



**ANALISIS SENTIMEN BELANJA ONLINE PADA SOSIAL MEDIA
TWITTER SELAMA PANDEMIC COVID-19 DENGAN
MENGUNAKAN NAIVE BAYES DAN SUPPORT VECTOR MACHINE.**

TUGAS AKHIR

UNIVERSITAS
MERCU BUANA

Anrizas Sitepu
41519120034

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



**ANALISIS SENTIMEN BELANJA ONLINE PADA SOSIAL MEDIA
TWITTER SELAMA PANDEMIC COVID-19 DENGAN
MENGUNAKAN NAIVE BAYES DAN SUPPORT VECTOR MACHINE.**

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

UNIVERSITAS
MERCU BUANA

Oleh:
Anrizas Sitepu
41519120034

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

LEMBAR PERNYATAAN ORISINALITAS

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM : 41519120034

Nama : Anrizas Sitepu

Judul Tugas Akhir : Analisis Sentimen Belanja Online Pada Sosial Media
Twitter Selama Pandemic Covid-19 Dengan Menggunakan
Naive Bayes dan Support Vector Machine.

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

Jakarta, 07 Juli 2022



Anrizas Sitepu

UNIVERSITAS
MERCU BUANA

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

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

Nama Mahasiswa : Anrizas Sitepu
NIM : 41519120034
Judul Tugas Akhir : Analisis Sentimen Belanja Online Pada Sosial Media Twitter Selama Pandemic Covid-19 Dengan Menggunakan Naive Bayes dan Support Vector Machine.

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non Eksklusif** (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Non Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalih media/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

Jakarta, 27 Juli 2022



Anrizas Sitepu

SURAT PERNYATAAN LUARAN TUGAS AKHIR

SURAT PERNYATAAN LUARAN TUGAS AKHIR

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

Nama Mahasiswa : Anrizas Sitepu
NIM : 41519120034
Judul Tugas Akhir : Analisis Sentimen Belanja Online Pada Sosial Media Twitter Selama Pandemi Covid-19 Dengan Menggunakan Naive Bayes dan Support Vector Machine.

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	: Jurnal edukasi dan penelitian informatika	
	ISSN	: 2548-9364	
	Link Jurnal	: https://jurnal.untan.ac.id/index.php/jepin	
	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

Afiyati, S.Si., MT

Jakarta, 27 Juli 2022

Anrizas Sitepu

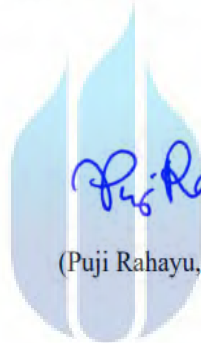
LEMBAR PERSETUJUAN PENGUJI

LEMBAR PERSETUJUAN PENGUJI

NIM : 41519120034
Nama : Anrizas Sitepu
Judul Tugas Akhir : Analisis Sentimen Belanja Online Pada Sosial Media Twitter Selama Pandemic Covid-19 Dengan Menggunakan Naive Bayes Dan Support Vector Machine.

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, 27 Juli 2022



(Puji Rahayu, Dr, MT)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

LEMBAR PERSETUJUAN PENGUJI

NIM : 41519120034
Nama : Anrizas Sitepu
Judul Tugas Akhir : Analisis Sentimen Belanja Online Pada Sosial Media Twitter Selama Pandemi Covid-19 Dengan Menggunakan Naive Bayes Dan Support Vector Machine.

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, 27 Juli 2022


(Wawan Gunawan, S.Kom, MT)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

LEMBAR PERSETUJUAN PENGUJI

NIM : 41519120034
Nama : Anrizas Sitepu
Judul Tugas Akhir : Analisis Sentimen Belanja Online Pada Sosial Media Twitter Selama Pandemi Covid-19 Dengan Menggunakan Naive Bayes Dan Support Vector Machine.

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, 27 Juli 2022



(Sabar Rudiarto, M.Kom)

UNIVERSITAS
MERCU BUANA

LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

NIM : 41519120034
Nama : Anrizas Sitepu
Judul Tugas Akhir : Analisis Sentimen Belanja Online Pada Sosial Media
Twitter Selama Pandemi Covid-19 Dengan Menggunakan
Naive Bayes Dan Support Vector Machine.

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, 27 Juli 2022

Menyetujui,



(Afiyati, S.Si., MT)
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

Puji syukur kita panjatkan panjatkan kepada ALLAH SWT yang telah memberikan nikmat dan rahmatnya sehingga penulis dapat menyelesaikan laporan tugas akhir ini yang berjudul “Analisis Sentimen Belanja Online Pada Sosial Media Twitter Selama Pandemic Covid-19 Dengan Menggunakan Naive Bayes dan Support Vector Machine. Tugas akhir ini disusun untuk memenuhi syarat memperoleh gelar sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.

Penulis menyadari tanpa bantuan dan bimbingan dari dosen pembimbing, orang tua serta teman-teman tidak akan terselesaikan dengan baik tugas akhir ini. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Ibu Afiyati,S.S., MT 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.
2. Ibu Dhanny Permatasari Putri, S.Kom, MT selaku dosen pembimbing akademik yang telah membimbing saya sejak awal mula semester hingga sampai saat ini dan selalu memberikan motivasi serta arahan agar dapat lulus tepat waktu
3. Orang tua yang selalu memberikan dukungan secara penuh dan doa sehingga dapat menyelesaikan semua laporan tugas akhir.
4. Teman-teman dan sahabat yang selalu memberikan dukungan dan motivasi dalam melakukan penulisan tugas akhir ini agar dapat terselesaikan dengan baik

Akhir kata, penulis berharap tugas akhir ini dapat menjadi bermanfaat bagi pembaca dan menambah wawasan pengetahuan semua pihak.

Jakarta, 27 April 2022

Penulis

XI

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	6
BAB 1. LITERATUR REVIEW	7
BAB 2. ANALISIS DAN PERANCANGAN	14
BAB 3. SOURCE CODE.....	18
BAB 4. DATASET	21
BAB 5. TAHAPAN EKSPERIMEN.....	23
BAB 6. HASIL SEMUA EKSPERIMEN	29
DAFTAR PUSTAKA	34
LAMPIRAN DOKUMEN HAKI	36
LAMPIRAN KORESPONDENSI	37



ANALISIS SENTIMEN BELANJA ONLINE PADA SOSIAL MEDIA TWITTER SELAMA PANDEMIC COVID-19 DENGAN MENGGUNAKAN NAIVE BAYES DAN SUPPORT VECTOR MACHINE.

Anrizas Sitepu^{#1}

[#]Jurusan Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana
Jl. Meruya Selatan No.1 Kembangan, Kota Jakarta Barat
*41519120034@student.mercubuana.ac.id

Abstrak— Analisis sentimen merupakan proses klasifikasi teks paling umum yang digunakan menganalisis pesan atau data untuk mendapatkan informasi sentimen dan memberi tahu apakah sentimen yang mendasarinya positif, negatif, netral. Umumnya masyarakat zaman sekarang menuangkan opini dengan membuat sebuah postingan di media social dengan berbagai macam topik, salah satu media social yang sering digunakan adalah Twitter. Topik yang diangkat pada penelitian ini adalah Analisis sentimen belanja online pada social media Twitter selama pandemi covid-19 menggunakan algoritma naive bayes dan Support Vector Machine. Pada penelitian ini data di ambil dari opini masyarakat melalui social media Twitter dengan melakukan proses pengumpulan data dengan memanfaatkan Twitter API dengan kata kunci "belanja online" kemudian melakukan pelabelan data manual yang akan di klasifikasikan menjadi 3 kelas. Dataset yang sudah dilabeli nantinya akan melewati tahapan preprocessing dan kemudian akan diklasifikasikan menjadi tiga kategori sentimen yaitu positif, negatif serta netral. Dari hasil pengujian dengan menerapkan metode klasifikasi naive bayes dan support vector machine untuk mengolah sentimen yang terdapat dalam suatu kalimat atau tweet secara otomatis yang dikategorikan dalam tiga kelas yaitu positif, negatif dan netral dengan jumlah data latih sebesar 400 dan data uji sebesar 100, didapatkan akurasi sebesar 72% untuk algoritma naive bayes dan 76% untuk algoritma support vector machine.

Kata kunci— Naive Bayes, Support vectors machine, Analisis sentimen, twitter, klasifikasi

I. PENDAHULUAN

Coronavirus Disease 2019 (Covid-19) adalah virus yang menyerang sistem pernafasan manusia dan menyebabkan kematian serta menyebar sangat cepat ke berbagai negara di dunia. Covid-19 pertama kali ditemukan di Kota Wuhan, Provinsi Hubei, China pada akhir Desember 2019. Virus Corona kemudian berkembang di enam puluh lima negara pada Februari 2020. Pada tanggal 30 Januari 2020 WHO telah menetapkan wabah ini sebagai ke darurat kesehatan masyarakat yang meresahkan dunia [1]. Pemerintah Indonesia telah melakukan berbagai cara untuk meminimalisir penyebaran virus corona, salah satunya yaitu program vaksinasi gotong royong, pembatasan sosial berskala besar (PSBB), dan aplikasi seperti peduli lindungi yang sangat membantu dalam mengurangi jumlah kasus positif di Indonesia. Program-program pemerintah ini tidak hanya berdampak ke ekonomi namun juga merubah kebiasaan dan gaya hidup masyarakat Indonesia salah satunya yaitu belanja online.

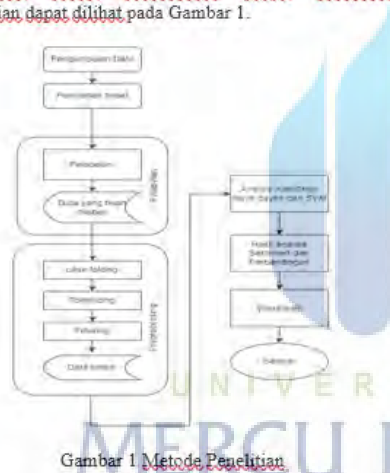
Belanja online di saat pandemic merupakan sebuah kebutuhan dan gaya hidup yang baru, namun hal ini menuai opini dari masyarakat dan mengekspresikannya di media social. Salah satu media sosial yang populer di masyarakat adalah Twitter. Twitter merupakan salah satu media sosial paling populer di Indonesia dan Indonesia merupakan negara dengan pengguna Twitter kelima teraktif di dunia [2]. Oleh sebab itu pada penelitian ini

penulis merasa perlu untuk membuat sebuah sistem menganalisa opini atau analisis sentimen sehingga bisa mengetahui dan dapat membantu memberikan informasi. Analisis sentimen merupakan proses untuk memahami dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini [3].

Penelitian ini akan melakukan analisis terhadap hasil pencarian tweet dengan kata kunci "belanja online" selama pandemi Covid19. Data yang diperoleh dari proses crawling. Crawling data di Twitter adalah suatu proses untuk mengambil atau mengunduh data dari server Twitter dengan bantuan Application Programming Interface[4] dan selanjutnya akan dibersihkan dan diklasifikasikan menjadi tiga kategori sentimen yaitu positif, negatif serta netral. Metode *Naive Bayes* dan *Support Vector Machine* yang digunakan dalam penelitian ini diharapkan mampu menghasilkan analisis data yang sesuai kategori dengan akurasi yang terbaik.

II. METODE PENELITIAN

Pada proses penelitian tahapan eksperimen yang digunakan secara keseluruhan dalam melakukan penelitian dapat dilihat pada Gambar 1.



Gambar 1 Metode Penelitian

Gambar diatas merupakan gambar dari metode penelitian yang dilakukan untuk mendapatkan hasil analisis dari algoritma *Naive Bayes* dan *Support Vector Machine*. Penulis akan memaparkan secara lebih jelas dari Gambar 1.

A. Pengumpulan Data

Data pada penelitian ini merupakan *tweet* berbahasa Indonesia yang di kumpulkan menggunakan API twitter dengan kata kunci "belanja online" dengan menggunakan Bahasa pemrograman python dengan bantuan library yang digunakan dalam pengumpulan data yaitu library *tweepy*

yang dapat mengakses API twitter. Dari proses pengumpulan data *tweet* melalui API tersebut diperoleh 3200 *tweet* yang di *crawling* dengan waktu yang berbeda. Crawling data di Twitter adalah suatu proses untuk mengambil atau mengunduh data dari server Twitter dengan bantuan Application Programming Interface (API) Twitter baik berupa data user maupun data *tweet* [5]. Pengumpulan data pertama kali di *crawling* pada bulan mei 2022, pengumpulan data yang kedua di awal bulan juni dan pengumpulan yang ketiga di akhir bulan juni 2022. Setelah *tweet* tersebut dikumpulkan selanjutnya penulis melakukan pembersihan duplikasi sehingga *tweet* yang dapat digunakan menjadi sebanyak 1640 data. Namun karena ketidakkonsistenan penulis dalam melakukan pelabelan maka dari itu penulis memutuskan untuk memilah *tweet* yang sudah dilakukan pelabelan menjadi 500 data dengan pelabelan yang sudah tepat.

Data yang tersimpan dari proses pengumpulan data terdiri dari tanggal pembuatan *tweet*, user Twitter, *id*, lokasi dan isi *tweet* mengenai belanja online yang di posting oleh pengguna Twitter, namun untuk melakukan proses klasifikasi data penulis hanya membutuhkan variable isi *tweet* untuk di analisa sentimennya.

B. Pelabelan Data

Setelah mengumpulkan data dan pemilahan *tweet* selesai dilakukan selanjutnya akan dilakukan pelabelan data secara manual pada setiap *tweet* yang dilakukan oleh penulis dan dengan bantuan 1 orang. Proses pelabelan dilakukan dengan memilih orientasi data dari *tweet* apakah bersifat positif, negatif dan netral. Pelabelan dilakukan untuk membentuk data latih dan dilakukan juga pelabelan pada data uji untuk menghitung akurasi dari hasil sentimen yang didapat secara otomatis.

Pada pelabelan data penulis membagi atau mengklasifikasikan data menjadi 3 kelas sentimen yaitu positif, negative dan netral. Dalam melakukan pelabelan penulis mengacu pada sentimen atau emosi yang tuliskan pada *tweet* yang telah dikumpulkan.

C. Preprocessing Data

Preprocessing merupakan proses untuk mempersiapkan data mentah sebelum dilakukan proses lain[6]. Tahapan preprocessing data perlu dilakukan karena beberapa kalimat *tweet* yang diperoleh tidak sepenuhnya menggunakan kata baku dan memiliki beberapa missing value, seperti hanya berisi emoji, dan juga tidak menggunakan bahasa indonesia yang baik dan benar. Selain meningkatkan akurasi, tahapan ini juga dapat memudahkan mesin dalam melakukan proses klasifikasi terhadap kalimat *tweet* yang ada. Preprocessing pada penelitian ini dilakukan menggunakan bantuan library pada bahasa pemrograman Python dan di bantu dengan menggunakan library pada python yaitu NLTK, dan Re. Preprocessing data dilakukan dengan beberapa tahap diantaranya, Case Folding, Tokenizing, Filtering,

Stemming yang akan dijelaskan lebih lanjut pada tahapan preprocessing data.

Berikut ini penjelasan lebih lanjut mengenai tahapan preprocessing data.

1) Case folding

Case folding adalah mengubah setiap kalimat tweet didalam teks menjadi huruf kecil. Tidak semua kata dalam teks konsisten dalam penggunaan huruf kapital [7] disinilah tujuan dilakukan case folding untuk mengkonversi setiap karakter dalam kata menjadi huruf kecil. Berikut merupakan contoh data penelitian dari case folding.

TABEL I
CASE FOLDING

Teks	Hasil
@CNNIndonesia: Menenal Bea Meterai Rp10 Ribu untuk Belanja Online https://t.co/Vfw1B5nZRw	@cnindonesia: menenal bea meterai rp10 ribu untuk belanja online https://t.co/vfw1b5nzw

2) Cleansing

Pada proses cleansing semua kalimat dibersihkan dari hashtag, emoji dan juga tanda baca.

TABEL II
CLEANSING

Teks	Hasil
@CNNIndonesia: Menenal Bea Meterai Rp10 Ribu untuk Belanja Online https://t.co/Vfw1B5nZRw	cnindonesia: menenal bea meterai rp10 ribu untuk belanja online

3) Tokenizing

Tokenisasi merupakan proses pemecahan kata pada suatu teks ke dalam satuan kata. Tokenisasi dilakukan untuk menghasilkan kumpulan kata yang berdiri sendiri. tokenisasi memecah teks yang semula berupa kalimat menjadi kata-kata. tokenisasi menghilangkan delimeter seperti titik (.), koma (,), spasi, dan karakter angka yang ada pada kata tersebut [8]. Dalam penelitian ini tokenisasi dilakukan untuk memecah sebuah kalimat menjadi kata perkata apabila di temukan suatu karakter pemisah kata serta melakukan penghapusan delimeter beserta karakter angka bersama tweet entity seperti hashtag, retweet dan mention. Berikut merupakan contoh hasil dari tokenisasi dapat dilihat pada table III.

TABEL III
TOKENIZING

Teks	Hasil
@cnindonesia: menenal bea meterai rp10 ribu untuk belanja online https://t.co/vfw1b5nzw	['menenal', 'bea', 'meterai', 'rp10', 'ribu', 'untuk', 'belanja', 'online']

4) Filtering

Filtering merupakan proses dalam text preprocessing setelah tokenisasi. filtering dilakukan untuk untuk mengambil kata penting hasil tokenisasi. Pada tahap filtering kata akan ditentukan apakah akan digunakan atau dibuang. Proses dalam filtering dalam membuang kata kata yang tidak digunakan atau dibuang [9]. Proses dalam filtering dalam membuang kata kata yang tidak digunakan atau stopword merupakan daftar kata-kata yang tidak mempresentasikan isi dari suatu dokumen teks. stopword dilakukan untuk menghilangkan kata atau term yang tidak memiliki arti. Daftar stoplist akan dibuat sebelum melakukan proses stopword removal, jika kata-kata terdapat dalam daftar stoplist, maka kata tersebut akan dihapus, sehingga kata-kata yang tersisa akan dianggap kata yang mencirikan isi suatu dokumen.

TABEL IV
FILTERING

Teks	Hasil
['menenal', 'bea', 'meterai', 'rp10', 'ribu', 'untuk', 'belanja', 'online']	menenal bea meterai ribu belanja online

Pada penelitian ini penulis tidak melakukan stemming dikarenakan pada kasus ini menggunakan stemming akan menurunkan tingkat akurasi dari algoritma yang telah di buat.

D. Klasifikasi Naive Bayes dan Support Vector Machine

Klasifikasi teks bisa didefinisikan sebagai proses untuk menentukan suatu dokumen teks ke dalam suatu kelas tertentu [10]. Dalam penelitian ini data yang sudah bersih akan di klasifikasikan ke dalam 3 kelas yaitu positif,netral, dan negatif dengan menggunakan algoritma naive bayes dan support vector machine.

Metode naive bayes atau sering disebut sebagai naive bayes classification, merupakan salah satu metode yang dapat mengklasifikasikan teks. Kelebihan Naive Bayes adalah sederhana tetapi memiliki akurasi yang tinggi [11].

Support Vector Machine (SVM) merupakan salah satu metode klasifikasi dengan menggunakan machine learning (supervised learning) yang memprediksi kelas berdasarkan model atau pola dari hasil proses training. Klasifikasi dilakukan dengan mencari hyperplane atau garis pembatas (decision boundary) yang memisahkan antara suatu kelas dengan kelas lain, yang dalam kasus ini garis tersebut berperan memisahkan tweet bersentimen positif (berlabel +1) dengan tweet bersentimen negatif (berlabel -1) [12].

E. Hasil Analisa Sentimen dan Perbandingan

Untuk mengetahui performa dari Algoritma Naive Bayes dan Support Vector Machine akan dilakukan

pengujian terhadap model yang telah dibuat. Hasil klasifikasi akan divisualisasi dalam bentuk confusion matrix. Confusion matrix merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi [13]. Nantinya pada penelitian ini juga akan dibandingkan performa dari klasifikasi *naive bayes* dan *support vector machine* yang mencakup nilai akurasi, *recall*, dan *precision*. Akurasi merupakan persentase dari total dokumen yang teridentifikasi secara tepat dalam proses klasifikasi [14]. Semakin besar nilai akurasinya, maka semakin baik performa dari model algoritma yang dibuat. Berikut persamaannya:

$$Akurasi = \frac{TP+TN}{(TP+FP+TN+FN)} \quad (1)$$

recall adalah nilai yang menunjukkan tingkat keberhasilan untuk mengetahui kembali sebuah informasi secara benar tentang data yang kelas negatif atau pun konten teks positif [15]. Berikut merupakan persamaan *recall*:

$$Recall = \frac{TP}{TP+FP} \quad (2)$$

Precision merupakan nilai perbandingan jumlah data yang diprediksi benar dibanding dengan keseluruhan data yang telah di klasifikasi. Berikut merupakan persamaan *precision*:

$$Precision = \frac{TP}{TP+FN} \quad (3)$$

F. Visualisasi

Pada penelitian ini penulis mencoba untuk divisualisasikan dan memprediksi kata baru dari model yang telah dibuat ke dalam aplikasi sederhana berbasis website yang bisa memprediksi kata baru ke dalam kelas sentimen positif, netral, dan negatif.

III. HASIL DAN PEMBAHASAN

A. Pengujian Algoritma

Pada penelitian ini dilakukan pengujian terhadap klasifikasi sentimen yang telah dilakukan. Perbandingan dataset yang dilakukan yaitu data latih 80% dan data uji 20% dari 500 data dengan metode klasifikasi *naive bayes* dan *support vector machine*. Awalnya penulis mencoba melakukan pengujian dengan menggunakan 1600 data yang telah dilabeli secara manual. Namun dikarenakan kurangnya pemahaman tentang pelabelan dan ketidakkonsistenan pada saat melakukan pelabelan 1640 data dan penulis menemukan banyak kesalahan pada data yang sudah dilabeli, dikarenakan kelas netral mengakibatkan bias sehingga menyebabkan tingkat akurasi data yang di klasifikasi menjadi rendah sekitar 60% untuk algoritma *naive bayes* dan 66% untuk algoritma *support vector machine*. Maka dari itu penulis memutuskan untuk memilah tweet yang sudah dilakukan

pelabelan menjadi 500 data dengan pelabelan yang sudah tepat. Setelah dilakukan pemilahan tweet menjadi 500 data dan dilakukan pelatihan dan pengujian. Pengujian pada penelitian ini dilakukan dengan menggunakan confusion matrix dengan ukuran 3x3 sebagai representatif kelas aktual dan kelas prediksi. Berdasarkan hasil pengujian yang dilakukan, penelitian ini menghasilkan confusion matrix sebagai berikut:

TABEL V
HASIL CONFUSION MATRIX SVM

Data aktual	Data prediksi		
	Negatif	Netral	Positif
Negatif	25	3	2
Netral	3	22	7
Positif	5	4	29

TABEL VI

HASIL CONFUSION MATRIX NAIVE BAYES

Data aktual	Data prediksi		
	Negatif	Netral	Positif
Negatif	22	2	6
Netral	6	16	10
Positif	3	1	34

Tabel diatas memperlihatkan Confusion Matrix dengan ukuran 3 x 3 yang mewakili setiap kelas yaitu negatif, netral, dan positif. Dari confusion matrix pada tabel V dapat dijelaskan bahwa model mengklasifikasikan secara benar sebesar data 25 sebagai negatif, 22 data sebagai netral, dan 29 sebagai data positif. Selain itu model salah dalam memprediksi, 2 data ke dalam kelas positif yang seharusnya negatif (False Negatif), serta salah dalam memprediksi 5 data ke dalam kelas negatif yang seharusnya positif (False positif).

TABEL VII
HASIL ANALISA SENTIMEN

Algoritma	Sentimen	Precision (%)	Recall (%)	F1 Score (%)	Accuracy (%)
SVM	Positif	76	83	79	76
	Negatif	76	69	72	
	Netral	76	76	76	
Naive Bayes	Positif	71	73	72	72
	Negatif	84	50	63	
	Netral	68	89	77	

Pengujian ini menghasilkan nilai akurasi 72% untuk algoritma *naive bayes* dan 76% untuk *support vector machine*, mengalami peningkatan akurasi dibandingkan dengan data sebelum dilakukan pemilihan tweet. Hasil akurasi pada penelitian ini dapat dilihat pada tabel VII

B. ANALISA PENGUNJIAN ALGORITMA

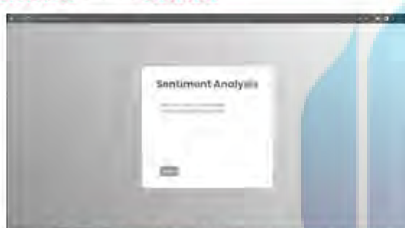
Pada tabel VII dapat dilihat bahwa algoritma Support Vector Machine memperoleh hasil yang lebih baik dengan tingkat akurasi rata-rata sebesar 76% dengan tingkat rata-

rata precision, recall, dan f1-score sebesar 76%, 76%, 75.6%. Sedangkan algoritma naive bayes hanya memperoleh akurasi sebesar 72% dengan tingkat rata-rata precision, recall, dan f1-score sebesar 74.3%, 70.6%, dan 70.6%.

Pada confusion matrix table V dan VI dapat diketahui nilai true positif dan true negatif. True positif adalah nilai data positif yang diklasifikasi dengan benar sesuai kelasnya yaitu positif. True negatif adalah nilai data sentimen yang diklasifikasi dengan benar sesuai kelas sentimennya, yaitu negatif. Dari tabel VII dapat disimpulkan bahwa algoritma *support vector machine* memiliki hasil analisa sentimen yang lebih baik dibandingkan dengan algoritma *naive bayes*.

C. Visualisasi

Setelah dilakukan proses pembersihan data dan klasifikasi algoritma selanjutnya didapatkan hasil sentimen. Selanjutnya akan dilakukan prediksi kata baru dan akan ditampilkan pada sebuah halaman web sederhana. Visualisasi dan prediksi sentimen pada penelitian ini dibuat dengan bantuan library python flask yang menyediakan libraries yang bisa digunakan untuk membangun website sederhana.



Gambar 2. Tampilan awal dan input data untuk di prediksi

Pada halaman ini pengguna dapat melakukan input data atau *tweet* sesuai keinginan namun tingkat keakuratan prediksi tergantung dari akurasi model yang telah dibuat. Penulis mencoba untuk melakukan pengujian prediksi data sentimen dengan menulis sebuah *tweet* yang berisi "Miskin amat negara ini sampai belanta pun di materai kemana kekayaan alam ini" dan kemudian model akan memprediksi dan menampilkan sentimen seperti pada gambar 3.



Gambar 3. Hasil prediksi sentimen

IV. KESIMPULAN

Berdasarkan dari penelitian yang sudah dilakukan maka penulis dapat menarik kesimpulan bahwasanya algoritma *support vector machine* memiliki hasil analisa sentimen yang lebih baik dalam melakukan analisis sentimen belanja online pada sosial media Twitter selama pandemic covid-19 dibandingkan dengan algoritma *Naive Bayes*. Hal itu dibuktikan dengan tingkat akurasi yang di dapatkan oleh algoritma *support vector machine* lebih tinggi yaitu dengan rata-rata sebesar 76% dengan tingkat rata-rata precision, recall, dan f1-score sebesar 76%, 76%, 75.6%. Sedangkan algoritma naive bayes hanya memperoleh akurasi sebesar 72% dengan tingkat rata-rata precision, recall, dan f1-score sebesar 74.3%, 70.6%, dan 70.6%.

- [1] N. K. Astini, Sari, "Pemanfaatan Teknologi Informasi dalam Pembelajaran Tingkat Sekolah Dasar pada Masa Pandemi Covid-19," *J. Lemb. Perencanaan Mutu STKIP Agama Hindu Anlapawa*, vol. 11, no. 2, pp. 13–25, 2020.
- [2] T. T. Widawati and M. Sadikin, "Analisis Sentimen Twitter terhadap Tokoh Publik dengan Algoritma Naive Bayes dan Support Vector Machine," *Simeris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 11, no. 2, pp. 626–636, 2021, doi: 10.24176/simet.v11i2.4568.
- [3] L. Ardiani, H. Sujani, and T. Tursina, "Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak," *J. Sist. dan Teknol. Inf.*, vol. 8, no. 2, p. 183, 2020, doi: 10.26418/justin.v8i2.36776.
- [4] B. Mas Puntoko and K. Muslim, "Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier," *e-Processing Eng.*, vol. 5, 2018.
- [5] J. Eka Sembodo, E. Budi Setiawan, and Z. Abdurrahman Baizal, "Data Crawling Otomatis pada Twitter," no. September, pp. 11–16, 2016, doi: 10.21108/indose.2016.111.
- [6] T. W. U. Imelya Susianti, Sri Soerya Ningsih, M. Al Haris, "Analisis Sentimen Pada Twitter Terkait New Normal Dengan Metode Naive Bayes Classifier," *Proc. Semin. Educaintech FMIPA LNMUS*, pp. 354–363, 2020, [Online]. Available: <https://prosiding.unimus.ac.id/index.php/edusaintek/article/view/576578>
- [7] S. Adi, "Perancangan Klasifikasi Tweet Berdasarkan Sentimen Dan Fitur Calon Gubernur DKI Jakarta 2017," *J. Inform. Pelita Nusantara*, vol. 3, no. 1, pp. 10–16, 2018.
- [8] D. Pakpahan and H. Widyastuti, "Aplikasi Opinion Mining dengan Algoritma Naive Bayes untuk Menilai Berita Online," *J. Integr.*, vol. 6, no. 1, pp. 1–10, 2014, [Online]. Available: <https://sci-hub.do/https://jurnal.polibatam.ac.id/index.php/JI/article/view/105>
- [9] E. E. Pratama and B. R. Trilaksono, "Klasifikasi Topik Keluhan Pelanggan Berdasarkan Tweet dengan Menggunakan Penggabungan Feature Hasil Ekstraksi pada Metode Support Vector Machine (SVM)," *J. Edukasi dan Penelit. Inform.*, vol. 1, no. 2, 2015, doi: 10.26418/jp.v1i2.11023.
- [10] A. Fathan Hidayatullah, M. Rifqi Ma, and arif Program Studi Manajemen Informatika STMIK Jenderal Achmad Yani Yogyakarta Jl Ringroad Barat, "Penerapan Text Mining dalam Klasifikasi Judul Skripsi," *Semin. Naz. Apl. Teknol. Inf. Agustus*, pp. 1907–5022, 2016.
- [11] A. Pranzono, R. Indriati, and A. Nugroho, "Sentiment Analysis Tokoh Politik pada Twitter," *Semin. Naz. Inov. Teknol.*, pp. 195–200, 2017.
- [12] L. Septiani and Y. Sibaroni, "Sentiment Analysis Terhadap Tweet Bermada Sarkasme Berbahasa Indonesia," *J. Linguist. Komputasional*, vol. 2, no. 2, pp. 62–67, 2019.
- [13] F. Satria, Z. Zamharri, and M. A. Syaripudin, "Prediksi

KERTAS KERJA

Ringkasan

Kertas kerja ini merupakan material kelengkapan artikel jurnal dengan judul di atas. Kertas kerja berisi semua material hasil penelitian Tugas Akhir yang tidak dimuat/atau disertakan di artikel jurnal. Di dalam kertas kerja ini tertera: literature review, Data Set, tahapan eksperimen, source code, dan hasil eksperimen secara keseluruhan. Literatur review berisi tentang jurnal-jurnal pendukung yang terkait dengan penelitian analisis sentimen yang telah dibuat. Pada bab ini dijelaskan masing-masing judul jurnal yang dikutip, topik yang dibahas pada setiap jurnal yang dikutip, metode yang digunakan dalam penelitian tersebut, dan hasil yang didapatkan dari pengujian algoritma yang digunakan. Pada bab Data Set dilampirkan variabel-variabel yang digunakan sebagai landasan pengujian algoritma, Data tersebut terbagi menjadi dua yaitu data latih dan data uji. Pada bab ini, dilampirkan pula sedikit sampel data yang digunakan dalam penelitian. Tahapan eksperimen merupakan bab yang membahas tentang eksperimen yang akan dilakukan untuk analisis sentimen data komentar menggunakan metode Support Vector Machine dan Naïve Bayes. Pada bab source code disajikan beberapa potongan code yang digunakan untuk melakukan pengujian algoritma dengan menggunakan bahasa program Python dan beberapa library pendukungnya. Kemudian yang terakhir yaitu bab hasil eksperimen secara keseluruhan membahas hasil yang didapat saat melakukan pengujian algoritma baik dari segi akurasi, presisi, recall dan f1-score. Selain itu, dijabarkan pula hasil dari perbandingan kedua algoritma yang digunakan tersebut dalam melakukan analisa sentimen sampai pada penarikan kesimpulan algoritma yang terbaik.

UNIVERSITAS
MERCU BUANA