

## ANALISIS SENTIMEN PELANGGAN WIFI.ID PADA TWITTER DENGAN SUPPORT VECTOR MACHINE

Dwatra Apriandi <sup>1)</sup>

M. Irwan Afandi <sup>2)</sup>

Eka Dyar Wahyuni <sup>3)</sup>

E-mail : <sup>1)</sup>[dwatra23@gmail.com](mailto:dwatra23@gmail.com), <sup>2)</sup>[andy\\_afandi@yahoo.com](mailto:andy_afandi@yahoo.com), <sup>3)</sup>[eka.dyar@yahoo.com](mailto:eka.dyar@yahoo.com)

<sup>1)</sup>Sistem Informasi, Fakultas Teknologi Industri, UPN “Veteran” Jatim

<sup>2)</sup>Sistem Informasi, Fakultas Teknologi Industri, UPN “Veteran” Jatim

<sup>3)</sup>Sistem Informasi, Fakultas Teknologi Industri, UPN “Veteran” Jatim

### Abstrak

Penelitian ini bertujuan untuk membangun suatu aplikasi yang mengimplementasikan *support vector machine* untuk menganalisis sentimen pelanggan wifi.id pada *twitter*. Tools yang digunakan untuk mendukung penelitian ini adalah RStudio. Dalam penelitian ini data yang digunakan sepenuhnya berasal dari *twitter* yang di dalamnya terdapat berbagai sentimen terhadap wifi.id. Algoritma yang digunakan adalah *support vector machine* dengan metode pembobotan TF-IDF. Kelas sentimen terdiri dari sentimen positif, negatif dan netral. Dari hasil pengujian model didapatkan presentase keakurasian model sebesar 70% dengan 60% data latih dan 40% data uji. Selanjutnya hasil analisis data akan di visualisasikan dalam *wordcloud* yang dikemas pada sebuah aplikasi berbasis web. Dengan adanya aplikasi yang dihasilkan dari penelitian ini diharapkan dapat dijadikan sebagai media untuk mengukur kualitas produk wifi.id berdasarkan sentimen pelanggannya pada *twitter*. Sehingga perusahaan dapat dengan tepat meningkatkan aspek kualitas produknya sesuai dengan harapan pelanggan.

**Kata kunci :** Analisis Sentimen, Twitter, Support Vector Machine, Wordcloud.

### 1. PENDAHULUAN

Analisis sentimen atau disebut juga dengan *opinion mining* adalah salah satu bagian dari *text mining*. Bidang ini melakukan studi mengenai opini orang-orang, sentimen, evaluasi, tingkah laku dan emosi terhadap suatu entitas seperti produk, layanan, organisasi, individu, permasalahan, topik, acara dan atribut-atributnya. Manfaat analisis sentimen dalam dunia usaha antara lain untuk melakukan pemantauan terhadap sebuah produk. Selain itu, dapat digunakan sebagai alat bantu untuk melihat respon masyarakat terhadap produk tersebut. Sehingga perusahaan dapat segera mengambil langkah-langkah strategis berikutnya untuk meningkatkan kualitas produknya. Besarnya pengaruh dan manfaat dari analisis sentimen ini menyebabkan penelitian ataupun aplikasi mengenai analisis sentimen berkembang pesat akhir-akhir ini.

Salah satu algoritma dari *text mining* yang bisa digunakan untuk menyelesaikan masalah analisis sentimen adalah *Support Vector Machine*. *Support Vector Machine* menunjukkan akurasi paling tinggi pada kategorisasi teks berbahasa Indonesia dengan presentase 92.5% [18]. Dengan pertimbangan tersebut maka *Support Vector Machine* digunakan dalam penelitian ini. *Support Vector Machine* digunakan sebagai fungsi pemisah (klasifier) optimal yang mampu memisahkan dua set data dari dua kelas yang berbeda. Yang nantinya akan diklasifikasikan sebagai sentimen positif atau sentimen negatif.

Pengguna internet di Indonesia pada tahun 2014 sudah mencapai 83 juta orang. Sebanyak 90% dari pengguna internet tersebut atau 66 juta orang menggunakan media sosial. Tahun lalu terdapat penambahan 10,5 juta orang pengguna media sosial. Angka ini berarti setiap menit, terdapat penambahan pengguna media sosial sebanyak 20 orang [14]. Oleh karena itu, media sosial merupakan salah satu sumber informasi yang digunakan sebagai sarana mengumpulkan berbagai macam opini masyarakat yang nantinya akan

dianalisis sentimennya. Penelitian ini memergunakan media sosial *twitter* sebagai sumber datanya, karena Indonesia adalah negara kelima dengan pengguna *twitter* terbanyak di dunia, yakni sebanyak 19,5 juta orang [13]. Sehingga media sosial *twitter* merupakan sebuah indikator yang baik untuk memberikan pengaruh dalam penelitian ini.

Melalui *twitter*, pengguna dapat mengirim dan membaca pesan, berbagi informasi, menjalin relasi bisnis, menuangkan isi hati dan pikiran dalam bentuk tulisan (sering disebut *tweet*), dengan kapasitas kata yang bisa diunggah dan ditampilkan pada *timeline* pengguna *twitter* mencapai 140 karakter. Sama halnya dengan situs jejaring sosial lain dalam *twitter* disediakan suatu mesin pencarian (*search engine*) yang berguna untuk mempermudah pengguna dalam menemukan informasi menggunakan kata kunci tertentu. Melalui *search engine* pengguna dapat menemukan lebih banyak informasi yang dibutuhkan terkait topik yang ingin dicari [3].

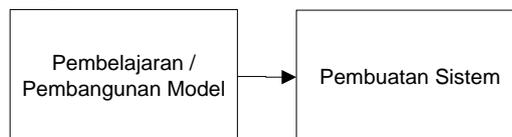
Penelitian ini bertujuan untuk membuat sebuah aplikasi yang nantinya akan mampu mengimplementasikan *support vector machine* untuk menganalisis sentimen pelanggan terkait status-statusnya di *twitter* terhadap salah satu produk telkom dalam hal ini *wifi.id*. Kemudian hasil analisis sentimen tersebut akan divisualisasikan dengan tujuan agar lebih interaktif dan mudah dipahami oleh manajemen tingkat atas, khususnya manager bagian pemasaran *wifi.id*.

Adapun hasil penelitian ini diharapkan dapat mempermudah, membantu serta memberikan pertimbangan serta gambaran kepada pihak manajemen tingkat atas telkom agar dapat meningkatkan kualitas produknya. Karena bagaimanapun juga hanya perusahaan yang berkualitas yang dapat bersaing dan menguasai pasar [4].

Berdasarkan kajian literatur yang ada, terdapat banyak macam *tools* yang dapat digunakan untuk proses analisis sentimen dengan *support vector machine*. *Tools-tools* yang sering digunakan diantaranya adalah WEKA, SPSS, Matlab, dan RStudio. Dalam penelitian ini diputuskan untuk menggunakan *tools* RStudio, karena kefleksibelannya mulai dari tahap pengumpulan data dari *twitter* hingga tahap mengklasifikasikan data yang ada. Karena di dalam *tools* tersebut telah disediakan berbagai macam *package library* yang siap digunakan sebagai sarana penunjang yang dibutuhkan dalam penelitian ini.

## 2. METODE PENELITIAN

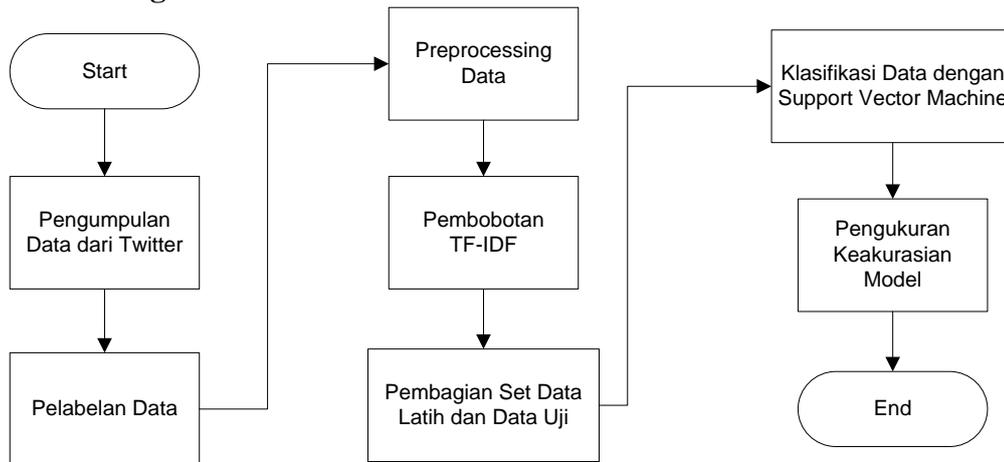
Untuk dapat menghasilkan sistem yang dimaksud, terdapat 2 proses besar yang harus dilakukan dalam penelitian ini. Proses-proses tersebut dapat dilihat pada gambar 1, diantaranya adalah pembelajaran atau pembangunan model klasifikasi, pembuatan sistem dan memvisualisasikan hasil.



Gambar 1 Analisis Sistem

Proses pembelajaran atau pembangunan model akan dijabarkan dalam subbab 2.1, sedangkan untuk proses pembuatan sistem dan memvisualisasikan hasil akan dijelaskan pada subbab 2.2.

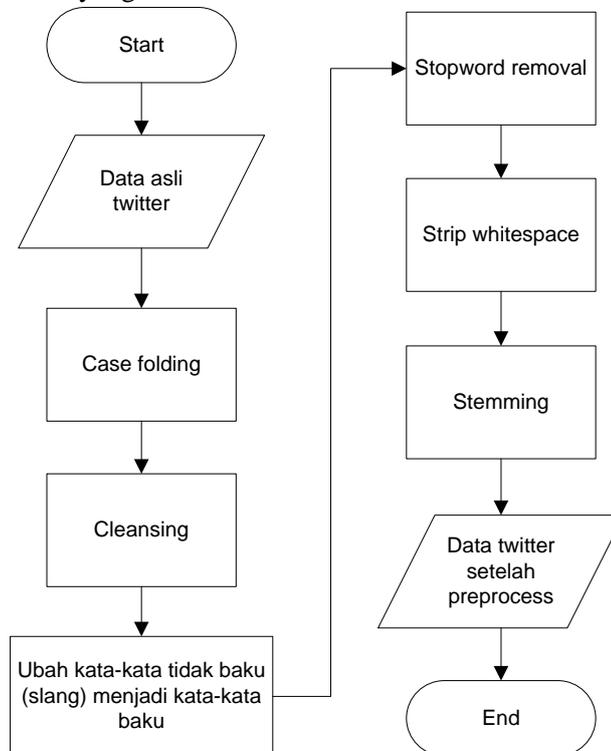
## 2.1 Pembangunan model



Gambar 2. Alur Pembangunan Model

Proses pertama yang akan dilakukan dalam pembangunan model adalah mengumpulkan data yang berkaitan dengan wifi.id pada *twitter*, lalu seluruh data yang telah terkumpul dilabeli secara manual berdasarkan kelas sentimennya. Dengan demikian maka setiap data akan dilabeli berdasarkan kelas sentimen positif, sentimen negatif, dan sentimen netral. Setelah melakukan penyaringan dan pemilahan terhadap data yang telah terkumpul pada proses sebelumnya, maka total keseluruhan data *tweet* yang diberi label berjumlah 1.155 buah data.

Tahap selanjutnya adalah *preprocessing* pada data. Tahapan *preprocessing* data bertujuan untuk menghilangkan *noise*, menyeragamkan bentuk kata dan mengurangi volume kata-kata yang tidak diperlukan, sehingga model klasifikasi dapat dengan baik mengklasifikasikan data yang akan diolah.



Gambar 3. Alur Proses pada tahap Preprocessing

Tahapan *preprocessing* dilakukan dalam *tools* RStudio dengan dukungan *package library text mining* (tm) yang terpasang di dalamnya. Gambar 3 menunjukkan alur proses pada tahapan preprocessing. Dari masukan data twitter dilakukan *case folding*. *Case folding* adalah proses penyeragaman *case* atau huruf yang terdapat dalam sebuah dokumen. Tahap berikutnya adalah *cleansing* yang bertujuan untuk membersihkan dokumen dari karakter-karakter yang tidak diperlukan dengan tujuan untuk mengurangi *noise*. Dalam penelitian ini proses *cleansing* terdiri atas beberapa tahapan, yaitu menghilangkan URL situs lain, menghilangkan URL gambar, menghilangkan nama akun, menghilangkan *hashtag*, menghilangkan karakter angka, menghilangkan tanda baca, menghilangkan huruf yang berulang, dan menghilangkan huruf tunggal.

Proses selanjutnya adalah mengubah kata tidak baku menjadi baku. Seperti kebanyakan sosial media lain, *twitter* memberikan keleluasaan kepada penggunanya untuk menulis apapun asalkan panjang karakternya tidak melebihi 140. Karena keterbatasan inilah, para pengguna *twitter* banyak yang mempergunakan singkatan informal dalam tweetnya. Untuk merubah kata tidak baku menjadi baku, yang dilakukan pertama kali adalah memasukkan kamus kata-kata tidak baku (*slang word*) ke dalam sistem. Kamus tersebut diambil dari <https://abdiansah.wordpress.com> yang mengalami perubahan isi, disesuaikan dengan kebutuhan data yang ada dalam penelitian ini.

Setelah kata tidak baku berubah menjadi kata baku, proses selanjutnya adalah *Stopword Removal*. *Stopword Removal* adalah proses untuk menghilangkan kata yang sering muncul namun tidak memiliki makna spesifik atau perannya tidak diperlukan dalam penelitian ini, contoh dari *stopword* adalah kata yang, lah, juga, dll. Setelah melalui berbagai tahapan *preprocess* sebelumnya seperti *cleansing* dan *stopword removal*, akan banyak ditemui karakter spasi berlebih pada data. *Stripping whitespace* adalah proses untuk menghilangkan spasi berlebih yang terdapat dalam keseluruhan dokumen.

Proses selanjutnya adalah *stemming*, yaitu proses untuk menemukan kata dasar dengan menghilangkan semua imbuhan (*affixes*) baik itu berupa awalan (*prefixes*), akhiran (*suffixes*), maupun kombinasi dari awalan dan akhiran (*confixes*) yang ada pada setiap kata dalam data. 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 menimbulkan *noise*. Perbedaan bahasa yang digunakan akan membuat perbedaan pula pada teknik *stemming* yang diimplementasikan.

Setelah tahap preprocessing selesai dilakukan, diperoleh *term* yang siap untuk diproses ke tahap pembobotan. Pembobotan data dilakukan untuk mendapatkan nilai data dari setiap kata atau *term* yang sudah melalui beberapa proses sebelumnya. Perhitungan bobot yang digunakan adalah *Term Frequency-Inverse Document Frequency* (TF-IDF), yang menggambarkan seberapa pentingnya sebuah *term* dalam sebuah data. Proses ini digunakan untuk menilai bobot relevansi *term* dari sebuah data terhadap keseluruhan dokumen yang ada.

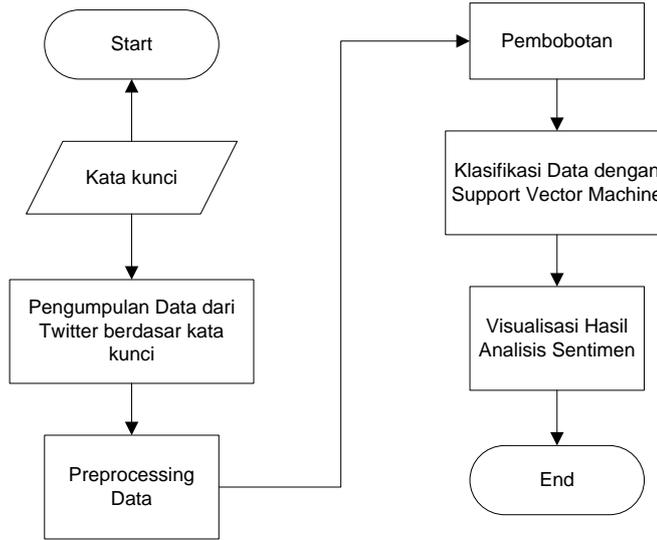
Untuk membangun sebuah model klasifikasi diperlukan pembagian set data menjadi dua macam, yakni data latih (*training*) dan data uji (*testing*). Setelah data dibagi menjadi dua, dapat dilakukan proses klasifikasi. Dalam penelitian ini, proses klasifikasi dilakukan dengan mempergunakan algoritma SVM. Data *tweet* yang ada nantinya akan diklasifikasikan ke dalam 3 kategori kelas yang berbeda yaitu kelas sentimen positif, negatif dan netral.

Untuk mengevaluasi performansi sebuah model yang dibangun dapat dilakukan dengan menghitung perbandingan jumlah dari data yang diprediksi secara benar atau salah oleh model tersebut dengan nilai dari atribut label kelas sentimen pada data uji. Perbandingan data yang diprediksi benar dengan keseluruhan data uji disebut akurasi. Sedangkan perbandingan data yang diprediksi salah dengan keseluruhan data uji disebut *error rate*.

Jika tingkat akurasi yang diperoleh dirasa cukup baik, maka model tersebut dapat digunakan dengan baik untuk mengklasifikasikan *record-record* data baru yang belum pernah dilatihkan atau diuji sebelumnya. Dengan kata lain model yang dibangun telah siap diimplementasikan ke dalam sistem untuk mengklasifikasikan *tweet* berdasarkan kelas sentimennya.

**2.2 Pembuatan Sistem**

Setelah didapatkan kombinasi antara data latih dan data uji yang memiliki keakurasian yang paling bagus, kombinasi ini menjadi model yang akan diimplementasikan ke dalam suatu aplikasi web untuk mengklasifikasikan *tweet* berdasarkan sentimennya. Gambar 4 menjelaskan tentang alur dari aplikasi web yang dikembangkan.



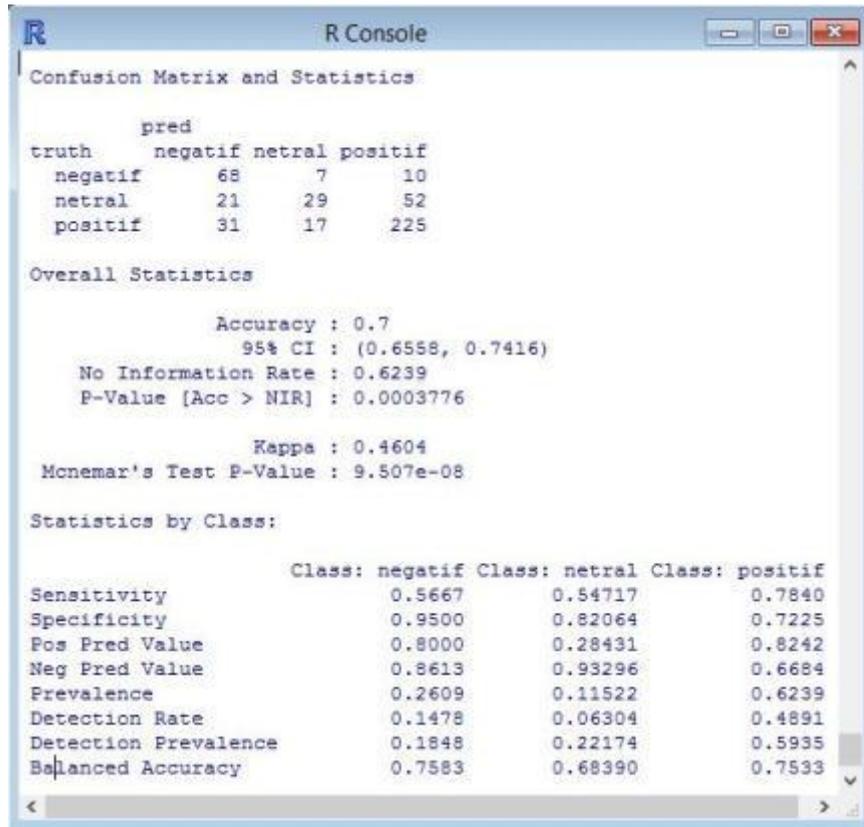
Gambar 4. Alur Sistem

Proses pertama dari aplikasi web yang dikembangkan adalah, user memasukan kata kunci kedalam suatu form. Form ini nantinya akan memanggil R script yang didalamnya terdapat beberapa proses untuk mengumpulkan data dari *twitter* yang berkaitan dengan kata kunci. Setelah data terkumpul, tahap berikutnya adalah *preprocessing* keseluruhan data. *Preprocessing* yang dilakukan pada tahap ini, langkah-langkahnya serupa dengan langkah *preprocessing* yang ada pada saat pembangunan model. Setelah itu dilanjutkan dengan proses pembobotan TF-IDF (*term frequency-inverse document frequency*) dan dilanjutkan dengan proses klasifikasi data dengan *support vector machine*. Kemudian proses terakhir adalah memvisualisasikan hasil dari klasifikasi tersebut mempergunakan *wordcloud*. Keseluruhan proses ini dikembangkan dengan mempergunakan PHP, HTML, R script, CSS, dan Javascript.

**3. HASIL DAN PEMBAHASAN**

**3.1 Uji coba model**

Pengujian model dilakukan dengan menggunakan data *tweet* yang telah dilabeli sebanyak 1.155 buah data. Kemudian data tersebut akan dibagi ke dalam data latih (*training*) sebanyak 60% dan data uji (*testing*) sebanyak 40% dari jumlah keseluruhan data. Pengujian model dilakukan dengan *tools* RStudio yang telah dilengkapi dengan beberapa *package library* seperti *text mining*, *RtextTools*, dan *caret*. Hasil dari uji coba model dapat dilihat pada gambar 5.

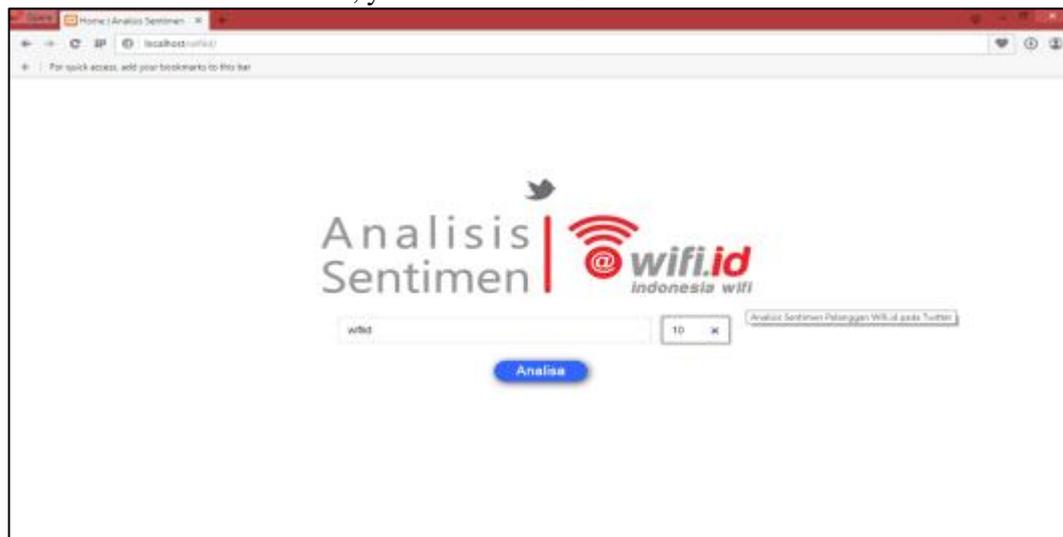


Gambar 5 Hasil Pengujian Model

Hasil dari pengujian model menunjukkan bahwa nilai akurasi dari model yang telah dibangun adalah 0,7 atau dalam presentase 70%. Dengan demikian maka performansi dari model klasifikasi yang telah dibangun dinilai cukup baik. Sehingga model yang dibangun telah siap diimplementasikan ke dalam sistem untuk mengklasifikasikan *tweet* berdasarkan kelas sentimennya.

### 3.2 Implementasi Sistem

Sistem yang dikembangkan dalam penelitian ini mempergunakan web dengan dua buah halaman antarmuka, yaitu halaman masukan dan halaman luaran



Gambar 6. Tampilan Halaman Masukan



Dengan hasil analisis tersebut maka, pihak wifi.id dapat dengan mudah mengetahui aspek kelebihan dan kekurangan produknya dari sudut pandang konsumennya secara tepat. Sehingga dalam contoh studi kasus di atas, dapat segera dilakukan penanganan terhadap aspek kualitas koneksi wifi.id. Hal ini tentu diharapkan dapat meningkatkan kualitas wifi.id di masyarakat dan meningkatkan jumlah pelanggan wifi.id.

#### 4. SIMPULAN DAN SARAN

Penelitian ini berhasil membangun suatu aplikasi yang mengimplementasikan *support vector machine* untuk analisis sentimen pelanggan wifi.id pada *twitter* dengan mengintegrasikan *tools* RStudio dan php. Dari hasil uji coba, keakurasian *support vector machine* dalam menganalisis sentimen pelanggan wifi.id pada *twitter* cukup baik dengan presentase 70%, menggunakan 60% data latih dan 40% data uji yang diambil dari *twitter* dalam periode bulan Januari hingga Oktober 2015. Untuk memvisualisasikan hasil analisis sentimen pelanggan wifi.id pada *twitter* dapat dilakukan dengan *word cloud* yang dikemas ke dalam aplikasi berbasis web. Sedangkan simpulan klasifikasi analisis sentimen positif, negatif dan netral ditampilkan dalam format teks berwarna.

Untuk penelitian berikutnya diharapkan sistem ini tidak hanya dapat digunakan untuk menganalisis sentimen pelanggan wifi.id pada *twitter* tetapi juga dapat digunakan untuk menganalisis sentimen produk lain ataupun seseorang tokoh tertentu. Diharapkan dapat dilakukan penelitian dengan mengimplementasikan model klasifikasi lain dan seleksi fitur yang lebih baik. Agar dapat meningkatkan keakurasian dari sebuah model klasifikasi pada penelitian selanjutnya dapat dilakukan dengan cara menambah data latih yang telah dipilah lebih banyak lagi. Diharapkan peneliti selanjutnya dapat memperkaya konten dari kamus kata tidak baku (*slang word*). Tidak hanya dapat menganalisis sentimen data dalam bahasa Indonesia saja, tetapi dapat juga digunakan untuk data dalam bahasa daerah atau bahasa asing.

#### 5. DAFTAR PUSTAKA

- [1] Anonymous. 2015. *About Us*. Available at: [www.wifi.id](http://www.wifi.id) [Accessed 17 September 2015].
- [2] Anonymous. 2012. *Telkom Indonesia*. Available at: [https://id.wikipedia.org/wiki/Telkom\\_Indonesia](https://id.wikipedia.org/wiki/Telkom_Indonesia) [Accessed 17 September 2015].
- [3] Adiyana, Imam dan Fajriya Hakim. 2009. Implementasi Text Mining pada Mesin Pencarian Twitter Untuk Menganalisis Topik – Topik Terkait “KPK dan Jokowi”. *Prosiding Seminar Nasional Matematika dan Pendidikan Matematika UMS 2015*.
- [4] Atmawati, R. dan Wahyudin. 2007. Analisis Pengaruh Kualitas Pelayanan Terhadap Kepuasan Konsumen Pada Matahari Departement Store Di Solo Grand Mall. *Jurnal Daya Saing, Program MM-UMS*. Surakarta.
- [5] Dehaff, M. 2010. *Sentiment Analysis, Hard But Worth It!*. Available at: [http://www.customerthink.com/blog/sentiment\\_analysis\\_hard\\_but\\_worth\\_it](http://www.customerthink.com/blog/sentiment_analysis_hard_but_worth_it) [Accessed 23 September 2015].
- [6] Liu, Bing. 2012. *Sentiment Analysis And Opinion Mining*. Chicago: Morgan & Claypool Publisher.
- [7] Manalu, B.U. 2014. *Analisis Sentimen Pada Twitter Menggunakan Text Mining*. Skripsi. Tidak dipublikasikan. Medan: Universitas Sumatera Utara.
- [8] Monarizqa, N., Nugroho, L.E. dan Hantono, B.S. 2014. Penerapan Analisis Sentimen Pada Twitter Berbahasa Indonesia Sebagai Pemberi Rating. *Jurnal Penelitian Teknik Elektro dan Teknologi Informasi*. 1 (3), 151-155.
- [9] Nomleni, Petrix. 2015. *Sentiment Analysis Menggunakan Support Vector Machine (SVM)*. Tesis. Tidak dipublikasikan. Surabaya: Institut Teknologi Sepuluh Nopember.

- [10] Nugroho, A.T. 2012. *Pemrograman game berbasis web menggunakan javacript dan HTML 5*. Yogyakarta : Andi Offset.
- [11] Prasetyo, Eko. 2012. *Data mining-konsep dan aplikasi menggunakan matlab*. Yogyakarta : Andi Offset.
- [12] Prasetyo, Eko. 2014. *Data mining-mengolah data menjadi informasi menggunakan matlab*. Yogyakarta : Andi Offset.
- [13] Semiocast. 2012. *Top 20 Cities By Number Of Tweets*. Available at: [http://semiocast.com/publications/2012\\_07\\_30\\_Twitter\\_reaches\\_half\\_a\\_billion\\_accounts\\_140m\\_in\\_the\\_US](http://semiocast.com/publications/2012_07_30_Twitter_reaches_half_a_billion_accounts_140m_in_the_US). [Accessed 21 Sep 2015].
- [14] Statista. 2015. *Countries with the highest number of internet users in 2014 (in millions)* Available at: <http://www.statista.com/statistics/262966/number-of-internet-users-in-selected-countries/>. [Accessed 21 Sep 2015].
- [15] Tjiptono, Fandi. 2001. *Kualitas Jasa: Pengukuran, Keterbatasan dan Implikasi Manajerial*, majalah Manajemen Usahawan Indonesia. Jakarta
- [16] Tjiptono, F. dan Chandra,G. 2007. *Service, Quality Satisfaction*. Yogyakarta : Andi Offset.
- [17] Wahana. 2009. *ShortCourse: PHP Programming*. Yogyakarta : Andi Offset.
- [18] Wulandini, F. & Nugroho, A. N. 2009. Text Classification Using Support Vector Machine for Webmining Based Spation Temporal Analysis of the Spread of Tropical Diseases. *International Conference on Rural Information and Communication Technology 2009*.