

Pengukuran Indeks Kebahagiaan Masyarakat Kota Bandung dari Jejaring Sosial Twitter Menggunakan Ontologi dengan Paradigma *Bottom-up*

Measuring Bandung Citizen's Happiness Index From Social Media Twitter Using Bottom-Up Paradigm Ontology

Anisa Herdiani¹, Nungki Selviandro², Dimas Ikhsanu Fitra³
123Prodi S1 Teknik Informatika, Fakultas Informatika, Universitas Telkom, Bandung
Jalan Telekomunikasi No. 1, Dayeuh Kolot, Bandung 40257
1aniseherdiani@gmail.com, 2nselviandro@gmail.com, 3dimas_if@yahoo.co.id

Abstrak

Kebahagiaan masyarakat dapat dijadikan parameter untuk menentukan pencapaian dari suatu negara. Dalam cakupan yang lebih kecil, hal ini dapat diaplikasikan juga ke sebuah kota. Pemerintah kota Bandung mengembangkan sistem yang dapat menghitung indeks kebahagiaan masyarakatnya. Pemerintah kota Bandung telah melakukan perhitungan indeks kebahagiaan dengan survey *sampling* secara acak kepada kepala keluarga atau pasangannya. Dalam pengambilan data dengan survey memerlukan usaha serta biaya yang besar. Dapat dilakukan upaya lain selain dari menggunakan survey dalam pengambilan data, yaitu dengan mengambil data berupa opini dari masyarakat dari media sosial. Pada penelitian ini dibuat suatu sistem yang mampu mengolah data berupa opini yang berasal dari media sosial Twitter yang sangat banyak digunakan masyarakat kota Bandung. Data *tweet* pada Twitter diklasifikasikan sesuai dengan kebutuhan perhitungan indeks kebahagiaan masyarakat kota Bandung. Klasifikasi dibantu ontologi yang dapat memetakan pengetahuan yang diketahui manusia untuk diketahui sistem. Ontologi dibangun dengan paradigma *bottom-up* yang pembangunannya bermula dari *term-term* pada *tweet*. Untuk mendapatkan sentimen dari data *tweet* digunakan klasifikasi sentimen menggunakan *classifier* pada Weka. Hasil klasifikasi ontologi dan sentiment digunakan untuk mengukur indeks kebahagiaan. Dilakukan pengujian terhadap sistem dengan mencari nilai F1 *measure* dan akurasi. Dari hasil pengujian ditemukan bahwa kelengkapan ontologi berpengaruh kepada performansi hasil klasifikasi yang dianalisis dari nilai akurasi dan F1 *measure* pada dua versi ontologi. Dengan data yang ada, jumlah data latih berpengaruh terhadap performansi *classifier* Weka dalam melakukan klasifikasi sentimen. Semakin besar jumlah data maka semakin besar nilai akurasi dan nilai F1 *measure*. Semakin besar akurasi maka semakin valid hasil perhitungan indeks kebahagiaan yang didapat.

Kata kunci: indeks kebahagiaan, pengambilan data, ontologi, twitter, analisis sentimen

Abstract

Citizen's happiness could be parameter for knowing how far some nation are growing. In smaller scope, this thing can be applied to the city. Bandung City's government are developing some system for measure their citizen's happiness index. They had measure the happiness index by random sampling survey to their citizen. In the mining data survey they must using big effort and big budget. We can mine some data by another way by mining data from social media that filled by opinion of the user. In this research we make system that can process opinion from Twitter that Bandung's citizen mostly use that represented by tweet. Tweet from Twitter classified by the system depend on Bandung Citizen's happiness index parameter. Classifying process helped by ontology that can mapping knowledge that human know to understood by system. Ontology construct by using bottom-up paradigm that start from tweet's terms. For getting sentiment of the tweet we use sentiment classifying by Weka classifier. Result from ontology classifying and sentiment classifying used for measuring happiness index. We must test the system's performance by finding accuracy and F1 measure from two version of ontology. From result of testing that measure accuracy and F1 measure we can learnt that completing ontology can make the performance of system increasing. From testing process of sentiment analysis we can learnt that total train data are affect the performance of sentiment analysis. More train data are good.

Keywords: happiness index, ontology, twitter, sentiment analysis, data mining

1. Pendahuluan

Kebahagiaan merupakan suatu hal penting yang dapat dijadikan suatu parameter untuk mengetahui pencapaian dari sebuah kota bahkan negara. Dengan mengukur kebahagiaan masyarakat suatu wilayah, kita dapat melakukan evaluasi dengan lebih efisien dan spesifik akan kinerja maupun dalam pembuatan kebijakan publik. Pemerintah Kota Bandung saat ini sedang bergerak pesat dalam menjalankan program-programnya yang bertujuan untuk memajukan kota Bandung. Untuk mengukur pencapaian yang telah dilakukan diadakan perhitungan indeks kebahagiaan. Pada tahun 2015, pemerintah kota Bandung telah melakukan perhitungan indeks kebahagiaan masyarakatnya yang dilakukan dengan cara melakukan survey ke 151 kelurahan di kota Bandung untuk mewakili seluruh wilayah kota Bandung. Survey dilakukan dengan melakukan *sampling* secara acak di setiap wilayah kepada kepala keluarga atau pasangannya [1]. Pengambilan data dengan metodologi survey membutuhkan usaha serta biaya yang besar dalam penyelenggaraannya. Salah satu cara yang minim usaha serta biaya adalah mendapatkan data dari media sosial yang sedang trend dalam masyarakat. Dengan ini kita mendapatkan data tanpa harus menanyakannya kepada masyarakat secara langsung. Kota Bandung menduduki peringkat ke-6 pengguna twitter terbanyak di seluruh dunia [2].

Dalam pengambilan data tersebut dapat digunakan suatu ontologi yang dirancang untuk menyaring setiap *tweet* yang ada agar dapat diambil data-data relevan sebagai bahan perhitungan indeks kebahagiaan. Ontologi sendiri merupakan suatu spesifikasi formal dan eksplisit dari sebuah konseptual [3]. Dari spesifikasi yang ada dari sebuah konseptual, dapat dipetakan kata-kata yang memiliki keterhubungan dengan konsep tersebut yang akan memudahkan pengambilan data yang relevan dari data yang sangat besar. Salah satu fungsi ontologi juga merupakan sebagai jembatan antara manusia dengan mesin dimana ontologi menghubungkan simbol-simbol yang dipahami manusia dan dirubah ke bentuk yang dapat dipahami oleh mesin [4] [5]. Dengan ontologi ini dapat dilakukan klasifikasi terhadap data *tweet* berdasarkan aspek atau bidang tertentu yang mempengaruhi kebahagiaan masyarakat kota Bandung. Ada beberapa paradigma dalam pembentukan suatu ontologi, yaitu *top-down*, *bottom-up*, dan kombinasi keduanya. Dalam perancangan ontologi, dapat terjadi ketidaksesuaian antara *vocabulary* pada *term-term* ontologi dengan data yang ada. Untuk mencegah hal tersebut maka paradigma pembentukan ontologi dengan *bottom-up* sangat cocok karena perancangan ontologi dilakukan dari *term-term* paling spesifik yang berasal dari dataset untuk kemudian dilakukan generalisasi ke arah yang lebih umum seperti *clustering*. Diharapkan dengan perancangan ontologi menggunakan paradigma *bottom-up* dapat ditemukan ontologi yang tepat untuk membantu perhitungan indeks kebahagiaan masyarakat kota Bandung. Selain ontologi, diperlukan pula suatu metode untuk memisahkan data yang mengarah ke penilaian positif ataupun negatif. Analisis sentimen merupakan salah satu metode yang dapat mengatasi permasalahan tersebut. Karena memang analisis sentimen merupakan suatu metode yang dapat mencari polaritas dari suatu kata ataupun frase [6] [7]. Dalam penelitian ini perhitungan indeks kebahagiaan akan dibantu dengan analisis sentimen memanfaatkan *classifier* pada Weka.

2. Dasar Teori dan Perancangan Sistem

2.1 Dataset

Dalam penelitian yang dilakukan terdapat dataset yang digunakan yang berupa data *tweet* dan ontologi yang dibangun dengan paradigma *bottom-up*.

2.1.1 Tweet

Dataset yang digunakan dalam penelitian ini adalah data *tweet* yang berasal dari kota Bandung yang diambil pada bulan Februari dan Maret 2016 dengan jumlah sebanyak 10.636 *tweet* yang memiliki karakteristik sebagai berikut:

1. *Tweet* diambil menggunakan API Twitter dengan *geocode* sebagai parameter pengambilan data.
2. Penulisan kalimat pada *tweet* cenderung tidak terstruktur
3. Kata pada *tweet* berasal dari aneka ragam bahasa.
4. Banyak kata yang tidak sesuai dengan kaidah penulisan yang tepat.

Data *tweet* yang didapatkan berjumlah 10.636 data. Dari ontologi versi pertama *tweet* yang terklasifikasi berjumlah 1.208. Dari angka tersebut dijadikan jumlah acuan untuk perhitungan akurasi klasifikasi ontologi. *Tweet* yang terklasifikasi oleh ontologi versi pertama diberikan label manual oleh peneliti. Contoh data *tweet* seperti pada Tabel 1.

Tabel 1: Contoh Data Tweet

No	Tweet
1	iya saya jadi pacar saja sudah siaga siap antar jaga apalagi sudah jadi suami
2	lagi lagi lagi banyak kesulitan bukannya mengurangi beban pikiran malah semakin membuat susah tidur ya allah mohon diberi kesabaran hambamu ini
3	susah tidak apa asal setia tapi orang susah sekarang juga susah setia apalagi kaya raya
4	ya allah berikanlah sisa umur yang bermanfaat aamiin

2.1.2 Ontologi

Ontologi pada pengujian memiliki dua versi dengan masing-masing versi memiliki perbedaan pada kelengkapan isi ontologi. Pada ontologi versi pertama *instance* yang dimiliki hanya 170 *term*. Sedangkan pada ontologi versi kedua memiliki *instance* sejumlah 1245 *term*. Kedua ontologi ini memiliki hierarki kelas yang sama dengan hanya berbeda pada jumlah *instance* pada masing-masing ontologi.

2.2 Pembuatan Ontologi

Dalam membangun ontologi, banyak cara yang dapat digunakan. Seperti dalam [8] dinyatakan bahwa tidak ada cara atau metodologi yang benar-benar tepat dalam membangun ontologi. Namun dari banyak cara yang pernah digunakan memiliki garis besar yang sama dalam pembuatan ontologi ini. Pada [8] juga dijelaskan bahwa dalam membangun ontologi pendekatan yang dilakukan adalah dengan *iterative* atau secara berulang-ulang. Maksudnya adalah dalam pembangunan ontologi sangat memungkinkan pengembangan yang terus berlanjut terhadap ontologi yang ada, tidak semata-mata jadi dan dianggap selesai. Noy dan McGuinness dalam [8] mengungkapkan step-step yang dilewati dalam pembangunan ontologi antara lain:

1. Menentukan Domain dan lingkup ontologi
2. Menentukan apakah akan menggunakan ontologi yang sudah ada atau tidak
3. Mendapatkan *term-term* yang penting terhadap ontologi
4. Mendefinisikan kelas dan hierarki kelasnya
5. Mendefinisikan properti dari kelas yang ada
6. Mendefinisikan aspek
7. Membuat *instance*

2.3 Paradigma *bottom-up*

Proses dalam perancangan ontologi dengan paradig *bottom-up* dimulai dengan cara mendefinisikan kelas yang paling spesifik, dari kelas-kelas yang ada dilakukan pengelompokan ke konsep yang lebih general [8]. Dalam membentuk kelas dari yang paling spesifik, dapat dibantu dengan adanya *term-term* yang didapat dari kumpulan kata, dalam penelitian ini didapatkan dari setiap *tweet*. Dari *term-term* tersebut dicari makna serta dikelompokkan berdasarkan makna yang didapat untuk menjadi sebuah kelas. *Term* yang ada juga dianalisis apakah dapat menjadi kelas atau menjadi *instance*. Seperti pada [9], ide utama paradigma *bottom-up* adalah membangun ontologi dari *instance* menjadi hierarki kelas dari kelas yang spesifik ke umum. Adapun informasi dasar didapatkan bukan dari konsep melainkan dari pelajaran yang diambil dari kehidupan manusia pembuat ontologi tersebut. Hal ini membuat paradigma *bottom-up* merupakan pendekatan yang paling menggunakan nalar dalam pembuatan konsepnya.

2.4 Evaluasi Klasifikasi

Untuk menguji sistem dilakukan perhitungan untuk mendapatkan akurasi dari performansi klasifikasi yang telah dilakukan sistem. Dalam evaluasi klasifikasi ada beberapa elemen yang menjadi bahan perhitungan, yaitu:

1. True Positive (TP) yang merupakan kondisi saat kelas diprediksi *positive* dan faktanya *true*
2. False Positive (FP) yang merupakan kondisi saat kelas diprediksi *positive* dan faktanya *false*
3. True Negative (TN) yang merupakan kondisi saat kelas diprediksi *negative* dan faktanya *true*
4. False Negative (FN) yang merupakan kondisi saat kelas diprediksi *negative* dan faktanya *false*

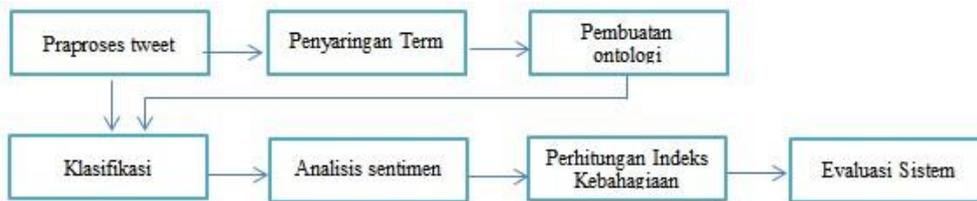
Dari elemen-elemen tersebut dapat didapatkan nilai-nilai *precision* yang merupakan nilai ketepatan sistem dalam perolehan *tweet* dengan kelas relevan pada pencarian *tweet* yang diterima dan *recall* yang merupakan ketepatan sistem dalam memperoleh *tweet* relevan pada pencarian seluruh *tweet* relevan yang ada. [10]

$$\begin{aligned} & \text{-----} & (1) \\ & \text{-----} & (2) \\ & \text{-----} & (3) \end{aligned}$$

Dari nilai *precision* dan *recall* dapat dicari nilai *F1 measure* dimana *F1 measure* merupakan kondisi dimana nilai *precision* dan *recall* memiliki *breakeven point* [11].

2.5 Perancangan Sistem

Dalam penelitian ini dibangun sistem yang dapat memanfaatkan ontologi untuk dijadikan parameter-parameter klasifikasi terhadap data *tweet*.



Gambar 1: Gambaran Umum Sistem

Berikut langkah-langkah sistem menurut gambar 1, sebagai berikut:

1. Praproses data *tweet* seperti *tokenization*, *stemming* dan *stopword removal*.
2. Penyaringan *term* pada data *tweet* untuk bahan pembentukan ontologi.
3. Menerima input ontologi
4. Klasifikasi *tweet* berdasarkan ontologi yang di-inputkan.
5. Analisis sentimen pada *tweet* yang terklasifikasi
6. Melakukan perhitungan indeks kebahagiaan
7. Evaluasi klasifikasi dengan *F1 measure*

3. Pembahasan

Dalam penelitian yang dilakukan, terdapat dua jenis klasifikasi, yaitu klasifikasi ontologi dan klasifikasi sentimen. Pada kedua klasifikasi menghasilkan kelas-kelas terhadap data *tweet*.

3.1 Hasil Klasifikasi Ontologi

Proses klasifikasi pada sistem menghasilkan *tweet-tweet* yang memiliki kelas berdasarkan kelas paling umum pada ontologi. Contoh *tweet-tweet* beserta kelas hasil klasifikasi ontologi seperti pada Tabel 2.

Tabel 2: Contoh Hasil Klasifikasi Ontologi

<i>Tweet</i>	Kelas
iya pacar siaga jaga suami	Hubungan_sosial
sulit beban pikir susah tidur al sabar hamba	Kesehatan
susah setia susah susah setia kaya raya	Hubungan_sosial
al berik sisa umur manfaat aamiin	Agama

3.2 Hasil Klasifikasi Sentimen

Proses analisis sentimen menghasilkan *tweet-tweet* yang telah memiliki kelas berdasarkan klasifikasi ontologi juga memiliki kelas sentimen. Contoh *tweet-tweet* yang telah melalui analisis sentimen seperti pada Tabel 3.

Tabel 3: Contoh Hasil Proses Analisis Sentimen

<i>Tweet</i>	Kelas	Sentimen
iya pacar siaga jaga suami	Hubungan_sosial	Positif
sulit beban pikir susah tidur al sabar hamba	Kesehatan	Negatif
susah setia susah susah setia kaya raya	Hubungan_sosial	Positif
al berik sisa umur manfaat aamiin	Agama	Positif

3.3 Hasil Evaluasi Klasifikasi

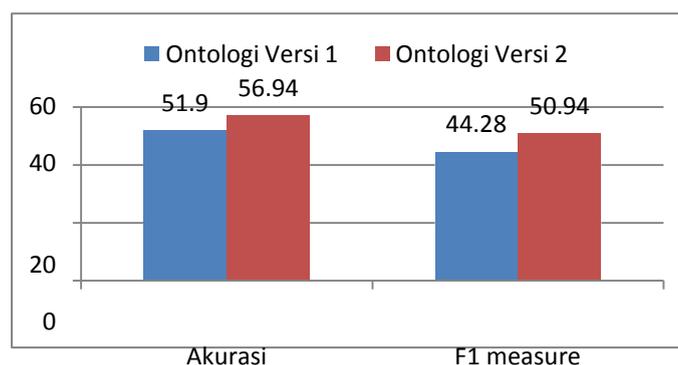
Berdasarkan pengujian terhadap sistem, didapatkan nilai akurasi serta F1 *measure* seperti pada Tabel 4.

Tabel 4: Performansi Sistem

No	Pengujian	Akurasi	F1 <i>measure</i>
1.	Ontologi versi 1	51.90%	44,28%
2.	Ontologi versi 2	56,94 %	50,94%
3.	Sentimen versi 1	55,3%	40%
4.	Sentimen versi 2	52,6%	37%
5.	Sentimen versi 3	47%	31,2%

3.4 Analisis Evaluasi Klasifikasi Ontologi

Dalam pengujian terhadap dua versi ontologi, didapatkan nilai akurasi dan F1 *measure* sebagai pembandingan seperti pada Gambar 2.

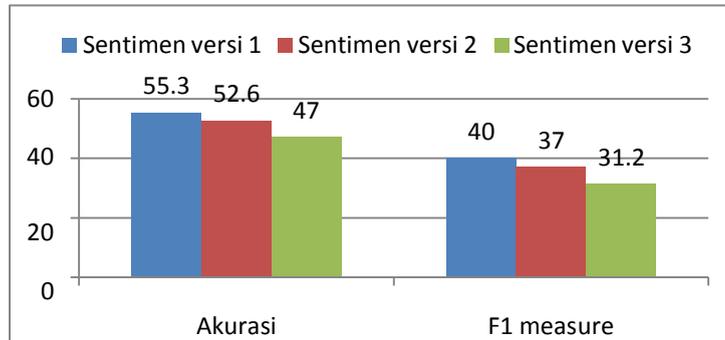


Gambar 2: Diagram Hasil Evaluasi Klasifikasi Ontologi

Berdasarkan hasil pengujian terhadap dua versi ontologi, didapatkan nilai akurasi dan nilai F1 *measure* yang lebih besar pada ontologi versi kedua dimana ontologi versi kedua lebih banyak jumlah *instance* yang dimiliki atau lebih lengkap ontologi yang dibuat. Jadi dapat disimpulkan ontologi yang lebih lengkap akan memberikan hasil klasifikasi yang lebih baik.

3.5 Analisis Evaluasi Klasifikasi Sentimen

Dalam pengujian terhadap *classifier* SVM didapatkan nilai *F1 measure* sebagai pembanding seperti pada gambar 3.



Gambar 3: Diagram Hasil Evaluasi Klasifikasi Sentimen

Dari Gambar 3 terlihat bahwa nilai akurasi serta *F1 measure* berbanding lurus dengan jumlah data yang digunakan. Semakin besar jumlah data maka akurasi dan *F1 measure* juga semakin besar. Jadi jumlah data sangat berpengaruh terhadap performansi dari klasifikasi sentimen, selain itu kualitas data latih ataupun label manual juga sangat berpengaruh.

4. Kesimpulan

Berdasarkan hasil evaluasi sistem, maka dapat ditarik kesimpulan berupa:

1. Kelengkapan ontologi mempengaruhi performansi klasifikasi ontologi berdasarkan nilai akurasi serta *F1 measure* yang dikeluarkan sistem. Semakin lengkap ontologi maka semakin baik performansi
2. Jumlah data latih berpengaruh pada performansi analisis sentimen dengan *classifier* Weka. Semakin banyak data latih maka semakin baik performansi analisis sentimen dengan *classifier* Weka.

Daftar Pustaka

- [1] Pemerintah Kota Bandung, "Indeks Kebahagiaan Kota Bandung Tahun 2015," Bandung, 2015.
- [2] Reza Wahyudi. (2012, July) <http://tekno.kompas.com>. [Online].
<http://tekno.kompas.com/read/2012/07/31/17362175/Jakarta.Juara.Tweet.Sedunia.Bandung.Peringkat.6>
- [3] Grigoris Antoniu and Frank van Harmelen, *A Semantic Web Primer.*, 2003.
- [4] M.A. Casanova, W.Truszkowski K. Breitman, *Semantic Web: Concepts, Technologies and Applications.*, 2007.
- [5] Profesor Dieter Fensel and Professor Frank van Harmelen Dr John Davies, *Towards The Semantic Web.*, 2003.
- [6] Vasileios Hatzivassiloglou and Kathleen R. McKeown, "Predicting the Semantic Orientation of Adjectives," 1997.
- [7] Andrea Esuli and Fabrizio Sebastiani, "SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining".

- [8] F. Noy Natalya and L. McGuinness Deborah, "Ontology Development 101: A Guide to Creating Your First Ontology".
- [9] Cheng-hsi Wei Charles, "Bottom-Up Ontology Creation with a Direct Instance Input Interface," California Polytechnic State University, PhD Thesis.
- [10] Torgo Luis and Ribeiro Rita, *Discovery Science*.
- [11] Feldman Ronen and Sanger James, *The text mining handbook: advanced approaches in analyzing unstructured data*.
- [12] Richard Layard and Jeffrey Sachs John Helliwell, *World Happiness Report 2015.*, 2015.
- [13] Julita Bermejo, "A Simplified Guide to Create an Ontology," May 2007.