

**PENERAPAN ALGORITMA *SUPPORT VECTOR*
MACHINE DENGAN METODE WRAPPER *GENETIC*
ALGORITHM DALAM RANGKA PEMILIHAN
FITUR UNTUK ANALISIS SENTIMEN
PADA TWITTER**



TESIS

Diajukan sebagai salah satu syarat untuk memperoleh gelar
Magister Ilmu Komputer (M.Kom)

DWI ANDINI PUTRI

14000898

**PROGRAM PASCASARJANA MAGISTER ILMU KOMPUTER
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
NUSA MANDIRI
JAKARTA
2014**

**PENERAPAN ALGORITMA *SUPPORT VECTOR
MACHINE* DENGAN METODE WRAPPER *GENETIC
ALGORITHM* DALAM RANGKA PEMILIHAN
FITUR UNTUK ANALISIS SENTIMEN
PADA TWITTER**



TESIS

Diajukan sebagai salah satu syarat untuk memperoleh gelar
Magister Ilmu Komputer (M.Kom)

DWI ANDINI PUTRI

14000898

**PROGRAM PASCASARJANA MAGISTER ILMU KOMPUTER
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
NUSA MANDIRI
JAKARTA
2014**

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini :

Nama : Dwi Andini Putri
NIM : 14000898
Program Studi : Magister Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : *Management Information System (MIS)*

Dengan ini menyatakan bahwa tesis yang telah saya buat dengan judul: "Penerapan Algoritma *Support Vector Machine* dengan Metode *Wrapper Genetic Algorithm* dalam Rangka Pemilihan Fitur untuk Analisis Sentimen pada Twitter" adalah hasil karya sendiri, dan semua sumber baik yang kutip maupun yang dirujuk telah saya nyatakan dengan benar dan tesis belum pernah diterbitkan atau dipublikasikan dimanapun dan dalam bentuk apapun.

Demikianlah surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila dikemudian hari ternyata saya memberikan keterangan palsu dan atau ada pihak lain yang mengklaim bahwa tesis yang telah saya buat adalah hasil karya milik seseorang atau badan tertentu, saya bersedia diproses baik secara pidana maupun perdata dan kelulusan saya dari Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri dicabut/dibatalkan.

Jakarta, 08 September 2014

METERAI
TEMPEL
INSTRUKSI PENCETAKAN
6000
Dwi Andini Putri

HALAMAN PENGESAHAN

Tesis ini diajukan oleh:

Nama : Dwi Andini Putri
NIM : 14000898
Program Studi : Magister Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : *Management Informations System (MIS)*
Judul Tesis : "Penerapan Algoritma Support Vector Machine dengan Metode Wrapper Genetic Algorithm dalam Rangka Pemilihan Fitur untuk Analisis Sentimen pada Twitter"

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, 08 September 2014
Pascasarjana Magister Ilmu Komputer
STMIK Nusa Mandiri
Direktur

Prof. Dr. Ir. Kaman Nainggolan, MS

DEWAN PENGUJI

Penguji I : Dr. Dana Indra Sensusu

Penguji II : Dr. Alimuddin, MM, MT

Penguji III /
Pembimbing : Dr. Ir. Prabowo Pudjo Widodo, MS

iv

Program Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri



LEMBAR KONSULTASI BIMBINGAN TESIS

PASCASARJANA SEKOLAH TINGGI MANAJEMEN INFORMATIKA & KOMPUTER NUSA MANDIRI

- N I M : 14000898
- Nama Lengkap : Dwi Andini Putri
- Dosen Pembimbing : Dr. Ir. Prabowo Pudjo Widodo, MS
- Judul Skripsi : Penerapan Algoritma *Support Vector Machine* dengan Teknik Optimasi *Genetic Algorithm* dalam Rangka Pemilihan Fitur untuk Analisis Sentimen pada Twitter

No	Tanggal Bimbingan	Pokok Bahasan	Paraf Dosen Pembimbing I
1	21 Mei 2014	Bimbingan Perdana dan Pengajuan BAB I	
2	25 Mei 2014	Revisi Bab I, Pengajuan Bab II	
3	11 Juni 2014	Revisi BAB II dan Pengujian data	
4	23 Juli 2014	Pengajuan BAB III dan Hasil pengujian	
5	14 Agustus 2014	Revisi Bab III dan Pengajuan Bab IV	
6	20 Agustus 2014	Revisi Bab IV dan Pengajuan Bab V	
7	27 Agustus 2014	Revisi ACC Keseluruhan	

Catatan untuk Dosen Pembimbing Bimbingan Tesis

- Dimulai pada tanggal : 21 Mei 2014
- Diakhiri pada tanggal : 27 Agustus 2014
- Jumlah pertemuan bimbingan : 7 kali pertemuan

Disetujui oleh,
Dosen Pembimbing

[Dr. Ir. Prabowo Pudjo Widodo, MS]

v

Program Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri

KATA PENGANTAR

Puji syukur alhamdulillah, penulis panjatkan kehadiran Allah, SWT, yang telah melimpahkan rahmat dan karunia-Nya, sehingga pada akhirnya penulis dapat menyelesaikan tesis ini tepat pada waktunya. Dimana tesis ini penulis sajikan dalam bentuk buku yang sederhana. Adapun judul tesis, yang penulis ambil sebagai berikut “Penerapan Algoritma *Support Vector Machine* dengan Metode Wrapper *Genetic Algorithm* dalam Rangka Pemilihan Fitur untuk Analisis Sentimen pada Twitter”.

Tujuan penulisan tesis ini dibuat sebagai salah satu untuk mendapatkan gelar Magister Ilmu Komputer (M.Kom) pada Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (PPs MIK STMIK Nusa Mandiri).

Tesis ini diambil berdasarkan data pada sebuah penelitian atau riset mengenai Analisis Sentimen Pada Twitter Menggunakan *Support Vector Machine* (SVM) yang penulis lakukan menggunakan data publik yang diambil dari situs halaman <http://rs.peoplebrowsr.com/>. Penulis juga lakukan mencari dan menganalisa berbagai macam sumber referensi, baik dalam bentuk jurnal ilmiah, buku-buku literatur, *internet*, dll yang terkait dengan pembahasan pada tesis ini.

Penulis menyadari bahwa tanpa bimbingan dan dukungan dari semua pihak dalam pembuatan tesis ini, maka penulis tidak dapat menyelesaikan tesis ini tepat pada waktunya. Untuk itu ijinilah penulis kesempatan ini untuk mengucapkan ucapan terima kasih yang sebesar-besarnya kepada :

1. Bapak Dr. Ir. Prabowo Pudjo Widodo, MS selaku pembimbing tesis yang telah menyediakan waktu, pikiran dan tenaga dalam membimbing penulis dalam menyelesaikan tesis ini.
2. Orang tua, kakak tercinta yang telah memberikan dukungan material dan moral kepada penulis

3. Seluruh staf pengajar (dosen) PPs MIK STMIK Nusa Mandiri yang telah memberikan pelajaran yang berarti bagi penulis selama menempuh studi.
5. Teman-Teman BSI (Bina Sarana Informatika) yang telah banyak memberikan dukungan mental.
5. Seluruh staf dan karyawan PPs MIK STMIK Nusa Mandiri yang telah melayani penulis dengan baik selama kuliah.

Serta semua pihak yang terlalu banyak untuk penulis sebutkan satu persatu sehingga terwujudnya penulisan tesis ini. Penulis menyadari bahwa penulisan tesis ini masih jauh sekali dari sempurna, untuk itu penulis mohon kritik dan saran yang bersifat membangun demi kesempurnaan penulisan karya ilmiah yang penulis hasilkan untuk yang akan datang.

Akhir kata semoga tesis ini dapat bermanfaat bagi penulis khususnya dan bagi para pembaca yang berminat pada umumnya.

Jakarta, September 2014

Dwi Andini Putri

Penulis

**SURAT PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH
UNTUK KEPENTINGAN AKADEMIS**

Yang bertanda tangan di bawah ini, saya :

Nama : Dwi Andini Putri
NIM : 14000898
Program Studi : Magsiter Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : *Management Informations System (MIS)*
Jenis Karya : Tesis

Demi pengembangan ilmu pengetahuan, dengan ini menyetujui untuk memberikan ijin kepada pihak Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri) **Hak Bebas Royalti Non-Eksklusif (*Non-exclusive Royalti-Free Right*)** atas karya ilmiah kami yang berjudul : “Penerapan Algoritma *Support Vector Machine* dengan Metode Wrapper *Genetic Algorithm* dalam Rangka Pemilihan Fitur untuk Analisis Sentimen pada Twitter” beserta perangkat yang diperlukan (apabila ada).

Dengan **Hak Bebas Royalti Non-Eksklusif** ini pihak STMIK Nusa Mandiri berhak menyimpan, mengalih-media atau *bentuk*-kan, mengelolanya dalam pangkalan data (*database*), mendistribusikannya dan menampilkan atau mempublikasikannya di *internet* atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari kami selama tetap mencantumkan nama kami sebagai penulis/pencipta karya ilmiah tersebut.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak STMIK Nusa Mandiri, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, September 2014

METERAI
TEMPEL
PALESTRA KEMENTERIAN
PENDIDIKAN DAN KEBUDAYAAN
3E30CACF590267245
ENAM RIBU RUPIAH
6000
Dwi Andini Putri

DAFTAR ISI

	Halaman
HALAMAN SAMPUL.....	i
HALAMAN JUDUL.....	ii
HALAMAN PERNYATAAN ORISINALITAS.....	iii
HALAMAN PENGESAHAN.....	iv
LEMBAR KONSULTASI.....	v
KATA PENGANTAR.....	vi
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS.....	viii
ABSTRAK.....	ix
ABSTRCT.....	x
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR.....	xiv
DAFTAR LAMPIRAN.....	xv
BAB 1. PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Permasalahan.....	3
1.2.1 Identifikasi Masalah	3
1.2.1 Batasan Masalah	3
1.2.1 Rumusan Masalah	4
1.3. Tujuan Penelitian.....	4
1.4. Manfaat Penelitian.....	4
1.5. Sistematika Penulisan.....	4
BAB 2. LANDASAN DAN KERANGKA PEMIKIRAN.....	6
2.1. Tinjauan Studi Pustaka.....	6
2.1.1 Analisis Sentimen (<i>Sentiment Analysis</i>).....	6
2.1.2 <i>Support Vector Machine</i>	7
2.1.3 Pemilihan Fitur.....	10
2.2. Tinjauan Studi Penelitian Terkait.....	14
2.2.1 Model Penelitian Passonneau.....	14
2.2.2 Model Penelitian Basari.....	16
2.2.3 Model Penelitian Ye.....	18
2.2.4 Rangkuman Penelitian Terkait.....	19
2.3. Tinjauan Obyek Penelitian.....	20
2.3.1 Sentimen pada Twitter.....	20
2.3.2 <i>Feature Selection</i> (Pemilihan Fitur).....	22
2.3.3 Validasi dan Evaluasi Algoritma Data Mining.....	22
2.4. Kerangka Pemikiran.....	23
2.5. Hipotesis.....	24
BAB 3. METODE PENELITIAN.....	25
3.1. Perancangan Penelitian.....	25
3.2. Pengumpulan Data.....	26
3.3. Pengolahan Awal Data.....	26
3.4. Metode Yang Diusulkan.....	27

3.5. EksperimendanPengujian Model.....	29
3.6. EksperimendanValidasi Hasil.....	30
BAB 4. HASIL PENELITIAN DAN PEMBAHASAN.....	31
4.1. Hasil.....	31
4.1.1KlasifikasiTeksMenggunakanAlgoritma Support Vector Machine.....	31
4.1.2Pengujian model dengan 10 Foldcross Validation.....	35
4.1.3Optimasi Model denganMetode Genetic Algoritma.....	36
4.1.4EksperimenTerhadapIndikator Model.....	36
4.2. Pembahasan.....	37
4.2.1Pengukurandengan <i>Confusion Matrix</i>	38
4.3. DesaindanImplementasi.....	41
4.4. ImpilasiPenelitian.....	44
BAB 5. PENUTUP.....	45
5.1. Kesimpulan.....	45
5.2. Saran.....	46
DAFTAR REFERENSI.....	47
LAMPIRAN.....	49

DAFTAR TABEL

	Halaman
Tabel 2.1. Perbandingan penelitian Terkait.....	18
Tabel 3.1. Spesifikasi computer yang digunakan.....	28
Tabel 3.2. <i>Confusion Matrix</i>	29
Tabel 4.1. Perbandingan teks sebelum dan sesudah dilakukan proses <i>tokennization</i>	30
Tabel 4.2. Perbandingan teks sebelum dan sesudah dilakukan proses <i>Stopwords</i>	31
Tabel 4.3. Perbandingan teks sebelum dan sesudah dilakukan proses <i>Steamming</i>	31
Tabel 4.4. Perbandingan teks sebelum dan sesudah dilakukan proses <i>Generate 2-Gram</i>	32
Tabel 4.5. Table dokumen yang dihitung nilai bobotnya.....	33
Tabel 4.6. Table indicator dan hasil pengujian.....	35
Tabel 4.7. Model algoritma <i>Support Vector Machine</i> (SVM) sebelum dan sesudah menggunakan metode <i>Genetic</i> <i>Algorithm</i>	37
Tabel 4.8. Model <i>Confusion Matrix</i> model <i>Support Vector Machine</i> (SVM) sebelum menggunakan metode <i>Genetic</i> <i>Algorithm</i>	37
Tabel 4.9. Model <i>Confusion Matrix</i> model <i>Support Vector Machine</i> (SVM) sesudah menggunakan metode <i>Genetic</i> <i>Algorith</i>	38

DAFTAR GAMBAR

	Halaman
Gambar 2.1. Margin Kecil dan Margin Besar.....	8
Gambar 2.2. Operasi Crossover dan Mutasi.....	11
Gambar 2.3. Model yang diusulkan oleh Passonneaus.....	14
Gambar 2.4. Model yang diusulkan oleh Basari.....	16
Gambar 2.5. Model yang diusulkan oleh Ye.....	17
Gambar 2.6. Kerangka Pemikiran.....	23
Gambar 3.1. Model yang diusulkan.....	27
Gambar 4.1. Desain model <i>Support Vector Machine</i> (SVM) menggunakan Rapid Miner.....	34
Gambar 4.2. Desain model <i>Support Vector Machine</i> (SVM) dan <i>Genetic Algorithm</i> menggunakan Rapid Miner.....	36
Gambar 4.3. Kurva ROC Model <i>Support Vector Machine</i> (SVM) sebelum menggunakan metode <i>Genetic Algorithm</i>	39
Gambar 4.4. Kurva ROC Model <i>Support Vector Machine</i> (SVM) sesudah menggunakan metode <i>Genetic Algorithm</i>	39
Gambar 4.5. Diagram alirahpan proses klasifikasi dengan <i>Algorithm Support Vector Machine</i> (SVM) menggunakan metode <i>Genetic Algorithm</i>	40
Gambar 4.6. Tampilan aplikasi sentiment analisis tweet.....	41
Gambar 4.6. Tampilan aplikasi sentiment analisis dengan keyword berbeda.	42

DAFTAR LAMPIRAN

	Halaman
Lampiran 1. Tabel dataset <i>tweetpositif</i>	48
Lampiran 2. Tabel dataset <i>tweetnegatif</i>	54

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Beberapa tahun terakhir ini telah terlihat lonjakan yang sangat menarik dalam sebuah metode komputasi yang memiliki pengaruh terhadap *opini mining*, deteksi subjectivitas, dan sentiment analisis terutama dalam penggunaan media social. Twitter telah menjadi salah satu platform mikroblog paling populer baru-baru ini. Jutaan pengguna dapat berbagi pikiran dan pendapat tentang berbagai aspek dan kegiatan untuk platform *micro blogging* mereka. Oleh karena itu twitter dianggap sebagai sumber yang kaya informasi untuk pengambilan keputusan dan analisis sentimen (Balahur, Mihalcea, & Montoyo, 2014). Munculnya media sosial telah memberikan tempat bagi para pengguna web untuk mengekspresikan berbagi pikiran serta pendapat mereka tentang topik yang berbeda dalam suatu peristiwa. Twitter telah hampir memiliki 600 juta pengguna dan lebih dari 250 juta pesan per hari. Hal ini dengan cepat telah menjadi tambang emas bagi organisasi untuk memantau reputasi dan merek mereka dengan mengekstraksi dan menganalisis sentimen dari *Tweets* yang diposting oleh publik tentang pasar mereka dan para pesaing (F. H. Khan, Bashir, & Qamar, 2014)

Area utama lain penelitian Twitter berfokus pada pengembangan pendekatan analisis sentimen yang dirancang khusus untuk tweet. Tweets adalah genre yang unik komunikasi, dan fitur unik dan sifat telah dipertanyakan dalam penerapan dan efektivitas dari pendekatan yang lebih tradisional untuk sentiment analisis. Tweets adalah unit yang sangat singkat teks, maksimal 140 karakter panjang, dan ditandai dengan santai, bahasa kompak dengan penggunaan luas gaul, singkatan, akronim, dan emoticon. *Tweets* juga mengandung hashtags, referensi pengguna, dan link tertanam ke situs lain yang berisi informasi direferensikan tambahan, lebih rumit analisis sentimen. (Ghiassi, Skinner, & Zimbra, 2013)

Terdapat beberapa penelitian yang telah dilakukan dalam hal klasifikasi sentimen pada twitter yang tersedia secara online diantaranya, analisis sentimen pada data twitter dengan menggunakan pengklasifikasian *Support Vector Machine* (SVM) (Passonneau, 2011). Analisa sentimen pada opini review film menggunakan pengklasifikasi *Support Vector Machine* (SVM) dan *Particle Swarm Optimization* (Basari et al., 2013). Sentimen kalsifikasi terhadap ulasan tempat tujuan *travel* dengan menggunakan *pendekatan Supervised Machine Learning* (Ye, Zhang, & Law, 2009).

Support Vector Machine (SVM) merupakan salah satu algoritma kalsifikasi linier yang memiliki prinsip utama untuk menentukan pemisah linear dalam ruang pencarian terbaik yang dapat memisahkan dua kelas yang berbeda. (Schoefegger, Tammet, & Granitzer, 2013) kelebihan pada algoritma SVM yaitu mampu mengidentifikasi *hyperplane* terpisah yang memaksimalkan margin antara dua kelas (Chou, Cheng, Wu, dan Pham, 2014). SVM adalah kasus khusus dari keluarga algoritma yang kita sebut sebagai pengatur metode klasifikasi linier dan metode yang kuat untuk minimalisasi resiko (Weiss, Indurkha, dan Zhang, 2010). Namun *Support Vector Machine* (SVM) memiliki kekurangan terhadap masalah pemilihan parameter yang sesuai. (Basari et al., 2013)

Seleksi fitur mempengaruhi beberapa aspek yaitu pola klasifikasi, akurasi klasifikasi, waktu yang diperlukan untuk pembelajaran fungsi klasifikasi, jumlah contoh yang dibutuhkan untuk pembelajaran dan biaya yang terkait dengan fitur menurut Yang dan Honavar dalam Zhao (Zhao, Fu, Ji, Tang, dan Zhou, 2011). Dalam masalah aplikasi tertentu, tidak semua fitur ini sama pentingnya. Kinerja yang lebih baik dapat dicapai dengan membuang beberapa fitur. Dengan demikian, kita dapat menghilangkan data yang *noise*, data yang tidak relevan dan berlebihan.

Selain seleksi fitur, pengaturan parameter *Support Vector Machine* (SVM) memiliki pengaruh penting pada akurasi klasifikasinya. Tidak sesuainya sebuah pengaturan parameter dapat menyebabkan hasil klasifikasi yang rendah menurut Keerthi & Lin dalam Zhao (Zhao et al., 2011). Akurasi klasifikasi optimal SVM diperoleh dengan mencari setting parameter yang optimal. Parameter yang harus dioptimalkan yaitu kesalahan parameter C dan parameter fungsi kernel seperti

Gaussian kernel parameter γ untuk fungsi kernel Gaussian. Pencarian grid merupakan alternatif dan pendekatan pencarian secara langsung. Namun, pencarian kemampuan dari pendekatan ini sangat rendah menurut Hsu, Chang, dan Lin dalam Zhao (Zhao et al., 2011). Selain itu, pencarian grid tidak dapat melakukan pilihan fitur. Dalam beberapa literatur, metode seleksi fitur berdasarkan *Genetic Algorithm* (GA) telah diusulkan oleh Raymer, Punch, Goodman, Kuhn, dan Jain dalam Zhao (Zhao et al., 2011).

Kecenderungan dalam beberapa tahun terakhir adalah untuk secara bersamaan mengoptimalkan bagian fitur dan parameter untuk *Support Vector Machine* (SVM), sehingga dapat meningkatkan akurasi klasifikasi pada *Support Vector Machine* (SVM). Algoritma Genetika 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 Machine* (SVM) dan metode *Genetic Algorithm* sebagai metode pemilihan fitur yang akan diterapkan untuk mengklasifikasikan *tweet* yang ada pada twitter untuk meningkatkan akurasi sentimen.

1.2 Permasalahan

1.2.1 Identifikasi Masalah

Berdasarkan latar belakang permasalahan tersebut, maka identifikasi permasalahannya yaitu pada teknik untuk melakukan analisis sentimen yang dapat digunakan dengan berbagai macam cara salah satunya yaitu menggunakan algoritma *Support Vector Machine* (SVM) sebagai algoritma klasifikasi linier yang memiliki prinsip utama untuk menentukan pemisah linear dalam ruang pencarian terbaik yang dapat memisahkan dua kelas yang berbeda. Namun, Namun *Support Vector Machine* memiliki kekurangan terhadap masalah pemilihan parameter yang sesuai yang mengakibatkan akurasi klasifikasi menjadi rendah.

1.2.2 Batasan Masalah

Berbagai macam teknik banyak digunakan untuk melakukan penelitian klasifikasi sentimen analisis. Diantaranya yaitu teknik menggunakan 3 jenis algoritma berbeda lalu ditentukan algoritma terbaik kemudian dipotimaskan

menggunakan metode optimasi. Ada pun teknik yang menggunakan optimasi terlebih dahulu kemudian menggunakan 3 algoritma dan ditentukan hasil yang terbaik. Membahas itu semua memungkinkan tetapi dalam penelitian ini hanya membahas mengenai penerapan algoritma *Support Vector Machine* (SVM) dengan teknik optimasi metode *Genetic Algorithm* untuk meningkatkan akurasi analisa sentimen pada twitter.

1.2.3 Rumusan Masalah

Berdasarkan batasan masalah tersebut, maka rumusan masalah dalam penelitian ini adalah:

- a. Apakah *Support Vector Machine* (SVM) mampu dalam menyelesaikan masalah dalam klasifikasi sentimen pada twitter?
- b. Apakah dengan metode *Genetic Algorithm* mampu meningkatkan akurasi analisa sentimen pada *Support Vector Machine* (SVM)?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah Penerapan algoritma klasifikasi linier *Support Vector Machine* (SVM) dan metode *Genetic Algorithm* dalam menganalisa sentimen pada twitter.

1.4 Manfaat Penelitian

Manfaat dari Penelitian ini adalah:

- a. Manfaat praktis dari hasil penelitian ini diharapkan dapat digunakan untuk menindaklanjuti suatu strategi perusahaan maupun organisasi dalam menentukan keputusan dari akun twitter mereka.
- b. Manfaat teoritis dari penelitian ini merupakan bukti empiris dalam perkembangan dan penerapan algoritma *Support Vector Machine* (SVM) dengan metode *Genetic Algorithm* dalam mengklasifikasikan analisa sentiment *tweet* pada twitter.

1.5 Sistematika Penulisan

dalam penulisan tesis ini terdapat beberapa bab yang dibahas, diantaranya adalah:

Bab 1 Pendahuluan

Bab ini menjelaskan tentang latar belakang penulisan, pemasalahan, identifikasi masalah, batasan masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penelitian.

Bab II Landasan Teori

Bab ini menjelaskan tentang teori yang melandasi penelitian yaitu mengenai sentimen analisis pada twitter, algoritma *Support Vector Machine* (SVM) dan metode *Genetic Algorithm*

Bab III Metode Penelitian

Bab ini menjelaskan tentang metode pengumpulan data, pengolahan awal data, metode yang diusulkan, eksperimen dan pengujian metode, dan evaluasi dan validasi hasil. Pada penelitian ini akan menguji algoritma *Support Vector Machine* dan metode *Genetic Algorithm* yaitu untuk meningkatkan akurasi untuk mengklasifikasikan *tweet* pada twitter.

Bab IV Hasil dan Pembahasan

Bab ini akan menampilkan hasil dari eksperimen, baik sebelum maupun . sesudah model diterapkan dan mengukur tingkat akurasi pada algoritma yang diterapkan.

Bab V Penutup

Bab ini membahas tentang kesimpulan dari pembahasan pada bab-bab sebelumnya dan saran-saran untuk penelitian selanjutnya.

BAB II

LANDASAN TEORI DAN KERANGKA PEMIKIRAN

2.1 Tinjauan Studi Pustaka

Dalam tesis ini, penulis menggunakan referensi dari buku dan jurnal untuk menjelaskan pengklasifikasi *Support Vector Machine* (SVM) dan pemilihan fitur *Wrapper* teknik optimasi *Genetic algorithm*.

2.1.1 Analisis Sentimen (*Sentiment Analysis*)

Menurut Haddi (Haddi, Liu, & Shi, 2013) Analisis Sentimen diperlakukan sebagai tugas dari sebuah klasifikasi untuk mengklasifikasikan orientasi teks menjadi positif atau negatif. Menurut Medhat (Medhat, Hassan, & Korashy, 2014) Analisis Sentimen dapat dianggap sebagai proses klasifikasi yang memiliki 3 tingkat klasifikasi utama yaitu tingkat pada dokumen, tingkat pada kalimat dan tingkat pada aspek. Pada tingkat dokumen analisis sentimen bertujuan untuk mengklasifikasikan dokumen opini sebagai pernyataan pendapat atau sentimen positif dan negatif. Pada tingkat kalimat analisis sentiment bertujuan untuk mengklasifikasikan sebuah sentimen yang dinyatakan dalam setiap kalimat. Langkah pertama yang dilakukan adalah mengidentifikasi kalimat tersebut apakah kalimat subjektif atau objektif. Pada tingkat kalimat sentimen analisis akan mengungkapkan pendapat positif atau negatif. Sedangkan pada tingkat aspek analisis sentiment bertujuan untuk mengklasifikasikan sentimen yang berhubungan dengan aspek yang spesifik dari sebuah objek. Langkah pertama yang dilakukan adalah mengidentifikasi objek dari topik yang dibahas. 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 mengidentifikasi alasan, opini atau penilaian dan kondisi kecenderungan (bagaimana si penulis ingin mempengaruhi pembaca).

Menurut Moraes (Moraes et al., 2013) langkah-langkah yang umumnya ditemukan pada klasifikasi teks analisa sentimen adalah:

- a. Definisikan domain dataset

Pengumpulan dataset yang melingkupi suatu domain, misalnya dataset review film, dataset review produk, dataset *tweet* dan lain sebagainya.

b. *Pre-processing*

Tahap pemrosesan awal yang umumnya dilakukan dengan proses *Tokenization*, *Stopwords removal*, dan *Stemming*.

c. *Transformation*

Proses representasi angka yang dihitung dari data tekstual. *Binary Representation* yang umumnya digunakan dan hanya 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. Proses yang umumnya digunakan yaitu *TF-IDF*, *Binary transformation*, dan *Frequency transformation*.

d. *Feature Selection*

Pemilihan fitur (*feature selection*) bisa membuat pengklasifikasi lebih efisien/efektif dengan mengurangi jumlah data untuk dianalisa dengan mengidentifikasi fitur yang relevan yang selanjutnya akan diproses. Metode pemilihan fitur yang biasanya digunakan adalah *Expert Knowledge*, *Minimum Frequency*, *Information gain*, *Chi-Square*, dan lain sebagainya.

e. *Classification*

Proses klasifikasi umumnya menggunakan pengklasifikasi seperti *Naïve Bayes*, *Support Vector Machine*, dan lain sebagainya.

f. *Interpretation/Evaluation*

Tahap evaluasi biasanya menghitung akurasi, *recall*, *precision*, dan *F-1*.

2.1.2 *Support Vector Machine (SVM)*

Menurut Widodo (2013) *Support Vector Machine (SVM)* merupakan algoritma yang bekerja menggunakan pemetaan nonlinier untuk mengubah data pelatihan asli ke dimensi yang lebih tinggi. Dalam hal ini dimensi baru, akan mencari *hyperlane* untuk memisahkan secara linier dan dengan pemetaan nonlinier yang tepat ke dimensi yang cukup tinggi, data dari dua kelas selalu dapat dipisahkan dengan *hyperlane* tersebut. SVM menemukan *hyperlane* ini menggunakan *Support Vector* dan *Margin*.

Support Vector Machine (SVM) merupakan algoritma yang mampu mengidentifikasi *hyperplane* terpisah yang memaksimalkan margin antara dua kelas. Memaksimalkan margin adalah masalah pemrograman kuadratik, yang dapat diselesaikan dari masalah ganda dengan memperkenalkan *Lagrangian multipliers*. Namun, mencari *hyperplane* yang cocok di ruang input sering terlalu ketat untuk penggunaan praktis. Salah satu solusi adalah dengan memetakan ruang input ke dalam ruang yang lebih tinggi fitur dimensi dan kemudian mencari *hyperplane* optimal. Bahkan tanpa pengetahuan pemetaan, SVM masih dapat menemukan *hyperplane* optimal dengan menggunakan fungsi produk dot dalam ruang fitur yang disebut kernel. *Hyperplane* yang optimal dapat dinyatakan sebagai kombinasi beberapa titik masukan yang disebut *support vector*. (Chou, Cheng, Wu, & Pham, 2014)

SVM adalah metode klasifikasi dua kelas dan teori ini didasarkan pada gagasan minimalisasi risiko struktural. SVM meminimalkan kesalahan generalisasi dan memaksimalkan margin geometris antara dua kelas, maka juga dikenal sebagai pengklasifikasi maksimum margin. (Aydin, Karakose, & Akin, 2011).

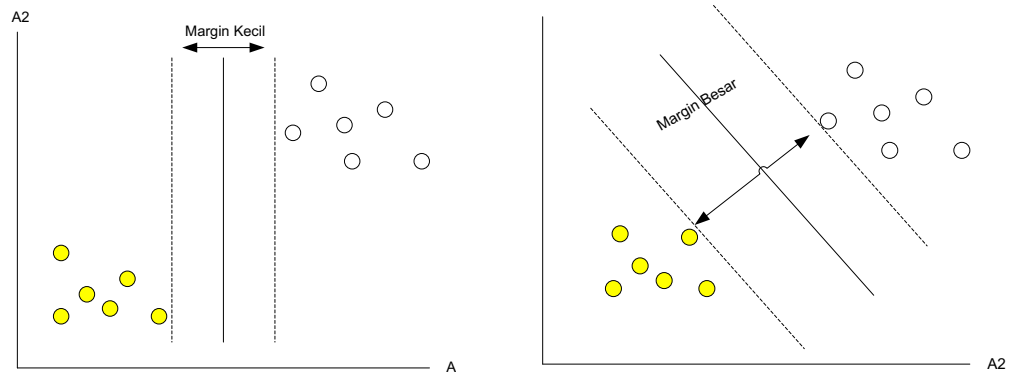
Menurut Widodo (2013) SVM mampu memecahkan masalah klasifikasi dengan mencari *hyperplane* marjinal maksimum di mana ada jumlah tidak terbatas *hyperplane* yang harus dicari mana yang terbaik.

Sebuah *hyperplane* dapat ditulis sebagai:

$$\mathbf{W} \cdot \mathbf{X} + \mathbf{b} = 0 \quad (1)$$

Di mana $\mathbf{W}$ adalah vektor bobot, yaitu $\mathbf{W} = \{w_1, w_2, \dots, w_n\}$, n adalah jumlah atribut, dan $\mathbf{b}$ adalah skalar yang sering disebut sebagai bias. Untuk membantu visualisasi, kita mempertimbangkan dua masukan atribut, A_1 dan A_2 seperti pada gambar 2.1. Tupel pelatihan 2-D, misalnya $\mathbf{X} = (X_1, X_2)$, di mana X_1 dan X_2 adalah nilai-nilai atribut A_1 dan A_2 , masing-masing untuk $\mathbf{X}$. Jika $\mathbf{b}$ sebagai bobot tambahan, $\mathbf{W}_0$ kita dapat menulis ulang *hyperplane* pemisah sebagai berikut:

$$\mathbf{W}_0 + \mathbf{W}_1\mathbf{X}_1 + \mathbf{W}_2\mathbf{X}_2 = 0 \quad (2)$$



Gambar 2.1 Margin Kecil dan Margin Besar (Widodo,2013)

Dengan demikian setiap titik yang terletak di atas *hyperplane* pemisah memenuhi:

$$W_0 + W_1X_1 + W_2X_2 > 0 \quad (3)$$

Dengan demikian pula, setiap titik yang terletak di dibawah *hyperplane* pemisah memenuhi:

$$W_0 + W_1X_1 + W_2X_2 < 0 \quad (4)$$

Bobot dapat disesuaikan sehingga *hyperplane* dapat didefinisikan “sisi” dari margin yang ditulis sebagai:

$$H1: W_0 + W_1X_1 + W_2X_2 \geq 1 \text{ untuk } Y_i = +1 \quad (5)$$

$$H2: W_0 + W_1X_1 + W_2X_2 \leq -1 \text{ untuk } Y_i = -1 \quad (6)$$

Artinya, setiap tuple yang jauh pada atau di atas H_1 , milik kelas +1 dan setiap tuple yang jatuh atau dibawah H_2 milik kelas -1.

Menurut Basari (Basari, Hussin, Ananta, & Zeniarja, 2013) dalam menentukan suatu nilai pembobotan kelas positif dan negatif dalam *Support Vector Machine (SVM)* ditentukan berdasarkan jika nilai dari bobot lebih besar sama dengan 0 maka diklasifikasikan kedalam kelas positif dan sebaliknya jika nilai bobot kurang dari 0 maka dapat diklasifikasikan kedalam kelas negatif.

Berikut adalah fungsi dari rumus perhitungan bobot pada *Support Vector Machine* (SVM) (Wu,2009) .

$$W^* = \sum_{i=1}^n \alpha_i^* y_i X_i$$

Variabel dan parameter Algoritma *Support Vector Machine* (SVM) untuk kalsifikasi:

1. χ = { $\chi_1, \chi_1, \dots, \chi_m$ } sebagai sampel training
2. y = { $y_1, \dots, y_m$ } $\in \{\pm 1\}$ sebagai label data training
3. kernel = jenis fungsi kernel
4. par = parameter kernel
5. C = komstanta cost
6. A = [$\alpha_1, \dots, \alpha_m$] sebagai multiplier dan bias

2.1.3. Pemilihan Fitur (Feature Selection)

Menurut Gorunescu (Gorunescu, 2011) pemilihan fitur digunakan untuk menghilangkan fitur yang tidak relevan dan berulang, yang mungkin menyebabkan kekacauan, dengan menggunakan metode tertentu. Menurut John, Kohavi, dan Pfleger dalam Chen (Chen et al., 2009) metode pemilihan fitur dalam machine learning salah satunya yaitu:

1. Wrapper

Menurut Chen (Chen et al., 2009) wrapper menggunakan akurasi klasifikasi dari beberapa algoritma sebagai fungsi evaluasinya. Karena wrapper harus menguji sebuah pengklasifikasi untuk tiap subset fitur untuk dievaluasi, biasanya lebih banyak memakan waktu khususnya ketika jumlah fiturnya tinggi. Menurut Kohavi dalam Yang (Yang, Liu, Zhou, Chawla, & Albert, 2010) wrapper mengevaluasi fitur secara berulang dengan mematuhi algoritma induktif. Oleh karena itu, fitur yang dipilih oleh pendekatan wrapper mungkin lebih cocok untuk algoritma induktif, dan menghasilkan akurasi klasifikasi yang tinggi. Menurut Gunal (Gunal, 2012) salah satu metode wrapper yang bisa digunakan dalam pemilihan fitur adalah Genetic algorithm

(GA). Penelitian ini akan dibahas mengenai Genetic algorithm lebih rinci sebagai berikut:

a. Algoritma Genetika (*Genetic Algorithm*)

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 Zukhri (Zukhri, 2014) Algoritma Genetika merupakan suatu metode heuristik yang dikembangkan berdasarkan prinsip genetika dan proses seleksi alamiah teori Evolusi Darwin. Metode optimasi dikembangkan oleh Jhon Holland sekitar tahun 1960-an dan dipopulerkan oleh salah seorang mahasiswanya, David Goldberg pada tahun 1980-an. Proses pencarian penyelesaian dalam algoritma ini berlangsung sama seperti terpilihnya individu untuk bertahan hidup dalam proses evolusi.

Menurut Suyanto (Suyanto, 2014) Algoritma Genetika merupakan algoritma pencarian yang didasarkan pada mekanisme seleksi alam dan genetika alamiah. Pada awalnya, Algoritma Genetika memang digunakan sebagai algoritma pencarian parameter-parameter yang optimal. Tetapi dalam perkembangannya, AG bisa diaplikasikan untuk berbagai masalah lain seperti *learning*, peramalan, pemrograman otomatis, dan sebagainya. Pada bidang *soft Computing*, AG banyak digunakan untuk mendapatkan nilai-nilai parameter yang optimal.

Tiga operator yang digunakan dalam algoritma genetika adalah (Larose, n.d.):

1. Seleksi

Operator seleksi mengacu pada metode yang digunakan untuk memilih mana kromosom akan bereproduksi. *Fitness function* mengevaluasi masing-masing

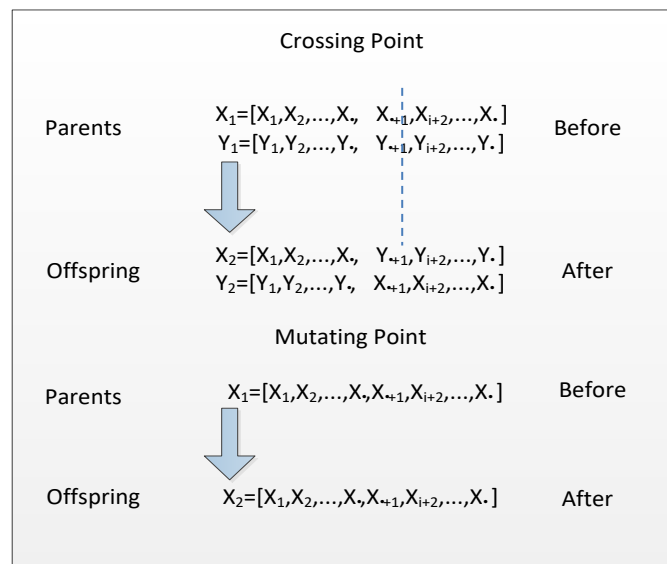
kromosom (kandidat solusi), dan kromosom terpilih akan semakin besar kemungkinan dipilih untuk mereproduksi.

2. *Crossover*

Operator *crossover* melakukan rekombinasi, menciptakan dua keturunan baru dengan cara acak memilih kedudukan dan bertukar setiap sisi ke kiri dan ke kanan yang kedudukannya antara dua kromosom dipilih selama seleksi. Misalnya, dalam representasi biner, dua string 11111111 dan 00000000 bisa menyeberang pada kedudukan keenam dalam masing-masing untuk menghasilkan dua keturunanyang baru 11111000 dan 00000111.

3. Mutasi

Operator mutasi acak mengubah bit atau digit pada kedudukan tertentu dalam kromosom, namun dengan probabilitas yang sangat kecil. Sebagai contoh, setelah Crossover, yaitu 11111000 bisa bermutasi pada sebuah dua kedudukan menjadi 10.111.000. Mutasi memperkenalkan informasi baru kedalam genetik dan melindungi terhadap konvergen yang terlalu cepat untuk optimum lokal.



Gambar 2.2 Operasi Crossover dan Mutasi (Zhao, Fu, Ji, Tang, & Zhou, 2011)

Proses dasar algoritma genetika adalah berikut (Zhao et al., 2011):

Langkah1 : acak menghasilkan populasi awal.

- Langkah2 : Perkirakan nilai fitness dari setiap kromosom dalam populasi.
- Langkah3 : Lakukan *crossover*, mutasi dan operasi, genetika seleksi.
- Langkah 4 : Hentikan algoritma jika kriteria penghentian dipenuhi,
Kembali ke Langkah 2 sebaliknya.

Algoritma genetika memberikan suatu pilihan bagi penentuan nilai parameter dengan meniru cara reproduksi genetika, pembentukan kromosom baru serta seleksi alami seperti yang terjadi pada makhluk hidup. Algoritma genetika mempunyai karakteristik-karakteristik yang perlu diketahui sehingga dapat terbedakan dari prosedur pencarian atau optimasi yang lain (Weise, 2009), yaitu:

1. Algoritma genetika dengan pengkodean dari himpunan solusi permasalahan berdasarkan parameter yang telah ditetapkan dan bukan parameter itu sendiri.
2. Algoritma genetika pencarian pada sebuah solusi dari sejumlah individu-individu yang merupakan solusi permasalahan bukan hanya dari sebuah individu.
3. Algoritma genetika informasi fungsi objektif (fitness), sebagai cara untuk mengevaluasi individu yang mempunyai solusi terbaik, bukan turunan dari suatu fungsi.
4. Algoritma genetika menggunakan aturan-aturan transisi peluang, bukan aturan-aturan deterministik.

Genetic Algorithm mudah disejajarkan dan telah digunakan untuk klasifikasi seperti masalah optimasi lainnya. Berdasarkan pengertian optimasi diatas, suatu masalah optimasi yang dapat dikonversi menjadi sebuah fungsi, sehingga dapat dengan mudah menyelesaikan masalah tersebut dengan menggunakan Algoritma Genetika. (Suyanto, 2014).

Menurut David dalam Wang (Wang, Shao, & Wu, 2013) Algoritma genetika (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

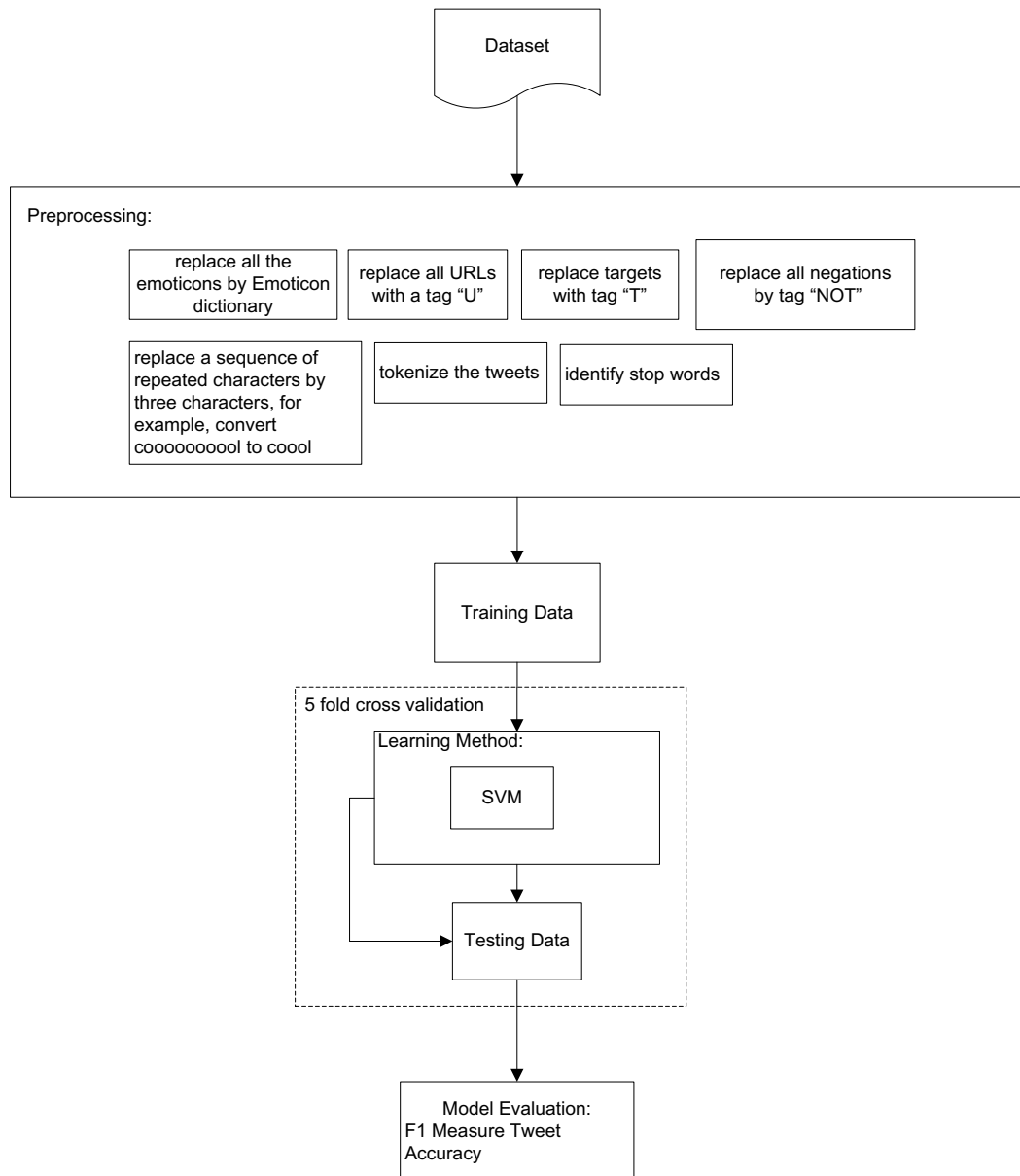
terbukti bahwa GA lebih cocok untuk kasus dengan parameter lainnya karena berkurangnya penggunaan memori dan waktu komputasi yang cepat.

2.2 Tinjauan Studi Penelitian Terkait

Ada beberapa penelitian yang menggunakan algoritma *Support Vector Machine* sebagai pengklasifikasi dalam klasifikasi teks sentimen pada twitter, diantaranya:

2.2.1 Model Penelitian Passonneau

Penelitian yang dilakukan oleh Passonneau mengenai sentimen analisis dari data twitter. Sumber data yang digunakan dalam penelitian Passonneau yaitu dengan <http://www.ngicorporation.com/>. Penelitian ini telah mengumpulkan data sebanyak 5127 *tweet* yang dibagi kedalam 3 kelas positif, negatif, dan netral yang masing masing berjumlah 1709 *tweet*. Langkah-langkah *preprocessing* yang dilakukan yaitu mengganti semua *emoticon* dengan polaritas sentimen mereka dengan melihat kamus *emoticon* (*replace all the emoticons by Emoticon dictionary*), mengganti semua URL dengan tag “U” (*replace all URLs with a tag “U”*), mengganti target (misalnya “@John”) dengan tag “T” (*replace targets with tag “T”*), mengganti semua negasi (*replace all negations by tag “NOT”*), mengganti urutan karakter yang berulang-ulang menjadi tiga karakter (*replace a sequence of repeated characters by three characters*) *tokenize tweets* dengan menggunakan Stanford *tokenizer* dan identifikasi kamus stop word (*stop word dictionary*) Setelah itu dilakukan proses pemisahan data (*splitting the data*) dengan cara membagi data kedalam data pelatihan (*training*) dan data set. Langkah selanjutnya yaitu klasifikasi dengan menggunakan pengklasifikasi menggunakan SVM (*Support Vector Machine*). Gambar 2.3 menunjukkan model yang diusulkan oleh Passonneau.



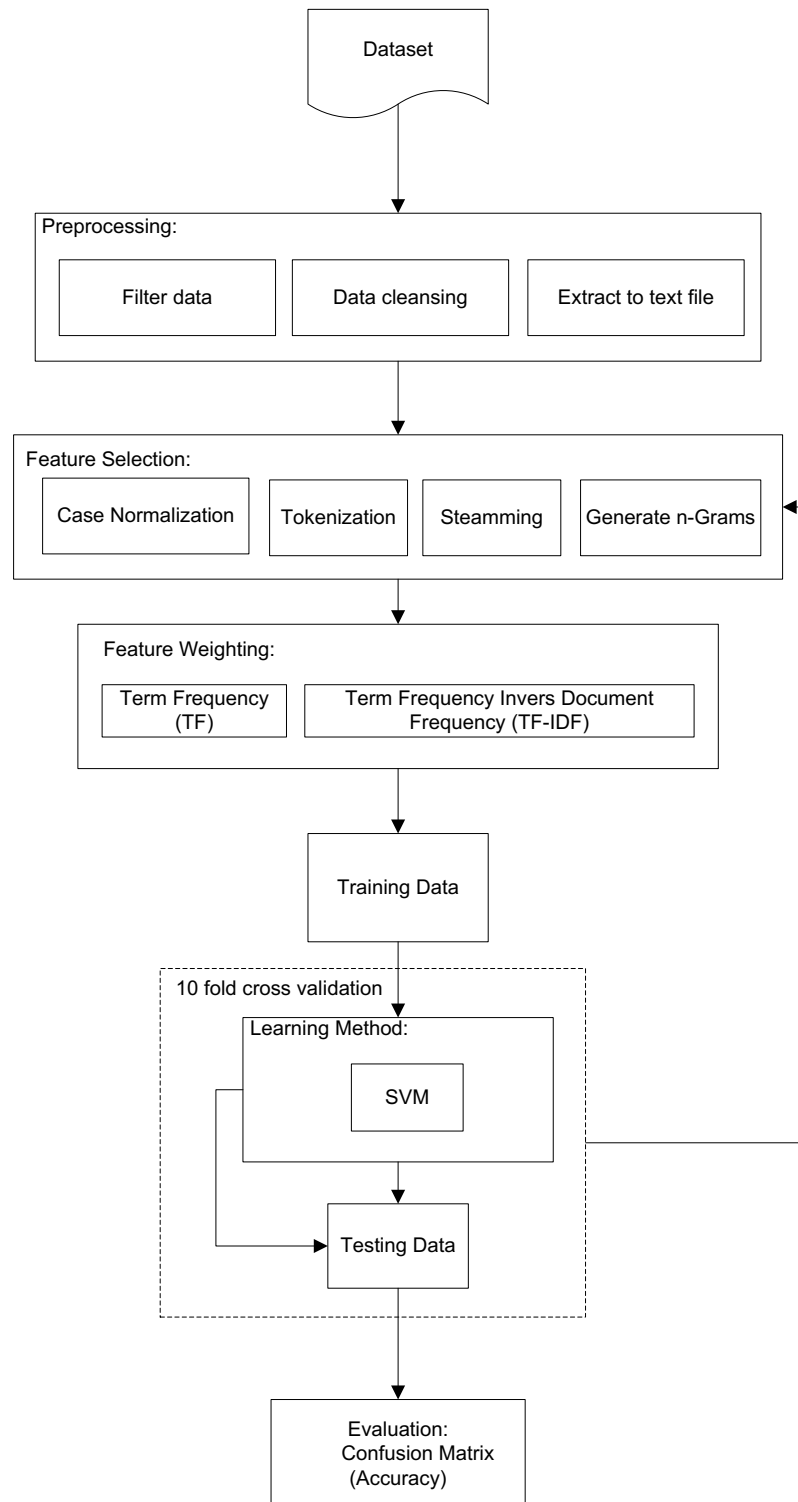
Gambar 2.3 Model yang diusulkan oleh Passonneau

Dari penelitian tersebut, *Support Vector Machine* (SVM) memiliki tingkat akurasi lebih tinggi dari pada pengklasifikasi lainnya (Passonneau, 2011).

2.2.2 Model Penelitian Basari

Penelitian yang dilakukan oleh Basari mengenai opinion mining dari review film menggunakan metode *Support Vector Machine* (SVM). Sumber data yang digunakan dalam penelitian Basari berdasarkan data yang diambil dari Website: <http://www.stanford.edu/~alecmgo/cs224n/trainingandtestdata.zip>.

Langkah-langkah *preprocessing* yang dilakukan yaitu *filter data*, *Data cleansing* sebagai pembersihan data, *extract to text file*. Langkah selanjutnya yaitu pemilihan fitur dan ekstraksi data dimulai dengan tahap *Case Normalization*, *Tokenization*, *Stemming (Snowball)*, *Generate 2-grams*. Kemudian dilakukan pembobotan fitur dengan menggunakan Term Frequency (TF) dan Term Frequency Inverse Document Frequency (TF-IDF). Langkah selanjutnya yaitu klasifikasi teks dengan menggunakan pengklasifikasian menggunakan SVM (Support Vector Machine) yang digabungkan dengan optimasi PSO. Gambar 2.4 menunjukkan model yang diusulkan oleh Basari.

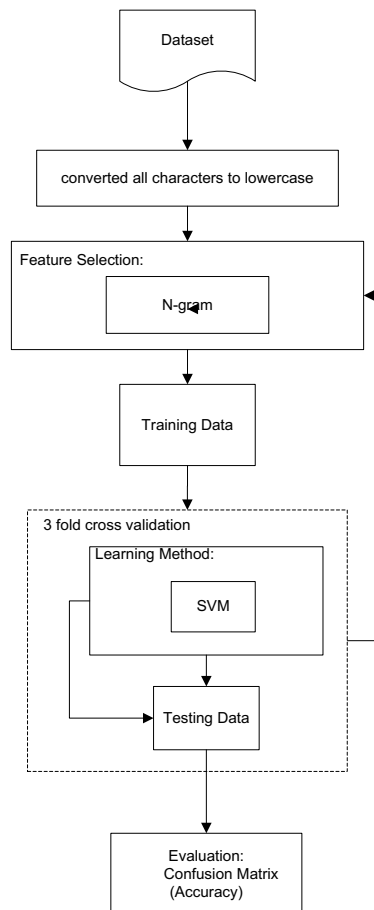


Gambar 2.4 Model yang diusulkan oleh Basari

Dari penelitian tersebut, *Support Vector Machine* (SVM) memiliki tingkat akurasi lebih tinggi dari pada pengklasifikasi lainnya (Basari et al., 2013).

2.2.3 Model Penelitian Ye

Penelitian yang dilakukan oleh Ye mengenai sentimen kalsifikasi terhadap ulasan tempat tujuan *travel* dengan menggunakan pendekatan Supervised Machine Learning. Sumber data yang digunakan dalam penelitian Ye yaitu dengan mengambil corpus dari ulasan wisatawan pada kolom perjalanan Yahoo.com (<http://www.travel.yahoo.com>). Langkah-langkah Untuk mempersiapkan dokumen dalam penelitian ini adalah mengkonversikan semua karakter menjadi huruf kecil baik dalam data pelatihan maupun data pengujian, kemudian menggunakan model n-Gram . Langkah selanjutnya yaitu klasifikasi teks dengan menggunakan pengklasifikasian menggunakan SVM (Support Vector Machine). Gambar 2.5 menunjukan model yang diusulkan oleh Ye



Gambar 2.5 Model yang diusulkan oleh Ye

Dari penelitian tersebut, *Support Vector Machine* (SVM) memiliki tingkat akurasi yang lebih tinggi (Ye, Zhang, & Law, 2009)

2.2.4 Rangkuman Penelitian Terkait

Ketiga penelitian terkait memiliki model yang berbeda, namun pengklasifikasi *Support Vector Machine* (SVM) memiliki akurasi yang paling tinggi diantara pengklasifikasi lainnya. Algoritma Genetika dapat digunakan sebagai model optimasi yang dapat digunakan untuk menghasilkan akurasi *Support Vector Machine* (SVM) yang lebih tinggi. Perbandingan ketiga penelitian terkait dapat dilihat pada tabel 2.1.

Tabel 2.1. Perbandingan Penelitian Terkait

Judul	Preprocessing	Feature Selection	Classifier	Accuracy
Sentiment Analysis of Twitter Data (Passonneau , 2011)	- Replace all the emoticons by emoticon dictionary - Replace all URLs with a Tag “U” - Replace targets with a Tag “T” - Replace all negations by Tag “NOT” - Replace a sequence of repeated characters - Tokenizing all the tweet - Dentify stop words	- POS-specific prior polarity features - tree kernel	Support Vector Machine (SVM)	75,39%
Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm	- fillter data - Data cleansing - xtract to text file	- Ase Normalization - okenizing - teamming - enerate n-Gram	Support Vector Machine (SVM)	77%

Optimization (Basari et al., 2013)				
Sentiment classification of online reviews to travel destinations by supervised machine learning approaches (Ye et al., 2009)	Convert all characters to lowercase	n-Gram	Support Vector Machine (SVM)	80%
Model yang diusulkan	- tokenization - identify stop words - stemming - generate 2-gram	- n-Gram - Genetic Algoritma	Support Vector Machine	?

Dari tinjauan studi diatas, dapat diketahui bahwa *Support Vector Machine* (SVM) merupakan pengklasifikasi terbaik untuk memecahkan masalah analisa sentimen. Pada penelitian ini digunakan sebuah pengklasifikasi algoritma *Support Vector Machine* (SVM) dan metode teknik optimasi yang digunakan adalah *Genetic algorithm* sebagai metode pemilihan fitur untuk meningkatkan tingkat akurasi dari pengklasifikasi.

2.3 Tinjauan Objek Penelitian

Objek penelitian yang dilakukan dalam tesis ini yaitu analisis sentimen pada twitter dan teknik optimasi dalam rangka pemilihan fitur dengan penjelasan sebagai berikut:

2.3.1. Sentimen pada Twitter

Menurut Khan (K. Khan, Baharudin, dan Khan, 2014) *Micro Blogging* situs Twitter merupakan sumber informasi yang kaya dan beragam. Hal tersebut disebabkan oleh sifat dari micro blog yang dimana orang-orang dapat mengirimkan pesan secara real time tentang pendapat mereka terhadap berbagai

topik, membahas isu-isu yang sedang populer, mereka juga dapat berpendapat tentang keluhan dan mengekspresikan sentiment positif bagi produk yang mereka gunakan dalam kehidupan sehari-hari. Bahkan beberapa perusahaan manufaktur produk telah mempelajari reaksi para pengguna lewat twitter. Namun luasnya data sentimen pada twitter tersebut menyebabkan sulit untuk menganalisa dan memahami sentimen tersebut untuk mengklasifikasikan *tweet* pada twitter sebagai sentimen positif, negatif atau netral secara real time.

Menurut Dehkharghani (Dehkharghani, Mercan, Javeed, & Saygin, 2014) Twitter adalah *microblogging* populer dan jaringan sosial situs web dengan basis pengguna terdaftar sekitar 650 juta per tahun 2013 yang memungkinkan penggunaannya untuk mengirim pesan teks paling banyak hanya 140 karakter (*tweet*). Pengguna Twitter mengirimkan atau membahas sebuah pesan (*tweet*) tentang subjek dalam kehidupan sehari-hari. Dan dapat dilihat bahwa dalam beberapa tahun terakhir ini twitter telah banyak digunakan partai politik untuk meluncurkan kampanye terhadap masyarakat.

Menurut Ráez (Montejo-Ráez et al., 2014) Twitter menerbitkan semua jenis informasi dan menyebarluaskan pandangan tentang berbagai topik diantaranya yaitu politik, bisnis, ekonomi dan sebagainya. Pengguna Twitter secara teratur mempublikasikan komentar mereka mengenai sebuah berita tertentu seperti jasa atau produk yang baru mereka beli dan segala sesuatu yang terjadi disekitar mereka. Ini telah menimbulkan kepentingan mengenai pengolahan Bahasa alami pada masyarakat, yang sudah mulai mempelajari teks-teks (*tweet*) yang diposting pada Twitter yang lebih khusus lagi berhubungan dengan tantangan analisis sentimen.

Menurut Ávila (Campo-Ávila, Moreno-Vergara, & Trella-López, 2013) twitter adalah sebuah jaringan informasi yang terdiri dari 140 karakter yang disebut *tweets*. Hal ini merupakan cara yang mudah untuk menentukan berita terbaru mengenai sesuatu topik tertentu atau pun mengenai orang-orang yang dianggap penting bagi kita. Alasan mengapa pengguna twitter tertarik pada platform ini pun beragam dan berbeda-beda. Dan banyak diantaranya yang memiliki tujuan untuk menjadi para pengguna twitter yang memiliki pengaruh, karena

pengaruh twitter dapat didefinisikan sebagai ukuran kemampuan untuk menyebabkan tindakan dan hasil yang diinginkan dan terukur.

2.3.2. *Feature Selection* (Pemilihan fitur)

Menurut Medhat (Medhat et al., 2014) tugas dari analisis sentimen telah dianggap sebagai masalah klasifikasi, maka langkah pertama dalam sebuah masalah klasifikasi sentimen adalah untuk mengekstrak dan memilih fitur pada teks. Berikut penjelasan dari beberapa pemilihan fitur saat ini:

a. *Terms Presence and Frequence*

Fitur-fitur ini adalah kata-kata individu atau N-Gram dan jumlah frekuensi yang sering muncul seperti memberikan bobot pada kata-kata menjadi nilai biner (nol jika kata tersebut muncul dan satu jika sebaliknya) atau menggunakan bobot frekuensi yaitu istilah untuk menunjukkan kepentingan yang relative pada fitur.

b. *Part Of Speech* (POS)

Penemuan kata sifat karena hal tersebut adalah indikator penting dari suatu pendapat.

c. *Opinion Words and Phrases*

Merupakan kata-kata yang umum digunakan untuk mengekspresikan pendapat termasuk pendapat yang baik atau buruk dan suka atau benci. Dari sisi lain beberapa ungkapan mengungkapkan suatu pendapat tanpa menggunakan kata-kata opini.

d. *Negations*

Munculnya kata-kata negatif yang dapat mengubah orientasi opini seperti tidak baik menjadi setara dengan buruk.

2.3.3 *Validasi dan Evaluasi Algoritma Data Mining*

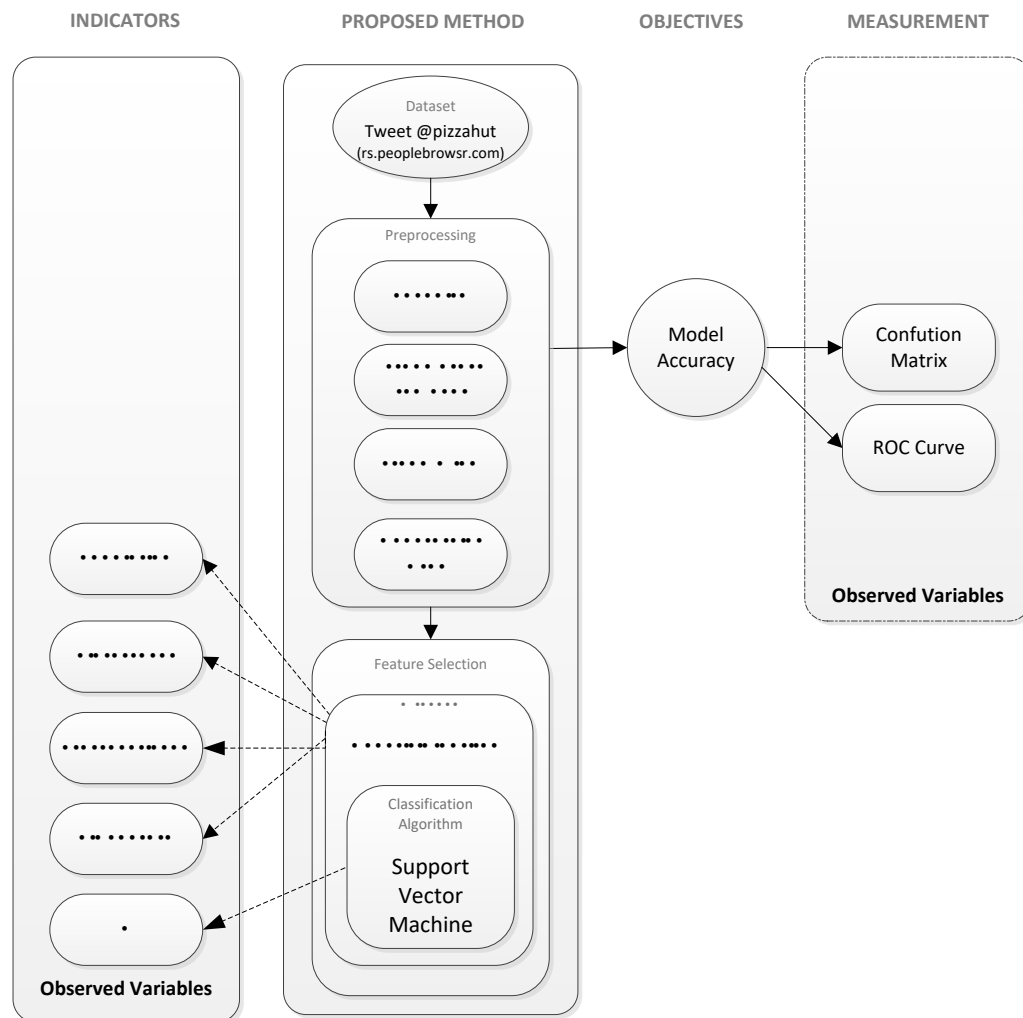
Menurut Gorunescu (Gorunescu, 2011) validasi adalah proses mengevaluasi akurasi prediksi dari suatu model. Ada banyak metode yang digunakan untuk memvalidasi suatu model berdasarkan data yang ada, seperti holdout, random sub-sampling, cross-validation, stratified sampling, bootstrap, dan lain sebagainya. Menurut Han (Han & Kamber, 2007) *confusion matrix*

adalah alat yang sangat berguna untuk menganalisa seberapa baik pengklasifikasi bisa mengenali tuple dari class yang berbeda. Dalam *confusion matrix* dikenal beberapa istilah seperti *True positive* yang merujuk pada tuple positif yang secara benar dilabeli oleh pengklasifikasi, sementara *True negative* adalah tuple negatif yang secara benar dilabeli oleh pengklasifikasi. Adapula *False positive* yang merupakan tuple negatif yang secara tidak benar dilabeli oleh pengklasifikasi, dan *False negative* yang merupakan tuple positif yang secara tidak benar dilabeli oleh pengklasifikasi.

Kurva ROC akan digunakan untuk mengukur AUC (Area Under Curve). ROC curve membagi hasil positif dalam sumbu y dan hasil negatif dalam sumbu x (Witten, Frank, & Hall, 2011). Sehingga semakin besar area yang berada dibawah kurva, semakin baik pula hasil prediksi.

2.4. Kerangka Pemikiran

Penelitian ini dimulai dari adanya masalah dalam kalsifikasi teks pada twitter yang terdiri dari kurang lebih 140 karakter menggunakan *Support Vector Machine* (SVM), di mana pengklasifikasian tersebut memiliki kekurangan terhadap masalah pemilihan parameter yang sesuai, karena dengan tidak sesuainya sebuah pengaturan parameter dapat menyebabkan hasil klasifikasi yang rendah. Dataset yang digunakan dalam penelitian ini yaitu mengambil data *tweet* pada twitter yang didapat dari <http://rs.peoplebrowsr.com/> yang terdiri dari 100 *tweet* positif dan 100 *tweet* negatif dengan waktu selama 5 bulan yang lalu. Preprocessing yang dilakukan dengan *tokenize*, *stopwords removal*, *Stemming*, *generate 2-Gram*. dan pemilihan fitur wrapper dengan metode *Genetic Algorithm*. Sedangkan pengkalsifikasi yang digunakan adalah *Support Vector Machine* (SVM). Pengujian 10 fold cross validation akan dilakukan, akurasi algorima 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 2.6 menggambarkan kerangka pemikiran yang diusulkan dalam penelitian ini.



Gambar 2.6 Kerangka Pemikiran

2.5 Hipotesis

Hipotesis dari penelitian ini adalah:

1. Diduga *Support Vector Machine* (SVM) mampu dalam menyelesaikan masalah dalam klasifikasi sentimen pada twitter.
2. Diduga teknik optimasi Genetic Algoritm mampu meningkatkan akurasi analisa sentiment pada *Support Vector Machine* (SVM).

BAB III

METODE PENELITIAN

3.1. 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 wrapper yaitu *Genetic Algorithm*.

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 Macromedia Dreamweaver 8 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*.

3.2. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini yaitu mengambil data *tweet* pada twitter yang diunduh dari <http://rs.peoplebrowsr.com/> yang terdiri dari 650 *tweet* positif dan 634 *tweet* negatif. Contoh *tweet* positif sebagai berikut:

“tweet this #BestFandom2014 Directioners and tag me please! make me happy I fucking love pizza and i choose pizza hut do it”.

Sedangkan contoh *tweet* negative sebagai berikut:

*“@pizzahut Fine then, ignore me. I'm going to Burger King anyways, phht. Your loss. Bye. I don't like you. Hmph. *Takes bite out of pizza*”*.

3.3 Pengolahan Awal Data

Untuk mengurangi lamanya waktu pengolahan data, pada penelitian ini hanya menggunakan 100 *Tweet* positif dan 100 *Tweet* negatif sebagai data training. Kemudian data *Tweet* tersebut melalui tahap preprosesing dikarenakan terdapat beberapa fitur sintaksis yang mungkin tidak berguna untuk diolah menggunakan algoritma *Machine Learning*, sehingga data *tweet* tersebut harus dibersihkan seperti tanda @ (at) untuk *link* ke nama pengguna, url atau *link* website (http, url, www), # (hashtag), RT (untuk *tweet* balik).

Dataset ini dalam tahap preprosesing harus melalui 4 proses, yaitu:

1. Tokenization

Yaitu mengumpulkan semua kata yang muncul dan menghilangkan tanda baca maupun simbol apapun yang bukan huruf.

2. Stopwords Removal

Yaitu penghapusan kata-kata yang tidak relevan, seperti *the, of, for, with* dan sebagainya.

3. Stemming

Yaitu mengelompokkan kata ke dalam beberapa kelompok yang memiliki kata dasar yang sama, seperti *drug, drugged*, dan *drugs* di mana kata dasar dari semuanya adalah kata *drug*.

4. Generate 2-Gram

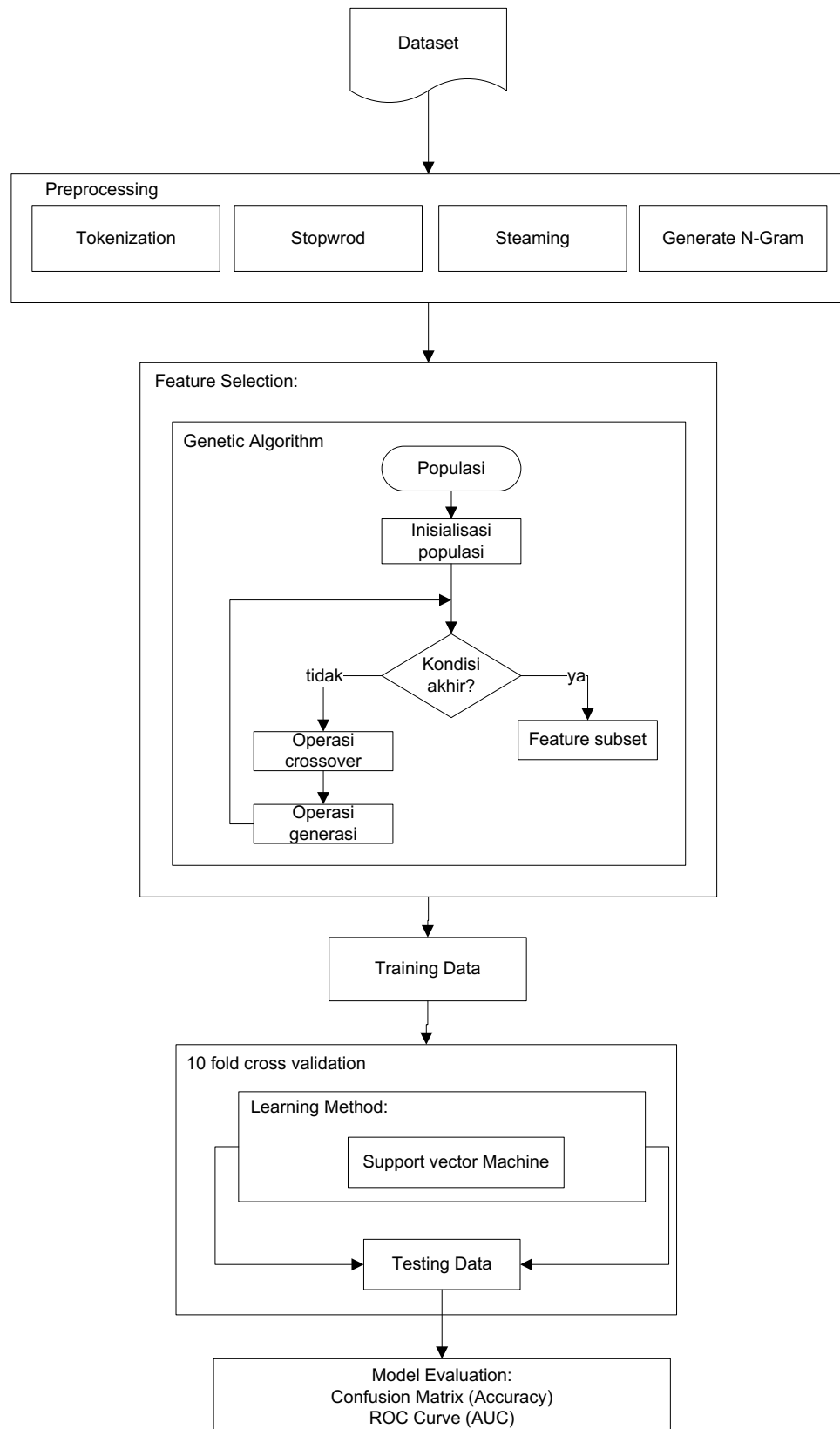
Karakter N-Gram adalah n angka terdekat dari urutan umpan balik yang diberikan. Sebagai contoh 3-gram dari kata “*TERM*” akan menjadi “_ _ _

T ”, “ $_T E$ ”, “ TER ”, “ ERM ”, “ $RM_$ ”, “ $M_ _$ ”. N-gram dengan 1 dimensi disebut dengan unigram, jika 2 dimensi disebut bigram, sedangkan 3 dimensi disebut trigram dan apabila lebih dari 3 dimensi pada dasarnya disebut N-gram.

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 didalam suatu dokumen juga digunakan sebagai skema pembobotan dari data tekstual.

3.4. Metode Yang Diusulkan

Metode yang usulkan dalam penelitian ini adalah menggunakan metode pemilihan fitur wrapper. Dari jenis wrapper yang digunakan yaitu *Genetic Algorithm* sebagai metode pemilihan fitur agar akurasi pengklasifikasi *Support Vector Machine* (SVM) bisa meningkat. Penulis menggunakan pengklasifikasi *Support Vector Machine* (SVM) karena sangat mampu mengidentifikasi *hyperplane* terpisah yang memaksimalkan margin antara dua kelas, efisien dan merupakan teknik machine learning yang populer untuk klasifikasi teks, serta memiliki performa yang baik. *Genetic Algorithm* yang penulis terapkan adalah menggunakan *Support Vector Machine* (SVM) yang diuji di dalam tahap wrapper. Lihat gambar 3.1 untuk model yang diusulkan secara lebih detail.



Gambar 3.1 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*. Menurut Gunal (Gunal, 2012) untuk *Genetic Algorithm*, parameter yang dianggap memberikan hasil yang optimal sebagai berikut *population size* diberi angka 15, *probability of initialize* diberi angka 0.9, *probability of crossover* diberi angka 0.9, dan *probability of Generate* diberi angka 1.0. Hasil yang dibandingkan adalah akurasi *Support Vector Machine* (SVM) sebelum menggunakan metode pemilihan fitur dengan akurasi *Support Vector Machine* (SVM) setelah menggunakan metode pemilihan fitur pada wrapper yaitu *Genetic Algorithm*.

3.5 Eksperimen dan Pengujian Model

Pada penelitian ini melakukan proses eksperimen dalam hal pengujian data dengan menggunakan aplikasi RapidMiner 5 dan melakukan *adjustment* pada beberapa parameter agar menghasilkan nilai akurasi yang terbesar. Menurut Rahman dalam wahyuni (2014) parameter yang membutuhkan *adjustment* yaitu *Population size*, *P Initialize*, *P Crossover*, *P Generate*. Sedangkan untuk pengujian model dilakukan menggunakan dataset *tweet* pada akun twitter @pizzahut. Aplikasi yang dibuat dalam penelitian ini menggunakan Macromedia Dreamweaver dengan bahasa PHP. Sedangkan spesifikasi komputer yang digunakan dalam penelitian ini dapat dilihat pada tabel 3.1.

Tabel 3.1. Spesifikasi komputer yang digunakan

Prosesor	Intel(R) Inside core(TM) i3
Memori	2 GB
Harddisk	250 GB
Sistem Operasi	Microsoft Windows 7
Perangkat Lunak	Microsoft Dreamwaver
Bahasa Pemrograman	PHP

T

Tahapan pengujian data untuk mengklasifikasikan *tweet* sebagai berikut:

1. Proses pelatihan dilakukan dengan menginput dataset yang sudah diketahui classnya.
2. Dengan antar muka aplikasi seperti di atas, input *keyword*.

3. Semua teks dari *tweet* yang telah diinput, dihilangkan terlebih dahulu simbol-simbol yang ada di dalamnya.
4. Setelah itu teks tersebut dipisahkan per kata.
5. Kata-kata yang berhubungan dengan sentimen dihitung bobotnya untuk ditentukan classnya.

3.6 Evaluasi dan Validasi Hasil

Validasi dilakukan menggunakan 10 fold cross validation. Sedangkan pengukuran akurasi diukur dengan *confusion matrix* dan kurva ROC untuk mengukur nilai AUC. Dengan *confusion matrix*, akurasi (SVM) *Support Vector Machine* sebelum menggunakan metode pemilihan fitur dan setelah menggunakan metode pemilihan fitur. Tabel 3.2 berikut adalah tampilan *confusion matrix* dan rumus perhitungannya menurut Gorunescu (Moraes, 2013):

Tabel 3.2 *Confusion Matrix*

	Perdicted	
	Positif documents	Negatif Documents
Actual positif documents	#True Positive Samples (TP)	#False Negative Sample (FN)
Actual negatif documents	#True Positive Samples (TP)	#True Negative Samples (TP)

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN}$$

BAB IV

HASIL DAN PEMBAHASAN

4.1. Hasil

4.1.1. Klasifikasi Teks Menggunakan Algoritma *Support Vector Machine*

Data training yang digunakan dalam pengklasifikasian teks ini terdiri dari 100 *tweet* positif dan 100 *tweet* negatif. Sebelum diklasifikasikan, data tersebut harus melalui beberapa tahapan proses agar bisa diklasifikasikan dalam proses selanjutnya, berikut adalah tahapan prosesnya:

1. Pengumpulan Data

Data *tweet* positif disatukan dalam folder dengan nama pos. Sedangkan data *tweet* 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 4.1 menunjukkan hasil perbandingan teks sebelum dan sesudah dilakukan proses *tokenization*.

Tabel 4.1. Perbandingan teks sebelum dan sesudah dilakukan proses *tokenization*

Teks data <i>tweet</i> sebelum dilakukan proses <i>tokenization</i>	<i>Craving my chicken and onion pizza - dont know where I got the combination from but I only like chicken and onion pizzas from @pizzahut</i>
Teks data <i>tweet</i> setelah dilakukan proses <i>tokenization</i>	<i>Craving my chicken and onion pizza dont know where I got the combination from but I only like chicken and onion pizzas from pizzahut</i>

b. *Stopwords*

Dalam proses ini, kata-kata yang tidak relevan akan dihapus, seperti kata *the, of, for, with* 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 4.2 menunjukkan hasil perbandingan teks sebelum dan sesudah dilakukan proses *stopword removal*.

Tabel 4.2. Perbandingan teks sebelum dan sesudah dilakukan proses *Stopwords*

Teks data <i>tweet</i> sebelum dilakukan proses <i>Stopwords Removal</i>	<i>Craving my chicken and onion pizza dont know where I got the combination from but I only like chicken and onion pizzas from pizzahut</i>
Teks data <i>tweet</i> setelah dilakukan proses <i>Stopwords Removal</i>	<i>Craving chicken onion pizza dont know I got combination I chicken onion pizzas pizzahut</i>

a. *Stemming*

Dalam proses ini kata-kata akan dikelompokkan ke dalam beberapa kelompok yang memiliki kata dasar yang sama, seperti *drug, drugged*, dan *drugs* di mana kata dasar dari semuanya adalah kata *drug*. Tabel 4.3 menunjukkan hasil perbandingan teks sebelum dan sesudah dilakukan proses *stemming*.

Tabel 4.3. Perbandingan teks sebelum dan sesudah dilakukan proses *Stemming*

Teks data <i>tweet</i> sebelum dilakukan proses <i>Stemming</i>	<i>Craving chicken onion pizza dont know I got combination I chicken onion pizzas pizzahut</i>
Teks data <i>tweet</i> setelah dilakukan proses <i>Stemming</i>	<i>crave chicken onion pizza dont know i got combin i chicken onion pizza pizzahut</i>

b. Generate 2-gram

Karakter *N-Gram* adalah *n* angka terdekat dari urutan umpan balik yang diberikan. Sebagai contoh 3-gram dari kata “*TERM*” akan menjadi “_ _ T”, “_ T E”, “*TER*”, “*ERM*”, “*RM* _”, “*M* _ _”. *N-gram* dengan 1 dimensi disebut dengan unigram, jika 2 dimensi disebut bigram, sedangkan 3 dimensi disebut trigram dan apabila lebih dari 3 dimensi pada dasarnya disebut *N-gram*.

Tabel 4.4. Perbandingan teks sebelum dan sesudah dilakukan proses Generate 2-gram

Teks data <i>tweet</i> sebelum dilakukan proses <i>Generate 2-gram</i>	crave chicken onion pizza dont know i got combin i chicken onion pizza pizzahut
Teks data <i>tweet</i> setelah dilakukan proses <i>Generate 2-gram</i>	<i>crave</i> <i>crave_chicken</i> <i>chicken</i> <i>chicken_onion</i> <i>onion</i> <i>onion_pizza</i> <i>pizza</i> <i>pizza_dont</i> <i>dont</i> <i>dont_know</i> <i>know</i> <i>know_i</i> <i>i</i> <i>i_got</i> <i>got</i> <i>got_combin</i> <i>combin</i> <i>combin_i</i> <i>i</i> <i>i_chicken</i> <i>chicken</i> <i>chicken_onion</i> <i>onion</i> <i>onion_pizza</i> <i>pizza</i> <i>pizza</i> <i>pizzahut</i> <i>pizzahut</i>

3. 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 4.5.

Tabel 4.5 Table dokumen yang dihitung nilai bobotnya.

Dokumen	Atribut						kelas
	Love_pizza	love	pizzahut	I_love	I_dislike	price	
Pos13.txt (@pizzahut I LOVE YOUR PIZZA)	0,758	0,354	0,003	0,412	0	0	positif
Neg94.txt (@pizzahut I dislike you're prices)	0	0	0,001	0	0,001	0,586	negatif

Berikut adalah perhitungan untuk contoh dokumen pos13.txt dengan fungsi sebagai berikut:

$$\begin{aligned}
 W_{13} &= Y_{13} \cdot X_{13} \\
 &= 1 \cdot (0,758+0,354+0,003+0,412) \\
 &= 1,527
 \end{aligned}$$

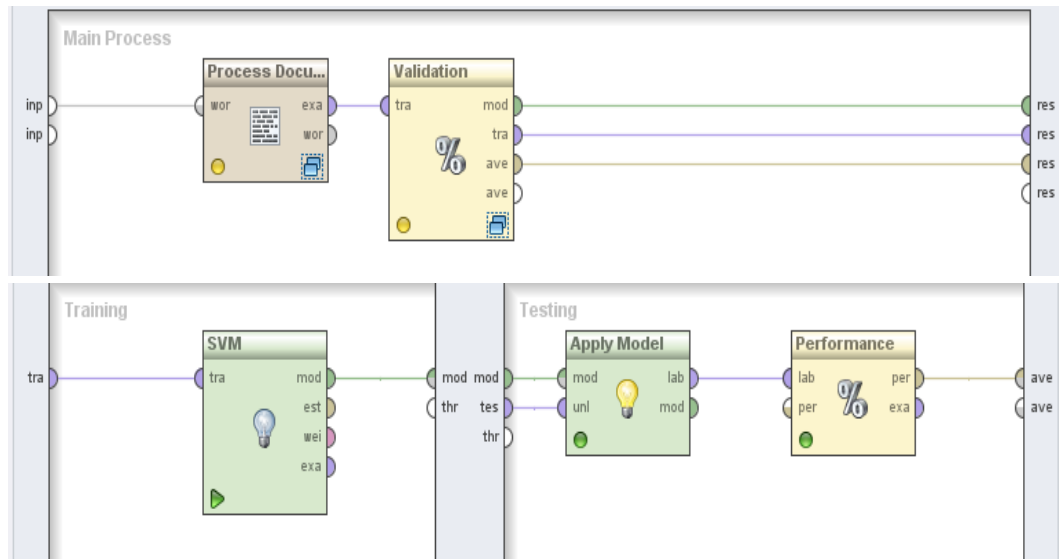
Di mana W_{13} adalah bobot untuk dokumen Pos13.txt yang menentukan klasifikasi kelas positif atau negatif, sedangkan Y_{13} merupakan nilai label untuk dokumen tersebut di mana Y hanya bernilai 1 untuk kelas positif dan -1 untuk kelas negatif. X_{13} nilai-nilai atribut yang ada pada sebuah dokumen yang ingin dihitung bobotnya dan ditentukan kelasnya. Bobot untuk dokumen Pos13.txt adalah 1,527, maka dokumen tersebut diklasifikasikan kedalam kelas positif.

Berikut adalah perhitungan untuk contoh dokumen Neg94.txt dengan fungsi sebagai berikut:

$$\begin{aligned}
 W_{94} &= Y_{94} \cdot X_{94} \\
 &= -1 \cdot (0,433+0,001+0,586) \\
 &= - 1,020
 \end{aligned}$$

Di mana W_{94} adalah bobot untuk dokumen pos94.txt yang menentukan klasifikasi kelas positif atau negatif, sedangkan Y_{94} merupakan nilai label untuk dokumen tersebut di mana Y hanya bernilai 1 untuk kelas positif dan -1 untuk kelas negatif. X_{94} nilai-nilai atribut yang ada pada sebuah dokumen

yang ingin dihitung bobotnya dan ditentukan kelasnya. Bobot untuk dokumen Neg94.txt adalah -1,02, maka dokumen tersebut diklasifikasikan kedalam kelas negatif. Perhitungan di atas dapat dibuat suatu model dengan menggunakan RapidMiner 5. Desain model *Support Vector Machine* (SVM) dapat dilihat pada gambar 4.2.



Gambar 4.1. Desain model *Support Vector Machine* (SVM) menggunakan RapidMiner

4.1.2. 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.

4.1.3. Optimasi Model dengan Metode *Genetic algorithm*

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

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

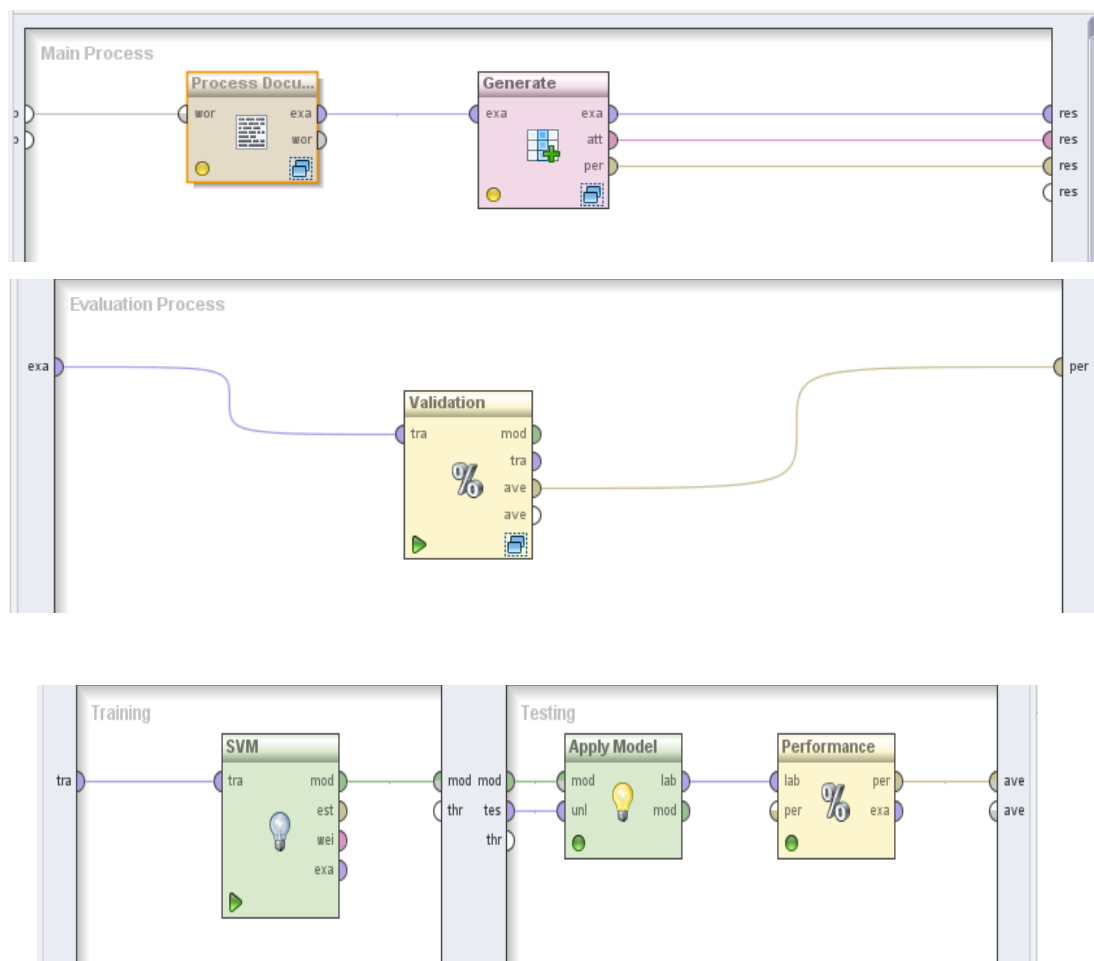
4.1.4. Eksperimen Terhadap Indikator Model

Berdasarkan percobaan yang dilakukan oleh Wahyuni (2014) percobaan penelitian ini dimulai dengan melakukan *adjustment* pada metode Genetik algoritma dengan nilai *Population size* yang berada pada nilai *default*, kemudian dilanjutkann dengan kelipatan 5 hingga nilai *Population size* yang menghasilkan akurasi paling tinggilah yang akan digunakan pada langkah percobaan selanjutnya. Kemudian dilakukan *adjustment* pada nilai *P Initialize*, *P Crossover*, *P Generate* dengan keliapan 0.2 untuk mendapatkan model yang baik. Sedangkan pada metode SVM (*Support Vector Machine*) berdasarkan penelitian Ilhan & Tezel (2013) *adjustment* dilakukan pada nilai parameter *C* sebagai parameter pengendali dengan nilai $=1.0$. Dan hasil akhir akurasi tertinggi dicapai pada saat nilai *Population size*=15, *P Initialize*=0.9, *P Crossover*=0.9, *P Generate*=1.0 dan . Tabel indikator dan hasil pengujian dapat dilihat pada tabel 4.6.

Tabel 4.6. Tabel indikator dan hasil pengujian

Population size	P Initialize	P Crossover	P Generate	C	Akurasi
5	0.5	0.5	0.1	1.0	80.56%
10	0.7	0.7	1.0	1.0	92.50%
15	0.5	0.5	1.0	1.0	86.50%
15	0.9	0.9	1.0	1.0	93.50%
15	1.1	1.1	1.0	1.0	92.50%
20	0.9	0.9	1.0	1.0	89.43%

Dalam penyesuaian indikator pada Genetic algorithm, akurasi paling tinggi diperoleh dengan kombinasi population size=15, p initialize=0.9, p crossover=0.9, dan p generate=1.0. Hasil akurasi mencapai 93.20%. 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 4.2.



Gambar 4.2. Desain model *Support Vector Machine* dan *Genetic Algorithm* didalam RapidMiner

4.2. Pembahasan

Dengan memiliki model klasifikasi teks *tweet* pada twitter , dapat digunakan untuk menindaklanjuti suatu strategi perusahaan maupun organisasi dalam menentukan keputusan dari akun twitter mereka, yaitu dengan

mengidentifikasi *tweet* tersebut sentimen yang positif maupun yang negatif. Dari data *tweet* 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 *tweet* 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 wrapper Genetic Algorithm, algoritma *Support Vector Machine* (SVM) sendiri sudah menghasilkan akurasi sebesar 63.50% dan nilai AUC 0.875. Akurasi tersebut masih kurang akurat, sehingga perlu ditingkatkan lagi menggunakan metode wrapper Genetic Algorithm. Akurasi algoritma *Support Vector Machine* (SVM) meningkat menjadi 93.50% dan nilai AUC 0.964 seperti yang bisa dilihat pada tabel 4.8.

Table 4.7 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 <i>tweet</i> positif	97	97
Sukses klasifikasi <i>tweet</i> negatif	30	90
Akurasi Model	63.50%	93.50%
AUC	0.903	0.940

4.2.1. 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 4.9 dan setelah ditambahkan metode Genetic algorithm yang bisa dilihat pada tabel 4.10.

Table 4.8 *Confusion Matrix* Model *Support Vector Machine* (SVM) sebelum penambahan metode Genetic Algorithm

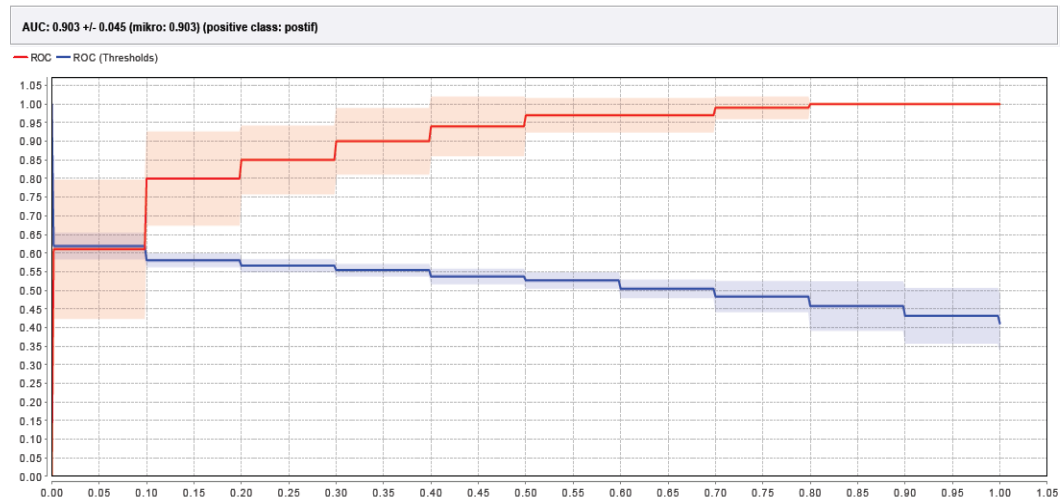
Akurasi <i>Support Vector Machine</i> (SVM): 63.50% + - 8.38% (mikro 63.50)			
	True negative	True positif	Kelas precision
Pred.negatif	30	3	90.91 %
Pred.Positif	70	97	58.08%
Kelas recall	30.00%	97.00%	

Table 4.9 Confusion Matrix Model Support Vector Machine (SVM) sesudah penambahan metode Genetic Algorithm

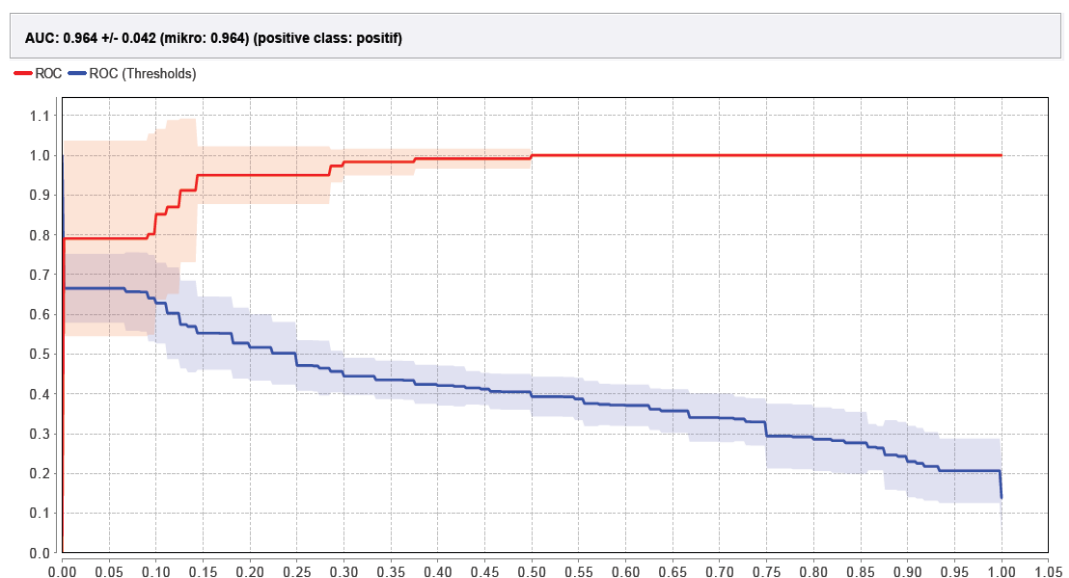
Akurasi <i>Support Vector Machine</i> (SVM): 93.50% + - 5.50% (mikro 93.50%)			
	True negative	True positif	Kelas precision
Pred.negatif	90	3	96.77%
Pred.Positif	10	97	90.65%
Kelas recall	2.14%	97.99%	

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} = \frac{97 + 90}{97 + 90 + 10 + 3} = \frac{187}{200} = 0.935$$

Berikut adalah tampilan kurva ROC yang akan dihitung nilai AUC-nya. Gambar 4.3 adalah kurva ROC untuk model *Support Vector Machine* (SVM) sebelum menggunakan metode Genetic Algorithm dan gambar 4.4 adalah kurva ROC untuk model *Support Vector Machine* (SVM) setelah menggunakan metode Genetic Algorithm.



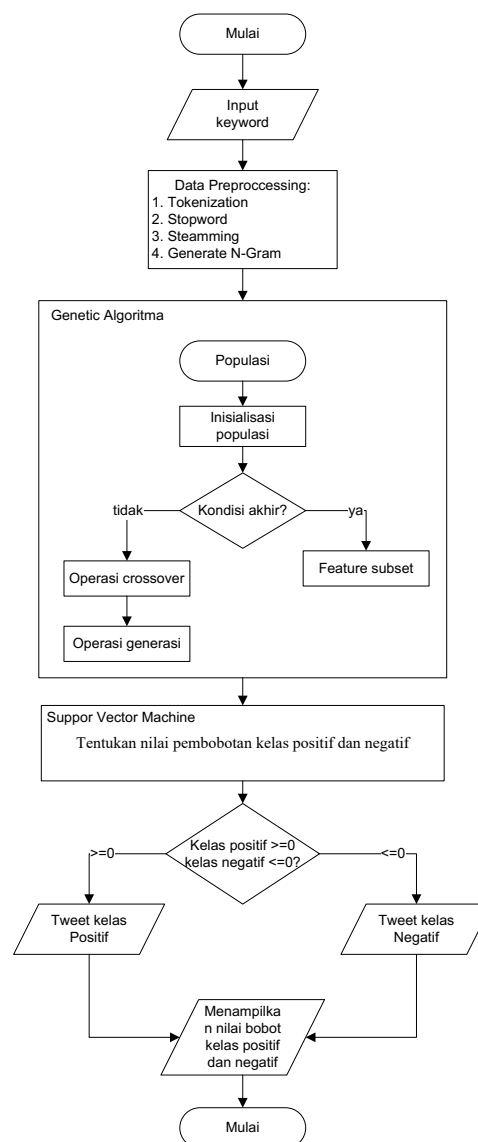
Gambar 4.3 Kurva ROC Model *Support Vector Machine* (SVM) sebelum menggunakan metode Genetik Algorithm



Gambar 4.4 Kurva ROC Model *Support Vector Machine* (SVM) setelah menggunakan metode Genetik Algorithm

4.3. 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 4.5 adalah diagram alir dari tahapan proses klasifikasi pada aplikasi yang penulis buat.

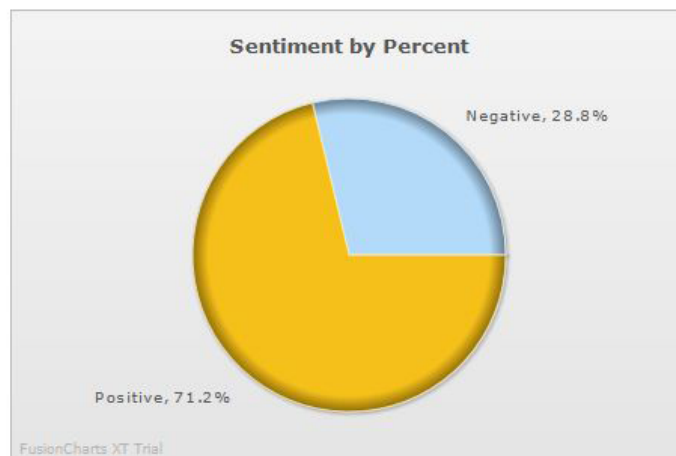


Gambar 4.5. Diagram alir tahapan proses klasifikasi dengan *Support Vector Machine* (SVM) menggunakan Metode Genetic algorithm

Berikut adalah tampilan rancangan aplikasi dan implementasinya. Gambar 4.6. adalah ketika aplikasi diinput *keyword* “@pizzahut” dan menampilkan sentimen analisis dalam diagram lingkaran beserta jumlah presentase *tweet* positif dan *tweet* negatif serta ditampilkan tweet-tweet yang berkaitan dengan *keyword* . Gambar 4.6 adalah ketika aplikasi diinput *keyword* “@pizzahut”.

SentTwimentAnalysis

Sentiment analysis for



Tweets about: "pizzahut"

Craving my chicken and onion pizza - dont know where I got the combination from but I only like chicken and onion pizzas from @pizzahut

RT @httpswifi: @pizzahut I love you still

@pizzahut i love you so much baby

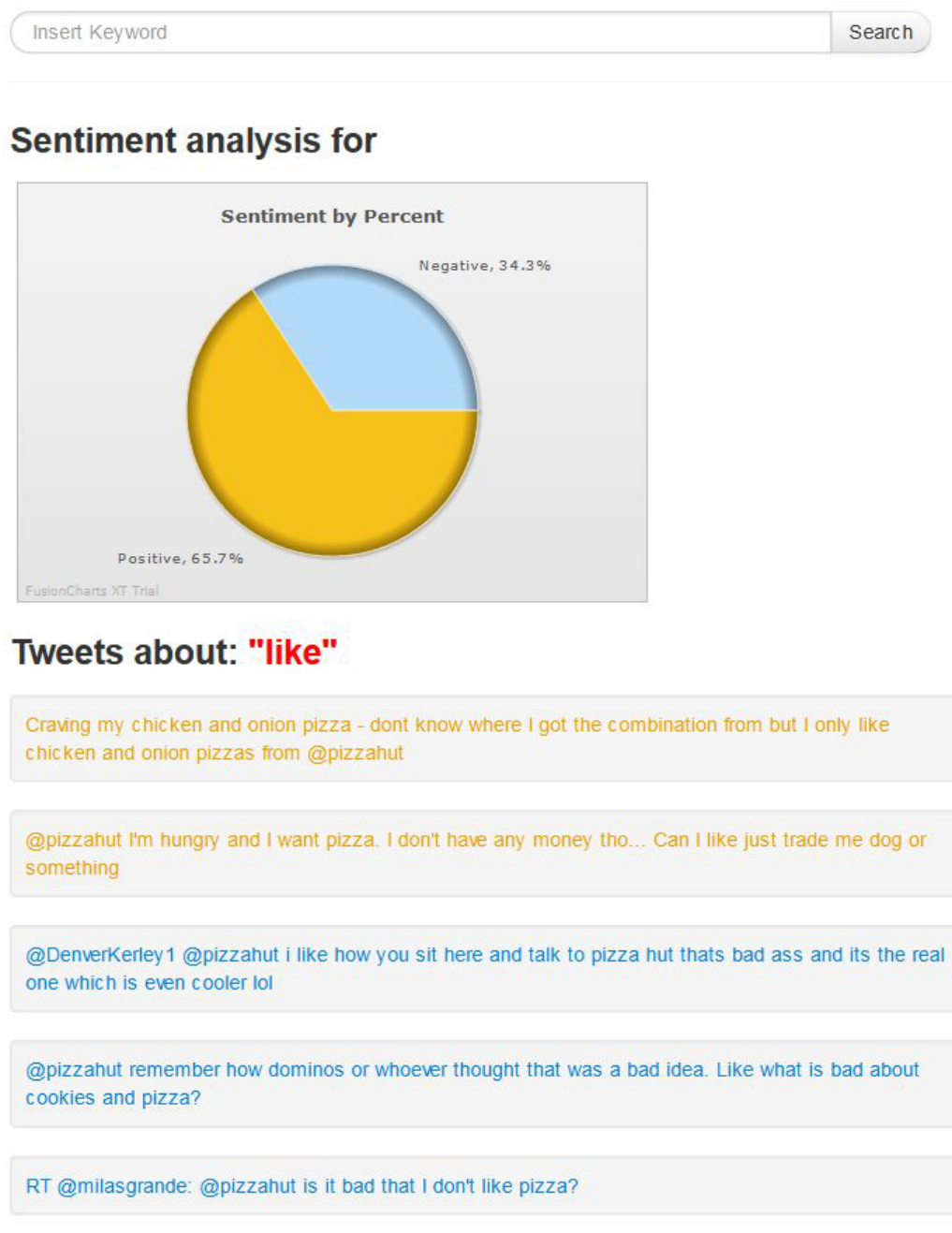
@pizzahut I love you so much

@pizzahut I love you, it makes me emotional to know

Gambar 4.6 Tampilan aplikasi sentimen analisi *tweet*

Pengujian aplikasi sentimen analisis *tweet* dengan keyword yang berbeda yaitu “like” hasil nya dapat dilihat dari gambar 4.7.

SentTwimentAnalysis



Gambar 4.6 Tampilan aplikasi sentimen analisi *tweet* dengan *keyword* berbeda

4.4 Implikasi Penelitian

Implikasi penelitian ini mencakup beberapa aspek, di antaranya:

1. Implikasi terhadap aspek sistem

Hasil evaluasi menunjukkan penerapan Genetic algorithm dapat meningkatkan akurasi *Support Vector Machine* (SVM) dan merupakan metode yang cukup baik dalam mengklasifikasi *tweet* pada twitter. Program aplikasi yang dibuat dalam penelitian ini menggunakan ketentuan spesifikasi komputer seperti prosesor Intel inside core i3, memori 2GB, harddisk 250GB, sistem operasi Windows 7, dengan menggunakan perangkat lunak Microsoft Dreamwaever dan bahasa pemrograman PHP.

2. Implikasi terhadap aspek Manajerial

Dengan demikian penerapan program aplikasi SentiTweetAnalysis ini dapat digunakan untuk menindaklanjuti suatu strategi perusahaan maupun organisasi dalam menentukan keputusan dari akun twitter mereka sebagai bahan evaluasi keputusan, yaitu dengan mengidentifikasi *tweet* tersebut sentimen yang positif maupun yang negatif. Dalam implementasi program aplikasi SentiTweetAnalysis ini perlu dibentuk SOP (*Standard Operating Procedure*) untuk mengetahui standar proses-proses yang ada dalam sistem, diadakan sosialisai atau penyuluhan untuk memberitahukan tata cara penggunaan aplikasi SentiTweetAnalysis tersebut, penunjukan terhadap administrator yang ditugaskan untuk mengelola aplikasi SentiTweetAnalysis, mengadakan pelatihan terhadap para pengguna aplikasi SentiTweetAnalysis agar dapat berjalan sesuai dengan SOP yang telah ditentukan, kemudian melakukan perawatan (*maintanace*) terhadap aplikasi

3. Implikasi terhadap aspek penelitian lanjutan

Penelitian seperti ini dapat dikembangkan oleh unit dan domain yang berbeda seperti review kamera, review hotel, dan sebagainya. Menggunakan pengklasifikasi lain yang mungkin di luar *Supervised Learning*. Sehingga bisa dilakukan penelitian yang berbeda dari umumnya yang sudah ada seperti menggunakan pengkalsifikasian Naive Bayes, Artificial Neural Network, Maximum Entropy dan lainnya.

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

1. Untuk mengklasifikasikan teks sentimen analisis dengan data berupa data *tweet* pada twitter, salah satu pengklasifikasi yang dapat digunakan adalah pengklasifikasi menggunakan algoritma *Support Vector Machine* (SVM). Hal ini dikarenakan *Support Vector Machine* (SVM) merupakan salah satu algoritma klasifikasi linier yang mampu mengidentifikasi *hyperplane* terpisah yang memaksimalkan margin antara dua kelas. Selain itu *Support Vector Machine* (SVM). juga sangat populer digunakan untuk klasifikasi teks dan memiliki performa yang baik.
2. Metode wrapper telah terbukti dapat meningkatkan akurasi pengklasifikasi *Support Vector Machine* (SVM) dalam hal pengolahan data. Data *tweet* dapat diklasifikasi dengan baik ke dalam bentuk positif dan negatif. Akurasi *Support Vector Machine* (SVM) sebelum menggunakan penggabungan metode wrapper Genetic Algorithm akurasi yang dihasilkan yaitu sebesar 63.50% dan nilai AUC 0.875. Sedangkan setelah menggunakan penggabungan metode Genetic algorithm, akurasinya meningkat hingga mencapai 93.50% dan nilai AUC 0.940. Peningkatan akurasi mencapai 30%. Untuk mendukung penelitian ini dikembangkan aplikasi sentimen analisis mengklasifikasikan *tweet* positif dan negatif menggunakan bahasa pemrograman PHP.
3. Model yang dibentuk dalam penelitian ini diterapkan pada akun twitter Pizza Hut, sehingga dapat dilihat secara langsung hasilnya dalam bentuk positif dan negatif. Hal ini dapat membantu Pizza Hut sebagai restoran makanan yang paling digemari banyak orang sehingga dapat dianalisa produk dan layanan mereka dari opini-opini yang pelanggan utarakan pada akun twitter mereka.

5.2. Saran

Walaupun pengklasifikasi *Support Vector Machine* (SVM) sudah sering digunakan dan mempunyai performa yang baik dalam mengklasifikasikan teks, namun ada beberapa hal yang dapat ditambahkan untuk penelitian selanjutnya:

1. Untuk membuat program aplikasi sentiTweetAnalysis dapat menggunakan spesifikasi komputer yang digunakan dalam penelitian ini, akan tetapi tidak menutup kemungkinan untuk menggunakan spesifikasi yang lebih rendah misal dengan menggunakan Intel pentium inside, memori 1GB, harddisk 160GB, sistem operasi Windows XP.
2. Dalam implementasi program aplikasi SentiTweetAnalysis ini perlu dibentuk SOP, sosialisai, penunjukan pengguna seperti administrator dan perawatan (*maintanace*).
3. Untuk meningkatkan hasil optimasi dapat dilakukan dengan menggunakan pengklasifikasi lain yang mungkin di luar *Supervised Learning* seperti Menggunakan Naive Bayes, Artificial Neural Network dan lainnya.

DAFTAR REFERENSI

- Aydin, I., Karakose, M., & Akin, E. (2011). A multi-objective artificial immune algorithm for parameter optimization in support vector machine. *Applied Soft Computing*, 11(1), 120–129. doi:10.1016/j.asoc.2009.11.003
- Balahur, A., Mihalcea, R., & Montoyo, A. (2014). Computational approaches to subjectivity and sentiment analysis: Present and envisaged methods and applications. *Computer Speech & Language*, 28(1), 1–6. doi:10.1016/j.csl.2013.09.003
- 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*, 53, 453–462. doi:10.1016/j.proeng.2013.02.059
- Campo-Ávila, J. Del, Moreno-Vergara, N., & Trella-López, M. (2013). Bridging the Gap Between the Least and the Most Influential Twitter Users. *Procedia Computer Science*, 19(Ant), 437–444. doi:10.1016/j.procs.2013.06.059
- Chen, J., Huang, H., Tian, S., & Qu, Y. (2009). Feature selection for text classification with Naïve Bayes. *Expert Systems with Applications*, 36(3), 5432–5435
- Chou, J.-S., Cheng, M.-Y., Wu, Y.-W., & Pham, A.-D. (2014). Optimizing parameters of support vector machine using fast messy genetic algorithm for dispute classification. *Expert Systems with Applications*, 41(8), 3955–3964. doi:10.1016/j.eswa.2013.12.035
- Dehkharghani, R., Mercan, H., Javeed, A., & Saygin, Y. (2014). Sentimental causal rule discovery from Twitter. *Expert Systems with Applications*, 41(10), 4950–4958. doi:10.1016/j.eswa.2014.02.024
- 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. doi:10.1016/j.eswa.2013.05.057
- Gorunescu, F. (2011). *Data Mining Concept Model Technique*.
- Gunal, S. (2012). Hybrid feature selection for text classification ", 20.
- Haddi, E., Liu, X., & Shi, Y. (2013). The Role of Text Pre-processing in Sentiment Analysis. *Procedia Computer Science*, 17, 26–32. doi:10.1016/j.procs.2013.05.005
- Han, J., & Kamber, M. (2007). *Data Mining Concepts and Techniques*

- Ilmiah, D. K., Akhir, T., Studi, P., Informatika, T., Komputer, F. I., Dian, U., Udinus, P. S. I. (2014). Dokumen Karya Ilmiah | Tugas Akhir | Program Studi Teknik Informatika - S1 | Fakultas Ilmu Komputer | Universitas Dian Nuswantoro Semarang | 2014, 0–1.
- Khan, F. H., Bashir, S., & Qamar, U. (2014). TOM: Twitter opinion mining framework using hybrid classification scheme. *Decision Support Systems*, 57, 245–257. doi:10.1016/j.dss.2013.09.004
- Khan, K., Baharudin, B., & 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
- Larose, D. T. (n.d.). *METHODS AND*.
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*. doi:10.1016/j.asej.2014.04.011
- Montejo-Ráez, A., Martínez-Cámara, E., Martín-Valdivia, M. T., & Ureña-López, L. A. (2014). Ranked WordNet graph for Sentiment Polarity Classification in Twitter. *Computer Speech & Language*, 28(1), 93–107. doi:10.1016/j.csl.2013.04.001
- 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.
- Passonneau, R. (2011). Sentiment Analysis of Twitter Data. *Proceedings of the Workshop on Languages in Social Media*, (ISBN: 978-1-932432-96-1), 30–38.
- Schoefegger, K., Tammet, T., & Granitzer, M. (2013). A survey on socio-semantic information retrieval. *Computer Science Review*, 8, 25–46. doi:10.1016/j.cosrev.2013.03.001
- Suyanto. (2014). *Artificial Intelligence Searching-Reasoning-Palnnng-Learning*. Bandung: Informatika Bandung.
- Wang, Z., Shao, Y.-H., & 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
- Weise, T. (2009). *Global Optimization Algorithms – Theory and Application – (Second Edi.)*. it-weise.de (self-published).
- Widodo, Prabowo Pudjo, Rahmadya Trias Handayanto dan Herlawati. (2013). Penerapan Data Mining dengan Matlab. Bandung: Rekayasa Sains.

- Witten, H. I., Frank, E., & Hall, M. A. (2011). *Data Mining Practical MachineLearning Tools And Technique*. Burlington: Elsevier Inc.
- Wu, Xindong, Vipin Kumar. (2009). *The Top Ten Algorithms in Data Mining*.
- Yang, P., Liu, W., Zhou, B. B., Chawla, S., & Albert, Y. (2010). Ensemble-based wrapper methods for feature selection and class imbalance learning, 1–12.
- Ye, Q., Zhang, Z., & Law, R. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Systems with Applications*, 36(3), 6527–6535. doi:10.1016/j.eswa.2008.07.035
- Zhao, M., Fu, C., Ji, L., Tang, K., & Zhou, M. (2011). 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
- Zukri, Zainudin. (2014) *Algoritma Genetika Metode Komputasi Evolusioner untuk Menyelesaikan Masalah Optimasi*. Yogyakarta: Andi Offset.

Lampiran 1. Tabel dataset *tweet* positif

No.	Dokumen	<i>Tweet</i>
1	Pos1	<i>Craving my chicken and onion pizza - dont know where I got the combination from but I only like chicken and onion pizzas from @pizzahut</i>
2	Pos2	<i>RT @httpswifi: @pizzahut I love you still</i>
3	Pos3	<i>@pizzahut i love you so much baby</i>
4	Pos4	<i>@pizzahut I love you so much</i>
5	Pos5	<i>@pizzahut I love you, it makes me emotional to know</i>
6	Pos6	<i>@pizzahut i love you more than i love my dog tbh</i>
7	Pos7	<i>@pizzahut can't wait. i love you ??</i>
8	Pos8	<i>Please follow me @pizzahut I love your food :)</i>
9	Pos9	<i>@pizzahut Can you follow me, this follow limit is pissing me off, Btw your pizza is the best. I love you</i>
10	Pos10	<i>@pizzahut i love meatlovers so much that i cry</i>
11	Pos11	<i>@pizzahut I would love to try the Hotwings.They look tasty.</i>
12	Pos12	<i>@pizzahut i love you man ????</i>
13	Pos13	<i>@pizzahut I LOVE YOUR PIZZA</i>
14	Pos14	<i>I'm in love with you @pizzahut And I know that love is just a shout into the void, and that oblivion is inevitable...& I ...</i>
15	Pos15	<i>I'm gonna sleep now I love @pizzahut more than my mom bye</i>
16	Pos16	<i>@pizzahut can you follow me just because I love your pizza</i>
17	Pos17	<i>@pizzahut i freaking love you</i>

18	Pos18	<i>@pizzahut i freaking love you</i>
19	Pos19	<i>@pizzahut PLEASE FOLLOW ME PLEASE PLEASE PLEASE PLEASE I LOVE YOUR PIZZA !</i>
20	Pos20	<i>@pizzahut i love your pizza pizzahut are the best small please follow me please</i>
21	Pos21	<i>@pizzahut I want pizza! I love you</i>
22	Pos22	<i>@pizzahut I love you guys more than I love myself</i>
23	Pos23	<i>@pizzahut OMG PIZZA HUT THANK YOU FOR MAKING MY NIGHT I LOVE YOU!??????</i>
24	Pos24	<i>I love @pizzahut but my mom never gets me it??</i>
25	Pos25	<i>@pizzahut tweet this #BestFandom2014 Directioners and tag me please! make me happy I FUCKING LOVE PIZZA AND I CHOOSE PIZZA HUT DO</i>
26	Pos26	<i>I kind of want @pizzahut to follow me bc I love pizza so much ????</i>
27	Pos27	<i>@pizzahut i am madly and deeply in love with your pizza</i>
28	Pos28	<i>@pizzahut I love you and your cookie pizza miracles</i>
29	Pos29	<i>I LOVE PIZZA! @pizzahut #TeenageMutantNinjaTurtles #Cowabunga</i>
30	Pos30	<i>@pizzahut hiya it's my birthday!! Can you please tweet me or something?? I love youuuuuu ??????</i>
31	Pos31	<i>@pizzahut I love Pizza Hut and Heavenly Steals! @pizzahut #WINGSTREET</i>
32	Pos32	<i>@pizzahut i love you so much & ' i cant live without</i>

		<i>you??????</i>
33	Pos33	<i>Eating the best pizza ever. I so love @PizzaHut</i>
34	Pos34	<i>@pizzahut I LOVE YOU, YOU'RE THE BEST #BestFandom2014 Beliebers</i>
35	Pos35	<i>@pizzahut @pizzahutuk I'm hungry and I love you, please help</i>
36	Pos36	<i>I had this beautiful pizza yesterday ?? thank you @pizzahut for a lovely meal????</i>
37	Pos37	<i>@d0egirl @pizzahut yeah! I love their crust!</i>
38	Pos38	<i>@pizzahut reply to me or follow me bc i love you i only order pizza hut pizza</i>
39	Pos39	<i>I love you @pizzahut you're my life</i>
40	Pos40	<i>@pizzahut #WingStreet #I love pizza hut and heavenly steals!</i>
41	Pos41	<i>@pizzahut thank you for existing. I love you</i>
42	Pos42	<i>@pizzahut I love you can I get a discount on your pizzas</i>
43	Pos43	<i>@pizzahut Pizza it's good I love pizza</i>
44	Pos44	<i>@pizzahut I FUCKING LOVE YOU</i>
45	Pos45	<i>@pizzahut I love pizza hut..my grand-daughters want to eat there every Wed. after Ballet class..</i>
46	Pos46	<i>@pizzahut i love you as much as i love bands and this is something really important</i>
47	Pos47	<i>@pizzahut "I love Pizza Hut and Heavenly Steals! @pizzahut #WINGSTREET".</i>
48	Pos48	<i>@pacsunharold @pizzahut I LOVE PIZZA HUT OKAY?!</i>

49	Pos49	<i>@pizzahut I love Pizza Hut, does Pizza Hut love me back?</i>
50	Pos50	<i>@pizzahut i love you a lottle</i>
51	Pos51	<i>Pizza makes me so happy ... I love you @pizzahut You taste so good btw ????</i>
52	Pos52	<i>I so love @pizzahut! #MarriedAtFirstSight</i>
53	Pos53	<i>@pizzahut @PizzaHutCares Yeah, terrible customer service. I love pizza...but, this ruined my experience. #shouldhavedominos</i>
54	Pos54	<i>I love you @pizzahut ? made my day</i>
55	Pos55	<i>i just...really really really like you. @pizzahut</i>
56	Pos56	<i>If you delivered like 3 boxes of free pizza to my house I would legit cry @pizzahut</i>
57	Pos57	<i>bought @pizzahut all week because im to lazy to cook, my gosh that pizza was delicious, I felt like I was in heaven, I deserve a follow</i>
58	Pos58	<i>@pizzahut hey my brothers and I only like Pizza Hut help</i>
59	Pos59	<i>@PizzaHut I would like to order something!</i>
60	Pos60	<i>@pizzahut I like pizza more than myself is that strange</i>
61	Pos61	<i>@pizzahut Smh. Terrible service. I like Pizza Hut, but I never get it delivered</i>
62	Pos62	<i>Major corporations like @pizzahut focus solely on profit. I was told they only had one driver on staff. One driver getting paid min. wage</i>
63	Pos63	<i>@pizzahut tweet this #BestFandom2014 Directioners and tag</i>

		<i>me please! make me happy I FUCKING LOVE PIZZA AND I CHOOSE PIZZA HUT DO IT</i>
64	Pos64	<i>i just...really really really like you. @pizzahut</i>
65	Pos65	<i>@pizzahut it's the place I like to store my hot dog, yes.</i>
66	Pos66	<i>My office smells like @pizzahut. Now I want Pizza Hut. ??</i>
67	Pos67	<i>@pizzahut i like the way you think, can my fave pizza follow me ??</i>
68	Pos68	<i>@pizzahut y'all still up or nah? I need a large chicken, sausage, and pineapple pan crust like ASAP.</i>
69	Pos69	<i>Why isn't @pizzahut open 24/7, like I crave pizza at 2am.</i>
70	Pos70	<i>@pizzahut bruh my mom worked for you like I'm obligated to get a follow. I love you</i>
71	Pos71	<i>@pizzahut why are you following everyone but me like I model my ice around you</i>
72	Pos72	<i>@pizzahut I THINK ILL MISS YOU FOREVER LIKE THE STARS MISS THE SUN IN THE NIGHT TIME SKY</i>
73	Pos73	<i>@pizzahut hi I would like 2 large pizzas. One with green peppers, mushrooms, and beef. The other is just cheese. Thank you.</i>
74	Pos74	<i>@pizzahut I went to pizza hut on i30 its nasty nasty people are rude district manager kathy dm said if you dont like it go somewhere else</i>
75	Pos75	<i>@pizzahut bc luke likes u and i like u AND luke so cough ya</i>
76	Pos76	<i>@pizzahut I like your crust</i>

77	Pos77	<i>@pizzahut have you heard "Whys it so hot?" By @Jarinus ? They mention Pizza Hut. I think you'd like it.</i>
78	Pos78	<i>I WANT A @pizzahut FOLLOW. THATS LIKE MY FAV PIZZA PLACE</i>
79	Pos79	<i>@pizzahut my friend told me today she doesn't like Pizza Hut so I dragged her to eat at yours anyways and she said she liked it. I WON</i>
80	Pos80	<i>@pizzahut I like your background and your pizza is delicious</i>
81	Pos81	<i>@pizzahut I would just like some breadsticks and marinara sauce</i>
82	Pos82	<i>@pizzahut a few days ago a guy i know said he didn't like pizza hut i hit him you owe me big time</i>
83	Pos83	<i>@pizzahut I eat pizza like every week, I think I deserve a follow</i>
84	Pos84	<i>@pizzahut I love you cus you are my fav food. Yummy!</i>
85	Pos85	<i>@pizzahut I'm hungry and I want pizza. I don't have any money tho... Can I like just trade me dog or something</i>
86	Pos86	<i>@pizzahut I like the inner decoration of your restaurants it's very relaxing with the light lights.</i>
87	Pos87	<i>@pizzahut and i feel like im on top of the world</i>
88	Pos88	<i>@pizzahut follow me because i like pizza</i>
89	Pos89	<i>@pizzahut @Cinnabon I KNOW ITS LIKE A PEICE OF ME WOULD BE MISSING</i>
90	Pos90	<i>whats your favorite kind of topping on pizza? I like pepperoni</i>

		<i>@pizzahut</i>
91	Pos91	<i>@pizzahut I bet, and it tastes like you got the recipe from Master Splinter! You make the best in the world and always have!</i>
92	Pos92	<i>And my night is complete thanks to #tmnt and @pizzahut I seriously feel like Michelangelo right now! #cowabunga</i>
93	Pos93	<i>@Dominos_UK still no response. Looks like I will be going to @pizzahut next time. #terriblecustomerservice</i>
94	Pos94	<i>izzahut here in Ottawa, Canada. Great pizza, folks! I don't even like pizza, but I enjoyed yours. Cheers!</i>
95	Pos95	<i>@Deadlox474 @pizzahut i was jk i just tweeted that like that i dont eat Pizza Hut</i>
96	Pos96	<i>@pizzahut Looks like I know what I am having for dinner.... Magic : 3</i>
97	Pos97	<i>@pizzahut do you love me like I love you????</i>
98	Pos98	<i>@PlZZASS @pizzahut I would like to see a sample of your pizza and test it out.</i>
99	Pos99	<i>@pizzahut I want pizza like right now.</i>
100	Pos100	<i>@pizzahut I HAD LIKE 7 SLICES AND THE HERSEY DESSERT</i>

Lampiran 2. Tabel dataset *tweet* Negatif

No.	Dokumen	<i>Tweet</i>
1	Neg1	<i>@pizzahut Fine then, ignore me. I'm going to Burger King anyways, phht. Your loss. Bye. I don't like you. Hmph. *Takes bite out of pizza*</i>
2	Neg2	<i>I really don't recommend @pizzahut in Elk City. If the service is enough to get my calm Mrs. worked up then it's screaming bad!</i>
3	Neg3	<i>@pizzahut omg I actually unfollowed you. This is worse and I feel so bad. Refollowed immediately :)</i>
4	Neg4	<i>@pizzahut is it bad that I'm having pizza I got last night for breakfast@ELOUGANTHARRY</i>
5	Neg5	<i>@pizzahut remember how dominos or whoever thought that was a bad idea. Like what is bad about cookies and pizza?</i>
6	Neg6	<i>RT @milasgrande: @pizzahut is it bad that I don't like pizza?</i>
7	Neg7	<i>@pizzahut is it bad that ive only tasted ur pizza like twice</i>
8	Neg8	<i>Is it bad that I'm waiting for @pizzahut to tweet and follow me?</i>
9	Neg9	<i>I think @pizzahut is bad at math.... http://t.co/4rYRztHC3</i>
10	Neg10	<i>@pizzahut I've had Pizza Hut two days in a row is this bad?</i>
11	Neg11	<i>@pizzahut having a buffet is a bad idea with me around - lets be havin yae wee pizza :)</i>
12	Neg12	<i>@DenverKerley1 @pizzahut i like how you sit here and talk</i>

		<i>to pizza hut thats bad ass and its the real one which is even cooler lol</i>
13	Neg13	<i>Is it bad when the @pizzahut employees recognize me?</i>
14	Neg14	<i>Why is the service so ridiculously bad in @pizzahut every single time?!</i>
15	Neg15	<i>is it bad that my goal for the year is to get @pizzahut to follow me.?</i>
16	Neg16	<i>is it bad that whenever @pizzahut tweets i receive it so i can know when to get discounts and shit</i>
17	Neg17	<i>is that bad that whenecer @pizzahut tweets i get them #totallynotobsessed</i>
18	Neg18	<i>Dear @pizzahut, while I appreciate the need for catchy email marketing, there is no excuse for bad math. http://t.co/brHTV5awgk</i>
19	Neg19	<i>Is it bad that I dont get excited when @pizzahut tweets me?</i>
20	Neg20	<i>If i get @pizzahut 2 days in a row is that bad? ??</i>
21	Neg21	<i>@cmcfall @lanniano10 @PizzaExpress @pizzahut @Dominos_UK hi pizza companies,chris had a bad salad experience & is looking some free pizza?</i>
22	Neg22	<i>@pizzahut the problem is I really shouldn't...bring a lunch every day, just want it so bad!!</i>
23	Neg23	<i>Is it bad that it's 11pm and I'm craving pizza @pizzahut</i>
24	Neg24	<i>@buteradinosaur @pizzahut to bad he is already with me??</i>
25	Neg25	<i>@Sydney_Nicolex3 @pizzahut I don't like pizza but they're</i>

		<i>the only pizza I like and doesn't make me sick!</i>
26	Neg26	<i>@drasticharry @pizzahut i don't like pizza hut i only like little caesars</i>
27	Neg27	<i>I don't like @pizzahut and I don't like @blakeshelton telling me what to eat or dinner</i>
28	Neg28	<i>@pizzahut @PizzaHutCares your Pizza Hut on wornall and 76th in Kansas City is really bad. You should fire everybody here, seriously.</i>
29	Neg29	<i>Hey @blakeshelton my dog really liked your @pizzahut pizza today... I don't blame him, it's pretty awesome. http://t.co/aax</i>
30	Neg30	<i>i feel like im betraying @pizzahut when i have to buy little caesars bc i don't have any money.</i>
31	Neg31	<i>@pizzahut is cheating on me. Don't act like I can't see your replies to everyone else. It's okay I still love you.??????</i>
32	Neg32	<i>@pizzahut I feel like my love for you only goes one way , as in you don't love me ...</i>
33	Neg33	<i>@pizzahut i'll just keep the pizzas bc i don't like football</i>
34	Neg34	<i>I follow @pizzahut on twitter and don't even eat pizza, I just like their tweets.</i>
35	Neg35	<i>@pizzahut If I don't like this BBQ pizza, I'm holding @blakeshelton personally responsible.</i>
36	Neg36	<i>Dear @pizzahut "you and i we don't want to be like them we can make it till the end no nothing can come between you and i ,,"</i>

37	Neg37	<i>@pizzahut , I don't like how the minimum for a delivery has to be \$9.99. it's not my fault I had a discount.. u r slacking.</i>
38	Neg38	<i>@lukeftpizza @pizzahut @Luke5SOS @Ashton5SOS I don't like coke, it makes my teeth feel weird</i>
39	Neg39	<i>@pizzahut ILY but I don't like pizza??</i>
40	Neg40	<i>@pizzahut I eat your pizza, like all the time, and yet, I still don't have a follow.</i>
41	Neg41	<i>@Pvptart @pizzahut good description moron. like how many pieces? I don't want no personal fucking 4piece pizza @at822pm</i>
42	Neg42	<i>@pizzahut i know you see me but you don't like me??</i>
43	Neg43	<i>@pizzahut why don't you love me Pizza Hut. I order your pizzas at least once a week to my school like please love me</i>
44	Neg44	<i>I just ate @pizzahut pizza with @dominos dip. I feel like I'm cheating on one of them. And I don't care. Morals are for sissies.</i>
45	Neg45	<i>Walk into @pizzahut: "sorry we don't have any dough" like I really just hate this place more every time I walk in</i>
46	Neg46	<i>@jenizzleeeee @pizzahut right?! I'm craving one right now. I just don't want to look like a moron trying to call and order one.</i>
47	Neg47	<i>is it really bad that I don't like @pizzahut pizzas but I like the ice cream</i>
48	Neg48	<i>RT @Ashuwee: @LAKings hockey and @pizzahut three</i>

		<i>cheese stuffed crust. It's like, I don't even care what happens the rest of the night.</i>
49	Neg49	<i>@LAKings hockey and @pizzahut three cheese stuffed crust. It's like, I don't even care what happens the rest of the night.</i>
50	Neg50	<i>@chrissyteigen @pizzahut I don't like it when Pizza Hut lies! Pizza Hut of LIES!!</i>
51	Neg51	<i>Yo @pizzahut apparently you got a lot of managers at your North Bergen location and they like to lie too, it's pretty cool cuz I hate food</i>
52	Neg52	<i>@tommytbcooper92 @SportsCenter @pizzahut and yaa I prolly did hate on him when he sat almost 2 whole seasons due to injuries</i>
53	Neg53	<i>@pizzahut my family hates your pizza but i love it how do you feel about that</i>
54	Neg54	<i>I fucken hate pizza hut @pizzahut fuck you</i>
55	Neg55	<i>@pizzahut i have a feeling you hate me</i>
56	Neg56	<i>"@Imafishwhale_: I fucking hate @pizzahut ."</i>
57	Neg57	<i>I hate you @pizzahut</i>
58	Neg58	<i>Same. I love tho bby @pizzahut "@taskforcekilla_: And I hate Little Caesars pizza"</i>
59	Neg59	<i>@pizzahut I hate that it's Ramadan and I can have some of you ?????</i>
60	Neg60	<i>The one thing I hate is a lot of sauce on my pizza @pizzahut @PapaJohns</i>

61	Neg61	<i>@smxlehemmings @pizzahut i don't i hate tesco....sorry tesco</i>
62	Neg62	<i>Shades of addict when I order @pizzahut: I feel awful afterwards and think "I knew this was a bad idea. Why did I do this? I hate myself."</i>
63	Neg63	<i>RT @Harmssy: @pizzahut RT @hannahohara_: Submitted this order at 7. It's 9:15. I'm pissed to the max. I hate Pizza Hut delivery.</i>
64	Neg64	<i>@pizzahut RT @hannahohara_: Submitted this order at 7. It's 9:15. I'm pissed to the max. I hate Pizza Hut delivery.</i>
65	Neg65	<i>RT @snookumzxo: @myarianathing @AdorbsGranday @pizzahut I hate Pizza Hut</i>
66	Neg66	<i>Dammit @pizzahut I hate how much power you have over me http://t.co/8lsTct6j07</i>
67	Neg67	<i>@Dominos_UK I fucking hate your pizzas. @pizzahut is the dominant faction of the two groups. #BelieveInPizzaHut</i>
68	Neg68	<i>@JonathanTato09 @pizzahut wings are the only wings I'll eat... I hate @BWWings I almost threw up when we went that one time with Nikk</i>
69	Neg69	<i>"@T_tooElite: I hate @pizzahut too bad rn ???. Where my food at ?!" lmfao , you so mad, you @ the company page ??????</i>
70	Neg70	<i>I hate @pizzahut too bad rn ???. Where my food at ?!</i>
71	Neg71	<i>@iNewBlade @idctesia @pizzahut u nasty nigga. I hate boogers.</i>

72	Neg72	<i>@pizzahut That's why I hate those kind of People</i>
73	Neg73	<i>@CassCamsModels @pizzahut everytime, I hate the flour mix @Dominos_UK slap on their base!</i>
74	Neg74	<i>RT @Iselaaaa___: I hate @pizzahut your pizza's make me sick now.</i>
75	Neg75	<i>@pizzahut @FacianeA @ProJared @RubberNinja #4thOfJuly Great day today and now I really hate rain because it got me really wet man</i>
76	Neg76	<i>@xKomradx @PizzaHut I hate their stuffed crust. I loved their pan pizza. But now I hate them forever.</i>
77	Neg77	<i>@xKomradx @PizzaHut I hate their stuffed crust. I loved their pan pizza. But now I hate them forever.</i>
78	Neg78	<i>@xKomradx @PizzaHut I hate them now.</i>
79	Neg79	<i>@pizzahut I hate sharing pizza ?????? http://t.co/AIp0hIMjIA</i>
80	Neg80	<i>@pizzahut I hate sharing pizza??????</i>
81	Neg81	<i>@pizzahut Hi i have a hate and love relationship with you.</i>
82	Neg82	<i>RT @hoezlovekaay: @_wizd0m_ @pizzahut no i hate Pizza Hut ! ???? & yesssssh ????</i>
83	Neg83	<i>I think @pizzahut hates me</i>
84	Neg84	<i>I hate you @pizzahut</i>
85	Neg85	<i>@pizzahut do you hate me because i love you</i>
86	Neg86	<i>@pizzahut i hate u ??</i>
87	Neg87	<i>@pizzahut I hate biting into my pasta and it is still cold!!!!</i>

		<i>#gross #getittogether</i>
88	Neg88	<i>@dunsbridges @pizzahut @dominos tbh I hate Pizza Hut they're horrible. Dominos and pats pizza are usually my top favorite</i>
89	Neg89	<i>It's a fuzzy pants, @pizzahut pasta, crampy kind of night. I hate being a girl sometimes. ??</i>
90	Neg90	<i>@pizzahut I don't like pizza, do you hate me?</i>
91	Neg91	<i>@pizzahut I hate you.</i>
92	Neg92	<i>i dislike you , @pizzahut .</i>
93	Neg93	<i>@pizzahut yeah I dislike you but the times I order they get my order on point!</i>
94	Neg94	<i>@pizzahut I dislike you're prices</i>
95	Neg95	<i>@pizzahut I actually dislike you, BYEEEEEE GOING TO DOMINOES</i>
96	Neg96	<i>@pizzahut I've come to the conclusion that you dislike me and there's nothing I can do</i>
97	Neg97	<i>I dislike the fact that my local @pizzahut does not participate in the current 50% off promo, as others as well. Looks like its Caseys</i>
98	Neg98	<i>If you dislike @pizzahut we can't be friends</i>
99	Neg99	<i>Can't believe manger got fired at pizza hut for trying to give employees time off! Dislike @pizzahut !! #UnfairPizza</i>
100	Neg100	<i>I hope that @pizzahut paid you a lot to promote the worst pizza I've ever eaten. #disappointed</i>

