

## ANALISIS SENTIMEN BERBAHASA INDONESIA DENGAN PENDEKATAN LEXICON BASED (STUDI KASUS : SOLUSI PENGELOLAAN SAMPAH)

Veronikha Effendy

Universitas Telkom

Jl. Telekomunikasi No. 1, Terusan Buah Batu, Bandung

Email : veffendy@telkomuniversity.ac.id

### ABSTRAK

Sampah diketahui menjadi problem yang dominan hampir di setiap kota besar, tidak terkecuali di Indonesia. Hampir setiap rumah tangga dan industri ikut berkontribusi terhadap penambahan jumlah sampah. Setiap hari, jumlah produksi sampah semakin tidak terkendali. Hal ini menjadi pemicu utama terjadinya bencana yang diakibatkan oleh tumpukan sampah, diantaranya : banjir, pencemaran tanah, air, udara. Saat ini, cukup banyak aktifitas para pencinta lingkungan dalam menanggulangi permasalahan sampah ini, seperti : melakukan kampanye pengelolaan sampah, memberikan pelatihan dan sebagainya. Aktifis mulai banyak melakukan kampanye dan menyampaikan informasi lewat berbagai social media. Cukup banyak ide kreatif solusi pengelolaan sampah berbetbaran di social media, terutama di twitter. Penelitian ini menggali informasi seputar solusi sampah dengan pendekatan analisis sentimen berdasar kamus pada data twitter. Sejumlah opini dari twitter diproses dan dikelompokkan ke dalam sentimen positif atau negatif. Sentimen positif diharapkan berisi solusi untuk pengolahan sampah. Beberapa diantaranya memiliki link web address yang diharapkan dapat lebih menjelaskan tentang solusi tersebut.

**Kata kunci** : pengelolaan sampah, analisis sentimen, kamus, social media, twitter

### 1. PENDAHULUAN

Penyelesaian permasalahan sampah yang hanya mengandalkan pemerintah dan swasta sangat kurang tepat. Pemerintah daerah sering mengalami kekurangan dana, sumber daya manusia dan pada akhirnya berujung pada ketidak mampuan mengelola sistem yang ada. Saat ini peran serta masyarakat yang tergabung dalam beberapa komunitas daur ulang menyumbang cukup signifikan terhadap pengurangan volume sampah. Mengurangi volume sampah sebaiknya dilakukan dari sumber, dan itu berarti melibatkan masyarakat [1]. Namun, kadang individu masyarakat masih

banyak yang bingung apa yang dapat dilakukan untuk mengurangi volume sampah dari limbah sekitarnya. Oleh karena itu untuk membantu individu masyarakat mencari ide/ solusi pengelolaan limbah sampahnya sendiri, maka dalam penelitian ini mencoba untuk membuat suatu framework bagaimana cara mendapatkan solusi pengelolaan sampah dengan cepat dan tepat.

Saat ini cukup banyak komunitas peduli sampah yang sering melakukan sosialisasi di berbagai daerah maupun melalui social media. Salah satu media social yang sering digunakan komunitas tersebut adalah twitter. Hal ini wajar karena twitter merupakan salah satu media social yang akrab di masyarakat Indonesia. Berdasarkan Semiocast, jumlah pemilik akun twitter di Indonesia merupakan terbesar kelima di dunia, dan berada pada posisi ketiga negara yang paling aktif mengirim tweet per hari [2].

Penelitian ini memanfaatkan data twitter terutama dengan topik pengelolaan sampah untuk mendapatkan solusi pengelolaan sampah yang pada akhirnya akan dapat diterapkan di masyarakat. Penelitian ini menggunakan pendekatan data mining yaitu dengan analisis sentimen. Analisis sentimen merupakan cabang ilmu dari data mining yang memiliki tujuan untuk menganalisis, memahami, mengolah dan mengekstrak data teks terhadap entitas seperti produk, servis, organisasi dan topik tertentu [3]. Dengan menggunakan pendekatan data mining, opini pada twitter tentang pengelolaan sampah dapat difilter untuk mendapatkan opini positif yang umumnya berisikan solusi pengelolaan sampah maupun link url tentang cara pengelolaan sampah. Dari opini positif tersebut kemudian diperoleh link url yang berisi solusi pengelolaan sampah. Isi dari URL tersebut kemudian diproses menggunakan text mining untuk mendapatkan kata kunci untuk pengolahan sampah. Pada akhirnya baik isi URL maupun kata kunci tersebut dapat menjadi referensi solusi pengelolaan sampah.

**2. DATA DAN METODE**

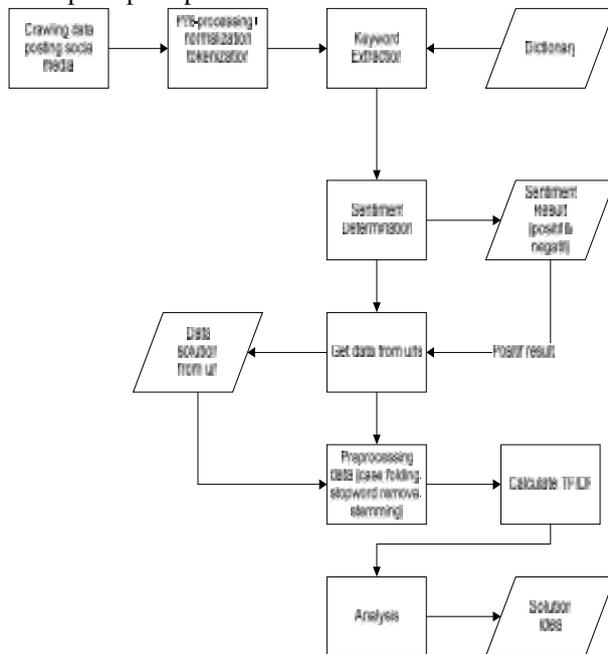
**2.1 Data**

Data diperoleh dari hasil crawling pada social media twitter karena karakteristik data tweeter yang lebih up-to-date dan dikenal masyarakat luas. Untuk keperluan penelitian ini, crawling data dilakukan selama 2 hari dan memperoleh 463 data tweet.

Karakteristik data tweet berbahasa Indonesia adalah banyaknya kata – kata “gaul”, singkatan yang tidak umum, dan makna konotasi dari tiap kata, khususnya pada kasus seputar “sampah”. Hal ini akan dicoba diatasi dengan pengayaan kamus.

**2.2 Metode yang Diajukan**

Gambar 1 menunjukkan Blok Proses yang akan diterapkan pada penelitian ini.



**Gambar 1. Blok Proses Penelitian**

**2.2.1 Crawling Data**

Proses ini merupakan proses untuk mengambil data – data posting dengan menggunakan bantuan API search pada twitter. Beberapa query yang dicoba antara lain :

- Pengelolaan sampah
- Bank sampah
- Sampah organik
- Sampah anorganik
- Sampah plastik
- Solusi sampah
- Pengelolaan sampah rumah
- Daur ulang plastik bekas

Opini–opini yang diperoleh kemudian dilabelkan secara manual untuk keperluan pengukuran akurasi dari proses sentimen analysis.

**2.2.2 Pre-processing**

Masalah utama dari pre-processing pada sentimen analisis yang berasal dari twitter adalah bahasanya yang kebanyakan bahasa “gaul” dan penuh singkatan. Oleh karena itu diperlukan proses untuk menormalkan kata – kata tersebut. Pada tugas ini, singkatan – singkatan dan bahasa gaul yang terdeteksi masih sangat terbatas dan akan diperlukan banyak tambahan lagi untuk mengantisipasi data – data baru yang kemungkinan akan muncul.

Pre-processing dilanjutkan dengan proses pemisahan kata per kata dari kalimat opini.

**2.2.3 Keyword Extraction**

Proses ini dilakukan dengan mendata token/kata yang sesuai dengan kamus. Kamus yang digunakan adalah kamus yang berisi daftar kata – kata positif dan negatif.

Permasalahan utama adalah penentuan kamus data, dimana cukup banyak kata – kata yang berhubungan dengan “sampah” yang secara umum memiliki kata yang merupakan ciri kalimat negatif, namun dalam kasus ini tidak bisa dianggap sebagai kata negatif. Contoh : “sampah”, cukup banyak opini – opini yang menggunakan kata “sampah” sebagai kata yang bermakna negatif, namun dalam kasus ini, tentunya kata sampah akan sering dijumpai. Emoticon yang pada umumnya dapat digunakan sebagai pendeteksi sentimen positif dan negatif dalam hal ini juga memiliki makna yang berbeda, misalnya : emoticon :D yang pada umumnya menggambarkan sentimen positif, tapi dalam kasus ini, biasanya emoticon ini banyak dijumpai pada opini yang mengandung candaan yang pada akhirnya opini tersebut dilabelkan negatif karena tidak mengandung solusi, contoh :

*@dentaafriyana Iye gue mungutin saMpah plastik,lo mungutin besi karat yehdent buat diloakin:D (canda yeh jgn marah)*

**2.2.4 Sentiment Determination**

Proses ini merupakan proses penentuan opini tersebut tergolong dalam kelompok positif atau negatif. Kelompok positif merupakan kandidat-kandidat yang berisi solusi yang dapat ditemukan.

Daftar solusi yang terbentuk dapat dijadikan acuan bagi masyarakat untuk dapat menerapkan solusi pengelolaan sampah yang dirasakan paling tepat, bahkan bisa juga menjadi peluang usaha. Idealnya solusi yang muncul dalam social media merupakan solusi yang sudah cukup dikenal di masyarakat, sehingga kemungkinan besar bisa diterapkan oleh masyarakat pada umumnya.

**2.2.5 Get Data URL**

Dari sejumlah data yang benar terprediksi sebagai solusi (sentimen positif), kemudian akan diambil isi dari link yang berupa berita, dengan detail statistik sebagai berikut:

Tabel 1. Statistik Data Sentimen Positif

<b>true Positif</b>	<b>248</b>
have link 1	73
no link	175
<b>distinct url</b>	<b>55</b>
<b>url contain distinct news</b>	<b>13</b>

Dalam proses ini, ditemukan beberapa URL yang ternyata sama persis isinya dari sumber tweet maupun url yang berbeda, sehingga pada akhirnya diperoleh sekitar 13 jenis berita maupun detail langkah – langkah solusi pengelolaan sampah.

**2.2.6 Preprocessing Data**

**2.2.6.1 Case Folding**

Merupakan proses untuk membuat huruf besar menjadi huruf kecil

**2.2.6.2 Stop Words Removal**

Proses ini menghilangkan kata – kata yang tidak penting. Daftar kata–kata yang termasuk tidak penting ini untuk sementara menggunakan daftar kata stop word yang umum digunakan pada pre-processing text mining untuk bahasa Indonesia.

**2.2.6.3 Stemming**

Proses untuk mengambil kata dasar dari suatu kata. Untuk bahasa Indonesia sendiri sudah ada beberapa algoritma untuk proses ini, salah satunya adalah algoritma Porters yang digunakan pada penelitian ini.

**2.2.7 Calculate TFIDF**

Ke-13 dokumen tersebut kemudian diproses lagi oleh sistem untuk dilakukan perhitungan TFIDF, dengan tujuan dapat memperoleh term – term penting sebagai kata kunci dari setiap dokumen tersebut. Proses untuk melakukan pembobotan terhadap setiap kata dasar yang terbentuk dari proses stemming (term). Pada penelitian ini dilakukan pembobotan dengan perhitungan TF-IDF. TF (*Term Frequency*) merepresentasikan jumlah kemunculan dari suatu term pada suatu dokumen, sedangkan DF (*Document Frequency*) merupakan jumlah document yang berisi suatu term. IDF merupakan invers dari DF. Pembobotan dengan menggunakan perhitungan TF.IDF

merepresentasikan tingkat kepentingan suatu term pada dokumen tersebut.

**2.2.8 Pengujian dan Analisis**

Tabel 2 menunjukkan confusion matriks dari hasil dari sentiment analysis :

Tabel 2. Confusion Matriks Analisis Sentimen

	<b>predict (1)</b>	<b>predict (-1)</b>	<b>unpredict (0)</b>
<b>actual(1)</b>	248	8	31
<b>actual(-1)</b>	36	92	66

Dari data tersebut dapat dihitung akurasi nya sebagai berikut :

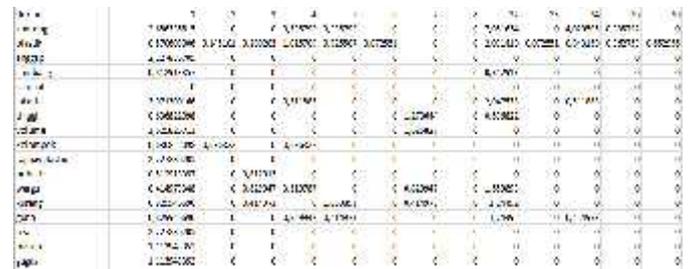
$$Accuracy = (248 +92)/(248+8+31+36+92+66) = 70,68 \%$$

$$True\ Positif\ rate = 248/(248+8+31) = 86,411\%$$

Secara umum, dengan akurasi diatas 70% dan TPR diatas 85% bisa dikatakan bahwa pendekatan sentimen analis bisa diterapkan dalam pemilahan sentimen solusi dan bukan solusi, khususnya dalam kasus pengelolaan sampah.

False prediction maupun unpredict sentimen disebabkan dari kamus yang belum dapat mengakomodir keseluruhan kata positif maupun negatif seputar pengelolaan sampah. Kamus yang umum digunakan tidak dapat mengakomodir sentimen dengan kasus pengelolaan sampah ini, dikarenakan cukup banyak kata – kata yang tidak sesuai ( jika secara umum dinyatakan sebagai kamus negatif, namun dalam kasus ini, tidak bisa diperlakukan sebaga kamus negatif, dan sebagainya).

Gambar 2 menunjukkan snapshot dari hasil perhitungan TFIDF dengan 13 data dari isi link url.



Gambar 2. Snapshot perhitungan TFIDF

Dari hasil perhitungan TFIDF terlihat beberapa term yang memiliki nilai tinggi dan term yang bernilai 0 pada setiap dokumen. Kombinasi nilai ini memperlihatkan bahwa setiap dokumen memang memiliki keunikan atau isi yang berbeda. Dari data

tersebut, dicoba untuk dilakukan analisa dengan memperhatikan term- term yang memiliki nilai tinggi (diambil nilai tfidf  $\geq 0,80$ ), sehingga diperoleh data pada Tabel 3.

Tabel 3. Analisa Dokumen

Dokumen id	Contoh term bernilai tinggi	Analisa
1	Indonesia jakarta kantong belanja gratis volume gagasan petisi payplastic imbau siar persnya .....	Adanya permasalahan kantong belanja di pusat perbelanjaan dan adanya gerakan <b>pay4plastic</b> dengan melakukan imbauan <b>diet</b> (diet penggunaan kantong belanja), <b>kampanye</b> , dan <b>penggunaan kantong alternatif</b>
2	Siswa sekolah lubukbasung bank rp negeri smpn ...	Adanya kegiatan di sekolah dalam bentuk pengelolaan <b>bank sampah</b> dengan inisiatif sendiri
3	Bank setor tugas cilebut untung tabung nasabah pameran jual uang ..	Pada dokumen ini terlihat lebih banyak ide pengelolaan <b>bank sampah</b> dengan mengelola sampah dapur menjadi barang yang bernilai ekonomis dengan menghasilkan omset yang lumayan.
4	...	Adanya <b>sosialisasi</b> oleh gubernur dalam hal memerangi permasalahan sampah ke masyarakat dan <b>ajakan</b> untuk hidup dengan <b>lingkungan yang sehat</b>
5	...	<b>Mengurangi</b> penggunaan/pembelian barang yang banyak menghasilkan sampah, Solusinya dengan membawa bekal sendiri dengan wadah yang aman dan dapat

		digunakan berulang-ulang.
6	...	Adanya <b>gerakan pungut sampah dan pembagian stiker</b> (kampanye) di semarang
7	...	Dokumen ini menjelaskan detail <b>pembuatan pupuk dari sampah</b> .
8	...	Pelatihan <b>pembuatan pupuk kompos</b>
22	...	Memaparkan kejadian yang membahayakan di Indonesia akibat dari sampah, seperti longsor (tumpukan sampah), pencemaran, banjir dsb, dan solusi untuk <b>reduce, diet</b> penggunaan sampah plastik dan <b>recycle</b> .
23	...	Solusi dengan melakukan <b>recycle</b> dari sampah <b>plastik</b>
24	...	Adanya solusi alternatif <b>kantong berbahan singkong</b> yang <b>degradable</b>
26	...	Adanya penemuan mesin pengolah limbah menjadi energi oleh ilmuwan jepang
30	...	Solusi permasalahan sampah di bali dengan pemilahan sampah yang baik

### 3. KESIMPULAN DAN SARAN

#### 3.1 Kesimpulan

Dari hasil pengujian yang telah dilakukan dapat diambil beberapa kesimpulan sebagai berikut:

1. Secara umum, dengan akurasi diatas 70% dan TPR diatas 85% bisa dikatakan bahwa pendekatan sentimen analis bisa diterapkan dalam pemilahan sentimen solusi dan bukan solusi, khususnya dalam kasus pengelolaan sampah. Namun akurasi ini sangat tergantung dengan kamus data, sehingga akurasi bisa diperbaiki dengan pengayaan kamus.
2. Pendekatan pencarian kata-kata kunci dalam setiap dokumen dengan perhitungan TFIDF, dapat diimplementasikan untuk menghasilkan kata – kata penting sehingga dapat dianalisa perkiraan solusi yang terdapat dalam tiap dokumen.

3. Kelemahan dari pendekatan tfidf ini adalah sistem tidak dapat mengeluarkan detail langkah – langkah yang ada. Kelemahan ini bisa diatasi dengan menyajikan link URL.

### 3.2 Saran

Saran untuk pengembangan berikutnya antara lain:

1. Perbaiki kamus data untuk dapat menghasilkan akurasi analisis sentimen yang lebih baik Penggunaan model persamaan yang lebih beragam untuk membangun fungsi prediksi.
2. Perlu dicoba untuk dilakukan penggalian informasi kata kunci dari tweet untuk sentimen yang tidak memiliki url
3. Bisa menggabungkan data dari internet ( baik dari social media (gabungan facebook dan twitter) maupun dari mesin *searching* seperti google)
4. Perlu penerapan teks summarization untuk dapat mendapatkan ringkasan dari dokumen solusi terkait agar detail langkah – langkah yang terdapat dalam dokumen dapat langsung disajikan.
5. Perlu dilakukan mekanisme secara otomatis untuk mendapatkan isi dari URL.

### DAFTAR PUSTAKA

- [1] H. R. Sudrajat, Solusi mengatasi masalah sampah kota dengan manajemen terpadu dan mengolahnya menjadi energi listrik & kompos, Bogor: Penebar Swadaya, 2006.
- [2] R. Asihh, "tekno.tempo.co," tempo, 2 Februari 2012. [Online]. Available: <http://tekno.tempo.co/read/news/2012/02/02/072381323/indonesia-pengguna-twitter-terbesar-kelima-dunia>. [Accessed 8 September 2015].
- [3] P.-N. Tan, M. Steinbach and V. Kumar, Introduction to Data Mining, Boston: Person Education, Inc., 2006.

