

Optimasi Algoritma Genetika Dalam Memprediksi Minat Baca Siswa Pada Perpustakaan SMK Negeri 1 Gantar Dengan Metode Decision Tree

Lina Yulita^{1,*}, Aswan S. Sunge², Nisa Nurhidayanti³

^{1,2}Teknik Informatika, Universitas Pelita Bangsa

¹yulita.lina@gmail.com

ABSTRAK

Siswa merupakan salah satu bagian dalam dunia pendidikan yang tidak bisa terlepas dari kegiatan membaca. Setiap sekolah tentunya berupaya untuk menyediakan fasilitas membaca seperti perpustakaan sekolah, sama halnya dengan perpustakaan yang dimiliki oleh SMK Negeri 1 Gantar bertujuan untuk dapat menumbuhkan dan meningkatkan minat baca siswa terhadap koleksi buku yang terdapat di perpustakaan sekolah. Tetapi saat ini perpustakaan SMK Negeri 1 Gantar cenderung minim pengunjung, hal tersebut bisa saja terjadi karena kurangnya kesadaran siswa terhadap pentingnya membaca buku atau ada faktor lain yang dapat mempengaruhi seperti pelayanan, jenis buku, kenyamanan, koleksi buku dan lain sebagainya. Maka dilakukan sebuah penelitian yang bertujuan untuk mengetahui seberapa besar minat baca siswa terhadap perpustakaan SMK Negeri 1 Gantar menggunakan optimasi algoritma genetika dengan metode decision tree. Data yang digunakan pada penelitian ini adalah data pengunjung yang dimiliki oleh perpustakaan SMK Negeri 1 Gantar sebanyak 290 data, proses pengujian metode menggunakan RapidMiner 9.2. Berdasarkan hasil pengujian terhadap penelitian dalam memprediksi minat baca siswa pada perpustakaan SMK Negeri 1 Gantar, diperoleh hasil dari algoritma C4.5 atau decision tree yaitu accuracy sebesar 84.48% dan setelah dioptimasi menggunakan algoritma genetika mengalami kenaikan accuracy sebesar 12.07% sehingga nilai accuracy yang diperoleh dari optimasi sebesar 96.55%. Maka dapat disimpulkan bahwa teknik optimasi algoritma genetika di nilai berhasil untuk meningkatkan accuracy dari algoritma C4.5 atau decision tree dalam memprediksi minat baca siswa pada perpustakaan SMK Negeri 1 Gantar.

Kata kunci: Minat baca, perpustakaan, algoritma C4.5, algoritma genetika.

ABSTRACT

Students are one part of the world of education that cannot be separated from reading activities. Each school certainly seeks to provide reading facilities such as school libraries, as well as libraries owned by SMK Negeri 1 Gantar aim to be able to foster and increase students' interest in reading books in the school library. But at the moment the library of SMK Negeri 1 Gantar tends to be minimal in number of visitors, this could be due to the lack of student awareness of the importance of reading books or there are other factors that can influence such as service, type of book, comfort, collection of books and so on. Then conducted a study that aims to find out how much interest in reading students of SMK Negeri 1 Gantar library using genetic algorithm optimization with the decision tree method. The data used in this study are visitor data owned by the library of SMK Negeri 1 Gantar as many as 290 data, the process of testing the method using RapidMiner 9.2. Based on the results of testing on research in predicting students' reading interest in the library of SMK Negeri 1 Gantar, the results obtained from the C4.5 algorithm or decision tree are accuracy by 84.48% and after being optimized using genetic algorithms the accuracy increases by 12.07% so that the accuracy value obtained from optimization of 96.55%. Then it can be concluded that the genetic algorithm optimization technique in value succeeded in increasing the accuracy of the C4.5 algorithm or decision tree in predicting students' interest in reading at the SMK Negeri 1 Gantar library.

Keyword: Reading interest, library, C4.5 algorithm, genetic algorithm.

PENDAHULUAN

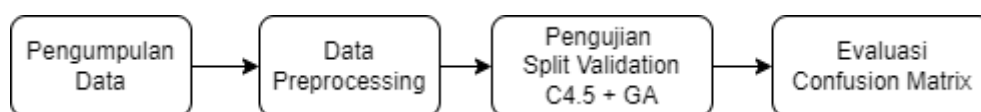
Membaca merupakan salah satu aspek keterampilan berbahasa yang bersifat aktif reseptif. Media yang digunakan dalam membaca berupa media bahas tulis. Membaca adalah melihat serta memahami isi dari apa yang tertulis, baik mengeja atau melafalkan apa yang ditulis (Yatun, 2015). Kegiatan membaca terlepas dari minat baca yaitu adanya suatu keinginan yang kuat untuk mengetahui isi dari suatu bacaan yang dilakukan secara berkelanjutan. Kegiatan membaca sering disosialisasikan dalam dunia pendidikan karena membaca merupakan suatu kegiatan yang sangat penting sebagai penunjang kegiatan belajar mengajar di sekolah (Utami, Wibowo, & Susanti, 2018). Siswa merupakan salah satu komponen dalam dunia pendidikan yang tidak bisa terlepas dari kegiatan membaca. Kegiatan membaca adalah kewajiban yang harus dilakukan oleh siswa, tetapi siswa yang melakukan kegiatan membaca belum bisa dikategorikan sebagai siswa yang mempunyai minat baca. Karena tidak sedikit siswa yang melakukan kegiatan membaca hanya pada saat kegiatan belajar di kelas saja. Dan ini membuat pihak sekolah harus menyediakan fasilitas membaca seperti perpustakaan sekolah agar siswa dapat melakukan kegiatan membaca bukan hanya pada kegiatan belajar di kelas saja tetapi bisa memanfaatkan waktu istirahat untuk melakukan kegiatan membaca di perpustakaan sekolah secara berkelanjutan. Jika kegiatan membaca bisa dilakukan secara berkelanjutan oleh siswa, diharapkan bisa menumbuhkan minat baca siswa pada perpustakaan sekolah.

Perpustakaan merupakan unit kerja dari suatu badan atau lembaga tertentu yang mengelola bahan pustaka berupa buku atau sumber bacaan lainnya seperti buku fiksi dan non-fiksi yang dikelola oleh seorang pustakawan sebagai penunjang pelayanan di dalam perpustakaan. Ada beberapa komponen yang harus diperhatikan dan ditingkatkan dalam sebuah perpustakaan yaitu seperti koleksi buku, kualitas buku, jenis buku, kenyamanan ruangan dan pelayanan yang diberikan oleh pustakawan terhadap anggota perpustakaan lainnya. Karena beberapa komponen tersebut bisa saja mempengaruhi daya minat siswa dalam melakukan kunjungan atau proses kegiatan pada perpustakaan sekolah.

Tujuan yang ingin dicapai pada penelitian ini yaitu diharapkan dapat menghasilkan sebuah informasi yang bermanfaat tentang tingkat prediksi minat baca siswa pada perpustakaan SMK Negeri 1 Gantar menggunakan pengujian metode optimasi algoritma genetika dengan algoritma C4.5 atau *decision tree*.

METODE PENELITIAN

Penelitian ini dilakukan pada perpustakaan SMK Negeri 1 Gantar yang beralamat di Jl. Raya Haurgeulis-Gantar KM. 06, Desa Gantar, Kecamatan Gantar, Kabupaten Indramayu, Provinsi Jawa Barat.



Gambar 1. Tahapan Penelitian

Pengumpulan Data

Metode pengumpulan data yang digunakan pada penelitian ini yaitu observasi langsung ke perpustakaan SMK Negeri 1 Gantar dan melakukan wawancara terhadap pustakawan untuk memperoleh data yang akan digunakan dalam penelitian ini. Sementara untuk jenis data terdapat 2 (dua) jenis yaitu data primer dan data sekunder. Data primer adalah data yang diperoleh secara langsung dari sumber objek penelitian tanpa perantara pihak lain. Sedangkan, data sekunder adalah data yang diperoleh secara tidak langsung dari objek penelitian atau melalui perantara pihak lain yang disebut *Kaggle* (Singestecia et al., 2018). Berikut adalah tampilan *dataset* yang berupa data primer karena diperoleh dari data perpustakaan SMK Negeri 1 Gantar sebanyak 290 data, terdapat 12 atribut yaitu tanggal kunjungan, nomor kunjungan, nama pengunjung, jenis kelamin, kelas, jurusan, kegiatan, jenis buku, koleksi buku, kualitas buku, kenyamanan, pelayanan yang ditunjukkan pada tabel 1.

Tabel 1. Data Perpustakaan

Tgl Kunjungan	No	Nama Pengunjung	Jenis Kelamin	Kelas	Jurusan	Kegiatan	Jenis Buku	Koleksi Buku	Kualitas Buku	Kenyamanan	Pelayanan
43563	1	Rihadatul Hasanah	Perempuan	X	RPL	Meminjam Buku	Nonfiksi	Cukup	Baik	Nyaman	Baik
43563	2	Fachmi Miftahillah	Laki-Laki	XI	TKRO	Membaca ditempat	Nonfiksi	Lengkap	Baik	Cukup	Cukup
43563	3	Yuliana Firdaus	Perempuan	XII	RPL	Membaca ditempat	Nonfiksi	Kurano	Baik	Nyaman	Cukup
43563	4	Andrian Baaskara	Laki-Laki	X	TKRO	Membaca ditempat	Nonfiksi	Kurana	Baik	Cukup	Baik
43563	5	Halimah	Perempuan	XI	RPL	Membaca ditempat	Fiksi	Kurano	Baik	Nyaman	Baik
43563	6	Aseo Satori	Laki-Laki	X	RPL	Membaca ditempat	Nonfiksi	Lengkap	Baik	Nyaman	Baik
43563	7	Yani Tresnawati	Perempuan	XI	RPL	Membaca ditempat	Nonfiksi	Lenoka	Baik	Cukup	Baik
43563	8	Nenena Sri Latun	Perempuan	XII	RPL	Membaca ditempat	Nonfiksi	Kurana	Kurana	Nyaman	Baik
43563	9	Oinda Putri	Perempuan	XI	TGB	Membaca ditempat	Nonfiksi	Lenoka	Kurano	Kurano	Baik
43563	10	Ado Gunawan	Laki-Laki	XII	TKRO	Membaca ditempat	Nonfiksi	Lengkap	Cukup	Kurana	Baik
43563	11	Anus Kusmana	Laki-Laki	XII	TKRO	Membaca ditempat	Nonfiksi	Lenoka	Baik	Nyaman	Baik

Pengolahan data awal

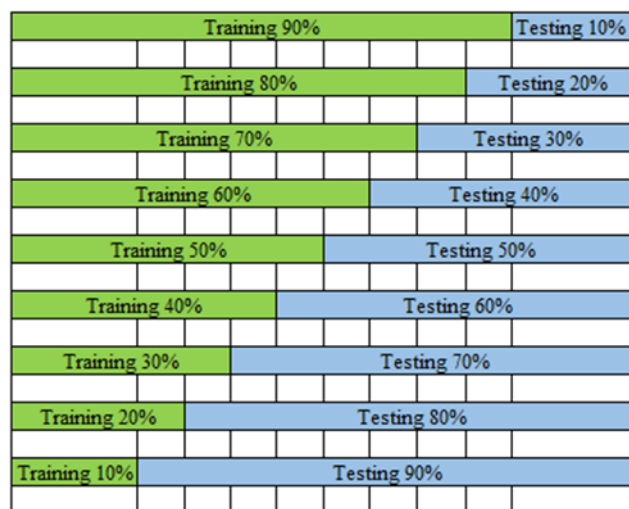
Data primer yang telah diperoleh dari data pengunjung yang dimiliki oleh perpustakaan SMK Negeri 1 Gantar, selanjutnya dilakukan proses *cleaning* data terlebih dahulu untuk melihat apakah ada data yang *missing value*, data hilang dan duplikasi data (Meilani & Asadulloh, 2015). Pada data perpustakaan tersebut memiliki atribut yang tidak terlalu diperlukan sehingga bisa dilakukan proses pembersihan atau dihilangkan seperti atribut tanggal kunjungan, nomor kunjungan dan nama pengunjung perpustakaan SMK Negeri 1 Gantar. Berikut adalah 10 atribut yang diusulkan atau digunakan pada penelitian ini yaitu jenis kelamin, kelas, jurusan, kegiatan, jenis buku, koleksi buku, kualitas buku, kenyamanan, pelayanan dan atribut prediksi hasil sebagai label target atau klasifikasi yang ditunjukkan pada tabel 2.

Tabel 2. Atribut Yang Digunakan

Atribut	Tipe Data	Nilai	Keterangan
Prediksi Hasil	Polynomial	Minat	Label Target
		Tidak < Minat	
Jenis Kelamin	Integer	2	Perempuan
		1	Laki-Laki
		3	XII
Kelas	Integer	2	XI
		1	X
Jurusan	Integer	1	Teknik Gambar Bangunan (TGB)
		2	Rekayasa Perangkat Lunak (RPL)
		3	Teknik Kendaraan Ringan Otomotif (TKRO)
Kegiatan	Integer	1	Membaca Ditempat
		2	Meminjam Buku
		3	Mengembalikan Buku
Jenis Buku	Integer	1	Fiksi
		2	Non Fiksi
Koleksi Buku	Integer	1	Kurang
		2	Cukup

Atribut	Tipe Data	Nilai	Keterangan
Kualitas Buku	Integer	3	Lengkap
		1	Kurang
		2	Cukup
		3	Baik
Kenyamanan	Integer	1	Kurang
		2	Cukup
		3	Nyaman
Pelayanan	Integer	1	Kurang
		2	Cukup
		3	Baik

Data primer yang sudah disiapkan untuk proses klasifikasi dibagi menjadi 2 (dua) bagian yang pertama untuk data *training* sebesar 80% dan data *testing* sebesar 20%. Proses pembagian data menjadi data *training* dan data *testing* menggunakan *split validation*. *Split validation* merupakan teknik validasi yang membagi data menjadi 2 (dua) bagian secara acak, sebagian untuk data *training* dan sebagian untuk data *testing* (Febriarini & Astuti, 2019). Dengan menggunakan *split validation* akan dilakukan proses pengujian terhadap penentuan data *training* berdasarkan *split ratio* yang sudah ditentukan sebelumnya, kemudian sisa dari *split ratio* data *training* akan dianggap sebagai data *testing*. Data *training* merupakan data yang akan digunakan dalam proses pelatihan (Raharjo & Mada, 2014), sedangkan data *testing* merupakan data yang belum pernah digunakan sebagai pelatihan dan berfungsi sebagai data pengujian kebenaran atau keakuratan hasil dari pelatihan (Galih, 2019). Berikut merupakan ilustrasi *split validation* yang ditunjukkan pada gambar 2.



Gambar 2. Split Validation

Algoritma C4.5

Algoritma C4.5 diperkenalkan oleh **Quinlan** (1993) yang mana algoritma C4.5 merupakan perkembangan dari algoritma *ID3*, algoritma C4.5 digunakan untuk pohon keputusan atau *decision tree* (Iriadi & Nuraeni, 2016). Pohon keputusan dinilai sebagai salah satu pendekatan yang sangat populer, dalam klasifikasi pada pohon keputusan terdiri dari *node* yang membentuk akar, *node* akar tidak memiliki masukan. *Node* lain yang tidak termasuk sebagai akar tetapi memiliki tepat satu masukan disebut *test node* atau *node internal*. Sedangkan, *node* yang lainnya disebut daun. Daun dapat mewakili sebuah nilai target yang sangat tepat dari salah satu *class* (Fatmawati, 2016)

Algoritma Genetika

Algoritma genetika (GA) merupakan sebuah metode data mining yang lebih efektif digunakan untuk proses pengenalan pola dan solusi masalah optimasi(Sugeha, Inkiriwang, & Pratas, 2019). Algoritma genetika biasanya digunakan pada proses seleksi alam, merupakan bagian dari kecerdasan buatan yang telah berkembang yaitu komputasi evolusioner. Algoritma genetika memiliki konsep yang hampir mirip dengan proses seleksi alam dimana yang lebih kuat akan menjadi pemenang. Pemenang terbaik merupakan hasil yang optimal dari proses genetika disebut *fitness*. Penyelesaian yang diperoleh dari algoritma genetika diimplementasikan pada populasi individu disebut kromosom. Nilai yang didapatkan dari setiap solusi dan setiap individu bergantung pada kromosom dan di evaluasi oleh fungsi *fitness* (Arsi & Somantri, 2018).

HASIL DAN PEMBAHASAN

Pengujian pada penelitian ini dilakukan dua tahap, yaitu pengujian algoritma c4.5 dan pengujian algoritma c4.5 dengan optimasi algoritma genetika untuk mengetahui perbandingan akurasi.

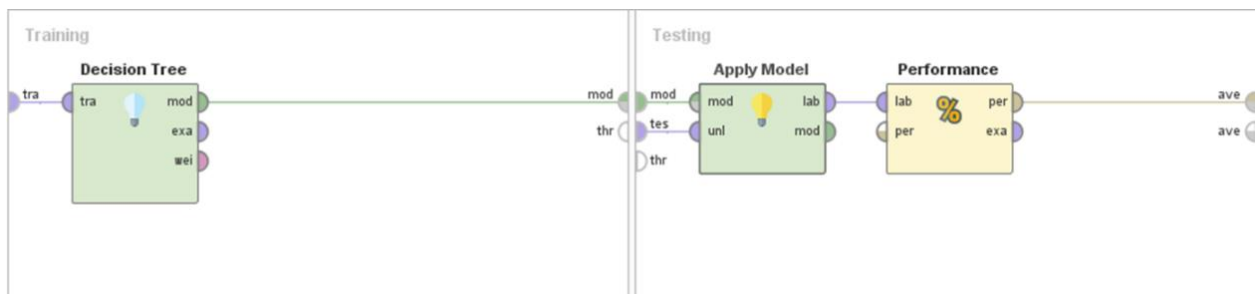
Proses awal yang dilakukan pada penelitian ini adalah mempersiapkan data yang telah dilakukan proses pengolahan data awal dimana prosesnya yaitu menghilangkan beberapa atribut yang dianggap tidak terlalu diperlukan dan kemudian ditransformasi ke dalam parameter berupa angka 1,2,3, yang diharapkan bisa membuat tampilan data menjadi lebih sederhana yang ditunjukkan pada tabel 3.

Tabel 3. Transformasi Data Perpustakaan

Jenis Kelamin	Kelas	Jurusan	Kegiatan	Jenis Buku	Koleksi Buku	Kualitas Buku	Kenyamanan	Pelayanan	PREDIKSI HASIL
2	1	2	2	2	2	3	3	3	Minat
1	2	1	1	2	3	3	2	2	Minat
2	3	2	1	2	1	3	3	2	TidakMinat
1	1	1	1	2	1	3	2	3	TidakMinat
2	2	2	1	1	1	3	3	3	Minat
1	1	2	1	2	3	3	3	3	TidakMinat
2	2	2	1	2	3	3	2	3	Minat
2	3	2	1	2	1	1	3	3	TidakMinat
2	2	3	3	2	3	1	1	3	TidakMinat
1	3	1	1	2	3	2	1	3	TidakMinat

Pengujian pertama Algoritma C4.5 atau *Decision Tree*

Proses pengujian terhadap penelitian ini menggunakan *RapidMiner 9.2*. Adapun operator yang digunakan pada pengujian algoritma C4.5 yaitu ditunjukkan pada Gambar 3.



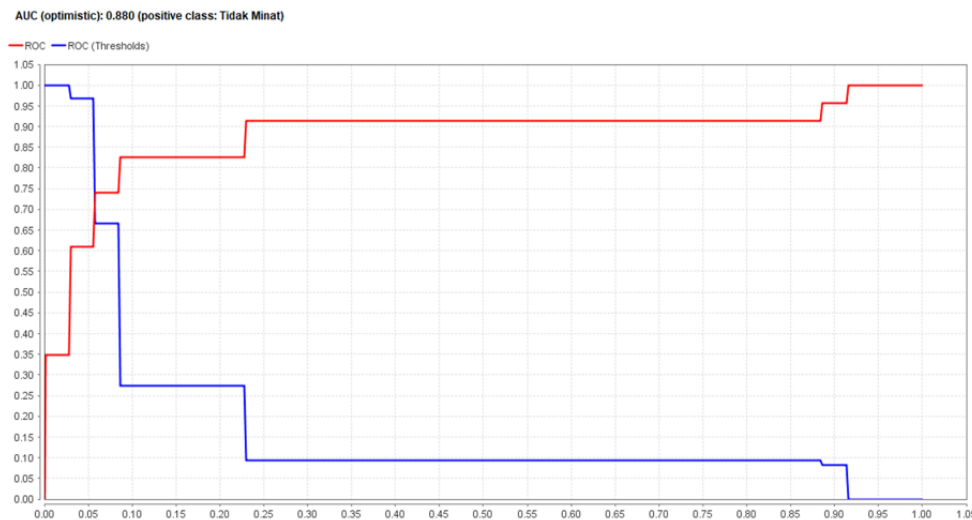
Gambar 3. Proses Pengujian C4.5

Dari proses yang sudah dijalankan pada beberapa tahapan operator diatas, maka diperoleh hasil pengujian sebagai berikut yang ditunjukkan pada tabel 4.

Tabel 4. Confusion Matrix Hasil Pengujian

	True Minat	True Tidak Minat	Class Precision
Pred. Minat	32	6	84.21%
Pred. Tidak Minat	3	17	85.00%
Class Recall	91.43%	73.91%	

Berdasarkan confusion matrix pada tabel x hasil accuracy yang diperoleh adalah 84.48%, nilai precision sebesar 85.00%, dan nilai recall sebesar 73.91%. Kurva ROC/AUC dapat dilihat pada gambar 4 pada pengujian algoritma C4.5 atau decision tree.



Gambar 4. Grafik ROC/AUC C4.5

Dari gambar x diatas menampilkan hasil nilai grafik ROC/AUC dari pengujian algoritma C4.5 atau *decision tree* sebesar 0.880 yang termasuk ke dalam kategori *good classification* karena berada diantara nilai 0.80 - 0.90.

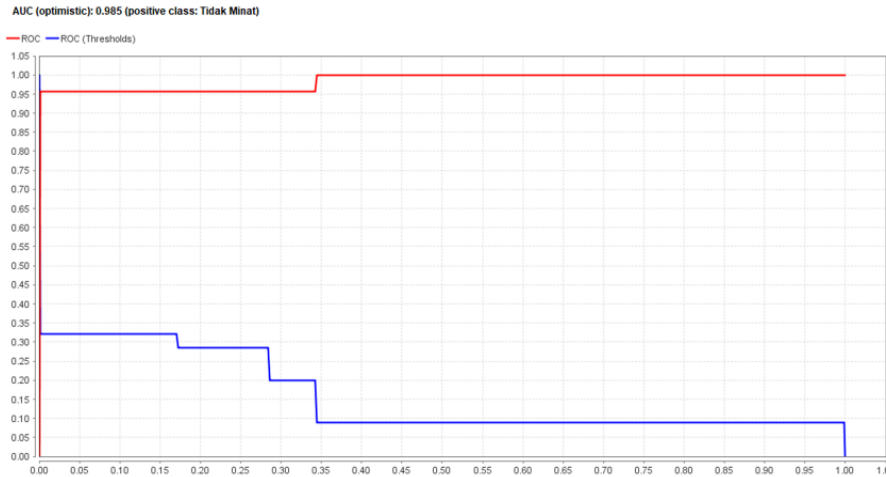
Pengujian Kedua Algoritma C4.5 Dengan Optimasi Algoritma Genetika

Pada pengujian kali ini menggunakan 2 (dua) algoritma yang digabungkan yaitu algoritma C4.5 atau *decision tree* yang di optimasi oleh algoritma genetika.

Tabel 5. Confusion Matrix Hasil Pengujian C4.5 + GA

	True Minat	True Tidak Minat	Class Precision
Pred. Minat	35	2	94.59%
Pred. Tidak Minat	0	21	100.00%
Class Recall	100.00%	91.30%	

Berdasarkan confusion matrix pada tabel 5 hasil accuracy yang diperoleh adalah 96.55%, nilai precision sebesar 100.00%, dan nilai recall sebesar 91.30%. Kurva ROC/AUC dapat dilihat pada gambar 5 pada pengujian algoritma C4.5 atau decision tree.

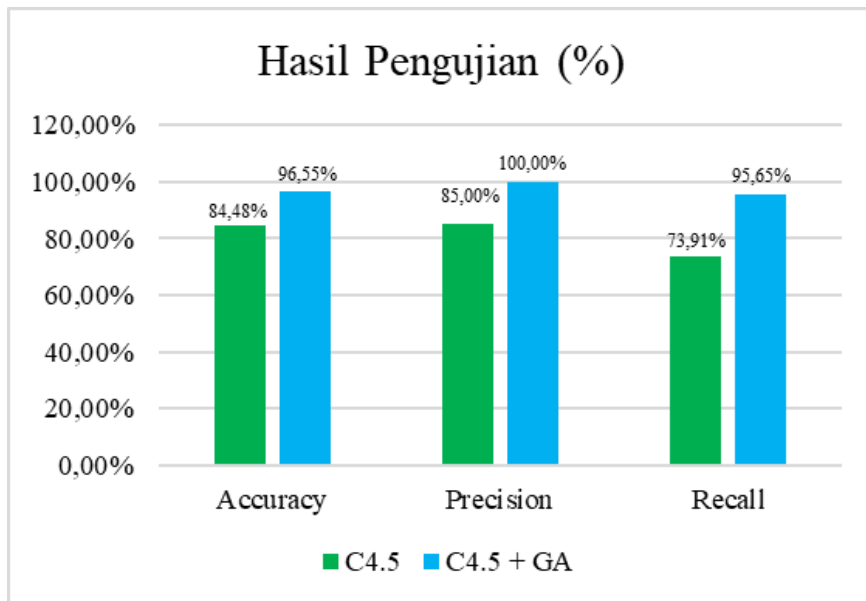


Gambar 5. Grafik ROC/AUC C4.5 + GA

Dari gambar 5 diatas menampilkan hasil nilai grafik ROC/AUC dari pengujian algoritma C4.5 dan algoritma genetika sebesar 0,985 yang termasuk ke dalam kategori *excellent classification* karena berada diantara nilai 0,90 - 1,00

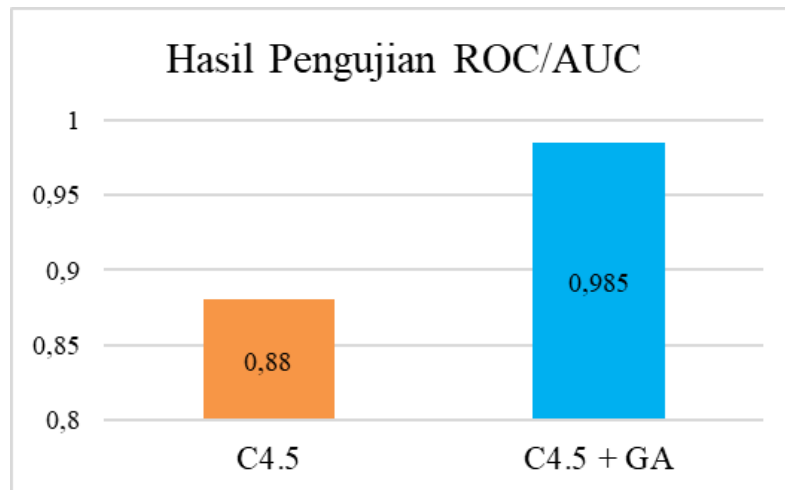
Hasil Pengujian

Berikut merupakan grafik yang menunjukkan perbandingan nilai *accuracy*, *precision*, *recall* yang diperoleh dari pengujian algoritma C4.5 tanpa di optimasi dengan algoritma C4.5 yang di optimasi dengan algoritma genetika, ditunjukkan pada Gambar 6.



Gambar 6. Grafik Hasil Pengujian

Terakhir yaitu gambar grafik ROC/AUC dari hasil pengujian algoritma C4.5 tanpa di optimasi dengan hasil pengujian algoritma C4.5 yang di optimasi dengan algoritma genetika yang ditunjukkan pada Gambar 7.



Gambar 7. Hasil Pengujian ROC/AUC

Perbandingan hasil pengujian terlihat pada tabel 6

Tabel 6. Perbandingan Hasil Pengujian

	C4.5	C4.5 + GA	Perbandingan nilai	Keterangan
Accuracy	84.48%	96.55%	12.07%	Naik
Precision	85.00%	100.00%	15.00%	Naik
Recall	73.91%	91.39%	17.39%	Naik
ROC/AVC (Optimistic)	088	0.985	0.105	Naik

KESIMPULAN

Berdasarkan hasil dari proses penelitian yang telah dilakukan menggunakan teknik optimasi algoritma genetika dengan metode *decision tree* atau algoritma C4.5 dalam memprediksi minat baca siswa pada perpustakaan SMK Negeri 1 Gantar dengan menggunakan data pengunjung yang dimiliki oleh perpustakaan SMK Negeri 1 Gantar sebanyak 290 data, maka dapat disimpulkan bahwa teknik optimasi dengan menggunakan algoritma genetika dinilai berhasil untuk meningkatkan *accuracy* dari algoritma C4.5 dalam memprediksi minat baca siswa pada perpustakaan SMK Negeri 1 Gantar. Hasil *accuracy* yang diperoleh dari algoritma C4.5 yaitu sebesar 84.48% dan setelah dioptimasi menggunakan algoritma genetika mengalami kenaikan *accuracy* sebesar 12.07% sehingga hasil *accuracy* yang diperoleh dari optimasi menjadi sebesar 96.55%.

SARAN

Perlu adanya pengujian dan pengembangan menggunakan metode seperti *CART*, *k-Nearest Neighbors*, *K-Means* dan lainnya, untuk membandingkan hasil tertinggi yang diperoleh dalam memprediksi minat baca siswa pada perpustakaan SMK Negeri 1 Gantar. Perlu adanya pengujian dan pengembangan dengan metode optimasi lain diantaranya seperti *Adaboost*, *Ant Colony Optimization (ACO)*, *Particle Swarm Optimization (PSO)* dan lain sebagainya, untuk mengetahui performa dalam meningkatkan *accuracy* setelah dilakukannya proses optimasi.

DAFTAR PUSTAKA

- Arsi, P., & Somantri, O. (2018). Deteksi Dini Penyakit Diabetes Menggunakan Algoritma Neural Network Berbasis Algoritma Genetika, *03(03)*, 290–294. <https://doi.org/10.30591/jpit.v3i3.1008>
- Fatmawati. (2016). Perbandingan Algoritma Klasifikasi Data Mining Model C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Diabetes, *XIII(1)*, 50–59.
- Febriarini, A. S., & Astuti, E. Z. (2019). Penerapan Algoritma C4.5 untuk Prediksi Kepuasan Penumpang Bus Rapid Transit (BRT) Trans Semarang. *Jurnal Eksplorasi Informatika*, *8(2)*, 95–103.

- Galih. (2019). Data Mining di Bidang Pendidikan untuk Analisa Prediksi Kinerja Mahasiswa dengan Komparasi 2 Model Klasifikasi pada STMIK Jabar, *2*(1), 23–30.
- Iriadi, N., & Nuraeni, N. (2016). kajian penerapan metode klasifikasi data mining algoritma C4. 5 untuk prediksi kelayakan kredit pada bank mayapada jakarta. *Jurnal Teknik Komputer Amik BSI*, *2*(1), 132–137.
- Meilani, B. D., & Asadulloh, M. (2015). Data Mining Untuk Menggali Pola Mahasiswa Baru Menggunakan Metode Frequent Pattern Growth (Studi Kasus : Institut Teknologi Adhi Tama Surabaya), 269–276.
- Raharjo, S., & Mada, U. G. (2014). Klasterisasi, klasifikasi dan peringkasan teks berbahasa indonesia, *8*(Kommit), 391–401.
- Singestecia, R., Handoyo, E., Isdaryanto, N., Politik, J., Sosial, F. I., & Negeri, U. (2018). Partisipasi Politik Masyarakat Tionghoa dalam Pemilihan Kepala Daerah di Slawi Kabupaten Tegal, *2*(1), 63–72.
- Sugeha, I. H., Inkiriwang, R. L., & Pratahis, P. A. K. (2019). Optimasi penjadwalan menggunakan metode algoritma genetika pada proyek rehabilitasi puskesmas minanga. *Jurnal Sipil Statik*, *7*(12).
- Utami, R. D., Wibowo, D. C., & Susanti, Y. (2018). Analisis minat membaca siswa pada kelas tinggi di sekolah dasar negeri 01 belitang. *Jurnal Pendidikan Dasar Perkebasa*, *4*(1), 179–188.
- Yatun, S. (2015). Menumbuhkan minat baca siswa melalui perpustakaan, *X*(2), 171–187.