

PENGEMBANGAN APLIKASI *SENTIMENT ANALYSIS*

MENGGUNAKAN METODE *NAÏVE BAYES*

(Studi Kasus Sentiment Analysis dari media Twitter)

Falahah¹⁾, Dyar Dwiki Adriadi Nur²⁾

^{1,2}Program Studi Teknik Informatika, Universitas Widyatama

Jl. Cikutra no.204 A Bandung

E-mail : falahah@widyatama.ac.id¹⁾, dyar.dwiki@widyatama.ac.id²⁾

Abstrak

Twitter adalah salah satu media sosial microblogging dimana pengguna dapat mencari topic tertentu dan membahas isu-isu terkini. Beberapa pesan singkat atau tweet dapat memuat opini terhadap produk dan layanan yang dirasakan oleh masyarakat. Pesan ini dapat menjadi sumber data untuk menilai opini public melalui Twitter.

Penelitian ini bertujuan untuk membangun aplikasi klasifikasi opini yang menerapkan pendekatan Naïve Bayes untuk mengklasifikasikan kata-kata dan difokuskan pada tweet dalam bahasa Indonesia. Aplikasi ini kemudian diterapkan untuk mengklasifikasikan opini publik pada Twitter terkait layanan pemerintah terhadap masyarakat, berdasarkan sentimen positif, negatif atau netral. Data latih diperoleh melalui aplikasi platform KNIME Analytic dan sumber text diperoleh dari akun Twitter Dinas pemerintah Kota Bandung. Proses klasifikasi dilakukan melalui serangkaian tahapan seperti preproses (case folding, parsing, dan transformasi) serta proses klasifikasi itu sendiri.

Kata Kunci: Twitter, KNIME, Naïve Bayes, Bahasa Indonesia.

Abstract

Twitter is a microblogging social media where the users can search for topics and discuss issues that recently happen. Some of short messages or tweets may contain opinions on products and services related to the community. It can be used as a source of data to assess the public opinion through Twitter.

The aim of this research is to build the application for opinion classification which implement Naïve Bayes approach to classify the words and focuses on tweet in Bahasa Indonesia. The application then implemented on classification of public opinion on Twitter according to government services to public, based on positive, negative, or neutral sentiment. Data trained is obtained using KNIME Analytics Platform, and the source of text is extracted from twitter account of government public services in the city of Bandung. The process is done through several stages such as preprocessing (case folding, parsing, and transformation) and classification process.

Keywords: Twitter, KNIME, Naïve Bayes, Indonesian.

1. PENDAHULUAN

Media sosial pada era globalisasi saat ini sudah sangat umum dan banyak digunakan untuk kepentingan masyarakat. Dalam implementasinya, media sosial lebih banyak digunakan untuk kegiatan jual beli, menyampaikan informasi, bahkan sebagai media untuk mengekspresikan diri. Pertumbuhan media sosial sangat cepat tidak hanya penggunaanya yang terus menerus meningkat, namun semakin banyaknya media sosial yang ditawarkan melalui aplikasi *mobile* ataupun *website*. Sehingga hal tersebut dimanfaatkan oleh Pemerintah kota Bandung untuk menyampaikan pesan dan informasi kepada masyarakat. Salah satu media sosial yang banyak digunakan adalah *Twitter*. *Twitter* merupakan suatu kumpulan kata yang berisikan maksimal 140 karakter. Didalam *text twitter*, masyarakat umum biasa menyebutkan dengan kata “*tweet*” atau kicauan. Kata yang terkandung dalam *twitter* adalah bahasa alami manusia yang merupakan bahasa dengan struktur kompleks.

Menurut Kamus Besar Bahasa Indonesia opini merupakan pendapat, pikiran, atau pendirian [1]. Semakin bertambah banyak kalimat kritik dan saran pada media sosial sehingga dapat membentuk opini masyarakat. Opini ini dapat dijadikan masukan terhadap penilaian kinerja layanan Pemerintahan dan Dinas di Kota Bandung terhadap masyarakat. Analisis sentimen atau *opinion mining* dapat digunakan untuk memperoleh gambaran

umum persepsi masyarakat terhadap kualitas layanan, apakah cenderung positif, negatif atau netral. Opini biasanya bernilai positif atau negatif tetapi dapat dikategorisasikan juga menjadi baik, sangat baik, buruk, dan sangat buruk [2].

Penelitian di bidang *sentiment analysis* atau *opinion mining* berbahasa Indonesia telah dilakukan oleh Yudi Wibisono [3]. Berdasarkan penelitian terkait dapat disimpulkan bahwa Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap nilai akurasi. Akurasi dihitung dengan menghitung persentase jumlah data latih dan data uji.

Penggunaan *tweet* berbahasa Indonesia terdapat salah satu masalah yaitu variasi sistem penulisan yang sangat tinggi, diantaranya banyak terdapat kalimat dengan kosa kata dan tata bahasa yang tidak baku. Sebagai contoh beberapa macam variasi penulisan kata tidak seperti "ndak", "nggak", "gak", "tdk" dan seterusnya masing masing memiliki arti yang sama [3]. Terbatasnya karakter dalam penulisan twitter membatasi pengguna agar dapat menulis secara singkat.

Sentiment pada twitter terhadap suatu topik tertentu dapat ditemukan salah satunya melalui *tweet* yang mengandung hastag dan pencarian menggunakan *acount (search)*. Hastag merupakan simbol pagar(#) yang digunakan untuk memberikan label dalam salah topik pembahasan seperti #bandung, #dinasbandung atau #kotabandung[2]. Sedangkan pencarian menggunakan *acount* merupakan *username* dengan simbol (@) yang ditulis di depan kata seperti @PemkotBandung, @Bandung_dinkes dan, @ridwankamil. Penelitian ini menggunakan metode *naive bayes classifier* untuk mengklasifikasi opini menjadi positif, negative dan netral. Metoda *naive bayes classifier* merupakan algoritma klasifikasi yang sederhana namun memiliki akurasi tinggi. Untuk itu perlu pembuktian dalam klasifikasi *tweet* untuk menghasilkan hasil yang optimal dibandingkan dengan metode lain.

2. SENTIMENT ANALYSIS

Analisa sentimen atau biasa disebut *opinion mining* merupakan salah satu cabang penelitian *Text Mining*. *Opinion mining* adalah riset komputasional dari opini, sentimen dan emosi yang diekspresikan secara tekstual. Jika diberikan suatu set dokumen teks yang berisi opini mengenai suatu objek, maka *opinion mining* bertujuan untuk mengekstrak atribut dan komponen dari objek yang telah dikomentasi pada setiap dokumen dan untuk menentukan apakah komentar tersebut bermakna positif atau negative [2]. *Sentiment Analysis* dapat dibedakan berdasarkan sumber datanya, beberapa level yang sering digunakan dalam penelitian *Sentiment Analysis* adalah *Sentiment Analysis* pada level dokumen dan *Sentiment Analysis* pada level kalimat [4]. Berdasarkan level sumber datanya *Sentiment Analysis* terbagi menjadi 2 kelompok besar yaitu [4]: *Coarse-grained Sentiment Analysis* dan *fined-grained Sentiment Analysis*

Pada *Sentiment Analysis Coarse-grained*, *Sentiment Analysis* yang dilakukan adalah pada level dokumen. Secara garis besar fokus utama dari *Sentiment Analysis* jenis ini adalah menganggap seluruh isi dokumen sebagai sebuah sentiment positif atau sentiment negatif. *Fined-grained Sentiment Analysis* adalah *Sentiment Analysis* pada level kalimat. Fokus utama *fined-greined Sentiment Analysis* adalah menentukan sentimen pada setiap kalimat.

2.1. Text Mining

Text mining dapat didefinisikan secara luas sebagai proses pengetahuan intensif dimana pengguna berinteraksi dengan koleksi dokumen dari waktu ke waktu dengan menggunakan seperangkat alat analisis. *Text mining* berusaha mengekstrak informasi yang berguna dari sumber data melalui identifikasi dan eksplorasi pola yang menarik. *Text mining* cenderung mengarah pada bidang penelitian *data mining*. Oleh karena itu, tidak mengherankan bahwa *text mining* dan *data mining* berada pada tingkat arsitektur yang sama [5]. Penambangan teks dapat dianggap sebagai proses dua tahap yang diawali dengan penerapan struktur terhadap sumber data teks dan dilanjutkan dengan ekstraksi informasi dan pengetahuan yang relevan dari data teks terstruktur ini dengan menggunakan teknik dan alat yang sama dengan penambangan data [5].

2.2. Text Preprocessing

Text Preprocessing merupakan tahapan dari proses awal terhadap teks untuk mempersiapkan teks menjadi data yang akan diolah lebih lanjut. Sebuah teks yang ada harus dipisahkan, hal ini dapat dilakukan dalam beberapa tingkatan yang berbeda. Suatu dokumen atau *tweet* dapat dipecah menjadi bab, sub-bab, paragraf, kalimat dan pada akhirnya menjadi potongan kata/token. Selain itu pada tahapan ini keberadaan digit angka, huruf kapital, atau karakter- karakter yang lainnya dihilangkan dan dirubah [5].

2.3. Naïve Bayes

Algoritma *naive bayes classifier* merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat [5]. Dalam penelitian ini yang menjadi data uji adalah dokumen *tweets*. 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 “ $X_1, X_2, X_3 \dots X_n$ ” dimana X_1 adalah kata pertama, X_2 adalah kata kedua dan seterusnya. Sedangkan V adalah himpunan kategori *Tweet*. Pada saat klasifikasi, algoritma akan mencari probabilitas tertinggi dari semua kategori dokumen yang diujikan (V_{map}), dimana persamaannya adalah sebagai berikut:

$$V_{\text{MAP}} = \underset{V_j \in V}{\arg \max} \frac{P(x_1, x_2, x_3, \dots, x_n | V_j) P(V_j)}{P(x_1, x_2, x_3, \dots, x_n)} \quad \dots\dots(1)$$

Untuk $P(X_1, X_2, X_3 \dots X_n)$ nilainya konstan untuk semua kategori (V_j) sehingga persamaan di atas dapat ditulis sebagai berikut:

$$V_{\text{MAP}} = \underset{V_j \in V}{\arg \max} P(x_1, x_2, x_3, \dots, x_n | V_j) P(V_j) \quad \dots\dots(2)$$

Persamaan diatas dapat disederhanakan menjadi sebagai berikut:

$$V_{\text{MAP}} = \underset{V_j \in V}{\arg \max} \prod_{i=1}^n P(x_i | V_j) P(V_j) \quad \dots\dots(3)$$

Keterangan:

V_j = kategori tweet $j = 1, 2, 3, \dots, n$ Dimana dalam penelitian ini

j_1 = kategori tweet sentimen negatif,

j_2 = kategori tweet sentimen positif,

j_3 = kategori tweet sentiment netral.

$P(X_i | V_j)$ = Probabilitas X_i pada kategori V_j

$P(V_j)$ = Probabilitas dari V_j

Untuk $P(V_j)$ dan $P(X_i | V_j)$ dihitung pada saat pelatihan dimana persamaannya adalah sebagai berikut :

$$P(V_j) = \frac{|docs\ j|}{|contoh|} \quad \dots\dots(4)$$

$$P(x_i | V_j) = \frac{n_k + 1}{n + |kosakata|} \quad \dots\dots(5)$$

Keterangan:

$|docs\ j|$ = jumlah dokumen setiap kategori j

$|contoh|$ = jumlah dokumen dari semua kategori

n_k = jumlah frekuensi kemunculan setiap kata

n = jumlah frekuensi kemunculan kata dari setiap kategori

$|kosakata|$ = jumlah semua kata dari semua kategori.

2.4. KNIME Analytics Platform

Pengembangan *KNIME* dimulai Januari 2004 oleh *Developer* perangkat lunak di *Konstanz University* sebagai produk *proprietary*. Tim pengembang asli yang dipimpin oleh *Michael Berthold* berasal dari sebuah perusahaan di *Silicon Valley* menyediakan perangkat lunak untuk industri farmasi. *KNIME* telah dikembangkan sejak awal melalui proses rekayasa perangkat lunak yang ketat, dan dari awal ditujukan untuk digunakan pada perusahaan-perusahaan skala besar. Tujuan awalnya adalah untuk menciptakan sebuah platform pengolahan data modular, sangat sederhana dan *opensource* yang memungkinkan untuk integrasi yang dapat digunakan pada banyak platform seperti *windows, linux, Mac OS*. Platform ini dimaksudkan sebagai media kolaborasi dan penelitian. Platform juga dapat berperan sebagai platform integrasi untuk berbagai proyek analisis data lainnya [6].

KNIME Analytics memungkinkan pengguna melakukan visualisasi aliran data masuk (*pipelines*), menjalankan tahapan analisis tertentu atau semua tahapan, dan melihat hasil dari pemodelan secara interaktif. *KNIME* ditulis dengan bahasa Java (dengan platform *eclipse*) dan dilengkapi dengan mekanisme plugin untuk kemudahan pengembangan dan penambahan fungsionalitas. Versi utamanya sudah meliputi ratusan modul integrasi data (file I/O, dukungan terhadap DBMS umum), transformasi data (*filter, converter, combiner*), dan metoda umum yang digunakan untuk analisis dan visualisasi data [6].

4. SENTIMEN ANALYSIS TWEET DINAS KOTA BANDUNG.

Permasalahan yang mendasari penelitian ini adalah sebagai berikut:

1. Pada umumnya masyarakat Indonesia memberikan keluhan, kritik maupun saran menggunakan *media sosial*, karena penggunaan *media sosial* khususnya *twitter* dapat digunakan dimana saja dan kapan saja.

Keluhan banyak terjadi apabila masyarakat merasa kurang nyaman atas kinerja sebuah perusahaan atau penyedia layanan barang dan jasa.

2. Sulitnya mengklasifikasi *tweet* masyarakat pada dinas kota Bandung sehingga membutuhkan aplikasi untuk mengetahui keluhan yang terdapat di masyarakat.
3. Menguji hasil keakuratan metode *naïve bayes* untuk penelitian opini masyarakat terhadap kinerja layanan dinas di kota Bandung, yang disampaikan melalui Twitter.

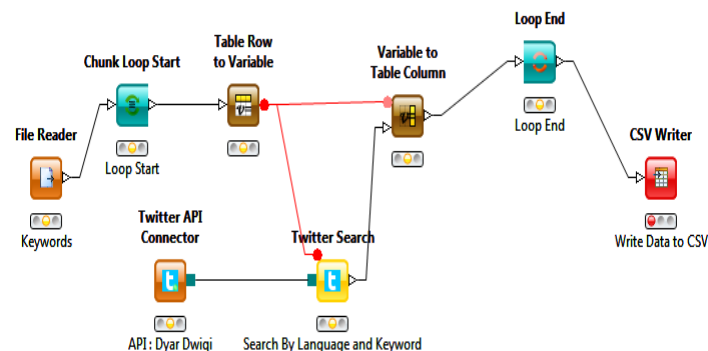
Proses sentimen analysis dilakukan dengan dua tahap, yaitu persiapan data dan proses text mining.

4.1. Persiapan Data

Secara garis besar tahap persiapan data dibagi menjadi kelompok yaitu *data understanding* dan *data preparation*. *Data understanding* meliputi proses pengumpulan data dan pembersihan data (*data cleansing*). *Data preparation* adalah penyiapan data hasil proses sebelumnya (data yang sudah dibersihkan) agar bebas dari *noise*. Data yang sudah siap selanjutnya dijadikan masukan pada proses text mining dan akan digunakan sebagai data training dan data testing

Data understanding dibagi lagi menjadi dua tahapan yaitu pengumpulan data dan pembersihan data sebagai berikut:

1. Proses pengumpulan data Tweet dilakukan dengan menggunakan *KNIME Analytics Platform 2.11.0*. Melalui aplikasi ini dapat dikumpulkan data *tweet* seperti *id*, *username*, *tweet*, *time*, *favorited*, *retweeted*, *retweet from* untuk dijadikan data yang akan diolah. *Workflow* proses pengambilan data menggunakan *KNIME Analytics Platform 2.11.0*, dapat dilihat pada Gambar 1.



Gambar 1 *Workflow* Proses Pengambilan *Tweet* menggunakan *Workflow*

Sumber data awal pada penelitian ini adalah tweet yang berasal dari akun twitter dinas/pemerintahan di kota Bandung.

2. *Tweet Cleansing*: merupakan proses memilah kalimat-kalimat *tweet* yang bertujuan mengurangi kalimat *noise* sehingga didapat kualitas data yang dibutuhkan. Kalimat yang disaring dalam proses *cleansing* yaitu *tweet* yang tidak mengandung kalimat *sentiment*.

Hasil tahap data preparation adalah terkumpulnya data tweet yang sudah siap diolah lebih lanjut yaitu tahapan data preparation. Pada tahapan ini dilakukan serangkaian pengolahan text dan *feature selection* yang meliputi langkah-langkah sebagai berikut :

1. *Text Preprocessing*: mengubah semua karakter menjadi huruf kecil, menghapus URL, mention dan hastag, dan delimiter.
2. *Case Folding*: mengubah semua huruf dalam dokumen menjadi huruf kecil [7].
3. *Tokenizing*: pemotongan string input berdasarkan tiap kata yang menyusunnya, sehingga jika kalimat : "telkomcare trm kasih kami mohon agar setelah selesai pekerjaan....", menjadi : telkomcare, trm, kasih, kami, mohon, agar, setelah
4. *Stopword R removal* atau *filtering* : mengambil kata-kata penting dari hasil token, dengan menggunakan algoritma stoplist (membuang kata yang kurang penting) atau wordlist (menyimpan kata penting). Stoplist / stopword adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan bag-of-words[7]. .
5. *Stemming* : mencari root kata dari tiap kata hasil filtering. Pada tahap ini dilakukan proses pengembalian berbagai bentukan kata ke dalam suatu representasi yang sama. Contoh hasil stemming misalnya kata "mhn" menjadi "mohon", "pekerjaan" menjadi "kerja".

4.2. Proses Perhitungan Text Mining

Setelah dilakukan *data preparation*, selanjutnya melakukan proses text mining itu sendiri. Proses ini terbagi menjadi tiga tahap yaitu: pengolahan teks (*text processing*, pada tahapan *text mining*), transformasi teks kedalam bentuk perantara (*text transformation / future generation*), dan penemuan pola (*pattern discovery*). Sistem akan

melakukan pengolahan terhadap data masukan berupa data teks dari *tweet* dan menghasilkan keluaran berupa pola sebagai hasil interpretasi. Berikut contoh perhitungan yang diambil dari satu tweet inputan sebagai berikut: “Ketika hujan deras turun Kamis sore, pertigaan Katamso-Pahlawan air menggenang ckp lumayan. Tanda drainase kurang berfungsi @dbmpkotabdg”

Hasil pengolahan text diatas dapat dilihat pada tabel 1. Contoh daftar kata yang berhasil disaring dari tweet inputan di atas dapat dilihat pada kolom paling kanan.

Tabel 1. Contoh Daftar Kata Hasil Penyaringan Tweet Inputan

Tweet (matriks)	Stoplist	stemming	Kata dasar	Sort kata	Rk (menghitung jml hasil)	Hasil
Ketika	DELETE	DELETE	DELETE	air	air	1
hujan	hujan	hujan	hujan	cukup	cukup	1
deras	deras	deras	deras	deras	deras	1
.....
sore	sore	sore	sore	hujan	hujan	1
pertigaan	pertigaan	pertigaan	DELETE	Kamis	Kamis	1
.....

Tahap selanjutnya yaitu melakukan proses perhitungan menggunakan metode *Naïve Bayes* dengan menggunakan data *tweet* diambil secara acak. Data dibagi menjadi 2 bagian yaitu data *training* yang terdiri dari 300 kata yang telah diketahui kategorinya masing masing terdiri dari 100 kata negatif, 100 kata positif, dan 100 kata netral. Data testing merupakan sebuah *tweet* yang belum diketahui kategorinya.

Berikutnya akan diberikan contoh penggunaan algoritma *Naïve bayes* untuk klasifikasi sentiment. Misalkan untuk tweet berikut [“Ketika hujan deras turun Kamis sore, pertigaan Katamso- Pahlawan air menggenang ckp lumayan.Tanda drainase kurang berfungsi @dbmpkotabdg”], yang sudah melalui proses *text mining* sehingga mendapatkan hasil *term frequency* (TF). Hasil perhitungan *Naïve Bayes* untuk *tweet* tersebut seperti terlihat pada Tabel 2.

Tabel 2. Contoh Hasil Perhitungan *Naïve Bayes*

No	Term Frekuensi (TF)	Training Neg	Training Pos	Training Net
1	air	1	1	1
2	cukup	1	1	1
4	drainase	2	1	1
5
10	turun	1	1	1

Dari Tabel 2 diketahui: Jumlah Term Frekuensi (TF) keseluruhan (n) = 10, Jumlah Kata Training = 100. Dari nilai-nilai tersebut, dapat dicari nilai nilai probabilitas keyword dengan menggunakan rumus $P(X_i|V_j)$ yaitu:

$$P(\text{drainase}|\text{negatif}) = (1+1)/(10+100) = 0.01$$

$$P(\text{n}|\text{positif}) = (1+1)/(10+100) = 0.01;$$

$$P(\text{n}|\text{netral}) = (1+1)/(10+100) = 0.01$$

Hasil perhitungan untuk setiap term frekuensi dapat dilihat pada tabel 3.

Tabel 3. Perhitungan probabilitas Tweet 1

No	Term Frekuensi (TF)	Training Neg	Probabilitas
1	Air	1	$(1+1)/(10+100) = 0.01$
2	Cukup	1	$(1+1)/(10+100) = 0.01$
5
10	Turun	1	$(1+1)/(10+100) = 0.01$

Berdasarkan nilai probabilitas pada tabel dapat dihitung:

$$= \arg \max_{V_j \in V} \prod_{i=1}^n P(x_i|V_j)P(V_j) = \arg \max_{(\text{positif})} \prod_{i=1}^{10} P(x_i|V_j)P(V_j)$$

$$= (0,01) (0,01) (0,01) (0,02) (0,01) (0,01) (0,01) (0,01) (0,01) (0,01) \times 100 = 2.0\text{E-}200$$

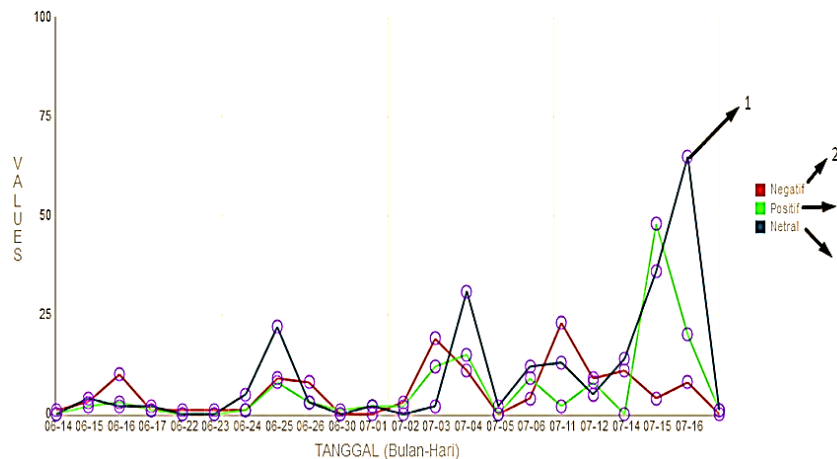
Hasil akhir perhitungan *Naïve bayes* untuk contoh tweet di atas dapat dilihat pada tabel 4.

Tabel 4. Hasil Perhitungan *Naïve Bayes*

No	Tweet	Hasil Probabilitas			Status
		term 1 negatif	Term 2 positif	term 3 netral	
1	Ketika hujan deras turun Kamis sore, pertigaan Katamso- Pahlawan air menggenang ckp lumayan.Tanda drainase kurang berfungsi @dbmpkotabdg	2.0E-200	1.0E-200	1.0E-200	Negatif

Perhitungan pada tabel 4 memperlihatkan proses penentuan klasifikasi sentimen didasari oleh hasil perhitungan probabilitas dengan nilai tertinggi yaitu 2.0E-200, sehingga *tweet* diatas dikategorikan sebagai *tweet* “negatif”.

Langkah-langkah perhitungan di atas kemudian diterapkan pada pengembangan aplikasi sentiment analysis sederhana berbasis web. Aplikasi ini membantu pelaksanaan langkah-langkah di atas menjadi lebih mudah dan otomatis. Gambar 2 menampilkan hasil sentimen analysis yang didapat dari eksekusi aplikasi ini pada data tweet yang dikumpulkan selama 1 bulan yaitu Juni-Juli 2014. Pada gambar 2 terlihat bahwa perbandingan antara opini negatif dan positif relatif berimbang dari hari ke hari, kecuali di akhir periode pengamatan yang didominasi oleh opini positif terhadap dinas/pemerintahan kota Bandung.



Gambar 2. Hasil Sentimen Analysis dalam Bentuk Grafik

Pengujian terhadap keakuratan klasifikasi sentimen yang dihasilkan oleh aplikasi dilakukan dengan membandingkan hasil klasifikasi sentimen dengan data testing berupa tweet yang diambil secara acak selama bulan Januari dan Maret 2015, masing-masing sebanyak 300 tweet. Tweet ini dikategorikan menjadi 3 kelompok (secara manual) yaitu tweet negatif, positif dan netral. Hasil pengujian sebanyak tiga kali menunjukkan bahwa persentase data yang sesuai sebanyak 73%, 68% dan 58%. Keakuratan hasil pengujian juga sangat bergantung pada jumlah data training yang dientrikan ke sistem.

5. KESIMPULAN

Berdasarkan uraian di atas, dapat disimpulkan hal-hal sebagai berikut:

1. Metoda *Naïve Bayes Classifier* dapat diterapkan sebagai metoda untuk melakukan klasifikasi *sentiment analysis*.
 2. Pengumpulan text berupa tweet dari akun-akun target dapat lebih mudah dan praktis dilakukan dengan menggunakan aplikasi KNIME.
 3. Aplikasi Sentimen Analysis yang dikembangkan dianggap cukup memadai, dikarenakan hasil uji akurasi klasifikasi yang dilakukan oleh aplikasi, pada sekumpulan data testing dan dilakukan dengan 3 variasi scenario, menghasilkan nilai akurasi tertinggi sebesar 73% dengan jumlah kata training 100 negatif, 100 positif, dan 100 netral.
 4. Proses klasifikasi akan semakin akurat apabila jumlah *keyword training* yang digunakan semakin banyak.
- Saran pengembangan penelitian ini yaitu menerapkan aplikasi untuk melakukan text mining pada sumber data dengan rentang waktu yang lebih lama (misalnya 3 bulan), serta melakukan pengembangan sistem untuk beberapa fitur tambahan seperti pembuatan laporan per periode dan penyajian grafik yang lebih interaktif.

5. DAFTAR RUJUKAN

- [1] ---, 2008, Kamus Besar Bahasa Indonesia, Jakarta: Pusat Bahasa Departemen Pendidikan Nasional, 2008.
- [2] Shelby.M.I., Warih.M., Adiwijaya, 2013, *Opinion Mining Pada Twitter Menggunakan Klasifikasi Sentimen pada Hastag berbasis Graf*
- [3] Harlili., Wibisono, Yudi, 2013, "*Sistem Analisis Opini Microblogging Berbahasa Indonesia.*" UPI Bandung. 2013
- [4] Clayton R. Fink, 2011, *Coarse- and Fine-Grained Sentiment Analysis of Social Media Text*. Johns hopkins apl technical digest, volume 30, number 1.
- [5] Feldman, Ronen dan Sanger, James. 2007. *The Text Mining Handbook Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, New York.
- [6] Bakos, Gabor 2013. *Perform Accurate Data Analysis Using the Power of KNIME*. KNIME Essentials. Birmingham, Mumbai.