



Analisis Sentimen Tokoh Politik pada Situs Berita Menggunakan NER. Studi Kasus: IMMC

Juan Ortega¹, Goenawan Brotosaputro²

¹Magister Ilmu Komputer, Program Pascasarjana, Universitas Budi Luhur

²Fakultas Teknologi Informasi, Universitas Budi Luhur

¹juan.invasi@gmail.com, ²goenawan.brotosaputro@budiluhur.ac.id

Abstract

In the political world, decisions made by the media can be a measuring instrument of the image of a character. In a news created by a news site, it can be categorized as positive and negative news. Currently there are only a few applications that can see and record the news of a character on a news site. Percentage count can be representative of news sites, it can show positive and negative results from rticle on those sites. To classify a news site, need sentiment analysis of each article on the news site. The sentiment analysis results of each article will affect the percentage count. In general, the analysis is done using preprocessing text which is compared with the word sentiment. However, the preprocessing process and the sentiment word are not appropriate if used to analyze the sentiments of an article with using bahasa. Named Entity Recognition (NER) is part of the word extracted from a collection of texts. NER can be used to extract positive and negative words. In this study, each article from a news site will be analyzed using Named Entity Recognition (NER). The results of sentiment analysis are validated by users. In this study, from 10 test data (articles), the accuracy of sentiment analysis with NER was 90%. While the sentiment analysis using sentiment word and Preprocessing is only 80%

Keywords: *Sentiment Word, Named Entity Recognition, NER*

Abstrak

Dalam dunia politik penilaian yang dibentuk oleh sebuah media dapat menjadi sebuah barometer citra dari seorang tokoh. Dalam sebuah berita yang dibuat oleh situs berita, dapat dikategorikan terdapat berita yang positif dan negatif. Saat ini masih sedikit aplikasi yang dapat melihat dan mencatat pemberitaan seorang tokoh pada sebuah situs berita. Kecenderungan situs berita dapat diukur dengan persentase positif dan negatif terhadap situs tersebut. Untuk melakukan persentase kecenderungan suatu situs berita, diperlukan sentimen analisis dari tiap artikel pada situs berita. Hasil sentimen analisis dari tiap artikel akan mempengaruhi nilai persentase yang dihitung. Pada umumnya sentimen analisis dilakukan menggunakan preprocessing text yang dibandingkan dengan sentimen word. Namun proses preprocessing dan sentimen word diduga tidak tepat bila digunakan untuk menganalisa sentimen suatu artikel berbahasa Indonesia. Named Entity Recognition (NER) adalah suatu bagian dari kata yang diekstraksi dari kumpulan teks. NER dapat digunakan untuk mengekstraksi kata positif dan negatif. Pada penelitian ini tiap artikel dari situs berita akan dianalisis menggunakan Named Entity Recognition (NER). Hasil dari sentimen analisis divalidasi oleh user. Pada penelitian ini dari 10 data uji (artikel) menunjukkan bahwa tingkat akurasi ketepatan sentimen analisis dengan NER sebesar 90%. Sedangkan tingkat akurasi ketepatan sentimen analisis untuk tiap artikel dengan sentimen Word dan Preprocessing hanya sebesar 80%.

Kata kunci: *Sentimen Word, Named Entity Recognition, NER*

1. Pendahuluan

Dalam dunia politik penilaian yang dibentuk oleh sebuah media dapat menjadi sebuah barometer citra dari seorang tokoh. Indonesia Media Monitoring Center (IMMC) adalah sebuah pusat kajian terhadap pola pemberitaan yang berkembang di berbagai media massa di Indonesia. IMMC melihat bahwa saat ini media monitoring juga harus mengikuti zaman dengan melakukan penilaian dengan media berupa situs berita. Saat ini ketika team IMMC melakukan sentimen analisis, mereka menghitung jumlah dari ekspresi kalimat positif dan ekspresi dari kalimat negatif. Sentimen analisis pada umumnya melakukan preproses seperti normalisasi, stemming dan stop wording. Proses ini akan menghilangkan imbuhan dan beberapa kata

yang tidak perlu dilakukan analisa. Proses ini berupaya memecahkan masalah untuk memahami bahasa alami manusia[1]. Dalam beberapa kalimat preproses dapat menghilangkan arti sebenarnya dari sebuah kalimat dalam bahasa Indonesia. Named Entity Recognition (NER) merupakan komponen dari ekstraksi informasi yang berfungsi untuk mengenali entitas nama (nama orang, lokasi, organisasi), ekspresi waktu (tanggal, waktu, durasi) dan ekspresi bilangan (uang, persen, numerik, kardinal) pada kumpulan teks. NER dapat mengenali bahasa manusia hingga 97% [2], namun hasil klasifikasi Named Entity memiliki nilai precision sebesar 34,22% bila digabung dengan preprocessing[3]. Dengan menerapkan NER untuk mengenali entitas kalimat positif dan negatif tanpa

menerapkan processing diharapkan program dapat melakukan analisa berita dengan hasil yang menyamai hasil analisa tim IMMC..

2. Metodologi Penelitian

Metodologi yang digunakan dalam penelitian ini adalah experimental research. Hal yang melandasi penelitian ini menggunakan metodologi experimental research karena adanya pembagian objek yang diteliti menjadi dua kelompok yaitu dengan metode NER dan sentimen word. Sentimen word menerapkan preprocessing dan digunakan sebagai metode pembandingan terhadap NER. Model yang dihasilkan dari kedua algoritma tersebut diukur secara kuantitatif kemudian dibandingkan performansinya. Pada akhir percobaan tiap metode di evaluasi dari akurasi yang didapatkan berdasarkan data validasi berupa kuesioner dari user. Penelitian ini menggunakan sampel yang diambil dari populasi yang jelas, dan teknik pengambilan sampel dari populasi dilakukan secara cluster random sampling. Cluster berdasarkan situs berita dan random dengan hanya mengambil 10 artikel. Hasil dari sentimen analisis terhadap artikel digunakan sebagai input untuk melakukan penghitungan kecenderungan secara persentase

2.1. Metode Pengumpulan Data

Data atau informasi yang diperoleh pada penelitian ini berasal dari data primer dan data sekunder. Untuk data primer, teknik pengumpulan data dilakukan dengan wawancara. Data dan informasi yang diperoleh melalui studi pustaka bersifat sekunder yaitu data yang diperoleh melalui studi literatur, tulisan ilmiah tentang sentimen analisis, dan tinjauan studi. Data yang dibutuhkan dalam penelitian ini dikumpulkan melalui:

a. Penelitian Kepustakaan (Library Research)

Metode ini dilakukan untuk mengumpulkan data dengan mencari dan membaca buku yang dapat menunjang penyusunan penelitian ini serta mencari data di internet guna memperoleh gambaran secara keseluruhan dari permasalahan yang nantinya data digunakan sebagai referensi.

b. Studi Literatur

Hal ini dilakukan dengan cara menelusuri data dan informasi yang ada dan menelaahnya secara tekun, dengan cara membaca buku-buku, jurnal, penelitian, maupun referensi lainnya yang dapat dijadikan acuan pembahasan dalam masalah ini. Metode yang akan digunakan adalah Named Entity Recognition (NER) dan sentiment word. Dari studi literatur peneliti juga akan mendapatkan contoh sampel data latih untuk kata positif dan negatif yang akan dibandingkan dengan hasil NLP.

c. Wawancara

Peneliti mewawancarai tim dari IMMC News (Indonesia Media Monitoring Center) untuk

mengetahui kebutuhan dan kriteria dalam melakukan sebuah sentimen analisis. Dari wawancara peneliti juga akan mendapatkan data latih untuk kalimat positif dan kalimat negatif. Naskah ditulis dalam ukuran kertas A4 dengan jumlah halaman minimum 6 halaman, maksimum 15 halaman, termasuk tabel dan gambar, serta dengan mengacu tata cara penulisan seperti telah yang disusun pada tulisan ini.

2.2. Metode Kuantitatif

Metode penelitian kuantitatif merupakan salah satu jenis penelitian yang spesifikasinya adalah sistematis, terencana, dan terstruktur dengan jelas sejak awal hingga pembuatan desain penelitiannya. Definisi lain menyebutkan penelitian kuantitatif adalah penelitian yang banyak menuntut penggunaan angka, mulai dari pengumpulan data, penafsiran terhadap data tersebut, serta penampilan dari hasilnya. Penelitian kuantitatif adalah suatu proses menemukan pengetahuan yang menggunakan data berupa angka sebagai alat menganalisis keterangan mengenai apa yang ingin diketahui [4]. Data Kuantitatif adalah data yang dapat dinyatakan dalam bentuk angka-angka [5].

2.3. Cluster Random Sampling

Cluster Sampling adalah teknik sampling secara berkelompok. Pengambilan sampel jenis ini dilakukan berdasar kelompok / area tertentu. Tujuan metode Cluster Random Sampling antara lain untuk meneliti tentang suatu hal pada bagian-bagian yang berbeda di dalam suatu object. Sampel adalah bagian dari jumlah dan karakteristik yang dimiliki oleh populasi tersebut [6]. Sampel adalah bagian dari populasi yang mempunyai ciri-ciri atau keadaan tertentu yang akan diteliti. Pengambilan sampel dalam penelitian ini dilakukan dengan metode "cluster random sampling" [6]. Yang mana metode cluster random sampling/area adalah teknik pengambilan sampel yang dipilih berdasarkan area-area tertentu [6]

2.4. Named Entity Recognition (NER)

NER adalah metode untuk melakukan ekstraksi dan klasifikasi nama ke dalam beberapa kategori dengan mengacu kepada makna yang tepat [7]. NER diimplementasikan dalam banyak bidang, antara lain dalam machine translation, question-answering machine system, indexing pada information retrieval, klasifikasi dan juga dalam automatic summarization. Tujuan yang diharapkan dari proses dalam NER adalah untuk melakukan ekstraksi dan klasifikasi nama ke dalam beberapa kategori dengan mengacu kepada makna yang tepat.

2.5. Pengukuran Sentimen Artikel

Dalam mengukur kecenderungan sentimen dalam sebuah artikel, situs immcnews.com melakukan pengukuran dan perhitungan dari jumlah kata yang memiliki sentimen. Cara menghitung sentimen dalam sebuah artikel yang dilakukan tim immcnews adalah:

Artikel positif = total kata positif >= (total kata positif + total negatif)*55%

Artikel negatif = total kata positif <= (total kata positif + total negatif)*45%.

Namun bila suatu artikel dikatakan tidak memiliki kata yang bersentimen maka artikel tersebut dikategorikan sebagai artikel bersentimen netral.

Tabel 1. Contoh Perhitungan Sentimen Berita

Artikel	Negative	Positive	Total	Persentase Positif	Sentimen
A	4	8	12	67%	Positive
B	2	3	5	60%	Positive
C	1	8	9	89%	Positive
D	5	4	9	44%	Negative
E	8	7	15	46%	Netral

Seperti terlihat dari tabel di atas, sentimen dari artikel dapat dibagi menjadi positive, negative dan netral.

2.6. Metode Pengukuran Kecenderungan Situs Berita

Dalam mengukur kecenderungan situs berita, situs immcnews.com melakukan pengukuran dan perhitungan dari jumlah artikel yg bersentimen. Immcnews akan mengambil artikel berdasarkan situs berita, tokoh dan kriteria yang diinput user. Cara menghitung sentimen dalam sebuah artikel yang dilakukan tim immcnews adalah:

Total Artikel = Total Berita bersentimen Positif + Total berita sesuai Kriteria yang bersentimen Positif + Total Berita bersifat Negatif + Total berita sesuai Kriteria yang bersentimen Negatif

Positif value = (Total Berita bersentimen Positif + Total berita sesuai Kriteria yang bersentimen Positif) / Total artikel

Negatif value = (Total Berita bersentimen Negatif + Total berita sesuai Kriteria yang bersentimen Negatif) / Total artikel.

3. Hasil dan Pembahasan

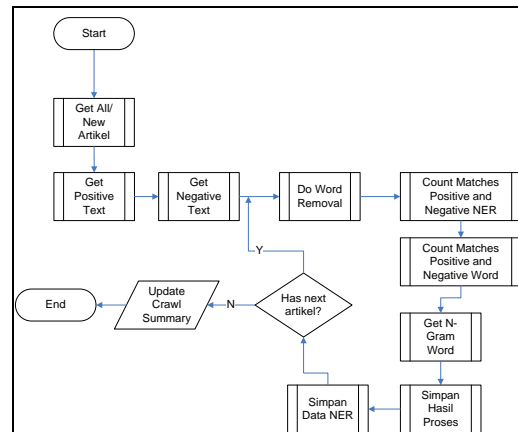
Pada penelitian ini, sistem akan menjalankan proses sentiment analisis menggunakan NER dan sentiment analisis menggunakan sentiment word secara terpisah kepada 10 artikel yang diambil secara acak namun dalam cluster tokoh tertentu.

Saat sistem dijalankan sistem akan melakukan proses penghitungan sentimen berdasarkan word dan NER

3.1. Hasil Sentimen Analisis

Pada penelitian ini akan dilakukan pengujian dari hasil sentimen analisis dengan NER dan sentimen analisis dengan sentimen word. Hasil analisis akan divalidasi dengan hasil analisis yang dilakukan oleh tim IMMC. Hasil dari analisis secara manual oleh tim IMMC didapatkan dari pengisian kuesioner. Dalam ketentuan tim IMMC suatu artikel dikatakan positif bila jumlah minimal nilai positif adalah 55% dan bila jumlah nilai positif tidak lebih dari 45% maka dikatakan artikel

tersebut bersentimen negatif. Data validasi didapat dari kuesioner yang diisi oleh 3 orang perwakilan tim IMMCNews.



Gambar 1. Flowchart Proses Berita

Untuk melakukan pengujian, data dari sentimen analisis perlu diproses menggunakan ketentuan penghitungan dari tim IMMC. Hasil dari proses penghitungan merupakan sentimen dari artikel.

Tabel 2. Nilai Sentimen Analisis dengan NER

Artikel No	Positive NER count	Negative NER count	Persentase NER Positive	Analisis result
1	5	7	41%	Negative
2	1	0	100%	Positif
3	4	1	80%	Positif
4	1	3	25%	Negative
5	9	15	38%	Negative
6	0	11	0%	Negative
7	1	2	33%	Negative
8	4	1	80%	Positif
9	2	1	67%	Positif
10	1	8	11%	Negative

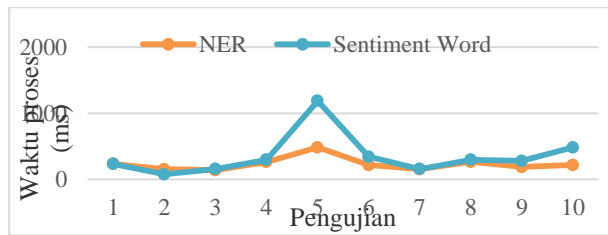
Sebagai metode pembandingan, akan dilakukan penghitungan nilai sentimen terhadap hasil dari sentimen analisis dengan menggunakan sentimen word. Output dari kedua tabel ini akan dibandingkan dengan data validasi dari user sehingga diperoleh hasil berupa persentase kesesuaian

Tabel 3. Nilai Sentimen Analisis dengan Sentiment Word

Artikel No	Positive Word count	Negative Word count	Persentase Word Positive	Analisis result
1	2	16	11%	Negative
2	0	0	50%	Netral
3	5	3	62%	Positif
4	9	6	60%	Positif
5	41	41	50%	Netral
6	0	23	0%	Negative
7	3	6	33%	Negative
8	13	6	68%	Positif
9	16	8	66%	Positif
10	16	22	42%	Negative

Dari 10 kali pengujian analisa sentimen, waktu untuk menjalankan proses bervariasi. Pada proses Sentimen word sistem menjalankan proses preprocessing,

penghitungan sentimen word dan pembuatan rekomendasi entitas. Pada proses NER sistem menggabungkan kata dari database hingga menjadi entitas sentimen. Gambar 2 merupakan grafik kinerja pengujian sentimen analisis yang dilakukan oleh sistem



Gambar 2. Statistik Performa Waktu Proses

3.2. Hasil Kuesioner User

Untuk melakukan validasi diperlukan data yang diisi oleh user secara manual. Kuesioner diberikan ke 3 user dari tim IMMC. Dari 3 user ini artikel akan dikategorikan sebagai kategori positif atau negative bila terjadi draw maka artikel tersebut dikategorikan sebagai artikel netral.

Tabel 4. Sentimen Analisis Melalui Kuesioner User 1

Artikel No	Positive count	Negative count	Persentase Positive	Analisis result
1	3	4	43%	Negative
2	0	0	50%	Netral
3	2	1	66%	Positif
4	0	4	0%	Netral
5	5	9	36%	Negative
6	0	7	0%	Negative
7	0	1	0%	Negative
8	3	0	100%	Positif
9	2	1	66%	Positif
10	0	5	0%	Negative

Tabel 5. Sentimen Analisis Melalui Kuesioner User 2

Artikel No	Positive count	Negative count	Persentase Positive	Analisis result
1	2	4	33%	Negative
2	0	0	50%	Netral
3	2	1	66%	Positif
4	0	6	0%	Negative
5	5	9	36%	Negative
6	0	5	0%	Negative
7	0	0	50%	Negative
8	3	0	100%	Positif
9	2	1	66%	Positif
10	0	2	0%	Negative

Tabel 6. Sentimen Analisis Melalui Kuesioner User 3

Artikel No	Positive count	Negative count	Persentase Positive	Analisis result
1	1	3	25%	Negative
2	1	0	100%	Positif
3	2	1	66%	Positif
4	0	2	0%	Negative
5	3	7	30%	Negative
6	1	7	12%	Negative
7	0	2	0%	Negative
8	2	0	100%	Positif
9	2	1	66%	Positif
10	0	5	0%	Negative

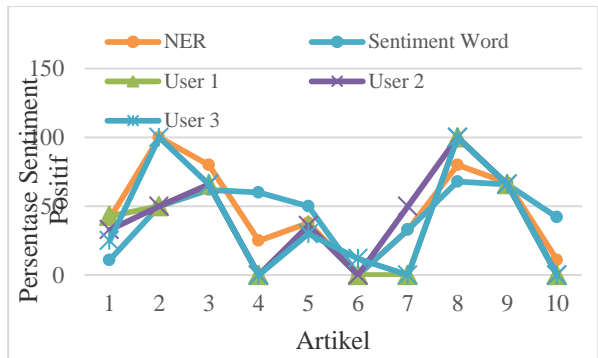
Dari hasil kuesioner tersebut akan dibuat data validasi dengan mempertimbangkan jumlah user yang paling banyak pada suatu sentimen untuk artikel tersebut. Hasil dari klasifikasi tersebut diambil dari musyawarah perolehan sentimen terbanyak. Data validasi dapat dilihat pada tabel 7

Tabel 7. Data Validasi

Artikel No	User 1	User 2	User 3	Data Validasi
1	Negative	Negative	Negative	Negative
2	Netral	Netral	Positif	Netral
3	Positif	Positif	Positif	Positif
4	Negative	Negative	Negative	Negative
5	Negative	Negative	Negative	Negative
6	Negative	Negative	Negative	Negative
7	Negative	Negative	Negative	Negative
8	Positif	Positif	Positif	Positif
9	Positif	Positif	Positif	Positif
10	Negative	Negative	Negative	Negative

3.3. Hasil Validasi Sentimen Analisis

Bila dilihat secara statistik kedua metode dapat menyamai hasil klasifikasi yang dilakukan secara manual, namun Named Entity Recognition memiliki perbedaan yang tidak terlalu jauh untuk setiap artikel. Kesalahan terjadi hanya pada artikel ke 2 dimana dengan selisih hanya 1 sentimen



Gambar 2. Statistik Kesesuaian Analisa Sentiment

Setelah data validasi didapatkan, hasil sentimen analisis dengan menggunakan NER dan sentimen word akan dibandingkan dengan data validasi. Perbandingan hasil analisis dapat dilihat pada tabel 8.

Dari tabel 8 dapat dilihat bahwa analisis berdasarkan NER memiliki nilai kecocokan sebesar 9/10 atau 90% dan analisis berdasarkan Sentimen Word memiliki nilai kecocokan sebesar 8/10 atau 80%. Angka ini dipengaruhi oleh jumlah dan ketepatan kamus data dan stop word list.

4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan maka dapat ditarik kesimpulan bahwa Named Entity Recognition (NER) dapat memberikan klasifikasi kecenderungan situs berita dengan keakuratan 90%. Hal ini terlihat dari hasil percobaan sentimen analisis terhadap 10 artikel. Named Entity Recognition

memiliki persentase persamaan mencapai 90% dengan error selisih hanya 1 entity. Persamaan yang paling sesuai adalah pada artikel ke 5 dimana hanya terjadi perbedaan sebesar 2% untuk artikel tersebut. Proses analisis tidak lebih dari 500ms sehingga memungkinkan untuk melakukan analisa berulang dengan parameter yang berbeda. Setiap ada penambahan entitas baru sistem dapat melakukan sentimen analisa kembali. Tingkat akurasi dari sentimen analisis dapat dikembangkan untuk mencapai tingkat akurasi yang lebih tinggi.

Tabel 8. Perbandingan Sentimen Analisis

No	NER	Sentimen Word	Data Validasi	Status analisis dengan NER	Status analisis dengan Word
1	Negative	Negative	Negative	Sesuai	Sesuai
2	Positif	Netral	Netral	Tidak Sesuai	Sesuai
3	Positif	Positif	Positif	Sesuai	Sesuai
4	Negative	Positif	Negative	Sesuai	Tidak Sesuai
5	Negative	Netral	Negative	Sesuai	Tidak Sesuai
6	Negative	Negative	Negative	Sesuai	Sesuai
7	Negative	Negative	Negative	Sesuai	Sesuai
8	Positif	Positif	Positif	Sesuai	Sesuai
9	Positif	Positif	Positif	Sesuai	Sesuai
10	Negative	Negative	Negative	Sesuai	Sesuai
Ketepatan				9/10	8/10

Ucapan Terimakasih

Peneliti mengucapkan terimakasih kepada Tim IMMC News yang bersedia menyisihkan waktu untuk membantu penelitian yang dilakukan.

Daftar Rujukan

- [1] Ratnawati F. 2018. Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. Politteknik Negeri Bengkalis, Riau.
- [2] Christianto D, Siswanto E, Chaniago R. 2015. Penggunaan named entity recognition dan artificial intelligence markup language untuk penerapan chatbot berbasis teks. Penggunaan named entity recognition dan artificial intelligence markup language untuk penerapan chatbot berbasis teks hal 61-68.
- [3] Kurniawati T, Indriarti, Adikara P.P. 2018. Implementasi named entity recognition pada factoid question answering system untuk cerita rakyat indonesia. Jurnal pengembangan teknologi informasi dan ilmu komputer vol. 2, no. 9, september 2018, hal. 3142-3149.
- [4] Kasiram, M. 2008. Metode Penelitian Kuantitatif-Kualitatif. Malang:UIN Malang Press.
- [5] Noor, Juliansyah. 2014. Metodologi Penelitian. Jakarta: Kencana Prenada Media Group
- [6] Sugiyono, 2013, Metodologi Penelitian Kuantitatif, Kualitatif Dan R&D. (Bandung: ALFABETA)
- [7] Dirgantara, et al.,2018. Penerapan Named Entity Recognition Untuk Mengenali Fitur Produk Pada E-Commerce Menggunakan Rule Template Dan Hidden Markov Model, Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer