

Analisis Sentimen Data Twitter menggunakan Naive Bayes dengan Negation Handling

Aswinda Prima Putra(G64144007)*, Julio Adisantoso

Abstrak/Abstract

Analisis Sentimen dalam penelitian ini akan menggunakan data twitter dengan term yang berkaitan dengan kementerian dan pendidikan. Klasifikasi yang akan digunakan dalam penelitian adalah *multinomial naive bayes*. Masalah utama yang dihadapi dalam klasifikasi sentimen adalah bagaimana cara menangani negasi. Dalam seleksi fitur pada penelitian ini akan ditambahkan metode *Negation Handling*. *Negation Handling* akan diimplementasikan dengan data Twitter berbahasa Indonesia.

Kata Kunci

analisis sentimen; bahasa indonesia; kementerian; negation handling; opini; pendidikan; twitter;

*Alamat Email: aswinda.pp@gmail.com

PENDAHULUAN

Latar Belakang

Microblogging saat ini telah menjadi alat komunikasi yang sangat populer di kalangan pengguna internet. Jumlah penggunaannya membagikan opini mereka tentang berbagai macam aspek dari kehidupan mereka ataupun membahas isu-isu saat ini. Oleh karena itu *microblogging* merupakan sumber data yang sangat kaya untuk melakukan *opinion mining* dan analisis sentimen Pak dan Paroubek (2010). Berbagai macam perusahaan sering menggunakan survey online atau dengan kertas untuk mengumpulkan opini dari penggunaannya. Namun dengan kemunculan sosial media, orang cenderung lebih memilih mengutarakan pendapatnya melalui facebook, twitter atau sosial media yang lainnya (Anwar Hridoy *et al.* 2015).

Salah satu situs *microblogging* yang populer adalah Twitter. Twitter merupakan situs web yang dioperasikan oleh Twitter, Inc yang memberikan jaringan sosial berupa *microblog* serta memiliki karakteristik dan format yang unik dengan simbol ataupun aturan khusus. Disebut *microblog* karena penggunaannya hanya dapat mengirim dan membaca pesan *blog* seperti pada umumnya dengan batas maksimal sejumlah 140 karakter, pesan tersebut dikenal dengan tweet (Zhang *et al.* 2011). Twitter digunakan dalam penelitian ini karena sifat dari semua tweetnya bersifat publik. Berbeda dengan sosial media lain seperti facebook yang karakteristik dari *post* nya dapat dirubah menjadi privat atau hanya di lingkaran teman nya saja.

Kementerian Komunikasi dan Informatika Indonesia menyampaikan bahwa Indonesia menempati peringkat ke lima terbesar di dunia dari jumlah penggunaannya, dan berdasarkan data dari PT Bakrie yang disampaikan oleh kominfo, Indonesia memiliki 19,5 juta pengguna. Data ini menjadi dasar untuk dilakukannya penelitian pada Twitter berbahasa Indonesia. Pada penelitian Institute for Development of Economics and Finance (Indef) pada tahun 2015, berhasil menjangkau 12 juta tweet terkait pemerintahan dan 150 ribu diantaranya memiliki tema pembangunan (Tempo 2015). Banyaknya jumlah tweet terkait pemerintahan khususnya dibidang kementerian dan pendidikan inilah yang mendorong dilakukannya penelitian ini dengan menyertakan kata tersebut tersebut sebagai kata kunci dalam pengumpulan data.

Sentimen analisis atau *opinion mining* adalah studi komputasional dari opini-opini orang, sentimen dan emosi melalui entitas dan atribut yang dimiliki yang diekspresikan dalam bentuk teks Liu (2012). Analisis sentimen akan mengelompokkan polaritas dari teks yang ada dalam kalimat atau dokumen untuk mengetahui pendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif, negatif atau netral Pang *et al.* 2002 (2002). Salah satu teknik pembelajaran mesin untuk analisis sentimen adalah *Naive Bayes classifier* (NBC). NBC merupakan teknik pembelajaran mesin yang berbasis probabilitas. NBC adalah metode sederhana tetapi memiliki akurasi serta performansi yang tinggi dalam pengklasifikasian teks (Routray *et al.* 2013). Xhemali *et al.* (Xhemali *et al.*) melakukan perbandin-

gan antara tiga metode. Metode-metode tersebut adalah *Naïve Bayes*, *Pohon Keputusan*, dan *Neural Networks*. Hasil penelitian secara keseluruhan menunjukkan bahwa *Naïve Bayes classifier* adalah pilihan terbaik untuk pelatihan domain. Penggunaan model *Naïve Bayes Classifier* ini dikarenakan proses yang sederhana dan mudah diaplikasikan pada berbagai keadaan sehingga tidak akan mengalami kegagalan secara keseluruhan pada hasilnya (Manning *et al.* 2008).

Dalam penelitian Narayanan *et al.* (Narayanan *et al.*) Negation Handling Diimplementasikan dalam klasifikasi sentiment menggunakan naive bayes dan hasilnya menunjukkan peningkatan akurasi sebesar 1%. Data yang digunakan dalam penelitian tersebut adalah data review film. Dalam penelitian ini akan berfokus pada metode Negation Handling pada data twitter berbahasa Indonesia.

Rumusan Masalah

Rumusan Masalah dari penelitian adalah:

1. Bagaimana mengimplementasikan Metode *Negation Handling* pada analisis sentimen Twitter berbahasa Indonesia ?
2. Apakah metode *Negation Handling* dapat meningkatkan akurasi analisis sentimen dibandingkan dengan tanpa menggunakan *Negation Handling* ?

Tujuan

Tujuan penelitian adalah:

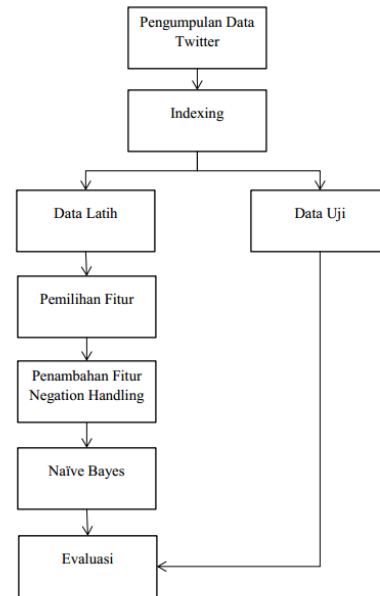
1. Mengimplementasikan analisis sentimen dengan menggunakan metode *Negation Handling* pada data Twitter berbahasa Indonesia.
2. Membandingkan akurasi dari analisis sentimen dengan *Negation Handling* dan tanpa menggunakan *Negation Handling*.

Manfaat

Hasil penelitian diharapkan dapat membantu kementerian pendidikan dalam menangani isu-isu negatif terkait dengan pendidikan dengan memanfaatkan sentimen analisis untuk mencari isu negatif tentang kementerian dan pendidikan.

METODE PENELITIAN

Penelitian yang dilakukan terbagi menjadi beberapa tahapan proses. Gambar 1 menunjukkan tahapan proses tersebut.



Gambar 1. Tahapan proses penelitian

Pengumpulan Data

Penelitian ini menggunakan data twitter bahasa Indonesia. Data yang akan diambil dari twitter adalah data dengan term “Kementrian”, “Menteri”, “Pendidikan”, “Sekolah” dan “Indonesia”. Data dengan term tersebut diambil karena di ranah kementerian sangat membutuhkan opini-opini masyarakat untuk meningkatkan kualitas dari kementerian terkhusus pada ranah Pendidikan. Pada tahap akuisisi data tweet, data diperoleh dari *tags.hawksey.info*. Data yang didapatkan berupa data excel dengan atribut seperti pada Tabel 1.

Tabel 1. Struktur Data Response Twitter

Atribut	Keterangan
id_str	id dari <i>post</i> twitter
from_user	<i>username</i> pemakai twitter
text	<i>post</i> twitter
created_at	tanggal dan waktu <i>post</i> dibuat
geo_coordinates	koordinat tempat <i>user</i>
source	tautan profil <i>user</i>
profile_image_url	gambar profil dari <i>user</i>
user_followers_count	jumlah <i>follower</i> <i>user</i>
user_friends_count	jumlah teman <i>user</i>
user_location	lokasi dari <i>user</i>
status_url	link dari <i>post</i> twitter

Dari struktur data yang didapatkan dari sistem tersebut, akan diambil atribut “text” sebagai data untuk dio-

lah sentimennya. Data yang diperoleh dari sistem masih berupa data mentah *post user* yang belum ada sentimennya. Untuk memberikan sentimen pada data tersebut yang akan digunakan sebagai data latih dilakukan dengan memberikan opini secara manual dengan bantuan manusia. Penelitian ini menggunakan data sebanyak 6000 data.

Indexing

Indexing merupakan proses persiapan yang dilakukan terhadap dokumen sehingga dokumen siap untuk diproses. Proses *indexing* dibagi menjadi dua proses, yaitu *document indexing* dan *term indexing*. Dari *term indexing* akan dihasilkan koleksi kata yang akan digunakan untuk meningkatkan performansi pencarian pada tahap selanjutnya. Selain itu, teknik *indexing* ini juga dilakukan agar hasil yang diperoleh lebih baik. Karena kebanyakan tweet hanya berisi tautan dan tidak menunjukkan sentimen tertentu, dan penulisannya ditulis dalam bahasa asing yang bukan bahasa Inggris (Parikh dan Movasate, 2014). Tahap-tahap yang dilakukan didalam proses *indexing* meliputi *tokenizing*, penghapusan *stopwords*, normalisasi kata, *stemming*, pembuatan *document term matrix*.

Tokenizing

Tokenizing adalah pengambilan kata-kata (*term*) dari kumpulan dokumen menjadi kumpulan *term* dan juga membuang beberapa karakter seperti tanda baca. Contoh dari tokenisasi adalah seperti pada Tabel 2.

Tabel 2. *Tokenizing*

Input	Data Twitter
Output	Data Twitter

Proses memotong dokumen atau kata menjadi bagian-bagian yang lebih kecil disebut *token*. *token* bisa berupa paragraf, kalimat, frasa kata tunggal sederhana dan konsep. teknik yang digunakan dalam proses tokenisasi adalah segmentasi dan memilah. Dalam tahap ini dokumen atau data *post twitter* diubah menjadi kumpulan *term* dengan cara menghilangkan *mention*, URL, tanda baca dan angka pada tweet. Semua huruf pada tweet diubah menjadi huruf kecil. Pada penelitian ini tokenisasi dilakukan dengan menggunakan kode dari Nette yang terdapat di <https://github.com/nette/tokenizer>.

Penghapusan Stopwords

Stopwords adalah sebuah kata-kata dalam bahasa tertentu yang sangat umum dan memiliki nilai informasi

nol (Feinerer *et al.* 2008). Stopwords didefinisikan sebagai term yang tidak berhubungan (*irrelevant*) dengan dokumen meskipun kata tersebut sering muncul di dalam dokumen. Contoh *stopwords* dalam bahasa Indonesia : yang, juga, dari, dia, kami, kamu, aku, saya, ini, itu, atau, dll. Penghapusan *stopwords* dilakukan untuk menghilangkan kata dalam daftar kata buang (*stopwords*). Kata tersebut merupakan kata yang jika dihapus tidak mengubah makna dari tweet. Daftar *stopwords* didapatkan dari penelitian Tala (2003) sebanyak 759 kata.

Normalisasi Kata

Menurut Aziz (2013) tahap normalisasi kata dilakukan dengan penggantian kata yang tidak baku menjadi baku, karena kata yang sudah baku akan cenderung lebih kecil ambiguitas dalam pelafalannya dibanding dengan kata yang tidak baku. Misalnya, kata dengan dapat ditulis dengan "dg" dan "dgn". Untuk itu perlu dilakukan normalisasi kata dengan cara mengganti kata yang tidak baku dengan kata yang sesuai konteksnya (Sproat *et al.* 2001). Sebelumnya sudah dibuat terlebih dahulu sebuah kamus yang tidak baku dengan kata bakunya, agar memudahkan dalam fungsi penggantian dan kemudian menggantinya dengan kata baku yang telah ada di dalam kamus tersebut. Dataset kata tidak baku dan kata baku yang digunakan dari penelitian Aziz (2013) sebanyak 3719 baris data.

Stemming

Stemming adalah proses konversi *term* ke bentuk umumnya. Dokumen dapat pula diekspansi dengan mencari sinonim bagi *term* tertentu di dalamnya. Sinonim adalah kata-kata yang mempunyai pengertian serupa tetapi berbeda dari sudut pandang morfologis. Seperti *stemming*, operasi ini bertujuan menemukan suatu kelompok kata terkait. *Stemming* merupakan salah satu cara yang digunakan untuk meningkatkan performa IR dengan cara mentransformasi kata-kata dalam sebuah dokumen teks ke kata dasarnya (Agusta 2009). Tahap *stemming* bertujuan untuk mengurangi jumlah kata dan mendapatkan kata dasar yang benar-benar sesuai. Tahap ini menggunakan algoritme Adriani *et al.* (2007) untuk menghapus berbagai variasi *prefix* (awalan) dan *suffix* (akhiran). Kamus kata dasar sebanyak 28.526 kata.

Pembuatan Document Term Matrix (DTM)

Menurut Nadilah (2016) tahap pembuatan *term document matrix* (TDM) dilakukan untuk membuat matriks jumlah kemunculan suatu kata pada dokumen. *Document Term Matrix* (DTM) merupakan cara yang paling

umum digunakan untuk merepresentasikan text. DTM dapat diekspor dari korpus dan digunakan sebagai mekanisme *bag-of-words*. Pendekatan ini menghasilkan matrik dengan id dokumen sebagai baris dan *term* sebagai kolom. Elemen matrik merupakan frekuensi. Sebagai contoh ada dua dokumen dengan id 1 dan 2 mempunyai kata yang sama yaitu “Nama saya budi dan nama ayah saya budi” dan “nama teman saya budi”. Term document matrix yang terbentuk adalah seperti pada Tabel 3.

Tabel 3. Document Term Matrix

ID	nama	saya	budi	dan	ayah	teman
1	2	2	2	1	1	0
2	1	1	1	0	0	1

Pada penelitian ini kolom matriks menunjukkan kata yang ada pada data *tweet*, sedangkan baris matriks menunjukkan indeks dari dokumen pada kumpulan korpus. Pada penelitian ini satu *tweet* menandakan satu dokumen.

Pembagian Data

Data yang dihasilkan setelah proses *indexing* dibagi menjadi dua subset data yaitu data latih dan data uji dengan perbandingan 70:30. Sebanyak 70 persen data latih dan 30 persen data uji. Data latih ini akan digunakan untuk tahapan selanjutnya sementara data uji digunakan untuk melakukan pengujian terhadap sistem klasifikasi yang telah dibuat dalam penelitian ini.

Pemilihan Fitur

Seleksi fitur merupakan proses pemilihan subset dari *term* pada data latih. Fitur yang terpilih pada seleksi fitur ini akan digunakan dalam klasifikasi teks. Tujuan dari seleksi fitur adalah membuat data latih yang digunakan *clasifier* lebih efisien dengan cara mengurangi ukuran kosakata yang efektif dan meningkatkan akurasi klasifikasi dengan menghilangkan fitur *noise* (Manning *et al.* 2008).

Seleksi fitur secara umum dibagi menjadi *unsupervised feature selection* dan *supervised feature selection* (Garnes 2009). *Unsupervised feature selection* adalah sebuah metode seleksi fitur yang tidak menggunakan informasi kelas dalam data pelatihan ketika memilih fitur untuk *classifier*. Salah satu contoh seleksi fitur yang tidak menggunakan informasi kelas dalam pemilihan fiturnya adalah IDF. Metode seleksi fitur selanjutnya adalah *supervised feature selection* yaitu metode yang

menggunakan informasi kelas dalam data latihnya. Untuk menggunakan seleksi fitur ini harus tersedia *pre-classified*. contoh dari *supervised feature selection* adalah MI dan *chi-square*.

Inverse document frequency (IDF)

DF merupakan banyaknya dokumen yang mengandung *term*. Ukuran nilai kepentingan suatu *term* dari dokumen yang digunakan sebagai penciri adalah nilai DF yang besar, namun nilai dari DF memiliki rentang nilai yang lebar. *Inverse document frequency* (IDF) adalah inverse dari nilai DF, sehingga ukuran kepentingan suatu *term* dari dokumen yang akan digunakan penciri yang memiliki nilai kecil dengan rentang yang tidak begitu jauh. Menurut Witten (1999) kata yang jarang atau paling sedikit muncul justru harus diperhatikan sebagai kata yang lebih penting dari pada kata yang paling sering muncul dalam dokumen. Nilai dari IDF disimbolkan dengan idf_t yang ditulis dengan formula [1]

$$idf_t = \log\left(\frac{N}{df_t}\right) \quad (1)$$

sedangkan N adalah banyaknya dokumen dan df_t adalah banyaknya dokumen didalam koleksi yang mengandung *term* tertentu.

Mutual information (MI)

MI menunjukkan seberapa besar informasi ada atau tidaknya kontribusi *term* untuk membuat klasifikasi yang tepat. Nilai dari MI disimbolkan dengan notasi I, dimana [2]

$$I(U;C) = \sum_{et \in \{1,0\}} \sum_{ec \in \{1,0\}} P(U = et, C = ec) \log_2 \frac{P(U = et, C = ec)}{P(U = et)P(C = ec)} \quad (2)$$

sedangkan U adalah variabel acak dengan nilai-nilai $et = 1$ (dokumen berisi *term* t) dan $et = 0$ (dokumen tidak mengandung t), dan C adalah variabel acak dengan nilai-nilai $ec = 1$ (dokumen di kelas c) dan $ec = 0$ (dokumen tidak di kelas c). Nilai dari I juga bisa dijabarkan menjadi [3]

$$\frac{N_{11}}{N} \log_2 \frac{NN_{11}}{N_1 N_1} + \frac{N_{01}}{N} \log_2 \frac{NN_{01}}{N_0 N_1} + \frac{N_{10}}{N} \log_2 \frac{NN_{11}}{N_1 N_0} + \frac{N_{00}}{N} \log_2 \frac{NN_{00}}{N_0 N_0} \quad (3)$$

dengan N adalah jumlah dokumen yang memiliki nilai-nilai et dan ec yang ditunjukkan oleh dua subscript.

Sebagai contoh, N_{10} adalah jumlah dokumen yang mengandung term t ($e_t = 1$) dan tidak dalam c ($e_c = 0$). $N_{11} = N_{10} + N_{11}$ adalah jumlah dokumen yang berisi term t ($e_t = 1$) dan untuk menghitung dokumen independen keanggotaan kelas ($e_c = 0, 1$). N adalah jumlah total dokumen atau $N = N_{00} + N_{01} + N_{10} + N_{11}$.

Chi-Square (X^2)

x^2 biasanya digunakan digunakan dalam menguji independensi dari dua variabel yang berbeda. Hipotesis nol jika kedua variabel saling bebas satu sama lain jika nilai dari x^2 tinggi maka hubungan kedua variabel tersebut semakin erat. Dalam seleksi fitur x^2 digunakan untuk mengukur independensi term t dan kelas c . hipotesis yang diuji adalah term dan kelas benar-benar independen, artinya fitur ini tidak berguna untuk mengelompokkan dokumen. Semakin tinggi nilai maka semakin rendah nilai independensinya. Persamaan dari x^2 dapat ditulis sebagai [4]

$$X^2(D, t, c) = \sum_{et \in \{1,0\}} \sum_{ec \in \{1,0\}} \frac{(N_{etec} - E_{etec})^2}{E_{etec}} \quad (4)$$

Sedangkan D adalah variabel acak dengan nilai-nilai $e_t = 1$ adalah dokumen berisi term t dan $e_t = 0$ adalah dokumen yang tidak mengandung t , $e_c = 1$ adalah dokumen di kelas c dan $e_c = 0$ adalah dokumen tidak di kelas c . N adalah frekuensi yang diamati dalam dokumen D dan E adalah frekuensi yang diharapkan. Pengambilan keputusan dilakukan berdasarkan nilai dari masing-masing kata. Kata yang memiliki nilai X^2 di atas nilai kritis pada taraf nyata adalah kata yang akan dipilih sebagai penciri dokumen. Kata yang dipilih sebagai penciri merupakan kata yang memiliki pengaruh terhadap kelas c . Nilai kritis X^2 untuk taraf nyata α yang digunakan dalam penelitian ini ditunjukkan pada Tabel 4.

Tabel 4. Nilai kritis untuk taraf nyata α

α	Nilai Kritis
0.050	3.840
0.010	6.630
0.005	7.880

Pada penelitian ini akan menggunakan IDF sebagai seleksi fiturnya. IDF dipilih karena metode ini efisien, mudah dan memiliki hasil yang akurat Robertson (2005).

Negation Handling

Negation Handling adalah salah satu faktor yang berkontribusi meningkatkan akurasi klasifikasi secara signifikan. Masalah utama yang dihadapi dalam klasifikasi sentimen adalah bagaimana cara menangani negasi. Karena setiap kata digunakan sebagai fitur kata "baik" dalam frase "tidak baik" akan mendapatkan sentimen positif daripada negatif karena kata "tidak" sebelumnya tidak dimasukkan dalam perhitungan.

Dalam penelitian Narayanan *et al.* (2013) *Negation Handling* diimplementasikan dalam klasifikasi sentimen menggunakan *naive bayes* dan hasilnya menunjukkan peningkatan akurasi sebesar 1%. Data yang digunakan dalam penelitian tersebut adalah data review film. Dalam penelitian tersebut kata "good" akan meningkatkan jumlah kelas positif dan kata "not good" yang sudah dilakukan negation handling masuk kedalam kelas negatif. Dalam penelitian ini, kata "baik" yang sebelumnya ada "tidak" akan diproses dulu dengan *Negation Handling* sehingga menjadi "tidak baik", jadi berbeda dengan penelitian yang dilakukan Narayanan *et al.* (2013), kata "tidak baik" akan diperlakukan sebagai satu term dan akan masuk ke dalam kelas negatif. Untuk menangani masalah ini dibuatlah algoritma untuk *Negation Handling* menggunakan representasi bentuk negasi dari Das & Chen (2001). Algoritma yang digunakan akan mengubah kata yang diikuti kata negatif menjadi bentuk "tidak_" + kata. Pseudocode pada [1] yang digunakan dalam penelitian ini.

```

for each word in document do
    if word before is negative then
        transform word to
        "negative-word-before_" + word
    end
    if a punctuation mark is encountered then
        continue to next word
    end
end

```

Algoritme 1: Pseudocode Negation Handling

Klasifikasi

Penelitian ini menggunakan tiga kelas sentimen yaitu, positif, negatif, dan netral. Data selanjutnya akan dikategorikan pada tahapan ini. Klasifikasi memiliki peranan yang penting untuk menganalisis sentimen terhadap data latih. Klasifikasi pada analisis sentimen bertujuan untuk mengkategorikan setiap data terhadap data pencirinya.

Pengolongan atau mengelompokan suatu dokumen ke dalam suatu kategori tertentu merupakan tujuan dari klasifikasi (Manning *et al.* 2008). Fungsi klasifikasi secara umum yaitu

$$\gamma: X \rightarrow C \quad (5)$$

merupakan fungsi untuk memetakan suatu dokumen ke dalam kategori tertentu, dengan X adalah kumpulan dokumen dan C merupakan kategori. Metode klasifikasi terbagi menjadi dua, yaitu metode klasifikasi berbasis vektor dan peluang. Metode klasifikasi yang digunakan pada penelitian ini adalah metode klasifikasi berbasis peluang yaitu klasifikasi *Multinomial Naïve Bayes*. Model klasifikasi ini digunakan karena proses yang sederhana dan pengaplikasiannya yang mudah pada berbagai keadaan (Manning *et al.* 2008)

Multivariate Bernoulli menyatakan bahwa dokumen diwakili oleh atribut biner yang menunjukkan bahwa ada dan tidak ada *term* (kata-kata) dalam dokumen. Frekuensi kemunculan *term* (kata-kata) dalam dokumen tidak ikut diperhitungkan. Sedangkan *Multinomial Naïve Bayes*, dokumen diwakili oleh kemunculan *term* dari dokumen. Pada model ini, sebelumnya dibuat asumsi jika kemunculan masing-masing *term* t bersifat independen antara satu *term* dengan yang lainnya. Dengan menggunakan nilai dari $P(c|d)$ peluang suatu dokumen d di dalam kelas c dapat ditulis sebagai (Manning *et al.* 2008)

$$P(C|d) \propto P(c) \prod_{1 \leq k \leq nd} P(t_k|c) \quad (6)$$

dengan $P(t_k|c)$ adalah peluang dari suatu term t_k muncul pada dokumen d yang diketahui memiliki kelas c . Pendugaan parameter $P(t_k|c)$ dihitung dengan cara

$$P(t_k|C) = \frac{T_{ct}}{\sum_{t' \in V} T_{ct'}} \quad (7)$$

dengan T_{ct} adalah jumlah kemunculan *term* t dalam dokumen *training* yang berada di kelas c , adalah jumlah seluruh *term* yang muncul berulang kali pada dokumen yang sama (Manning *et al.* 2008).

Term tidak selalu muncul pada salah satu kelas saat dilakukan

klasifikasi sehingga nilai $P(t_k|c)$ yang dihasilkan adalah nol. Untuk mengatasi permasalahan tersebut, digunakan *laplace smoothing*, yaitu menambahkan frekuensi term sebanyak 1 sehingga perhitungan dari $P(t_k|c)$ menjadi

$$P(t_k|C) = \frac{T_{ct}}{\sum_{t' \in V} T_{ct'} + B} \quad (8)$$

Pada penelitian ini *Multinomial Naïve Bayes* akan digunakan sebagai metode klasifikasinya.

Evaluasi

Tahap evaluasi dilakukan untuk mengetahui tingkat akurasi dari hasil penggunaan *indexing*, pemilihan fitur, serta klasifikasi pada data uji. Pengujian penelitian ini dilakukan pada data uji terhadap fungsi klasifikasi yang sudah di *training*. *Token* dari hasil seleksi fitur, akan dihitung peluangnya berdasarkan kelas-kelasnya dari dokumen Twitter. Setelah itu, membandingkannya dengan kelas aktual dari data uji dan kelas hasil prediksi dengan menggunakan *confusion matrix*. Isi dari *confusion matrix* adalah jumlah kasus-kasus yang telah diklasifikasikan dengan benar dan kasus-kasus yang salah diklasifikasikan.

Evaluasi digunakan untuk mengukur efektivitas dari sistem IR. Dua pengukuran yang sering digunakan dalam IR adalah *precision* dan *recall*.

Precision

Precision merupakan teknik untuk evaluasi yang didefinisikan sebagai presentase dokumen yang di-*retrieve* yang benar-benar relevan. Formula *Precision* didapatkan dengan

$$\begin{aligned} Precision &= \frac{\#(\text{relevant items retrieved})}{\#(\text{retrieved items})} \\ &= P(\text{relevant}|\text{retrieved}) \end{aligned} \quad (9)$$

dengan jumlah *item* relevan yang didapatkan dibagi dengan jumlah *item* yang didapatkan.

Recall

Recall adalah teknik evaluasi untuk menemukan semua *item* yang relevan dari dalam koleksi dokumen dan didefinisikan sebagai presentase dokumen yang relevan. [10]

$$\begin{aligned} Recall &= \frac{\#(\text{relevant items retrieved})}{\#(\text{relevant items})} \\ &= P(\text{retrieved}|\text{relevant}) \end{aligned} \quad (10)$$

Konsep tersebut dapat diperjelas pada tabel 5.

Tabel 5. Tabel Kontigensi

aasd	Relevant	Not Relevant
Retrieved	True positive (tp)	False positives (fp)
Not Retrieved	False negatives (fn)	True negatives (tn)

Maka didapatkan rumus seperti berikut [11]

$$\begin{aligned} P &= tp/(tp + fp) \\ R &= tp/(tp + fn) \end{aligned} \quad (11)$$

Accuracy

Selain Precision dan Recall, keakuratan sentiment analisis juga dinilai dari akurasi, untuk menghitung akurasi digunakan rumus seperti berikut dan mengacu tabel 5 [12]

$$accuracy = (tp + tn) / (tp + fp + fn + tn) \quad (12)$$

F-measure

Pengukuran yang lain adalah F-measure yang merupakan *weighted harmonic mean* dari *precision* dan *recall*. [13]

$$F = \frac{1}{\alpha \frac{1}{p} + (1 - \alpha) \frac{1}{r}} = \frac{(\beta^2 + 1)PR}{\beta^2 P + R} \quad (13)$$

where $\beta^2 = \frac{1 - \alpha}{\alpha}$

Dimana $\alpha \in [0, 1]$ dan $\beta^2 \in [0, \Psi]$

Balanced F-measure menyamakan bobot dari precision dan recall, yang berarti membuat $\alpha = 1/2$ atau $\beta = 1$. Ketika menggunakan $\beta = 1$, formula dapat disederhanakan sebagai berikut [14]

$$F_\beta = \frac{2PR}{P + R} \quad (14)$$

DAFTAR PUSTAKA

- Adityawan, E. 2014. "Analisis Sentimen Dengan Klasifikasi Naïve Bayes Pada Pesan Twitter Menggunakan Data Seimbang". Skripsi. Departemen Ilmu Komputer, Institut Pertanian Bogor.
- Adriani, Mirna *et al.* 2007. "Stemming Indonesian: A Confix-stripping Approach". 6 (4), pp. 1–33. ISSN: 1530-0226. DOI: 10.1145/1316457.1316459.
- Agusta, Ledy. 2009. "Perbandingan Algoritma Stemming Porter dengan algoritma Nazief & Adriani untuk Stemming Dokumen Teks Bahasa Indonesia". Universitas Kristen Satya Wacana.
- Alkhatib, K, Najadat H, Hmeidi I, dan Shatnawi MKA. 2013. "Stock Price Prediction Using k-Nearest Neighbor (kNN) Algorithm".
- Anwar Hridoy, Syed Akib *et al.* 2015. "Localized twitter opinion mining using sentiment analysis" dalam: *Decision Analytics* 2 (1), pp. 1–19. ISSN: 2193-8636. DOI: 10.1186/s40165-015-0016-4.
- Aziz, ATA. 2013. "Sistem pengklasifikasian entitas pada pesan twitter menggunakan ekspresi reguler dan naïve Bayes". Skripsi. Departemen Ilmu Komputer, Institut Pertanian Bogor.
- Das, Sanjiv R. *et al.* 2001. "Yahoo! for amazon: Sentiment extraction from small talk on the web". *8th Asia Pacific Finance Association Annual Conference*.
- Dimastyo, JG. 2014. "Pengukuran Kinerja Spam Filter dengan Seleksi Fitur yang berbeda menggunakan Fungsi Klasifikasi Multinomial Naïve Bayes". Skripsi. Departemen Ilmu Komputer, Institut Pertanian Bogor.
- Feinerer, Ingo, Kurt Hornik, dan David Meyer. 2008. "Text Mining Infrastructure in R" dalam: *Journal of Statistical Software* 25 (5), pp. 1–54.
- Garnes. 2009. "Feature Selection for Text Categorisation". Norwegia (NO): NTNU.
- Liu, Bing. 2010. "Sentiment Analysis and Subjectivity, in Handbook of Natural Language Processing". Chicago (US): University of Illinois.
- 2012. *Sentiment Analysis and Opinion Mining*. Morgan dan Claypool Publishers. [Internet]. [Diunduh tanggal 11/08/2016]. Dapat diunduh dari: <https://www.cs.uic.edu/~liub/FBS/SentimentAnalysisand-OpinionMining.pdf>.
- Manning, Christopher D., Prabhakar Raghavan, dan Hinrich Schütze. 2008. *An Introduction to Information Retrieval*. Cambridge University Press Cambridge, England.
- Mudinas, Andrius, Dell Zhang, dan Mark Levene. 2012. "Combining Lexicon and Learning Based Approaches for Concept-level Sentiment Analysis". *Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining*. WISDOM '12. Beijing, China: ACM, 5:1–5:8. ISBN: 978-1-4503-1543-2. DOI: 10.1145/2346676.2346681.
- Nadilah. 2016. "Asosiasi dan Geovisualisasi Antara Data Tweet Terkait Kebakaran Hutan dengan Data Cuaca di Provinsi Riau dan Kepulauan Riau". Institut Pertanian Bogor.
- Narayanan, V, I Arora, dan A Bhatia. 2013. "Fast and accurate sentiment classification using an enhanced Naive Bayes model". Department of Electronics Engineering, Indian Institute of Technology (BHU), Varanasi, India.
- Pak, Alexander dan Patrick Paroubek. 2010. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining". *Proceedings of the Seventh International Conference on Language Resources and Evaluation*

- (LREC'10). Ed. by Nicoletta Calzolari (Conference Chair) *et al.* Valletta, Malta: European Language Resources Association (ELRA). ISBN: 2-9517408-6-7.
- Pang, Bo, Lillian Lee, dan Shivakumar Vaithyanathan. 2002. "Thumbs Up? Sentiment Classification Using Machine Learning Techniques". *Proceedings of EMNLP*, pp. 79–86.
- Robertson, Stephen. 2004. "Understanding inverse document frequency: On theoretical arguments for IDF" dalam: *Journal of Documentation* 60, p. 2004.
- Routray, Preeti, Chinmaya Kumar Swain, dan Smita Praya Mishra. 2013. "Article: A Survey on Sentiment Analysis" dalam: *International Journal of Computer Applications* 76 (10). Full text available, pp. 1–8.
- Seiberg, Nathan dan Edward Witten. 1999. "String theory and noncommutative geometry" dalam: *Journal of High Energy Physics* 1999 (09), p. 032.
- Sproat, Richard *et al.* 2001. "Normalization of non-standard words." dalam: *Computer Speech & Language* 15 (3), pp. 287–333.
- Tala, Fadillah Z. 2003. "A study of stemming effects on information retrieval in Bahasa Indonesia" dalam: *Institute for Logic, Language and Computation Universiteit Van Amsterdam*.
- Xhemali, Daniela, Christopher J. HINDE, dan Roger G. STONE. 2009. *Naive Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages*.
- Zhang *et al.* 2011. "Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis".