

## Penerapan Algoritma Genetika Untuk Seleksi Fitur Pada Analisis Sentimen Review Jasa Maskapai Penerbangan Menggunakan Naive Bayes

Risa Wati

Program Studi Manajemen Informatika, AMIK BSI Tasikmalaya  
risawati06@gmail.com

**Abstract** - The quality of airline services can be seen from any opinions or reviews on passengers before. This reviewer classification grouped into positive opinion and a negative opinion. Data mining classification algorithm used is Naive Bayes are widely used in research because it serves well as a text classifier method however has the disadvantage that is very sensitive in the selection of features. Genetic Algorithm (GA) is one of the optimization algorithms, which is invented to mimic some of the processes observed in natural evolution. After testing the two models, namely models Naive Bayes algorithm and Naive Bayes algorithm based on the results obtained GA is Naive Bayes algorithm produces an accuracy of 60.00% while for Naive Bayes algorithm based on GA value amounted to 89.50% accuracy. Difference in value by 29.5% accuracy and included into the category of excellent classification.

**Keywords:** Sentiment Analysis, Review, Naive Bayes, Text Classification

**Abstrak** - Kualitas layanan maskapai dapat dilihat dari pendapat atau review penumpang sebelumnya. Klasifikasi resensi ini dikelompokkan menjadi opini positif dan pendapat negatif. algoritma klasifikasi data mining yang digunakan adalah Naive Bayes secara luas digunakan dalam penelitian karena berfungsi juga sebagai metode classifier teks namun memiliki kelemahan yang sangat sensitif dalam pemilihan fitur. Algoritma genetik (GA) merupakan salah satu algoritma optimasi, yang diciptakan untuk meniru beberapa proses yang diamati dalam evolusi alam. Setelah menguji dua model, yaitu model algoritma Naive Bayes dan algoritma Naive Bayes berdasarkan hasil yang diperoleh GA adalah algoritma Naive Bayes menghasilkan akurasi 60,00% sedangkan untuk algoritma Naive Bayes berdasarkan nilai GA sebesar akurasi 89,50%. Selisih nilai dengan akurasi 29,5% dan termasuk ke dalam kategori klasifikasi sangat baik.

**Kata Kunci:** Analisis Sentimen, Ulasan, Naif Bayes, Klasifikasi Teks

### A. PENDAHULUAN

Seiring perkembangan dunia usaha sekarang ini, pertumbuhan dibidang perekonomian baik jasa maupun industri mengalami perkembangan yang sangat pesat, contohnya pada jasa maskapai penerbangan. Menurut Sari dkk (2014, 2), Jasa penerbangan adalah salah satu bentuk jasa yang melayani kebutuhan manusia untuk bermigrasi dari satu tempat ke tempat lain yang berjauhan dengan waktu yang singkat. Menurut Departemen Perhubungan (2015), Berdasarkan data statistik lalu lintas angkutan udara pada Kementerian Perhubungan Republik Indonesia Menunjukkan bahwa jumlah penumpang angkutan udara setiap tahunnya mengalami peningkatan baik tujuan domestik maupun tujuan internasional.

Kualitas layanan jasa maskapai penerbangan dapat diketahui dari pendapat atau *review* para penumpang sebelumnya. Saat ini konsumen yang menulis opini dan pengalaman secara *online* terus meningkat. Membaca *review* tersebut secara keseluruhan bisa memakan waktu, namun jika hanya sedikit *review* yang dibaca evaluasi akan bias. Menurut Tan & 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 pernah dilakukan mengenai klasifikasi sentimen diantaranya adalah penelitian yang dilakukan oleh Ting et al (2011, 43) mengenai klasifikasi kategori teks dengan nilai akurasi mencapai 97,00%. Penelitian yang dilakukan oleh Zhang et al (2011, 7677) mengenai analisis sentimen *review* restoran dengan nilai akurasi mencapai 95,67%. Penelitian yang dilakukan oleh Hidayatullah dan Azhari (2014, A-7) mengenai analisis sentimen dan klasifikasi kategori terhadap tokoh publik pada twitter, dengan nilai akurasi mencapai 73,81%.

Menurut Zhang et al (2011, 2160) Naive 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. Sedangkan menurut Ting et al (2011, 37) Naive Bayes merupakan algoritma paling sederhana dari pengklasifikasi probabilitistik. Seperti yang sudah disebutkan diatas bahwa Naive Bayes sangat sederhana dan efisien, disisi lain menurut Chen et al (2009, 5432) Naive Bayes sangat sensitif terhadap

pemilihan fitur seleksi, maka dari itu pemilihan fitur yang sesuai sangat diperlukan.

Menurut 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 Guo et al (2010, 2990) Algoritma Genetika merupakan salah satu algoritma optimasi, yang diciptakan untuk meniru beberapa proses yang diamati dalam evolusi alam. Algoritma Genetika juga merupakan algoritma *stochastic* yang kuat berdasarkan prinsip-prinsip seleksi alam dan natural genetik yang cukup berhasil diterapkan dalam masalah *machine learning* dan optimasi.

## B. TINJAUAN PUSTAKA

### 1. Penelitian Terkait

Beberapa penelitian terdahulu yang dijadikan rujukan terkait dengan topik penelitian.

Tabel 1. Rangkuman Penelitian Terkait

Peneliti	Text Processing	Feature Selection	Classifier	Accuracy
S.L. Ting, W.H. Ip, Albert H.C. Tsang (2011)	<ul style="list-style-type: none"> <li>• Stopwords</li> <li>• Removing</li> <li>• missing value interpretation</li> <li>• Stemming</li> </ul>	<ul style="list-style-type: none"> <li>• Chi-Square</li> </ul>	Naïve Bayes	97,0 %
Ziqiong Zhang, Qiang Ye, Zili Zhang, dan Yijun Li (2011)	<ul style="list-style-type: none"> <li>• Substitution</li> <li>• N-gram</li> </ul>	<ul style="list-style-type: none"> <li>• Information Gain</li> </ul>	Naïve Bayes	95,67 %
Ahmad Fathan Hidayatullah dan Azhari SN (2014)	<ul style="list-style-type: none"> <li>• Remove URL</li> <li>• Convert Emoticon</li> <li>• Remove Special Character</li> <li>• Remove Symbol</li> <li>• Tokenize</li> <li>• Case Folding</li> <li>• Stemming</li> <li>• Stopwords</li> </ul>	<ul style="list-style-type: none"> <li>• Unigram</li> <li>• Negation</li> <li>• Term Frequency</li> <li>• TF-IDF</li> </ul>	Naïve Bayes	73,81 %
Juan Ling, Putu Eka N. Kençana, IJokorda Bagus Oka (2014)	<ul style="list-style-type: none"> <li>• Tokenization</li> <li>• Stemming</li> <li>• Stopwords</li> </ul>	<ul style="list-style-type: none"> <li>• Chi-Square</li> </ul>	Naïve Bayes	83 %
Alaa Saleh Alithenyan & Mohamed El Bachir Menaj (2014)	<ul style="list-style-type: none"> <li>• Normalization</li> <li>• Stopwords</li> <li>• Elongation</li> <li>• Stemming</li> </ul>	<ul style="list-style-type: none"> <li>• Term Frequency</li> </ul>	Naïve Bayes	82,30 %
Risa Wati (2016)	<ul style="list-style-type: none"> <li>• Tokenization</li> <li>• Stopwords</li> <li>• Removing</li> <li>• Stemming</li> </ul>	<ul style="list-style-type: none"> <li>• Algoritma Genetika</li> </ul>	Naïve Bayes	?

### 2. Review Analisis Sentimen

Menurut Medhat et al (2014, 1093) 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.

### 3. Pemilihan Fitur

Metode seleksi fitur memainkan peran penting dalam analisis sentimen, sama seperti dalam tugas *text mining* lainnya. Menurut Koncz & Paralic (2011, 358) Penggunaan yang tepat dari metode seleksi fitur membantu juga memahami atribut yang relevan untuk kelas tertentu, serta meningkatkan akurasi klasifikasi.

### 4. Naive Bayes

*Bayesian Classification* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar. Menurut Gorunescu (2011,186) rumus Bayes dapat ditulis sebagai berikut:

$$Posterior = \frac{likelihood \times prior \ probability}{evidence}$$

### 5. Algoritma Genetika

Menurut Guo et al (2010, 2990) Algoritma Genetika merupakan salah satu algoritma optimasi, yang diciptakan untuk meniru beberapa proses yang diamati dalam evolusi alam. Algoritma Genetika juga merupakan algoritma *stochastic* yang kuat berdasarkan prinsip-prinsip seleksi alam dan natural genetik yang cukup berhasil diterapkan dalam masalah *machine learning* dan optimasi.

### 6. Evaluasi dan Validasi Klasifikasi

Untuk melakukan evaluasi pada algoritma Naïve Bayes dan algoritma Naïve Bayes dengan Algoritma Genetika dilakukan beberapa pengujian menggunakan *confusion matrix* dan kurva ROC. Menurut Gorunescu (2011, 320) ketika dataset hanya memiliki dua kelas, yaitu *class* positif dan *class* negatif, maka dapat dibuatkan tabel seperti dibawah ini:

Tabel 2 Class pada Confusion Matrix

Classification	Predicted Class		
	Class = Yes	Class = No	
Observed Class	Class = Yes	a true positive – (TP)	b false negative – (FN)
	Class = No	c false positive – (FP)	d true negative – (TN)

Keterangan:

True Positive (TP) : Proporsi positif dalam data set yang diklasifikasikan positif

True Negative (TN) : Proporsi negative dalam data set yang diklasifikasikan negatif

False Positive (FP) : Proporsi negatif dalam data set yang diklasifikasikan positif

False Negative (FN) : Proporsi negative dalam data set yang diklasifikasikan negatif

Berikut adalah persamaan model *Confusion Matrix*:

a) Nilai *Accuracy* adalah proporsi jumlah prediksi yang benar. Dapat dihitung dengan menggunakan persamaan:

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

b) *Sensitivity* digunakan untuk membandingkan proporsi TP terhadap tupel yang positif, yang dihitung dengan menggunakan persamaan:

$$Sensitivity = \frac{TP}{TP + FN}$$

c) *Specificity* digunakan untuk membandingkan proporsi TN terhadap tupel yang negatif, yang dihitung dengan menggunakan persamaan:

$$Specificity = \frac{TN}{TN + FP}$$

d) PPV (*Positive Predictive Value*) adalah proporsi kasus dengan hasil diagnosa positif, yang dihitung dengan menggunakan persamaan:

$$PPV = \frac{TP}{TP + FP}$$

e) NPV (*Negative Predictive Value*) adalah proporsi kasus dengan hasil diagnosa negatif, yang dihitung dengan menggunakan persamaan:

$$NPV = \frac{TN}{TN + FN}$$

Pedoman umum untuk mengklasifikasikan keakuratan pengujian menggunakan AUC menurut Gorunescu (2011, 325-326) :

- 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 dalam penelitian ini seperti yang dijelaskan pada Gambar 1.



Sumber: Peneliti

Gambar 1. Langkah Metode Penelitian

#### 1. Pengumpulan Data

Penelitian ini diawali dengan melakukan pengumpulan data. Data yang diperoleh dari *website* berupa kumpulan opini masyarakat yang sudah banyak tersedia. Kemudian dari kumpulan opini tersebut diintegrasikan kedalam dataset.

#### 2. Pengolahan Data Awal

Pada tahap pengolahan data, data yang sudah terkumpul, terlebih dahulu diolah dengan melakukan penyeleksian data yang meliputi pembersihan data, mentransformasikan data kedalam bentuk yang dibutuhkan.

#### 3. Metode yang Diusulkan

Setelah melakukan pengumpulan dan pengolahan data tahap selanjutnya adalah menentukan metode, dimana metode ini merupakan gambaran dari rangkaian kegiatan dan membagi data kedalam data *training* dan data *testing*.

#### 4. Eksperimen dan Pengujian Metode

Menjabarkan bagaimana eksperimen yang dilakukan sampai terbentuknya model. Melakukan perhitungan dengan masing-masing algoritma yang akan diulang beberapa kali sampai menghasilkan nilai akurasi tertinggi.

#### 5. Evaluasi dan Validasi Hasil

Tahap evaluasi merupakan tahap akhir dari rangkaian kegiatan penelitian ini. Setelah melakukan tahap pengujian model maka akan menghasilkan nilai akurasi dan AUC. Kemudian dari hasil itu dievaluasi, dari hasil evaluasi itu dapat ditarik kesimpulan dari hasil penelitian ini.

### D. HASIL DAN PEMBAHASAN

#### 1. Pengumpulan Data

Dataset yang digunakan pada penelitian ini berupa *review* jasa maskapai penerbangan yang diperoleh dari

<http://www.airlinequality.com> terdiri dari 100 review positif dan 100 review negatif.

2. Pengolahan Data Awal

a) *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.

b) *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.

c) *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*.

d) *N-gram*

Teknik *n-gram* didasarkan pada pemisahan teks menjadi string dengan panjang n mulai dari posisi tertentu dalam suatu teks. Posisi *n-gram* berikutnya dihitung dari posisi yang sebenarnya bergeser sesuai dengan offset yang diberikan. Nilai *offset* bergantung pada pembagian yang digunakan dalam *n-gram*.

3. Model dengan Metode Klasifikasi Menggunakan Naïve Bayes

Proses klasifikasi adalah proses untuk menentukan kalimat tersebut sebagai *class* positif atau negatif berdasarkan nilai perhitungan probabilitas Sebagai contoh penulis menampilkan 10 dokumen dari 200 data training dan tiga kata yang berhubungan dengan sentimen positif serta tiga kata yang berhubungan dengan sentimen negatif, yaitu *bad*, *delay*, *disappoint*, *excel*, *good* dan *recommend*. Jika kata tersebut muncul dalam dokumen maka akan diberi nilai 1 dan jika tidak muncul maka pada dokumen akan diberi nilai 0.

Tabel 3 Tabel *Vector* Dokumen Boolean dengan Label *Class* Hasil Klasifikasi

Dokumen	Bad	Delay	Disappoint	Excel	Good	Recommend	Class
c_positif1.txt	0	1	1	1	1	0	Positif
c_positif2.txt	0	0	0	1	0	0	Positif
c_positif3.txt	0	0	0	1	0	0	Positif
c_positif4.txt	0	0	0	1	0	1	Positif
c_positif5.txt	0	0	0	0	0	0	Positif
c_negatif16.txt	1	0	1	0	0	0	Negatif
c_negatif17.txt	0	1	0	0	0	1	Negatif
c_negatif18.txt	0	1	0	0	0	0	Negatif
c_negatif19.txt	0	1	0	0	1	0	Negatif
c_negatif20.txt	0	1	0	0	0	0	?

Sumber: Peneliti

Probabilitas Bayes yang akan penulis jabarkan adalah probabilitas untuk dokumen c\_negatif20.txt.

a) Hitung probabilitas bersyarat (*likelihood*) dokumen c\_negatif20.txt pada *class* positif dan *class* negatif.

1) *Class* Positif

$$\begin{aligned}
 P(c\_negatif20|positif) &= P(Bad=1|positif) \times P(Delay=0|positif) \\
 &\times P(Disappoint=1|positif) \times P(Excel=0|positif) \\
 &\times P(Good=1|positif) \times P(Recommend=0|positif) \\
 P(c\_negatif20|positif) &= 0/5 \times 4/5 \times 1/5 \\
 &\times 1/5 \times 1/5 \times 4/5 \\
 &= 0 \times 0.8 \times 0.2 \times 0.2 \times 0.2 \times 0.8 \\
 &= 0
 \end{aligned}$$

2) *Class* Negatif

$$\begin{aligned}
 P(c\_negatif20|negatif) &= P(Bad=1|negatif) \times P(Delay=0|negatif) \\
 &\times P(Disappoint=1|negatif) \times P(Excel=0|negatif) \\
 &\times P(Good=1|negatif) \times P(Recommend=0|negatif) \\
 P(c\_negatif20|negatif) &= 1/4 \times 1/4 \times 1/4 \times 5/4 \\
 &\times 1/4 \times 4/4 \\
 &= 0.25 \times 0.25 \times 0.25 \times 1.25 \times 0.25 \times 1 \\
 &= 0.00488
 \end{aligned}$$

b) Probabilitas *prior* dari *class* positif dan *class* negatif dihitung dengan proporsi dokumen pada tiap *class*:

$$P(positif) = 5/9 = 0.55$$

$$P(negatif) = 4/9 = 0.44$$

c) Hitung probabilitas *posterior* dengan memasukkan rumus Bayes dan menghilangkan penyebut P(c\_negatif20.txt):

$$P(positif|c\_negatif20) = \frac{(0)(0.55)}{P(c\_negatif20)} = 0$$

$$P(negatif|c\_negatif20) = \frac{(0.00488)(0.44)}{P(c\_negatif20)} =$$

$$0.00215$$

Berdasarkan probabilitas diatas, maka dapat disimpulkan bahwa dokumen c\_negatif20.txt merupakan class negatif karena  $P(\text{negatif} | c\_negatif20)$  lebih besar dari  $P(\text{positif} | c\_negatif20)$ .

Hasil Eksperimen pengujian metode Naive Bayes ditunjukkan pada tabel 4 nilai akurasi nya 67.00% dan AUC 0.512 Berikut Hasil Akurasi pada metode Naive Bayes:

Tabel 4 Model Confusion Matrix untuk Metode Naive Bayes

accuracy: 60.00% +/- 7.75% (mikro: 60.00%)			
	true Data_Negatif	true Data_Positif	class precision
pred. Data_Negatif	72	52	58.06%
pred. Data_Positif	28	48	63.16%
class recall	72.00%	48.00%	

Nilai accuracy dari confusion matrix adalah sebagai berikut:

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

$$Accuracy = \frac{72 + 48}{72 + 28 + 48 + 52} = \frac{120}{200} = 0.60 = 60.00\%$$

$$b) Sensitivity = \frac{TP}{TP+FN}$$

$$Sensitivity = \frac{48}{48 + 28} = \frac{48}{76} = 0.6316 = 63.16\%$$

$$c) Specificity = \frac{TN}{TN+FP}$$

$$Specificity = \frac{72}{72 + 52} = \frac{72}{124} = 0.5806 = 58.06\%$$

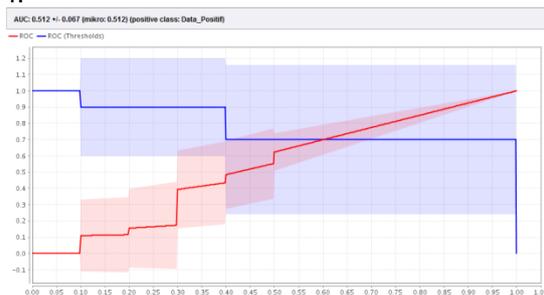
$$d) PPV = \frac{TP}{TP+FP}$$

$$PPV = \frac{48}{48 + 52} = \frac{48}{100} = 0.48 = 48.00\%$$

$$e) NPV = \frac{TN}{TN+FN}$$

$$NPV = \frac{72}{72 + 28} = \frac{72}{100} = 0.72 = 72.00\%$$

Kurva ROC pada gambar 2 mengekspresikan confusion matrik dari tabel 4.



Gambar 2. Kurva ROC Naive Bayes

#### 4. Model dengan Metode Klasifikasi Menggunakan Naive Bayes dan Pemilihan Fitur Algoritma Genetika

Nilai training dalam penelitian ini ditentukan dengan cara melakukan uji coba memasukkan nilai parameter *Maximum Number of Generation* adjustment dimulai dari 30-60, untuk *population size* dimulai dari 5-50 dan untuk *p crossover* 0.1-1.0.

Tabel 5. Rencana Eksperimen

Maximum Number of Generation	Population Size	P Crossover	Accuracy	AUC
30-60	5-50	0.1-1.0	?	?

Berikut ini adalah hasil dari percobaan yang telah dilakukan untuk menentukan nilai training, hasil percobaan yang ada ditabel 6 adalah hasil percobaan yang nilai akurasi dan AUC tertinggi:

Tabel 6. Eksperimen Penentuan Nilai Training NB berbasis GA

Maximum Number of Generation	Population size	p crossover	Naive Bayes + GA	
			Accuracy	AUC
30	25	0.5	82.00 %	0.765
30	30	0.5	83.00 %	0.854
30	40	0.5	84.00 %	0.815
45	5	0.5	84.00 %	0.919
30	5	0.9	80.00 %	0.763
30	5	0.5	77.00 %	0.774
45	40	0.9	89.50 %	0.919

Hasil Eksperimen pengujian metode Naive Bayes berbasis GA ditunjukkan pada tabel 10 nilai akurasi nya 89.50 % dan AUC 0.919.

Berikut Hasil Akurasi pada metode Naive Bayes berbasis GA:

Tabel 7. Model Confusion Matrix untuk Metode Naive Bayes Berbasis GA

accuracy: 89.50% +/- 6.10% (mikro: 89.50%)			
	true Data_Negatif	true Data_Positif	class precision
pred. Data_Negatif	86	7	92.47%
pred. Data_Positif	14	93	86.92%
class recall	86.00%	93.00%	

Nilai accuracy dari confusion matrix adalah sebagai berikut:

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

$$Accuracy = \frac{93 + 86}{93 + 86 + 7 + 14} = \frac{179}{200} = 0.8950 = 89.50\%$$

$$b) Sensitivity = \frac{TP}{TP+FN}$$

$$Sensitivity = \frac{93}{93 + 14} = \frac{93}{107} = 0.8692 = 86.92\%$$

$$c) Specificity = \frac{TN}{TN+FP}$$

$$Specificity = \frac{86}{86 + 7} = \frac{86}{93} = 0.9247 = 92.47\%$$

$$d) PPV = \frac{TP}{TP+FP} = \frac{93}{93 + 7} = \frac{93}{100} = 0.93 = 93.00\%$$

$$e) NPV = \frac{TN}{TN+FN} = \frac{86}{86 + 14} = \frac{86}{100} = 0.86 = 86.00\%$$

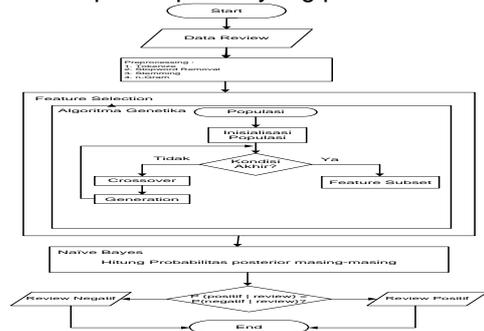
Kurva ROC pada gambar 3 mengekspresikan *confusion matrik* dari tabel 7.



Gambar 3 Kurva ROC Naive Bayes berbasis GA

### 5. Implementasi

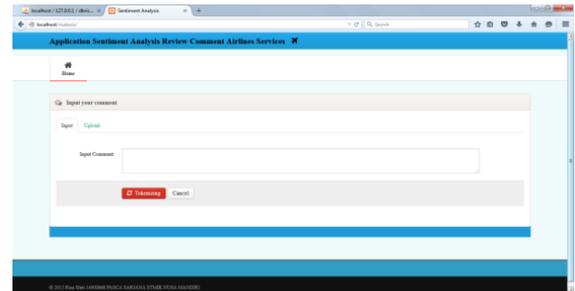
Peneliti membuat aplikasi untuk menguji model yang sudah ada menggunakan dataset yang berbeda dan belum diketahui classnya. Aplikasi dibuat menggunakan *dreamweaver* dengan bahasa pemrograman PHP. Gambar 4 adalah diagram alir dari tahapan proses klasifikasi pada aplikasi yang peneliti buat.



Gambar 4. Diagram alir tahapan proses klasifikasi

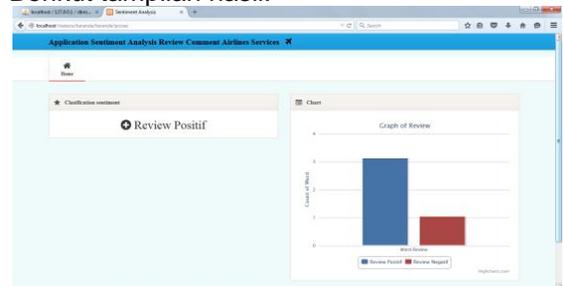
Berikut tampilan rancangan aplikasi dan implementasi:

Tampilan halaman *index* pada aplikasi pengklasifikasi teks dapat dilihat pada gambar 5, dimana ada dua buah tab yaitu tab *input* dan *upload*. Jika memilih tab *input*, dapat mengetikkan langsung komentar-komentar tentang maskapai penerbangan, sedangkan tab *upload* dapat mengupload file yang berisi komentar.



Gambar 5. Tampilan *Index Home* pada Aplikasi *Review*

Setelah komentar *diinput* atau diambil dari sebuah *file* maka komentar tersebut akan melalui tahapan *preprocessing* kemudian komentar akan diproses untuk mengetahui apakah komentar tersebut termasuk kedalam komentar positif ataupun komentar negatif. Berikut tampilan hasil:



Gambar 6. Tampilan Hasil *Review Positif*



Gambar 7. Tampilan Hasil *Review Negatif*

### 6. Implikasi Penelitian

Implikasi penelitian ini mencakup beberapa aspek, di antaranya:

- a) Implikasi terhadap aspek sistem  
Hasil evaluasi menunjukkan penerapan Algoritma Genetika untuk seleksi fitur dapat meningkatkan akurasi Naive Bayes dan merupakan metode yang cukup baik dalam mengklasifikasi teks *review* jasa maskapai penerbangan sehingga dapat mempermudah pengguna untuk menentukan maskapai penerbangan yang akan digunakan.
- b) Implikasi terhadap aspek manajerial  
Membantu para pengembang sistem yang berkaitan dengan *review* jasa maskapai penerbangan.
- c) Implikasi terhadap aspek penelitian

lanjutan Penelitian selanjutnya bisa menggunakan metode pemilihan fitur ataupun dataset dari domain yang berbeda, seperti *review* produk, *review* restoran dan sebagainya.

#### E. KESIMPULAN

1. Pada penelitian ini penulis melakukan pengklasifikasian teks *review* jasa maskapai penerbangan dengan pengklasifikasi Naive Bayes, dikarenakan Naive Bayes dapat berfungsi dengan baik sebagai metode pengklasifikasi teks.
2. Pada Penelitian ini terbukti bahwa penggabungan metode Naive Bayes dengan pemilihan fitur Algoritma Genetika dapat meningkatkan akurasi.
3. Penggunaan Metode Naive Bayes dalam penelitian ini menghasilkan akurasi 60.00% dan AUC sebesar 0.512 sedangkan akurasi setelah menggunakan pemilihan fitur Algoritma Genetika sebesar 89.50 dan AUC sebesar 0.919 termasuk kedalam *excellent classification* terjadi peningkatan akurasi sebesar 29.5% dan AUC sebesar 0.407.
4. Pada penelitian ini peneliti mengembangkan aplikasi *review* jasa maskapai penerbangan berbasis web yang dapat menampilkan hasil *review* dalam bentuk *review* positif dan negatif, sehingga dapat membantu konsumen dalam memilih jasa maskapai penerbangan yang baik.

#### DAFTAR PUSTAKA

- [1] Chen, J. et al., 2009. Expert Systems with Applications Feature selection for text classification with Naive Bayes. *Expert Systems With Applications*, 36(3), pp.5432–5435.
- [2] Dephub. 2015. Lalu Lintas Angkutan Udara [Online] Tersedia di: <http://hubud.dephub.go.id/?id/llu/index/filter:category,2;tahun,0;bulan,0;airpor t,0>
- [3] Gorunescu, F., 2011. Data mining: concepts and techniques.
- [4] Guo, P., Wang, X. & Han, Y., 2010. The Enhanced Genetic Algorithms for the Optimization Design. , (Bmei), pp.2990–2994.
- [5] Hidayatullah, A.F. et al., 2014. Analisis sentimen dan klasifikasi kategori terhadap tokoh publik pada twitter. , 2014(semnasIF), pp.1–8.
- [6] Koncz, P. & Paralic, J., 2011. An approach to feature selection for sentiment analysis. 2011 15th IEEE

- [7] International Conference on Intelligent Engineering Systems, pp.357–362. Medhat, W., Hassan, A. & Korashy, H., 2014. Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), pp.1093–1113.
- [8] Sari, E. P., & Apriatni, E. P. 2014. Analisis Tingkat Kepuasan Pelanggan Dilihat Dari Dimensi Kualitas Pelayanan Dan Harga PT. Garuda Indonesia Airlines (Persero), Tbk Kantor Cabang Semarang. *Diponegoro journal of social and politic*.
- [9] Tan, S. & Zhang, J., 2008. An empirical study of sentiment analysis for chinese documents. , 34, pp.2622–2629.
- [10] Ting, S.L., Ip, W.H. & Tsang, A.H.C., 2011. Is Naive Bayes a Good Classifier for Document Classification ? , 5(3), pp.37–46.
- [11] Zhang, W., & Gao, F. 2011. An improvement to naive bayes for text classification. *Procedia Engineering*, 15, 2160–2164
- [12] *Sardiarinto ., Sa'diyah Noor Novita Alfisahrin, Anik Andriani, Rancang Bangun Sistem Reservasi Hotel Menggunakan Metode Waterfall Studi kasus: Hotel Bizz Yogyakarta, Vol 3, No 1 (2015): Bianglala 2015*
- [13] *Muhammad Tabrani, IMPLEMENTASI SISTEM INFORMASI RESERVASI PENGINAPAN PADA ARGOWISATA GUNUNG MAS CISARUA BOGOR, Vol 2, No 1 (2014): Bianglala 2014*