

PENERAPAN MODEL *SUPPORT VECTOR MACHINE* TEXT MINING PADA KOMENTAR REVIEW SMARTPHONE ANDROID VS BLACKBERRY DENGAN TEKNIK OPTIMASI GENETIC ALGORITHM

AGUS DARMAWAN

agus.darmawan@ymail.com,
Program Studi Teknik Informatika
Universitas Indraprasta PGRI

Abstrak. Pasar *smartphone* saat ini semakin banyak, penjualannya tidak hanya secara konvensional tetapi sudah merambah di *online shop*. Namun tidak semua *smartphone* memiliki kualitas yang baik untuk menunjang kebutuhan konsumen dan hal ini yang harus diperhatikan oleh para konsumen. Sebelum konsumen memutuskan untuk membeli *smartphone android* atau *belackberry* sebaiknya konsumen mengetahui dengan detail spesifikasi dan fungsi dari *smartphone* tersebut, hal ini dapat dipelajari dari testimoni dan opini atau hasil komentar *review* dari pengguna perbedaan *smartphone android* atau *belackberry*. Membaca komentar *review* tersebut secara keseluruhan dapat memakan waktu, namun jika hanya sedikit komentar *review* yang dibaca evaluasi akan menjadi bias. Dari beberapa teknik tersebut yang paling sering digunakan untuk klasifikasi data adalah *Support Vector Machines* (SVM). SVM memiliki kelebihan yaitu mampu mengidentifikasi dengan *hyperplane* terpisah yang memaksimalkan margin antara dua kelas yang berbeda. Pemilihan fitur sekaligus penyetingan parameter di SVM secara signifikan mempengaruhi hasil akurasi klasifikasi. Oleh karena itu, dalam penelitian ini digunakan penggabungan metode seleksi fitur, yaitu *Genetic Aloritm* (GA) agar bisa meningkatkan akurasi pengklasifikasi *Support Vector Machines*. Penelitian ini menghasilkan klasifikasi teks dalam bentuk positif atau negatif dari review produk *Smartphone*. Pengukuran berdasarkan akurasi *Support Vector Machines* sebelum dan sesudah penambahan metode seleksi fitur. Sedangkan pengukuran akurasi diukur dengan *confusion matrix* dan kurva ROC. Hasil penelitian menunjukkan peningkatan akurasi *Support Vector Machines* dari 71.00 % dengan penambahan *Genetic Aloritm* menjadi 78.02%.

Kata Kunci: Komentar, Review, SVM, *Genetic Aloritm* (GA), *Confusion Matrix*, Kurva ROC

Abstract. The smartphone market is now more and more, not only in the conventional penjualannya but have penetrated in the online shop. But not all smartphones have good quality to support the needs of consumers and it is to be noticed by the consumer. Before consumers decide to buy android smartphone or belackberry consumer should know the details of the specifications and functions of the smartphone, it can be learned from the testimony and opinion or the results of a user review comments or belackberry difference android smartphone. Reading the comments of the review as a whole can be time consuming, but if only a few comments are read reviews on evaluations will be biased. Of some of the techniques most often used for data classification is Support Vector Machines (SVM). SVM has the advantage of being able to identify the separate hyperplane that maximizes the margin between two different classes. Selection of features at once setup parameters in SVM significantly influence the results of classification accuracy. Therefore, in this study used the incorporation of feature selection methods, namely Genetic Encryption (GA) in order to improve the accuracy of the classifier Support

Vector Machines. This research resulted in the classification of text in a positive or negative form of product reviews Smartphone. Measurement is based on Support Vector Machines accuracy before and after the addition of feature selection methods. While the measurement accuracy is measured by the confusion matrix and ROC curves. The results showed an increase in Support Vector Machines accuracy of 71.00% with the addition of Genetic Encryption be 78.02%

Keywords : Comments , Review, SVM , Genetic Encryption (GA) , Confusion Matrix , ROC curve

PENDAHULUAN

Perkembangan teknologi informasi saat ini sangat pesat salah satu adalah ponsel, ponsel yang pintar sering di sebut dengan *smartphone* dengan sifatnya yang canggih dan mudah digunakan. Menjamurnya penggunaan *smartphone* di beberapa tahun belakangan ini dikarenakan beberapa faktor penunjang dalam pekerjaan, sifat dari *smartphone* yang *multi tasking*. Hal ini yang menjadikan *smartphone* banyak diminati oleh siapapun dan oleh kalangan manapun. Pasar *smartphone* saat ini pun semakin banyak, penjualannya tidak hanya secara konvensional tetapi sudah merambah di *online shop*. Namun tidak semua *smartphone* memiliki kualitas yang baik untuk menunjang kebutuhan konsumen. Sebelum konsumen memutuskan untuk membeli *smartphone* Android atau Blackberry sebaiknya konsumen membaca hasil komentar *review* perbedaan android dan blackberry dari pengguna *smartphone*. Membaca komentar *review* tersebut secara keseluruhan dapat memakan waktu, namun jika hanya sedikit komentar *review* yang dibaca evaluasi akan menjadi bias. Analisa sentimen atau *opinion mining* adalah studi komputasi mengenai pendapat, perilaku dan emosi seseorang terhadap entitas. Entitas tersebut dapat menggambarkan individu, kejadian atau topik. Topik tersebut kemungkinan besar dapat berupa *review* [Medhat et al., 2014]. Teknik klasifikasi yang biasa digunakan untuk analisis sentimen *review* diantaranya *Naïve Bayes*, *Support Vector Machines* (SVM) dan *K-Nearest Neighbor* (KNN) [Dehkharghani et al., 2014]. Dari beberapa teknik tersebut yang paling sering digunakan untuk klasifikasi data adalah *Support Vector Machines* (SVM). Seleksi fitur sekaligus penyetingan parameter di SVM secara signifikan mempengaruhi hasil akurasi klasifikasi [Zhao et al., 2011]. Kecenderungan dalam beberapa tahun terakhir adalah untuk secara bersamaan mengoptimalkan bagian fitur untuk *Support Vector Machine* (SVM), sehingga dapat meningkatkan akurasi klasifikasi pada *Support Vector Machine* (SVM). metode seleksi fitur *Algoritma Genetika* (GA) memiliki potensi untuk menghasilkan fitur yang lebih baik dan menjadi parameter yang optimal pada waktu yang sama. [Zhao et al., 2011]. Pada penelitian ini algoritma *Support Vector Machines* dan Optimasi *Genetic Algorithm* (GA) metode seleksi fitur akan diterapkan oleh penulis untuk mengklasifikasikan teks pada komentar *review* produk *smartphone* android atau blackberry dalam rangka meningkatkan akurasi analisa sentimen.

TINJAUAN PUSTAKA

Review Produk Smartphone

Situs *web* adalah wadah bagi beragam pendapat. Salah satu bentuk opini yang memiliki kredibilitas adalah *review* produk. Situs *web* seperti amazon.com mendorong pengguna untuk memberikan ulasan (*review*) [Weiss et al., 2010]. Hasil penambangan teks dari *review* tersebut dapat diklasifikasikan dengan tiga kategori, yaitu positif, negatif dan netral. Menurut Zhang dan Liu dalam [Khan et al., 2014] analisis mendalam dari setiap aspek produk berdasarkan pendapat konsumen adalah sama pentingnya bagi masyarakat, para

pedagang dan produsen. Selain di amazon.com, epinions.com maupun gsmarena.com merupakan jenis situs lain yang berisikan *review* produk.

Analisa Sentimen (*Sentiment Analysis*)

Menurut Tang dalam Haddi [Haddi et al., 2013], analisa sentimen pada *review* adalah proses menyelidiki *review* produk di internet untuk menentukan opini atau perasaan terhadap suatu produk secara keseluruhan. Menurut Thelwall dalam Haddi [Haddi et al., 2013], analisa sentimen diperlakukan sebagai suatu tugas klasifikasi yang mengklasifikasikan orientasi suatu teks ke dalam positif atau negatif. Menurut Mejova dalam Basari [Basari et al., 2013], tujuan dari analisa sentimen adalah untuk menentukan perilaku atau opini dari seorang penulis dengan memperhatikan suatu topik tertentu. Perilaku bisa mengindikasikan alasan, opini atau penilaian, kondisi kecenderungan (bagaimana si penulis ingin mempengaruhi pembaca).

Pemilihan Fitur (*Feature Selection*)

Menurut [Vercellis, 2009] tujuan seleksi fitur adalah untuk pengurangan fitur, untuk menghilangkan dari dataset subset dari variabel yang dianggap tidak relevan untuk tujuan dari kegiatan data mining dan fitur metode seleksi dapat diklasifikasikan ke dalam tiga kategori utama yaitu:

1. Metode *filter*

Metode Filter adalah memilih atribut yang relevan sebelum pindah ke tahap pembelajaran berikutnya, atribut yang dianggap paling penting yang dipilih untuk pembelajar, sedangkan sisanya dikecualikan.

2. Metode *wrapper*

Metode wrapper menilai sekelompok variabel dengan menggunakan klasifikasi yang sama atau algoritma regresi digunakan untuk memprediksi nilai dari variabel target.

3. Metode *embedded*

Untuk metode *embedded*, proses seleksi atribut terletak di dalam algoritma pembelajaran, sehingga pemilihan set optimal atribut secara langsung dibuat selama fase generasi model.

Konsep Genetic Algorithm (GA)

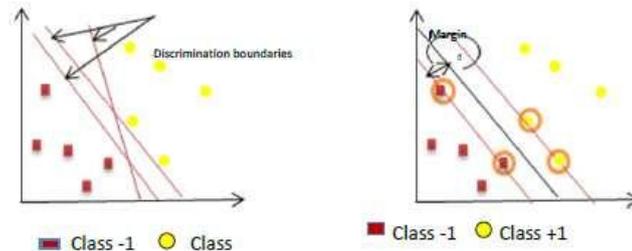
Menurut Zukhri [Zukhri, 2014] Optimasi adalah proses menyelesaikan suatu masalah tertentu supaya berada pada kondisi yang paling menguntungkan dari suatu sudut pandang. Masalah yang diselesaikan berkaitan erat dengan data-data yang dapat dinyatakan dalam satu atau beberapa variable.

Menurut David dalam Wang [Wang et al., 2013] Algoritma genetik (GA) telah berhasil disajikan untuk pilihan model SVM baik untuk pemilihan parameter dan seleksi fitur. Dengan GA, SVM telah mencapai kinerja yang lebih baik dari segi waktu komputasi dan generalisasi dari grid dan beberapa teknik lainnya. Dan menurut Huang dalam Wang [Wang et al., 2013] telah terbukti bahwa GA lebih cocok untuk kasus dengan parameter lainnya karena berkurangnya penggunaan memori dan waktu komputasi yang cepat.

Konsep Support vector machine

Konsep SVM dapat dijelaskan secara sederhana sebagai usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah class pada input space. Untuk dimensional space, input data x ($i=1, \dots, k$), dimana milik kelas 1 atau kelas 2 dan label yang terkait menjadi -1 untuk kelas 1 dan +1 untuk kelas 2. Gambar di bawah ini memperlihatkan beberapa pattern yang merupakan anggota dari dua buah class: positif (dinotasikan dengan +1) dan negatif (dinotasikan dengan -1). *Pattern* yang tergabung

pada class negatif disimbolkan dengan kotak, sedangkan *pattern* pada *class* positif, disimbolkan dengan lingkaran. Jika data input dapat dipisahkan secara linear, pemisahan *hyperplane* dapat diberikan dalam:



Gambar 1. SVM berusaha menemukan *Hyperplane* terbaik yang memisahkan kedua class negatif dan positif [Nugroho, 2008]

Tujuan dari SVM adalah untuk memisahkan data kelas dengan cara maksimal margin *hyperplane*. Dengan demikian, SVM menjamin untuk memaksimalkan jarak antara data yang paling dekat dengan *hyperplane*. Jika input data dapat dipisahkan secara linear, pemisahan *hyperplane* dapat diberikan dalam persamaan:

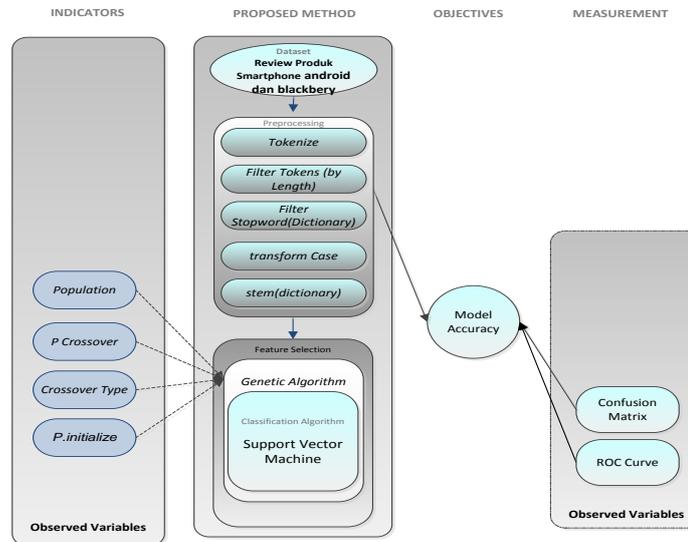
$$f(X) = w^T x + b \quad (2.1)$$

dimana w adalah n- mensi bobot vektor dan b adalah pengali skalar atau nilai bias. Persamaan ini menemukan maksimum margin untuk memisahkan kelas dari kelas positif dari kelas negatif. Fungsi keputusan ditunjukkan dalam persamaan. Contoh untuk data linear terpisah ditunjukkan pada Gambar 1:

$$y_i(w \cdot x_i + b) \geq 1 \quad i=1. \dots k \quad (2.2)$$

Kerangka Pemikiran

Penelitian ini dimulai dari adanya masalah dalam klasifikasi teks pada komentar *review* produk menggunakan pengklasifikasi Support Vector Machine (SVM), dimana pengklasifikasian tersebut memiliki kekurangan terhadap masalah pemilihan parameter yang sesuai, karena dengan tidak sesuainya sebuah pengaturan parameter dapat menyebabkan hasil klasifikasi menjadi rendah. Sumber data yang digunakan dalam penelitian ini yaitu mengambil data komentar *review* produk perbedaan *smartphone android* dan *blackberry* pada situs penyedia *review* produk yang didapat dari <http://www.kaskus.co.id/thread/50b74b657b1243c665000077/> perbedaan-android-dan-blackberry yang terdiri dari 50 komentar *review* positif dan 50 komentar *review* negatif. *Preprocessing* yang dilakukan dengan *Tokenization (Tokenize)*, *filtering (Filter Tokens (by Length))*, *filtering (Filter Stopword(Dictionary))*, *Stemming (stem(dictionary))*, *Transformation (transform Case)*. Metode pembobotan Fitur yang akan digunakan adalah *Term Frequency Invers Document Frequency (TF-IDF)* dan pemilihan fitur dengan metode *Genetic Algorithm*. Sedangkan pengklasifikasi yang digunakan adalah *Support Vector Machine (SVM)*. Pengujian 10 fold cross validation akan dilakukan, akurasi algoritma akan diukur menggunakan *confusion matrix* dan kurva ROC. RapidMiner Versi 5.3 digunakan sebagai alat bantu dalam mengukur akurasi data eksperimen yang dilakukan dalam penelitian. Gambar 2 menggambarkan kerangka pemikiran yang diusulkan dalam penelitian ini.



Gambar 2. Kerangka Pemikiran

METODE

Perancangan Penelitian

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

1. Pengumpulan Data

Data untuk eksperimen ini dikumpulkan, lalu diseleksi dari data yang tidak sesuai.

2. Pengolahan Awal Data

Model dipilih berdasarkan kesesuaian data dengan metode yang paling baik dari beberapa metode pengklasifikasian teks yang sudah digunakan oleh beberapa peneliti sebelumnya. Model yang digunakan adalah algoritma *Support Vector Machine* (SVM).

3. Metode Yang Diusulkan

Untuk meningkatkan akurasi dari Algoritma *Support Vector Machine* (SVM), maka dilakukan penambahan dengan menggunakan metode optimasi *Genetic Algorithm*.

3. Eksperimen dan Pengujian Metode

Untuk eksperimen data penelitian, penulis menggunakan RapidMiner untuk mengolah data. Sedangkan untuk pengujian metode, penulis membuat aplikasi menggunakan bahasa pemrograman PHP dengan perangkat lunak Adobe Dreamweaver sebagai editor.

4. Evaluasi dan Validasi Hasil

Evaluasi dilakukan untuk mengetahui akurasi dari model algoritma *Support Vector Machine* (SVM). Validasi digunakan untuk melihat perbandingan hasil akurasi dari model yang digunakan dengan hasil yang telah ada sebelumnya. Teknik validasi yang digunakan adalah *cross validation*.

Analisis Data

Pengumpulan Data

Penulis menggunakan data komentar *review* produk smartphone yang dikumpulkan dari situs <http://www.kaskus.co.id/thread>. Data terdiri dari 50 komentar *review* positif dan 50 komentar *review* negatif. Penulis mengunduh data tersebut dari

<http://www.kaskus.co.id/thread/50b74b657b1243c665000077/> perbedaan-android-dan-blackberry. Contoh komentar review positif sebagai berikut:

“Kalau menurut saya:

- bb lebih ke segi komunikasi

- android lebih ke segi hiburan dan entertainment“

Sedangkan contoh komentar review negatif sebagai berikut:

“setuju gan!

Tp android buat komunikasi oke juga kok.

Kapok ane pake blackbery”

Pengolahan Awal Data

Untuk mengurangi lamanya waktu pengolahan data, penulis hanya menggunakan 50 komentar *review* positif dan komentar 50 komentar *review* negatif sebagai data training. Dataset ini dalam tahap *preprocessing* harus melalui 5 (tiga) proses, yaitu:

1. *Tokenization*

Dalam proses *tokenization* ini, semua kata yang ada di dalam tiap dokumen dikumpulkan dan dihilangkan tanda bacanya, serta dihilangkan jika terdapat simbol atau apapun yang bukan huruf.

2. Filter Tokens (by Length)

Dalam proses ini, kata-kata yang memiliki pajang kurang dari 4 dan lebih dari 25 akan dihapus, seperti kata *yg,ane, gan* yang merupakan kata-kata yang tidak mempunyai makna tersendiri jika dipisahkan dengan kata yang lain dan tidak terkait dengan kata sifat yang berhubungan dengan sentimen.

3. Stopwords Removal

Dalam proses ini, kata-kata yang tidak relevan akan dihapus, seperti kata *tetapi, untuk, dengan,* yang merupakan kata-kata yang tidak mempunyai makna tersendiri jika dipisahkan dengan kata yang lain dan tidak terkait dengan kata sifat yang berhubungan dengan sentimen

4. Transform Case

Dalam proses ini, kata-kata yang tidak relevan akan diubah, seperti kata yang mengandung huruf besar yang diubah menjadi huruf kecil sehingga dapat saling berhubungan dengan sentimen.

5. Stemming

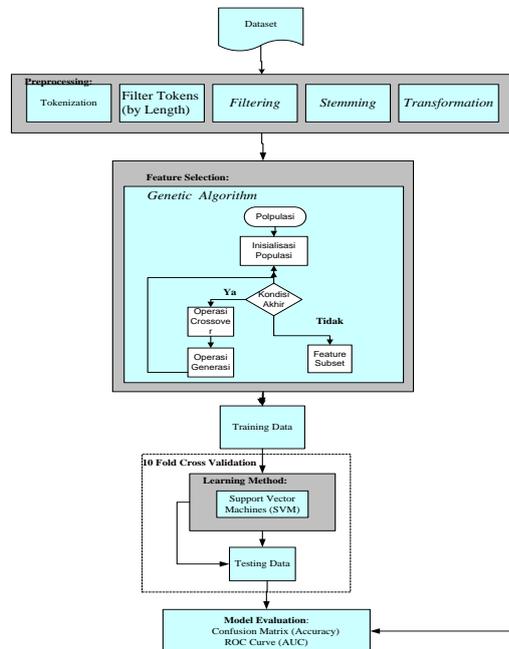
Dalam proses ini kata-kata akan dikelompokkan ke dalam beberapa kelompok yang memiliki kata dasar yang sama, seperti *hibur, hiburan, menghibur,* di mana kata dasar dari semuanya adalah kata *hibur*.

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.

Metode Yang Diusulkan

Metode yang penulis usulkan adalah penggunaan 1 (satu) jenis metode pemilihan fitur, yaitu *Genetic Algorithm (GA)* yang digunakan sebagai metode pemilihan fitur agar akurasi pengklasifikasi Support Vector Machines (SVM) bisa meningkat. Penulis menggunakan pengklasifikasi Support Vector Machines karena merupakan teknik *machine learning* yang populer untuk klasifikasi teks, serta memiliki performa yang baik pada banyak domain. SVM memiliki kelebihan yaitu mampu mengidentifikasi *hyperplane* terpisah yang memaksimalkan margin antara dua kelas yang berbeda. SVM

menjamin untuk memaksimalkan jarak antara data yang paling dekat dengan *hyperplane*. Jika input data dapat dipisahkan secara linear. Lihat gambar 3. untuk model yang diusulkan secara lebih detail.



Gambar 3. Model yang diusulkan

Data harus melalui tahap *preprocessing* terlebih dahulu agar didapatkan kata-kata relevan untuk diklasifikasi. Proses evaluasi dilakukan menggunakan *10 Fold Cross Validation*. Pengukuran akurasi diukur dengan *Confusion Matrix*. Hasil yang dibandingkan adalah akurasi *Support Vector Machines (SVM)* sebelum menggunakan metode pemilihan fitur dengan akurasi *Support Vector Machines (SVM)* setelah menggunakan metode seleksi fitur yaitu *Genetic Algorithm (GA)*

HASIL DAN PEMBAHASAN

Klasifikasi Teks Menggunakan *Support Vector Machine*

Data training yang digunakan dalam pengklasifikasian teks ini terdiri dari 50 komentar *review* positif dan 50 komentar *review* negatif, berikut adalah tahapan prosesnya:

1. Pengumpulan Data

Data komentar *review* positif disatukan dalam folder dengan nama pos. Sedangkan data komentar *review* negatif disatukan penyimpanannya dalam folder dengan nama neg. Tiap dokumen berekstensi .txt yang dapat dibuka menggunakan aplikasi Notepad.

2. Pengolahan Awal Data

a. *Tokenization*

Dalam proses *tokenization* ini, semua kata yang ada di dalam tiap dokumen dikumpulkan dan dihilangkan tanda bacanya, serta dihilangkan jika terdapat simbol atau apapun yang bukan huruf. Berikut adalah contoh hasil dari proses *tokenization* dalam RapidMiner. Tabel 1 menunjukkan hasil perbandingan teks sebelum dan sesudah dilakukan proses *tokenization*.

Tabel 1. Perbandingan teks sebelum dan sesudah dilakukan proses *tokenization*

Teks data <i>komentar review</i> sebelum dilakukan proses <i>tokenization</i>	Kalau menurut <i>ane gan</i> : - <i>bb</i> lebih untuk komunikasi - tetapi <i>Android</i> lebih untuk Hiburan dan Entertainment
Teks data <i>komentar review</i> setelah dilakukan proses <i>tokenization</i>	Kalau menurut <i>ane gan bb</i> lebih untuk komunikasi tetapi <i>Android</i> lebih untuk Hiburan dan Entertainment

b. Filter Tokens (by Length)

Dalam proses ini, kata-kata yang memiliki pajang kurang dari 4 dan lebih dari 25 akan dihapus, seperti kata *yg,ane, gan* yang merupakan kata-kata yang tidak mempunyai makna tersendiri jika dipisahkan dengan kata yang lain dan tidak terkait dengan kata sifat yang berhubungan dengan sentimen. Tabel.2 menunjukkan hasil perbandingan teks sebelum dan sesudah dilakukan proses *Filter Tokens*.

Tabel 2. Perbandingan teks sebelum dan sesudah dilakukan proses *filter tokens*

Teks data <i>komentar review</i> sebelum dilakukan proses <i>filter tokens</i>	Kalau menurut <i>ane gan bb</i> lebih untuk komunikasi tetapi <i>Android</i> lebih untuk Hiburan dan Entertainment
Teks data <i>komentar review</i> setelah dilakukan proses <i>filter tokens</i>	Kalau menurut lebih untuk komunikasi tetapi <i>Android</i> lebih untuk Hiburan Entertainment

c. Stopwords

Dalam proses ini, kata-kata yang tidak relevan akan dihapus, seperti kata *tetapi, untuk, dengan,* yang merupakan kata-kata yang tidak mempunyai makna tersendiri jika dipisahkan dengan kata yang lain dan tidak terkait dengan kata sifat yang berhubungan dengan sentimen. Tabel 3 menunjukkan hasil perbandingan teks sebelum dan sesudah dilakukan proses *stopword removal*.

Tabel 3. Perbandingan teks sebelum dan sesudah dilakukan proses *Stopwords*

Teks data <i>komentar review</i> sebelum dilakukan proses <i>Stopwords Removal</i>	Kalau menurut lebih untuk komunikasi tetapi <i>Android</i> lebih untuk Hiburan Entertainment
Teks data <i>komentar review</i> setelah dilakukan proses <i>Stopwords Removal</i>	menurut komunikasi <i>Android</i> Hiburan Entertainment

d. Transform Cases

Dalam proses ini, kata-kata yang tidak relevan akan diubah, seperti kata yang mengandung huruf besar yang diubah menjadi huruf kecil sehingga dapat saling berhubungan dengan sentimen. Tabel 4 menunjukkan hasil perbandingan teks sebelum dan sesudah dilakukan proses *transform cases*

Tabel 4. Perbandingan teks sebelum dan sesudah dilakukan proses *transform cases*

Teks data komentar <i>review</i> sebelum dilakukan proses <i>transform cases</i>	<i>menurut komunikasi Android Hiburan Entertainment</i>
Teks data komentar <i>review</i> setelah dilakukan proses <i>transform cases</i>	<i>menurut komunikasi android hiburan entertainment</i>

e. Stemming

Dalam proses ini kata-kata akan dikelompokkan ke dalam beberapa kelompok yang memiliki kata dasar yang sama, seperti *hibur*, *hiburan*, *menghibur*, di mana kata dasar dari semuanya adalah kata *hibur*. Tabel 5 menunjukkan hasil perbandingan teks sebelum dan sesudah dilakukan proses *stemming*.

Tabel 5. Perbandingan teks sebelum dan sesudah dilakukan proses *Stemming*

Teks data komentar <i>review</i> sebelum dilakukan proses <i>Stemming</i>	<i>menurut komunikasi android hiburan entertainment</i>
Teks data komentar <i>review</i> setelah dilakukan proses <i>Stemming</i>	<i>turut komunikasi android hibur entertainment</i>

Klasifikasi

Proses klasifikasi pada penelitian ini adalah untuk menentukan sebuah kalimat sebagai anggota kelas positif atau kelas negatif berdasarkan nilai perhitungan pembobotan dari rumus SVM yang bernilai 0. Jika nilai dari bobot lebih besar sama dengan 0 maka diklasifikasikan kedalam kelas positif dan sebaliknya jika nilai bobot kurang dari sama dengan 0 maka dapat diklasifikasikan kedalam kelas negatif. Dokumen yang akan dihitung nilai bobotnya dapat dilihat pada table 6.

Berikut adalah perhitungan untuk contoh dokumen pos44.txt dengan fungsi sebagai berikut:

$$\begin{aligned} W_{44} &= Y_{44} \cdot X_{44} \\ &= 1 \cdot (0,667+0,667+0+0+0+ 0,667) \\ &= 2,001 \end{aligned}$$

Tabel 6. Tabel dokumen yang dihitung nilai bobotnya.

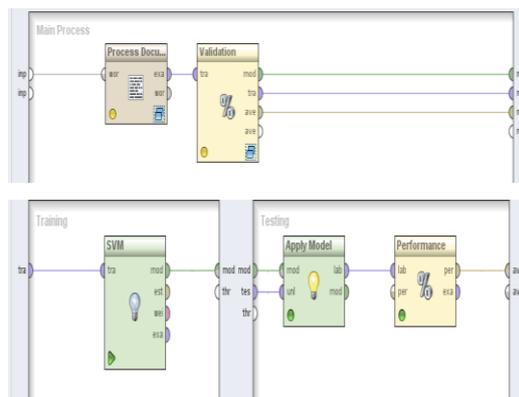
Dokumen	Atribut						Kelas
	android	cinta	blackberry	milih	benci	murah	
Pos44.txt (saya cinta android gan karena murah)	0,667	0,667	0	0	0	0,667	positif
Neg3.txt (tidak cinta benci android lebih memilih blackberry)	0,667	0,667	0,667	0,667	0,667	0	negatif

Di mana W_{44} adalah bobot untuk dokumen Pos44.txt yang menentukan klasifikasi kelas positif atau negatif, sedangkan Y_{44} merupakan nilai label untuk dokumen tersebut di mana Y hanya bernilai 1 untuk kelas positif dan -1 untuk kelas negatif. X_3 nilai-nilai atribut yang ada pada sebuah dokumen yang ingin dihitung bobotnya dan ditentukan kelasnya. Bobot untuk dokumen Pos44.txt adalah **2,001**, maka dokumen tersebut diklasifikasikan ke dalam kelas positif.

Berikut adalah perhitungan untuk contoh dokumen Neg3.txt dengan fungsi sebagai berikut:

$$\begin{aligned}
 W_3 &= Y_3 \cdot X_3 \\
 &= -1 \cdot (0,667+0+0,667+0,667+0,667+0) \\
 &= -3,336
 \end{aligned}$$

Di mana W_3 adalah bobot untuk dokumen pos44.txt yang menentukan klasifikasi kelas positif atau negatif, sedangkan Y_{44} merupakan nilai label untuk dokumen tersebut di mana Y hanya bernilai 1 untuk kelas positif dan -1 untuk kelas negatif. X_3 nilai-nilai atribut yang ada pada sebuah dokumen yang ingin dihitung bobotnya dan ditentukan kelasnya. Bobot untuk dokumen Neg44.txt adalah **-3,336**, maka dokumen tersebut diklasifikasikan ke dalam kelas negatif. Perhitungan di atas dapat dibuat suatu model dengan menggunakan RapidMiner 5. Desain model *Support Vector Machine* (SVM) dapat dilihat pada gambar 4.



Gambar 4. Desain model *Support Vector Machine* (SVM) menggunakan RapidMiner

Pengujian Model dengan 10 Fold Cross Validation

Pada penelitian ini, penulis 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.

Optimasi Model dengan Metode *Genetic algorithm*

Dalam penelitian ini metode yang digunakan adalah *Genetic Algorithm* untuk meningkatkan akurasi pengklasifikasi *Support Vector Machine* (SVM).

1. Menghasilkan populasi awal yang terdiri dari *population_size* perorangan (individu). Setiap atribut diaktifkan dengan *p_initialize*.
2. Memilih dua individu dari populasi dan melakukan *crossover* dengan probabilitas *p_crossover*.
3. probabilitas masing-masing atribut untuk diseleksi untuk sebuah generasi (keturunan)
4. jika generasi membaik, kembali ke langkah ke a.

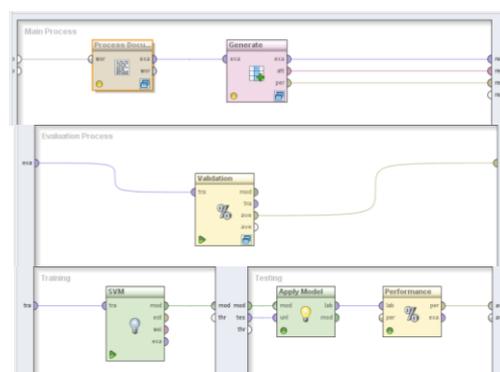
Eksperimen Terhadap Indikator Model

Untuk mendapatkan model yang baik, beberapa indikator disesuaikan nilainya agar didapatkan hasil akurasi yang tinggi. Dalam penelitian ini *Genetic Algorithm*, indikator yang disesuaikan adalah *population size=15*, *p initialize=0.9*, *p crossover=0.9*, dan *p generate=1.0*. Tabel indikator dan hasil pengujian dapat dilihat pada tabel 7

Tabel 7. Tabel indikator dan hasil pengujian

Population size	P Initialize	P Crossover	P Mutation	C	Akurasi
15	0.5	0.5	1.0	1.0	74.96%
10	0.7	0.7	1.0	1.0	74.03%
20	0.2	0.2	1.0	1.0	76.08%
15	0.2	0.2	1.0	1.0	78.02%
15	0.9	0.9	1.0	1.0	73.11%

Dalam penyesuaian indikator pada *Genetic algorithm*, akurasi paling tinggi diperoleh dengan kombinasi *population size=15*, *p initialize=0.2*, *p crossover=0.2*, dan *p generate=1.0*. Hasil akurasi mencapai **78.02%**. Jika indikator lainnya turut diubah nilainya, dapat menyebabkan proses pengolahan data menjadi semakin lama. Desain model *Support Vector Machine* (SVM) dengan *Genetic Algorithm* dapat dilihat pada gambar 5.



Gambar 5. Desain model *Support Vector Machine* dan *Genetic Algorithm* di dalam RapidMiner

Pembahasan

Dengan memiliki model klasifikasi komentar *review*, dapat digunakan untuk menindak lanjuti suatu strategi konsumen memutuskan untuk membeli *smartphone* Android atau Blackberry, yaitu dengan mengidentifikasi komentar *review* tersebut sentimen yang positif maupun yang negatif. Dari data komentar *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 komentar *review* tersebut positif atau negatif. Dalam penelitian ini, hasil pengujian model akan dibahas melalui *confusion matrix* untuk menunjukkan seberapa baik model yang terbentuk. Tanpa menggunakan metode Genetic Algorithm, algoritma *Support Vector Machine* (SVM) sendiri sudah menghasilkan akurasi sebesar 60.00% dan nilai AUC 0.664. Akurasi tersebut masih kurang akurat, sehingga perlu ditingkatkan lagi menggunakan metode wrapper Genetic Algorithm. Akurasi algoritma *Support Vector Machine* (SVM) meningkat menjadi 71.00% dan nilai AUC 0.964 seperti yang bisa dilihat pada tabel 8

Table 8 Model algoritma *Support Vector Machine* (SVM) sebelum dan sesudah menggunakan metode *Genetic Algorithm*

	Algoritma <i>Support Vector Machine</i> (SVM)	Algoritma <i>Support Vector Machine</i> (SVM) + Genetic Algorithm
Sukses klasifikasi komentar <i>review</i> positif	44	43
Sukses klasifikasi komentar <i>review</i> negatif	27	35
Akurasi Model	71.00%	78,02%
AUC	0.862	0.855

Pengukuran dengan *Confusion Matrix*

Pengukuran dengan *confusion matrix* di sini akan menampilkan perbandingan dari hasil akurasi model *Support Vector Machine* (SVM) sebelum ditambahkan metode *Genetic algorithm* yang bisa dilihat pada tabel 9 dan setelah ditambahkan metode Genetic algorithm yang bisa dilihat pada tabel 10.

Tabel 9. *Confusion Matrix* Model *Support Vector Machine* (SVM) sebelum penambahan metode Genetic Algorithm

Akurasi <i>Support Vector Machine</i> (SVM): 71.00% + - 11.36% (mikro 71.00%)			
	True negative	True positif	Kelas precision
Pred.negatif	27	6	81.82 %
Pred.Positif	23	44	65.67%
Kelas recall	54.00%	88.00%	

Tabel 10. *Confusion Matrix* Model *Support Vector Machine* (SVM) sesudah penambahan metode Genetic Algorithm

Akurasi <i>Support Vector Machine</i> (SVM): 78.02% + - 5.07% (mikro 78.00%)

	True negative	True positif	Kelas precision
Pred.negatif	35	7	83.33%
Pred.Positif	15	43	74.14%
Kelas recall	70.00%	86.00%	

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} = \frac{35+43}{35+43+15+7} = \frac{78}{100} = 0.780=78.00\%$$

$$Precision = \frac{TP}{TP+FP} = \frac{43}{43+7} = \frac{43}{50} = 0.860=86.00\%$$

$$Recall = \frac{TP}{TP+FN} = \frac{35}{35+15} = \frac{35}{50} = 0.700 = 70.00\%$$

Berikut adalah tampilan kurva ROC yang akan dihitung nilai AUC-nya. Gambar 6 adalah kurva ROC untuk model *Support Vector Machine* (SVM) sebelum menggunakan metode Genetic Algorithm dan gambar IV.4 adalah kurva ROC untuk model *Support Vector Machine* (SVM) setelah menggunakan metode Genetic Algorithm.



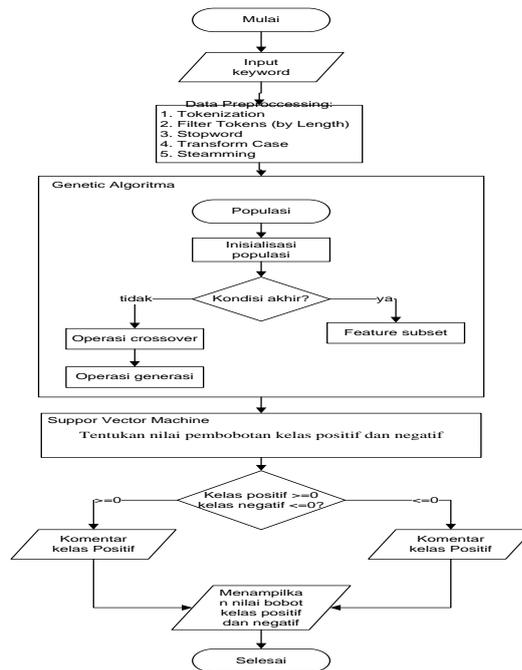
Gambar 6. Kurva ROC Model *Support Vector Machine* (SVM) sebelum menggunakan metode Genetik Algorithm



Gambar 7. Kurva ROC Model *Support Vector Machine* (SVM) setelah menggunakan metode Genetik Algorithm

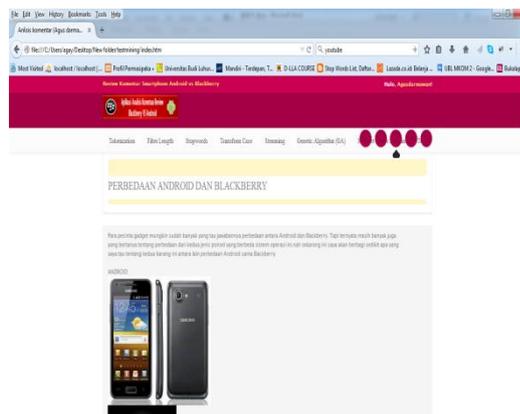
Desain dan Implementasi

Penelitian ini akan membuat aplikasi untuk menguji model yang sudah ada menggunakan dataset yang berbeda dan belum diketahui kelasnya. Aplikasi dibuat menggunakan perangkat lunak *Macromedia Dreamweaver* dengan bahasa pemrograman PHP. Gambar 6 adalah diagram alir dari tahapan proses klasifikasi pada aplikasi yang penulis buat.

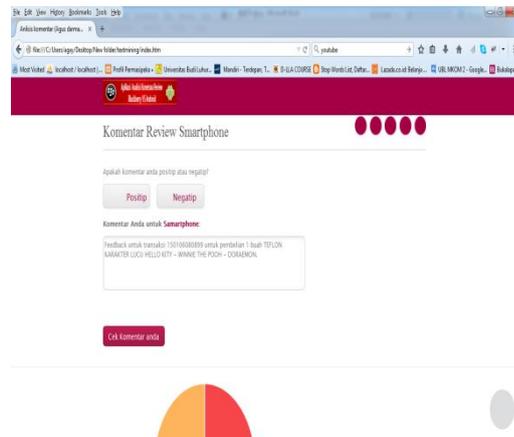


Gambar 8. Diagram alir tahapan proses klasifikasi dengan *Support Vector Machine* (SVM) menggunakan Metode Genetic algorithm

Berikut adalah tampilan rancangan aplikasi dan implementasinya. Gambar 9. adalah ketika aplikasi diinput *komentar dan dilakukan pengecekan komentar* maka akan menampilkan analisis komentar dalam diagram lingkaran beserta jumlah presentase *komentar review* positif dan *komentar review* negatif serta ditampilkan beberapa kometar terbaru .



Gambar 9. Tampilan aplikasi sentimen analisis *komentar*



Gambar 10. Tampilan aplikasi sentimen analisis *komentar*

PENUTUP

Kesimpulan

Dalam penelitian ini dilakukan pengujian model dengan menggunakan Support Vector Machine dan Support Vector Machine berbasis Optimasi *Genetic Algorithm (GA)* dengan menggunakan data komentar *review* produk *smartphone* android dan blackberry yang positif maupun negatif. Model yang dihasilkan diuji untuk mendapatkan nilai *accuracy*, *precision*, *recall* dan AUC dari setiap algoritma sehingga didapat pengujian dengan menggunakan Support Vector Machine nilai *accuracy* adalah 71.00%. Sedangkan pengujian dengan menggunakan Support Vector Machine berbasis Optimasi *Genetic Algorithm GA* menghasilkan nilai *accuracy* adalah 78.02%. Maka dapat pengujian data komentar *review* produk *smartphone* android dan blackberry menggunakan Support Vector Machine berbasis Optimasi *Genetic Algorithm (GA)* lebih baik daripada Support Vector Machine sendiri. Dengan demikian dari hasil pengujian model diatas dapat disimpulkan bahwa Support Vector Machine berbasis Optimasi *Genetic Algorithm (GA)* memberikan pemecahan untuk permasalahan klasifikasi komentar *review* produk *smartphone* android dan blackberry lebih akurat. Dari hasil *review* komentar di dapatkan bawah peminat dari android lebih banyak dari blackberry.

Saran

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

1. Menggunakan metode pemilihan fitur yang lain, seperti Chi Square, Gini Index, Mutual Information, dan lain-lain agar hasilnya bisa dibandingkan
2. Menggunakan data komentar *review* dari domain yang berbeda, misalnya komentar *review* restoran, komentar *review* film, *review* saham dan lain sebagainya.

DAFTAR PUSTAKA

- Basari, A. S. H dan Hussin, B., Ananta, I. G. P., dan Zeniarja, J. **Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization.** *Procedia Engineering*, 53, 453–462. doi:10.1016/j.proeng.2013.02.059, 2013
- Dehkharghani, R. dan Mercan, H. dan Javed, A., Saygin, Y. **Sentimental causal rule discovery from Twitter.** *Expert Systems with Applications*, 41(10), 4950–4958. doi:10.1016/j.eswa.2014.02.024, 2014
- Haddi, E., Liu, X., dan Shi, Y. **The Role of Text Pre-processing in Sentiment Analysis.** *Procedia Computer Science*, 17, 26–32. doi:10.1016/j.procs.2013.05.005, 2013.

- Huang, K., Yang, H., King, I., dan Lyu, M. *Machine Learning Modeling Data Locally And Globally*. Berlin Heidelberg: Zhejiang University Press, 2008.
- Khan, K., Baharudin, B., dan Khan, A. **Mining Opinion Components from Unstructured Reviews: A Review**. *Journal of King Saud University - Computer and Information Sciences*. doi:10.1016/j.jksuci.2014.03.009, 2014.
- Liu, Y., Wang, G., Chen, H., Dong, H., Zhu, X., dan Wang, S. **An Improved Particle Swarm Optimization for Feature Selection**. *Journal of Bionic Engineering*, 8(2), 191–200. doi:10.1016/S1672-6529(11)60020-6, 2011.
- A.S. Nugroho, A. W, “**Application of Support Vector Machine in Bioinformatics**”, *Indonesian Scientific Meeting in Central Japan*, 2008.
- Medhat, W. ,Hassan, A., dan Korashy, H. **Sentiment analysis algorithms and applications: A survey**. *Ain Shams Engineering Journal*. doi:10.1016/j.asej.2014.04.011, 2014
- Vercellis, C. *Business Intelligence Data Mining And Optimization For Decision Making* .United Kingdom: A John Wiley And Sons, Ltd.,Publication, 2009.
- Wang, Z., Shao, Y.-H., dan Wu, T.-R. 2013. *A GA-based model selection for smooth twin parametric-margin support vector machine*. *Pattern Recognition*, 46(8), 2267–2277. doi:10.1016/j.patcog.2013.01.023, 2013
- Weiss, S. M., Indurkha , Nitin dan Zhang, Tong. *Fundamentals of Predictive Text Mining*. London: Springer-Verlag,2010
- Zhang, Z., Ye, Q., Zhang, Z., dan Li, Y. **Sentiment classification of Internet restaurant reviews written in Cantonese**. *Expert Systems with Applications*, 38(6), 7674–7682. doi:10.1016/j.eswa.2010.12.147, 2011.
- Zhao, M., Fu, C., Ji, L., Tang, K., dan Zhou, M.**Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes**. *Expert Systems with Applications*, 38(5), 5197–5204. doi:10.1016/j.eswa.2010.10.041, 2011.
- Zukri, Zainudin. *Algoritma Genetika Metode Komputasi Evolusioner untuk Menyelesaikan Masalah Optimasi*. Yogyakarta: Andi Offset,2014.