

**ANALISIS SENTIMEN ETIKA BERINTERNET PADA MEDIA SOSIAL  
DENGAN MENGGUNAKAN ALGORITMA NAIVE BAYES DAN  
PARTICLE SWARM OPTIMIZATION**



**TESIS**

**SOPIAN AJI  
14001106**

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

## SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini :

Nama : Sopian Aji  
NIM : 14001106  
Program Studi : Magister Ilmu Komputer  
Jenjang : Strata Dua (S2)  
Konsentrasi : *Management Information System*

Dengan ini menyatakan bahwa tesis yang telah saya buat dengan judul: **“Analisis Sentimen Etika Berinternet Pada Media Sosial Dengan Menggunakan Algoritma Naive Bayes Dan Particle Swarm Optimization”** 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 Inbentukika dan Komputer Nusa Mandiri dicabut/dibatalkan.

Jakarta, 01 Desember 2015

Yang menyatakan,

  
  
Sopian Aji

## HALAMAN PENGESAHAN

Tesis ini diajukan oleh :

Nama : Sopian Aji  
NIM : 14001106  
Program Studi : Magister Ilmu Komputer  
Jenjang : Strata Dua (S2)  
Konsentrasi : *Management Information System*  
Judul Tesis : "Analisis Sentimen Etika Berinternet Pada Media Sosial Dengan Menggunakan Algoritma Naive Bayes Dan Particle Swarm Optimization"

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, 01 Desember 2015  
Pascasarjana Magister Ilmu Komputer  
STMIK Nusa Mandiri  
Direktur

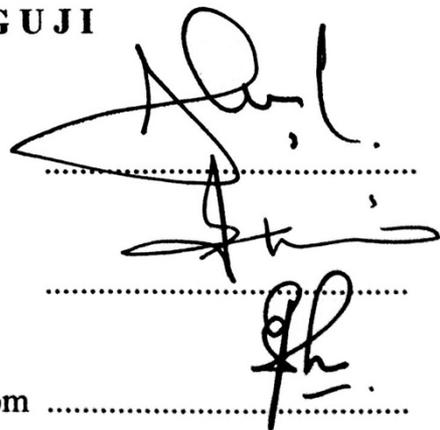
Prof. Dr. Ir. R. Eko Indrajit, MSC, MBA

## DEWAN PENGUJI

Penguji I : Dr. Windu Gata, M. Kom

Penguji II : Dr. Sfenrianto, M. Kom

Pembimbing : Dr. Sularso Budilaksono, M. Kom

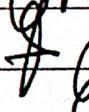
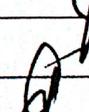


.....  
.....  
.....

	<b>LEMBAR KONSULTASI BIMBINGAN TESIS</b>
	<b>SEKOLAH TINGGI MANAJEMEN INFORMATIKA &amp; KOMPUTER NUSA MANDIRI</b>

- NIM : 14001106
- Nama Lengkap : Sopian Aji
- Dosen Pembimbing : Dr. Sularso Budilaksono
- Judul Tesis : Analisis Sentimen Etika Berinternet Dimedia Sosial Dengan Menggunakan Algoritma Naive Bayes Dan Particle Swarm Optimization



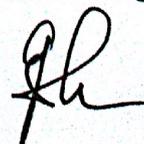
No	Tanggal Bimbingan	Pokok Bahasan	Paraf Dosen Pembimbing
1	26 Okt '15	Bimbingan perdana	
2	02 Nop '15	Pengajuan Bab I & Bab II	
3	06 Nop '15	Pengajuan Bab III	
4	11 Nop '15	Pengajuan Bab IV	
5	25 Nop '15	Pengajuan Bab V	
6	30 Nop '15	Periksa keseluruhan & Acc Ujian Sidang	

Catatan untuk dosen pembimbing

**Bimbingan Tesis**

- Dimulai pada tanggal : 26 Oktober 2015
- Diakhiri pada tanggal : 30 Nopember 2015
- Jumlah pertemuan bimbingan : 6 Kali Bimbingan

Disetujui Oleh,  
Dosen Pembimbing



Dr. Sularso Budilaksono

## 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. Tujuan penulisan tesis ini dibuat sebagai salah satu syarat 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). Penulis mengangkat topik *Knowledge Management* dalam rangka meningkatkan *competitive advantage* pada perusahaan, dengan judul **“Analisis Sentimen Etika Berinternet Pada Media Sosial Dengan Menggunakan Algoritma Naive Bayes Dan Particle Swarm Optimization”**.

Pada kesempatan ini, penulis ingin menyampaikan penghargaan dan rasa terima kasih yang sebesar-besarnya kepada berbagai pihak yang telah memberikan bantuan, petunjuk, saran, nasehat, bimbingan, doa, dan semangat, sehingga memungkinkan penulis untuk dapat menyelesaikan Tesis ini:

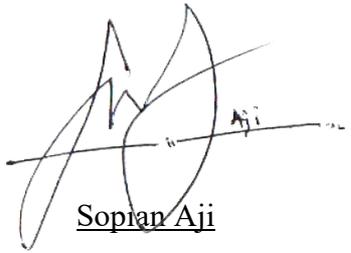
1. Bapak Prof. Dr. Ir. R. Eko Indrajit, MSC, MBA selaku direktur STMIK Nusa Mandiri Pascasarjana.
2. Bapak Dr. Sularso Budilaksono, M. Kom selaku dosen pembimbing, yang telah banyak menyediakan waktu dan pikiran untuk penulis dalam penyusunan tesis ini.
3. Orangtua tercinta yaitu Maman S. dan Masfuah, Dan Mertua tercinta yaitu Kurdi dan Kusti
4. Istriku Ida Rosmaidah yang solehah yang selalu menjadikan penyejuk Hati dan Anakku yang soleh Hafizhar Abidiandra Maika yang telah banyak memberikan dukungan material dan moral kepada penulis.
5. Sahabat-sahabat staf pengajar Bina Sarana Informatika Kampus Cengkareng yang telah memberikan dukungan bagi penulis selama penyusunan Tesis ini.
6. Rekan-rekan mahasiswa Program Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri Jakarta, atas dukungan, motivasi serta bantuan yang telah diberikan.

Serta semua pihak yang terlalu banyak untuk penulis sebutkan satu persatu sehingga atas bantuan semua pihak tersebut penulis dapat menyelesaikan penulisan Tesis ini. Penulis menyadari bahwa dalam penulisan Tesis ini masih terdapat banyak kekurangan dan kekeliruan, untuk

itu penulis mengharapkan kritik dan saran yang membangun agar penulisan karya ilmiah yang penulis hasilkan dimasa yang akan datang menjadi lebih baik.

Akhir kata, semoga Tesis ini dapat bermanfaat bagi penulis khususnya dan bagi para pembaca pada umumnya.

Jakarta, 01 Desember 2015

A handwritten signature in black ink, consisting of several overlapping loops and a horizontal line extending to the right. The signature is written over the printed name 'Sopian Aji'.

Sopian Aji

Penulis

## SURAT PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini, saya:

Nama : Sopian Aji  
NIM : 14001106  
Program Studi : Magister Ilmu Komputer  
Jenjang : Strata Dua (S2)  
Konsentrasi : *Management Information System*  
Jenis Karya : Tesis

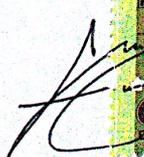
Demi pengembangan ilmu pengetahuan, dengan ini menyetujui untuk memberikan ijin kepada pihak Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri) **Hak Bebas Royalti Non-Eksklusif (*Non-exclusive Royalti-Free Right*)** atas karya ilmiah kami yang berjudul : **“Prototipe *Knowledge Management System* pada Perusahaan Jasa dan Produk: Studi Kasus pada PT. Jane Indah Perkasa Jakarta”** 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, 01 Desember 2015  
Yang menyatakan,

  
  
Sopian Aji

## DAFTAR ISI

	Halaman
HALAMAN JUDUL .....	i
HALAMAN PERNYATAAN ORISINALITAS .....	ii
HALAMAN PENGESAHAN .....	iii
LEMBAR KONSULTASI BIMBINGAN TESIS.....	iv
KATA PENGANTAR.....	v
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS .....	vii
DAFTAR ISI .....	viii
DAFTAR TABEL .....	ix
DAFTAR GAMBAR.....	x
ABSTRACT .....	xi
1. PENDAHULUAN.....	1
2. PENELITIAN TERKAIT.....	4
3. METODE PENELITIAN .....	10
4. HASIL PENELITIAN .....	12
5. KESIMPULAN .....	16
6. DAFTAR PUSTAKA.....	17

## DAFTAR TABEL

	Halaman
Tabel 1. Tabel Kontingensi Evaluasi Kinerja Sistem.....	7
Tabel 2. Text Processing Menggunakan Tokenize, Transform Case, Stopwords(Dictionar) .....	12
Tabel 3. Accuracy Algoritma Naïve Bayes .....	13
Tabel 4. Accuracy Algoritma Naïve Bayes Berbasis Particle Swarm Optimization ....	15

## DAFTAR GAMBAR

	Halaman
Gambar 1. Diagram Proses Klasifikasi Teks .....	5
Gambar 2. Klasifikasi Naïve Bayes Sebagai Jaringan Bayes Dengan Atribut Prediksi (P1, P2, . . . .Pk) Dan Kelas C .....	6
Gambar 3. Kerangka Pemikiran .....	10
Gambar 4. Kurva ROC Naïve Bayes .....	14
Gambar 5. Kurva ROCNaïve Bayes Berbasis Particle Swarm Optimization .....	15

## **ABSTRACT**

*Name* : Sopian Aji  
*NIM* : 14001106  
*Studi of Program* : Magister Ilmu Komputer  
*Level* : Strata Dua (S2)  
*Concentration* : *Management Information System*  
*Title* : Analisis Sentimen Etika Berinternet Pada Media Sosial Dengan Menggunakan Algoritma Naive Bayes Dan Particle Swarm Optimization

*Sentiment analysis is process that contains text-based datasets which are positive, negative or neutral. Social media has provided a place for web users to express sharing of thoughts, opinions and convey news on different topics in an event. Haters in various media, including social media, can be punished. In the circular, stated that hate speech issue has been getting attention both national and international community concern for the protection of human rights (HAM). Classification of algorithms such as Naive Bayes (NB) and Particle Swarm Optimization (PSO) was proposed by many researchers to be used in the analysis of text sentiment. Naive Bayes's algorithms and methods, will be tested with two inputs using tokenize and Transform Cases's comments are positive (100 text comments) and negative (100 comments text), it obtained experimental results accuracy: 62.50 % +/- 7.50 % (micro : 62.50 %). The result will be increased if the results of the experiment combine with Particle Swarm Optimization (PSO), it obtained experimental results were better accuracy: +/- 74.00 % 7:68 % (micro : 74.00 %). The results showed that Naive Bayes (NB) will get the best results if it combined with Particle Swarm Optimization (PSO).*

*Keywords : sentiment analysis, internet ethics, social media, Naive Bayes, particle swarm optimization*

## 1. PENDAHULUAN

Banyak sekali media pemberita yang menyampaikan informasinya kepada masyarakat mulai dari pemberitaan melalui media Televisi, Surat Kabar, Majalah atau Radio, kini sudah mulai menggunakan sistem berbasis web kemudian media Jejaring sosial. Media Jejaring Sosial terdiri dari berbagai media online seperti blog, forum, review produk, situs media jejaring sosial, aplikasi media jejaring sosial, game sosial, dan media online lainnya. Macam-macam media jejaring sosial yang disediakan: *Path, Line, Instagram, Whatsapp, Blackberry Messenger, Twitter, LinkedIn, Facebook, Wikipedia, Google Plus, youtube* dan masih banyak lagi lainnya. Pada umumnya berita yang disampaikan dalam portal tersebut terdiri dari beberapa kategori seperti berita politik, olahraga, ekonomi, kesehatan, dan lain-lain. Tentu saja informasi yang didapat bukan saja pemberitaan yang bersifat positif saja tetapi ada juga yang bersifat negatif. Pemberitaan yang negatif ini tentu saja berbahaya karena semua kalangan mendapatkan pemberitaan tersebut misalnya dari segi usia, golongan, suku, agama bahkan antar-negara, contoh dari pemberitaan yang bersifat negatif di media sosial yang banyak beredar pornografi, penghinaan, pencemaran nama baik, penistaan, perbuatan tidak menyenangkan, memprovokasi, menghasut, menyebarkan berita bohong dan semua tindakan memiliki tujuan atau bisa berdampak pada tindak diskriminasi, kekerasan, penghilangan nyawa dan konflik sosial.

Surat Edaran (SE) Kapolri soal penanganan ujaran kebencian atau hate speech Nomor SE/06/X/2015. Surat tersebut diteken Jenderal Badrodin Haiti pada 8 Oktober 2015 lalu dan telah dikirim ke Kepala Satuan Wilayah (Kasatwil) seluruh Indonesia. Dalam surat edaran tersebut, disebutkan bahwa persoalan ujaran kebencian semakin mendapatkan perhatian masyarakat baik nasional atau internasional seiring meningkatnya kepedulian terhadap perlindungan hak asasi manusia (HAM). Berikut poin-poin krusial dalam SE tersebut: **Bentuk Ujaran Kebencian** disebutkan bahwa "ujaran kebencian dapat berupa tindak pidana yang diatur dalam Kitab Undang-Undang Hukum Pidana (KUHP) dan ketentuan pidana lainnya di luar KUHP, yang berbentuk antara lain: (1) Penghinaan, (2) Pencemaran nama baik, (3) Penistaan, (4) Perbuatan tidak menyenangkan, (5) Memprovokasi, (6) Menghasut, (7) Menyebarkan berita bohong dan semua tindakan di atas memiliki tujuan atau bisa berdampak pada tindak diskriminasi, kekerasan, penghilangan nyawa, dan atau konflik sosial". **Aspek Ujaran Kebencian** Selanjutnya, bahwa ujaran kebencian sebagaimana

dimaksud di atas bertujuan untuk menghasut dan menyulut kebencian terhadap individu dan atau kelompok masyarakat dalam berbagai komunitas yang dibedakan dari aspek: (1) Suku, (2) Agama, (3) Aliran keagamaan, (4) Keyakinan atau kepercayaan, (5) Ras, (6) Antargolongan, (7) Warna kulit, (8) Etnis, (9) Gender, (10) Kaum difabel, (11) Orientasi seksual. **Media Ujaran Kebencian** disebutkan bahwa "ujaran kebencian sebagaimana dimaksud di atas dapat dilakukan melalui berbagai media, antara lain:

(1) Dalam orasi kegiatan kampanye, (2) Spanduk atau banner, (3) Jejaring media sosial, (4) Penyampaian pendapat di muka umum (demonstrasi), (5) Ceramah keagamaan, (6) Media massa cetak atau elektronik, (7) Pamflet. Dengan memperhatikan pengertian ujaran kebencian di atas, perbuatan ujaran kebencian apabila tidak ditangani dengan efektif, efisien, dan sesuai dengan ketentuan peraturan perundang-undangan, akan berpotensi memunculkan konflik sosial yang meluas, dan berpotensi menimbulkan tindak diskriminasi, kekerasan, dan atau penghilangan nyawa.

Pada Penelitian ini penulis menggunakan algoritma *Naïve Bayes* dan *Particle Swarm Optimization* yaitu untuk mendapatkan hasil akurasi kedua model tersebut. Identifikasi permasalahan pada penelitian ini adalah:

1. Banyaknya pemberitaan dimedia sosial yang bersifat negatif sehingga perlunya pembatasan atau penyaringan pada pemberitaan
2. Bagaimana tingkat akurasi analisa sentimen yang dihasilkan algoritma *Naïve Bayes* dan *Particle Swarm Optimization*

Dan adapun manfaat dari penelitian ini adalah:

1. Manfaat dari penelitian ini adalah membantu para pembaca atau penerima informasi yang lebih sehat dan bersifat positif
2. Manfaat teoritis penelitian ini adalah memberikan bukti secara empiris untuk teori yang berkaitan sehingga dapat dijadikan sumbangan pemikiran untuk pengembangan teori berikutnya.

Sentimen terutama mengacu pada perasaan, emosi, pendapat atau sikap. Dengan meningkatnya *World Wide Web*, orang sering mengungkapkan perasaan mereka di melalui internet melalui media sosial, blog, rating dan reviews. Karena ini meningkatkan data tekstual, ada kebutuhan untuk menganalisa konsep mengekspresikan sentimen dan menghitung wawasan

untuk menjelajahi bisnis(Tripathy, Agrawal, and Rath, 2015). Analisis sentimen, adalah proses bertujuan untuk menentukan apakah polaritas korpus tekstual (dokumen, kalimat, paragraf dll) cenderung ke arah positif, negatif atau netral(Kontopoulos, Berberidis, Dergiades and Bassiliades, 2013).

Opini mining tidak memperhatikan topik dari teks tersebut tetapi lebih fokus kepada ekspresi yang digambarkan dari teks opini tersebut. Hal ini menentukan komentar dalam forum online, blog, atau komentar yang berkaitan dengan topik tertentu (produk, buku, film, dan lain-lain) termasuk opini positif, negatif atau netral(Martinez, Sanchez, Garcia, Garcia, Moreno, Fraga and Cervantes, 2014). Pada analisis sentiment Para penulis menggunakan empat kelompok yang berbeda: *Naive Bayes*, *Maximum Entropy*, *SVM*, and *the character based n-gram language model* (Dehkharghani, Mercan, Javeed, Saygin, 2015). Metode untuk melakukan analisis sentimen review restoran dengan menggunakan *Support Vector Machine*(SVM) maupun *Naive Bayes*(NB) (Kang, Yoo, Han- Senti, 2012). Ada berbagai pendekatan pembelajaran mesin untuk otomatis klasifikasi sentimen, misalnya Teori *Naive Bayes*, *Maximum Entropy*(ME), *Support Vector Machine*(SVM), dll. *Naive Bayes classifier*(NBC) adalah metode sederhana tetapi memiliki akurasi serta performansi yang tinggi dalam pengklasifikasian teks(Routray, Swain and Mishra, 2013)

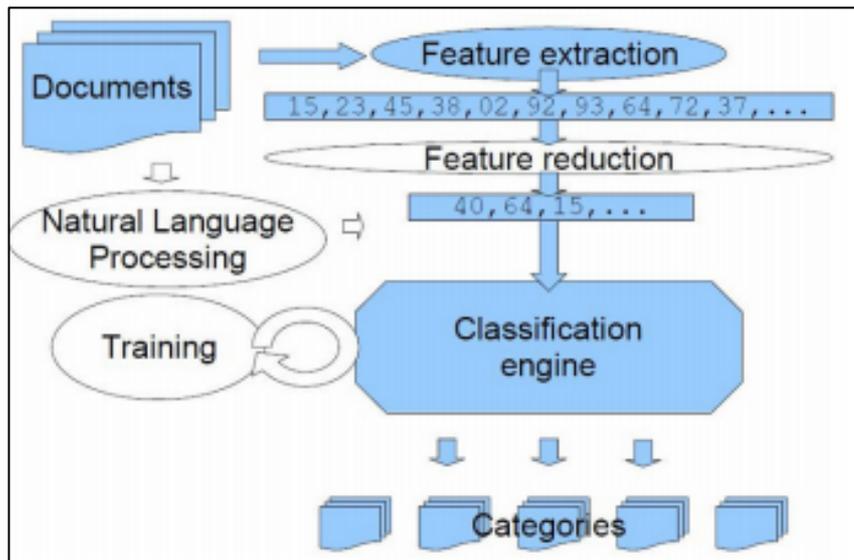
## 2. PENELITIAN TERKAIT

Banyak ide-ide telah muncul selama bertahun-tahun tentang bagaimana mencapai hasil yang berkualitas dari sistem klasifikasi Web, sehingga ada pendekatan yang berbeda yang dapat digunakan untuk permasalahan analisis sentimen, seperti *Naïve Bayes* (NB) dan *Bayesian Networks*, NNs, DTs, *Support Vector Machines* (SVM) dll. *Model (Naïve Bayes)* NB yang populer dalam aplikasi machine learning, karena kesederhanaan mereka dalam membiarkan setiap atribut berkontribusi terhadap keputusan akhir yang sama dan independen dari atribut lainnya (Xhemali, Hinde and Stone, 2009). Berbagai metode *machine learning* telah digunakan untuk penentuan atribut dengan menggunakan *naïve Bayes* (Altheneyan and Menai, 2014)

Proses klasifikasi teks dapat dibagi ke dalam dua *fase* (Darujati and Gumelar, 2012) yaitu:

1. *Fase information retrieval* (IR) untuk mendapatkan data numerik dari dokumen teks. Langkah pertama yang dilakukan pada fase ini adalah feature extraction. Pendekatan yang umum digunakan adalah distribusi frekuensi kata. Nilai numerik yang diperoleh dapat berupa berapa kali suatu kata muncul di dalam dokumen, 1 jika kata ada di dalam dokumen atau 0 jika tidak ada (biner), atau jumlah kemunculan kata pada awal dokumen. *Feature* yang diperoleh dapat direduksi agar dimensi vektor menjadi lebih kecil. Beberapa pendekatan feature reduction dapat diterapkan seperti menghapus *stop-words*, *stemming*, *statistical filtering*. Teknik lebih lanjut seperti SVD dan genetic algorithm akan menghasilkan vektor berdimensi lebih rendah
2. *Fase* klasifikasi utama ketika suatu algoritma memroses data numerik tersebut untuk memutuskan ke kategori mana teks ditempatkan. Terdapat beberapa algoritma klasifikasi yang merupakan kajian di bidang statistika dan machine learning yang dapat diterapkan pada fase ini, di antaranya adalah *Naive Bayesian*, *Rocchio*, *Decision Tree*, *kNearest Neighbor*, *Neural Network*, dan *Support Vector Machines*. Teknik-teknik tersebut berbeda dalam mekanisme pembelajaran dan representasi model yang dipelajari.

proses klasifikasi teks melibatkan banyak teknik IR mulai dari preprocessing, pengukuran kemiripan selama operasional klasifikasi sampai dengan evaluasi kinerja algoritma



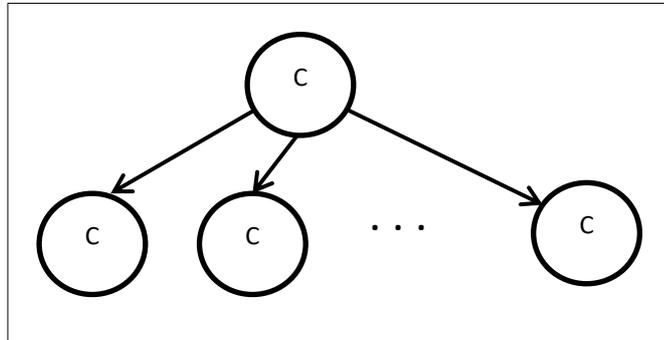
**Gambar 1. Diagram proses klasifikasi teks (Darujati and Gumelar, 2012)**

*Pre-Processing* pada pengujiannya Analisis Sentimen Menggunakan Metode *Naïve Bayes Classifier* Dengan Seleksi Fitur Chi Square (Ling, Kencana and Oka, 2014) sebagai berikut:

1. **Tokenization** adalah tugas pemotongan urutan karakter dan sebuah set dokumen yang diberikan menjadi potongan-potongan kata atau karakter yang sesuai dengan kebutuhan sistem. Potongan-potongan tersebut dikenal dengan istilah token.
2. **Stemming** merupakan salah satu proses dari mengubah token yang berlebihan menjadi kata dasar, dengan menghilangkan semua imbuhan yang ada pada token tersebut. Pentingnya stemming dalam proses pembuatan sistem adalah untuk menghilangkan imbuhan pada awalan dan akhiran. Berdasarkan hasil proses tersebut, akan didapatkan sebuah informasi mengenai banyaknya fitur yang muncul dalam sebuah dokumen.
3. **Stopwords** dapat diartikan sebagai menghilangkan karakter, tanda baca, serta kata-kata umum yang tidak memiliki makna atau informasi yang dibutuhkan. *Stopwords* umumnya digunakan dalam pengambilan informasi salah satu contohnya adalah mesin pencari *Google*. Pengurangan ukuran indeks dalam teks dengan penghilangan beberapa kata kerja, kata sifat, dan kata keterangan lainnya dapat dimasukkan ke dalam daftar *stopwords*.

Klasifikasi–klasifikasi *Bayes* adalah klasifikasi statistik yang dapat memprediksi kelas suatu anggota probabilitas. Untuk klasifikasi *Bayes* sederhana yang lebih dikenal sebagai *naïve Bayesian Classifier* dapat diasumsikan bahwa efek dari suatu nilai atribut

sebuah kelas yang diberikan adalah bebas dari atribut-atribut lain. Dalam *Naïve Bayes* diasumsikan prediksi atribut adalah tidak tergantung pada kelas atau tidak dipengaruhi atribut lain (Darujati and Gumelar, 2012), dengan gambaran sebagai berikut:



**Gambar 2. Klasifikasi Naïve Bayes sebagai jaringan bayes dengan atribut prediksi (P1, P2,.....Pk) dan kelas (C). (Darujati and Gumelar, 2012)**

C adalah adalah anggota kelas dan X adalah variabel acak sebuah vektor sebagai atribut nilai yang diamati. c mewakili nilai label kelas dan x mewakili nilai atribut vector yang diamati. Jika diberikan sejumlah x tes untuk klasifikasi maka probabilitas tiap kelas untuk atribut prediksi

vektor yang diamati adalah:

$$p(C = c|X = x) = \frac{p(C = c)p(X = x|C = c)}{p(C = c)}$$

X = x adalah mewakili kejadian dari Jumlah dari untuk semua kelas adalah 1

$$X_1 = x_1 \wedge X_2 = x_2 \wedge \dots \wedge X_k = x_k$$

$$p(C = c|X = x)$$

Pada tahap evaluasi sistem, perhitungan menggunakan tabel kontingensi yang diberikan (Ling, Kencana and Oka, 2014):

**Tabel 1. Kontingensi Evaluasi Kinerja Sistem(Ling, Kencana and Oka, 2014)**

#	Relevant	Not Relevant
Retrieved	True Positive (TP)	False Positive(FP)
Not Retrieved	False Negatif (FN)	True Negatif(TN)

Alternatif yang jelas terlintas pada pikiran pembaca dalam menilai sebuah system adalah dengan akurasi. Akurasi adalah ketepatan suatu sistem melakukan klasifikasi yang benar. Perhitungan untuk Presisi (P), Recall (R), dan akurasi dapat dikalkulasi sebagai berikut:

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

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

Berikut adalah persamaan model confusion matrix (Indrayuni and Wahyudi, 2015) sebagai berikut:

- a. Nilai akurasi (acc) adalah proporsi jumlah prediksi yang benar.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- b. Sensitivity digunakan untuk membandingkan proporsi tp terhadap tupel yang positif.

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

- c. Specificity digunakan untuk membandingkan proporsi tn terhadap tupel yang negatif.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

- d. PPV (positive predictive value) adalah proporsi kasus dengan hasil diagnosa positif.

$$\text{ppv} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

- e. NPV (negative predictive value) adalah proporsi kasus dengan hasil diagnosa negatif.

$$\text{npv} = \frac{\text{TN}}{\text{TN} + \text{FN}}$$

Tahapan *text processing* menurut (Darmawan,2015) sebagai berikut:

1. *Tokenize*

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

2. *Transform Cases*

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

3. Stopwords(Dictionar)

Dalam proses ini, kata-kata yang tidak relevan akan dihapus, seperti kata benci, bodoh, lebay, yang merupakan kata-kata yang negatif. Dengan menggunakan kamus manual yang mengandung arti negatif yang bersifat sentimen disimpan pada notepad, kemudian diberi nama misalnya kamus.txt sehingga jika ada teks yang sama dengan kamus yang telah kita buat kata tersebut akan dihilangkan.

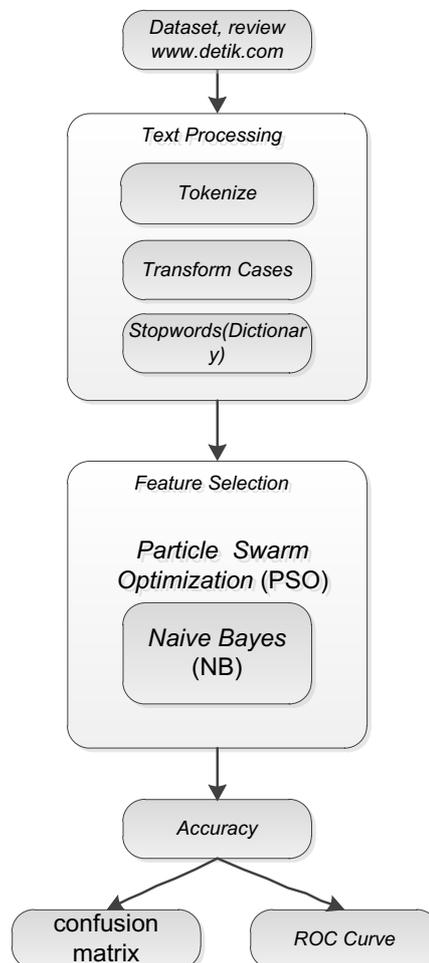
*Particle Swarm Optimization* (PSO) adalah teknik optimasi berbasis populasi yang dikembangkan oleh Eberhart dan Kennedy pada tahun 1995, yang terinspirasi oleh perilaku sosial kawanan burung atau ikan (Park, Lee & Choi, 2009). Karakteristik *particle swarm optimization* adalah interaksi sosial yang mempromosikan pembagian informasi antara partikel yang akan membantu dalam pencarian solusi yang optimal (Park, Lee, & Choi, 2009). Particle swarm optimization memiliki beberapa parameter seperti posisi, kecepatan, kecepatan

maksimum, percepatan konstanta dan berat inersia. Dalam teknik Particle swarm optimization terdapat beberapa cara untuk melakukan pengoptimasian diantaranya: meningkatkan bobot atribut (attribute weight) terhadap semua atribut atau variabel yang dipakai, menyeleksi atribut (attribute selection), dan feature selection. *Particle swarm optimization* memiliki perbandingan lebih atau bahkan pencarian kinerja lebih unggul untuk banyak masalah optimasi dengan lebih cepat dan tingkat konvergensi yang lebih stabil (Park, Lee & Choi, 2009).

### 3. METODE PENELITIAN

Penelitian yang dilakukan pada analisa sentimen dengan menggunakan Algoritma *Naive Bayes*(NB) dan *Naive Bayes*(NB) berbasis *Particle Swarm Optimization* (PSO) pada pemberitaan “Surat Edaran (SE) Kapolri soal penanganan ujaran kebencian atau hate speech Nomor SE/06/X/2015” dengan mengukur akurasi data eksperimen yang dilakukan dalam penelitian dengan menggunakan aplikasi *RapidMiner Studio 6.5.002*.

Pada gambar 3 menggambarkan kerangka pemikiran yang peneliti usulkan dalam penelitian ini adalah sebagai berikut:



Sumber: Hasil Penelitian(2015)

**Gambar 3. Kerangka Pemikiran**

Untuk melakukan eksperimen peneliti menggunakan data komentar *review* <http://news.detik.com/berita/3058647/awas-penebar-kebencian-di-media-sosial-bisa-diancam-pidana>. Maka didapat 200 data *review* yang terdiri dari 100 *review* positif dan 100 *review* negatif. Data *review* positif disatukan dalam folder dengan nama Positif. Sedangkan data *review* negatif disatukan penyimpanannya dalam folder dengan nama Negatif. Tiap dokumen berekstensi .txt yang dapat dibuka menggunakan aplikasi *Notepad*.

Kemudian setelah mendapatkan data positif dan negatif dalam bentuk .txt dengan jumlah data yang sama, yaitu 100 *review* data positif maka yang bersifat negatif harus didapatkan 100 *review* data negatif, data akan diproses menggunakan Aplikasi *RapidMiner*. Pada aplikasi *RapidMiner* untuk *text processing* yang peneliti gunakan *Tokenize*, *Transform Cases*, *Stopwords(Dictionar)*. Pengujian yang dilakukan dengan memilih *feature selection Naive Bayes(NB)* berbasis *Particle Swarm Optimization (PSO)* dengan bertujuan untuk mendapatkan nilai *accuracy*. Dari nilai *accuracy* persamaan model confusion matrix proporsi jumlah nilai prediksi didapatkan dan *ROC Curve* sehingga kita bisa melihat grafik *ROC* dengan nilai *AUC (Area Under Curve)*.

#### 4. HASIL PENELITIAN

Penelitian yang dilakukan menggunakan Aplikasi *RapidMiner Studio 6.5.002* dengan menggunakan spesifikasi komputer CPU *Intel Core i3 1.90GHz*, RAM 4GB, dan *system operasi Microsoft Windows 7 Professional 64-bit*. Hasil pengujian *review* Pemberitaan “*Awas, Penebar Kebencian di Media Sosial Bisa Diancam Pidana*” dengan menggunakan teks bahasa Indonesia.

Hasil dari penelitian menggunakan *Text Processing* pada table 2 yang dilalui terdiri dari *Tokenize, Transform Cases, Stopwords(Dictionar)*. Hasil pengolahan sebagai berikut:

**Tabel 2. Text Processing menggunakan Tokenize, Transform Cases, Stopwords(Dictionar)**

<i>Text Review</i>	<i>Tokenize</i>	<i>Transform Cases</i>	<i>Stopwords(Dictionar)</i>
Satu kata "LEBAY" kapan mau maju indo?? Klo peraturannya makin lama makin "GA PENTING"	Satu kata LEBAY kapan mau maju indo Klo peraturannya makin lama makin GA PENTING	Satu kata lebay kapan mau maju indo Klo peraturannya makin lama makin ga penting	Satu kata kapan mau maju indo Klo peraturannya makin lama makin ga penting
SAATNYA JADI ORANG MUNAFIK,,MARI BICARA YANG BAIK-BAIK MAU SUMPAAH SERAPAH CUKUP DALAM HATI AJA..wk..wk..	SAATNYA JADI ORANG MUNAFIKMARI BICARA YANG BAIK BAIK MAU SUMPAAH SERAPAH CUKUP DALAM HATI AJA wk wk	saatnya jadi orang munafikmari bicara yang baik baik mau sumpah serapah cukup dalam hati aja wk wk	saatnya jadi orang mari bicara yang baik baik mau sumpah serapah cukup dalam hati aja wk wk
YESSS !!!	YESSS	yesss penjara semakin ramai makin	yesss semakin ramai makin

PENJARA SEMAKIN RAMAI !!!	PENJARA SEMAKIN RAMAI makin		
---------------------------------	-----------------------------------	--	--

Sumber: Hasil Penelitian(2015)

Pada tabel 2 terlihat *Text Processing* dari *text review* yang bersifat negatif akan berubah menjadi *text review* yang bersifat positif. Pada *text review* melalui *Tokenize* seluruh tanda baca akan dihilangkan, kemudian *Transform Cases* seluruh *text review* yang menggunakan huruf besar akan berubah menjadi huruf kecil dan yang terakhir *Stopwords(Dictionar)* kosa kata atau kata yang bersifat negatif akan dihilangkan.

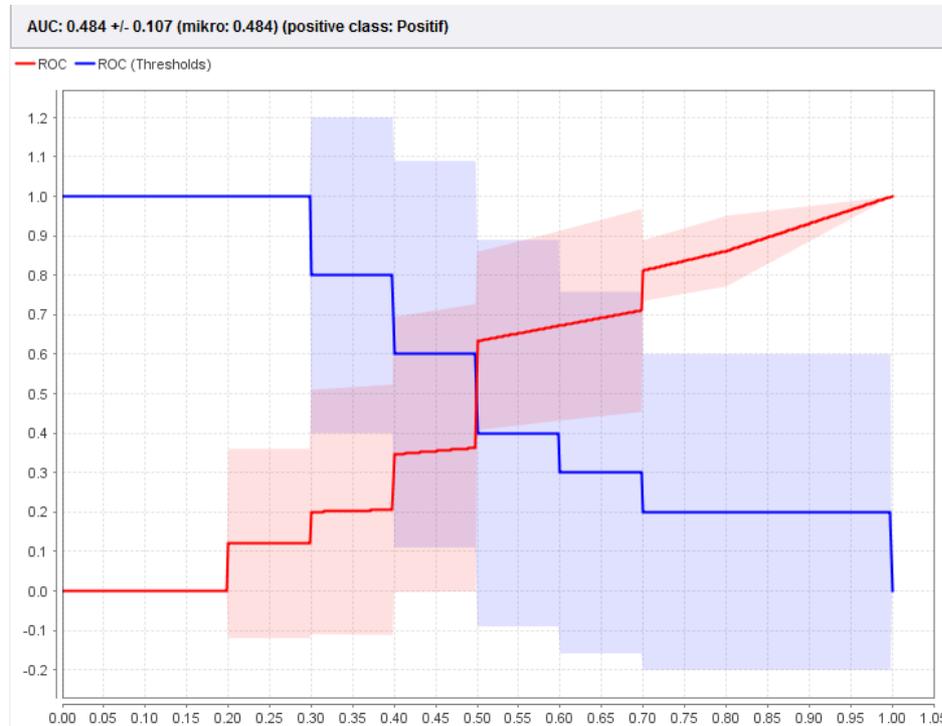
Pengujian pertama penelitian dilakukan menggunakan algoritma *Naive Bayes* untuk mendapatkan *accuracy* dapat kita lihat pada tabel 3 dengan nilai *accuracy* 62.50%, sebagai berikut:

**Tabel 3. Accuracy Algoritma Naive Bayes**

	true Negatif	true Positif
pred. Negatif	58	33
pred. Positif	42	67

Sumber: Hasil Penelitian(2015)

Sedangkan dari penelitian *accuracy* menggunakan algoritma *Naive Bayes* dapat kita lihat pada gambar 4 dengan grafik ROC dengan nilai *AUC*:  $0.484 \pm 0.107$  (*mikro*:  $0.484$ ) (*positive class*: *Positif*) sebagai berikut:



Sumber: Hasil Penelitian(2015)

**Gambar 4. Kurva ROC *Naive Bayes***

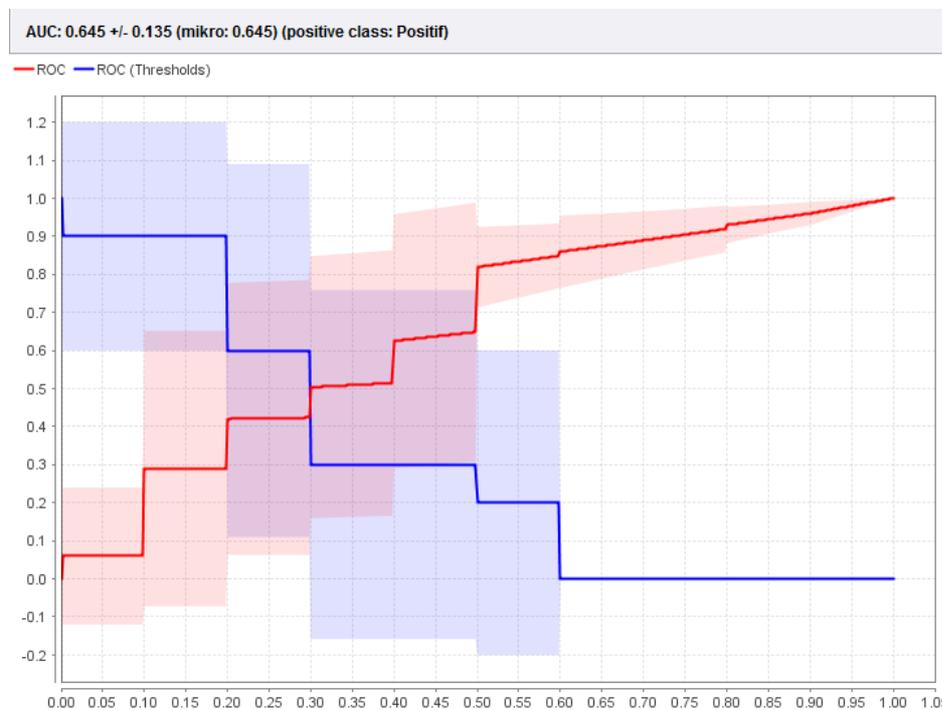
Hasil dari pengujian pertama *accuracy* atau akurasi menggunakan algoritma *Naive Bayes* dengan nilai *accuracy* 62.50% dan *AUC*:  $0.484 \pm 0.107$  (*mikro*:  $0.484$ ) (*positive class*: *Positif*). Kemudian pengujian kedua dilakukan yaitu dengan menggunakan algoritma *Naive Bayes* berbasis *Particle Swarm Optimization* yang dapat kita lihat pada tabel 4 dengan mendapatkan hasil *accuracy* terbaik 74.00%.

**Tabel 4. Accuracy Algoritma Naive Bayes berbasis Particle Swarm Optimization**

	true Negatif	true Positif
pred. Negatif	76	28
pred. Positif	24	72

Sumber: Hasil Penelitian(2015)

Dan sedangkan dari penelitian *accuracy* menggunakan algoritma *Naive Bayes* berbasis *Particle Swarm Optimization* dapat kita lihat pada gambar 5 dengan grafik ROC dengan nilai terbaik *AUC: 0.645 +/- 0.135 (mikro: 0.645) (positive class: Positif)* sebagai berikut:



Sumber: Hasil Penelitian(2015)

**Gambar 5. Kurva ROC Naive Bayes berbasis Particle Swarm Optimization**

Pada hasil pengujian kedua ini yaitu algoritma *Naive Bayes* berbasis *Particle Swarm Optimization* berdasarkan pengamatan *accuracy* menjadi lebih baik dengan nilai *accuracy* 74.00% dan *AUC: 0.645 +/- 0.135 (mikro: 0.645) (positive class: Positif)*. Dengan demikian hasil dari *review* analisis sentimen *Naive Bayes* (NB) dapat meningkatkan nilai akurasi setelah menggabungkan dengan *Particle Swarm Optimization* (PSO).

## 5. KESIMPULAN

Pengujian model dengan menggunakan *Naive Bayes*(NB) dan *Naive Bayes*(NB) berbasis *Particle Swarm Optimization* (PSO) dengan menggunakan data *review* analisis sentimen Etika berinternet dimedia sosial dari 200 data *review* dalam teks bahasa indonesia yang terdiri dari 100 *review* positif dan 100 *review* negatif. Model yang dihasilkan mendapatkan hasil *accuracy* pengujian pertama dilakukan dengan menggunakan *Naive Bayes*(NB) yaitu mendapatkan nilai *accuracy* 62.50% dan *AUC*: 0.484 +/- 0.107 (*mikro*: 0.484) (*positive class*: *Positif*), yang kemudian diuji kembali dengan pengujian yang kedua dengan menggabungkan *Naive Bayes*(NB) dengan *Particle Swarm Optimization* (PSO) sehingga didapat nilai yang terbaik dengan *accuracy* 74.00% dan *AUC*: 0.645 +/- 0.135 (*mikro*: 0.645) (*positive class*: *Positif*). maka dapat disimpulkan pengujian analisa sentiment dengan menggunakan *Naive Bayes*(NB) akan lebih baik lagi hasilnya dengan menggabungkan *Naive Bayes*(NB) dengan *Particle Swarm Optimization* (PSO) sehingga kita dapat memecahkan masalah analisa sentimen dengan memilih pemberitaan atau informasi dimedia sosial khususnya lebih akurat.

## 6. DAFTAR PUSTAKA

- Altheneyan, A. S. , Menai, M. E. B. (2014). Naïve Bayes classifiers for authorship attribution of Arabic texts. *Journal of King Saud University –Computer and Information Sciences*, 473-484. doi.org/10.1016/j.jksuci.2014.06.006
- Darmawan, Agus. (2015). Penerapan Model Support Vector Machine Text Mining Pada Komentar Review Smartphone Android Vs Blackberry Dengan Teknik Optimasi Genetic Algorithm. 100-115, ISSN: 1979-276X
- Darujati, Cahyo. Gumelar, A., B., (2012) . Pemanfaatan Teknik Supervised untuk Klasifikasi Teks Bahasa Indonesia. ISSN 1858-4667
- Dehkharghani, R., Mercan, H., Javeed, A., & Saygin, Y. (2014). Sentimental causal rule discovery from Twitter. *Expert Systems with Applications*, 4950–4958. doi:10.1016/j.eswa.2014.02.024
- Indrayuni, E., and Wahyudi, M., (2015). Penerapan Character N-Gram Untuk Sentiment Analysis Review Hotel Menggunakan Algoritma Naive Bayes. ISBN 978-602-72850-0-2
- Kang, H., Yoo, S. J., & Han, D.(2012). Senti lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews. *Expert Systems with Applications*, 6000–6010. doi:10.1016/j.eswa.2011.11.107
- Kontopoulos, E., Berberidis, C., Dergiades, T., & Bassiliades, N. (2013). Ontology-based sentiment analysis of twitter posts. *Expert Systems with Applications*, 40(10), 4065–4074. doi:10.1016/j.eswa.2013.01.001
- Martinez, I. P., Sanchez, F. G., Garcia, R. V., Moreno, V., Fraga, A., Cervantez, J. L. S. (2014). Feature-based opinion mining through ontologies. *Expert Systems with Applications*, 5995-6008. doi:10.1016/j.eswa.2014.03.022
- Park, T. S., Lee, J. H., & Choi, B. (2009). Optimization for Artificial Neural Network with Adaptive inertial weight of particle swarm optimization. *Cognitive Informatics, IEEE International Conference* , 481-485
- Routray, P., Swain, C. K. & Mishra, S. P., 2013. A Survey on Sentiment Analysis. *International Journal of Computer Applications*, Agustus, pp. 1-8.
- Tripathy, A., Agrawal, A., Rath, S. K. (2015). Classification of Sentimental Reviews Using Machine Learning Techniques. *ScienceDirect*, 57 ( 2015 ) 821 – 829. doi: 10.1016/j.procs.2015.07.523

Xhemali, D., J. Hinde, C. & G. Stone, R., 2009. Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages. *International Journal of Computer Science Issues*, pp. 16-23