

Analisis User Opinion Twitter Pada Level Fine-grained Sentiment Analysis Terhadap Tokoh Publik

***Twitter User Opinion Analysis At Fine-grained
Sentiment Analysis Level Toward Public Figure***
Kadek Yogi Setiawan¹, Hetti Hidayati², Alfian Akbar Gozali³

^{1,2,3} School of Computing, Telkom University

Jalan Telekomunikasi No.1, Dayeuh Kolot, Bandung 40257

¹kadekvogisetiawan@gmail.com, ²htt@ittelkom.ac.id, ³alfian.akbar.gozali@gmail.com

Abstrak

Berbagai situs *social media*, sering kali menimbulkan feedback dari para *follower* terkait user yang mereka *follow* sebelumnya. *Review* dari *follower* dapat dijadikan pertimbangan sebagai sumber informasi yang berguna baik untuk pihak *follower* itu sendiri maupun untuk pihak user tersebut. Namun untuk user yang memiliki jumlah *follower* yang sangat banyak. Ini dapat menyulitkan *follower* lain jika harus membaca satu per satu semua *review* yang ada. Maka dari itu diperlukan suatu sistem yang dapat menilai *review* atau opini dari *follower* berdasarkan *tweet*. Salah satu solusi yang diusulkan adalah melakukan suatu penilaian terhadap opini dari user. *Heterogenous graph model* adalah salah satu metode berorientasikan *graph* dimana pemodelan datanya dimodelkan dalam bentuk *graph* dalam hal ini user yang terhubung langsung secara langsung atau *directed* yang akan dipakai dalam penentuan klasifikasi user yang akan diambil. Hasil pengujian menunjukkan bahwa nilai akurasi berdasarkan *directed graph* lebih baik jika dibandingkan penilaian berdasarkan komentar asli saja dengan rata-rata akurasi yang dihasilkan pada kisaran 69%. Besarnya nilai akurasi ini dipengaruhi oleh jumlah kata dan varian kata pada tiap *review* user

Kata kunci : *Social media, Sentiment analysis, Opinion mining, Twitter*

Abstract

Various social media sites, often give feedback from the follower with a user that they follow previously. Review of online followers may be one consideration as a source of valuable information both for the follower itself and for the topic. However, for users who has many followers. It can be difficult for another follower if it must read one by one of all the existing reviews. Therefore we need a system that can summarize a review of the user based on tweet. One proposed solution is to conduct an assessment of the users' opinion. Graph-based is one of method that graph-oriented data where the modeling data is modeled in graph form in this case the user is directly connected which will be used in determining the classification of a user to be taken. The test results show that the accuracy value based on directed graph is better than compared with assessment based on the original commentary with average accuracy produced in the range of 69%. The value of this accuracy is affected by the number of words and word variants for each user review.

Keywords: *Social media, Sentiment analysis, Opinion mining, Twitter*

1. Pendahuluan

Sentiment analysis didefinisikan sebagai tugas meneliti pendapat tentang suatu entitas tertentu. Proses pengambilan keputusan dari seseorang dipengaruhi oleh pendapat yang dibentuk oleh para user. Ketika seseorang ingin membeli produk secara online dia biasanya akan mulai dengan mencari ulasan dan opini yang ditulis oleh orang lain pada berbagai macam tawaran. Sistem yang secara otomatis menentukan sudut pandang yang akan memungkinkan pengguna untuk memahami sebagian besar pendapat yang diungkapkan di internet, mulai

dari ulasan produk sampai posisi pada kasus politik[15]. Media jejaring sosial banyak dimanfaatkan para politisi untuk berkampanye dan meningkatkan popularitas mereka menjelang pemilihan umum.

Salah satu media jejaring sosial yang telah dimanfaatkan dalam pemilihan umum adalah Twitter. Hal tersebut dikarenakan Twitter merupakan salah satu media jejaring sosial dengan pengguna terbanyak diantara beberapa situs jejaring sosial yang ada. Analisis sentimen terhadap tokoh publik dari Twitter menjelang adanya pemilihan umum sangat bermanfaat dalam memberikan tambahan wawasan

serta gambaran bagi masyarakat tentang tokoh publik yang menjadi kandidat dalam pemilihan umum.

Oleh karena itu, digunakanlah Twitter sebagai dasar untuk percobaan klasifikasi sentimen, dengan cara memasukkan informasi antar pengguna dan relasinya, sebagai berikut. Pertama memanfaatkan model yang didasarkan pada jaringan pengikut/ yang diikuti yang memiliki ketergantungan tidak hanya antara pendapat pengguna dan pendapat yang diutarakan dalam *tweet* nya, tetapi juga antara pendapat pengguna dan orang-orang yang diikutinya. Varian "@" berbasis jaringan dipertimbangkan untuk digunakan, dimana pengguna memiliki ketergantungan antara pendapat pengguna lain dan pendapat dari orang-orang yang dapat di *mention* melalui "@". Alasan mengapa perlu diadopsi pendekatan semi-supervised adalah bahwa akuisisi untuk sejumlah data yang besar dan relevan dengan sentimen-label dapat menjadi rawan akan kesalahan dan dapat memakan waktu yang lama. Penelitian ini berfokus pada level opini karena tujuan akhir bagi banyak pengguna teknologi *opinion-mining* adalah untuk mengetahui apa yang orang pikir serta menentukan sentimen yang dinyatakan dalam pesan suatu individu. Selain itu, banyak kasus di mana beberapa *tweet* pengguna yang benar-benar ambigu (mungkin karena *tweet*nya yang sangat pendek dan dibatasi maksimal 140 karakter) tetapi pendapat pengguna secara keseluruhan dapat ditentukan dengan melihat koleksi *tweets*nya dan kepada siapa seseorang dapat terhubung. Pengklasifikasian pada kata sifat juga dilakukan karena banyaknya kata ambigu yang memiliki lebih dari satu klasifikasi kata atau makna, dengan dilakukannya pengklasifikasian kata berdasarkan kata sifat, dapat membantu dalam pengertian maksud dari *tweet* seseorang. Alasan lain menggunakan twitter adalah karena twitter memiliki keterbatasan dalam hal penyampaian kata-kata tidak seperti media social lainnya yang bebas mengekspresikan kata-kata tanpa batasan dan twitter sendiri dapat menimbulkan sentimen yang berbeda tergantung dari sudut pandang dan pendapat user lain terhadap user/obyek tertentu yang akan diamati.

2. Landasan Teori

2.1 Social network

Social network adalah sebuah media untuk bersosialisasi satu sama lain dan dilakukan secara online yang memungkinkan manusia untuk saling berinteraksi tanpa dibatasi ruang dan waktu. Seorang individu dapat terpengaruh di lingkungan jaringan sosial mereka dengan kejadian yang terjadi di sekitar lingkungan sekitar mereka. *Social network* pertama diawali dengan peluncuran SixDegrees.com pada tahun 1997 dimana user atau pengguna dapat

membuat profil dan mendaftarkan teman mereka, sementara untuk menjelajahi pertemanan pertama kali dimulai pada tahun 1998. Pada era 2000an, *social network* mulai banyak bermunculan dimulai dari

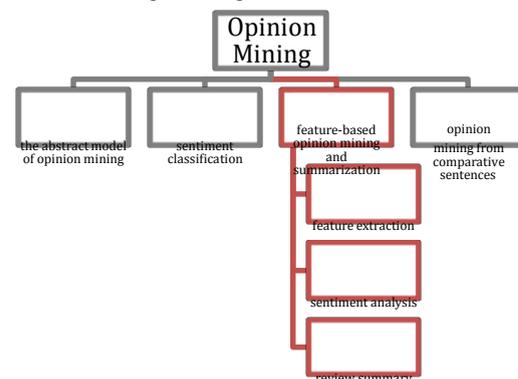
Friendster, Match.com, MySpace, twitter, hingga Facebook yang memiliki fitur bagi developer di luar Facebook untuk membangun aplikasi yang memperbolehkan user mempersonalisasi profil mereka[7]

2.1.1 Twitter

Twitter, sebuah layanan microblogging, telah muncul sebagai media baru dan tengah menjadi sorotan. Pengguna Twitter mem*follow* atau *difollow* user lain. Tidak seperti pada kebanyakan situs jejaring sosial online, seperti Facebook atau MySpace, dimana pada twitter hubungan antar user tidak memerlukan konfirmasi pertemanan untuk berinteraksi. Seorang pengguna dapat mem*follow* pengguna lain, dan user yang sedang *difollow* tidak harus mem*follow* kembali. Menjadi follower di Twitter berarti bahwa pengguna menerima semua pesan (disebut *tweet*) dari user yang mem*follow*nya. Praktek umum menanggapi *tweet* telah berkembang ke dalam budaya markup didefinisikan dengan baik: RT singkatan dari reply *tweet*, '@' diikuti dengan tanda alamat atau id user tertentu, dan '#' biasa disebut hashtag diikuti oleh kata tertentu. Kosakata markup ini didefinisikan dengan baik dikombinasikan dengan batas yang ketat yakni 140 karakter per postingan kemudahan pengguna dengan singkat dalam ekspresi. Mekanisme *tweet* balik memberdayakan pengguna untuk menyebarkan informasi pilihan mereka tanpa harus mendapat informasi aslinya[10].

2.2 Opinion Mining

Informasi merupakan data yang telah diolah sedemikian rupa sehingga berguna bagi *user* yang membutuhkan dalam membantu mengambil keputusan. Informasi sendiri dapat dibedakan menjadi dua, yaitu fakta dan opini. Fakta merupakan pernyataan objektif terhadap sesuatu yang sudah terjadi dan biasanya disertai bukti-bukti, sedangkan opini lebih bersifat subjektif bagaimana seseorang mengekspresikan diri mereka terhadap segala sesuatu yang terjadi sesuai dengan persepsi dan asumsi mereka masing-masing.



Gambar 2-1 Model opinion mining [3]

Opinion mining dikenal juga dengan sebutan *sentiment analysis*. *Opinion mining* merupakan salah

satu cabang dari *data mining* dimana *opinion mining* ini mengasumsikan secara umum informasi yang bersifat subjektif dan mengklasifikasikannya menjadi opini positif atau opini negatif. Pada umumnya *opinion mining* dapat dibagi menjadi beberapa model[3] :

1. *The abstract model of opinion mining*
Berkfokus untuk mengekstrak suatu kalimat dan menentukan yang termasuk fitur atau opini didalamnya.
2. *Sentiment classification*
Berkfokus untuk menentukan suatu kalimat opini apakah kalimat tersebut termasuk opini positif atau opini negatif.
3. *Featured-based opinion mining and summarization*
Berkfokus untuk membuat hasil ringkasan kalimat opini yang sudah dikelompokkan berdasarkan fitur produk dan orientasinya. Merupakan model yang akan dibahas pada tugas akhir ini.
4. *Opinion mining from comparative sentences*
Berkfokus untuk menentukan orientasi opini yang bersifat membandingkan yang satu dengan yang lain.

2.2.1 Sentiment Analysis

Sentiment analysis disebut juga *opinion mining* adalah tugas mencari pendapat tentang suatu entitas tertentu[15 yang artinya menentukan apakah suatu teks tergolong kalimat opini atau bukan. *Sentiment analysis* bertujuan untuk mengetahui sikap pembicara atau penulis sehubungan dengan beberapa topik atau polaritas kontekstual keseluruhan dokumen. Sikap mungkin nya penilaian atau evaluasi, keadaan afektif (artinya, keadaan emosi penulis saat menulis), atau komunikasi emosional (yaitu, efek emosional penulis yang ingin ditanamkan pada si pembaca). Sebuah cabang penelitian di domain Text Mining yang mulai booming pada awal tahun 2002-an. Riset-nya mulai marak semenjak paper dari B.Pang dan L.Lee keluar. Secara umum, *Sentiment analysis* ini dibagi menjadi 2 kategori[6]:

2.2.1.1 Coarse-Grained Sentiment Analysis

Coarse-Grained merupakan proses analisis yang dilakukan pada level dokumen. Singkatnya adalah mencoba mengklasifikasikan orientasi sebuah dokumen secara keseluruhan. Orientasi ini ada 3 jenis : Positif, Netral, Negatif. Akan tetapi, ada juga yang menjadikan nilai orientasi ini bersifat kontinu / tidak diskrit[6].

2.2.1.2 Fined-Grained Sentiment Analysis

Objek yang ingin diklasifikasi pada fine-grained bukan berada pada level dokumen melainkan sebuah kalimat pada suatu dokumen. Orientasi ini ada 2 jenis: Positif dan Negatif. Informasi berupa paper, artikel, jurnal dan lain sebagainya berisikan informasi tekstual yang bersifat objektif, yaitu informasi yang tidak menggambarkan perasaan atau pendapat

penulis terhadap hal yang ditulisnya. Informasi-informasi objektif tersebut cenderung bersifat netral, sedangkan informasi seperti opini, komentar atau pendapat masing-masing memiliki nilai atau makna yang diungkapkan oleh penulisnya, makna tersebut dapat bernilai positif atau negatif. Nilai positif dan negatif yang terkandung dalam kalimat subjektif ini disebut dengan polaritas. Kalimat opini, komentar atau pendapat yang menggambarkan penilaian atau perasaan seseorang terhadap suatu hal merupakan sentimen. Polaritas sentimen ini yang menjadi objek penelitian yang akan dilaksanakan[6].

2.2.1.3 Sentiment Analysis Dengan Klasifikasi Adjective

Dengan menggunakan dataset yang telah dibangun, sistem mengklasifikasikan arti dari kata sifat menjadi positif atau negatif sentimen polaritas sesuai dengan konteks tertentu. Disambiguating sentimen terhadap kata sifat yang ambigu termasuk dalam NLP. Penelitian yang pernah dilakukan sebelumnya sebagian besar terfokus pada rasa kata (*word sense*) disambiguasi daripada sentimen disambiguasi. Meskipun kedua masalah terlihat mirip, namun pada kasus sentimen disambiguasi penilaiannya lebih subjektivitas. Untuk menyelesaikannya, tidak hanya berurusan dengan semantik konteks, tetapi juga dengan aspek psikologis dari persepsi manusia emosi dari teks tertulisp[17].

2.3 Klasifikasi

Klasifikasi merupakan salah satu tahapan *preprocessing* pada *data mining* yang bertujuan untuk memprediksi kelas dari setiap data yang ada, yang nantinya diberi label sesuai dengan akurasi yang mendekati. Klasifikasi dibagi menjadi yaitu *testing* dan *key*. Pada proses *learning*, terjadi pembangunan model oleh *data testing* yang digunakan untuk prediksi kelas. Sedangkan pada data *key*, terjadi pengujian manual yang bertujuan untuk menentukan seberapa besar akurasi dari model yang sudah dibangun[14].

2.4 Wordnet

WordNet merupakan sebuah leksikal database elektronik. WordNet dikembangkan untuk bahasa Inggris oleh Universitas Princeton di Amerika. WordNet adalah sistem referensi leksikal online yang rancangannya terinspirasi oleh teori psikolinguistik dari memori leksikal manusia. Kata benda, kata kerja, kata sifat dan kata keterangan dalam bahasa Inggris diorganisir menjadi himpunan sinonim, dimana masing masing merepresentasikan satu konsep leksikal. Relasi yang berbeda menghubungkan himpunan sinonim[14].

2.4.1 Sentiwordnet

SentiWordNet merupakan kamus *opini lexicon* yang berasal dari pengembangan *database WordNet* dimana berisi kumpulan kata-kata yang telah memiliki nilai numerik yang digunakan untuk

menunjukkan apakah kata tersebut memiliki sentiment positif atau negatif. Kata opini yang sudah terekstrak akan dicari nilainya pada *database* untuk

selanjutnya ditentukan orientasinya[14].

2.5 NodeXL

NodeXL adalah sebuah aplikasi gratis berbasis open-source template untuk Microsoft Excel © 2007, 2010 dan 2013 yang membuatnya mudah untuk menjelajahi grafik suatu jaringan.

Dengan NodeXL, user dapat membuat grafik hierarki secara kustom. user bisa membuat grafik jejaring sosial tanpa menggunakan aplikasi dan pemrograman yang rumit. user bisa mengimpor follower pada Twitter atau YouTube atau Flickr, lalu membuat grafik yang merepresentasikan koneksi pada jejaring

sosial user, semua tersedia dalam jendela Excel yang sudah familiar dipakai dalam membuat suatu

data[10].

2.6 Natural Language Processing

Natural language processing secara teori merupakan suatu teknik komputasi untuk menganalisis dan menggambarkan teks secara alami dalam satu atau lebih tingkat analisis linguistic dengan tujuan untuk mencapai manusia seperti pengolahan bahasa untuk berbagai tugas dan aplikasi. *Natural language processing* menyediakan teori maupun implementasinya untuk berbagai aplikasi. Contoh beberapa aplikasi yang menerapkan NLP yakni *Summarization* (yang akan menjadi fokus pada pengerjaan tugas akhir ini), *Machine Translation*, *Dialogue System*, *Information Retrieval (IR)*, *Information Extraction (IE)*, *Question-Answering*[16].

2.7 Performance Evaluation

Evaluasi performansi yang digunakan pada tugas akhir ini adalah akurasi. Untuk pengklasifikasiannya ada empat kelas :

Istilah positif dan negatif mengacu pada prediksi *classifier* (biasa disebut *observation*), sedangkan istilah *true* dan *false* mengacu pada prediksi dengan penilaian eksternal (biasa disebut *expectation*). Berikut tabel matriks contingency dan actual :

Tabel 2-2 Matriks Contingency dan Actual

X/Y	Predicted Class		
Actual Class		<i>class = yes</i>	<i>class = no</i>
	<i>class = yes</i>	TP	FN
	<i>class = no</i>	FP	TN

2.7.1 Accuracy

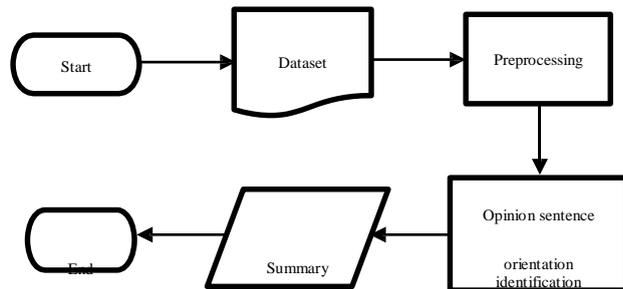
Accuracy merupakan perbandingan dari

total semua prediksi yang diklasifikasikan ke dalam kelas tersebut.

Rumus *accuracy* :

$$Accuracy = \frac{TP + TN}{(TP + FN) + (FP + TN)} \quad (2.3)$$

3. Perancangan Sistem



jumlah ketepatan prediksi tiap kelas dengan jumlah

Gambar 3 Skema Umum Sistem

1. Sistem menerima input berupa dataset yang berisi *review* dari konsumen. Dataset ini terdiri dari 4 jenis topik yang berbeda.
2. Dataset ini akan dibagi menjadi dua tipe, yaitu data testing dan data kunci.
3. Pada dataset kemudian dilakukan *preprocessing*.
4. *Opinion words* yang telah didapatkan sebelumnya kemudian dicari orientasinya dengan menggunakan *SentiWordNet*. Kalimat opini yang mengandung fitur dan opini kemudian dihitung skornya untuk menentukan orientasi dari kalimat tersebut.

Perbedaan sistem ini dengan sistem sebelumnya yang telah diriset adalah terletak pada dataset dan bahasa yang digunakan. Dengan struktur bahasa yang berbeda diperlukan penanganan yang berbeda pula, mulai dari *preprocessing*, proses orientasi *opinion words* dan kalimat opini, hingga proses *summarization*

4. Analisis dan Pengujian

4.1 Pengujian Sistem

Tujuan pengujian pada tugas akhir ini adalah untuk melihat hasil akurasi dari proses sistem *opinion summarization*. Hasil keluaran yang akan diuji antara adalah jumlah akurasi dari kalimat opini. Hal ini dilakukan bertujuan untuk mengetahui tingkat keberhasilan sistem dalam melakukan tahapan pada proses *opinion summarization* ini.

4.1.1 Tujuan Pengujian

Tujuan dilakukan pengujian ini adalah :

1. Menganalisis akurasi komentar berdasarkan *performance evaluation* dengan menggunakan klasifikasi adjective maupun tidak.

4.1.2 Dataset

Dataset yang digunakan adalah data twitter yang diambil dari berbagai user. Jumlah user yang digunakan di dalam pengujian ini 4 akun dimana masing-masing akun diambil sample yang berjumlah 500 tweet. Berikut adalah rincian *dataset* yang digunakan :

Tabel 4-1 Dataset opini topik

Dataset	Rincian data
@Prabowo08	500 kalimat opini (minggu 1) 200 kalimat opini (minggu 3)
@hattarajasa	500 kalimat opini (minggu 1) 200 kalimat opini (minggu 3)
@jokowi_do2	500 kalimat opini (minggu 1) 200 kalimat opini (minggu 3)
@Pak_JK	500 kalimat opini (minggu 1) 200 kalimat opini (minggu 3)

Dataset ini kemudian dibagi menjadi *data testing* dan *data key* yang dipilih secara acak dengan perbandingan antara *data testing* dan *data testing* yaitu 1 : 1. Dilihat dari *review* pada *dataset* rata-rata memiliki jumlah kata 8-25.

Kalimat opini yang digunakan pada dataset memiliki beberapa aturan diantaranya :

1. Kalimat opini yang hanya bersifat eksplisit menyebutkan karakteristik seseorang yang dikomentari oleh konsumen dan menggunakan bahasa Indonesia.
2. Antara satu kalimat opini dengan yang lain dipisahkan oleh baris yang berbeda.

4.2 Skenario Pengujian

Berikut adalah skenario uji yang akan dilakukan untuk penilaian sistem *opinion summarization* ini :

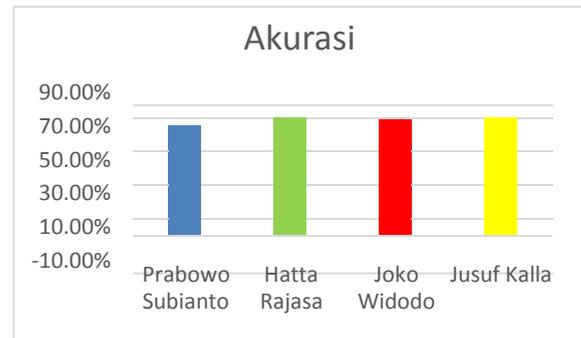
1. Prediksi akurasi kalimat opini.
Pengujian ini dilakukan untuk mengetahui keberhasilan sistem dalam menentukan akurasi dari kalimat opini.
2. Prediksi akurasi kalimat opini dengan tambahan kata *adjective*
Pengujian ini dilakukan untuk mengetahui keberhasilan sistem dalam menentukan akurasi dari kalimat opini dengan menambahkan pengklasifikasian kata-kata sifat/*adjective* yang sering dianggap banyak orang sebagai kata benda/*noun* karena terlalu seringnya diucapkan.

4.3 Analisa Hasil Pengujian

Berikut akan dijelaskan beberapa hasil pengujian yang telah dilakukan :

4.3.1 Analisis Akurasi User Berdasarkan Karakteristik Data

Dari hasil pengujian data dengan empat jenis *dataset* berdasarkan karakteristik *dataset* tersebut maka diperoleh nilai *akurasi* karakteristik *user* yang terekstrak untuk semua data, yaitu :



Gambar 4-1 Grafik akurasi opini

Dari gambar 4-1 dapat diketahui bahwa perbandingan nilai akurasi Data pertama adalah Prabowo Subianto yakni yang paling rendah dibanding ketiga data lainnya yakni bernilai 65,09%. Data kedua adalah Hatta Rajasa dimana tingkat akurasinya yakni bernilai 72,15%. Data ketiga adalah Joko Widodo yang memiliki tingkat akurasi sebesar 66,28%. Dan data yang terakhir adalah Jusuf Kalla yang memiliki tingkat akurasi paling tinggi dibanding ketiga data lainnya sebesar 74,82%. Jika pada satu *dataset* memiliki kalimat dengan banyak kata memenuhi aturan yang sudah ditetapkan sentiwordnet maka peluang tingkat akurasi lebih tinggi dapat dicapai.

4.3.2 Analisis Sentimen User Berdasarkan Akurasi Data

Karena user yang diambil oleh penulis merupakan heterogenous directed graph dari topik yang bersangkutan, maka analisis awal untuk opini setiap user bernilai positif. Maka hasil sentimen dapat diketahui dengan cara menggunakan perhitungan hasil akurasi data diatas dimana *accuracy* (a) yang didapatkan dikalikan dengan total tweet (t) yang ada pada setiap topic yakni 700 tweet di bagi dengan 100% total akurasi. $\frac{a}{100\%} \times 700$ maka didapatkan hasil sentimen sebagai berikut:

Tokoh Publik	Sentimen Positif	Sentimen Negatif
Prabowo Subianto	456	244
Hatta Rajasa	505	195
Joko Widodo	464	236
Jusuf Kalla	524	176

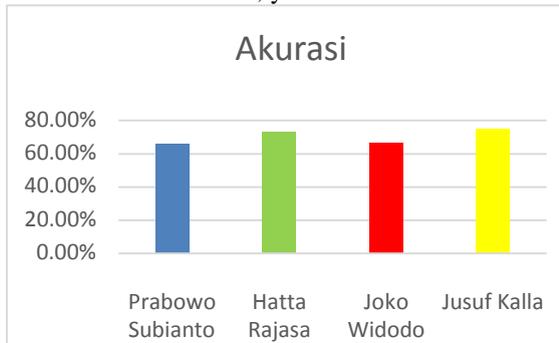
Tabel 4-2 Hasil sentimen opini user

Dari tabel 4-2 diketahui Jusuf Kalla memiliki sentimen positif tertinggi yakni sebanyak 524 sentimen, diikuti dengan Hatta Rajasa yang memiliki sentimen positif sebanyak 505 sentimen, lalu Joko

Widodo dengan 464 sentimen positif, dan yang terakhir Prabowo Subianto dengan 456 sentimen positif.

4.3.3 Analisis Akurasi User Berdasarkan Karakteristik Data dengan Klasifikasi Adjective

Dari hasil pengujian data dengan empat jenis *dataset* berdasarkan karakteristik *dataset* tersebut maka diperoleh nilai *akurasi* karakteristik *user* yang terekstrak untuk semua data, yaitu :



Gambar 4-2 Grafik akurasi opini

Dari gambar 4-2 dapat diketahui bahwa perbandingan nilai akurasi dengan melakukan pengklasifikasian kata-kata adjective pada data pertama adalah Prabowo Subianto yakni yang paling rendah dibanding ketiga data lainnya yakni bernilai 65,97%. Data kedua adalah Hatta Rajasa dimana tingkat akurasinya bernilai 73,70%. Data ketiga adalah Joko Widodo yang memiliki tingkat akurasi sebesar 66,93%. Dan data yang terakhir adalah Jusuf Kalla yang memiliki tingkat akurasi paling tinggi dibanding ketiga data lainnya yakni sebesar 75,05%. Jika pada satu *dataset* memiliki kalimat dengan banyak kata memenuhi aturan yang sudah ditetapkan sentiwordnet maka peluang tingkat akurasi lebih tinggi dapat dicapai.

4.3.4 Analisis Sentimen User Berdasarkan Akurasi Data dengan Klasifikasi Adjective

Karena user yang diambil oleh penulis merupakan heterogenous directed graph dari topik yang bersangkutan, maka analisis awal untuk opini setiap user bernilai positif. Maka hasil sentimen dapat diketahui dengan cara menggunakan perhitungan hasil akurasi data diatas dimana *accuracy* (a) yang didapatkan dikalikan dengan total tweet (t) yang ada pada setiap topic yakni 700 tweet di bagi dengan 100% total akurasi. $\frac{a}{100\%} \times 700$ maka didapatkan hasil sentimen sebagai berikut:

Tokoh Publik	Sentimen Positif	Sentimen Negatif
Prabowo Subianto	462	238
Hatta Rajasa	513	187
Joko Widodo	469	231
Jusuf Kalla	525	175

Tabel 4-3. Hasil sentimen opini user

Dari tabel 4-3 diketahui Hatta Rajasa memiliki sentimen positif tertinggi yakni sebanyak 382 sentimen, diikuti dengan Jusuf Kalla yang memiliki sentimen positif sebanyak 361 sentimen, lalu Joko

terakhir Prabowo Subianto dengan 299 sentimen positif.

4.3.5 Analisis Hasil Pengujian

Perbedaan diantara analisis akurasi dengan pengklasifikasian dan tanpa pengklasifikasian *adjective* adalah tingkat akurasi pada keempat dataset yang berpengaruh terhadap sentimen positif keempat dataset tersebut.

Dari keempat dataset yang penulis gunakan, penulis memilih salah satu dataset sebagai perbandingan antara dataset yang diuji tanpa menambahkan kata *adjective* dan yang telah ditambahkan kata *adjective*.

Tabel 4-4 contoh tweet dengan klasifikasi *adjective* dan bukan *adjective*

Tweet	Kata adjective
presiden2014com RT @kajieko: Indonesia Hebat, Karyawan @Metro_TV dukung @Prabowo08 dan @hattarajasa ini baru jujur dan adil @Presiden2014com @TrioMacan2000 . relawandibopiss RT"@hermaniueo: Aamiin RT @DonyWenas: Berita "@antaranews: @Prabowo08 - @hattarajasa Diprediksi Menang Telak http://t.co/1mc7AKMawt" Mantap" .	"jujur", "bagus", "adil", "gila", "sombong", "keren", "angkuh", "boros", "selamat", "sensitif", "jelek", "munafik", "demokratis", "optimis", "maksimal", "juara", "muda", "fitnah", "cepat", "kaku", "tegas", "sederhana", "maju", "tangguh", "luar biasa", "lembut", "lengkap", "bosan", "prima", "bodoh", "jijik", "miskin", "jauh", "unik", "pandai", "cerdas", "hemat", "malas", "halus"

Dari tabel diatas diambil dua contoh tweet dimana tweet pertama memiliki kata adjective dan tweet kedua tidak memiliki kata adjective yang telah diklasifikasikan. Dimana kata "jujur" pada data

sentiwordnet yang penulis pakai berada pada klasifikasi r(adverb) bukan pada klasifikasi a(adjective) maka dari itu penulis melakukan pengklasifikasian terhadap tweet yang memiliki klasifikasi yang masih ambigu. Ini disebabkan oleh data sentiwordnet yang penulis gunakan, karena data sentiwordnet yang penulis gunakan merupakan terjemahan dari bahasa inggris ke bahasa indonesia dimana didalam bahasa inggris satu buah kata dapat memiliki lebih dari satu arti sedangkan dalam bahasa indonesia satu buah kata hanya memiliki satu buah arti. Seperti contoh diatas "jujur". jadi tingkat akurasi kata yang masih dianggap memiliki lebih dari satu arti atau klasifikasi akan bernilai lebih besar.

5. Kesimpulan dan Saran

5.1 Kesimpulan

Berdasarkan hasil pengujian yang didapat, maka dapat ditarik kesimpulan sebagai berikut:

1. Dari 4 sampel tokoh publik yang diberikan maka dapat diketahui user Jusuf Kalla paling direkomendasikan untuk difollow dikarenakan memiliki sentimen positif terbanyak diantara sampel lainnya.
2. *user opinion* pada *twitter* dengan *sentiment analysis* berdasarkan klasifikasi kata sifat memiliki nilai positif yang lebih baik dibandingkan yang tidak menggunakan klasifikasi kata sifat/*adjective*.

5.2 Saran

Berikut adalah beberapa saran yang dapat diberikan:

1. Menggunakan data twitter yang lebih banyak
2. Menggunakan rentang waktu yang lebih bervariasi dalam pengambilan data.
3. Mengidentifikasi kalimat opini yang bersifat implisit.
4. Proses pengambilan *dataset* dapat dilakukan langsung oleh sistem tanpa harus meng-copy *paste*.

Daftar Pustaka:

- [1] Apoorv Agarwal, Boyi Xie, Ilya Vovsha, Owen Rambow, Rebecca Passonneau, *Sentiment Analysis of Twitter Data*. New York, USA, 2010.
- [2] Bing Liu, *A Survey Of Opinion Mining and Sentiment Analysis*. University of Illinois at Chicago. Chicago, IL, 2012.
- [3] Bing Liu, *Opinion Mining*. Department of Computer Science, University of Illinois at Chicago, 851 S. Morgan Street, Chicago, IL 60607-0753, 2007.
- [4] Bo Pang and Lillian Lee, *Opinion Mining and Sentiment Analysis*. Foundations and TrendsR in Information Retrieval, USA, Vol. 2, Nos. 1–2 (2008) pp. 1–135.

[5] Chenhao Tan, Jie Tang, Lillian Lee, Long Jiang, Ming Zhou, Ping Li. *User-Level Sentiment Analysis Incorporating Social Networks* Microsoft Research Asia, San Diego, California, USA. August 21–24, 2011.

[6] Clayton R. Fink, Danielle S. Chou, Jonathon J. Kopecky and Ashley J. *Coarse- and Fine-Grained Sentiment Analysis of Social Media Text*. Llorens. JOHNS HOPKINS APL TECHNICAL DIGEST, 2011, VOLUME 30, NUMBER 1 .

[7] Danah m. boyd, Nicole B. Ellison. *Social Network Sites: Definition, History, and Scholarship*. USA. 2008.

[8] Daniel P. Huttenlocher, Pedro F. Felzenszwalb. *Efficient Graph-Based Image Segmentation*. Massachusetts. 2002.

[9] Mike Thelwall, Rudy Prabowo. *Sentiment Analysis: A Combined Approach*. University of Wolverhampton, Wolverhampton, UK, 2009 .

[10] Nick Gramsky, *The visualization in NodeXL*. University of Maryland, College Park, USA. 2012.

[11] Patrick Paroubek, Pak Alexander, Twitter Based System: *Using Twitter for Disambiguating Sentiment Ambiguous Adjectives*. Universit´e de Paris-Sud, Laboratoire LIMSI-CNRS, Bˆatiment 508, F-91405 Orsay Cedex, France 2010.

[12] Reinhard Diestel, *Graph Theory*. Electronic Edition 2000. Springer-Verlag New York 1997, 2000, Vol. 173 pp. 2.

[13] Ronen Feldman, *Techniques and Applications for Sentiment Analysis*. Communications of the ACM, Vol. 56 No. 4, pp. 82-89. 2013.

[14] Tierney Brendan, Bruno Ohana, Sentiment classification of reviews using SentiWordNet. School of Computing 9th. IT & T Conference, Dublin Institute of Technology 2009.

[15] W Maharani, Demanti, Restia. *Opinion Summarization Fitur Produk dalam Bahasa Indonesia dengan Metode Maximum Entropy*. Fakultas Teknik Informatika Institut Teknologi Telkom Bandung. 2013.

[16] Wicaksono, Alfan Farizki. *HMM Based Part-of-Speech Tagger for Bahasa Indonesia*. Bandung : ITB. 2010.

[17] Yongzheng (Tiger) Zhang, Dan Shen*. *Sentiment Analysis in Practice*. 2011.