

# Perbandingan Metode Lexicon-based dan SVM untuk Analisis Sentimen Berbasis Ontologi pada Kampanye Pilpres Indonesia Tahun 2019 di Twitter

Ahmad Choirun Najib <sup>1)</sup>\*, Akhmad Irsyad <sup>2)</sup>, Ghiffari Assamar Qandi <sup>3)</sup>, Nur Aini Rakhmawati <sup>4)</sup>

Prodi Sistem Informasi, Institut Teknologi Sepuluh Nopember Surabaya <sup>1),2),3),4)</sup>  
[ahmadchoirunnajib@gmail.com](mailto:ahmadchoirunnajib@gmail.com) <sup>1)</sup>\*, [sentinel\\_irsyad@hotmail.com](mailto:sentinel_irsyad@hotmail.com) <sup>2)</sup>, [ghiffari1996@gmail.com](mailto:ghiffari1996@gmail.com) <sup>3)</sup>  
[nur.aini@is.its.ac.id](mailto:nur.aini@is.its.ac.id) <sup>4)</sup>

## Abstrak

Penggunaan media sosial semakin hari semakin meningkat. Salah satu media sosial yang populer saat ini adalah Twitter. Menjelang pemilihan Presiden Republik Indonesia semakin banyak tweet yang membahas tentang kegiatan tersebut. Hal ini menyebabkan topik kampanye pemilu memiliki peluang yang baik untuk dilakukan proses analisis sentimen. Saat ini, mayoritas analisis sentimen di Indonesia dilakukan hanya menilai sentimen dari kalimat tanpa mengetahui apa entitas yang ada dalam kalimat. Tujuan penelitian ini yaitu melakukan analisis sentimen dengan pendekatan berbasis ontologi. Ontologi digunakan dalam menyaring data yang akan digunakan. Ontologi dalam penelitian ini adalah ekonomi dengan atribut finansial, lapangan kerja, dan kesejahteraan. Proses analisis sentimen dilakukan dengan metode Lexicon-based dan Support Vector Machine (SVM). Proses akuisisi data diperoleh sejumlah 700.000 tweet. Koleksi tersebut diseleksi berdasarkan ontologi ekonomi menghasilkan 16.998 tweet dan dilakukan pelabelan manual sebanyak 1.600. Kemudian dilakukan pengolahan data hingga diperoleh dataset final sejumlah 1.050 tweet. Berdasarkan hasil penelitian yang dilakukan akurasi yang diperoleh berdasarkan metode Lexicon-based adalah 39% dan metode SVM sebesar 83%. Dari penelitian ini diketahui bahwa SVM mempunyai performa yang lebih baik dibandingkan dengan Lexicon-based. Hasil Lexicon-based menunjukkan bahwa sentimen pada mayoritas atribut berupa netral. Sedangkan hasil SVM menunjukkan bahwa sentimen pada mayoritas atribut (finansial dan kesejahteraan) berupa positif, sisanya (lapangan kerja) berupa netral. Selanjutnya, proses ekstraksi dan pembuatan ontologi Bahasa Indonesia secara semi-otomatis pada dataset perlu untuk dikembangkan pada penelitian berikutnya untuk menyempurnakan ontologi.

**Kata kunci:** Analisis Sentimen, Twitter, Ontology, SVM, Lexicon

## Abstract

**[Comparison of the Lexicon-based and SVM Method for Ontology-Based Analysis of the 2019 Presidential Election Campaign on Twitter]** The use of social media is increasing. One of the most popular social media is Twitter. Towards the election of the President of the Republic of Indonesia, election topic tweets discussed almost every day. Hence, it is suitable for the sentiment analysis process. Nowadays, the sentiment analysis is only evaluating the sentence without knowing what the entity is in the sentence. To overcome this drawback, we propose a sentiment analysis based on ontology. Ontology is used to filter the data to be used. The ontology used in this study is economics with attributes, i.e., financial employment, and welfare. The sentiment analysis process is carried out using the Lexicon and Support Vector Machine (SVM) based methods. The process of acquiring data obtained 700,000 tweets. The collection was selected based on economic ontology to produce 16,998 tweets, and 1,600 manual labels were labelled. Then, the number of the final dataset is 1,050 tweets. The results show that the accuracy of the Lexicon-based method is 39%, and the SVM method is 83%. The SVM has better performance than Lexicon-based. Lexicon-based results show that the sentiment on the majority attributes is neutral. While the SVM results show that the sentiment on the majority attributes (financial and welfare) is positive, the rest (employment) is neutral. A semi-automatic ontology extraction and development for Bahasa Indonesia is necessary for the future works to make a comprehensive ontology and provide better results.

**Keywords:** Sentiment Analysis, Twitter, Ontology, SVM, Lexicon

## 1. PENDAHULUAN

Di era website 2.0 memunculkan sebuah media untuk berbagi dimana saja dan kapan saja melalui internet. Salah satu media sosial yang populer adalah

Twitter. Menurut data statistik Kominfo, jumlah pengguna Twitter Indonesia pada tahun 2013 diperkirakan sebanyak 19,5 juta [1]. *Tweet* adalah konten dari pengguna berbentuk teks dengan panjang

karakter maksimal sebesar 140 karakter. Dengan adanya batasan karakter, setiap *tweet* membuat pengguna cenderung lebih ekspresif dibandingkan dengan media sosial lainnya. Sehingga konten tersebut memiliki nilai emosi atau sentimen dan dikategorikan sebagai sumber yang bagus untuk analisis sentimen. Analisis sentimen merupakan sebuah proses untuk menentukan polaritas atau kecenderungan sebuah kalimat memiliki nilai positif, negatif atau netral.

Jejaring sosial media Twitter memiliki karakteristik berupa data *tweet* yang mutakhir dengan jumlah yang banyak dan topik yang bervariasi. Sehingga, hal tersebut memiliki potensi untuk dilakukan analisis sentimen. Analisis sentimen pada jejaring sosial Twitter yang menggunakan Bahasa Indonesia telah banyak dilakukan pada topik yang bermacam-macam.

Pada penelitian sebelumnya, Sunni dan Widyantoro, 2012 melakukan analisis sentimen dan ekstraksi topik penentu sentimen pada opini terhadap tokoh publik [2]. Penelitian ini mengusulkan metode F3 (F3 is Factor Finder) yang bertujuan untuk menangani permasalahan karakteristik model bahasa pengguna Twitter Indonesia yang berbeda dengan pengguna luar Indonesia. Nurhuda et al., 2013 melakukan analisis sentimen masyarakat terhadap calon Presiden Indonesia 2014 berdasarkan opini dari Twitter menggunakan metode Naive Bayes Classifier [3]. Hidayatullah dan SN, 2014 melakukan analisis sentimen dan klasifikasi kategori terhadap tokoh publik pada Twitter [4]. Penelitian ini menggunakan metode NBC untuk proses klasifikasi. Monarizqa et al., 2014 melakukan analisis sentimen pada kalimat berbahasa Indonesia di Twitter [5]. Penelitian ini menggunakan algoritma SVM untuk proses klasifikasi kelas sentimen pada *tweet* berbahasa Indonesia. Novantirani, et al., 2015 melakukan analisis sentimen pada twitter untuk mengenai penggunaan transportasi umum darat dalam kota dengan metode Support Vector Machine [6]. Kemudian, Buntoro, 2017 melakukan analisis sentimen calon Gubernur DKI Jakarta 2017 di Twitter [7]. Penelitian ini menggunakan metode *Naive Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM) yang bertujuan untuk klasifikasi kelas dari sentimen.

Namun, mayoritas analisis sentimen yang telah dilakukan hanya melihat polaritas dari sebuah kalimat tanpa memperhatikan entitas yang ada pada sebuah *tweet*. Pada penelitian ini dilakukan proses analisis sentimen berbasis ontologi dengan melakukan ekstraksi entitas yang ada pada *tweet* terlebih dahulu dan menghitung nilai sentimennya. Penelitian ini mengambil studi kasus topik kampanye Pemilihan Presiden Republik Indonesia tahun 2019.

Proses ekstraksi entitas dilakukan dengan membuat ontologi dan melakukan seleksi terhadap koleksi *tweet* berdasarkan ontologi yang telah dibuat. Kemudian, penghitungan sentimen dilakukan menggunakan pendekatan *Lexicon-based* dan *machine learning SVM*. Metode *Lexicon-based* memberikan hasil berupa penilaian kuantitatif (digit angka).

Sedangkan SVM memberikan hasil berupa penilaian kualitatif (klasifikasi positif, negatif atau netral).

Sejauh pengetahuan penulis, di Indonesia belum ada yang melakukan penelitian tentang analisis sentimen berbasis ontologi. Sehingga, penelitian analisis sentimen berbasis ontologi layak untuk dilakukan. Tujuan penelitian ini yaitu melakukan analisis sentimen dengan pendekatan berbasis ontologi. Harapan penulis ini menjadi awal bagi penelitian lain untuk menerapkan ontologi kedalam analisis teks.

Telah ada beberapa penelitian sebelumnya yang menggabungkan ontologi dan sentimen analisis. Kontopoulos et al., 2013 melakukan identifikasi ontologi berdasarkan data dari Twitter [8]. Peneliti mencoba membuat ontologi yang berhubungan merk handphone dimana setiap merk handphone akan di teliti sentimennya berdasarkan beberapa atribut komponen seperti prosesor, micro usb, kamera, dan baterai. Proses sentimen dilakukan dengan WordNET dan Lexicon.

Di Indonesia terdapat beberapa penelitian tentang analisis sentimen baik yang menggunakan *Lexicon* ataupun machine learning. Kurniawan et al., 2019 meneliti dengan mengumpulkan data dari Twitter yang mengandung #2019GantiPresiden dan #2019TetapJokowi, dari data yang diperoleh dilakukan labeling dengan metode K-Means. Selanjutnya dilakukan pra proses, lalu pelatihan model dengan metode Naive Bayes. Hasil dari proses pemodelan ini akan diperoleh klasifikasi apakah *tweet* memiliki sentimen negatif atau positif [9].

Lestari et al., 2019 meneliti tentang sentimen analisis menggunakan *Lexicon-based*, data yang digunakan adalah *tweet* yang dilakukan oleh orang-orang pada saat debat pertama pemilihan presiden Indonesia 2019. Dalam menentukan sentimen dilakukan dengan mencari score sentimen dimana jumlah nilai kata positif dikurangi jumlah nilai kata negatif. Apabila score sentimen bernilai positif maka *tweet* tersebut akan dianggap memiliki sentimen positif, atapun sebaliknya [10]. Dalam penelitian ini peneliti akan mencoba membandingkan hasil dari sentimen analisis yang dilakukan dengan *Lexicon-based* dan machine learning SVM berdasarkan data hasil ontologi.

## 2. BAHAN DAN METODE

### 2.1 Bahan

#### 2.1.1 Ontologi

Ontologi didefinisikan sebagai sebuah konseptualisasi yang dapat digunakan secara bersama-sama dengan karakteristik yang jelas dan dapat dibaca oleh mesin [11]. Ontologi digunakan untuk memodelkan istilah atau sesuatu pada domain tertentu beserta relasinya. Ontologi biasanya menyediakan vocabulary untuk mendeskripsikan domain area tertentu beserta spesifikasi dan makna dari terminologi yang digunakan pada vocabulary tersebut [12]. Ontologi dapat dianalogikan seperti halnya model konseptual seperti klasifikasi dan skema basis data.

### 2.1.2 Formal Concept Analysis (FCA)

*Formal Concept Analysis* adalah teori analisis data matematika [13]. Karakteristik utamanya adalah metodologi yang menerapkan langkah-langkah yang berdasarkan pengguna untuk membuat domain model. Keuntungan menggunakan FCA didalam ontology sebagai berikut [8]:

- Ukuran ontology yang sesuai  
Ontologi domain terus berkembang setiap tahap menyesuaikan dengan dataset (*tweet*). Sehingga dapat menghindari redundansi data dari pengguna dan dari sisi ontology tidak kekurangan dari segi konsep.
- Desain ontology yang lebih baik.  
Konsep hirarkinya tersirat, akan tetapi ditandai secara dinamis menggunakan properti. Ini membuat desain ontologinya semakin baik dan membuat ragam konsep klasifikasinya semakin berbeda-beda.
- Ontologi dengan domain yang spesifik.  
Dalam membuat domain ontology memiliki kecenderungan menggunakan ontology yang sudah ada dan kemudian mendeskripsikan secara detail apa yang dibutuhkan. Namun, tujuannya tidak mendeskripsikan secara lengkap, akan tetapi mengembangkan ontology yang menyesuaikan dengan dataset (*tweet*).

### 2.1.3 Lexicon

Lexicon adalah kumpulan kata yang sering digunakan untuk mengekspresikan sentimen positif

atau negatif yang diberikan nilai-nilai pada setiap kata. Barasa adalah sebuah kamus data sentimen Indonesia yang didalamnya berisi kumpulan kata-kata dalam Bahasa Indonesia yang telah diberikan nilai berupa angka yang diasosiasikan dengan sentimen positif atau negatif [14].

### 2.1.4 Support Vector Machine (SVM)

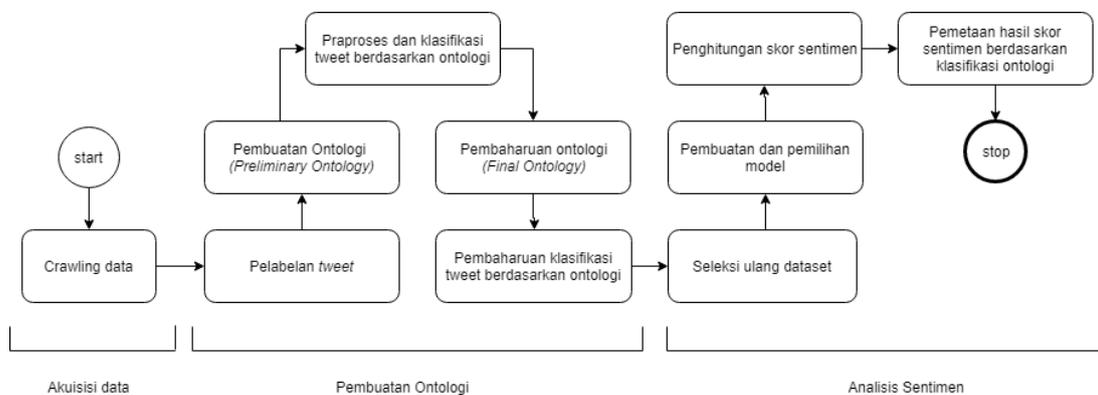
*Support Vector Machine* (SVM) adalah salah satu Teknik dalam machine learning yang berdasarkan teori struktural pembelajaran statistik. SVM mempunyai performa yang lebih baik apabila dibandingkan dengan metode *machine learning* lain [3]. Proses utama dari SVM adalah menemukan batas yang memisahkan tiap kelas (*hyperlane*).

### 2.2 Metode

Pada bagian ini dijelaskan mengenai dasar teori dan metodologi penelitian. Kami membagi proses analisis sentimen menjadi tiga bagian proses utama yaitu, akuisisi data, pembuatan ontology, dan analisa sentimen. Detail metodologi diilustrasikan pada Gambar 1.

#### 2.2.1 Akuisisi Data

Pada tahap ini dilakukan proses *Crawling data* untuk memperoleh data dari Jejaring sosial media Twitter. *Tweet* yang dikumpulkan adalah *tweet* yang dibuat di antara tanggal 16 Maret 2019 sampai 16 April 2019, data yang bertepatan dengan proses kampanye Presiden Indonesia 2019. Berdasarkan kueri pencarian Jokowi dan Prabowo diperoleh sebanyak 700.000 *tweet*, data yang diperoleh selanjutnya dilakukan praproses.



**Gambar 1.** Metodologi Penelitian

### 2.2.2 Pembuatan Ontologi

Pada tahap ini dilakukan pembuatan ontology dengan pendekatan *Formal Concept Analysis* (FCA). Pembuatan ontology dilakukan secara bertahap untuk memperoleh ontology yang utuh sesuai dengan dataset yang diperoleh pada proses akuisisi data.

#### a) Pelabelan *tweet* (Labelling)

*Labelling* melakukan klasifikasi sebuah *tweet*, apakah *tweet* itu mempunyai sentimen yang bernilai positif, netral, dan negatif. Proses ini dilakukan secara manual, proses ini dilakukan oleh tiga orang. *Tweet* yang dilabeli ini adalah *tweet* asli yang belum diolah.

Pelabelan dilakukan hingga mencapai 600 *tweet* sentimen positif, 600 *tweet* sentimen negatif dan 600 *tweet* sentimen netral dengan total 1.800.

#### b) Pembuatan ontology awal (*Preliminary Ontology*)

Menentukan atribut yang akan disematkan kepada masing-masing *tweet* dengan dasar ontologinya adalah ekonomi. Atribut awal yang kami tentukan adalah ekonomi, harga, pertumbuhan, lapangan kerja, pekerjaan, penghasilan, gaji, pajak, pengangguran, nganggur, kemiskinan, kesejahteraan.

c) Praproses dan klasifikasi *tweet* berdasarkan ontologi

Praproses diimplementasikan untuk menghindari data yang tidak lengkap, gangguan data, dan data yang tidak konsisten [14]. Tahapan praproses dalam penelitian ini meliputi:

- Menghapus URL (<http://www.situs.com>), email ([nama@situs.com](mailto:nama@situs.com)) dan alamat gambar unggahan di Twitter ([pic.twitter.com/slug](http://pic.twitter.com/slug)).
- Menghapus karakter Twitter khusus, Proses ini dilakukan dengan menghapus karakter Twitter khusus seperti *hashtag*, nama pengguna (@namapengguna), dan karakter khusus (misalnya RT, yang menunjukkan bahwa pengguna me-retweet sesuatu).
- Menghapus Simbol. Langkah ini dilakukan untuk menghapus simbol dan tanda baca di *tweet*.
- Menghapus Stopwords. Stopwords adalah kata-kata yang tidak mempengaruhi proses klasifikasi.
- Tokenisasi, proses tokenisasi adalah proses pemotongan string masukan berdasarkan tiap kata yang menyusunnya.

Setelah proses praproses, didapatkan hasil akhir berupa dataset yang lengkap dengan data yang konsisten dan *tweet* sudah bersih dari kata-kata yang tidak mempengaruhi proses klasifikasi. *Tweet* tersebut kemudian dikategorikan berdasarkan atribut awal yang telah ditentukan melalui fungsi *contains*.

d) Pembaharuan ontologi

Untuk melakukan validasi ontologi yang sudah dibuat di tahapan awal. Diterapkan pembuatan *Wordcloud* untuk melihat fitur ontologi yang dipilih sudah sesuai. *Wordcloud* adalah bentuk penyajian kata-kata dalam bentuk gambar yang sering digunakan

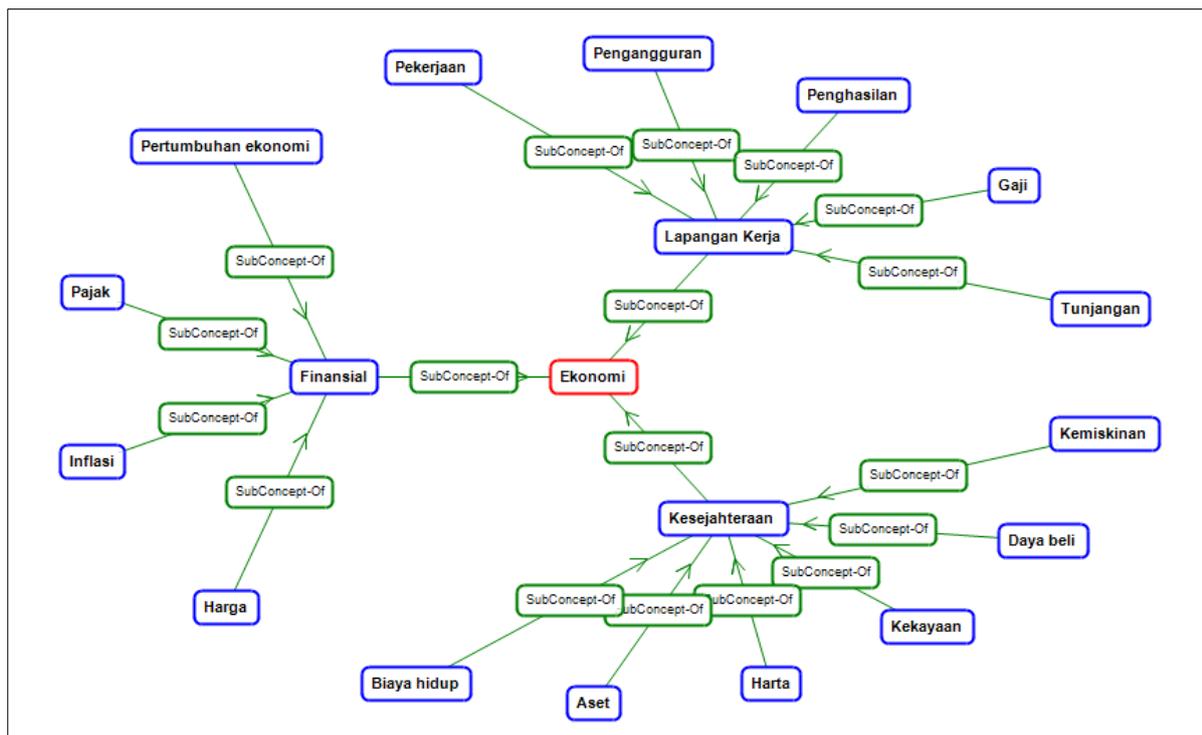
untuk melakukan analisis [15]. Hal ini membuat peluang adanya pembaharuan didalam ontologi yang sudah ditentukan sebelumnya. Kata yang muncul pada *Wordcloud* dan memiliki korelasi dengan topik ekonomi akan pertimbangan untuk memperbaharui ontologi hingga tercapai atribut final. Koleksi kata-kata yang sejenis pada *Wordcloud* akan dilakukan proses penggabungan pada atribut yang lebih luas untuk menyederhanakan atribut. Kata-kata yang sejenis dengan atribut final dihubungkan dengan properti *SubConcept-Of*. Atribut final diilustrasikan pada Gambar 2. Ketiga atribut final akan dijadikan sebagai acuan dalam melakukan pembaharuan koleksi *tweet*.

Atribut final dari ontologi ekonomi adalah sebagai berikut:

- Kesejahteraan: mencakup kata miskin, kemiskinan, kaya, kekayaan, harta, aset, biaya hidup, daya beli.
- Finansial: mencakup kata harga, terjangkau, inflasi, pajak, pertumbuhan ekonomi.
- Lapangan kerja: mencakup kata lapangan kerja, pekerjaan, pengangguran, gaji, penghasilan, tunjangan.

e) Pembaharuan klasifikasi *tweet*

Dengan adanya perubahan pada ontologi, dilakukan kembali penyesuaian dataset terhadap ontologi baru. Masing-masing *tweet* kembali diklasifikasikan berdasarkan atribut yang baru dengan melakukan penghitungan banyaknya jumlah kata dari *tweet* bersih yang mengandung atribut pada ontologi baru. Hal ini dilakukan untuk menangani permasalahan kemungkinan sebuah *tweet* terklasifikasi lebih dari dua kelas yang berbeda.



Gambar 2. Visualisasi ontologi final menggunakan Ontogen

### 2.2.3 Analisis Sentimen

Pada tahap ini, *tweet* yang telah diklasifikasikan berdasarkan ontologi dihitung skor sentimennya dengan menggunakan pendekatan Lexicon dan SVM.

#### a) Seleksi ulang dataset

Setelah diklasifikasikan ulang, untuk mencapai keselarasan jumlah sentimen positif, negatif dan netral dilakukan kembali seleksi ulang terhadap dataset. Jumlah keseluruhan *tweet* yang akan digunakan disesuaikan dengan jumlah *tweet* dengan atribut klasifikasi paling sedikit yaitu pada atribut kesejahteraan yang berjumlah 350. Dataset akhir berjumlah 1.050 dari yang semula berjumlah 1.800.

#### b) Pembuatan dan pemilihan model

Pembuatan dan pemilihan model dilakukan menggunakan pendekatan SVM untuk menemukan model terbaik. Model terbaik ditemukan dengan cara membandingkan performa klasifikasi SVM dengan klasifikasi pelabelan manual. Konfigurasi model yang terbaik akan dipilih sebagai dasar penghitungan skor sentimen menggunakan SVM.

#### c) Penghitungan skor sentimen

Pada tahap ini dua pendekatan digunakan untuk menghitung nilai sentimen dari dataset yang digunakan. Kedua pendekatan tersebut adalah pendekatan berbasis Lexicon dan SVM. Setelah melakukan proses penghitungan nilai sentimen, masing-masing pendekatan akan diuji nilai akurasi, presisi, nilai F1 dan *recall*.

#### d) Pemetaan hasil skor sentimen berdasarkan klasifikasi ontologi

Setelah didapatkan nilai sentimen dari masing-masing pendekatan. Hasil tersebut dikelompokkan sesuai dengan atribut dari klasifikasi ontologi akhir.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Penghitungan Lexicon

Dalam penghitungan Lexicon yang mempengaruhi adalah jumlah nilai kata baik yang bernilai positif ataupun negatif. Dalam melakukan proses klasifikasi nilai skor Lexicon apakah negatif, netral atau positif dilakukan uji coba dengan berbagai batas *threshold* pada batas bawah dan batas atas. Jika nilai skor Lexicon dibawah batas bawah, maka *tweet* tersebut akan memiliki sentimen negatif. Jika nilai skor Lexicon diatas batas atas, maka *tweet* tersebut akan

memiliki sentimen positif. Jika nilai skor Lexicon diantara batas bawah dan batas atas, *tweet* tersebut memiliki sentimen netral. Hasil uji coba menunjukkan bawa hasil performa klasifikasi terbaik terdapat pada konfigurasi batas bawah berupa kuartil bawah dengan nilai (-0,004680), dan batas atas berupa kuartil atas dengan nilai (0,019266). Sebagai contoh pada Tabel 1 *tweet* mengandung kata-kata negatif seperti goblok dan kampret. Sehingga akan menghasilkan nilai Lexicon yang bernilai minus (-0,0125910890830138). Selanjutnya *tweet* diklasifikasikan memiliki sentimen negatif.

### 3.2 Penghitungan SVM

Pada pendekatan SVM model yang sudah buat lalu dilakukan test menggunakan *tweet* yang telah disiapkan. Kemudian data tersebut dihitung probabilitas sentimennya. Dalam uji coba pembuatan model, ditemukan bahwa performa terbaik terdapat pada konfigurasi sebagai berikut:

- Transformasi dataset kedalam vector SVM menggunakan TFIDFVector
- Perbandingan data training dengan data testing yaitu 60%:40%
- Penggunaan random state yaitu 0
- Kernel RBF dengan nilai koefisien gamma 0,01 dan C 100

Skor terbesar pada sebuah kelas akan menentukan sentimen pada *tweet* tersebut. Contoh pada Tabel 2 menampilkan nilai skor SVM pada kelas negatif merupakan skor terbesar (0,55948534), oleh karena itu sentimen pada sampel *tweet* yaitu negatif.

### 3.3 Pengukuran Performa metode Lexicon dan SVM

Skor pengukuran performa dihitung berdasarkan nilai presisi, recall dan f1-score dari masing-masing kelas (negative, positif, netral) serta nilai akurasi secara keseluruhan. Masing-masing skor diilustrasikan pada Tabel 3 dan Tabel 4.

### 3.4 Pemetaan Hasil Skor Sentimen Berdasarkan Klasifikasi Ontologi

Berdasarkan skor sentimen yang dihasilkan oleh metode Lexicon dan model SVM, selanjutnya sentimen dipetakan berdasarkan atribut ontologi final. Hasil pemetaan diilustrasikan pada Tabel 5.

**Tabel 1.** Hasil prediksi Lexicon

Sampel <i>tweet</i>	Skor Lexicon	Sentimen
pantes kampret orang goblok coba belajar pakar pakar ekonomi kasih pemaparan biar tau gimana berbisnis	-0,0125910890830138	negatif

**Tabel 2.** Hasil prediksi SVM

<i>Tweet</i>	SVM	Sentimen
pantes kampret orang goblok coba belajar pakar pakar ekonomi kasih pemaparan biar tau gimana berbisnis	0,55948534 (negatif) 0,22888234 (positif) 0,21163232 (netral)	negatif

**Tabel 3.** Presisi, Recall, F1-skor dan Akurasi metode Lexicon

Kelas	Presisi	Recall	F1-skor	Akurasi
negatif	0,426	0,329	0,338	
netral	0,349	0,534	0,422	0,399
positif	0,471	0,338	0,394	

**Tabel 4.** Presisi, Recall, F1-skor dan Akurasi metode SVM

Kelas	Presisi	Recall	F1-skor	Akurasi
negatif	0,840	0,865	0,852	
netral	0,826	0,816	0,821	0,839
positif	0,850	0,837	0,843	

**Tabel 5.** Hasil Pemetaan Sentimen Berdasarkan Atribut Ontologi

Metode	Atribut	Sentimen	Jumlah
Lexicon	Finansial	negatif	79
		netral	179
		positif	92
	Lapangan Kerja	negatif	91
		netral	166
		positif	93
	Kesejahteraan	negatif	93
		netral	179
		positif	78
SVM	Finansial	negatif	111
		netral	108
		positif	131
	Lapangan Kerja	negatif	118
		netral	143
		positif	89
	Kesejahteraan	negatif	121
		netral	88
		positif	141

### 3.5 Analisa

Terdapat beberapa faktor yang mempengaruhi kemampuan klasifikasi diantaranya adalah penggunaan bahasa yang tidak baku, menyingkat suatu kata seperti “yg”, “bln”, dan “sgt” yang seharusnya ditulis “yang”, “bulan”, “sangat” dan terdapat kata berbahasa inggris sehingga kata-kata tersebut tidak ada dalam kamus data sentimen Bahasa Indonesia. Meskipun penghilangan stopwords telah dilakukan, masih terdapat beberapa kata yang tidak terhilangkan.

Selain banyaknya penulisan yang tidak sesuai dengan kamus, penyebab kurangnya performa dari metode *Lexicon* adalah kemunculan kata-kata yang bernilai positif lebih banyak dibandingkan dengan kata bernilai negative dalam *tweet* yang seharusnya bernilai negative ataupun sebaliknya. Pada contoh di Tabel 6 terdapat kata positif “tulus”, “menang” dan terdapat kata negatif “bodoh”, apabila dihitung berdasarkan metode *Lexicon* maka akan di peroleh sentimen bernilai positif, padahal apabila dilihat secara *manual* sentimen seharusnya bernilai negatif. Hal ini menunjukkan bahwa pemilihan dataset mempengaruhi performa analisis sentimen [8].

Tabel 5 menunjukkan pemetaan hasil akhir analisis sentimen kedua pendekatan dengan dataset,

dapat dilihat analisis menggunakan *Lexicon* memiliki kecenderungan memberikan nilai netral Hal ini disebabkan oleh hasil uji menggunakan pendekatan *Lexicon* terhadap dataset yang dilakukan tidak ditemukan nilai sentimen yang tepat pada angka 0. Seperti contoh pada Tabel 7 yang menunjukkan perbandingan klasifikasi dari pelabelan manual, *Lexicon* dan SVM. Nilai *Lexicon* yang dihasilkan adalah netral dengan skor sebesar 0,013132005571713328 angka tersebut merupakan nilai positif, akan tetapi pada metode *Lexicon* tidak ada aturan pasti mengenai pemberian nilai netral. Hal ini menyebabkan kesenjangan di nilai netral, karena pada pelabelan *manual* klasifikasi dibagi 3 kelas yaitu positif, negatif dan netral, begitu juga dengan SVM yang memberikan hasil dalam bentuk pembagian 3 kelas. Oleh karena itu peneliti menggunakan nilai kuartil bawah dan kuartil atas untuk menentukan batas bawah dan batas akhir nilai untuk menklasifikasikan nilai sentimen *Lexicon*. Sehingga pada akhir dari uji analisis ditemukan bahwa nilai akurasi untuk analisis sentimen menggunakan *Lexicon* sebesar 39,9% dan SVM memiliki akurasi sebesar 83,9%.

### 3.6 Diskusi

Apabila dibandingkan dengan hasil dari penelitian sebelumnya. Penelitian yang dilakukan memiliki beberapa kelebihan. G. A. Buntoro, 2017 yang menjadikan nilai lexicon sebagai input proses *machine learning*, sehingga apabila terdapat kata yang tidak terdapat pada kamus maka akan mempengaruhi hasil dari klasifikasi [7]. Selain itu, jumlah *dataset* pada penelitian tersebut masih sedikit, sehingga menyebabkan proses *learning* menjadi kurang maksimal. Sedangkan pada penelitian ini, kami menggunakan feature TF-IDF yang dimana setiap kata

memiliki bobot sesuai dengan kemunculan kata pada *term* atau token terkait.

Penelitian sebelumnya yang menggunakan metode yang sama yaitu menggunakan SVM [2]–[6] dan beberapa diantaranya menggunakan pembobotan menggunakan TF-IDF. Yang membedakan penelitian ini dengan penelitian sebelumnya adalah penggunaan metode ontologi. Dengan ontologi, maka akan memungkinkan untuk mendapatkan entitas yang termuat dalam sebuah *tweet*. Sehingga, masing-masing entitas yang terdapat pada *tweet* akan diketahui sentimennya.

**Tabel 6.** Contoh error sentimen *Lexicon*

Tweet Asli	Praproses	Sentimen asli	Sentimen <i>Lexicon</i>
Perbodohan macam apa lagi ini. . . kebelet bat pen jadi presiden ge, santay ae kali pa Janji Tak Ambil Gaji Jika Menang Pilpres, Prabowo: Bukti Kita Tulus <a href="http://detik.id/6ceGXIA,Å">http://detik.id/6ceGXIA,Å</a>	bodoh kebelet bat pen presiden ge santay kali janji ambil gaji menang pilpres prabowo bukti tulus	negatif	positif

**Tabel 7.** Perbandingan hasil klasifikasi

Tweet Asli	Praproses	Sentimen asli	Sentimen <i>Lexicon</i>	Sentimen SVM
#UASdiftinahKejiBalasDiTPS kerjasama ekonomi bilateral, multilateral, multilateral kita doakan di TPS sambil dengarkan suara hati... #UASdiftinahKejiBalasDiTPS	kerjasama ekonomi bilateral multilateral doakan tps dengarkan suara hati	Positif	Netral	Positif

## 4. KESIMPULAN

Kesimpulan yang diperoleh dari penelitian ini adalah metode *Lexicon* dan SVM dapat digunakan dalam melakukan analisis sentimen. Penggunaan metode SVM memiliki performa yang lebih baik dibandingkan metode *Lexicon*. Selain itu ditemukan juga terdapat beberapa factor yang mempengaruhi performa dari sentimen analisis dengan pendekatan *Lexicon*-based diantaranya penggunaan bahasa yang tidak baku, kurangnya kata-kata yang tidak ada dalam kamus, kemunculan kata-kata yang bernilai positif lebih banyak dibandingkan dengan kata bernilai negative dalam *tweet* yang seharusnya bernilai negative ataupun sebaliknya. Untuk SVM faktor yang mempengaruhi akurasi dari analisis sentimen adalah pemilihan dataset. Hasil kedua penghitungan sentimen diklasifikasikan berdasarkan ontologi ekonomi dengan atribut finansial, lapangan kerja, dan kesejahteraan. Hasil *Lexicon* menunjukkan bahwa sentimen pada mayoritas atribut berupa netral. Sedangkan hasil SVM menunjukkan bahwa sentimen pada mayoritas atribut (finansial dan kesejahteraan) berupa positif, sisanya (lapangan kerja) berupa netral. Pada penelitian berikutnya, diperlukan proses ekstraksi dan pembuatan ontologi secara semi-otomatis dari dataset untuk penyempurnaan ontologi yang telah dibuat. Hal ini akan memperbanyak jumlah atribut yang mungkin belum ditemukan pada pembuatan ontologi secara

manual. Sehingga, ontologi yang komprehensif akan mempengaruhi hasil analisis sentimen.

## 5. DAFTAR PUSTAKA

- [1] “Kementerian Komunikasi dan Informatika.” [Online]. Available: [https://kominfo.go.id/content/detail/2366/indonesia-peringkat-lima-pengguna-twitter/0/sorotan\\_media](https://kominfo.go.id/content/detail/2366/indonesia-peringkat-lima-pengguna-twitter/0/sorotan_media). [Accessed: 21-Oct-2019].
- [2] I. Sunni and D. H. Widyantoro, “Analisis sentimen dan ekstraksi topik penentu sentimen pada opini terhadap tokoh publik,” *J. Sarj. ITB Bid. Tek. Elektro dan Inform.*, vol. 1, no. 2, 2012.
- [3] F. Nurhuda, S. W. Sihwi, and A. Doewes, “Analisis sentimen masyarakat terhadap calon Presiden Indonesia 2014 berdasarkan opini dari Twitter menggunakan metode Naive Bayes Classifier,” *ITSMART J. Teknol. dan Inf.*, vol. 2, no. 2, pp. 35–42, 2014.
- [4] A. F. Hidayatullah and A. S. N. Azhari, “Analisis sentimen dan klasifikasi kategori terhadap tokoh publik pada twitter,” in *Seminar Nasional Informatika (SEMNASIF)*, 2015, vol. 1, no. 1.
- [5] N. Monarizqa, L. E. Nugroho, and B. S. Hantono, “Penerapan Analisis Sentimen Pada Twitter Berbahasa Indonesia Sebagai Pemberi

- Rating,” *J. Penelit. Tek. Elektro dan Teknol. Inf.*, vol. 1, no. 3, 2014.
- [6] A. Novantirani, M. K. Sabariah, and V. Effendy, “Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine,” *eProceedings Eng.*, vol. 2, no. 1, 2015.
- [7] G. A. Buntoro, “Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter,” *INTEGER J. Inf. Technol.*, vol. 2, no. 1, 2017.
- [8] E. Kontopoulos, C. Berberidis, T. Dergiades, and N. Bassiliades, “Ontology-based sentiment analysis of twitter posts,” *Expert Syst. Appl.*, 2013.
- [9] I. Kurniawan and A. Susanto, “Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019,” *Eksplora Inform.*, 2019.
- [10] A. Lestari and D. Karolita, “ Summarizing Netizens’ Sentiments Towards the 1 st Indonesian Presidential Debate using Lexicon Sentiment Analysis ,” *IOP Conf. Ser. Mater. Sci. Eng.*, 2019.
- [11] R. Studer, V. R. Benjamins, and D. Fensel, “Knowledge Engineering: Principles and methods,” *Data Knowl. Eng.*, 1998.
- [12] J. Euzenat and P. Shvaiko, *Ontology matching*. 2007.
- [13] Bernhard Ganter and R. Wille, *Formal Concept Analysis: Mathematical Foundations*. 1999.
- [14] T. A. Le, D. Moeljadi, Y. Miura, and T. Ohkuma, “Sentiment Analysis for Low Resource Languages: A Study on Informal Indonesian Tweets,” in *Proceedings of the 12th Workshop on Asian Language Resources (ALR12)*, 2016.
- [15] F. Heimerl, S. Lohmann, S. Lange, and T. Ertl, “Word cloud explorer: Text analytics based on word clouds,” in *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2014.