

Analisa sentimen dan Klasifikasi Komentar Positif Pada Twitter dengan Naïve Bayes Classification

Sentiment Analysis and Classification of Positive Comments on Twitter with Naive Bayes Classification

Akhmad Pandhu Wijaya¹⁾ Diki Wardhani²⁾

Program Studi Teknologi Informasi ¹⁾

Program Studi Teknologi Bisnis Digital ²⁾

Institut Teknologi dan Bisnis Bank Rakyat Indonesia ^{1,2)}

Jl RM Harsono No.2 Ragunan, Jakarta Selatan

pandhujobs@bri-institute.ac.id, diky.wardhani@bri-institute.ac.id

Naskah Diterima Tanggal 10 Oktober 2019 dan Naskah Disetujui tanggal 8 Desember 2019

Abstract

The existence of social media at this time is becoming a trend in the wider community, even being like the necessities of life that must be met. The last few years social media use has spread very widely without exception, used by adult consumers and even children. While the problem faced is the spread of writings that smelled negatively and had a negative impact on the psychic of its readers, especially children. In various moments and situations always appear related news but negative comments also appear that are very possible to herd the opinions of readers, it is necessary to overcome and identify early on negative comments that are possible to guide readers opinions. Therefore this study seeks to analyze the negative content that exists on Indonesian social media Twitter. This analysis is done by attempting to classify tweets that contain negative and positive comments using the method Naïve Bayes Classifier. While the Naïve Bayes Classifier method is combined with a feature that is calculated using TF-IDF. The tweet classification in this study consisted of negative and positive sentiment classes. The test results in this study were obtained by using RapidMiner tools which showed that the classification accuracy added by the TF-IDF feature produced level of accuracy 55.80%.

Keywords: classification, naïve bayes, sentiment analysis, term

Abstrak

Keberadaan sosial media pada dewasa ini menjadi trend dimasyarakat luas, bahkan berlaku seperti layaknya kebutuhan hidup yang harus dipenuhi. Beberapa tahun terakhir penggunaan sosial media merebak sangat luas tanpa terkecuali, dimanfaatkan oleh konsumen dewasa bahkan anak-anak. Sementara permasalahan yang dihadapi adalah merebaknya tulisan-tulisan yang berbau negatif dan berdampak buruk bagi psikis pembacanya, terutama anak-anak. Di berbagai momen dan situasi selalu muncul berita-berita yang terkait dengannya namun bermunculan pula komentar-komentar negatif yang sangat memungkinkan untuk menggiring opini para pembacanya, maka perlu adanya penanggulangan dan pengidentifikasian sejak dini terhadap komentar-komentar negatif yang dimungkinkan untuk penggiringan opini pembacanya. Maka dari itu penelitian ini berupaya menganalisis konten negatif yang ada pada sosial media twitter berbahasa Indonesia. Analisis ini dilakukan dengan upaya mengklasifikasi tweet yang berisi komentar negatif dan positif menggunakan metode Naïve Bayes Classifier. Sementara metode Naïve Bayes classifier dikombinasikan dengan fitur yang dihitung bobotnya menggunakan TF-IDF. Klasifikasi tweet pada penelitian ini terdiri dari kelas sentiment negatif dan positif. Hasil pengujian pada penelitian didapatkan dengan menggunakan tools RapidMiner yang memperlihatkan bahwa akurasi klasifikasi dengan ditambahkan fitur TF-IDF menghasilkan tingkat akurasi sebesar 55.80%.

Kata kunci: klasifikasi, naïve bayes, analisis sentiment, pembobotan

1. PENDAHULUAN

Perkembangan teknologi pada saat ini mengalami peningkatan yang sangat signifikan dan mengalami pertumbuhan konsumsi pengguna yang sangat tinggi, hal ini berbanding lurus dengan pertumbuhan luasan pengguna portal berita elektronik atau *e-news* sebagai alternatif yang simpel dan memberikan informasi yang lengkap secara online (Irfani, Fauzi, & Sari, 2018). Peningkatan kualitas konten pada sosial media sebagai salah satu upaya untuk menghadirkan suatu informasi yang relevan, dalam hal ini twitter merupakan salah satu media sosial yang dipilih karena memiliki konten dengan konsistensi tinggi dalam menyajikan berita baik secara resmi dari lembaga tertentu maupun perseorangan.

Pada kenyataannya sosial media twitter tidak menyaring atau bahkan mengelompokkan informasi yang ada di dalamnya berdasarkan sentimen positif dan negatif, semua informasi dapat dibaca secara terbuka baik informasi yang menggambarkan suatu argumen positif dan argument negatif. Serta kebebasan yang diberikan oleh penggunanya dalam mengekspresikan setiap pikiran dan situasi emosi pada saat menulis (Sari & Adikara, 2019). Sese kali *tweet* yang ada mengalami strukturisasi Bahasa yang mungkin tidak mudah untuk difahami baik karena banyaknya penyingkatan maupun kata-kata yang tidak baku. Maka dari itu teks mining merupakan suatu solusi yang dapat dilakukan untuk mengetahui adanya sentimen konten pada twitter, sebagai dasar dari dilakukannya pemilihan konten dengan dataset teks.

Text mining sendiri merupakan teknik yang digunakan untuk memenuhi kebutuhan dalam proses klasifikasi dokumen dengan konten apapun (Sudiantoro et al., 2018). Cara kerja dari teknik ini merupakan upaya dalam memunculkan variasi dari kumpulan data yang tersedia dalam jumlah besar, untuk kemudian berusaha menemukan pola yang sesuai dengan apa yang diharapkan dari kumpulan data teks yang ada (Wiliyani, Sani, & Andyanto, 2019). Analisa sentiment menjadi proses yang sangat penting dalam memahami isi data dengan tujuan mengekstrak informasi atau menggali pengetahuan dari sumber data tekstual yang ada dalam jumlah besar secara otomatis. Hal ini diperuntukan pada dataset dengan opini positif dan negatif pada *tweet* berbahasa Inggris yang diperoleh dari portal <http://help.sentiment140.com>. Website tersebut menyediakan berbagai dataset yang dibutuhkan untuk penelitian penambangan data, termasuk data twitter berbahasa Inggris yang digunakan pada penelitian ini. Sedangkan model yang digunakan dalam menguji data tersebut adalah Metode *Naïve Bayes Classifier*, algoritma tersebut dianggap cocok digunakan untuk menganalisa sentiment yang ada pada dataset.

Model ini pula yang telah diusulkan oleh peneliti sebelumnya dalam melakukan klasifikasi data berupa teks (Pandhu Wijaya & Agus Santoso, 2018). Pada penelitian tersebut penulis memaparkan bahwa *Naïve Bayes Classification* merupakan model yang sesuai untuk melakukan klasifikasi dengan data teks dalam jumlah besar. Maka dari itu pada penelitian ini diambil metode *NBC* untuk mengklasifikasikan sentimen twitter pada *tweet* berbahasa Inggris. Pada penelitian tersebut (Pandhu Wijaya & Agus Santoso, 2018) memaparkan bahwa bahasa memiliki struktur masing-masing terutama bahasa Indonesia, maka konsep yang ada pada model *text mining* sangat tepat untuk mengatasi permasalahan analisis sentimen. Metode *NBC* yang diujikan pula menunjukkan progress yang baik dengan peningkatan performa sebesar 1.06% dengan dataset sebanyak 600 file pada masing-masing kelas yang diuji. Penelitian tersebut berfokus pada peningkatan akurasi pada klasifikasi berita *hoax* yang memiliki pola mirip dengan klasifikasi sentiment positif dan negatif pada Twitter yang penulis lakukan. Hasil yang didapat pada penelitian tersebut adalah

meningkatnya akurasi pada proses klasifikasi yang dilakukan dengan penambahan metode *Particle Swarm Optimization* dalam menyeleksi fitur yang digunakan dalam klasifikasi. Klasifikasi tunggal yang dilakukan menggunakan metode *Naïve Bayes* menghasilkan nilai akurasi sebesar 91,17%, sedangkan dengan penambahan metode *Particle Swarm Optimization* pada seleksi fitur menghasilkan akurasi sebesar 92,33%. Maka, dapat disimpulkan bahwa PSO efektif dalam meningkatkan akurasi klasifikasi pada metode *Naïve Bayes*. *Naïve Bayes* pun memiliki performa yang baik dalam melakukan klasifikasi tanpa adanya penambahan metode lain, hal ini yang menjadikan landasan bagi penulis untuk memilih metode *Naïve Bayes* pada penelitian ini.

Hal yang serupa pun dikatakan pada penelitian lain yang menggunakan data dari sosial media twitter menggunakan metode *Naïve Bayes* dengan Quert Expansion Hipernim-Hiponim (Irfani et al., 2018). Dengan latar belakang yang telah disampaikan pada awal pendahuluan yang menyatakan keberagaman media sosial Twitter dan tidak memiliki pengelompokan yang dimungkinkan untuk menjadi cara dalam mengidentifikasi suatu berita dengan kategori tertentu. Kesulitan terjadi pada pembaca yang ingin mengonsumsi bacaan pada sosial media tersebut dengan maksud mengambil informasi pada kategori tertentu, sesuai dengan apa yang pembaca kehendaki. Pada proses klasifikasi yang dilakukan menggunakan metode *Naïve Bayes* dengan penambahan *query expansion* guna menambahkan *term* yang menjadi indikator optimalnya proses klasifikasi yang dilakukan. *Term* merupakan indikator penting pada klasifikasi dengan *short text* seperti data pada penelitian yang penulis lakukan yaitu data sentiment sosial media Twitter. Tingkat akurasi yang dihasilkan pada penelitian tersebut sebesar 72% dengan klasifikasi menggunakan *Naïve Bayes* tanpa adanya penambahan *query expansion*, sedangkan penambahan *query expansion* mengakibatkan penurunan akurasi dari kemampuan klasifikasi. Pengujian yang dilakukan dengan menggunakan penambahan hiponim dan hipernim menghasilkan akurasi sebesar 65,75%, dan sebesar 66,3% untuk pengujian dengan hanya menambahkan hiponim, sedangkan untuk penambahan hipernim saja menghasilkan akurasi sebesar 67,5%. Maka dapat disimpulkan bahwa Metode *Naïve Bayes* memiliki performas yang baik dalam melakukan klasifikasi meski tanpa penambahan metode lain sebagai pendukung klasifikasi.

2. METODOLOGI PENELITIAN

2.1 Analisis Data

Mengacu kepada latar belakang yang telah disampaikan sebelumnya, penelitian ini mengarah kepada proses klasifikasi dokumen teks menggunakan *Naïve Bayes Classification*, dengan harapan bahwa dokumen dokumen yang diunduh dari media sosial twitter yaitu berupa sentimen positif dan negatif berbahasa Inggris dapat terpecahkan oleh konsep penelitian *Text Mining*. Berita yang diolah pada proses *text mining* ini adalah data pada tabel 1

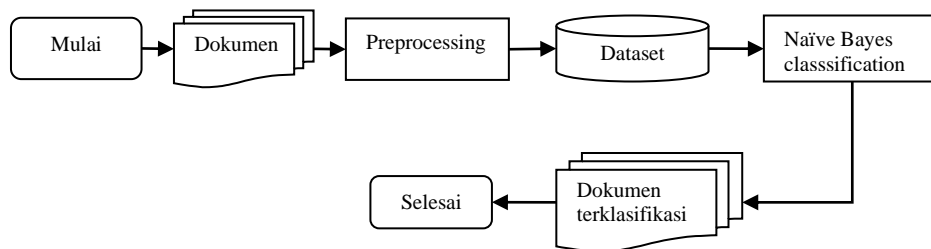
Tabel 1 Contoh Dataset

No	Data	Kelas
1	"@switchfoot http://twitpic.com/2y1zl - Awww, that's a bummer. You shoulda got David Carr of Third Day to do it. ;D"	Negatif
2	is upset that he can't update his Facebook by texting it... and might cry as a result School today also. Blah!	Negatif
3	@Kenichan I dived many times for the ball. Managed to save 50% The rest go out of bounds	Negatif
4	my whole body feels itchy and like its on fire	Negatif
5	"@nationwideclass no, it's not behaving at all. i'm mad.	Negatif

	why am i here? because I can't see you all over there. "	
6	I LOVE @Health4UandPets u guys r the best!!	Positif
7	im meeting up with one of my besties tonight! Cant wait!! - GIRL TALK!!	Positif
8	"@DaRealSunisaKim Thanks for the Twitter add, Sunisa! I got to meet you once at a HIN show here in the DC area and you were a sweetheart. "	Positif
9	"Being sick can be really cheap when it hurts too much to eat real food Plus, your friends make you soup"	Positif
10	@LovesBrooklyn2 he has that effect on everyone	Positif

Sumber : Sentiment140

Tabel 1 merupakan sample dataset yang digunakan pada penelitian ini yaitu berupa sentiment positif dan negatif pada twitter berbahasa Inggris yang diunduh dari situs Sentiment140. Data twitter pada dasarnya merupakan konten yang sangat bisa diandalkan dalam rangka pengambilan keputusan dikarenakan data yang tersimpan di dalamnya merupakan himpunan variasi dari berbagai opinin pada topik yang terkini (Gandharv, Richhariya, & Richhariya, 2017).



Gambar 1 Diagram Alir Metode

Data yang dialokasikan untuk penelitian ini sebanyak 2000 data yang dibagi rata dengan masing-masing pada berita positif sebanyak 1000 data teks dan 1000 data pada kategori negatif, dan data tersebut hanya teks tidak mengandung gambar. File yang telah diunduh kemudian disimpan dalam bentuk excel sebagai penyimpanan awal, akan tetapi jika data disimpan dalam bentuk excel penulis kesulitan untuk melakukan *preprocessing* dikarenakan *preprocessing* teks dilakukan secara *random* pada file teks secara utuh. Maka dari itu dataset yang tersedia harus dirubah terlebih dahulu kedalam format txt, sehingga diperlukan langkah *parsing* data pada teks tersebut.

Penelitian ini akan mengusulkan proses klasifikasi dokumen *tweet* yang dilakukan menggunakan metode *Naïve Bayes* dengan *tf-idf* sebagai langkah awal yang dilakukan untuk menghitung *term frequency* untuk kemudian dilanjutkan dengan proses klasifikasi menggunakan metode *Naïve Bayes*. Gambar 1 merupakan langkah dari proses yang akan dilakukan pada penelitian, dimulai dengan pengumpulan dokumen sebagai data latih yang akan menjadi acuan dalam membentuk pola baik data positif maupun data negatif. Setelah data terkumpul maka dilakukan *Preprocessing* data yang bertujuan membersihkan data yang semula memiliki banyak *noise* dan memiliki banyak tag sehingga akan berpengaruh pada proses klasifikasi termasuk pengaplikasian metode *tf-idf* dalam melakukan pembobotan sehingga

menghasilkan data yang memiliki nilai bobot sebagai bahan masukan dalam klasifikasi menggunakan metode *Naïve Bayes*.

2.2 Text Mining

Merupakan proses *mining* atau menambang suatu informasi dari data yang tersaji dalam jumlah besar, dalam hal ini adalah teks. Proses ini dilakukan dalam rangka penggalan, pengolahan, serta pengaturan pada informasi dengan menganalisa keterkaitan antara informasi satu dengan yang lainnya (Sudiantoro et al., 2018). Proses *text mining* ini sendiri merupakan hal yang berdasarkan kombinasi yang baik serta kata, dan frase yang sesuai, serta melibatkan proses ekstraksi fitur yang menjadi kunci efektif utama dalam tahap pelatihan atau yang sering disebut dengan *training* (Handayani & Pribadi, 2015).

2.3. Text Preprocessing

Struktur pada Data yang didapatkan merupakan data dengan kualitas apa adanya, sehingga diperlukan pra-proses yang menjadikan data tersebut mudah dalam komputasinya (Dellia, 2017). Perubahan data teks yang tidak terstruktur sangat diperlukan sehingga perlu adanya *Text Preprocessing*.

2.3.1. Tokenizing

Proses yang dilakukan pada *tokenizing* merupakan awal yang sangat penting, salah satunya adalah *text to lowercase* (Pandhu & Agus, 2016). Selebihnya proses yang dilakukan adalah menguraikan setiap kalimat yang ada pada data menjadi per kata, yang nantinya akan menjadi indikator keterkaitan antara satu kata denganyang lain.

2.3.2. Stopword filtering

Proses *text mining* membutuhkan langkah melakukan *indexing*, proses yang dimaksud adalah *stopword fltering* (Rahutomo & Ririd, 2019). Pada Bahasa Inggris terdapat banyak list kata-kata yang harus dihilangkan yang tidak sesuai dengan kaidah Bahasa Inggris, hal ini dilakukan dengan tujuan penyesuaian dokumen denga kaidah-kaidah Bahasa.

Tabel 1 Stopword Filtering

a	as	changes
a's	aside	clearly
able	ask	co
about	asking	com
above	associated	come
according	at	comes
accordingly	available	concerning
across	away	consequently
actually	awfully	consider
after	b	considering
afterwards	be	contain
again	became	containing
against	because	contains
ain't	become	corresponding
all	becomes	could

allow	becoming	couldn't
allows	been	course
almost	before	currently
alone	beforehand	d
along	behind	definitely
already	being	described
also	believe	despite
although	below	did

2.2.3 Stemming

Proses yang dilakukan pada tahap ini merupakan suatu tahapan mentransformasi data yang telah diolah menjadi kata-kata pada tahap sebelumnya, semula masing-masing kata memiliki imbuhan dan akhiran pada Bahasa Inggris dijadikan kata dasar (Pramudita, 2014).

2.2.4 Transform Case

Merupakan proses yang dilakukan guna merubah seluruh karakter dengan huruf kapital menjadi huruf standar guna mendukung klasifikasi (Jaka, 2015), hal tersebut dilakukan karena komentar pada twitter sebagian besar merupakan kata-kata sentiment yang mengandung huruf kapital.

2.3 Term Frequency

Perhitungan bobot pada tiap kata sebagai acuan dalam melakukan klasifikasi merupakan langkah yang penting, dengan menggunakan metode *TF-IDF* yang diaplikasikan pada fitur dapat diketahui bobot masing-masing kata terhadap dokumen (Amrizal, 2018). *Term Frequency* merupakan teknik dalam memberikan bobot pada kata terhadap dokumen tertentu, *TF-IDF* merupakan keterkaitan antara *term frequency* (TF) dan *invers document frequency* (IDF) dengan persamaan sebagai berikut :

Berlaku (W) untuk menghitung bobot

$$W_{dt} = tf_{dt} * Id_{ft} \quad (1)$$

Dimana :

- W_{dt} = bobot pada dokumen ke-d terhadap kata ke-t
- tf_{dt} = banyaknya kata yang dicari pada sebuah dokumen
- Id_{ft} = *Inversed Document Frequency* ($\log(N/df)$)
- N = total dokumen
- df = banyak dokumen yang mengandung kata yang dimaksud

2.4 Naïve Bayes

Klasifikasi merupakan tahapan yang bertujuan mengelompokkan objek berdasarkan data latih yang telah ditentukan kelasnya terlebih dahulu, kemudian proses ini mengelompokkan data baru berdasarkan probabilitas yang mendekati kelas yang sudah ada. Metode yang digunakan dalam melakukan klasifikasi adalah *Naïve Bayes*. Metode ini merupakan pendekatan yang dilakukan berlandaskan statistik (Annur, 2018), dengan persamaan sebagai berikut :

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Dimana :

- X = Data dengan Kelas yang belum diketahui
- H = Representasi data X merupakan Kelas yang spesifik
- $P(H|X)$ = Probabilitas H berdasarkan kondisi X
- $P(H)$ = Probabilitas
- $P(X|H)$ = Probabilitas X terhadap kondisi H
- $P(X)$ = Probabilitas X

3. HASIL DAN PEMBAHASAN

Pengujian metode Klasifikasi terhadap data dilakukan menggunakan rapidminer dengan porsi data yang seimbang pada masing – masing kelas, sehingga menghasilkan klasifikasi yang sepadan. Data teks dengan sentimen positif digunakan sebanyak 100 *record*, dan data sentimen negatif digunakan sebanyak 1000 *record* pula.

Row No.	label	metadata	filemetadata	p	metadata	d	aaaaahhhh	aaaaand	aaminahh	aaronva	aaru	abay	abbi	abigalloves	abi	about	abroad	absolu
1	negatif	negatif_1.bt	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	negatif	negatif_10.bt	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	negatif	negatif_100.bt	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	negatif	negatif_1000	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	negatif	negatif_101	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	negatif	negatif_102	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	negatif	negatif_103	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	negatif	negatif_104	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	negatif	negatif_105	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	negatif	negatif_106	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0.321	0	0
11	negatif	negatif_107	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	negatif	negatif_108	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	negatif	negatif_109	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	negatif	negatif_11.bt	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	negatif	negatif_110	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	negatif	negatif_111	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	negatif	negatif_112	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	negatif	negatif_113	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	negatif	negatif_114	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	negatif	negatif_115	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	negatif	negatif_116	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	negatif	negatif_117	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Gambar 2 Record data positif

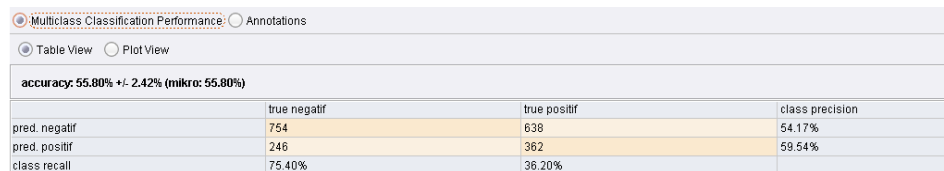
Gambar 2 menampilkan record data positif yang digunakan sebagai dataset yang diproses pada *tools* sesuai dengan kategori yang telah ditentukan oleh *repository* dengan data seperti pada Tabel 1.

Row No.	label	metadata	filemetadata	p	metadata	d	aaaaahhhh	aaaaand	aaminahh	aaronva	aaru	abay	abbi	abigalloves	abi	about	abroad	absolu
1001	positif	positif_1.bt	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1002	positif	positif_10.bt	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1003	positif	positif_100.bt	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1004	positif	positif_1000	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1005	positif	positif_101	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1006	positif	positif_102	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1007	positif	positif_103	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1008	positif	positif_104	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1009	positif	positif_105	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1010	positif	positif_106	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0.161	0	0
1011	positif	positif_107	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1012	positif	positif_108	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1013	positif	positif_109	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1014	positif	positif_11.bt	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0.191	0	0
1015	positif	positif_110	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1016	positif	positif_111	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1017	positif	positif_112	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1018	positif	positif_113	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1019	positif	positif_114	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1020	positif	positif_115	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1021	positif	positif_116	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1022	positif	positif_117	E\hatespee	04-Dec-2011	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Gambar 3 Record data negatif

Gambar 3 menampilkan record data negatif yang digunakan sebagai dataset yang diproses pada *tools* sesuai dengan kategori yang telah ditentukan oleh *repository* dengan data seperti pada Tabel 1.

Setelah seluruh data dipisah dan sebelumnya telah dilakukan *preprocessing* maka dilakukan klasifikasi pada data yang bersih dan siap untuk diklasifikasi, dengan metode *Naïve Bayes Classification* dihasilkan akurasi sebesar 55.80%.



Multiclass Classification Performance: Annotations			
Table View Plot View			
accuracy: 55.80% +/- 2.42% (mikro: 55.80%)			
	true negatif	true positif	class precision
pred. negatif	754	638	54.17%
pred. positif	246	362	59.54%
class recall	75.40%	36.20%	

Gambar 4 Akurasi klasifikasi *Naïve Bayes Classification*

Klasifikasi yang dilakukan pada 2000 record dataset *Tweeter* berbahasa Inggris dengan komposisi yang seimbang, dihasilkan nilai akurasi seperti pada Gambar 4. Hal tersebut dipengaruhi oleh beberapa faktor seperti *stopword*, *stem* proses yang harus dilengkapi dikarenakan struktur Bahasa yang beragam pada dataset tersebut disebabkan oleh gaya menulis pengguna twitter sangat beragam.

4. SIMPULAN

Klasifikasi konten twitter dengan sentimen positif dan negatif dilakukan dengan baik oleh *tools* dengan metode *Naïve Bayes Classifier*, namun beberapa aspek penting perlu ditingkatkan seperti keberagaman *stemmer*, dan *stop word list* guna meningkatkan akurasi pengklasifikasian dengan berbagai jenis pola penulisan berbahasa Inggris. Atau metode Klasifikasi lain dapat diterapkan pada penelitian serupa dan dataset yang sama.

DAFTAR PUSTAKA

- Amrizal, V. (2018). Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim). *Jurnal Teknik Informatika*, 11(2), 149–164. <https://doi.org/10.15408/jti.v11i2.8623>
- Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. *ILKOM Jurnal Ilmiah*, 10(2), 160. <https://doi.org/10.33096/ilkom.v10i2.303.160-165>
- Dellia, P. (2017). Kombinasi TF-IDF dan Pemilihan Fitur Markov Random Field (MRF) Dalam Perbaikan Kinerja Pengelompokan Twitter Keluhan Pajak, 109.
- Gandharv, S., Richhariya, V., & Richhariya, V. (2017). Real Time Text Mining on Twitter Data. *International Journal of Computer Applications*, 178(3), 24–28. <https://doi.org/10.5120/ijca2017915779>
- Handayani, F., & Pribadi, S. (2015). Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110, 7(1).
- Irfani, F. F., Fauzi, M. A., & Sari, Y. A. (2018). Klasifikasi Berita pada Twitter Menggunakan Metode Naive Bayes dan Query Expansion Hipernim-Hiponim. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 2(12), 6093–6099.
- Jaka, A. T. (2015). Preprocessing Text untuk Meminimalisir Kata yang Tidak

- Berarti dalam Proses Text Mining. *Informatika UPGRIS*, 1, 1–9.
- Pandhu, A., & Agus, H. (2016). Naive Bayes Classification pada Klasifikasi Dokumen Untuk Identifikasi Konten E-Government. *Journal of Applied Intelligent System*, 1(1), 48-55–55.
- Pandhu Wijaya, A., & Agus Santoso, H. (2018). Improving the Accuracy of Naïve Bayes Algorithm for Hoax Classification Using Particle Swarm Optimization. *Proceedings - 2018 International Seminar on Application for Technology of Information and Communication: Creative Technology for Human Life, ISemantic* 2018, 482–487.
<https://doi.org/10.1109/ISEMANTIC.2018.8549700>
- Pramudita, H. (2014). Penerapan Algoritma Stemming Nazief & Adriani Dan Similarity Pada Penerimaan Judul Thesis. *Data Manajemen Dan Teknologi Informasi (DASI)*, 15(4), 15.
- Rahutomo, F., & Ririd, A. R. T. H. (2019). Evaluasi Daftar Stopword Bahasa Indonesia. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 6(1), 41.
<https://doi.org/10.25126/jtiik.2019611226>
- Sari, Y. A., & Adikara, P. P. (2019). Klasifikasi Hate Speech Berbahasa Indonesia di Twitter Menggunakan Naive Bayes dan Seleksi Fitur Information Gain dengan Normalisasi Kata, 3(5), 4914–4922.
- Sudiantoro, A. V., Zuliarso, E., Studi, P., Informatika, T., Informasi, F. T., Stikubank, U., & Mining, T. (2018). Analisis Sentimen Twitter Menggunakan Text Mining Dengan Algoritma NAÏVE BAYES CLASSIFIER, 398–401.
- Wiliani, N., Sani, A., & Andyanto, A. T. (2019). Klasifikasi Kerusakan Dengan Jaringan Syaraf Backpropagation Pada Permukaan Solar Panel. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, 5(1), 89–94.
<https://doi.org/10.33480/jitk.v5i1.662>