

## Penerapan Particle Swarm Optimization Untuk Seleksi Fitur Pada Analisis Sentimen Review Perusahaan Penjualan Online Menggunakan Naïve Bayes

Siti Ernawati

Program Studi Sistem Informasi, STMIK Nusa Mandiri Jakarta  
rna2103@gmail.com

**Abstract** - Review of company sales made online is a channel that connects consumers with each other. They can pour their opinion of a company which has been making purchases. Consumer reviews online good influence whether or not a company's sales. The main problem in text classification is higher dimensions of space features, it is often the case in the text that have tens of thousands of features. Most of these features are irrelevant and not useful for text classification can even reduce the level of accuracy. Therefore, in this study using Naïve Bayes classifier with Particle Swarm Optimization as a feature selection method is applied to classify the text on the review of online sales company to improve accuracy. This research resulted in the classification of the text in the form of positive and negative. Measurement is based on the accuracy of Naïve Bayes before and after the addition of feature selection methods. Validation is performed using 10-fold cross validation. While the measurement accuracy is measured by the confusion matrix and ROC curves. The results showed that an increase of 79.50% and AUC 0.500 become 86.88% and AUC 0.705. It can be concluded that the application of Particle Swarm Optimization for feature selection in the Naïve Bayes able to improve accuracy.

**Keywords:** Company Review Online Sales, Fiture Selection, Naïve Bayes, Particle Swarm Optimization, Sentimen Analisis

**Abstrak** - Ulasan penjualan perusahaan dilakukan secara online adalah saluran yang menghubungkan konsumen dengan satu sama lain. Mereka bisa menuangkan pendapat mereka dari perusahaan yang telah melakukan pembelian. ulasan konsumen pengaruh secara online baik apakah penjualan perusahaan atau tidak. Masalah utama dalam klasifikasi teks dimensi yang lebih tinggi dari fitur ruang, yang sering terjadi dalam teks yang memiliki puluhan ribu fitur. Kebanyakan dari fitur ini tidak relevan dan tidak berguna untuk klasifikasi teks bahkan dapat mengurangi tingkat akurasi. Oleh karena itu, dalam penelitian ini menggunakan Naïve Bayes classifier dengan Optimization Particle Swarm sebagai metode seleksi fitur diterapkan untuk mengklasifikasikan teks pada review dari perusahaan penjualan online untuk meningkatkan akurasi. Penelitian ini menghasilkan klasifikasi teks dalam bentuk positif dan negatif. Pengukuran didasarkan pada keakuratan Naïve Bayes sebelum dan sesudah penambahan metode seleksi fitur. Validasi dilakukan dengan menggunakan 10-fold cross validasi. Sementara akurasi pengukuran diukur dengan matriks kebingungan dan kurva ROC. Hasil penelitian menunjukkan bahwa peningkatan 79,50% dan AUC 0,500 menjadi 86,88% dan AUC 0,705. Dapat disimpulkan bahwa penerapan Particle Swarm Optimization untuk seleksi fitur di Naïve Bayes dapat meningkatkan akurasi.

**Kata Kunci:** Perusahaan Ulasan Penjualan Online, Fiture Seleksi, Naïve Bayes, Particle Swarm Optimization, Sentimen Analisis

### A. PENDAHULUAN

Saat ini, tersedia banyak *review* dalam web, *review* yang tersedia merupakan sumber yang sangat berguna untuk berbagai bidang, seperti bidang pemasaran, sosial dan lain-lain. *Review* tersebut menggambarkan pandangan terhadap sesuatu misalnya *review* terhadap produk tertentu, *review* restoran dan lain sebagainya. Adapun contoh *review* menurut Tang et al. (2009, 10760) apakah *review* suatu produk positif atau negatif, bagaimana suasana hati yang dirasakan kalangan blogger, bagaimana cerminan publik terhadap urusan politik.

Banyak konsumen yang menuangkan *review* atau pengalaman mereka melalui media sosial seperti facebook, twitter atau situs media lainnya. *Review* perusahaan

penjualan yang dibuat secara *online* adalah saluran yang menghubungkan konsumen yang satu dengan yang lainnya, mereka dapat menuangkan opini tentang perusahaan dimana mereka telah melakukan transaksi pembelian. *Review* konsumen secara *online* mempengaruhi bagus atau tidaknya sebuah perusahaan penjualan tersebut. Menurut Tan dan Zhang (2008, 2622) Masalah dalam analisis sentimen adalah klasifikasi sentimen, dimana dokumen diberi label sebagai label positif (*'thumbs up'*) atau label negatif (*'thumbs down'*).

Terdapat beberapa penelitian yang sudah dilakukan dalam klasifikasi sentimen terhadap *review* secara *online* diantaranya, Analisis sentimen untuk *review* restoran menggunakan algoritma Naïve Bayes menurut Kang et al.

(2012, 6000). Klasifikasi sentimen *review online* tujuan perjalanan menggunakan algoritma Naïve Bayes, Support Vector Machines, dan character based N-gram menurut Ye et al. (2009, 6527). Klasifikasi sentimen *review* restoran di internet menggunakan bahasa Canton menggunakan algoritma Naïve Bayes dan Support Vector Machines menurut Zhang et al. (2011, 7674). Analisis sentimen dengan Naïve Bayes untuk melihat persepsi masyarakat terhadap batik menurut Apriliyanti (2015, 833).

Naïve Bayes adalah algoritma yang sering digunakan dalam pengkategorian teks. Ide dasarnya adalah menggabungkan probabilitas kata-kata dan kategori untuk memperkirakan probabilitas dari kategori sebuah dokumen menurut Zhang et al. (2011, 7676). Naïve Bayes merupakan algoritma paling sederhana dari pengklasifikasi probabilistik Ting et al. (2011, 38). Seperti yang sudah disebutkan diatas bahwa Naïve Bayes sangat sederhana dan efisien, disisi lain Naïve Bayes sangat sensitif terhadap pemilihan fitur seleksi, maka dari itu pemilihan fitur yang sesuai sangat diperlukan Chen et al. (2009, 5432).

Masalah utama dalam klasifikasi teks adalah dimensi tinggi dari ruang fitur, hal ini sering terjadi pada teks yang memiliki puluhan ribu fitur. Sebagian besar fitur ini tidak relevan dan tidak bermanfaat bagi klasifikasi teks bahkan dapat mengurangi tingkat akurasi menurut Chen et al. (2009, 5432).

Seleksi fitur merupakan langkah penting dalam klasifikasi teks dan mempengaruhi secara langsung terhadap performa klasifikasi. Dalam rangka untuk meningkatkan efek seleksi fitur, banyak penelitian mencoba untuk menambahkan algoritma optimasi cerdas dalam metode seleksi fitur. Menurut Lu, et al. (2015, 629) Jika dibandingkan dengan *ant colony algorithm* dan *genetic algorithms*, algoritma Particle Swarm Optimization adalah algoritma paling sederhana dan cepat dalam proses pengaplikasiannya untuk menemukan nilai optimasi.

Dalam penelitian ini menggunakan pengklasifikasi Naïve Bayes dengan *Particle Swarm Optimization* sebagai seleksi fitur yang akan diterapkan untuk klasifikasi teks pada *review* perusahaan penjualan *online*. Adapun Manfaat dari penelitian ini adalah:

1. Manfaat praktis dari penelitian ini adalah sebagai bahan evaluasi untuk perusahaan agar lebih meningkatkan kualitas dari produk ataupun pelayanan.
2. Manfaat kebijakan dari penelitian ini adalah dapat digunakan sebagai bahan pertimbangan dalam pengambilan

keputusan dalam menentukan perusahaan penjualan *online* yang dipercaya.

Tujuan dari penelitian ini adalah untuk melihat seberapa besar pengaruh Particle Swarm Optimization sebagai seleksi fitur dalam menganalisa sentimen *review* perusahaan penjualan *online* dengan pengklasifikasi Naïve Bayes yang nantinya digunakan pengguna untuk mengambil keputusan dalam menentukan perusahaan penjualan *online* yang dipercaya.

Pokok permasalahan yang dibahas dalam penelitian ini dibatasi dalam ruang lingkup penelitian yaitu mencakup pada analisis sentimen *review* perusahaan penjualan *online* berdasarkan *review* yang ada di [http://www.consumeraffairs.com/online/online\\_sales.htm](http://www.consumeraffairs.com/online/online_sales.htm). menggunakan Naïve Bayes dengan menerapkan Particle Swarm Optimization sebagai seleksi fitur untuk meningkatkan akurasi.

## B. TINJAUAN PUSTAKA

### 1. Text Mining

*Text Mining* dapat didefinisikan secara luas sebagai proses pengetahuan intensif, dimana pengguna berinteraksi dengan koleksi dokumen dari waktu ke waktu dengan menggunakan seperangkat alat analisis menurut Feldman dan Sanger (2007, 1). *Text mining* mencakup kategori teks, deteksi topik, pencarian dan pengambilan, clustering dokumen dan lain-lain, setiap teknik ini dapat digunakan dalam mencari beberapa informasi dari dokumen.

### 2. Review Analisis Sentimen

Analisis sentimen adalah suatu bidang yang sedang berlangsung dalam penelitian berbasis teks. Analisis sentimen atau opini mining adalah kajian tentang cara untuk memecahkan masalah dari opini masyarakat, sikap dan emosi suatu entitas, dimana entitas tersebut dapat mewakili individu, peristiwa atau topik Medhat et al. (2014, 1093).

*Review* konsumen secara *online* mempengaruhi bagus atau tidaknya sebuah perusahaan penjualan *online*. Internet terus menjadi bagian penting dari kehidupan sehari-hari. Sekarang, tidak hanya dari anggota keluarga dan teman-teman, tetapi juga dari orang asing yang berlokasi diseluruh dunia yang mungkin telah menggunakan produk tertentu, belanja *online* disitus tertentu, mengunjungi tempat atau tujuan tertentu dan melihat film tertentu dapat menuangkan pendapat mereka secara *online*. Terdapat contoh situs *review* produk seperti

www.amazon.com. Situs ini memberikan informasi mengenai *review* produk-produk tertentu, lengkap dengan ulasan yang mampu menunjukkan kualitas dari produk tersebut.

### 3. Seleksi Fitur

Dalam metode untuk analisis sentimen berdasarkan pendekatan pembelajaran mesin (*Machine Learning*) biasanya ada ruang fitur yang sangat besar. Dengan ruang fitur besar ini, maka datanglah masalah. Metode seleksi fitur memainkan peran penting dalam analisis sentimen, sama seperti dalam tugas *text mining* lainnya. Penggunaan yang tepat dari metode seleksi fitur membantu juga memahami atribut yang relevan untuk kelas tertentu, serta meningkatkan akurasi klasifikasi menurut Koncz et al. (2011, 358).

Menurut Jhon, Kohavi, dan Pflieger dalam Chen et al. (2009, 359) ada dua jenis metode seleksi fitur dalam pembelajaran *machine learning*, yaitu *wrappers* dan *filters*.

#### a) *Wrappers*

*Wrappers* menggunakan akurasi klasifikasi beberapa algoritma sebagai fungsi evaluasinya. *Wrappers* harus menguji pengklasifikasi untuk setiap fitur bagian yang akan dievaluasi, biasanya lebih banyak waktu ketika jumlah fitur tinggi.

#### b) *Filters*

Berbeda dengan *wrappers*, *filters* melakukan seleksi fitur yang menggunakan fitur yang dipilih. Dalam mengevaluasi fitur, *filters* menggunakan matrix evaluasi yang mengukur kemampuan fitur untuk membedakan masing-masing kelas. Metode filter terdiri dari *Information gain*, *Term frequency*, *Chi-square*, *Expected cross entropy*, *Odds ratio*, *The weight of evidence of text*, *Mutual information*, dan *Gini index*.

### 4. Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) pertama kali diusulkan oleh Eberhart dan Kennedy pada tahun 1995, Particle swarm optimization (PSO) adalah jenis algoritma kecerdasan yang berasal dari perilaku kawanan burung mencari makan menurut Lu et al. (2015, 629).

Particle Swarm Optimization dapat diasumsikan sebagai kelompok burung secara mencari makanan disuatu daerah. Burung tersebut tidak tahu dimana makanan tersebut berada, tapi mereka tahu seberapa jauh makanan itu berada, strategi terbaik untuk menemukan makanan tersebut adalah dengan mengikuti burung yang terdekat dari makanan

tersebut menurut Salappa et al. (2007, 1). Particle swarm optimization banyak digunakan untuk memecahkan masalah optimasi, serta sebagai masalah seleksi fitur menurut Liu et al. (2012, 1).

### 5. Algoritma Naïve Bayes

*Bayesian Classification* didasarkan pada teorema *Bayes* yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network*. *Bayesian Classification* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar.

Teorema Bayes memiliki bentuk umum sebagai berikut:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Dalam hal ini:

$X$  = Data dengan *class* yang belum diketahui

$H$  = Hipotesis data  $X$  merupakan suatu *class* spesifik

$P(H|X)$  = Probabilitas hipotesis  $H$  berdasar kondisi  $X$  (*posteriori probability*)

$P(H)$  = Probabilitas hipotesis  $H$  (*prior probability*)

$P(X|H)$  = Probabilitas  $X$  berdasar kondisi pada hipotesis  $H$

$P(X)$  = Probabilitas dari  $X$

### 6. Evaluasi dan Validasi Klasifikasi

Untuk melakukan evaluasi pada algoritma Naïve Bayes dan Particle Swarm Optimization dilakukan beberapa pengujian menggunakan *confusion matrix* dan kurva ROC.

#### a) *Confusion Matrix*

Matriks yang menginformasikan hasil prediksi secara keseluruhan dari nilai akurasi dan untuk melihat kinerja pengklasifikasi, yaitu seberapa sering kasus *class X* yang benar diklasifikasikan sebagai *class X* atau kesalahan klasifikasi *class* yang lainnya. Ketika dataset hanya memiliki dua kelas, yaitu *class* positif dan *class* negatif, maka dapat dibuatkan tabel seperti dibawah ini menurut Bramer (2007, 174).

Tabel 1. *Class* pada *confusion matrix*

Correct classification	Classified as	
	+	-
+	true positives	false negatives
-	false positives	true negatives

Sumber: Bramer (2007, 174)

*True* positif (tp) merupakan jumlah *record* positif dalam dataset yang

diklasifikasikan positif. True negatif (tn) merupakan jumlah *record* negatif dalam dataset yang diklasifikasikan negatif. *False* positif merupakan jumlah *record* negatif dalam dataset yang diklasifikasikan positif. *False* negatif (fn) merupakan jumlah *record* positif dalam dataset yang diklasifikasikan negatif.

b) Kurva ROC (*Receiver Operating Characteristic*)

Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. Kurva ROC mengekspresikan confusion matrix. ROC adalah grafik dua dimensi dengan *false* positif sebagai garis horizontal dan *true* positif sebagai garis vertikal. Pedoman umum untuk mengklasifikasikan keakuratan pengujian menggunakan AUC menurut Gorunescu (2011, 325):

- 0.90 - 1.00 = *Excellent Classification*;
- 0.80 - 0.90 = *Good Classification*;
- 0.70 - 0.80 = *Fair Classification*;
- 0.60 - 0.70 = *Poor Classification*;
- 0.50 - 0.60 = *Failure*.

**C. METODE PENELITIAN**

Dalam penelitian ini menggunakan metode eksperimen, dimana dilakukan beberapa langkah seperti yang dijelaskan pada Gambar di bawah ini:



Sumber: Peneliti

Gambar 1. Langkah Metode Penelitian

1. Pengumpulan Data

Penelitian ini menggunakan data yang berasal dari situs web [http://www.consumeraffairs.com/online/online\\_sales.htm](http://www.consumeraffairs.com/online/online_sales.htm). Banyak *review* yang tersedia dari situs tersebut mengenai ulasan para pelanggan terhadap perusahaan belanja *online*. Data yang digunakan dalam penelitian ini sebanyak

400 data yang terdiri dari 200 data *review* positif dan 200 data *review* negatif.

2. Pengolahan Data Awal

Tahap selanjutnya adalah pengolahan data awal. Dataset yang digunakan sebanyak 400 data, 200 *review* positif dan 200 *review* negatif yang dijadikan sebagai data *training*. Dataset ini dalam tahap *preprocessing* harus melalui 3 proses, yaitu:

- a) *Tokenization*
- b) *Stopword Removal*
- c) *Stemming*

3. Metode yang Diusulkan

Metode yang penulis usulkan adalah penggunaan metode pemilihan fitur yaitu Particle Swarm Optimization (PSO) yang digunakan untuk meningkatkan akurasi pada pengklasifikasi Naïve Bayes.

4. Eksperimen dan Pengujian Metode

Penulis melakukan proses eksperimen menggunakan RapidMiner. Data *training* yang digunakan adalah dataset *review* perusahaan penjualan *online* pada situs [http://www.consumeraffairs.com/online/online\\_sales.htm](http://www.consumeraffairs.com/online/online_sales.htm), dimana dataset telah dikelompokkan menjadi 2 bagian yaitu *review* positif dan *review* negatif sedangkan untuk pengujian model dilakukan menggunakan dataset *review* perusahaan penjualan *online*. Spesifikasi komputer yang digunakan oleh penulis dapat dilihat pada tabel berikut ini:

Tabel 2. Spesifikasi Komputer yang Digunakan

Processor	AMD A6-3420M APU with Radeon™ HD Graphics 1.50 GHz
Memori	4.00 GB
Harddisk	650 GB
Sistem Operasi	Microsoft Windows 7
Aplikasi	RapidMiner 5.2

Sumber: Peneliti

5. Evaluasi dan Validasi Hasil

Pada penelitian ini validasi dilakukan dengan menggunakan *10 fold cross validation*. Akurasi diukur dengan *confusion matrix* yaitu membandingkan akurasi Naïve Bayes sebelum menggunakan pemilihan fitur dengan Naïve Bayes berbasis Particle Swarm Optimization (PSO) sebagai pemilihan fitur. Kurva ROC digunakan untuk mengukur nilai AUC.

**D. HASIL DAN PEMBAHASAN**

1. Pengumpulan Data

Penelitian ini menggunakan data yang berasal dari situs web [http://www.consumeraffairs.com/online/online\\_sales.htm](http://www.consumeraffairs.com/online/online_sales.htm). Banyak *review* yang tersedia dari situs tersebut mengenai *review* para pelanggan terhadap perusahaan penjualan *online*. Data yang digunakan dalam penelitian ini sebanyak 400 data yang terdiri dari 200 *review* positif dan 200 *review* negatif. Data tersebut masih berupa sekumpulan teks yang terpisah dalam bentuk dokumen. Data *review* positif disatukan dalam satu folder dan diberi nama positif, sedangkan data *review* negatif disatukan dalam satu folder dan diberi nama negatif.

a) Pengolahan Data Awal

1) *Tokenization*

Dalam proses ini, semua kata yang ada didalam setiap dokumen dikumpulkan dan dihilangkan tanda bacanya, serta dihilangkan jika terdapat simbol atau apapun yang bukan huruf. Berikut adalah contoh hasil dari proses *tokenization* dalam RapidMiner.

Tabel 3. Perbandingan teks sebelum dan sesudah dilakukan proses *tokenization*

Sebelum Proses <i>Tokenization</i>	<i>This is the first time I purchased anything from ItsHot and I am not happy at all. I place an order and I was told it was on back order and it would take 4 to 6 weeks to get. I decided to wait on the items. It is week nine and I still don't have my order. I have talked to different people and for 3 weeks I was told it would be in this week, and still no order. Then every time I ask for my money back, they change the subject. I had to get married last week with no rings. This does not make any sense.</i>
Setelah Proses <i>Tokenization</i>	<i>This is the first time I purchased anything from ItsHot and I am not happy at all I place an order and I was told it was on back order and it would take to weeks to get I decided to wait on the items It is week nine and I still don t have my order I have talked to different people and for weeks I was told it would be in this week and still no order Then every time I ask for my money back they change the subject I had to get married last week with no rings This does not make any sense</i>

Sumber: Peneliti

2) *Stopwords Removal*

Dalam proses ini, kata-kata yang tidak relevan akan dihapus, seperti kata *the, of, for, with* yang merupakan kata-kata yang tidak mempunyai makna tersendiri jika dipisahkan dengan kata yang lain dan tidak terkait dengan dengan kata sifat yang berhubungan dengan sentimen.

Tabel 4. Perbandingan teks sebelum dan sesudah dilakukan proses *stopwords removal*

Sebelum Proses <i>Stopwords Removal</i>	<i>This is the first time I purchased anything from ItsHot and I am not happy at all. I place an order and I was told it was on back order and it would take 4 to 6 weeks to get. I decided to wait on the items. It is week nine and I still don't have my order. I have talked to different people and for 3 weeks I was told it would be in this week, and still no order. Then every time I ask for my money back, they change the subject. I had to get married last week with no rings. This does not make any sense.</i>
Setelah Proses <i>Stopwords Removal</i>	<i>time I purchased ItsHot I happy I place order I told order take weeks get I decided wait items week nine I t order I talked people weeks I told week order time I ask money change subject I get married week rings make sense</i>

Sumber: Peneliti

3) *Stemming*

Merupakan salah satu proses dari mengubah token yang berimbuhan menjadi kata dasar, dengan menghilangkan semua imbuhan yang ada pada token tersebut. Seperti *drug, drugged, dan drugs* dimana kata dasar dari semuanya adalah kata *drug*. Pentingnya *stemming* dalam proses pembuatan sistem adalah untuk menghilangkan imbuhan pada awalan dan akhiran. Berdasarkan hasil proses tersebut, akan didapatkan sebuah informasi mengenai banyaknya fitur yang muncul dalam sebuah dokumen.

Tabel 5. Perbandingan teks sebelum dan sesudah dilakukan proses *stemming*

Sebelum Proses Stemming	<i>This is the first time I purchased anything from ItsHot and I am not happy at all. I place an order and I was told it was on back order and it would take 4 to 6 weeks to get. I decided to wait on the items. It is week nine and I still don't have my order. I have talked to different people and for 3 weeks I was told it would be in this week, and still no order. Then every time I ask for my money back, they change the subject. I had to get married last week with no rings. This does not make any sense.</i>
Setelah Proses Stemming	<i>time i purchas itshot i happi i place order i told order take week get i decid wait item week nine i t order i talk peopl week i told week order time i ask money chang subject i get marri week ring make sens</i>

Sumber: Peneliti

2. Model dengan Metode Klasifikasi Menggunakan Naïve Bayes

Proses klasifikasi disini adalah untuk menentukan sebuah kalimat sebagai anggota *class* positif atau *class* negatif berdasarkan nilai perhitungan probabilitas dari rumus Bayes. Jika hasil probabilitas kalimat tersebut untuk *class* positif lebih besar dari pada *class* negatif, maka kalimat tersebut termasuk ke dalam *class* positif. Jika probabilitas untuk *class* positif lebih kecil dari pada *class* negatif, maka kalimat tersebut termasuk ke dalam *class* negatif.

Peneliti hanya menampilkan 10 dokumen dari keseluruhan 200 data. 5 kata yang berhubungan dengan sentimen dan yang paling sering muncul yaitu *recommend*, *disappoint*, *horrible*, *good* dan *great*.

Tabel 6. Proses Klasifikasi Penentuan Sebuah Kalimat Sebagai Anggota Class Positif atau Class Negatif

Dokumen Ke-	Recommend	disappoint	horrible	good	great	Class
1	1	0	0	0	1	Positif
2	1	0	0	0	1	Positif
3	0	0	0	0	0	Positif
4	1	0	0	1	0	Positif
5	1	0	0	1	0	Positif
101	0	0	1	0	0	Negatif
102	0	0	0	1	0	Negatif
103	1	1	0	0	1	Negatif
104	0	1	0	0	1	Negatif

105	0	1	1	0	0	?
-----	---	---	---	---	---	---

Sumber: Peneliti

Probabilitas Bayes yang akan dijelaskan adalah probabilitas untuk dokumen ke 105.

1. Hitung probabilitas bersyarat (*likelihood*) dokumen ke 105 pada *class* positif dan negatif.

Untuk *class* positif:

$$P(105|positif) = P(\text{recommend}=1|positif) \times P(\text{dissappoint}=0|positif) \times P(\text{horrible}=0|positif) \times P(\text{good}=1|positif) \times P(\text{great}=1|positif)$$

$$P(105 | positif) = 4/5 \times 0/5 \times 0/5 \times 2/5 \times 1/5 = 0,8 \times 0 \times 0 \times 0,4 \times 0,2 = 0$$

Untuk *class* negatif:

$$P(105|negatif) = P(\text{recommend}=1|negatif) \times P(\text{dissappoint}=1|negatif) \times P(\text{horrible}=1|negatif) \times P(\text{good}=1|negatif) \times P(\text{great}=1|negatif)$$

$$P(105 | negatif) = 1/4 \times 3/4 \times 2/4 \times 1/4 \times 2/4 = 0,25 \times 0,75 \times 0,5 \times 0,25 \times 0,5 = 0,012$$

2. Probabilitas prior dari *class* positif dan negatif dihitung dengan proporsi dokumen pada tiap *class*:

$$P(\text{positif}) = 5/9 = 0,56$$

$$P(\text{negatif}) = 4/9 = 0,44$$

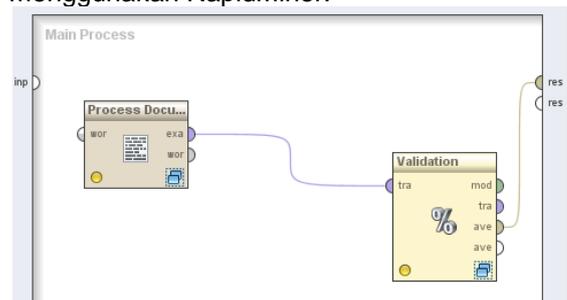
3. Hitung probabilitas posterior dengan memasukkan rumus Bayes dan menghilangkan penyebut  $P(105)$ :

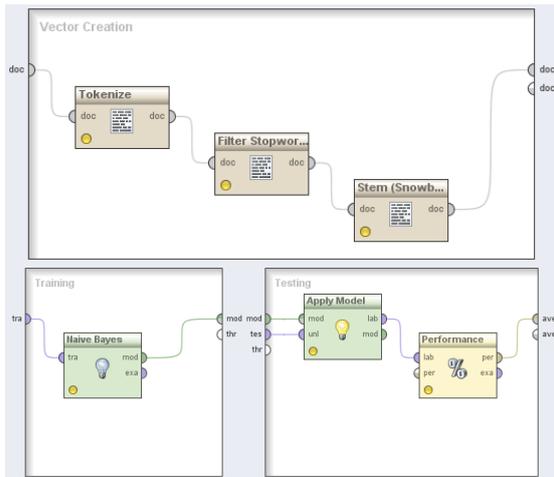
$$P(\text{positif}|105) = \frac{(0)(0,56)}{P(105)} = 0$$

$$P(\text{negatif}|105) = \frac{(0,012)(0,44)}{P(105)} = 0,00528$$

Berdasarkan probabilitas diatas dapat disimpulkan bahwa nilai dari  $P(\text{positif}|105)$  lebih kecil dari nilai  $P(\text{negatif}|105)$  maka untuk dokumen ke 105 masuk ke dalam *class* negatif.

Berikut gambar model Naïve Bayes menggunakan Rapidminer.





Sumber: Peneliti  
 Gambar 2. Desain model Naive Bayes menggunakan RapidMiner

Tabel 7. Hasil akurasi menggunakan algoritma Naive Bayes

Accuracy: 79,50% +/- 3,67% (mikro: 79,50%)			
	true negatif	true positif	class precision
pred.negatif	168	50	77,06%
pred.positif	32	150	82,42%
class recall	84,00%	75,00%	

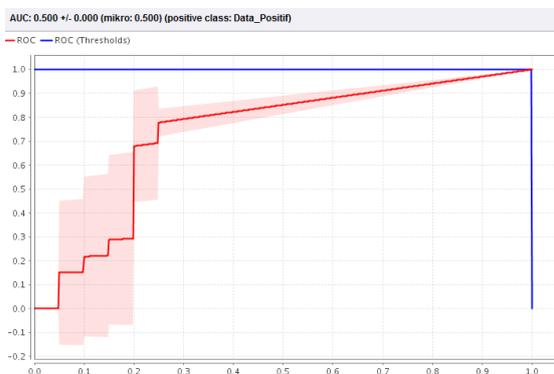
Sumber: Peneliti

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

$$Accuracy = \frac{(TN + TP)}{(TN + FN + TP + FP)}$$

$$Accuracy = \frac{(168 + 150)}{(168 + 32 + 150 + 50)}$$

$$Accuracy = \frac{318}{400} = 0,795 = 79,50\%$$



Sumber: Peneliti  
 Gambar 3. Grafik Area Under Curve (AUC) menggunakan Algoritma Naive Bayes

### 3. Hasil Pengujian Model Naive Bayes berbasis PSO

Nilai *training* dalam penelitian ini ditentukan dengan cara melakukan uji coba memasukkan nilai parameter *Population Size* dan *Inertia Weight*. Berikut ini adalah hasil dari percobaan yang telah dilakukan untuk menentukan nilai *training*:

Tabel 8. Rencana Eksperimen

Population size (Q)	Inertia Weight (w)	Naive Bayes + PSO	
		Accuracy	AUC
5-12	0.1-1.0	?	?

Sumber: Peneliti

Tabel 9. Hasil Eksperimen Menggunakan Algoritma Naive Bayes berbasis PSO dengan Mengubah Nilai dari *Population size*

Population size (Q)	Inertia Weight (w)	Naive Bayes + PSO	
		Accuracy	AUC
5	0,1	83,63	0,630
6	0,1	85,91	0,715
7	0,1	82,37	0,638
8	0,1	84,63	0,617
9	0,1	84,40	0,614
<b>10</b>	<b>0,1</b>	<b>86,12</b>	<b>0,761</b>
11	0,1	83,38	0,723
12	0,1	82,89	0,638

Sumber: Peneliti

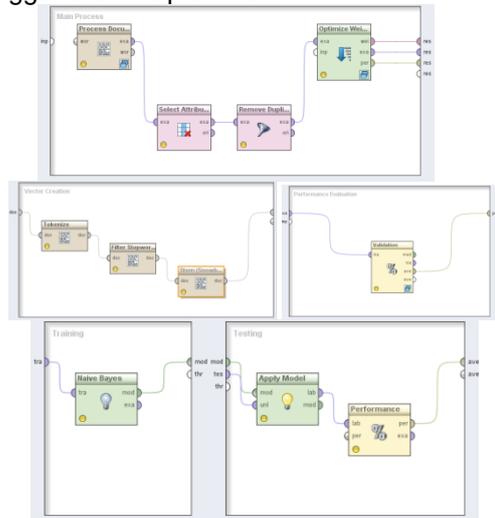
Tabel 10. Hasil Eksperimen Menggunakan Algoritma Naive Bayes berbasis PSO dengan Mengubah Nilai dari *Inertia Weight*

Population size (Q)	Inertia Weight (w)	Naive Bayes + PSO	
		Accuracy	AUC
<b>10</b>	<b>0,2</b>	<b>86,88</b>	<b>0,705</b>
10	0,3	86,65	0,662
10	0,4	85,13	0,687
10	0,5	84,13	0,760
10	0,6	84,14	0,654
10	0,7	86,20	0,740
10	0,8	83,87	0,678
10	0,9	84,65	0,678
10	1,0	84,88	0,672

Sumber: Peneliti

Hasil terbaik pada eksperimen Naive Bayes berbasis PSO adalah dengan nilai *Population size* = 10 dan nilai *Inertia Weight* = 0.3, *Accuracy* yang dihasilkan adalah 86.88% dan AUC yang dihasilkan sebesar 0.705.

Berikut gambar model Naïve Bayes berbasis Particle Swarm Optimization (PSO) menggunakan Rapidminer.



Sumber: Peneliti  
Gambar 4. Model Pengujian Naïve Bayes Berbasis PSO

Tabel 11. Model *Confusion Matrix* untuk Metode Naïve Bayes Berbasis PSO

Accuracy: 86,88% +/- 6,26% (mikro: 86,90%)			
	true negatif	true positif	class precision
pred.negatif	181	33	84,58%
pred.positif	19	164	89,62%
class recall	90,50%	83,25%	

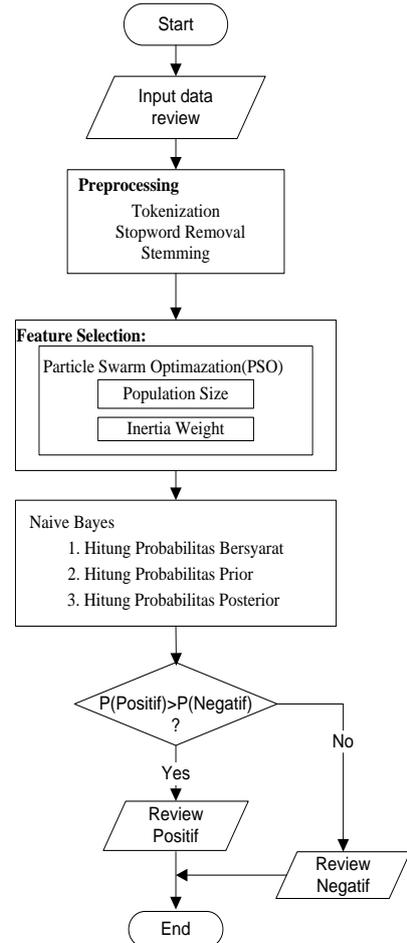
Sumber: Peneliti



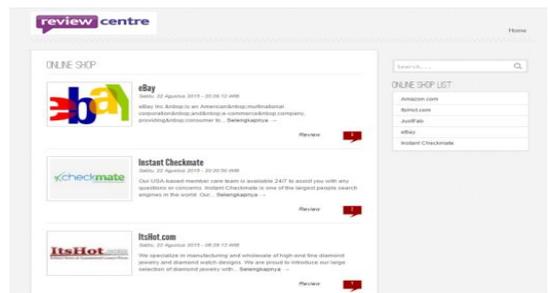
Sumber: Peneliti  
Gambar 5. Kurva ROC Naïve Bayes berbasis PSO

#### 4. Desain dan Implementasi

Dalam penelitian ini penulis membuat aplikasi untuk menguji model dengan dataset yang berbeda dan *class* yang belum diketahui. Aplikasi dibuat menggunakan Adobe Dreamweaver dengan bahasa pemrograman php dan HTML. Diagram alir pada aplikasi proses pengklasifikasi teks dapat dilihat pada gambar berikut ini:



Sumber: Peneliti  
Gambar 6. Diagram Alir Proses Klasifikasi Menggunakan Naïve Bayes berbasis PSO



Sumber: Peneliti  
Gambar 7. Tampilan rancangan aplikasi



Sumber: Peneliti  
Gambar 8. Tampilan Rancangan Aplikasi Mengklasifikasi *Review* Positif

## 5. Implikasi Penelitian

Implikasi penelitian ini mencakup beberapa aspek, diantaranya:

- a) Implikasi terhadap aspek sistem  
Hasil evaluasi menunjukkan bahwa Particle Swarm Optimization sebagai seleksi fitur dapat meningkatkan nilai akurasi Naïve Bayes dan merupakan metode yang baik sebagai pengklasifikasi teks *review* perusahaan penjualan *online* sehingga dapat mempermudah pengguna untuk membaca *review* perusahaan penjualan *online*.
- b) Implikasi terhadap aspek manajerial  
Membantu para pengembang sistem yang berkaitan dengan *review* perusahaan penjualan *online*.
- c) Implikasi terhadap aspek penelitian lanjutan  
Pada penelitian selanjutnya dapat menggunakan metode pemilihan fitur yang lain seperti *Genetic Algorithm*, *Information Gain*, *Chi Square* dan lain-lain dan juga menggunakan dataset yg berbeda seperti *review* tempat pariwisata dan lain-lain.

## E. KESIMPULAN

Dari pengolahan data yang sudah dilakukan, penggunaan metode pemilihan fitur yaitu Particle Swarm Optimization dapat meningkatkan akurasi pengklasifikasi Naïve Bayes. Data *review* perusahaan penjualan *online* dapat diklasifikasi kedalam bentuk positif dan negatif dengan baik.

Akurasi Naïve Bayes sebelum menggunakan penggabungan dengan metode pemilihan fitur mencapai 79.50%, dengan nilai AUC sebesar 0.500. Setelah menggunakan penggabungan metode pemilihan fitur akurasi meningkat menjadi 86.88%, dengan nilai AUC sebesar 0.705. Peningkatan akurasi mencapai 7.38%. Sehingga Naïve Bayes berbasis Particle Swarm Optimization pada permasalahan klasifikasi *review* perusahaan penjualan *online* terbukti memberikan nilai akurasi yang lebih akurat. Untuk mendukung penelitian, penulis mengembangkan aplikasi *review* perusahaan penjualan *online* untuk mengklasifikasikan *review* positif dan *review* negatif menggunakan bahasa pemrograman php.

## DAFTAR PUSTAKA

- [1] Apriliyanti, A. *Sentiment Analysis*. 21, 1834–1848. 2010.
- [2] F. Gorunescu. *Data mining: concepts and techniques*. 2011.
- [3] H. Kang, S. J. Yoo, and D. Han. *Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews*. *Expert Syst. Appl.*, vol. 39, no. 5, pp. 6000–6010. 2012.
- [4] H. Tang, S. Tan, and X. Cheng. *Expert Systems with Applications A survey on sentiment detection of reviews*. *Expert Syst. Appl.*, vol. 36, no. 7, pp. 10760–10773. 2009.
- [5] J. Chen, H. Huang, S. Tian, and Y. Qu, *Expert Systems with Applications Feature selection for text classification with Naïve Bayes*. *Expert Syst. Appl.*, vol. 36, no. 3, pp. 5432–5435. 2009.
- [6] Liu, B. *Sentiment Analysis and Opinion Mining*. *Synthesis Lectures on Human Language Technologies*, 5(May), 1–167. 2012.
- [7] Lu, Y., Liang, M., Ye, Z., & Cao, L. *Improved particle swarm optimization algorithm and its application in text feature selection*. *Applied Soft Computing*, 35, 629–636. 2015.
- [8] L. Yonghe, L. Minghui, Y. Zeyuan, and C. Lichao. *Improved particle swarm optimization algorithm and its application in text feature selection*. *Appl. Soft Comput. J.*, vol. 35, pp. 1–8. 2015.
- [9] M. Bramer, *Principle of Data Mining*. 2007.
- [10] P. Koncz and J. Paralic. *An approach to feature selection for sentiment analysis*. *2011 15th IEEE Int. Conf. Intell. Eng. Syst.*, pp. 357–362. 2011.
- [11] Q. Ye, Z. Zhang, and R. Law. *Sentiment classification of online reviews to travel destinations by supervised machine learning approaches*. *Expert Syst. Appl.*, vol. 36, no. 3 PART 2, pp. 6527–6535. 2009.
- [12] Salappa, A., Doumpos, M., & Zopounidis, C. *Feature Selection Algorithms in Classification Problems: An Experimental Evaluation*. *Systems Analysis, Optimization and Data Mining in Biomedicine*. 199-212. 2007.
- [13] S. Tan and J. Zhang. *An empirical study of sentiment analysis for chinese documents*. vol. 34, pp. 2622–2629. 2008.
- [14] S. L. Ting, W. H. Ip, and A. H. C. Tsang. *Is Naïve Bayes a Good Classifier for*

- Document Classification ?*. vol. 5, no. 3, pp. 37–46. 2011.
- [15] W. Medhat, A. Hassan, and H. Korashy. *Sentiment analysis algorithms and applications: A survey*. *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113. 2014.
- [16] Z. Zhang, Q. Ye, Z. Zhang, and Y. Li, *Sentiment classification of Internet restaurant reviews written in Cantonese*. *Expert Syst. Appl.*, vol. 38, no. 6, pp. 7674–7682. 2011.