OPTIMASI MODEL KLASIFIKASI C4.5 DAN PARTICLE SWARM OPTIMIZATION UNTUK PREDIKSI SISWA BERMASALAH

Hayati Noor

Fakultas Teknologi Informasi
Universitas Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin
email: adhienk.athie@gmail.com

Abstrak

Sekolah adalah lembaga pendidikan kedua bagi seorang anak yang memiliki peranan sangat strategis yang akan menjadi pusat-pusat kegiatan pendidikan untuk menumbuhkan dan mengembangkan potensi anak sebagai makhluk individu, sosial, susila dan religius. Deteksi dini dapat juga mendeteksi siswa dengan masalah belum serius, sehingga pihak sekolah memberikan dukungan dan perhatian yang tepat sebelum kondisi ini memburuk. Dalam menentukan apakah seorang siswa bermasalah maka pendidik (orang tua, wali siswa, wali kelas, dan guru) harus memperhatikan kekhasan perilaku anak dan perlu memahami tahapan perkembangan anak dalam segala aspek. Berdasarkan kondisi tersebut teknik data mining yang tepat digunakan adakah klasifikasi. Salah satu teknik klasifikasi data mining adalah C4.5. Dalam penelitian ini, membandingkan algoritma C4.5 dengan C4.5 berbasis PSO (Particle Swarm Optimazion) yang diterapkan pada data siswa bermasalah. C4.5 dan Particle Swarm Optimization menjadi lebih baik dalam memprediksi nilai akurasi daripada menggunakan hanya metode C4.5 saja, yang mampu meningkatkan nilai akurasi cukup tinggi yaitu sebesar 35.7%. Optimasi Decision Tree C4.5 dapat diterapkan untuk prediksi siswa berpotensi bermasalah dengan tingkat akurasi 99,08%. Dalam prediksi siswa berpotensi bermasalah, akurasi optimasi Algoritma C4.5 dan Particle Swarm Optimization lebih baik dari pada Algoritma decision Tree C.4.5 saja, dengan perbedaan yang cukup besar.

Keyword : *Siswa bermasalah, data mining, C4.5. PSO,*

I. PENDAHULUAN

Lembaga penelitian telah menunjukkan bahwa identifikasi awal, penilaian, dan intervensi untuk masalah emosional dan perilaku anak-anak sampai remaja dapat membantu mencegah masalah yang lebih serius, seperti kegagalan pendidikan, penyalahgunaan narkoba, keterlibatan dalam tindak kriminal, atau bunuh diri [1]. Data terbaru menunjukkan bahwa, pada masa tertentu, antara 5 dan 9 persen dari anak-anak dan pengalaman remaja mengalami masalah emosional yang serius, yang dapat menghambat kemampuan mereka untuk menjalani kehidupan yang penuh dan produktif [2].

Deteksi dini dapat membantu mengurangi dampak yang signifikan terhadap anak-anak dan remaja yang mungkin mengalami masalah perilaku dan kebiasaan bermasalah. Deteksi dini dapat juga mendeteksi siswa dengan masalah belum serius, sehingga pihak sekolah memberikan dukungan dan perhatian yang tepat sebelum kondisi ini memburuk. Dengan demikian, prediksi siswa bermasalah adalah langkah pertama dalam deteksi dini, mengenali masalah emosional dan perilaku dan memberikan bantuan pada titik awal dan efektif; jika masalah terdeteksi, penilaian dan evaluasi lebih lanjut dapat menentukan perhatian dan layanan yang sesuai yang diperlukan [3].

Dalam menentukan apakah seorang siswa bermasalah maka pendidik (orang tua, wali siswa, wali kelas, dan guru) harus memperhatikan kekhasan perilaku anak dan perlu memahami tahapan perkembangan anak dalam segala aspek. Pemahaman tersebut dapat membantu menganalisis dan mengelompokkan anak pada kategori bermasalah atau tidak. Apakah frekuensi tingkah laku yang menyimpang tersebut terlihat setiap waktu? Apakah perilaku tersebut mengganggu aktivitas

siswa baik di dalam maupun di luar kelas? Jika tingkah laku tersebut tidak diatasi dengan segera apakah akan menimbulkan masalah dalam perkembangan anak secara menyeluruh? Jika semua pertanyaan tersebut dijawab "ya", maka besar kemungkinan anak tersebut bermasalah. Daftar kekhasan perilaku dan kebiasaan ini telah buat secara sistematis, di antaranya *Student Risk Screening Scale* (SRSS) dan Kredit Point Sanksi Pelanggaran tata Tertib Siswa

Data Point Siswa berdasarkan daftar cek kredit point tersedia, namun perlu suatu proses untuk mendapatkan informasi yang berguna sesuai dengan tujuan tertentu. Proses ini dapat menggunakan klasifikasi prediksi [4]. Tujuannya agar dapat diketahui pola-pola universal dari data yang ada. Faktor-faktor yang dapat mempengaruhi perilaku siswa inilah yang akan dianalisa dan di klasifikasi untuk kemudian di prediksi sehingga membentuk kelompok-kelompok yang memiliki kesamaan tersembunyi dari data yang telah dianalisa [5]. Jadi, klasifikasi merupakan suatu proses menemukan kumpulan pola atau fungsi yang mendeskripsikan serta memisahkan kelas data yang satu dengan yang lainnya untuk menyatakan objek tersebut masuk pada kategori tertentu yang sudah ditentukan [6].

Klasifikasi data dapat digunakan untuk mengetahui tingkat hubungan antar data tersebut adalah dengan metode *decision tree*. Metode ini populer karena mampu melakukan klasifikasi sekaligus menunjukkan hubungan antar atribut. Algoritma klasifikasi *decision tree* C4.5, C5.0, CART, daan CHAID dapat digunakan untuk menganalis kredit [7], penjualan dan sebarannya [8]. Hongxing He, dkk membandingkan algoritma C4.5 dengan *Logistic Regression* serta *Risk Pattern Mining* (RPM) ketika mempelajari faktor-faktor yang menyebabkan untuk memutuskan apakah menyusui bayi baru lahir secara langsung atau tidak menyusui bayi yang baru lahir. Mereka menyatakan bahwa C4.5 lebih baik dibandingkan *logistic regression* serta *Risk Pattern Mining*.

Berdasarkan latar tersebut, dalam penelitian ini, algoritma C4.5 dan *Particle Swarm Optimization* akan digunakan untuk membangun model klasifikasi prediksi siswa bermasalah dan dibandingkan dengan hasil algoritma CART.

Perumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah dalam bentuk pertanyaan penelitian (*research questions*), adalah: "Bagaimana model algoritma klasifikasi C4.5 dan *Particle Swarm Optimization* dapat membantu pendidik memprediksi siswa bermasalah dengan akurat?"

Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah diatas, maka penelitian ini bertujuan untuk mengoptimasi penerapan algoritma klasifikasi C4.5 dan *Particle Swarm Optimization* untuk prediksi siswa bermasalah.

Manfaat Penelitian

a. Manfaat praktis dari hasil penelitian ini adalah diharapkan agar dapat menjadi referensi untuk penerapan model klasifikasi untuk prediksi.
b. Manfaat teoritis dari hasil penelitian ini diharapkan dapat membantu pendidik (orang tua siswa, wali siswa, wali kelas, dan guru) dalam memperhatikan perkembangan pendidikan dan karakter siswa.
c. Manfaat kebijakan yaitu diharapkan agar metode klasifikasi untuk prediksi mampu menjadi deteksi dini siswa bermasalah dan dapat mempengaruhi dan mengembangkan pola asuh terhadap siswa.

II. METODE PENELITIAN

Metode penelitian yang dilakukan adalah metode penelitian eksperimen, dengan tahapan penelitian sebagai berikut :



Gambar 2.1 Tahapan Metode Penelitian Eksperimen

Data yang digunakan adalah data yang diperoleh dari SMAN 2 Banjarbaru. Pada proses awal data tersebut diolah kembali untuk menentukan objek yang akan dijadikan bahan pengujian. Setelah dianalisa dan diolah, data kemudian diuji dengan menggunakan metode yang telah ditentukan agar dapat diketahui hasil dari pengujian data tersebut.

Pengumpulan Data

Data didapatkan dari Sekolah Menengah Atas Negeri 2 Banjarbaru. Data diambil adalah tahun ajaran 2016-2017 siswa kelas XI IPA dan IPS.

- Data Sekunder
  Data siswa meliputi: Nama, NIPD, Tangal Lahir, Jenis Kelamin, Umur, Kelas, Asal Sekolah, Agama, Anak Ke, Jumlah Saudara, Status Anak, Tempat Belajar, Nama Ayah, Tahun Lahir Ayah, Jenjang Pendidikan Ayah, Pekerjaan Ayah, Penghasilan Ayah, Nama Ibu, Tahun Lahir Ibu, Jenjang Pendidikan Ibu, Pekerjaan Ibu, Penghasilan Ibu, Poin. Jadi ada 23 atribut.
- Data Primer
  Data primer yang digunakan adalah data hasil komputasi algoritma klasifikasi.

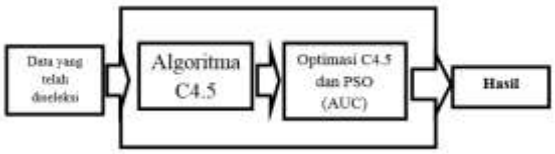
Pengolahan Data Awal

Data yang dikumpulkan kemudian dipilah dan disiapkan untuk menghasilkan data yang siap untuk diproses oleh Algoritma. Data siswa meliputi: NIPD, Umur, Jenis Kelamin, Kelas, Asal Sekolah, Agama, Anak ke, Jumlah Saudara, Status Anak, Tempat Belajar, Umur Ayah, Pendidkan Ayah, Pekerjaan Ayah, Penghasilan Ayah, Umur Ibu, Pendidikan Ibu, Pekerjaan Ibu, Penghasilan Ibu, Poin. Jadi hanya 19 atribut.

Data yang pilih adalah data yang seluruh attribut yang diperlukan lengkap (22 atribut). Jadi data siswa yang tidak lengkap, otomatis diabaikan.

Metode/Model yang diusulkan:

Metode yang digunakan adalah perbandingan akurasi prediksi yang dihasilkan oleh Algoritma Klasifikasi C4.5 dan *Particle Swarm Optimization*. Data diproses menggunakan C4.5 dan *Particle Swarm Optimization* yang diaplikasikan menggunakan *RapidMiner*.



Gambar 2.2 Metode yang diusulkan

Eksperimen dan Pengujian Model/Metode

Data akan diproses oleh model C4.5 kemudian dicatat akurasinya. Kemudian, dilakukan beberapa perubahan setting, dan hasilnya dicatat. Kemudian data yang sama akan diproses secara sama dengan model Optimasi C4.5 dan PSO, kemudian hasilnya dibandingkan.

Evaluasi dan Validasi

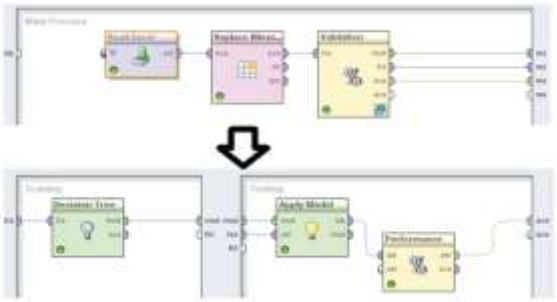
Evaluasi dan validasi dilakukan dengan mengamati hasil prediksi dibandingkan dengan data awal yang diolah. Hasil akan diperoleh setelah proses eksperimen dan pengujian dilaksanakan.

Pengukuran kinerja dilakukan dengan menghitung rata-rata error yang terjadi melalui besaran akurasi (AUC). Semakin besar akurasi menyatakan semakin dekat nilai prediksi dengan nilai sebenarnya. Dengan demikian dapat diketahui tingkat akurasi algoritma C4.5 yang dioptimasi dengan PSO. Sehingga dapat diketahui apakah algoritma C4.5 dan PSO dapat digunakan untuk memprediksi anak berpotensi bermasalah di SMA khususnya SMAN 2 Banjarbaru.

III. HASIL DAN PEMBAHASAN

Hasil Eksperimen dan Pengujian Metode Eksperimen C4.5

Untuk desain eksperimen dan pengujian penelitian dengan menggunakan C4.5 adalah seperti gambar 3.1 :



Gambar 3.1 Desain eksperimen C4.5

Pada gambar 4.1 di atas terlihat desain model eksperimen yang digunakan peneliti untuk memperoleh tingkat akurasi prediksi siswa bermasalah berpotensi bermasalah dengan menggunakan algoritma C4.5, data yang kosong dhilangkan dengan menggunakan *replace missing*. Pengujian model dengan merubah nilai validasinya. Pengujian data nilai validasi dimulai 2, 3, 4, 5, 6, 7, 8, 9 dan 10 kemudian dievaluasi dan dibandingkan dengan beberapa algoritma lain dengan hasil sebagai berikut :

Tabel 3.1 Hasil Pengujian C4.5

Validasi	2	3	4	5	6	7	8	9	10
Accuracy	63.08%	62.15%	60.92%	58.46%	62.46%	61.23%	58.77%	63.38%	62.15%
Precision	63.46%	63.46%	63.05%	62.02%	64.96%	65.60%	62.88%	66.28%	64.52%
Recall	97.06%	93.63%	91.18%	87.25%	87.25%	80.39%	84.80%	84.80%	88.24%
AUC	0.464	0.489	0.505	0.498	0.506	0.520	0.536	0.531	0.541

Dari hasil percobaan seperti tabel 3.1 hasil pengujian sebanyak 9 percobaan dengan hasil tingkat akurasi tertinggi sebesar 63.38% dengan nilai validasi = 9.

Hasil Akurasi C4.5 Nilai Validasi 2

Tahap pertama pengujian dilakukan dengan data nomer akurasi validasi 2. Ketika diimplementasikan menghasilkan data sebagai berikut :

Tabel 3.2 Hasil Akurasi C4.5

accuracy: 63.08% +/- 0.11% (mikro: 63.08%)			
	True berpotensi Bermasalah	true Tidak Bermasalah	class precision
pred. Berpotensi Bermasalah	7	6	53.85%
Pred. Tidak Bermasalah	114	198	63.46%
class recall	5.79%	97.06%	

Dari hasil pengujian pada tabel 3.2 menggunakan nilai validasi dengan data original tersebut didapatkan nilai akurasi 63.08%.

Hasil Akurasi C4.5 Nilai Validasi 3

Tahap kedua pengujian dilakukan dengan data nomer akurasi validasi 3. Ketika di implementasikan menghasilkan data sebagai berikut :

Tabel 3.3 Hasil Akurasi C4.5

accuracy: 62.15% +/- 1.40% (mikro: 62.15%)			
	True berpotensi Bermasalah	True Tidak Bermasalah	class precision
pred. Berpotensi Bermasalah	11	13	45.83%
Pred. Tidak Bermasalah	110	191	63.46%
class recall	9.09%	93.63%	

Dari hasil pengujian pada tabel 3.3 menggunakan nilai validasi dengan data original tersebut didapatkan nilai akurasi 62.15%.

Hasil Akurasi C4.5 Nilai Validasi 4

Tahap ketiga pengujian dilakukan dengan data nomer akurasi validasi 4. Ketika di implementasikan menghasilkan data sebagai berikut :

Tabel 3.4 Hasil Akurasi C4.5

accuracy: 60.93% +/- 2.32% (mikro: 60.92%)			
	True berpotensi Bermasalah	True Tidak Bermasalah	class precision
pred. Berpotensi Bermasalah	12	18	40.00%
Pred. Tidak Bermasalah	109	186	63.05%
class recall	9.92%	91.18%	

Dari hasil pengujian pada tabel 3.4 menggunakan nilai validasi dengan data original tersebut didapatkan nilai akurasi 60.92%.

Hasil Akurasi C4.5 Nilai Validasi 5

Tahap keempat pengujian dilakukan dengan data nomer akurasi validasi 5. Ketika diimplementasikan menghasilkan data sebagai berikut :

Tabel 3.5 Hasil Akurasi C4.5

accuracy: 58.46% +/- 4.46% (mikro: 58.46%)			
	true berpotensi Bermasalah	True Tidak Bermasalah	class precision
pred. Berpotensi Bermasalah	12	26	31.58%
Pred. Tidak Bermasalah	109	178	62.02%
class recall	9.92%	87.25%	

Dari hasil pengujian pada tabel 3.5 menggunakan nilai validasi dengan data original tersebut didapatkan nilai akurasi 58.46%.

Hasil Akurasi C4.5 Nilai Validasi 6

Tahap kelima pengujian dilakukan dengan data nomer akurasi validasi 6. Ketika diimplementasikan menghasilkan data sebagai berikut :

Tabel 3.6 Hasil Akurasi C4.5

accuracy: 62.44% +/- 3.50% (mikro: 62.46%)			
	true berpotensi Bermasalah	true Tidak Bermasalah	class precision
pred. Berpotensi Bermasalah	25	26	49.02%
Pred. Tidak Bermasalah	96	178	64.96%
class recall	20.66%	87.25%	

Dari hasil pengujian pada tabel 3.6 menggunakan nilai validasi dengan data original tersebut didapatkan nilai akurasi 62.46%.

Hasil Akurasi C4.5 Nilai Validasi 7

Tahap keenam pengujian dilakukan dengan data nomer akurasi validasi 7. Ketika diimplementasikan menghasilkan data sebagai berikut :

Tabel 3.7 Hasil Akurasi C4.5

accuracy: 61.23% +/- 8.10% (mikro: 61.23%)			
	true berpotensi Bermasalah	true Tidak Bermasalah	class precision
pred. Berpotensi Bermasalah	35	40	46.67%
Pred. Tidak Bermasalah	86	164	65.60%
class recall	28.93%	80.39%	

Dari hasil pengujian pada tabel 3.7 menggunakan nilai validasi dengan data original tersebut didapatkan nilai akurasi 61.23%.

Hasil Akurasi C4.5 Nilai Validasi 8

Tahap ketujuh pengujian dilakukan dengan data nomer akurasi validasi 8. Ketika diimplementasikan menghasilkan data sebagai berikut :

Tabel 3.8 Hasil Akurasi C4.5

accuracy: 58.77% +/- 4.23% (mikro: 58.77%)			
	true berpotensi Bermasalah	true Tidak Bermasalah	class precision
pred. Berpotensi Bermasalah	18	31	36.73%
Pred. Tidak Bermasalah	103	173	62.68%
class recall	14.88%	84.80%	

Dari hasil pengujian pada tabel 3.8 menggunakan nilai validasi dengan data original tersebut didapatkan nilai akurasi 58.77%.

Hasil Akurasi C4.5 Nilai Validasi 9

Tahap kedelapan pengujian dilakukan dengan data nomer akurasi validasi 9. Ketika diimplementasikan menghasilkan data sebagai berikut :

Tabel 3.9 Hasil Akurasi C4.5

accuracy: 63.38% +/- 5.22% (mikro: 63.38%)			
	true berpotensi Bermasalah	true Tidak Bermasalah	class precision
pred. Berpotensi Bermasalah	33	31	51.56%
Pred. Tidak Bermasalah	88	173	66.28%
class recall	27.27%	84.80%	

Dari hasil pengujian pada tabel 3.9 menggunakan nilai validasi dengan data original tersebut didapatkan nilai akurasi 63.38%.

Hasil Akurasi C4.5 Nilai Validasi 10

Tahap kesembilan pengujian dilakukan dengan data nomer akurasi validasi 10. Ketika diimplementasikan menghasilkan data sebagai berikut :

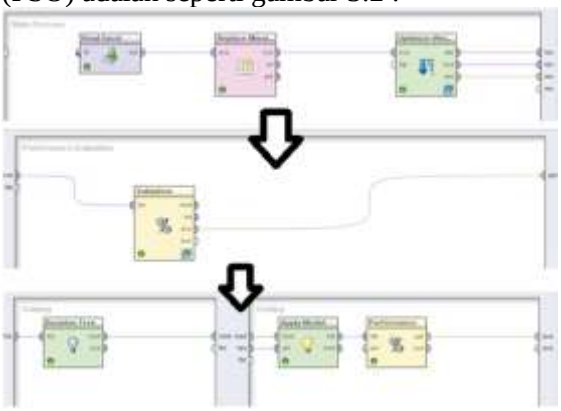
Tabel 3.10 Hasil Akurasi C4.5

accuracy: 62.15% +/- 7.25% (mikro: 62.15%)			
	true berpotensi Bermasalah	true Tidak Bermasalah	class precision
pred. Berpotensi Bermasalah	22	24	47.83%
Pred. Tidak Bermasalah	99	180	64.52%
class recall	18.18%	88.24%	

Dari hasil pengujian pada tabel 3.10 menggunakan nilai validasi dengan data original tersebut didapatkan nilai akurasi 62.15%.

Eksperimen Optimasi C4.5 dan Particle Swarm Optimization

Untuk desain eksperimen dan pengujian penelitian dengan menggunakan C4.5 berbasis *Particle Swarm Optimization* (PSO) adalah seperti gambar 3.2 :



Gambar 4.2 Desain eksperimen Optimasi C4.5 dan PSO

Pada gambar 4.2 di atas terlihat desain model eksperimen yang digunakan peneliti untuk memperoleh tingkat akurasi prediksi siswa berpotensi bermasalah dengan menggunakan algoritma C4.5 dan PSO, data yang kosong dhilangkan dengan menggunakan *replace missing*. Algoritma PSO diterapkan untuk dapat memperoleh nilai bobot variable yang terbaik pada C4.5 sehingga terjadinya peningkatan akurasi. Pengujian model dengan merubah nilai validasinya. Pengujian data nilai validasi dimulai 2, 3, 4, 5, 6, 7, 8, 9 dan 10 kemudian dievaluasi dan dibandingkan dengan beberapa algoritma lain dengan hasil sebagai berikut :

Tabel 3.11 Hasil Pengujian optimasi C4.5 dan particle swarm optimization

Validasi	2	3	4	5	6	7	8	9	10
Accuracy	95.69%	63.38%	99.08%	99.08%	98.15%	97.23%	97.23%	93.54%	99.08%
Precision	97.98%	63.41%	99.02%	99.02%	99.01%	97.10%	97.10%	94.20%	99.02%
Recall	95.10%	98.53%	99.51%	99.51%	98.04%	98.53%	98.53%	95.59%	99.51%
AUC	0.977	0.498	0.500	0.500	0.739	0.697	0.662	0.798	0.500
Time	25 s	40 s	52 s	1m10s	1m20s	2m47s	2m35s	3m21	2m03s

Dari hasil percobaan seperti tabel 3.11 hasil pengujian sebanyak 9 percobaan dengan hasil tingkat akurasi tertinggi sebesar 99.08% dengan nilai validasi = 4.

Hasil Optimasi C4.5 dan *Particle Swarm Optimization* Nilai Validasi 2

Tahap kesepuluh pengujian dilakukan dengan data nomer akurasi validasi 2. Ketika diimplementasikan menghasilkan data sebagai berikut :

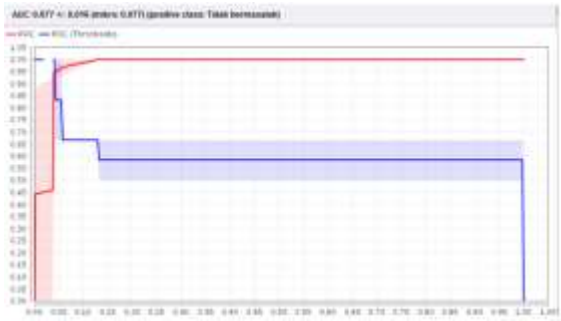
Tabel 3.12 Hasil Optimasi C4.5 dan Particle Swarm Optimization

accuracy: 95.70% +/- 2.45% (mikro: 95.69%)

	true berpotensi Bermasalah	True Tidak Bermasalah	class precision
pred. Berpotensi Bermasalah	117	10	92.13%
Pred. Tidak Bermasalah	4	194	97.98
class recall	96.69%	85.10%	

Dari hasil pengujian pada tabel 3.12 menggunakan nilai validasi dengan data original tersebut didapatkan nilai akurasi 95.69%. Dan akurasi prediksi data berpotensi bermasalah 96,69% sedangkan tidak bermasalah hasil akurasinya 85,10%.

Hasil perhitungan diatas dapat divisualisasikan dengan kurva ROC perbandingan kedua kelas bisa dilihat pada gambar 3.3 yang merupakan kurva ROC untuk optimasi algoritma C4.5 dan *Particle Swarm Optimization* dengan nilai validasi 2. Garis horizontal adalah *false positives* dan garis vertikal adalah *true positives*. Menghasilkan nilai AUC (*Area Under Curve*) sebesar 0.977 dengan nilai akurasi cukup (*fair classification*) pada dataset kredit point siswa berpotensi bermasalah.



Gambar 3.3 Kurva AUC

Hasil Optimasi C4.5 dan *Particle Swarm Optimization* Nilai Validasi 3

Tahap kesebelas pengujian dilakukan dengan data nomer akurasi validasi 3. Ketika diimplementasikan menghasilkan data sebagai berikut :

Tabel 3.13 Hasil Optimasi C4.5 dan Particle Swarm Optimization

accuracy: 63.39% +/- 6.84% (mikro: 63.38%)

	true berpotensi Bermasalah	True Tidak Bermasalah	class precision
pred. Berpotensi Bermasalah	5	3	62.50%
Pred. Tidak Bermasalah	116	201	63.41%
class recall	4.13%	98.53%	

Dari hasil pengujian pada tabel 3.13 menggunakan nilai validasi dengan data original tersebut didapatkan nilai akurasi 63.38%. Dan akurasi prediksi data berpotensi bermasalah 4,13% sedangkan tidak bermasalah hasil akurasinya 98,53%.

Hasil perhitungan diatas dapat divisualisasikan dengan kurva ROC perbandingan kedua kelas bisa dilihat pada gambar 3.4 yang merupakan kurva ROC untuk optimasi algoritma C4.5 dan *Particle Swarm Optimization* dengan nilai validasi 3. Garis horizontal adalah *false positives* dan garis vertikal adalah *true positives*. Menghasilkan nilai AUC (*Area Under Curve*) sebesar 0.498 dengan nilai akurasi cukup (*fair classification*) pada dataset kredit point siswa berpotensi bermasalah.



Gambar 3. 4Kurva AUC

Hasil Optimasi C4.5 dan *Particle Swarm Optimization* Nilai Validasi 4

Tahap keduabelas pengujian dilakukan dengan data nomer akurasi validasi 4. Ketika diimplementasikan menghasilkan data sebagai berikut :

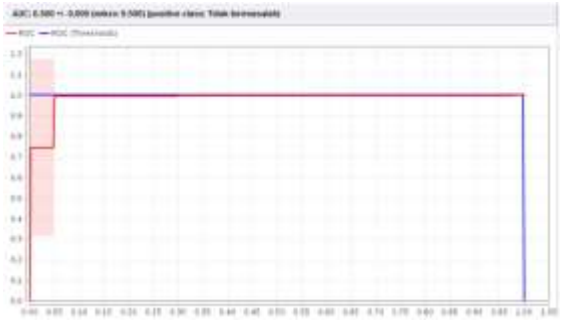
Tabel 3.14 Hasil akurasi Optimasi C4.5 dan Particle Swarm Optimization

accuracy: 99.07% +/- 1.02% (mikro: 99.08%)

	true berpotensi Bermasalah	True Tidak Bermasalah	class precision
pred. Berpotensi Bermasalah	119	1	99.17%
Pred. Tidak Bermasalah	2	203	99.02%
class recall	98.35%	99.51%	

Dari hasil pengujian pada tabel 3.14 menggunakan nilai validasi dengan data original tersebut didapatkan nilai akurasi 99.08%. Dan akurasi prediksi data berpotensi bermasalah 98,35% sedangkan tidak bermasalah hasil akurasinya 99,51%.

Hasil perhitungan diatas dapat divisualisasikan dengan kurva ROC perbandingan kedua kelas bisa dilihat pada gambar 3.5 yang merupakan kurva ROC untuk optimasi algoritma C4.5 dan *Particle Swarm Optimization* dengan nilai validasi 4. Garis horizontal adalah *false positives* dan garis vertikal adalah *true positives*. Menghasilkan nilai AUC (*Area Under Curve*) sebesar 0.500 dengan nilai akurasi cukup (*fair classification*) pada dataset kredit point siswa berpotensi bermasalah.



Gambar 3.5 Kurva AUC

Hasil Optimasi C4.5 dan *Particle Swarm Optimization* Nilai Validasi 5

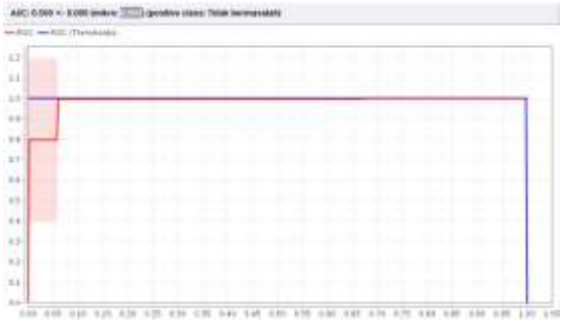
Tahap ketigabelas pengujian dilakukan dengan data nomer akurasi validasi 5. Ketika diimplementasikan menghasilkan data sebagai berikut :

Tabel 3.15 Hasil Akurasi Optimasi C4.5 dan Particle Swarm Optimization

accuracy: 99.07% +/- 1.02% (mikro: 99.08%)			
	true berpotensi Bermasalah	True Tidak Bermasalah	class precision
prod. Berpotensi Bermasalah	119	1	99.17%
Prod. Tidak Bermasalah	2	203	99.02%
class recall	98.35%	99.51%	

Dari hasil pengujian pada tabel 3.15 menggunakan nilai validasi dengan data original tersebut didapatkan nilai akurasi 99.08%. Dan akurasi prediksi data berpotensi bermasalah 98,35% sedangkan tidak bermasalah hasil akurasinya 99,51%.

Hasil perhitungan diatas dapat divisualisasikan dengan kurva ROC perbandingan kedua kelas bisa dilihat pada gambar 3.6 yang merupakan kurva ROC untuk optimasi algoritma C4.5 dan *Particle Swarm Optimization* dengan nilai validasi 5. Garis horizontal adalah *false positives* dan garis vertikal adalah *true positives*. Menghasilkan nilai AUC (*Area Under Curve*) sebesar 0.500 dengan nilai akurasi cukup (*fair classification*) pada dataset kredit point siswa berpotensi bermasalah.



Gambar 3.6 Kurva AUC

Hasil Optimasi C4.5 dan *Particle Swarm Optimization* Nilai Validasi 6

Tahap keempatbelas pengujian dilakukan dengan data nomer akurasi validasi 6. Ketika diimplementasikan menghasilkan data sebagai berikut :

Tabel 3.16 Hasil Akurasi Optimasi C4.5 dan Particle Swarm Optimization

accuracy: 98.15% +/- 1.51% (mikro: 98.15%)			
	true berpotensi Bermasalah	True Tidak Bermasalah	class precision
prod. Berpotensi Bermasalah	119	4	96.75%
Prod. Tidak Bermasalah	2	200	99.01%
class recall	98.35%	98.04%	

Dari hasil pengujian pada tabel 3.16 menggunakan nilai validasi dengan data original tersebut didapatkan nilai akurasi 98.15%. Dan akurasi prediksi data berpotensi bermasalah 98,35% sedangkan tidak bermasalah hasil akurasinya 98,04%.

Hasil perhitungan diatas dapat divisualisasikan dengan kurva ROC perbandingan kedua kelas bisa dilihat pada gambar 3.7 yang merupakan kurva ROC untuk optimasi algoritma C4.5 dan *Particle Swarm Optimization* dengan nilai validasi 6. Garis horizontal adalah *false positives* dan garis vertikal adalah *true positives*. Menghasilkan nilai AUC (*Area Under Curve*) sebesar 0.739 dengan nilai akurasi cukup (*fair classification*) pada dataset kredit point siswa berpotensi bermasalah.



Gambar 3.7 Kurva AUC

Hasil Optimasi C4.5 dan *Particle Swarm Optimization* Nilai Validasi 7

Tahap kelimabelas pengujian dilakukan dengan data nomer akurasi validasi 7. Ketika diimplementasikan menghasilkan data sebagai berikut :

Tabel 3.17 Hasil Akurasi Optimasi C4.5 dan Particle Swarm Optimization

accuracy: 97.21% +/- 6.46% (mikro: 97.23%)			
	true berpotensi Bermasalah	True Tidak Bermasalah	class precision
pred. Berpotensi Bermasalah	115	3	97.46%
Pred. Tidak Bermasalah	6	201	97.10%
class recall	95.04%	98.53%	

Dari hasil pengujian pada tabel 3.17 menggunakan nilai validasi dengan data original tersebut didapatkan nilai akurasi 97.23%. Dan akurasi prediksi data berpotensi bermasalah 85,04% sedangkan tidak bermasalah hasil akurasinya 98,53%.

Hasil perhitungan diatas dapat divisualisasikan dengan kurva ROC perbandingan kedua kelas bisa dilihat pada gambar 3.8 yang merupakan kurva ROC untuk optimasi algoritma C4.5 dan *Particle Swarm Optimization* dengan nilai validasi 7. Garis horizontal adalah *false positives* dan garis vertikal adalah *true positives*. Menghasilkan nilai AUC (*Area Under Curve*) sebesar 0.697 dengan nilai akurasi cukup (*fair classification*) pada dataset kredit point siswa berpotensi bermasalah.



Gambar 3.8 Kurva AUC

Hasil Optimasi C4.5 dan *Particle Swarm Optimization* Nilai Validasi 8

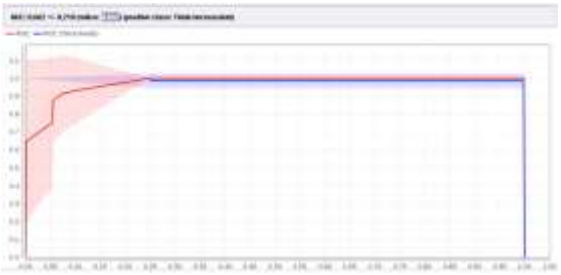
Tahap keenambelas pengujian dilakukan dengan data nomer akurasi validasi 8. Ketika diimplementasikan menghasilkan data sebagai berikut :

Tabel 3.18 Hasil Optimasi C4.5 dan Particle Swarm Optimization

accuracy: 97.23% +/- 4.30% (mikro: 97.23%)			
	true berpotensi Bermasalah	True Tidak Bermasalah	class precision
pred. Berpotensi Bermasalah	115	3	97.46%
Pred. Tidak Bermasalah	6	201	97.10%
class recall	95.04%	98.53%	

Dari hasil pengujian pada tabel 3.18 menggunakan nilai validasi dengan data original tersebut didapatkan nilai akurasi 97.23%. Dan akurasi prediksi data berpotensi bermasalah 95,04% sedangkan tidak bermasalah hasil akurasinya 98,53%.

Hasil perhitungan diatas dapat divisualisasikan dengan kurva ROC perbandingan kedua kelas bisa dilihat pada gambar 3.9 yang merupakan kurva ROC untuk optimasi algoritma C4.5 dan *Particle Swarm Optimization* dengan nilai validasi 8. Garis horizontal adalah *false positives* dan garis vertikal adalah *true positives*. Menghasilkan nilai AUC (*Area Under Curve*) sebesar 0.662 dengan nilai akurasi cukup (*fair classification*) pada dataset kredit point siswa berpotensi bermasalah.



Gambar 3.9 Kurva AUC

Hasil Optimasi C4.5 dan *Particle Swarm Optimization* Nilai Validasi 9

Tahap ketujuhbelas pengujian dilakukan dengan data nomer akurasi validasi 9. Ketika diimplementasikan menghasilkan data sebagai berikut :

Tabel 3.19 Hasil Akurasi Optimasi C4.5 dan Particle Swarm Optimizatio

accuracy: 93.54% +/- 9.88% (mikro: 93.54%)			
	true berpotensi Bermasalah	True Tidak Bermasalah	class precision
pred. Berpotensi Bermasalah	109	9	92.37%
Pred. Tidak Bermasalah	12	195	94.20%
class recall	90.08%	95.59%	

Dari hasil pengujian pada tabel 3.19 menggunakan nilai validasi dengan data original tersebut didapatkan nilai akurasi 93.54%. Dan akurasi prediksi data berpotensi bermasalah 90,08% sedangkan tidak bermasalah hasil akurasinya 95,59%.

Hasil perhitungan diatas dapat divisualisasikan dengan kurva ROC perbandingan kedua kelas bisa dilihat pada gambar 3.10 yang merupakan kurva ROC untuk optimasi algoritma C4.5 dan *Particle Swarm Optimization* dengan nilai validasi 9. Garis horizontal adalah *false positives* dan garis vertikal adalah *true positives*. Menghasilkan nilai AUC (*Area Under Curve*) sebesar 0.798 dengan nilai akurasi cukup (*fair classification*) pada dataset kredit point siswa berpotensi bermasalah.



Gambar 4.10 Kurva AUC

Hasil Optimasi C4.5 dan *Particle Swarm Optimization* Nilai Validasi 10

Tahap kedelabanbelas pengujian dilakukan dengan data nomer akurasi validasi 10. Ketika diimplementasikan menghasilkan data sebagai berikut :

Tabel 3.20 Hasil Akurasi Optimasi C4.5 dan Particle Swarm Optimization

accuracy: 99.08% +/- 1.95% (mikro: 99.08%)

	true berpotensi Bermasalah	true Tidak Bermasalah	class precision
pred. Berpotensi Bermasalah	119	1	99.17%
Pred. Tidak Bermasalah	2	203	99.02%
class recall	98.35%	99.51%	

Dari hasil pengujian pada tabel 3.20 menggunakan nilai validasi dengan data original tersebut didapatkan nilai akurasi 99.08%. Dan akurasi prediksi data berpotensi bermasalah 98,35% sedangkan tidak bermasalah hasil akurasinya 99,51%.

Hasil perhitungan diatas dapat divisualisasikan dengan kurva ROC perbandingan kedua kelas bisa dilihat pada gambar 3.11 yang merupakan kurva ROC untuk optimasi algoritma C4.5 dan *Particle Swarm Optimization* dengan nilai validasi 10. Garis horizontal adalah *false positives* dan garis vertikal adalah *true positives*. Menghasilkan nilai AUC (*Area Under Curve*) sebesar 0.500 dengan nilai akurasi cukup (*fair classification*) pada dataset kredit point siswa berpotensi bermasalah.



Gambar 3.11 Kurva AUC

Perbandingan C4.5 dengan optimasi C.4.5 dan *Particle Swarm Optimization*

Tabel 3.21 Perbandingan C4.5 dengan optimasi C4.5 dan Particle Swarm Optimization

METODE	C4.5	Optimasi C4.5 dan *Particle Swarm Optimization*
Accuracy	63.38%	99.08%
Precision	66.28%	99.02%
Recall	84.80%	99.51%
AUC	0.531	0.500
Time		52 s

Dari tabel 3.21 dapat terlihat bahwa dengan menggunakan optimasi C4.5 dan *Particle Swarm Optimization* dapat meningkatkan hasil akurasi C4.5 dari 63,38 % sampai dengan 99.08%. Dapat dilihat jelas adanya kenaikan yang dihasilkan, kenaikannya cukup tinggi yaitu sebesar 35.7%.

Evaluasi dan Validasi Hasil

Setelah melakukan pengujian metode algoritma C4.5 dengan menggunakan *folds cross validation*, kemudian dilihat hasil–hasil dari pengujian tersebut yang bisa kita lihat pada tabel 3.21 dapat disimpulkan metode C4.5 berbasis *Particle Swarm Optimization* memprediksi lebih akurat.

Implementasi Penelitian

Dengan adanya penerapan optimasi C4.5 dan *Particle Swarm Optimization* bisa memberikan solusi dan dapat membantu pihak terkait dalam hal untuk mengidentifikasi dan memprediksi siswa berpotensi bermasalah atau tidak bermasalah khususnya sekolah-sekolah lainnya.

IV. Kesimpulan

Berdasarkan hasil pengujian klasifikasi dan pembahasan, maka diperoleh kesimpulan:

1. Optimasi Decision Tree C4.5 dan *Particle Swarm Optimization* menjadi lebih baik dalam memprediksi nilai akurasi dari menggunakan hanya metode C4.5 saja, yang mampu meningkatkan nilai akurasi cukup tinggi yaitu sebesar 35.7%.
2. Optimasi Decision Tree C4.5 dapat diterapkan untuk prediksi siswa berpotensi bermasalah dengan tingkat akurasi 99,08%. Dalam prediksi siswa berpotensi bermasalah, akurasi optimasi Algoritma C4.5 dan *Particle Swarm Optimization*

lebih baik dari pada Algoritma decision Tree C.4.5 saja, dengan perbedaan yang cukup besar.

V. Saran

Berdasarkan kesimpulan di atas, maka ada beberapa saran yang dapat disampaikan :

1. Untuk menghasilkan prediksi yang lebih baik dan akurat penelitian berikutnya memerlukan data inputan yang lebih lengkap
2. Perbanyak pelatihan konfigurasi sehingga mendapatkan model yang menghasilkan error sekecil-kecilnya.
3. Dapat dikombinasikan dengan algoritma optimasi lain seperti *Ant Colony Optimization* (ACO), *Adoboost*, atau algoritma optimasi lainnya.

Daftar Pustaka

[1] Eklund Katie et al., "*Early Identification of Behavioral and Emotional Problems in Youth: Universal screening versus Teacher-Referral Identification*," *The California School of Psychologist*, vol. 14, pp. 89-95, 2009.

[2] Wallbro wn Amber Rae, "*Concurrent and Predictive Validity of the Behavior and Emotional Screening System*," *Graduate College of Marshall University*, *Theses Paper* 587, 2013.

[3] Drummond T, "*The Student Risk Screening Scale* (SRSS)," Josephine *County Mental Health Program*, Grants Pass - Oregon, 1994.

[4] Ulf Johansson, Henrik Boström, and Tuve Löfström, "*Conformal Prediction Using Decision Trees*," *IEEE 2013*, vol. 1550-4786/13, 2013.

[5] K Anuradha, Radha N Tulasi, and Kumar Pavan, "*A Decision Tree Algorithm For Uncertain Data*," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. ISSN: 2249 – 8958 Volume-1, no. 4, 2012.

[6] I H Witten and Eibe Frank, *Data Mining : Practical Machine Learning Tools and Techniques Second Edition*. San Francisco: Morgan Kauffman, 2005.

[7] W Yogi Yusuf, "Perbandingan Performance Algoritma Decision Tree C5.0, CART, Dan CHAID: Kasus Prediksi Status Resiko Kredit Di Bank X," in *Seminar Nasional Aplikasi Teknologi Informasi (SNATI 2007)*, Yogyakarta, 2007

[8] Rong Cao and Lizhen Xu, "*Improved C4.5 Algorithm for the Analysis of Sales*," in *2009 Sixth Web Information Systems and Applications Conference*, 2009.