

(1) WhatsApp x PKP ANALISIS SENTIMEN MASYARAKAT x +

Not secure | journal.poltekanika.ac.id/index.php/inf/article/view/201/185


INFORMANIKA

P-ISSN : 2407 - 1730
 E-ISSN : 2775 - 5762
 Jl. Kol. H. Burlian KM. 7 Kota Palembang

Publisher :
 Politeknik Anika
<http://poltekanika.ac.id/journal>

HOME ABOUT LOGIN REGISTER SEARCH CURRENT ARCHIVES

Home > JURNAL INFORMANIKA, VOL.7 NO.01 JANUARI-JUNI 2021 > Adha Oktarini Saputri

Download this PDF file

Page: 1 of 8 Automatic Zoom:

Jurnal Informanika, Volume 7 No.1, Januari-Juni 2021 ISSN :2407-1730

ANALISIS SENTIMEN MASYARAKAT TERHADAP PILPRES 2019
BERDASARKAN OPINI DARI TWITTER MENGGUNAKAN
METODE NAIVE BAYES CLASSIFIER

Nurul Adha Oktarini Saputri¹, Khoirul Zuhri²
Teknik Informatika, Fakultas Ilmu Komputer, Universitas Bina Dharma
Email:nuruladhaos@binadarma.ac.id¹, khoirulzuhri014@gmail.com²

ABSTRAK

OPEN JOURNAL SYSTEMS

TEMPLATE



ADDITIONAL MENU

- Author Guidelines
- Focus And Scope
- Online Submission
- Publication Ethics
- Editorial Team
- Peer Reviewers
- Copyright Notice
- Open Access Policy
- Peer Review Process
- Publication Frequency

jurnal.pdf | 201-373-1-SM.pdf | jurnal.pdf

Type here to search

ENG 9:30
INTL 08/03/2021

(1) WhatsApp x PKP ANALISIS SENTIMEN MASYARAKAT x +

Not secure | journal.poltekanika.ac.id/index.php/inf/article/view/201/185

Nurul Adha Oktarini Saputri¹, Khoirul Zuhri²
Teknik Informatika, Fakultas Ilmu Komputer, Universitas Bina Dharma
Email:nuruladhaos@binadarma.ac.id¹, khoirulzuhri014@gmail.com²

ABSTRAK

Twitter merupakan media sosial yang sedang populer saat ini, disini publik bebas berkomentar dan menulis apapun. Tidak jarang publik berkomentar dengan kata – kata kasar bahkan ujaran kebencian. Pemilihan presiden 2019 menuai banyak komentar,ada yang memuji,mengkritik dan menghina.Untuk dapat menggali informasi dan melakukan klasifikasi sebuah teks diperlukan analisis sentimen. Dalam penelitian ini analisis sentimen merupakan proses klasifikasi dokumen tekstual ke dalam dua kelas, yaitu kelas sentimen negatif dan positif. Data opini diperoleh dari jejaring sosial Twitter berupa tweet. Data yang digunakan berjumlah 3337 tweet terdiri dari 80% data latih dan 20% merupakan data training. Data latih merupakan data yang telah diketahui sentimennya. Penelitian ini bertujuan untuk menemukan tweet merupakan tweet positif atau negatif yang disampaikan di Twitter dalam Bahasa Indonesia. Pengklasifikasian data tweet menggunakan algoritma naive bayes classifier. Hasil klasifikasi pada data uji

REFBACKS

- There are currently no refbacks.



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Editorial Team

Peer Reviewers

Copyright Notice

Open Access Policy

Peer Review Process

Publication Frequency

Archiving Policy

Author Fees

License

Plagiarism Policy

Publishing System

Copyright Transfer Form

VISITORS

Visitors

ID 3,552	HK 7
US 107	CA 5
MY 18	NL 4
SG 10	TL 3
RU 8	KR 3

Pageviews: 8,956
Flags Collected: 30

FLAG counter

INDEX BY




jurnal.pdf | 201-373-1-SM.pdf | jurnal.pdf

Type here to search

ENG 9:51
INTL 08/03/2021

(1) WhatsApp x ANALISIS SENTIMEN MASYARAKAT x +

Not secure | journal.poltekanika.ac.id/index.php/inf/article/view/201/185



JOURNAL HELP

USER

Username

Password

Remember me

NOTIFICATIONS

View

Subscribe

LANGUAGE

Select Language

English

JOURNAL CONTENT

Search

Search Scope

All

jurnal.pdf ^ 201-373-1-SM.pdf ^ jurnal.pdf ^ x

Type here to search

ENG 9:51
INTL 08/03/2021

ANALISIS SENTIMEN MASYARAKAT TERHADAP PILPRES 2019 BERDASARKAN OPINI DARI *TWITTER* MENGGUNAKAN METODE *NAIVE BAYES CLASSIFIER*

Nurul Adha Oktarini Saputri¹, Khoirul Zuhri²

Teknik Informatika, Fakultas Ilmu Komputer, Universitas Bina Darma
Email:nuruladhaos@binadarma.ac.id¹, khoirulzuhri014@gmail.com²

ABSTRAK

Twitter merupakan media sosial yang sedang populer saat ini, disini publik bebas berkomentar dan menulis apapun. Tidak jarang publik berkomentar dengan kata – kata kasar bahkan ujaran kebencian. Pemilihan presiden 2019 menuai banyak komentar,ada yang memuji,mengkritik dan menghina.Untuk dapat menggali informasi dan melakukan klasifikasi sebuah teks diperlukan analisis sentimen. Dalam penelitian ini analisis sentimen merupakan proses klasifikasi dokumen tekstual ke dalam dua kelas, yaitu kelas sentimen negatif dan positif. Data opini diperoleh dari jejaring sosial *Twitter* berupa *tweet*. Data yang digunakan berjumlah 3337 *tweet* terdiri dari 80% data latih dan 20% merupakan data training. Data latih merupakan data yang telah diketahui sentimennya. Penelitian ini bertujuan untuk menentukan *tweet* merupakan *tweet* positif atau negatif yang disampaikan di *Twitter* dalam Bahasa Indonesia. Pengklasifikasian data *tweet* menggunakan algoritma *naïve bayes classifier*. Hasil klasifikasi pada data uji menunjukkan,algoritma *Naïve Bayes Classifier* memberikan nilai akurasi sebesar71%. Untuk nilai akurasi tiap sentimennya yaitu 71% untuk sentimen negatif dan 70% untuk sentiment positif.

Kata Kunci : *Analisis Sentimen, Klasifikasi, Naïve Bayes, Twitter, Pilpres 2019.*

I. PENDAHULUAN

Seiring dengan semakin majunya teknologi telekomunikasi, pengguna internet di dunia juga mengalami kemajuan yang cukup pesat. Hal ini dikarenakan semakin mudahnya aktifitas sehari-hari yang bisadilakudenganmemanfaatkaninternet,baikituhanyasekedarberkomunikasimelaluimediiasosial, mencari informasi secara *up-to-date* dan bahkan melakukan transaksi jual beli. Menurut data dari *Internet World Status*, Indonesia merupakan salah satu Negara dengan pengguna internet terbesar di dunia yang menempati posisi ke-5 dengan 143.260.000 pengguna. Perilaku pengguna internet lebih didominasi untuk

mengaksesmediasosialseperti*Facebook, Twitter, Instagram, YouTube* sebesar97,4%dari totalpengguna internet.

Menurut (Odevole, 2017) media sosial didefinisikan sebagai alat (*tools*) teknologi informasi *online* yang memungkinkan setiap pengguna melakukan komunikasi dengan mudah melalui internet dalam berbagi informasi seperti pesan teks, audio, video, gambar dan lain sebagainya [1]. Media sosial sekarang ini tidak hanya digunakan sebagai sarana pertemanan, mencari teman, tapi sudah banyak untuk kegiatan lain, seperti pengenalan suatu produk ataupun untuk sarana masyarakat dalam mengomentari hasil pilpres 2019.

Pemilihan Presiden 2019 sudah selesai dilakukan, dan banyak menghasilkan bermacam-macam opini dari masyarakat terutama melalui media sosial *Twitter*. Opini melalui *tweet* inilah yang dapat dimanfaatkan untuk melihat bagaimana sentimen yang dimunculkan oleh masyarakat mengenai Pilpres 2019. Penentuan sentimen positif, negatif sebenarnya bisa dilakukan secara manual, tetapi karena data yang sangat banyak maka menjadi semakin banyak tentunya waktu dan usaha yang dibutuhkan untuk mengklasifikasikan opini-opini tersebut. Oleh sebab itu peneliti tertarik dalam menerapkan metode pembelajaran mesin untuk mengklasifikasikan opini dari data yang sangat banyak tersebut. Untuk melakukan klasifikasi data tersebut peneliti menggunakan metode *Naive Bayes Classifier*.

Metode *Naive Bayes Classifier* merupakan salah satu metode *machine learning* yang menggunakan perhitungan probabilitas. Cara kerja dari metode *Naive Bayes Classifier* menggunakan perhitungan probabilitas. Konsep dasar yang digunakan oleh *Naive Bayes* adalah *Teorema Bayes*, yaitu teorema yang digunakan dalam statistika untuk menghitung suatu peluang, *Bayes Optimal Classifier* menghitung peluang dari satu kelas dari masing-masing kelompok atribut yang ada, dan menentukan kelas mana yang paling optimal.

II. TINJAUAN PUSTAKA

2.1 Analisis Sentimen

Menurut (beineke et al. 2004) Analisis sentimen adalah deteksi sikap-sikap (*attitudes*) terhadap objek atau orang [2]. Dari miliaran data cewitan di

twitter, anda dapat melakukan analisis sentimen untuk menemukan berapa persentase sentiment positif dan berapa persentase sentiment negative terhadap seseorang, perusahaan, institusi, kelompok, atau sebuah situasitertentu

2.2 Naive Bayes

Menurut (Dr. Ghayda A et al. 2013) *Naive Bayes* adalah metode terbaru yang digunakan untuk mengklasifikasikan sekumpulan dokumen. Algoritma ini memanfaatkan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya [4]. Keuntungan penggunaan metode *Naive Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian.

III. METODOLOGI PENELITIAN

3.1 Metode Penelitian

Dalam penelitian ini peneliti menggunakan metode deskriptif (*Deskriptive Research*). Menurut (Sugiyono, 2015) deskriptif adalah metode yang berfungsi untuk mendeskripsikan atau menggambarkan terhadap objek yang diteliti melalui data atau sampel yang telah terkumpul tanpa bermaksud membuat kesimpulan yang berlaku umum atau generalisasi [3].

3.2 Metode Pengumpulan Data

Peneliti menggunakan metode *crawling* untuk mengumpulkan data. *Crawling* adalah suatu teknik yang digunakan untuk mengumpulkan informasi yang ada dalam web.

Crawling berkerja secara otomatis, dimanainformasiyangdikumpulkanberdasarkanataskatakunciyangdiberikan olehpengguna. Alatyang digunakan untuk *crawling* disebut *crawler*. *Crawler* berbentuk program yang diprogram dengan algoritma tertentu sehingga dapat melakukan pemindaian ke halaman-halaman *web*, sesuai dengan alamat *web* ataupun kata kunci yang diberikan pengguna. Peneliti menggunakan *Python* untuk *crawling* data *twitter* ini.

IV. HASIL DAN PEMBAHASAN

4.1 Mengumpulkan Data Tweet

Pengumpulan data dilakukan dengan teknik *crawling*, Tahap awal untuk melakukan proses pengambilan data dari *Twitter*, peneliti harus memiliki *keynumber* dan *secret number* dari *API Key Twitter*. Mendapatkan *API Key Twitter* peneliti harus mendaftarkan atau melakukan pengajuan terhadap pihak *developer* atau pengembang *Twitter* untuk mendapatkan *API Key Twitter* tersebut. Hasil proses *crawling* data, menggunakan *API Key Twitter* maka didapatkan data dengan jumlah 3.337 *record Tweet*, dengan ketentuan data latih sebesar **70%** dan data training sebesar**30%**.



Gambar 1. Hasil Crawling

Hasil dari proses *crawling* data di atas yang nantinya akan dilakukan proses *labelling* data untuk menentukan klasifikasi pendapat atau pandangan dari hasil *tweet* yang telah di *crawling* tadi. Pada proses *labeling* ini dibedakan menjadi dua kelas.Yaitu *class positif* dan *class negative*. Contoh dari proses *labelling* data seperti di bawah ini.

Tabel 1. Data Hasil labelling

Tweet	Pemberitahuan Text	Label
b'RT @hd_bijau : jokowi yessssssss #pilpres#jokowi#prabowo #pemuh#indonesia#prabowosand #pileg	Jokowi yes	0
B'Prabowo insya Allah jadi presiden 2014 :), yang like berarti setuju. Hehehe #pilpres2014#pemuh	prabowo insya allah jadi presiden 2019, yang like berarti setuju. hehe	0
B'@detikcom : Tolak prabowo jd presiden.masa lalunya yg penuh dosa sampai mantan mertuanya bilang prabowo itu penghianat jd para pendukung prabowo ga usah banyak omong lah #jokowi#prabowo#pilpres #jokowipresiden	tolak prabowo jd presiden.masa lalunya yg penuh dosa sampai mantan mertuanya bilang prabowo itu penghianat jd para pendukung prabowo ga usah banyak omong lah	1
B'Moga Bapak Jokowi Menjadi Presiden NKRI And Tentunya Tegap, Ramah Dan Bertanggung jawab. #pilpres#jokowi	moga bapak jokowi menjadi presiden nkn, tentunya tegap, ramah, dan bertanggung jawab	0
b'@detikcom Kamret Semubar HIDUP JOKOWI kamret n#JokowiPresidenRT	kamret semubar, pembunuh mau jadi presiden. hidup jokowi	1

Dalam kasus ini *class positif* label 1 menyatakan bahwa *tweet* tersebut adalah kata-kata yang mengandung unsur *hatespeech* atau ujaran kebencian, sedangkan *class negative* berlabelkan 0 adalah kata-kata yang netral atau tidak mengandung unsur *hatespeech*.

4.2 Preprocessing Data

Preprocessing adalah tahapan proses untuk membersihkan data dari kata-kata atau *tweet* yang tidak di perlukan serta kata-kata yang tidak memiliki makna.

a. Cleaning

Cleaning adalah proses penghapusan simbol,tandabaca,huruf capital dan bilangan angka yang sering muncul pada *tweet* pengguna *Twitter* sehingga

data tersebut menjadi data yang tidak efektif dan tidak memiliki arti. Penerapan proses *cleaning* dapat dilihat seperti Tabel di bawah ini

Tabel 2. Data Hasil Cleaning.

Tweet Sebelum Cleaning	Tweet Setelah Cleaning
b7RT @hd_hisau : jokowi seessssss #pilpres#jokowi#prabowo #pemub#indonesia#prabowosandi #pileg	jokowi yes
B Prabowo insya Allah ia di presiden 2014 -, yang like berarti setuju. Hehe #pilpres2014#pemalu	prabowo insya allah ia di presiden 2019, yang like berarti setuju. hehe
B @detikcom : Tolak prabowo jd presiden.masa lahuya yg pemuh dosa.sampai2 mantan mertuanya bilang prabowo itu penghianat jd para pendukung prabowo ga usah banyak omong lah #jokowi#prabowo#pilpres #jokowi#presideni	tolak prabowo jd presiden.masa lahuya yg pemuh dosa.sampai2 mantan mertuanya bilang prabowo itu penghianat jd para pendukung prabowo ga usah banyak omong lah
B Moza Bapak Jokowi Menjadi Presiden NKRI.And tentunya Tegas, Ramah, Dan Bertanggung jawab. :) #pilpres#jokowi	moza bapak jokowi menjadi presiden nkn. tentunya tegas, ramah, dan bertanggung jawab
b/@detikcom Kampret Sesumbar. HIDUP JOKOWI kampret us#JokowiPresidenR!	kampret sesumbar. pembunuh mau jadi presiden. hidup jokowi

b. Remove Stopword

Remove Stopword adalah proses pephapusan kata-kata yang kurang bermakna atau kata yang tidak memiliki arti seperti kata dan, atau, kamu, saya. Proses penerapan pada tahap *Remove Stopword* dapat dilihat pada Tabel di bawah ini.

Tabel 3. Data Hasil Remove Stopword

Tweet Sebelum Remove Stopword	Tweet Setelah Remove Stopword
jokowi yes	jokowi yes
prabowo insya allah ia di presiden 2019, yang like berarti setuju. hehe	prabowo insya allah presiden 2019, berarti setuju.
tolak prabowo jd presiden.masa lahuya yg pemuh dosa.sampai2 mantan mertuanya bilang prabowo itu penghianat jd para pendukung prabowo ga usah banyak omong lah	tolak prabowo presiden, masa pemuh dosa.sampai2 mantan mertuanya bilang prabowo penghianat para pendukung prabowo ga usah banyak omong
moza bapak jokowi menjadi presiden nkn. tentunya tegas, ramah, dan bertanggung jawab	bapak jokowi presiden nkn. tegas, ramah, bertanggung jawab
kampret sesumbar. pembunuh mau jadi presiden. hidup jokowi	kampret sesumbar. pembunuh presiden. hidup jokowi

c. Tokenization

Tokenization adalah proses untuk memecahkan kalimat untuk menjadi beberapa bagian yang dinamakan *token*. Sebuah *token* dapat dianggap menjadi satu bentuk sebuah kata, frasa, atau suatu elemen yang berarti. Proses pada tahap *tokenization* dapat dilihat pada Tabel di bawah ini.

Tabel 4. Data Hasil Tokenization

Tweet Sebelum Tokenization	Tweet Setelah Tokenization
jokowi yes	[jokowi, 'yes']
prabowo insya allah presiden 2019, like berarti setuju	['prabowo', 'insya', 'allah', 'presiden', '2019', 'berarti', 'setuju']
tolak prabowo presiden. masa pemuh dosa.sampai2 mantan mertuanya bilang prabowo penghianat para pendukung prabowo ga usah banyak omong	['tolak', 'prabowo', 'presiden', 'masa', 'pemuh', 'dosa', 'sampai', 'mantan', 'mertuanya', 'bilang', 'prabowo', 'penghianat', 'para', 'pendukung', 'prabowo', 'banyak', 'omong']
bapak jokowi presiden nkn. tegas, ramah, bertanggung jawab	['bapak', 'jokowi', 'presiden', 'nkn', 'tegas', 'ramah', 'bertanggung', 'jawab']
kampret sesumbar. pembunuh presiden. hidup jokowi	['kampret', 'sesumbar', 'pembunuh', 'presiden', 'hidup', 'jokowi']

d. Stemming
Stemming adalah proses untuk mengubah kata kedalam bentuk dasar dengan menghapus kata imbuhan di depan maupun di belakang kata. Proses penetapan *stemming* dapat dilihat pada Tabel di bawah ini.

Tabel 5. Tabel Hasil Stemming

Tweet Sebelum Stemming	Tweet Setelah Stemming
[jokowi, 'yes']	jokowi yes
['prabowo', 'insya', 'allah', 'presiden', '2019', 'like', 'berarti', 'setuju']	prabowo insya allah presiden 2019 berarti setuju.
['tolak', 'prabowo', 'presiden', 'masa', 'pemuh', 'dosa', 'sampai', 'mantan', 'mertuanya', 'bilang', 'prabowo', 'penghianat', 'para', 'pendukung', 'prabowo', 'banyak', 'omong']	tolak prabowo presiden masa pemuh dosa sampai mantan mertua bilang prabowo khianat para dukung prabowo banyak omong
['bapak', 'jokowi', 'presiden', 'nkn', 'tegas', 'ramah', 'bertanggung', 'jawab']	bapak jokowi presiden nkn tegas ramah tanggung jawab
['kampret', 'sesumbar', 'pembunuh', 'presiden', 'hidup', 'jokowi']	kampret sesumbar bunuh presiden hidup jokowi

4.3 Ekstraksi Fitur

Pada proses ekstraksi fitur, proses pertama yang dilakukan oleh sistem setelah *tokenization* yaitu mengubah dataset menjadi suatu representasi *vector*

(DTweet1) "Jokowi Yes"

(DTweet2) "Prabowo InsyaAllah Presiden 2019 Berarti Setuju"

(DTweet3) "Kampret Sesumbar Bunuh Presiden Hidup Jokowi"

Setelah sistem melakukan *preprocessing* terdapat 4 jumlah kata baku dari 3 kalimat di atas yaitu "Presiden", "Setuju", "Hidup", dan "Bunuh".

Setelah tahapan di atas dari setiap dokumen ditampilkan mejadi sebuah *vector* dengan elemen, ketika kata tersebut terdapat di dalam dokumen maka diberikan nilai 1, jika tidak ada maka diberikan nilai 0.

Tabel 6. Pembuatan Word Vector

	Presiden	Setuju	Hidup	Bunuh
DTweet1	1	1	0	0
DTweet2	0	0	2	1
DTweet3	1	0	0	1

Dokumen yang telah diubah menjadi *word vector* selanjutnya akan dihitung menggunakan rumus *TF-IDF*, dengan menggunakan rumus ini maka akan menghasilkan *word vector* yang memiliki nilai yang sudah terbobot. Berikut adalah proses perhitungan *TF-IDF*

Tabel 7. Perhitungan TF (Term Frequency)

	D _{Tweet1}	D _{Tweet2}	D _{Tweet3}
Presiden	1	0	1
Setuju	1	0	0
Hidup	0	2	0
Bunuh	0	1	1

Setelah proses perhitungan bobot *TF* selesai selanjutnya dilakukan proses menentukan *DF* atau *Document Frequency* yaitu dengan banyaknya *term (t)* muncul dalam semua dokumen. Maka akan memperoleh hasil sebagai berikut

Tabel 8. Perhitungan DF (Document Frequency)

T (Term)	DF (Document Frequency)
Presiden	2
Setuju	1
Hidup	2
Bunuh	2

Dilanjutkan menghitung nilai *IDF* (*Inverse Document Frequency*) dengan cara menghitung nilai dari log hasil D atau jumlah dokumen dalam contoh kasus ini ada 3 tweet, dari 3 dokumen tersebut dibagi dengan nilai *DF* (*Document Frequency*).

Tabel 9. Proses IDF (Inverse Document Frequency)

T (Term)	DF (Document Frequency)	D/DF	IDF (Inverse Document Frequency)
Presiden	2	1.5	$\log 1,5 = 0,176$
Setuju	1	3	$\log 3 = 0,477$
Hidup	2	1.5	$\log 1,5 = 0,176$
Bunuh	2	1.5	$\log 1,5 = 0,176$

Setelah mendapatkan nilai *IDF* (*Inverse Document Frequency*), selanjutnya dilanjutkan dengan menghitung *TF-IDF*.

Tabel 10. Proses Perhitungan TF-IDF

Q	TF			D	D/D	IDF	IDF+	D _{Tweet1}	D _{Tweet2}	D _{Tweet3}
	D _{Tweet1}	D _{Tweet2}	D _{Tweet3}							
Presiden	1	0	1	2	1.5	0.477	1.176	1.176	0	1.176
Setuju	1	0	0	1	3	0.176	1.477	1.477	0	0
Hidup	0	2	0	2	1.5	1.176	0	2.352	0	0
Bunuh	0	1	1	2	1.5	0.176	0	1.176	1.176	1.176
								2.653	3.528	2.352

Hasil dari *word vector* yang sudah mendapatkan bobot dapat dilihat di bawah ini.

Tabel 11. Word Vector yang Sudah Dibobotkan

	Presiden	Setuju	Hidup	Bunuh
D _{Tweet1}	1.176	1.477	0	0
D _{Tweet2}	0	0	2.352	1.176
D _{Tweet3}	1.176	0	0	1.176

4.4 Implementasi Klasifikasi Naive Bayes Pada Python

Pada proses ekstraksi fitur dan proses pengklasifikasian *Naive Bayes* yang nantinya akan di compress menjadi satu *class pipeline vectorizer => transformer => classifier*. Proses pengklasifikasian tersebut berjalan dengan bantuan *library* pada bahasa pemrograman Python3 yang mempunyai nama *library scikit-learn* untuk proses pengklasifikasian, selain itu terdapat *library numpy* dan juga *panda* sebagai pembacaan data.

Untuk *library scikit-learn* disini yang digunakan adalah *Pipeline, CountVectorizer, Naive Bayes, MultinomialNB, Confusion Matrix, TfidfTransformer*, dan *f1 Score*.

Untuk langkah awal pengerjaan proses ekstraksi fitur dan klasifikasi adalah dilakukan proses menginstal *library* yang diperlukan. Selanjutnya setelah semua *library* terinstal maka dilanjutkan ke proses deklarasi

semua library yang akan digunakan. Adapun kode program untuk deklarasi pada Gambar dibawah.

```
import pandas as pd
import numpy as np
from sklearn.naive_bayes import MultinomialNB
from sklearn.svm import LinearSVC, SVC
from sklearn.feature_extraction.text import CountVecorizer
from sklearn.feature_extraction.text import TfidfTransformer
from sklearn.pipeline import Pipeline
from sklearn.model_selection import train_test_split
from sklearn.model_selection import cross_val_score
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score, f1_score, precision_score, recall_score
```

Gambar 2. Fungsi pendeklarasian *library* yang digunakan

Setelah selesai mendeklarasi *library* dilanjutkan dengan proses mengambil *dataset* yang akan dipergunakan sebagai data *training* menggunakan *library pandas*. Untuk kode program tersebut pada Gambar di bawah.

```
data = pd.read_excel('analissentimenpilpres2019.CSV', encoding='latin-1')
len(data)
```

Gambar 3. Fungsi memanggil data set

Selanjutnya proses pembuatan *class pipeline* yang di dalamnya terdapat 3 tahapan yaitu mengubah *dataset* yang hasil *crawling data Twitter* menjadi *representasi vector* (mengubah huruf menjadi angka) menggunakan *library Count Vectorizer* dengan pembobotan menggunakan *word vector* dalam *library Tfidf Transformer*, tahapan terakhir dilakukan klasifikasi dengan menggunakan *library Multinomial Naive Bayes*. Proses dari pengimplementasian dari tiga proses pembuatan *class pipeline* pada Gambar di bawah ini.

```
pipeline_mnb = Pipeline([
    ('vect', CountVectorizer()),
    ('tfidf', TfidfTransformer(use_idf=True, smooth_idf=True)),
    ('clf', MultinomialNB(alpha=1))
])

txt = data['cleartext'].values.astype('U')
X_train, X_test, y_train, y_test = train_test_split(data['cleartext'], data['label'], test_size=0.33, random_state = 0)
X_train, X_test, y_train, y_test = train_test_split(txt, data['label'], test_size=0.33, random_state = 0)
pipeline_mnb.fit(X_train, y_train)
```

Gambar 4. Proses Pengimplementasian *Class Pipeline*

4.5 Uji Model

Untuk mengetahui tingkatan dari performa *Algoritme Naive Bayes*, maka peneliti melakukan pengujian terhadap model. Hasil dari klasifikasi nantinya akan ditampilkan dalam bentuk *confusionmatrix*. Tabel yang ditampilkan di dalam *confusion matrix* ini terdiri dari kelas *predicted* dan juga kelas *actual*. Model dari *confusion matrix* ini dapat dilihat pada Tabelberikut.

Tabel 12. Model Confusion Matrix

		Predict Class	
		Class A	Class B
Actual Class	Positif	TP	FP
	Negatif	FN	TN

Untuk mengetahui nilai dari akurasi model diperoleh dari banyak jumlah data yang tepat hasil klarifikasi dibagi dengan total dari data, seperti pada Gambar di bawah ini.

$$Akurasi = \frac{TP + TN}{TP + FP + TN + FN}$$

Gambar 5. Rumus Akurasi

Pada saat proses pengujian model maka akan mendapatkan hasil dari nilai akurasi dan *confusion matrix* 2x2 sebesar 71.0% yang proses perhitungannya berdasarkan jumlah nilai dari diagonal *confusion matrix* dibagi dengan seluruh jumlah data. Karena jumlah pada data setiap kelas data *training* tidak seimbang, maka besarnya nilai akurasi bukanlah terpenting.

4.6 Evaluasi Model

Dalam proses evaluasi model ini dilakukan setelah uji model telah selesai dilakukan. Evaluasi model berguna sebagai menghitung performa dari metode yang dipilih. Pada proses uji model ini akan menghasilkan *confusion matrix* dengan ukuran 2x2.

Tabel 13. Hasil *Confusion Matrix*

		Predict Class	
		Positif	Negatif
Actual Class	Positif	558	105
	Negatif	261	334

Seperti pada Tabel 13. diatas, *confused matrix* matriks yang berukuran 2x2 setiap kolomnya mewakili nilai dari setiap kelas yaitu kelas positif, dan kelas negatif. Nilai presisi pada keseluruhan sistem bernilai sebesar **0.704** dan untuk nilai dari *recall* keseluruhan sistem berupa **0.615** sedangkan untuk nilai dari *f-1 Score* untuk pengevaluasian dalam informasi temu kembali yang dihitung mengombinasi nilai dari *presisi* dan *recall* yaitu sebesar **0.656**.

Dengan diketahuinya nilai dari *precision*, *recall*, dan *f-1 Score* dalam kinerja di keseluruhan sistem, maka dapat mengetahui kemampuan dari sistem untuk mencari ketepatan atau kebenaran dari informasi yang diminta oleh pengguna dengan hasil jawaban yang dikeluarkan oleh sistem dan memberitahu tingkat keberhasilan dari suatu system dalam menentukan kembali suatu informasi atau nilai *accuracy* sebesar **71%**. Setelah proses diatas selesai, untuk performa dari metode pengklasifikasian dari setiap kelas dapat diketahui dengan *precision*, *recall*, dan *f-1 Score* di setiap kelasnya. Hasil dari *precision*, *recall*, dan *f-1 Score* memiliki ukuran penilaian sebesar 0-1. Semakin tinggi nilai maka semakin baik, dalam artian semakin mendekati tingkat nilai dari 0 maka system semakin baik. Hasil dari

proses pengevaluasian model keseluruhan sistem ini terdapat pada Gambar di bawah.

Gambar 6. Hasil dari Proses Pengevaluasian Model

Hasil dari nilai *precision*, *recall*, dan *f-1*

Accuracy: 0.7103				
Confusion matrix:				
[[358 96]				
[143 228]]				
	precision	recall	f1-score	support
0	0.71	0.79	0.75	454
1	0.70	0.61	0.66	371
avg / total	0.71	0.71	0.71	825

Score di setiap kelas terdapat pada Tabel di bawah ini.

Tabel 14. Hasil dari Nilai *Precision*, *Recall*, dan *F-1 score*

Klasifikasi	<i>Precision</i>	<i>Recall</i>	<i>F-1 Score</i>
Positif	0.71	0.79	0.75
Negatif	0.70	0.61	0.66

Dapat dilihat dari hasil evaluasi model Tabel 4.14 dapat dilihat nilai *precision*, dan *recall* dari setiap kelas dapat dilihat tingkat kemampuan pemrosesan system dalam mencari tingkat ketepatan antara informasi yang diinginkan oleh pengguna sebagai kelas positif adalah “71%”, dan untuk kelas *negative* adalah “70%”. Tingkat keberhasilan dari pemrosesan sistem dalam memperoleh kembali informasi kelas positif adalah “79%”, untuk kelas negatif adalah “61%”.

Berdasarkan hasil *Tweet* dengan data 3.337, dengan ketentuan data latih sebesar 70% dan data *training* sebesar 30%. Dengan hasil yang didapat peneliti menyimpulkan dalam analisis sentiment pemilihan presiden 2019 nilai positif lebih tinggi dengan selisih sebesar 18% dengan tingkat akurasi sangat tinggi yaitu sebesar 71%.

Dengan nilai-nilai tersebut dapat dikatakan kinerja system dari keberhasilan system untuk menemukan kembali suatu informasi yang bernilai positif dan negatif dalam dokumen sangatlah tinggi.

V. KESIMPULAN

Berdasarkan penelitian yang dilakukan dan hasil pembahasan yang diuraikan, maka dapat ditarik kesimpulan sebagai berikut:

- a. Metode *Naive Bayes* dapat mengklasifikasikan data berupa teks, terutama teks yang berasal dari *Twitter (tweet)*.
- b. Jumlah kata dalam tiap kelas *training* sangat mempengaruhi hasil klasifikasi pada data *testing*, oleh karena itu keseimbangan data perlu dijaga.
- c. Kosakata tidak baku dapat mempengaruhi hasil klasifikasi terhadap suatu kelas *testing* apabila suatu kelas *training* memiliki data jumlah kata tidak baku lebih banyak dibanding kelas *training* lain.

DAFTAR PUSTAKA

- [1] Odewole. (2017) The Role of Librarian in Using Social Media Tools to Promote the Research Output of HIS/HER Clientes. *Journal of Education and Practice*, 8 (27),109-113.
- [2] Beineke, P., Hastie, T., Manning, C., & Vaithyanathan, S. 2004. Exploring Sentiment Summarization. In Y. Qu, J. Shanahan, & J. Wiebe (eds) *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and*

Applications, AAAI Press.

- [3] Sugiono (2016). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Alfabeta. Bandung.
- [4] Dr. Ghayda A. Al-Talib1, Hind S. Hassan, A . (2013). *Study on Analysis of SMS Classification Using TF-IDF Weighting*

SURAT KETERANGAN

Nomor: 004/SK/LPPM-UBD/III/2021

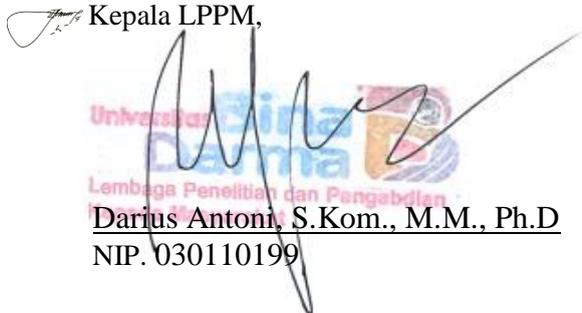
Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) Universitas Bina Darma menerangkan bahwa :

No	Nama	Jabatan
1	Nurul Adha Oktarini Saputri, M.Kom.	Dosen Program Studi Teknik Informatika

Adalah benar telah dipublikasikan artikel atau paper karya ilmiah dengan judul daftar terlampir.

Palembang, 24 Maret 2021

Kepala LPPM,



Darius Antoni, S.Kom., M.M., Ph.D
NIP. 030110199

DAFTAR JUDUL ARTIKEL ILMIAH

No.	Karya Ilmiah	Judul	Identitas Karya Ilmiah (ISBN/ISSN/Edisi/Tahun Terbit/Penerbit)	Alamat Unggah Online
1	Jurnal	ANALISIS SENTIMEN MASYARAKAT TERHADAP PILPRES 2019 BERDASARKAN OPINI DARI TWITTER MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER	JURNAL INFORMANIKA. P-ISSN: 2407-1730, E-ISSN: 2775-5762. VOL.7 NO.01 JANUARI-JUNI 2021. Manajemen Informatika, Politeknik Anika Palembang. Jurnal Nasional Tidak Terakreditasi	http://journal.poltekanka.ac.id/index.php/inf/article/view/201/185
2	Jurnal	IMPLEMENTASI PENGAMANAN DATA DAN INFORMASI DI BALAI DESA TANDING MARGA DENGAN METODE STEGANOGRAFI LSB DAN ALGORITMA KRIPTOGRAFI AES	JURNAL INFORMANIKA. P-ISSN: 2407-1730, E-ISSN: 2775-5762. VOL.7 NO.01 JANUARI-JUNI 2021. Manajemen Informatika, Politeknik Anika Palembang. Jurnal Nasional Tidak Terakreditasi	http://journal.poltekanka.ac.id/index.php/inf/article/view/202/186

Kepala LPPM



Darius Antoni, S.Kom., M.M., Ph.D
 NIP. 030110199