

SENTIMENT ANALYSIS TOKOH POLITIK PADA TWITTER

Agung Pramono¹, Rini Indriati², Arie Nugroho³,

^{1,2,3}Sistem Informasi, Fakultas Teknik, Universitas Nusantara PGRI Kediri

E-mail: ¹pramonoagung0741@gmail.com, ²rini.indriati@unpkediri.ac.id,

³arienugroho@unpkediri.ac.id

Abstrak – *Sentiment analysis* atau *opinion mining* digunakan menganalisis opini publik kepada tokoh politik berdasarkan data yang didapatkan dari twitter. Pengguna twitter melakukan *posting* atau *tweet* tentang pendapat mereka kepada tokoh politik digunakan sebagai sumber data untuk menilai sentimen kepada tokoh politik. Metode *Naïve Bayes Classifier* digunakan untuk melakukan klasifikasi pada twitter untuk mengetahui *sentiment* atau *opinion* pada *tweet* dan dikelompokkan menjadi positif, negatif atau netral. Proses pengklasifikasian metode *Naïve Bayes Classifier* ada beberapa tahap. Tahap pertama Proses Pembelajaran *Naïve Bayes Classifier* tahap kedua Proses Klasifikasi *Naïve Bayes Classifier*. Hasil proses klasifikasi berdasarkan nilai *Vmap* diambil nilai terbesar yang menjadi kategori dari tweet kelas positif, negatif dan netral. Pada proses klasifikasi nilai akurasi tergantung banyaknya data *training*. Pada data *training* positif 100, negatif 100, netral 100 dan data *test* 100, nilai akurasi 61% dan error 39% . Pada data *training* positif 700, negatif 700 dan netral 700 dengan menggunakan data test 100 akurasi 91% dan error 9%. Rekomendasi penelitian selanjutnya menggunakan bahasa daerah atau bahasa asing, mengklasifikasi tentang produk atau hal lain, mengklasifikasi *tweet* yang terdapat singkatan.

Kata Kunci — *Sentiment Analysis, Opinion Mining, Naïve Bayes Classifier, Twitter*

ABSTRACT – Sentiment analysis or opinion mining is used to analyze public opinion to the political figure is based on data obtained from twitter. Twitter user to post or tweet about their opinion to political figures. Therefore, the data tweet can be used as a source of data to assess the sentiment to political figures. In the opinion mining methods are needed Naïve Bayes classifier to perform classification on twitter for sentiment or opinion on a tweet and grouped into positive, negative or neutral. The process of classification methods Naïve Bayes classifier there are several stages. The first stage Learning Process Naïve Bayes classifier second stage classification process Naïve Bayes classifier. The results of the classification process based on the value VMAP taken the largest value into the category of class tweet positive, negative and neutral. In the process of classification accuracy value depends on the amount of training data, the training data 100 positive, 100 negative, neutral 100 and 100 test data, accuracy value of 61% and 39% error. On the positive training data 700, 700 negative and neutral 700 using data 100 test accuracy of 91% and 9% error. Recommendations subsequent studies using the local language or a foreign language, classification of a product or something else, classifying tweets contained abbreviations.

Keywords: Sentiment Analysis, Opinion Mining, Naïve Bayes classifier, Twitter.

1. PENDAHULUAN

Text Mining merupakan subyek riset yang tergolong baru. *Text Mining* dapat memberikan solusi dari permasalahan seperti : pemrosesan, pengelompokkan

dan menganalisa *unstructured text* dalam jumlah besar. Proses *text mining* yang khas meliputi kategorisasi teks, *text clustering*, ekstraksi konsep atau entitas, produksi taksonomi granular, *sentiment analysis*, penyimpulan dokumen, dan pemodelan relasi entitas (pembelajaran hubungan antara entitas).

Dalam penelitian ini melakukan proses atau mengkategorikan komentar (*tweet*) berdasarkan kelas positif, negatif dan netral dalam *text mining* yang disebut *sentiment analysis*. *Sentiment analysis* adalah studi komputasi mengenai sikap, emosi, pendapat, penilaian, pandangan dari sekumpulan teks yang fokus pada mengekstraksi, mengidentifikasi atau menemukan karakteristik sentimen. Tugas dasar dari analisis sentimen itu sendiri adalah mengelompokkan polaritas dari teks yang ada dalam dokumen, kalimat dan lain sebagainya, apakah pendapat yang dikemukakan dalam dokumen, kalimat tersebut bersifat positif, negatif atau netral.

Untuk menentukan apa komentar (*tweet*) tersebut bersifat positif, negatif atau netral diperlukan proses klasifikasi. Metode yang digunakan untuk klasifikasi adalah metode *Naive Bayes* yaitu *Naive Bayes Classifier* (NBC). Metode ini merupakan salah satu metode yang digunakan pada *text mining* yang sederhana tetapi memiliki keakuratan yang tinggi dalam mengklasifikasi. *Naive Bayes Classifikasi* (NBC) merupakan metode yang cepat dalam proses klasifikasi. Metode *Naive Bayes Classifier* (NBC) merupakan salah satu metode yang digunakan untuk menyelesaikan masalah *sentiment analysis*.

2. METODE PENELITIAN

2.1. Pengumpulan Data Training

Data Latih atau *training data* yang digunakan dalam penelitian ini

diambil dari twitter yang berdasarkan hasil pencarian tentang pendapat para pengguna twitter tentang tokoh politik. Data Latih yang didapatkan dengan memanfaatkan API (*Application Programming Interface*) yang disediakan oleh *Twitter*. Data Latih yang diperoleh dari twitter kemudian disimpan kedalam *Database*. Skema dari proses pengambilan *Tweet* dapat dilihat pada gambar 2.1.



Gambar 2.1. Skema pengambilan *tweet*

Data yang diambil dari twitter untuk data *training* adalah sebanyak 2100 data *tweet*. Dimana data yang diambil adalah data *tweet* yang mengandung sentimen terhadap tokoh politik. Data *training* dikategorikan secara manual yang dilakukan oleh user dan memilih sentimen pada *tweet* tersebut dan menandai *tweet* tersebut menjadi 3 kategori sentimen yaitu *tweet* yang mengandung sentimen positif, negatif dan netral.

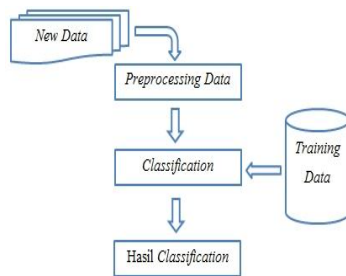
Tabel 1. Data *Training*

<i>Tweet Sentiment</i>	Positif	Negatif	Netral
<i>Tweet Tokoh Politik</i>	700	700	700

Data Latih yang sudah diperoleh disimpan dalam database yang nantinya digunakan untuk data *training*.

2.2. Desain Sistem

Desain sistem terdiri atas beberapa langkah seperti pada gambar 2. 2.



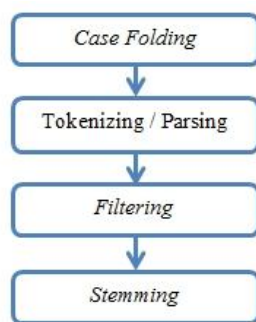
Gambar 2. 2 Diagram Alir Kerja

a. Pengambilan Data Baru

Pengambilan Data *tweet* baru pada proses ini dimana user mengambil data dari *twitter* dengan memanfaatkan API (*Application Programming Interface*) dan pengguna *twitter* yang melakukan *tweet* tentang tokoh politik. User memasukan *Keyword* tokoh politik mana yang akan diklasifikasi

b. Preprocessing Data

Preprocessing merupakan tahapan awal dalam mengolah data *Input* sebelum memasuki proses tahapan utama. *Preprocessing text* dilakukan untuk penyeragaman. *Preprocessing* terdiri dari beberapa tahapan: *case folding*, *tokenizing / parsing*, *filtering*, *Stemming*. Berikut adalah diagram alir *preprocessing*.



Gambar 3. *Preprocessing*

1) *Case Folding*

Case folding merupakan tahapan yang mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf 'a' sampai dengan 'z' yang diterima.

2) *Tokenizing / Parsing*

Tokenizing / parsing adalah tahap pemotongan string *input* berdasarkan tiap kata yang menyusunnya. Pada prinsipnya proses ini adalah memisahkan kalimat atau dokumen menjadi kata per kata.

3) *Filtering*

Filtering pada tahap ini akan dilakukan pembersihan *tweet* dari spesial karakter, *URL link*, *username*, serta *emoticon*. Contoh *stopword* adalah "yang", "dan", "di", "dari" dan lain – lain.

4) *Stemming*

Stemming merupakan suatu proses yang terdapat dalam sistem *Information Retrieval* yang mentransformasi kata-kata dalam suatu dokumen ke kata-kata akarnya (*root word*) menggunakan aturan-aturan tertentu. *Stemming* untuk proses bahasa Indonesia memiliki struktur imbuhan yang rumit atau kompleks sehingga lebih sulit untuk diolah. Pada proses *Stemming* menggunakan algoritma *porter*

c. Classification

Pada tahap klasifikasi adalah tahap penentuan keterhubungan antar kata-kata pada data. Tahap ini menggunakan sebuah algoritma *Naïve Bayes Classifier*. *Naïve Bayes Classifier* terdiri dari 2 proses dalam proses klasifikasi datanya. Kedua proses itu adalah proses pembelajaran *Naïve Bayes Classifier* dan proses klasifikasi *Naïve Bayes Classifier*.

1. Proses Pembelajaran *Naïve Bayes Classifier*

Perhitungan $P(V_j)$

Pada perhitungan pada setiap kelas yang ada. Rumusnya sebagai berikut:

$$P(V_j) = \frac{N_c}{N}$$

$P(V_j)$ Probalitas kelas yang dimiliki

N_c Jumlah dari kelas yang ada

N Jumlah dari data *training*

Perhitungan $P(x_i|V_j)$

Tahap ini melakukan perhitungan pada kata yang terdapat data test dan rumusnya adalah sebagai berikut

$$P(x_i|V_j) = \frac{nk + 1}{n + |Vocabulary|}$$

$P(x_i|V_j)$: Probabilitas x_i pada kelas V_j , x_i adalah kata dalam data test dan V_j kelas yang ada seperti positif, negatif dan netral
 nk : Jumlah kata yang terdapat pada setiap kelas
 n : Jumlah data *training* pada kata
Vocabulary : Jumlah semua kata dari semua kategori

2. Proses Klasifikasi *Naïve Bayes Classifier*

Secara umum proses ini menentukan kelas dari data test dengan menggunakan rumus sebagai berikut:

$$V_{MAP} = \underset{V_j \in V}{\operatorname{argmax}} P(V_j) \prod_{i=1}^n P(x_i|V_j)$$

d. Hasil Classification

Berdasarkan nilai V_{map} diambil nilai terbesar yang menjadi kelas dari tweet tersebut. Pada contoh ini didapatkan nilai V_{map} positif adalah nilai V_{map} yang terbesar dibandingkan dengan hasil klasifikasi yang lain maka contoh tweet tadi diklasifikasikan sebagai kelas positif.

3. HASIL DAN PEMBAHASAN

1. Proses Klasifikasi

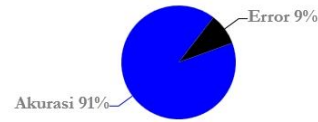
Berdasarkan metodologi penelitian, maka dilakukan implementasi sistem klasifikasi *sentiment analysis* pada twitter dengan metode *text mining* menggunakan *naïve bayes classifier*.

Proses klasifikasi adalah proses dimana menentukan sebuah kalimat (*tweet*) sebagai anggota kelas opini positif, negatif dan netral berdasarkan nilai perhitungan probabilitas Bayes yang lebih besar.

2. Pengujian Sistem

Pada Pengujian Sistem / *Testing* dilakukan pengujian terhadap sistem *sentiment analysis* yang berfungsi untuk mengetahui kinerja sistem dan untuk mengetahui berapa persen (%) *error testing* dalam melakukan proses klasifikasi *tweet* menggunakan 100 *tweet* untuk data *test*

Sentiment Analysis Testing



Gambar 4. Pengujian Sistem

```

Prior:
p(positif) = 628/1748 = 0.36
p(negatif) = 655/1748 = 0.35
p(netral) = 515/1748 = 0.29

Probabilitas:
p(positif|banyak) = (6+1)/(2031+770) = 0.0025
p(negatif|banyak) = (6+1)/(1941+770) = 0.0037
p(netral|banyak) = (6+1)/(1721+770) = 0.0026
p(positif|sukung) = (95+1)/(2031+770) = 0.0343
p(negatif|sukung) = (10+1)/(1941+770) = 0.0041
p(netral|sukung) = (9+1)/(1721+770) = 0.0040

Choosing Class:
V positif = 0.36 x 0.0025 = 0.0009000000000000000
V negatif = 0.35 x 0.0037 x 0.0041 = 0.000005309500000000000
V netral = 0.29 x 0.0026 x 0.0040 = 0.000004176000000000000

Positif
    
```

Gambar 5. Perhitungan *Naïve Bayes Classifier*

3. Hasil Pengujian

Pada pengujian ini dilakukan untuk mengetahui *error* pada proses klasifikasi, dengan menguji data *testing* pada beberapa kali proses klasifikasi dengan menggunakan data latih 100 sampai 1000 *tweet*, menggunakan 100 *tweet* untuk *testing*

Tabel 2. Hasil Pengujian

Data Latih			Data Testin g	Akurasi	Error
Positif	Negatif	Netral			
100	100	100	100	61%	39%
200	200	200	100	70%	30%
300	300	300	100	72%	28%
400	400	400	100	72%	28%
500	500	500	100	77%	23%
600	600	600	100	87%	13%
700	700	700	100	91%	9%

Dari tabel hasil pengujian dapat

diketahui bahwa metode *Naïve Bayes Classifier*, semakin banyak data latih yang digunakan akan semakin tinggi akurasi. Hasil pengujian dilihat pada tabel tersebut pada data latih positif 100, negatif 100, netral 100 dan data test 100 nilai akurasi 61% dan error 39%. Pada data latih positif 700,

negatif 700, netral 700 dan data test 100 akurasi 91% dan error 9%.

4. SIMPULAN

Kesimpulan dari pembahasan yang telah dikemukakan adalah sebagai berikut :

1. Proses pengklasifikasian metode *Naïve Bayes Classifier* ada beberapa tahap. Tahap pertama Proses Pembelajaran *Naïve Bayes Classifier* tahap kedua Proses Klasifikasi *Naïve Bayes Classifier*.
2. Dalam proses mengklasifikasikan *Sentiment Analysis* Berdasarkan nilai *Vmap* tersebut kemudian diambil nilai terbesar yang menjadi kategori dari tweet tersebut kelas positif, negatif dan netral.
3. Dalam proses klasifikasi akan semakin akurat tergantung banyaknya data latih dan saat menggunakan data latih positif 100, negatif 100 dan netral 100 dengan menggunakan data test 100 akurasi 61% dan error 39% dan saat menggunakan data latih positif 700, negatif 700 dan netral 700 dengan menggunakan data test 100 akurasi 91% dan error 9%.

5. SARAN

Saran pengembangan penelitian lebih lanjut sistem pengklasifikasian Tweet sebagai berikut:

1. Menggunakan bahasa daerah atau bahasa asing seperti bahasa Inggris dan bahasa asing lainnya.
2. Mengklasifikasi tentang produk atau tentang hal yang lain.
3. mengklasifikasi *tweet* yang terdapat kata singkatan.

DAFTAR PUSTAKA

- [1] Liu, Bing. (Ed.). 2012. *Sentiment Analysis and Opinion Mining*. Graeme Hirst, University of Toronto
- [2] Retnawiyati, Eka / Fatoni, M.M.,M.Kom., / Surya, Edi

Negara, M.Kom. *Analisis Sentimen Pada Data Twitter dengan Menggunakan Text Mining terhadap Suatu Produk*.

- [3] Nurhuda, Faishol / Widya, Sari Sihwi / Doewes, Afrizal. *Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naïve Bayes Classifier*.
 - [4] Fahrur, Imam Rozi / Hadi, Sholeh Pramono / Achmad, Erfan Dahlan. *Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi*
 - [5] Agusta, Ledy. 2009. *Perbandingan Algoritma Stemming Porter Dengan Algoritma Nazief & Adriani Untuk Stemming Dokumen Teks Bahasa Indonesia*.
 - [6] Kini, Mahesh M / Devi, Saroja H / G Desai, Prashant / Chiplunkar, Nirranjan. 2016. *Text Mining Approach to Classify Technical Research Documents using Naïve Bayes*
 - [7] Wayan, Ni Sumartini Saraswati. 2011. *Text Mining Dengan Metode Naïve Bayes Classifier Dan Support Vector Machines Untuk Sentiment Analysis*.
 - [8] Nomleni, Petrix / Hariadi, Mochamad / Ketut, Eddy, I Purnama *Sentiment Analysis Berbasis Big Data*.
 - [9] Feldman, Ronen / James, Israel Sanger. 2007. *The Text Mining Handbook Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press
- Mujilahwati, Siti. 2016. *Pre-Processing Text Mining Pada Data Twitter*.
- Raschka, Sebastian. 2014. *Naive Bayes and Text Classification I*

Seminar Nasional Inovasi Teknologi
UN PGRI Kediri, 22 Februari 2017

ISBN : 978-602-61393-0-6
e-ISSN : 2549-7952

Halaman ini sengaja dikosongkan