

Sarcasm in Sentiment Analysis of Indonesian Text: A Literature Review

Sherly Christina^{a,1*}

^aUniversitas Palangka Raya, Jl. Yos Sudarso Kel. Palangka Kec. Jekan Raya, Palangka Raya, Kalimantan Tengah

¹sherly.christina.upr@gmail.com*

*corresponding author

ARTICLE INFO

Keywords

Opinion
Indonesian Text
Sarcasm
Sentiment analysis

ABSTRACT

Social media, blogs and online groups become a forum that makes it easy for the Indonesian people to express their opinions, suggestions, complaints and even criticisms of a subject liberally. Sentiment analysis is a method for classifying positive, neutral, and negative polarity of the opinions that expressed by the internet users. Sarcasm is one of the challenges to classifying the sentiments of an opinion. This research is a literature review to examine several studies to find out the methods for detecting sarcasm and to know the effect of sarcasm on the sentiment classification accuracy. The result of this literature review can be used as a reference for developing the sarcasm detection methods.

1. Pendahuluan

Internet dan teknologi komunikasi memberikan kemudahan bagi setiap orang untuk menyampaikan pendapat, aspirasi, ide maupun kritik secara *online*. Media sosial, blog, maupun forum-forum online menjadi antarmuka bagi pengguna internet untuk menanggapi suatu peristiwa, mengomentari suatu produk, menyampaikan persepsi terhadap suatu objek secara bebas dan mudah. Seiring dengan berjalannya waktu dan besarnya intensitas aktivitas pengguna internet pada media sosial online atau berbagai fitur komunikasi online, maka inputan dari para pengguna tersebut berakumulasi menjadi *big data* yang potensial bagi pihak-pihak yang berkepentingan.

Tersedianya *big data* ini membuka era baru bagi *sentiment analysis*, yang bertujuan untuk menganalisis pendapat atau sikap seseorang. *Sentiment analysis* memungkinkan mesin komputer untuk mengkategorikan secara otomatis polaritas suatu pendapat mengenai suatu subyek menjadi positif, netral, atau negatif [1].

Sentiment analysis memiliki banyak kesulitan dalam penerapannya, salah satu diantaranya adalah mendeteksi sarkasme dari suatu teks. Sarkasme adalah bentuk penyampaian pendapat atau sikap seseorang yang disampaikan dalam gaya bahasa ironi. Sarkasme mengubah polaritas dari suatu teks menjadi kebalikan dari isi teks tersebut [2]. Kegagalan mengidentifikasi sarkasme di dalam teks dapat mempengaruhi hasil pengklasifikasian *sentiment analysis* dari suatu himpunan data.

Jika seseorang menyampaikan komentar yang berisikan sarkasme secara langsung dihadapan orang lain, maka sentimen dari komentar tersebut dapat dipahami dengan mudah dengan mengamati ekspresi wajah dan gerak tubuh dari pembicara. Tetapi memahami komentar sarkasme yang disampaikan dalam bentuk teks digital adalah pekerjaan yang sulit [3].

Sarkasme adalah gaya bahasa yang mengandung ejekan maupun kritikan [4]. Para pengguna internet di Indonesia yang disebut netizen biasa menggunakan sarkasme untuk menyampaikan

pendapat atau sikap yang cenderung negatif terhadap topik pembicaraan. Berikut ini adalah contoh beberapa contoh komentar negatif netizen dalam akun instagram Lambe Turah [5].

- *Sinetron jaman skg gda yg bermutu.... Mending balik k jaman KELUARGA CEMARA*
- *mending nonton FTV yg judulnya kayak kereta panjang mbulet .. drpd nonton ini*
- *Muka sudah cakep2 di permak jadi ninja hatori.. sungguh miris*
- *mukanya kaya prisma sgitiga apa kubus trapezium. Ini dokternya ga ngerti estetika asal tarik tirus aja gitu ya*

Contoh-contoh tersebut menunjukkan sarkasme adalah cara seseorang mengungkapkan emosinya melalui pesan yang tersirat. Biasanya sarkasme yang disampaikan dengan berhadapan muka dengan pendengar mudah untuk dipahami karena diiringi dengan ekspresi wajah, gerakan tubuh dan nada suara yang menyiratkan makna negatif. Sedangkan sarkasme yang disampaikan dalam bentuk teks dapat menyebabkan pembaca keliru memahami sentimen yang dimaksud oleh penulis. Ciri khas dari sarkasme menjadi tantangan bagi mesin komputer untuk menganalisis sentimen suatu teks secara otomatis dan akurat. Oleh karena itu untuk meningkatkan performa *sentiment analysis*, beberapa penelitian [6][7][8][9][10] telah dilakukan untuk mengembangkan metode pengidentifikasi maupun pengklasifikasi sarkasme dari data berbentuk teks.

Penelitian ini adalah sebuah studi literatur yang mengkaji beberapa penelitian terkait sarkasme di dalam sentimen analysis pada data set berbahasa Indonesia. Penelitian ini mengkaji metode-metode yang telah diterapkan serta hasil-hasil yang dicapainya. Hasil dari penelitian ini dapat menjadi bahan acuan untuk mengembangkan metode-metode pendeteksi sarkasme dalam pengklasifikasian analisis sentimen.

2. Sarkasme

Sarkasme adalah jenis sentimen dimana seseorang mengungkapkan emosi atau tanggapan negatif menggunakan kata-kata yang positif bahkan hiperpositif atau berlebihan. Penyampaian sarkasme dalam bentuk teks seringkali kompleks sehingga bisa saja menyebabkan pembaca salah memahami polaritas dari suatu teks. Gaya bahasa sarkasme yang kompleks terdiri atas beberapa jenis [10], yaitu:

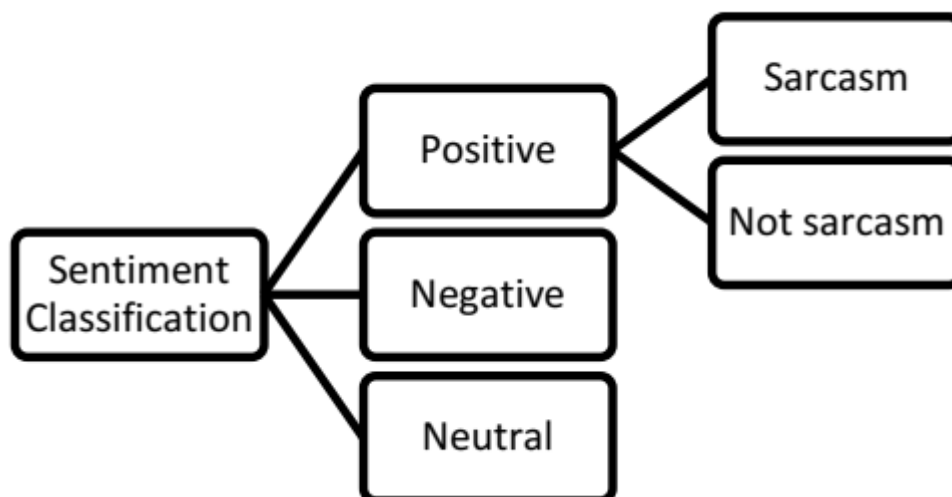
- *Self-deprecating sarcasm*, sarkasme yang menunjukkan rasa tidak penting dan inferioritas.
- *Brooding sarcasm*, sarkasme yang disampaikan dengan cara yang baik dengan nada yang menjengkelkan atau membuat sakit hati.
- *Deadpan sarcasm*, sarkasme yang disampaikan tanpa ekspresi atau emosi, sehingga sulit untuk dipahami.
- *Polite sarcasm*, sarkasme tidak langsung, yaitu sarkasme yang semula terdengar sopan dan otentik, tetapi pada akhirnya menjadi jelas maksudnya.
- *Obnoxious sarcasm*, sarkasme yang disampaikan dengan nada suara cengeng.
- *Manic sarcasm*, sarkasme yang mengekspresikan kegembiraan dengan cara yang tidak wajar.
- *Raging sarcasm*, sarkasme yang disampaikan secara hiperbola dan berisi ancaman kekerasan.

Struktur, karakteristik maupun gaya bahasa Indonesia yang unik dibanding bahasa Inggris maupun bahasa asing lainnya merupakan salah satu tantangan untuk mengidentifikasi sarkasme di dalam teks berbahasa Indonesia. Di dalam bahasa Indonesia, sarkasme merupakan gaya bahasa yang rumit dan juga seringkali disampaikan dengan kata-kata yang tidak formal. Sarkasme biasanya digunakan oleh netizen di Indonesia untuk mengkritik seseorang, produk, peristiwa, aturan bahkan kebijakan yang berlaku. Pada subbab berikut akan dikaji beberapa penelitian yang telah dilakukan terhadap data set berbahasa Indonesia.

3. Mendeteksi Sarkasme dalam Teks Berbahasa Indonesia

Sarkasme adalah gaya bahasa yang memiliki polaritas negatif serta sulit dideteksi secara otomatis oleh mesin komputer, bahkan oleh manusia. Beberapa penelitian telah dilakukan untuk mendeteksi sarkasme dalam analisis sentimen pada teks berbahasa Indonesia. Lunando dan Purwarianti [2] mengusulkan beberapa fitur untuk mendeteksi sarkasme setelah pengklasifikasian sentimen dilaksanakan pada data teks yang diperoleh dari media sosial Indonesia. Fitur-fitur tersebut diekstrak setelah dilakukan pra pemrosesan teks, beberapa fitur pendeteksi sarkasme tersebut adalah:

- a. *Unigram*, menurut Lunando dan Purwarianti [2] ekstraksi *unigram* dari teks sosial media lebih tepat, karena seringkali tata bahasa Indonesia yang digunakan netizen bervariasi dan tidak formal. *Unigram* yang diambil hanya istilah-istilah yang terjemahnya terdapat di dalam Sentiwordnet. Pada saat menggunakan *unigram* untuk mendeteksi sarkasme, Lunando dan Purwarianti menemukan beberapa fenomena, yaitu:
 - Kata negasi seperti “tidak” cenderung mengubah sentimen dari kata tertentu. Contohnya “Televisi jaman sekarang tidak begitu bagus karena mahal”, kata “bagus” memiliki sentimen positif tetapi pada contoh kalimat tersebut sentimen kata “bagus” berubah menjadi negatif karena didahului dengan kata “tidak”.
 - Pada bahasa Indonesia sentimen suatu kata bisa berubah tergantung dari konteksnya. Contohnya, kata “mahasiswa” adalah kata dengan sentimen netral, tetapi jika didahului oleh kata “harga” menjadi “harga mahasiswa” maka sentimen kata “mahasiswa” berubah menjadi positif.
 - Imbuhan pada bahasa Indonesia bisa mengubah sentimen suatu kata. Contohnya, kata “murah” yang memiliki sentimen positif akan berubah sentimennya menjadi negatif bila diberi akhiran -an menjadi “murahan”.
- b. *Negativity*, adalah fitur yang merepresentasikan persentase sentimen negatif dari suatu topik.
- c. *Number of interjection words*, fitur yang menunjukkan jumlah kata seru di dalam teks, contoh kata-kata seru seperti: ‘bah’, ‘nah’, ‘wew’, ‘wow’, ‘yay’, ‘uh’.
- d. *Question word*, fitur kata tanya seperti ‘siapa’, ‘apa’, ‘kapan’, ‘bagaimana’, ‘dimana’, dan ‘kenapa’. Fitur kata tanya mengindikasikan bahwa sentimen suatu teks adalah netral.



Gambar 1. Model Pengklasifikasian Sentimen dari Penelitian Lunando dan Purwarianti

Pada Gambar 1 ditunjukkan pendeteksian sarkasme dilakukan pada data teks yang telah diklasifikasikan bersentimen positif. Lunando dan Purwarianti mengimplementasikan algoritma *machine learning* yaitu Naive Bayes, Maximum Entropy dan Support Vector Machine pada model pengklasifikasian sentimen. Proses pengklasifikasian sentimen dibantu oleh Sentiwordnet, sehingga data set yang berupa teks dari media sosial Indonesia harus diterjemahkan lebih dulu ke dalam bahasa Inggris. Hasil dari penelitian Lunando dan Purwarianti menunjukkan bahwa penambahan fitur untuk mendeteksi sarkasme telah meningkatkan akurasi sebesar 6% pada hasil klasifikasi sentimen.

Menurut Afiyati dkk [11] mengidentifikasi sarkasme dapat membantu para pelaku bisnis memahami kecenderungan dan pendapat sesungguhnya dari konsumen ketika konsumen menyampaikan keluhan maupun umpan balik. Dan untuk politisi yang sedang melakukan kampanye politik, mengidentifikasi sarkasme dapat menjadi bahan evaluasi diri.

Afiyati dkk [11] menganalisis percakapan dari grup Whatsapp dan mengusulkan beberapa fitur untuk mengklasifikasikan pernyataan yang sarkastis dan tidak sarkastis. Data yang dianalisis adalah data berupa teks dari grup whatsapp berbahasa Indonesia. Fitur-fitur yang digunakan untuk mengklasifikasikan pernyataan sarkastis dan tidak sarkastis adalah:

- Fitur *Semantic-related*, fitur ini diproses secara manual. Fitur *semantic-related* mengklasifikasikan teks menjadi kata-kata bersentimen positif dan kata-kata bersentimen negatif. Hasil klasifikasi akan menunjukkan nilai suatu teks yang mengandung sarkasme.
- Fitur *Punctuation-related*, adalah fitur-fitur jumlah tanda seru, tanda tanya, tanda titik, semua huruf kapital, tanda petik, dan jumlah huruf vokal yang dituliskan lebih dari dua kali. Tanda seru dan tanda tanya diindikasikan sebagai sarkasme. Hasil penelitian Afiyati dkk, menunjukkan bahwa fitur *Punctuation-related* bekerja kurang optimal untuk mengidentifikasi sarkasme.
- Fitur *syntactic dan semantic*, adalah fitur-fitur yang menghitung nilai kata-kata yang tidak biasa seperti kata-kata tidak formal dan juga termasuk kata-kata bahasa asing seperti bahasa Inggris, jumlah kata seru dan emotikon tertawa. Hasil penelitian Afiyati dkk, menunjukkan bahwa fitur *syntactic dan semantic* juga bekerja kurang optimal untuk mengidentifikasi sarkasme.
- Fitur *Pattern-related*, adalah fitur karakteristik sarkasme yang diekstrak dari data set. Fitur ini akan memeriksa keberadaan karakteristik dari teks percakapan grup whatsapp yang terkait dengan sarkasme. Hasil penelitian Afiyati dkk, menunjukkan bahwa fitur *pattern-related* bekerja paling optimal dibandingkan fitur lain untuk mendeteksi sarkasme.

Walaupun beberapa fitur yang digunakan dapat mengidentifikasi sarkasme di dalam teks, namun penelitian yang dilakukan oleh Afiyati dkk tidak menyebutkan berapa nilai akurasi yang diperoleh dari fitur-fitur pendeteksi sarkasme. Afiyati dkk juga menyatakan bahwa keakuratan metode tidak memadai, karena hanya menggunakan data sampel dan analisis manual percobaan.

Berikutnya penelitian untuk mendeteksi sarkasme dari komentar netizen Indonesia dari media sosial twitter. Yunitasari dkk [12] menggunakan unigram dan empat set fitur Boazizi yang terdiri dari fitur *sentiment-relate*, fitur *punctuation-relate*, fitur *lexical* dan *syntactic*, dan fitur *top words*. Metode untuk mendeteksi sarkasme menggunakan algoritma Random Forest. Kemudian untuk mengekstrak fitur dari data teks bagi proses analisis sentimen, menggunakan metode TF-IDF. Pengklasifikasian polaritas sentimen dari teks menggunakan algoritma Naive Bayes.

Sebuah model dibangun untuk menganalisis sentimen dari kumpulan data *tweet* dengan memperhatikan probabilitas adanya konten sarkasme di dalam suatu *tweet*. Tahapan proses dari model yang telah dibangun oleh Yunitasari dkk adalah :

- i. *Input tweet labeled linguists*
- ii. *Preprocessing*
- iii. *Sarcasm feature extraction*
- iv. *Sarcasm detection*
- v. *Testing algorithm result and linguist*
- vi. *The result of the sarcasm detection test.*

Jika model mengklasifikasikan sentimen sebuah *tweet* positif, tetapi di dalamnya terdeteksi mengandung sarkasme, maka polaritas dari sentimen *tweet* diubah menjadi negatif.

Hasil dari penelitian Yunitasari dkk menunjukkan model pendeteksi sarkasme meningkatkan kinerja dari model analisis sentimen sekitar 5,49%. Kinerja dari model pendeteksi sarkasme menunjukkan akurasi sebesar 80,4 % dengan nilai presisi 83,2% dan recall 91,3%. Tabel 1 berikut berisi rangkuman dari penelitian yang dilakukan oleh [2][11][12].

Tabel 1. Analisis Perbandingan dari Beberapa Penelitian Terhadap Sarkasme dalam *Sentiment Analysis* Teks Berbahasa Indonesia

Peneliti	Data Set	Metode Pendeteksi Sarkasme	Hasil Pengujian Metode
Edwin Lunando, Ayu Purwarianti	Data teks dari media sosial Indonesia	Menambahkan fitur-fitur pendeteksi sarkasme: <ul style="list-style-type: none"> • Unigram • Negativity • Number of interjection words • Question words 	Fitur-fitur pendeteksi sarkasme meningkatkan akurasi pengklasifikasian sentimen sebesar 6%.
Afiyati, Edi Winarko, Anis Cherid	Data teks dari grup whatsapp Indonesia	Menggunakan fitur-fitur: <ul style="list-style-type: none"> • <i>Semantic-related</i> • <i>Punctuation-related</i> • <i>syntactic dan semantic</i> • <i>Pattern-related</i> 	Fitur-fitur dapat mendeteksi sarkasme tetapi akurasinya belum memadai.
Yessi Yunitasari, Aina Musdholifah, Anny Kartika Sari	Data teks dari media sosial Indonesia, tweeter	Menggunakan fitur-fitur: unigram dan fitur sentiment-relate, fitur punctuation-relate, fitur lexical dan syntactic, dan fitur top words.	Fitur-fitur pendeteksi sarkasme meningkatkan kinerja pengklasifikasian sentimen sebesar 5,49%. Dan kinerja dari fitur-fitur pendeteksi sarkasme menunjukkan akurasi sebesar 80,4 % dengan nilai presisi 83,2% dan recall 91,3%.

4. Kesimpulan

Hasil pengujian dari beberapa penelitian yang dibahas pada studi literatur ini menunjukkan sarkasme di dalam teks berbahasa Indonesia dapat mempengaruhi akurasi dari sentimen analisis. Metode-metode yang telah diuji pada penelitian [2][11][12] dapat berkontribusi untuk mengidentifikasi sarkasme di dalam teks berbahasa Indonesia dan dapat dikembangkan lagi untuk meningkatkan optimasi dari pendeteksi sarkasme.

Daftar Pustaka

- [1] S. Christina, "Opinion Mining Framework Berbasis Leksikal," *J. Teknol. Inf.*, vol. 12, no. 1, 2018.
- [2] E. Lunando and A. Purwarianti, "Indonesian Social Media Sentiment Analysis with Sarcasm Detection," *J. Sarj. Inst. Teknol. Bandung Bid. Tek. Elektro dan Inform.*, 2013.
- [3] S. Masroor and M. S. Husain, "Sarcasm Analysis Using Social Media : A Literature Review," *Int. J. Eng. Manag. Res.*, pp. 26–31, 2018.
- [4] A. Aflikah O., "Gaya Bahasa Sarkasme dan Kekhasan Bahasa Penulis pada Judul Rubrik

- Kriminal di Surat Kabar Harian Meteor Edisi April 2012,” Universitas Muhammadiyah Surakarta, 2012.
- [5] E. Inderasari, F. Achsani, and B. Lestari, “BAHASA SARKASME NETIZEN DALAM KOMENTAR AKUN INSTRAGRAM ‘ LAMBE TURAH ,”” *J. Semant.*, vol. 8, no. 1, 2019.
- [6] S. K. Bharti, K. S. Babu, and S. K. Jena, “Parsing-based Sarcasm Sentiment Recognition in Twitter Data,” in *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining Parsing-based*, 2015, pp. 1373–1380.
- [7] M. Bouazizi and T. Ohtsuki, “Opinion mining in Twitter: How to make use of sarcasm to enhance sentiment analysis,” in *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2015.
- [8] D. K. Tayal, S. Yadav, K. Gupta, B. Rajput, and K. Kumari, “Polarity detection of sarcastic political tweets,” in *2014 International Conference on Computing for Sustainable Global Development (INDIACom)*, 2014.
- [9] D. Al-Ghadhban, E. Alnkhilan, L. Tatwany, and M. Alrazgan, “Arabic sarcasm detection in Twitter,” in *2017 International Conference on Engineering & MIS (ICEMIS)*, 2017.
- [10] K. Rajeswari and P. Shanthibala, “Recognition of Sarcastic Emotions of Individuals on Social Network,” *Int. J. Pure Appl. Math.*, vol. 118, no. 7, pp. 253–259, 2018.
- [11] Afiyati, E. Winarko, and A. Cherid, “Recognizing the Sarcastic Statement on WhatsApp Group with Indonesian Recognizing the Sarcastic Statement on WhatsApp Group with Indonesian Language Text,” in *2017 International Conference on Broadband Communication, Wireless Sensors and Powering (BCWSP)*, 2018, no. May.
- [12] Y. Yunitasari, A. Musdholifah, and A. K. Sari, “Sarcasm Detection For Sentiment Analysis in Indonesian Tweets,” *Indones. J. Comput. Cybern. Syst.*, vol. 13, no. 1, pp. 53–62, 2019.