Identifikasi Provokasi pada Sosial Media Twitter Terhadap Calon Presiden 2019 Menggunakan Metode Ontologi

Hisyam Bagas Prakoso[1], Anisa Herdiani, S.T., M.T.[2], Ibnu Asror, S.T., M.T.[3]
[1,2,3]Fakultas Informatika, Universitas Telkom, Bandung
[1]hisambp@student.telkomuniversity.ac.id,
[2]anisaherdiani@telkomuniversity.ac.id, [3]iasror@telkomuniversity.ac.id

Abstrak

Pemilihan presiden merupakan kegiatan rutin yang dilaksanakan setiap 5 tahun sekali. Banyak kampanye yang disuarakan secara langsung melalui sosial media. Twitter salah satu sosial media yang dimanfaatkan para pasangan calon (paslon) presiden untuk berkampanye. Sayangnya, kampanye disosial media khususnya pada Twitter dirusak oleh sejumlah oknum yang sengaja ingin menjatuhkan paslon presiden lain. Menimbulkan iklim suasana pemilihan presiden 2019 pada sosial media Twitter menjadi kurang baik. Berdasarkan permasalahan di atas, maka diperlukan sebuah sistem yang mampu mendeteksi adanya *tweet* yang mengandung unsur provokasi. Sehingga dapat membedakan mana *tweet* yang mengandung unsur provokasi dengan *tweet* yang tidak mengandung unsur provokasi. Pada penelitian ini dibagi menjadi beberapa tahapan, tahapan pertama adalah pengambilan data twitter menggunakan Twitter API untuk mendapatkan *dataset,* kemudian *preprocessing*, pembangunan ontologi tambahan, pembangunan ontologi provokasi, klasifikasi, dan tahapan terakhir adalah evaluasi. Penggunaan metode ontologi sebagai klasifikasi calon presiden 2019 dan klasifikasi provokasi dan non provokasi dari data setiap calon dalam pemilihan presiden 2019. Dari hasil pengujian pada ontologi tanpa *Rule-Based Classifier* didapatkan rata-rata *f1-score* sebesar 80,5%. Sedangkan pengujian pada ontologi dengan *Rule-Based Classifier* didapatkan rata-rata *f1-score* sebesar 87,7%, maka rata-rata *f1-score* naik sebesar 7,2%

Kata kunci: Ontologi, provokasi, *tweet*, pemilihan presiden 2019

Abstract

The presidential election is a routine activity that is held every 5 years. Many campaigns are voiced directly through social media. Twitter is one of the social media used by the presidential candidate pairs to attract the people of Indonesia with their routine activities. Unfortunately, the social media campaign especially on Twitter was damaged by a number of individuals who deliberately wanted to bring down another presidential candidate pair. The climate of the 2019 presidential election on Twitter's social media is not good. Based on the above problems, we need a system that can detect tweets that contain provocation elements. So as to distinguish between tweets that contain provocation elements from tweets that do not contain provocation elements. In this study divided into several stages, the first stage is retrieval of Twitter data using Twitter API to get the dataset, then preprocessing, development of additional ontologies, construction of provocation ontologies, classification, and the last stage is evaluation. The use of ontology methods as the classification of 2019 presidential candidates and the classification of provocation and non-provocation from the data of each candidate in the 2019 presidential election. From the results of testing on ontologies without a Rule-Based Classifier obtained an average f1-score of 80,5%. While testing on ontology with Rule-Based Classifier obtained an average f1-score of 87,7%, then the average f1-score rises by 7,2%

Keywords: Ontology, provocation, lines, tweets, 2019 presidential election

1. Pendahuluan

Pada bagian ini berisi penjelasan mengenai latar belakang, batasan, dan tujuannya, serta sistematika penulisan penelitian.

1.1 Latar Belakang

Pemilihan presiden (pilpres) merupakan kegiatan rutin yang dilaksanakan setiap 5 tahun sekali [1]. Kegiatan ini bertujuan untuk memilih pemimpin yang diharapkan dapat mengelola pemerintahan menjadi lebih baik dan membawa perubahan bagi kehidupan bernegara dan bermasyarakat [2]. Dalam jangka waktu dekat masyarakat Indonesia akan ikut serta dalam menyuarakan hak pilihnya melalui pemilu presiden 2019. Seperti di tahun 2014 kemarin, Media sosial akan terus memainkan peran penting dalam kampanye politik termasuk pemilihan presiden 2019 [3]. Twitter salah satu sosial media yang di manfaatkan para pasangan calon presiden untuk berkampanye [3]. Sayangnya, kampanye di sosial media khususnya pada Twitter dirusak oleh sejumlah provokator yang sengaja ingin memprovokasi atau menjatuhkan pasangan calon presiden lain. Menimbulkan iklim suasana pemilihan presiden 2019 pada sosial media Twitter menjadi kurang baik. Definisi provokasi sendiri dalam *KBBI Daring* [4] Perbuatan untuk membangkitkan kemarahan; tindakan menghasut; penghasutan; pancingan. Selain itu provokasi juga memiliki efek yang kuat yaitu sifat merendahkan atau

menghina [5].

Berdasarkan permasalahan di atas, maka diperlukan sebuah sistem yang mampu mendeteksi adanya *tweet* yang mengandung unsur provokasi, sehingga dapat membedakan mana *tweet* yang mengandung unsur provokasi dengan *tweet* yang tidak mengandung unsur provokasi. Teknik yang digunakan untuk pengelompokkan ini menggunakan ontologi dan dikombinasikan dengan metode *rule-based classifier*. Alasan menggunakan metode ontologi ini karena proses klasifikasi menjadi lebih efisien dan sederhana karena data dipisahkan berdasarkan domain tertentu yang telah didefinisikan sebelumnya [6]. Penggunaan metode *rule-based classifier* memberikan aturan sesuai dengan kebutuhan sistem [7].

1.2 Topik dan Batasan

Berdasarkan latar belakang yang sudah dituliskan, maka rumusan masalah dalam penelitian ini adalah bagaimana mengidentifikasi *tweet* yang mengandung unsur provokasi pada *tweet* terkait pemilihan presiden 2019 menggunakan metode ontologi yang dikombinasikan dengan metode *rule-based classifier*, serta mengetahui tingkat performansi dari hasil klasifikasi dari kombinasi ontologi yang dibangun dengan *rule-based*.

Adapun batasan masalah yang digunakan untuk membatasi penelitian ini diantaranya adalah bentuk dari provokasi yang terdapat pada penelitian ini hanya provokasi berupa hinaan, identifikasi provokasi hanya dapat dilakukan berupa teks/tulisan dalam bahasa indonesia, tidak dapat menangani tulisan berupa singkatan dan *typo*.

1.3 Tujuan

Tujuan penelitian ini adalah untuk mengimplementasikan kombinasi dari metode ontologi dengan *rule-based classifier* dalam identifikasi *tweet* yang mengandung provokasi terkait pemilihan presiden 2019. Selain itu, mengetahui hasil performansi yang didapat dari kombinasi metode ontologi dengan *rule-based classifier* tersebut.

1.4 Sistematika Penulisan

Sistematika penulisan jurnal ini sebagai berikut: Pada bagian 2 (Studi terkait) menunjukan hasil studi terkait penelitian ini. Pada penelitian ini sistem yang akan diajukan yaitu Identifikasi Provokasi Pada Sosial Media Twitter Terhadap Calon Presiden 2019 Menggunakan Metode Ontologi yang akan dijelaskan pada bagian 3. Pada bagian 4 akan dijelaskan mengenai pengujian, hasil pengujian dan evaluasi hasil sistem. Kemudian pada bagian 5 akan dijelaskan kesimpulan dan saran.

2. Studi Terkait

2.1 Ontologi

Ontologi adalah sebuah representasi primitif yang digunakan untuk memodelkan sebuah domain, dimana ontologi juga berfungsi sebagai jembatan antara pengguna dengan perangkat lunak [8]. Keuntungan utama dari ontologi yaitu sekali dikembangkan akan memberikan ketertarikan tersendiri dan dapat digunakan kembali di semua aplikasi, serta dapat memungkinkan untuk berbagi lebih banyak data dan menggunakan tag yang sederhana untuk memberikan informasi semantik [6]. Pada umumnya komponen utama dalam ontologi yaitu *class*, instance dan relation.

2.2 *Rule-Based Classification*

Rule-Based Classification merupakan teknik yang terdiri dari *antecendent* dan *consequent* yang terkait dengan sebuah hubungan *`if-then* [9]:

$$R{=}antecendent{\rightarrow}consequent \quad (1)$$

Antecendent mendefinisikan suatu kondisi dan terdiri dari token atau urutan token yang disambung oleh operator. Token dapat berupa kata '?' mewakili kata benda yang tepat, atau '#' mewakili istilah target. Sebuah target termis adalah istilah yang mewakili konteks di mana satu set dokumen terjadi, seperti nama seorang politisi, rekomendasi kebijakan, nama perusahaan, merek produk atau judul film. *Consequent* mewakili suatu kata atau frasa termasuk provokasi atau bukan provokasi, dan merupakan hasil dari pemenuhan kondisi yang didefinisikan oleh *antecendent*.

2.3 Evaluasi Performasi

Dalam melakukan evaluasi terhadap hasil prediksi pada klasifikasi menggunakan gagasan *precision* dan *recall*, dimana gagasan ini lebih diterima pada distribusi kelas [10]. *Precision* dan *recall* digunakan bersama untuk mendapatkan nilai bobot, seperti misalnya *F1-Score.* Sebelum dapat menentukan *precision* dan *recall*, maka harus ditentukan dokumen yang relevan dan tidak relevan. Dokumen relevan merupakan *event* target, sedangkan dokumen tidak relevan merupakan bukan *event* target. Untuk lebih jelasnya dapat dilihat pada Tabel 2.1.

Tabel 2. 1 *Confusion Matrix [11]*

		Nilai Aktual	
		Positive	*Negative*
Nilai Prediksi	*Positive*	*True Positive* (TP)	*False Positive* (FP)
	Negative	*False Negative* (FN)	*True Negative* (TN)

TP (*True Positive*) merupakan prediksi positif dan nilai sebenarnya positif, TN (*True Negative*) merupakan prediksi negatif dan nilai sebenarnya negatif, FP (*False Positive*) merupakan prediksi positif dan nilai sebenarnya negatif dan FN (*False Negative*) merupakan prediksi negatif dan nilai sebenarnya positif. Rumus dari *precision*, *recall*, dan *f1-score* dapat dilihat pada persamaan berikut :

1. *Precision*

   Precision adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Rumus pernyataan *precision* sebagai berikut.

$$Precision = \frac{TP}{TP+FP} *100\% \quad (2)$$

2. *Recall*

   Recall merupakan tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Rumus pernyataan *recall* sebagai berikut.

$$Recall = \frac{TP}{TP+FN} *100\% \quad (3)$$

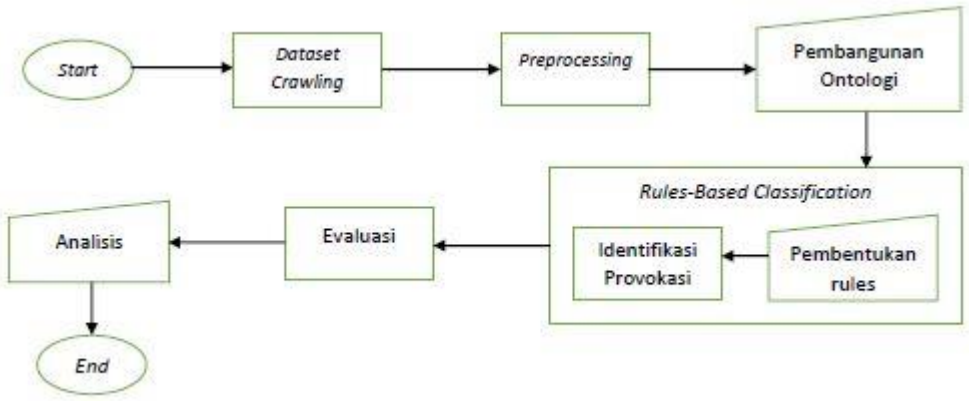
3. *F1-Score*

   F1-score merupakan hasil akhir dari perhitungan nilai dari *precision* dan *recall* untuk mencari nilai rata-rata dari *precision* dan *recall*. *F1-score* juga diartikan sebagai penyetaraan nilai *precision* dan *recall*. Rumus pernyatan *F1-score* sebagai berikut.

$$\text{F1-}Score = \frac{2 \text{ x } precision \text{ x } recall}{precision + recall} *100\% \quad (4)$$

3. Metodologi Penelitian

Gambaran umum dari metode penelitian ini dapat dilihat pada *flowchart* Gambar 3.1



Gambar 3. 1 Metodologi Penelitian

Keterangan :

: *start/end*

: proses manual

: proses sistem

Berikut merupakan penjelasan dari Gambar 3.1 yang menjelaskan tentang alur proses/tahapan pada penelitian ini.

3.1 *Crawling data*

Crawling data dalam penelitian ini menggunakan Twitter API sebagai layanan untuk mendapatkan *dataset* pada twitter. Pembuatan kode program *crawling* menggunakan bahasa pemrograman *Python*. Dalam

pengambilan *dataset* menggunakan kata kunci Pilpres 2019, Prabowo, Jokowi. *Crawling* dimulai pada tanggal 4 april 2019 sampai dengan 19 april 2019. Data yang didapatkan berisi konten berupa tanggal *crawling* dan isi dari *tweet* itu sendiri.

Data yang didapatkan dari hasil *crawling* sebanyank 928 *tweets*. Setelah itu dilakukan validasi oleh seorang psikolog. Data dengan label yang benar sebanyak 370 *tweets* dan data dengan label salah sebanyak 558 *tweets*. Berikut contoh *tweet* yang telah di *crawling*.

Tabel 3. 1 Contoh *Dataset*

Tweet
@jokowi MANTAP...!!! biarkan anjing menggonggong pak Jokowi.. Insya Allah masih lebih banyak rakyat yg mendukungmu

3.2 *Preprocessing*

Langkah selanjutnya setelah labeling adalah preprocessing untuk merapikan *dataset*. Dalam penelitian ini *preprocessing* yang digunakan adalah *removal punctuation*, *case folding*, *stopword removal*.

1. *Data cleaning*

   Pada tahap ini untuk menghapus *username*, tanda baca, simbol dan tautan. Hal ini bertujuan untuk mengurangi data kotor yang akan masuk ke dalam perhitungan klasifikasi.

Tabel 3. 2 *Removal Punctuation*

Tweet sebelum	*Tweet* sesudah
@jokowi MANTAP...!!! biarkan anjing menggonggong pak Jokowi.. Insya Allah masih lebih banyak rakyat yg mendukungmu	MANTAP biarkan anjing menggonggong pak Jokowi Insya Allah masih lebih banyak rakyat yg mendukungmu

2. *Case folding*

   Case folding digunakan untuk mengganti huruf besar menjadi huruf kecil.

Tabel 3. 3 *Case Folding*

Tweet sebelum	*Tweet* sesudah
MANTAP biarkan anjing menggonggong pak Jokowi Insya Allah masih lebih banyak rakyat yg mendukungmu	mantap biarkan anjing menggonggong pak jokowi insya allah masih lebih banyak rakyat yg mendukungmu

3. *Tokenization*

   Pada tahap ini berfungsi untuk memisahkan suatu kalimat menjadi masing-masing kata.

Tabel 3. 4 *Tokenization*

Tweet sebelum	*Tweet* sesudah
mantap biarkan anjing menggonggong pak jokowi insya allah masih lebih banyak rakyat yg mendukungmu	mantap, biarkan, anjing, menggonggong, pak, jokowi, insya, allah, masih, lebih, banyak, rakyat, yg, mendukungmu

4. *Stopword Removal*

   Stopword Removal yaitu tahapan dalam menghilangkan kata penghubung atau kata kata yang sering muncul di teks tetapi tidak terlalu bermakna. Seperti: ini, itu, dan, yaitu, oleh, dan lain lain [10].

Tabel 3. 5 *Stopword Removal*

Tweet sebelum	*Tweet* sesudah
mantap, biarkan, anjing, menggonggong, pak, jokowi, insya, allah, masih, lebih, banyak, rakyat, yg, mendukungmu	mantap, biarkan, anjing, menggonggong, jokowi, insya, allah, rakyat, yg, mendukungmu

3.3 Pembangunan Ontologi

Pada pembangunan ontology ini mengacu kepada referensi sebuah paper dengan judul "*Ontology Development 101 : A Guide to Create Your First Ontology"* oleh Natalya F. Noy dan Deborah L. McGuinness [12]. Berikut merupakan tahapannya.

1. Menentukan domain serta ruang lingkup ontologi.

   Domain yang dibangun pada tahapan ini adalah provokasi dan tambahan. Dimana domain ini nantinya akan berfungsi untuk penentu suatu *tweet* termasuk pada provokasi atau bukan.
2. Menentukan penggunaan ontologi yang sudah ada.

   Pembangunan ontologi provokasi dibuat berdasarkan referensi yang sudah ada sebelumnya. Referensi tersebut mengacu kepada paper terkait bentuk-bentuk kata hinaan atau umpatan dengan judul "*Penggunaan kata umpatan di Twitter berdasarkan gender*" oleh L. S. Anggreni, R. A. Nugroho, H. S. Luthfi, I. M. Kresna dan T. B. Santoso [13] dan "*Bentuk Dan Fungsi Kata Umpatan*

Pada Komunikasi Informal Di Kalangan Siswa SMA Negeri 3 Surabaya : Kajian Sosiolinguistik" oleh R. R. Putra [14] yang terdiri dari binatang, jenis anggota tubuh, jenis makanan, kata asing, kata benda, kata kekerabatan, kata kerja, kata keterangan, kata sifat, makhluk halus, dan profesi.

3. Mengidentifikasi istilah penting
   Identifikasi istilah penting dilakukan dimulai dari mendeskripsikan terlebih dahulu domain dan *scope* yang sudah dibuat pada tahap sebelumnya. Istilah penting berhubungan dengan domain provokasi yang mengacu pada referensi sebuah paper terkait bentuk-bentuk kata hinaan atau umpatan [13] dan [14]. Serta domain tambahan yang diperoleh dari kata netral yang seringkali mengikuti *tweet* hinaan berupa binatang.
4. Mendefinisikan kelas dan hirarki kelas
   Pada tahap ini dilakukan pemetaan terhadap istilah yang sudah ditentukan sebelumnya. Istilah-istilah penting yang sudah ditentukan kemudian dimasukan ke dalam kelas yang saling berhubungan. Pemetaan istilah dilakukan berdasarkan pengertian yang terdapat pada kbbi. Pada pembangunan ontologi ini menggunakan pendekatan kombinasi antara *top-down* dan *bottom-up*. Pendekatan *top-down* dipilih karena ontologi memiliki acuan yang pasti untuk menentukan kelas. Pendekatan *bottom-up* digunakan untuk menentukan *instance*, pendekatan ini akan mengambil istilah yang spesifik ke yang paling umum. Ontologi yang dibangun pada penelitian ini memiliki 11 kelas terkait provokasi yaitu kelas binatang, jenis anggota tubuh, jenis makanan, kata asing, kata benda, kata kekerabatan, kata kerja, kata keterangan, kata sifat, makhluk halus, dan profesi, serta 4 kelas tambahan yaitu kelas berlebihan, kata ganti, pengandaian, dan sanggahan.

Gambar 3. 2 Kelas Pada Ontologi

5. Membuat *instance*
   Berdasarkan kelas dan hirarki kelas yang sudah ditentukan kemudian kata-kata penting yang sudah dipilih dapat dijadikan sebuah *instance*. Pada Gambar 3.3 merupakan salah satu contoh dari *instance*.

Gambar 3. 3 Contoh *Instance* Dari Kelas

Tabel 3.6 dan Tabel 3.7 merupakan gambaran statistik ontologi provokasi dan ontologi tambahan

Tabel 3. 6 Gambaran Statistik Ontologi Provokasi

Kelas	**Jumlah *instance***
Binatang	13 *instance*
Jenis Anggota Tubuh	9 *instance*
Jenis Makanan	3 *instance*
Kata Asing	4 *instance*
Kata Benda	85 *instance*
Kata Kekerabatan	3 *instance*
Kata Kerja	14 *instance*
Kata Keterangan	2 *instance*
Kata Sifat	79 *instance*
Makhluk Halus	4 *instance*
Profesi	14 *instance*

Tabel 3. 7 Gambaran Statistik Ontologi Tambahan

Kelas	**Jumlah *instance***
Berlebihan	6 *instance*
Kata Ganti	9 *instance*
Pengandaian	7 *instance*
Sanggahan	8 *instance*

Dari Tabel 3.6 dan Tabel 3.7 dapat diketahui bahwa jumlah total *instance* untuk masing-masing kelas pada ontologi provokasi adalah sebanyak 230 *instance* dan total *instance* untuk masing-masing kelas pada ontologi tambahan adalah sebanyak 30 *instance*

3.4 *Rules-Based Classification*

Pada penelitian ini penetapan aturan *rule-based* menggunakan aturan *if-then* untuk klasifikasi [15]. Klausa *if* pada aturan *rule-based* merupakan *antecedent* atau *precondition*, dapat terdiri satu atau lebih token untuk memperlihatkan kondisi. Untuk klausa *then* (*conclusion/ rule consequent*) pada aturan *rule-based* merupakan keputusan dari kondisi yang sudah dibuat pada klausa *if*. Untuk penanganan klasifikasi provokasi diperlukan *rules*, termasuk mengatasi keambiguan kata/frasa.

Rules dibuat berdasarkan hasil *review* dari 928 *dataset* yang sudah di*crawling* dan di*preprocessing*. Hasil *dataset review* berupa pola dari struktur kalimat provokasi yang sering dilontarkan oleh *Twitter Users* itu sendiri. Dari pola struktur kalimat provokasi yang sudah didapatkan, kemudian dapat diklasifikasikan kalimat lontaran *Twitter Users* ke dalam *class* provokasi. Klasifikasi tersebut meliputi kondisi berapa kali jumlah hinaan yang dilontarkan, struktur kalimat yang dapat diindikasikan sebagai hinaan, menangani kalimat sanggahan hinaan, serta yang terakhir mengklasifikasikan beberapa kategori hinaan yang dipastikan sebagai kalimat hinaan. Setelah dilakukan *Review* untuk mendapatkan *Rules* yang dibutuhkan. Kemudian diilakukan proses validasi oleh psikolog. Validasi berupa daftar tabel berisi *rules* yang sudah dibuat. Psikolog melakukan ceklis kebenaran terhadapa setiap *rules* yang sudah dijelaskan sebelumnya. Berikut *rules* yang telah dibuat :

R1 = (Binatang = Yes) $> 1 \rightarrow$ Provokasi
R2 = ((Kata Ganti = Yes) $\vee$ (Pengandaian = Yes)) $\wedge$ (Binatang = Yes) $\rightarrow$ Provokasi
R3 = (Binatang = Yes) $\wedge$ ((Berlebihan = Yes) $\vee$ (Kata Ganti = Yes)) $\rightarrow$ Provokasi
R4 = (Sanggahan = Yes) $\wedge$ (Binatang = Yes) $\rightarrow$ Bukan Provokasi
R5 = (Jenis Anggota Tubuh = Yes) $\vee$ (Jenis Makanan = Yes) $\vee$ (Kata Asing = Yes) $\vee$ (Kata Benda = Yes) $\vee$ (Kata Kekerabatan = Yes) $\vee$ (Kata Kerja = Yes) $\vee$ (Kata Keterangan = Yes) $\vee$ (Kata Sifat = Yes) $\vee$ (Makhluk Halus = Yes) $\vee$ (Profesi = Yes) $\rightarrow$ Provokasi

Keterangan :

R1 : Jika kata/frasa hinaan dari kelas binatang dilontarkan lebih dari satu kali, maka hasil menunjukan kalimat provokasi.

R2 : Jika terdapat kata tambahan dari kelas kata ganti atau kata pengandaian yang diikuti dengan kata/frasa hinaan,dari kelas binatang, maka hasil menunjukan kalimat provokasi.

R3 : Jika terdapat kata/frasa hinaan dari kelas binatang diikuti dengan kata tambahan dari kelas berlebihan atau kata ganti, maka hasil menunjukan kalimat provokasi.

R4 : Jika terdapat kata tambahan dari kelas sanggahan diikuti dengan kata/frasa hinaan dari kelas binatang, maka hasil menunjukan kalimat bukan provokasi.

R5 : Jika terdapat kata/frasa hinaan dari kelas manapun kecuali kelas binatang, maka hasil menunjukan kalimat provokasi.

Berikut merupakan contoh *tweet* yang diklasifikasikan sebagai provokasi maupun bukan provokasi berdasarkan masing-masing *rules* yang telah dibuat.

Tabel 3. 8 Contoh *Tweet* Provokasi dengan *Rule* 1

Rule	Hasil Prediksi	*Tweet*
1	Provokasi	coba besok kubu kodok buduk dapat separonya aja pake tuh duit cina perasan bumn apbn **babi anjing kodok**

Berdasarkan Tabel 3.8 dapat diketahui bahwa *tweet* tersebut diklasifikasikan sebagai *tweet* provokasi karena terdapat kata "babi", "anjing", "kodok". Dimana kata-kata tersebut masuk dalam klasifikasi provokasi dengan *rule based classifier* dan termasuk pada *rule* pertama.

Tabel 3. 9 Contoh *Tweet* Provokasi dengan *Rule* 2

Rule	Hasil Prediksi	*Tweet*
2	Provokasi	kau bilang otak dungu kayak kau orang paling pintar aja kau itu sama begonya **seperti keledai**

Berdasarkan Tabel 3.9 dapat diketahui bahwa *tweet* tersebut diklasifikasikan sebagai *tweet* provokasi karena terdapat kata "seperti keledai". Dimana kata-kata tersebut masuk dalam klasifikasi provokasi dengan *rule based classifier* dan termasuk pada *rule* kedua.

Tabel 3. 10 Contoh *Tweet* Provokasi dengan *Rule* 3

Rule	Hasil Prediksi	*Tweet*
3	Provokasi	**dasar cebong** ntar loe team sukses mengantar wiwi pulang solo habis pilpres ya

Berdasarkan Tabel 3.10 dapat diketahui bahwa *tweet* tersebut diklasifikasikan sebagai *tweet* provokasi karena terdapat kata "dasar cebong". Dimana kata-kata tersebut masuk dalam klasifikasi provokasi dengan *rule based classifier* dan termasuk pada *rule* ketiga.

Tabel 3. 11 Contoh *Tweet* bukan Provokasi dengan *Rule* 4

Rule	Hasil Prediksi	*Tweet*
4	Bukan Provokasi	saya sebut anda **bkn monyet** yaa jwb sendiri

Berdasarkan Tabel 3.11 dapat diketahui bahwa *tweet* tersebut diklasifikasikan sebagai *tweet* bukan provokasi karena terdapat kata "bkn monyet". Dimana kata-kata tersebut masuk dalam klasifikasi bukan provokasi menggunakan *rule based classifier* dan termasuk pada *rule* keempat.

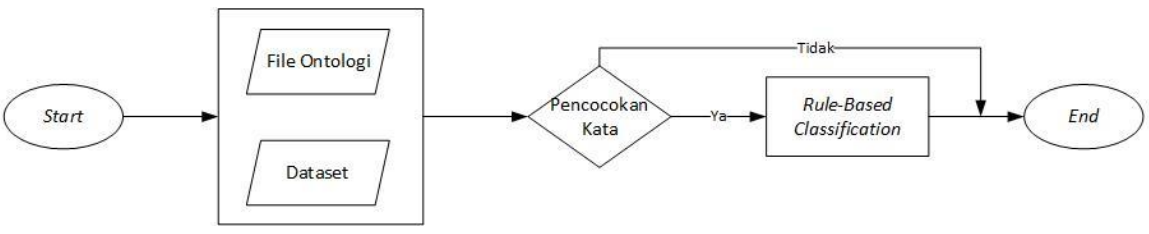
Tabel 3. 12 Contoh *Tweet* Provokasi dengan *Rule* 5

Rule	Hasil Prediksi	*Tweet*
5	Provokasi	bahaya nih si **kurus rakus** batok kepalanya ngurusin duit gakkan paham kedaulatan krn otaknya gakkan nyampe

Berdasarkan Tabel 3.12 dapat diketahui bahwa *tweet* tersebut diklasifikasikan sebagai *tweet* provokasi karena terdapat kata "kurus", dan "rakus". Dimana kata-kata tersebut masuk dalam klasifikasi provokasi dengan *rule based classifier* dan termasuk pada *rule* kelima.

3.5 Identifikasi Provokasi

Pada tahapan Identifikasi Provokasi ini menggunakan sistem deteksi provokasi yang dibuat menggunakan bahasa pemrograman *python* dan menggunakan *tools Jupyter Notebook* yang disediakan oleh *Anaconda Navigator*. Berikut merupakan gambaran alur sistem untuk pengidentifikasi provokasi.



Gambar 3. 4 Identifikasi Provokasi

Pada Gambar 3.4 di atas merupakan tahap-tahap bagaimana cara mengidentifikasikan kalimat provokasi. Berikut penjelasan untuk setiap tahapannya :

1. Tahap pertama. Memasukkan file ontologi dan juga file dataset. File ontologi dalam penelitian ini terdiri dari dua file, yaitu file prov_2.owl yang berisi kelas-kelas provokasi, dan juga tambahan.owl yang berisi kelas-kelas tambahan. Kedua file ini nantinya akan dikombinasikan dengan file prov_2.owl untuk keperluan *rules.* Dalam membaca file ontologi pada bahasa pemrograman *python* menggunakan *library* yang sudah disediakan oleh *Anaconda Navigator* yaitu *Owlready2*
2. Tahap kedua, ketika sistem membaca kata pada setiap tweet, program mengidentifikasi apakah kata tersebut terdapat pada file prov_2.owl dan file tambahan.owl . jika iya, maka dilakukan identifikasi terhadap *rule-based classification*, untuk dicek apakah tweet tersebut mengandung provokasi atau bukan. Jika tidak, maka tweet tersebut mutlak tidak mengandung provokasi. Contoh identifikasi provokasi ontologi dengan *rule based classifier* dan tanpa *rule based classifier* dapat dilihat pada Tabel 3.13 dan Table 3.14

Tabel 3. 13 Contoh Identifikasi Provokasi Ontologi dengan *Rule Based Classifier*

Hasil Prediksi	Hasil Aktual (*labelling*)	*Tweet*
Bukan Provokasi	Bukan Provokasi	jokowi mantap biarkan **anjing** menggonggong jokowi insya allah rakyat yg mendukungmu

Berdasarkan hasil klasifikasi provokasi ontologi dengan *rule-based classifier* didapatkan, bahwa *tweet* di atas termasuk kalimat bukan provokasi. Karena sistem menyocokkan setiap kata pada *tweet* tersebut dengan ontologi yang sudah dibuat. Hasilnya memang ada kata yang cocok dengan ontologi yang sudah dibuat yaitu kata "anjing". Namun ternyata kata tersebut tidak memenuhi aturan/*rule* yang sudah dibuat. Sehingga dapat disimpulkan bahawa kalimat pada *tweet* di atas bukan kalimat provokasi.

Tabel 3. 14 Contoh Identifikasi Provokasi Ontologi tanpa *Rule Based Classifier*

Hasil Prediksi	Hasil Aktual (*labelling*)	*Tweet*
Provokasi	Bukan Provokasi	jokowi mantap biarkan **anjing** menggonggong jokowi insya allah rakyat yg mendukungmu

Berdasarkan hasil klasifikasi provokasi dengan ontologi didapatkan, bahwa *tweet* di atas termasuk kalimat provokasi. Karena sistem langsung menyocokkan setiap kata pada *tweet* tersebut dengan ontologi yang sudah dibuat. Ternyata terdapat kata "anjing" pada ontologi.

Berikut merupakan *pseudocode* untuk identifikasi provokasi menggunakan ontologi dengan *rule based classifier.*

```
for setiap kelas pada seluruh kelas
   for instance pada setiap kelas
      cek hinaan pada tweet
            if hinaan = binatang > 1, maka
               provokasi ← TRUE
            else
                if hinaan = binatang , maka
                   cek2 ← cek kembali kata pada tweet
                   if cek2=subjek atau cek2=kata_ganti atau cek2=pengandaian
                      provokasi ← TRUE
                   else if cek2=berlebihan atau cek2=kata_ganti
                        provokasi ← TRUE
                 else
                       cek 2 ← cek kembali kata pada tweet
                        if cek2 = sanggahan
                           provokasi ← FALSE
               if hinaan=jenis anggota tubuh atau hinaan=jenis makanan atau hinaan=kata asing atau
               hinaan=kata benda atau hinaan=kata kekerabatan atau hinaan=kata kerja atau
               hinaan=kata keterangan atau hinaan=kata sifat atau hinaan=makhluk halus atau
               hinaan=profesi , maka
                    provokasi ← TRUE
               else
                    provokasi ← FALSE
output provokasi
```

3.6 Evaluasi Performansi

Langkah selanjutnya setelah semua *tweet* berhasil diklasifikasikan adalah dengan melakukan perhitungan *recall, presicion*, dan juga *f1-score*. Hasil dari evaluasi akan dibahas pada bagian berikutnya.

4. Analisis dan Pengujian

Setelah sistem selesai dibuat, maka tahap selanjutnya adalah pengujian terhadap sistem. Pengujian ini bertujuan untuk mengetahui seberapa baik sistem yang telah dibuat dengan mengetahui hasil performansi dari sistem. Pengujian metode ini dilakukan pada *dataset* yang sudah didapat dari *hashtags* terkait pilpres 2019.

4.1 Skenario Pengujian

Tabel 4.1 menunjukan beberapa skenario pengujian untuk identifikasi provokasi. Adapun skenario pengujian dari 2 pengujian yang dilakukan adalah sebagai berikut.

Tabel 4. 1 Skenario Pengujian

Nomor	Skenario Pengujian
1.	*Rule-Based Classifier*
2.	Tanpa *Rule-Based Classifier*

4.2 Analisis Pengujian *Rule-Based Classifier* pada Ontologi

Hasil pengujian pada ontologi dengan *Rule-Based Classifier* didapatkan nilai performansi dari *precision, recall,* dan *f1-score* dapat dilihat pada tabel 4.2.

Tabel 4. 2 Hasil Uji Performansi dengan *Rule-Based Classifier*

Performansi			
Metode	*Precision*	*Recall*	*F1-Score*
Rule-Based Classifier	92,8%	83,2%	87,7%
Tanpa *Rule-Based Classifier*	91,1%	72,2%	80,5%

Berdasarkan Tabel 4.2 hasil uji performansi pada ontologi dengan *Rule-Based Classifier* memiliki tingkat presentase yang berbeda-beda. Dari hasil pengujian pada ontologi tanpa *Rule-Based Classifier* didapatkan rata-rata *f1-score* sebesar 80,5%. Sedangkan pengujian pada ontologi dengan *Rule-Based Classifier* didapatkan rata-rata *f1-score* sebesar 87,7%, Dengan demikian, metode *Rule-Based Classifier* dapat meningkatkan rata-rata *f1-score* sebesar 7,2%. Hal ini menunjukan bahwa klasifikasi provokasi menggunakan metode ontologi dengan metode *Rule-Based Classifier* dapat menangani masalah tertentu terkait tweet yang terdapat kata binatang. Berikut merupakan penyebab yang dapat mengurangi *f1-score* ketika menggunakan metode Ontologi dengan *Rule-Based Classifier* yaitu ketidaksesuaian hasil klasifikasi yang disebabkan oleh beberapa *tweet* yang *typo* sehingga kata tersebut tidak terdeteksi oleh sistem. Pada Tabel 4.3 merupakan salah satu contoh *tweet* yang terdapat *typo* atau kesalahan penulisan.

Tabel 4. 3 Contoh hasil data uji

Hasil Prediksi	Hasil Aktual (*labelling*)	*Tweet*
Bukan Provokasi	Provokasi	sesungguhnya orang cerdas pandai sayangnya terpaksa membela si **dunggu** ya kelabakan.

Pada kata tersebut “sesungguhnya orang cerdas pandai sayangnya terpaksa membela si dunggu ya kelabakan.”, kata “dunggu” yang seharusnya “dungu” membuat sistem menilai bahwa *tweet* tersebut bukan provokasi dikarenakan kata “dungu” tersebut merupakan salah satu istilah pada ontologi. Hal tersebut juga menyebabkan nilai *false negative* bertambah sehingga berpengaruh terhadap perhitungan nilai *recall*.

5. Kesimpulan dan Saran

5.1 Kesimpulan

Berdasarkan hasil pengujian dan analisis yang telah dilakukan maka dapat diambil kesimpulan sebagai berikut:

1. Metode Ontologi dengan Rule-based classifier dapat digunakan untuk mengidentifikasi provokasi pada social media twitter terhadap calon presiden 2019.

2. Klasifikasi provokasi menggunakan kombinasi metode Ontologi dengan *Rule-Based Classifier* menghasilkan nilai rata-rata *precision* sebesar 92,8%, *recall* sebesar 83,2%, serta *f1-score* sebesar 87,7%.
3. Penggunaan kombinasi dari metode Ontologi dengan *Rule-Based Classifier* untuk klasifikasi provokasi mampu menaikan *f1-score* sebesar 7,2% daripada penggunaan dari metode ontologi tanpa *Rule-Based Classifier*. Hal ini disebabkan karena *Rule-Based Classifier* yang sudah dibuat mampu menangani perbedaan beberapa kata yang belum bisa dibedakan maknanya oleh penggunaan metode Ontologi biasa.

5.2 Saran

Adapun saran yang didapatkan dari penelitian ini untuk pengembangan sistem selanjutnya adalah penggunaan algoritma *Levenstein Distance* untuk membantu mengidentifikasi jika terjadi *typo* atau salah penulisan.

Daftar Pustaka

[1] UNDANG-UNDANG DASAR REPUBLIK INDONESIA, vol. 3, INDONESIA, 2001.

[2] Z. Arifin and T. D. Hariyana, "PERILAKU PEMILIH (VOTERS BEHAVIOR) PEMILU PRESIDEN," *Journal Diversi*, vol. 1, no. 02, p. 152, September 2015.

[3] B. Ardha, "SOCIAL MEDIA SEBAGAI MEDIA KAMPANYE PARTAI POLITIK 2014," *Jurnal Visi Komunikasi*, vol. 13, no. 01, Mei 2014.

[4] D. Sunendar, "KBBI Daring," Badan Pengembangan dan Pembinaan Bahasa, Kemdikbud, [Online]. Available: https://kbbi.kemdikbud.go.id/entri/provokasi. [Accessed Jumat Oktober 2018].

[5] T. Maradona, "Analisis Isi Pesan Provokatif Dalam Film Buatan Amerika Serikat Tentang Korea Utara (Analisis Hermeneutika Pada Film "The Interview Dan Red Dawn")," Universitas Lampung, Bandar Lampung, 2016.

[6] S, Murugan; P, S Bala; G, Aghila, "An ontology for exploring knowledge in computer networks," *International Journal on Computational Science & Applications*, vol. 3, Agustus 2013.

[7] P. Chikersal, S. Poria, Cambria dan S. E, "Sentiment Analysis Of Tweets By Combining A Rule-Based Classifier With Supervised Learning," dalam *In Proceedings Of The 9th International Workshop On Semantic Evalution*, 2015.

[8] Thakora, Pratik ; Sasi, Sreela;, "Ontology-based Sentiment Analysis," *Procedia Computer Science*, vol. 53, p. 203, 2015.

[9] R. Prabowo dan M. Thelwall, "Sentiment analysis: A combined approach," *Journal of Informetrics*, pp. 143-157, 2009.

[10] Torgo, Luis; Ribeiro, Rita, "Precision and Recall for Regression," *International Conference on Discovery Science*, vol. 5808, 2009.

[11] D. L. Olson and D. Delen, Advanced Data Mining Techniques, Berlin, Heidelberg: Springer, 2008.

[12] N. F. noy and D. L. Mcguinness, A Guide to Create Your First Onthology, 2001.

[13] L. S. Anggreni, R. A. Nugroho, H. S. Luthfi, I. M. Kresna dan T. B. Santoso, "Penggunaan kata umpatan di Twitter berdasarkan gender," *Jurnal Kajian Komunikasi*, vol. 7, no. 1, pp. 121-132, 2019.

[14] R. R. Putra, "Bentuk Dan Fungsi Kata Umpatan Pada Komunikasi Informal Di Kalangan Siswa SMA Negeri 3 Surabaya : Kajian Sosiolinguistik," pp. 93-105, 2012.

[15] P. Naido, P. R. Munhurrun and P. Seegoolam, "An Assessment of Visitor Satisfaction with Nature-Based Tourism Attractions," *International Journal of Management and Marketing Research*, vol. 4, 2011.

[16] V. S, I. J and Nithya, "Preprocessing Techniques for Text Mining - An Overview," *International Journal of Computer Science & Communication Networks*, vol. 5, no. 01, pp. 7-16, 2015.

[17] R. N. T and S. R, "Ontology based clustering algorithm for information retrieval," *Research*, vol. 1, no. 07, p. 1, juli 2013.

[18] B. Qin, Y. Xia, S. Prabhakar and Y. Tu, "A rule-based classification algorithm for uncertain data," in *In Data Engineering*, IEEE, 2009, pp. 1633-1640.