

**PENERAPAN PRINCIPAL COMPONENT ANALYSIS
DAN GENETIC ALGORITHM UNTUK PEMILIHAN
FITUR PADA ANALISIS SENTIMEN REVIEW
PENGIRIMAN BARANG MENGGUNAKAN
ALGORITMA SUPPORT VECTOR
MACHINE**



TESIS

HILDA RACHMI

14000896

**PROGRAM PASCASARJANA MAGISTER ILMU KOMPUTER
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
NUSA MANDIRI
JAKARTA
2015**

HALAMAN PENGESAHAN

Tesis diajukan oleh:

Nama : Hilda Rachmi
NIM : 14000896
Program Studi : Magister Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : *Management Information System*
Judul Tesis : Penerapan Principal Component Analysis dan Genetic Algorithm untuk Pemilihan Fitur pada Analisis Sentimen Review Pengiriman Barang menggunakan Algoritma Support Vector Machine

Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Magister Ilmu Komputer (M.Kom) pada Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri).

Jakarta, 18 Agustus 2015
Pascasarjana Magister Ilmu Komputer
STMIK Nusa Mandiri
Direktur

Prof. Dr. Ir. R. Eko Indrajit, M.Sc, MBA

DEWAN PENGUJI

Penguji I : Dr. Sularso Budilaksono



.....

Penguji II : Dr. Sfenrianto, M.Kom



.....

Penguji III/
Pembimbing : Dr. Mochamad Wahyudi, MM, M.Kom, M.Pd



.....

	Lembar Konsultasi Bimbingan Tesis
	Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri

NIM : 14000896
 Nama Lengkap : Hilda Rachmi
 Dosen Pembimbing : Dr. Mochamad Wahyudi, MM, M.Kom, M.Pd
 Judul Tesis : Penerapan Principal Component Analysis dan Genetic Algorithm untuk Pemilihan Fitur pada Analisis Sentimen Review Pengiriman Barang menggunakan Algoritma Support Vector Machine

No	Tanggal Bimbingan	Materi Bimbingan	Paraf Dosen Pembimbing
1	23 Juni 2015	Bimbingan perdana dan periksa Bab I	
2	07 Juli 2015	Acc Bab I dan periksa Bab II	
3	28 Juli 2015	Acc Bab II dan pengajuan Bab III	
4	05 Agustus 2015	Acc Bab III dan pengajuan Bab IV	
5	10 Agustus 2015	Acc Bab IV dan periksa Bab V	
6	15 Agustus 2015	Acc Bab V dan koreksi keseluruhan	
7	18 Agustus 2015	Acc keseluruhan dan daftar sidang tesis	

Bimbingan dimulai pada tanggal : 23 Juni 2015
 Bimbingan diakhiri pada tanggal : 18 Agustus 2015
 Jumlah Pertemuan : 7 Kali Pertemuan

Jakarta, 18 Agustus 2015

Dosen Pembimbing



Dr. Mochamad Wahyudi, MM, M.Kom, M.Pd

ABSTRAK

Nama : Hilda Rachmi
NIM : 14000896
Program Studi : Magister Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : *Management Information System*
Judul Tesis : Penerapan Principal Component Analysis dan Genetic Algorithm untuk Pemilihan Fitur pada Analisis Sentimen Review Pengiriman Barang menggunakan Algoritma Support Vector Machine

Analisis sentimen menganalisis pendapat orang-orang, sentimen, sikap, penilaian, evaluasi, dan emosi terhadap entitas seperti produk, jasa, peristiwa, organisasi, individu, masalah, topik dan atributnya. Analisis sentimen dari data pelanggan melalui website sangat penting dalam skenario bisnis saat ini karena setiap kali harus membuat keputusan seseorang perlu mendengar pendapat orang lain. Informasi ini dapat digunakan untuk riset pasar, umpan balik produk dan menganalisis efektivitas layanan pelanggan. Masalah pada penelitian ini yaitu pemilihan fitur untuk meningkatkan akurasi *Support Vector Machine* serta mencari nilai parameter untuk mendapatkan akurasi yang paling tinggi pada analisis sentimen review pengiriman barang serta menghasilkan klasifikasi hasil review negatif dan positif dengan tepat. Penulis menggunakan *Principal Component Analysis* dan *Genetic Algorithm* sebagai optimasi untuk meningkatkan akurasi pada metode Support Vector Machine. Akurasi yang dihasilkan pada algoritma *Support Vector Machine* sebesar 78,67%, setelah dioptimasi dengan menggunakan *Principal Component Analysis* dan *Genetic Algorithm* akurasi mengalami peningkatan menjadi 98,33%. Dapat disimpulkan bahwa penerapan optimasi dapat meningkatkan akurasi dan model Support Vector Machine berbasis *Principal Component Analysis* dan *Genetic Algorithm* dalam memberikan solusi terhadap permasalahan klasifikasi analisis sentimen review pengiriman barang yang menghasilkan review negatif dan positif dengan lebih akurat.

Kata Kunci : Analisis Sentimen, Review Pengiriman Barang, Support Vector Machine

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN PENGESAHAN.....	iii
ABSTRAK	iii
DAFTAR ISI.....	iii
BAB 1 PENDAHULUAN	1
BAB 2 LANDASAN TEORI.....	3
BAB 3 METODE PENELITIAN.....	5
BAB 4 HASIL PENELITIAN DAN PEMBAHASAN	6
BAB 5 KESIMPULAN DAN SARAN.....	15
DAFTAR REFERENSI	16

BAB 1

PENDAHULUAN

Perkembangan ilmu pengetahuan dan teknologi komputer memberikan pengaruh yang sangat besar di dunia teknologi informasi, dengan demikian mendorong kemunculan berbagai jenis aplikasi, seperti desktop, web, atau *mobile*. Di antara tiga aplikasi, web adalah yang paling cepat maju saat ini, sehingga membuat internet menjadi kebutuhan primer (Made et al., 2013). Web adalah tempat yang baik bagi orang-orang untuk mengekspresikan pendapat mereka, pada berbagai topik. Bahkan pemberi opini secara profesional memiliki blog dimana publik dapat mengomentari dan merespon apa yang mereka pikirkan (Wayan et al., 2013).

Pada umumnya review yang diberikan oleh customer memiliki rating yang dapat ditentukan oleh pelanggan itu sendiri. Namun, rating yang diberikan tidak selalu menunjukkan isi dari review yang disampaikan. Pembaca review pada umumnya tidak dapat menilai sebuah review hanya melalui rating. Penilaian sebuah review harus dilakukan dengan membaca keseluruhan isi dari review tersebut. Hal ini sangat merepotkan apabila review yang diberikan berjumlah banyak. Membaca *review* tersebut secara keseluruhan dapat membingungkan atau memakan waktu, namun jika hanya beberapa review yang dibaca evaluasi akan menjadi bias (Zhang, Ye, Zhang, dan Li, 2011). Oleh karena itu perlu adanya sistem yang bisa mengklasifikasikan review secara otomatis.

Analisis sentimen dapat digunakan untuk melihat karakter sifat pengguna terhadap suatu topik atau isu yang menarik berdasarkan identifikasi dari pola reaksi yang dapat ditemukan dalam data berbasis teks yang diposting dan dapat dikumpulkan (Ghiassi, Skinner, dan Zimbra, 2013). Pola sentimen tersembunyi di dalam komentar, umpan balik dan kritik yang sering memberikan informasi berguna yang dapat dimanfaatkan untuk tujuan yang berbeda (Kontopoulos, Berberidis, Dergiades, dan Bassiliades, 2013). Salah satu contohnya adalah data yang dibagikan oleh pengguna internet mencerminkan sentimen mereka terhadap layanan atau produk untuk membantu meningkatkan kualitas produk atau jasa (Wang, 2015).

Berbagai jenis metode klasifikasi telah diterapkan seperti Support Vector Machine (SVM), Naive Bayesian classifier, Decision trees, Entropy, Fuzzy logic, KNN (k-nearest-neighbour) dan masih banyak lagi (Swati, 2015). Pada penelitian sebelumnya, Songbo Tan dan Jin Zhang (Tan dan Zhang, 2008) membandingkan empat metode seleksi fitur (MI, IG, CHI dan DF) dan lima metode pembelajaran (centroid classifier, K-nearest neighbor, winnow classifier, Naive Bayes dan SVM) yang diteliti pada sentimen corpus Cina dengan jumlah 1.021 dokumen. Hasil percobaan menunjukkan bahwa IG merupakan seleksi fitur yang terbaik dan SVM menunjukkan kinerja terbaik untuk klasifikasi sentimen. Peneliti lain melakukan analisis sentiment dengan menggunakan Support Vector Machine dan Artificial Neural Network (Moraes, Valiati, dan Gavião Neto, 2013)

SVM merupakan algoritma yang sensitif terhadap pemilihan parameter yang digunakan. Parameter C dianggap dapat meningkatkan akurasi pada SVM

Namun, sulitnya menentukan parameter C tetap menjadi kendala dalam meningkatkan akurasi SVM. Pemilihan nilai parameter C akan dipilih dengan melihat implikasi terhadap hasil akurasi yang paling tinggi terhadap nilai C yang diberikan (Kurniawan, 2013). Pemilihan parameter C yang tepat diharapkan dapat diterapkan dalam penelitian ini.

Analisis Komponen Utama atau *Principal Component Analysis* (PCA) adalah teknik yang digunakan untuk menyederhanakan suatu data, dengan cara mentransformasi linier sehingga terbentuk sistem koordinat baru dengan variansi maksimum. PCA dapat digunakan untuk mereduksi dimensi suatu data tanpa mengurangi karakteristik data tersebut secara signifikan (Puspitaningrum, 2014). Menurut (Morchid, Dufour, Bousquet, Linarès, dan Torres-Moreno, 2014) Tujuan dari PCA adalah untuk mengekstrak informasi yang paling penting dari tabel data, kompres ukuran dataset dengan hanya menjaga informasi yang penting, menjelaskan dan menyederhanakan deskripsi dataset, serta menganalisis struktur pengamatan dan variabel.

Algoritma genetika merupakan salah satu algoritma optimasi seleksi fitur. Menurut (Gunal, 2012) salah satu metode *wrapper* yang bisa digunakan dalam pemilihan fitur adalah *Genetic algorithm* (GA). Proses seleksi pada GA adalah dengan mengambil beberapa individu terbaik. Selain itu juga dapat dilakukan dengan proses pengambilan acak proporsional, dengan proporsi setara dengan proporsi kualitasnya. artinya individu yang kualitasnya lebih baik memiliki peluang terpilih lebih besar dan pengambilan dilakukan dengan pemulihan (Sartono, 2010).

Berdasarkan penelitian sebelumnya, *Support Vector Machine* merupakan algoritma klasifikasi yang banyak digunakan oleh para peneliti dibidang analisa sentimen. Untuk meningkatkan akurasi *Support Vector Machine* selain dengan mengubah nilai parameter, maka pada penelitian ini *Principal Component Analysis* dan *Genetic Algorithm* sebagai metode pemilihan fitur akan diterapkan untuk mengklasifikasikan teks pada komentar dari review pengiriman barang.

BAB 2 LANDASAN TEORI

Text Mining

Pendekatan manual *text mining* secara intensif dalam laboratorium pertama muncul pada pertengahan 1980-an, namun kemajuan teknologi telah memungkinkan ranah tersebut untuk berkembang selama dekade terakhir. *Text mining* adalah bidang interdisipliner yang mengacu pada pencarian informasi, 30 pertambangan data, pembelajaran mesin, statistik, dan komputasi linguistik. Dikarenakan kebanyakan informasi (perkiraan umum mengatakan lebih dari 80%) saat ini disimpan sebagai teks, *text mining* diyakini memiliki potensi nilai komersial tinggi. Text mining adalah proses ekstraksi pola (informasi dan pengetahuan yang berguna) dari sejumlah besar data tak terstruktur. Masukkan untuk pengembangan teks adalah data yang tidak (atau kurang) terstruktur, seperti dokumen Word, PDF, kutipan teks, sedangkan masukkan untuk pengembangan data adalah data yang terstruktur (Feldman, 2007)

Review Pengiriman Barang

Ulasan atau review yang terdapat di internet sangat banyak namun tidak diolah menjadi sebuah informasi yang bermanfaat. Kini konsumen semakin meningkat sehingga mereka dapat memberikan opini dan pengalaman yang tersedia secara online (Zhang et al., 2011)

Algoritma Support Vector Machine

SVM disampaikan pada tahun 1992 oleh Vladimir Vapnik dan rekannya, Bernhard Boser dan Isabelle Guyon, meskipun dasar untuk SVM telah ada sejak 1960-an. Metode ini menjadikan SVM sebagai metode baru yang menjanjikan untuk klasifikasi data, baik linier maupun nonlinier. Langkah awal suatu algoritma SVM adalah pendefinisian persamaan suatu *hyperplane* pemisah yang dituliskan dengan:

$$W \cdot X + b = 0 \dots\dots\dots 1)$$

W adalah suatu bobot vektor, yaitu $W = \{ W_1, W_2, \dots, W_n \}$; n adalah jumlah atribut dan b merupakan suatu skalar yang disebut dengan bias. Jika berdasarkan pada atribut A_1, A_2 dengan permisalan tupel pelatihan $X = (x_1, x_2)$, x_1 dan x_2 merupakan nilai dari atribut A_1 dan A_2 , dan jika b dianggap sebagai suatu bobot tambahan w_0 , maka persamaan suatu *hyperplane* pemisah dapat ditulis ulang sebagai berikut:

$$w_0 + w_1 x_1 + w_2 x_2 = 0 \dots\dots\dots 2)$$

Pemilihan Fitur

Feature selection adalah sebuah proses yang bias digunakan pada machine learning dimana sekumpulan dari features yang dimiliki data digunakan untuk pembelajaran algoritma. Subset yang baik memiliki sedikitnya dimensi angka yang paling banyak berkontribusi untuk akurasi dan nantinya akan dibuang sisa dari dimensi yang tidak berkepentingan (Wahyuni &

Luthfiarta, 2004). *Forward selection* dimulai tanpa variabel dan menambahkan mereka satu persatu, pada setiap langkah ditambahkan variable yang menurunkan error paling banyak, sampai semua error dihilangkan.

Principal Component Analysis

Penggunaan seleksi fitur sangat berguna untuk mencegah terjadinya efek dimensionalitas, mengurangi jumlah waktu dan memori yang dibutuhkan oleh algoritme yang digunakan, memudahkan dalam memvisualisasikan data, dan membantu mengurangi fitur-fitur yang tidak relevan (derau) tanpa kehilangan informasi dari data aslinya (Hermawati, 2013). PCA mentransformasikan sejumlah besar variabel yang berkorelasi, menjadi beberapa variabel yang tidak berkorelasi tanpa menghilangkan informasi penting di dalamnya. Cara untuk mendapatkan *principal component* adalah dengan mencari nilai *eigenvalue* dan *eigenvector* dari matriks *covariance*-nya (Santosa, 2007). Kovarian digunakan untuk mengukur bagaimana sebaran data dengan dimensi yang bervariasi terhadap nilai tengahnya yang berkaitan dengan hubungan antar data (Hermawati, 2013). Nilai covarian didapatkan menggunakan persamaan:

$$cov(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)}$$

Genetic Algorithm

Algoritma Genetika biasa digunakan untuk klasifikasi dan juga masalah optimisasi. Metode ini diawali dengan data set solusi atau populasi. Individual pada suatu populasi merepresentasikan solusi untuk menyelesaikan suatu masalah. Individu pada suatu populasi disebut *chromosom*, yang merepresentasikan suatu solusi atas permasalahan. *Chromosom* yang sesuai dengan tujuan memiliki kemungkinan tertinggi untuk dipilih. Setelah beberapa generasi, *Genetic Algorithm* menghasilkan kromosom-kromosom terbaik yang diharapkan dapat menjadi solusi optimal atau suboptimal atas permasalahan untuk optimisasi.

Evaluasi dan Validasi

Untuk melakukan evaluasi pada algoritma *Support Vector Machine* dan algoritma *Support Vector Machine* yang sudah dioptimasi dengan *Principal Component Analysis* dan *Genetic Algorithm* dilakukan beberapa pengujian menggunakan *confusion matrix* dan kurva ROC (*receiver operating characteristic*).

BAB 3 METODE PENELITIAN

Penjelasan mengenai tahapan-tahapan yang dilakukan adalah:

1. Pengumpulan data (*Data gathering*)
Tahapan ini menjelaskan tentang sumber data untuk penelitian ini, cara pengumpulan data dan bagaimana cara pengolahan data yang sudah dikumpulkan agar dapat dipergunakan dalam eksperimen. Data diambil dari satu *website* yang menyediakan *review* untuk berbagai kategori. Data *review* dari *website* tersebut diseleksi dari data yang tidak sesuai dan dikumpulkan ke dalam aplikasi notepad untuk diolah dalam pengujian data.
2. Pengolahan data awal (*Data pre-processing*)
Tahapan ini menjelaskan bagaimana pengolahan data awal untuk mendapatkan data yang sudah ditransformasikan, agar sesuai dengan bentuk yang diinginkan pada metode yang diusulkan serta dapat digunakan dalam proses pembelajaran (*training*) dan pengujian (*testing*). Pengolahan data awal bertujuan untuk mengurangi data yang tidak relevan, atau data dengan atribut yang hilang. Pengolahan juga dapat berupa konversi nilai-nilai redundan atau nilai yang terlalu beragam kedalam kelompok yang lebih kecil untuk mempermudah pembentukan model. Model yang digunakan pada penelitian ini adalah algoritma Support Vector Machine. Algoritma Support Vector Machine dipilih berdasarkan metode yang paling baik dari beberapa metode pengklasifikasian teks yang sudah digunakan oleh beberapa peneliti sebelumnya.
3. Metode yang diusulkan (*Proposed Method*)
Pada penelitian ini penulis mengusulkan metode Support Vector Machine dengan penambahan optimasi agar dapat meningkatkan nilai akurasi. Penambahan optimasi dilakukan dengan menggabungkan metode pemilihan fitur *Principal Component Analysis* (PCA) dan *Genetic Algorithm* (GA).
4. Eksperimen dan pengujian metode (*Experiment and testing method*)
Tahapan ini menjelaskan bagaimana eksperimen dan pengujian model dilakukan. Eksperimen dilakukan dengan menggunakan RapidMiner 5.3 untuk mengolah data sehingga menghasilkan nilai akurasi yang tinggi. Pengujian metode dilakukan melalui aplikasi yang dibuat dengan bahasa pemrograman PHP dan HTML.
5. Evaluasi dan validasi hasil (*Evaluation and validation result*)
Tahapan ini merupakan tahap akhir dari penelitian untuk melakukan evaluasi hasil dari experiment dan pengujian model yang digunakan oleh penulis. Evaluasi berfungsi untuk mengetahui akurasi dari model algoritma yang diusulkan. Akurasi algoritma akan diukur menggunakan *Confusion Matrix* dan akan ditampilkan dalam bentuk *Curve ROC* (*Receiver Operating Characteristic*). Validasi digunakan untuk melihat perbandingan hasil akurasi dari model yang digunakan dengan hasil yang telah ada sebelumnya. Teknik validasi yang digunakan adalah *Cross Validation*.

BAB 4

HASIL PENELITIAN DAN PEMBAHASAN

Pada bagian ini akan dijelaskan hasil dari pengukuran kinerja model dan validasi hasil penelitian. Hasil pengukuran kinerja model akan menjabarkan pengukuran kinerja model untuk analisa sentimen review pengiriman barang dengan menerapkan *Principal Component Analysis* dan *Genetic Algorithm* sebagai pemilihan fitur menggunakan algoritma *Support Vector Machine*. Data training yang digunakan pada saat pengujian data diambil dari www.reviewcentre.com. Pengujian data, dilakukan dengan 300 *data training* dengan merujuk pada penelitian sebelumnya (Puspaningrum dan Syidada, 2010), yang terdiri dari 150 review negatif dan 150 review positif. Kemudian dilakukan *training dan testing dataset* sehingga didapatkan nilai *accuracy* dan nilai AUC.

Tabel 4.1 Grid Search Nilai Akurasi Untuk SVM (%)

C	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.0	78.67	81.33	79.33	79.67	82.33	79.33	82.00	82.33	83.33	83.00	50.00
0.1	86.00	85.33	84.67	83.67	83.33	83.00	83.33	84.33	83.00	83.00	50.00
0.2	86.00	85.33	84.67	83.67	83.33	83.00	83.33	84.33	83.00	83.00	50.00
0.3	86.00	85.33	84.67	83.67	83.33	83.00	83.33	84.33	83.00	83.00	50.00
0.4	86.00	85.33	84.67	83.67	83.33	83.00	83.33	84.33	83.00	83.00	50.00
0.5	86.00	85.33	84.67	83.67	83.33	83.00	83.33	84.33	83.00	83.00	50.00
0.6	86.00	85.33	84.67	83.67	83.33	83.00	83.33	84.33	83.00	83.00	50.00
0.7	86.00	85.33	84.67	83.67	83.33	83.00	83.33	84.33	83.00	83.00	50.00
0.8	86.00	85.33	84.67	83.67	83.33	83.00	83.33	84.33	83.00	83.00	50.00
0.9	86.00	85.33	84.67	83.67	83.33	83.00	83.33	84.33	83.00	83.00	50.00
1.0	86.00	85.33	84.67	83.67	83.33	83.00	83.33	84.33	83.00	83.00	86.00

Sumber: Hasil Penelitian (2015)

Pada penelitian ini *Principal Component Analysis* diterapkan ke dalam algoritma *Support Vector Machine*. Pada RapidMiner, *Principal Component Analysis* memiliki parameter *variance threshold* yang dapat diubah untuk meningkatkan akurasi. Perubahan nilai parameter *threshold* mempengaruhi nilai akurasi yang dihasilkan. Disini parameter *Genetic Algorithm* masih menggunakan nilai default yang ada. Adapun hasil dari perubahan nilai akurasi yang dihasilkan setelah perubahan parameter *threshold* disajikan dalam Tabel 4.2.

Tabel 4.2. Hasil eksperimen setelah menggunakan *Principal Component Analysis*

SVM		PCA	GA					SVM + PCA + GA	
C	Epsilon	Threshold	Population	Maximum Number	P initialize	P crossover	P generate	Accuracy	AUC
0.0	0.0	1.0	5	30	0.5	0.5	0.1	95.67	0.983
1.0	0.0	1.0	5	30	0.5	0.5	0.1	95.33	0.984
0.0	1.0	1.0	5	30	0.5	0.5	0.1	50.00	0.500
1.0	1.0	1.0	5	30	0.5	0.5	0.1	50.00	0.500
0.0	0.0	0.9	5	30	0.5	0.5	0.1	95.33	0.986
1.0	0.0	0.9	5	30	0.5	0.5	0.1	96.00	0.987
0.0	1.0	0.9	5	30	0.5	0.5	0.1	50.00	0.500
1.0	1.0	0.9	5	30	0.5	0.5	0.1	50.00	0.500
0.0	0.0	0.8	5	30	0.5	0.5	0.1	95.33	0.986
1.0	0.0	0.8	5	30	0.5	0.5	0.1	97.00	0.990
0.0	1.0	0.8	5	30	0.5	0.5	0.1	50.00	0.500
1.0	1.0	0.8	5	30	0.5	0.5	0.1	50.00	0.500
0.0	0.0	0.7	5	30	0.5	0.5	0.1	95.33	0.986
1.0	0.0	0.7	5	30	0.5	0.5	0.1	96.00	0.987
0.0	1.0	0.7	5	30	0.5	0.5	0.1	50.00	0.500
1.0	1.0	0.7	5	30	0.5	0.5	0.1	50.00	0.500
0.0	0.0	0.6	5	30	0.5	0.5	0.1	95.33	0.986
1.0	0.0	0.6	5	30	0.5	0.5	0.1	96.00	0.987
0.0	1.0	0.6	5	30	0.5	0.5	0.1	50.00	0.500
1.0	1.0	0.6	5	30	0.5	0.5	0.1	50.00	0.500
0.0	0.0	0.5	5	30	0.5	0.5	0.1	95.33	0.986
1.0	0.0	0.5	5	30	0.5	0.5	0.1	96.00	0.987
0.0	1.0	0.5	5	30	0.5	0.5	0.1	50.00	0.500
1.0	1.0	0.5	5	30	0.5	0.5	0.1	50.00	0.500
0.0	0.0	0.4	5	30	0.5	0.5	0.1	95.00%	0.987
1.0	0.0	0.4	5	30	0.5	0.5	0.1	92.33%	0.974

Sumber: Hasil Penelitian (2015)

Eksperimen dengan menggunakan *Genetic Algorithm* dimulai dengan melakukan *adjustment* pada nilai *maximum number of generation*, yaitu dimulai dari 10-100 dengan kelipatan nilai 10, untuk menentukan jumlah generasi yang menghasilkan akurasi paling tinggi. Ketika *maximum of number generation* di-*adjustment*, nilai *population*, *p initialize*, *p crossover*, dan *p mutation* berada pada nilai default, yaitu 5 untuk *population*, 0.5 untuk *p mutation*, 0.5 untuk *p crossover*, dan 0.1 untuk *p mutation*. Setelah didapatkan jumlah maksimal generasi yang menghasilkan akurasi paling tinggi, kemudian dilanjutkan dengan melakukan *adjustment* pada nilai *pop population* yang dimulai dari 5-100 dengan kelipatan nilai 5. Nilai *population* yang menghasilkan akurasi paling tinggilah yang akan digunakan pada langkah percobaan selanjutnya. Setelah itu, dilakukan *adjustment* pada *p initialize* dengan range 0.1-1.0 dengan kelipatan nilai 0.1. Dilanjutkan dengan *adjustment* pada *p crossover* dengan range 0.1-1.0 dengan kelipatan nilai 0.1. Dan yang terakhir dilakukan *adjustment* pada *p mutation* dengan range 1.0-1.0 dengan kelipatan nilai 0.1.

Berikut *adjustment* pada nilai *maximum number of generation*.

Tabel 4.3. Adjustment pada nilai *maximum number of generation*

SVM		PCA	GA					SVM + PCA + GA	
C	Epsilon	Threshold	Maximum Number of Population	Population	P initialize	P cross over	P mutation	Accuracy	AUC
0.0	0.0	0.3	10	5	0.5	0.5	0.1	96.67	0.993
0.0	0.0	0.3	20	5	0.5	0.5	0.1	96.67	0.993
0.0	0.0	0.3	30	5	0.5	0.5	0.1	97.67	0.993
0.0	0.0	0.3	40	5	0.5	0.5	0.1	97.33	0.993
0.0	0.0	0.3	50	5	0.5	0.5	0.1	97.00	0.992
0.0	0.0	0.3	60	5	0.5	0.5	0.1	97.33	0.993
0.0	0.0	0.3	70	5	0.5	0.5	0.1	97.00	0.992
0.0	0.0	0.3	80	5	0.5	0.5	0.1	96.00	0.993
0.0	0.0	0.3	90	5	0.5	0.5	0.1	96.00	0.993
0.0	0.0	0.3	100	5	0.5	0.5	0.1	96.00	0.993

Sumber: Hasil Penelitian (2015)

Akurasi tertinggi dicapai pada saat jumlah generasi 30. Nilai akurasi dan AUC yang diperoleh adalah 97.675% dan 0.993, Semakin banyak jumlah generasi, maka waktu komputasi yang dibutuhkan juga akan semakin bertambah. Nilai *maximum number of generation* yang digunakan untuk percobaan selanjutnya adalah 30. Selanjutnya dilakukan *adjustment* pada nilai *population* mulai dari 5-100.

Tabel 4.4. Adjustment pada nilai *population*

SVM		PCA	GA					SVM + PCA + GA	
C	Epsilon	Threshold	Maximum Number of Population	Population	P initialize	P cross over	P mutation	Accuracy	AUC
0.0	0.0	0.3	30	5	0.5	0.5	0.1	97.67	0.993
0.0	0.0	0.3	30	10	0.5	0.5	0.1	97.33	0.988
0.0	0.0	0.3	30	15	0.5	0.5	0.1	97.33	0.991
0.0	0.0	0.3	30	20	0.5	0.5	0.1	97.67	0.992
0.0	0.0	0.3	30	25	0.5	0.5	0.1	97.33	0.989
0.0	0.0	0.3	30	30	0.5	0.5	0.1	97.67	0.992
0.0	0.0	0.3	30	35	0.5	0.5	0.1	98.00	0.993
0.0	0.0	0.3	30	40	0.5	0.5	0.1	97.33	0.992
0.0	0.0	0.3	30	45	0.5	0.5	0.1	97.67	0.993
0.0	0.0	0.3	30	50	0.5	0.5	0.1	97.67	0.993
0.0	0.0	0.3	30	55	0.5	0.5	0.1	97.67	0.993
0.0	0.0	0.3	30	60	0.5	0.5	0.1	97.67	0.992

Sumber: Hasil Penelitian (2015)

Berdasarkan proses *adjustment* pada nilai *population* di atas, diperoleh akurasi tertinggi yang tercapai adalah 98,33% dengan nilai AUC 0,992. Akurasi ini dicapai oleh nilai *population* 95. Kemudian berikutnya akan dilakukan *adjustment* pada nilai *probability initialize* mulai dari 0.1 hingga 1.0.

Tabel 4.5. Adjustment pada nilai *p initialize*

SVM		PCA	GA					SVM + PCA + GA	
C	Epsilon	Threshold	Maximum Number of Population	Population	P initialize	P crossover	P mutation	Accuracy	AUC
0.0	0.0	0.3	30	95	0.1	0.5	0.1	97.00	0.992
0.0	0.0	0.3	30	95	0.2	0.5	0.1	97.67	0.992
0.0	0.0	0.3	30	95	0.3	0.5	0.1	97.67	0.994
0.0	0.0	0.3	30	95	0.4	0.5	0.1	98.33	0.992
0.0	0.0	0.3	30	95	0.5	0.5	0.1	98.33	0.992
0.0	0.0	0.3	30	95	0.6	0.5	0.1	97.67	0.995
0.0	0.0	0.3	30	95	0.7	0.5	0.1	97.67	0.993
0.0	0.0	0.3	30	95	0.8	0.5	0.1	97.67	0.994
0.0	0.0	0.3	30	95	0.9	0.5	0.1	97.67	0.991
0.0	0.0	0.3	30	95	1.0	0.5	0.1	98.00	0.994

Sumber: Hasil Penelitian (2015)

Pada proses *adjustment* nilai *p initialize* di atas, akurasi tertinggi yang tercapai adalah 98.33% dengan nilai AUC 0,992. Akurasi ini dicapai oleh nilai *p initialize* 0,4 dan 0,5. Semakin tinggi nilai *p initialize* waktu komputasi yang dibutuhkan juga semakin lama. Karena ke-empat nilai *p initialize* memiliki akurasi yang sama, maka dipilih nilai *p initialize* terkecil yaitu 0,4 untuk eksperimen selanjutnya. Kemudian dilakukan *adjustment* pada nilai *probability crossover* mulai dari 0.1 hingga 1.0.

Tabel 4.6. Adjustment pada nilai *p crossover*

SVM		PCA	GA					SVM + PCA + GA	
C	Epsilon	Threshold	Maximum Number of Population	Population	P initialize	P crossover	P mutation	Accuracy	AUC
0.0	0.0	0.3	30	95	0.4	0.1	0.1	97.67	0.995
0.0	0.0	0.3	30	95	0.4	0.2	0.1	98.00	0.991
0.0	0.0	0.3	30	95	0.4	0.3	0.1	97.67	0.990
0.0	0.0	0.3	30	95	0.4	0.4	0.1	97.67	0.992
0.0	0.0	0.3	30	95	0.4	0.5	0.1	98.33	0.992
0.0	0.0	0.3	30	95	0.4	0.6	0.1	98.00	0.993
0.0	0.0	0.3	30	95	0.4	0.7	0.1	97.67	0.991
0.0	0.0	0.3	30	95	0.4	0.8	0.1	98.33	0.994
0.0	0.0	0.3	30	95	0.4	0.9	0.1	98.00	0.996
0.0	0.0	0.3	30	95	0.4	1.0	0.1	98.33	0.992

Sumber: Hasil Penelitian (2015)

Dari hasil *adjustment* nilai *p crossover* di atas, diperoleh nilai *p crossover* 0,5, 0,8, dan 1,0 yang dapat menghasilkan akurasi paling tinggi yaitu sebesar 98,33 %. Namun AUC yang dihasilkan berbeda. AUC paling tinggi dihasilkan saat nilai *p crossover* 0,8. Oleh karena itu untuk langkah selanjutnya adalah melakukan *adjustment* pada nilai *probability mutation* dengan menggunakan nilai *p crossover* 0,8. Berikut adalah hasil *adjustment* pada nilai *p mutation* dari 0,1 hingga 1,0:

Tabel 4.7. Adjustment pada nilai *p mutation*

SVM		PCA	GA					SVM + PCA + GA	
C	Epsilon	Threshold	Maximum Number of Population	Population	P initial size	P crossover	P mutation	Accuracy	AUC
0.0	0.0	0.3	30	95	0.4	0.8	0.1	98.33	0.994
0.0	0.0	0.3	30	95	0.4	0.8	0.2	98.33	0.994
0.0	0.0	0.3	30	95	0.4	0.8	0.3	98.33	0.994
0.0	0.0	0.3	30	95	0.4	0.8	0.4	98.33	0.994
0.0	0.0	0.3	30	95	0.4	0.8	0.5	98.33	0.994
0.0	0.0	0.3	30	95	0.4	0.8	0.6	98.33	0.994
0.0	0.0	0.3	30	95	0.4	0.8	0.7	98.33	0.994
0.0	0.0	0.3	30	95	0.4	0.8	0.8	98.33	0.994
0.0	0.0	0.3	30	95	0.4	0.8	0.9	98.33	0.994
0.0	0.0	0.3	30	95	0.4	0.8	1.0	98.33	0.994

Sumber: Hasil Penelitian (2015)

Pada proses *adjustment* pada nilai *p mutation* di atas, diperoleh akurasi yang sama untuk seluruh nilai parameter yang diambil sesuai range 0,1-1,0. Akurasi yang dihasilkan adalah 98.33% dengan nilai AUC 0,994. Semakin tinggi nilai *p mutation*, waktu komputasi yang dibutuhkan juga semakin lama. Karena seluruh nilai *p mutation* memiliki akurasi yang sama, maka dipilih nilai *p mutation* terkecil yaitu 0,1. Hasil pengujian data dengan Metode *Support Vector Machine* berbasis *Principal Component Analysis* dan *Genetic Algorithm* menghasilkan Akurasi 98,33% dan AUC sebesar 0.994. Hal ini menunjukkan bahwa dengan menggunakan optimasi *Principal Component Analysis* dan *Genetic Algorithm* dapat meningkatkan akurasi yang lebih baik.

Untuk melakukan akurasi pada Algoritma *Support Vector Machine* dilakukan pengujian dengan menggunakan Kurva ROC dan *Confusion Matrix*.

Kurva ROC

Kurva ROC (*Receiver Operating Characteristic*) adalah cara lain untuk mengevaluasi akurasi dari klasifikasi secara visual. Sebuah grafik ROC adalah plot dua dimensi dengan proporsi positif salah pada sumbu X dan positif benar pada sumbu Y. Hasil perhitungan pada kurva ROC, menggambarkan kurva ROC untuk algoritma *Support Vector Machine* dengan penambahan pemilihan fitur *Principal Component Analysis* dan *Genetic Algorithm*. Dapat disimpulkan

bahwa satu point pada kurva ROC adalah lebih baik dari pada yang lainnya jika arah garis melintang dari kiri bawah ke kanan atas didalam grafik. Kurva ROC *Support Vector Machine* memiliki **AUC** sebesar **0.994** yang termasuk ke dalam kategori *excellent classification*. Berikut dapat dilihat Kurva ROC *Support Vector Machine* pada Gambar 4.8.



Sumber: Hasil Penelitian (2015)

Gambar 4.1 Kurva ROC SVM + PCA + GA

Keterangan:

Akurasi 0.90 – 1.00 = *Excellent classification*

Akurasi 0.80 – 0.90 = *Good classification*

Akurasi 0.70 – 0.80 = *Fair classification*

Akurasi 0.60 – 0.70 = *Poor classification*

Akurasi 0.50 – 0.60 = *Failure*

Confusion Matrix

Memberikan keputusan yang diperoleh dalam *training* dan *testing*, *confusion matrix* memberikan penilaian *performance* klasifikasi berdasarkan objek benar atau salah. *Confusion matrix* berisi informasi aktual (*actual*) dan prediksi (*predicted*) pada sistem klasifikasi. Berikut merupakan persamaan model *Confusion Matrix* untuk menghitung akurasi pada percobaan *Support Vector Machine* sebelum adanya penyesuaian nilai parameter dan sebelum dilakukan optimasi dengan pemilihan fitur.

Tabel 4.7 Confusion Matrix Support Vector Machine

CLASSIFICATION	PREDICTED CLASS	
	Class=Yes	Class=No
Class=Yes	148	4
Class=No	2	146

Sumber: Hasil Penelitian (2015)

$$\text{Acc (Accuracy)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \frac{148 + 146}{148 + 2 + 4 + 146} = \frac{294}{300} = 0.98$$

Akurasi yang diperoleh yaitu 97.00% dari 150 data review negatif dan 150 data review positif. Data review yang dihasilkan RapidMiner dengan model SVM + PCA + GA menunjukkan bahwa klasifikasi untuk review positif yang sesuai prediksi yaitu **148**. Data review negatif yang termasuk ke dalam prediksi positif yaitu **2**. Data review positif yang termasuk ke dalam prediksi negatif yaitu **4** dan data review negatif yang termasuk ke dalam prediksi negatif yaitu **146**.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{148}{148 + 4} = \frac{148}{152} = 0.973$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} = \frac{146}{146 + 2} = \frac{146}{148} = 0.986$$

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{148}{148 + 2} = \frac{148}{150} = 0.986$$

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}} = \frac{146}{146 + 4} = \frac{146}{150} = 0.973$$

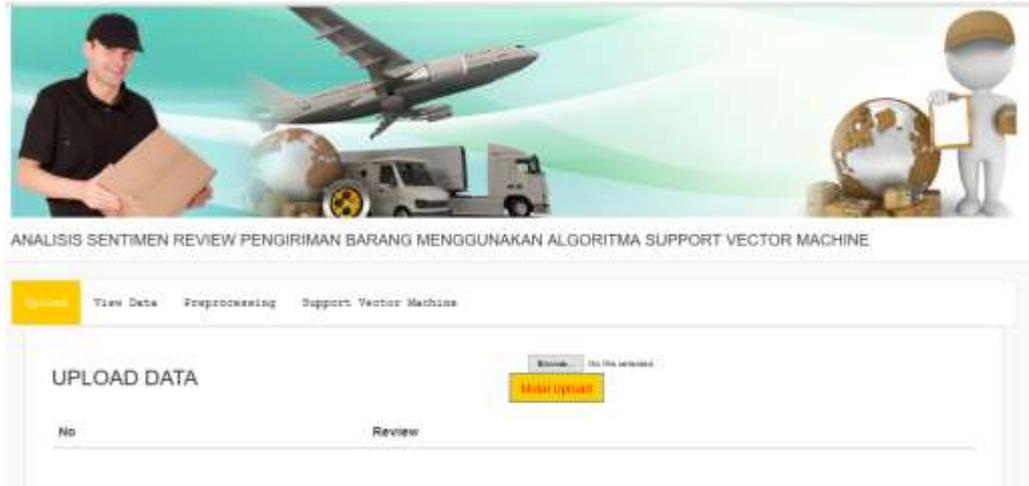
Pengembangan Aplikasi

Selanjutnya dari hasil ekeperimen yang sudah dilakukan, maka dibuatkan model aplikasi analisis sentimen *review* pengiriman barang. Data review bersifat tidak terstruktur dan memiliki banyak *noise*, oleh karena itu dibutuhkan suatu aplikasi untuk memproses data tersebut menjadi data terstruktur sehingga nantinya akan menjadi informasi yang berguna. Berikut tampilan aplikasi yang dirancang:



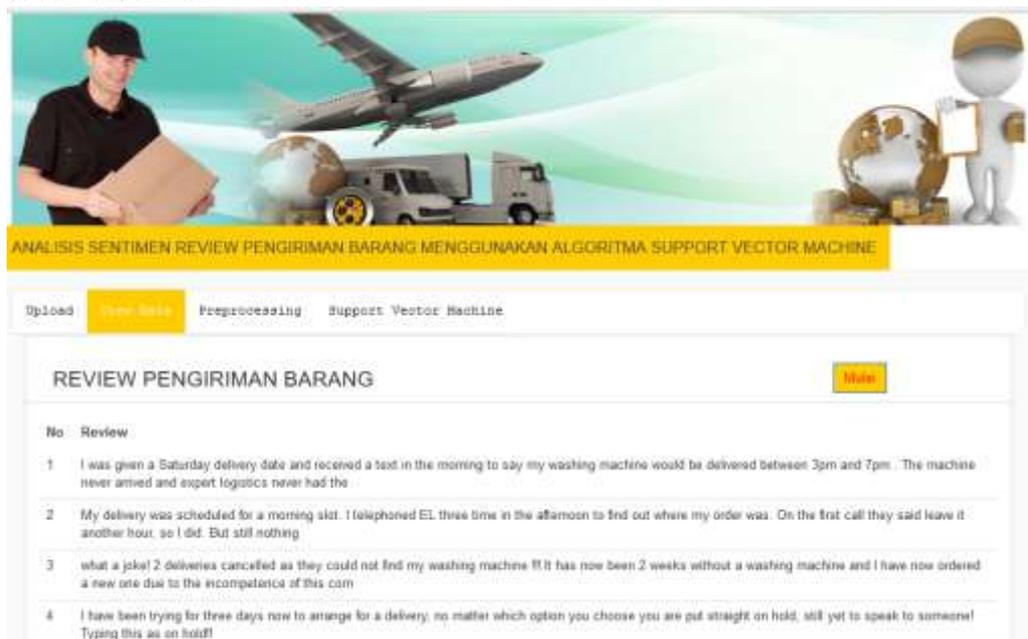
Gambar 4.2 Tampilan awal *website* analisis *review* pengiriman barang

Tampilan diatas menunjukkan petunjuk penggunaan web sesuai dengan menu yang disediakan.



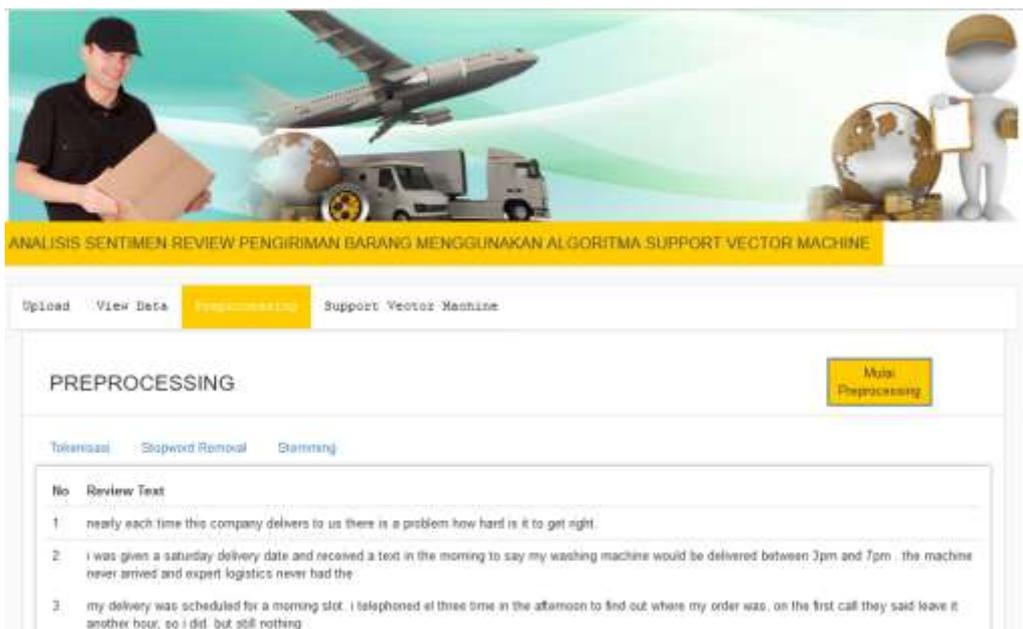
Gambar 4.2 Tampilan Menu Upload

Menu Upload digunakan untuk meng-*input* data review yang akan dianalisa sehingga dihasilkan klasifikasi positif dan negative dari review yang sudah diberikan.



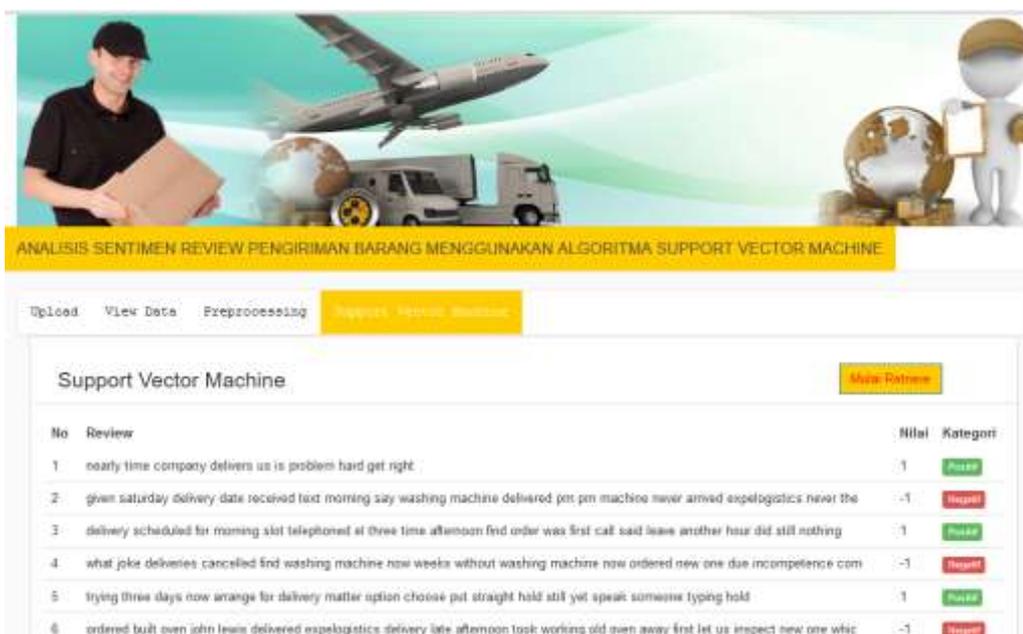
Gambar 4.3 Tampilan Menu View Data

Menu View Data digunakan untuk menampilkan data review yang akan sudah diinput pada menu upload. Untuk menampilkan data, gunakan tombol mulai yang ada di kanan atas website.



Gambar 4.4 Tampilan Menu Preprocessing

Menu *Support Vector Machine* menampilkan hasil klasifikasi dari review yang suda diinput. Kategori menunjukkan apakah review tersebut masuk ke dalam review negative atau positif. Nilai 1 menunjukkan bahwa review yang diinput masuk ke dalam kategori positif, sedangkan nilai 0 menunjukkan bahwa review tersebut masuk ke dalam kategori negatif.



Gambar 4.5 Tampilan Menu Support Vector Machine

BAB 5 KESIMPULAN DAN SARAN

Kesimpulan

Semakin besar nilai parameter C yang dipengaruhi oleh nilai ϵ pada *Support Vector Machine*, maka performa cenderung makin menurun. Penelitian ini menunjukkan bahwa untuk mengklasifikasikan teks dengan data berupa review pengiriman barang, salah satu pengklasifikasi yang dapat digunakan adalah *Support Vector Machine*. *Support Vector Machine* sangat populer digunakan untuk klasifikasi teks dan memiliki performa yang baik pada banyak domain. Penelitian ini menghasilkan akurasi pada algoritma *Support Vector Machine* 78,67% dan AUC sebesar 0.975. Hasil akurasi ini didapatkan hanya dengan menggunakan algoritma *Support Vector Machine* tanpa adanya penyesuaian parameter maupun penambahan optimasi. Kemudian dibandingkan dengan *Support Vector Machine* yang dioptimasi dengan *Principal Component Analysis* dan *Genetic Algorithm* dengan akurasi 98,33% dan AUC sebesar 0.994. Pada penelitian ini nilai parameter yang digunakan untuk *Support Vector Machine* adalah 0.0, baik untuk parameter C maupun parameter ϵ . Sedangkan nilai parameter threshold untuk *Principal Component Analysis* adalah 0,3. Dan untuk *Genetic algorithm*, nilai parameter yang digunakan adalah *Maximum Number of Generation* dengan nilai 30, *Population* dengan nilai 95, p initialize dengan nilai 0,4, p crossover dengan nilai 8,8, dan p mutation dengan nilai 0,1. Model *Support Vector Machine* berbasis *Principal Component Analysis* dan *Genetic Algorithm* dapat memberikan solusi terhadap permasalahan klasifikasi review negatif maupun positif pada pengiriman barang agar lebih akurat dan optimal. Model ini dapat digunakan untuk mengetahui hasil klasifikasi pada review, apakah review tersebut termasuk ke dalam review negatif atau review positif. Hal ini juga dapat membantu seseorang untuk menghemat waktu saat mencari review pengiriman barang dengan pelayanan yang baik sesuai reviewnya.

Saran

Metode *Support Vector Machine* sudah banyak digunakan dan mempunyai performa yang baik dalam klasifikasi teks, namun ada beberapa hal yang dapat ditambahkan untuk penelitian selanjutnya:

1. Menggunakan pemilihan fitur lain, seperti *Chi Square*, *Mutual Information*, *Information Gain*, dan lain-lain. Penggunaan pemilihan fitur lain dapat dilakukan dengan cara penggabungan ataupun terpisah. Hasil dari penggunaan pemilihan fitur lain dapat dibandingkan dengan metode yang sudah digunakan.
2. Menggunakan pengklasifikasi lain yang mungkin di luar *Supervised learning*. Sehingga bisa dilakukan penelitian yang berbeda dari umumnya yang sudah ada.
3. Menggunakan data review dari domain yang berbeda, misalnya review film, restoran, review produk, dan lain sebagainya.

DAFTAR REFERENSI

- Basari, A. S. H., Hussin, B., Ananta, I. G. P., & Zeniarja, J. 2013. Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization. *Procedia Engineering*, 453–462. <http://doi.org/10.1016/j.proeng.2013.02.059>
- Chandani, V. 2015. Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film, *Journal of Intelligent Systems*, 55–59.
- Dawson, C. W. 2009. *Projects in Computing and Information Systems*. England: Pearson Education Limited.
- Elkan, C. 2013. Predictive analytics and data mining. Technical report, Cato Institute.
- Feldman. 2007. *The Text Mining Handbook*. New York: Cambridge Univesity Press.
- Ghiassi, M., Skinner, J., & Zimbra, D. 2013. Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with Applications*, 40(16), 6266–6282. <http://doi.org/10.1016/j.eswa.2013.05.057>
- Gorunescu, F. 2011. *Data mining: concepts and techniques*. Springer-Verlag Berlin Heidelberg. <http://doi.org/10.1007/978-3-642-19721-5>
- Gunal. 2012. Hybrid feature selection for text classification, *turk J Elec Eng & Comp Sci*, 20. <http://doi.org/10.3906/elk-1101-1064>
- Kontopoulos, E., Berberidis, C., Dergiades, T., & Bassiliades, N. 2013. Ontology-based sentiment analysis of twitter posts. *Expert Systems with Applications*, 40(10), 4065–4074. <http://doi.org/10.1016/j.eswa.2013.01.001>
- Kurniawan. 2013. Optimasi Algoritma Support Vector Machine (SVM), *Jurnal Teknologi Informasi*, 9 (April), 38–49.
- Liu, B. 2010. Sentiment Analysis and Subjectivity, 1–38. Chicago: Handbook of Natural Language Processing
- Liu, B. 2012. Sentiment Analysis and Opinion Mining. Chicago: Morgan & Claypool Publishers.
- Made, N., Lestari, A., Gede, I. K., Putra, D., Ketut, A. A., & Cahyawan, A. 2013. Personality Types Classification for Indonesian Text in Partners Searching Website Using Naïve Bayes Methods, *10(1)*, 1–8.

- Miner, G., Delen, D. 2012. The Seven Practice Areas of Text Analytics, (January), 29–41. Elsevier Inc. <http://doi.org/10.1016/B978-0-12-386979-1.00002-5>
- Moraes, R., Valiati, J. F., & Gavião Neto, W. P. 2013. Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Systems with Applications*, 40(2), 621–633. <http://doi.org/10.1016/j.eswa.2012.07.059>
- Morchid, M., Dufour, R., Bousquet, P.-M., Linarès, G., & Torres-Moreno, J.-M. 2014. Feature selection using Principal Component Analysis for massive retweet detection. *Pattern Recognition Letters*, 49, 33–39. <http://doi.org/10.1016/j.patrec.2014.05.020>
- Noviati. 2015. Analisis Pengaruh Seleksi Fitur pada Klasifikasi Konsentrasi. Semarang. Jurnal Nasional Sains dan Teknologi.
- Pang, B., & Lee, L. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*. 2 (1).
- Pratama, E. E., & Trilaksono, B. R. 2014. Klasifikasi Keluhan Pelanggan Berdasarkan Tweet dengan Menggunakan Metode Support Vector Machine (SVM). Bandung: Jurnal Teknik Elektro dan Informatika.
- Puspaningrum, E. Y., & Syidada, S. 2010. Reduksi Fitur Untuk Kategorisasi Text Dengan, 2–5. Surabaya: Jurnal Teknik Informatika.
- Puspitaningrum. 2014. Dampak Reduksi Sampel Menggunakan Principal Component Analysis (PCA) Pada Pelatihan Jaringan Saraf Tiruan Terawasi - 2014. Bengkulu: Jurnal Pseudocode.
- Sartono. 2010. Pengenalan Algoritma Genetik Untuk Pemilihan Peubah Penjelas Dalam Model Regresi Menggunakan SAS / IML (An Introduction to Variables Selection in Regression Analysis using Genetic Algorithm with SAS / IML) Bagus Sartono Peneliti di Data Mining Center – , 15(2), 10–15.
- Singh, P., & Pandi, G. S. 2015. Opinion Mining Techniques for Social Network Analysis : A Survey, 2(12), 350–354. International Journal for Scientific Research & Development.
- Swati. 2015. SVM Based Improvement In KNN For Text. *International Journal of Engineering Research and General Science*. 3(4), 214–220.
- Tan, S., & Zhang, J. 2008. An empirical study of sentiment analysis for chinese documents. *Expert Systems with Applications*, 34(4), 2622–2629. <http://doi.org/10.1016/j.eswa.2007.05.028>

- Thakor, P., & Sasi, S. 2015. Ontology-based Sentiment Analysis Process for Social Media Content. *Procedia Computer Science*, 53, 199–207. <http://doi.org/10.1016/j.procs.2015.07.295>
- Vercellis, C. 2009. *Business Intelligence: Data Mining and Optimization for Decision Making*. United Kingdom: A John Wiley and Sons, Ltd., Publication.
- Wahyuni, D. T., & Luthfiarta, A. 2004. Prediksi Hasil Pemilu Legislatif DKI Jakarta Menggunakan Naïve Bayes Dengan Algoritma Genetika Sebagai Fitur Seleksi. Semarang: Jurnal Teknik Informatika FIK UDINUS.
- Wan, C. H., Lee, L. H., Rajkumar, R., & Isa, D. 2012. A hybrid text classification approach with low dependency on parameter by integrating K-nearest neighbor and support vector machine. *Expert Systems with Applications*, 39(15), 11880–11888. <http://doi.org/10.1016/j.eswa.2012.02.068>
- Wang, Z. 2015. *Enhancing Machine-Learning Methods for Sentiment Classification of Web Data* Enhancing Machine-Learning Methods for Sentiment, (September). Switzerland: Springer International Publishing.
- Wayan, N., Saraswati, S., Prodi, S., Informatika, T., Denpasar, S., Tukad, J., & No, B. 2013. Naïve Bayes Classifier Dan Support Vector Machines Untuk Sentiment Analysis, 2–4. Bali: Seminar Nasional Sistem Informasi Indonesia.
- Zhang, Z., Ye, Q., Zhang, Z., & Li, Y. 2011. Sentiment classification of Internet restaurant reviews written in Cantonese. *Expert Systems with Applications*, 38(6), 7674–7682. <http://doi.org/10.1016/j.eswa.2010.12.147>