekitarnya seperti keluarga dan orang lain karena akan memberikan sebuah perubahan hidup. Tingkat stres seseorang sangatlah berbeda-beda, ini menjadikan dirinya dapat berubah-ubah dan bahkan apabila tidak dilakukan kontrol serta pengobatan secara psikis akan mengakibatkan buruknya perilaku serta keadaan kehidupan dan kesehatan mental dirinya terganggu [1], [2]. Siswa sekolah merupakan seseorang berada pada masa usia remaja dengan karakteristik yang khas, dimana usia ini siswa mempunyai adrenalin tantangan tinggi, rasa ingin tahu, dan rasa mencari jati diri tapi disisi lain *attitude* sangatlah diperlukan [3], [4]. Keadaan emosi dan kepribadian baik merupakan sebuah kebutuhan dan keharusan secara psikis bagi anak usia sekolah sehingga dapat meningkatkan kepercayaan dirinya [5]-[7].

Banyak diantara kita yang kurang begitu menyadari akan karakter dari siswa sekolah meskipun terkadang kita seperti sudah mengetahuinya. Beban psikis yang dialami oleh setiap siswa sangat berbeda dan ini tergantung dari keadaan situasi pribadi dan keadaan lingkungan disekitarnya. Deteksi penilaian terhadap kebutuhan dukungan kepribadian secara psikis siswa sekolah sangat diperlukan, hal ini adalah upaya untuk meminimalisir keadaan psikis siswa agar tidak terpuruk dan lebih buruk lagi, selain itu upaya meningkatkan kepercayaan diri siswa menjadi hal yang sangat penting [8]. Saat ini deteksi kebutuhan seseorang membutuhkan bantuan konseling atau tidak masih dilakukan hampir kebanyakan manual yaitu dilakukan secara langsung oleh seorang psikiater. Berkembangnya ilmu pengetahuan teknologi informasi memungkinkan dapat dibuatkannya model *data mining* yang dapat membantu untuk mengklasifikasikan data siswa dalam deteksi dini rekomendasi konseling. Pentingnya deteksi dini untuk mendapatkan rekomendasi konseling bagi siswa dapat menjadikannya sebagai upaya pencegahan dalam meminimalisir buruknya keadaan psikis siswa yang dikawatirkan dapat berpengaruh terhadap kehidupan pribadi yang dijalani seperti *attitude*, keadaan emosi, dan kepercayaan diri. Perkembangan teknologi khususnya dalam bidang kecerdasan buatan yang berbasis *machine learning* dengan menerapkan data mining saat ini sangat memungkinkan untuk digunakan pada permasalahan yang dihadapi khususnya pada sebuah klasifikasi data untuk deteksi dini rekomendasi konseling bagi siswa.

Data mining dengan kemampuannya yang dapat memungkinkan dalam menganalisis sebuah klasifikasi data [9], [10], maka dapat memberikan solusi dalam mendeteksi perlu tidaknya seseorang siswa mendapatkan

sebuah konseling dari seorang psikiater. Saat ini teknologi *machine learning* memberikan sebuah solusi yang dapat diterapkan pada model klasifikasi data, seperti diantaranya *neural network*, *support vector machine*, *deep learning*, *k-NN* dan metode lainnya dalam berbagai bidang penerapannya seperti pendidikan dan kesehatan [11]. Salah satu metode yang saat ini dikembangkan dengan kelebihanannya yaitu metode *decision tree* (DT) dapat diterapkan pada model untuk klasifikasi deteksi dini rekomendasi konseling bagi siswa sekolah.

Beberapa penelitian terdahulu terkait dengan analisis dan identifikasi mengenai konseling siswa sekolah, seperti diantaranya adalah penelitian mengenai pengembangan dan validasi rubrik konseling perawatan dini (SCCR) untuk menilai keterampilan konseling siswa [12]. Penelitian lainnya adalah dilakukan analisis penilaian kinerja siswa pada sesi konseling [13]. Sedikit berbeda dengan penelitian lainnya dimana metode *linear mixed regression model* (LMM) dan metode *Jacobson and Truax* (JT) dalam penerapannya untuk konseling karir untuk mahasiswa [14]. DT untuk evaluasi kesehatan mental siswa sekolah menggunakan algoritma C.5.0 [15]. Penelitian selanjutnya adalah terkait dengan penelitian pendekatan data science menggunakan *data mining* untuk identifikasi kebutuhan konseling pada tahun pertama siswa sekolah [16]. *Decision tree* merupakan metode yang dikembangkan berbasis sebuah pohon keputusan yang dalam kerjanya mempunyai kemampuan untuk mengklasifikasikan data dan hampir kebanyakan adalah tipe data nominal. Sistem kerja model DT ini sangat dipengaruhi oleh tipe data dan jenis data serta jumlah variabel atribut yang digunakan sehingga memerlukan upaya yang tepat dapat proses penyiapan data sebelum diterapkan pada model.

Pada penelitian dalam artikel ini bagaimanapun model DT masih memerlukan sebuah upaya optimasi karena *performance* akurasi yang dihasilkan masih memerlukan peningkatan dan masih memiliki tingkat akurasi yang belum memuaskan. Selain itu, pada proses tahapan *praprocessing* data masih memiliki permasalahan pada proses pemilihan fitur masih dianggap belum optimal sehingga diperlukan upaya yang dapat mengatasi pada proses pemilihan fitur yang berpengaruh terhadap model sehingga tingkat akurasi model yang diusulkan mendapatkan nilai yang terbaik. Pada penelitian ini, untuk mengatasi permasalahan yang ada maka diusulkan sebuah optimasi menggunakan *feature selection* dengan harapan dapat meningkatkan akurasi yang sudah ada. Tujuan dari penelitian ini adalah mencari model terbaik untuk deteksi dini rekomendasi konseling bagi siswa sekolah dengan menggunakan algoritma *decision tree*, selain itu pada artikel ini diusulkan sebuah metode *feature selection* [17] menggunakan *forward selection* dan algoritma genetika (GA) untuk dapat meningkatkan tingkat akurasi dari model yang sudah ada.

2. METODOLOGI PENELITIAN

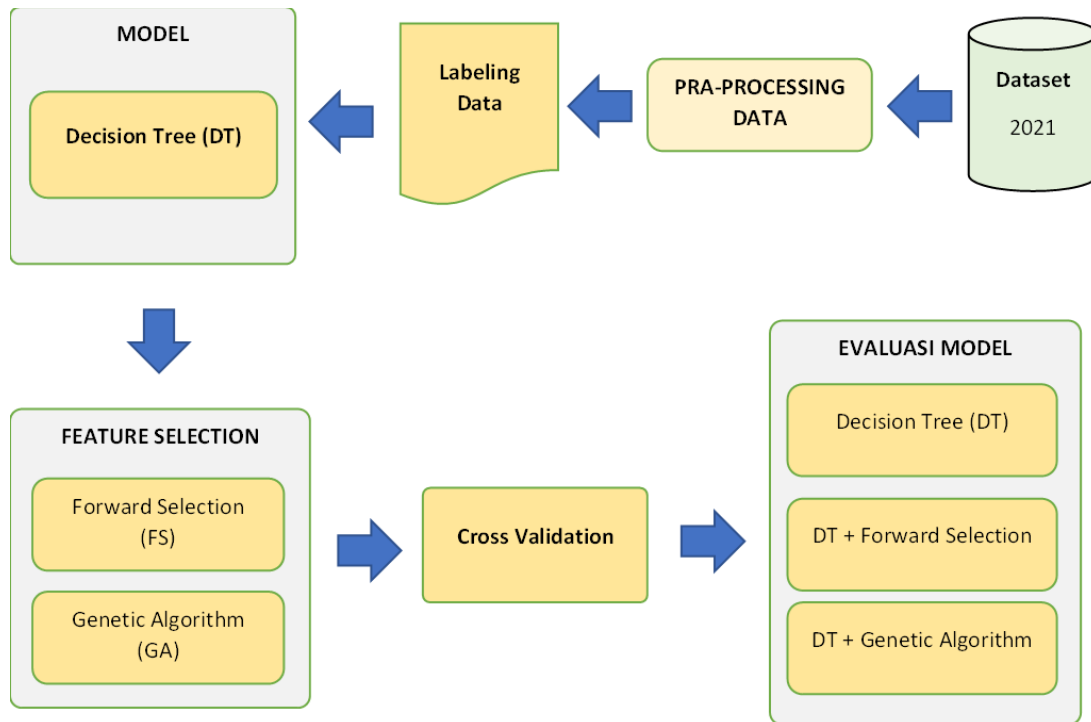
2.1 Data Penelitian & Pra-processing data

Data penelitian yang digunakan merupakan data yang diambil pada tahun 2021 yang berasal dari data identifikasi penilaian dini terhadap siswa sekolah yang dilakukan oleh psikiater sekolah. Jumlah data yang digunakan sebanyak 388 *record* yang terdiri dari dua label klasifikasi yaitu “Diharuskan Konseling” dan “Tidak Diharuskan Konseling”. Pada proses analisis pencarian model yang diterapkan pada penelitian ini menggunakan software Rapidminer studio 9.0 yang dijalankan pada sistem operasi windows 11 memiliki processor intel i7 dan memori 8GB. Untuk mendapatkan data yang sesuai dengan harapan, maka terlebih dahulu dilakukan *pra-processing* dimana data yang telah diperoleh dilakukan beberapa proses seperti *cleansing data*, *remove atribut*, *labelling data*.

2.2 Framework Penelitian

Pada proses pencarian model yang diinginkan sesuai dengan tujuan yang ditetapkan pada penelitian yang diusulkan ini, proses eksperimen dilakukan dengan melalui beberapa tahapan kegiatan diantaranya proses tahapan *pra-processing* data, penerapan metode dan algoritma, validasi data, serta tahapan terakhir adalah evaluasi model. Pada tahapan *pra-processing* data ini dilakukan proses penentuan data berdasarkan variabel yang didapatkan dimana proses ini ditentukan role tiap variabel yang terdiri dari ID, atribut, dan label. *Labelling* data di tentukan menjadi dua *class* label, dimana setiap proses *labelling* data sudah ditentukan oleh ahlinya sesuai dengan data yang sudah didapatkan.

Proses tahapan sebelum dilakukan analisis menggunakan model yang diusulkan, data terlebih dahulu dibagi menjadi dua yaitu *data training* dan *data testing*. Pembagian data ini terdiri dari 90% *data training* dan 10% *data testing* dari keseluruhan data yang digunakan pada eksperimen ini. Algoritma *decision tree* pada penelitian ini diusulkan dan diterapkan, dengan harapan dapat memberikan mode terbaik dengan tingkat akurasi yang terbaik dalam upaya deteksi dini rekomendasi konseling bagi siswa sekolah dan selanjutnya adalah dilakukan optimasi dengan menggunakan *feature selection* [18] yang menggunakan dua algoritma yaitu *forward selection* (FS) dan algoritma genetika (GA). Tahapan selanjutnya adalah proses validasi data, dimana pada tahapan ini menggunakan metode *cross validation* yang kemudian pada prosesnya dilakukan penentuan nilai *k-fold* terlebih dahulu. Tahapan terakhir dalam eksperimen ini adalah evaluasi model, dalam prosesnya dilakukan komparasi model anatara model *decision tree* klasik dengan model DT yang telah dioptimasi yaitu DT dengan menggunakan *forward selection* dan DT berbasis algoritma genetika. Penggambaran *framework* keseluruhan tahapan penelitian yang dilakukan pada penelitian dalam artikel ini diperlihatkan pada Gambar 1.



Gambar 1. Framework Penelitian yang diusulkan

Selain itu, berdasarkan pada Gambar 1 evaluasi model dilakukan untuk memberikan sebuah penilaian terhadap perbandingan nilai tingkat akurasi yang didapatkan antara model yang sudah ada dengan model yang diusulkan pada artikel ini. Proses evaluasi model yang dilakukan adalah dengan melakukan komparasi antara model menggunakan metode *decision tree* (DT), metode DT + *Forward Selection*, dan metode DT + *Genetic Algorithm*. Untuk setiap model yang diusulkan ada artikel ini untuk indikator penilaian *performance* menggunakan nilai *performance* yang terdiri dari tiga pengukuran yaitu *accuracy*, *recall* dan *precision*. Untuk pengukuran *performance* dari model yang diusulkan adalah dengan menggunakan metode *confusion matrix* [19], [20], seperti pada persamaan (1), persamaan (2) dan persamaan (3).

$$Precision = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP)+False\ Positive\ (FP)} \quad (1)$$

$$Recall = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP)+False\ Negative\ (FN)} \quad (2)$$

$$Accuracy = \frac{True\ Positive\ (TP)+True\ Negative\ (TN)}{True\ Positive\ (TP)+True\ Negative\ (TN)+False\ Positive\ (FP)+False\ Negative\ (FN)} \quad (3)$$

2.3 Decision Tree (DT)

Decision tree adalah sebuah metode machine learning dengan kemampuan digunakan untuk membuat sebuah keputusan yang divisualisasikan dalam bentuk pohon keputusan [21]. Untuk mendapatkan hasil dari model DT maka menggunakan persamaan (1) dan persamaan (2).

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (4)$$

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{S} * Entropy(S_i) \quad (5)$$

Pada persamaan (4) dimana S = himpunan kasus, n = jumlah partisi S , p_i = proporsi dari S_i terhadap S . Untuk persamaan (5) dimana A = atribut, n = jumlah atribut A , $|S_i|$ = jumlah kasus pada partisi ke- i , $|S|$ = jumlah kasus dalam S .

2.4 Forward Selection (FS) & Algoritma Genetika

Feature selection merupakan sebuah metode yang digunakan untuk optimalisasi model yang dalam prosesnya dilakukan pada saat praprocessing data. Proses FS cara kerjanya adalah mencoba untuk menemukan sebuah subhimpunan dari himpunan fitur model yang ada untuk dapat meningkatkan *performance* dari model yang diterapkan. Selain itu, FS dapat memungkinkan untuk dapat menghilangkan fitur yang tidak relevan dan berlebihan pada model yang digunakan sehingga model yang digunakan dapat optimal.

Forward selection (FS) merupakan salah satu metode untuk optimalisasi model dengan cara menemukan kombinasi peubah yang terbaik, selain itu pada metode ini memasukan sebuah variabel bebas yang memiliki

korelasi paling erat dengan variabel tak bebas lainnya [22]. Metode FS memiliki kelebihan yang dapat mengurangi kemungkinan terjadi adanya multikolinearitas pada model yang dihasilkan. *Forward selection* merupakan salah satu model terbaik dalam regresi dengan eliminasi variabel bebas yang pada model secara bertahap [23].

Algoritma genetika (GA) adalah sebuah metode optimasi yang yang dapat melakukan pencarian untuk menyelesaikan masalah yang berbasis pada mekanisme seleksi alam dan genetika [24]. Kelebihan yang dimiliki oleh GA sebagai algoritma optimasi, hal ini memungkinkan digunakan untuk digunakan untuk metode *feature selection* [25]. Secara garis besar persamaan yang digunakan pada algoritma GA adalah seperti pada persamaan (6).

$$m(H, t + 1) \geq m(H, t) \frac{\bar{f}(H,t)}{\bar{f}(t)} \left(1 - p_c \frac{\delta(H)}{\ell-1} - p_m o(H) \right) \tag{6}$$

Pada persamaan (6) dimana $m(H,t)$ adalah nomor dari instance pada H pada waktu t , $\bar{f}(H,t)$ adalah rata-rata nilai fitness dari nilai instance H pada waktu t , $\bar{f}(t)$ adalah rata-rata nilai fitness pada saat waktu t , $\delta(H)$ adalah mendefinisikan panjang schema H , p_c adalah nilai probabilitas *crossover*, p_m adalah nilai probabilitas *mutasi*, ℓ adalah panjang *string*, dan $o(H)$ adalah *order* dari *schema* H .

3. HASIL DAN PEMBAHASAN

3.1 Penerapan Decision Tree (DT)

Tahapan penerapan model yang diusulkan adalah model dengan menggunakan algoritma *Decision Tree* (DT). Model DT diimplementasikan dengan sebelumnya dilakukan *setting* dari parameter yang digunakan model ini. Nilai parameter yang diterapkan ada model DT mempunyai pengaruh yang signifikan terhadap *performance* nilai akurasi yang dihasilkan. Eksperimen awal dilakukan dengan menetapkan model *Decision Tree* (DT) dengan menggunakan *fold=10* dan metode *sampling stratified*, hasilnya dapat dilihat pada Tabel 1, Tabel 2 dan Tabel 3.

Tabel 1. Decision Tree menggunakan *stratified sampling*

Fold	Criterion	Accuracy	Precision	Recall
10	gain_ratio	95.37%	64.83%	65.00%
10	information_gain	95.37%	64.83%	65.00%
10	gini_index	95.37%	64.83%	65.00%
10	accuracy	94.60%	45.00%	45.00%

Tabel 2. Decision Tree menggunakan *shuffled sampling*

Fold	Criterion	Accuracy	Precision	Recall
10	gain_ratio	94.06%	45.83%	52.38%
10	information_gain	95.63%	57.69%	71.43%
10	gini_index	95.10%	54.17%	61.90%
10	accuracy	94.57%	48.33%	42.86%

Tabel 3. Decision Tree menggunakan *linear sampling*

Fold	Criterion	Accuracy	Precision	Recall
10	gain_ratio	92.82%	26.67%	19.05%
10	information_gain	93.59%	35.71%	23.81%
10	gini_index	93.33%	33.33%	23.81%
10	accuracy	93.33%	30.77%	19.05%

Pada Tabel 2 menunjukkan model decision tree dengan tingkat akurasi paling tinggi yaitu sebesar 95,63%. Model yang diperoleh merupakan model yang memiliki *performance* lebih baik dibandingkan dengan model lainnya yang memiliki nilai *precision* 57,69% dan *recall* 71,43%. Eksperimen lanjutan yang dilakukan adalah dengan menetapkan nilai *fold=8*, dan menetapkan parameter *criterion* sesuai dengan tipe yang digunakan, menghasilkan nilai *performance model* seperti pada Tabel 4. Hasil eksperimen menunjukkan nilai *performance* dengan menggunakan *fold=8* dan *criterion=gain_ratio* memiliki tingkat akurasi paling tinggi sebesar 95,64%. Model lain dihasilkan dengan menggunakan DT, dimana parameter yang ditetapkan adalah dengan menggunakan *fold=6*, hasil *performance* diperlihatkan pada Tabel 5.

Tabel 4. Decision Tree menggunakan *fold=8*

No.model	Sampling	Criterion	Accuracy	Precision	Recall
1	statified	gain_ratio	95.64%	60.42%	60.42%
2	statified	information_gain	95.64%	60.42%	60.42%
3	statified	gini_index	95.12%	60.42%	52.08%
4	statified	accuracy	94.60%	52.08%	45.83%

No.model	Sampling	Criterion	Accuracy	Precision	Recall
5	shuffled	gain_ratio	95.11%	46.88%	57.14%
6	shuffled	information_gain	95.62%	57.69%	71.43%
7	shuffled	gini_index	95.61%	59.09%	61.90%
8	shuffled	accuracy	94.34%	46.25%	47.62%
9	linear	gain_ratio	93.81%	41.18%	33.33%
10	linear	information_gain	93.81%	42.11%	38.10%
11	linear	gini_index	93.81%	41.18%	33.33%
12	linear	accuracy	93.82%	40.00%	28.57%

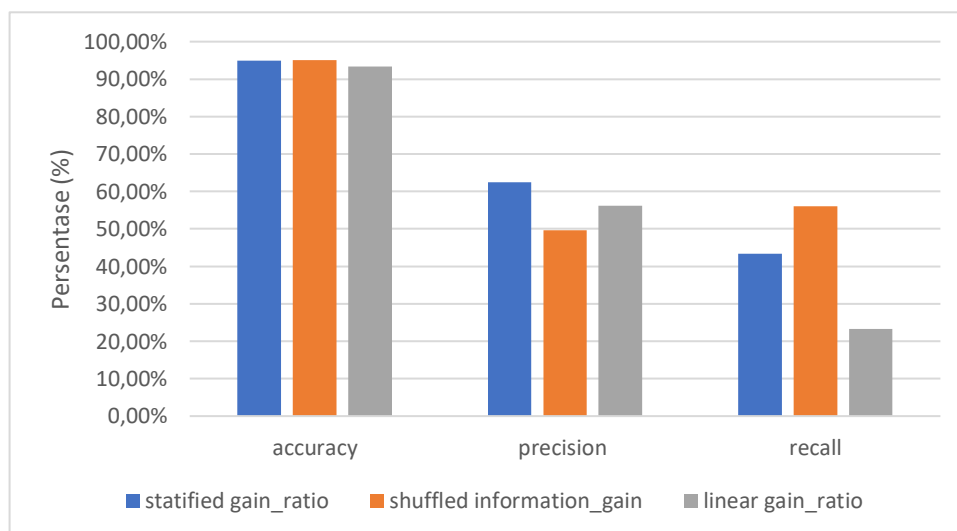
Tabel 5. Decision Tree menggunakan fold=6

No.model	Sampling	criterion	accuracy	precision	recall
1	statified	gain_ratio	93.07%	28.57%	19.05%
2	statified	information_gain	94.32%	57.50%	47.22%
3	statified	gini_index	94.32%	57.50%	47.22%
4	statified	accuracy	94.07%	52.54%	43.06%
5	shuffled	gain_ratio	93.56%	44.44%	31.55%
6	shuffled	information_gain	94.60%	50.00%	50.99%
7	shuffled	gini_index	94.33%	46.67%	34.33%
8	shuffled	accuracy	93.05%	31.25%	31.94%
9	linear	gain_ratio	94.06%	54.72%	47.22%
10	linear	information_gain	93.07%	31.25%	23.81%
11	linear	gini_index	93.07%	28.57%	19.05%
12	linear	accuracy	93.33%	30.77%	19.05%

Berdasarkan pada Tabel 5 nilai akurasi tertinggi adalah sebesar 96,60% dengan menggunakan parameter *criterion = information_gain*. Nilai *performance* terendah adalah sebesar 93,07% menggunakan parameter *fold=6*, dan *criterion=accuracy* sehingga dapat dilihat nilai akurasi ini sedikit jauh berbeda dengan yang lainnya. Upaya lain yang dilakukan adalah dengan melakukan perubahan nilai parameter *fold* pada model yang diusulkan dengan menggunakan *fold=4*. Hasil eksperimen menunjukkan pada Tabel 6 dan gambar 2 model dengan menggunakan *sampling=shuffled* dan *information_gain* memiliki akurasi tertinggi yaitu sebesar 95,10%.

Tabel 6. Decision Tree menggunakan fold=4

Sampling	criterion	accuracy	precision	recall
statified	gain_ratio	94,85%	62,50%	43,33%
shuffled	information_gain	95,10%	49,58%	56,07%
linear	gain_ratio	93,30%	56,25%	23,30%



Gambar 2. Performance model DT menggunakan fold=4

Pada Gambar 2 terlihat untuk nilai *performance* dengan menggunakan parameter metode *gain_ratio* pada model DT yang menerapkan nilai *fold=4* lebih baik dibandingkan dengan metode *information_gain*. Untuk tingkat akurasi terendah adalah sebesar 93,30%, pada model ini menggunakan metode *sampling=linear* dan *criterion=gain_ratio*.

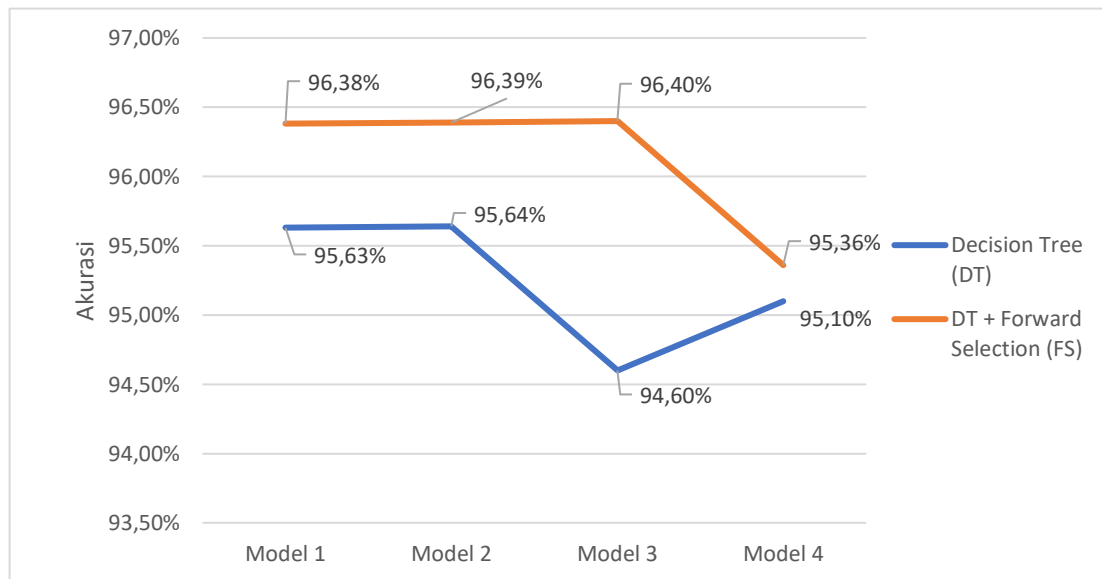
3.2 Optimasi Model Froward Selection

Optimalisasi yang pertama dilakukan untuk meningkatkan tingkat akurasi model pada Decision Tree yang telah diperoleh sebelumnya adalah dengan menerapkan metode forward selection. Pada model ini dilakukan optimasi dengan menetapkan parameter DT terbaik yang telah diperoleh. Percobaan yang dilakukan adalah dengan menetapkan parameter Fold yang berbeda yaitu 10, 8, 6 dan 4. Hasil eksperimen untuk optimasi FS+DT secara lengkap diperlihatkan oleh Tabel 7.

Tabel 7. Penerapan Forward Selection pada Decision Tree (DT+FS)

Fold	criterion	accuracy	precision	recall
10	information_gain	96,38%	68,42%	61,90%
8	gain_ratio	96,39%	68,75%	66,67%
6	information_gain	96,40%	68,42%	59,72%
4	information_gain	95,36%	62,50%	43,33%

Pada Tabel 7 memperlihatkan tingkat akurasi klasifikasi yang diperoleh rata-rata adalah sebesar 96%, dan ini menggunakan parameter criterion yang berbeda yaitu *information_gain* dan *gain_ratio*. Model dengan tingkat kurasi tertinggi pada optimasi FS ini adalah sebesar 96,40% dengan menggunakan *fold=6* dan *criterion=information_gain*. Pada hasil percobaan ini tingkat akurasi terkecil didapatkan sebesar 95,36% dengan nilai *precision* sebesar 62,5% dan *recall* 43,33%. Apabila melihat dari hasil keseluruhan maka hampir semua model yang didapatkan diatas 95%, ini diakibatkan optimasi yang dilakukan merupakan model terbaik *decision tree* yang telah didapatkan sebelumnya sehingga nilai akurasi tersebut sudah sedikit lebih baik.

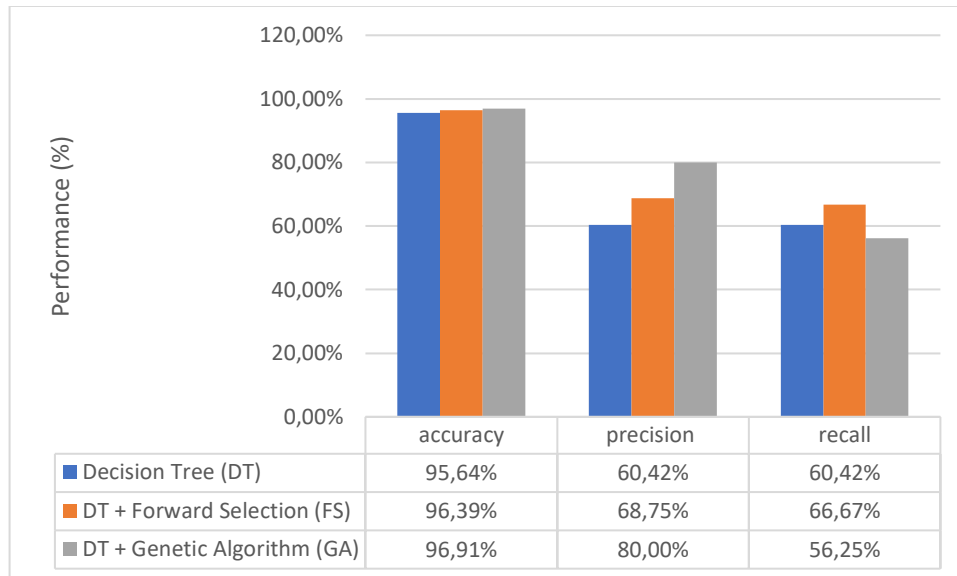


Gambar 3. Evaluasi performance model

Pada Gambar 3 merupakan grafik penggambaran dari perbandingan nilai *performance* model yang sudah dilakukan optimasi *forward selection* dengan model yang belum dilakukan optimasi atau dengan kata lain hanya dengan menggunakan decision tree saja. Berdasarkan hasil yang didapatkan hampir semua model yang dilakukan optimasi seluruhnya mengalami peningkatan nilai akurasi. Peningkatan yang signifikan adalah terjadi pada pengujian optimasi model 3, dimana peningkatannya menghasilkan model terbaik dengan tingkat akurasi 96,40% yang semula 94,60%. Untuk model yang mendapatkan akurasi terkecil dihasilkan pada model 4, dimana akurasi yang dihasilkan hanya sebesar 95,36% yang sebelumnya adalah 95,10%. Apabila kita melihat dari hasil yang diperlihatkan pada Gambar 3, perubahan peningkatan akurasi ini dipengaruhi oleh jumlah nilai fold dan jenis criterion yang digunakan sehingga tidak semua model yang dihasilkan merupakan model yang dapat meningkatkan akurasi.

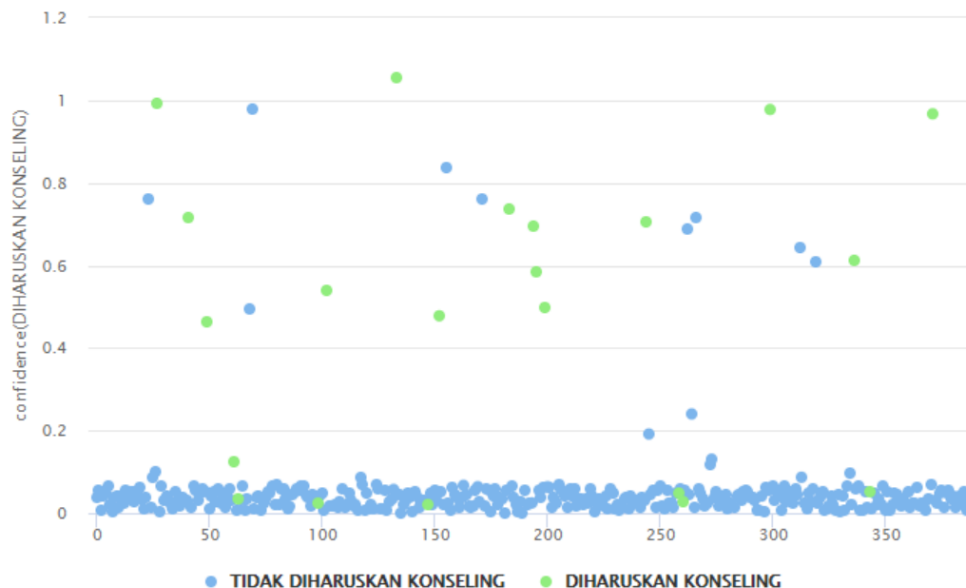
3.3 Penerapan Algoritma Genetika (GA)

Upaya lain yang dilakukan untuk meningkatkan akurasi adalah dengan menggunakan algoritma genetika (GA). Penerapan GA pada model decision tree yang digunakan adalah untuk melakukan pemilihan fitur terbaik pada model sehingga tingkat akurasi yang dihasilkan model pada decision tree mendapatkan performance yang lebih baik. Pada model DT+GA menggunakan parameter *population=5*, *selection_scheme=tournament*, dan *crossover=uniform*.



Gambar 4. Evaluasi performance model

Pada Gambar 4, berdasarkan hasil eksperimen yang telah dilakukan dengan menggunakan *decision tree*, menghasilkan tingkat akurasi yang masih dapat ditingkatkan. Tingkat akurasi tertinggi yang dihasilkan dari model DT klasik adalah sebesar 95,64% yang menggunakan parameter *sampling stratified*, *criterion=gain_ratio*, dan *fold=8*. Upaya peningkatan *performance* model yang diusulkan pada artikel ini adalah dengan menggunakan *feature selection*. Metode yang diusulkan adalah dengan menggunakan metode *forward selection* dan algoritma genetika (GA).



Gambar 5. Visualisasi dataset pada proses validasi

Hasil eksperimen memperlihatkan tingkat akurasi yang dihasilkan pada saat menerapkan metode *forward selection* adalah menghasilkan *performance* tingkat akurasi sebesar 96,39%. Parameter yang ditetapkan pada model ini adalah dengan menggunakan *stratified sampling* dan *criterion gain_ratio*. Sedikit berbeda dengan menggunakan algoritma genetika, menghasilkan nilai akurasi sebesar 96,91%, nilai *performance* ini lebih tinggi dibandingkan dengan yang lainnya. Visualisasi data pada model yang diusulkan pada saat validasi seperti diperlihatkan pada Gambar 5.

4. KESIMPULAN

Upaya peningkatan akurasi pada model untuk klasifikasi deteksi dini rekomendasi konseling bagi siswa sekolah dengan menggunakan algoritma *decision tree* (DT) telah didapatkan. Metode *feature selection* merupakan upaya yang dilakukan untuk mengatasi permasalahan masih rendahkan *performance* dari model DT yang dihasilkan.

Peningkatan akurasi telah dihasilkan dengan metode FS terbaik yaitu dengan menggunakan algoritma genetika yang menghasilkan tingkat akurasi terbaik sebesar 96,91%. Pada penelitian selanjutnya model DT masih dapat diupayakan untuk dioptimalkan menjadi lebih baik lagi sehingga dapat menghasilkan akurasi yang lebih baik seperti dapat dilakukan dengan menerapkan metode feature weight yang menggunakan berbagai algoritma di dalamnya. Upaya *pra-processing* data pada penelitian ini juga sangat berpengaruh sehingga pada proses ini harus benar-benar sudah tepat.

REFERENCES

- [1] C. A. W. Morris, K. L. Wester, C. T. Jones, and S. Fantahun, "School Counselors and Unified Educator–Counselor Identity: A Data-Informed Approach to Suicide Prevention," *Prof. Sch. Couns.*, vol. 24, no. 1_part_3, p. 2156759X2110119, Jan. 2021, doi: 10.1177/2156759X211011909.
- [2] G. Ping, "Application of Decision Tree Algorithm in Mental Health Evaluation," 2022, pp. 524–529.
- [3] O. W. A. Wilson, C. M. Bopp, Z. Papalia, M. Duffey, and M. Bopp, "College Students' Experiences and Attitudes Toward Physical Activity Counseling," *J. Nurse Pract.*, vol. 16, no. 8, pp. 623–628, Sep. 2020, doi: 10.1016/j.nurpra.2020.06.006.
- [4] J. M. Faro *et al.*, "U.S. medical students personal health behaviors, attitudes and perceived skills towards weight management counseling," *Prev. Med. Reports*, vol. 27, p. 101814, Jun. 2022, doi: 10.1016/j.pmedr.2022.101814.
- [5] C. Carter, J. Harnett, I. Krass, and I. Gelissen, "Attitudes, behaviours, and self-reported confidence of Australian pharmacy students and interns towards nutritional counselling," *Curr. Pharm. Teach. Learn.*, vol. 14, no. 11, pp. 1411–1419, Nov. 2022, doi: 10.1016/j.cptl.2022.09.028.
- [6] V. Kuryluk, J. McAuley, and M. Maguire, "Naloxone counseling: Confidence and attitudes of student pharmacists after a volunteer syringe exchange experience," *Curr. Pharm. Teach. Learn.*, vol. 12, no. 4, pp. 429–433, Apr. 2020, doi: 10.1016/j.cptl.2019.12.027.
- [7] H. Lim and J. C. Barner, "Impact of a pilot workshop on student pharmacists' confidence and comfort in counseling patients at risk for maternal mortality," *Curr. Pharm. Teach. Learn.*, vol. 14, no. 1, pp. 71–82, Jan. 2022, doi: 10.1016/j.cptl.2021.12.001.
- [8] S. Chircu, "Career Counseling Needs for Students – A Comparative Study," *Procedia - Soc. Behav. Sci.*, vol. 127, pp. 549–553, Apr. 2014, doi: 10.1016/j.sbspro.2014.03.308.
- [9] A. Naik and L. Samant, "Correlation Review of Classification Algorithm Using Data Mining Tool: WEKA, Rapidminer, Tanagra, Orange and Knime," *Procedia Comput. Sci.*, vol. 85, pp. 662–668, 2016, doi: 10.1016/j.procs.2016.05.251.
- [10] M. Gordan *et al.*, "State-of-the-art review on advancements of data mining in structural health monitoring," *Measurement*, vol. 193, p. 110939, Apr. 2022, doi: 10.1016/j.measurement.2022.110939.
- [11] O. Somantri, "An Optimize Weights Naïve Bayes Model for Early Detection of Diabetes," *Telematika*, vol. 15, no. 1, scholar.archive.org, 2022, doi: 10.35671/telematika.v15i1.1307.
- [12] A. M. H. Chen, S. Cailor, T. Franz, N. Fox, P. Thornton, and M. Norfolk, "Development and validation of the self-care counseling rubric (SCCR) to assess student self-care counseling skills," *Curr. Pharm. Teach. Learn.*, vol. 11, no. 8, pp. 774–781, Aug. 2019, doi: 10.1016/j.cptl.2019.04.006.
- [13] K. B. Garza, N. S. Hohmann, J. Kavookjian, and E. L. Kleppinger, "Assessment of student performance on a mock new prescription counseling session and an objective structured clinical examination across five years," *Curr. Pharm. Teach. Learn.*, vol. 12, no. 9, pp. 1046–1055, Sep. 2020, doi: 10.1016/j.cptl.2020.04.018.
- [14] L. Carvalho, L. Mourão, and C. Freitas, "Career counseling for college students: Assessment of an online and group intervention," *J. Vocat. Behav.*, p. 103820, Nov. 2022, doi: 10.1016/j.jvb.2022.103820.
- [15] M. Li and H. Yang, "Decision Tree Algorithm in College Students' Health Evaluation System," 2021, pp. 705–710.
- [16] A. Bottcher, V. Thurner, T. Hafner, and J. Hertle, "A Data Science-based Approach for Identifying Counseling Needs in first-year Students," in *2021 IEEE Global Engineering Education Conference (EDUCON)*, Apr. 2021, pp. 420–429, doi: 10.1109/EDUCON46332.2021.9454042.
- [17] Y. Xue, H. Zhu, J. Liang, and A. Słowik, "Adaptive crossover operator based multi-objective binary genetic algorithm for feature selection in classification," *Knowledge-Based Syst.*, vol. 227, p. 107218, Sep. 2021, doi: 10.1016/j.knosys.2021.107218.
- [18] P. Agrawal, H. F. Abutarboush, T. Ganesh, and A. W. Mohamed, "Metaheuristic Algorithms on Feature Selection: A Survey of One Decade of Research (2009-2019)," *IEEE Access*, vol. 9, pp. 26766–26791, 2021, doi: 10.1109/ACCESS.2021.3056407.
- [19] I. Markoulidakis, I. Rallis, I. Georgoulas, G. Kopsiaftis, A. Doulamis, and N. Doulamis, "Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem," *Technologies*, vol. 9, no. 4, p. 81, Nov. 2021, doi: 10.3390/technologies9040081.
- [20] C. S. Hong, "Confusion plot for the confusion matrix," *J. Korean Data Inf. Sci. Soc.*, vol. 32, no. 2, pp. 427–437, Mar. 2021, doi: 10.7465/jkdi.2021.32.2.427.
- [21] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *J. Appl. Sci. Technol. Trends*, vol. 2, no. 01, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.
- [22] V. Kotu and B. Deshpande, "Feature Selection," in *Data Science*, Elsevier, 2019, pp. 467–490.
- [23] A. Meyer-Baese and V. Schmid, "Feature Selection and Extraction," in *Pattern Recognition and Signal Analysis in Medical Imaging*, Elsevier, 2014, pp. 21–69.
- [24] E. Wirsansky, *Hands-on genetic algorithms with Python : applying genetic algorithms to solve real-world deep learning and artificial intelligence problems*. Packt Publishing Ltd, 2020.
- [25] F. Buontempo, *Genetic Algorithms and Machine Learning for Programmers: Create AI Models and Evolve Solutions (Pragmatic Programmers) 1st Edition*. 2019.