

OPINION MINING FRAMEWORK BERBASIS LEMKASIKAL

Sherly Christina^{a,1,*}

^a Universitas Palangka Raya, Jl. H. Timang

¹ sherly.christina.upr@gmail.com *

* corresponding author

ARTICLE INFO

Keywords

Opinion Mining
Sentiment Analysis
FrameWork
Lexicon

ABSTRACT

Opinion Mining and Sentiment Analysis can contribute in improving system performance to meet the user's needs. The text that represents the user's opinion may express the emotions or opinions implicitly or explicitly. The words that compose the sentence of opinion can have many different meanings, related to the subject being discussed or the words that compose the sentence. This study, conducted a literature study on several researches that related to opinion mining, and designed an opinion mining framework based lexicon.

1. Pendahuluan

Salah satu informasi yang dibutuhkan untuk membuat strategi dan mengambil keputusan adalah pendapat orang lain. Apa yang dipikirkan oleh orang lain terhadap suatu subjek dapat memberikan kontribusi terhadap optimalnya strategi, pengambilan keputusan atau tindakan terkait subyek tersebut. Media *online* sepertimedia sosial, kolom opini, *customer review* atau buku tamu pada suatu website menyediakan alternatif untuk mengumpulkan opini publik terhadap suatu subjek[9]. Gambar 1 adalah contoh opini publik dalam bentuk *customer review* terhadap layanan hotel yang diambil dari website www.tripadvisor.com. Pendapat, sentimen, evaluasi, penilaian, sikap dan emosi yang disajikan dalam bentuk teks dapat dinyatakan secara implisit maupun eksplisit, sehingga dibutuhkan pengetahuan untuk mengidentifikasi tujuan dan kecenderungan atau polaritas dari opini tersebut. Polaritas suatu opini terhadap suatu topik atau entitas dapat dikategorikan menjadi *Good*, *Neutral* atau *Bad* terkait dengan subyek yang sedang dibahas.

Setiap opini yang diberikan oleh publik terhadap suatu topik kemudian dikumpulkan dan disimpan sehingga menjadi koleksi data teks yang besar dan kompleks. Untuk mengolah dan menganalisis koleksi data tersebut dibutuhkan usaha dan sumber daya yang tidak kecil. Selain masalah jumlah koleksi data teks, dibutuhkan pengetahuan manusia untuk memahami pendapat, sentimen, evaluasi, penilaian, sikap dan emosi seseorang yang direpresentasikan dalam bentuk teks oleh publik.

Sentiment Analysis atau *Opinion Mining* adalah disiplin ilmu *Natural Language Processing* yang mempelajari komputasi terhadap pendapat publik, sikap dan emosi terhadap suatu entitas. Entitas dapat berupa individu, peristiwa atau topik tertentu [1][12]. Walaupun begitu ada penelitian yang menyatakan bahwa *Sentiment Analysis* dan *Opinion Mining* berbeda. *Opinion Mining* adalah salah satu disiplin ilmu dalam *Natural Language Processing* yang bertujuan untuk mengekstrak dan menganalisis pendapat, sikap, atau emosi seseorang terhadap suatu subjek. Subjek tersebut bisa berupa orang, tempat, kejadian atau suatu topik. Sedangkan *Sentiment Analysis* bertujuan untuk mengidentifikasi sentimen, emosi atau sikap seseorang terhadap suatu objek yang disajikan dalam bentuk teks.

Pada penelitian ini dilakukan studi literatur terhadap beberapa penelitian terhadap *opinion mining* dan *sentiment analysis* dan merancang *opinion mining framework* untuk mengidentifikasi

polaritas dari dari opini yang direpresentasikan oleh publik dalam bentuk teks di media online. Kerangka kerja *opinion mining* yang dirancang pada penelitian ini menggunakan pendekatan berbasis lexical.



Gambar 1. *Customer Reviewer* terhadap kamar hotel pada website www.tripadvisor.com.

1.1 Tinjauan Pustaka

Beberapa penelitian telah mengembangkan berbagai metode *Opinion Mining* dan *Sentiment Analysis* yang diimplementasikan dalam berbagai bidang layanan terhadap masyarakat seperti pemerintahan, pertanian, industri, kesehatan, *e-commerce*, atau pariwisata. *Opinion Mining* dan *Sentiment Analysis* dapat berkontribusi terhadap sistem e-governance untuk menindaklanjuti opini yang dikumpulkan dari masyarakat, sehingga meningkatkan kinerja pemerintah dalam pembangunan dan pelayanan terhadap kebutuhan masyarakat [7][8][9].

Pada penelitian yang dilakukan oleh Ravi Anurachalam dkk [8] dinyatakan dalam judul penelitiannya bahwa disiplin ilmu *Sentiment Analysis* adalah *New eye of Government*. Jika instansi pemerintah dapat mengamati dan mengikuti detak kehidupan warganya secara konstan, maka dapat terbuka jalan untuk terbentuknya sistem pemerintahan yang lebih baik, karena publik tentu saja mengharapkan pelayanan yang cepat, tindakan yang tepat dan tanggapan yang segera terhadap setiap keluhan dan kebutuhan yang disampaikan lewat media online. *Opinion Mining* dan *Sentiment Analysis* menyediakan kesempatan bagi pemangku kepentingan untuk menemukan sumber permasalahan lebih cepat sebelum masalah itu terjadi.

Haji Binalli dkk [3] menyatakan bahwa emosi sangat mempengaruhi pengambilan keputusan. Menurut Wordnet Search [13] *emotion* memiliki makna “*any strong feelings*”. Penelitian oleh [3] melakukan deteksi emosi yang tertuang dalam teks dengan menggunakan arsitektur hybrid dan algoritma *Support Vector Machine* untuk menguji arsitektur tersebut.

Vinodhini Gopalakrishnan[5] menggunakan neural network untuk melakukan *opinion mining* pada website kesehatan untuk memprediksi kepuasan pasien terhadap obat-obatan yang telah digunakan. Hasil penelitian [5] menunjukkan nilai *precision*, *recall* dan *f-score* dari *neural network* lebih baik dari pada *support vector machine*.

Cristian Bucur [2] mengekstrak dan mengikhtisarkan opini yang disampaikan oleh *user* mengenai kamar hotel. Sistem yang dikembangkan oleh [2] terdiri atas 2 modul, yaitu modul untuk akuisisi konten berisi *review* dari *user* dan modul untuk menganalisis konten, yang berupa tahapan untuk prapemrosesan data yang telah diekstrak dan berikutnya melakukan *opinion mining* yang dibantu oleh SentiWordNet untuk mengidentifikasi polaritas data. Metode yang digunakan oleh [2] memberikan hasil yang cukup akurat dan bersifat domain *independent* sehingga tidak membutuhkan sumber daya yang mahal untuk dioperasikan.

Stavros Valsamidis[10] merancang *framework* untuk melakukan *opinion mining* dari blog pertanian berbahasa Yunani. *Opinion Mining* pada kerangka kerja [10] melakukan *intent mining* dan *sentiment analysis* terhadap teks yang telah diekstrak dari blog untuk menyelidiki opini dari para pembaca blog terhadap suatu topik yang sedang dibahas dalam bidang pertanian. *Framework* berisi

tahapan prapemrosesan teks, dan kemudian pada tahapan *opinion mining* didukung oleh SentiWordNet untuk mengidentifikasi polaritas dari data teks. Untuk meningkatkan performa *framework*, Stavros Valsamidis[10] menggunakan Naive Bayes untuk mengidentifikasi dan mengkategorikan data berdasarkan polaritasnya. *Framework* yang dirancang[10] diharapkan dapat mengelola masalah dan pendapat petani mengenai suatu topik pertanian, serta mengevaluasi sikap petani terhadap suatu topik pertanian menggunakan text mining. Keakuratan dari *framework* yang dirancang oleh [10] membutuhkan metode *Word Sense Disambiguation* sehingga makna dari setiap kata bisa diidentifikasi dan skor synset bisa diperoleh.

Jan Prichystal[4] merancang dan membangun aplikasi untuk perangkat mobile yang berfungsi mengidentifikasi perasaan seseorang yang direpresentasikan dalam pesan teks dari *customer review*. *Customer* biasanya mencari dan membaca *review* dari customer-customer lain yang telah membeli atau menggunakan produk yang dimaksud, sebelum memutuskan untuk membeli produk tersebut. Penelitian [4] berfokus pada *review* negatif dan positif terhadap suatu produk, sehingga membantu user mengidentifikasi bahwa produk tersebut menarik untuk dibeli. Opini yang diperoleh dari *review* customer terhadap produk diolah menjadi *Word Clouds*. *Word Clouds* adalah interpretasi grafis dari pengkategorian penting atau tidaknya kata. *Word Clouds* berisi himpunan kata dengan berbagai ukuran dan bentuk. Kata yang paling besar frekuensinya adalah *word cloud* dengan ukuran dan bentuk yang lebih besar. Sedangkan kata-kata yang dengan frekuensi yang lebih kecil adalah *word cloud* dengan bentuk dan ukuran yang lebih kecil. *Word Clouds* menunjukkan topik apa yang paling sering didiskusikan mengenai suatu produk. Jika *word cloud* yang berukuran besar adalah positif, maka *user* akan menemukan opini yang positif, dan jika *cloud word* yang berukuran besar adalah negatif maka *user* akan menemukan opini yang negatif. Pendekatan yang digunakan [4] fokus terhadap perolehan hasil yang cepat, sehingga kurang akurat dalam menganalisis. Pendekatan [4] tidak mampu menentukan polaritas kata yang memiliki banyak makna atau ekspresi.

Beberapa penelitian lain yang telah dibahas sebelumnya memberikan ide bagi penelitian ini untuk membangun *framework Opinion Mining* dengan menggabungkan pendekatan frekuensi kata pada teks dan pendekatan leksikal. Pada penelitian ini dibutuhkan database kamus untuk menyimpan kata-kata untuk menentukan polaritas dari opini.

2. Metode Penelitian

Penelitian ini melakukan studi literatur dari berbagai artikel dan jurnal seperti yang dicantumkan di subbab referensi. Ide yang dituangkan dalam *framework* yang dirancang diperoleh dari hasil pembelajaran terhadap beberapa penelitian yang telah dipublikasikan sebelumnya.

Pada beberapa penelitian sebelumnya yang menggunakan pendekatan dengan metode *machine learning* memberikan hasil *precision* dan *recall* yang cukup tinggi tetapi membutuhkan sumber daya yang besar juga. Penelitian lain dengan pendekatan leksikal juga mampu mengidentifikasi polaritas suatu opini, dan kinerjanya dapat ditingkatkan lagi dengan menerapkan *word sense disambiguation* untuk mengidentifikasi makna dari kata-kata dalam data set opini.

3. Hasil Dan Pembahasan

Opinion Mining framework pada penelitian ini terdiri atas 7 proses, seperti ditunjukkan oleh Gambar 2. Proses pertama adalah melakukan akuisisi data set menggunakan *web crawler* untuk mengambil konten halaman web yang berisi *review* dari pengguna web. Pada proses pertama, opini yang dituliskan oleh pengguna web akan diekstrak dari halaman html kemudian disimpan dalam *database* untuk dilakukan prapemrosesan teks pada beberapa proses berikutnya, yaitu tokenisasi, membuang *stopwords* dan *stemming*.

Proses Tokenisasi merupakan proses pemotongan *string* berdasarkan tiap kata yang menyusunnya. Sebelum melakukan tokenisasi, setiap huruf dalam dokumen harus diubah ke dalam huruf kecil. Hanya huruf “a” sampai dengan “z” yang diterima. Karakter selain huruf alfabet dihilangkan dan dianggap tanda baca.



Proses Ketiga adalah membuang *stopword*. *Stopword* adalah kata-kata tidak deskriptif yang dapat dibuang dalam pendekatan *bag-of-words*. Pendekatan *bag-of-words* merupakan penyederhanaan asumsi yaitu, sebuah teks direpresentasikan sebagai sebuah koleksi kalimat yang tidak berurutan, serta tidak mengindahkan tata bahasa dan urutan kata. *Stopwords* perlu dihilangkan karena dua alasan, yaitu menghemat tempat dan waktu pencarian. Pada proses ketiga kata yang menyebabkan negasi seperti “*not*” tidak akan dibuang. Kata yang menyebabkan negasi dapat menyebabkan perubahan polaritas dari opini, contohnya kata “*not good*” memiliki makna sama dengan “*bad*”.

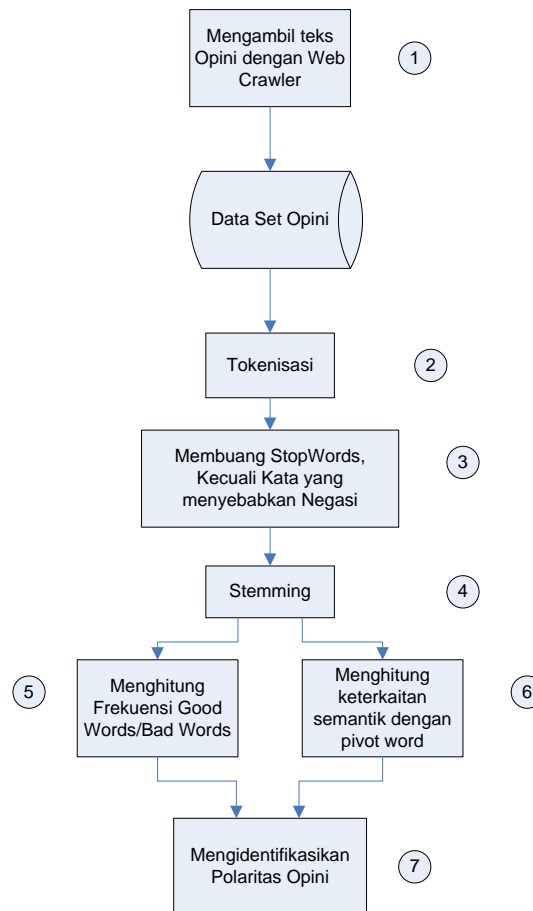
Pada proses keempat dilakukan *stemming*. *Stemming* adalah tahap mencari akar kata dari tiap kata hasil penyaringan. Pada tahap ini dilakukan proses pengembalian berbagai bentukan kata ke dalam suatu representasi yang sama.

Pada proses kelima akan dihitung jumlah kata yang merupakan *good words* dan *bad words* berdasarkan koleksi kata *good words* dan *bad words* yang disimpan dalam *database*. Setelah itu dihitung nilai rata-rata *good words* adalah jumlah *good words* dibagi dengan total kata di dalam teks dan nilai rata-rata *bad words* adalah jumlah *bad words* dibagi total kata di dalam teks.

Pada proses kelima akan dihitung nilai keterkaitan semantik antara kata-kata di dalam teks dengan pivot words, pivot words yang akan digunakan pada kerangka kerja ini adalah “*good*” dan “*bad*”. Proses kelima akan dibantu dengan database WordNet sebagai database leksikal untuk melakukan proses *word sense disambiguation*. Pada proses kelima akan diperoleh dua nilai keterkaitan semantik, yaitu nilai keterkaitan teks dengan semantik “*good*” dan nilai keterkaitan teks dengan semantik “*bad*”.

Proses terakhir adalah mengidentifikasi polaritas opini, pada proses ini nilai yang diperoleh dari proses keempat dikalkulasikan dengan nilai dari proses kelima. Untuk menghitung nilai polaritas “*good opinion*” dilakukan dengan menjumlahkan nilai rata-rata *good words* dengan nilai keterkaitan semantik *good* kemudian dibagi dua. Kemudian untuk menghitung nilai polaritas “*bad opinion*” dilakukan dengan menjumlahkan nilai rata-rata *bad words* dengan nilai keterkaitan semantik *bad* kemudian dibagi dua. Polaritas dari opini akan diperoleh dengan membandingkan nilai polaritas *good opinion* dan *bad opinion*. Bila nilai polaritas *good opinion* lebih besar dari pada *bad opinion* maka diidentifikasi bahwa opini yang direpresentasikan dalam teks adalah baik, atau memiliki kecenderungan emosi yang positif terkait topik yang disampaikan oleh teks tersebut. Sedangkan bila nilai polaritas *bad opinion* lebih besar dari *good opinion* maka diidentifikasi bahwa opini yang direpresentasikan dalam teks adalah buruk, atau memiliki kecenderungan emosi yang negatif terkait topik yang disampaikan oleh teks tersebut.

Untuk mengidentifikasi opini yang dengan polaritas netral, diperlukan nilai tengah dari nilai polaritas *good opinion* dan *bad opinion*. Oleh karena itu akan dilakukan pengujian dari sejumlah kasus pengidentifikasi polaritas opini.



Gambar 2. *Opinion mining framework.*

4. Kesimpulan

Opinion mining atau *sentiment analysis* adalah disiplin ilmu *natural language processing* yang dapat berkontribusi untuk mengoptimalkan kinerja sistem untuk menindaklanjuti kebutuhan pengguna dengan efisien. Kerangka kerja yang dirancang pada penelitian ini berbasiskan leksikal diharapkan mampu memberi hasil identifikasi opini yang akurat, mengingat pentingnya makna kata dalam teks, yang dapat membuat konotasi dari suatu opini menjadi kompleks.

Penelitian selanjutnya adalah mengembangkan aplikasi dari kerangka kerja ini dan melakukan pengujian untuk mengevaluasi kinerja dari framework.

Daftar Pustaka

- [1] Chetashri Bhadanea, Hardi Dalalb, Heenal Doshic , “*Sentiment analysis: Measuring opinions*”, International Conference on Advanced Computing Technologies and Applications (ICACTA-2015)
- [2] Cristian Bucur, “*Using Opinion Mining Techniques in Tourism*”, 2nd GLOBAL CONFERENCE on BUSINESS, ECONOMICS, MANAGEMENT and TOURISM, 30-31 October 2014, Prague, Czech Republic
- [3] Haji Binali, Chen Wu, Vidyasagar Potdar, *Computational Approach for Emotion Detection in Text*, 2010. 4th IEEE International Conference on Digital Ecosystems and Technologies (IEEE DEST 2010)
- [4] Jan Prichsytal, *Mobile Application for Customers' Reviews Opinion Mining*, 19th International Conference Enterprise and Competitive Environment 2016, ECE 2016, 10-11 March, Brno, Czech Republic

- [5] Vinodhini Gopalakrishnan, Chandrasekaran Ramaswamy, *Patient opinion mining to analyze drugs satisfaction using supervised learning*, Journal of Applied Research and Technology 15(2017)311-319
- [6] Vinodhini Gopalakrishnan, R.M. Chandrasekaran, *Sentiment Analysis and Opinion Mining: A Survey*, 2012. International Journal of Advanced Research in Computer Science and Software Engineering Vol.2
- [7] Pankaj Kumar, Alisha Zaidi, Ankita Srivastava, Geet Jain, *Evolving Sentiment towards E-Governance using Opinion Mining*, 2015. International Journal of Computer Applications(0975-8887), International Conference on Distributed Computing and Internet Technology
- [8] Ravi Arunachalam, Sandipan Sarkar, 2013. *The New Eye of Government: Citizen Sentiment Analysis in Social Media*, 2013. Workshop on Natural Language Processing for Social Media (Social NLP) pages 23-28, Nagoya, Japan
- [9] Sherly Christina, Deddy Ronaldo, *Peran Opinion Mining dan Sentiment Analysis untuk Mengidentifikasi Sentimen Publik dalam Sistem E-Governance*, Jurnal Teknologi Informasi Vol. 10. No. 2 Agustus 2016 ISSN 1907-896X, Palangka Raya, Kalimantan Tengah
- [10] Stavros Valsamidis, Theodosios Theodosioua, Ioannis Kazanidisa, Michael Nikolaidis, *A Framework for Opinion Mining in Blogs for Agriculture*, 6th International Conference on Information and Communication Technologies in Agriculture, Food and Environment (HAICTA 2013)
- [11] Testsuya Nasukawa, Jeonghee Yi, *Sentiment Analysis: Capturing Favorability Using Natural Language Processing*, 2003. Proceedings of The 2nd International Conference on Knowledge Capture Pages 70-77, ACM ©2003 ISBN:1-58113-583-1
- [12] Walaa Medhat, Ahmed Hasan, Hoda Korashy, *Sentiment Analysis Algorithms and Applications: A Survey*, Ain Sham Engineering Journal (2014) 5, 1093-1113
- [13] Wordnet 2.1 Browser