



**ANALISA SENTIMEN TERHADAP PEMBELAJARAN OFFLINE DI ERA  
NEW NORMAL MENGGUNAKAN ALGORITMA SVM DAN NAÏVE**

**BAYES**

*Tugas Akhir*

Diajukan Untuk Melengkapi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer

Oleh:

Azhar Fauzan

41517010010

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2021

## LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM : 41517010010

Nama : Azhar Fauzan

Judul Tugas Akhir : Analisa Sentimen Terhadap Pembelajaran Offline Di Era New Normal Menggunakan Algoritma SVM Dan Naïve Bayes

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.



UNIVERSITAS  
MERCU BUANA

## SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

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

Nama Mahasiswa : Azhar Fauzan  
NIM : 41517010010  
Judul Tugas Akhir : Analisa Sentimen Terhadap Pembelajaran Offline Di Era New Normal Menggunakan Algoritma SVM Dan Naïve Bayes

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.

MERCU BUANA



## SURAT PERNYATAAN LUARAN TUGAS AKHIR

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

Nama Mahasiswa : Azhar Fauzan  
 NIM : 41517010010  
 Judul Tugas Akhir : Analisa Sentimen Terhadap Pembelajaran Offline Di Era New Normal Menggunakan Algoritma SVM Dan Naïve Bayes

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	
Disubmit/dipublikasikan di :	Nama Jurnal	: Jurnal Teknik Informatika dan Sistem Informasi	
	ISSN	: 2407-4322	
	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

  
Dr. Ida Nuthaida, S.T.M.T.

Jakarta,

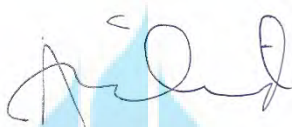
  
 Azhar Fauzan

## LEMBAR PERSETUJUAN PENGUJI

NIM : 41517010010  
Nama : Azhar Fauzan  
Judul Tugas Akhir : Analisa Sentimen Terhadap Pembelajaran Offline  
Di Era New Normal Menggunakan Algoritma SVM  
Dan Naïve Bayes

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, 18 Agustus 2021



(Anis Cherid, SE, MTI)

UNIVERSITAS  
MERCU BUANA

## LEMBAR PERSETUJUAN PENGUJI

NIM : 41517010010  
Nama : Azhar Fauzan  
Judul Tugas Akhir : Analisa Sentimen Terhadap Pembelajaran Offline  
Di Era New Normal Menggunakan Algoritma SVM  
Dan Naïve Bayes

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, 18 Agustus 2021



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

UNIVERSITAS  
MERCU BUANA

## LEMBAR PERSETUJUAN PENGUJI

NIM : 41517010010  
Nama : Azhar Fauzan  
Judul Tugas Akhir : Analisa Sentimen Terhadap Pembelajaran Offline  
Di Era New Normal Menggunakan Algoritma SVM  
Dan Naïve Bayes

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, 18 Agustus 2021



(Wawan Gunawan, S.Kom., MT)

UNIVERSITAS  
MERCU BUANA

## LEMBAR PENGESAHAN

NIM : 41517010010  
Nama : Azhar Fauzan  
Judul Tugas Akhir : Analisa Sentimen Terhadap Pembelajaran Offline Di Era New Normal Menggunakan Algoritma SVM Dan Naïve Bayes

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, 18 Agustus 2021

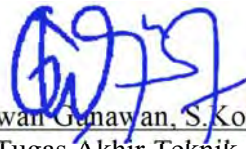
Menyetujui,



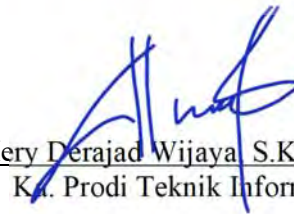
(Dr. Ida Nurhaida, S.T.M.T.)  
Dosen Pembimbing

UNIVERSITAS  
MERCU BUANA

Mengetahui,



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



(Hery Derajad Wijaya, S.Kom. MM)  
Ka. Prodi Teknik Informatika



## ABSTRAK

Nama : Azhar Fauzan  
NIM : 41517010010  
Pembimbing TA : Dr. Ida Nurhaida,S.T,M.T.  
Judul : Analisa Sentimen Terhadap Pembelajaran Offline  
Di Era New Normal Menggunakan Algoritma SVM  
Dan Naïve Bayes

Dari peristiwa yang terjadi belakangan ini dimana semakin luasnya penyebaran COVID-19 di Indonesia, pemerintah Indonesia mengeluarkan kebijakan-kebijakan guna mencegah penyebaran tersebut. Salah satunya adalah kebijakan yang diterapkan pada bidang pendidikan, dimana proses belajar mengajar dilakukan secara online atau daring. Sekarang, Indonesia telah melaksanakan masa tanggap darurat penanganan yang kemudian pemerintah mulai menjajaki kehidupan normal yang baru atau New Normal. Karena hal tersebut pada siaran pers pemerintah memperbolehkan kembali pelaksanaan pembelajaran tatap muka pada tahun ajaran baru di masa pandemi Covid-19 dengan tetap memperhatikan protokol kesehatan. Hal tersebut menyebabkan pro-kontra pada masyarakat. Pada penelitian ini dilakukan perbandingan Algoritma SVM dengan Naive Bayes dalam menganalisa sentimen mengenai pelaksanaan pembelajaran offline di masa new normal berdasarkan data tweet masyarakat. Data yang digunakan pada penelitian ini sebanyak 2.708 data yang telah melewati proses preprocessing, labeling yang dilakukan secara otomatis, resampling dan TF-IDF. Hasil yang didapat menggunakan SVM precision, recall dan F1-Score sebesar 91%, 91%, 91% dan 91.5% sedangkan dengan Naïve Bayes diperoleh hasil sebesar 79%, 83%, 79% dan 78,5% dengan percentage split sebesar 80%:20%. Dari hasil tersebut dapat disimpulkan bahwa Algoritma SVM lebih baik dibandingkan dengan Naive Bayes dalam menganalisa sentimen mengenai pembelajaran offline di era new normal.

Kata kunci:

New Normal, SVM, Naive Bayes

## ABSTRACT

Name : Azhar Fauzan  
Student Number : 41517010010  
Counsellor : Dr. Ida Nurhaida,S.T,M.T.  
Title : Analisa Sentimen Terhadap Pembelajaran Offline  
Di Era New Normal Menggunakan Algoritma SVM  
Dan Naïve Bayes

From the recent events where the spread of COVID-19 in Indonesia is getting wider, the Indonesian government has issued policies to prevent the spread. One of them is the policy applied to the field of education, where the teaching and learning process is carried out online or offline. Now, Indonesia has carried out the emergency response period, after which the government began to explore a new normal life or New Normal. Because of this, in a press release, the government allowed the implementation of face-to-face learning in the new school year during the Covid-19 pandemic while still paying attention to health protocols. This causes pros and cons in society. In this study, a comparison of the SVM Algorithm with Naive Bayes is carried out in analyzing sentiment regarding the implementation of offline learning in the new normal period based on community tweet data. The data used in this study were 2,708 data that had passed the preprocessing process, automatic labeling, resampling and TF-IDF. The results obtained using SVM precision, recall and F1-Score are 91%, 91%, 91% and 91.5% while with Naïve Bayes the results are 79%, 83%, 79% and 78.5% with a percentage split of 80% :20%. From these results it can be concluded that the SVM algorithm is better than Naive Bayes in analyzing sentiment regarding offline learning in the new normal era.

Key words:  
New Normal, SVM, Naive Bayes

## KATA PENGANTAR

Puji syukur atas kehadiran Tuhan Yang Maha Esa yang telah melimpahkan nikmat, serta hidayah-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul “Analisa Sentimen Terhadap Pembelajaran Offline Di Era New Normal Menggunakan Algoritma SVM Dan Naïve Bayes” dengan baik. Laporan Tugas Akhir ini merupakan salah satu persyaratan wajib untuk menyelesaikan Program Studi Strata Satu (S1) pada jurusan Informatika, Universitas Mercu Buana.

Dalam penyelesaian laporan Tugas Akhir ini, penulis mendapatkan banyak dukungan, bantuan, serta sumbangan ide maupun pikiran dari berbagai pihak. Untuk itu, penulis menyampaikan ucapan terimakasih kepada :

1. Ibu Dr. Ida Nurhaida,S.T,M.T., selaku dosen pembimbing Tugas Akhir yang telah meluangkan waktunya untuk memberikan bimbingan serta arahan dalam penyusunan Tugas Akhir ini.
2. Bapak Herry Derajat Wijaya, S.Kom., MM, selaku Ka. Prodi Teknik Informatika Universitas Mercu Buana.
3. Bapak Drs. Ahmad Kodar, MT, selaku Pembimbing Akademik yang telah membimbing dan membantu dalam memberikan informasi dan penjelasan mengenai proses akademik selama perkuliahan berlangsung.
4. Orang tua yang senantiasa memberikan doa dan dukungan.
5. Teman-teman yang selalu memberi semangat dan memberi motivasi kepada penulis selama pelaksanaan tugas akhir.
6. Semua pihak yang telah membantu dalam penyusunan Tugas Akhir ini yang tidak dapat disebutkan satu persatu.

Penulis menyadari bahwa penyusunan laporan Tugas Akhir ini masih jauh dari kesempurnaan dan masih terdapat banyak kekurangan. Walaupun demikian, penulis telah berusaha semaksimal mungkin untuk mendapatkan hasil yang terbaik. Penulis berharap tugas akhir ini dapat bermanfaat bagi pembaca guna menambah pengetahuan dan wawasan serta pembaca dapat memberikan kritik maupun saran.

Jakarta, 04 Juni 2021  
Azhar Fauzan

## 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 .....	xii
NASKAH JURNAL .....	1
KERTAS KERJA.....	13
BAB 1. LITERATUR REVIEW .....	14
BAB 2. SOURCE CODE .....	27
BAB 3. DATASET.....	40
BAB 4. TAHAPAN EKSPERIMEN.....	42
BAB 5. HASIL SEMUA EKSPERIMEN.....	47
DAFTAR PUSTAKA .....	51
LAMPIRAN DOKUMEN HAKI.....	53
LAMPIRAN KORESPONDENSI .....	54

## NASKAH JURNAL

■ |

### Analisa Sentimen Terhadap Pembelajaran *Offline* di Era *New Normal* Menggunakan Algoritma SVM dan Naïve Bayes

Azhar Fauzan<sup>1</sup>, Ida Nurhaida<sup>2</sup>

<sup>1,2</sup>Universitas Mercu Buana:

Jl. Meruya Selatan, Jakarta 11650, Indonesia

e-mail: <sup>1</sup>41517010010@student.mercubuana.ac.id, <sup>2</sup>ida.nurhaida@mercubuana.ac.id

#### Abstrak

Dari peristiwa yang terjadi belakangan ini dimana semakin luasnya penyebaran COVID-19 di Indonesia, pemerintah Indonesia mengeluarkan kebijakan-kebijakan guna mencegah penyebaran tersebut. Salah satunya adalah kebijakan yang diterapkan pada bidang pendidikan, dimana proses belajar mengajar dilakukan secara online atau daring. Sekarang, Indonesia telah melaksanakan masa tanggap darurat penanganan yang kemudian pemerintah mulai menjajaki kehidupan normal yang baru atau *New Normal*. Karena hal tersebut pada siaran pers pemerintah memperbolehkan kembali pelaksanaan pembelajaran tatap muka pada tahun ajaran baru di masa pandemi Covid-19 dengan tetap memperhatikan protokol kesehatan. Hal tersebut menyebabkan pro-kontra pada masyarakat. Pada penelitian ini dilakukan perbandingan Algoritma SVM dengan Naïve Bayes dalam menganalisa sentimen mengenai pelaksanaan pembelajaran *offline* di masa *new normal* berdasarkan data tweet masyarakat. Data yang digunakan pada penelitian ini sebanyak 2.708 data yang telah melewati proses *preprocessing*, *labeling* yang dilakukan secara otomatis, *resampling* dan *TF-IDF*. Hasil yang didapat menggunakan SVM *precision*, *recall* dan *F1-Score* sebesar 91%, 91%, 91% dan 91.5% sedangkan dengan Naïve Bayes diperoleh hasil sebesar 79%, 83%, 79% dan 78.5% dengan *percentage split* sebesar 80%:20%. Dari hasil tersebut dapat disimpulkan bahwa Algoritma SVM lebih baik dibandingkan dengan Naïve Bayes dalam menganalisa sentimen mengenai pembelajaran *offline* di era *new normal*.

**Kata kunci**— *New Normal*, SVM, Naïve Bayes

#### Abstract

From the recent events where the spread of COVID-19 in Indonesia is getting wider, the Indonesian government has issued policies to prevent the spread. One of them is the policy applied to the field of education, where the teaching and learning process is carried out online or online. Now, Indonesia has carried out the emergency response period, after which the government began to explore a new normal life or *New Normal*. Because of this, in a press release, the government allowed the implementation of face-to-face learning in the new school year during the Covid-19 pandemic while still paying attention to health protocols. This causes pros and cons in society. In this study, a comparison of the SVM Algorithm with Naïve Bayes is carried out in analyzing sentiment regarding the implementation of *offline* learning in the new normal period based on community tweet data. The data used in this study were 2,708 data that had passed the *preprocessing* process, *automatic labeling*, *resampling* and *TF-IDF*. The results obtained using SVM *precision*, *recall* and *F1-Score* are 91%, 91%, 91% and 91.5% while with Naïve Bayes the results are 79%, 83%, 79% and 78.5% with a *percentage split* of 80% :20%. From these results it can be concluded that the SVM algorithm is better than Naïve Bayes in analyzing sentiment regarding *offline* learning in the new normal era.

**Keywords**— *New Normal*, SVM, Naive Bayes

## 1. PENDAHULUAN

Setelah WHO menetapkan Covid-19 menjadi pandemic global, karena tingkat penyebarannya yang semakin luas dan cepat[1], guna mencegah hal tersebut pemerintah Indonesia mengeluarkan kebijakan-kebijakan selama masa pandemi Covid-19 berlangsung.

Salah satu kebijakan yang di keluarkan oleh Presiden Indonesia, Bapak Joko Widodo, yaitu dengan melakukan Pembatasan Sosial Berskala Besar (PSBB). Pembatasan ini berdampak pada ruang publik, termasuk kantor, sekolah dan juga kampus. Dalam pendidikan strategi ini melarang proses belajar tatap muka dan mengharuskan semua institusi pendidikan menghentikan proses belajar mengajar di kelas dan menggantinya dengan pembelajaran daring atau online [2].

Indonesia telah melaksanakan masa tanggap darurat penanganan dan PSBB selama kurang lebih 3 bulan, kemudian pemerintah Indonesia mulai menjajaki penerapan kehidupan normal yang baru atau New Normal dan melonggarkan PSBB [3].

Pada siaran Pers Nomor 137/sipers/A6/VI/2020, mengenai penyelenggaraan pembelajaran tahun ajaran dan tahun akademik baru di masa pandemi Covid-19 yang salah satu point pentingnya yaitu penyelenggaraan pembelajaran tatap muka boleh dilakukan, dengan ketentuan daerah berada pada zona hijau dan tetap memperhatikan protokol kesehatan [4].

Analisa sentimen merupakan salah satu cabang ilmu dari text mining, natural language program, dan artificial intelligence. Proses yang dilakukan oleh analisis sentimen untuk memahami, mengekstrak, dan mengolah data teks secara otomatis sehingga menjadi suatu informasi yang bermanfaat [5]. Isi dari teks atau konten sebagian besar adalah opini, sentiment, penilaian sikap dan emosi yang merupakan inti dari Analisa Sentimen [6]. Ada berbagai macam Algoritma yang dapat digunakan untuk melakukan Analisa Sentimen, seperti k-Nearest Neighbor, Naive Bayes dan lain sebagainya.

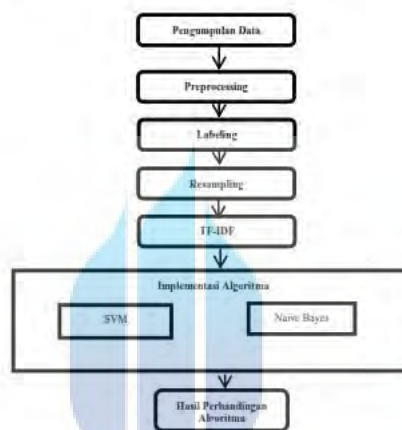
Support Vector Machine merupakan metode klasifikasi yang pada proses kerjanya menggunakan ruang hipotesis yang terdiri dari fungsi linear bersifat dua arah dalam sebuah ruang fitur yang berdimensi tinggi sehingga SVM pada umumnya selalu digunakan untuk pengklasifikasian data yang hanya memiliki dua kelas saja [7]. Pada penelitian yang dilakukan oleh [8] dapat disimpulkan bahwa pendeteksi sarkasme pada proses analisa sentiment dapat dilakukan dengan menggunakan metode multikelass SVM yaitu SVM OAO dan SVM OAA. Dari penelitian tersebut didapat hasil recall sebesar 92% dengan SVM OAA.

Naive Bayes merupakan metode yang digunakan untuk mengklasifikasi data berdasarkan kategori yang sudah ada [9]. Salah satu penelitian yang dilakukan oleh [10] membuktikan bahwa Algoritma Naive Bayes lebih unggul dibandingkan dengan SVM dalam mengklasifikasi opini public mengenai kandidat gubernur Jawa Barat periode 2018-2023, dengan hasil akurasi sebesar 94% sedangkan SVM sebesar 75.50%.

Dalam penelitian ini peneliti melakukan perbandingan metode klasifikasi yang dilakukan menggunakan Algoritma Support Vector Machine dan Naive Bayes dengan menggunakan dataset yang sama. Hasil yang didapat dari kedua metode klasifikasi tersebut akan dibandingkan untuk mengetahui nilai tingkat akurasi metode klasifikasi mana yang paling efektif dalam menganalisa sentimen. Penelitian ini juga bertujuan untuk membantu pemerintah dalam melihat tanggapan masyarakat mengenai pembelajaran tatap muka yang akan di lakukan kembali pada semester baru.

## 2. METODE PENELITIAN

Pada penelitian ini dilakukan perbandingan hasil kinerja algoritma klasifikasi Support Vector Machine (SVM) dan Naive Bayes pada analisa sentiment twitter. Rangkaian dari tahapan pada penelitian ini yaitu sebagai berikut, pada tahap pengumpulan data dilakukan dengan teknik crawling data menggunakan API twitter dengan bantuan tools rapid miner. Setelah pengumpulan data kemudian dilakukan proses *preprocessing*. Setelah data dipreprocessing kemudian data diberikan label atau *labelling*. Setelah data diberi label kemudian dilakukan proses *resampling*. Tahap selanjutnya adalah pembobotan menggunakan TF-IDF. Tahapan ini dilakukan untuk mengubah data yang berbentuk teks menjadi angka agar dapat diolah dan dibaca oleh komputer. Langkah selanjutnya adalah klasifikasi. Pada penelitian ini metode klasifikasi dilakukan dengan menggunakan algoritma Support Vector Machine (SVM) dan Naive Bayes yang kemudian dilakukan perbandingan hasilnya. Dibawah ini gambar tahapan penelitian yang dilakukan.



Gambar 1. Tahapan Penelitian

### 2.1 Pengumpulan Data

Pengumpulan data diambil melalui teknik crawling dengan API Twitter menggunakan bantuan tools Rapid Miner. Data yang diambil adalah data opini di Twitter yang dituliskan dalam Bahasa Indonesia Mulai tanggal 26 November 2020 – 01 Mei 2021. Dengan keyword Pembelajaran Offline, Pembelajaran Tatap Muka dan Pembelajaran Luring [11]. Data yang terkumpul sebanyak 2.514 data tweet. Contoh data tweet ada pada Tabel 1.

Tanggal	Text	Id
2020-11-27 02:40:11	Ditengah ketidak jelasan covid-19 ini, pemerintah malah membolehkan pembelajaran offline	133204651449178 1120
2020-11-26 22:44:18	"bsk januari sudah mulai pembelajaran offline". jiwa introvert & otakku : ????	133198715256187 2896

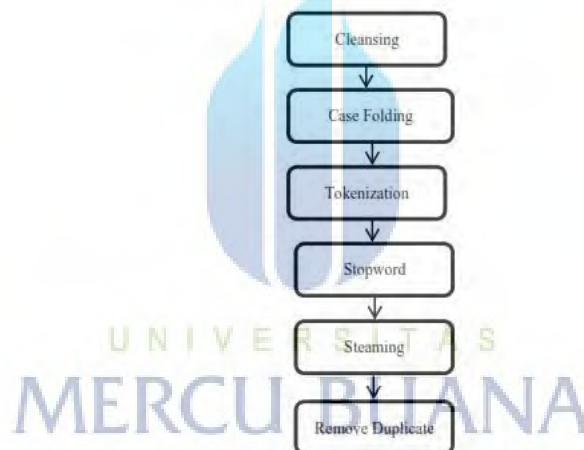
4

2021-05-01 12:13:14	Sambut Hardiknas, KPAI: Pembelajaran Tatap Muka Tidak Dapat Diseragamkan di Seluruh Indonesia <a href="https://t.co/WIPUUbkTYI">https://t.co/WIPUUbkTYI</a> <a href="https://t.co/TqAtZ92uEa">https://t.co/TqAtZ92uEa</a>	138836084576836 4032
2021-05-01 02:58:08	RT @baygajh: Bulan Juli Kita pembelajaran tatap muka ya Kalau tatap kamu nya aku tak bisa Takut di amuk Meta ?? ?? ?? ?? [ #blsd #bbrighvc...	138822114723420 1603

Tabel 1. Contoh dataset

## 2.2 Preprocessing

Proses *preprocessing* ini bertujuan untuk mengubah data mentah menjadi data yang siap digunakan untuk proses selanjutnya. *Preprocessing* yang dilakukan meliputi beberapa tahap, yaitu: *cleansing*, *case folding*, *tokenization*, *stopword*, *steaming* dan *remove duplicate*. *Preprocessing* juga mempunyai arti mengekstraksi kata-kata kunci dari data yang diberikan sedemikian rupa sehingga memberikan beberapa informasi yang bermakna atau analitis mengenai peristiwa, objek, atau hal yang diberikan. Data pra-pemrosesan berarti menghapus spasi, kata henti, simbol khusus (!, @, \*), Tag hash, kata duplikat, emoji. Membersihkan data penting karena tweet berisi banyak kata sintaksis yang tidak berguna untuk analisis. Setelah data dibersihkan, data tersebut digunakan untuk analisis data. Juga di media sosial kami dapat membagikan URL untuk mengidentifikasi tautan semacam itu dan menggantinya dengan teks atau tag URL [12].



Gambar 2. Preprocessing Data

- *Cleansing* adalah tahapan penghapusan tanda baca atau karakter yang tidak diperlukan seperti tanda baca koma (,), titik (.), tagar (#), Username dan karakter URL pada tweet. Contoh “ #PembelajaranTatapMukaBukaMulaiJuli ” menjadi “ PembelajaranTatapMukaBukaMulaiJuli ” [13].



- *Case Folding* pada tahapan ini adalah perubahan semua jenis huruf ke dalam huruf kecil [14].
- *Tokenization* Dalam penelitian ini memisahkan kalimat menjadi potongan kata disebut token, tokenize bisa juga untuk kata-kata yang tidak relevan akan dihapus, seperti kata untuk hanya dengan dan sebagainya yang merupakan kata-kata yang tidak mempunyai makna tersendiri jika dipisahkan dengan kata yang lain dan tidak terkait dengan kata sifat yang berhubungan dengan sentiment [15].
- *Stopword* Pada Proses *stopword removal* merupakan proses penghapusan term yang tidak memiliki arti atau tidak relevan [16].
- *Steaming* Setelah langkah tokenisasi, langkah selanjutnya adalah *stemming*. Langkah *stemming* adalah mengubah kata menjadi bentuk aslinya (yaitu, bentuk akar untuk mengurangi jumlah jenis atau kelas kata dalam data) [17].
- *Remove Duplicate* pada penelitian ini digunakan fitur tambahan yaitu *remove duplicate* yang bertujuan untuk menghapus data yang mempunyai kata-kata yang serupa atau duplikat.

Data yang telah dilakukan preprocessing menjadi 2.375 data. Dibawah ini merupakan beberapa contoh data yang telah dilakukan tahap preprocessing.

id	text	label
100	... ..	...
101	... ..	...
102	... ..	...
103	... ..	...
104	... ..	...
105	... ..	...
106	... ..	...
107	... ..	...
108	... ..	...
109	... ..	...
110	... ..	...
111	... ..	...
112	... ..	...
113	... ..	...
114	... ..	...
115	... ..	...
116	... ..	...
117	... ..	...
118	... ..	...
119	... ..	...
120	... ..	...

Gambar 3. Hasil Preprocessing

### 2.3 Labeling

Label berarti karakter atau himpunan karakter yang digunakan untuk mengidentifikasi suatu variabel atau bagian dari data atau berkas [16]. Pada tahap ini labeling dilakukan dengan dua skenario yaitu labeling otomatis dan labeling manual. Dilakukan dua skenario labeling ini untuk membandingkan labeling manakah yang lebih baik. Labeling otomatis data yang sudah bersih dan hanya memiliki kata yang bermakna diolah dengan bahasa R di google collaborative. Kemudian data yang sudah bersih dan bermakna dicocokkan ke dalam kamus corpus bahasa Indonesia sentiment negatif dan positif yang disiapkan sebelumnya dalam bentuk csv. Jika mempunyai kata yang cocok pada tweet dengan corpus sentimen positif maka diberi nilai 1, jika ada kata pada tweet yang cocok dengan corpus sentimen negatif maka diberi nilai negatif -1, dan jika tidak ditemukan kata tersebut di korpus negatif dan positif maka diberi nilai 0. Selanjutnya dilakukan penghitungan skor dengan menjumlahkan semua nilai dari kata. Jika skor bernilai  $<0$  maka tweet dilabeli negatif dan jika jumlah skor  $\geq 0$  maka tweet dilabeli positif. Ada berbagai Contoh dari kelas positif yang berisi pujian, saran, masukan, dan cerminan emosi positif seperti puas, senang, dan bahagia. Sedangkan kelas negatif berisi keluhan, kalimat sindiran, kritik, dan cerminan emosi negatif seperti amarah, kesal, dan kecewa [2]. Sedangkan labeling manual dilakukan dengan memberikan label pada setiap tweet. Kalimat yang berisi pujian akan diberi label positif, sedangkan untuk kalimat yang berisi kata kasar dan sarkas akan diberikan label negatif. Dibawah ini merupakan hasil dari dua skenario pemberian label. Untuk

gambar yang berada disebelah kiri merupakan hasil labeling otomatis, sedangkan untuk gambar yang berada disebelah kanan merupakan hasil labeling manual.

tweet	tweet	Klasifikasi	0	Positif	a ku kump jam jam
0	a ku kump jam jam	Positif	1	Negatif	a thread dia open gue pribadi apa eh emm leka
1	a thread dari omi gua pribadi apa eh emm leka...	Positif	2	Positif	aaa aks lgi udah sekolah tatap muka
2	aaa aks lgi udah sekolah tatap muka	Negatif	3	Negatif	aaa banget ga bimbel online offline bingung ki
3	aaa banget ga bimbel online offline bingung ki...	Negatif	4	Positif	aaa gua tatap muka
4	aaa gua tatap muka	Negatif	5	Negatif	aaaaa ga ga sekolah tatap muka maal ga jammel yg
5	aaaaa ga ga sekolah tatap muka maal ga jammel yg...	Negatif	6	Negatif	abis luring itu tdlr boles dondam semalem tidu
6	abis luring itu tdlr boles dondam semalem tidu...	Negatif	7	Negatif	abis ngit kursoner kin dang agi nyesei ya...
7	abis ngit kursoner kin dang agi nyesei ya...	Negatif	8	Negatif	absen genap pra us mik ipa nya tatap muka stress
8	absen genap pra us mik ipa nya tatap muka stress...	Negatif	9	Negatif	ada diem minggu gw tatap muka
9	ada diem minggu gw tatap muka	Negatif	10	Positif	acara diskusi publik bagi seni selinggara kul
10	acara diskusi publik bagi seni selinggara kul...	Positif	11	Positif	acara makan kartin sekolah mendokud radiem
11	acara makan kartin sekolah mendokud radiem...	Positif	12	Negatif	acara searung tuduh quina kokain searung hangka
12	acara searung tuduh quina kokain searung hangka...	Negatif	13	Positif	acu statement ide des rencana sekolah tatap mu...
13	acu statement ide des rencana sekolah tatap mu...	Negatif	14	Positif	ada insialff yg dibal sekolah tatap muka tik
14	ada insialff yg dibal sekolah tatap muka tik...	Positif	15	Positif	adadosen ku blang semester udah masuk dang...
15	adadosen ku blang semester udah masuk dang...	Positif	16	Positif	adek adek yg manz ajar tatap muka ysh moge su...
16	adek adek yg manz ajar tatap muka ysh moge su...	Positif	17	Negatif	adek ak sht mund sma d srg sama sht jarang s...
17	adek ak sht mund sma d srg sama sht jarang s...	Negatif	18	Negatif	adek gue temenya ga tatap kali ya pindah maget
18	adek gue temenya ga tatap kali ya pindah maget...	Negatif	19	Negatif	adek gue udah sekolah tatap muka masuk que anj...
19	adek gue udah sekolah tatap muka masuk que anj...	Negatif	20	Negatif	adek gw fs tatap muka ngakak bgt gw asa bgt pa...
20	adek gw fs tatap muka ngakak bgt gw asa bgt pa...	Negatif			

Gambar 4. Hasil Labeling

#### 2.4 Resampling

Resampling mempunyai dua cara untuk menyelesaikan masalahnya yaitu data dari kelas yang jumlahnya sedikit ditambah yaitu upsampling. Dan pemangkasan data yang berlebih dari suatu kelas yaitu downsampling. Upsampling dilakukan jika jumlah data yang diolah sedikit sehingga mempengaruhi pada model yang dibuat dan begitu pula sebaliknya pada downsampling[18].

Pada penelitian ini resampling yang digunakan yaitu Upsampling. Hal tersebut dikarenakan agar data yang di training tidak memiliki kecenderungan untuk memprediksi kepada salah satu kelas yang jumlah datanya lebih banyak.

#### 2.5 TF-IDF

Pembobotan Term Frequency-Inverse Document Frequency (TF-IDF) adalah salah satu proses dari teknik ekstraksi fitur dengan proses memberikan nilai pada masing-masing kata yang ada pada tweets latih (data latih). Untuk mengetahui seberapa penting sebuah kata mewakili sebuah kalimat, akan dilakukan pembobotan atau perhitungan. Pemberian skor dalam TF-IDF berdasarkan frekuensi munculnya kata dalam dokumen[19].

$$TF = \frac{w}{d} \quad IDF = \log \left( \frac{n}{df} \right) + 1 \quad TF-IDF = \left( \frac{w}{d} \right) \times \left( 1 + \log \left( \frac{n}{df} \right) \right)$$

$d$  = dokumen ke  $d$

$t$  = kata ke  $t$  dari kata kunci

$w$  = bobot dokumen ke  $d$  terhadap kata ke  $t$

$tf$  = Banyaknya kata yang dicari pada sebuah dokumen

$D$  = total dokumen

$df$  = banyaknya dokumen yang mengandung kata yang dicari

### 2.5 Klasifikasi

Pada penelitian menggunakan metode klasifikasi yang digunakan adalah algoritma Support Vector Machine (SVM) dan Naïve Bayes (NB). Support Vector Machine (SVM) adalah pengklasifikasi linier biner non probabilistik. SVM digunakan untuk menganalisis data vektor berdasarkan kata untuk menentukan hyperplane. Hyperplane terbaik dapat ditentukan dengan mengukur margin hyperplane dan menemukan titik maksimum. Margin adalah jarak antara hyperplane dengan pola terdekat dari tiap kelas. Pola terdekat disebut vektor dukungan [20]. Algoritma klasifikasi Naïve Bayes merupakan salah satu algoritma teknik klasifikasi, algoritma ini memanfaatkan metode probabilitas dan statistik. Algoritma ini merupakan salah satu metode klasifikasi yang populer dan termasuk dalam sepuluh algoritma teratas dalam data mining [21]. Metode Naïve Bayes adalah metode klasifikasi untuk penambangan teks yang digunakan dalam analisis sentimen. Pendekatan ini secara teoritis baik dalam hal konsistensi data dan klasifikasi perhitungan. Naïve Bayes biasanya digunakan dalam teknik klasifikasi, khususnya Twitter [22].

## 3. HASIL DAN PEMBAHASAN

Pada penelitian ini, peneliti berhasil melakukan eksperimen pengklasifikasian data yang diambil melalui media sosial twitter dengan bantuan tools Rapid Miner. Proses pengambilan data itu sendiri dilakukan dengan teknik crawling data. Setelah itu dilakukan preprocessing yang kemudian data diberikan label. Labeling pada penelitian ini dilakukan dengan dua cara yaitu otomatis dan manual, selesai proses labeling dilakukan proses resampling, TF-IDF. Pada penelitian ini digunakan algoritma SVM dan Naïve Bayes untuk mengklasifikasi data yang kemudian dibandingkan hasilnya. Penelitian ini juga digunakan untuk membantu pemerintah dalam melihat tanggapan masyarakat mengenai pembelajaran tatap muka yang akan di lakukan kembali pada semester baru.

Data yang digunakan pada penelitian ini sebanyak 2.514 data yang kemudian dilakukan *preprocessing*. Tahapan *preprocessing* yang dilakukan antara lain *cleansing*, *case folding*, *tokenization*, *stopword*, *stemming* dan *remove duplicate*. Data yang di didapat dari hasil *preprocessing* sebanyak 2375 data dengan data positif sebanyak 1021 dan data negatif sebanyak 1354 setelah dilakukan labeling otomatis, sedangkan untuk labeling manual data di dapatkan sebanyak 1359 label positif dan 1016 label negatif. Hasil dari pelebelan kemudian dilakukan proses resampling sehingga menghasilkan data akhir sebanyak 2.708 data untuk labeling otomatis dan 2.718 untuk labeling manual. Resampling dilakukan untuk meningkatkan hasil dari akurasi.

Pada penelitian ini dilakukan 3 skenario *percentage split*. Arti dari *percentage split* merupakan pemisahan data train dan data testing berdasarkan perbandingannya. Contoh nya 70% : 30% yang berarti 70% nya merupakan data training dan 30 % merupakan data testing. Pada penelitian ini dilakukan *percentage split* dengan 3 skenario yaitu 70% : 30%, 80% : 20 % dan 90% : 10%. Akurasi dari masing-masing skenario mempunyai hasil yang berbeda. Model yang telah terbentuk akan dilakukan evaluasi menggunakan konsep confusion matrix. Pada confusion matrix terdapat istilah TP, TN, FP, dan FN. TP berarti hasil prediksi kelas positif yang dinyatakan prediksinya benar dengan kelas aslinya, dan begitu juga pada TN untuk prediksi kelas negatif yang dinyatakan prediksinya benar sesuai kelas aslinya. Kemudian FP adalah hasil prediksi kelas positif yang dinyatakan salah prediksi atau tidak sesuai dengan kelas aslinya, dan begitu pula untuk FN pada kelas negatif. Setelah didapatkan jumlah dari TP, TN, FP, dan FN maka barulah model dapat diketahui tingkat akurasi.

TP = True Positif  
 TN = True Negatif  
 FP = False Positif  
 FN = False Negatif

$$\text{Akurasi} = \frac{TP+FN}{TP+TN+FP+FN} \times 100$$

Hasil kinerja model dari algoritma Naïve Bayes dan Support Vector Machine dengan menggunakan labeling otomatis pada skenario *percentage split* 70% : 30% untuk algoritma Naïve Bayes menghasilkan akurasi sebesar 78% sedangkan algoritma Support Vector Machine menghasilkan nilai akurasi sebesar 88%. Dari hasil kinerja percobaan tersebut SVM memiliki nilai akurasi yang lebih baik. Berikut merupakan hasil *percentage split* 70% : 30% dari masing-masing algoritma.

0.17645977613776133 [[395 11]] [[171 246]]					0.8014883488348 [[359 45]] [[ 46 355]]				
	precision	recall	f1-score	support		precision	recall	f1-score	support
0	0.77	0.91	0.84	404	0	0.83	0.98	0.90	402
1	0.98	0.58	0.73	407	1	0.88	0.82	0.85	405
accuracy			0.78	811	accuracy			0.88	813
macro avg	0.84	0.75	0.77	811	macro avg	0.86	0.88	0.88	813
weighted avg	0.83	0.74	0.77	811	weighted avg	0.88	0.88	0.88	813

Gambar 5. Naïve Bayes – SVM 70% : 30%

Hasil kinerja model dari algoritma Naïve Bayes dan Support Vector Machine dengan menggunakan labeling otomatis pada skenario *percentage split* 80% : 20% untuk algoritma Naïve Bayes memiliki nilai akurasi sebesar 79% sedangkan algoritma Support Vector Machine memiliki nilai akurasi sebesar 91%. Dari hasil kinerja percobaan tersebut, SVM memiliki nilai akurasi yang lebih baik. Berikut merupakan hasil kinerja *percentage split* 80% : 20% dari masing-masing algoritma.

0.1933579335793358 [[262 10]] [[188 369]]					0.9134301143301433 [[232 23]] [[ 25 262]]				
	precision	recall	f1-score	support		precision	recall	f1-score	support
0	0.72	0.96	0.83	272	0	0.90	0.91	0.91	255
1	0.98	0.62	0.78	270	1	0.92	0.91	0.92	287
accuracy			0.79	542	accuracy			0.91	542
macro avg	0.83	0.79	0.78	542	macro avg	0.91	0.91	0.91	542
weighted avg	0.83	0.79	0.78	542	weighted avg	0.91	0.91	0.91	542

Gambar 6. Naïve Bayes – SVM 80% : 20%

Hasil kinerja model dari algoritma Naïve Bayes dan Support Vector Machine dengan menggunakan labeling otomatis pada skenario *percentage split* 90% : 10% untuk algoritma Naïve Bayes memiliki nilai akurasi sebesar 78% sedangkan algoritma Support Vector Machine memiliki nilai akurasi sebesar 90%. Dari hasil kinerja percobaan tersebut, SVM memiliki nilai akurasi yang lebih baik. Berikut merupakan hasil kinerja *percentage split* 90% : 10% dari masing-masing algoritma.

0.1785977859778598 [[327 4]] [[ 58 363]]					0.8564785647856479 [[102 3]] [[ 19 102]]				
	precision	recall	f1-score	support		precision	recall	f1-score	support
0	0.69	0.87	0.81	131	0	0.84	0.92	0.88	753
1	0.95	0.60	0.74	140	1	0.94	0.88	0.91	149
accuracy			0.78	271	accuracy			0.90	271
macro avg	0.82	0.73	0.77	271	macro avg	0.89	0.90	0.89	271
weighted avg	0.81	0.73	0.77	271	weighted avg	0.89	0.90	0.89	271

Gambar 7. Naïve Bayes – SVM 90% : 10%

Hasil kinerja model dari algoritma Naïve Bayes dan Support Vector Machine dengan menggunakan labeling manual pada skenario *percentage split* 70% : 30% untuk algoritma Naïve Bayes menghasilkan akurasi sebesar 73% sedangkan algoritma Support Vector Machine menghasilkan nilai akurasi sebesar 85%. Dari hasil kinerja percobaan tersebut SVM memiliki nilai akurasi yang lebih baik. Berikut merupakan hasil *percentage split* 70% : 30% dari masing-masing algoritma.

0.7303921568627453 (1256 163) ( 51 346)					0.8529411764705882 (1352 85) ( 55 344)				
	precision	recall	f1-score	support		precision	recall	f1-score	support
0	0.83	0.60	0.70	119	0	0.86	0.74	0.80	417
1	0.67	0.87	0.76	397	1	0.84	0.66	0.75	399
accuracy			0.73	516	accuracy			0.82	816
macro avg	0.75	0.73	0.73	816	macro avg	0.85	0.70	0.77	816
weighted avg	0.75	0.73	0.73	816	weighted avg	0.85	0.69	0.76	816

Gambar 8. Naïve Bayes – SVM 70% : 30%

Hasil kinerja model dari algoritma Naïve Bayes dan Support Vector Machine dengan menggunakan labeling manual pada skenario *percentage split* 80% : 20% untuk algoritma Naïve Bayes memiliki nilai akurasi sebesar 78% sedangkan algoritma Support Vector Machine memiliki nilai akurasi sebesar 87%. Dari hasil kinerja percobaan tersebut, SVM memiliki nilai akurasi yang lebih baik. Berikut merupakan hasil kinerja *percentage split* 80% : 20% dari masing-masing algoritma.

0.7757922941176471 (1377 30) ( 52 245)					0.8676470588235294 (1056 34) ( 32 214)				
	precision	recall	f1-score	support		precision	recall	f1-score	support
0	0.85	0.66	0.74	261	0	0.83	0.78	0.80	280
1	0.73	0.88	0.80	277	1	0.86	0.65	0.74	284
accuracy			0.78	544	accuracy			0.87	564
macro avg	0.79	0.77	0.77	544	macro avg	0.87	0.72	0.81	544
weighted avg	0.79	0.78	0.77	544	weighted avg	0.87	0.67	0.77	544

Gambar 9. Naïve Bayes – SVM 80% : 20%

Hasil kinerja model dari algoritma Naïve Bayes dan Support Vector Machine dengan menggunakan labeling manual pada skenario *percentage split* 90% : 10% untuk algoritma Naïve Bayes memiliki nilai akurasi sebesar 77% sedangkan algoritma Support Vector Machine memiliki nilai akurasi sebesar 86%. Dari hasil kinerja percobaan tersebut, SVM memiliki nilai akurasi yang lebih baik. Berikut merupakan hasil kinerja *percentage split* 90% : 10% dari masing-masing algoritma.

0.7483823529411765 (121 21) ( 32 141)					0.8613705882352941 (1140 59) ( 18 81)				
	precision	recall	f1-score	support		precision	recall	f1-score	support
0	0.86	0.64	0.74	140	0	0.89	0.88	0.88	161
1	0.70	0.81	0.75	130	1	0.80	0.84	0.83	111
accuracy			0.77	272	accuracy			0.86	272
macro avg	0.78	0.77	0.77	272	macro avg	0.86	0.86	0.86	272
weighted avg	0.75	0.77	0.77	272	weighted avg	0.86	0.86	0.86	272

Gambar 10. Naïve Bayes – SVM 90% : 10%

Dari hasil diatas, dapat dilihat bahwa hasil perbandingan dua kinerja algoritma SVM dan Naïve Bayes dengan menggunakan *percentage split* 80%:20% lebih baik dari 70%:30% dan 90%:10%. Dibawah ini merupakan tabel hasil perbandinganya.

Algoritma	70 : 30	80 : 20	90 : 10
SVM	88%	91%	90%
NB	78%	79%	78%

**Tabel 2.** Perbandingan Algoritma SVM dan NB Labeling Otomatis

Algoritma	70 : 30	80 : 20	90 : 10
SVM	85%	87%	86%
NB	73%	78%	77%

**Tabel 3.** Perbandingan Algoritma SVM dan NB Labeling Manual

Dari hasil diatas dapat dilihat bahwa algoritma SVM lebih baik dibandingkan dengan algoritma Naïve Bayes dalam mengklasifikasi sentimen masyarakat mengenai pembelajaran *offline* di era new normal. Dengan *percentage split* 80% : 20% yang terbaik, hal tersebut karena proporsi data 80% dan 20% dari dataset pelatihan dan pengujian memiliki hasil yang lebih baik karena memberikan nilai evaluasi yang mendekati keseimbangan. Percobaan ini menunjukkan bahwa semakin banyak jumlah training dataset maka akan semakin baik nilai evaluasi yang akan diperoleh karena akan banyak proses pembelajaran yang terjadi pada training dataset.

#### 4. KESIMPULAN

Penelitian ini berhasil melakukan pengklasifikasian data sentiment masyarakat yang didapat dari media sosial twitter mengenai pembelajaran *offline* di era new normal. Klasifikasi yang dilakukan menggunakan algoritma SVM dan Naïve Bayes yang kemudian dilakukan perbandingan hasil klasifikasinya. Dari eksperimen yang dilakukan dapat disimpulkan bahwa hasil akurasi dari labeling otomatis lebih baik dibandingkan dengan labeling manual, dengan *percentage split* sebesar 80%:20% menjadi yang terbaik dibandingkan dengan 70%:30% dan 90%:10%. Dengan hasil akurasi, *precision*, *recall* dan *F1-Score* dari masing-masing algoritma Support Vector Machine sebesar 91%, 91%, 91% dan 91.5% sedangkan dengan Naïve Bayes diperoleh hasil sebesar 79%, 83%, 79% dan 78.5%. Dari hasil tersebut juga dapat diketahui bahwa algoritma SVM lebih baik dibandingkan dengan Naïve Bayes dalam mengklasifikasi sentimen masyarakat mengenai pembelajaran *offline* di era new normal. Dari hasil yang diperoleh dapat disimpulkan bahwa dari komentar atau pendapat masyarakat mengenai pembelajaran *offline* di era new normal ini masih terdapat pro dan kontra. Dari hal tersebut diharapkan dapat menjadi pertimbangan untuk pemerintah dalam melaksanakan pembelajaran *offline* atau tatap muka kedepanya.

#### 5. SARAN

Dari penelitian yang telah dilakukan, saran dari penulis untuk penelitian selanjutnya ialah dapat melakukan penambahan jumlah data yang akan digunakan serta metode yang lebih beragam seperti pada proses *preprocessing* maupun klasifikasinya, seperti menggunakan metode *deep learning*.

#### UCAPAN TERIMA KASIH

Terima kasih kepada kedua orang tua saya yang selalu mendukung serta memberi doa terbaik untuk saya. Kepada Ibu Dr. Ida Nurhaida, selaku pembimbing saya selama melakukan

penelitian ini. Terima kasih juga kepada rekan-rekan saya, yang sudah memberi support serta menjadi rekan diskusi saya selama penelitian ini dilakukan. Dan terima kasih sebesar-besarnya saya ucapkan kepada Tuhan Yang Maha Esa.

#### DAFTAR PUSTAKA

- [1] V. Ayumi and I. Nurhaida, "Klasifikasi Chest X-Ray Images Berdasarkan Kriteria Gejala Covid-19 Menggunakan Convolutional Neural Network," vol. 4, no. 2, pp. 147–153, 2021.
- [2] Samsir, Ambiyar, U. Verawardina, F. Edi, and R. Watrianthos, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes," *J. Media Inform. Budidarma*, vol. 5, pp. 157–163, 2021, doi: 10.30865/mib.v5i1.2604.
- [3] K. Perencanaan *et al.*, "Covid-19, New Normal, dan Perencanaan Pembangunan di Indonesia," *J. Perenc. Pembang. Indones. J. Dev. Plan.*, vol. 4, no. 2, pp. 240–252, 2020, doi: 10.36574/jpp.v4i2.118.
- [4] S. F. Nissa and A. Haryanto, "Implementasi Pembelajaran Tatap Muka Di Masa Pandemi Covid-19," *J. IKA PGSD (Ikatan Alumni PGSD) UNARS*, vol. 8, no. 2, p. 402, 2020, doi: 10.36841/pgsdunars.v8i2.840.
- [5] S. Bhanap and S. Kawthekar, "SENTIMENT ANALYSIS OF MOBILE DATASETS USING NAÏVE BAYES ALGORITHM," *Int. J. Adv. Res. Comput. Sci.*, vol. 9, no. 2, pp. 742–744, 2018.
- [6] K. Huda, "Classification Technique for Sentiment Analysis of Twitter Data," vol. 8, no. 5, pp. 2551–2555, 2017.
- [7] S. Styawati and K. Mustofa, "A Support Vector Machine-Firefly Algorithm for Movie Opinion Data Classification," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 13, no. 3, p. 219, 2019, doi: 10.22146/ijccs.41302.
- [8] D. Alita, "Multiclass SVM Algorithm for Sarcasm Text in Twitter," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 1, pp. 118–128, 2021, doi: 10.35957/jatisi.v8i1.646.
- [9] M. Wongkar and A. Angdresey, "Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter," *Proc. 2019 4th Int. Conf. Informatics Comput. ICIC 2019*, pp. 1–5, 2019, doi: 10.1109/ICIC47613.2019.8985884.
- [10] D. A. Kristiyanti, A. H. Umam, M. Wahyudi, R. Amin, and L. Marlinda, "Comparison of SVM Naïve Bayes Algorithm for Sentiment Analysis Toward West Java Governor Candidate Period 2018-2023 Based on Public Opinion on Twitter," *2018 6th Int. Conf. Cyber IT Serv. Manag. CITSM 2018*, no. Citsm, pp. 1–6, 2019, doi: 10.1109/CITSM.2018.8674352.
- [11] R. Ardiansyah, "Analisis Sentimen Calon Presiden Dan Wakil Presiden Periode 2019-2024 Pasca Debat Pilpres Di Twitter," *Sci. Comput. Sci. Informatics J.*, vol. 2, no. 1, p. 21, 2019, doi: 10.22487/j26204j18.2019.v2i1.13068.
- [12] P. Deshpande, "Review of Sentiment Analysis on Twitter Data Using Python," *Int. J. Adv. Res. Comput. Sci.*, vol. 8, no. 9, pp. 284–286, 2017, doi: 10.26483/ijarcs.v8i9.4956.
- [13] B. T. Dewo and I. Nurhaida, "Analisis Sentimen Twitter Terhadap Kepindahan Ibu Kota Menggunakan Metode Klasifikasi K-Nearest Neighbor, Decision Tree, dan Support Vector Machine Twitter Sentiment Analysis of Kepindahan Ibu Kota using of K-Nearest Neighbor, Decision Tree, and Suppo," pp. 1–7, doi: 10.14710/jtsiskom.x.x.xxxx.xx-xx.
- [14] M. R. A. Nasution and M. Hayaty, "Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter," *J. Inform.*, vol. 6, no. 2, pp. 226–235, 2019, doi: 10.31311/ji.v6i2.5129.
- [15] A. Taufik, "Optimasi Particle Swarm Optimization Sebagai Seleksi Fitur Pada Analisis Sentimen Review Hotel Berbahasa Indonesia Menggunakan Algoritma Naïve Bayes," *J. Tek. Komput. AMIK BSI*, vol. III, no. 2, pp. 40–47, 2017.

- 
- [16] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 4, no. 2, p. 113, 2018, doi: 10.26418/jp.v4i2.27526.
- [17] X. Zhang, H. Saleh, E. M. G. Younis, R. Sahal, and A. A. Ali, "Predicting Coronavirus Pandemic in Real-Time Using Machine Learning and Big Data Streaming System," *Complexity*, vol. 2020, 2020, doi: 10.1155/2020/6688912.
- [18] W. Wijanarto and S. P. Brilianti, "Peningkatan Performa Analisis Sentimen Dengan Resampling dan Hyperparameter pada Ulasan Aplikasi BNI Mobile," *J. Eksplora Inform.*, vol. 9, no. 2, pp. 140–153, 2020, doi: 10.30864/eksplora.v9i2.333.
- [19] P. P. A. Arsyia Monica Pravina, Imam Cholissodin, "Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2789–2797, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/4793>.
- [20] 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. 1–10, 2021, doi: 10.1088/1742-6596/1776/1/012058.
- [21] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.
- [22] Pristiyono, M. Ritonga, M. A. Al Ihsan, A. Anjar, and F. H. Rambe, "Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1088, no. 1, p. 012045, 2021, doi: 10.1088/1757-899x/1088/1/012045.



UNIVERSITAS  
MERCU BUANA



## KERTAS KERJA

### Ringkasan

Kertas kerja ini merupakan material kelengkapan artikel jurnal dengan judul “Analisa Sentimen Terhadap Pembelajaran Offline Di Era New Normal Menggunakan Algoritma SVM Dan Naïve Bayes”. 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 yang menjadi dasar atau landasan dalam penelitian ini. Bagian II menjelaskan mengenai source code yang digunakan pada penelitian ini. Bagian III menjelaskan mengenai dataset yang digunakan, meliputi penjelasan, cara perolehan data, parameter data, dan penyesuaian data akhir yang siap untuk diolah. Bagian IV memuat tahapan eksperimen yang disajikan dalam gambar dengan penjelasan dari setiap tahapan. Bagian V merupakan bagian terakhir dari kertas kerja ini yang menjelaskan hasil keseluruhan dari eksperimen yang telah dilakukan, meliputi penjelasannya.

