**TESIS**

**OPINION MINING PADA TWITTER UNTUK BAHASA INDONESIA DENGAN METODE SUPPORT VECTOR MACHINE DAN METODE BERBASIS LEXICON**

**OPINION MINING ON TWITTER FOR INDONESIAN USING SUPPORT VECTOR MACHINE AND BASED LEXICON METHOD**

Diajukan untuk memenuhi salah satu syarat memperoleh derajat

Master of Computer Science

****

**JAN KRISTANTO WIBISONO**

**10/306141/PPA/03230**

**PROGRAM STUDI S2 ILMU KOMPUTER**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS GADJAH MADA**

**YOGYAKARTA**

**2013**

**HALAMAN PERNYATAAN**

Dengan ini saya menyatakan bahwa Tesis ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar Master di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 2, Juni 2013

JAN KRISTANTO WIBISONO

**HALAMAN PENGESAHAN**

**tesis**

**OPINION MINING PADA TWITTER UNTUK BAHASA INDONESIA DENGAN METODE SUPPORT VECTOR MACHINE DAN METODE BERBASIS LEXICON**

**oleh**

**Nama Mahasiswa**

**Nomor Pokok Mahasiswa (NPM)**

telah dipertahankan di depan Tim Penguji

pada tanggal ........(lengkapi *tanggal bulan tahun*)

**Susunan Tim Penguji**

**Ketua, Sekretaris,**

**Nama .......... Nama...................**

**NIP. .... NIP. ....**

**Penguji Utama, Penguji Sekretaris,**

**Nama .......... Nama...................**

**NIP. .... NIP. ....**

Diterima dan dinyatakan memenuhi syarat kelulusan pada tanggal .....

**Ketua Jurusan .....**

**Fakultas ....**

**Nama.............**

**NIP.................**

# KATA PENGANTAR

Segala puji bagi Allah SWT., Pencipta dan Pemelihara alam semesta, shalawat serta salam semoga terlimpah bagi Muhammad SAW., keluarga dan para pengikutnya yang setia hingga akhir masa.

Atas rahmat Allah SWT., akhirnya Penulis dapat menyelesaikan tesis ini, meskipun proses belajar sesungguhnya tak akan pernah berhenti. Tesis ini sesungguhnya bukanlah sebuah kerja individual dan akan sulit terlaksana tanpa bantuan banyak pihak yang tak mungkin Penulis sebutkan satu persatu, namun dengan segala kerendahan hati, Penulis mengucapkan terima kasih kepada …….. selama melaksanakan studi dan menyelesaikan tesis ini.

…..

Akhirnya, Penulis berharap semoga penelitian ini menjadi sumbangsih yang bermanfaat bagi dunia sains dan teknologi di Indonesia, khususnya disiplin keilmuan yang Penulis dalami.

Yogyakarta, Juni 2013

Penyusun

# DAFTAR ISI

[KATA PENGANTAR iv](#_Toc352620917)

[DAFTAR ISI v](#_Toc352620918)

[DAFTAR TABEL vii](#_Toc352620919)

[DAFTAR GAMBAR viii](#_Toc352620920)

[Intisari ix](#_Toc352620921)

[ABSTRACT x](#_Toc352620922)

[BAB I 1](#_Toc352620923)

[PENDAHULUAN 1](#_Toc352620924)

[1.1 Latar Belakang 1](#_Toc352620925)

[1.2 Rumusan Masalah 2](#_Toc352620926)

[1.3 Batasan Masalah 3](#_Toc352620927)

[1.4 Tujuan Penelitian 3](#_Toc352620928)

[1.5 Manfaat Penelitian 3](#_Toc352620929)

[1.6 Metode Penelitian 3](#_Toc352620930)

[1.7 Sistematika Penulisan 6](#_Toc352620931)

[BAB II 8](#_Toc352620932)

[TINJAUAN PUSTAKA 8](#_Toc352620933)

[2.1 Kajian Pustaka 8](#_Toc352620934)

[2.1.1 Subbab Pertama Derajat Kedua 8](#_Toc352620935)

[2.1.2 Subbab Pertama Derajat Kedua 8](#_Toc352620936)

[2.2 Kerangka Pemikiran 8](#_Toc352620937)

[2.3 Hipotesis 9](#_Toc352620938)

[BAB III 10](#_Toc352620939)

[LANDASAN TEORI 10](#_Toc352620940)

[3.1 Subbab Pertama 10](#_Toc352620941)

[3.1.1 Subbab Pertama Derajat Kedua 10](#_Toc352620942)

[3.1.2 Subbab Pertama Derajat Kedua 10](#_Toc352620943)

[3.2 Subbab Kedua 10](#_Toc352620944)

[3.2.1 Subbab Pertama Derajat Kedua 10](#_Toc352620945)

[3.2.2 Subbab Pertama Derajat Kedua 10](#_Toc352620946)

[3.3 Subbab Kedua 11](#_Toc352620947)

[3.3.1 Subbab Pertama Derajat Kedua 11](#_Toc352620948)

[3.3.2 Subbab Pertama Derajat Kedua 11](#_Toc352620949)

[BAB IV 12](#_Toc352620950)

[ANALISIS DAN RANCANGAN SISTEM 12](#_Toc352620951)

[4.1 Analisis Sistem 12](#_Toc352620952)

[4.1.1 Deskripsi Sistem 12](#_Toc352620953)

[4.1.2 Subbab Pertama Derajat Kedua 19](#_Toc352620954)

[4.2 Subbab Kedua 19](#_Toc352620955)

[4.2.1 Subbab Pertama Derajat Kedua 19](#_Toc352620956)

[4.2.2 Subbab Pertama Derajat Kedua 19](#_Toc352620957)

[4.3 Subbab Kedua 19](#_Toc352620958)

[4.3.1 Subbab Pertama Derajat Kedua 20](#_Toc352620959)

[4.3.2 Subbab Pertama Derajat Kedua 20](#_Toc352620960)

[BAB V 21](#_Toc352620961)

[IMPLEMENTASI 21](#_Toc352620962)

[5.1 Kesimpulan 21](#_Toc352620963)

[5.2 Saran 22](#_Toc352620964)

[BAB VI 23](#_Toc352620965)

[HASIL PENELITIAN DAN PEMBAHASAN 23](#_Toc352620966)

[BAB VII 24](#_Toc352620967)

[KESIMPULAN DAN SARAN 24](#_Toc352620968)

[DAFTAR PUSTAKA 25](#_Toc352620969)

\*) untuk meng-update isi daftar lakukan klik kanan pada salah satu judul, kemudian klik **Update Field,** selanjutnya **Update Entire Table**, lalu klik **OK**.

# DAFTAR TABEL

[Tabel 0.1. Contoh nomor dan judul tabel 9](#_Toc353190176)

[Tabel 0.2. Contoh nomor dan judul tabel 9](#_Toc353190177)

[Tabel 4.1 Daftar Konversi 16](#_Toc353190178)

[Tabel 4.2 Contoh Kopnversi Angka 17](#_Toc353190179)

[Tabel 4.3 Contoh menghapus karakter berualng 17](#_Toc353190180)

[Tabel 4.4 Contoh Pengantian kata tidak baku 18](#_Toc353190181)

[Tabel 4.5 Contoh SpellCorrection 19](#_Toc353190182)

[Tabel 4.6 Opinion Rules 21](#_Toc353190183)

\*) untuk meng-update isi daftar lakukan klik kanan pada salah satu judul, kemudian klik **Update Field,** selanjutnya **Update Entire Table**, lalu klik **OK**.

# DAFTAR GAMBAR

[Gambar II.1. Contoh untuk nomor dan judul gambar 8](#_Toc244585044)

[Gambar III.1. Contoh untuk nomor dan judul gambar 11](#_Toc244585045)

[Gambar IV.1. Contoh untuk nomor dan judul gambar 14](#_Toc244585046)

\*) untuk meng-update isi daftar lakukan klik kanan pada salah satu judul, kemudian klik **Update Field,** selanjutnya **Update Entire Table**, lalu klik **OK**.

# Intisari

Tulislah isi abstrak dalam bahasa Indonesia di sini. Tulislah isi abstrak dalam bahasa Indonesia di sini. Tulislah isi abstrak dalam bahasa Indonesia di sini. Tulislah isi abstrak dalam bahasa Indonesia di sini. Tulislah isi abstrak dalam bahasa Indonesia di sini. Tulislah isi abstrak dalam bahasa Indonesia di sini. Tulislah isi abstrak dalam bahasa Indonesia di sini.

# ABSTRACT

Tulislah isi abstrak dalam bahasa Inggris di sini. Tulislah isi abstrak dalam bahasa Inggris di sini. Tulislah isi abstrak dalam bahasa Inggris di sini. Tulislah isi abstrak dalam bahasa Inggris di sini. Tulislah isi abstrak dalam bahasa Inggris di sini. Tulislah isi abstrak dalam bahasa Inggris di sini. Tulislah isi abstrak dalam bahasa Inggris di sini.

# 

# PENDAHULUAN

## Latar Belakang

Opini adalah pendapat pribadi yang tidak obyektif dan tidak melalui proses verifikasi (Quirk et al., 1985). Opini orang lain tentang suatu menjadi penting dalam pengambilan keputusan. Misalkan seseorang akan membeli sebuah mobil biasanya dia akan bertanya kepada orang lain bagaimana pendapat mereka tentang mobil tersebut. Atau sering juga ketika kita membeli produk tertentu, kita dipersilakan untuk mengisi kuisoner yang berkaitan dengan produk tersebut. Sekarang ini tidak perlu kita langsung bertanya tentang pendapat orang lain secara langsung. Seiring dengan perkembangan internet dan web, telah banyak orang yang berbagi informasi tentang pendapat dan pengalaman bahkan kritik mereka terhadap suatu produk tertentu.

Sekarang ini *social media* telah berkembang dengan pesat, dimulai dari awal Friendster sampai era sekarang ini Facebook, Twitter, Google+ dan Foursquare. Menurut <http://socialbakers.com> pengguna Facebook di Indonesia telah mecapai 42.5 juta, yang menjadikan Indonesia menjadi peringkat ke empat pengguna Facebook setelah Amerika, India dan Brazil. Pengguna Twitter di Indonesia juga tidak kalah besar jumlahnya yang mencapai 19.5 juta ([http://semiocast.com](http://semiocast.com/)) yang menempati peringkat kelima setelah Amerika, Brazil, Japan dan U.K. Perkembangan yang begitu pesat ini membuat orang-orang berkomunikasi dengan mudah. Mereka saling berbagi informasi, baik berbagi tentang aktifitas sehari-hari mereka ataupun memberikan pemberitahuan pada komunitas mereka, bahkan keluhan mereka terhadap suatu hal.

Begitu populernya *social media*, pengguna *social media* sekarang ini bukan hanya kalangan pribadi atau perseorangan saja. Tapi kalangan *corporate* juga memanfaatkan *social media* sebagai media komunikasinya dengan masyarakat. Seperti pengenalan produk baru, penawaran produk, bahkan para legislative juga mempromosikan diri mereka melalui *social media*. Selain yang dianggap lebih murah dari segi biaya, juga lebih tepat pada sasaran.

Berdasarkan pada paparan diatas terdapat potensi besar pada social media untuk dilakukan pengekstrakan opini atau yang disebut *opinion mining,* untuk mengetahui pendapat orang lain tentang suatu hal yang dapat dimanfaatkan selama proses pengambilan keputusan. Beberapa penelitian yang sudah ada sekarang ini kebanyakan menggunakan *machine learning* baik dengan naïve bayes (aliandu., 2012 , Pang 2002, Franky., 2008, Go et al., 2009, Pak dan Paurobek., 2009) , Support Vector Machine (Pang., 2002 , Franky., 2008 , Zhang., 2008, Go et al., 2009) dan Maximum Entropy(Go et al., 2009) ataupun yang berbasiskan lexicon (komansilan., 2012). Pada penelitian ini akan dilakukan *opinion mining* dengan mengkombinasikan antara pendekatan berbasis *machine learning* dan berbasiskan lexicon. Metode yang akan digunakan untuk mewakili pendekatan *machine learning* adalah Support Vector Machine (SVM), yang menurut paper Zhang (2008) sering memberikan akurasi yang terbaik dibanding dengan metode learning yang lainnya. Dengan melakukan kombinasi antara metode Support Vector Machine (SVM) dan berbasiskan lexicon diharapkan dapat meningkatkan akurasi pada proses pengklasifikasian opini.

## Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dipaparkan diatas dapat dirumuskan hal-hal sebagai berikut :

1. Bagaimana melakukan *preprocessing* terhadap kalimat tidak baku.
2. Bagaimana melakukan *opinion mining* pada *social media* untuk bahasa Indonesia.
3. Bagaimana mengkombinasikan Support Vector Machine (SVM) dengan Lexicon Based untuk menentukan suatu kalimat beropini positif, negative atau netral.

## Batasan Masalah

Adapun yang menjadi batasan masalah pada penelitian ini adalah :

1. Bahasa yang digunakan dalam *opinion mining* dalam penelitian ini adalah bahasa Indonesia.
2. Sumber dari *social media* yang digunakan adalah dari Twitter.
3. Untuk menentukan apakah suatu kalimat merupakan kalimat positif, negatif, maupun netral adalah dengan mengkombinasikan antara metode *lexicon-based* dan *learning-based*

## Tujuan Penelitian

Tujuan yang ingin dicapai pada penelitian ini adalah menciptakan mesin yang dapat melakukan *opinion mining* dalam bahasa Indonesia pada *social media* dengan mengkombinasikan metode *lexicon-based* dan *learning-based*.

## Manfaat Penelitian

Penelitian ini diharapkan dapat menghasil mesin yang bisa mengekstrak opini orang lain yang dapat menjadi pertimbangan dalam proses pengambilan keputusan.

## Metode Penelitian

Metode penelitian yang akan dilakukan terdiri atas beberapa tahap sebagai berikut :

1. Studi Kepustakaan

Melakukan berbagai macam pengumpulan bahan referensi, seperti jurnal penelitian, tesis, buku-buku teori dan sumber lain termasuk informasi yang diperoleh dari internet. Mempelajari litelatur dan teori pendukung penelitian yaitu mengenai klasifikasi khususnya SVM (Support Vector Machine), Twitter API.

1. Analisis Sistem

Melakukan analisa terhadap requirement yang dibutuhkan dalam pengembangan sistem.

1. Perancangan

Pada tahap ini akan dilakukan perancangan metode berdasar lexicon, dan *machine learning*, yaitu Support Vector Machine (SVM) yang akan diimplementasikan untuk *Opinion mining*. Data-data yang diambil dalam penetian ini diambil dari twitter dengan memanfaatkan Twitter API yang sudah disediakan, selanjutnya data tersebut disimpan dalam basis data. Langkah pertama dalam proses besar penelitian ini adalah melakukan *preprocessing* data tweet. Setelah dilakukan *preprocessing*, sistem akan mencari kalimat yang mengandung frasa opini, sehingga kalimat yang tidak mengandung frasa opini akan langsung dimasukan dalam kelas netral. Kemudian sistem akan menentukan sentimen tweet yang memiliki frase opini tersebut menggunakan metode *lexicon based* apakah positif, negatif atau netral. Hasil dari proses lexicon based ini akan menjadi data training pada proses penentuan sentimen tweet baru yang kali ini akan dilakukan dengan *machine learning* yaitu dengan metode Support Vector Machine (SVM).

Secara garis besar gambaran umum dari penelitan ini ditunjukan seperti gambar 1:

Crawl tweets

Preprocessing

Machine Learning Method

Lexicon based method

Classifier sentiment

Gambar 1. Gambaran Umum Sistem yang akan dibangun

1. Implementasi

Pada tahap ini akan dilakukan pembuatan kode program sampai proses pembangunan basis data.

1. Pengujian

Pada tahap ini, dilakukan perhitungan akurasi dari implementasi yang telah dilakukan. Metode yang digunakan adalah *precision , recall* dan *F-score*. Menurut Risbergen (1979), *Precision* adalah perbandingan antara jumlah kalimat yang mengandung opini baik sentimen positif maupun negatif yang dapat dikelompokan oleh sistem dengan jumlah seluruh kalimat yang diambil oleh system. Sedangkan *Recall* adalah perbandingan antara jumlah kalimat yang mengandung opini, baik sentimen positif atau negatif yang dapat dikelompokan oleh sistem dengan jumlah kalimat yang mengandung opini, baik positif maupun negatif yang berada dalam koleksi dokumen. *F-score* adalah 2 kali perbandingan antara perkalian nilai *Precision* dan nilai *Recall* dengan penjumlahan antara *Precision* dan *Recall*.

## Sistematika Penulisan

BAB I PENDAHULUAN

Bab ini memuat latar belakang masalah, rumusan masalah, batasan masalah,tujuan penelitian, manfaat penelitian, metode penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini memuat penelitian-penelitian terdahulu yang berkaitan dengan penelitian yang dilakukan penulis. Beberapa penelitian terkait Opinion Miningatau *Sentiment analysis* dibahas secara singkat dan dibandingkan dengan penelitian penulis.

BAB III LANDASAN TEORI

Bab ini membahas teori-teori yang berkaitan dengan penelitian. Bagian pertama membahas tentang Opinion Mining, permasalahannya, dan beberapa pendekatan yang digunakan. Bagian kedua membahas tentang metode *lexicon based* untuk Opinion Mining beserta contohnya. Bagian ketiga membahas tentang Support Vector Machine (SVM)dan bagaimana model tersebut bekerja. Bagian terakhir masih dipikirkan.

BAB IV ANALISIS DAN RANCANGAN SISTEM

Bab ini membahas analisis kebutuhan data, analisis model beserta rancangan sistemnya.

BAB V IMPLEMENTASI

Bab ini membahas implementasi opinion mining dalam Support Vector Machine sesuai rancangan pada bab sebelumnya.

BAB VI HASIL DAN PEMBAHASAN

Bab ini membahas percobaan yang dilakukan pada proses pelatihan dan penentuan kelas beserta uraian mengenai hasil dan perbandingannya.

BAB VII KESIMPULAN DAN SARAN

Bab ini memuat kesimpulan-kesimpulan dari hasil penelitian dan saran-saran yang berguna untuk penelitian selanjutnya.

# 

# TINJAUAN PUSTAKA

## Kajian Pustaka

Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini.

### Subbab Pertama Derajat Kedua

Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini.

### Subbab Pertama Derajat Kedua

Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini.

j0217698

Gambar .. Contoh untuk nomor dan judul gambar

## Kerangka Pemikiran

Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini.

## Hipotesis

Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini.

Tabel .. Contoh nomor dan judul tabel

|  |  |  |
| --- | --- | --- |
| No. | Isi tabel | Isi Tabel |
| 1 | Isi tabel | Isi Tabel |
| 2 | Isi tabel | Isi Tabel |
| 3 | Isi tabel | Isi Tabel |

Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini.

Tabel .. Contoh nomor dan judul tabel

|  |  |  |
| --- | --- | --- |
| No. | Isi tabel | Isi Tabel |
| 1 | Isi tabel | Isi Tabel |
| 2 | Isi tabel | Isi Tabel |
| 3 | Isi tabel | Isi Tabel |

##### (II.1)

##### (.2)

##### (.3)

# 

# LANDASAN TEORI

## Subbab Pertama

Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini.

### Subbab Pertama Derajat Kedua

Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini.

### Subbab Pertama Derajat Kedua

Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini.

## Subbab Kedua

Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini.

### Subbab Pertama Derajat Kedua

Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini.

### Subbab Pertama Derajat Kedua

Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini.

## Subbab Kedua

Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini.

### Subbab Pertama Derajat Kedua

Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini.

### Subbab Pertama Derajat Kedua

Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini.

j0217698

Gambar .. Contoh untuk nomor dan judul gambar

##### (III.1)

##### (III.2)

# 

# ANALISIS DAN RANCANGAN SISTEM

## Analisis Sistem

*Opinion mining* merupakan sebuah sistem yang dapat digunakan untuk menentukan sentimen atau opini secara otomatis pada sebuah dokumen. Penelitian ini akan mengambil data dari *twitter* yang berbahasa indonesia. Sistem yang dibangun dan dirancang adalah sistem yang memiliki kemampuan untuk menentukan sentimen dari *tweet* hasil *query user* atas *term* pada aplikasi yang terhubung dengan Twitter Search API. Metode yang digunakan untuk menentukan sentimen adalah metode gabungan *unsupersive* yang berbasis pada analisa bahasa dan metode *supervise* yaitu *Support Vector Machine* (SVM).

*Data training* yang digunakan adalah tweet yang juga digunakan dalam penelitian paulina. Data tersebut akan dikategorikan ke dalam kelas sentimen positif, negatif dan netral dengan menggunakan analisis bahasa (*unsupervise*). Hasil pengkategorian *data training* digunakan untuk menentukan sentiment pada *tweet* yang baru dengan menggunakan metode *supervise* yaitu *Supoort Vector Machine* (SVM).

### Deskripsi Sistem

Sistem Opinion Mining ini memiliki dua buah proses utama, yaitu pelatihan dan penentuan kelas sentimen pada data baru. Pelatihan digunakan untuk mendapatkan model yang akan digunakan untuk menentuan kelas sentimen pada data baru. Proses-proses yang terjadi dalam sistem ini diperlihatkan pada gambar.

Pada proses pelatihan dengan menggunakan masukan berupa corpus yang digunakan pada penelitian paulina yang telah dilakukan pelabelan dengen metode lexicon based. Langkah awal dari pelabelan dengan menggunakan lexicon based adalah melakukan preprocessing terhadap data. Dalam preprocessing terdapat beberapa tahap, secara berurutan adalah membersihkan yang bukan termasuk karater UTF-8, merubah semua karakter menjadi huruf kecil, menghapus simbol-simbol, memproses angka-angka, menghapus perulangan karakter, melakukan penggantian terhadap kata tidak baku, melakukan spelling correction.



Gambar Proses-proses dalam Opinion Mining

Setelah selesai tahap preprocessing, selanjut akan masuk dalam tahan penentuan label. Dalam tahap ini juga terdapat beberapa proses diantaranya Postagging, mengalasis tweet dengan rules yang sudah ada, mengecek negasi. Sampai disini telah dihasil corpus yang labelnya ditentukan dengan menggunakan lexicon based dan siap digunakan untuk proses training.

Proses berikutnya adalah proses inti pelatihan, dalam penelitian ini terdapat 25000 data pelatihan yang terdiri dari data 12500 data berlabel positif dan 12500 data berlabel negatif. Proses pelatihan ini nantinya akan menghasilkan svm model. Yaitu sebuah model yang nantinya digunakan untuk menentukan sentiment dengan metode SVM.

Proses utama yang kedua adalah penentuan label tweet baru dengan metode SVM. Langkah pertama pada proses utama ini adalah sistem akan mencari tweet yang berhubungan dengan kata kunci yang diinginkan. Selanjutnya akan masuk dalam tahap preprocessing yang telah dijelaskan sebelumnya. Kemudian dilakukan pengecekan indikasi keberadaan sentimen pada tweet yang telah dipreprocessing. Selanjutnya untuk tweet yang terindikasi memiliki sentimen, sentimennya akan ditentukan dengan metode SVM.

## Analisis Data

Data yang dibutuhkan pada penelitian ini terdiri atas dua jenis data, yaitu data yang digunakan untuk pelatihan dan data pengujian pemberian label. Data yang digunakan untuk pelatihan berupa dokumen tweet berbahasa indonesia yang diberi label dengan metode lexicon based. Dokumen ini diambil dari penelitian paulina yang jumlah 81500 tweet. Namun setelah diproses menggunakan lexicon based data yang digunakan untuk pelatihan hanya 25000.

Berbeda dengan data yang digunakan untuk pelatihan , data yang digunakan pada proses pelabelan menggunakan tweet yang belum memiliki label. Dokumen tweet tersebut didapatkan dengan memanfaatkan API twitter yang dapat mencari tweets yang ada pada twitter dengan parameter tertentu.

## Preprocessing

Preprocessing sangat menentukan dalam proses pelabelan secara lexicon based. Dimana pelabelan lexicon based nantinya akan menentukan proses pelabelan yang dilakukan dengan metode SVM. Preprocessing juga digunakan dalam kedua proses utama, baik proses pelatihan maupun proses penentuan label dengan SVM. Jadi preprocessing ini cukup penting peranannya dalam penelitian ini.

Tahap preprocessing sendiri terdiri atas beberapa proses yang akan dibahas satu per satu secara lebih mendetail. Proses-proses tersebut diantaranya adalah pembersihkan karater yang bukan termasuk UTF-8 (clear invalid UTF8), merubah ke huruf kecil (casefolding), menghapus simbol-simbol(remove symbol), memproses angka-angka(processing numbers), menghapus perulangan karakter (remove repeat), melakukan penggantian terhadap kata tidak baku(replace slang), melakukan spelling correction(spelling correction).

### Clear Invalid UTF-8

Proses ini adalah bagian dari tahap preprocessing. Pada intinya proses ini akan menghilakan karakter-karakter yang tidak termasuk dalam UTF-8. Secara teknik proses ini dilakukan dengan fitur pada bahasa pemrograman yang sudah disediakan.

### Casefolding

Proses ini hanyalah merupan semua karakter yang ada pada tweet menjadi karakter dalam huruf kecil semua. Hal ini dilakukan karena tidak ada kaitan antara huruf kecil dan huruf besar terhadap sentimen, serta mempermudah proses-proses selanjutnya.

### Remove Symbol

Proses ini adalah proses selanjutnya dari tahap preprocessing. Pada proses ini tanda baca seperti titik(.) , koma(,), seru(!), tanya (?) dan lain sebagainya akan dihilangkan. Tanda/simbol ini dapat dihilangkan karena tidak banyak berpengaruh pada penentuan label. Secara teknik ini akan dilakukan dengan bahasa pemrograman dengan mencari simbol-simbol kemudian menghapusnya.

### Processing Numbers

Proses selanjuatnya dalam tahap preprocessing adalah pemrosesan terhadap angka-angka. Kita tahu dalam sosial media, pengguna sering menggunakan angka-angka baik untuk menyingkat kata, atau untuk bergaya atau untuk menyebutkan nilai. Angka-angka yang terdapat dalam sebuah kata akan di convert menurut tabel.

Tabel Daftar Konversi

|  |  |
| --- | --- |
| Angka | Konversi |
| 0 | O |
| 1 | i |
| 2 | Kopi karakter-karakter sebelumnya |
| 3 | e |
| 4 | a |
| 5 | s |
| 6 | g |
| 7 | T |
| 8 | b |
| 9 | g |

Berikut adalah beberapa contoh kata yang akan di convert oleh proses ini bila ditemukan.

Tabel Contoh Kopnversi Angka

|  |  |
| --- | --- |
| Masukan | Keluaran |
| ga2l | gagal |
| p3rg1 | pergi |
| Hati2 | Hati-hati |
| Ba9u5 | Bagus |

### Remove Repeat

Proses ini salah satu proses penting dalam tahap prepocessing. Proses ini akan menghapus karakter yang berulang lebih dari dua kali. Seperti diketahui dalam bahasa indonesia perulangan karakter yang berurutan hanya dua karakter, maka jika lebih akan dihapus dan hanya akan disisakan 2 karakter yang sama yang berurut saja. Untuk memperjelas perhatikan tabel berikut

Tabel 4. Contoh menghapus karakter berualng

|  |  |
| --- | --- |
| Input | Ouput |
| Tidaaaaaaaaaak | Tidaak |
| Cemunguuuuudh | Cemunguudh |
| eaaaaaa | eaa |

### Replace Slang

Proses ini merupakan proses penggantian kata tidak baku menjadi kata baku. Proses ini dilakukan dengan menggunakan bantuan daftar kata-kata informal dan padanannya dalam kata-kata formal. Proses ini akan mengecek kata terdapat dalam daftar kata informal atau tidak. Jika tidak maka kata tidak akan diganti tetapi bila ya maka kata akan diganti dengan padanannya yang terdapat dalam daftar kata baku. Daftar kata baku dan kata tidak baku ini menggunakan daftar yang digunakan pada pelatihan edmond, yang sudah ditambah secara manual berdasarkan pengamatan dalam proses pengembangan sistem ini. Contoh dari proses ini akan ditunjukan pada tabel berikut

Tabel Contoh Pengantian kata tidak baku

|  |  |
| --- | --- |
| Kata tidak baku | Kata baku |
| Nggak | Tidak |
| Lu | Kamu |
| Ngadu | mengadu |

### Spelling Correction

Spelling Correct adalah proses terakhir pada tahap preprocessing. Proses ini akan memperbaiki kesalahan penulisan yang ada data tweet. Seperti kita tahu pengguna sosial media sering sengaja menyingkat kata untuk memenuhi 140 karakter (jumlah maksimal karakter pada twitter) atau secara tidak sengaja salah menuliskan kata. Metode yang digunakan akan dengan memanfaatkan daftar kata baku yang sudah ada dan dengan digabungkan dengan fungsi levenshtein. Fungsi levenshtein adalah fungsi yang dapat menghitung kedekatan/kemiripan antar kata. Secara teknis kata yang telah melalui proses sebelumnya dan tidak terdapat dalam daftar kata baku, maka kata ini akan dikenai proses ini. Berikut adalah contoh kata-kata yang dapat ditangani oleh proses ini.

Tabel Contoh SpellCorrection

|  |  |
| --- | --- |
| Input | Ouput |
| Bapk | Bapak |
| Menghitumg | menghitung |
| Yng | yang |

Walaupun proses ini memberikan manfaat yang cukup besar. Namun terkadang proses ini memberikan kelemahan. Hal ini terjadi ketika proses ini diberikan input suatu frase atau kata serapan baru yang belum terdapat dalam dafatar kata baku. Misalkan BLBI oleh karena BLBI tidak terdapat dalam daftar kata baku maka BLBI akan digantikan dengan kata yang terdekat dengan kata BLBI. Padahal yang dimaksud dengan kata ini adalah kata serapan baru.

## Penentuan Label Data Training

Tahap selanjutnya setelah preprocessing data adalah penentuan label dengan metode lexicon based. Tahap ini masih dalam proses utama yang pertama yaitu proses pelatihan. Pada tahap penentuan label ini terdapat beberapa proses di dalamnya. Seperti ditunjukan pada gambar, proses-proses tersebut diantaranya PosTagging, Opinion Rules dan check negation.

### PosTagging

Setelah tahap preprocessing selesai berarti data training sudah bersih dan telah siap untuk dilakukan proses penentuan label secara lexicon based. Proses pertama yang dilakukan pada tahap ini adalah melakukan PosTagging pada data tersebut.

PosTagging adalah proses untuk menentukan kelas kata dari setiap kata-kata yang terdapat dalam kalimat. Proses PosTagging dalam penelitian ini menggunakan metode dictionary. Misalnya terdapat satu tweet “Kasus penembakan yang dilakukan oleh polisi sangat rumit” maka dengan PosTagging akan didapatkan “Kasus/NN penembakan/VB yang/CK dilakukan/VB oleh/IN polisi/NN sangat/RB rumit/JJ”.

Langkah pertama adaalah melakukan tokenisasi, yaitu pross yang memecah-mecah tweet menjadi satuan kata. Dimana sistem akan melihat setiap spasi yang ada dlam tweet dan berdasarkan spasi tersebut maka kata-kata dapat dipecah.

Langkah kedua yang dilakukan adalah melakukan penyesuaian dengan koleksi kata yang terdapat dalam database. Setiap kata dari hasil tokenisasi tersemua kemudian dicek satu-persatu kelas katanya sesuai dengan koleksi kata dalam database. Jika kata tersebut terdapat dalam database maka akan langsung ditentukan kelas katanya. Namun apabila kata tersebut tidak ditemukan dalam database, maka kata tersebut akan dianggap sebagai kata benda(NN).

Apabila dilihat dari cara kerja PosTagging ini, ada kata-kata yang tidak bisa dideteksi dengan benar, karena karakteristiknya tidak bisa menangani kata-kata yang memiliki konteks dan maknna ganda. Misalkan pada tweet “Mobil polisi menarik mobil sedang yang kecelakaan.” Kata menarik dalam kali ini akan dideteksi menjadi kata sifat padahal bila dilihat dari konteks nya kata ‘menarik’ disini adalah kata kerja.

### Opinion Rules

Seperti yang telah dijelaskan sebelumnya, dalam penentuan label data training akan menggunakan metode lexicon based, dimana penentuan labelnya menggunakan rule-rule tertentu. Rule-rule ini mengacu pada penelitian yang dilakukan oleh Edmond.

Secara teknis langkah yang dilakukan adalah dengan melihat sekuens dari kata-kata hasil PosTagging sebelumnya yang sudah menyertakan kelas kata dari masing-masing kata. Rule yang dipakai untuk mendeteksi frasa yang terdiri dari dua kata saja dan kata-kata yang berdiri sendiri. Pada tabel berikut ditampilkan daftar rule frasa opini yang digunakan dalam sistem ini.

Tabel 4. Opinion Rules

|  |  |
| --- | --- |
| NO | Rule |
| 1 | RB JJ |
| 2 | RB VB |
| 3 | CK VB |
| 4 | NN JJ |
| 5 | NN VB |
| 6 | VB JJ |
| 7 | JJ VB |

Dalam penentuan frase ini terlebih dahulu mendeksi kehadiran dari kata sifat atau kata kerja. Setelah sistem dapat mendeteksi kehadiran dari kata kerja atau kata sifat dalam kalimat, maka sistem akan melihat kata sebelum kata kerja atau kata sifat dan melihat kelas katanya. Sehingga apabila kedua kelas kata dari kedua kata tesebut diurutkan makan akan dicocokan dengan rule yang ada. Jika sama dengan rule yang ada maka frase tersebut bisa dikatakan memiliki opini atau sentimen.

Misal terdapat suatu kalimat “Film itu sangat bagus”, oleh PosTagging akan ditandai sebagai “Film/NN itu/DT sangat/RB bagus/JJ”. Dari hasil dapat dilihat ada frasa yang sesua dengan rule yaitu ‘sangat bagus’ (RB JJ) karena kata ‘bagus’ memiliki sentimen positif maka kalimat tersebut diberi label positif.

### Check Negation

Proses terakhir pada tahap penentuan rule ini adalah mengecek negasinya. Setalah tweet dilakukan PosTagging dan dicocokan dengan rule-rule yang ada. Apabila terdapat sentimen dari rule-rule dari maka selanjutnya akan dicek apakah terdapat negasi dari hasil rule-rule tadi.

Konsepnya mencari kata negasi sebelum kata sifat atau kata kerja. Dimana daftar kata negasi diambil dari database. Toleransi jarak yang digunakan adalah 5 kata dari kata sifat atau kata benda. Misalkan terdapat tweet “Film itu tidak bagus”. Pertama akan dilakukan PosTagging sehingga menghasilkan seperti berikut “Film/NN itu/DT tidak/MD bagus/JJ”. Kemudian akan dilakukan opinion rule yang akan menghasilkan frase positif yang dihasilkan dari kata ‘bagus’. Proses check negation akan melakukan pengecekan kata-kata sebelum frase tersebut. Jika terdapat kata negasi maka labelnya akan dibalik. Jadi penentuan dari tweet tersebut akan menghasilkan label negatif.

## Pelatihan

Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini.

### Subbab Pertama Derajat Kedua

Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini.

### Subbab Pertama Derajat Kedua

Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini. Tuliskan isi subbab di sini.

j0217698

Gambar .. Contoh untuk nomor dan judul gambar

Untuk membuat bab baru lakukan dengan Next Page pada menu Break dalam tab Page layout.

# 

# IMPLEMENTASI

## Kesimpulan

Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini.

## Saran

Tuliskan saran di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini. Tuliskan kesimpulan dari penelitian di sini.

# 

# HASIL PENELITIAN DAN PEMBAHASAN

# 

# KESIMPULAN DAN SARAN

# DAFTAR PUSTAKA

1. Brown, Steven and Zvonko Vranesic. *Fundamentals of Digital Logic with VHDL Design.* Vol. II. New York: Mc Graw-Hill, 2005.
2. Daemen, Joan and Vincent Rijmen. "AES Proposal: Rijndael." 09 Maret 1999. *http:// www. esat.kuleuven.ac.be/ ~rijmen/ rijndael.pdf.* 09 Juni 2007 <http:// www. esat.kuleuven.ac.be>.
3. Hamblen, James O., Tyson S. Hall and Michael D. Furman. *Rapid Prototyping of Digital Systems Quartus II Edition.* New York: Springer, 2006.
4. Krishna, C. M. and Kang G. Shin. *Real-Time Systems.* New York: McGraw-Hill, 1997.
5. Stallings, William. *Cryptography and Network Security Principles and Practices.* Fourth Edition. New Jersey: Prentice Hall, 2005.
6. Sutikno, Sarwono and Yusuf Kurniawan. "The Cryptanalysis of Block Cipher BC2." 2006. 4 Juni 2008 <http://www.jakarta2006.engage-ist.org/fileadmin/engage/public\_jakarta2006/presentations/ A4\_04\_Sarwono.pdf>.
7. Agustini, Sari and Danang Jaya. "Kriteria Teknik Kriptoragrafi untuk Image Stenografi berbasis Singular Value Decompresition." *SNIKTI VII.* Jakarta: Fakultas Ilmu Komputer UI, 2007. 500-506.
8. Langi, Armien Z. R. "A Rural Next Generation Network (NGN) and Its Testbed." *Itb journal of Information and Communication Technology* (2007): 1-15.
9. Pressman, Roger S. *Software Engineering: A Practitioner's Approach.* New York: McGraw-Hill, 2001.
10. Venkatraman, Dheera. "Methods and implementation of quantum cryptography." 27 April 2004*. http://dheera.net.* 13 Oktober 2008 <http://dheera.net/sci/qcrypt.pdf >.