

­­

**ANALISIS SENTIMEN OTOMATIS DI MEDIA SOSIAL TERHADAP KINERJA PEMERINTAHAN KABUPATEN JEMBER BERBASIS TEXT MINING**

Proposal Skripsi

Diajukan guna melengkapi salah satu syarat

untuk melengkapi tugas akhir

oleh:

Vananda Rahadika 132410101079

**PROGRAM STUDI SISTEM INFORMASI**

**UNIVERSITAS JEMBER**

**2016**

* 1. **Judul**

Analisis Sentimen Otomatis Di Media Sosial Terhadap Kinerja Pemerintahan Kabupaten Jember Berbasis Text Mining

* 1. **Latar Belakang**

Pembuatan.

Berdasarkan survei yang dilakukan oleh APJII (Asosiasi Penyelenggara Jasa Internet Indonesia) pada tahun 2016, jumlah pengguna internet Indonesia mencapai 88.1 juta jiwa, dimana media sosial mendapatkan peringkat tertinggi untuk jenis konten yang sering diakses (129.2 juta), dimana Facebook merupakan situs yang paling dikunjungi yaitu sebesar 71.6 juta kali.

Hal ini tentu saja membuat data yang berada di internet menjadi melimpah

* 1. **Rumusan Masalah**
  2. **Batasan Masalah**
  3. **Tujuan**
  4. **Tinjauan Pustaka**

F.1 Penelitian Terdahulu

Penelitian sebelumnya mengenai analisis sentimen dilakukan oleh Falahah dengan judul Pengembangan Aplikasi Sentimen Analysis Menggunakan Metode Naïve Bayes pada tahun 2015. Pada jurnal tersebut penulis berhasil melakukan analisis sentimen pada kota Bandung menggunakan data training di twitter dan melakukan analisis sentimen menggunakan Naïve Bayes. Hasil yang didapat dari penelitian itu menunjukkan bahwa metode Naïve Bayes Classifier dapat diterapkan sebagai metode untuk melakukan klasifikasi Sentiment Analysis. Pengumpulan data yang dilakukan menggunakan aplikasi KNIME, dimana akurasi yang didapat mencapai 73%.

Penelitian lain dilakukan oleh Nuke Y.A. Faradhillah, dengan judul jurnal Eksperimen Sistem Klasifikasi Analisa Sentimen Twitter Pada Akun Resmi Pemerintah Kota Surabaya Berbasis Pembelajaran Mesin. Pada jurnal tersebut penulis melakukan analisis sentiment pada akun twitter @e100ss dan @sapawargaSby selama periode 1 September 2015 sampai dengan 13 Oktober 2015. Klasifikasi dilakukan dengan dua metode yaitu memakai Support Vector Machine (SVM) dan Naïve Bayes Classifier. Hasil dari klasifikasi ini memiliki akurasi sebesar 79.81%.

Berdasarkan beberapa pustaka diatas dapat diambil kesimpulan bahwa Naïve Bayes Classifier adalah metode yang sering banyak dipakai dalam mengklasifikan sentiment dan memiliki akurasi lebih dari 70%.

F.2 Text Mining

*Text mining* (penambangan teks) adalah penambangan yang dilakukan oleh komputer untuk mendapatkan sesuatu yang baru, sesuatu yang tidak diketahui sebelumnya atau menemukan kembali informasi yang tersirat secara implisit, yang berasal dari informasi yang diekstrak secara otomatis dari sumber-sumber data teks yang berbeda-beda (Feldman & Sanger, 2007). *Text mining* merupakan teknik yang digunakan untuk menangani masalah klasifikasi, *clustering*, *information extraction* dan *information retrival* (Berry & Kogan, 2010)*.*

Pada dasarnya proses kerja dari *text mining* banyak mengapdopsi dari penelitian *Data Mining* namun yang menjadi perbedaan adalah pola yang digunakan oleh *text mining* diambil dari sekumpulan bahasa alami yang tidak terstruktur sedangkan dalam *Data Mining* pola yang diambil dari *database* yang terstruktur (Han & Kamber, 2006). Tahap-tahap t*ext mining* secara umum adalah *text preprocessing* dan *feature selection* (Feldman & Sanger 2007, Berry & Kogan 2010) . Dimana penjelasan dari tahap-tahap tersebut adalah sebagai berikut :

*F.2.1 Text Preprocessing*

Tahap *text preprocessing* adalah tahap awal dari *text mining*. Tahap ini mencakup semua rutinitas, dan proses untuk mempersiapkan data yang akan digunakan pada operasi *knowledge discovery* sistem *text mining* (Feldman & Sanger, 2007). Tindakan yang dilakukan pada tahap ini adalah *toLowerCase*, yaitu mengubah semua karakter huruf menjadi huruf kecil dan *Tokenizing* yaitu proses penguraian deskripsi yangsemula berupa kalimat-kalimat menjadi kata-kata dan menghilangkan delimiter-delimiter seperti tanda titik (.), koma (,), spasi dan karakter angka yang ada pada kata tersebut (Weiss et al, 2005).

*F.1.2 Feature Selection*

Tahap seleksi fitur (*feature selection*) bertujuan untuk mengurangi dimensi dari suatu kumpulan teks, atau dengan kata lain menghapus kata-kata yang dianggap tidak penting atau tidak menggambarkan isi dokumen sehingga proses pengklasifikasian lebih efektif dan akurat (Do et al, 2006., Feldman & Sanger, 2007., Berry & Kogan 2010). Pada tahap ini tindakan yang dilakukan adalah menghilangkan *stopword* (*stopword removal*) dan *stemming* terhadap kata yang berimbuhan (Berry & Kogan 2010., Feldman & Sanger 2007).

*Stopword* adalah kosakata yang bukan merupakan ciri (kata unik) dari suatu dokumen. Misalnya “di”, “oleh”, “pada”, “sebuah”, “karena” dan lain sebagainya. Sebelum proses *stopword removal* dilakukan, harus dibuat daftar *stopword* (*stoplist*). Jika termasuk di dalam *stoplist* maka kata-kata tersebut akan dihapus dari deskripsi sehingga kata-kata yang tersisa di dalam deskripsi dianggap sebagai kata-kata yang mencirikan isi dari suatu dokumen atau *keywords*.

Setelah melalui proses *stopword removal* tindakan selanjutnya adalah yaitu proses *stemming*. *Stemming* adalah proses pemetaan dan penguraian berbagai bentuk (*variants*) dari suatu kata menjadi bentuk kata dasarnya *(stem)* ( Tala, 2003)*.* Tujuan dari proses *stemming* adalah menghilangkan imbuhan-imbuhan baik itu berupa prefiks, sufiks, maupun konfiks yang ada pada setiap kata. Jika imbuhan tersebut tidak dihilangkan maka setiap satu kata dasar akan disimpan dengan berbagai macam bentuk yang berbeda sesuai dengan imbuhan yang melekatinya sehingga hal tersebut akan menambah beban *database*. Hal ini sangat berbeda jika menghilangkan imbuhan-imbuhan yang melekat dari setiap kata dasar, maka satu kata dasar akan disimpan sekali walaupun mungkin kata dasar tersebut pada sumber data sudah berubah dari bentuk aslinya dan mendapatkan berbagai macam imbuhan. Karena bahasa Indonesia mempunyai aturan morfologi maka proses *stemming* harus berdasarkan aturan morfologi bahasa Indonesia.

Berdasarkan penelitian sebelumnya, ada beberapa algoritma stemming yang bisa digunakan untuk *stemming* bahasa Indonesia diantaranya algoritma *confix-stripping*, algoritma Porter *stemmer* bahasa Indonesia, algoritma Arifin dan Sutiono, dan Algorima Idris (Tala 2003, Agusta 2009, Asian et al 2005, Adriani et al 2007). Dimana, Algoritma *confix-stripping* adalah algoritma yang akurat dalam *stemming* bahasa Indonesia ( Tala 2003, Agusta 2009, Asian et al 2005, Adriani et al 2007).

F.3 Sentiment Analysis

*Sentiment analysis* atau *opinion mining* mengacu pada bidang yang luas dari pengolahan bahasa alami, komputasi linguistik dan *text mining* yang bertujuan menganlisa pendapat, sentimen, evaluasi, sikap, penilaian dan emosi seseorang apakah pembicara atau penulis berkenaan dengan suatu topik , produk, layanan, organisasi, individu, ataupun kegiatan tertentu (Liu, 2011).

Tugas dasar dalam analisis sentimen adalah mengelompokkan teks yang ada dalam sebuah kalimat atau dokumen kemudia menentukan pendapat yang dikemukakan dalam kaliamat atau dokumen tersebut apakah bersifat positif, negatif atau netral. *Sentiment analysis* juga dapat menyatakan perasaan emosional sedih, gembira, atau marah.

Kita dapat mencari pendapat tentang produk-produk, merek atau orang-orang dan menentukan apakah mereka dilihat positif atau negatif di web. Hal ini memungkinkan kita untuk mencari informasi tentang:

a. Deteksi Flame (rants buruk)

b. Persepsi produk baru.

c. Persepsi Merek.

d. Manajemen reputasi.

Ekspresi atau *sentiment* mengacu pada fokus topik tertentu, pernyataan pada satu topik mungkin akan berbeda makna dengan pernyataan yang sama pada *subject* yang berbeda. Oleh karena itu pada beberapa penelitian, terutama pada review produk, pekerjaan didahului dengan menentukan elemen dari sebuah produk yang sedang dibicarakan sebelum memulai proses *opinion mining.*

F.4 Naïve Bayes Classifier

Algoritma *naive bayes classifier* merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat (Feldman & Sanger 2007). Dalam penelitian ini yang menjadi data uji adalah dokumen . Ada dua tahap pada klasifikasi dokumen. Tahap pertama adalah pelatihan terhadap dokumen yang sudah diketahui kategorinya. Sedangkan tahap kedua adalah proses klasifikasi dokumen yang belum diketahui kategorinya.

Dalam algoritma *naïve bayes classifier* setiap dokumen direpresentasikan dengan pasangan atribut “x1, x2, x3,...xn” dimana x1 adalah kata pertama, x2 adalah kata kedua dan seterusnya. Sedangkan V adalah himpunan kategori sentimen. Pada saat klasifikasi algoritma akan mencari probabilitas tertinggi dari semua kategori dokumen yang diujikan (VMAP), dimana persamaannya adalah sebagai berikut :

Untuk P(x1, x2, x3,...xn) nilainya konstan untuk semua kategori (Vj) sehingga persamaan dapat ditulis sebagai berikut :

Persamaan diatas dapat disederhanakan menjadi sebagai berikut :

Keterangan :

: Kategori sentimenj =1, 2, 3,…n. Dimana dalam penelitian ini =

kategori *tweet* sentimen negatif, = kategori *tweet* sentimen positif, dan = kategori *tweet* sentiment netral

: Probabilitas xi pada kategori Vj

: Probabilitas dari Vj

Untuk dan dihitung pada saat pelatihan dimana persamaannya adalah sebagai berikut :

Keterangan :

|docs j| : jumlah dokumen setiap kategori j

|contoh| : jumlah dokumen dari semua kategori

nk : jumlah frekuensi kemunculan setiap kata

n : jumlah frekuensi kemunculan kata dari setiap kategori

|kosakata| : jumlah semua kata dari semua kategori

* 1. **Metodologi Penelitian**

G.1 Jenis Penelitian

Jenis penelitian yang dilakukan merupakan penelitian kualitatif dan kuantitatif. Penggunaan metode kualitatif digunakan karena penelitian ini menganalisa sentimen masyarakat di media sosial terhadap kinerja pemerintahan Kabupaten Jember. Penggunaan metode kuantitatif dalam penulisan ini adalah penerapan serta pengkajian teori yang sudah ada sebelumnya.

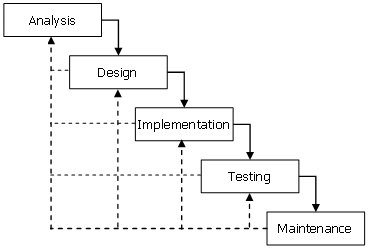
G.2 Tempat dan Waktu Penelitian

Penelitian ini dilakukan di media sosial dengan lingkup Jember. Penelitian dilakukan selama 5 bulan yaitu mulai dari bulan Januari 2017 sampai dengan Mei 2017.

G.3 Tahap Penelitian

Penelitian ini akan dilakukan dalam beberapa tahap yan disesuaikan dengan metode *Software Development Life Cycle* (SDLC) *waterfall* yang dibagi menjadi beberapa tahapan, yaitu analisis kebutuhan, desain sistem, implementasi, pengujian, dan pemeliharaan.

Pembangunan perangkat lunak pada penelitian ini adalah model *waterfall. Waterfall* Model Waterfall SDLC adalah proses pengembangan perangkat lunak sekuensial di mana perkembangan terlihat seperti mengalir semakin ke bawah (mirip dengan air terjun) melalui daftar tahapan yang harus dijalankan agar berhasil membangun sebuah perangkat lunak komputer. Awalnya, model Waterfall diusulkan oleh Winston W. Royce pada tahun 1970 untuk menggambarkan praktek rekayasa perangkat lunak yang memungkinkan. Model *waterfall* mendefinisikan beberapa fase yang berurutan yang harus diselesaikan satu demi satu dan pindah ke tahap berikutnya hanya ketika fase sebelumnya yang benar-benar selesai dilakukan (Bassil, 2012).



Gambar 5 Model Waterfall

G.3.1 Analisis Kebutuhan (Requirement)

Tahap pertama pada proses perancangan sistem adalah tahap analisis kebutuhan. Pada tahap ini dilakukan penelitian mencari permasalahan yang ada untuk dapat dianalisis kebutuhan yang diperlukan sebagai solusi dari pemasalahan yang muncul. Data-data yang telah didapat kemudian dikelompokkan menjadi kebutuhan fungsional dan non-fungsional. Teknik pengumpilan data yang dilakukan antara lain :

1. Studi pustaka

Teknik ini dilakukan dengan tujuan sebagai dasar pembaasan penyusunan dasar teori yang digunkan dalam penelitian. Sumber yang digunakan dalam studi pustaka berupa buku, jurnal, karya ilmiah, penelitian sebelumnya dan situs website.

1. Wawancara

Wawancara merupakan teknik mengumpulkan data dengan cara mengajukan pertanyaan-pertanyaan kepada narasumber untuk memperoleh data yang dibutuhkan guna menyelesaikan penelitian ini.

G.3.2 Desain Sistem

Tahap desain merupakan proses yang berfokus pada struktur arsitektur perangkat lunak, representasi *interface*, dan detail algoritma. Tahapan ini akan menghasilkan dokumen yang disebut software requirement langkah ini dilakukan oleh pengembang sistem. Perancangan sistem dengan konsep perancangan berorientasi objek, dengan menggunakan UML(*Unified Modeling Language*) yang digunakan antara lain:

1. *Business Process* digunakan untuk mendefinisikan aktifitas dan proses.

2. *Use Case Diagram* digunakan untuk mendefinisikan fungsional sistem.

3. *Scenario* digunakan untuk menjelaskan fitur sistem.

4. *Sequence Diagram* digunakan untuk menunjukkan rangkaian pesan yang sikirim antar object juga interaksi antar object.

5. *Class Diagram* digunakan untuk menggambar struktur statis class dalam sistem.

6. *Entity Realtionship Diagram* digunakan untuk menunjukkan relasi antar object.

G.3.3 Implementasi

Setelah perancangan sistem telah selesai dilakukan maka selanjutnya dilakukan tahap pembuatan sistem dan implementasi. Pembuatan sistem meliputi pembuatan desain interface, coding dan database. Menggunakan tools PHPStorm sebagai editor dengan bahasa PHP dan tools Xampp for windows untuk manajemen database.

G.3.4 Testing dan Evaluasi

Testing dan evaluasi digunakan untuk mengetahui sejauh mana sistem ini dapat berjalan. Testing berfungsi untuk mengetahui apakah sistem ini dapat berfungsi dengan baik sesuai dengan yang diharapkan. Serta untuk mengetahui letak kekurangan yang ada pada sistem. Pengujian dilakukan oleh tim penguji dari developer. Selanjutnya dilakukan evaluasi serta perbaikan terhadap kekurangan-kekurangan yang ada pada sistem ini. dilakukan dua metode untuk pengujian ini yakni :

1. White box testing

White box testing Merupakan cara pengujian dengan melihat modul yang telah dibuat dengan program-program yang ada. Pengujian ini, dilakukan oleh developer. Jika ada modul yang menghasilkan output yang tidak sesuai maka baris-baris program, variabel dan parameter yang terlibat pada unit tersebut satu persatu akan dicek dan diperbaiki, kemudian akan di compile ulang. Teknik pengujian ini menggunakan pengujian jalur dasar (*basis path testing*) dimana kompleksitas dari perangkat lunak yang dibangun akan dihitung menggunakan *Cyclomatic Complexity.*

1. Black box testing

Pengujian Black Box melibatkan pengguna/*User*, dimana hanya memperhatikan fungsionalitas yang berkaitan dengan masukan/keluaran (I/O) apakah sesuai dengan sistem yang dijalankan. Pengujian dilakukan oleh beberapa mahasiswa yang diambil secara acak, menggunakan kuesioner.

G.3.5 Pemeliharaan

Pemeliharaan diadakan untuk mengatasi masalah pada sistem dilain waktu ketika aplikasi sudah dapat digunakan oleh *user*. Selama *user* menemui bug pada sistem, maka *user* langsung konfirmasi kepada *developer* untuk segera ditangani.

* 1. **Luaran Yang Diharapkan**

Luaran yang diharapkan dari penelitian adalah

1. Skripsi
2. Jurnal yang dipublikasikan
3. Sistem Informasi Analisis Sentimen Otomatis Di Media Sosial Menggunakan Text Mining
   1. **Jadwal Kegiatan**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No | Jadwal Kegiatan | Des | Jan | Feb | Mar | Apr | Mei |
| 1 | Penyusunan dan Pengajuan Proposal |  |  |  |  |  |  |
| 2 | Seminar Proposal |  |  |  |  |  |  |
| 3 | Studi Literatur |  |  |  |  |  |  |
| 4 | Analisis Requirement |  |  |  |  |  |  |
| 5 | Perancangan Aplikasi |  |  |  |  |  |  |
| 6 | Implementasi |  |  |  |  |  |  |
| 7 | Pengujian |  |  |  |  |  |  |
| 8 | Penulisan Skripsi |  |  |  |  |  |  |
| 9 | Pemeriksaan Skripsi |  |  |  |  |  |  |
| 10 | Penyempurnaan Skripsi |  |  |  |  |  |  |
| 11 | Seminar Tugas Akhir |  |  |  |  |  |  |

Tabel 1 Jadwal Kegiatan

**DAFTAR PUSTAKA**

Asian, J., Williams, H.E., Tahaghoghi, S.M.M. 2005. Stemming Indonesia. Proceedings of the Twenty-eighth Australasian conference on Computer Science. Vol. 38, hal. Australia : Association for Computing Machinery.

Adriani, M., Asian, J., Nazief, B., Tahaghoghi, S.M.M., Williams, H.E. 2007. Stemming Indonesian : A Confix-Stripping Approach. Transaction on Asian Langeage Information Processing. Vol. 6, No. 4, Articel 13. Association for Computing Machinery : New York .

Berry, M.W. & Kogan, J. 2010. Text Mining Aplication and theory. WILEY : United Kingdom.

Dragut, E., Fang, F., Sistla, P., Yu, S. & Meng, W. 2009. Stop Word and Related Problems in Web Interface Integration. http://www.vldb.org/pvldb/2/vldb09-384.pdf. Diakses tanggal 13 Desember 2016.

Feldman, R & Sanger, J. 2007. The Text Mining Handbook : Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press : New York.

Han, J & Kamber, M. 2006 Data Mining: Concepts and Techniques Second Edition. Morgan Kaufmann publisher : San Francisco.

Liu, Bing. 2012. *Sentiment Analysis And Opinion Mining*. Chicago: Morgan & Claypool Publisher. http://www.dcc.ufrj.br/~valeriab/DTMSentiment-AnalysisAndOpinionMining-BingLiu.pdf. Diakses tanggal 10 Januari 2014.

Tala, Fadillah Z. 2003. A Study of Stemming Efects on Information Retrieval in Bahasa Indonesia. Institute for Logic, Language and ComputationUniversiteit van Amsterdam The Netherlands. http://www.illc.uva.nl/Research/Reports/MoL-2003-02.text.pdf. Diakses tanggal 29 September 2011.

Weiss, S.M., Indurkhya, N., Zhang, T., Damerau, F.J. 2005. *Text Mining : Predictive Methods fo Analyzing Unstructered Information*. Springer : New York