JURNAL TEKNOLOGI INFORMASI
Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika

[E-ISSN 2656-0321]
[Vol. 12 No. 2]
[Agustus 2018]

A SURVEY OF SENTIMENT ANALYSIS USING SENTIWORDNET ON BAHASA INDONESIA

Sherly Christina [a,1,*], **Deddy Ronaldo** [b,2]
[a] Universitas Palangka Raya, Jl. H. Timang
[b] Universitas Palangka Raya, Jl. H. Timang
[1] sherly.christina.upr@gmail.com *; [2]deddy.ronaldo@gmail.com
* corresponding author

ARTICLE INFO

Keywords
user opinion
Indonesian
sentiwordnet
sentiment analysis

ABSTRACT

High internet usage in Indonesia brings a number of data user opinions. User opinion data can be processed into information that is useful for decision making, data searching and researching for production and marketing strategies. This research conducted a survey on some Indonesian Sentiment Analysis research articles. Some of these studies have different topics and data sources, butuse Sentiwordnet as a lexical database. The results of this survey show the performance of Sentiwordnet can improve classification accuracy in the Sentiment analysis system for Indonesian text.

1. Pendahuluan

Pengguna internet di Indonesia setiap tahun terus meningkat. Menurut media berita *online* Kompas.com pada 22 Februari 2018, jumlah pengguna internet di Indonesia adalah 50% dari populasi penduduk Indonesia. Masyarakat Indonesia menggunakan Internet tidak hanya untuk berkomunikasi, tetapi juga untuk berbelanja, memesan transportasi, berbisnis, maupun berkarya.

Penggunaan Internet yang meningkat juga membawa sejumlah data yang berguna untuk analisis. Data dari pengguna Internet yang disajikan dalam bentuk *user opinion*, dapat diolah menjadi informasi yang dapat membantu pengambilan keputusan, pencarian data, dan riset bagi pemasaran. *User opinion* diperoleh dari setiap inputan oleh pengguna internet berupa tanggapan, pendapat, kritik, saran maupun komentar yang disampaikan secara online melalui media internet.

User opinion dapat dianalisis menjadi informasi yang lebih berguna menggunakan teknik *opinion mining* atau disebut juga *sentiment analysis.* Sentiment Analysis adalah proses untuk mengidentifikasikan dan mengklasifikasikan pendapat yang diekspresikan melalui teks. Sentiment Analysis mengidentifikasikan sikap dari penulis terhadap suatu entitas, produk atau topik tertentu. Sehingga *sentiment analysis* mencoba memberikan informasi mengenai emosi yang tertuang di dalam teks[4].

Penelitian ini melakukan survey pada beberapa artikel yang melaporkan hasil penelitian mengenai *sentiment analysis* terhadap data set berbahasa Indonesia. Ketiga penelitian yang dibahas pada artikel ini berbasiskan leksikon dan menggunakan Sentiwordnet di dalam kerangka kerja sistemnya. Survey pada penelitian memberikan inspirasi untuk mengaplikasikan metode yang terbaik pada aplikasi Sentiment Analysis bagi data berbahasa Indonesia.

1.1 Sentiwordnet

Sentiwordnet adalah basis data leksikal yang dibangun untuk mendukung pengklasifikasian *sentiment* dan proses-proses *opinion mining* pada aplikasi [5]. Sentiwordnet berisi klasifikasi

JURNAL TEKNOLOGI INFORMASI
Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika
[E-ISSN 2656-0321]
[Vol. 12 No. 2]
[Agustus 2018]

*sentiment*positif, negatif dan netral dari seluruh *synset* Wordnet. Setiap synset(s)di dalam Wordnet memiliki skor numerik dalam klasifikasi sentiment. Skor *sentiment* pada Sentiwordnet terdiri atas skor Pos(s), Neg(s) dan Obj(s) yang mengindikasikan nilai seberapa positif, negatif dan objektif (netral) istilah-istilah yang terdapat dalam suatu synset.

Contohnya pada Sentiwordnet 1.0 pada *synset* kata sifat [estimable(J,3)], terkait makna "may be computed or estimated" memiliki skor *Obj*=1, sedangkan pada synset [estimable(J,1)], terkait makna "deserving of respect or high regard" memiliki skor *Pos*=0.75 , skor *Neg*=0 dan skor *Obj*=0.25. Kisaran skor *Pos, Neg* dan *Obj* adalah [0 sampai 1] dan jumlah ketiga skor tersebut adalah 1 untuk setiap *synset.*

2. Metode Penelitian

2.1 Sentiment Analysis Pada *Review* Produk Berbahasa Indonesia

Pamungkas dan Putri[1] mengimplementasikan*sentiment analysis*berbasikan leksikal pada koleksi data teks berbahasa Indonesia yang diperoleh dari sejumlah *review* terhadap produk dari Google Play dan Apple App Store. Hasil *sentiment analysis* dari opini pengguna terhadap suatu produk dapat mempengaruhi pengambilan keputusan produksi maupun strategi bisnis suatu produk.

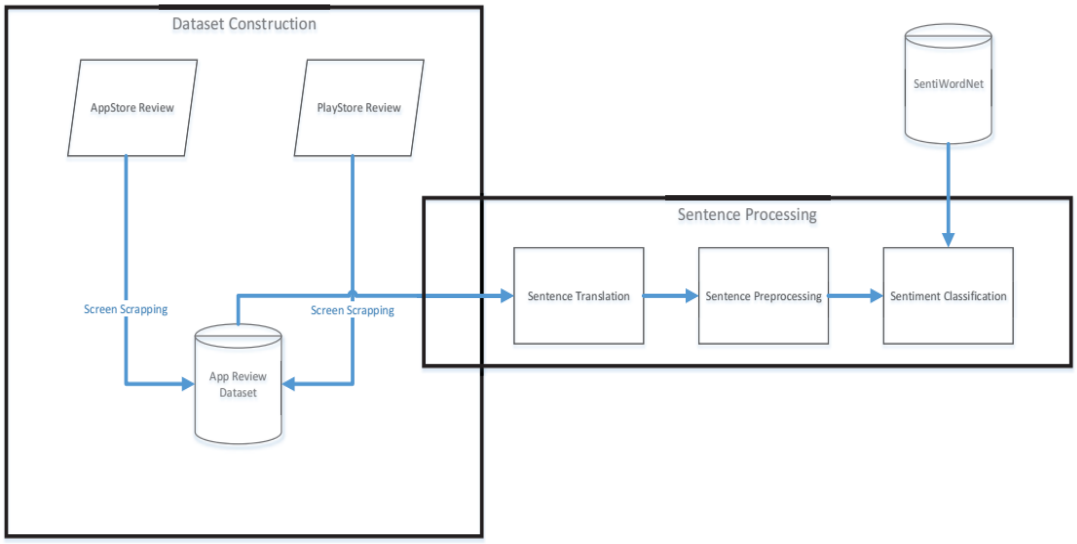
Penelitian berbasiskan leksikal oleh Pamungkas dan Putri mencoba untuk menghindari masalah yang kerap timbul dalam penggunaan metode terawasi (*supervised method*), seperti terbatasnya subyek atau topik dari *sentiment analysis*, serta pelabelan data secara manual yang memerlukan waktu dan usaha yang cukup besar.

Metode *Sentiment analysis* yang digunakan oleh Pamungkas dan Putri berbasiskan leksikal yang didukung oleh Sentiwordnet. Sentiwordnet tidak mendukung data teks berbahasa Indonesia, sehingga sebelum dianalisa, data set harus di terjemahkan lebih dulu menjadi bahasa Inggris. Berikut ini beberapa tahapan penelitian yang dilakukan oleh Pamungkas dan Putri i.

a. Data set yang diperoleh dari Google Play Store dan Apple App Store diterjemahkan ke bahasa Indonesia. Data yang telah diterjemah tidak dicek lagi kebenarannya, karena diasumsikan sudah benar.
b. Berikutnya data hasil terjemahan melewati tahap prapemrosesan teks, berupa tokenisasi dan *part of speech tagging.*
c. Setelah prapemrosesan teks, tahap berikutnya adalah mengklasifikasikan *sentiment* dari tiap data opini/*review*. Proses pengklasifikasian menggunakan Sentiwordnet sebagai basis data leksikal. *Part of speech* (jenis kata kerja, kata sifat, kata keterangan) dari setiap kata digunakan untuk memilih *synset* yang relevan. *Synset* yang paling sering muncul adalah *synset* yang akan dipilih. Nilai *sentiment* dihitung setelah skor *sentiment* untuk setiap kata dalam data opini diperoleh. Kemudian klasifikasi *sentiment* untuk setiap kalimat atau data opini diperoleh dari perbandingan total skor positif dan skor negatif. Bila skor positif lebih banyak dari skor negatif maka kalimat tersebut diklasifikasikan memiliki *sentiment* positif.

Gambar 1 menunjukkan metode berbasiskan leksikal yang digunakan oleh Pamungkas dan Putri.

JURNAL TEKNOLOGI INFORMASI
Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika
[E-ISSN 2656-0321]
[Vol. 12 No. 2]
[Agustus 2018]



Gambar 1. Diagram Sistem Sentiment Analysis oleh Pamungkas dan Putri[1]

Gambar 1. Diagram Metode Sentiment Analysis berbasikan Leksikal oleh Buyung dan Putri

Pamungkas dan Putri membangun sistem berbasiskan website dan melakukan percobaan terhadap 553 data berupa review atau opini pengguna produk. Hasil pengujiannya menunjukkan nilai akurasi rata-rata dari seluruh data adalah 0,68.

2.2 *Sentiment Analysis* Pada *Public Complaint* Berbahasa Indonesia

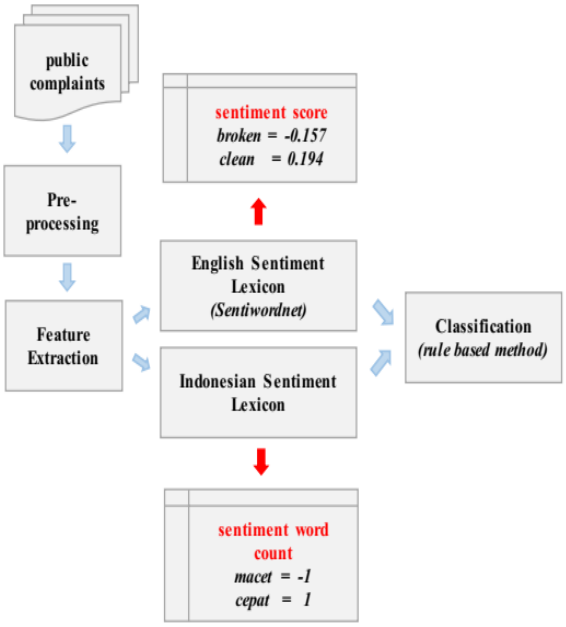
Lailiyah dkk[2] melakukan penelitian untuk menganalisis *sentiment* dari opini masyarakat terhadap pelayanan publik maupun kebijakan-kebijakan Pemerintahan. Opini berupa keluhan, pendapat atau pujian dari masyarakat dapat berkontribusi bagi pengambilan kebijakan di berbagai sektor pemerintahan. Untuk mengumpulkan tanggapan atau opini dari masyarakat, berbagai instansi pemerintahan di Indonesia telah menggunakan media sosial*online* maupun website resmi guna menampung opini masyarakat.

Basis data leksikal berisi koleksi *sentiment* berbahasa Indonesia dan berbahasa Inggris yaitu sentiwordnet, adalah komponen utama sistem *sentiment analysis* pada penelitian Lailiyan dkk[2]. Penggunaan basis data leksikal berbahasa Indonesia pada penelitian ini memberikan gambaran keunikan cara masyarakat menyampaikan opini atau emosinya di dalam teks. Beberapa orang Indonesia menggunakan bahasa yang sopan untuk menyampaikan keluhan atau kekecewaannya, sehingga menyebabkan polaritas *sentiment* menjadi ambigu. Ambiguitas ini terjadi karena sebuah kalimat bisa memiliki lebih dari satu polaritas *sentiment*.

Keunikan juga terjadi pada pilihan kata yang digunakan oleh masyarakat pada saat menggunakan media sosial *online* dan website resmi pemerintah untuk menyampaikan opini. Pada media sosial *online* seperti Facebook atau Twitter, masyarakat seringkali menggunakan bahasa yang tidak resmi, tidak jelas topiknya dan cenderung berisi *sentiment* negatif. Sedangkan pada forum website resmi pemerintah, masyarakat cenderung menggunakan bahasa yang resmi, topik pembicaraan yang jelas dan menggunakan kata-kata yang sopan.

Basis data leksikal berbahasa Indonesia yang digunakan pada penelitian ini belum memiliki skor polaritas (Pos, Neg, Obj) seperti di dalam Sentiwordnet. Lailiyah dkk[2] menggunakan pendekatan semantic di dalam penelitiannya. Gambar 2 menunjukkan kerangka kerja dari sistem Sentiment Analysis pada penelitian ini.

JURNAL TEKNOLOGI INFORMASI
Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika
[E-ISSN 2656-0321]
[Vol. 12 No. 2]
[Agustus 2018]



Gambar 2. Diagram Sistem Sentiment Analysis oleh Lailiyah, dkk. [2]

Pada penelitian Lailiyah dkk, data set berisi koleksi *public complaint* melewati prapemrosesan teks terlebih dahulu. Kemudian hasil ekstraksi prapemrosesan teks diproses menggunakan dua database leksikal, untuk menghitung jumlah skor Pos, Neg dan Obj dari Sentiwordnet dan menghitung jumlah kata *sentiment* positif, negatif dan netral dari basis data leksikal berbahasa Indonesia. Proses terakhir adalah pengklasifikasian sentiment dari kalimat menggunakan *rule based method.*

Hasil analisa dari pengujian sistem menunjukkan akurasi Sentiwordnet untuk mengklasifikasikan *sentiment* pada public complaint adalah 47% pada media sosial online Twitter dan 56,85% pada website resmi pemerintah. Sedangkan pada basis data leksikal berbahasa Indonesia menunjukkan akurasi 65,4% pada media Twitter dan 81,4 % pada website resmi pemerintah.

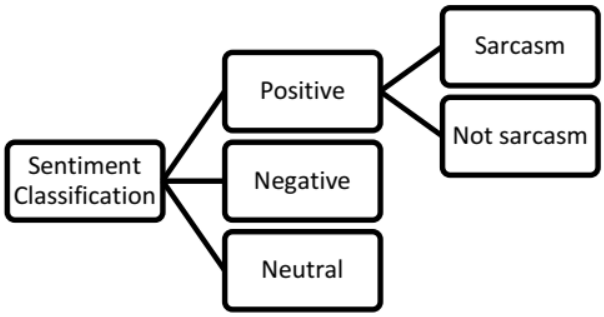
2.3 Sentiment Analysis Untuk Mendeteksi Sarkasme Pada Teks Berbahasa Indonesia

Menurut penelitian Lunando dan Purwarianti[3], masyarakat Indonesia cenderung menggunakan majas sarkasme di media sosial *online* ketika mengkritik suatu topik. Sehingga mendeteksi sarkasme adalah salah satu masalah yang rumit dalam sentiment analysis, karena sarkasme seringkali direpresentasikan dalam bentuk ironi, yaitu kalimat yang bermakna sebaliknya.

Lunando dan Purwariati[3] menggunakan unigram dan faktor pragmatis seperti *smileys* atau *emoticon* untuk mendeteksi sarkasme di dalam teks berbahasa Indonesia. Lunando dan Purwarianti menggunakan Sentiwordnet sebagai basis data leksikal dan beberapa algoritma *machine learning* seperti Naive Bayes, Maximum Entropy dan Support Vector Machine untuk pengklasifikasian.

Gambar 3 menunjukkan mekanisme sistem pengklasifikasian untuk mengidentifikasikan sarkasme di dalam teks berbahasa Indonesia.

JURNAL TEKNOLOGI INFORMASI
Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika
[E-ISSN 2656-0321]
[Vol. 12 No. 2]
[Agustus 2018]



Gambar 3. Komponen Pengklasifikasian *Sentiment* oleh Lunando dan Purwarianti[3]

Pada penelitian Lunando dan Purwariati[3], pengklasifikasian *sentiment* dilakukan dalam dua tahap. Tahap pertama mengklasifikasikan teks ke dalam tiga kelas sentiment: *Positive*, *Negative* dan *Neutral*. Tahap kedua adalah mengklasifikasikan teks sarkasme di dalam kelas *Positive*.

Hasil dari salah satu percobaan dalam penelitian Lunando dan Purwariati[3] menunjukkan performa Sentiwordnet dalam pengklasifikasian. Skor *sentiment* dari Sentiwordnet menunjukkan akurasi yang lebih tinggi dalam pengklasifikasian dibanding hanya menggunakan basis data leksikal. Skor dari Sentiwordnet dapat membedakan kata dengan skor rendah dan kata dengan skor tinggi dalam polaritas sentiment (Pos, Neg, Obj). Tabel 1 menunjukkan akurasi pengklasifikasian dengan basis data leksikal saja dan basis data leksikal Sentiwordnet pada ketiga algoritma *machine learning.* Akurasi menggunakan Skor sentiment rata-rata meningkat 4,3% dibandingkan menggunakan basis data leksikal saja.

Tabel 1. HasilPengklasifikasian Dengan Basis Data Leksikal Saja Dan Basis Data Leksikal Sentiwordnet[3].

Algorithm	Lexical value	Sentiment Score
Naïve bayes	73.1%	77.4%
Maximum Entropy	73.2%	78.4%
Support Vector Machine	74.3%	77.8%

3. Hasil dan Pembahasan

Tabel 2 menunjukkan perbandingan hasil survey pada tiga artikel penelitian. Ketiga penelitian yang menjadi obyek survey, menggunakan basis data leksikal Sentiwordnet, tetapi memiliki jumlah data, sumber data, topik dan hasil pengujian yang berbeda.

Tabel 2. Perbandingan Ketiga Penelitian

Sita-si Arti-kel	**Topik**	**Data**	**Sumber Data**	**Akurasi**
[1]	*Product review*	553 data user opinion	Google Play Store dan Apple App Store	0,679
[2]	*Public Complaint*	1357 data	Twitter dan Website Resmi Pemerintah	47 % dari media Twitter 56,85% dari Website Resmi Pemerintah
[3]	*Sarcasm*	502 training data	Twitter	Rata-rata 77,87%

	300 testing data

Tabel 2 menunjukkan nilai akurasi yang tinggi untuk penelitian [1][3] walaupun desain sistem dan topik pada Sentiment Analysis teks berbahasa Indonesis pada penelitian-penelitian itutidak saling terkait. Akurasi dari kinerja Sentiwordnet untuk mengklasifikasikan *sentiment* teks berbahasa Indonesia masih perlu dianalisa lagi. Walaupun begitu hasil survey ini menunjukkan kinerja yang baik dari basis data leksikal dengan skor polaritas sentiment. Skor polaritas positif, negatif dan netral dari suatu kata dapat berkontribusi untuk meningkatkan akurasi, sehingga hal ini dapat memotivasi untuk melakukan penelitian lebih lanjut guna mengembangkan basis data leksikal untuk menganalisis sentiment pada teks berbahasa Indonesia.

4. Kesimpulan

Penelitian ini telah melakukan survey dan membandingkan hasil pengujian dari tiga artikel penelitian. Hasil survey ini menunjukkan kinerja dari Sentiwordnet sebagai basis data leksikal untuk mendukung sentiment analysis pada teks berbahasa Indonesia. Kinerja dari Sentiwordnet menunjukkan bahwa pemberian skor sentiment pada suatu kata dapat berkontribusi untuk meningkatkan akurasi klasifikasi. Sehingga hasil survey ini memberikan ide untuk mengembangkan lanjut basis data leksikal bagi sentiment analysis teks berbahasa Indonesa.

Daftar Pustaka

[1] Endang Wahyu Pamungkas, Divi Galih Prasetyo Putri, *An Experimental Study of Lexicon-based Sentiment Analysis on Bahasa Indonesia*, 2016. Proceedings of 6th Annual Engineering Seminar (InAES), Yogyakarta, Indonesia

[2] Lailiyah M, Sumpeno S, Purnama I.K.E, *Sentiment Analysis of Public Complaint Using Lexical Resources Between Indonesian Sentiment and Sentiwordnet,* 2017.Proceedings of International Seminar on Intelligent Technology and Its Application 2017, IEEE

[3] Edwin Lunando, Ayu Purwarianti, 2013, *Indonesian Social Media Sentiment Analysis with Sarcasm Detection,* 2013. Jurnal Sarjana Institut Teknolodi Bandung Bidang Teknik Elektro dan Informatika, Juni 2013

[4] Anju Rose G. Punneliparambi, Latest Trends in Sentiment Analysis-A Surver, 2017. International Journal of Computer Science and Engineering Communication.

[5] Stefano Baccianella, Andrea Esuli, Fabrizio Sebastiani, *Sentiwordnet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining*. 2010, Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May 2010, Valletta, Malta