

BAB IV HASIL DAN PEMBAHASAN

4.1 *Bussiness Understanding* atau Pemahaman Bisnis

Tujuan dari penelitian ini adalah untuk membandingkan model algoritma yang memiliki akurasi sangat tinggi menggunakan metode algoritma C4.5 dan Naïve Bayes

4.2 *Data Preparation* atau Persiapan Data

Pengolahan data awal yang dilakukan pada data mining untuk mengubah data mentah (*raw data*) yang dikumpulkan dari berbagai sumber menjadi informasi yang lebih bersih yang kemudian digunakan untuk proses pengolahan data selanjutnya. Berdasarkan sebaran kuisioner yang dilakukan diperoleh hasil raw data sebanyak 1101 data. Potongan dataset bisa dilihat pada gambar 4.1 berikut:

future_care... <i>integer</i>	social_sup... <i>integer</i>	peer_press... <i>integer</i>	extracurric... <i>integer</i>	bullying <i>integer</i>	stress_level <i>integer label</i>
3	2	3	3	2	1
5	1	4	5	5	2
2	2	3	2	2	1
4	1	4	4	5	2
2	1	5	0	5	1
5	1	4	4	5	2
1	3	2	2	1	0
4	1	4	4	5	2
3	3	3	2	2	1
5	1	5	3	4	1
4	1	4	4	5	2

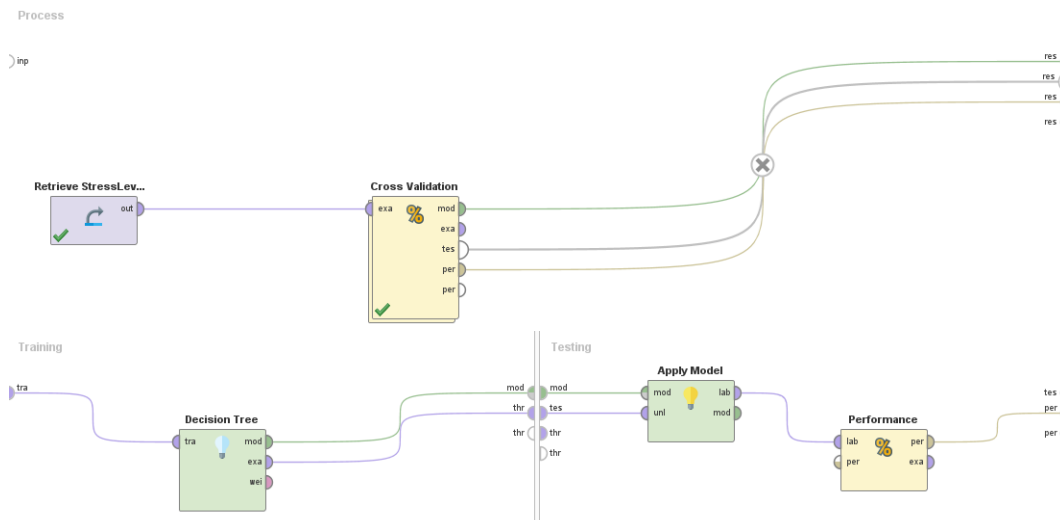
Gambar 4.1 Potongan Dataset

Pada tahap ini dilakukan pengubahan *type* variabel dan perubahan target/label.

4.3 Pemodelan

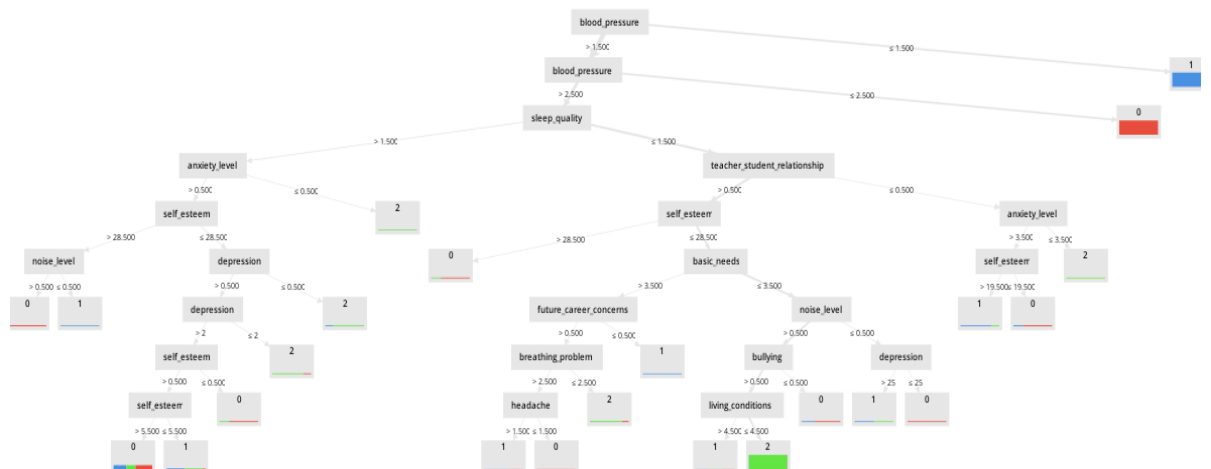
4.3.1 Penelitian Menggunakan Algoritma C4.5

Pada algoritma, record yang sudah di import ke *Rapid Miner* digunakan untuk menentukan pola pohon keputusan. Penerapan data pada *Rapid Miner* digunakan untuk Klasifikasi penerima beasiswa menggunakan algoritma C4.5 ditunjukkan pada gambar 4.2 dibawah ini:



Gambar 4.2 Model Klasifikasi Algoritma C4.5

Setelah melakukan beberapa langkah diatas dalam proses klasifikasi metode algoritma C4.5 maka akan diperoleh model yang terbentuk dari proses pengklasifikasian algoritma C4.5 berupa pohon keputusan seperti gambar 4.3 dibawah ini.



Gambar 4.3 Pohon Keputusan

Gambar 4.3 merupakan gambar pohon keputusan yang merupakan output dari proses klasifikasi menggunakan algoritma C4.5. Pohon keputusan terbentuk berdasarkan node. Node dalam pohon keputusan merupakan variabel-variabel yang digunakan dalam penelitian. Berdasarkan gambar 4.3 maka faktor dari variabel yang paling dominan dalam mempengaruhi klasifikasi penerima beasiswa yaitu kepemilikan rumah. Selain pohon keputusan seperti pada gambar 4.3 juga terdapat text view yang menjelaskan berdasarkan pohon keputusan seperti dibawah ini

Tree

Tree

blood_pressure > 1.500

| blood_pressure > 2.500

| | sleep_quality > 1.500

| | | anxiety_level > 0.500

| | | | self_esteem > 28.500

| | | | | noise_level > 0.500: 0 {1=0, 2=0, 0=8}

| | | | | noise_level ≤ 0.500: 1 {1=2, 2=0, 0=0}

| | | | self_esteem ≤ 28.500

| | | | | depression > 0.500

| | | | | | depression > 2

| | | | | | | self_esteem > 0.500

| | | | | | | | self_esteem > 5.500: 0 {1=30, 2=22, 0=39}

| | | | | | | | self_esteem ≤ 5.500: 1 {1=10, 2=10, 0=1}

| | | | | | | | self_esteem ≤ 0.500: 0 {1=0, 2=1, 0=3}

| | | | | | | depression ≤ 2: 2 {1=0, 2=4, 0=1}

| | | | | | depression ≤ 0.500: 2 {1=1, 2=4, 0=0}

| | | anxiety_level ≤ 0.500: 2 {1=0, 2=2, 0=0}

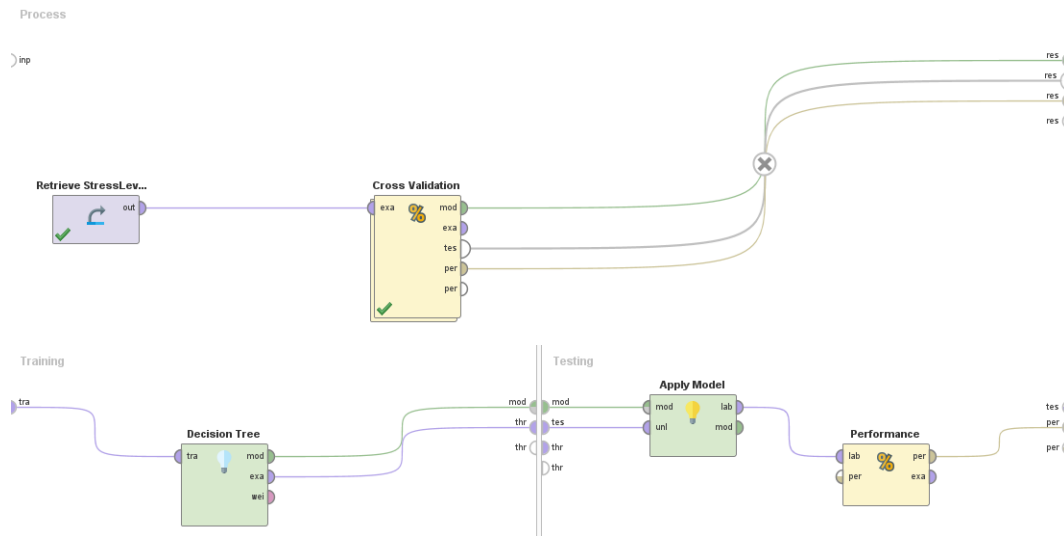
| | sleep_quality ≤ 1.500

| | | teacher_student_relationship > 0.500

| | | | self_esteem > 28.500: 0 {1=0, 2=1, 0=3}
 | | | | self_esteem ≤ 28.500
 | | | | | basic_needs > 3.500
 | | | | | | future_career_concerns > 0.500
 | | | | | | | breathing_problem > 2.500
 | | | | | | | | headache > 1.500: 1 {1=2, 2=0, 0=1}
 | | | | | | | | headache ≤ 1.500: 0 {1=0, 2=0, 0=4}
 | | | | | | | | breathing_problem ≤ 2.500: 2 {1=0, 2=5, 0=1}
 | | | | | | | | future_career_concerns ≤ 0.500: 1 {1=2, 2=0, 0=0}
 | | | | | | basic_needs ≤ 3.500
 | | | | | | | noise_level > 0.500
 | | | | | | | | bullying > 0.500
 | | | | | | | | | living_conditions > 4.500: 1 {1=1, 2=1, 0=1}
 | | | | | | | | | living_conditions ≤ 4.500: 2 {1=2, 2=315, 0=1}
 | | | | | | | | | | bullying ≤ 0.500: 0 {1=1, 2=0, 0=2}
 | | | | | | | | | | noise_level ≤ 0.500
 | | | | | | | | | | | depression > 25: 1 {1=1, 2=1, 0=0}
 | | | | | | | | | | | depression ≤ 25: 0 {1=0, 2=0, 0=2}
 | | | | teacher_student_relationship ≤ 0.500
 | | | | | anxiety_level > 3.500
 | | | | | | self_esteem > 19.500: 1 {1=4, 2=1, 0=0}
 | | | | | | self_esteem ≤ 19.500: 0 {1=2, 2=0, 0=6}
 | | | | | | | anxiety_level ≤ 3.500: 2 {1=0, 2=2, 0=0}
 | | blood_pressure ≤ 2.500: 0 {1=0, 2=0, 0=300}
 | blood_pressure ≤ 1.500: 1 {1=300, 2=0, 0=0}

4.3.2 Algoritma C4.5 Dengan *Cross Validation*

Penerapan algoritma C4.5 pada rapidminer dengan menggunakan *Cross validation* dengan nilai akurasi, precision, confusion matrix atau nilai recall, alur pada rapidminer dapat dilihat pada gambar berikut.



Gambar 4.4 Proses C4.5 dan *Cross Validation*

Pengujian dilakukan dengan confusion matrix yang terdiri dari accuracy, precision dan recall dilakukan pada dataset sebanyak 1101 data yang diolah dengan menggunakan C4.5. Pengujian *confusion matrix* untuk dataset yang diolah menggunakan C4.5 dapat dilihat pada gambar dibawah ini.

accuracy: 88.64% +/- 3.67% (micro average: 88.64%)

	true 1	true 2	true 0	class precision
pred. 1	311	11	16	92.01%
pred. 2	17	322	15	90.96%
pred. 0	30	36	342	83.82%
class recall	86.87%	87.26%	91.69%	

Gambar 4.5 Hasil Proses C4.5

Pengujian *confusion matrix* untuk dataset yang diolah menggunakan C4.5 dapat dilihat pada tabel dibawah ini.

Tabel 4.1 Hasil dari C4.5 dan *Cross Validation*

No	C4.5	Hasil Nilai
1	<i>Accuracy</i>	88,64%
2	<i>Precision</i>	90,96%
3	<i>Recall / Confusion Matrix</i>	87,26%

Nilai akurasi dari *confusion matrix* tersebut adalah sebagai berikut:

Menghitung *Accuracy* : Akurasi = $\frac{(TN+TP)}{(TN+FN+FP+TP)}$

$$\begin{aligned} \text{Akurasi} &= \frac{(311 + 322 + 342)}{(311 + 322 + 342 + 17 + 11 + 30 + 11 + 36 + 16)} \\ &= \frac{975}{1096} \\ &= 88,64\% \end{aligned}$$

Menghitung *Precision*, Rumus : $P:precision = \frac{TP}{(TP+FP)}$

$$\begin{aligned} \text{Presisi} &= \frac{322}{(322+15+17)} \\ &= \frac{322}{354} \\ &= 90,96\% \end{aligned}$$

Menghitung *Recall*, Rumus : $Recall = \frac{TP}{(TP+FN)}$

$$\begin{aligned} \text{Reccal} &= \frac{322}{(322+11+36)} \\ &= \frac{322}{369} \\ &= 87,26\% \end{aligned}$$

4.3.3 Atribut Sebelum Dilakukan PSO

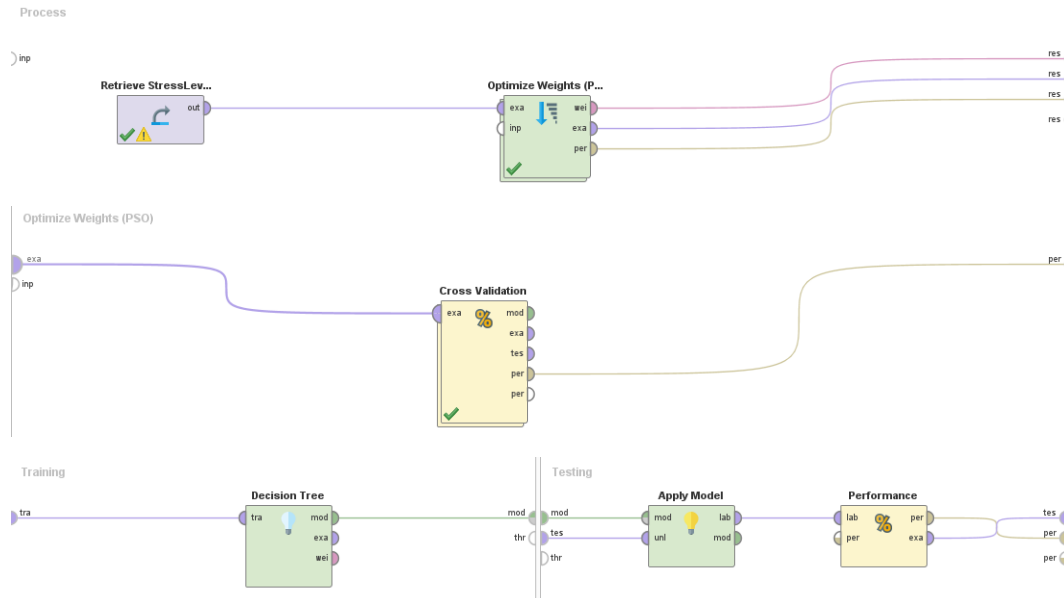
attribute	weight
anxiety_l...	0.041
sleep_q...	0.040
peer_pre...	0.031
self_est...	0.221
academi...	0.045
teacher_...	0.039
blood_pr...	0.133
headache	0.131
depressi...	0.121
mental_...	0.107
living_co...	0.091

Gambar 4.6 Atribut Sebelum Dilakukan PSO

Atribut sebelum dilakukan proses PSO maka atribut tersebut belum ada nilai maksimal yang mengidentifikasi atribut sangat berpengaruh. Nilai yang dihasilkan sebelum proses PSO hampir rata dan sama. Maka dari itu nantinya dilakukan proses PSO untuk melihat atribut yang terseleksi mempunyai nilai terbesar dan berpengaruh.

4.3.4 *Patricle Swarm Optimization (PSO)* dan C4.5

Metode ini adalah untuk meningkatkan akurasi dalam klasifikasi dengan menggunakan Teknik seleksi fitur pembobotan atribut dengan menggunakan PSO dengan C4.5. Penerapan metode ini dapat dilihat pada gambar dibawah ini.



Gambar 4.7 Proses PSO dan C4.5

Pengujian dilakukan dengan confusion matrix yang terdiri dari accuracy, precision dan recall dilakukan pada dataset sebanyak 1101 data yang diolah dengan menggunakan C4.5 + PSO. Pengujian confusion matrix untuk dataset yang diolah menggunakan C4.5 dapat dilihat pada gambar dibawah ini.

accuracy: 89.18% +/- 2.86% (micro average: 89.18%)

	true 1	true 2	true 0	class precision
pred. 1	325	19	23	88.56%
pred. 2	9	323	17	92.55%
pred. 0	24	27	333	86.72%
class recall	90.78%	87.53%	89.28%	

Gambar 4.8 Hasil Proses C4.5 + PSO

Pengujian *confusion matrix* untuk dataset yang diolah menggunakan C4.5 dapat dilihat pada tabel dibawah ini

Tabel 4.2 Hasil dari C4.5 + PSO

No	C4.5 + PSO	Hasil Nilai
1	Accuracy	89,18%
2	Precision	92,55%
3	Recall / Confusion Matrix	87,53%

4.3.5 Atribut Setelah Dilakukan PSO

attribute	weight
anxiety_l...	1
self_est...	0.910
mental_...	1
depressi...	1
headache	0.806
blood_pr...	1
sleep_q...	1
breathin...	0.355
noise_le...	0
living_co...	0.673
safety	0.111
basic_n...	0
academi...	0
study_lo...	0
teacher_...	0.273
future_c...	0
social_s...	0.633
peer_pre...	0
extracurri...	1
bullying	0

Gambar 4.9 Atribut Setelah dilakukan PSO

Atribut setelah dilakukan proses PSO maka terseleksi beberapa atribut yang mempunyai nilai maksimal yaitu “1”. Dari nilai maksimal tersebut maka ada 6 atribut paling mempengaruhi diantaranya *anxiety_level*, *mental_health_history*, *depression*, *blood_pressure*, *sleep_quality* dan *extracurricular_activities*.

4.3.6 Hasil Pengolahan Dataset

Dari tabel yang diberikan, kita memiliki dua set nilai untuk metrik evaluasi kinerja model: satu menggunakan metode C4.5 saja dan yang lainnya menggunakan kombinasi antara metode C4.5 dan *PSO (Particle Swarm Optimization)*.

Berikut adalah kesimpulan dari tabel tersebut:

1. *Accuracy* (Akurasi):

- Dalam metode C4.5, akurasi mencapai 88,64%.
- Ketika menggunakan kombinasi C4.5 + PSO, akurasi meningkat menjadi 89,18%.
- Kesimpulannya, penggunaan PSO menghasilkan peningkatan yang kecil tetapi signifikan dalam akurasi model.

2. *Precision* (Presisi):

- Presisi model C4.5 saja adalah 90,96%.
- Dengan penambahan PSO, presisi meningkat menjadi 92,55%.
- Dengan adanya PSO, model menjadi lebih presisi dalam mengklasifikasikan hasil positif.

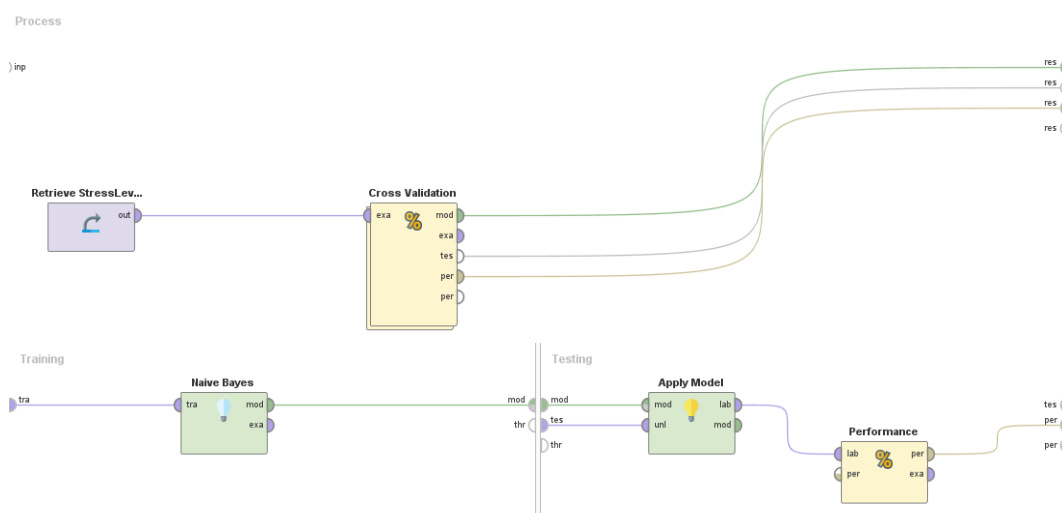
3. *Recall / Confusion Matrix* (Recall / Matriks Kebingungan):

- *Recall* untuk metode C4.5 adalah 87,26%.
- Dengan PSO, recall meningkat sedikit menjadi 87,53%.
- Hal ini menunjukkan bahwa penambahan PSO memperbaiki kemampuan model dalam mengingat kelas yang sebenarnya.

Dengan demikian, dapat disimpulkan bahwa penggunaan PSO sebagai tambahan pada C4.5 telah memberikan peningkatan kinerja model dalam hal akurasi, presisi, dan *recall*. Meskipun peningkatannya tidak signifikan, namun setiap peningkatan performa model dalam tugas klasifikasi dapat memberikan dampak yang positif dalam aplikasi praktisnya.

4.3.7 Algoritma Naïve Bayes Dengan Cross Validation

Penerapan algoritma Naïve Bayes pada rapidminer dengan menggunakan *cross validation* dengan nilai akurasi, *precision*, *confusion matrix* atau nilai *recall*, alur pada rapidminer dapat dilihat pada gambar berikut.



Gambar 4.10 Proses Naïve Bayes

Pengujian *confusion matrix* untuk dataset yang diolah menggunakan algoritma Naïve Bayes dapat dilihat pada gambar dibawah ini.

accuracy: 88.27% +/- 4.19% (micro average: 88.27%)

	true 1	true 2	true 0	class precision
pred. 1	304	2	2	98.70%
pred. 2	44	352	56	77.88%
pred. 0	10	15	315	92.65%
class recall	84.92%	95.39%	84.45%	

Gambar 4.11 Hasil Proses Naïve Bayes

Pengujian *confusion matrix* untuk dataset yang diolah menggunakan Naïve Bayes dapat dilihat pada tabel dibawah ini.

Tabel 4.3 Hasil dari Naïve Bayes

No	Naïve Bayes	Hasil Nilai
1	<i>Accuracy</i>	88,27%
2	<i>Precision</i>	77,88%
3	<i>Recall / Confusion Matrix</i>	95,39%

Nilai akurasi dari *confusion matrix* tersebut adalah sebagai berikut:

$$\text{Menghitung Accuracy : Akurasi} = \frac{(TN+TP)}{(TN+FN+FP+TP)}$$

$$\text{Akurasi} = \frac{(304 + 352 + 315)}{(304 + 352 + 315 + 44 + 2 + 10 + 2 + 15)}$$

$$= \frac{971}{1044}$$

$$= 88,27\%$$

Menghitung *Precision* : $Precision = \frac{TP}{(TP+FP)}$

$$:Precision = \frac{352}{(352+56+44)}$$

$$= \frac{352}{456}$$

$$= 77,88\%$$

Menghitung *Recall* : $Recall = \frac{TP}{(TP+FN)}$

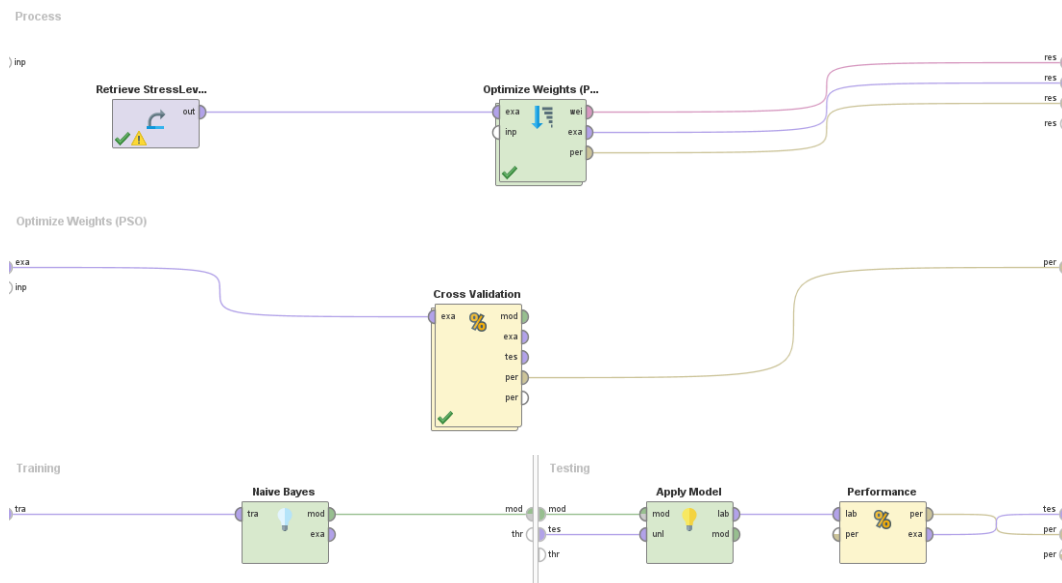
$$Recall = \frac{352}{(352+15+2)}$$

$$= \frac{352}{369}$$

$$= 95,39\%$$

4.3.8 Particle Swarm Optimization (PSO) dan Naïve Bayes

Metode ini adalah untuk meningkatkan akurasi dalam klasifikasi dengan menggunakan Teknik seleksi fitur pembobotan atribut dengan menggunakan PSO dengan Naïve Bayes serta perbandingan *Cross Validation*. Penerapan metode ini dapat dilihat pada gambar dibawah ini.



Gambar 4.12 Proses Naïve Bayes + PSO

Pengujian *confusion matrix* untuk dataset yang diolah menggunakan algoritma Naïve Bayes dapat dilihat pada gambar dibawah ini.

accuracy: 89.55% +/- 2.88% (micro average: 89.55%)

	true 1	true 2	true 0	class precision
pred. 1	305	0	1	99.67%
pred. 2	45	359	51	78.90%
pred. 0	8	10	321	94.69%
class recall	85.20%	97.29%	86.06%	

Gambar 4.13 Hasil Proses Naïve Bayes + PSO

Pengujian *confusion matrix* untuk dataset yang diolah menggunakan Naïve Bayes dapat dilihat pada tabel dibawah ini.

Tabel 4.4 Hasil dari Naïve Bayes + PSO

No	NaïveByaes dan PSO	Hasil Nilai
1	<i>Accuracy</i>	89.55%
2	<i>Precision</i>	78.90%
3	<i>Recall / Confusion Matrix</i>	97,29%

Berdasarkan tabel yang diberikan, terdapat perbandingan antara hasil dari dua metode: Naïve Bayes dan kombinasi antara Naïve Bayes dan *PSO* (*Particle Swarm Optimization*). Berikut adalah kesimpulan dari perbandingan tersebut:

1. *Accuracy* (Akurasi): Naïve Bayes: 88,27%, Naïve Bayes dan PSO: 89,55%
Kesimpulan: Kombinasi antara Naïve Bayes dan PSO memberikan peningkatan akurasi yang lebih tinggi dibandingkan dengan Naïve Bayes tunggal.
2. *Precision* (Presisi): Naïve Bayes: 77,88%, Naïve Bayes dan PSO: 78,90%
Kesimpulan: Terdapat peningkatan kecil dalam tingkat presisi saat menggunakan PSO bersama dengan Naïve Bayes. Namun, peningkatan ini tidak signifikan.
3. *Recall / Confusion Matrix* (*Recall* / Matriks Kebingungan): Naïve Bayes: 95,39% Naïve Bayes dan PSO: 97,29%.

Dari kesimpulan di atas, bahwa penggunaan PSO bersama dengan Naïve Bayes berhasil meningkatkan akurasi dan recall dalam klasifikasi data. Namun, presisi mengalami peningkatan yang kurang signifikan. Hal ini menunjukkan bahwa penggunaan PSO dapat memberikan nilai tambah terutama dalam meningkatkan kemampuan model untuk mengidentifikasi instance positif yang relevan.

4.1 Laporan

Dalam analisis ini, peneliti membandingkan kinerja dua metode klasifikasi utama, yaitu C4.5 dan Naïve Bayes, baik secara tunggal maupun dengan pengoptimalan menggunakan *Particle Swarm Optimization (PSO)*. Evaluasi dilakukan berdasarkan tiga metrik utama: Akurasi, Presisi, dan *Recall* (dalam bentuk *Confusion Matrix*). Data yang digunakan untuk evaluasi telah dipecah menjadi data pelatihan dan data pengujian.

C4.5 vs. C4.5 dengan PSO

1. C4.5:
 - Akurasi: 88,64%
 - Presisi: 90,96%
 - Recall: 87,26%

2. C4.5 dengan PSO:

- Akurasi: 89,18%
- Presisi: 92,55%
- Recall: 87,53%

Analisis:

Kombinasi C4.5 dengan PSO menunjukkan peningkatan kecil dalam akurasi, namun peningkatan signifikan terlihat pada tingkat Presisi dan Recall. PSO mampu meningkatkan kemampuan C4.5 dalam mengklasifikasikan instance positif secara tepat, meskipun pengaruhnya terhadap akurasi keseluruhan relatif kecil.

Naïve Bayes vs. Naïve Bayes dengan PSO

1. Naïve Bayes:

- Akurasi: 88,27%
- Presisi: 77,88%
- Recall: 95,39%

2. Naïve Bayes dengan PSO:

- Akurasi: 89,55%
- Presisi: 78,90%
- Recall: 97,29%

Analisis:

Penggunaan PSO bersama dengan Naïve Bayes meningkatkan akurasi secara signifikan. Terdapat peningkatan yang terlihat dalam Presisi dan *Recall* saat menggunakan Naïve Bayes dengan PSO, menunjukkan PSO berkontribusi pada peningkatan kemampuan model dalam mengenali instance positif.

Tabel 4.5 Hasil dari PSO, C4.5 dan Naïve Bayes

No	Keterangan	Naïve Bayes	Naïve Bayes +PSO	C.45	C4.5+PSO
1	<i>Accuracy</i>	88,27%	89,55%	88,64%	89,18%
2	<i>Precision</i>	77,88%	78,90%	90,96%	92,55%
3	<i>Recall</i>	95,39%	97,29%	87,26%	87,53%

Dari tabel yang diberikan, terdapat perbandingan antara empat skenario yang melibatkan penggunaan algoritma *Machine Learning* (Naïve Bayes dan C4.5) dengan dan tanpa pengoptimalan menggunakan *Particle Swarm Optimization (PSO)*. Berikut adalah analisis mengapa PSO dapat meningkatkan akurasi, presisi, dan *recall* dari model-model tersebut:

1. Naïve Bayes vs. Naïve Bayes + PSO:

1. Accuracy: Naïve Bayes + PSO memberikan akurasi yang sedikit lebih tinggi (89,55% vs. 88,27%). Ini bisa disebabkan oleh PSO membantu menyesuaikan parameter-model Naïve Bayes dengan lebih baik, misalnya, memilih distribusi probabilitas yang lebih sesuai dengan data.
2. Precision: Kenaikan presisi dari Naïve Bayes + PSO (78,90% vs. 77,88%) menunjukkan bahwa PSO membantu mengurangi false positive, membuat model lebih tepat dalam mengklasifikasikan kelas yang positif.
3. Recall: Naïve Bayes + PSO memiliki recall yang lebih tinggi (97,29% vs. 95,39%), menandakan bahwa model ini lebih baik dalam mengidentifikasi sebagian besar instance positif.

2. C4.5 dan C4.5 + PSO:

1. Accuracy: C4.5 + PSO memiliki akurasi yang sedikit lebih tinggi (89,18% vs. 88,64%). PSO mungkin membantu dalam pemilihan atribut dan pembentukan aturan yang lebih optimal dalam pembangunan pohon keputusan.
2. Precision: Kenaikan presisi dari C4.5 + PSO (92,55% vs. 90,96%) menunjukkan bahwa PSO membantu dalam mengurangi false positive, membuat keputusan yang lebih tepat.
3. Recall: Meskipun C4.5 + PSO memiliki recall yang sedikit lebih rendah (87,53% vs. 87,26%), namun perbedaannya tidak signifikan. Hal ini bisa disebabkan oleh penyesuaian yang dilakukan oleh PSO terhadap parameter-parameter algoritma C4.5.

Alasan utama mengapa PSO dapat meningkatkan kinerja model adalah karena kemampuannya dalam menemukan solusi-solusi yang lebih optimal

dalam ruang pencarian parameter-parameter model. PSO dapat menyesuaikan parameter-parameter ini secara dinamis berdasarkan evaluasi kinerja model pada setiap iterasi, sehingga memungkinkan model untuk mencapai titik optimal yang lebih baik daripada menggunakan parameter default atau pemilihan parameter secara manual. Selain itu, PSO juga dapat membantu dalam mengatasi masalah overfitting atau underfitting dengan memperbaiki generalisasi model melalui optimasi parameter-parameter yang relevan dengan kompleksitas model. Hal ini terutama terlihat dari peningkatan presisi dan recall pada model-model yang dioptimalkan dengan PSO.