



**PERBANDINGAN ALGORITMA SUPPORT VECTOR MACHINE DAN  
NAIVE BAYES UNTUK MENILAI KEPUASAN MASYARAKAT  
TERHADAP PELAYANAN KESEHATAN MELALUI MEDIA SOSIAL  
TWITTER**

*TUGAS AKHIR*



Eren Zosa  
41518010024

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MERCU BUANA  
JAKARTA  
2022**



**PERBANDINGAN ALGORITMA SUPPORT VECTOR MACHINE DAN  
NAIVE BAYES UNTUK MENILAI KEPUASAN MASYARAKAT  
TERHADAP PELAYANAN KESEHATAN MELALUI MEDIA SOSIAL  
TWITTER**

*Tugas Akhir*

Diajukan Untuk Melengkapi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer

Oleh:

Eren Zosa  
41518010024

UNIVERSITAS  
MERCU BUANA  
PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MERCU BUANA  
JAKARTA  
2022

## LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:  
NIM : 41518010024  
Nama : Eren Zosa  
Judul Tugas Akhir : Perbandingan Algoritma Support Vector Machine dan Naive Bayes Untuk Menilai Kepuasan Masyarakat Terhadap Pelayanan Kesehatan Melalui Media Sosial Twitter

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan didalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.

Jakarta, 5 Juli 2022



MERCUBUANA  
PEPEL  
60B4FA0X943834273  
Eren Zosa



## SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Eren Zosa  
NIM : 41518010024  
Judul Tugas Akhir : Perbandingan Algoritma Support Vector Machine dan Naive Bayes Untuk Menilai Kepuasan Masyarakat Terhadap Pelayanan Kesehatan Melalui Media Sosial Twitter

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 05 Juli 2022



Eren Zosa

UNIVERSITAS  
MERCU BUANA

## SURAT PERNYATAAN LUARAN TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Eren Zosa  
NIM : 41518010024  
Judul Tugas Akhir : Perbandingan Algoritma Support Vector Machine dan Naive Bayes Untuk Menilai Kepuasan Masyarakat Terhadap Pelayanan Kesehatan Melalui Media Sosial Twitter

Menyatakan bahwa :

1. Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis	Status
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi	
		Jurnal Nasional Terakreditasi	✓
		Jurnal International Tidak Bereputasi	
		Jurnal International Bereputasi	Diterima
Disubmit/dipublikasikan di :	Nama Jurnal	: JATISI (Jurnal Teknik Informatika dan Sistem Informasi)	
	ISSN	: 25032933	
	Link Jurnal	: <a href="https://jurnal.mdp.ac.id/index.php/jatisi/">https://jurnal.mdp.ac.id/index.php/jatisi/</a>	
	Link File Jurnal Jika Sudah di Publish		

2. Bersedia untuk menyelesaikan seluruh proses publikasi artikel mulai dari submit, revisi artikel sampai dengan dinyatakan dapat diterbitkan pada jurnal yang dituju.
3. Diminta untuk melampirkan scan KTP dan Surat Pernyataan (Lihat Lampiran Dokumen HKI), untuk kepentingan pendaftaran HKI apabila diperlukan

Demikian pernyataan ini saya buat dengan sebenarnya.

Mengetahui  
Dosen Pembimbing TA



Sabar Rudiarto, M.Kom

Jakarta, 5 Juli 2022



Eren Zosa

## LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010024  
Nama : Eren Zosa  
Judul Tugas Akhir : Perbandingan Algoritma Support Vector Machine dan Naive Bayes Untuk Menilai Kepuasan Masyarakat Terhadap Pelayanan Kesehatan Melalui Media Sosial Twitter

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 04 Agustus 2022



## LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010024  
Nama : Eren Zosa  
Judul Tugas Akhir : Perbandingan Algoritma Support Vector Machine dan Naive Bayes Untuk Menilai Kepuasan Masyarakat Terhadap Pelayanan Kesehatan Melalui Media Sosial Twitter

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 04 Agustus 2022

  


(Vina Ayumi, S.Kom., M.Kom)

UNIVERSITAS  
MERCU BUANA

## LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010024  
Nama : Eren Zosa  
Judul Tugas Akhir : Perbandingan Algoritma Support Vector Machine dan Naive Bayes Untuk Menilai Kepuasan Masyarakat Terhadap Pelayanan Kesehatan Melalui Media Sosial Twitter

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 04 Agustus 2022



(Adi Hartanto, ST, M.Kom)

UNIVERSITAS  
MERCU BUANA

## LEMBAR PENGESAHAN

NIM : 41518010024

Nama : Eren Zosa

Judul Tugas Akhir : Perbandingan Algoritma Support Vector Machine dan Naive Bayes Untuk Menilai Kepuasan Masyarakat Terhadap Pelayanan Kesehatan Melalui Media Sosial Twitter

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 4 Agustus 2022

Menyetujui,



(Sabar Rudiarto, M.Kom)  
Dosen Pembimbing

Mengetahui,



(Wawan Gunawan, S.Kom, MT)  
Koord. Tugas Akhir Teknik Informatika



(Ir. Emil R. Kaburuan, Ph.D., IPM.)  
Ka. Prodi Teknik Informatika

## KATA PENGANTAR

Terimakasih dan rasa syukur saya panjatkan kepada Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, dikarenakan Tugas Akhir yang berjudul “Perbandingan Algoritma Support Vector Machine dan Naive Bayes Untuk Menilai Kepuasan Masyarakat Terhadap Pelayanan Masyarakat Melalui Media Sosial Twitter” dapat diselesaikan dalam jangka waktu yang sudah ditentukan. Laporan Tugas Akhir ini dibuat sebagai syarat untuk LULUS sebagai sarjana Ilmu Komputer dari Universitas Mercu Buana.

Penulis menyadari bahwa pembuatan Tugas Akhir ini tidak terlepas dari bantuan dan bimbingan berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Kedua orang tua. Ayah dan Ibu, yang tak pernah lelah mendukung dan selalu percaya bahwa saya bisa menyelesaikan kuliah dengan baik, juga tak pernah luput mendoakan yang terbaik untuk proses meraih gelar sarjana bagi saya.
2. Bapak Dr. Harwikarya, MT selaku Dosen Pembimbing Akademik
3. Bapak Sabar Rudianto, M.Kom selaku Dosen Pembimbing Tugas Akhir yang telah memberikan masukan saat bimbingan dan meluangkan waktu sebagian besarnya untuk melakukan bimbingan dalam penyusunan tugas akhir ini hingga selesai.
4. Seluruh Dosen Program Studi Teknik Informatika yang sudah memberikan ilmu yang bermanfaat selama kuliah berlangsung. Memberi kesempatan untuk belajar, berkarya dan juga berkembang.
5. Sahabat dan kerabat, yang telah percaya bahwa saya bisa melewati dan menyelesaikan Tugas Akhir ini dan mendapatkan gelar sarjana dengan baik, juga tidak pernah bosan memberi dukungan dan doa.
6. Seluruh Staff Administrasi dan Tata Usaha yang telah banyak membantu dan memberikan kemudahan, terima kasih atas semua pelayanan dan arahannya.

7. Semua pihak dan personal yang tidak dapat disebutkan satu per satu yang terlibat dalam pembuatan Tugas Akhir ini sehingga dapat selesai dengan baik.

Akhir kata, hasil Tugas Akhir ini masih jauh dari sempurna. Masih terdapat kekurangan dalam eksperimen, cara penjelasan maupun kekeliruan penulisan. Untuk itu, kritik dan saran pembaca sangat dihargai dan diharapkan. Semoga Tugas Akhir ini dapat memberikan manfaat bagi para pembaca.

Jakarta, 5 Juli 2022

Eren Zosa



## DAFTAR ISI

HALAMAN SAMPUL.....	i
HALAMAN JUDUL .....	i
LEMBAR PERNYATAAN ORISINALITAS .....	ii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR... iii	
SURAT PERNYATAAN LUARAN TUGAS AKHIR.....	iv
LEMBAR PERSETUJUAN PENGUJI .....	v
LEMBAR PENGESAHAN .....	viii
ABSTRAK .....	ix
ABSTRACT.....	x
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xiii
NASKAH JURNAL .....	1
KERTAS KERJA.....	11
BAB 1. LITERATUR REVIEW.....	12
BAB 2. SOURCE CODE.....	20
BAB 3. DATASET.....	23
BAB 4. TAHAPAN EKSPERIMEN.....	25
BAB 5. HASIL SEMUA EKSPERIMEN.....	31
DAFTAR PUSTAKA .....	35
LAMPIRAN DOKUMEN HAKI.....	37
LAMPIRAN KORESPONDENSI .....	40

## NASKAH JURNAL

■1

### PERBANDINGAN ALGORITMA SVM & NAIVE BAYES TERHADAP PELAYANAN MASYARAKAT KESEHATAN PADA TWITTER

Eren Zosa<sup>1</sup>, Sabar Rudiarto<sup>2</sup>

<sup>1,2</sup>Universitas Mercu Buana; Jl. Raya, RT.4/RW.1, Meruya Sel., Kec. Kembangan, Jakarta, Daerah Khusus Ibukota Jakarta 11650, (021) 5857722  
Jurusan Ilmu Komputer, Informatika, Jakarta  
e-mail: <sup>1</sup>41518010024@student.mercubuana.ac.id, <sup>2</sup>sabar.rudiarto@mercubuana.ac.id.

#### Abstrak

Selama beberapa tahun ini, kita dihadapkan berbagai penyakit yang tidak tau akan dapat membahayakan nyawa kita atau tidak. Dan juga saat ini masih adanya virus corona (Covid-19) yang masih perlu diwaspadai agar tidak terkena dampak dari penyakit tersebut. Banyak hal yang telah diupayakan oleh pemerintah agar masyarakat tidak terkena dampak dari penyakit yang ada, dengan cara menyediakan sarana kesehatan seperti puskesmas, rumah sakit, ataupun klinik menjadi tempat kita memeriksa kesehatan. Tetapi ada juga masyarakat yang kesulitan dengan fasilitas yang disediakan, karena sistem yang mementingkan keuangan dibandingkan nyawa seseorang. Hal tersebut menyebabkan adanya pembicaraan yang memberikan dampak positif maupun negatif pada apa yang dialami oleh masyarakat menggunakan fasilitas tersebut. Untuk masalah tersebut banyak dibicarakan melalui media sosial Twitter, yang sering dijumpai berbagai macam tanggapan masyarakat yang mereka sampaikan mengenai pelayanan masyarakat kesehatan tersebut baik negatif maupun positif. Pada penelitian ini dilakukan suatu analisa mengenai perbandingan algoritma untuk mengetahui pendapat masyarakat terhadap pelayanan masyarakat terhadap kesehatan ini. Metode yang digunakan adalah algoritma klasifikasi Naive Bayes dan Support Vector Machine dengan dibantu dengan Google Collab dan menggunakan bahasa Python. Hasil eksperimen menunjukkan bahwa algoritma Support Vector Machine memberikan nilai akurasi paling tinggi yaitu 62,36% dan diikuti oleh algoritma Naive Bayes yaitu 40,76% untuk dataset yang diuji mengenai pelayanan masyarakat terhadap kesehatan.

**Kata kunci** — Machine Learning, Support Vector Machine, Naive Bayes, TF-IDF, Sarana Kesehatan

#### Abstract

During these few years, we are faced with various diseases that will not endanger our lives or not. And also at this time there is still a corona virus (Covid-19) that still needs to be watched out so as not to be affected by the disease. Many things have been tried by the government so that the community is not affected by the existing disease, by providing health facilities such as health centers, hospitals, or clinics as places for health checks. But there are also people who have difficulty with the facilities provided, because the system is more concerned with finances

than one's life. This has led to discussions that have both positive and negative impacts on what people experience using these facilities. Many of these problems were conveyed through social media Twitter, which often found various kinds of public responses that they conveyed about the health services, both negative and positive. In this study, an analysis of the comparison of the agortima was carried out to determine the public's opinion on this public service for health. The method used is the Naive Bayes classification algorithm and the Support Vector Machine assisted by Google Collab and using the Python language. The experimental results show that the Support Vector Machine algorithm provides the highest accuracy value, which is 62.36% and is followed by the Naive Bayes algorithm, which is 40.76% for the tested dataset regarding public services to health.

**Keywords** — Machine Learning, Support Vector Machine, Naive Bayes, TF-IDF, Health facility

## 1. PENDAHULUAN

Selama beberapa tahun ini, kita dihadapkan berbagai penyakit yang tidak tau akan dapat membahayakan nyawa kita atau tidak. Dan juga saat ini masih adanya virus corona (Covid-19) yang masih perlu diwaspadai agar tidak terkena dampak dari penyakit tersebut. Banyak hal yang telah diupayakan oleh pemerintah agar masyarakat tidak terkena dampak dari penyakit yang ada, dengan cara menyediakan sarana kesehatan seperti puskesmas, rumah sakit, ataupun klinik menjadi tempat kita memeriksa kesehatan. Tetapi ada juga masyarakat yang kesulitan dengan fasilitas yang disediakan, karena sistem yang mementingkan keuangan dibandingkan nyawa seseorang. Hal tersebut menyebabkan adanya pembicaraan yang memberikan dampak positif maupun negatif pada apa yang dialami oleh masyarakat menggunakan fasilitas tersebut.

Machine Learning adalah cabang kecerdasan buatan yang digunakan untuk mengajari mesin cara menangani data dengan lebih efisien. Dengan banyaknya dataset yang tersedia, machine learning sering digunakan untuk banyak industry dari kedokteran hingga militer untuk mengekstrak informasi yang relevan.[1] Penelitian ini menggunakan model *Machine Learning* untuk menguji keakuratan dalam klasifikasi yang dilakukan. Untuk dataset yang digunakan ialah menggunakan data crawling yang diperoleh melalui media sosial Twitter.

Untuk masalah tersebut banyak dibicarakan melalui media sosial Twitter, yang sering dijumpai berbagai macam tanggapan masyarakat yang mereka sampaikan mengenai pelayanan masyarakat kesehatan tersebut baik negatif maupun positif.[2] Pada penelitian ini dilakukan suatu analisa mengenai perbandingan agortima untuk mengetahui pendapat masyarakat terhadap pelayanan masyarakat terhadap kesehatan ini. Metode yang digunakan adalah algoritma klasifikasi Naive Bayes dan Support Vector Machine dengan dibantu dengan Google Collab dan menggunakan bahasa Python

Pada penelitian sebelumnya yang dilakukan oleh (R. Ardianto, T. Rivanie, Y. Alkhalifi, F. S. Nugraha, and W. Gata 2020) dengan judul "Sentiment Analysis on E-Sports for Education Curriculum Using Naive Bayes and Support Vector Machine". Penelitian ini menggunakan data time series berdasarkan edukasi terhadap olahraga e-sport. Hasil yang didapatkan, ialah hasil Perbandingan

tingkat akurasi, presisi dan recall antara Naïve Bayes dengan SVM menggunakan metode SMOTE. [4]

## 2. METODE PENELITIAN

Metode penelitian ini menjelaskan tentang proses mengumpulkan data sampai pada tahap evaluasi dan validasi hasil, sehingga dapat menghasilkan hasil perbandingan dari masing-masing algoritma.



Gambar 1. Tahapan Penelitian

Tahapan pertama yang dilakukan ialah Pengumpulan data yang nantinya akan digunakan dalam pengujian. Selanjutnya pada tahapan berikutnya dilakukan tahapan preprocessing untuk membersihkan dataset mentah menjadi data yang dapat diolah untuk mendapatkan hasil dari pengujian. Setelah data selesai dilakukan preprocessing. Setelah itu dilakukan tahapan pembobotan TF-IDF untuk memberikan bobot pada setiap kata yang terdapat di dalam dokumen, dan juga dalam tahap ini kita melakukan data split untuk nantinya data dapat dilakukan pengujian menggunakan algoritma yang ditentukan. Selanjutnya, dilakukan metode klasifikasi dengan menggunakan algoritma Naïve Bayes dan Support Vector Machine.

### 2.1 Pengumpulan Data

Data yang digunakan merupakan data yang didapat melalui media social Twitter. Total keseluruhan data yang digunakan sebanyak 1434 untuk studi kasus. Seluruh data yang didapat dibantu menggunakan Google Collab dengan bantuan API Twitter dalam mengcrawling dataset.

### 2.2 Labeling Data

Labeling dataset dilakukan secara manual yang bertujuan akan menghasilkan akurasi yang baik dalam tahap pelabelan. Labeling data dibagi menjadi 3 katogeri yaitu, netral, positif, dan negatif. Pada proses sistem data tersebut dilabeli dengan menggunakan 0 – 2 (netral = 0, positif = 1, negatif = 2).

4

teks	label
@limellimonia lh di aku belum bisa masa, aku na...	0
Keren f Puskesmas Kota Baru Bentuk Duta Hiperf...	1
Salah satu inovasi unggulan Puskesmas Taman Sa...	1
Pada kesempatan ini juga diberikan edukasi ten...	0
Pagi hari tadi, pengunjung Puskesmas Kecamatan...	0
...	...
Bener gak? 🤔🤔🤔\n\nhttps://t.co/QqntmMP6d4	0
@missufe bismillah bt byr bpjs	0
@dMiraCie0 @jokowi Gimana rakyat g ngeluh coba...	2
@rubyoetta akibat opias pke bpjs, mngenkn 🤔🤔🤔	0
Minta rujukan bpjs bisa 2x dalam bulan yang sa...	0

Gambar 2. Labeling Data Manual

Setelah melakukan Labeling data secara manual, penulis mendapatkan label yang berupa kategori yang disebutkan sebelumnya. Jumlah record data terdapat pada tabel 1.

	Netral	Positif	Negatif
Dataset	823	340	270

Tabel 1. Jumlah data Labeling Manual

### 2.3 Preprocessing Data

Preprocessing data bertujuan untuk membersihkan dataset dari data mentah menjadi data yang siap pakai agar memudahkan proses klasifikasi menjadi positif dan negatif. Data preprocessing adalah teknik data mining yang melibatkan transformasi data mentah menjadi format yang mudah dimengerti. Pada penelitian ini digunakan lima langkah dalam tahapan preprocessing. Berikut adalah penjelasan dari langkah preprocessing yang dilakukan dalam berbentuk tabel 2:

No	Proses	Deksripsi	Input	Output
1	Cleansing	Membersihkan atau menghapus kata-kata yang terdapat usernam, hastag, ataupun url pada tweets	@barasurvivor Sudah saya tanya katanya gaboleh karena saya baru 4 bulan. Ribet puskesmas ga jelas banget.	Sudah saya tanya katanya gaboleh karena saya baru bulan Ribet puskesmas ga jelas banget

2	Case Folding	Pada tahap ini, seluruh data yang berupa tweet akan diubah menjadi huruf kecil atau nonkapital.	Sudah saya tanya katanya gaboleh karena saya baru bulan Ribet puskesmas ga jelas banget	sudah saya tanya katanya gaboleh karena saya baru bulan ribet puskesmas ga jelas banget
3	Stopword	Proses penghapusan kata yang dianggap tidak memiliki arti. Hal ini disesuaikan dengan keinginan peneliti.	sudah saya tanya katanya gaboleh karena saya baru bulan ribet puskesmas ga jelas banget	saya gaboleh karena saya ribet puskesmas ga jelas banget
4	Stemming	Proses pemotongan kata – kata yang mengandung imbuhan menjadi kata dasar.	saya gaboleh karena saya ribet puskesmas ga jelas banget	saya karena saya ribet puskesma ga jelas banget
5	Tokenization	Proses pemisahan atau pemotongan setiap kata yang menyusun suatu dokumen.	saya karena saya ribet puskesma ga jelas banget	['saya', 'karena', 'saya', 'ribet', 'puskesmas', 'ga', 'jelas', 'banget']

Tabel 2. Proses PreProcessing Data

#### 2.4 TF-IDF

Untuk mengetahui seberapa penting sebuah kata mewakili sebuah kalimat, dilakukan pembobotan atau perhitungan. Pemberian skor dalam TF-IDF berdasarkan frekuensi munculnya kata dalam dokumen [5].

#### 2.5 Implementasi Algoritma

Setelah semua data telah di proses dan mendapatkan data bersih maka selanjutnya data akan di implementasikan menggunakan 2 algoritma yaitu algoritma Support Vector Machine dan Algoritma Naïve Bayes. Dalam implementasinya nilai x dan y ditentukan yaitu variabel independent dan variabel dependent.

##### 1. Algoritma Support Vector Machine

Dalam Algoritma ini mengelaskan data baru mengelompokkan data dengan memisahkan berdasarkan hyperlane dalam ruang N-dimensi (Jumlah fitur) yang

secara jelas mengklasifikasikan titik suatu data. Support Vector Machine ialah teknik untuk penggunaan suatu klasifikasi data ataupun regresi, SVM sendiri mempunyai prinsip dasar dalam linier classifier, yaitu adanya kasus klasifikasi yang secara linier dapat dipisahkan, tetapi dalam SVM yang telah dikembangkan agar dapat beroperasi pada masalah kasus yang non-linier dengan memasukkan konsep kernel pada ruang kerja berdimensi tinggi. Pengklasifikasi SVM melakukan klasifikasi biner, yaitu memisahkan sekumpulan vektor pelatihan untuk dua kelas yang berbeda  $(x_1, y_1)$ ,  $(x_2, y_2)$ , ...,  $(x_m, y_m)$ , di mana  $x_i \in R^d$  menunjukkan vektor dalam dimensional ruang fitur dan  $y_i \in \{-1, +1\}$  adalah label kelas.[6] SVM bekerja untuk menemukan hyperplane. Hyperplane adalah batas keputusan dalam sebuah fitur yang memisahkan dua kelas dengan margin terbesar. SVM juga memiliki akurasi yang tinggi dengan kesalahan yang relatif kecil dan bisa jadi digunakan untuk membuat prediksi.

## 2. Algoritma Naive Bayes

Proses yang dilakukan dalam algoritma naive bayes algoritma yang dipelajari probabilitas suatu objek dengan ciri-ciri tertentu yang termasuk dalam kelompok/kelas tertentu dalam data yang akan dimasukkan akan dihitung label atau kelasnya. Kemudian dari kelas tersebut akan dihitung probabilitas dari setiap kelas. Setelah didapat maka hasil kelas akan dibandingkan. Algoritma Naive Bayes disebut "naif" karena membuat asumsi bahwa kemunculan fitur tertentu tidak tergantung pada kemunculan fitur lainnya. [7]

Keuntungan menggunakan naive Bayes dibandingkan model lainnya adalah model ini cocok untuk sejumlah kecil data pelatihan untuk menentukan estimasi setiap parameter yang diperlukan dalam proses klasifikasi. Naive bayes memiliki teori dasar dari teorema bayes yang biasa dipakai dalam klasifikasi di bidang ilmu *machine learning*.

## 2.6 Evaluasi Data

Evaluasi diperlukan untuk menganalisa dan mengukur sejauh mana keakuratan hasil yang telah dicapai oleh model dengan menggunakan Confusion Matrix. Lalu hasil evaluasi akan diberikan dengan menggunakan classification report yang terdiri dari Accuracy, Precision dan Recall.[8]

### 1. Akurasi

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

### 2. Precision

$$precision = \frac{TP}{TP + FP}$$

### 3. Recall

$$recall = \frac{TP}{TP + FN}$$

### 3. HASIL DAN PEMBAHASAN

Pada bagian ini akan menampilkan hasil semua eksperimen yang telah dilakukan dalam penelitian. Untuk pengujian dilakukan dengan menggunakan 2 algoritma, yaitu Support Vector Machine dan Naive Bayes. Tujuan dalam penggunaan dalam kedua algoritma tersebut ialah, untuk membandingkan manakah hasil pengujian yang memiliki hasil terbaik dari algoritma yang diuji.

Naive Bayes (NB)

```
[[85 31 53]
 [37 15 12]
 [19 18 17]]
```

	precision	recall	f1-score	support
0	0.60	0.50	0.55	169
1	0.23	0.23	0.23	64
2	0.21	0.31	0.25	54
accuracy			0.41	287
macro avg	0.35	0.35	0.34	287
weighted avg	0.45	0.41	0.42	287

nilai akurasinya adalah 40.76655052264808

Support Vector Machine (SVM)

```
[[161 7 1]
 [ 52 12 0]
 [ 43 5 6]]
```

	precision	recall	f1-score	support
0	0.63	0.95	0.76	169
1	0.50	0.19	0.27	64
2	0.86	0.11	0.20	54
accuracy			0.62	287
macro avg	0.66	0.42	0.41	287
weighted avg	0.64	0.62	0.54	287

nilai akurasinya adalah 62.36933797909408

Setelah melakukan evaluasi terhadap kedua algoritma yang diuji, dapat kita ketahui dari akurasi yang ada dapat dilihat bahwa nilai akurasi untuk Algoritma Naive Bayes ialah 40,76%, sedangkan algoritma Support Vector Machine nilai akurasi 62,36%. Algoritma SVM yang telah diuji memiliki nilai akurasi yang lebih baik dibandingkan dengan Algoritma Naive Bayes, Karena algoritma NaiveBayes tidak termanfaatkannya probabilitas bersyarat variabel input yang dimiliki

naive bayes membuat performa naive bayes di bawah algoritma yang lain. Naive bayes akan berjalan baik pada data yang bersifat independent. Sedangkan data yang diolah memiliki variabel yang dependent. Yang menghasilkan Akurasi yang lebih kecil dibanding algoritma SVM. Dan Algoritma SVM dapat menghasilkan lebih baik, karena dalam algoritma ini dalam pengujian mengelompokkan data-data yang ada dengan cara memisahkannya berdasarkan nilai ruang dimensi yang secara jelas dapat mengklasifikasi titik suatu data.

Dari hasil yang telah diketahui proses evaluasi menghasilkan algoritma Support Vector Machine dalam akurasi, presisi dan recall yang lebih baik dibandingkan algoritma Naive Bayes, tetapi yang berbeda dari 2 algoritma ini hanya di waktu eksekusi pengujian, yang dimana untuk algoritma support vector machine lebih sedikit lebih lambat 1 detik dibandingkan dengan algoritma Naive Bayes dalam penelitian yang dilakukan.

#### 4. KESIMPULAN

Berdasarkan pengujian yang telah dilakukan berdasarkan dataset tweet yang diambil dengan cara mengcrawling data melalui media sosial Twitter menggunakan bantuan tools Google Collabs yang dapat disimpulkan dari tujuan dari penelitian untuk membandingkan tingkat akurasi tiap algoritma yang digunakan dengan studi kasus Pelayanan Masyarakat terhadap Kesehatan. Untuk model yang diuji yaitu algoritma Naive Bayes dan algoritma Support Vector Machine. Berdasarkan pengetahuan yang dipelajari dan diuji, Penulis dapat memberikan beberapa poin untuk disimpulkan yang dijelaskan sebagai berikut :

- Untuk pengujian yang dilakukan Algoritma Support Vector Machine memiliki akurasi yang lebih baik dibandingkan algoritma Naive Bayes.
- Akurasi yang didapat pada kedua algoritma ialah 62,36% untuk akurasi algoritma Support Vector Machine dan 40,76% untuk algoritma Naive Bayes
- Dalam waktu eksekusi dalam pengujian algoritma Naive Bayes memiliki waktu lebih dari 1 detik dibandingkan algoritma Support Vector Machine

#### 5. SARAN

Dalam penelitian penulis juga tidak dapat sepenuhnya mendapatkan hasil yang sesuai harapan penulis, disini akan dijelaskan saran apa saja yang harus dilakukan agar mendapatkan hasil yang lebih baik di berikut ini :

- Pengujian yang dilakukan jika bisa diuji lebih dari 1 kali.
- Melakukan pemilahan data labeling hanya menggunakan positif dan negatif.
- Menguji dengan algoritma lain dengan tujuan agar mendapat lebih tau kelebihan dan kekurangan dari tiap algoritma machine learning.

## UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Tuhan Yang Maha Esa serta pihak dosen Universitas Mercu Buana yang memberikan saya kesempatan dalam pembuatan jurnal ini dan juga membantu menyelesaikan jurnal ini sampai ke tahap ini. Tidak lupa penulis pun mengucapkan kepada seluruh pihak keluarga telah memberikan seluruh bentuk dukungan baik berupa moril maupun materil sehingga kegiatan pengabdian ini dapat terselesaikan.

## DAFTAR PUSTAKA

- [1] H. D. Wijaya and S. Dwiasnati, "Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat," *J. Inform.*, vol. 7, no. 1, pp. 1–7, 2020, doi: 10.31311/ji.v7i1.6203.
- [2] T. Rivanie, R. Pebrianto, T. Hidayat, A. Bayhaqy, W. Gata, and H. B. Novitasari, "Analisis Sentimen Terhadap Kinerja Menteri Kesehatan Indonesia Selama Pandemi Covid-19," *J. Inform.*, vol. 21, no. 1, pp. 1–13, 2021, doi: 10.30873/ji.v21i1.2864.
- [3] N. R. Wardani and A. Erfina, "Konsultasi Dokter Menggunakan Algoritma Naive," *SISMATIK (Seminar Nas. Sist. Inf. dan Manaj. Inform.)*, pp. 11–18, 2021.
- [4] R. Ardianto, T. Rivanie, Y. Alkhalifi, F. S. Nugraha, and W. Gata, "Sentiment Analysis on E-Sports for Education Curriculum Using Naive Bayes and Support Vector Machine," *J. Ilmu Komput. dan Inf.*, vol. 13, no. 2, pp. 109–122, 2020, doi: 10.21609/jiki.v13i2.885.
- [5] A. Widyanto, "Data Mining Predicts the Need for Immunization Vaccines Using the Naive Bayes Method," *J. Appl. Data Sci.*, vol. 2, no. 3, pp. 93–101, 2021, doi: 10.47738/jads.v2i3.39.
- [6] D. Ramayanti and U. Salamah, "Complaint Classification Using Support Vector Machine for Indonesian Text Dataset," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 3, no. 7, pp. 179–184, 2018.
- [7] W. Yulita *et al.*, "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier," *Jdmsi*, vol. 2, no. 2, pp. 1–9, 2021.
- [8] E. Indrayuni, "Komparasi Algoritma Naive Bayes Dan Support Vector Machine Untuk Analisa Sentimen Review Film," *J. Pilar Nusa Mandiri*, vol. 14, no. 2, p. 175, 2018, doi: 10.33480/pilar.v14i2.918.
- [9] Sourabh\* and P. V. Mansotra, "3335 Retrieval Number: D8118118419/2019©BEIESP DOI:10.35940/ijrte.D8118.118419 Journal Website: www.ijrte.org Published By: Blue Eyes Intelligence Engineering & Sciences Publication A Model for Accurate Prediction of Child Immunization Data for Knowledge Discovery using Bayesian TAN and Naive Bayes Classifiers," *Int. J. Recent Technol. Eng.*, vol. 8, no. 4, pp. 3335–3343, 2019, doi: 10.35940/ijrte.d8118.118419.
- [10] H. A. Santoso, E. H. Rachmawanto, A. Nugraha, A. A. Nugroho, D. R. I. M. Setiadi, and R. S. Bastuki, "Hoax classification and sentiment analysis of Indonesian news using Naive Bayes optimization," *Telkonnika (Telecommunication Comput. Electron. Control.)*, vol. 18, no. 2, pp. 799–806, 2020, doi: 10.12928/TELKOMNIKA.V18I2.14744.
- [11] M. Nur Habibi and Sunjana, "Analysis of Indonesia Politics Polarization before 2019 President Election Using Sentiment Analysis and Social Network Analysis," *Int. J. Mod. Educ. Comput. Sci.*, vol. 11, no. 11, pp. 22–30, 2019, doi: 10.5815/ijmeecs.2019.11.04.
- [12] R. Ahuja, A. Chug, S. Kohli, S. Gupta, and P. Ahuja, "The impact of features extraction on the sentiment analysis," *Procedia Comput. Sci.*, vol. 152, pp. 341–348, 2019, doi: 10.1016/j.procs.2019.05.008.
- [13] N. Ruhjana, "Analisis Sentimen terhadap Penerapan Sistem Plat Nomor Gnjil/Genap pada Twitter dengan Metode Klasifikasi Naive Bayes," *J. IKRA-ITH Inform.*, vol. 3, no. 1, pp. 94–99, 2019.

- [Online]. Available: [www.situs.com](http://www.situs.com).
- [14] A. Nugroho, "Analisis Sentimen Pada Media Sosial Twitter Menggunakan Naive Bayes Classifier Dengan Ekstraksi Fitur N-Gram," *J-SAKTI (Jurnal Sains Komput. dan Inform.,* vol. 2, no. 2, p. 200, 2018, doi: 10.30645/j-sakti.v2i2.83.
- [15] R. Wongso, F. A. Luwinda, B. C. Trisnajaya, O. Rusli, and Rudy, "News Article Text Classification in Indonesian Language," *Procedia Comput. Sci.,* vol. 116, pp. 137–143, 2017, doi: 10.1016/j.procs.2017.10.039.
- [16] R. A. Ariyanto and N. Chamidah, "Sentiment Analysis for Zoning System Admission Policy Using Support Vector Machine and Naive Bayes Methods," *J. Phys. Conf. Ser.,* vol. 1776, no. 1, pp. 0–9, 2021, doi: 10.1088/1742-6596/1776/1/012058.



## KERTAS KERJA

### Ringkasan

Kertas kerja ini merupakan material kelengkapan artikel jurnal yang telah terlampir sebelumnya dengan judul “Perbandingan Algoritma SVM dan Naive Bayes Terhadap Pelayanan Masyarakat Terhadap Kesehatan Menggunakan Data Crawling Melalui Twitter”. Kertas kerja ini berisi semua material hasil penelitian Tugas Akhir. Di dalam kertas kerja ini disajikan beberapa bagian yang terdiri dari literature review, dataset yang digunakan, tahapan eksperimen, dan hasil eksperimen secara keseluruhan.

Bagian I membahas mengenai literature review yang berisi artikel jurnal sebelumnya yang menjadi dasar atau landasan dalam penelitian ini. Bagian II menjelaskan tentang source code yang digunakan pada penelitian ini. Bagian III menjelaskan mengenai dataset yang digunakan. Bagian IV memuat tahapan eksperimen yang disajikan dalam gambar beserta penjelasan dari tiap tahapan. Bagian V merupakan bagian terakhir dari kertas kerja ini yang menjelaskan hasil keseluruhan dari eksperimen yang telah dilakukan, meliputi penjelasannya.

UNIVERSITAS  
MERCU BUANA