

## **KOMPARASI *FEATURE SELECTION PARTICLE SWARM OPTIMIZATION (PSO)* DENGAN *GENETIC ALGORITHM (GA)* TERHADAP ALGORITMA *NAÏVE BAYES* PADA ANALISIS SENTIMENT *TWITTER***

**Karno Juni Prayoga\*, Agung Nugroho, dan Tri Ngudi Wiyatno**

Fakultas Teknik, Universitas Pelita Bangsa, Kabupaten Bekasi

\*E-mail: karnojuni23@mhs.pelitabangsa.ac.id

### **Abstract**

Sentiment analysis is a field of study that analyzes opinions, views, evaluations, assessments, attitudes, and emotions towards an entity and its aspects expressed through text. Twitter is one of the media that is often used as research objects. The main problem in the sentiment analysis process is how to choose and use the best feature selection to get maximum results. One of the most often used classification methods is Naïve Bayes. Naïve Bayes is very sensitive to feature selection. In the test, a comparison of Particel Swarm Optimization and Genetic algorithm feature selection are performed to improve the accuracy of the Naïve Bayes Algorithm. Measurements were made by comparing the testing before and after using feature selection. Validation used cross validation techniques while accuracy was measured by a confussion matrix. The results showed an increasing in the accuracy of the Naïve Bayes Algorithm by using the Particle Swarm Optimization feature selection from 67.71% to 71.84%. In contrast to increasing the accuracy of the Naïve Bayes Algorithm by using the Genetic Algorithm feature selection from 67.71% to 68.96%. Thus, the best feature selection was Particle Swarm Optimization with an accuracy increase of 4.00%.

**Keywords:** Sentiment Analysis, Twitter, Naïve Bayes, Feature Selection, Particle Swarm Optimization, Genetic Algorithm.

### **Abstrak**

*Analisis sentiment merupakan bidang studi yang menganalisa pendapat, pandangan, evaluasi, penilaian, sikap, dan emosi terhadap entitas dan aspek-aspeknya yang diekspresikan melalui teks. Twitter merupakan salah satu media yang sering dijadikan obyek penelitian. Masalah utama dalam proses analisis sentiment adalah bagaimana memilih dan menggunakan feature selection terbaik untuk mendapatkan hasil maksimal. Salah satu metode klasifikasi yang sering digunakan Naïve Bayes. Naïve Bayes sangat sensitif terhadap feature selection. Pada pengujian dilakukan komparasi feature selection Particle Swarm Optimization dan Genetic Algorithm untuk meningkatkan kinerja akurasi Algoritma Naïve Bayes. Pengukuran dilakukan dengan membandingkan antara pengujian sebelum dan sesudah menggunakan feature selection. Validasi menggunakan Teknik Cross Validation, sedangkan akurasi diukur dengan confussion matrix. Hasilnya menunjukkan peningkatan akurasi Algoritma Naïve*

*Bayes dengan menggunakan feature selection Particle Swarm Optimization dari 67.71% menjadi 71.84%. Berbeda dengan peningkatan akurasi Algoritma Naïve Bayes dengan menggunakan feature selection Genetic Algorithm dari 67.71% menjadi 68.96%. Dengan demikian, feature selection terbaik adalah Particle Swarm Optimization dengan peningkatan akurasi sebesar 4.00%.*

**Kata kunci:** Analisis Sentiment, Twitter, Naïve Bayes, Feature Selection, Particle Swarm Optimization, Genetic Algorithm.

## 1. PENDAHULUAN

Menurut Kementerian dan Komunikasi dan Informatika (Kemenkominfo) mengungkapkan pengguna internet di Indonesia saat ini mencapai 63 juta orang. Dari angka tersebut, 95 persennya menggunakan internet untuk mengakses jejaring sosial. “Hal ini menyebabkan banyaknya jumlah pesan atau *post* di media sosial sangat banyak setiap harinya dengan bermacam variasi. Dengan jumlah *post* yang banyak dan beragam tersebut tentu sangat sulit jika ingin mengetahui opini atau penilaian yang diungkapkan oleh setiap individu. Hal ini dapat diatasi dengan melakukan penyaringan atau mengambil entitas tertentu pada sebuah pesan atau *post* yang biasa disebut ekstraksi informasi.

Ekstraksi informasi merupakan teknik mengumpulkan informasi dari kumpulan teks yang tidak terstruktur. Untuk memperoleh informasi dari data yang tidak terstruktur perlu didefinisikan informasi target sebagai informasi terstruktur yang akan diekstrak (Zy & Nugroho, 2018). Selanjutnya dari data tersebut dapat diambil informasi tentang opini dari pengguna media sosial terhadap entitas tertentu.

Agar semua data opini bermanfaat, maka dapat dilakukan pengolahan data menggunakan analisis sentiment. “*Sentiment Analysis* atau *Opinion Mining* adalah bidang studi yang menganalisa pendapat, pandangan, evaluasi, penilaian, sikap, dan emosi terhadap entitas dan aspek-aspeknya

yang diekspresikan melalui teks”. Tugas dasar dalam analisis sentiment adalah mengelompokkan polaritas dari teks yang ada dalam dokumen, kalimat, atau fitur atau tingkat aspek apakah yang dikemukakan dalam dokumen, kalimat atau fitur entitas baik bersifat positif, negatif atau netral. (Saraswati, 2013).

Dari beberapa Teknik klasifikasi yang paling sering digunakan adalah Metode *Naïve Bayes*. Menurut penelitian yang dilakukan (Bilal et al., 2016) telah melakukan perbandingan antara Algoritma *Naïve Bayes*, *Decision Tree*, dan *K Nearest Neighbour* untuk klasifikasi sentiment yang menghasilkan kesimpulan bahwa Algoritma *Naïve Bayes* adalah yang terbaik dari segi akurasi, presisi, *recall* dan F-measure. Begitu juga Sigit Suryono dan Ema Utami, dkk telah melakukan analisis sentiment pada twitter terhadap pemerintahan Presiden Joko Widodo dengan menggunakan Metode *Naïve Bayes Classifier*. Untuk analisis sentiment dengan metode memperoleh hasil tertinggi pada percobaan keempat 66,79%. *Naive Bayes* merupakan metode klasifikasi yang kuat dan efisien yang dapat membuat tugas klasifikasi tampil secara konsisten dan baik. (Kundu et al., 2015) (Solecha, 2019).

Dalam pengklasifikasian *Naive Bayes* sangat sederhana dan efisien (Ayu Muthia, 2013; Chen et al., 2009). Selain itu, *Naïve Bayes* adalah Teknik *Machine Learning* yang populer untuk klasifikasi teks, serta memiliki performa yang baik pada banyak domain

(Ayu Muthia, 2013; Ye, Zhang, & Law, 2009). Namun, salah satu kekurangan *Naïve Bayes* yaitu sangat sensitif dalam pemilihan fitur (Chen et al., 2009). Pemilihan fitur merupakan salah satu faktor yang paling penting yang dapat mempengaruhi tingkat akurasi klasifikasi. (Cahyono, 2017).

Penelitian (Wardhani dkk, 2018) telah melakukan klasifikasi sentiment artikel menggunakan Algoritma *Naïve Bayes* dan *Support Vector Machine (SVM)* dengan *Particle Swarm Optimization (PSO)* yang menghasilkan bahwa Algoritma *Naïve Bayes* dengan *PSO* adalah yang terbaik dari segi akurasi dengan nilai 92.00% dan AUC 0.550%. Penelitian yang dilakukan (Ernawati dkk, 2018) telah melakukan analisis sentiment terhadap *review* perusahaan online menggunakan Algoritma *Naïve Bayes* dengan *feature selection* menggunakan Algoritma Genetika yang menghasilkan nilai akurasi sebesar 87.50% dan AUC 0.819%. Begitu juga Wati (2016) yang melakukan analisis sentiment terhadap *review* jasa Maskapai Penerbangan menggunakan *Naïve Bayes* dengan Algoritma Genetika sebagai seleksi fitur menghasilkan nilai akurasi Algoritma Genetika 89.50%. Berdasarkan hal tersebut, penelitian dilakukan untuk mengetahui pemilihan fitur terbaik untuk meningkat kinerja *Naïve Bayes* terhadap analisis sentiment *twitter*.

## 2. METODE PENELITIAN

### 2.1 Objek Penelitian

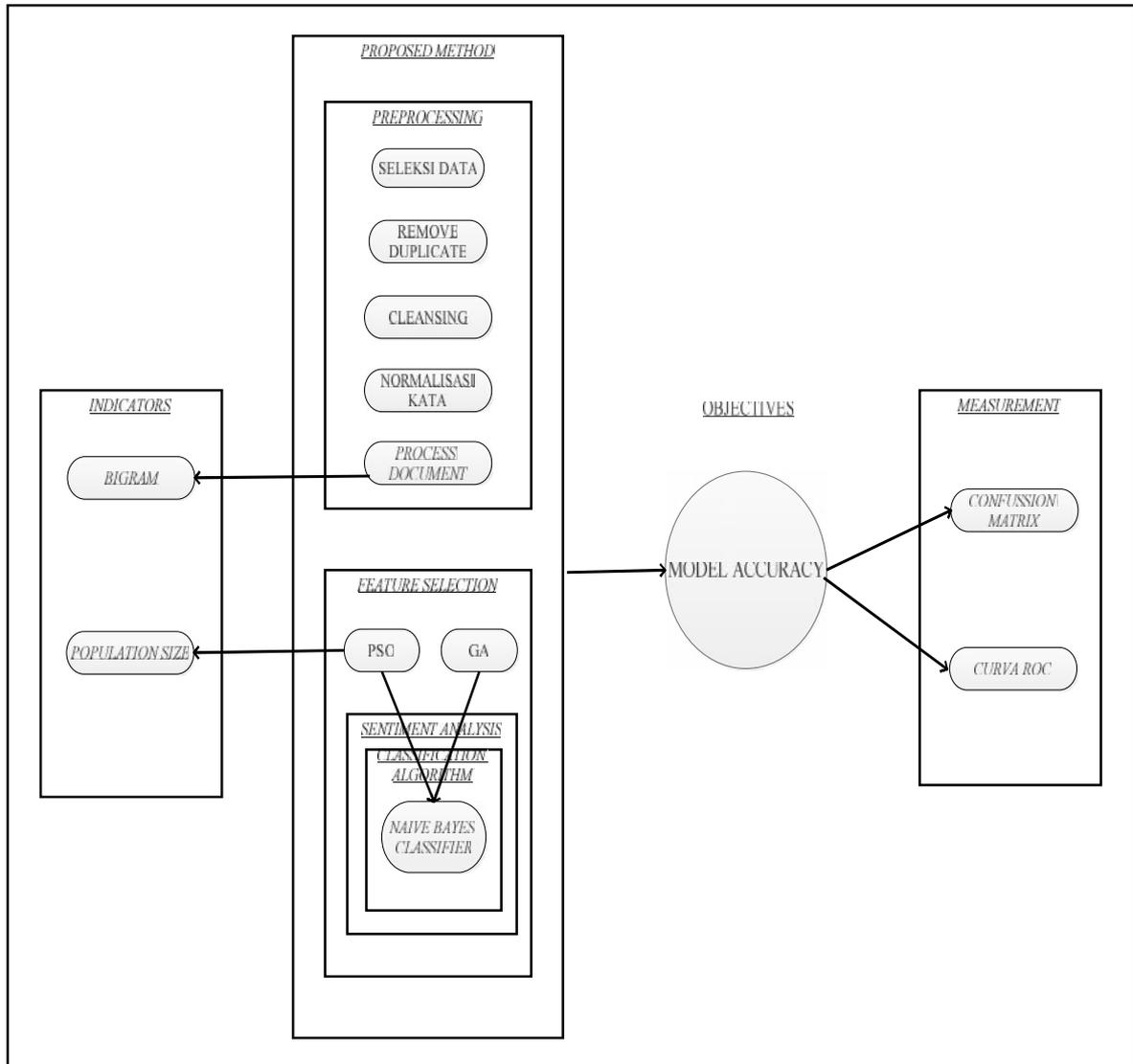
Objek penelitian yang diteliti adalah komentar yang berasal dari akun @hendralm. Pilih objek didasari karena akun tersebut berani mengungkapkan praktik jual beli data E-KTP. Sehingga menyita banyak perhatian masyarakat di media sosial khususnya *twitter*

dan banyaknya jumlah komentar yang di berikan baik dukungan maupun penolakan.

### 2.2 Kerangka Pemikiran

Pada penelitian ini, data set yang digunakan adalah data set yang diambil dari *twitter*. Data yang diambil berdasarkan akun yang berhubungan dengan @hendralm. Akun adalah akun yang sempat viral karena berani mengungkapkan sindikat jual beli data pribadi yaitu E-KTP. Data yang digunakan adalah komentar *tweet* terhadap akun tersebut. Pemrosesan data awal yaitu *preprocessing* dengan melakukan proses *cleansing* untuk menghilangkan *noise*, setelah itu proses normalisasi kata untuk menghilangkan kata-kata yang tidak baku, kemudian *process document* menggunakan *RapidMiner* yang meliputi *transform cases*, *tokenize*, *filter tokens (by length)*, *filter stopwords*, *stem*, *generate n-grams*. Metode yang diusulkan (**Proposed Method**) adalah proses klasifikasi dengan menggunakan Algoritma *Naïve Bayes*, model fitur seleksi yang akan digunakan pertama adalah *PSO* sedangkan fitur seleksi kedua menggunakan *GA*, indikator (*indicators*) yang akan dilakukan observasi adalah penggunaan *n-gram* yaitu *bigram* dan *population size* pada fitur seleksi.

Obyektif (*Objectives*) pada penelitian ini adalah untuk mengetahui hasil klasifikasi berdasar model yang diusulkan yaitu dengan Teknik *Cross Validation* dengan indikator yang ditentukan dan telah dijelaskan sebelumnya. Kemudian dilakukan pengukuran (*measurement*) evaluasi menggunakan *confussion matrix* dan kurva ROC untuk hasil akurasi algoritma. Kerangka pemikiran dapat dilihat pada Tabel 1.



Gambar 1. Kerangka Pemikiran

### 2.3 Tahapan Penelitian

Tahapan Penelitian (Gambar 2) yang akan dilakukan adalah sebagai berikut:

#### a. Pengumpulan Data

Pengumpulan data dilakukan dengan cara *crawling* dengan menggunakan perangkat lunak *Orange*. Data yang dikumpulkan merupakan teks berbahasa Indonesia yang diambil dari akun media sosial *twitter* yaitu @hendralm. Teks yang diperoleh berisikan *tweet* atau komentar terhadap akun tersebut.

#### b. Preprocessing

*Preprocessing* merupakan proses untuk mendapatkan data bersih agar dapat

dilakukan proses selanjutnya, berikut ini adalah tahapannya:

#### 1. Seleksi Data

Tahapan ini bertujuan untuk menyeleksi data *tweet* yang masih terdiri dari data-data yang tidak diperlukan.

#### 2. Remove Duplicate

Proses ini bertujuan untuk menyeleksi komentar *tweet* yang berulang. Sebab pada *twitter* terdapat fitur *retweet* yang memberikan dampak banyaknya teks berulang dengan topik dan isi yang sama.

#### 3. Cleansing

Tahapan *cleansing* pada penelitian ini adalah untuk menghapus kata, karakter

dan simbol yang tidak diperlukan dalam proses penelitian.

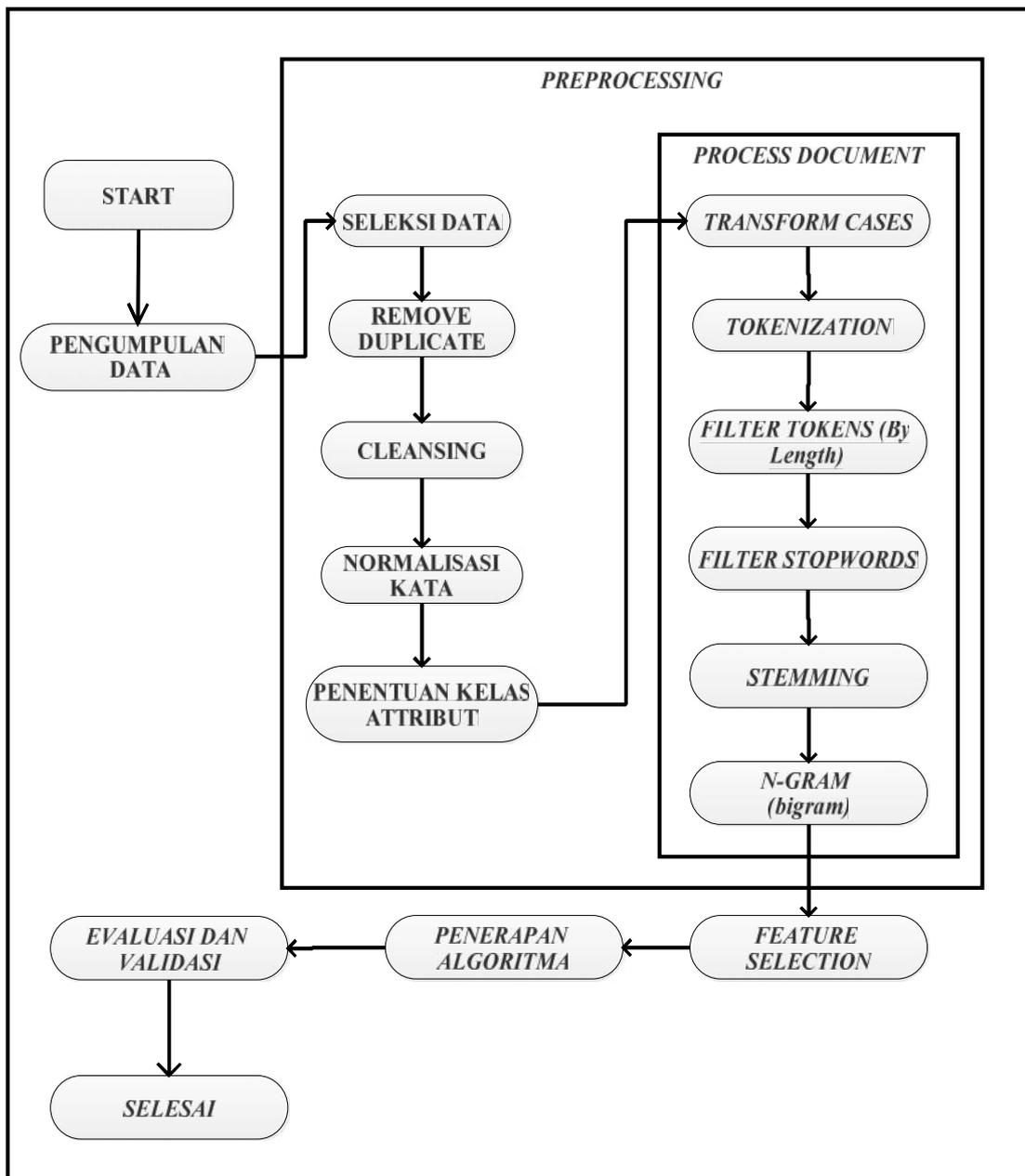
**4. Normalisasi Kata**

Tahapan ini bertujuan untuk mengganti kata-kata yang tidak baku yang terdapat pada data komentar atau *tweet*. Diantaranya bahasa gaul (*slang word*),

kata singkatan, merenggangkan kata, dan menghapus huruf yang berulang.

**5. Penentuan Kelas Atribut**

Penentuan kelas atribut diberikan sesuai dengan subyektifitas peneliti. Pembagian kelas pada penelitian ini dibagi menjadi positif dan negatif.



**Gambar 2.** Tahapan Penelitian

**c. Process Document**

*Preprocessing* pada tahap ini dilakukan dengan menggunakan *tool Rapidminer*,

dimana proses menggunakan beberapa rangkaian *operator* sub proses dalam *process document from data* pada rangkaiannya.

*Preprocessing* dengan menggunakan *RapidMiner* ini diterapkan pada data *training* maupun *data testing*. Berikut penjelasan:

### 1. *Transform Cases*

*Operator* ini bertujuan untuk mengubah atau menyeragamkan semua bentuk huruf kecil semua (*lowercases*), karena pada komentar *twitter* memiliki bentuk huruf yang tidak seragam.

### 2. *Tokenization*

Pada proses ini, semua kata yang ada didalam tiap dokumen dipisahkan dan dihilangkan tanda bacanya, serta dihilangkan jika terdapat simbol atau apapun yang bukan huruf.

### 3. *Filter Tokens (by Length)*

Pada proses ini, bertujuan untuk menghilangkan atau menghapus kata berdasarkan panjang karakter yang ditentukan. Pada penelitian ini minimum panjang karakter yang digunakan adalah empat karakter pada setiap kata. Parameter yang digunakan ada dalam proses ini adalah *min chars* = 3 dan *max chars* = 25

### 4. *Filter Stopwords*

Kata-kata yang tidak relevan pada tahap

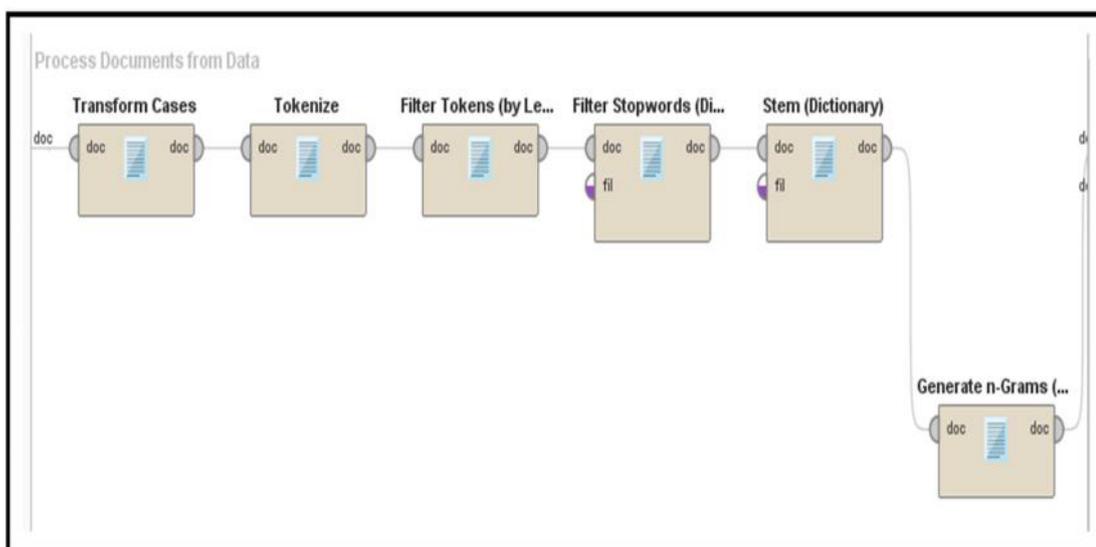
ini akan dihapus, kata-kata yang tidak mempunyai makna tersendiri jika dipisahkan dengan kata yang lain dan tidak terikat dengan kata sifat yang berhubungan dengan sentiment.

### 5. *Stemming*

*Stemming* merupakan suatu proses pencarian kata dasar dengan menghilangkan imbuhan. Dalam proses ini kata-kata akan dikelompokkan ke dalam beberapa kelompok yang mempunyai kata dasar sama (Nugroho, 2018).

### 6. *Generate N-gram*

*Generate N-gram* merupakan proses menghitung probabilitas yang bersyarat untuk sebuah kata dari urutan kata sebelumnya. Sebuah *n-gram* adalah sekumpulan kata dengan masing-masing panjang n kata. Dalam penelitian ini *n-gram* yang digunakan adalah *bigram*. Karena pada penelitian sebelumnya dapat meningkatkan kinerja Algoritma *Naive bayes*. Salah satu keunggulan *n-gram* adalah bahwa *n-gram* tidak terlalu sensitif terhadap kesalahan dalam penulisan kata (Nugroho, 2018).

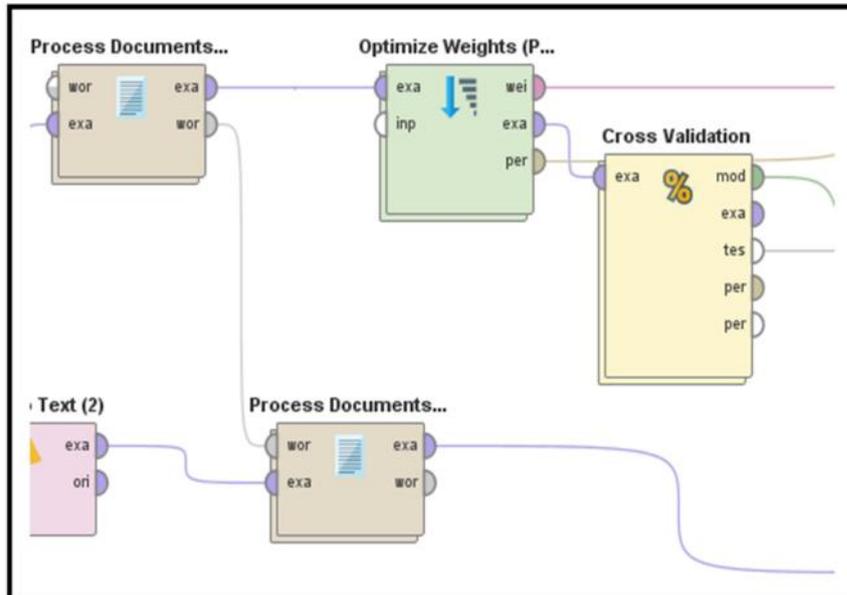


Gambar 3. *Process Document*

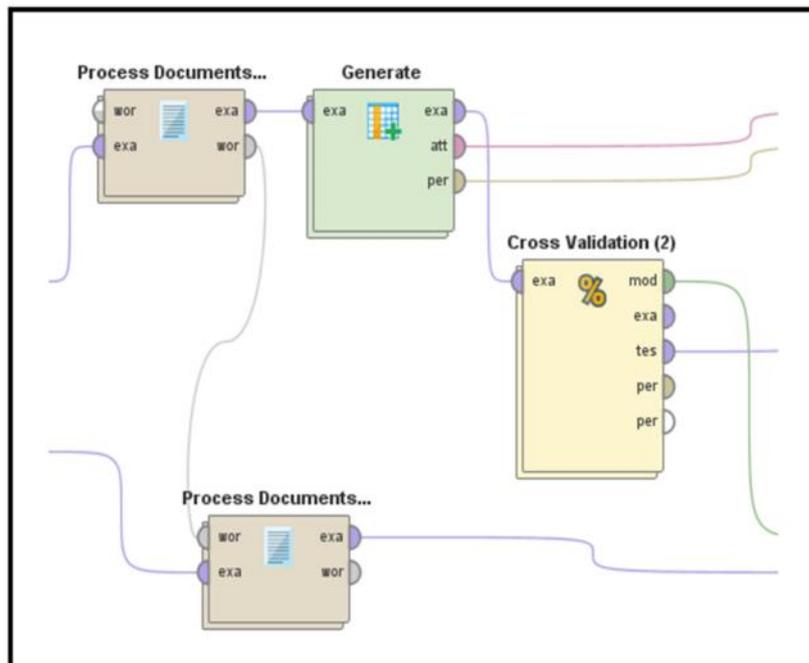
**d. Feature Selection**

Pada penelitian ini *feature selection* yang akan digunakan adalah *PSO* dan *GA* dan cara

yang dilakukan yaitu dengan menambahkan salah satu pemilihan fitur pada *subprocess*.



**Gambar 4.** Desain dengan *PSO*

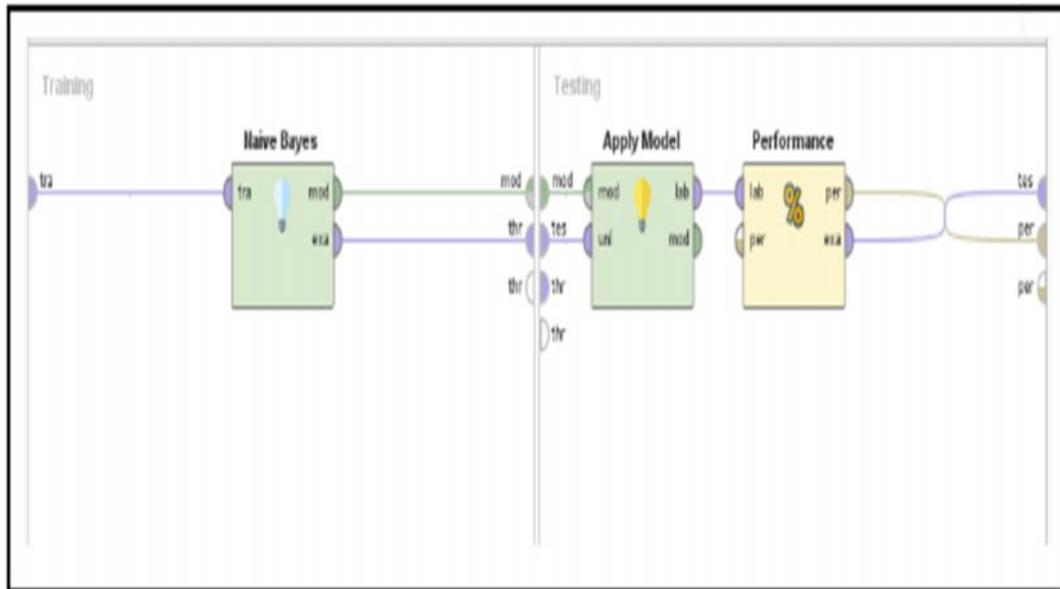


**Gambar 5.** Desain dengan *GA*

**e. Penerapan Algoritma**

Algoritma yang akan di pakai dalam penelitian ini yaitu *Naïve Bayes*. Metode

*Naïve Bayes* merupakan metode yang sangat sederhana yang digunakan untuk menghitung performa klasifikasi.



**Gambar 6.** Desain Penerapan Algoritma

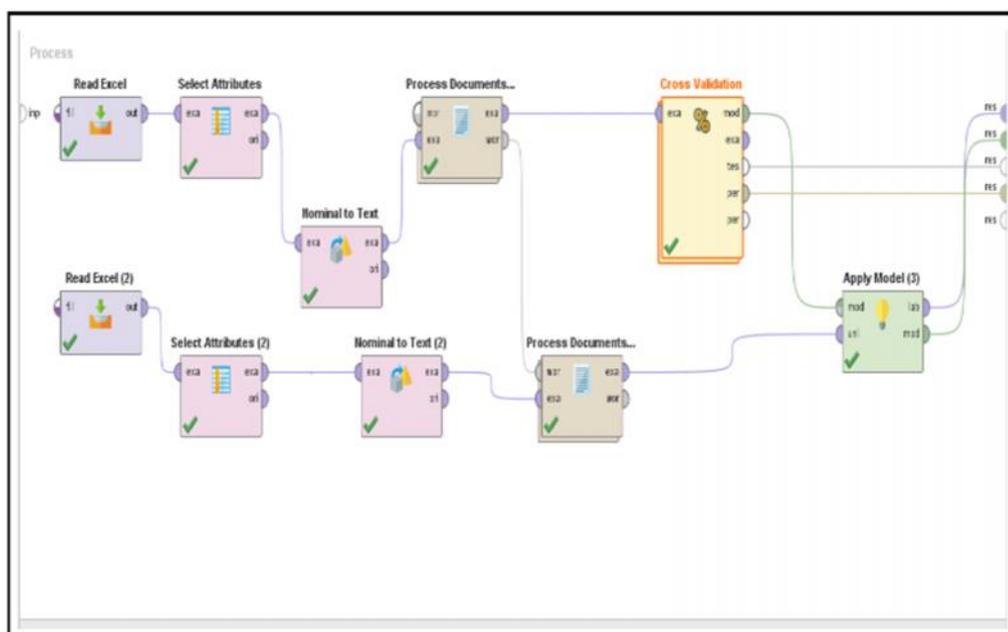
**f. Evaluasi dan Validasi**

Pada penelitian ini, evaluasi yang dilakukan untuk mengetahui akurasi dan kinerja dari *feature selection* terhadap Algoritma *Naïve Bayes*. Validasi bertujuan untuk melihat perbandingan hasil akurasi dari metode atau model yang digunakan dengan hasil yang telah ada sebelumnya. Teknik validasi yang digunakan dalam penelitian ini adalah *Cross Validation*, akurasi algoritma akan diukur menggunakan *Confusion Matrix*.

**3. HASIL DAN DISKUSI**

**3.1 Proses Pengujian**

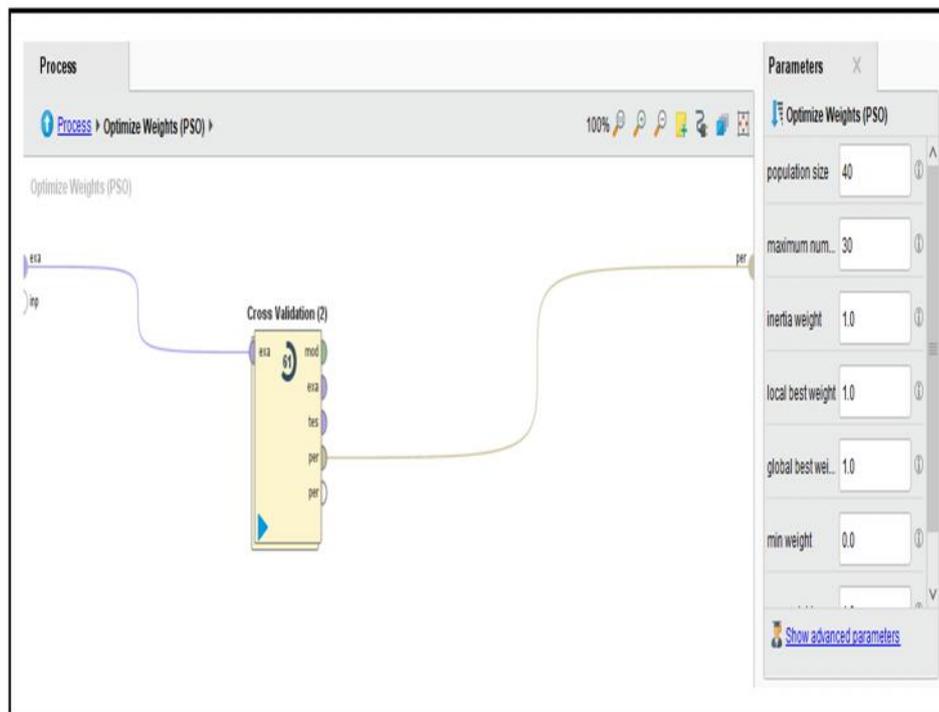
Proses pengujian ini menggunakan *RapidMiner* dengan merangkai beberapa operator proses yang sesuai yaitu *read excel*, *select attributes*, *nominal to text*, *process document form data*, *cross validation*, dan *apply model* serta *feature selection*. Gambar utama beberapa proses eksperimen dapat dilihat pada Gambar 7.



**Gambar 7.** Desain Pengujian Utama

Proses selanjutnya melakukan observasi atau pengujian terhadap nilai *population size* sebagai indikator eksperimen untuk penggunaan *feature selection* terhadap Algoritma *Naïve Bayes*. Berdasarkan penelitian sebelumnya bahwa nilai *population size* sangat membantu untuk meningkatkan kinerja *feature selection*. Proses penentuan nilai *population size*

menggunakan pengujian terhadap *feature selection PSO* dengan Algoritma *Naïve Bayes*. Nantinya nilai *population size* yang menghasilkan akurasi tertinggi akan digunakan dalam pengujian menggunakan *cross validation* pada setiap proses pengujian. Gambaran proses pengujian dapat dilihat pada Gambar 8.



**Gambar 8.** Pengujian Nilai *Population Size*

Dari pengujian diatas, beberapa nilai akurasi dari *population size* dimasukkan ke dalam *parameters*. Hasil nilai akurasi dapat dilihat pada Tabel 1.

Berdasarkan pengujian tersebut diperoleh nilai akurasi *population size* tertinggi yang terdapat pada nilai *population size* 40 dengan nilai akurasi 70.50% dan AUC sebesar 0.573. Sehingga nilai *population size* tersebut akan digunakan untuk meningkatkan akurasi *feature selection* pada pengujian Algoritma *Naïve Bayes* dengan *feature selection* yang telah ditentukan sebelumnya.

**Tabel 1.** Nilai Akurasi Pengujian *Population Size*

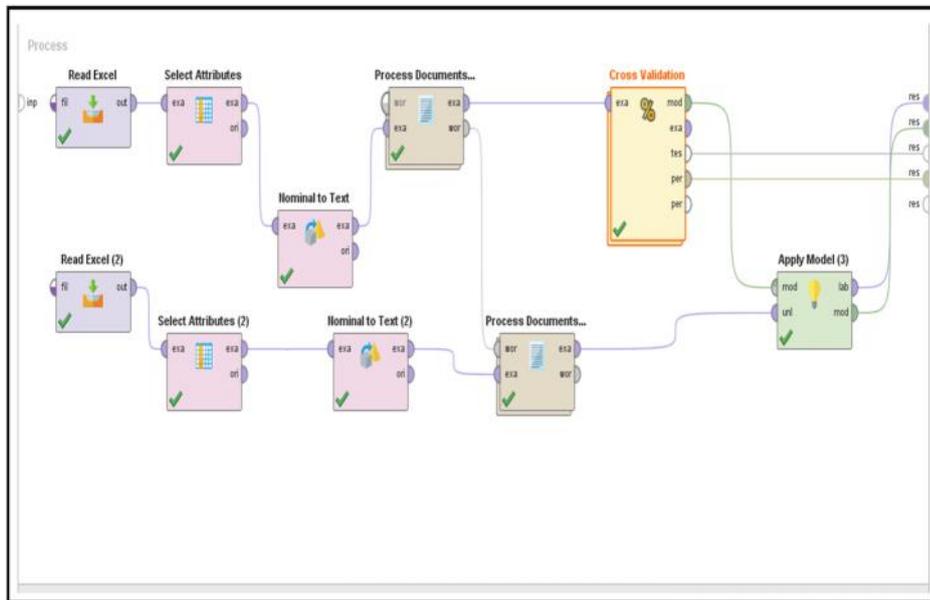
<i>Population Size</i>	<i>Accuraccy</i>	<i>Curva ROC</i>
5	69.25%	0.539
10	68.75%	0.557
20	69.25%	0.550
40	70.50%	0.573

### 3.2 Pengujian dengan Algoritma *Naïve Bayes*

Pada pengujian ini dilakukan eksperimen dengan menggunakan Algoritma

*Naïve Bayes* tanpa *feature selection*. Gambar proses pengujian tersebut dapat dilihat pada Gambar 9.

Proses pengujian dilakukan dengan Teknik *Cross Validation* dengan tujuan untuk memperoleh nilai akurasi terbaik.



**Gambar 9.** Desain Pengujian dengan *Naïve Bayes*

Hasil pengujian dapat dilihat pada Tabel 2. Berdasarkan pengujian tersebut diperoleh bahwa nilai akurasi tertinggi yang terdapat pada pengujian ke-10 dengan nilai akurasi 67.71% dan AUC sebesar 0.540.

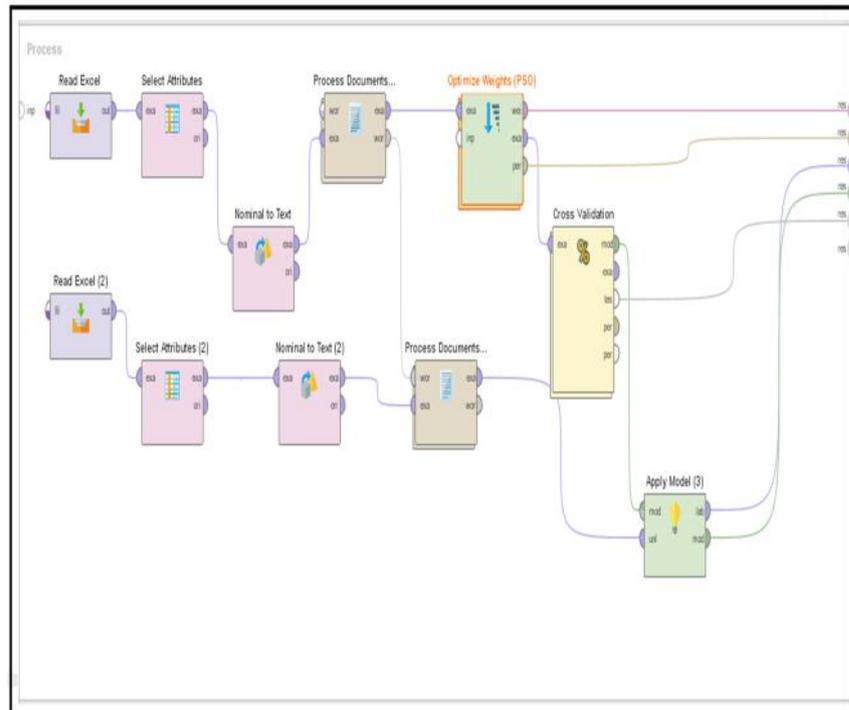
Pengujian	Accuracy	AUC
Uji 9	63.45%	0.467
Uji 10	67.71%	0.540

**Tabel 2.** Nilai Akurasi Pengujian dengan *Naïve Bayes*

Pengujian	Accuracy	AUC
Uji 1	67.12 %	0.544
Uji 2	63.20%	0.589
Uji 3	65.70%	0.549
Uji 4	65.83%	0.343
Uji 5	63.21%	0.589
Uji 6	62.96%	0.532
Uji 7	65.83%	0.592
Uji 8	63.19%	0.385

### 3.3 Pengujian *Naïve Bayes* dengan *PSO*

Pada pengujian ini dilakukan lakukan eksperimen dengan menggunakan Algoritma *Naïve Bayes* dengan *feature selection PSO*. Tujuan untuk mengetahui akurasi dari Algoritma *Naïve Bayes* dengan *feature selection PSO* akan di bandingkan dengan pengujian sebelumnya. Pada pengujian di gunakan nilai *population size* yang telah di tentukan sebelumnya. Nilai *population size* yang digunakan yaitu 40. Proses pengujian dapat dilihat pada Gambar 10 dan masih sama seperti sebelumnya dengan menggunakan Teknik *Cross Validation*.



**Gambar 10.** Pengujian *Naive Bayes* dengan *PSO*

Hasil pengujian ini dapat dilihat pada Tabel 3. Berdasarkan pengujian tersebut diperoleh bahwa nilai akurasi tertinggi yang terdapat pada pengujian ke-7 dengan nilai akurasi 71.85% dan AUC sebesar 0.590.

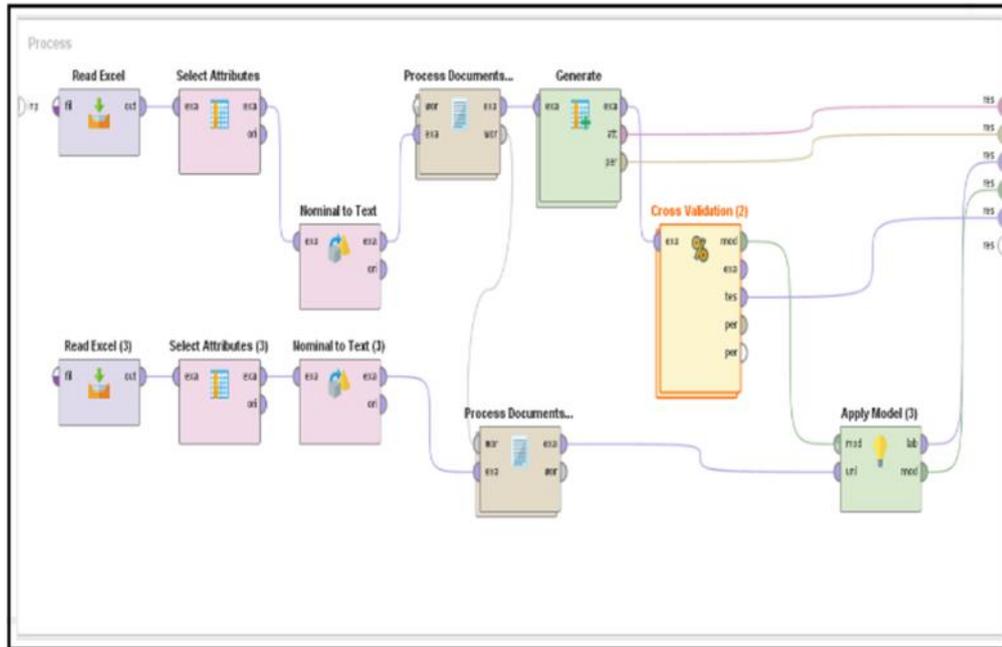
**Tabel 3.** Nilai Akurasi Pengujian *Naive Bayes* dengan *PSO*

Pengujian	Accuracy	AUC
Uji 1	70.50 %	0.573
Uji 2	69.21%	0.575
Uji 3	70.40%	0.628
Uji 4	69.60%	0.420
Uji 5	69.22%	0.575
Uji 6	68.20%	0.625
Uji 7	71.84%	0.590

Pengujian	Accuracy	AUC
Uji 8	68.95%	0.444
Uji 9	68.70%	0.612
Uji 10	69.59%	0.562

### 3.4 Pengujian *Naive Bayes* dengan *GA*

Pada pengujian ini dilakukan eksperimen dengan menggunakan Algoritma *Naive Bayes* dengan *feature selection GA*. Tujuan untuk mengetahui akurasi dari Algoritma *Naive Bayes* dengan *feature selection GA* yang akan dibandingkan dengan pengujian sebelumnya. Pada pengujian digunakan nilai *population size* yang telah ditentukan sebelumnya. Nilai *population size* yang digunakan yaitu 40. Desain proses pengujian dapat dilihat pada Gambar 11.



**Gambar 11.** Pengujian *Naive Bayes* dengan *GA*

Proses pengujian ini masih sama seperti sebelumnya yaitu dengan menggunakan Teknik *Cross Validation*. Hasil pengujian ini dapat dilihat pada Tabel 4.

**Tabel 4.** Nilai Akurasi Pengujian *Naive Bayes* dengan *GA*

Pengujian	Accuracy	AUC
Uji 1	68.25 %	0.407
Uji 2	68.59%	0.600
Uji 3	68.47%	0.532
Uji 4	68.09%	0.391
Uji 5	68.59%	0.600
Uji 6	65.33%	0.627
Uji 7	68.96%	0.525
Uji 8	66.72%	0.515
Uji 9	65.09%	0.421
Uji 10	68.58%	0.467

Berdasarkan pengujian tersebut diperoleh bahwa nilai akurasi tertinggi yang terdapat pada pengujian ke-7 dengan nilai akurasi 68.96% dan AUC sebesar 0.525.

### 3.5 Analisa Akurasi

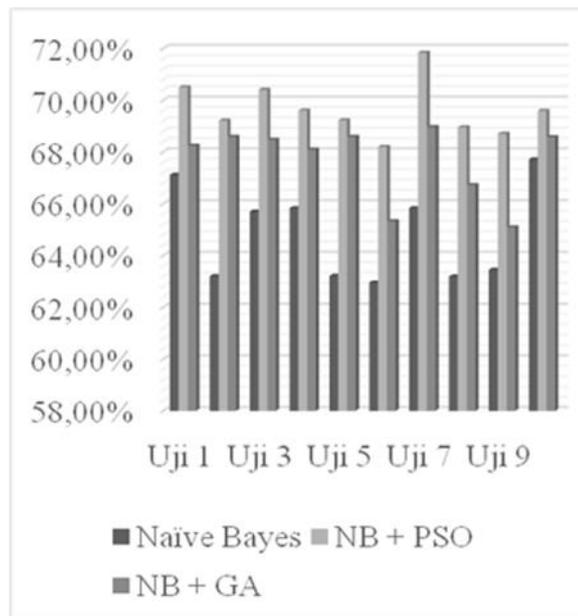
Performa atau akurasi merupakan tingkat kedekatan atau seberapa dekat nilai prediksi dengan nilai sebenarnya. Berdasarkan pengujian di atas dapat dilihat nilai performa dari pengujian menggunakan Algoritma *Naive Bayes* dengan *feature selection*. Tabel 5 merupakan nilai perbandingan dari setiap masing-masing pengujian yang dilakukan dengan menggunakan Teknik *Cross Validation*.

**Tabel 5.** Pengujian Nilai Akurasi

Pengujian	AKURASI		
	<i>Naive Bayes</i>	<i>NB + PSO</i>	<i>NB + GA</i>
Uji 1	67.12%	70.50%	68.25 %
Uji 2	63.20%	69.21%	68.59%

Pengujian	AKURASI		
	<i>Naïve Bayes</i>	<i>NB + PSO</i>	<i>NB + GA</i>
Uji 3	65.70%	70.40%	68.47%
Uji 4	65.83%	69.60%	68.09%
Uji 5	63.21%	69.22%	68.59%
Uji 6	62.96%	68.20%	65.33%
Uji 7	65.83%	71.84%	68.96%
Uji 8	63.19%	68.95%	66.72%
Uji 9	63.45%	68.70%	65.09%
Uji 10	67.71%	69.59%	68.58%

Berdasarkan pengujian tersebut di peroleh nilai akurasi tertinggi dari pengujian ke-7 dengan nilai akurasi sebesar 71.84%. Nilai tersebut diperoleh dari pengujian dengan menggunakan Algoritma *Naïve Bayes* dengan *feature selection PSO*. Pengujian dengan menggunakan *feature selection PSO* dapat meningkatkan akurasi 6.00% dari pengujian tanpa menggunakan *feature selection*, serta meningkatkan akurasi sebesar 3.00% dari pengujian menggunakan *feature selection GA*. Hal ini menunjukkan bahwa *feature selection PSO* dapat meningkatkan akurasi dari Algoritma *Naïve Bayes* walaupun tidak terlalu signifikan. *Chart* akurasi dari masing-masing pengujian ditampilkan pada Gambar 12.



**Gambar 12.** Grafik Pengujian Akurasi

#### 4. KESIMPULAN

Pengujian terbaik dari semua eksperimen yang telah dilakukan dalam penelitian ini adalah Model Algoritma *Naïve Bayes* yang dikombinasikan dengan *feature selection PSO* dengan selisih akurasi dari pengujian Algoritma *Naïve Bayes* sebesar 4.00% dan selisih dari pengujian dengan

Model Algoritma *Naïve Bayes* yang dikombinasikan dengan *feature selection GA* sebesar 3.00%.

Dengan demikian bahwa *PSO* merupakan *feature selection* terbaik untuk membantu meningkatkan kinerja dari Algoritma *Naïve Bayes* pada analisis sentiment *twitter*.

## Saran

Pada penelitian ini masih memiliki banyak kekurangan, oleh karena itu penulis memberikan saran untuk penelitian selanjutnya yaitu:

1. Menggunakan Algoritma Klasifikasi yang berbeda.
2. Menggunakan *feature selection* yang berbeda.
3. Meningkatkan nilai *population size* serta menggunakan indikator yang berbeda untuk meningkatkan kinerja dari *feature selection PSO* ataupun *GA*.

## DAFTAR PUSTAKA

- Ayu Muthia, D. 2013. Integrasi Algoritma Genetika dan Information Gain Untuk Seleksi Fitur Pada Analisis Sentimen Review Film Menggunakan Algoritma Naïve Bayes. *Jurnal Teknik Komputer AMIK BSI (JTK)*, 4(1), 186–193.
- Bilal, M., Israr, H., Shahid, M., & Khan, A. 2016. Sentiment classification of Roman-Urdu opinions using Naïve Bayesian, Decision Tree and KNN classification techniques. *Journal of King Saud University - Computer and Information Sciences*, 28(3), 330–344. <https://doi.org/10.1016/j.jksuci.2015.11.003>
- Cahyono, Y. 2017. Analisis Sentiment pada Sosial Media Twitter Menggunakan Na ve Bayes Classifier dengan Feature Selection Particle Swarm Optimization dan Term Frequenc. *Jurnal Informatika Universitas Pamulang*, 2, 14--19.
- Chen, J., Huang, H., Tian, S., & Qu, Y. (2009). Feature selection for text classification with Naïve Bayes. *Expert Systems with Applications*, 36 (3 PART 1), 5432–5435. <https://doi.org/10.1016/j.eswa.2008.06.054>
- Ernawati, S., Yulia, E. R., Frieyadie, & Samudi. 2018. Implementation of The Naïve Bayes Algorithm with Feature Selection using Genetic Algorithm for Sentiment Review Analysis of Fashion Online Companies. 2018 6th International Conference on Cyber and IT Service Management (CITSM), (Citsm), 1–5. <https://doi.org/10.1109/CITSM.2018.8674286>
- Kundu, A., Jain, V., Kumar, S., & Chandra, C. 2015. A journey from normative to behavioral operations in supply chain management: A review using Latent Semantic Analysis. *Expert Systems with Applications*, 42(2), 796–809. <https://doi.org/10.1016/j.eswa.2014.08.035>
- Nugroho, A. 2018. Analisis Sentimen pada Media Sosial Twitter Menggunakan Naive Bayes Classifier dengan Ekstraksi Fitur N-Gram. *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, 2(2), 200. <https://doi.org/10.30645/j-sakti.v2i2.83>
- Saraswati, N. W. S. 2013. Naïve Bayes Classifier dan Support Vector Machines untuk Sentiment Analysis. *Seminar Nasional Sistem Informasi Indonesia*, 586–591. Retrieved from [http://is.its.ac.id/pubs/oajis/index.php/file/download\\_file/512](http://is.its.ac.id/pubs/oajis/index.php/file/download_file/512)
- Solecha, K. 2019. Analisa Sentimen dengan Algoritma Naïve Bayes Classifier Berbasis Particle Swarm Optimization untuk Review Restoran. 11(1).
- Wardhani, N. K., Rezkiani, Kurniawan, S., Setiawan, H., Gata, G., Tohari, S., ... Wahyudi, M. 2018. Sentiment Analysis Article News Coordinator Minister of Maritime Affairs using Algorithm Naive Bayes and Support Vector

- Machine with Particle Swarm Optimization. *Journal of Theoretical and Applied Information Technology*, 96(24), 8365–8378.
- Wati, R. 2016. Penerapan Algoritma Genetika untuk Seleksi Fitur pada Analisis Sentimen Review Jasa Maskapai Penerbangan Menggunakan Naive Bayes. *Evolusi*, 4(1), 25–31.
- Ye, Q., Zhang, Z., & Law, R. 2009. Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Systems with Applications*, 36(3 PART 2), 6527–6535.  
<https://doi.org/10.1016/j.eswa.2008.07.035>
- Zy, A. T., & Nugroho, A. 2018. Comparison Algorithm Classification Naive Bayes, Decision Tree, and Neural Network for Analysis Sentiment. *International Conference on Economic, Business, and Accounting*, 1(c), 115–115.

HALAMAN INI SENGAJA DIKOSONGKAN