

# KOMPARASI FITUR SELEKSI PADA ALGORITMA SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN REVIEW

Yoseph Tajul Arifin

AMIK BSI Bogor  
Jl. Merdeka No. 168, Bogor Jawa Barat.  
e-mail: yosephtajularifin@gmail.com

## ABSTRAK

*The main problem in the process sentiment analysis review is how to choose and use the best feature selection to get the maximal result. The accuracy of the use of algorithm in analysis sentiment review also have been an important role in the determination results of the analysis. Analysis of the sentiment is a study computing on an opinion, behavior and emotion of a person to an entity. This research also discussed comparative studies, technique classification and combining method of the feature selection to comparison result of the people opinion about tourist destination. The classifications technique to analyze sentiment review of the tourist destinations, using support vector machine algorithm (svm) and a model of the features selection will be compared between a particle swarm optimization and genetic algorithm to increase the accuracy classifications of support vector machines algorithm. The measurement of were based on accuracy support vector machines before and after the addition of features. The evaluation uses 10 cross fold validation. While the measurement of accuracy measured by confusion the matrix and a curve roc. The result showed an increase in accuracy support vector machines of 75.33 % to 88.67 %.*

**Keywords:** Sentimen Review, Support Vector Machine, Analysis Review, Feature Selection.

## 1. Pendahuluan

Opini-opini yang tersebar di berbagai media internet seperti *website*, atau situs jejaring sosial belakangan ini sering dijadikan salah satu acuan seseorang dalam membandingkan dan menentukan keputusan yang akan diambil. Banyaknya *review* yang disajikan dapat menjadi tolak ukur atau pembanding bagi si pembaca *review*. Akan tetapi dengan membaca *review* secara keseluruhan dapat memakan waktu, namun jika sedikit *review* yang dibaca hasil evaluasi dari *review* tersebut akan menjadi bias. Salah satu tujuan dari analisa sentimen adalah untuk mengatasi masalah tersebut dengan mengelompokkan/mengklasifikasikan *review* secara otomatis kedalam positif atau negatif (Z. Zhang et al., 2011).

Analisa sentimen atau *opinion mining* adalah studi komputasi mengenai pendapat, perilaku dan emosi seseorang terhadap entitas. Entitas tersebut dapat menggambarkan individu, kejadian atau topik. Topik tersebut kemungkinan besar

dapat berupa *review* (Medhat, Hassan & Korashy et al., 2014). Teknik klasifikasi yang biasa digunakan untuk analisis sentimen *review* diantaranya *Naïve Bayes*, *Support Vector Machines* (SVM) dan *K-Nearest Neighbor* (KNN) (Dehkharghani et al., 2014). Beberapa penelitian yang telah dilakukan dalam proses klasifikasi sentimen terhadap beberapa *review* yang tersedia secara *online* diantaranya, Analisa sentimen pada *review* pengguna ponsel (Zhang et al., 2014). Analisa sentimen pada opini *review* film menggunakan pengklasifikasi *Support Vector Machine* dan *Particle Swarm Optimization* (Basari et al., 2013). Klasifikasi sentimen pada *review online* tempat tujuan perjalanan menggunakan pengklasifikasi *Naïve Bayes*, *Support Vector Machines* dan *Character Based N-gram Model* (Ye, Zhang, & Law, 2009). Analisa sentimen pada *review* film dan beberapa produk dari Amazon.com menggunakan pengklasifikasi *Support Vector Machines* dan *Artificial Neural Network* (Moraes, Valiati, & Neto., 2013). Pengklasifikasian sentimen pada *review*

restoran di internet yang ditulis dalam bahasa Canton menggunakan pengklasifikasi *Naïve Bayes* dan *Support Vector Machines* (Z. Zhang et al., 2011). Analisa sentimen pada sosial media Republik Ceko menggunakan *Supervised Machine Learning* (Habernal, Ptáček, & Steinberger, 2013).

Dari beberapa penelitian tersebut diatas algoritma *Support Vector Machines* (SVM) adalah model yang sering digunakan. SVM merupakan metode *supervised learning* yang menganalisa data dan mengenali pola-pola yang digunakan untuk klasifikasi (Basari et al., 2013).

Pada penelitian ini penulis memilih 2 (dua) model fitur seleksi untuk meningkatkan hasil akurasi penelitian, yaitu *Particle Swarm Optimization* (PSO) dan *Genetic Algorithm* (GA) yang akan penulis bandingkan satu per satu terhadap algoritma *Support Vector Machine* (SVM) untuk diterapkan dalam proses klasifikasi teks pada *review* tempat wisata, yang hasilnya dapat menentukan fitur seleksi manakah yang terbaik dalam penelitian ini.

### 1.1. Teks review

Penambahan Opini merupakan area yang menarik dari penelitian karena mencakup di beberapa bidang. Mengumpulkan pendapat orang-orang tentang produk tertentu, kondisi sosial dan politik serta masalah melalui situs *web* menjadi semakin populer setiap hari. Pendapat pengguna bermanfaat untuk umum dan bagi para pemangku kepentingan dalam membuat keputusan tertentu. Penambahan opini adalah cara untuk mengambil informasi melalui mesin pencari, situs blog dan sosial media. Karena besar jumlah tinjauan dalam bentuk teks yang tidak terstruktur adalah mustahil untuk meringkas informasi tersebut secara manual. Dengan demikian, metode komputasi dibutuhkan untuk pertambahan dan meringkas *review* dari *corpuses* dan dokumen *web* secara efisien (Khan et al., 2014).

### 1.2. Analisa sentiment

Analisa sentimen pada *review* adalah proses menyelidiki *review* produk di internet untuk menentukan opini atau perasaan terhadap suatu produk secara keseluruhan (Haddi, Liu, & Shi, 2013). Analisa sentimen diperlakukan sebagai suatu tugas klasifikasi yang mengklasifikasikan orientasi suatu teks ke dalam positif atau negatif (Haddi, Liu, &

Shi, 2013), tujuan dari analisa sentimen adalah untuk menentukan perilaku atau opini dari seorang penulis dengan memperhatikan suatu topik tertentu. Perilaku bisa mengindikasikan alasan, opini atau penilaian, kondisi kecenderungan bagaimana si penulis ingin mempengaruhi pembaca (Basari et al., 2013).

### 1.3. Pemilihan Fitur

Seleksi fitur adalah salah satu faktor yang paling penting yang dapat mempengaruhi tingkat akurasi klasifikasi karena jika dataset berisi sejumlah fitur, dimensi ruang akan menjadi besar, merendahkan tingkat akurasi klasifikasi (Liu et al., 2011). Seleksi fitur mempengaruhi beberapa aspek yaitu pola klasifikasi, akurasi klasifikasi, waktu yang diperlukan untuk pembelajaran fungsi klasifikasi, jumlah contoh yang dibutuhkan untuk pembelajaran dan biaya yang terkait dengan fitur menurut Yang dan Honavar dalam Zhao (Zhao, et al., 2011).

### 1.4. Support Vector Machine

*Support Vector Machines* (SVM) adalah seperangkat metode yang terkait untuk suatu metode pembelajaran, untuk kedua masalah klasifikasi dan regresi (Maimon, 2010). Dengan berorientasi pada tugas, kuat, sifat komputasi yang mudah dikerjakan, SVM telah mencapai sukses besar dan dianggap sebagai *state-of-the-art classifier* saat ini (Huang, et al., 2008). Dua kelas data yang digambarkan sebagai lingkaran dan padat titik-titik yang disajikan di angka ini. Secara intuitif diamati, ada banyak keputusan *hyperplanes* yang dapat digunakan untuk memisahkan kedua kelompok data.

### 1.5. Tinjauan Studi

Beberapa penelitian yang telah dilakukan menggunakan algoritma yang sama yaitu *Support Vector Machine* sebagai pengklasifikasi dalam klasifikasi teks sentimen pada *review online*. Tiga penelitian yang dapat dijadikan sebagai penelitian terkait diantaranya penambahan opini *review* film menggunakan metode *Support Vector Machines* dan *Particle Swarm Optimization* (Basari et al., 2013), mengoptimalkan parameter *Support Vector Machine* menggunakan algoritma genetika untuk klasifikasi kecenderungan sengketa di tahap awal proyek kemitraan publik-swasta (Chou et al., 2014), Klasifikasi sentimen pada *review online* tempat tujuan perjalanan menggunakan pengklasifikasi *Naïve Bayes*,

*Support Vector Machines* dan *Character Based N-gram Model* (Ye, Zhang, & Law, 2009). Model dan hasil ketiga penelitian ini akan dijadikan landasan dalam penelitian yang dilakukan, lebih jelasnya akan penulis uraikan sebagai berikut:

Tabel 1. Perbandingan Penelitian Terkait

Judul	Pre Processing	Feature Selection	Classifier	Accuracy
Optimal Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization (Basri et al., 2013)	a. Edit Data b. Data Cleaning c. Extract to Text File	a. Case Normalization b. Tokenization c. Stemming d. Generate N-Grams e. Particle Swarm Optimization (PSO)	Support Vector Machine (SVM)	74,20 %
Optimizing parameters of support vector machine using fast multi genetic algorithm for aspect classification (Chen et al., 2014)	a. Initialize competitive template b. Periodical initialization	a. Threshold selection b. Building blocks filter c. Genetic Algorithm	Support Vector Machine (SVM)	89,30 %
Sentiment classification of online reviews to travel destinations by supervised machine learning approach (Ye et al., 2009)	Converted all characters to lowercase	N-Grams	Support Vector Machine (SVM)	73,07 %
Model yang diusulkan	a. Tokenization b. Stopword Removal c. Stemming d. Generate N-Gram	Particle Swarm Optimization (PSO) Genetic Algorithm (GA)	Support Vector Machine (SVM)	?

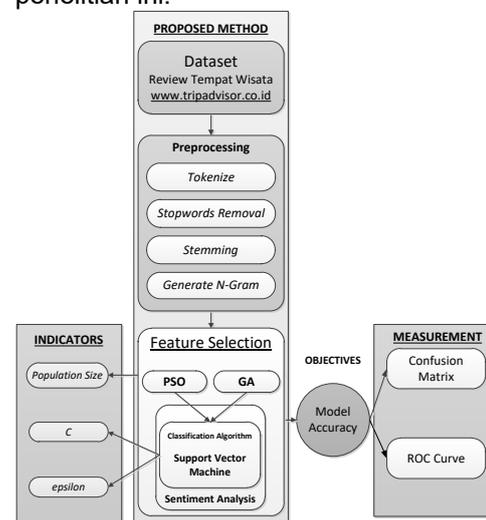
## 1.6. Kerangka Pemikiran

Pada penelitian ini terdapat beberapa tahapan berikut :

- Tahap pertama pada penelitian ini dimulai dari pengumpulan data *review* dari website [www.tripadvisor.co.id](http://www.tripadvisor.co.id). Data diambil secara random dan selanjutnya dikelompokkan kedalam masing-masing data *review* positif sebanyak 150 data dan data *review* negatif sebanyak 150 data.
- Tahap selanjutnya yaitu pengolahan data awal yang diawali dengan tahap *Preprocessing* yang dilakukan dengan proses *Tokenization*, *Stemming*, *Stopword Removal* dan *Generate N-Grams*.
- Tahap berikutnya proses klasifikasi dengan model algoritma *Support Vector Machine* (SVM), model ini dipilih berdasarkan kesesuaian data dengan metode yang paling baik dari beberapa metode pengklasifikasian teks yang sudah digunakan oleh beberapa peneliti sebelumnya.
- Tahap selanjutnya menambahkan model pembobotan dengan *Term Frequency Invers Document Frequency* (TF-IDF) dan pemilihan fitur seleksi. Karena proses pengklasifikasian tersebut memiliki kekurangan terhadap masalah pemilihan parameter yang

sesuai. Karena dengan tidak sesuainya sebuah pengaturan parameter dapat menyebabkan hasil klasifikasi menjadi rendah. Model fitur seleksi yang akan digunakan pertama kali adalah *Particle Swarm Optimization* (PSO) sedangkan seleksi fitur kedua menggunakan *Genetic Algorithm* (GA). Klasifikasi yang digunakan adalah *Support Vector Machine* dengan pengujian *10 Fold Cross Validation*.

- Hasil akurasi algoritma akan digambarkan ke dalam *Confusion Matrix* dan kurva ROC. Rapid Miner Versi 5 digunakan sebagai alat bantu dalam mengukur akurasi data eksperimen yang dilakukan dalam penelitian.
- Gambar 1 menggambarkan kerangka pemikiran yang penulis usulkan dalam penelitian ini.



Gambar 1. Kerangka Pemikiran

## 2. Metode Penelitian

### 2.1. Rancangan Penelitian

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

- Pengumpulan Data.**  
Data untuk eksperimen penelitian ini dikumpulkan dari situs [www.tripadvisor.com](http://www.tripadvisor.com), kemudian dikelompokkan dari hasil penyeleksian *dataset* yang diambil secara acak.
- Pengolahan Awal Data.**  
Model dipilih berdasarkan kesesuaian data dengan metode yang paling baik dari beberapa metode pengklasifikasian teks yang sudah digunakan oleh beberapa peneliti sebelumnya. Pada penelitian ini model yang digunakan

- adalah algoritma *Support Vector Machine* (SVM).
- c. Metode Yang Diusulkan  
Untuk meningkatkan akurasi dari algoritma *Support Vector Machine* (SVM), maka dilakukan penambahan model seleksi fitur *Particle Swarm Optimization* (PSO), dan *Genetic Algorithm* (GA) pada Algoritma tersebut. Kedua seleksi fitur tersebut penulis uji satu per satu dengan algoritma *Support Vector Machine* (SVM) dan kemudian penulis bandingkan hasil keduanya untuk diketahui metode peningkatan optimasi (seleksi fitur) terbaik yang dapat digunakan untuk meningkatkan akurasi pada penelitian ini.
  - d. Eksperimen dan Pengujian Metode.  
Dalam hal eksperimen data penelitian, penulis menggunakan Rapid Miner versi 5 untuk mengolah data *review* dan sebagai alat bantu dalam mengukur akurasi data eksperimen yang dilakukan dalam penelitian.
  - e. Evaluasi dan Validasi Hasil.  
Evaluasi dilakukan untuk mengetahui akurasi dari model algoritma *Support Vector Machine* (SVM). Validasi digunakan untuk melihat perbandingan hasil akurasi dari model yang digunakan dengan hasil yang telah ada sebelumnya. Teknik validasi yang digunakan adalah *Cross Validation*, akurasi algoritma akan diukur menggunakan *Confusion Matrix* dan hasil olahan data dalam bentuk kurva ROC.

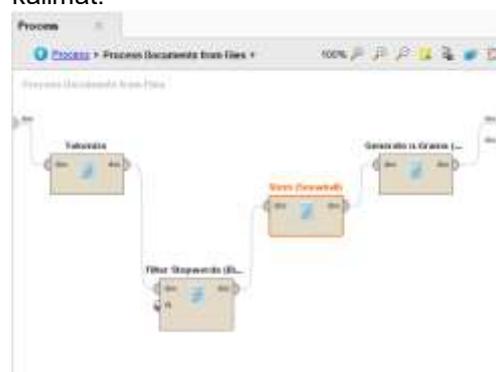
## 2.2. Pengolahan Awal Data

Pada penelitian ini penulis hanya menggunakan 150 *review* positif dan 150 *review* negatif sebagai data *training*. Tahap pertama pengolahan data adalah tahap *preprocessing* yang mana pada tahap ini ada 4 (empat) proses tahapan pengolahan, diantaranya:

- a. *Tokenization*  
*Tokenization* yaitu mengumpulkan semua kata yang muncul dan menghilangkan tanda baca maupun simbol apapun yang bukan huruf.
- b. *Stopword Removal*  
Yaitu penghapusan kata-kata yang tidak relevan, seperti kata sambung dan lain lain seperti *the*, *of*, *for*, *with* dan sebagainya.
- c. *Stemming*  
Yaitu mengelompokkan kata ke dalam beberapa kelompok yang memiliki kata

dasar yang sama. Sedangkan untuk tahap *transformation* dengan melakukan pembobotan TF-IDF pada masing-masing kata. Di mana prosesnya menghitung kehadiran atau ketidakhadiran sebuah kata di dalam dokumen. Berapa kali sebuah kata muncul di dalam suatu dokumen juga digunakan sebagai skema pembobotan dari data tekstual.

- d. *Generate N-Gram*  
*Generate N-Gram* yaitu proses menghitung probabilitas bersyarat untuk sebuah kata dari urutan kata sebelumnya. Sebuah *n-gram* adalah sebuah kumpulan kata dengan masing-masing memiliki panjang *n* kata. *N-gram* ukuran 1 (satu) disebut sebagai *unigram*, ukuran 2 (dua) sebagai *bigram*, ukuran 3 (tiga) sebagai *trigram*, dan seterusnya. Proses ini dimulai dengan memecah kata per kata dan mengelompokkan hasil pemecahan kata tersebut kedalam *n-gram* ukuran 1 (satu), *n-gram* ukuran 2 (dua), *n-gram* ukuran 3 (tiga) dan seterusnya untuk dilakukan perhitungan dari kata yang sering muncul pada suatu kalimat.



Gambar 2. Tahap Preprocessing

## 3. Hasil dan Pembahasan

### 3.1. Tokenization

Proses *tokenization* berfungsi untuk menghilangkan tanda baca, *symbol* dan katakter yang bukan berupa huruf pada setiap dokumen *review*.

Tabel 2.

Perbandingan teks sebelum dan sesudah dilakukan proses *Tokenization*

Teks sebelum dilakukan proses tokenization	Bagusnya di Tidung ini adalah perjalanan menuju ke Tidungnya itu sendiri walaupun terasa lama juga (sekitar dua jam ) namun mengasyikkan. Sayangnya di Tidung sudah ramaaaii sekali baik ramai orangnya maupun ramai kendaraan. Jadinya tidak terlalu nyaman
Teks setelah dilakukan proses tokenization	Bagusnya di Tidung ini adalah perjalanan menuju ke Tidungnya itu sendiri walaupun terasa lama juga sekitar dua jam namun mengasyikkan. Sayangnya di Tidung sudah ramaaaii sekali baik ramai orangnya maupun ramai kendaraan. Jadinya tidak terlalu nyaman

Perbandingan teks sebelum dan sesudah dilakukan proses *Generate N Gram*

Teks sebelum dilakukan proses generate n gram	bagusnya tidung perjalanan tidungnya jam mengasyikkan sayangnya tidung ramaaaii ramai orangnya ramai kendaraan nyaman
Teks setelah dilakukan proses generate n gram	bagusnya bagusnya tidung tidung tidung perjalanan perjalanan perjalanan tidungnya tidungnya tidungnya jam jam jam mengasyikkan mengasyikkan mengasyikkan sayangnya sayangnya sayangnya tidung tidung tidung ramaaaii ramaaaii ramaaaii ramai ramai ramai orangnya orangnya orangnya ramai ramai ramai kendaraan kendaraan kendaraan nyaman nyaman nyaman

3.2. **Stopword Removal**

Yaitu penghapusan kata-kata yang tidak relevan, seperti kata sambung dan lain lain seperti *the, of, for, with* dan sebagainya.

Tabel 3.

Perbandingan teks sebelum dan sesudah dilakukan proses *Stopword Removal*

Teks sebelum dilakukan proses stopword removal	Bagusnya di Tidung ini adalah perjalanan menuju ke Tidungnya itu sendiri walaupun terasa lama juga sekitar dua jam namun mengasyikkan. Sayangnya di Tidung sudah ramaaaii sekali baik ramai orangnya maupun ramai kendaraan. Jadinya tidak terlalu nyaman
Teks setelah dilakukan proses stopword removal	Bagusnya Tidung perjalanan Tidungnya jam mengasyikkan. Sayangnya Tidung ramaaaii ramai orangnya ramai kendaraan nyaman

3.3. **Stemming**

Proses pengelompokkan kata ke dalam beberapa kelompok yang memiliki kata dasar yang sama dan melakukan *transformation* untuk proses pembobotan dengan melakukan penghitungan terhadap kehadiran atau ketidakhadiran sebuah kata di dalam dokumen.

Tabel 4.

Perbandingan teks sebelum dan sesudah dilakukan proses *Stemming*

Teks sebelum dilakukan proses stemming	Bagusnya Tidung perjalanan Tidungnya jam mengasyikkan. Sayangnya Tidung ramaaaii ramai orangnya ramai kendaraan nyaman
Teks setelah dilakukan proses stemming	bagusnya tidung perjalanan tidungnya jam mengasyikkan. sayangnya tidung ramaaaii ramai orangnya ramai kendaraan nyaman

3.4. **Generate n-gram**

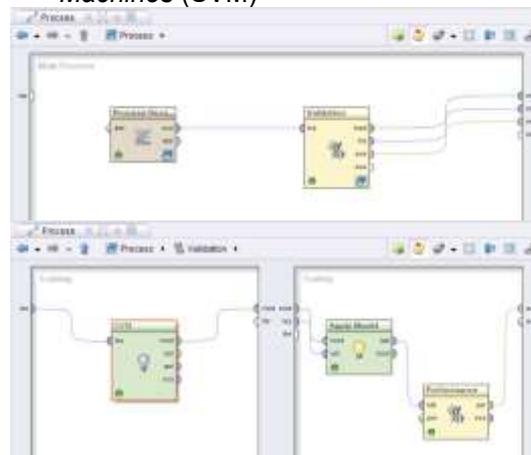
*Generate N-Gram* yaitu proses menghitung probabilitas bersyarat untuk sebuah kata dari urutan kata sebelumnya. Sebuah *n-gram* adalah sebuah kumpulan kata dengan masing-masing memiliki panjang n kata. *N-gram* ukuran 1 disebut sebagai unigram, ukuran 2 sebagai bigram, ukuran 3 sebagai trigram, dan seterusnya.

Tabel 5.

3.5. Hasil Eksperimen Pengujian Metode.

a. **Model Support Vector Machine (SVM)**

1) Model Klasifikasi *Support Vector Machines (SVM)*



Gambar 3. Model Klasifikasi SVM

- 2) Hasil terbaik pada eksperimen SVM di atas adalah dengan  $C = 0.0$  dan  $Epsilon = 0.7$  serta  $population\ size = 5$  dihasilkan *Accuracy* 75.33% dan *AUC* = 0.975 .
- 3) Hasil Eksperimen Nilai *accuracy, precision* dan *recall* dari data *training* dapat dihitung dengan menggunakan RapidMiner.

Tabel 6. Eksperimen Penentuan Nilai *Training Cycles SVM*

C	Epsilon	SVMINDO		
		Accuracy	AUC	Time
0.0	0.0	64.67%	0.976	00:00:21
0.0	0.1	73.00%	0.974	00:00:21
0.0	0.2	67.33%	0.976	00:00:22
0.0	0.3	71.00%	0.973	00:00:23
0.0	0.4	72.00%	0.976	00:00:28
0.0	0.5	74.67%	0.971	00:00:26
0.0	0.6	74.00%	0.974	00:00:26
0.0	0.7	75.33%	0.975	00:00:24
0.0	0.8	73.67%	0.974	00:00:26
0.0	0.9	72.33%	0.975	00:00:25
0.0	1.0	50.00%	0.500	00:00:23

- 4) **Confusion Matrix**  
Hasil uji terbaik pada pengklasifikasian data review menggunakan Algoritma

Support Vector Machine (SVM) dapat dilihat pada gambar berikut :

Tabel 7. Model *Confusion Matrix* Untuk Metode SVM

Accuracy : 75.33 %, +/- 8.19 % (Mikro : 75.33 %)			
	True Positif	True Negatif	Class Precision
Prediksi Positif	79	3	96.34 %
Prediksi Negatif	71	147	67.43 %
Class Recall	52.67%	98.00 %	

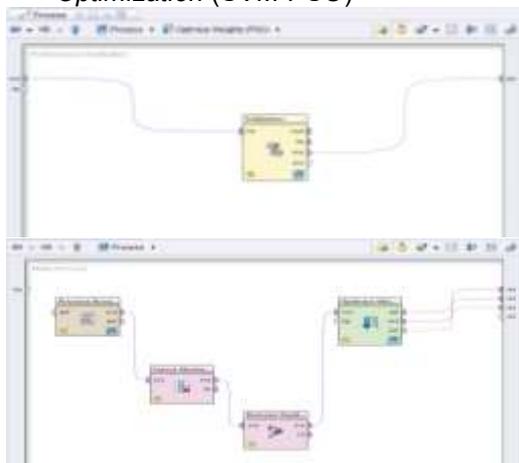
Hasil perhitungan yang divisualisasikan dengan kurva ROC dapat dilihat pada gambar 4.2 yang mengekspresikan confusion matrix dari tabel 4.6. Garis horizontal adalah *false positive* dan garis vertikal *true positive*.



Gambar 4. Kurva ROC SVM

**b. Model SVM-PSO**

1) Model Klasifikasi *Support Vector Machines* berbasis *Particle Swarm Optimization* (SVM-PSO)



Gambar 5. Model Klasifikasi SVM-PSO

Hasil terbaik pada eksperimen SVM berbasis PSO di atas adalah dengan C=0.1 dan Epsilon=0.8 serta population size=5 yang dihasilkan accuracy 88.67% dan AUC nya 0.959.

2) Hasil Eksperimen Nilai accuracy, precision dan recall dari data training dapat dihitung dengan menggunakan RapidMiner.

Tabel 8. Eksperimen Penentuan Nilai *Training Cycles* SVM-PSO

C	Epsilon	SVM PSO		
		Accuracy	AUC	Time
0.0	0.0	72.33%	965	01:01:53
0.0	0.1	85.00%	961	01:13:07
0.0	0.2	78.67%	974	01:00:06
0.0	0.3	80.33%	960	01:09:57
0.0	0.4	82.00%	964	01:05:26
0.0	0.5	80.33%	969	01:13:12
0.0	0.6	79.33%	970	01:06:16
0.0	0.7	80.33%	970	01:03:12
0.0	0.8	86.67%	967	01:00:53
0.0	0.9	83.67%	943	01:04:19
0.1	0.8	88.67%	959	01:00:21

3) *Confusion Matrix*

Hasil uji terbaik pada pengklasifikasian data review menggunakan Algoritma *Support Vector Machine* (SVM) dapat dilihat pada gambar berikut :

Tabel 9. Model *Confusion Matrix* Untuk Metode SVM-PSO

Accuracy : 88.67 %, +/- 7.48 % (Mikro : 86.67 %)			
	True Positif	True Negatif	Class Precision
Prediksi Positif	123	7	94.62 %
Prediksi Negatif	27	143	84.12 %
Class Recall	82.00%	95.33 %	

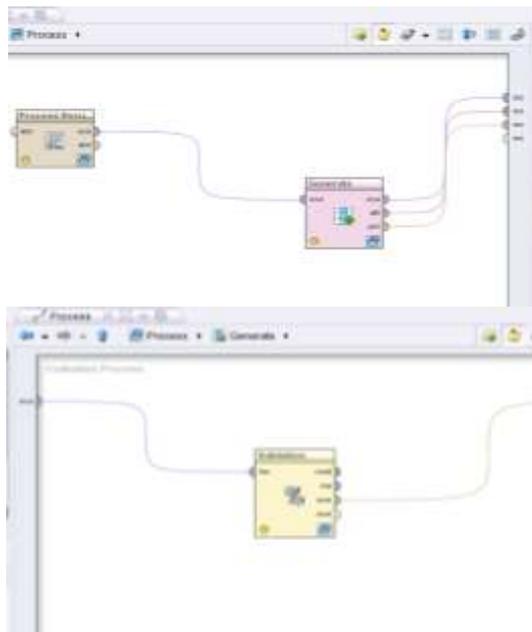
Hasil perhitungan yang divisualisasikan dengan kurva ROC dapat dilihat pada gambar 4.4 yang mengekspresikan confusion matrix dari tabel 4.8. Garis horizontal adalah *false positive* dan garis vertikal *true positive*.



Gambar 6. Kurva ROC SVM-PSO

**c. Model SVM-GA**

1) Model Klasifikasi *Support Vector Machines* berbasis *Generate Alogarithm* (SVM-GA)



Gambar 7. Model Klasifikasi SVM-GA

Hasil terbaik pada eksperimen SVM berbasis GA di atas adalah dengan  $C=0.0$  dan  $Epsilon=0.9$  serta  $population\ size=5$  yang dihasilkan  $accuracy\ 84.33\%$  dan  $AUC\ nya\ 0.936$ .

- 2) Hasil Eksperimen  
 Nilai *accuracy*, *precision* dan *recall* dari data *training* dapat dihitung dengan menggunakan RapidMiner. Hasil pengujian dengan menggunakan *model Support Vector Machine* berbasis *Genetic Algorithm* didapatkan hasil pada tabel 4.9.

Tabel 10.  
 Eksperimen Penentuan Nilai *Training Cycles* SVM – GA

C	Epsilon	SVM GA		
		Accuracy	AUC	Time
0.0	0.0	80.33%	0.896	00:31:07
0.0	0.1	84.00%	0.938	00:35:20
0.0	0.2	78.00%	0.928	00:27:31
0.0	0.3	76.33%	0.875	00:26:50
0.0	0.4	80.33%	0.884	00:30:29
0.0	0.5	79.33%	0.912	00:27:34
0.0	0.6	80.00%	0.890	00:29:23
0.0	0.7	79.33%	0.953	00:28:45
0.0	0.8	80.33%	0.888	00:30:20
0.0	0.9	84.33%	0.936	00:39:31
0.0	1.0	50.00%	0.500	00:22:16

- 3) *Confusion Matrix*  
 Hasil uji terbaik pada pengklasifikasian data *review* menggunakan Algoritma *Support Vector Machine* (SVM) berbasis *Genetic Algorithm* (GA) dapat di lihat pada gambar berikut:

Tabel 11.

Model *Confusion Matrix* Untuk Metode SVM Berbasis GA

Accuracy : 84.33 % , +/- 5.97 % (Mikro : 84.33 %)			
	True Positif	True Negatif	Class Precision
Prediksi Positif	108	5	95.58%
Prediksi Negatif	42	145	77.54 %
Class Recall	72.00%	96.67 %	

Hasil perhitungan yang divisualisasikan dengan kurva ROC dapat di lihat pada gambar 4.6 yang mengekspresikan *confusion matrix* dari tabel 4.10. Garis horizontal adalah *false positive* dan garis vertikal *true positive*.



Gambar 8. Kurva ROC SVM-GA

### 3.6. Analisis Evaluasi Hasil dan Validasi Model

Dari hasil pengujian di atas, pengukuran akurasi menggunakan *confusion matrix* dan kurva ROC terbukti bahwa hasil pengujian algoritma SVM berbasis PSO memiliki nilai akurasi yang lebih tinggi dibandingkan dengan algoritma SVM dan SVM berbasis GA. Setelah dilakukan penyesuaian pada parameter C dan Epsilon didapatkan nilai akurasi untuk hasil pengujian data *review* dengan model algoritma SVM sebesar 75.33%, untuk nilai akurasi dari hasil pengujian data *review* dengan model algoritma SVM berbasis PSO sebesar 88.67%, nilai akurasi untuk hasil pengujian data *review* dengan model algoritma SVM berbasis GA sebesar 88.67%.

Perbandingan hasil pengujian terbaik dapat di lihat pada table 4.11 dan gambar 4.6 di bawah ini :

Tabel 12. Hasil Pengujian Terbaik

Algoritma	Accuracy	AUC
SVM	75.33%	0.975
SVM Berbasis PSO	88.67%	0.959
SVM Berbasis GA	84.33%	0.936



Gambar 9.  
Perbandingan Hasil Pengujian Terbaik

#### 4. Kesimpulan

Hasil pengujian dan perbandingan model algoritma dan fitur seleksi menunjukkan bahwa model algoritma *Support Vector Machine* menghasilkan tingkat akurasi pengujian sebesar 75.33%, sedangkan model algoritma *Support Vector Machine* yang telah dikombinasikan dengan fitur seleksi *Particle Swarm Optimization* menunjukkan tingkat akurasi sebesar 88.67%, lain halnya dengan model algoritma *Support Vector Machine* yang dikombinasikan dengan fitur seleksi *Genetic Algorithm* menunjukkan hasil akurasi sebesar 84.33%. Hal ini berarti bahwa hasil pengujian terbaik dari semua pengujian yang telah dilakukan dalam penelitian ini adalah model Algoritma *Support Vector Machine* yang dikombinasikan dengan fitur seleksi *Particle Swarm Optimization* dengan selisih akurasi dari pengujian model algoritma *Support Vector Machine* sebelumnya yaitu 13.34%.

Dengan demikian model algoritma *Support Vector Machine* berbasis *Particle Swarm Optimization* (SVM-PSO) adalah model algoritma terbaik dalam penelitian ini dan dapat memberikan hasil maksimal dalam pengujian dan pengklasifikasian analisis *sentiment review* dengan objek yang diteliti adalah opini/review tempat wisata dibandingkan dengan model algoritma *Support Vector Machine* (SVM) dan *Support Vector Machine Berbasis Genetic Algorithm* (SVM-GA) yang telah diujikan sebelumnya.

##### 4.1. Saran

Agar penelitian ini bisa ditingkatkan, berikut adalah saran-saran yang diusulkan:

- Menggunakan metode pemilihan fitur yang lain, seperti *Chi Square*, *Information Gain*, *Gini Index*, *Mutual Information*, dan lain-lain agar hasilnya bisa dibandingkan

- Menggunakan pengklasifikasi lain yang mungkin di luar *Supervised learning*. Sehingga bisa dilakukan penelitian yang berbeda dari umumnya yang sudah ada.
- Menggunakan data *review* dari domain yang berbeda tetapi tetap berkaitan, seperti *review hotel*, *review transportasi wisata*, *review kuliner tempat wisata*, dan sebagainya.

#### Referensi

- Basari, A. S. H., Hussin, B., Ananta, I. G. P., & Zeniarja, J. (2013). Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization. *Procedia Engineering*, 53, 453–462. doi:10.1016/j.proeng.2013.02.059
- Chou, J.-S., Cheng, M.-Y., Wu, Y.-W., & Pham, A.-D. (2014). Optimizing parameters of support vector machine using fast messy genetic algorithm for dispute classification. *Expert Systems with Applications*, 41(8), 3955–3964. doi:10.1016/j.eswa.2013.12.035
- Dehkharghani, R., Mercan, H., Javeed, A., & Saygin, Y. (2014). Sentimental causal rule discovery from Twitter. *Expert Systems with Applications*, 41(10), 4950–4958. doi:10.1016/j.eswa.2014.02.024
- Habernal, I., Ptáček, T., & Steinberger, J. (2014). Supervised sentiment analysis in Czech social media. *Information Processing & Management*, 50(5), 693–707. doi:10.1016/j.ipm.2014.05.001
- Haddi, E., Liu, X., & Shi, Y. (2013). The Role of Text Pre-processing in Sentiment Analysis. *Procedia Computer Science*, 17, 26–32. doi:10.1016/j.procs.2013.05.005
- Han, J., & Kamber, M. (2007). *Data Mining Concepts and Techniques*. San Francisco: Diane Cerra.
- Haupt, R. L., & Haupt, S. E. (2004). *Practical Genetic Algorithms*. United States Of America: A John Wiley & Sons Inc Publication.
- Huang, K., Yang, H., King, I., & Lyu, M. (2008). *Machine Learning Modeling*

- Data Locally And Globally*. Berlin Heidelberg: Zhejiang University Press, Hangzhou And Springer-Verlag GmbH.
- Khan, F. H., Bashir, S., & Qamar, U. (2014). TOM: Twitter opinion mining framework using hybrid classification scheme. *Decision Support Systems*, 57, 245–257. doi:10.1016/j.dss.2013.09.004
- Khan, K., Baharudin, B., & Khan, A. (2014). Mining Opinion Components from Unstructured Reviews: A Review. *Journal of King Saud University - Computer and Information Sciences*. doi:10.1016/j.jksuci.2014.03.009
- Litchfield, Steve. (2010). Defining the Smartphone. (n.d.)
- Liu, Y., Wang, G., Chen, H., Dong, H., Zhu, X., & Wang, S. (2011). An Improved Particle Swarm Optimization for Feature Selection. *Journal of Bionic Engineering*, 8(2), 191–200. doi:10.1016/S1672-6529(11)60020-6
- Maimon, O. (2010). *Data Mining And Knowledge Discovery Handbook*. New York Dordrecht Heidelberg London: Springer.
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*. doi:10.1016/j.asej.2014.04.011
- Moraes, R., Valiati, J. F., & Gavião Neto, W. P. (2013). Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Systems with Applications*, 40(2), 621–633. doi:10.1016/j.eswa.2012.07.059
- Review Text. (n.d). Juli 15, 2015. <http://www.ef.co.id/englishfirst/englishstudy/pengertian-dan-struktur-review-text.aspx>
- Trend Bisnis 2014. (2015). <http://www.ekonomiholic.com/2014/01/trend-bisnis-2014.html?m=0>
- Vercellis, C. (2009). *Business Intelligence Data Mining And Optimization For Decision Making*. United Kingdom: A John Wiley And Sons, Ltd., Publication.
- Weiss, S. M., Indurkha, Nitin & Zhang, Tong. (2010). *Fundamentals of Predictive Text Mining*. London: Springer-Verlag
- Witten, H. I., Frank, E., & Hall, M. A. (2011). *Data Mining Practical Machine Learning Tools And Technique*. Burlington: Elsevier Inc
- 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), 6527–6535. doi:10.1016/j.eswa.2008.07.035
- Zhang, L., Hua, K., Wang, H., Qian, G., & Zhang, L. (2014). Sentiment Analysis on Reviews of Mobile Users. *Procedia Computer Science*, 34, 458–465. doi:10.1016/j.procs.2014.07.013
- Zhang, Z., Ye, Q., Zhang, Z., & Li, Y. (2011). Sentiment classification of Internet restaurant reviews written in Cantonese. *Expert Systems with Applications*, 38(6), 7674–7682. doi:10.1016/j.eswa.2010.12.147
- Zhao, M., Fu, C., Ji, L., Tang, K., & Zhou, M. (2011). Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes. *Expert Systems with Applications*, 38(5), 5197–5204. doi:10.1016/j.eswa.2010.1