

ANALISA MODEL SUPPORT VECTOR MACHINE TEXTMINING PADA KOMENTAR POSITIF DAN NEGATIF UNTUK REVIEW PERBANDINGAN WHATSAPP VS BBM

Agus darmawan¹, Syamsiah²

Program Studi Teknik Informatika, Universitas Indraprasta PGRI ^{1,2}

Email: agus.darmawan@ymail.com¹

Email: ncham.unindra08@gmail.com²

Abstrak

Sebagai pengguna *smartphone*, akhir-akhir ini kita dibuat pusing oleh munculnya beberapa layanan instant messaging lintas platform mobile yang kelihatannya dalam waktu singkat dapat menjadi sangat populer. Mulai dari *Whatsapp* dan *BBM* rasanya saat ini sedang bersaing ketat dalam memperebutkan perhatian para pengguna *smartphone*. Sebelum kita memutuskan untuk menggunakan instant messaging, ada hal dapat dipelajari dari hasil komentar review perbandingan *Whatsapp* dan *BBM*. 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. Penelitian ini menghasilkan klasifikasi teks dalam bentuk positif atau negatif dari *review* layanan *instant messaging* lintas *platform mobile*. Pengukuran berdasarkan akurasi *Support Vector Machines*, sedangkan pengukuran akurasi diukur dengan *confusion matrix* dan kurva ROC. Hasil penelitian menunjukkan akurasi *Support Vector Machines* dari 71.00 %.

Kata Kunci: Komentar, SVM, BBM, *Whatsapp*, *Confusion Matrix*, Kurva ROC

Pendahuluan

Persaingan layanan *chatting* di *smartphone* belakangan ini mencapai puncaknya setelah salah satu *producent smartphone* berhasil menjual produknya ke seluruh dunia lantaran aplikasi *messaging* yang ditanamkan didalamnya memberi sensasi berbeda dari pada layanan *chatting* yang beredar di internet selama ini. Menjamurnya penggunaan aplikasi *messaging* di beberapa tahun belakangan ini dikarenakan beberapa faktor diantaranya Fenomena *social messaging*, sifat multi platform, sebagai alat komunikasi yang sederhana dan mudah dipakai serta dibawa kemana saja. Hal ini yang menjadikan Aplikasi *messaging* akhir-akhir ini menjadi aplikasi wajib install (harus ada) bagi pengguna *smartphone*. Di tanah air, setidaknya sampai saat ini ada empat aplikasi yang bersaing ketat. Mulai dari *BBM*, *Whatsapp*, *Line*, *Kakao Talk* dan *WeChat*. Memang masih ada beberapa lagi layanan sejenis. Namun tidak semua Aplikasi *messaging* memiliki kualitas yang baik untuk menunjang kebutuhan konsumen dan hal ini yang harus diperhatikan oleh para konsumen. Sebelum konsumen memutuskan untuk menggunakan Layanan Aplikasi *messaging* yang baik, sebaiknya konsumen mengetahui dengan Keunggulan dari Layanan Aplikasi *messaging* tersebut, hal ini dapat dipelajari dari testimoni dan opini atau hasil komentar *review* perbandingan dari *whatsapp* dan *BBM*. Saat ini konsumen yang menulis opini dan pengalaman secara *online* semakin meningkat. Membaca komentar *review* tersebut secara keseluruhan dapat memakan waktu, namun jika hanya sedikit komentar *review* yang dibaca evaluasi akan menjadi bias. Klasifikasi sentimen bertujuan untuk mengatasi masalah ini dengan secara otomatis mengelompokkan *review* pengguna menjadi opini positif atau negatif (Z. Zhang et al., 2011). Dari beberapa teknik klasifikasi sentimen tersebut yang paling sering digunakan untuk klasifikasi data adalah *Support Vector Machines* (SVM). SVM merupakan metode *supervised learning* yang menganalisa data dan mengenali pola-pola yang digunakan untuk klasifikasi (Basari et all: 2013). *Support Vector Machines* (SVM) adalah kasus khusus

dari keluarga algoritma yang disebut sebagai *regularized* metode klasifikasi linier dan metode yang kuat untuk meminimalisasi resiko (Weiss et al., 2010). SVM memiliki kelebihan yaitu mampu mengidentifikasi *hyperplane* terpisah yang memaksimalkan margin antara dua kelas yang berbeda (Chou et al., 2014). Namun *Support Vector Machine* (Basari et al., 2013). Pada penelitian ini algoritma *Support Vector Machines* diterapkan oleh penulis untuk mengklasifikasikan teks pada komentar *review* perbandingan antara whatsapp dan BBM dalam rangka meningkatkan akurasi analisa sentimen.

Tinjauan Pustaka

Review Perbandingan BlackBerry Messenger Vs WhatsApp

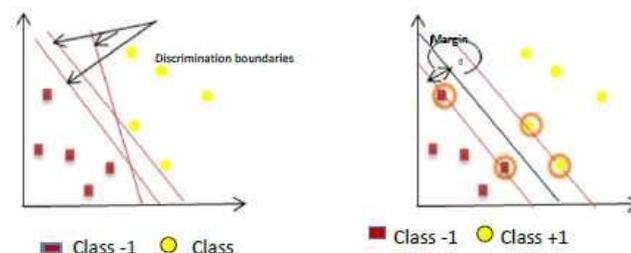
Situs *web* adalah wadah bagi beragam pendapat. Salah satu bentuk opini yang memiliki kredibilitas adalah *review* produk. Situs *web* seperti amazon.com mendorong pengguna untuk memberikan ulasan (*review*) (Weiss et al., 2010). Hasil penambangan teks dari *review* tersebut dapat diklasifikasikan dengan 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 Kaskus.com merupakan jenis situs lain yang berisikan Perbandingan BlackBerry Messenger Vs WhatsApp.

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).

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 ($i=1. . .k$), 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:

$$f(X) = w^T x + b \quad (1)$$

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:

$$y_i(w \cdot x_i + b) \geq 1 \quad i=1. \dots k \quad (2)$$

Metodologi Penelitian

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 tahapan *preprocessing*.

4. 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.

5. 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/523c00edfcca174a4a00000f/blackberry-messenger-vs-whatsapp-bagus-mana/>. Contoh komentar *review* positif sebagai berikut:

“Kalau menurut saya:

- *bbm lebih ke segi komunikasi*

- *whatsapp lebih ke segi hiburan dan entertainment*”

Sedangkan contoh komentar *review* negatif sebagai berikut:

“setuju gan!

Tp whatsapp buat komunikasi oke juga kok.

Kapok ane pake bbm”

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.

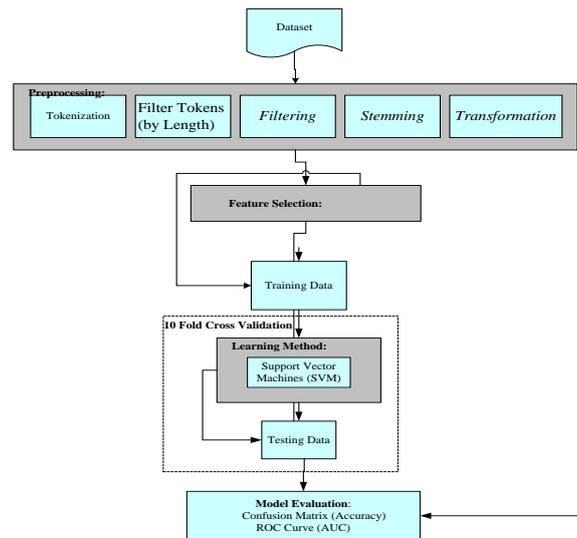
5. *Stemming*

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) metode pemilihan fitur agar akurasi pengklasifikasi Support Vector Machines (SVM) bisa digunakan. 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 2. 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 adalah akurasi *Support Vector Machines (SVM)*.

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*

No	Teks Sebelum	Teks Sesudah
1.	Teks data komentar <i>review</i> sebelum dilakukan proses <i>tokenization</i>	<i>Kalau menurut ane gan:</i> - <i>whatsapp</i> lebih untuk komunikasi - tetapi <i>BBM</i> lebih untuk Hiburan dan Entertainment
2.	Teks data komentar <i>review</i> setelah dilakukan proses <i>tokenization</i>	<i>Kalau menurut ane gan whatsapp</i> lebih untuk komunikasi tetapi <i>BBM</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*

No	Teks Sebelum	Teks Sesudah
1.	Teks data komentar <i>review</i> sebelum dilakukan proses <i>filter tokens</i>	Kalau menurut ane gan whatsapp lebih untuk komunikasi tetapi BBM lebih untuk Hiburan dan Entertainment
2.	Teks data komentar <i>review</i> setelah dilakukan proses <i>filter tokens</i>	Kalau menurut whatsapp lebih untuk komunikasi tetapi BBM 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*

No	Teks Sebelum	Teks Sesudah
1.	Teks data komentar <i>review</i> sebelum dilakukan proses <i>Stopwords Removal</i>	Kalau menurut whatsapp lebih untuk komunikasi tetapi BBM lebih untuk Hiburan Entertainment
2.	Teks data komentar <i>review</i> setelah dilakukan proses <i>Stopwords Removal</i>	menurut whatsapp komunikasi BBM 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*

No	Teks Sebelum	Teks Sesudah
1.	Teks data komentar <i>review</i> sebelum dilakukan proses <i>transform cases</i>	menurut whatsapp komunikasi BBM Hiburan Entertainment
2.	Teks data komentar <i>review</i> setelah dilakukan proses <i>transform cases</i>	menurut whatsapp komunikasi bmm hiburan entertainment

e. *Stemming*

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*

No	Teks Sebelum	Teks Sesudah
1.	Teks data komentar <i>review</i> sebelum dilakukan proses <i>Stemming</i>	Menurut whatsapp komunikasi bmm hiburan entertainment
2.	Teks data komentar <i>review</i> setelah dilakukan proses <i>Stemming</i>	turut whatsapp komunikasi bmm hibur entertainment

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 tabel 6.

Tabel 6. Tabel Dokumen yang Dihitung Nilai Bobotnya

Dokumen	Atribut						kelas
	bbm	cinta	whatsapp	milih	benci	murah	
Pos44.txt (saya cinta bbm gan karena murah)	0,667	0,667	0	0	0	0,667	positif
Neg3.txt (tidak cinta benci bbm lebih memilih whatsapp)	0,667	0,667	0,667	0,667	0,667	0	negatif

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

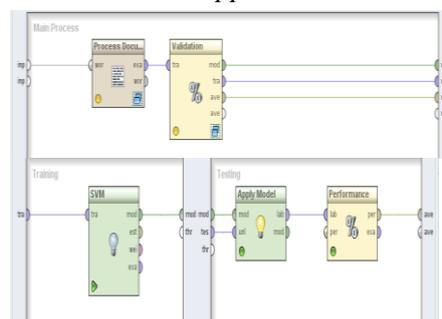
$$\begin{aligned}
 W_{44} &= Y_{44} \cdot X_{44} \\
 &= 1 \cdot (0,667+0,667+0+0+0+ 0,667) \\
 &=2,001
 \end{aligned}$$

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 kedalam 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 kedalam kelas negatif. Perhitungan Desain model *Support Vector Machine* (SVM) dapat dilihat pada gambar 4.



Gambar 3. 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 fold 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.

Pembahasan

Dengan memiliki model klasifikasi komentar *review*, dapat digunakan untuk menindak lanjuti suatu strategi konsumen memutuskan untuk menggunakan whatsapp atau BBM, 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. algoritma *Support Vector Machine* (SVM) sendiri sudah menghasilkan nilai akurasi 71.00% dan nilai AUC 0.862 seperti yang bisa dilihat pada tabel 7.

Tabel 7. Model Algoritma Support Vector Machine (SVM)

No	Algoritma Support Vector Machine (SVM)	Nilai
1.	Sukses klasifikasi komentar <i>review</i> positif	44
2.	Sukses klasifikasi komentar <i>review</i> negatif	27
3.	Akurasi Model	71.00%
4.	AUC	0.862

Pengukuran dengan Confusion Matrix

Pengukuran dengan *confusion matrix* di sini akan menampilkan perbandingan dari hasil akurasi model *Support Vector Machine* (SVM) bisa dilihat pada tabel 9.

Table 9. Confusion Matrix Model Support Vector Machine (SVM)

No	Name	True negative	True positif	Kelas precision
1.	Pred.negatif	27	6	81.82 %
2.	Pred.Positif	23	44	65.67%
3.	Kelas recall	54.00%	88.00%	

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} = \frac{27+44}{27+44+23+6} = \frac{71}{100} = 0.710=71.00\%$$

$$Precision = \frac{TP}{TP+FP} = \frac{44}{44+6} = \frac{44}{50} = 0.880=88.00\%$$

$$Recall = \frac{TP}{TP + FN} = \frac{27}{27 + 23} = \frac{27}{50} = 0.540 = 54.00\%$$

Berikut adalah tampilan kurva ROC yang akan dihitung nilai AUC-nya. Gambar 5 adalah kurva ROC untuk model *Support Vector Machine* (SVM).



Gambar 4. Kurva ROC Model *Support Vector Machine* (SVM)

Kesimpulan dan Saran

Kesimpulan

Dalam penelitian ini dilakukan pengujian model dengan menggunakan *Support Vector Machine* 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%. Dengan demikian dari hasil pengujian model diatas dapat disimpulkan bahwa *Support Vector Machine* dapat memberikan pemecahan untuk permasalahan klasifikasi komentar *review* perbandingan whatsapp dan BBM lebih akurat.

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. (2013). 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.
- Haddi, E., Liu, X., dan Shi, Y. (2014). The Role of Text Pre-processing in Sentiment Analysis. *Procedia Computer Science*, 17, 26–32. doi:10.1016/j.procs.2013.05.005.
- Khan, K., Baharudin, B., dan Khan, A. (2014). 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.
- Nugroho, A.W. (2008). *Application of Support Vector Machine in Bioinformatics. Indonesian Scientific Meeting in Central Japan*.
- Weiss, S. M., Indurkha, Nitin dan Zhang, Tong. (2010). *Fundamentals of Predictive Text Mining*. London: Springer-Verlag
- Zhang, Z., Ye, Q., Zhang, Z., dan Li, Y. (2011). Sentiment classification of Internet restaurant reviews written in Cantonese. *Expert Systems with Applications*, 38(6), 7674–7682. doi:10.1016/j.eswa.2010.12.14.