

Opinion Mining Pada Review Produk Kecantikan Menggunakan Algoritma Naïve Bayes

Dinda Ayu Muthia

Abstract— In recent years many sentiment analysis and opinion mining applications have been developed to analyze opinions, feelings and attitudes about products, brands, and news, etc. These applications mine opinions from different sources like online forums and news sites and from movie, product and hotel reviews. The Naïve Bayes algorithm is a popular machine learning technique for opinion mining, as it is very simple, efficient and performs well on many domains. However, Naïve Bayes has a deficiency that is very sensitive to features that are too numerous, resulting in low classification accuracy. Therefore, this research used Genetic Algorithm feature selection method to improve the accuracy of Naïve Bayes. This study produces text classification in the form of positive or negative from beauty product reviews. Measurements based on Naive Bayes accuracy before and after the addition of feature selection methods. The evaluation was performed using 10 fold cross validation. Measurement accuracy is measured with confusion matrix and ROC curve. The results showed an increase in the accuracy of Naïve Bayes from 65.50% to 83%.

Intisari— Dalam beberapa tahun terakhir banyak analisis sentimen dan aplikasi *opinion mining* telah dikembangkan untuk menganalisis pendapat, perasaan dan sikap tentang produk, merek, dan berita, dan sejenisnya. Aplikasi ini menambah pendapat dari berbagai sumber seperti forum online dan situs berita dan dari ulasan film, produk dan hotel. Algoritma *Naïve Bayes* adalah teknik *machine learning* yang populer untuk *opinion mining*, 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 *Genetic Algorithm* agar bisa meningkatkan akurasi *Naïve Bayes*. Penelitian ini menghasilkan klasifikasi teks dalam bentuk positif atau negatif dari review produk kecantikan. Pengukuran berdasarkan akurasi *Naive Bayes* sebelum dan sesudah penambahan metode pemilihan fitur. Evaluasi dilakukan menggunakan 10 fold cross validation. Pengukuran akurasi diukur dengan *confusion matrix* dan kurva ROC. Hasil penelitian menunjukkan peningkatan akurasi *Naïve Bayes* dari 65.50% menjadi 83%.

Kata Kunci— Algoritma, *Naive Bayes*, Review, *Opinion Mining*

I. PENDAHULUAN

Dalam beberapa tahun terakhir banyak analisis sentimen dan aplikasi *opinion mining* telah dikembangkan untuk menganalisis pendapat, perasaan dan sikap tentang produk, merek, dan berita, dan sejenisnya. Aplikasi ini menambah

Jurusan Manajemen Informatika AMIK BSI Bekasi, Jl. Cut Mutiah No. 88 Bekasi 17113 INDONESIA (tlp: 021-82425634; fax: 021-82425634; e-mail: dinda.dam@bsi.ac.id)

pendapat dari berbagai sumber seperti forum online dan situs berita dan dari ulasan film, produk dan hotel [1].

Penelitian-penelitian yang sudah dilakukan dalam bidang *opinion mining* atau analisis sentimen pada review atau ulasan yang tersedia di internet diantaranya, penelitian yang dilakukan oleh Basari dan kawan-kawan pada tahun 2013, yaitu *opinion mining* pada review film yang diambil dari Twitter menggunakan pengklasifikasi *Support Vector Machine* dan *Particle Swarm Optimization* di mana proses *Generate N-grams* yang digunakan adalah unigram [2]. Penelitian yang dilakukan oleh Z. Zhang dan kawan-kawan pada tahun 2011, yaitu *opinion mining* pada review restoran di internet yang ditulis dalam bahasa Canton menggunakan pengklasifikasi *Naïve Bayes* dan *Support Vector Machine*, di mana dalam penelitian tersebut dilakukan perbandingan proses *Generate N-grams* (unigram, bigram, dan trigram) [3]. Penelitian yang dilakukan oleh Moraes dan kawan-kawan pada tahun 2013, yaitu analisis sentimen pada review film dari IMDB, review produk seperti GPS, buku, dan kamera dari Amazon.com menggunakan pengklasifikasi *Support Vector Machine* dan *Artificial Neural Network* [4].

Salah satu proses yang dianalisis di beberapa penelitian dalam bidang *opinion mining* yaitu penggunaan fitur n-gram (*Generate N-grams*). Fitur ini berfungsi untuk menggabungkan kata agar bisa dilihat perbedaan sentimen dari tiap gabungan kata. Penggunaan fitur ini rupanya berpengaruh dengan akurasi yang dihasilkan pengklasifikasi seperti yang telah diteliti oleh Zhang dan kawan-kawan pada tahun 2011 [5]. Jika fitur yang digunakan 2-gram (penggabungan dua kata) atau 3-gram (penggabungan tiga kata), maka otomatis jumlah kata yang diproses akan semakin banyak dan bisa mempengaruhi akurasi klasifikasi.

Berdasarkan penelitian, pengklasifikasi *Naïve Bayes* terbukti sangat sederhana dan efisien [6]. Di samping kesederhananya, *Naïve Bayes* adalah salah satu teknik *machine learning* yang populer untuk klasifikasi teks, dan mempunyai performa yang baik pada banyak domain [7]. Akan tetapi, *Naïve Bayes* ternyata mempunyai kekurangan, yaitu sangat sensitif dalam pemilihan fitur [6]. Jumlah fitur yang terlalu banyak dalam proses klasifikasi, tidak hanya meningkatkan waktu penghitungan tetapi juga menurunkan akurasi [8].

Pada beberapa penelitian di bidang *opinion mining*, tidak jarang ditemukan perbandingan hasil dari metode pemilihan fitur. Pemilihan fitur terbukti dapat membuat pengklasifikasi menjadi lebih efisien dan efektif dengan mengurangi jumlah data yang dianalisis, maupun mengidentifikasi fitur yang sesuai untuk dipertimbangkan

dalam proses pembelajaran. Terdapat dua macam metode pemilihan fitur dalam *machine learning*, yaitu *wrapper* dan *filter*. *Wrapper* menggunakan akurasi klasifikasi dari beberapa algoritma sebagai fungsi evaluasinya. *Wrapper* mengevaluasi fitur secara berulang dan menghasilkan akurasi klasifikasi yang tinggi. Salah satu metode *wrapper* yang bisa digunakan adalah metode *Genetic Algorithm* (GA).

Tujuan dari penelitian ini adalah untuk menggabungkan metode pemilihan fitur *Genetic Algorithm* dalam menganalisis sentimen pada review produk kecantikan menggunakan algoritma *Naïve Bayes*.

II. TINJAUAN PUSTAKA

A. Opinion Mining

Opinion mining berfungsi untuk mendapatkan opini dari penulis tentang suatu entitas. Menurut Tang dalam Haddi [9], analisis sentimen terhadap review yaitu suatu proses menganalisis review produk yang ada di internet guna menentukan opini terhadap suatu produk. Menurut Thelwall dalam Haddi, analisis sentimen adalah mengklasifikasikan suatu teks ke dalam bentuk positif atau negatif [9].

Menurut Mejova dalam Basari [2], analisis sentimen bertugas untuk menentukan perilaku ataupun opini dari penulis terhadap suatu topik atau produk. Perilaku penulis dapat menentukan alasan, opini atau penilaian terhadap produk tertentu dan dapat mempengaruhi penilaian orang lain yang membaca opini tersebut [10].

B. Pemilihan Fitur

Pemilihan fitur berfungsi untuk menghilangkan fitur yang tidak sesuai yang mungkin menimbulkan kekacauan dalam klasifikasi, dengan menggunakan suatu metode. Dalam *machine learning*, terdapat dua macam metode pemilihan fitur, yakni *wrapper* dan *filter* [11].

Metode pemilihan fitur *wrapper* menggunakan akurasi klasifikasi dari beberapa algoritma sebagai fungsi evaluasinya [6]. *Genetic Algorithm* (GA) merupakan metode *wrapper* yang bisa digunakan dalam pemilihan fitur [12].

1) Genetic Algorithm

Genetic algorithm berupaya untuk menyatukan ide-ide evolusi alam [13]. Umumnya, pembelajaran *genetic* dimulai sebagai berikut:

- Sebuah populasi awal dibuat dan terdiri dari aturan acak. Setiap aturan dapat diwakili oleh string bit. Sebagai contoh, sampel dalam satu set pelatihan yang diberikan dijelaskan oleh dua atribut Boolean, A1 dan A2, dan bahwa ada dua kelas, C1 dan C2. Aturan "If A1 And Not A2 Then C2" bisa dikodekan sebagai string bit "100," di mana dua bit paling kiri mewakili atribut A1 dan A2, masing-masing, dan bit paling kanan mewakili kelas. Demikian pula, aturan "If Not A1 And Not A2 Then C1" dapat dikodekan sebagai "001." Jika atribut memiliki nilai-nilai k, di mana $k > 2$, maka k bit dapat digunakan untuk

mengkodekan nilai-nilai atribut tersebut. Kelas dapat dikodekan dengan cara yang sama.

- Berdasarkan gagasan ketahanan dari yang paling sesuai, populasi baru terbentuk, yang terdiri dari aturan yang paling sesuai dalam populasi saat ini, serta keturunan aturan ini. Biasanya, fitness aturan dinilai dengan akurasi klasifikasi pada satu set sampel pelatihan.
- Keturunan dihasilkan dengan menerapkan operator genetika seperti *crossover* dan mutasi. Dalam *crossover*, substring dari sepasang aturan ditukar untuk membentuk pasangan aturan baru. Dalam mutasi, bit yang dipilih secara acak dalam aturan string dibalik.
- Proses menghasilkan populasi baru berdasarkan aturan populasi sebelumnya berlanjut sampai populasi, P, berkembang di mana setiap aturan dalam P memenuhi ambang batas *fitness* yang sudah ditentukan.

Genetic Algorithm kini telah digunakan untuk klasifikasi seperti masalah optimasi lainnya. Dalam *data mining*, algoritma genetika bisa digunakan untuk mengevaluasi fitness algoritma lainnya.

C. Naïve Bayes

Seperti yang sudah dijelaskan sebelumnya, *Naïve Bayes* merupakan algoritma yang sederhana. Tahapan-tahapan algoritma *Naïve Bayes* terdiri dari [14]:

- 1) Hitung probabilitas bersyarat/*likelihood*:

$$P(x | C) = P(x_1, x_2, \dots, x_n | C) \quad (1)$$

$C = \text{class}$

$x = \text{vektor dari nilai atribut } n$

$P(x_i | C) = \text{proporsi dokumen dari class } C \text{ yang mengandung nilai atribut } x_i$

- 2) Hitung probabilitas *prior* untuk tiap class:

$$P(C) = \frac{N_j}{N} \quad (2)$$

$N_j = \text{jumlah dokumen pada suatu class}$

$N = \text{jumlah total dokumen}$

- 3) Hitung probabilitas *posterior* dengan rumus:

$$P(C|x) = \frac{P(x|C) P(C)}{P(x)} \quad (3)$$

Dengan kata lain, rumus Bayes bisa diberikan sebagai berikut:

$$\text{Posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}} \quad (4)$$

III. HASIL DAN PEMBAHASAN

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

a. Pengumpulan Data

Penulis mengumpulkan data review produk kecantikan Wardah BB Cream yang diambil dari situs khusus

review produk kecantikan di Indonesia, yaitu www.femaledaily.com. Review yang dikumpulkan terdiri dari 100 review positif dan 100 review negatif.

b. Pengolahan Awal Data

Dalam tahap *preprocessing*, terdiri dari dua proses, yaitu:

1) *Tokenization*

Mengumpulkan semua kata yang ada dalam review dan menghilangkan semua tanda baca maupun simbol apapun yang bukan huruf.

2) *Generate N-grams*

Menggabungkan kata sifat yang sering muncul untuk menunjukkan sentimen, seperti kata “sangat” dan kata “bagus”. Kata “bagus” memang sudah menunjukkan sentimen bentuk opini positif. Kata “sangat” tidak akan berarti jika berdiri sendiri. Namun jika dua kata tersebut digabung menjadi “sangat bagus”, maka akan sangat menguatkan opini positif tersebut. Penulis hanya menggunakan penggabungan dua kata, yang disebut 2-grams (bigrams).

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.

c. Metode Yang Diusulkan

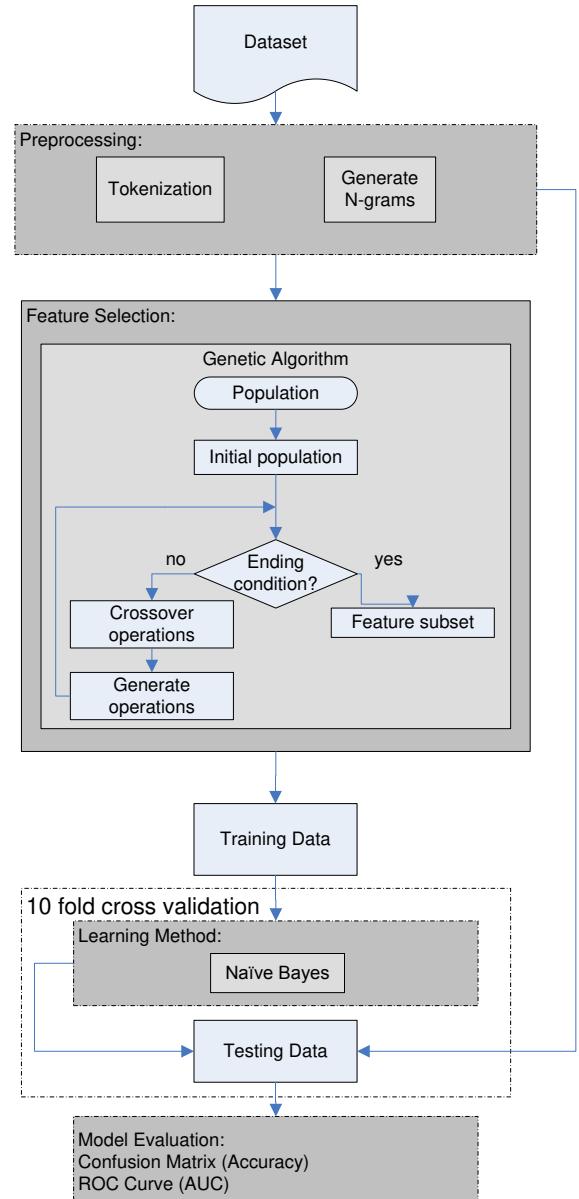
Metode yang diusulkan yakni dengan menerapkan *Genetic Algorithm* sebagai metode pemilihan fitur agar akurasi algoritma *Naïve Bayes* bisa meningkat. *Genetic Algorithm* yang penulis terapkan adalah menggunakan *Naïve Bayes* yang diuji di dalam tahap *wrapper*. Lihat pada Gbr 1 untuk model yang diusulkan secara lebih detail.

d. Eksperimen dan Pengujian Metode

Untuk eksperimen data penelitian, penulis menggunakan Rapid Miner Studio untuk mengolah data.

e. Evaluasi dan Validasi Hasil

Validasi dilakukan menggunakan 10 *fold cross validation*. Sedangkan pengukuran akurasi diukur dengan *confusion matrix* dan kurva ROC untuk mengukur nilai AUC. Dengan *confusion matrix*, akurasi *Naïve Bayes* sebelum dan sesudah menggunakan metode pemilihan fitur.



Sumber : Hasil Penelitian

Gbr 1. Model yang Diusulkan

A. Hasil

1) Klasifikasi Teks Menggunakan Algoritma *Naïve Bayes*

Data training yang digunakan dalam pengklasifikasian teks ini terdiri dari 100 review positif dan 100 review negatif dari suatu produk kecantikan. Sebelum masuk ke proses pengklasifikasian, data tersebut harus melalui beberapa tahapan proses, diantaranya:

a) Pengumpulan Data

Data review positif disatukan dalam folder dengan nama POS. Data review negatif disimpan dalam folder dengan nama NEG. Tiap dokumen berekstensi .txt yang dapat dibuka menggunakan aplikasi Notepad.

b) Pengolahan Awal Data

Proses yang dilalui terdiri dari *tokenization* dan *generate N-grams*. Hasil pengolahan awal data dapat dilihat pada tabel 1 dan 2.

TABEL 1
PROSES TOKENIZATION

Review	Tokenization
suka sama produk ini walaupun harganya murah tp kualitasnya ga jelek2 amat. hehe shadanya juga pas buat kulit fair kaya aku teksturnya juga ringan dikulit, coveragenya medium lumayan mengcover bekas2 jerawat overall i love it ??	suka sama produk ini walaupun harganya murah tp kualitasnya ga jelek amat hehe shadanya juga pas buat kulit fair kaya aku teksturnya juga ringan dikulit coveragenya medium lumayan mengcover bekas jerawat overall i love it

Sumber : Hasil Penelitian

TABEL 2
HASIL PENGOLAHAN 2-GRAM

Review setelah Tokenization	Generate 2-grams
suka sama produk ini walaupun harganya murah tp kualitasnya ga jelek amat hehe shadanya juga pas buat kulit fair kaya aku teksturnya juga ringan dikulit coveragenya medium lumayan mengcover bekas jerawat overall i love it	suka suka_sama sama sama_produk produk produk_ini ini_ini_walaupun walaupun_walaupun_harganya harganya_harganya_murah murah_murah_tp_tp_tp_kualitasnya kualitasnya_kualitasnya_ga ga_ga_jelek jelek_jelek_amat_amat_amat_hehe_hehe_hehe_shadanya shadanya_shadanya_juga juga_juga_pas pas_pas_buat buat_buat_kulit kulit_kulit_fair fair_fair_kaya kaya_kaya_aku_aku_aku_teksturnya teksturnya_juga juga_juga_ringan ringan_dikulit dikulit_dikulit_coveragenya coveragenya coveragenya_medium medium_medium_lumayan lumayan_lumayan_mengcover mengcover_mengcover_bekas bekas_bekas_jerawat jerawat_jerawat_overall_overall_overall_i i_love_love_love_love_it_it

Sumber : Hasil Penelitian

c) Klasifikasi

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

negatif. Penulis mendapatkan 6 kata dan 2 gabungan kata yang berhubungan dengan sentimen yang paling sering muncul, yaitu “Cocok”, “Bagus”, “Suka”, “Oke”, “Gak” yang berarti “Tidak”, “Kurang”, “Aku Suka” dan “Ga Cocok” yang berarti “Tidak Cocok”.

2) Pengujian Model dengan 10 Fold Cross Validation

Pada penelitian ini, pengujian model dilakukan dengan menggunakan teknik 10 *cross validation*. 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. Kemudian proses akurasi dihitung dengan melihat seberapa banyak data yang sudah terkласifikasi dengan benar.

3) Optimasi Model dengan Menambahkan Metode Pemilihan Fitur

Penulis menambahkan metode pemilihan fitur *Genetic Algorithm* dari *wrapper*, bertujuan untuk meningkatkan akurasi pengklasifikasi *Naïve Bayes*. Penambahan metode ini tentu akan menambah lama durasi proses klasifikasi, namun hasil dari Algoritma *Naïve Bayes* dan *Genetic Algorithm* terbukti meningkatkan tingkat akurasi klasifikasi.

4) Eksperimen Terhadap Indikator Model

Untuk mendapatkan model yang terbaik, beberapa indikator disesuaikan nilainya agar didapatkan hasil akurasi yang tinggi. Dalam penyesuaian indikator pada *Genetic Algorithm*, akurasi paling tinggi diperoleh dengan kombinasi *population size*=30, *maximum number of generations*=50, *p initialize*=0.5, *p crossover*=0.5, dan *p generate*=0.1. Hasil akurasi mencapai 83%. Jika indikator lainnya turut diubah nilainya, dapat menyebabkan peningkatan durasi dalam mengolah data.

B. Pembahasan

Dengan memiliki model klasifikasi teks pada review, pembaca dapat dengan mudah mengidentifikasi mana review yang positif maupun yang negatif. Dari data review yang sudah ada, dipisahkan menjadi kata-kata, lalu diberikan bobot pada masing-masing kata tersebut. Dapat dilihat kata mana saja yang berhubungan dengan sentimen yang sering muncul dan mempunyai bobot paling tinggi. Dengan demikian dapat diketahui review tersebut positif atau negatif. Dalam penelitian ini, hasil pengujian model akan dibahas melalui *confusion matrix* untuk menunjukkan seberapa baik model yang terbentuk.

1) Pengukuran dengan *Confusion Matrix*

Pengukuran dengan *confusion matrix* akan menampilkan perbandingan dari hasil akurasi model *Naïve Bayes* sebelum ditambahkan metode pemilihan fitur yang bisa dilihat pada tabel 3 dan setelah ditambahkan metode pemilihan fitur *Genetic algorithm* yang bisa dilihat pada tabel 4.

TABEL 3
CONFUSION MATRIX MODEL NAÏVE BAYES SEBELUM
PENAMBAHAN METODE PEMILIHAN FITUR

Akurasi Naïve Bayes: 65.50% +/- 10.11% (mikro:
65.50%)

	<i>True Positive</i>	<i>True Negative</i>	<i>Class precision</i>
<i>Pred. Positive</i>	77	46	62.60%
<i>Pred. Negative</i>	23	54	70.13%
<i>Class recall</i>	77.00%	54.00%	

Sumber : Hasil Penelitian

Tanpa menggunakan metode pemilihan fitur, algoritma *Naïve Bayes* dengan fitur 2-gram sudah menghasilkan akurasi sebesar 65.50%. Akurasi tersebut masih berada pada level *poor classification*, masih kurang akurat, sehingga perlu ditingkatkan lagi menggunakan metode pemilihan fitur. Untuk pengujian data selanjutnya, akan dilanjutkan dari pengujian data menggunakan fitur 2-gram. Setelah menggunakan metode pemilihan fitur *wrapper*, yaitu *Genetic Algorithm*, akurasi algoritma *Naïve Bayes* meningkat menjadi 83% dan masuk ke level *good classification*, seperti yang bisa dilihat pada tabel 4.

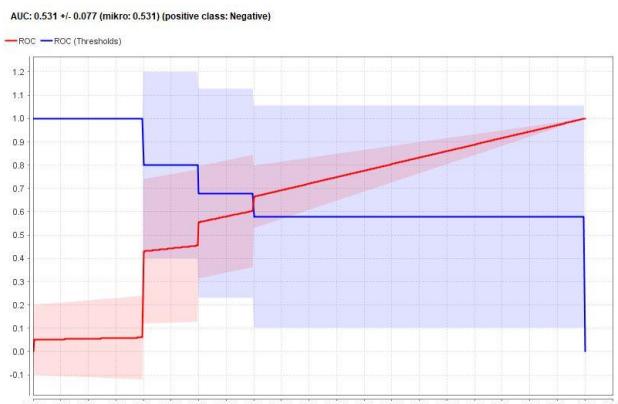
TABEL 4
CONFUSION MATRIX MODEL NAÏVE BAYES SETELAH
PENAMBAHAN METODE PEMILIHAN FITUR GENETIC
ALGORITHM

Akurasi Naïve Bayes: 83.00% +/- 9.27% (mikro:
83.00%)

	<i>True Positive</i>	<i>True Negative</i>	<i>Class precision</i>
<i>Pred. Positive</i>	80	14	85.11%
<i>Pred. Negative</i>	20	86	81.13%
<i>Class recall</i>	80.00%	86.00%	

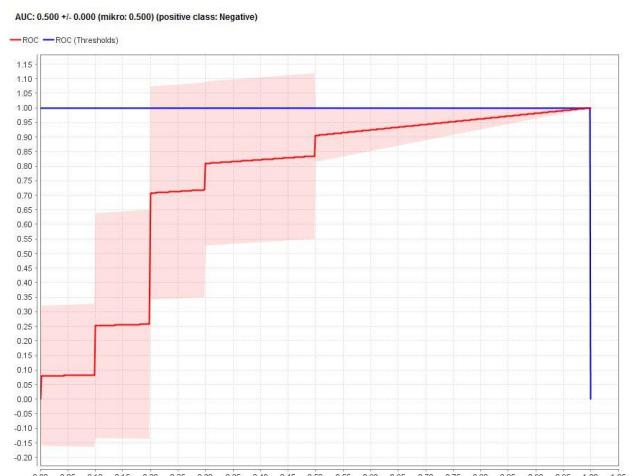
Sumber : Hasil Penelitian

Perbandingan kurva ROC pada penelitian ini bisa dilihat pada gambar 2 dan 3 di bawah ini.



Sumber : Hasil Penelitian

Gbr 2. Kurva ROC Model Naïve Bayes Sebelum Penambahan Metode Pemilihan Fitur



Sumber : Hasil Penelitian

Gbr 3. Kurva ROC Model Naïve Bayes Setelah Penambahan Metode Pemilihan Fitur

IV. KESIMPULAN

Penerapan metode pemilihan fitur *Genetic Algorithm* terbukti bisa meningkatkan akurasi algoritma *Naïve Bayes*. Data review produk kecantikan dapat diklasifikasi dengan baik ke dalam bentuk positif dan negatif. Akurasi *Naïve Bayes* saat sebelum menggunakan metode pemilihan fitur hanya mencapai 65.50% Sedangkan setelah menggunakan metode pemilihan fitur *Genetic Algorithm*, akurasinya meningkat hingga mencapai 83%. Peningkatan akurasi mencapai 17.50%.

Model yang dihasilkan bisa diterapkan pada seluruh data review produk kecantikan dengan teks bahasa Indonesia. Model ini bisa membantu seseorang untuk menghemat waktu saat mencari review suatu produk kecantikan yang kini banyak tersedia online.

Saran dari peneliti untuk pengembangan penelitian selanjutnya adalah dengan menambahkan atau membandingkan metode pemilihan fitur, baik itu dari jenis *filter* maupun *wrapper*. Selain itu, bisa juga menggunakan data dari domain yang berbeda seperti review produk elektronik, film, buku, restoran dan lain-lain yang berasal dari sumber lainnya.

UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada Tim JSI yang telah membantu dalam rangka menerbitkan jurnal ini. Tidak lupa kepada institusi STMIK Antar Bangsa yang telah memberikan waktu untuk menyelesaikan penelitian ini.

REFERENSI

- [1] I. Maks and P. Vossen, "A lexicon model for deep sentiment analysis and opinion mining applications," *Decis. Support Syst.*, vol. 53, no. 4, pp. 680–688, Nov. 2012.
- [2] A. Samad, H. Basari, B. Hussin, I. G. Pramudya, and J. Zeniarja, "Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization,"

- [3] *Procedia Eng.*, vol. 53, pp. 453–462, 2013.
- [4] Z. Zhang, Q. Ye, Z. Zhang, and Y. Li, “Sentiment classification of Internet restaurant reviews written in Cantonese,” *Expert Syst. Appl.*, vol. 38, no. 6, pp. 7674–7682, Jun. 2011.
- [5] R. Moraes, J. F. Valiati, and W. P. Gavião Neto, “Document-level sentiment classification: An empirical comparison between SVM and ANN,” *Expert Syst. Appl.*, vol. 40, no. 2, pp. 621–633, Feb. 2013.
- [6] W. Zhang and F. Gao, “An Improvement to Naive Bayes for Text Classification,” *Procedia Eng.*, vol. 15, pp. 2160–2164, Jan. 2011.
- [7] J. Chen, H. Huang, S. Tian, and Y. Qu, “Feature selection for text classification with Naïve Bayes,” *Expert Syst. Appl.*, vol. 36, no. 3, pp. 5432–5435, Apr. 2009.
- [8] Q. Ye, Z. Zhang, and R. Law, “Expert Systems with Applications Sentiment classification of online reviews to travel destinations by supervised machine learning approaches,” *Expert Syst. Appl.*, vol. 36, no. 3, pp. 6527–6535, 2009.
- [9] A. K. Uysal and S. Gunal, “A novel probabilistic feature selection method for text classification,” *Knowledge-Based Syst.*, vol. 36, pp. 226–235, Dec. 2012.
- [10] E. Haddi, X. Liu, and Y. Shi, “The Role of Text Pre-processing in Sentiment Analysis,” *Procedia Comput. Sci.*, vol. 17, pp. 26–32, Jan. 2013.
- [11] R. Feldman, “Techniques and applications for sentiment analysis,” *Commun. ACM*, vol. 56, no. 4, p. 82, Apr. 2013.
- [12] F. Gorunescu, *Data Mining Concept Model Technique*. 2011.
- [13] S. Gunal, “Hybrid feature selection for text classification .,” vol. 20, 2012.
- [14] J. Han and M. Kamber, *Data Mining Concepts and Techniques*. 2007.
- [15] Z. Markov and T. Daniel, *Uncovering Patterns in*. 2007.



Dinda Ayu Muthia, M.Kom. Lahir di Jakarta, 8 Juli 1988. Tahun 2011 lulus dari Program Strata Satu (S1) Program Studi Sistem Informasi STMIK Nusa Mandiri Jakarta. Tahun 2013 lulus dari Program Strata Dua (S2) Program Studi Magister Ilmu Komputer STMIK Nusa Mandiri Jakarta. Tahun 2015 sudah memiliki Jabatan Fungsional Akademik Asisten Ahli di AMIK BSI Bekasi. Tahun 2016 sudah menjadi dosen tersertifikasi. Jurnal yang pernah diterbitkan di antaranya, Analisis Sentimen Pada Review Buku Menggunakan

Algoritma *Naïve Bayes* pada tahun 2014 di Jurnal Paradigma dan Opinion Mining Pada Review Buku Menggunakan Algoritma *Naïve Bayes* di Jurnal Teknik Komputer pada tahun 2016. Proceeding yang pernah diseminarkan di ISSIT 2014 dengan judul *Sentiment Analysis of Hotel Review Using Naïve Bayes Algorithm and Integration of Information Gain and Genetic Algorithm as Feature Selections Methods* dan pada Konferensi Nasional Ilmu Sosial dan Teknologi (KNIST) 2017 dengan judul Penerapan Algoritma Naïve Bayes Pada Analisis Sentimen Review Hotel.