

BAB IV

HASIL DAN PEMBAHASAN

Pada bab ini penulis akan membahas mengenai perhitungan data yang digunakan. Data tersebut akan dihitung menggunakan seleksi fitur dengan menggunakan algoritma naïve bayes.

4.1. Persiapan Data

Data yang digunakan dalam penelitian ini adalah dataset dari Kaggle yaitu data pasien covid -19 (<https://www.kaggle.com/hemanthhari/symptoms-and-covid-presence>).

Jumlah record data sebanyak 5.434 data yang merupakan banyaknya pasien yang mengalami gejala dan jumlah atribut dataset sebanyak 21 atribut yang merupakan gejala-gejala yang dialami pasien covid 20. Atribut-atribut tersebut dijelaskan pada table 4.1 dibawah ini:

Tabel 4.1 Dataset Penelitian

No	Breathing Problem	Fever	Dry Cough	Sore throat	Running Nose	Asthma	Chronic Lung Disease	Headache	Heart Disease	Diabetes	Hypertension	Fatigue	Gastrointestinal	Abroad travel	Contact with COVID Patient	Attended Large Gathering	Visited Public Exposed Places	Family working in Public Exposed Places	Wearing Masks	Sanitization from Market	COVID-19
1	Yes	Yes	Yes	Yes	Yes	No	No	No	No	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes	No	No	Yes
2	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	No	Yes	No	No	No	Yes	Yes	No	No	No	Yes
3	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	No	No	No	No	No	No	Yes
4	Yes	Yes	Yes	No	No	Yes	No	No	Yes	Yes	No	No	No	Yes	No	Yes	Yes	No	No	No	Yes
5	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No	Yes	No	No	No	Yes
6	Yes	Yes	Yes	No	No	No	No	No	Yes	No	Yes	No	No	No	No	No	No	No	No	No	Yes
7	Yes	Yes	Yes	No	No	No	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No	Yes
8	Yes	Yes	Yes	No	Yes	Yes	No	No	No	Yes	Yes	No	Yes	Yes	No	No	Yes	No	No	No	Yes
9	Yes	Yes	Yes	No	Yes	No	Yes	No	No	Yes	No	Yes	No	Yes	Yes	Yes	No	No	No	No	Yes

10	Yes	Yes	Yes	No	No	Yes	No	No	No	Yes	Yes	Yes	No	No	No	No	Yes	No	No	No	Yes
----	-----	-----	-----	----	----	-----	----	----	----	-----	-----	-----	----	----	----	----	-----	----	----	----	-----

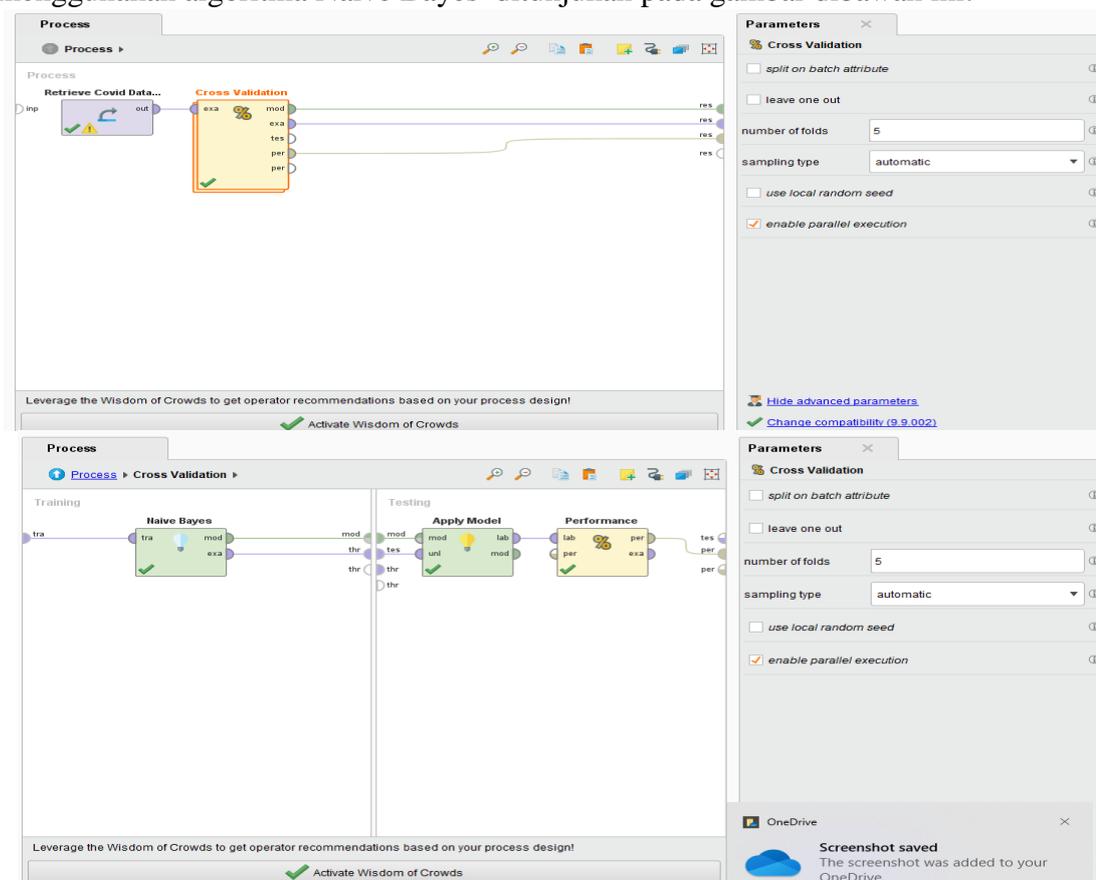
Tabel diatas menunjukkan karakter data pada penelitian ini, data tersebut memiliki label pada atribut covid-19 yang berarti data ini merupakan data supervised karena memiliki label, label pada data ini adalah Yes/No Covid 19. Alat yang digunakan pada penelitian ini adalah aplikasi Rapidminer.

4.2. Pemodelan

Pemilihan dan penerapan teknik pemodelan yang sesuai dilakukan pada tahapan ini.. Pemodelan pada penelitian ini menggunakan teknik data mining klasifikasi.

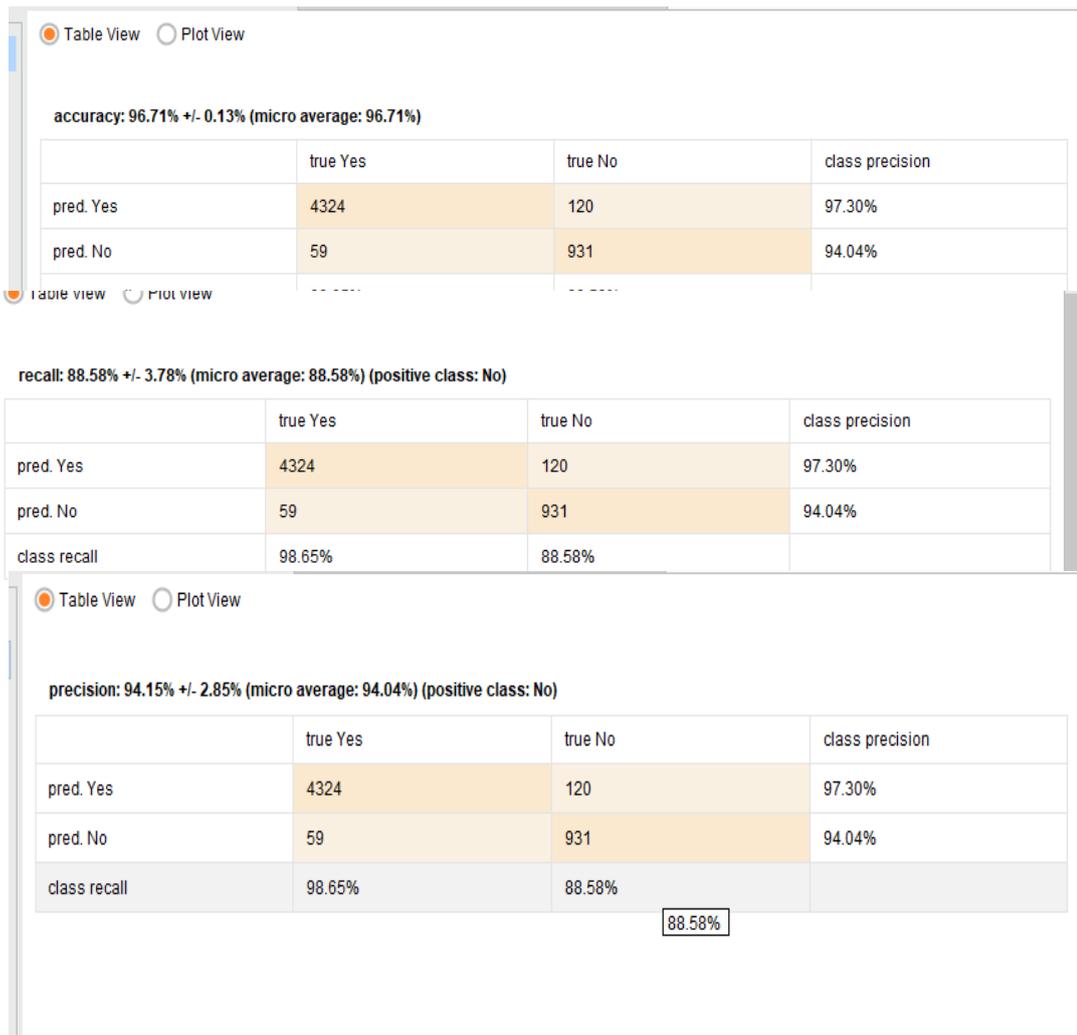
4.2.1. Penelitian Menggunakan Algoritma Naïve Bayes

Penerapan data pada Rapidminer untuk Klasifikasi Penyakit Covid 19 menggunakan algoritma Naïve Bayes ditunjukkan pada gambar dibawah ini:



Gambar 4.1 Penerapan Data Klasifikasi Covid 19 Menggunakan Algoritma Naïve Bayes pada Rapidminer

Pada gambar 4.1 data yang sudah disiapkan diterapkan pada aplikasi rapidminer kalsifikasi dengan melakukan eksperimen menggunakan cross validation yang secara langsung dapat membagi data menjadi data training dan data testing karena data yang digunakan adalah supervised dan algoritma yang digunakan adalah naïve bayes. Hasil eksperimen dapat kita lihat pada gambar 4.2 dibawah ini.



Gambar 4.2 adalah confusion matrix yang menunjukkan hasil eksperimen, didalam confusion matrix kita dapat melihat hasil akurasi, class presisi, dan class recall. Akurasi yang dihasilkan adalah 96,71 % class recall 88,58 % dan class presisi 94,15 %. Hasil akurasi pada penelitian sebelumnya dengan eksperimen penelitian ini dengan menggunakan algoritma naïve bayes untuk klasifikasi penyakit covid 19 ditunjukkan pada tabel dibawah ini:

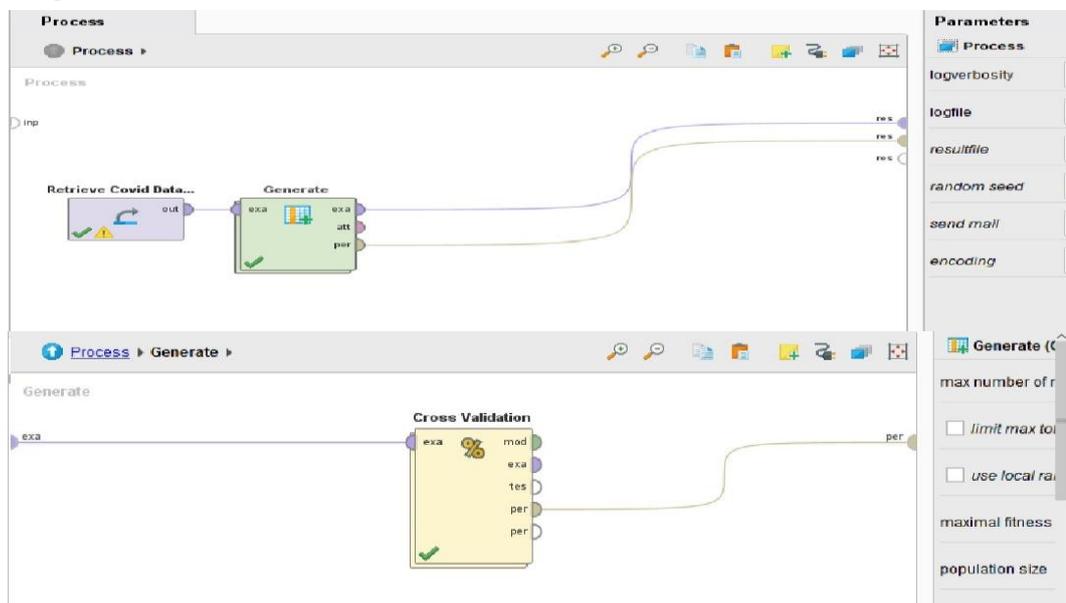
Tabel. 4.2 Perbandingan Hasil akurasi Penelitian sebelumnya dengan penelitian ini:

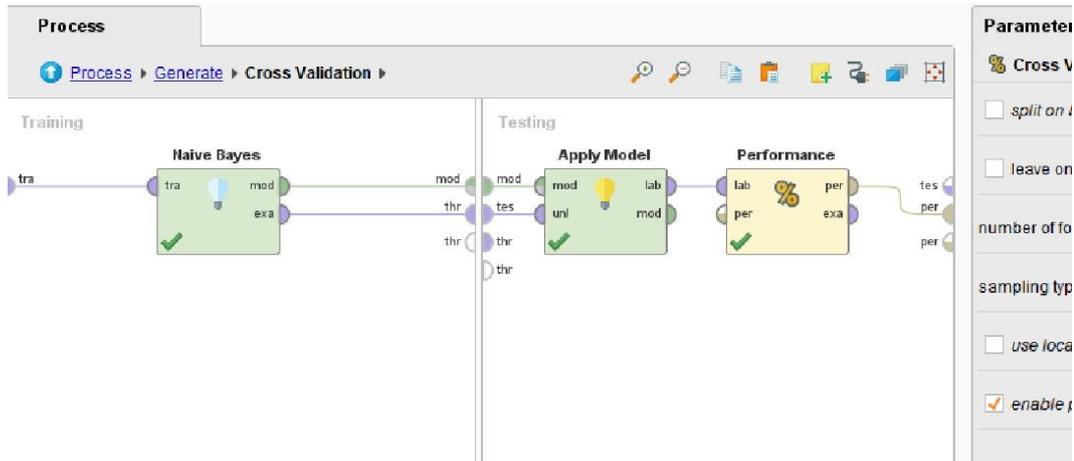
Penelitian	Tingkat Akurasi
Penelitian sebelumnya	96,67%
Penelitian saat Ini	96,71 %

Eksperimen penelitian ini dengan pengujian K-Fold 2 pada cross validation menghasilkan tingkat akurasi paling tinggi dan tingkat akurasi lebih tinggi dari penelitian sebelumnya yang dilakukan oleh Dewi Ynati Liliana dkk, akan tetapi masih dapat ditingkatkan tingkat akurasinya.

4.2.2. Penelitian Menggunakan Genetic Algorithm (GA) dan Algoritma Naïve Bayes

Metode yang dapat meningkatkan tingkat akurasi dalam klasifikasi adalah penggunaan teknik optimasi. Eksperimen lanjutan yang dilakukan pada penelitian ini adalah penggunaan teknik optimasi seleksi fitur menggunakan metode Genetic Algorithm (GA)/ Algoritma Genetika. Penerapan pada rapidminer klasifikasi penyakit Covid 19 menggunakan GA dan Algoritma Naïve Bayes dapat kita lihat pada gambar 4.3 dibawah ini:





Gambar 4.3 Penerapan Rapidminer menggunakan GA dan Naïve Bayes

Ekperimen pada gambar 4.3 teknik validasi untuk membagi data training dan data testing menggunakan teknik validasi cross validation. Dari eksperimen tersebut menndapatkan hasil yang dapat kita lihat pada gambar 4.4.

Criterion: accuracy, precision, recall, AUC (optimistic), AUC (pessimistic)

Table View Plot View

accuracy: 95.42% +/- 0.39% (micro average: 95.42%)

	true Yes	true No	class precision
pred. Yes	4292	158	96.45%
pred. No	91	893	90.75%
class recall	97.92%	84.97%	

Criterion: accuracy, precision, recall, AUC (optimistic), AUC (pessimistic)

Table View Plot View

recall: 84.97% +/- 3.26% (micro average: 84.97%) (positive class: No)

	true Yes	true No	class precision
pred. Yes	4292	158	96.45%
pred. No	91	893	90.75%
class recall	97.92%	84.97%	

Criterion: accuracy, precision, recall, AUC (optimistic), AUC (pessimistic)

Table View Plot View

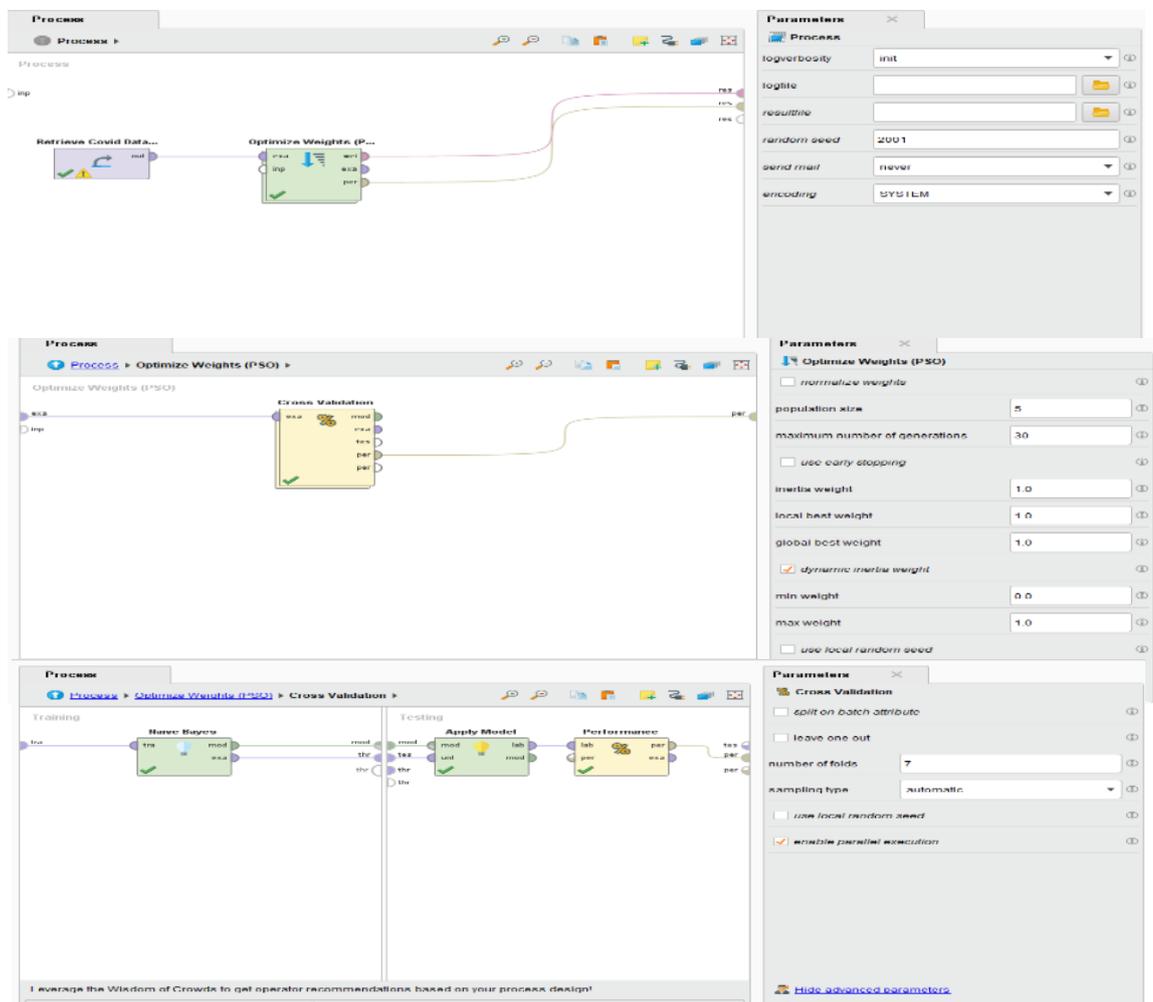
	true Yes	true No	class precision
pred. Yes	4292	158	96.45%
pred. No	91	893	90.75%
class recall	97.92%	84.97%	

Gambar 4.4 Confision Matrix Hasil Klasifikasi Covid 19 Menggunakan Metode GA Dan NB

Hasil dari eksperimen dapat diketahui number of fold yang menghasilkan baik akurasi, recall dan precision terbaik adalah penggunaan number of fold 3. Dimana hasil akurasi 95,42 %, Recall 84,97 %, Precision 90,81 %. Hasil yang didapatkan pada ekperimen ini tidak lebih baik dari ekperimen yang hanya menggunakan algoritma naïve bayes.

4.2.3. Penelitian Menggunakan Particle Swarm Optimization (PSO) dan Algoritma Naïve Bayes

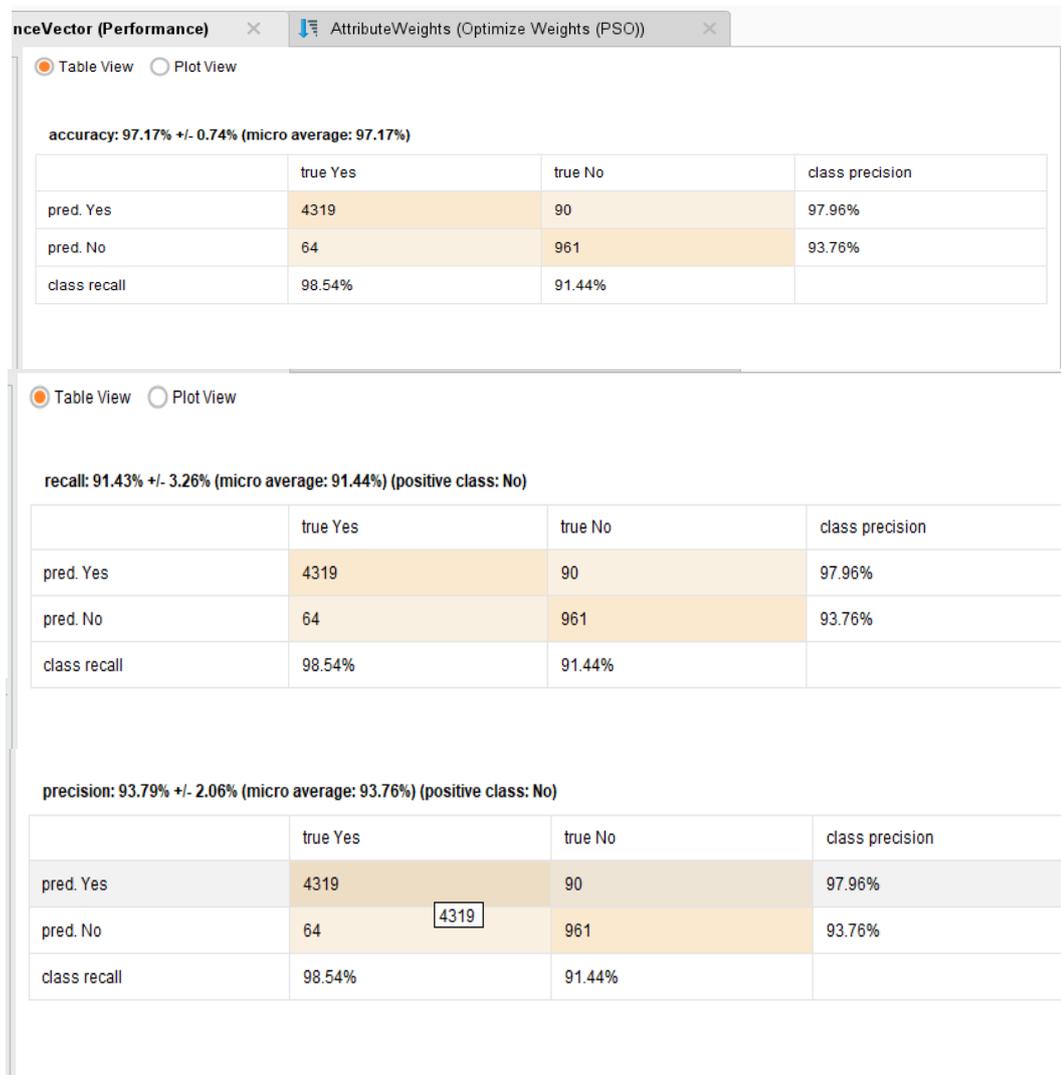
Metode optimasi seleksi fitur yang dapat dilakukan selain GA adalah metode PSO dengan algoritma yang sama yaitu naïve bayes. Penerapan Metode dan algoritma pada rapidminer ditunjukkan pada gambar 4.5



Gambar 4.5 Penerapan Data Klasifikasi Covid 19 Menggunakan Metode PSO dan Algoritma Naïve Bayes pada Rapidminer

Pada gambar 4.5 data yang sudah disiapkan diterapkan pada aplikasi rapidminer dalam mengkalsifikasi penyakit covid 19 dengan melakukan

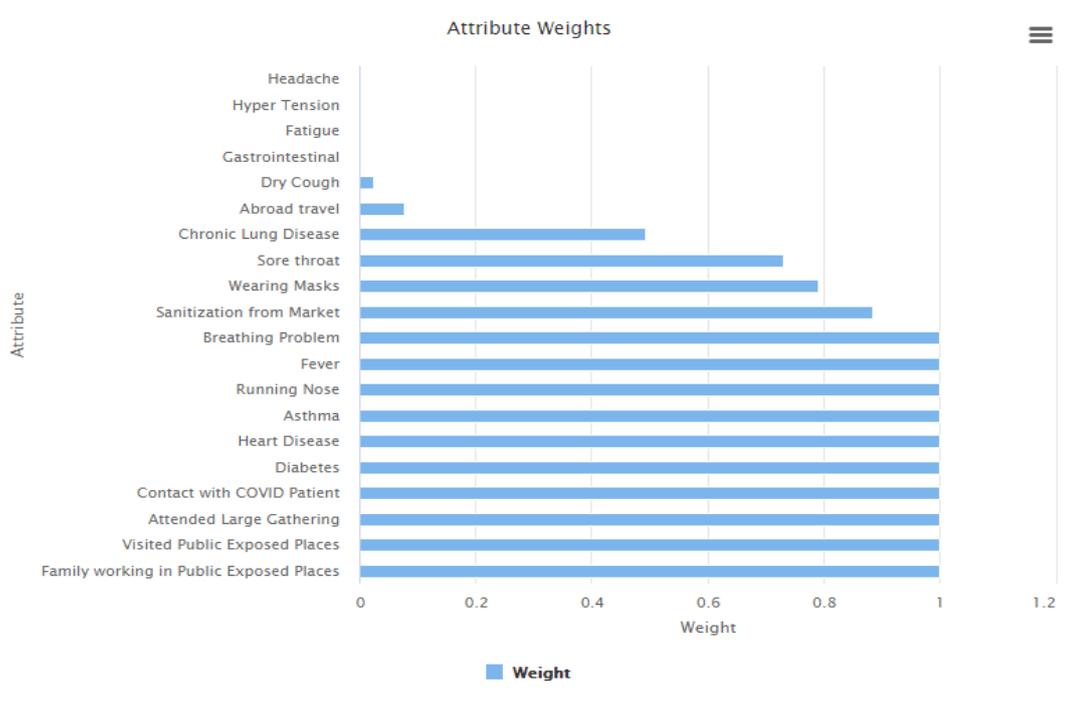
eksperimen menggunakan cross validation dan penggunaan teknik optimasi seleksi fitur Particle Swarm Optimization (PSO) dan algoritma yang digunakan tetap sama adalah naïve bayes. Hasil eksperimen dapat kita lihat pada gambar 4.6



4.6. Confusion Matrix Hasil Klasifikasi Covid 19 Menggunakan Metode PSO dan Algoritma Naïve Bayes pada Rapidminer

Gambar 4.6 adalah confusion matrix yang menunjukkan hasil eksperimen lanjutan, didalam confusion matrix kita dapat melihat hasil akurasi, class presisi, dan class recall. Akurasi yang dihasilkan adalah 97,17 % presisi 93,79 % . dan recall 91,43 % .

Merujuk dari model proses yang telah dilakukan, maka dapat diperoleh data seleksi pembobotan atribut pada gambar 4.7



Gambar 4.7 Hasil Seleksi Atribut

Dari gambar diatas dapat dilihat hasil pembobotan atribut-atribut, terdapat 4 atribut yang memiliki nilai 0 yang artinya atribut tersebut otomatis terseleksi dan tidak perlu digunakan karena tidak memiliki pengaruh dalam proses. Atribut-atribut terpilih untuk selanjutnya diklasifikasi pada tabel 4.3 dibawah ini:

Tabel 4.3. Atribut Terpilih

No	Atribut
1	Dry Cough
2	Abroad Travel
3	Chronic Long Desease
4	Sore Throat
5	Wearing Mask
6	Sanization from Market
7	Breathing Problem
8	Fever

9	Running Nose
10	Asthma
11	Hearth Desease
12	Diabetes
13	Contact With Covid Patient
14	Attended Large Gathering
15	Visited Public Exposed Places
16	Family Working in Public Exposed Places

4.2.4. Perbandingan Penelitian Tidak Menggunakan PSO dan Menggunakan PSO

Perbandingan hasil eksperimen yang tidak menggunakan metode PSO dan eksperimen yang menggunakan PSO ditunjukkan pada tabel dibawah ini:

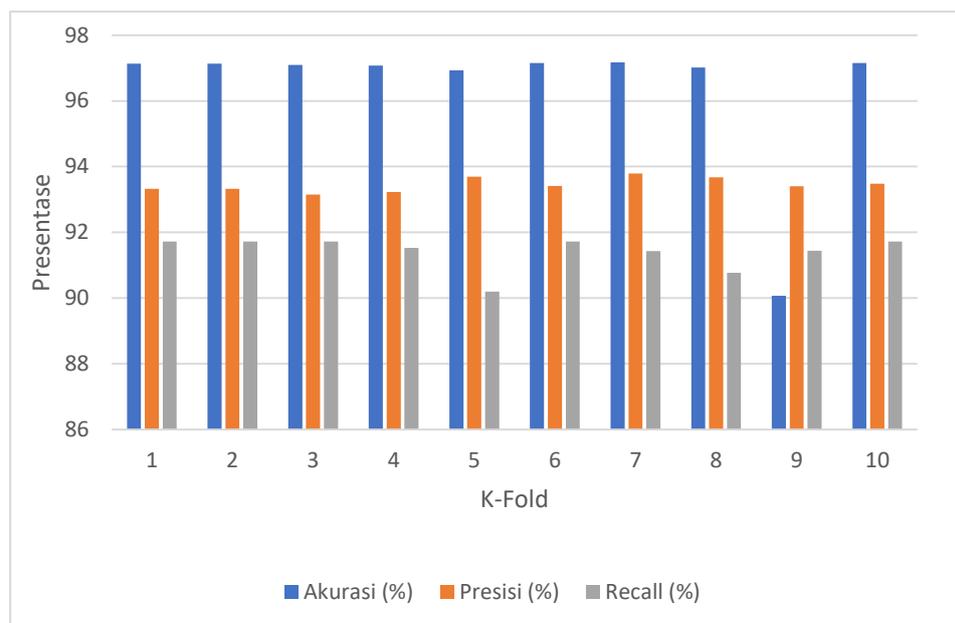
Tabel 4.4 Perbandingan Hasil Akurasi Penggunaan PSO

Penelitian	Tingkat Akurasi
Eksperimen Tidak Menggunakan Metode PSO	96,71%
Eksperimen Yang Menggunakan Metode PSO	97,17 %

Hasil dari eksperimen ini untuk menghasilkan tingkat akurasi yang tinggi sebesar 97,17% menggunakan K-Fold 7, ini merupakan hasil akurasi paling tinggi dalam pengujian. Perbandingan akurasi, dan recall dapat kita lihat pada table dan gambar diagram dibawah ini:

Tabel 4.5 Perbandingan K-Fold Penggunaan PSO+NB

Metode	Presentase Kinerja	K-Fold									
		1	2	3	4	5	6	7	8	9	10
NB+PSO	Akurasi (%)	97,13	97,13	97,09	97,07	96,93	97,15	97,17	97,02	90,07	97,15
	Presisi (%)	93,32	93,32	93,15	93,23	93,69	93,41	93,79	93,67	93,4	93,48
	Recall (%)	91,72	91,72	91,72	91,53	90,2	91,72	91,43	90,77	91,44	91,72



Gambar 4.8 Perbandingan K-Fold Penggunaan NB+PSO

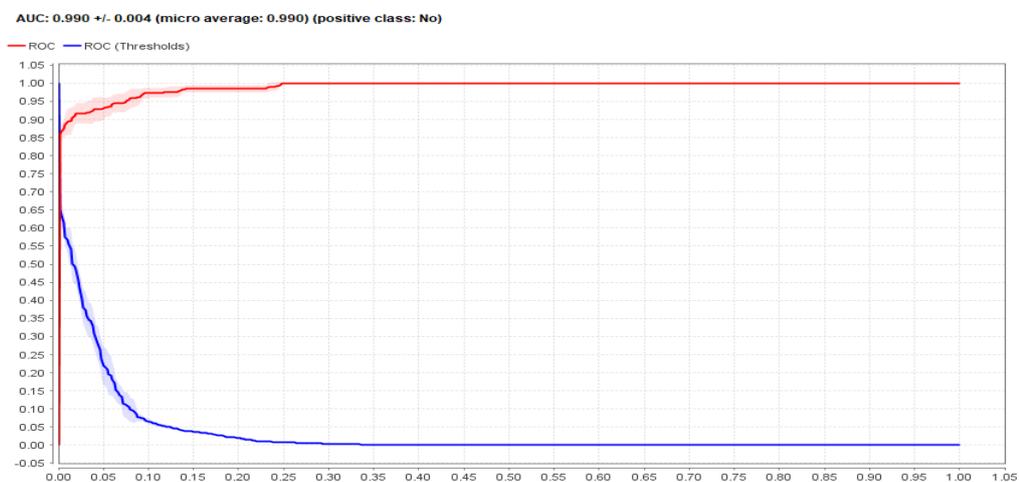
Berdasarkan tabel dan gambar diagram diatas, menunjukkan bahwa perubahan jumlah K-Fold pada Cross Validation dalam melakukan klasifikasi akan menghasilkan akurasi yang berbeda sehingga kita dapat menghasilkan akurasi yang terbaik. Dapat kita lihat pada K-Fold 7 hasil akurasi dan presisi merupakan yang paling tinggi berbeda dengan hasil recall yang tidak lebih tinggi dari penggunaan K-Fold lainnya.

Selain Confusion matrix untuk mengetahui kinerja dari eksperimen ini kita dapat mengandalkan kurva ROC-AUC yang dihasilkan. Perbandingan hasil Kurva ROC-AUC pada penelitian yang tidak menggunakan PSO dan yang



Gambar 4.9 Kurva ROC-AUC Hasil Eksperimen Tidak Menggunakan PSO

Gambar 4.9 menunjukkan kinerja algoritma dalam mengklasifikasi penyakit covid19, kinerja algoritma pada eksperimen ini sangat baik dengan menghasilkan Area Under Curva (AUC) 99% termasuk dalam kategori *Excelent Clasification*..



Gambar 4.10 Kurva ROC-AUC Hasil Eksperimen Menggunakan PSO

Gambar 4.10 menunjukkan kinerja algoritma dalam mengklasifikasi penyakit covid 19 menggunakan Particle Swarm Optimization (PSO), kinerja algoritma pada eksperimen sama baiknya dari eksperimen sebelumnya dengan menghasilkan Area Under Curva (AUC) 99% termasuk dalam kategori *Excelent Clasification*.

4.5. Evaluasi

Pemodelan yang telah dilakukan dengan melakukan eksperimen menggunakan seleksi fitur Algorithm Genetic/ Algorithma Genetika (GA) tidak menghasilkan akurasi yang lebih baik. Setelah penggunaan seleksi fitur Particle Swarm Optimization (PSO) berhasil meningkatkan akurasi dalam klasifikasi penyakit Covid 19 yang menggunakan 20 atribut data. Penggunaan PSO +Naïve Bayes memperoleh akurasi yang tinggi yaitu 97,17 %. Merujuk dari model proses yang telah dilakukan, maka dapat diperoleh data seleksi pembobotan atribut terdapat 4 atribut yang memiliki nilai 0 yang artinya atribut tersebut otomatis terseleksi dan tidak perlu digunakan karena tidak memiliki pengaruh dalam proses. Selain itu kinerja algoritma pada penelitian ini juga menunjukkan performa yang sangat baik sekali termasuk dalam kategori *Excelent Clasification* berdasarkan hasil AUC yang memiliki nilai 99%.