*Jurnal Teknik Komputer,*
*Vol. III, No. 2, Agustus 2017*

# Optimasi Particle Swarm Optimization Sebagai Seleksi Fitur Pada Analisis Sentimen Review Hotel Berbahasa Indonesia Menggunakan Algoritma Naïve Bayes

**Andi Taufik**
Program Studi Sistem Informasi
STMIK Nusa Mandiri Jakarta
Jl. Damai No. 8 Warung Jati Barat Jakarta Selatan
a.taufik30@gmail.com

*Abstract*— Currently visitors who wrote an opinion to share experiences online continues to increase. Each visitor will need to make a decision while on vacation before ordering a hotel for an overnight stay, usually reading the results of the review of the visitor before, certainly requires quite a long time when reading the review as a whole, however, if just a little review that read, the information obtained will be biased. Sentiment analysis aims to address this problem by automatically classify user review be opinions positive or negative. Naïve Bayes classification machine learning technique is popular for text classification, because it is very simple, efficient and have good performance in many domains. However, Naïve Bayes has a shortage that is very sensitive on the features too much, resulting in a lower classification accuracy. Therefore, in this study used methods the selection of features, i.e. Particle Swarm Optimization in order to improve the accuracy of classification of Naïve Bayes. This research resulted in the classification of texts in the form of a positive review or a negative review of a hotel review taken from the website www.Tripadvisor.com. The measurements accuracy based on Naive Bayes method before and after the addition of the selection of features. The evaluation was conducted using a 10 fold cross validation. While the measurement accuracy is measured by the confusion matrix and ROC curves. The results showed an increase in the accuracy of Naïve Bayes from 90.50% to 96.92%.

**Keywords: Analysis Sentiment, Reviews Hotel, Naïve Bayes, Particle Swarm Optimization, Text Classification, Selection Feature**

*Abstrak* – Saat ini pengunjung yang menulis *review* untuk berbagi pengalaman secara *online* terus meningkat. Setiap pengunjung perlu untuk membuat keputusan saat berlibur sebelum memesan hotel untuk menginap, biasanya menbaca hasil review dari pengunjung sebelumnya, tentunya membutuhkan waktu yang cukup lama apabila membaca review tersebut secara keseluruhan namun, jika hanya sedikit review yang dibaca, informasi yang didapatkan akan bias. Analisa sentimen bertujuan untuk mengatasi masalah ini dengan secara otomatis mengelompokkan review pengguna menjadi opini positif atau negatif . Pengklasifikasi Naïve Bayes adalah teknik *machine learning* yang populer untuk klasifikasi teks, karena sangat sederhana, efisien dan memiliki performa yang baik pada banyak domain. Namun, Naïve Bayes memiliki kekurangan yaitu sangat sensitif pada fitur yang terlalu banyak, yang mengakibatkan akurasi klasifikasi menjadi rendah. Oleh karena itu, dalam penelitian ini digunakan metode pemilihan fitur, yaitu Particle Swarm Optimization agar bisa meningkatkan akurasi pengklasifikasi Naïve Bayes. Penelitian ini menghasilkan klasifikasi teks dalam bentuk review positif atau review negatif dari review hotel yang diambil dari situs www. Tripadvisor.com.Pengukuran berdasarkan akurasi Naive Bayes sebelum dan sesudah penambahan metode pemilihan fitur. Evaluasi dilakukan menggunakan 10 *fold cross validation*. Sedangkan pengukuran akurasi diukur dengan *confusion matrix* dan kurva ROC. Hasil penelitian menunjukkan peningkatan akurasi Naïve Bayes dari 90.50% menjadi 96.92%

**Kata Kunci***:* Analisa Sentimen, *Review* Hotel, Naïve Bayes, *Particle Swarm Optimization*, Klasifikasi teks, pemilihan fitur.

## I. PENDAHULUAN

Dengan memanfaatkan perkembangan teknologi informasi melalui pengguna jejaringan sosial mengenai *review* hotel menyediakan *review* pengunjung yang digunakan untuk berinteraksi dengan pengunjung lain nya, *platform* digunakan sebagai wadah untuk membuat dan mendengar pendapat pengunjung yang menghasilkan ulasan perjalanan dan jasa perhotelan yang telah dikunjungi pada saat liburan menjadi sumber informasi yang sangat penting bagi pengunjung (Duan, Cao dan Yu).

Informasi yang sangat berguna saat ini, karena orang cenderung mencari informasi yang cepat dalam pemesanan. Lebih banyak pengguna yang mencari informasi melalui pendapat orang lain di media sosial, blog dan situs-situs *review*. Pentingnya ulasan hotel sebagai sumber informasi khusus untuk pemesanan hotel (Markopoulos, Mikros dan Iliadi)

Memungkinkan para pengelola dunia pariwisata untuk memberikan informasi lebih detail tentang produk pariwisata yang ditawarkan. Banyak orang yang memeriksa pendapat dari pembeli lain sebelum membeli produk untuk membuat pilihan yang tepat. Hotel merupakan salah satu produk pariwisata yang sangat penting untuk dipertimbangkan baik dari segi fasilitas, pelayanan ataupun jarak tempuh perjalanan wisata (Taylor, Velasquez dan Marquez)

Setiap orang perlu untuk membuat keputusan saat berlibur sebelum memesan hotel untuk menginap, biasanya mereka meminta pendapat orang lain, hal ini dapat diperoleh dengan membaca opini atau hasil *review* dari pengalaman pengunjung sebelumnya yang tentunya membutuhkan waktu yang cukup lama.

Terdapat beberapa penelitian yang sudah dilakukan dalam hal pengklasifikasian analisis sentimen terhadap *review yang* tersedia, diantaranya adalah penelitian yang dilakukan oleh (Suardika) sentimen analisis dilakukan menggunakan metode *Naïve Bayes* yang mencari hubungan peringkat antar hotel pada situs *Tripadvisor* dengan hasil klasifikasi dalam sentimen positif, sentiment negatif dan sentimen netral. Dengan metode naïve bayes nilai akurasi rata-ratanya adalah 81% dan menghasilkan analisis korelasi membuktikan hipotesis bahwa

 p-ISSN: 2442-2436, e-ISSN: 2550-0120

semakin rendah peringkat hotel, semakin besar persentasi sentimen negatif. Lalu Penelitian yang dilakukan oleh (Zhang, Ye dan Li), Pengklasifikasian sentimen pada review restoran di internet yang ditulis dalam bahasa Canton menggunakan pengklasifikasi Naive Bayes dan Support Vector Machines. Sedangkan penelitian dari (Markopoulos, Mikros dan Iliadi) dimana dalam membuat *classifier sentiment* yang menerapkan *Support Vector Mechines* dengan *fiture Unigram* pada *review* hotel dalam bahasa Yunani modern yang membandingkan dua metodologi yang berbeda.

Menurut (Duan, Cao dan Yu) *Naïve Bayes* merupakan klasifikasi sederhana dan efektif . Namun *Naïve Bayes* sebagai klasifikasi yang sangat sederhana dan efesien serta sangat sensitif dalam pemilihan fitur (Chen, Huang dan Tian)

Menurut (Lu, Liang dan Ye) 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. Sedangkan menurut (Basari, Hussin dan Ananta). PSO banyak digunakan untuk memecahkan masalah optimasi serta masalah seleksi fitur. *Particle Swarm Optimization* (PSO) adalah suatu teknik optimasi yang sangat sederhana untuk menerapkan dan memodifikasi beberapa parameter.

Pada penelitian ini menggunakan klasifikasi *Naïves Bayes* dengan *Particle Swarm Optimization* sebagai metode pemilihan fitur pada komentar dari *review* hotel berbahasa Indonesia sebagai teknik untuk meningkatkan nilai akurasi analisa sentiment

Rumusan masalah pada penelitian ini adalah untuk melihat apakah terjadi peningkatan akurasi klasifikasi *Naïve Bayes* apabila *Particle Swarm Optimization* untuk seleksi fitur pada analisis sentiment *review* hotel diterapkan

Tujuan penelitian ini adalah untuk mengetahui seberapa besar meningkatnya akurasi pengklasifikasi Naïve Bayes dengan menggunakan Particle Swarm Optimization sebagai seleksi fitur pada analisis sentimen *review* hotel berbahasa Indonesia

Berdasarkan tujuan penelitian, maka manfaat dari penelitian ini adalah :

1. Manfaat dari penelitian ini adalah membantu dalam menggambil keputusan saat ingin melakukan pemesanan hotel yang sesuai dangan keinginan agar lebih efisien dan efektif dibandingkan jika harus membaca *review* yang memakan waktu cukup lama.
2. Memberikan kontribusi keilmuan pada penelitian yang berkaitan dengan analisa sentimen atau Opinion Mining yang menerapkan pengklasifikasi Naïve Bayes Dengan mengunakan pemilihan Fitur *Particle Swarm Optimization* dalam pengklasifikasian review atau opini sehingga dapat dijadikan sebagai pemikiran untuk pengembangan teori berikutnya.

## A. Tinauan Pustaka

### 1. Data Mining

Menurut (Gorunescu) Data mining dapat didefenisikan sebagai sebuah proses untuk menemukan pola data.

Menurut (Witten, Frank dan Hall) *Data mining* merupakan perpaduan dari ilmu statistik, kecerdasan buatan (sistem pakar) dan penelitian dalam bidang *database*, untuk itu diperlukan penyaringan melalui sejumlah besar material data atau melakukan penyelidikan dengan cerdas tentang keberadaan suatu data yang memiliki nilai *Daryl Pregibons*.

2. Klasifikasi, validasi dan Evaluasi Algoritma Data Mining

Berikut ini adalah persamaan model *Confusion Matrix* (Han dan Kamber)

1. Nilai akurasi (acc) adalah proporsi jumlah predikasi yang benar
   ……(2.1)
2. *Sensitivity* digunakan untuk membandingkan *proporsi* tp terhadap *tupel* yang positif.
   $Sensitivity = \frac{TP}{TP+FN}$……(2.2)
3. *Specifity* digunakan untuk membandingkan proporsi tn terhdap tupel yang negatif
   $Specifity = \frac{TP}{TN+FF}$……..(2.3)
4. PPV (*Positive Predictive Value* ) adalah proporsi kasus dengan diagnosa positif
   $PPV = \frac{TP}{TP+FF}$……(2.4)
5. NPV(*negative Predictive Value* ) adalah *proporsi* kasus dengan diagnosa nega$Accuracy = \frac{TP+TN}{TP+FN+FP+TN}$tif
   $NPV = \frac{TN}{TN+FN}$……(2.5)

### 3. Text Mining

Menurut (Bramer), teks merupakan sesuatu yang umum dalam melakukan pertukaran informasi. Syarat umum data dan *teks mining* adalah informasi yang diambil dan dapat menjadi data yang berguna.*Text mining* merupakan proses menganalisa teks untuk menjadi informasi yang berguna untuk tujuan tertentu. Informasi yang diambil harus jelas dan *eksplisit*, karena *text mining* merubah menjadi bentuk yang dapat digunakan oleh *computer* atau orang yang tidak memiliki waktu untuk membaca *full teks.*

*Text mining* adalah penemuan dari pengetahuan yang menarik pada dokumen teks.Hal ini merupakan tantangan untuk menemukan pengetahuan yang akurat pada teks dokumen untuk menolong pengguna untuk menemukan yang diinginkan. Penemuan pengetahuan dapat menjadi efektif digunakan dan memperbaharui pola penemuan dan menerapkannya ke *text mining* (Charjan dan Pun ).

### 4. Sentimen Analisis

Analisis atau opini mining merupakan kaian tentang cara untuk memecahkan masalah opini masyarakat, sikap dan emosi suatu entita, yang dapat mewakili individu, peristiwa atau topik (Medhat, Hassan dan Korashy).

Menurut (Kontopoulos, Berberidis dan Dergiades), *Opinion mining* atau juga dikenal sebagai analisa sentimen adalah proses yang bertujuan untuk menentukan apakah polaritas kumpulan teks tulisan (dokumen, kalimat, paragrap, dll) cenderung ke arah positif, negatif, atau netral.

### 5. Review

Ulasan wisata dari konsumen lain mempengaruhi setengah dari semua keputusan pembelian hotel (Duan, Cao dan Yu).

Hotel merupakan salah satu produk pariwisata yang sangat penting untuk dipertimbangkan baik dari segi fasilitas, pelayanan ataupun jarak tempuh perjalanan wisata. Saat ini sudah banyak website wisata yang menyediakan fasilitas untuk pengguna internet menuliskan opini dan pengalaman pribadinya secara *online.* Banyak orang yang memeriksa pendapat dari pembeli lain sebelum membeli produk untuk membuat pilihan yang tepat. Yang memungkinkan para pengelola dunia pariwisata untuk memberikan informasi lebih detail tentang produk pariwisata yang ditawarkan (Taylor, Velasquez dan Marquez).

6. Pre-Processing

Menurut (Haddi, Liu dan Shi), *Preprocessing* data adalah proses pembersihan dan mempersiapkan teks untuk klasifikasi . Seluruh proses melibatkan beberapa langkah: membersihkan teks *online*, penghapusan ruang *spasi*, memperluas singkatan, kata dasar (*stemming)*, penghapusan kata henti *(stopword removal)*, penanganan negasi dan terakhir seleksi fitur.

*N-gram* didefinisikan sebagai sub-urutan n karakter dari kata diberikan. Misalnya, "*mountain*" dapat diwakili dengan *character* n-gram (Gencosman, Ozmutlu dan Ozmutlu)

7. TF-IDF (*Term Frequency-Inverse Document Frequency*)

Metode ini akan menghitung nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) pada setiap kata di setiap dikomen dalam korpus.

1. Rumus umum untuk pembobotan TF-IDF menurut (Robertson) :
   $W = tf * idf$ ……..(2.6)
   $W = tf * log(\frac{N}{df})$ ……(2.7)
2. Berdasarkan rumus (2.7), berapapun besarnya nilai tf, apabila N = df dimana sebuah kata/*term* muncul di semua dokumen, maka akan didapatkan hasil 0 (nol) untuk perhitungan idf, sehingga perhitungan bobotnya diubah menjadi sebagai berikut:
   $W = tf * (log\left(\frac{N}{df}\right) + 1)$……..(2.8)
3. Rumus (2.8) dapat dinormalisasi dengan rumus (2.9) dengan tujuan menstandarisasi nilai bobot (wtd) ke dalam interval 0 s.d. 1 Menurut (Intan dan Defeng) :
   $W = \frac{tf*(log\left(\frac{N}{df}\right)+1)}{\sqrt{\sum_{k=1}^{t}(tf)^2*\left(log\left(\frac{N}{df}\right)+1\right)^2}}$.....(2.10)

8. Pemilihan Fitur

Menurut (Maimon dan Rokach) Seleksi fitur untuk mengidentifikasi beberapa fitur dalam kumpulan data yang sama penting dan membuang semua fitur lain seperti informasi yang tidak *relevan* dan berlebihan. Proses seleksi fitur mengurangi dimensi dari data dan memungkinkan algoritma *learning* untuk beroperasi lebih cepat dan lebih efektif. menurut Yang dan Honavar dalam (Zhao, Fu dan Ji), Seleksi *fitur* merupakan proses optimasi untuk mengurangi satu set besar *fitur* besar sumber asli agar subset *fitur* yang relatif kecil yang signifikan untuk meningkatkan akurasi klasifikasi cepat dan efektif.

Menurut John, kohavi dan pfleger dalam (Chen, Huang dan Tian) ada dua jenis metode seleksi fitur dalam pembelajaran *machine learning*, yaitu itu *wrappers* dan *filters.*

9. Particle Swarm Optimization (PSO)

Menurut (Lu, Liang dan Ye) *Particle Swarm Optimization* dirumuskan pertama kali oleh Edward dan kennedy pada tahun 1995. Proses pemikiran dibalik algoritma ini terinspirasi dari perilaku sosial hewan. Seperti burung yang berkelompok atau sekelompok ikan.

*Particle Swarm Optimization* sering digunakan dalam penelitian, karena PSO memiliki kesamaan sifat dengan Genetic Algorithm (GA). PSO banyak digunakan untuk memecahkan masalah optimasi dan sebagai pemecah masalah seleksi fitur menurut (Liu). Tidak seperti GA, PSO tidak memiliki operator seperti crossover dan mutasi. Baris dalam metric disebut particle (sama dengan kromosom GA). Setiap partikel bergerak dipermukaan partikel dengan kecepatan, setiap pembaharuan kecepatan dan posisi berdasarkan lokasi terbaik dari lokal dan global.

10. Naïve Bayes

Naïve bayes merupakan klasifikasi data dengan menggunakan probabilitas dan static.Menurut (Han dan Kamber) tahapan dalam algoritma Naïves Bayes:

1. Perhatikan D adalah record training dan ketetapan label-label kelasnya dan masing-masing record dinyatakan n atribut (n field ) $X = (X_1, X_2 \ldots., X_n)$
2. Misalkan terdapat m kelas $C_1, C_2 \ldots., C_m)$
3. Klasifikasi adalah diperoleh maximum posteriori yaitu maximum P(Ci|X)
4. Ini diperoleh dari teorema Bayes
   $P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}$……..(2.11)

Karena P(X) adalah konstan untuk semua kelas, hanya perlu dimaksimalkan.

$P(C_i|X) = P(X|C_i)P(C_i)$…….(2.12)

## II. METODOLOGI PENELITIAN

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

1. Pengumpulan Data
   Pengumpulan data ditentukan berdasarkan data yang akan diproses yaitu berupa *review* positif maupun *review* negatif. Data tersebut kemudian dioptimasikan didalam *dataset.*
2. Pengolahan Data Awal
   Dilakukan penyeleksian data. Data dibersihkan dan dioptimasikan kedalam bentuk yang dinginkan sebelum dilakukan pembuatan model.
3. Metode yang diusulkan
   Data yang diteliti dan dianalisa kemudian dikelompokkan ke variabel mana yang berhubungan dengan satu sama lainnya, lalu dibuatkan model yang sesuai dengan jenis data. Pembagian data kedalam data latihan (*training* data) dan data uji (*testing* data) juga diperlukan untuk pembuatan model. Dengan penambahkan metode pemilihan fitur *Particle Swarm Optimization* untuk meningkatkan akurasi pada klasifikasi *Naïve Bayes*
4. Eksperimen dan Pengujian Metode
   Eksperimen pada model yang akan dilakukan dengan menggunakan *RapidMiner* 5.3 untuk mengolah data. Model diuji untuk melihat hasil yang akan dimanfaatkan untuk mengambil keputusan hasil penelitian
5. Evaluasi Dan Validasi Hasil
   Pada sebuah penelitian dilakukan evaluasi terhadap model untuk mengetahui akurasi dari model yang telah digunakan. Validasi hasil digunakan untuk melihat perbandingan dari model yang digunakan dengan hasil yang telah dilakukan sebelumnya.

## III. HASIL DAN PEMBAHASAN

Sebelum diklasifikasikan, dataset harus melalui beberapa tahapan proses agar bisa diklasifikasikan dalam proses selanjutnya, berikut ini adalah tahapan prosesnya :

### *A.* Pengumpulan Data

Pada penelitian ini menggunakan data *review* hotel yang diambil dari situs http://www.tripadvisor.com . *Review* hotel yang digunakan hanya 200 *review* hotel yang terdiri dari 100 *review* positif dan 100 *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


p-ISSN: 2442-2436, e-ISSN: 2550-0120

*folder* dan diberi nama negatif. Tiap dokumen berekstensi .txt yang dapat dibuka dengan menggunakan aplikasi *Notepad*

## *B.* Pengolahan Data Awal

### *1. Tokenization*

Dalam proses *tokenization* ini, semua kata yang ada didalam setiap dokumen dikumpulkan dan di hilangkan tanda bacanya, serta dihilangkan juga apabila ada simbol yang bukan huruf. Berikut adalah contoh hasil dari proses *tokenization* dalam *RapidMiner*.

### *2. Tokenization*

Dalam proses ini, kata-kata yang tidak relevan akan dihapus , seperti kata untuk hanya dengan dan sebagainya yang merupakan kata –kata yang tidak mempunyai makna tersendiri jika dipisahkan dengan kata yang lain dan tidak terkait dengan kata sifat yang berhubungan dangan sentimen.

### *3. N-Gram (Bi-Gram)*

Dalam proses ini, potongan 2 karakter dalam suatu *string* tertentu atau potongan 2 kata dalam suatu kalimat tertentu. Contoh pemotongan 2 kata Bi-gram dalam kata Hotel cukup nyaman : "hotel", "hotel_cukup", "cukup", "cukup_nyaman", "nyaman".

Sedangkan untuk tahap *transformation* dengan melakukan pembobotan TF-IDF (*Term Frequency - Inverse Document Frequency*) pada masing-masing kata.Dimana prosesnya menghitung kehadiran atau ketidak hadiran sebuah kata didalam sebuah dokumen. Beberapa kali sebuah kata muncul didalam suatu dokumen juga digunakan sebagai skema pembobotan dari kata tekstual.

Pada tabel 1menunujukan hasil *preprocessing* dari *tokenization, stopword removal dan N-gram (Bi-Gram)*

Tabel 1. Hasil *Preprocessing, Tokenization, Stopword* dan *Bi-Gram*

| *Review* | *Tokenization* | *Stopword* | *Bi-Gram* |
|---|---|---|---|
| Berharap dapat datang kembali dengan keluarga. Namun untuk dapat berbelanja sebaiknya menggunakan kendaraan karena aga jauh dari pusat pertokoan. Suasana hotel sangat nyaman, bersih, dan pelayanan yang ramah | berharap dapat datang kembali dengan keluarga Namun untuk dapat berbelanja sebaiknya menggunakan kendaraan karena aga jauh dari pusat pertokoan Suasana hotel sangat nyaman bersih dan pelayanan yang ramah | berharap datang kembali keluarga untuk berbelanja sebaiknya menggunakan kendaraan aga jauh pusat pertokoan Suasana hotel nyaman bersih pelayanan ramah | berharap berharap_dat ang datang datang_kemb ali kembali kembali_kelu arga keluarga keluarga_unt uk untuk untuk_berbel anja berbelanja berbelanja_se baiknya sebaiknya sebaiknya_m enggunakan menggunaka n menggunaka n_kendaraan kendaraan kendaraan_a ga aga aga_jauh jauh jauh_pusat pusat pusat_pertok oan pertokoan pertokoan_S uasana Suasana Suasana_hote l hotel hotel_nyama n nyaman nyaman_bers ih bersih bersih_pelay anan pelayanan pelayanan_ra mah ramah |

Sumber : Peneliti

## *C.* Metode Yang di Usulkan

Metode yang peneliti usulkan adalah menunggunkan metode pemilihan fitur yaitu *Particle Swarm Optimization,* yang digunakan untuk meningkatkan akurasi dari pengkasifikasi Naïve Bayes. Penelitian ini mengenai *review* hotel dengan menggunakan pengkasifikasi Naïve Bayes merupakan salah satu algoritma yang memiliki kecepatan yang tinggi saat diaplikasikan kedalam *database* dengan data yang besar. *Dataset* yang digunakan berasal dari www.tripadvisor.com yang terdiri dari 100 *review* positif dan 100 *review* negatif. Untuk *preprocessing* dilakukan *tokenization, stopword removal* dan *N-Grams*. Penelitian ini nantinya menghasilkan akurasi dan nilai *AUC* dengan menggunakan aplikasi *RapidMiner versi* 5.3 untuk hasil evaluasi.

## *D.* Model Dengan Metode Klasifikasi Naïve Bayes

Proses pengklasifikasian ini adalah menentukan *class* untuk setiap kalimat sebagai anggota *class* positif atau *class* negatif. Penentuan *class* pada setiap kalimat ditentukan melalui perhitungan probabilitas dari rumus Naïve Bayes. *Class* diberikan nilai Positif apabila nilai probabilitas pada dokumen tersebut untuk nilai *class* positifnya lebih besar dibandingkan dengan *class* negatif. Dan suatu kalimat dikatakan *class* negatif apabila nilai probabilitas pada dokumen tersebut untuk nilai *class* negatifnya lebih besar dibandingkan dengan *class* positifnya.

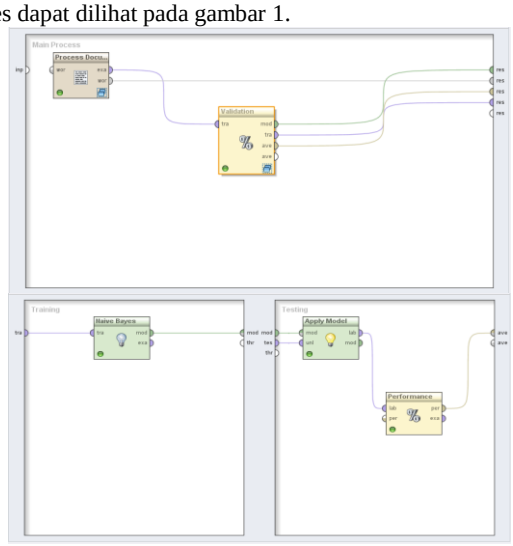
Peneliti hanya menampilkan 10 dokumen sentimen dari keseluruhan 200 data *training* dan 4 kata yang berhubungan dengan kata sentimen, yaitu bagus, nyaman, kotor dan buruk. Kehadiran kata dalam suatu kalimat akan diwakili oleh angka 1 dan angka 0 jika kata tersebut tidak muncul dalam kalimat pada dokumen.

Tabel 2. Hasil Klasifikasi Text

| Dokumen | Bagus | Nyaman | Kotor | Buruk | *Class* |
|---|---|---|---|---|---|
| Positif-1.txt | 1 | 1 | 0 | 0 | Positif |
| Positif-2.txt | 0 | 1 | 0 | 0 | Positif |
| Positif-3.txt | 1 | 0 | 0 | 0 | Positif |
| Positif-4.txt | 0 | 1 | 0 | 0 | Positif |
| Positif-5.txt | 1 | 0 | 0 | 0 | Positif |
| Negatif-1.txt | 0 | 0 | 1 | 0 | Negatif |
| Negatif-2.txt | 1 | 0 | 1 | 0 | Negatif |
| Negatif-3.txt | 0 | 0 | 1 | 0 | Negatif |
| Negatif-4.txt | 0 | 1 | 0 | 1 | Negatif |
| Negatif-5.txt | 0 | 0 | 1 | 0 | Negatif |

Sumber: Peneliti

Perhitungan diatas dapat dibuat suatu model dengan *RapidMiner* 5.3 Desain model arsitektur klasifikasi Naïve Bayes dapat dilihat pada gambar 1.



Sumber : Peneliti

Gambar 1. *Desain* Model *Arsitektur* Klasifikasi Naive Bayes

*E.* Hasil Eksperimen Menggunakan Algoritma Naïve Bayes

Dari data sebanyak 200 data *review* hotel yang terdiri dari 100 data *review* positif dan 100 data *review* negatif. Sebanyak 96 data di prediksi *class* negatif sesuai yaitu termasuk kedalam prediksi *class* negatif dan sebanyak 4 data di prediksi *class* negatif ternyata termasuk kedalam *class* positif, 85 data di prediksi *class* positif sesuai yaitu termasuk kedalam prediksi *class* positif dan sebanyak 15 data di prediksi *class* positif ternyata termasuk kedalam *class* negatif. Hasil yang diperoleh dengan menggunakan algoritma Naïve Bayes menggunakan *RapidMiner* 5.3 mendapatkan nilai *Accuracy* = 90.50% seperti pada tabel 3 dan mendapatkan nilai *AUC* : 0.500.

Tabel 3. *Confusion Matrix* Algoritma Naïve Bayes

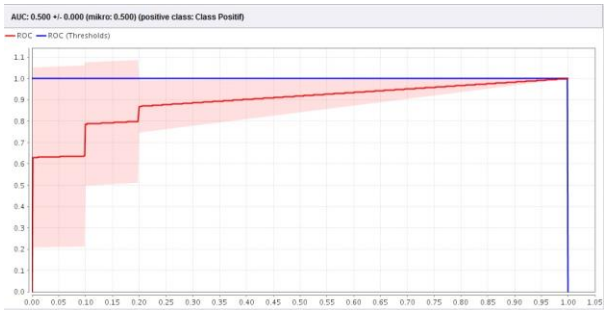
| *Accuracy* : 90.50% +/- 9.07% (mikro :90.50%) | | | |
|---|---|---|---|
| | *True Class Negatif* | *True Class Positif* | *Class Precision* |
| *Pred. Class Negatif* | 96 | 15 | 86.49% |
| *Pred. Class Positif* | 4 | 85 | 95.51% |
| *Class Recall* | 96.00% | 85.00% | |

Sumber : Peneliti

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

$$Accuracy = \frac{96 + 85}{95 + 15 + 4 + 85}$$

$$Accuracy = \frac{181}{200} = 0.905 = 90.50\%$$



Sumber : Peneliti

Gambar 2. *Grafik Area Under Curve (AUC)* Naïve Bayes

Pada penelitian ini, peneliti melakukan pengujian model dengan menggunakan teknik 10 *cross validation*, di mana proses ini membagi data secara acak ke dalam 10 bagian. Proses pengujian dimulai dengan pembentukan model dengan data pada bagian pertama. Model yang terbentuk akan diujikan pada 9 bagian data sisanya. Setelah itu proses akurasi dihitung dengan melihat seberapa banyak data yang sudah terklasifikasi dengan benar.

Teknik 10 *fold Cross Validation* ditentukan berdasarkan dari hasil uji coba peneliti untuk mendapatkan hasil akurasi yang tinggi, dalam hal ini yang akan di uji coba untuk meningkatkan akurasi adalah nilai *validation*. Tabel indikator dan hasil pengujian dapat dilihat pada tabel 4 pengujian model 10 *fold cross Validation*

Tabel 4. Pengujian Model 10 *Fold Cross Validation*

| *Validation* | *Accuracy (%)* |
|---|---|
| 1 | 85.00% |
| 2 | 85.00% |
| 3 | 87.52% |
| 4 | 85.00% |
| 5 | 89.00% |
| 6 | 88.47% |
| 7 | 89.06% |
| 8 | 87.50% |
| 9 | 90.47% |
| **10** | **90.50%** |

Sumber: Peneliti

*F.* Hasil Eksperimen Menggunakan Algoritma *Naïve Bayes* berbasis *Particle Swarm Optimization*

Untuk mendapatkan model yang terbaik peneliti mencoba meyesuaikan beberapa nilai parameter agar mendapatkan hasil akurasi yang tinggi. Pada model klasifikasi *Naïve Bayes* dan *Particle Swarm Optimization*, pertama dilakukan uji coba dengan dengan merubah nilai parameter *Population Size* dari 1-10 dengan nilai *inertia* nya 0.1 dan *maximum number of generation* 30 bernilai tetap. Berikut adalah hasil dari percobaan yang telah dilakukan untuk hasil nilai *Accuracy* dan AUC


p-ISSN: 2442-2436, e-ISSN: 2550-0120

Tabel 5. Hasil Eksperimen Menggunakan Algoritma Naïve Bayes Berbasis *Particle Swarm Optimization* Dengan Merubah Nilai Parameter Dari *Population Size*

| *Population Size (Q)* | *Inertia Weight (W)* | *Naïve Bayes* dan PSO | |
|---|---|---|---|
| | | *Accuracy* | AUC |
| 1 | 0.1 | 90.92% | 0.536 |
| 2 | 0.1 | 90.45% | 0.570 |
| 3 | 0.1 | 90.95% | 0.500 |
| 4 | 0.1 | 90.39% | 0.555 |
| 5 | 0.1 | 92.45% | 0.500 |
| 6 | 0.1 | 93.89% | 0.543 |
| 7 | 0.1 | 93.97% | 0.550 |
| 8 | 0.1 | 93.45% | 0.535 |
| 9 | 0.1 | 93.97% | 0.544 |
| **10** | **0.1** | **94.97%** | **0.578** |

Sumber : Peneliti

Dalam uji coba merubah nilai parameter *Population Size* pada *Particle Swarm Optimization*, akurasi dan AUC yang paling tinggi diperoleh dengan nilai *populalation size* 10. Percobaan kedua peneliti melakukan uji coba dengan merubah nilai *parameter Maximum Number Of Generation* dari 10-100, dengan nilai parameter *Population size* 10 dan nilai parameter *Inertia Weight 1.0.*

Tabel 6. Hasil Eksperimen Menggunakan Algoritma *Naïve Bayes* Berbasis *Particle Swarm Optimization* Dengan Merubah Nilai Parameter Dari *Maximum Number Of Generation*

| Population Size (Q) | Maximum Number Of Generation | Inertia Weight (W) | Naïve Bayes dan PSO | |
|---|---|---|---|---|
| | | | Accuracy | Auc |
| 10 | 10 | 1.0 | 91.42% | 0.529 |
| 10 | 20 | 1.0 | 91.95% | 0.616 |
| 10 | 30 | 1.0 | 93.95% | 0.634 |
| 10 | 40 | 1.0 | 95.97% | 0.589 |
| 10 | 50 | 1.0 | 93.45% | 0.530 |
| 10 | 60 | 1.0 | 94.97% | 0.500 |
| 10 | 70 | 1.0 | 94.97% | 0.550 |
| **10** | **80** | **1.0** | **96.00%** | **0.550** |
| 10 | 90 | 1.0 | 95.00% | 0.500 |
| 10 | 100 | 1.0 | 95.92% | 0.500 |

Sumber : Peneliti

Pada percobaan Kedua dengan mengubah nilai *Maximum Number Of Generation*, nilai akurasi dan AUC yang paling tinggi diperoleh dengan nilai *population Size* 10 dan *Maximum Number Of Generation* 80. Kemudian uji coba dilanjutkan dengan mengubah nilai parameter *Inertia Weight* dari 0.1-1.0.

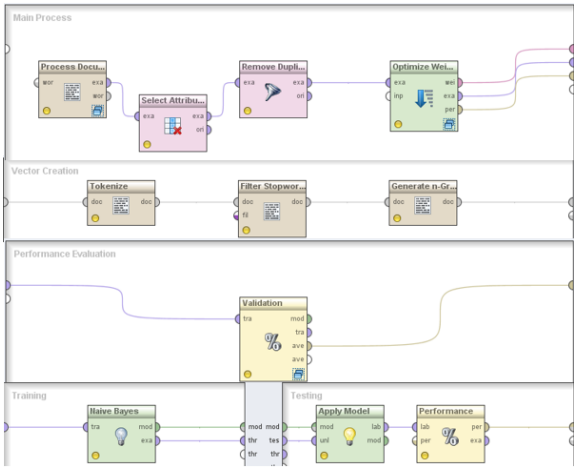
Tabel 7. Hasil Eksperimen Menggunakan Algoritma Naïve Bayes Berbasis *Particle Swarm Optimization* Dengan Merubah Nilai Parameter Dari *Inertia Weight*

| Population Size (Q) | Maximum Number Of Generation | Inertia Weight (W) | Naïve Bayes dan PSO | |
|---|---|---|---|---|
| | | | Accuracy | Auc |
| **10** | **80** | **0.1** | **96.92%** | **0.590** |
| 10 | 80 | 0.2 | 93.47% | 0.500 |
| 10 | 80 | 0.3 | 93.97% | 0.575 |
| 10 | 80 | 0.4 | 94.97% | 0.500 |
| 10 | 80 | 0.5 | 95.00% | 0.500 |
| 10 | 80 | 0.6 | 96.92% | 0.544 |
| 10 | 80 | 0.7 | 95.89% | 0.544 |
| 10 | 80 | 0.8 | 95.95% | 0.500 |
| 10 | 80 | 0.9 | 95.45% | 0.500 |
| 10 | 80 | 1.0 | 96.00% | 0.550 |

Sumber : Peneliti

Dalam percobaan hasil terbaik pada eksperimen *Naïve Bayes* dan *Particle Swarm Optimization* sebagai pemilihan fitur dengan merubah nilai parameter nilai *populalation size* 1-10, nilai *maximum number of generation* 10-100 dan nilai *inertia weight* nya 0.1-1.0.

Hasil akurasi dan AUC tertinggi pada saat nilai parmeter *Population Size* 10, *Maximum Number of Generation* 80 dan *Inertia Weight* 0.1 mencapai 96.92 % dan Nilai AUC 0.590 berikut ini desain model *Naïve Bayes* Dan pemilihan Fitur *Particle Swarm Optimization* ini dapat dilihat pada gambar 3




Sumber: Peneliti

Gambar 3 Desain Model *Naïve Bayes* Dan *Particle Swarm Optimization* Menggunakan *RapidMiner*

Tabel 8 Model Confusion Matrix Untuk Algoritma *Naïve Bayes* Berbasis *Particle Swarm Optimization*

| *Accuracy* : 95,92% +/- 5.08% (Mikro :95,96%) | | | |
|---|---|---|---|
| | *True Class Negatif* | *True Class Positif* | *Class Precision* |
| *Pred.Class* Negatif | 94 | 4 | 95.92% |
| *Pred.Class* Positif | 4 | 96 | 96.00% |
| *Class Recall* | 95.92% | 96.00% | |

Sumber: Peneliti

Berikut ini adalah tampilan kurva ROC yang akan dihitung nilai AUC nya dari review positif dan 100 review negatif yang diambil dari situs www.tripadvisor.com setelah menggunakan metode pemilihan fitur *Particle Swarm Optimization*


Sumber : Peneliti

Gambar 4 Kurva ROC *Naïve Bayes* Dan *Particle Swarm Optimization*

Hasil pengujian semua algoritma Naïve Bayes sebelum dan sesudah menggunakan metode pemilihan fitur *Particle Swarm Optimization* dapat dilihat pada tabel sebagai berikut.

Tabel 9 Hasil Ekperimen Algoritma *Naïve Bayes* dan *Particle Swarm Optimization*

| | Naïve Bayes | Naïve Bayes dan *PSO* |
|---|---|---|
| *Accuracy* | 90,50% | 96,92% |
| AUC | 0,500 | 0,590 |

Sumber :Peneliti

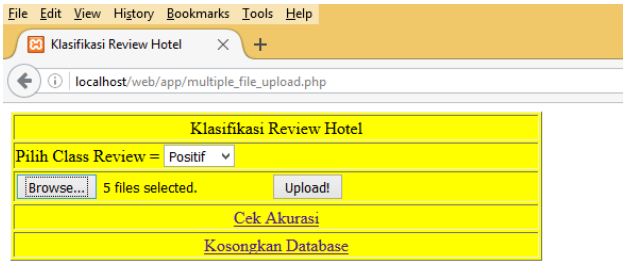
Berdasarkan hasil evaluasi menggunakan *Confusion Matrix* maupun *ROC Curve* terbukti bahwa optimasi *Particle Swarm Optimization* pada proses optimasi metode dapat meningkatkan nilai akurasi algoritma *Naïve Bayes*. Percobaan yang telah dilakukan memperoleh nilai akurasi *Naïve Bayes* 90.50 % sedangkan nilai akurasi *Naïve Bayes* setelah menggunakan pemilihan fitur *Particle Swarm Optimization* paling tinggi saat nilai parameter *Population Size* 10 dengan *Maximum Number Of Generation* 80 dan *Inertia Weight* 0.1 sebesar 96.92 %. Pada ROC Curve dapat dilihat Nilai AUC untuk algoritma Naïve Bayes sebesar 0.500, sedangkan setelah menggunakan *Particle Swarm Optimization* menjadi 0.590.

Nilai Akurasi ini mengalami peningkatan sebesar 6.42 % dari penggunaaan Naïve Bayes Sebelem menambahkan metode pemilihan fitur *Particle Swarm Optimization*

### *G.* Implementasi

Peneliti membuat aplikasi untuk menghitung nilai akurasi, menguji model yang sudah ada menggunakan *dataset* dalam *review* hotel. Hasil akurasi dari penelitian akan diterapakan kedalam pembuatan aplikasi untuk klasifikasi *review* hotel menggunakan perangkat lunak *dreamweaver* CS 3 menggunakan bahasa pemrograman php, sehingga dapat mengetahui nilai akurasi dari jumlah *review* menggunakan Algoritma *Naïve Bayes.*

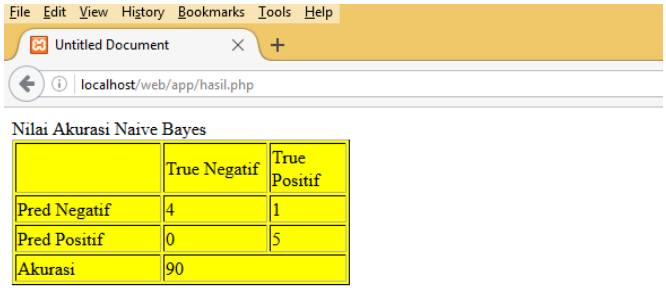
Desain aplikasi dibuat untuk memudahkan pengguna dalam menganalisa *review* hotel berdasarkan pengalaman orang lain. Seperti berikut :




Sumber : Peneliti

Gambar 5. Tampilan Aplikasi *Input Review* Positif dan *Review* Negatif

Pengguna dapat melakukan analisa review text dengan menyimpan sebuah review ke dalam bentuk file berekstensi .txt, maka pengguna dapat langsung mengupload file dengan memilih *class review* terlebih dahulu kemudian cari *review* yg telah disimpan lalu *upload*, setelah mengupload *review* positif dan *review* negatif kemudian klik cek akurasi maka akan tampil hasil akurasi seperti gambar dibawah ini.




sumber : Peneliti

Gambar 6. Tampilan Aplikasi Hasil Analisa Sentimen Review

## IV. KESIMPULAN

Berdasarkan klasifikasi teks dengan data *review* hotel, yang terdiri dari 100 *review* positif dan 100 *review* negatif yang diambil dari situs www.tripadvisor.com salah satu metode klasifikasi yang dapat digunakan adalah pengklasifikasi Naïve Bayes. Dalam hal ini Algoritma Naïve Bayes merupakan metode klasifikasi yang sangat sederhana dan efisien. Selain itu Naïve Bayes merupakan pengklasifikasi teks yang sangat popular yang memiliki performa yang sangat baik pada banyak domain baik dalam klasifikasi teks.

Dari pengolahan data yang sudah dilakukan, menggunakan metode pemilihan fitur yaitu *Particle Swarm Optimization* terbukti dapat meningkatkan akurasi pada pengklasifikasi Naïve Bayes. Data *review* hotel berbahasa Indonesia diklasifikasi dengan baik kedalam *review* positif maupun *review* negatif. Akurasi model Naïve Bayes sebelum menggunakan metode pemilihan fitur *Particle Swarm Optimization* mencapai 90.50%, sedangkan setelah menggunakan metode pemilihan fitur *Particle Swarm Optimization* akurasinya meningkat menjadi 96.92%, dapat meningkatkan akurasi sebesar 6.42%, Dalam mendukung klasifikasi teks berbahasa Indonesia, peneliti mengembangkan


p-ISSN: 2442-2436, e-ISSN: 2550-0120

aplikasi *review* hotel untuk mengklasifikasi *review* positif dan *review* negatif menggunakan bahasa pemrograman PHP.

Model yang dibentuk dapat diterapkan pada seluruh *review* hotel, sehingga dapat langsung hasilnya dalam mengklasifikasi teks pada *review* termasuk kedalam *review* positif atau *review* negatif. Sehinga dapat membantu pengunjung atau pemesan hotel dalam mengambil keputusan dangan cepat dan efisien saat memesan penginapan tanpa harus khawatir adanya pemberian rating yang tidak sesuai dengan *review*nya dan dapat memberikan informasi dalam menentukan kamar yang disediakan sesuai dengan keinginan pengunjung hotel, untuk meningkatkan kenyamanan dan pelayanan hotel kedepanya.

## REFERENSI

Basari, A. S. H., et al. "Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization." *Procedia Engineering, 53,* (2013.): 453-462. doi:10.1016/j.proeng. 2013.02.059.

Bramer, Max. *Principles of Data Mining*. London: Springer, 2007.

Charjan, Miss Dipti S. and Mukesh A. Pun . "Pattern Discovery For Text Mining Using Pattern Taxonomy." *International Journal* (2013).

Chen, J., et al. "Feature selection for text classification with Naïve Bayes." *Expert Systems with Applications, 36, no 3 pp.* (2009): 5432– 5435.

Duan, W, et al. "Mining Online User-Genenrated Content : Using Sentimen Analisis Technique to Study Hotel Quality." (2013).

Gencosman, B. C., H. C. Ozmutlu and S. Ozmutlu. "Character n-gram application for automatic new topic identification." *Information Processing and Management, 50,* (2014): 821-856. doi:10.1016/j.ipm.2014.06.005.

Gorunescu, F. *Data Mining: Concepts, Models and Techniques*. Berlin:: Springer, 2011.

Haddi, E., X. Liu and Y. Shi. "The Role of Text Pre-processing in Sentiment Analysis." *Procedia Computer Science, 17,* (2013): 26-32. doi:10.1016/j.procs.2013.05.005.

Han, J. and M. Kamber. *Data Mining Concepts and Techniques.* 2007.

Intan, R. and A. Defeng. *Subject Based Search Engine Menggunakan TF-IDF dan Jaccard's Coefficient,*. Surabaya: Universitas Kristen Petra, 2006.

Kontopoulos, E., et al. "Ontology-based sentiment analysis of twitter post." *Expert Systems with Applications, 40* (2013): 4065-4074. doi:10.1016/j.eswa.2013.01.001.

Liu, B. "Sentiment Analysis and Opinion Mining." *Synthesis Lectures on Human Language Technologies, 5 (May),* (2012): 1–167.

Lu, Y , et al. "Improved particle swarm optimization algorithm and its application in text feature selection." *Applied Soft Computing, 35,* (2015): 629–636.

Maimon , O. and L. Rokach. *Data Mining and Knowledge Discovery Handbook, Second.* Boston: MA: Springer US, 2010.

Markopoulos, G, et al. "Sentiment Analysis of Hotel Reviews in Greek: A Comparison of Unigram Features." *Springer Proceeding in Business and Economics* (2015): DOI 10.1007/978-3-319-15859-4_3.

Medhat, W., A. Hassan and H. Korashy. "Sentiment analysis algorithms and applications: A survey." *Ain Shams Engineering Journal, 5(4), pp.* (2014): 1093–1113.

Robertson, S. "Understanding Inverse Document Frequency: On Theoritical Arguments for IDF." *Journal of Documentation; 2004; 60, 5; ABI/INFORM Global.* (2014).

Suardika, I. G. "Sentiment Analysis System And Correlation Analysis On Hospitality In Bali." *Journal Of Theoretical and Applied Information Technology (Vol.84. No.1).* (2016).

Taylor, E. M., et al. "Indentifying Customer Preferences about Tourism Products using an Aspect-Based Opinion Mining Approach." *Procedia Computer Science, 22,* (2013): 182-191. doi:10.1016/j.procs.2013.09.094.

Witten, H. I., E. Frank and M. A. Hall. *Data Mining Practical Machine*. 2011.

Zhang, Z., Q. Ye and Y. Li. "Sentiment classification of Internet restaurant reviews written in Cantonese." *Expert Systems with Applications, 38(6),* (2011): 7674–7682. doi:10.1016/j.eswa.2010.12.147.

Zhao, M., et al. "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),* (2011): 5197–5204. doi:10.1016/j.eswa.2010.10.041.

### PROFIL PENULIS

Andi Taufik Lahir di Bogor 30 November 1991. Lulus Pasca Sarjana STMIK Nusa Mandiri tahun 2016.