Terakreditasi SINTA Peringkat 2
Surat Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti No. 10/E/KPT/2019
masa berlaku mulai Vol. 1 No. 1 tahun 2017 s.d Vol. 5 No. 3 tahun 2021

Terbit online pada laman web jurnal: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 3 No. 2 (2019) 141 - 147 ISSN Media Elektronik: 2580-0760

Pengaruh *Semantic Expansion* pada *Naïve Bayes Classifier* untuk Analisis Sentimen Tokoh Masyarakat

Muhamad Satria Adhi[1], Muhammad Zidny Naf'an[2], Elisa Usada[3]
[1,2,3]Program Studi Informatika, Fakultas Teknologi Industri dan Informatika, Institut Teknologi Telkom Purwokerto
[1]15102027@st3telkom.ac.id, [2]zidny@ittelkom-pwt.ac.id, [3]elisa@ittelkom-pwt.ac.id

Abstract

Sentiment analysis is a field of study that analyzes one's opinions, sentiments, evaluations, attitudes and emotions that are conveyed in written text. There are several factors that cause low accuracy results from sentiment analysis. These factors such as less optimal stemming process, word negation process that does not produce maximum results, writing errors in the dataset, and others. These problems can be overcome by optimizing the process of normalizing words, negation, stemming, and adding methods of semantic expansion. The purpose of adding the Semantic Expansion method and improvement in the process is to increase the accuracy value of the Sentiment Analysis process. This study aims to create a sentiment analysis model from public comments on a public figure (Ridwan Kamil) using the Naïve Bayes Classifier algorithm. Based on the test results in the sentiment analysis model using the Naïve Bayes Classifier method with the addition of the semantic expansion method it is proven that it can improve accuracy. The accuracy obtained using the semantic expansion method is 72%. While the value of accuracy without semantic expansion is 70%.

Keywords: accuration, naïve bayes classifier, semantic expansion, sentiment analysis, text preprocessing.

Abstrak

Analisis sentimen merupakan sebuah bidang studi yang menganalisis pendapat, sentimen, evaluasi seseorang, sikap dan emosi seseorang yang disampaikan dalam bentuk teks tertulis. Terdapat beberapa faktor yang menyebabkan rendahnya hasil akurasi dari analisis sentimen. Faktor-faktor tersebut seperti proses *stemming* yang kurang optimal, proses negasi kata yang tidak menghasilkan hasil yang maksimal, kesalahan penulisan pada dataset, dan lainnya. Permasalahan tersebut dapat diatasi dengan melakukan optimalisasi pada proses normalisasi kata, negasi, *stemming*, dan penambahan metode *semantic expansion*. Tujuan dari penambahan metode *semantic expansion* dan perbaikan pada proses tersebut adalah untuk meningkatkan nilai akurasi dari proses Analisis Sentimen. Penelitian ini bertujuan membuat model analisis sentimen dari komentar masyarakat terhadap seorang tokoh (Ridwan Kamil) menggunakan algoritma *Naïve Bayes Classifier*. Berdasarkan hasil pengujian yang dilakukan penulis dalam melakukan analisis sentimen menggunakan metode *Naïve Bayes Classifier*, penambahan metode *semantic expansion* terbukti dapat meningkatkan akurasi. Akurasi yang didapatkan dengan menggunakan metode *semantic expansion* adalah sebesar 72% Sedangkan nilai akurasi tanpa *semantic expansion* adalah 70%.

Kata kunci: akurasi, *analisis sentimen, naïve bayes classifier, text preprocessing, semantic expansion*

1. Pendahuluan

Internet menjadi pilihan utama masyarakat untuk melakukan sebuah komunikasi, karena biaya yang terjangkau dan dapat mengirim informasi dengan cepat dan dapat dijangkau oleh orang lain yang jaraknya jauh. Dengan menggunakan internet, masyarakat dapat mengolah, menerima, mencari, mengirim, menyimpan, dan menyebarluaskan data atau informasi secara cepat dan mudah. Media sosial merupakan salah satu wadah atau tempat untuk mendapatkan informasi yang dibagikan oleh masyarakat secara umum. *Facebook, Twitter, Instagram* merupakan contoh dari media sosial yang tersedia saat ini di internet. Aplikasi-aplikasi tersebut biasa digunakan masyarakat untuk memberikan kritik, menyampaikan sebuah opini dan saran terhadap suatu hal, mencari data untuk dijadikan informasi, atau hanya sekedar memberikan informasi dan berbagi pengalaman tertentu. Selain pencarian data secara manual di media sosial, pengguna dapat mencari

Diterima Redaksi : 22-05-2019 | Selesai Revisi : 22-07-2019 | Diterbitkan Online : 01-08-2019

atau menggali informasi dari data teks media sosial dengan menggunakan sebuah sistem yang di dalamnya sudah diterapkan algoritma dan metode tertentu. Teknik penggalian informasi dari teks tersebut biasa disebut dengan *text mining*. *Text mining* merupakan metode klasifikasi bentuk variasi dari *data mining* berusaha menemukan pola yang menarik dari sekumpulan data tekstual yang berjumlah besar [1]. Bentuk dari data yang dimasukkan pada *text mining* merupakan data-data yang tidak terstruktur, seperti dokumen XML, PDF, *Microsoft office* dan sejenisnya. Analisis sentimen merupakan salah satu penerapan dari metode *text mining*. Analisis sentimen merupakan bidang studi yang menganalisis pendapat seseorang, sentimen seseorang, evaluasi seseorang, sikap seseorang dan emosi seseorang dalam bentuk teks tertulis [2]. Komentar masyarakat di media sosial merupakan salah satu sumber data untuk analisis sentimen.

Analisis sentimen adalah suatu bidang ilmu yang mengintegrasikan *natural language processing*, komputasi linguistik, dan analisis teks yang bertujuan untuk mengidentifikasi sentimen teks sehingga mengidentifikasi opini mengenai suatu produk yang disampaikan oleh masyarakat [3]. Tujuan dari analisis sentimen adalah untuk menganalisis penilaian dari masyarakat terhadap suatu benda, baik itu benda mati ataupun benda hidup. Sebuah komentar masyarakat di sosial media atau artikel-artikel di internet dapat dijadikan sebagai data yang digunakan untuk menganalisis.

Analisis sentimen memiliki beberapa tahapan, tahap pertama yaitu pengumpulan data sebelum dilakukannya proses *preprocessing*. Data yang dimasukkan diambil dari artikel di internet seperti komentar masyarakat terhadap suatu objek di media sosial. Pada tahap *preprocessing* terdapat 5 proses seperti *case folding*, *tokenizing*, *filtering*, *stemming*, dan *convert negation*. setelah tahap *preprocessing*, selanjutnya data hasil *preprocessing* diklasifikasikan menjadi 2 kelas sentimen yaitu kelas positif atau negatif.

Tujuan dilakukannya *text preprocessing* yaitu untuk menghilangkan derau pada data, menyeragamkan bentuk kata, dan mengurangi volume kata [4]. Hasil dari *text preprocessing* adalah sebuah dataset yang informatif dan efektif sehingga dapat dianalisis dengan mudah.

Menurut [5] setidaknya terdapat tiga permasalahan pada tahap *preprocessing* dari sentimen analisis teks media sosial, yaitu normalisasi kata, pembentukan kata negasi dan proses *stemming*. Pada [5] menggunakan teknik *query expansion* untuk mengatasi permasalahan tersebut. Data yang digunakan oleh [5] adalah komentar masyarakat mengenai aplikasi *mobile*, dengan metode klasifikasi Naïve Bayes menghasilkan akurasi setelah menambahkan *query expansion* sebesar 98%, dimana tanpa *query expansion* hanya mnghasilkan akurasi sebesar 95%.

Selain [5], terdapat beberapa penelitian lain yang membahas tentang analisis sentimen. Afshoh [6] melakukan penelitian tentang analisis sentimen terhadap persepsi masyarakat mengenai kenaikan harga jual rokok. Data yang digunakan sebanyak 350 teks *tweet*. Pada hasil akhir dapat diketahui bahwa nilai sentimen positif yang paling banyak terbentuk dalam menanggapi topik kenaikan harga jual rokok sebesar 53%. Sedangkan untuk evaluasi kinerja sistem Algoritma *Naïve Bayes Classifier* menunjukkan hasil klasifikasi yang lebih baik daripada metode *Lexicon Based* [6].

Antinasari, dkk [7] melakukan penelitian tentang analisis sentimen terhadap opini masyarakat mengenai perfilman di Indonesia. Metode klasifikasi *Naive Bayes* dengan perbaikan kata tidak baku dapat diterapkan pada proses analisis sentimen tentang opini film pada dokumen Twitter berbahasa Indonesia. Data latih dan data uji dilakukan proses *preprocessing* terlebih dahulu, yang mana pada *preprocessing* [7] melakukan perbaikan kata tidak baku menggunakan kamus kata baku yang dilakukan setelah *case folding*.

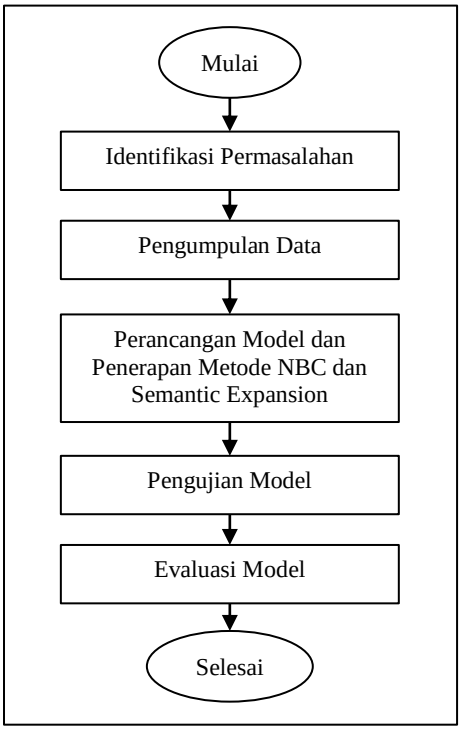
Mentari, dkk[8] melakukan penelitian analisis sentimen pada *tweets* yang menanggapi kebijakan diberlakukannya Kurikulum 2013. Penelitian Analisis sentimen ini memiliki 4 proses utama yang dilakukan sistem yaitu, *text preprocessing*, perhitungan term *weighting* (TFIDF) pada data uji dan latih, *feature selection* dengan menggunakan metode *Query Expansion Ranking*, dan klasifikasi dengan *K-Nearest Neighbor* pada setiap dokumen uji apakah termasuk kelas "opini positif" atau "opini negatif".

Fannisa, dkk[9] melakukan penelitian analisis sentimen komentar masyarakat tentang pariwisata Kota Malang yang diambil dari sebuah *website* pariwisata bernama TripAdvisor dan diklasifikasikan menjadi dua kelas yaitu positif dan negatif. Pemberian seleksi fitur dalam proses analisis sentimen bertujuan untuk mengurangi dimensi fitur, dan metode *Query Expansion Ranking* bekerja dengan baik bersama metode *Multinomial Naive Bayes*. Berdasarkan pengujian, algoritma *Query Expansion Ranking* menghasilkan akurasi tertinggi sebesar 86.6% pada seleksi fitur 75%.

Berdasarkan tinjauan pustaka yang telah penulis sampaikan, penulis melakukan penelitian analisis sentimen pada komentar masyarakat terhadap sosok Ridwan Kamil yang disampaikan melalui media sosial Twitter. Penulis juga akan meneliti apakah ada pengaruh penggunaan *semantic expansion* dalam analisis sentimen dengan Algoritma *Naïve Bayes Classifier*. Diharapkan dengan adanya perbaikan pada tahap *preprocessing* dan penambahan metode *semantic expansion* tersebut dapat meningkatkan akurasi dalam hasil penelitian.

2. Metode Penelitian

Analisis sentimen yang dilakukan pada penelitian ini terdiri dari 5 tahap. Gambar 1 menggambarkan sebuah alur dari tahapan-tahapan analisis sentimen yang dilakukan.



Gambar 1. Tahapan-tahapan penelitian

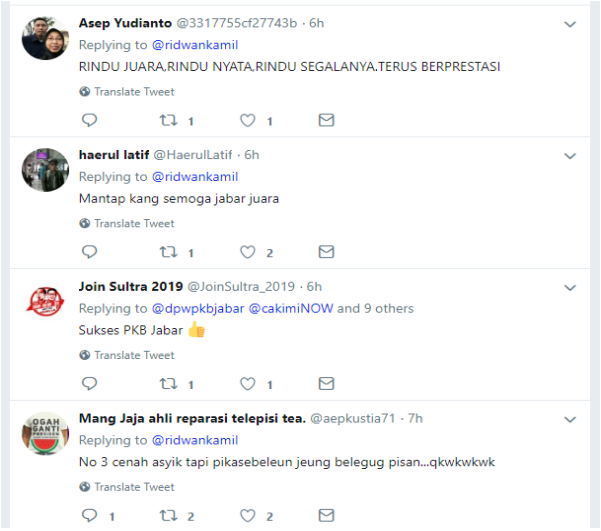
2.1. Pengumpulan Data

Pada tahap pengumpulan data, penulis menggunakan teknik *web scraping*, dimana dengan teknik ini penulis dapat mengambil data semi-terstruktur dari laman web Twitter. Proses *Web scraping* dilakukan menggunakan salah satu modul Twitterscraper yang dapat diunduh dari laman https://github.com/taspinar/twitterscraper.

Tweet yang diambil berisi opini atau penilaian masyarakat tentang Ridwan Kamil saat menjabat sebagai Wali Kota Bandung. Pencarian data tersebut diawali dengan eksplorasi pada aplikasi Twitter dengan cara mencari kata kunci yang berhubungan dengan Ridwan Kamil atau mencari *tweet* dengan *hash tag* tertentu mengenai Ridwan Kamil. Jumlah *tweet* yang dikumpulkan pada penelitian ini adalah sebesar 544 *tweet*. Gambar 2 merupakan contoh *tweet* mengenai Ridwan Kamil. Data yang dikumpulkan selanjutnya diberi label kelas oleh *expert* bidang kebahasaan.

2.2. Perancangan Model dan Penerapan Metode Naïve Bayes Classifier dan Semantic Expansion

Model yang dibuat pada penelitian ini adalah untuk mengklasifikasikan *tweets* dari pengguna Twitter mengenai Ridwan Kamil menjadi dua kelas sentimen yaitu kelas positif dan kelas negatif. *Naive Bayes Classifier* dan *Semantic Expansion* merupakan metode yang digunakan untuk melakukan pengklasifikasian pada analisis sentimen. *Semantic Expansion* digunakan dengan cara menambahkan kata tambahan dari kata adjektif berupa sinonim yang diambil dari API Kateglo. Gambar 3 menunjukkan tahapan pada perancangan model dan penerapan *Naïve Bayes Classifier*.



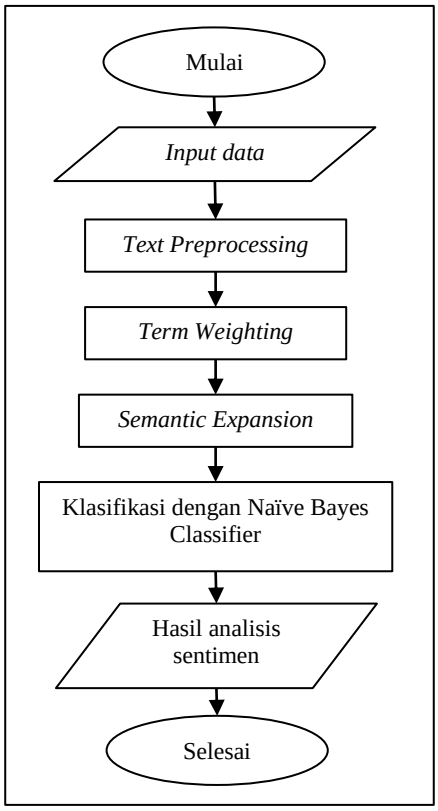
Gambar 2. Contoh isi *tweet* tentang Ridwan Kamil

Selanjutnya, penulis mengolah seluruh teks pada *data set* melalui tahap *preprocessing*. Terdapat 6 proses yang digunakan pada tahap *preprocessing*, yaitu:

1. *Case folding* untuk merubah teks menjadi *lower case*.
2. *Tokenizing*, yaitu proses memotong setiap kata dalam bentuk string dari suatu kalimat pada *data set.*
3. *Lexicon based stopwords removal*, untuk menghilangkan kata yang tidak memiliki pengaruh pada topik analisis sentiment berdasarkan kamus kata yang masuk daftar stopword.
4. *Stemming* menggunakan algoritma Nazief-Adriani yang telah diimplementasikan pada modul sastrawi (https://github.com/sastrawi/sastrawi/wiki/Stemming-Bahasa-Indonesia).
5. *Text cleaning*, untuk menghilangkan angka, hashtag, alamat website, email, *username*, dan simbol dalam suatu kalimat.
6. Membangkitkan antonim dari kata sifat yang sebelumnya terdapat kata negasi. Kata yang dikonversi hanya kata negasi (tidak, kurang, dan belum) yang diikuti kata sifat setelahnya. Penulis menggunakan API Kateglo.com untuk membangkitkan antonim kata sifat.

Selain 6 tahapan tersebut, pada tahap *preprocessing* penulis menambahkan proses

semantic expansion yang akan dijelaskan pada Subbab 2.4.



Gambar 3. Diagram alur proses perancangan model

Kemudian semua teks hasil *preprocessing* masuk pada tahap *term weighting*, dimana pada tahap ini sistem memberikan bobot setiap *term* pada teks dengan cara melakukan perhitungan bobot *Term Frequency- Invers Document Frequency* (TF-IDF) pada setiap *term*-nya. Nilai bobot yang telah diperoleh tersebut digunakan pada proses klasifikasi.

Term Frequency dinotasikan dengan $\mathbf{tf_{(t,d)}}$ yang menyatakan jumlah kemunculan term pada dokumen **d** [10]. Tahap selanjutnya yaitu mencari nilai IDF(*Inverse Document Frequency).* Metode *Inverse Document Frequency* digunakan untuk memperhitungkan faktor-faktor yang menyangkut penyebaran suatu term dalam sekumpulan dokumen [11]. $\mathbf{df_t}$ merupakan dokumen yang mengandung term $\mathbf{t_i}$ dalam sebanyak **N** dokumen. Berdasarkan rumus mencari IDF, berapapun nilai *Term Frequency*, apabila nilai $\mathbf{N{=}df_t}$ maka akan didapatkan hasil 0 (nol) untuk perhitungan IDF. Untuk itu dapat ditambahkan nilai 1(satu) pada sisi IDF, sehingga perhitungan bobotnya menjadi sebagai berikut[10]:

$$idf(t_i) = \log \frac{N}{df_t} + 1 \quad (1)$$

Keterangan :

$idf(t_i)$ = nilai idf term t_i
N = jumlah seluruh dokumen
df = dokumen yang mengandung term $\mathbf{t_i}$

Persamaan $idf(t_i)$ digunakan untuk mengantisipasi dimana terdapat term $\mathbf{t_i}$ yang ada pada seluruh koleksi dokumen. Tujuan dari penambahan nilai 1 pada pencarian nilai $idf(t_i)$ adalah untuk menghindari hasil 0 atau *zero*. Nilai dari TF dan IDF digunakan untuk mencari bobot suatu kata pada kelas tertentu atau yang biasa disebut dengan TF-IDF. Berikut merupakan persamaan TF-IDF:

$$tfidf_{t,d} = tf_{t,d} * idf_t \quad (2)$$

Keterangan :

$tfidf_{t,d}$ = bobot term ke-t pada kelas tertentu.
$tf_{t,d}$ = jumlah kemunculan term ke-t pada kelas tertentu.
idf(t) = nilai idf dari term ke-t.

2.3. Klasifikasi teks menggunakan Naive Bayes Classifier

Naive Bayes Classifier pada penelitian ini digunakan untuk mengklasifikasi teks. Alasan penggunaan metode tersebut adalah karena proses pengklasifikasian dengan metode tersebut memiliki tingkat efektifitas yang tinggi, dan mempunyai tingkat akurasi yang tinggi.

Naïve Bayes classifier merupakan suatu metode klasifikasi yang menggunakan perhitungan probabilitas [12]. Metode *Naive Bayes Classifier* berfungsi untuk mencari nilai probabilitas munculnya suatu kejadian berdasarkan pengaruh yang dihasilkan dari hasil observasi. Metode tersebut berasal dari pengembangan Teorema Bayes. Metode tersebut memiliki asumsi bahwa setiap atribut bersifat bebas (independen). Metode *Naive Bayes Classifier* dapat diimplementasikan pada proses klasifikasi gambar dan klasifikasi teks. Secara umum rumus dasar proses pengklasifikasian *Naive Bayes* dapat dilihat seperti berikut, dengan asumsi bahwa atribut tidak saling terkait dan data sudah memiliki kelas pada *dataset*. Pada proses klasifikasi, algoritma akan mencari probabilitas tertinggi dari semua kategori dokumen yang diuji sebagaimana yang ditunjukkan pada Persamaan 3 [13]:

$$h_{MAP} = \underset{h \in H}{argmax}\ P(D|h)\ P(h) \quad (3)$$

Keterangan :

h_{MAP} = Nilai hasil klasifikasi *Naïve Bayes Classifier.*
h = Kategori atau kelas pada dokumen.
$P(h)$ = Probabilitas dari h.
D = Term atau kata pada dokumen.
$P(D|h)$ = Probabilitas suatu kata pada kelas tertentu.

Untuk mencari nilai dari $P(D|h)$ atau probabilitas suatu kata pada kelas tertentu dapat dilakukan dengan Persamaan 4.

$$P(D|h) = \frac{n_k + 1}{n + |kosakata|} \quad (4)$$

Keterangan :

n_k = Jumlah fekuensi kemunculan kata(w_i) pada kelas tertentu.
n = Jumlah frekuensi kemunculan seluruh kata pada kelas tertentu.
|kosakata| = Jumlah seluruh kata dari semua kelas.
$P(D|h)$ = Probabilitas suatu kata pada kelas tertentu.

Prior adalah perhitungan untuk mencari nilai probabilitas munculnya suatu kelas pada semua dokumen. Perhitungan *prior* ditunjukkan pada Persamaan 5.

$$\mathrm{P}(h) = \frac{|N_j|}{|N|} \quad (5)$$

Keterangan :

N_i = banyaknya dokumen pada kelas j.
N = jumlah seluruh dokumen.

Pada penelitian ini, mengimplementasikan metode Multinomial Naive Bayes (MNB) yang terdapat pada modul sklearn pada python. MNB merupakan algoritma yang naïve karena mengasumsikan indepedensi diantara kemunculan kata-kata dalam dokumen, tanpa memperhitungkan urutan kata dan informasi konteks dalam kalimat atau dokumen secara umum. Selain itu metode tersebut memperhitungkan jumlah kemunculan kata dalam dokumen [14]. Persamaan 6 menunjukkan persamaan MNB menggunakan pembobotan kata TF-IDF[15]:

$$P(t_n|c) = \frac{Wct + 1}{(\sum w^I \in VW^I ct) + B^I} \quad (6)$$

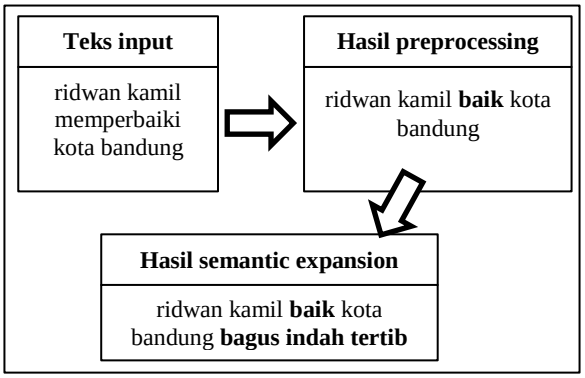
Keterangan:

Wct = nilai hasil dari pembobotan TF-IDF atau bobot dari kata t pada kelas c.
$(\sum w^I \in VW^I ct)$ = jumlah total bobot dari semua kata yang berada di kelas c.
B^I = jumlah W atau bobot kata unik dari semua dokumen.

2.4. *Semantic Expansion*

Pada tahap ini seluruh kata sifat yang terdeteksi pada data latih dan data uji akan diperluas dengan cara menambahkan beberapa kata yang memiliki makna sama. Penulis menggunakan teknik *automatic semantic expansion* [16]. Dengan menggunakan teknik tersebut, penulis tidak perlu melakukan perluasan kata secara manual. Gambar 5 menunjukkan proses *semantic expansion*.

Dalam penelitian ini digunakan API **kateglo.com** untuk membangun *atomatic semantic expansion* berdasarkan sinonim kata. Gambar 4 adalah contoh hasil dari proses *Semantic Expansion*.

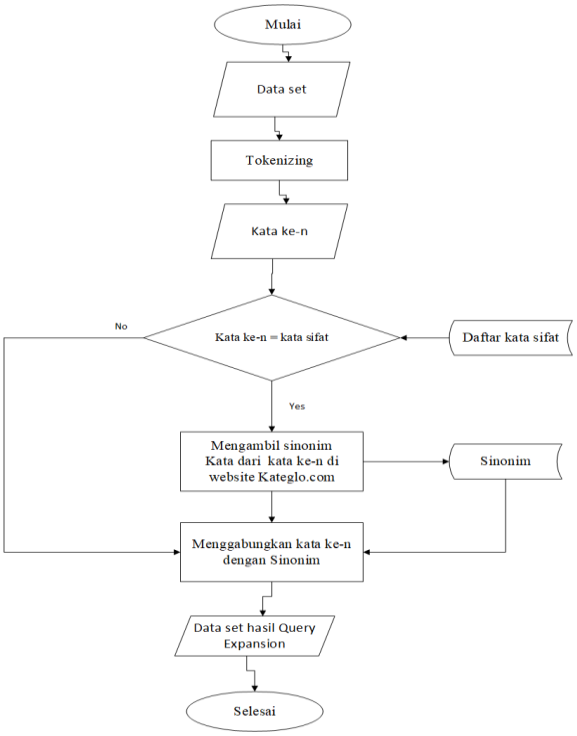


Gambar 4. Contoh hasil *semantic expansion*

Berdasarkan Gambar 5, program mendeteksi sebuah kata adjektif. Kata adjektif tersebut adalah kata "baik". Kata tersebut diperluas dengan cara menambahkan beberapa kata baru dengan makna yang sama atau sinonim.

2.5. *Pengujian Model*

Pengujian model dilakukan untuk mengetahui apakah model tersebut dapat berjalan dengan baik sesuai dengan yang diharapkan. Tahap pengujian *model* ini diawali dengan *input* data set. Selanjutnya *model* akan melakukan pengolahan terhadap data set tersebut dengan menerapkan metode *Naive Bayes Classifier* dan *Semantic Expansion* untuk mengklasifikasikan sentimen. Jika hasil yang dikeluarkan dari tahap ini menghasilkan nilai akurasi yang tinggi, maka *Model* dikatakan berhasil.



Gambar 5. Alur *semantic expansion*

2.6. Evaluasi Model

Penulis melakukan evaluasi terhadap keluaran yang dihasilkan dari tahap pengujian model. Keluaran yang dihasilkan dari *model* tersebut berupa data yang sudah diklasifikasikan. Setelah mendapatkan hasil klasifikasi data dari model, tahap selanjutnya adalah membandingkan hasil klasifikasi dari model dengan hasil klasifikasi data secara manual yang dilakukan oleh penulis. Apabila *output* yang diklasifikasikan oleh model sama dengan *output* yang diklasifikasikan secara manual maka model tersebut dapat berjalan dengan baik. Langkah selanjutnya adalah menghitung akurasi. Penghitungan akurasi dilakukan untuk mendapatkan nilai seberapa tepat model dalam melakukan klasifikasi data. Nilai akurasi didapatkan melalui pengujian secara *Cross Validation*.

3. Hasil dan Pembahasan

Nilai akurasi yang dihasilkan dari pengujian Analisis Sentimen dengan menggunakan metode Naïve Bayes Classifier dan Semantic Expansion lebih tinggi dibandingkan tanpa menggunakan metode Semantic Expansion. Penggunaan metode Naïve Bayes Classifier dan Semantic Expansion pada penelitian ini mendapatkan nilai akurasi sebesar 72%, dan tanpa menggunakan semantic expansion sebesar 70%.

Penambahan metode *semantic expansion* pada *dataset* sangat mempengaruhi hasil dari klasifikasi data, karena terdapat beberapa kata baru (*Out of Vocabulary*) yang ada pada data uji pada saat proses klasifikasi, namun kata baru tersebut tidak terdapat pada data latih. Secara istilah kata baru tersebut memiliki makna yang sama dengan kata tertentu pada data latih, tetapi pada saat proses klasifikasi data uji kata baru tersebut dianggap sebagai kata baru yang tidak memiliki nilai bobot. Misalkan terdapat kata "seret" pada data uji, namun kata tersebut tidak terdapat pada data latih. Hal tersebut merupakan penyebab terjadinya ketidaksempurnaan proses klasifikasi dalam mendeteksi kata "seret" sebagai kata baru yang belum memiliki kelas. Sedangkan kata "seret" pada data uji merupakan kata yang memiliki makna sama dengan kata "macet" pada data latih yang masuk ke dalam kelas negatif. Tabel 1 merupakan contoh hasil dari *semantic expansion*.

Dengan *semantic expansion*, perluasan kata dilakukan sebelum proses klasifikasi. Sehingga kata yang akan diklasifikasi sudah ditambah dengan kata baru yang memiliki makna sama. Seperti pada pengujian *Multinomial Naïve Bayes* dengan *Semantic Expansion* diatas, kata "baik" pada data uji diperluas dengan penambahan kata "bagus", "indah", dan "tertib".

Tabel 1. Contoh hasil *semantic expansion*

Dokumen	Tweet	Kelas
1	bandung macet kalau tidak s olusi kongkrit pergi kantor s ubuh sendat serat	-
2	formasi baik persib mata rid wan formasi baik pernah mil ik persib versi bapak emil ba gus indah tertib	+
3	bapak rangka merdeka kaya bagus kalau lomba baik inda h elok	+
4	idiiiih malu deh hina jakarta minta ganti rugi nye jakarta cowo bukan buruk ceroboh k eji	-

Hasil akurasi dari penelitian sebelumnya jauh lebih baik daripada hasil akurasi yang dilakukan pada penelitian ini. Hal tersebut disebabkan karena faktor-faktor berikut:

1. Tidak semua *tweet* yang diambil pada saat proses pengambilan data mengandung sentimen terhadap Ridwan Kamil. Dalam data set terdapat tweet "rumahnya kang ridwan kamil terbuat dari ribuan botol kratingdeng". *Tweet* tersebut merupakan contoh *tweet* yang bersifat netral. *Tweet* yang bersifat netral merupakan *tweet* yang tidak mengandung sentimen negatif ataupun positif terhadap Ridwan Kamil. Berbeda dengan penelitian sebelumnya, dimana komentar masyarakat seluruhnya mengandung sentimen dan hanya tertuju kepada sebuah aplikasi *mobile.*
2. Terdapat beberapa *tweet* di data set yang menggunakan Bahasa selain Bahasa Indonesia. Salah satu contohnya seperti "Pa @ridwankamil nahanya wayah kieu saya sok lapar??". Pada penelitian sebelumnya hampir seluruh data setnya menggunakan Bahasa Indonesia.
3. Proses normalisasi kata yang masih belum berjalan secara optimal. Memperbaiki kata seperti penulisan singkatan, salah penulisan dan penggunaan kata tidak baku pada proses normalisasi kata belum menghasilkan hasil yang maksimal.
4. Proses pelabelan yang dilakukan oleh *expert judgement* tidak seluruhnya tepat, terdapat beberapa kesalahan dalam pemberian label pada data set. Salah satu contohnya seperti tweet "RT : @ridwankamil Selamat istirahat kawan. Suatu hari kita semua akan pulang. pic.twitter.com/sUvpLbPqtw". Tweet tersebut mengandung sentimen positif, namun oleh Expert Judgement tweet tersebut diberi label negatif. Kesalahan tersebut sangat mempengaruhi hasil akurasi dari proses klasifikasi.

4. Kesimpulan dan Saran

4.1. Kesimpulan

Berdasarkan hasil pengujian yang dilakukan penulis dalam melakukan analisis sentimen pengguna Twitter terhadap Ridwan Kamil selama masa jabatannya sebagai Wali Kota Bandung dapat disimpulkan bahwa penambahan metode *semantic expansion* dalam proses klasifikasi teks dapat meningkatkan akurasi. Hal tersebut diketahui dari hasil pengujian yang dilakukan oleh penulis mendapatkan nilai akurasi sebesar 70% jika hanya menggunakan metode *Naïve Bayes Classifier* dalam melakukan klasifikasi. Sedangkan nilai akurasi yang didapat jika menggunakan metode *semantic expansion* adalah sebesar 72%. Dari hasil tersebut dapat membuktikan bahwa dengan adanya penambahan metode *semantic expansion*, Metode Naïve Bayes Classifier menjadi lebih baik dalam melakukan klasifikasi teks.

4.2. Saran

Berdasarkan penelitian yang telah dilakukan, berikut merupakan saran yang diberikan penulis untuk pengembangan penelitian selanjutnya:

1. Penulis sudah membuat sebuah kamus yang berisi perbaikan kata dari kata-kata modern, slang, dan kesalahan penulisan kata atau kata ambigu. Namun, masih ada beberapa kata ambigu yang tidak terdeteksi oleh kamus tersebut. Perlu adanya pengembangan lebih lanjut dalam pembuatan kamus untuk mengatasi permasalahan ini. Penulis menyarankan untuk menggunakan kamus pada *website* **kitabgaul.com** dalam mencari kata untuk memperbaiki penggunaan kata ambigu.
2. Pembuatan kelas baru selain kelas negatif dan kelas positif. Kelas baru yang dibutuhkan merupakan kelas netral. Kelas tersebut untuk menempatkan data set yang tidak mengandung sentimen positif dan tidak mengandung sentimen negatif. Tujuannya untuk meningkatkan akurasi dari proses klasifikasi.
3. Data set yang digunakan sebagai bahan pengujian seharusnya hanya menggunakan satu Bahasa saja. Hal tersebut untuk menghindari terjadinya *Out of Vocabulary*.

Daftar Rujukan

[1] R. Feldman dan J. Sanger, *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. New York: Cambridge University Press, 2007.

[2] N. Saputra, T. B. Adji, dan A. E. Permanasari, "Analisis Sentimen Data Presiden Jokowi dengan Preprocessing Normalisasi dan Stemming menggunakan Metode Naive Bayes dan SVM," *J. Din. Inform.*, vol. 5, no. 1, 2015.

[3] G. Vinodhini dan R. M. Chandrasekaran, "A comparative performance evaluation of neural network based approach for sentiment classification of online reviews," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 28, no. 1, hal. 2–12, 2016.

[4] A. M. Imelda Muis, "Penerapan Metode Support Vector Machine (SVM) Menggunakan Kernel Radial Basis Sunction (RBF) Pada Klasifikasi Tweet," *J. Sains, Teknol. dan Ind.*, vol. 12, no. 2, hal. 189–197, 2015.

[5] R. F. N. Firmansyah, M. A. Fauzi, dan T. Afirianto, "Sentiment Analysis pada Review Aplikasi Mobile Menggunakan Metode Naive Bayes dan Query Expansion," *DORO PTIIK*, vol. 8, hal. 14, 2016.

[6] F. Afshoh, "Analisa Sentimen Menggunakan Naïve Bayes Untuk Melihat Persepsi Masyarakat Terhadap Kenaikan Harga Jual Rokok Pada Media Sosial Twitter," *Dr. Diss. Univ. Muhammadiyah Surakarta*, 2017.

[7] P. Antinasari, R. S. Perdana, dan M. A. Fauzi, "Analisis Sentimen Tentang Opini Film pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes dengan Perbaikan Kata Tidak Baku," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 12, hal. 1733–1741, 2017.

[8] N. D. Mentari, M. A. Fauzi, dan L. Muflikhah, "Analisis Sentimen Kurikulum 2013 Pada Sosial Media Twitter Menggunakan Metode K-Nearest Neighbor dan Feature Selection Query Expansion Ranking," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 8, hal. 2739–2743, 2018.

[9] S. Fanissa, M. A. Fauzi, dan S. Adinugroho, "Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 8, hal. 2766–2770, 2018.

[10] C. D. Manning, P. Raghavan, dan H. Schütze, *An Introduction to Information Retrieval*. Cambridge: Cambridge University Press, 2009.

[11] S. Robertson, "Understanding inverse document frequency: On theoretical arguments for IDF," *J. Doc.*, vol. 60, no. 5, hal. 503–520, 2004.

[12] J. Aldrich, "R. A. Fisher on Bayes and Bayes' Theorem. Bayesian Analysis," *Bayesian Anal.*, vol. 3, no. 1, hal. 161–170, 2008.

[13] R. Zacharski, *A Programmer's Guide to Data Mining: The Ancient Art of the Numerati*, 1 ed. Beijing: The People's Posts and Telecommunications Press, 2015.

[14] Destuardi dan S. Sumpeno, "Klasifikasi Emosi Untuk Teks Bahasa Indonesia," *Semin. Nas. Pascasarj. IX Inst. Teknol. Sepuluh Novemb.*, 2009.

[15] A. Rahman, "Online News Classification Using Multinomial Naive Bayes," *Itsmart*, vol. 6, no. 1, hal. 32–38, 2017.

[16] E. N. Efthimiadis, *Query Expansion*, 4 ed., vol. 31. Medford: Annual Review of Information Science and Technology (ARIST), 1996.