

Pemanfaatan Ontologi dengan Paradigma Pembangunan *Combined Hierarchy* dalam Perhitungan Indeks Kebahagiaan Masyarakat Bandung

Anisa Herdiani¹, Nungki Selviandro², Muhammad Faris Adri Azka³

^{1,2,3}Prodi S1 Teknik Informatika, Fakultas Informatika, Universitas Telkom, Bandung

Jalan Telekomunikasi No. 1, Dayeuhkolot, Bandung 40257

¹anisaherdiani@gmail.com, ²selviandro@yahoo.co.uk, ³mfarisazka@gmail.com

Abstrak

Pemerintah kota Bandung saat ini sedang mengembangkan sebuah sistem untuk mengukur indeks kebahagiaan masyarakatnya. Pengukuran indeks kebahagiaan tersebut menggunakan parameter-parameter tertentu yang disesuaikan dengan hal-hal yang dapat membuat masyarakat Bandung menjadi bahagia. Sering kali orang-orang menyampaikan apa yang ada di pikirannya melalui sosial media. Penelitian ini mengusulkan sistem untuk mengukur indeks kebahagiaan tersebut menggunakan data yang ada di media sosial kemudian dilakukan perhitungan sesuai dengan parameter-parameter yang ada.

Penelitian ini akan terbagi atas beberapa tahap. Tahap pertama adalah melakukan *crawling* data dari media sosial twitter dengan pembatasan area *crawling* hanya dari kota Bandung. Tahap kedua adalah melakukan proses *text processing* untuk mendapatkan data tweet yang bersih. Tahap ketiga adalah melakukan pembuatan ontologi yang dijadikan dasar klasifikasi data tweet berdasarkan parameter-parameter indeks kebahagiaan. Tahap keempat adalah mengklasifikasikan tweet dengan menggunakan ontologi yang telah dibuat pada tahap sebelumnya. Tahap kelima adalah klasifikasi sentimen dari data hasil klasifikasi ontologi yang telah dilakukan. Tahap terakhir adalah menghitung indeks kebahagiaan dari hasil klasifikasi sentimen.

Keluaran dari penelitian ini adalah nilai indeks kebahagiaan masyarakat kota Bandung. Indeks kebahagiaan yang didapatkan pada penelitian ini adalah sebesar 58%. Hasil dari penelitian ini dapat digunakan sebagai alternatif perhitungan indeks kebahagiaan Kota Bandung disamping perhitungan indeks kebahagiaan secara manual.

Kata kunci : Indeks Kebahagiaan, Ontologi, Analisis Sentimen, Twitter

Abstract

Bandung city government is currently developing a system to measure the happiness index of their citizen. Happiness index measurement are using some parameters that can make Bandung citizen be happy. People often express what they are thinking on their mind through social media. This study proposes a system to measure the happiness index using existing data in social media then calculated in accordance with the parameters that has been declared.

This study will be divided into several stages. The first phase is crawling of data from social media twitter with crawling area only in Bandung city. The second stage is the process of text processing to get the clean data tweet. The third stage is to create ontology that used for the basis for the tweet classification using the parameters of happiness index. The fourth stage is to classify tweets using ontologies that have been made in the previous stage. The fifth stage is the sentiment classification from the outcome data of ontology classification has been done. The last stage is to calculate the happiness index from the sentiment classification results.

The output of this research is the happiness index value of the Bandung city. Happiness index obtained in this study was 58%. The results of this study can be used as an alternative calculation happiness index Bandung besides happiness index manual calculations.

Keyword : Happiness Index, Ontology, Sentiment Analysis, Twitter, Data Crawling

I. Pendahuluan

Kebahagiaan Nasional Bruto (KNB) merupakan sebuah indeks baru yang banyak diukur oleh berbagai negara yang dianggap lebih penting dibanding dengan Produk Domestik Bruto [1]. Skala perhitungan indeks kebahagiaan yang lebih kecil dapat di terapkan didalam sebuah kota. Pemerintah Kota (Pemkot) Bandung merencanakan untuk melakukan perhitungan indeks kebahagiaan masyarakatnya. Tentunya dengan mengetahui aspek-aspek apa saja yang mempengaruhi indeks kebahagiaan masyarakatnya pemkot Bandung dapat membuat sebuah kebijakan ataupun merubah kebijakan yang sudah ada sebelumnya agar dapat meningkatkan poin dari sebuah aspek.

Bandung termasuk kedalam 20 kota yang paling aktif menggunakan Twitter pada tahun 2012 [2]. Twitter identik dengan penyampaian opini, ataupun perasaan yang sedang dimiliki oleh penulis tweetnya. Maka dari itu pada penelitian ini diusulkan sebuah sistem yang dapat menilai indeks kebahagiaan dari media sosial twitter.

Pengetahuan tentang indeks kebahagiaan dituliskan ke dalam ontologi agar dapat diketahui oleh komputer. Ontologi digunakan karena dapat menyampaikan sebuah pengetahuan dan informasi secara eksplisit, sehingga dapat diinterpretasikan oleh mesin komputer [3]. Paradigma yang digunakan dalam pembuatan ontologi adalah paradigma *combined hierarchy* yang menggabungkan paradigma *top-down* dan *bottom-up*.

Sistem yang dibuat diawali dari crawling data tweet yang berasal dari twitter dengan mengambil tweet dari daerah kota Bandung saja. Setelah data terkumpul maka langkah selanjutnya adalah pembuatan ontologi. Kelas utama dalam ontologi yang merepresentasikan indeks kebahagiaan diambil dari parameter BPS, karena parameter ini yang paling cocok diterapkan di Indonesia. Kelas-kelas dan instan dibawah parameter utama diambil dari twitter, dan sumber lain seperti thesaurus, kamus, dan wordnet. Pembuatan hierarki kelas disesuaikan dengan paradigma pembangunan *combined hierarchy*. Langkah selanjutnya adalah klasifikasi data tweet dengan ontologi. Klasifikasi ontologi pada penelitian ini membandingkan dua teknik yang berbeda yaitu *increase weighting* dan *decrease weighting*. Setelah klasifikasi ontologi dilakukan maka langkah selanjutnya adalah klasifikasi sentimen. Klasifikasi sentimen pada penelitian ini menggunakan algoritma SVM. Langkah terakhir adalah menghitung indeks kebahagiaan dari hasil klasifikasi sentimen.

II. Dasar Teori

1. Ontologi

Ontologi merupakan kumpulan representatif objek yang memodelkan sebuah domain dari pengetahuan [4]. Ontologi memiliki beberapa komponen pembentuknya. Komponen-komponen ini antara lain adalah :

- *Class*, merupakan komponen utama dalam pembentukan ontologi. *Class* dapat merepresentasikan kumpulan dari beberapa individu atau instan yang ada didalam sebuah konsep. Contoh dari sebuah kelas adalah “hewan” yang merupakan sebuah representasi dari individu-individu seperti “sapi, kucing, ayam, anjing”
- *Instances*, merupakan sebuah hal yang paling dasar dalam pembentukan ontologi, *instance* dapat menjelaskan ruang lingkup yang dibahas didalam sebuah ontologi. *Instance* ini merupakan bagian-bagian dari sebuah kelas.
- *Relation*, merupakan hubungan yang menjelaskan keterkaitan antara sebuah *instance* dengan *instance* lainnya atau sebuah kelas dengan kelas lainnya. *Relation* yang digunakan pada penelitian ini hanya sebatas *SubClassOf* [3]

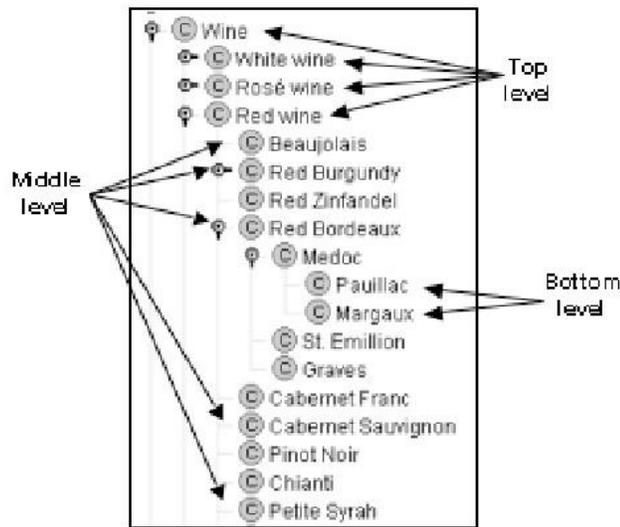
2. Noy McGuinness Method

Metodologi Noy & McGuinness merupakan metode yang dapat digunakan dalam pembangunan ontologi. Pembangunan ontologi dilakukan dalam beberapa langkah, antara lain :

1. Mendefinisikan domain dan cakupan dari ontologi
2. Mempertimbangkan penggunaan ontologi yang sudah ada
3. Mengumpulkan kata-kata penting yang berhubungan dengan cakupan ontologi
4. Mendefinisikan kelas, dan kelas hierarki
5. Mendefinisikan properti dari kelas
6. Mendefinisikan facet dan slot
7. Mendefinisikan instan-instan dari kelas [5].

3. Combined Hierarchy Ontology Development

Combined Hierarchy Ontology Development merupakan sebuah metode yang digunakan dalam pembentukan hierarki kelas-kelas yang terdapat di dalam ontologi. Proses pembangunan ini merupakan kombinasi dari pendekatan *top-down* dan *bottom-up*. Dengan menggunakan metode kombinasi ini kita dapat memulai pembuatan dari mana saja, baik itu *top-level* maupun *bottom-level*. Kelas-kelas yang berada di *Top-level* dan *bottom-level* dapat dihubungkan oleh *middle-level*. Proses pembangunan ini dapat memungkinkan munculnya beberapa *middle-level class* yang berfungsi untuk menghubungkan kedua kelas di tingkat yang berbeda [6]. Contoh hierarki kelas ontologi dapat dilihat pada gambar 1.



Gambar 1 Hierarki Ontologi

4. Training-less Ontology-based using Depth and Weight Classification (TRODEW)

Training-less Ontology-based Text Categorization adalah dasar dari metode baru yaitu *Training-less Ontology-based using Depth and Weight Classification* (TRODEW) dalam mengklasifikasikan data teks ke dalam parameter-parameter tertentu. Perbedaan TRODEW dengan *Training-less Ontology-based Text Categorization* adalah dengan adanya pembobotan dari setiap kelas yang ada didalam ontologi. Terdapat dua metode yaitu *increase weighting* dan *decrease weighting*. *Increase weighting* merupakan teknik pembobotan yang apabila hierarki sebuah kelas didalam ontologi semakin dalam maka nilai bobotnya semakin tinggi, sedangkan *decrease weighting* memberikan bobot tertinggi kepada kelas yang paling atas.

5. Precision, Recall, F1-Measure, dan Akurasi

Precision dan recall merupakan teknik yang dapat digunakan untuk menghitung nilai akurasi dari sistem pemrosesan teks. Perhitungan precision dan recall membutuhkan empat komponen yaitu TP(True positive) data *true* diklasifikasikan *true*, TN(True Negative) data *false* diklasifikasikan *false*, FP(False Positive) data *false* diklasifikasikan *true*, FN(False Negative) data *true* diklasifikasikan *false*. Jika digambarkan dalam tabel maka empat komponen tersebut adalah seperti pada tabel 1 berikut :

Tabel 1 Matriks Kontingensi

	Predicted class (YES)	Predicted class (NO)
Actual class (YES)	TP	FN
Actual class (NO)	FP	TN

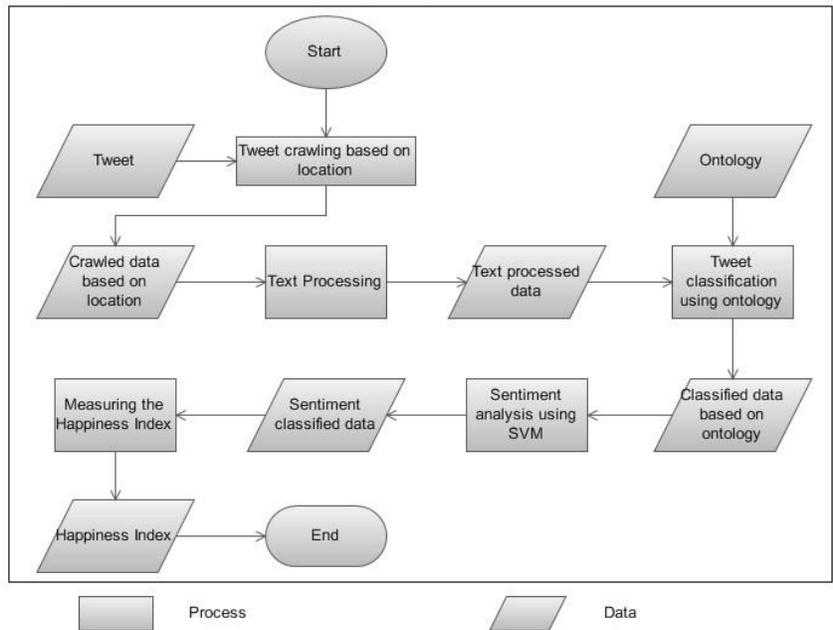
- *Precision* merupakan perhitungan jumlah kemunculan data yang diklasifikasikan true dengan benar dibanding seluruh jumlah data yang diklasifikasikan true, sedangkan recall adalah jumlah data yang diklasifikasikan true dengan benar dibanding dengan jumlah data yang diklasifikasikan true.
- Recall merupakan jumlah kelas yang diklasifikasikan dengan benar pada sebuah kelas dibanding seluruh data hasil klasifikasi sistem pada kelas tersebut.
- F1-measure merupakan hasil rata-rata antara precision dan recall. F1-measure ini muncul untuk menyetarakan nilai precision dan recall yang sering terpaut jauh.
- Accuracy merupakan perhitungan data yang diklasifikasikan dengan benar dibanding dengan seluruh data.

Rumus dari masing-masing perhitungan dapat dilihat pada persamaan berikut

	1
	2
	3
	4

III. Perancangan Sistem

Rancangan sistem dalam penelitian ini adalah seperti pada gambar 2 di bawah ini.



Gambar 2 Perancangan Sistem

Proses-proses yang terjadi didalam sistem pengukuran indeks kebahagiaan masyarakat Bandung adalah :

- a. *Crawling data* twitter, crawling data twitter ini dilakukan dengan menggunakan geocode dari kota Bandung sehingga didapatkan data yang akurat berasal dari kota Bandung.
- b. Pemrosesan teks dilakukan setelah didapatkan data tweet dari kota Bandung untuk mendapatkan data teks yang mudah untuk dilakukan analisis.
- c. Klasifikasi data tweet dilakukan untuk mengelompokkan data tweet kedalam kelas parameter kebahagiaan yang sesuai.
- d. Analisis sentiment berfungsi untuk klasifikasi kedalam kelas tweet positif ataupun negative dari data tweet yang sudah terklasifikasi sebelumnya kedalam parameter
- e. Pengukuran indeks kebahagiaan dilakukan untuk mengetahui prosentase dari indeks kebahagiaan masyarakat Bandung.

IV. Pembahasan

Sistem dibangun sesuai rancangan yang telah dibuat. Pembahasan terdiri hasil dari implementasi sistem

Hasil pengujian Teknik Klasifikasi Ontologi dengan Pembobotan *Increase Weighting*

Hasil klasifikasi ontologi dari kedua metode menghasilkan jumlah yang sama yaitu sebanyak 2191 tweet dari 10798 tweet hasil *preprocessing* sebelumnya. Pada grafik dibawah dapat dilihat bahwa parameter yang paling banyak muncul adalah parameter hubungan sosial, sedangkan yang paling sedikit muncul adalah kondisi rumah dan aset. Keseluruhan hasil klasifikasi dapat dilihat pada tabel 2 dibawah ini.

Tabel 2 Jumlah per Parameter Klasifikasi Increase Weighting

Parameter	Jumlah
hubungan_sosial	645
keadaan_lingkungan	555
kesehatan	367
pendidikan	305
pekerjaan	107
waktu_luang	57
keharmonisan_rumah_tangga	49
pendapatan_rumah_tangga	47
kondisi_keamanan	43
kondisi_rumah_dan_aset	16

Nilai akurasi dari sistem klasifikasi menggunakan metode *increase weight* berada pada angka 67.0%. Nilai akurasi tersebut didapatkan dengan membagi tweet yang diklasifikasikan dengan benar dengan seluruh tweet yang dijadikan sampel untuk pengujian akurasi sistem ini yakni sebanyak 170 tweet. Data sampel diambil secara acak dari seluruh tweet yang diklasifikasikan oleh sistem. Untuk nilai precision, recall dan F1-measure dapat dilihat pada tabel 3 dibawah ini.

Tabel 3 Evaluasi Increase Weighting

Parameter	Precision	Recall	F1-measure
hubungan sosial	71.7%	66.0%	68.8%
keadaan_lingkungan	78.7%	68.5%	73.3%
keharmonisan_rumah_tangga	100.0%	54.5%	70.6%
kesehatan	45.7%	94.1%	61.5%
kondisi_keamanan	50.0%	20.0%	28.6%
kondisi_rumah_dan_aset	100.0%	66.7%	80.0%
pekerjaan	50.0%	50.0%	50.0%
pendapatan_rumah_tangga	100.0%	66.7%	80.0%
pendidikan	56.0%	87.5%	68.3%
waktu_luang	100.0%	25.0%	40.0%

Hasil pengujian Teknik Klasifikasi Ontologi dengan Pembobotan *Decrease Weighting*

Hasil klasifikasi ontologi dengan teknik pembobotan *decrease weighting* per parameter dapat dilihat pada tabel 4.2 di bawah ini.

Tabel 4 Jumlah per Parameter Klasifikasi Decrease Weighting

Parameter	Jumlah
hubungan_sosial	634
keadaan_lingkungan	545
kesehatan	361
pendidikan	311
pekerjaan	126
waktu_luang	66
keharmonisan_rumah_tangga	48

Parameter	Jumlah
kondisi_keamanan	43
pendapatan_rumah_tangga	41
kondisi_rumah_dan_aset	16

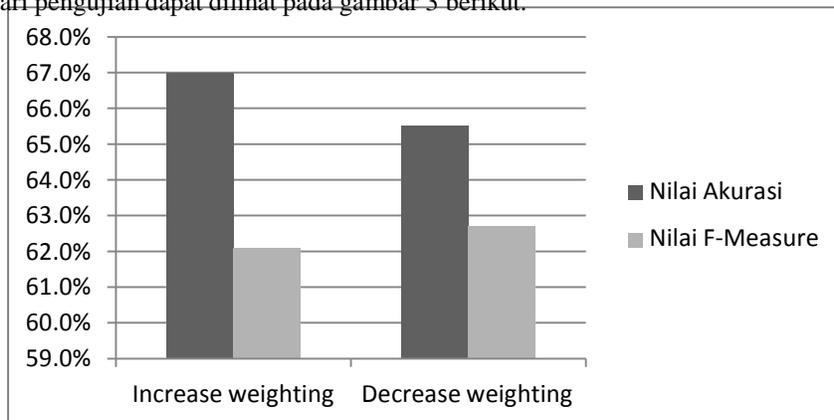
Hasil akurasi untuk klasifikasi dengan metode decrease weight adalah sebesar 65.5%. Cara perhitungannya adalah dengan membandingkan jumlah tweet yang diklasifikasikan benar dengan seluruh jumlah tweet yang dijadikan sampel yaitu 170 tweet. Nilai precision, recall, dan f1-measure dapat dilihat pada tabel 4.4 berikut ini.

Tabel 5 Evaluasi Decrease Weighting

Parameter	Precision	Recall	F1-Measure
hubungan_sosial	69.6%	64.0%	66.7%
keadaan_lingkungan	78.3%	66.7%	72.0%
keharmonisan_rumah_tangga	85.7%	54.5%	66.7%
kesehatan	48.6%	100.0%	65.4%
kondisi_keamanan	50.0%	20.0%	28.6%
kondisi_rumah_dan_aset	100.0%	66.7%	80.0%
pekerjaan	42.9%	50.0%	46.2%
pendapatan_rumah_tangga	100.0%	66.7%	80.0%
pendidikan	56.5%	81.3%	66.7%
waktu_luang	100.0%	37.5%	54.5%

Analisis Hasil Pengujian Klasifikasi Ontologi

Hasil Klasifikasi ontologi dengan menggunakan metode *decrease weight* memiliki hasil yang tidak jauh berbeda dibandingkan dengan metode *increase weight*. Nilai akurasi dan f-measure dari kedua teknik tersebut merupakan hasil dari pengujian dapat dilihat pada gambar 3 berikut.



Gambar 3 Hasil akurasi teknik klasifikasi ontologi

Gambar 4 Grafik Perbandingan Akurasi Teknik Klasifikasi Ontologi

Perbedaan yang sedikit bisa disebabkan karena kebanyakan tweet hanya membicarakan sebuah topik didalamnya. Sebuah tweet kemungkinan memiliki beberapa topik sekaligus, pada saat itulah fungsi weighting digunakan. Sebuah parameter yang memiliki weight yang lebih tinggi pada sebuah tweet akan ditetapkan sebagai parameter tweet tersebut. Jika ternyata sebuah tweet hanya memiliki sebuah topik yang diperbincangkan, maka sebuah parameter tidak memiliki pembanding yang lain sehingga sebuah tweet tidak akan berubah parameternya. Contohnya adalah seperti pada tabel 6 :

Tabel 6 Tweet dengan satu parameter

Tweet	Kecocokan term ontologi
jakartaaa kota kejam kesayangann	Kota weight 3
rancaekek banjir lalin arah tasik gmana situasi nuhun	Banjir weight 3

Sebuah parameter dapat berubah ketika setidaknya memiliki dua buah term yang cocok dengan term yang ada di dalam ontologi. Contohnya pada tabel dibawah ini jika menggunakan teknik klasifikasi increase maka tweet akan di kategorikan kepada parameter induk dari kata “libur” yakni “ketersediaan_waktu_luang”, sedangkan dengan teknik klasifikasi decrease tweet akan dikategorisasikan pada parameter induk kata “pening” yakni “kesehatan”, contoh weighting dapat dilihat pada tabel 7:

Tabel 7 Tweet dengan dua parameter

Tweet	Increase Weighting	Decrease Weighting
Libur pening pala	Libur weight 4 Pening weight 3	Libur weight 1 Pening weight 2

Teknik klasifikasi terbaik menghasilkan nilai akurasi sebesar 67.0% yang masih cukup jauh dari 100%. Terdapat beberapa hal yang mempengaruhi turunnya nilai hasil klasifikasi ontologi, hal-hal tersebut antara lain :

1. Terdapat banyak kata-kata yang tidak baku dan salah ejaan yang tidak dapat ditangani oleh proses *text processing*. Contohnya sepeeti pada tweet “jakartaaa kota kejam kesayangann” terdapat kata “jakartaa” dan “kesayangann” yang tidak dapat dikenali oleh sistem
2. Ontologi yang kurang lengkap. Ontologi di dalam sistem ini menjadi filter pertama dalam sistem indeks kebahagiaan, sehingga apabila terdapat kata-kata di tweet yang tidak terdapat di dalam ontologi sedangkan kata tersebut mempengaruhi indeks kebahagiaan maka tweet tersebut tidak dapat terklasifikasi didalam ontologi.
3. Ambiguitas dari sebuah kalimat. Contohnya pada tweet “barbie pusing tidur beso kuliah pagi kuk bikin cemas kembar cuci muka sikat gigi salat isya tidur”, inti dari kalimat tersebut sulit dipahami oleh manusia, sedangkan komputer menghitung dari bobot tertinggi dari kemunculan term.
4. Kesalahan labeling oleh manusia. Perhitungan akurasi membandingkan hasil keluaran sistem dengan hasil labeling manual oleh manusia, jika proses pelabelan manual tersebut memiliki pemahaman yang berbeda, maka dapat menurunkan nilai akurasinya.
5. Sistem yang tidak dapat menangani frasa. Kata-kata seperti rumah sakit, nasi goreng mafia, tidak dapat dideteksi oleh sistem, sehingga dapat berubah parameter dari parameter asli yang dipahami manusia.

Hasil Klasifikasi Sentimen Menggunakan SVM

Proses klasifikasi SVM menggunakan data hasil dari klasifikasi ontologi yang berjumlah. Data latih dan data uji berjumlah 2191 data. Data-data tersebut diberi label secara manual, kemudian akan dilakukan *cross-validation* untuk menemukan tingkat validasinya dengan menghitung rata-rata hasil pengujian yang dilakukan. Hasil akurasi *cross-validation* yang dilakukan sebanyak 10 kali dari pengujian ini adalah 58.2% data berhasil diklasifikasikan dengan benar. Hasil klasifikasi ini nantinya akan digunakan untuk perhitungan nilai *happiness index* masyarakat kota Bandung. Untuk perhitungan precision, recall, dan f1-measure dari masing-masing jumlah data dapat dilihat pada tabel 8 dibawah ini :

Tabel 8 Hasil Klasifikasi Sentimen

Jumlah data	Precision	Recall	F1-Measure
500	60.8%	53.2%	51.9%
1000	59.5%	55.3%	54.2%
1500	61.0%	57.1%	55.6%
Semua data	61.6%	58.2%	56.5%

Hasil pengujian tersebut menunjukkan bahwa semakin banyak data yang digunakan dalam proses *training* maka nilai F-measure akan semakin meningkat. Hal ini dikarenakan dengan semakin banyaknya data latih maka pengetahuan sistem akan semakin sesuai dengan pengetahuan aktual. Nilai f1-measure terbaik adalah pada 56.5% yang masih jauh dari 100%. Rendahnya nilai klasifikasi sentimen bisa disebabkan karena banyaknya kata-kata yang tidak baku, disingkat, ataupun pengejaan yang salah.

Perhitungan Indeks Kebahagiaan

Perhitungan indeks kebahagiaan Hasil perhitungan indeks kebahagiaan dari kedua teknik klasifikasi tentunya beda. SVM menghasilkan nilai indeks kebahagiaan 58.0%.

V. Kesimpulan

Berdasarkan hasil pengujian dan analisis yang telah dilakukan maka dapat disimpulkan beberapa poin sebagai berikut :

1. Parameter indeks kebahagiaan dapat dibangun menggunakan ontologi dengan paradigma pembangunan *combined hierarchy*
2. Ontologi dapat digunakan untuk mengklasifikasikan data tweet dengan cara mencocokkan *term* yang muncul di dalam tweet dengan *term* yang terdapat di dalam ontologi
3. Akurasi dari klasifikasi ontologi dengan teknik pembobotan *increase weighting* adalah sebesar 67.0%, sedangkan akurasi dengan teknik pembobotan *decrease weighting* didapatkan akurasi sebesar 65.5%. Nilai akurasi klasifikasi sentimen dengan menggunakan SVM adalah sebesar 56.5%.
4. Hasil penelitian yang telah dilakukan menghasilkan indeks kebahagiaan masyarakat kota Bandung adalah sebesar 58%

Daftar Pustaka

- [1] The Centre for Bhutan Studies, "Gross National Happiness Index," p. 6, 2012.
- [2] SuaraJakarta. SuaraJakarta. [Online]. <http://suarajakarta.co/lifestyle/teknologi/semiocast-warga-jakarta-paling-aktif-di-twitter/>
- [3] Natalya F. Noy and Deborah L. McGuinness, "Ontology Development 101: A Guide to Creating Your First Ontology," Stanford University, Stanford,.
- [4] Phillip Lord. (2010, January) Ontogenesis. [Online]. <http://ontogenesis.knowledgeblog.org/514>
- [5] Deborah L. McGuinness Natalya F. Noy, *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford: Standford University.
- [6] RAJENDRA AKERKAR TERJE AABERGE, "ONTOLOGY AND ONTOLOGY CONSTRUCTION: BACKGROUND AND PRACTICES," *International Journal of Computer Science and Applications*, p. 6, 2012.