Volume IV No. 2 Agustus 2018
P-ISSN 2442-2436, E-ISSN: 2550-0120
http://ejournal.bsi.ac.id/ejurnal/index.php/jtk

JURNAL TEKNIK KOMPUTER AMIK BSI

Komparasi Algoritma Text Mining Untuk Klasifikasi Review Hotel

Andi Taufik

STMIK Nusa Mandiri Jakarta
Jl. Damai No. 8 Warung Jati Barat Jakarta Selatan
e-mail: a.taufik30@gmail.com

Cara Sitasi: Taufik, A. (2018). Komparasi Algoritma Text Mining Untuk Klasifikasi Review Hotel. Jurnal Teknik Komputer, IV(2), 112-118. doi:10.31294/jtk.v4i2.3461

Abstract - The use of information technology using the internet is very easy to find information, so its users do not have to go to the place to find information, website and mobile applications internet users can directly get the information you want. Every Manager world tourism now provides the details about the tourism products offered. Very useful information at this time because people tend to look for quick information in the booking through the review of others in social media, blogs and websites. The importance of the review of the hotel as a source of information for travelers will plan trips. Currently known methods of classification of the highest accuracy in classifying hotels Indonesia-language review. So I need to know how Naïve Bayes algorithms of accuracy, Support Vector Machine, Decission Tree (C4.5) and Naïve Bayes Method with Particle Swarm Optimization Feature Selection. The results obtained from the comparison of four methods of such algorithms, a better level of accuracy in the classification review of hotel indonesian using an algorithm Decission Tree (C4.5) 96.94% While achieving the fit method of optimization of the Nave bayes by using Particle Swarm Optimization feature Selection of 95.91%, accuracy using Naive Bayes Algorithm of 89.98% and the accuracy of the model of Support Vector Machine of 89.86%.

Keywords: Reviews Hotel, Naive Bayes, PSO, SVM, Decission Tree, C4.5

PENDAHULUAN

Kemudahan dalam penggunaan teknologi informasi khusunya internet membuat para pengguna internet dengan mudah mendapatkan informasi yang diinginkan. Para pengguna teknologi informasi kini banyak memanfaatkan informasi yang disediakan tidak terkecuali informasi mengenai hotel seperti identifikasi hotel, rekomendasi hotel dan kemudahan pemesanan hotel baik dengan website ataupun dengan mobile. Review dan perbandingan harga hotel sangat mempengaruhi pertumbuhan pemesanan kamar hotel dan tamu yang datang (Marrese-Taylor, Velásquez, Bravo-Marquez, & Matsuo, 2013).

TripAdvisor merupakan salah satu situs wisata yang dikenal di dunia yang dapat membantu para wisatawan merencanakan dan memesan perjalanan wisata. Pada situs ini para wisatawan dapat memberikan penilai baik berupa rating bintang tinggi, menengah dan rendah atau memberikan testimoni baik maupun buruk dengan tingkat kepuasan pelanggan dalam pelayanan, kenyamanan dan fasilitas yang telah diberikan.

Dengan memanfaatkan perkembangan teknologi informasi melalui pengguna jejaring sosial mengenai review hotel, yang menyediakan ulasan untuk pengunjung dan dapat digunakan untuk berinteraksi dengan pengunjung lainnya, *platform* ini digunakan sebagai wadah untuk membuat dan mendengar pendapat pengunjung yang menghasilkan ulasan perjalanan dan jasa perhotelan yang sudah dikunjungi saat liburan dan menjadi sumber informasi yang penting bagi pengunjung lainnya (Duan, Cao, Yu, & Levy, 2013)

Pengelola dunia pariwisata dalam memberikan informasi dapat lebih detail mengenai produk pariwisata yang ditawarkan. Saat ini sebelum memesan, seseorang akan memeriksa pendapat dari web review yang ada. Hotel merupakan salah satu produk dari pariwisata yang sangat dipertimbangkan baik dari segi fasilitas, pelayanan ataupun jarak tempuh perjalanan wisata (Marrese-Taylor et al., 2013).

Untuk memudahkan para wisatawan mengetahui hotel yang cocok untuk mereka, para wisatawan dapat melihat review dari pengalaman pengunjung sebelumnya yang ada pada situs TripAdvisor. Tetapi untuk membaca review atau opini yang ada pada situs tersebut membutuhkan waktu yang cukup lama, namun jika hanya sedikit review yang dibaca evaluasi akan bias (Ziqiong, Qiang, Zili, & Yijun, 2011).

 Diterima: 2018-04-05, Direvisi: 2018- 06-26, Disetujui: 2018-07-20

Terdapat penelitian yang sudah dilakukan sebelumnya dalam hal pengklasifikasian analisis sentiment terhadap review, yaitu:

a. Penelitian yang dilakukan oleh Suardika dengan judul *Sentimen Analysis System and Correlation Analisys on Hospitality in* Bali. Pada penelitian ini menggunakan metode Naïve Bayes untuk mencari hubungan peringkat antar hotel pada TripAdvisor dan menggunakan *text processing* tokenization dan *remove all token*. Pada penelitian ini mendapatkan nilai *accuracy* 81% (Gede Suardika, 2016).
b. Penelitian yang dilakukan oleh Zhang, Ye dan Li dengan judul *Sentiment classification of Internet restaurant reviews written in Cantonese*. Pada penelitian ini dilakukan pengklasifikasian sentiment pada review restoran di internet yang ditulis dalam Bahasa Canton. Pada penelitian ini penelitian dengan menggunakan algoritma Support Vector Machine (SVM) dengan Bigram_freq mendapatkan nilai akurasi 94.83% sedangkan dengan algoritma Naïve Bayes (NB) dengan Bigram mendapataan nilai akurasi 95.67 (Ziqiong et al., 2011).
c. Penelitian yang dilakukan oleh Markopoulos, Mikros dan Iliadi dengan judul *Sentiment Analysis of Hotel Reviews in Greek: A Comparison of Unigram Features*. penelitian ini membuat classifier sentiment yang menerapkan Support Vector Machine dengan fitur Unigram pada review hotel dalam Bahasa Yunani. Dalam penelitian ini membandingkan dua metodologi yang berbeda yaitu dengan TF IDF dan The Term Occurrence (TO). Dengan menggunakan TF IDF mendapatkan nilai akurasi 95.78% sedangkan dengan metode TO mendapatkan nilai akurasi 71.76% (Markopoulos, Mikros, & Iliadi, 2015).
d. Penelitian yang dilakukan Taufik dengan judul Optimasi *Particle Swarm Optimization* Sebagai Seleksi Fitur Pada Analisis Sentimen *Review* Hotel Berbahasa Indonesia Menggunakan Algoritma Naïve Bayes. Pada penelitian ini dilakukan pengklasifikasian sentiment pada review hotel berbahasa Indonesia pada TripAdvisor. Pada penelitian ini dengan metode Naïve Bayes mendapatkan nilai akurasi 90.50% dan metode Naïve Bayes dengan pemilihan fitur Particle Swarm Optimization mendapatkan nilai akurasi 96.92% (Taufik, 2017).

Saat ini belum diketahui metode klasifikasi yang paling tinggi akurasinya dalam mengklasifikasikan review hotel berbahasa Indonesia. Sehingga perlu diketahui bagaimana akurasi dari algortima Naïve Bayes, Support Vector Machine, *Decission Tree* dan Metode Naïve Bayes dengan Pemilihan Fitur Particle Swarm Optimization.

Maksud dari penelitian ini adalah untuk mengetahui algoritma mana yang mendapatkan nilai akurasi yang terbaik antara algoritma Naïve Bayes, Support Vector Machine, *Decission Tree* dan Metode Naïve Bayes dengan Pemilihan Fitur Particle Swarm Optimization.

1. Tinjauan Pustaka
   a. Data Mining
      Menurut (Witten, Frank, & Hall, 2011) Data mining merupakan perpaduan dari ilmu statistik, kecerdasan biatan (sistem pakar) dan penelitian dalam bidang database, untuk itu dibutuhkan penyaringan melalui sejumlah besar material data atau melakukan penyelidikan dengan cerdas tentang keberadaan suatu data yang memiliki nilai *Daryl Pregibons*.

   b. Text Mining
      Text mining adalah penemuan dari pengetahuan yang menarik pada dokumen teks. Hal ini merupakan tantangan untuk menemukan pengetahuan yang akurat pada dokumen teks untuk menolong pengguna dalam menemukan apa yang diinginkan. Penemuan pengetahuan dapat menjadi efektif digunakan dan memperbaharui pola penemuan dan menerapkannya ke *text mining* (Charjan & Pund, 2013).

   c. Sentimen Analisis
      Menurut (Kontopoulos, Berberidis, Dergiades, & Bassiliades, 2013), *Opinion mining* atau dikenal sebagai analisa sentiment adalah proses yang bertujuan untuk menentukan apakah polaritas kumpulan teks tulisan (dokumen, kalimat, paragrap, dan lain-lain) cenderung ke arah positif, negatif, atau netral.

   d. Pre-Processing
      Menurut (Haddi, Liu, & Shi, 2013), Preprocessing adalah merupakan proses pembersihan dan mempersiapkan teks untuk klasifikasi. Seluruh proses melibatkan beberapa langkah: membersihkan teks online, penghapusan ruang spasi, memperluas singkatan dasar (*stemming*), penghapusan kata henti (*Stopword removal*), penganganan negasi dan terakhir seleksi fitur.

      N-gram didefiniskan sebagai sub-urutan n karakter dari kata diberikan (Gencosman & Ozmutlu, Huseyin C., 2014).

   e. TF-IDF
      Metode ini akan menghitung nilai *Term Frequency* (TF) dan *Inverse Document*

Frequency (IDF) pada setiap kata di setiap dikomen dalam korpus.

a) Rumus umum untuk pembobotan TF-IDF :
$W = tf * idf$ ..... (1)
$W = tf * log(\frac{N}{df})$ .....(2)

b) Berdasarkan rumus (2), 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:
$W = tf * (log\left(\frac{N}{df}\right) + 1)$.....(3)

c) Rumus (3) dapat dinormalisasi dengan rumus (4) dengan tujuan menstandarisasi nilai bobot (wtd) ke dalam interval 0 s.d. 1 :
$W = \frac{tf*(log\left(\frac{N}{df}\right)+1)}{\sqrt{\sum_{k=1}^{t}(tf)^2*\left(log\left(\frac{N}{df}\right)+1\right)^2}}$..... (4)

f. Pemilihan Fitur
Menurut (Tsoumakas, Katakis, & Vlahavas, 2010), 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 John, kohavi dan pfleger dalam (Chen, Jingnian, Houkuan Huang, Shengfeng Tian, 2009), ada dua jenis metode seleksi fitur dalam pembelajaran *machine learning*, yaitu itu *wrappers* dan *filters.*

g. *Naïve Bayes*
Naïve bayes merupakan klasifikasi data dengan menggunakan probabilitas dan static. Menurut (Han, Kamber, & Pei, 2012) tahapan dalam algoritma Naïves Bayes:

a) Perhatikan D adalah record training dan ketetapan label-label kelasnya dan masing-masing record dinyatakan n atribut (n field)
$X = (X_1, X_2 \ldots, X_n)$..... (5)

b) Misalkan terdapat m kelas
$C_1, C_2 \ldots, C_m$).....(6)

c) Klasifikasi adalah diperoleh maximum posteriori yaitu maximum P(Ci|X)

d) Ini diperoleh dari teorema Bayes
$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}$..... (7)

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

$P(C_i|X) = P(X|C_i)P(C_i)$..... (8)

h. *Support Vector Machine (*SVM*)*
Support Vector Machine merupakan *machine learning* yang termasuk dalam model *supervised* learning atau pembelajaran dengan pengawasan yang berhubungan dengan analisis data dan pengenalan pola (Putra & Irawati, 2015).

Fungsi keputusan klasifikasi sign (f(x)):
$f(x) = wx + b$.....(9)
atau
$fx = \sum_{i=1}^{m} a_i y_i\ K(x, x_i) + b$.....(10)

Keterangan:
N : Banyaknya data
n : dimensi data atau banyaknya fitur
Ld : Dualitas Lagrange Multipier
αi : nilai bobot setiap titik data
C: nilai konstanta
M :jumlah support vector/titik data yang memiliki αi > 0
K(x,xi) : fungsi kernel

i. *Decission Tree (*C4.5*)*
Merupakan metode klasifikasi yang melibatkan konstruksi pohon keputusan, koleksi node keputusan, terhubung oleh cabang-cabang, memperpanjang bawah dari simpul akar sampai berakhir di node daun (Sukardi & Supriyanto, 2014).

Tahapan dalam membuat sebuah pohon keputusan dengan algoritma C4.5.

1) Mempersiapkan data training
2) Menghitung total entropy sebelum di cari masing-masing entropy class
$H(T) = -\sum_j Pj\ log2(Pj)$.....(11)
Keterangan:
H: Himpunan kasus
T: Atribut
Pj: proposi dai Hj terhadap H
3) Hitung nilai Gain dengan information gain dengan rata-rata

$Gain\ average = H(T) - Hsaving(T)$...(12)

Keterangan:
H(T)=Total Entropy
Hsaving(T)=Total Gain information untuk masing-masing atribut

j. *Particle Swarm Optimization (*PSO)
Menurut (Lu, Liang, Ye, & Lichao, 2015), *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.

METODOLOGI PENELITIAN

A. Metode Penelitian

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

1. Pengumpulan Data
   Pengumpulan data menggunakan review yang ada pada TripAdvisor dengan menggunakan 50 data review positif dan 50 data review negatif
2. Pengolahan Data Awal
   Dalam pengolahan data awal dilakukan tahap preprocessing dengan melalui beberapa proses yaitu, *tokenisasi, stopword removal, steeming dan N-grams*.
3. Metode yang diusulkan
   untuk mengetahui metode klasifikasi data mining yang paling akurat pada review hotel. Metode yang digunakan yaitu Naïves Bayes, Support Vector Machine, C4.5 dan dan metode Naïve Bayes dengan pemilihan fitur Particle Swarm Optimization.
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.

B. Eksperimen dan Pengujian Model

Proses eksperimen yang dilakukan dalam melakukan pengujian model menggunakan RapidMiner 5.3. dengan dataset review hotel. Tahapan dalam melakukan pengujian review hotel sebagai berikut:

1. Menyiapkan dataset untuk eksperimen yang sudah diketahui classnya
2. Mendesain arsitektur algoritma klasifikasi Naïve Bayes, Support Vector Machine, C4.5 dan metode pemilihan fitur Particle Swarm Optimization pada klasifikasi Naïve Bayes.
3. Melakukan training dan testing terhadap algoritma Naives Bayes, Support Vector Machine, C4.5 dan metode Naïve Bayes dengan pemilihan fitur Particle Swarm Optimization, kemudian mencatat hasil accuracy dan AUC

C. Evaluasi dan Validasi Hasil

Validasi dilakukan menggunakan validation tertinggi dalam tiap algoritma. Sedangkan pengukuran akurasi diukur dengan confusion matrix dan kurva ROC (Receiver Operating Characteristics) untuk mengukur nilai AUC.

HASIL DAN PEMBAHASAN

A. Hasil Pengujian menggunakan Algoritma Decission Tree (C4.5)

Pengujian Algoritma C4.5 Untuk mendapatkan nilai akurasi yang paling tinggi maka peneliti melakukam pengujian dengan merubah nilai validasi nya dari 1sampai 10, berikut ini pengujiannya

Tabel 1 Pengujian Algoritma C4.5

Validation	Accuracy	AUC
1	93,00%	0.682
2	93,00%	0.682
3	94,00%	0.794
4	96,00%	0.950
5	95,00%	0.870
6	94,91%	0.887
7	**96,94%**	**0.959**
8	94,95%	0.939
9	96,04%	0.952
10	95,00%	0.959

Sumber : Peneliti

Berdasarkan hasil pengujian pada tabel 1 dengan merubah angka validasi 1 sampai dengan 10 untuk mendapatkan nilai akurasi tertinggi didapatkan pada nilai validasi 7 sebesar 96.94 % dengan nilai AUC 0.959

B. Hasil Pengujian Menggunakan Metode Naïve Bayes dengan Pemilihan Fitur Particle Swarm Optimization

Untuk mendapatkan nilai akurasi tertinggi pada Metode Optimasi *Naive Bayes* dengan pemilihan *Fitur Particle Swarm Optimization*, peneliti melakukan beberapa pengujian dengan merubah nilai parameter.

Pengujian pertama dilakukan dengan merubah nilai Validasi dari 1 sampai 10 dangan nilai Population Saze 10, Maximum Number Of Generation 80 dan nilai Ineteria Weight 1.0 bernilai tetap. Berikut ini

hasil pengujianya hanya menampilkan akurasi tertinggi yaitu pada validasi 6 sampai 10

Tabel 2 Pengujian Metode Optimasi Naive Bayes dan Particle Swarm Optimization Dengan merubah nilai Validasi 1-10

Validasi	Population Size (P)	Maximum Number Of Generation	*Ineteria Weight (w)*	Accuracy
6	10	80	1.0	92.89%
7	10	80	1.0	93.88%
8	10	80	1.0	**95.83%**
9	10	80	1.0	94.85%
10	10	80	1.0	93.89%

Sumber : Peneliti

Dari Tabel 2 bisa dilihat nilai akurasi paling tinggi pada nilai vaklidasi 8 sebesar 95.83%

Pengujian ke dua dilakukan untuk mendapatkan nilai akurasi lebih tinggi lagi dengan merubah nilai parameter Population Size 1 Sampai 10, nilai Validasi 8, *Maximal Of Generation* 80 dan nilai *Ineteria Weight* 1.0 bernilai tetap berikut ini pengujianya.

Tabel 3 Pengujian Metode Optimasi Naive Bayes dan Particle Swarm Optimization Dengan merubah nilai Population Size 1-10

Validasi	Population Size (P)	Maximum Number Of Generation	*Ineteria Weight (w)*	Accuracy
8	6	80	1.0	94,87%
8	7	80	1.0	95,91%
8	8	80	1.0	94,95%
8	9	80	1.0	94,95%
8	10	80	1.0	95,83%

Pada tabel 3 hasil pengujian ke dua dapat meningkatkan nilai akurasi dengan merubah parameter nilai Population Size, nilai akurasi tertinggi dapat dilihat pada nilai population size 7 dengan validasi 8, *Maximal Of Generation* 80 dan nilai *Ineteria Weight* 1.0 bernilai tetap

C. Hasil Pengujian Menggunakan Algoritma Naïve Bayes

Pengujian Algoritma Naive bayes menggunakan teknik 10 *fold Cross Validation* untuk mendapatkan nilai akurasi yang tinggi berikut ini adalah pengujianya.

Tabel 4 Pengujian Algoritma Naïve Bayes

Validation	Accuracy	AUC
1	83,00%	0.682
2	83,00%	0.682
3	79,06%	0.500
4	86,00%	0.500
5	84,00%	0.550
6	83,95%	0.500
7	85,03%	0.500
8	86,86%	0.500
9	**89,98%**	**0.500**
10	84,00%	0.500

Sumber : Peneliti

Hasil pengujian Naïve Bayes nilai akurasi paling tinggi terdapat pada Validasi 9 sebesar 89.98%

D. Hasil Pengujian Menggunakan Algoritma Support Vector Machine

Untuk mendapatkan Hasil yang baik peneliti mencoba merubah beberapa parameter agar mendapatkan hasil akurasi yang tinggi. Model klasifikasi menggunakan Algoritma SVM , uji coba yang pertama dilakukan adalah merubah nilai Validasi 1-10, nilai C 0.0 dan nilai Epsilion 0.0 bernilai tetap, berikut ini adalah hasil pengujian yang telah dilakukan hanya menampilkan pengujian diambil dari yang paling besar pada validasi 6 sampai dengan 10

Tabel 5 Hasil pengujian Menggunakan Algoritma Support Vector Machine Dengan nilai C 0.0 dan nalai Epsilion 0.0

Validasi	C	Epsilion	*Accuracy*	AUC
6	0.0	0.0	66,97%	0.968%
7	0.0	0.0	65,51%	0.977%
8	0.0	0.0	67.63%	0.970%
9	0.0	0.0	66,08%	0.989%
10	0.0	0.0	60,00%	0.960%

Sumber : Peneliti

Dari tabel 5 hasil pengujian Algoritma Support Vector Machine masih sangat rendah nilai akurasi tertinggi pada validasi 8 sebesar 67,63%.

Pengujian ke dua untuk mencari nilai akurasi lebih tinggi peneliti merubah angka Validasi 1 sampai 10 dengan nilai parameter C 0.1 dan epsilion 0.1

Tabel 6 Hasil pengujian Menggunakan Algoritma Support Vector Machine Dengan nilai C 0.1 dan nilai Epsilion 0.1

Validasi	C	Epsilion	*Accuracy*	AUC
6	0.1	0.0	85.11%	0,965
7	0.1	0.0	88.84%	0.968
8	0.1	0.0	85.90%	0,967
9	0.1	0.0	88.05%	0.985
10	0.1	0.0	87.00%	0,956

Sumber : Peneliti

Hasil pengujian ke dua yang telah dilakukan mengalami peningkatan peningkatan pada nilai akurasi 7 dengan ni;ai C 0.1 dan Epsilion 0.1 sebesar 88.84%.

Pengujian ke tiga untuk mendapatkan nilai akurasi lebih baik lagi maka peneliti melakukan uji coba dengan merubah parameter nilai validasi 1-10, nilai C 1.0 dan nilai Epsilion 0.0 bernilai tetap, dibawah ini adalah hasil pengujianya

Tabel 7 Hasil Eksperimen Menggunakan Algoritma *Support Vector Machine* dengan nilai C 1.0 dan nilai *Epsilion* 0.0

Validasi	C	Epsilion	*Accuracy*	AUC
6	1.0	0.0	85,11%	0,965
7	1.0	0.0	**89.86%**	**0.951**
8	1.0	0.0	84.86%	0,967
9	1.0	0.0	87,04%	0,985
10	1.0	0.0	87,00%	0,952

Sumber : Peneliti

Dari Tabel 7 Hasil pengujian yang telah dilakukan oleh peneliti untuk mendapatkan nilai akurasi tertinggi didapatkan pada nilai akurasi 7 dengan nilai parameter C 1.0 dan nilai Epsilion 0.0 sebesar 89.86%.

Hasil pengujian komparasi Metode Algoritma Text Mining pada klasifikasi Review Hotel adalah Sebagai Berikut

Tabel 8 Hasil Komparasi Metode Pada Klasifikasi Review Hotel

Metode	*Accuracy*	AUC
C4.5	**96,94%**	**0.959**
NB dan PSO	95,91%	0.500
Naive Bayes	89,98%	0.500
SVM	89.86%	0.951

Sumber :Peneliti

Berdasarkan Tabel 8 hasil komparasi menggunakan *Confusion Matrix* Maupun *ROC Curve* maka akurasi yang didapatkan untuk metode menggunakan algoritma C45 sebesar 96,94%, sedangkan akurasi menggunakan Metode *Optimasi Naive Bayes* dengan pemilihan fitur *Particle Swarm Optimization* sebesar 95,91, akurasi menggunakan Algoritma Naive bayes sebesar 89.98% dan akurasi Algoritma *Support Vector Machine* sebesar 89.86%. Pada ROC Curve dapat dilihat AUC Algoritrma C45 sebesar 0.959 sedangkan AUC metode Optimasi Naive Bayes dan PSO sebesar 0.500, AUC menggunakan Algoritma Naive Bayes sebesar 0.500 dan AUC Algoritma Support Vector Machine sebesar 0.951.

Nilai Akurasi dan AUC C45 lebih unggul dibandingkan dengan tiga metode lain nya.

KESIMPULAN

Dari hasil pengujian komparasi empat metode yang berbeda yaitu menggunakan Algoritma Support Vector Machine, Naive Bayes, Optimasi Naive Bayes dengan Particle Swarm Optimization dan Algoritma C45 untuk mengetahui prediksi metode yang lebih akurat dalam algoritma text mining pada klasifikasi review hotel. Akurasi Algoritma C45 mencapai 96,94 % Sedangkan Metode Optimasi Nave bayes dengan menggunakan Pemilihan fitur *Particle Swarm Optimization* sebesar 95,91%, akurasi menggunakan Algoritma *Naive Bayes* sebesar 89,98% dan Akurasi model *Support Vector Machine* sebesar 89,86%.

Dengan pengolahan hasil pengujian komparasi metode Algoritma Support Vector Machine, Naive Bayes, OptimasiNaive Bayes dengan pemilihan fitur Particle Swarm Optimization dan Algoritma C4.5 dapat sebuah kesimpulan yaitu algoritma C4.5 lebih unggul dalam memprediksi klasifikasi review hotel.

Dengan menerapkan Algoritma C4.5 pada klasifikasi review hotel dapat membantu pengunjung situs online yang mencari tempat penginapan atau hotel dalam pengambilan keputusan yang lebih cepat dan efisien tanpa harus membaca satu persatu review hotel dari pengunjung sebelumnya dan langsung bisa membandingkan fasilitas dan harga yang diinginkan pengunjung

REFERENSI

Charjan, D. S., & Pund, P. M. A. (2013). Pattern Discovery For Text Mining Using Pattern Taxonomy, *4*(10), 4550–4555.

Chen, Jingnian, Houkuan Huang, Shengfeng Tian, Y. Q. (2009). Feature selection for text classification with Naïve Bayes. *Expert Systems with Applications: An International Journal*, *36*(3), 5432– 5435. https://doi.org/10.1016/j.eswa.2008.06.054

Duan, W., Cao, Q., Yu, Y., & Levy, S. (2013). Mining Online User-Generated Content: Using Sentiment Analysis Technique to Study Hotel Service Quality. *2013 46th Hawaii International Conference on System Sciences*, 3119–3128. https://doi.org/10.1109/HICSS.2013.400

Gede Suardika, I. (2016). Sentiment analysis system and correlation analysis on hospitality in Bali, *84*(1), 88–95.

Gencosman, B. C., & Ozmutlu, Huseyin C., S. O. (2014). Character n-gram application for automatic new topic identification. *Information Processing and Management*, *50*(6), 821–856. https://doi.org/doi.org/10.1016/j.ipm.2014.06.005

Haddi, E., Liu, X., & Shi, Y. (2013). The role of text pre-processing in sentiment analysis. *Procedia Computer Science*, *17*, 26–32. https://doi.org/10.1016/j.procs.2013.05.005

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. *San Francisco, CA, itd: Morgan Kaufmann*. https://doi.org/10.1016/B978-0-12-381479-1.00001-0

Kontopoulos, E., Berberidis, C., Dergiades, T., & Bassiliades, N. (2013). Ontology-based sentiment analysis of twitter posts. *Expert Systems with Applications*, *40*(10), 4065–4074. https://doi.org/10.1016/j.eswa.2013.01.001

Lu, Y., Liang, M., Ye, Z., & Lichao, C. (2015). Improved particle swarm optimization algorithm and its application in text feature selection. *Applied Soft Computing*, *35*, 629–636. https://doi.org/doi.org/10.1016/j.asoc.2015.07.005

Markopoulos, G., Mikros, G., & Iliadi, A. (2015). Cultural Tourism in a Digital Era, 373–383. https://doi.org/10.1007/978-3-319-15859-4

Marrese-Taylor, E., Velásquez, J. D., Bravo-Marquez, F., & Matsuo, Y. (2013). Identifying customer preferences about tourism products using an aspect-based opinion mining approach. *Procedia Computer Science*, *22*, 182–191. https://doi.org/10.1016/j.procs.2013.09.094

Putra, C., & Irawati, E. (2015). Algoritma Support Vector Machine Untuk Mendeteksi Sms Spam Berbahasa Indonesia, 109–116.

Sukardi, A. S., & Supriyanto, C. (2014). Klasifikasi Spam Email Menggunakan Algoritma C4.5 Dengan Seleksi Fitur. *Jurnal Teknologi Informasi*, *10*(1), 19–30. Retrieved from http://research.pps.dinus.ac.id/lib/jurnal/Vol 10.1 019-030.pdf

Taufik, A. (2017). Optimasi Particle Swarm Optimization Sebagai Seleksi Fitur Pada Analisis Sentimen Review Hotel Berbahasa Indonesia Menggunakan Algoritma Naïve Bayes. *Jurnal Teknik Komputer*, *III*(2), 40–47.

Tsoumakas, G., Katakis, I., & Vlahavas, I. (2010). *Data Mining and Knowledge Discovery Handbook*. *Journal of Chemical Information and Modeling*. https://doi.org/10.1017/CBO9781107415324.004

Witten, I. H., Frank, E., & Hall, M. a. (2011). *Data Mining: Practical Machine Learning Tools and Techniques (Google eBook)*. *Complementary literature None*. https://doi.org/0120884070, 9780120884070

Ziqiong, Z., Qiang, Y., Zili, Z., & Yijun, L. (2011). Sentiment classification of Internet restaurant reviews written in Cantonese. *Expert Systems with Applications*, *38*(6), 7674–7682. https://doi.org/10.1016/j.eswa.2010.12.147

PROFIL PENULIS

Andi Taufik Lahir di Bogor 30 November 1991. Lulus Sariana 2014 dan Lulus Pasca Sarjana STMIK Nusa Mandiri tahun 2016