

ANALISIS SENTIMEN PENGGUNA TWITTER MENGUNAKAN METODE SUPPORT VECTOR MACHINE BERBASIS CLOUD COMPUTING

Rizky Maulana, Sri Redjeki

Jurusan Teknik Informatika STMIK AKAKOM

Jl. Raya Janti No 143 Yogyakarta

Telp. (0274)486664 website: www.akakom.ac.id

rizmaulana@live.com, dzecky@akakom.ac.id

ABSTRAK

Twitter merupakan jejaring sosial dengan pertumbuhan tercepat sejak tahun 2006 menurut MIT Technology Review (2013), Indonesia menempati Negara ketiga penyumbang *tweet* terbanyak dengan jumlah 1 milyar *tweet*. Fakta tersebut menjadikan Twitter menjadi salah satu sumber data text yang dapat digali dan dimanfaatkan untuk berbagai keperluan melalui metode-metode pengambilan data teks atau text mining, salah satunya adalah analisis sentimen pengguna terhadap tokoh-tokoh publik Indonesia. Penelitian ini membuat sebuah sistem yang dapat melakukan analisis sentimen pengguna twitter terhadap tokoh publik secara *real time* dengan menggunakan Twitter Streaming API dan metode *Support Vector Machine* (SVM) memanfaatkan pustaka libSVM sebagai salah satu *machine learning* untuk *text classification*. Algoritma Porter digunakan dalam proses *stemming* untuk ekstraksi fitur dan metode *Term Frequency* untuk pembobotan. Perangkat lunak dibangun dengan menggunakan bahasa pemrograman PHP untuk sisi *server* yang berjalan pada platform *cloud* Windows Azure dan Java untuk sisi *client* yang berjalan pada platform Android. Dari hasil penelitian dengan 1.400 *tweet* pada dataset dan 200 data uji didapatkan akurasi sebesar 79,5%.

Kata Kunci : Analisis Sentimen, Cloud Computing, Real Time, SVM, Tokoh Publik.

1. PENDAHULUAN

1.1 Latar Belakang Masalah

Indonesia menempati Negara ketiga penyumbang *tweet* terbanyak dengan jumlah 1 milyar *tweet*, di bawah Amerika Serikat (3,7 milyar) dan Jepang (1,8 milyar). Bahkan, Jakarta menjadi *Twitter Capital City*, yakni kota dengan jumlah *tweet* terbanyak dan teraktif di dunia. Fakta tersebut menjadikan Twitter menjadi salah satu sumber data text yang dapat digali dan dimanfaatkan untuk berbagai keperluan melalui metode-metode pengambilan data teks atau *text mining*.

Text mining merupakan bagian dari data mining dimana proses yang dilakukan utamanya adalah melakukan ekstraksi pengetahuan dan informasi dari pola-pola yang terdapat dalam sekumpulan dokumen teks menggunakan alat analisis tertentu (R. Feldman, 2006). *Text mining* dapat diolah untuk berbagai macam keperluan diantaranya adalah untuk *summarization*, pencarian dokumen teks dan *sentiment analysis*. *Sentiment analysis* merupakan bidang yang 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 (B. Lui, 2012).

Analisis sentimen sangatlah berguna untuk menganalisis komentar-komentar di Twitter untuk kemudian diterjemahkan menjadi sesuatu yang lebih bermakna, salah satunya dalam bentuk statistik sederhana mengenai presentasi sentimen positif dan negatif terhadap tokoh publik di Indonesia, seperti pejabat, pemuka agama, pelaku industri hiburan dan sebagainya. Berdasarkan referensi [Nurvirta

Monarizqa pada tahun 2014 dengan judul] Penerapan Analisis Sentimen Pada Twitter Berbahasa Indonesia Sebagai Pemberi Rating yang dikembangkan berbasis web dengan menggunakan metode *support vector machine* mempunyai akurasi sebesar 73.43%. Berdasarkan referensi [Ahmad Fathan Hidayatullah pada tahun 2014] diperoleh analisis Sentimen dan Klasifikasi Kategori Terhadap Tokoh Publik Pada Twitter menggunakan metode *Support Vector Machine* dengan bantuan software RapidMiner dengan akurasi sebesar 79.68%.

1.2 Rumusan Masalah

Rumusan masalah yang akan dibahas pada penelitian kali ini adalah :

- Bagaimana melakukan pengambilan data di twitter dengan menggunakan API yang sudah disediakan oleh twitter.
- Bagaimana memproses kicauan di twitter untuk mengetahui sentimen pengguna twitter terhadap tokoh publik di Indonesia dengan menggunakan pustaka SVM pada *cloud*.
- Bagaimana mengirimkan, menerima dan memvisualisasikan hasil pemrosesan data dari *cloud* kedalam bentuk *chart* atau grafik pada perangkat *mobile*.

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah :

- Pengambilan data aplikasi melalui jejaring sosial Twitter dengan menggunakan Twitter Streaming API dengan menggunakan pustaka TwitterAPIExchange.

- Sumber data yang digunakan dalam penelitian ini adalah kalimat opini yang akan dianalisis melalui tweet berkaitan dengan sentimen pengguna twitter terhadap tokoh-tokoh publik.
- Metode pemrosesan kalimat dengan metode *Support Vector Machine* menggunakan pustaka LibSVM.
- Pemrosesan kalimat dilakukan pada *cloud* dengan memanfaatkan *Virtual Machine* Windows Azure.
- Penerimaan data dari *cloud* ke perangkat *mobile* untuk dilakukan visualisasi data dengan menggunakan format pertukaran data JSON (*Javascript Object Notation*).
- Pembuatan grafik berupa *pie chart* dilakukan dengan menggunakan pustaka HelloChart untuk Android.

1.4 Tujuan Penelitian

Penelitian ini diharapkan dapat memberikan gambaran mengenai bagaimana sentimen pengguna jejaring sosial twitter terhadap tokoh-tokoh publik untuk berbagai kepentingan.

1.5 Manfaat Penelitian

Hasil penelitian ini diharapkan dapat memberikan manfaat antara lain :

- Pengoptimalan informasi jejaring sosial untuk kepentingan publik
- Pengembangan metode Support Vector Machine
- Pengembangan teknologi cloud computing

2. TINJAUAN PUSTAKA

2.1 Penelitian Sebelumnya

Beberapa pustaka yang dijadikan tinjauan adalah Nurvirta Monarizqa yang membahas penerapan Analisis Sentimen Pada Twitter Berbahasa Indonesia Sebagai Pemberi Rating. Penelitian ini menggunakan metode *support vector machine* dengan bantuan *software SVMLight*. Data dari penelitian ini adalah *tweet* dari pengguna twitter dengan beberapa kata kunci tempat dan tokoh sehingga dapat diketahui seberapa populer tempat atau tokoh tersebut di kalangan pengguna twitter dalam bentuk persentase dengan akurasi sebesar 73.43%. Tinjauan pustaka yang lain Ahmad Fathan Hidayatullah yang membahas Analisis Sentimen Dan Klasifikasi Kategori Terhadap Tokoh Publik Pada Twitter. Pengambilan data menggunakan *cron job* dan menggunakan metode *Support Vector Machine* dengan bantuan *software RapidMiner*. Hasil dari penelitian ini menghasilkan akurasi sebesar 79.68%.

2.2 Sentiment Analysis

Sentiment analysis atau *opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Analisis sentimen dilakukan untuk

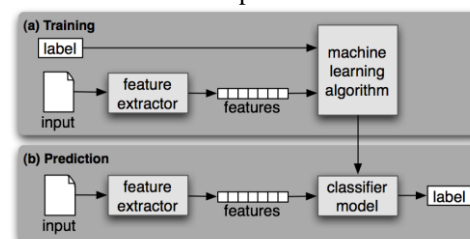
melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung berpandangan atau beropini negatif atau positif (B. Liu. 2010).

2.3 Twitter Streaming API

API (*Application Programming Interface*) yang digunakan adalah Twitter Streaming API, Twitter Streaming API biasa digunakan untuk penggalian data pada Twitter, karena melalui API ini informasi bisa didapatkan secara realtime dengan dengan berbagai macam *query* yang dibutuhkan. Pada penelitian ini sendiri, Twitter Streaming API akan digunakan untuk mengumpulkan data penelitian. Twitter Streaming API dapat digunakan untuk melakukan pengambilan data secara realtime dengan kata kunci tertentu dengan menggunakan API GET search/tweets yang dapat memberikan data realtime sebanyak 100 tweet terakhir dari kata kunci yang ditentukan.

2.4 Support Vector Machine

Support Vector Machine (SVM) merupakan salah satu metode klasifikasi dengan menggunakan *machine learning* (*supervised learning*) yang memprediksi kelas berdasarkan model atau pola dari hasil proses training



Gambar 1. Supervised Machine Learning

Sumber : <http://ravikiranj.net/posts/2012/code/how-build-twitter-sentiment-analyzer/>

Klasifikasi dilakukan dengan mencari *hyperplane* atau garis pembatas (*decision boundary*) yang memisahkan antara suatu kelas dengan kelas lain yang dalam kasus ini garis tersebut berperan memisahkan *tweet* bersentimen positif (berlabel +1) dengan *tweet* bersentimen negatif (berlabel -1). SVM melakukan pencarian nilai *hyperlane* yang paling maksimal dengan menggunakan *support vector* dan nilai *margin* (J. Han, 2006).

2.4 Algoritma Porter

Algoritma Porter merupakan algoritma yang akan digunakan untuk melakukan *stemming* pada *tweet*. *Stemming* merupakan sebuah proses ekstraksi dari sebuah kata yang digunakan untuk mendapatkan kata dasar dari suatu kata, proses *stemming* dilakukan setelah proses *preprocessing* pada data mentah yang diambil dari twitter.

Proses *stemming* dilakukan dengan menerapkan algoritma *Porter Stemmer for Bahasa Indonesia* pada bahasa pemrograman PHP. Adapun algoritma

dari *Porter Stemmer for Bahasa Indonesia* dikembangkan oleh Fadillah Z. Tala pada tahun 2003. Implementasi *Porter Stemmer for Bahasa Indonesia* berdasarkan *English Porter Stemmer* yang dikembangkan oleh W.B. Frakes pada tahun 1992. Karena Bahasa Inggris datang dari kelas yang berbeda, beberapa modifikasi telah dilakukan untuk membuat Algoritma Porter dapat digunakan sesuai dengan bahasa Indonesia (Gregorius S. Budhi dkk, 2012).

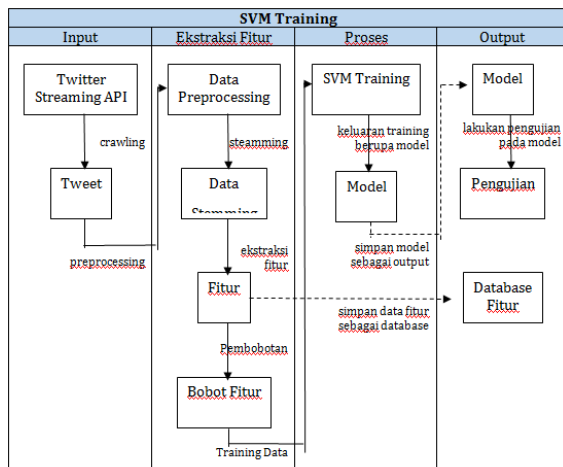
3. METODE PENELITIAN

3.1 Kebutuhan Data

Kebutuhan data yang digunakan pada penelitian ini berupa dataset *tweet* dari pengguna twitter berbahasa Indonesia mengenai sentimen terhadap tokoh-tokoh publik. Seluruh data *tweet* tersebut akan disimpan dalam sebuah dataset dan diberikan label secara manual untuk memberikan atribut apakah suatu data *tweet* dianggap positif atau negatif. Jumlah dataset yang akan di *crawling* dari twitter yang berkenaan dengan subjek penelitian adalah sebanyak 2.000 data *tweet*. 80 persen dari data tersebut akan digunakan sebagai data latih dan 20 persen akan digunakan sebagai data uji untuk mengetahui seberapa besar akurasi dari proses *training* dalam memprediksi suatu data.

3.2 Model Perancangan

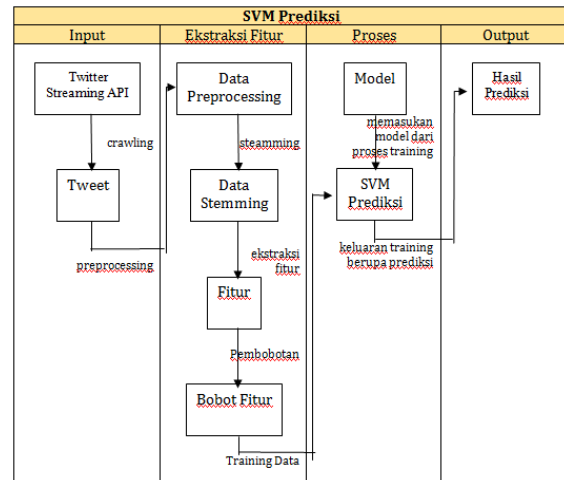
Secara keseluruhan sistem dibagi menjadi 2 yaitu sistem *training* dan sistem *testing* yang terlihat pada gambar 2 dan gambar 3.



Gambar 2. Blok Diagram Sistem Training

Pada proses *training* data dikumpulkan dengan Twitter Streaming API, kemudian dari data tersebut dilakukan *preprocessing* dan *stemming* sebagai bagian dari ekstraksi fitur, setelah dilakukan *preprocessing* dan *stemming* maka didapatkan data fitur yang kemudian disimpan kedalam database. Setelah mendapatkan fitur maka dilakukan pembobotan pada masing-masing fitur berdasarkan indeks, apabila sudah dilakukan pembobotan maka

dilakukan *training* oleh SVM dengan LibSVM. Validasi dilakukan secara manual, untuk menghitung keakuratan data



Gambar 3. Blok Diagram Sistem Klasifikasi

3.3 Analisis Sistem

Tahapan analisa sistem *training* dan *testing*, yaitu :

a. Proses Training

Proses *training* dilakukan untuk memberikan pembelajaran kepada *machine learning* agar dapat melakukan prediksi terhadap data yang diujikan. Terdapat dua buah keluaran dari proses *training*, yaitu fitur dan model. Fitur didapatkan melalui proses ekstraksi fitur yaitu dengan melakukan *preprocessing* dan *stemming* pada *tweet*. Pada proses *training* data diberi label secara manual, data diberi label dengan -1 sebagai data negatif dan 1 sebagai data positif dan dirubah dalam format data yang sesuai dengan format data libSVM., yaitu :

<label> <index fitur ke-1>:<bobot> <index fitur ke-2>:<bobot> <index fitur ke-n>:<bobot>.

Contoh :

Data Positif : 1 92:1 102:1 107:2 117:1

Data Negatif : -1 109:1 121:2 158:1

Pembobotan dilakukan dengan menggunakan TF (*term frequency*), yaitu dengan menghitung seberapa banyak suatu fitur muncul dalam suatu *tweet*.

b. Proses Prediksi

Proses prediksi dilakukan untuk melakukan prediksi terhadap data masukan berdasarkan model dan fitur yang didapatkan melalui proses *training*. Pada proses ini dilakukan *crawling* data secara *real time* dari twitter. Data *tweet* tersebut akan di *crawling* oleh aplikasi yang berada pada sisi server, untuk kemudian dianalisis sentimen dari data *tweet* tersebut.

Tahapan setelah data *real time* tersebut didapatkan adalah *preprocessing*. *Preprocessing* adalah tahapan untuk menghilangkan atribut yang dianggap tidak

perlu, seperti *hashtag*, *username*, *link* sampai dengan penghilangan *stop words*.

Setelah dilakukan *preprocessing* maka selanjutnya dilakukan *stemming* untuk mengetahui akar kata dari sebuah kata didalam *tweet*. Setelah proses *stemming* inilah data *tweet* yang sudah bersih akan dicocokkan dengan *database* fitur yang didapat melalui proses *training* sebelumnya untuk kemudian diberikan bobot pada masing-masing fitur berdasarkan indeks fiturnya.

Contoh : 279:1 281:1 390:2

Maka fitur dengan indeks 279 dan 281 mempunyai bobot 1, fitur dengan indeks 390 mempunyai bobot 2, sedangkan fitur yang tidak ada didalam data tersebut dianggap mempunyai bobot 0. Dari data tersebut berdasarkan model yang sudah didapat program dapat melakukan klasifikasi apakah data yang diujikan termasuk dalam kelas positif atau negatif. Sedangkan sentimen netral didapat apabila dalam proses ekstraksi fitur tidak didapatkan fitur. Kemudian data berupa persentase sentimen positif, negatif, netral dan daftar *tweet* yang di *crawling* akan dikirimkan ke perangkat *mobile* untuk kemudian ditampilkan dalam bentuk *pie chart*.

4. IMPLEMENTASI DAN PEMBAHASAN

Proses pengambilan data Twitter menggunakan format data JSON dengan query kata kunci dan penyaringan berupa bahasa dan jumlah *tweet* yang diambil dalam sekali *query*.

Twitter menyediakan *public API* berupa *search/tweet* yang dapat digunakan untuk mengambil data yang bersifat publik, sehingga *tweet* dari akun-akun yang di-*private* tidak akan muncul atau tidak akan terambil dengan menggunakan *public API*, selain itu twitter juga melakukan pembatasan pengambilan data setiap kali *query* dengan *public API search/tweet* yaitu 100 data *real time* terakhir.

Permasalahan yang muncul ketika melakukan pengambilan data pada twitter adalah adanya duplikasi data *tweet* atau yang biasa disebut dengan *spamming*, terutama dengan kata kunci yang lebih formal yang dilakukan oleh portal-portal berita online. Sebagai contoh pada gambar 4 ketika pengguna memasukkan kata kunci 'Joko Widodo' maka akan mendapatkan lebih banyak duplikasi *tweet* dibandingkan dengan kata kunci 'Jokowi', atau 'Ridwan Kamil' dibandingkan dengan 'Kang Emil' dan sebagainya.

<p>Kata Kunci : Jokowi</p> <p>Jangan Setengah Hati Gunakan #DanaDesa https://t.co/1A40P9Bf1 via @chrisstory @KemenDesa @HermanJafar @getahajid @Jokowi Nyakin den PDIP merangkul, lihat aja Jokowi "Kalau Sama Teman Ahok Bisa Raju, Ngapain Lagi Hego Sama Partai?" https://t.co/rj8 RT @RyT11ta: Para tokoh Golkar pggaduh adalah kelompok Jokowi-JK. Hreka ibin mmentingin kluasaan drpd mlah rjkt & neg. https:// @Jokowi kita mau nanan sahan tetapi pendapatn saya gak nentu, @yulingsih12228 @sutrisonorachai @B8Vudhoyono @HeryTanoa @RPI_RI @Jokowi itu yang lain mbak, kalau HT saya percaya dia punya hat Di Rakernas PDIP, Jokowi Laporan Kinerja Pemerintah https://t.co/FDg126p1Y</p>
<p>Kata Kunci : Jokowi Widodo</p> <p>Saya Bangga Dengan PDI Perjuangan: Presiden Joko Widodo merupakan kader PDI Perjuangan. Dia menyampaikan... https://t.co/25</p> <p>Saya Bangga Dengan PDI Perjuangan: Presiden Joko Widodo merupakan kader PDI Perjuangan. Dia menyampaikan... https://t.co/18</p> <p>: #FollowMe Jokowi: Saya Bangga Dengan PDI Perjuangan: Presiden Joko Widodo merupakan kader PDI Perjuangan. Dia menyampaikan... https://t.co/ehq4e10Hk #NewsUpdate</p> <p>Saya Bangga Dengan PDI Perjuangan: Presiden Joko Widodo merupakan kader PDI Perjuangan. Dia menyampaikan... https://t.co/UD</p> <p>Saya Bangga Dengan PDI Perjuangan: Presiden Joko Widodo merupakan kader PDI Perjuangan. Dia menyampaikan... https://t.co/3o</p>

Gambar 4. Perbandingan Streaming Data Dengan Kata Kunci 'Jokowi' dengan 'Joko Widodo'

4.1 Proses Training dan Validasi

Data yang disiapkan berjumlah 2.000 *tweet* yang dikumpulkan sejak tanggal 7 September 2015 sampai tanggal 31 Desember 2015. Dari sebanyak 2.000 data, 1.600 diantaranya digunakan sebagai data latih dan 400 digunakan sebagai data uji untuk mengetahui seberapa besar akurasi dari proses *training* yang sudah dilakukan. *Tweet* yang sudah dikumpulkan kemudian diproses seperti tabel 1.

Tabel 1. Contoh Pemrosesan Data

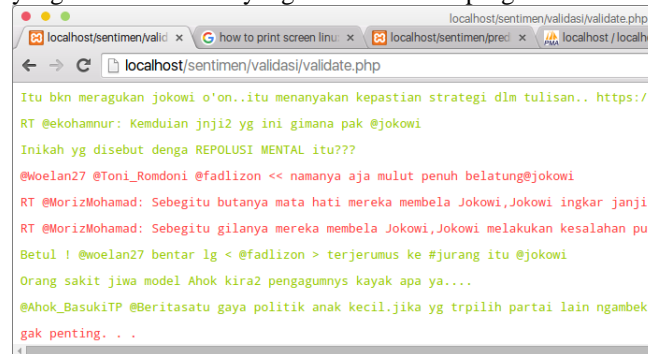
Data Mentah	Apapun bentuk responnya, saya mengucapkan terima kasih kepada presiden @jokowi. ini sebuah kehormatan untuk rakyat kecil seperti saya https://t.co/sRUK7PvdsN
<i>Preprocessing</i>	bentuk responnya mengucapkan terima kasih presiden kehormatan rakyat kecil
<i>Stemming</i>	bentuk respon ucap terima kasih presiden hormat rakyat kecil
Pembobotan dan konversi data vektor	5:1 20:1 96:1 109:1 110:1 178:1 390:1 412:1 419:1
Pelabelan	1 5:1 20:1 96:1 109:1 110:1 178:1 390:1 412:1 419:1

Data dalam bentuk vektor dengan label tersebut akan dimasukkan sebagai data training bagi SVM, dimana hasil dari *training* tersebut adalah *database* fitur dan sebuah model, kemudian dari model tersebut dilakukan pengujian dengan dataset pengujian yang sudah disiapkan seperti gambar 5.

<p>datavalidasi.txt X</p> <p>395 mungkin gak jufus kalla kongkalikong dengan said didu yang satu daerah asal? #saveDPRPecatpapanova</p> <p>396 jufus kalla harusnya mendinginkan suasana, bukannya malah jadi tukang kompor mledug #saveDPRPecatpapanova</p> <p>397 persamaan jufus kalla dan said didu ternyata di sepak bola... pantes seneng gocak-gocak... #saveDPRPecatpapanova</p> <p>398 "@dzatnursir: Ngeri, Kasus RJ Lino Bisa Menyeret Jusuf Kalla https://t.co/h5yLvpzhfy https://t.co/SbbV9tneox"</p> <p>399 RT @didienAZHAR: Sebelum Dipolisikan, Pose Ustadz Maulana Naik Mimbar di Masjid Bikin Heboh https://t.co/1ZDMUsW92q</p> <p>400 terserah mau omong apa tad,yg penting ada azab dan balasan Allah loh,Ustadz Maulana Berulah Lagi https://t.co/MXbPKIo4HR lewat @rimanews</p>
<p>Plain Text ▾ Tab Width: 8 ▾ Ln 400, Col 137 INS</p>

Gambar 5. Dataset Uji

Proses pengujian dilakukan secara manual dengan melakukan penghitungan berapa jumlah prediksi yang salah dan benar yang dilakukan oleh program.



Gambar 6. Hasil Pengujian dengan Dataset Uji

Dengan 400 data uji yang terdiri dari 200 data positif dan 200 data negatif, terdapat sebanyak 82 kesalahan yang dilakukan oleh program, sehingga dapat dihitung akurasi sebagai berikut :

$$\text{Akurasi} = \frac{400 - 82}{400} = 79.5\%$$

Akurasi sebesar 79.5% dianggap cukup baik untuk dijadikan sebagai model dan *database* fitur untuk perangkat lunak yang dikembangkan. Adapun beberapa kesalahan prediksi *tweet* oleh program seperti terlihat pada tabel 2.

Tabel 2. Contoh Kesalahan Prediksi

Tweet	Prediksi	Seharusnya
Inikah yang disebut dengan REPOLUSI MENTAL itu?	Positif	Negatif
Orang sakit jiwa model Ahok kira-kira pengagumnya kayak apa ya ...	Positif	Negatif
RT @Tiemin24: Jakarta kota mafia byk hewan buas, lebih cocok dipipin ahok yg tanpa kompromi #tegas. https://t.co/CqF8bwYUkX	Negatif	Positif
Sya ttp mndkungmu pak "@detikcom Relawan Jokowi : Sudirmain Said dianggap gagal me... http://t.co/t5DW4x4Ln1"	Negatif	Positif

Beberapa kesalahan prediksi umumnya dikarenakan kalimat *tweet* berupa sindirian baik berupa pertanyaan maupun pernyataan seperti pada kesalahan baris pertama dan kedua, sehingga makna yang ditulis berbeda dengan makna sebenarnya.

Selain itu kesalahan juga sering muncul ketika program tidak mengetahui konteks pembicaraan pada *tweet* yang dimaksud seperti pada baris ketiga.

Adanya kedua fitur sentimen pada sebuah *tweet* juga membuat program melakukan kesalahan prediksi seperti pada baris keempat.

4.2 Implementasi Mobile

Perangkat lunak *client* diimplementasikan pada platform Android. Halaman utama perangkat lunak berisi sebuah *text field* dan *button* yang digunakan pengguna untuk memasukan nama tokoh publik sebagai kata kunci dan mengirimkannya ke *server* untuk melakukan *request* data.

Data yang diterima oleh perangkat *client* ditampilkan dalam bentuk persentase, *pie chart* dan *list* dari *tweet* yang dianalisis. Pada *pie chart* potongan berwarna hijau menandakan presentase *tweet* positif, ungu untuk *tweet* netral dan merah untuk *tweet* negatif gambar 7.



Gambar 7. Menampilkan Data JSON

Pada *list tweet* yang dianalisa juga diberikan warna yang berbeda berdasarkan hasil analisis sentimen oleh perangkat lunak, tweet berwarna hijau berarti positif, ungu berarti netral dan merah berarti negatif.

5. PENUTUP

5.1 Kesimpulan

Berdasarkan pembahasan diatas, didapatkan beberapa kesimpulan, antara lain :

1. Pengambilan data melalui Twitter Streaming API dengan *public API search/tweets* dapat memberikan data *tweet public* secara *real time* dengan filter atau penyortiran yang sangat baik.
2. Tweet yang sudah melalui tahapan *preprocessing* dan *stemming* serta perubahan format data yang sesuai dengan LibSVM dapat dipisahkan dengan cukup baik oleh LibSVM menggunakan linear kernel dengan akurasi 79.5% dari pengujian dengan menggunakan data uji.
3. Akurasi 79,5% pada proses pengujian dianggap sudah cukup baik mengingat objek penelitian kali ini adalah kalimat berbahasa Indonesia dengan

struktur dan variasi kalimat dan bahasa yang lebih rumit.

5. JSON dapat digunakan untuk melakukan pertukaran data pada dua platform yang berbeda. Pada proses implementasi JSON dapat dibuat dan dikirimkan oleh PHP pada *server* dan dapat diterima oleh Java pada *client*, sehingga data yang diterima dapat ditampilkan dan divisualisasikan pada *pie chart* dengan pustaka HelloChart for Android.

5.2 Saran

Berdasarkan hasil evaluasi terhadap perangkat lunak yang dikembangkan agar pada pengembangan selanjutnya didapatkan hasil yang lebih baik lagi. Berikut adalah saran-saran yang dipandang perlu untuk dipertimbangkan untuk penelitian selanjutnya :

1. Melakukan deteksi pada *tweet spam* sehingga mengurangi duplikasi *tweet*.
2. Pemilihan data latih dilakukan lebih selektif sehingga fitur pada masing-masing kelas yang berbeda dapat terekstraksi dengan lebih baik.
3. Menentukan topik pada bagian analisis sehingga prediksi dapat dilakukan dengan lebih akurat.

DAFTAR PUSTAKA

- [1] Ahmad Hidayatullah. 2014. *Analisis Sentimen dan Klasifikasi Kategori Terhadap Tokoh Publik Pada Twitter*. Universitas Pembangunan Veteran. Yogyakarta.
- [2] B. Liu. 2012. *Sentiment Analysis and Opinion Mining*, Morgan & Claypool Publishers.
- [3] B. Liu. 2010. *Handbook of Natural Language Processing, chapter Sentiment Analysis and Analysis, 2nd Edition*. Chapman & Hall / CRC Press. Prancis.
- [4] Ferry Yuwono. 2012. *Algoritma Porter Stemmer For Bahasa Indonesia Untuk Pre-Processing Text Mining Berbasis Metode Market Basket Analysis*. Universitas Kristen Petra.
- [5] Janardhana, Ravikiran. 2012. *How to build a twitter sentiment analyzer?*, (<http://ravikiranj.net/posts/2012/code/how-build-twitter-sentiment-analyzer/>, diakses 18 September 2015).
- [6] J. Han, M. Kamber, and J. Pei. 2006. *Data Mining: Concepts and Techniques, Second Edition, 2nd ed.* Morgan Kaufmann. California.
- [7] Kowalczyk, Alexandre. 2015. *How to prepare your data for text classification?*, (<http://www.svm-tutorial.com/text-classification-prepare-data/>, diakses 10 September 2015).
- [8] Noviah Dwi Purtanti. 2014. *Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support*

Vector Machine. Universitas Gadjah Mada. Yogyakarta.

- [9] Nurvirta Monarizqa. 2014. *Penerapan Analisis Sentimen Pada Twitter Berbahasa Indonesia Sebagai Pemberi Rating*. Universitas Gadjah Mada. Yogyakarta.
- [10] R. Feldman dan J. Sanger. 2006. *The Text Mining Handbook*. Cambridge University Press. Cambridge.