



**PERBANDINGAN ALGORITMA SVM DAN NAÏVE BAYES DALAM
ANALISA SENTIMEN TERHADAP PENGGUNA *TWITTER* MENGENAI
VAKSIN *BOOSTER***

TUGAS AKHIR

Muhammad Ramzy Alfinrizq Ramadhan
41518010042

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



**PERBANDINGAN ALGORITMA SVM DAN NAÏVE BAYES DALAM
ANALISA SENTIMEN TERHADAP PENGGUNA *TWITTER* MENGENAI
VAKSIN *BOOSTER***

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:

Muhammad Ramzy Alfinrizq Ramadhan
41518010042

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

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini: .

NIM : 41518010042

Nama : Muhammad Ramzy Alfinrizq Ramadhan

Judul Tugas Akhir : PERBANDINGAN ALGORITMA *SVM* DAN *NAÏVE BAYES* DALAM ANALISA SENTIMEN TERHADAP PENGGUNA *TWITTER* MENGENAI VAKSIN *BOOSTER*

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, 11 Agustus 2022



Muhammad Ramzy Alfinrizq Ramadhan

UNIVERSITAS
MERCU BUANA

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

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

Nama Mahasiswa : Muhammad Ramzy Alfinrizq Ramadhan
NIM : 41518010042
Judul Tugas Akhir : PERBANDINGAN ALGORITMA *SVM* DAN *NAÏVE BAYES* DALAM ANALISA SENTIMEN TERHADAP PENGGUNA *TWITTER* MENGENAI VAKSIN *BOOSTER*

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, 11 Agustus 2022



Muhammad Ramzy Alfinrizq Ramadhan

SURAT PERNYATAAN LUARAN TUGAS AKHIR

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

Nama Mahasiswa : Muhammad Ramzy Alfinrizq Ramadhan
NIM : 41518010042
Judul Tugas Akhir : PERBANDINGAN ALGORITMA SVM DAN NAÏVE BAYES DALAM ANALISA SENTIMEN TERHADAP PENGGUNA TWITTER MENGENAI VAKSIN BOOSTER

Menyatakan bahwa :

1. Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis		Status		
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi		Diajukan	✓	
		Jurnal Nasional Terakreditasi	✓			
		Jurnal International Tidak Bereputasi		Diterima		
		Jurnal International Bereputasi				
	Disubmit/dipublikasikan di :	Nama Jurnal	: Jurnal Informatika: Jurnal Pengembangan IT			
		ISSN	: 2477-5126			
		Link Jurnal	: ejournal.poltektegal.ac.id/index.php/informatika			
		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.

Jakarta, 11 Agustus 2022



Muhammad Ramzy Alfinrizq Ramadhan

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010042
Nama : Muhammad Ramzy Alfinrizq Ramadhan
Judul Tugas Akhir : PERBANDINGAN ALGORITMA *SVM* DAN *NAÏVE BAYES* DALAM ANALISA SENTIMEN TERHADAP PENGGUNA *TWITTER* MENGENAI VAKSIN *BOOSTER*

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, 11 Agustus 2022



LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010042
Nama : Muhammad Ramzy Alfinrizq Ramadhan
Judul Tugas Akhir : PERBANDINGAN ALGORITMA *SVM* DAN
NAÏVE BAYES DALAM ANALISA SENTIMEN
TERHADAP PENGGUNA *TWITTER* MENGENAI
VAKSIN BOOSTER

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, 11 Agustus 2022



(Anis Cherid, SE, MTI)



UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010042
Nama : Muhammad Ramzy Alfinrizq Ramadhan
Judul Tugas Akhir : PERBANDINGAN ALGORITMA *SVM* DAN
NAÏVE BAYES DALAM ANALISA SENTIMEN
TERHADAP PENGGUNA *TWITTER* MENGENAI
VAKSIN BOOSTER

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, 11 Agustus 2022



(Ir. Emil R. Kaburuan, Ph.D., IPM.)



UNIVERSITAS
MERCU BUANA

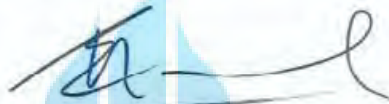
LEMBAR PENGESAHAN

NIM : 41518010042
Nama : Muhammad Ramzy Alfinrizq Ramadhan
Judul Tugas Akhir : PERBANDINGAN ALGORITMA *SVM* DAN *NAÏVE BAYES* DALAM ANALISA SENTIMEN TERHADAP PENGGUNA *TWITTER* MENGENAI VAKSIN *BOOSTER*

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, 11 Agustus 2022

a.n. Menyetujui,

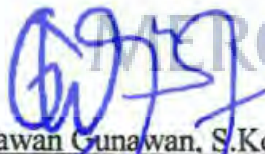


(Dr. Leonard Goeirmanto, ST, M.Sc)
Dosen Pembimbing

Mengetahui,

UNIVERSITAS

MERCU BUANA



(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 kehadir Allah SWT yang telah memeberikan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul “PERBANDINGAN ALGORITMA *SVM* DAN *NAÏVE BAYES* DALAM ANALISA SENTIMEN TERHADAP PENGGUNA *TWITTER* MENGENAI VAKSIN *BOOSTER*” untuk memenuhi syarat memperoleh gelar sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana dalam jangka waktu yang telah ditentukan.

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

1. Bapak Dr. Leonard Goeirmanto, ST, M. Sc selaku dosen pembimbing Tugas Akhir yang telah memberikan masukan serta semangat dan meluangkan waktu sebagian besarnya untuk melakukan bimbingan dalam penyusunan tugas akhir dari awal hingga selesai.
2. Bapak Dr. Harwikarya, MT selaku dosen pembimbing akademik yang telah membimbing sejak awal penulis berkuliah hingga saat ini dan selalu memberikan motivasi dan semangat agar dapat lulus tepat waktu.
3. Seluruh Dosen Program Studi Teknik Informatika, Fakultas Ilmu Komputer, dan Universitas Mercu Buana yang telah memberikan ilmu yang bermanfaat selama berkuliah.
4. Seluruh Staff Administrasi dan Tata Usaha yang telah banyak membantu atas pelayanan dan arahannya.
5. Orang tua serta keluarga penulis yang selalu memberikan dukungan, motivasi, dan semangat dalam berkuliah dan menyelesaikan tugas akhir.
6. Teman-teman dan sahabat yang selalu memberikan dukungan dalam mengerjakan tugas akhir agar dapat terselesaikan dengan baik.

Akhir kata, penulis berharap tugas akhir ini dapat menjadi referensi untuk dikembangkan menjadi lebih baik, karena penulis menyadari tugas akhir ini masih

jauh dari kata sempurna dan terdapat kekurangan dalam melakukan eksperimen maupun penulisan. Karena itu, kritik dan saran pembaca sangat dihargai. Semoga tugas akhir ini dapat memberikan manfaat bagi para pembaca.

Jakarta, 11 Agustus 2022

Penulis



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.....	9
BAB 1. LITERATUR REVIEW	10
BAB 2. ANALISIS DAN PERANCANGAN.....	22
BAB 3. SOURCE CODE	25
BAB 4. <i>DATASET</i>	56
BAB 5. TAHAPAN EKSPERIMEN.....	58
BAB 6. HASIL SEMUA EKSPERIMEN.....	67
DAFTAR PUSTAKA	78
LAMPIRAN DOKUMEN HAKI.....	82
LAMPIRAN KORESPONDENSI	84

NASKAH JURNAL

PERBANDINGAN ALGORITMA SVM DAN NAÏVE BAYES DALAM ANALISA SENTIMEN TERHADAP PENGGUNA TWITTER MENGENAI VAKSIN BOOSTER

Muhammad Ramzy Alfinrizq Ramadhan^{1*)}, Leonard Goeirmanto²

^{1,2}Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana, Jakarta

² Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana, Jakarta

^{1,2}Jln. Raya, Meruya Sel., Kec. Kembangan, Jakarta, Daerah Khusus Ibukota Jakarta, 11650, Indonesia
email: ¹41518010042@student.mercubuana.ac.id, ²leonard@mercubuana.ac.id

Abstract – *The implementation of the 3rd dose of vaccination given to health workers, assistant health workers, and supporting personnel working in health care facilities by the Ministry of Health, which has been carried out since July 16, 2021, has generated positive and negative responses from twitter users. Efforts to implement the 3rd dose of vaccine or booster vaccine aim to reduce the number of health workers infected with Covid-19. The responses from these twitter users will then be used as data in sentiment analysis in this study to find out how the sentiments of twitter users are towards giving the 3rd dose of vaccine or booster vaccine. This study uses the Support Vector Machine (SVM) and Naïve Bayes algorithms as a comparison to determine a better algorithm in classifying sentiment on tweet data about booster vaccines. Using 5,318 tweet data collected, the results obtained from the comparison of the SVM and Naïve Bayes algorithms using GridSearchCV and cross validation in sentiment analysis in the form of SVM to be a better algorithm with an average precision of 85.42% in SVM scheme 1, average recall 99.76%, an average f1-score of 91.45%, and an average accuracy of 84.04% in SVM scheme 2.*

Abstrak – Pelaksanaan vaksinasi dosis ke-3 yang diberikan kepada tenaga kesehatan, asisten tenaga kesehatan, dan tenaga penunjang yang bekerja di fasilitas pelayanan kesehatan oleh Kementerian Kesehatan yang dilakukan sejak 16 Juli 2021 memunculkan tanggapan positif dan negatif dari pengguna *twitter*. Upaya pelaksanaan pemberian vaksin dosis ke-3 atau vaksin *booster* bertujuan untuk mengurangi angka tenaga kesehatan yang terinfeksi *Covid-19*. Tanggapan-tanggapan dari pengguna *twitter* ini kemudian akan dijadikan data dalam analisa sentimen pada penelitian ini untuk mengetahui bagaimana

sentimen pengguna *twitter* terhadap pemberian vaksin dosis ke-3 atau vaksin *booster*. Penelitian ini menggunakan algoritma *Support Vector Machine (SVM)* dan *Naïve Bayes* sebagai perbandingan untuk menentukan algoritma yang lebih baik dalam melakukan klasifikasi sentimen terhadap data *tweet* mengenai vaksin *booster*.

Menggunakan 5.318 data *tweet* yang dikumpulkan, didapatkan hasil dari perbandingan algoritma *SVM* dan *Naïve Bayes* menggunakan *GridSearchCV* dan *cross validation* dalam analisa sentimen berupa *SVM* menjadi algoritma yang lebih baik dengan nilai rata-rata *precision* 85,42% pada *SVM* skema 1, rata-rata *recall* 99,76%, rata-rata *f1-score* 91,45%, dan rata-rata akurasi 84,04% pada *SVM* skema 2.

Kata Kunci – *Support Vector Machine, Naïve Bayes, Analisa Sentimen.*

I. PENDAHULUAN

Pelaksanaan vaksinasi dosis ke-3 yang diberikan kepada tenaga kesehatan, asisten tenaga kesehatan, dan tenaga penunjang yang bekerja di fasilitas pelayanan kesehatan oleh Kementerian Kesehatan yang dilakukan sejak 16 Juli 2021 memunculkan tanggapan positif dan negatif dari pengguna *twitter*. Upaya pelaksanaan pemberian vaksin dosis ke-3 atau vaksin *booster* bertujuan untuk mengurangi angka tenaga Kesehatan yang terinfeksi *Covid-19*. Tanggapan-tanggapan dari pengguna *twitter* ini kemudian akan dijadikan data dalam analisa sentimen pada penelitian ini untuk mengetahui bagaimana sentimen pengguna *twitter* terhadap pemberian vaksin dosis ke-3 atau vaksin *booster*.

Analisa Sentimen digunakan sebagai metode dalam menganalisa data agar diketahui polaritas dari data *tweet* yang digunakan untuk mendapatkan informasi bernilai dari data *tweet* [1]. Dengan polaritas yang didapatkan kemudian dikelompokkan untuk menentukan sentimen data *tweet* yang dianalisa termasuk ke dalam tanggapan positif, netral, atau

*) penulis korespondensi: Muhammad Ramzy Alfinrizq RamadhanR
Email: 41518010042@student.mercubuana.ac.id

negatif [2][3][4]. Pada penelitian mengenai analisa sentimen yang sudah dilakukan menggunakan algoritma *SVM* dan *Naïve Bayes* dilakukan analisa sentimen masyarakat tentang upaya vaksinasi dalam mengatasi pandemi Covid-19 dan menghasilkan nilai akurasi rata-rata sebesar 84,41% untuk algoritma *SVM* dan 85,59% untuk algoritma *Naïve Bayes* [5]. Algoritma *SVM* dan *Naïve Bayes* juga pernah dilakukan dalam penelitian analisa sentimen lainnya, yaitu mengenai analisa sentimen pengguna *twitter* terhadap kebijakan PSBB dan menghasilkan nilai akurasi sebesar 95,94% [6].

Penelitian-penelitian sebelumnya membuktikan bahwa algoritma *SVM* dan *Naïve Bayes* dapat digunakan dalam melakukan analisa sentimen. Berdasarkan hal tersebut, penulis melakukan penelitian mengenai perbandingan algoritma *SVM* dan *Naïve Bayes* untuk analisa sentimen terhadap pemberian vaksin dosis ke-3 atau vaksin *booster* dengan data yang diambil dari *twitter* dengan tujuan untuk mengetahui perbandingan algoritma *Support Vector Machine (SVM)* dan *Naïve Bayes* dalam analisa sentimen terhadap pengguna *twitter* mengenai pemberian vaksin *booster*.

II. PENELITIAN YANG TERKAIT

Dalam penyusunan penelitian ini, terdapat beberapa penelitian yang berkaitan dengan latar belakang masalah ataupun algoritma yang digunakan. Berikut ini beberapa penelitian terdahulu mengenai analisa sentimen.

Dalam penelitian berjudul “Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19”, peneliti menggunakan algoritma *Support Vector Machine* dan *Naïve Bayes* mendapatkan hasil sentimen positif yang paling dominan dengan kata kunci “vaksinmerahputih” sebesar 98% pada algoritma *SVM* dan 89% pada *Naïve Bayes* [5].

Dalam Penelitian lain yang berjudul “Analisa Sentimen Pengguna Twitter Terhadap Kebijakan Psbb” dengan algoritma *Support Vector Machine* mendapatkan hasil analisa sentimen yang cenderung positif terhadap kebijakan PSBB di Indonesia, model klasifikasi *SVM* dengan kernel *RBF* mendapatkan hasil yang paling baik secara keseluruhan dengan nilai akurasi 95.94%, *precision* 94.41%, *recall* 97.8%, dan *f1-score* 96.08% [6].

Pada penelitian lain yang berjudul “Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm”, peneliti menggunakan algoritma *Naïve Bayes*, *Decision Tree*, dan *Random Forest* mendapatkan hasil yang menunjukkan algoritma *Naïve Bayes* menghasilkan nilai akurasi tertinggi sebesar 83,43% dibandingkan *Decision Tree* dan *Random Forest* yang sama-sama mendapatkan hasil akurasi 82,91%. Penelitian ini juga menunjukkan bahwa pengguna media sosial di Indonesia bersikap netral terhadap kampanye *anti-LGBT* [7].

Pada penelitian berjudul “Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes” peneliti melakukan analisa sentimen menggunakan algoritma *Decision Tree*, *K-Nearest Neighbor*, dan *Naïve Bayes* mengenai sentimen pengguna *twitter* terhadap e-commerce tokopedia dan bukalapak. Hasil

yang didapatkan menunjukkan bahwa *Naïve Bayes* mampu menghasilkan akurasi paling tinggi dengan nilai 77%[8].

Dalam penelitian yang berjudul “Optimization of Sentiment Analysis for Indonesian Presidential Election using Naïve Bayes and Particle Swarm Optimization” menggunakan algoritma *Naïve Bayes* dan *Particle Swarm Optimization* mendapatkan hasil penelitian berupa algoritma *Naïve Bayes* dengan *Particle Swarm Optimization (PSO)* memiliki akurasi yang lebih tinggi sekitar 4,12% dengan nilai 90,74% dibandingkan tanpa menggunakan *PSO* yang hanya menghasilkan 86,62% [9].

Penelitian lain dengan judul “Seleksi Fitur Support Vector Machine pada Analisis Sentimen Keberlanjutan Pembelajaran Daring” yang menggunakan algoritma *Support Vector Machine* dengan dua seleksi fitur dan variabel *k-Fold* pada *Cross Validation* menghasilkan evaluasi rata-rata nilai akurasi 83,90% untuk seleksi fitur *Term Frequency* dan 83,50% untuk seleksi fitur *TF-IDF* [10].

Pada penelitian lain yang berjudul “Analisis Sentimen Terhadap Penerapan Sistem Plat Nomor Ganjil / Genap Pada Twitter Dengan Metode Klasifikasi Naive Bayes”, peneliti melakukan analisa sentimen menggunakan algoritma klasifikasi *Naïve Bayes* dan menghasilkan menghasilkan nilai akurasi sebesar 86,67%, *precision* 71,43% dan *recall* 80,00% [11].

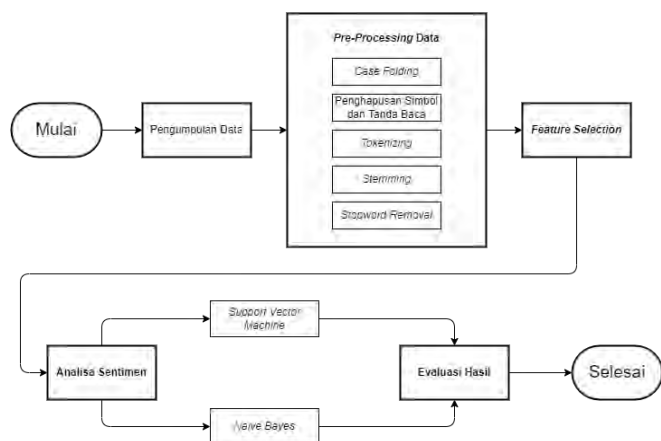
Dalam penelitian mengenai analisa sentimen dengan judul “Sentiment Analysis on Covid19 Vaccines in Indonesia: From the Perspective of Sinovac and Pfizer”, peneliti melakukan analisa sentimen menggunakan algoritma *Naïve Bayes*, *Support Vector Machine*, dan *Random Forest* terhadap 2 dataset dengan keyword “Sinovac” dan “Pfizer”. Penelitian menunjukkan bahwa *Support Vector Machine* memberikan nilai akurasi tertinggi dalam analisa sentimen menggunakan kedua dataset, yaitu 85% dengan dataset vaksin *Sinovac* dan 78% dengan dataset vaksin *Pfizer* [12].

Pada penelitian dengan judul “Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization” peneliti menggunakan algoritma *Support Vector Machine* Berbasis *Particle Swarm Optimization (PSO)* dan hasil penelitian menunjukan model *SVM* mampu menghasilkan akurasi sebesar 95,46% dan *AUC* 0,979, sedangkan model *SVM* dengan *PSO* memberikan hasil akurasi yang lebih besar yaitu 96,04% dan nilai *AUC* 0,993 [13].

Dalam penelitian lain yang berjudul “Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier” menunjukan hasil perbandingan berupa algoritma *Support Vector Machine* mampu memberikan hasil akurasi yang lebih baik dengan nilai 81,67% dibandingkan menggunakan metode klasifikasi *Naive Bayes Classifier* yang hanya mampu menghasilkan akurasi sebesar 67.20% [14].

III. METODE PENELITIAN

Tahapan yang dilakukan dalam penelitian ini dapat dilihat pada gambar di bawah ini.



Gbr. 1 Tahapan Penelitian

Berikut ini penjelasan mengenai tahapan penelitian yang dilakukan dalam analisa sentimen menggunakan algoritma *Support Vector Machine (SVM)* dan *Naive Bayes* berdasarkan gambar 1:

A. Pengumpulan Data

Pada tahap ini, peneliti melakukan proses pengumpulan data dengan cara mengambil data dari twitter. Dataset yang digunakan pada penelitian ini adalah data *tweet* para pengguna *twitter* mengenai vaksin *booster*. Dataset yang digunakan pada penelitian ini adalah data *tweet* para pengguna *twitter* mengenai vaksin *booster*. Dataset didapatkan dengan cara melakukan *crawling* data dari *twitter* menggunakan bahasa pemrograman *python* menggunakan library *tweepy* dan *API twitter*.

Pengambilan data yang dilakukan di *twitter* dalam penelitian ini menggunakan beberapa *keyword*, yaitu *keyword* “vaksin booster”, “vaksin dosis ke-3”, “vaksin dosis ketiga” dan “vaksin nakes” dalam jangka waktu 1 September – 30 November 2021. Melalui proses pengambilan data menggunakan *keyword-keyword* di atas didapatkan data sebanyak 4032 data untuk *keyword* “vaksin booster”, 213 data untuk *keyword* “vaksin dosis ke-3”, 770 data untuk *keyword* “vaksin dosis ketiga”, dan 303 data untuk *keyword* “vaksin nakes” dengan total jumlah data secara keseluruhan berjumlah 5.318 data. Data yang sudah dikumpulkan digabung menjadi 1 *dataset* keseluruhan. Kemudian, data *tweet* yang sudah dikumpulkan dilakukan *filtering* untuk menyaring data dari *tweet* yang mengandung *keyword* pencarian tetapi tidak membahas mengenai vaksin *booster* seperti berita atau pengumuman jadwal pelaksanaan vaksinasi. Sehingga total jumlah *dataset* yang dipakai untuk melakukan analisa sentimen berjumlah 4.632 data. Setelah *dataset* dilakukan penyaringan, data *tweet* diberikan label sentimen pada kolom “*Sentiment*” secara manual sebagai acuan untuk algoritma yang digunakan dalam melakukan proses analisa sentimen.

Pada *dataset* terdapat 4 kolom, yaitu kolom “*CreatedAt*” sebagai kolom yang menunjukkan kapan *tweet* diunggah, kolom “*UserName*” yang berisi nama pengguna *twitter*, kolom “*Tweet*” yang berisi cuitan atau *tweet* dari pengguna *twitter*, dan kolom “*Sentiment*” yang digunakan untuk label sentimen dari isi *tweet*.

	CreatedAt	UserName	Tweet	Sentiment
0	2021-09-10 08:30:30	springbnn	ditawari booster sama atasan, tapi kepikiran e...	Unlabeled
1	2021-09-10 12:43:55	SaintChyrl	Apa beda isi vaksin pertama, kedua dan booster?	Unlabeled
2	2021-09-10 14:28:46	arifbsantoso	Ada jalan keluar buat para penabung titer,\n/n...	Unlabeled
3	2021-09-10 09:10:52	traxex87	@abdul_mallek90 @khaidirsalimah @IKJAWMY @KKMP...	Unlabeled
4	2021-09-10 08:06:48	DonVapridass	Vaksin booster bikin meriang sama lemes sehari...	Unlabeled
...
4627	2021-11-01 04:12:50	imanr81	@_Banyoe @StopPlandemit Selesai PCR muncul gel...	Unlabeled
4628	2021-11-01 08:34:19	nagtira	@Khairikyj Nak log in mysejahtera pun tak boleh...	Unlabeled
4629	2021-11-01 13:47:09	mytippytoe	Dokter Vet bilang karena umurnya uda 7 bulan m...	Unlabeled
4630	2021-11-01 04:19:51	ambargedon	@CNNIndonesia Padahal jarak tdk menjamin seseo...	Unlabeled
4631	2021-10-31 23:51:36	lemonAZS	And then sekarang, dh ada vaksin penggalak (bo...	Unlabeled

4632 rows x 4 columns

Gbr. 2 Dataset yang Digunakan

B. Pre-Processing Data

Pada tahap *pre-processing* data, data yang telah dilakukan penyaringan dan diberi label sentimen, selanjutnya dilakukan *pre-processing* pada *dataset* yang digunakan untuk membersihkan data dari kolom yang tidak memiliki nilai, tanda baca yang tidak diperlukan, simbol-simbol dan *emoticon* yang tidak dibutuhkan, kata-kata yang tidak baku melalui proses *Case Folding*, *Penghapusan simbol dan tanda baca*, *Tokenizing*, *Stemming*, dan *Stopword Removal*. Proses ini diperlukan untuk membuat *dataset* menjadi lebih rapih untuk proses klasifikasi analisa sentimen [15][16].

Berikut ini penjelasan tahapan yang dilakukan pada *pre-processing* terhadap *dataset* pada penelitian ini:

• Case Folding

Case Folding dilakukan untuk mengubah huruf-huruf kapital menjadi huruf kecil pada kalimat *tweet* pada *dataset*. *Case Folding* bertujuan untuk menghindari perbedaan persepsi dari 1 kata yang sama antara kata yang menggunakan huruf kapital dengan yang tidak. Berikut ini contoh hasil *Case Folding*.

TABEL I. CASE FOLDING

Fase	Tweet
Sebelum	Pemerintah Indonesia saat ini harusnya memfokuskan vaksinasi pada mereka yang belum tersentuh, yang sulit dijangkau, dan yang paling merasa keparahan dari pandemi, alih-alih mempromosikan akses booster pada mereka yang tidak alami hambatan ataupun kesulitan mengakses vaksin.
Sesudah	pemerintah Indonesia saat ini harusnya memfokuskan vaksinasi pada mereka yang belum tersentuh, yang sulit dijangkau, dan yang paling merasa keparahan dari pandemi, alih-alih mempromosikan akses booster pada mereka yang tidak alami hambatan ataupun kesulitan mengakses vaksin.

• Penghapusan simbol dan tanda baca

Pada tahap penghapusan simbol dan tanda baca, simbol-simbol dan tanda baca yang tidak dibutuhkan dihapus untuk meminimalisir kesalahan persepsi pada proses analisa sentimen *tweet*. Berikut ini contoh hasil penghapusan simbol dan tanda baca.

TABEL II. PENGHAPUSAN SIMBOL DAN TANDA BACA

Fase	<i>Tweet</i>
Sebelum	pemerintah Indonesia saat ini harusnya memfokuskan vaksinasi pada mereka yang belum tersentuh, yang sulit dijangkau, dan yang paling merasa keparahan dari pandemi, alih-alih mempromosikan akses booster pada mereka yang tidak alami hambatan ataupun kesulitan mengakses vaksin.
Sesudah	pemerintah Indonesia saat ini harusnya memfokuskan vaksinasi pada mereka yang belum tersentuh yang sulit dijangkau dan yang paling merasa keparahan dari pandemi alih alih mempromosikan akses booster pada mereka yang tidak alami hambatan ataupun kesulitan mengakses vaksin

- *Tokenizing*

Tokenisasi dilakukan untuk memecah kalimat tweet menjadi kata per kata. *Tweet* dipecah per karakter pemisah seperti spasi dan enter. Berikut ini contoh hasil *Tokenizing*.

TABEL III. TOKENIZING

Fase	<i>Tweet</i>
Sebelum	pemerintah Indonesia saat ini harusnya memfokuskan vaksinasi pada mereka yang belum tersentuh yang sulit dijangkau dan yang paling merasa keparahan dari pandemi alih alih mempromosikan akses booster pada mereka yang tidak alami hambatan ataupun kesulitan mengakses vaksin
Sesudah	['pemerintah', 'indonesia', 'saat', 'ini', 'harusnya', 'memfokuskan', 'vaksinasi', 'pada', 'mereka', 'yang', 'belum', 'tersentuh', 'yang', 'sulit', 'dijangkau', 'dan', 'yang', 'paling', 'merasa', 'keparahan', 'dari', 'pandemi', 'alih', 'alih', 'mempromosikan', 'akses', 'booster', 'pada', 'mereka', 'yang', 'tidak', 'alami', 'hambatan', 'ataupun', 'kesulitan', 'mengakses', 'vaksin']

- *Stemming*

Tahapan *stemming* dilakukan untuk mengubah kata-kata dalam *tweet* pengguna *twitter* menjadi kata dasar berdasarkan Kamus Besar Bahasa Indonesia (KBBI). Berikut ini contoh hasil tahapan *stemming*.

TABEL IV. STEMMING

Fase	<i>Tweet</i>
Sebelum	['pemerintah', 'indonesia', 'saat', 'ini', 'harusnya', 'memfokuskan', 'vaksinasi', 'pada', 'mereka', 'yang', 'belum', 'tersentuh', 'yang', 'sulit', 'dijangkau', 'dan', 'yang', 'paling', 'merasa', 'keparahan', 'dari', 'pandemi', 'alih', 'alih', 'mempromosikan', 'akses', 'booster', 'pada', 'mereka', 'yang', 'tidak', 'alami', 'hambatan', 'ataupun', 'kesulitan', 'mengakses', 'vaksin']
Sesudah	['perintah', 'indonesia', 'saat', 'ini', 'harus', 'fokus', 'vaksinasi', 'pada', 'mereka', 'yang', 'belum', 'sentuh', 'yang', 'sulit', 'jangkau', 'dan', 'yang', 'paling', 'rasa', 'parah', 'dari', 'pandemi', 'alih', 'alih', 'promosi', 'akses', 'booster', 'pada', 'mereka', 'yang', 'tidak', 'alami', 'hambat', 'atau', 'sulit', 'akses', 'vaksin']

- *Stopword Removal*

Stopword Removal bertujuan untuk menghilangkan bagian kata yang tidak berkontribusi pada kata tersebut seperti kata sambung. Dalam tahapan ini penulis membuat *list stopwords* berdasarkan *library Sastrawi* untuk diterapkan pada *dataset*. Berikut ini contoh hasil tahapan *stopword removal*.

TABEL V. STOPWORD REMOVAL

Fase	<i>Tweet</i>
Sebelum	['perintah', 'indonesia', 'saat', 'ini', 'harus', 'fokus', 'vaksinasi', 'pada', 'mereka', 'yang', 'belum', 'sentuh', 'yang', 'sulit', 'jangkau', 'dan', 'yang', 'paling', 'rasa', 'parah', 'dari', 'pandemi', 'alih', 'alih', 'promosi', 'akses', 'booster', 'pada', 'mereka', 'yang', 'tidak', 'alami', 'hambat', 'atau', 'sulit', 'akses', 'vaksin']
Sesudah	['perintah', 'indonesia', 'fokus', 'vaksinasi', 'sentuh', 'sulit', 'jangkau', 'parah', 'pandemi', 'alih', 'alih', 'promosi', 'akses', 'booster', 'alami', 'hambat', 'sulit', 'akses', 'vaksin']

- *Feature Selection*

Setelah melakukan *pre-processing*, *dataset* yang akan digunakan dalam melakukan pengolahan analisa sentimen tetap berjumlah 4.632 data *tweet*. Selanjutnya, dilakukan *features selection* menggunakan metode *Term Frequency*

Inverse Document Frequency (TF-IDF) untuk memudahkan teks pada *dataset* dibaca oleh komputer dengan mengubah data teks menjadi angka. *Feature Selection* mengubah data teks menjadi angka-angka dalam melakukan pengolahan *dataset*, *TF-IDF* memberikan pembobotan pada tiap kata yang ada pada *dataset* berdasarkan seberapa sering kata muncul pada dokumen [17][18][16]. Semakin sering sebuah kata muncul maka semakin besar bobot dari kata tersebut [8].

Perhitungan TF-IDF dapat dilakukan dengan menghitung nilai TF menggunakan persamaan (1) dan menghitung nilai IDF menggunakan persamaan (2), kemudian menggunakan persamaan (3) untuk dapat menemukan nilai TF-IDF [19].

$$TF_{(t,d)} = \frac{f_{(t,d)}}{n_d} \quad (1)$$

$$IDF_{(t,d)} = \log \frac{N}{df_{(t,d)}} \quad (2)$$

$$TF-IDF_{(t,d)} = tf_{(t,d)} \cdot idf_{(t,d)} \quad (3)$$

D. Proses Analisa Sentimen Menggunakan Algoritma SVM dan Naïve Bayes

Pada tahapan ini, dilakukan proses analisa sentimen menggunakan data yang telah *di-pre-processing* sebelumnya. Algoritma *Support Vector Machine (SVM)* dan *Naïve Bayes* digunakan untuk melakukan analisa sentimen dan melakukan validasi terhadap hasil pengolahan data yang dilakukan.

Algoritma *Support Vector Machine* adalah salah satu algoritma *machine learning* bertipe *supervised learning* yang implementasinya membutuhkan tahap pelatihan menggunakan *sequential training SVM* dan diikuti proses pengujian [13]. Prinsip *SVM* dalam bekerja adalah menentukan ruang pemisah pada ruang pencarian yang dapat memisahkan kelas-kelas yang berbeda atau disebut *hyperplane*[20], semakin besar *hyperplane*, maka tingkat kesalahan yang terjadi saat klasifikasi akan semakin kecil[14][10]. Proses analisis dimulai dengan cara mengubah data teks yang ada menjadi sebuah data vektor, kemudian akan dikombinasikan dengan menggunakan *TF-IDF* untuk dilakukan pembobotan [5]. Dalam *SVM*, *hyperplane* dapat digambarkan dalam persamaan (4) [21].

$$x \cdot w + b = 0 \quad (4)$$

Pada Kemudian, dalam memisahkan kelas positif dan negatif. *Hyperplane* menggunakan pertidaksamaan (5) [21].

$$y_i(x_i \cdot w + b) \geq 1, \forall_i \quad (5)$$

Dengan nilai $y_i = 1$ untuk kelas positif dan $y_i = -1$ untuk kelas negatif

Sedangkan, *Naïve Bayes* merupakan salah satu model *machine learning* yang dapat digunakan dalam melakukan klasifikasi data dan dapat digunakan untuk melakukan analisa sentimen [22][11]. *Naïve Bayes* merupakan model *machine learning* yang menggunakan konsep probabilitas berdasarkan teorema *bayes* dengan konsep probabilitas bersyarat dan hasil probabilitas masa lalu dalam melakukan klasifikasi [8] [23]. Teorema pada *Naïve Bayes* dapat digambarkan dalam persamaan 4 sebagai berikut.

$$P(H|X) = \frac{P(H) \cdot P(X|H)}{P(X)} \quad (6)$$

E. Evaluasi Hasil

Pada tahap ini, evaluasi dilakukan untuk melihat hasil prediksi dalam menganalisa sentimen tweet menggunakan algoritma *Support Vector Machine (SVM)* dan *Naïve Bayes*. Evaluasi dilakukan dengan membandingkan nilai akurasi, *precision*, *recall*, dan *f1-score*.

Nilai akurasi berdasarkan seberapa banyak algoritma *Support Vector Machine* dan *Naïve Bayes* dalam memprediksi sentimen dengan benar atau sesuai dengan label aslinya. Semakin banyak data *tweet* yang diprediksi dengan benar maka nilai akurasi semakin besar. Nilai akurasi dapat ditemukan berdasarkan persamaan berikut [8]:

$$Akurasi = \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad (7)$$

Nilai *Precision* didapatkan dengan menghitung seberapa banyak data *tweet* yang diprediksi dengan benar berdasarkan labelnya. Nilai *precision* didapatkan menggunakan persamaan di bawah ini:

$$Precision = \frac{TP}{(TP + FP)} \quad (8)$$

Nilai *Recall* diperoleh menggunakan persamaan (9) di bawah ini. *Recall* digunakan untuk mengevaluasi seberapa besar algoritma yang digunakan dapat memprediksi *True Positive* dari data *tweet*.

$$Recall = \frac{TP}{(TP + FN)} \quad (9)$$

Sedangkan, nilai *F1-Score* adalah nilai yang didapatkan dari pembobotan *Recall* dan *Precision*. Persamaan di bawah ini digunakan untuk mendapatkan nilai *F1-Score*:

$$F1-Score = \frac{2 \cdot (Precision \cdot Recall)}{(Precision + Recall)} \quad (10)$$

IV. HASIL DAN PEMBAHASAN

Pada Pengujian Algoritma, dilakukan beberapa kali percobaan berdasarkan proporsi data latih dan data uji pada algoritma *Support Vector Machine* dan *Naïve Bayes* dalam melakukan analisa sentimen data *tweet* mengenai vaksin booster.

TABEL VI. PROPORSI DATA PENGUJIAN

Data Latih	Data Uji
50%	50%
60%	40%
70%	30%
80%	20%
90%	10%

Pada pengujian algoritma *SVM* dan *Naïve Bayes* dilakukan beberapa *tuning hyperparameter* seperti “C”, “gamma” pada algoritma *SVM* dan “alpha” pada *naïve bayes* menggunakan *GridSearchCV* dan validasi menggunakan *cross validation* sebesar 30 [21][24][25][26].

TABEL VII. PARAMETER SVM

Parameter	Nilai
C	0.01, 0.1, 0.5, 1.0, 10.0
gamma	1, 0.1, 0.01, 0.001, 0.0001
Kernel Skema 1	rbf

Kernel Skema 2	<i>linear</i>
cross validation	30

Pengujian pertama dilakukan menggunakan *GridSearchCV* menggunakan detail parameter seperti pada tabel VII. Setelah melakukan *tuning*, didapatkan parameter terbaik “C” dengan nilai “10.0” dan “gamma” dengan nilai “0.1”. Sehingga detail model *SVM* yang digunakan pada skema pengujian pertama dapat dilihat pada tabel 6.3 di bawah ini.

TABEL VIII. *SVM* SKEMA 1

Parameter	Nilai
<i>C</i>	10.0
<i>gamma</i>	0.1
kernel	<i>rbf</i>
<i>cross validation</i>	30

Menggunakan skema 1 di atas, didapatkan hasil prediksi dalam analisa sentimen sebagai berikut:

TABEL IX. AKURASI KESELURUHAN *SVM* SKEMA 1

Proporsi Data	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	Akurasi
50% - 50%	85,63	96,75	90,85	83,66
60% - 40%	86,08	96,85	91,15	84,17
70% - 30%	85,65	96,85	90,89	83,73
80% - 20%	85,65	96,82	90,89	83,05
90% - 10%	84,1	96,31	89,79	82,07
Rata-rata	85,42	96,72	90,71	83,37

Berdasarkan tabel IX di atas, pengujian algoritma *SVM* menggunakan skema 1 di atas mendapatkan hasil rata-rata akurasi sebesar 83,37%. Akurasi, *precision*, *recall*, dan *f1-score* terbaik terdapat pada pengujian skema 1 dengan distribusi pembagian data sebesar 60% - 40%, di mana pembagian data latih dan uji menggunakan proporsi data tersebut mendapatkan akurasi sebesar 84,17%, *precision* 86,08%, *recall* 96,85%, *f1-score* 91,15% dan akurasi, *precision*, *recall*, dan *f1-score* terendah didapatkan pada pengujian menggunakan proporsi data 90% - 10% dengan nilai akurasi 82,07%, *precision* 84,1%, *recall* 96,31%, *f1-score* 89,79%.

Pengujian kedua juga dilakukan menggunakan *GridSearchCV* menggunakan detail parameter seperti pada tabel VII. Parameter terbaik yang didapatkan setelah melakukan *tuning hyperparameter* dan model *SVM* yang digunakan untuk melakukan analisa sentimen dalam pengujian skema 2 dapat dilihat pada tabel X di bawah ini.

TABEL X. *SVM* SKEMA 2

Parameter	Nilai
<i>C</i>	1.0
<i>gamma</i>	1
kernel	<i>linear</i>

cross validation	30
-------------------------	----

TABEL XI. AKURASI KESELURUHAN *SVM* SKEMA 2

Proporsi Data	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	Akurasi
50% - 50%	84,52	99,64	91,46	84,4
60% - 40%	84,78	99,74	91,65	84,71
70% - 30%	84,46	99,83	91,5	84,45
80% - 20%	84,13	99,87	91,33	84,13
90% - 10%	82,53	99,74	90,32	82,51
Rata-rata	84,48	99,76	91,45	84,04

Hasil pengujian algoritma *SVM* menggunakan skema 2 dapat dilihat pada tabel 6.6 di atas. *SVM* skema 2 mendapatkan hasil rata-rata akurasi sebesar 84,04%. Nilai *precision*, *f1-score*, dan akurasi tertinggi terdapat pada pengujian skema 2 dengan proporsi data sebesar 60% - 40% dengan nilai 84,78% *precision*, 91,65% *f1-score*, dan 84,71% akurasi. Nilai *Recall* tertinggi didapat menggunakan proporsi data 80% - 20% yaitu 99,87%.

Berdasarkan 2 skema pengujian algoritma *SVM* yang telah dilakukan didapatkan hasil bahwa skema 2 memiliki rata-rata yang lebih tinggi pada akurasi, *recall*, dan *f1-score* dibanding skema 1 yaitu 84.04% akurasi, 99,76% *recall*, dan 91,45% *f1-score*. Sedangkan *SVM* skema 1 hanya lebih baik pada *precision* dibanding *SVM* skema 2 dengan nilai rata-rata *precision* 85,42% untuk skema 1.

Pada pengujian algoritma *naïve bayes* juga dilakukan beberapa pengujian menggunakan *GridSearchCV* untuk melakukan *tuning* parameter “alpha” dengan detail sebagai berikut.

TABEL XII. PARAMETER *NAÏVE BAYES*

Parameter	Nilai
<i>alpha</i>	0.01, 0.1, 0.5, 1.0, 10.0
<i>fit_prior</i> skema 1	<i>False</i>
<i>fit_prior</i> skema 2	<i>True</i>
<i>cross validation</i>	30

Pengujian pertama dilakukan menggunakan *GridSearchCV* menggunakan detail parameter seperti pada tabel 6.7. Parameter “alpha” terbaik yang didapatkan adalah “10.0”. Detail model *Naïve Bayes* yang digunakan pada skema pertama digambarkan pada tabel XIII di bawah ini.

TABEL XIII. *NAÏVE BAYES* SKEMA 1

Parameter	Nilai
<i>alpha</i>	10.0
<i>fit_prior</i>	<i>False</i>
<i>cross validation</i>	30

TABEL XIV. AKURASI KESELURUHAN *NAÏVE BAYES* SKEMA 1

Proporsi Data	Precision	Recall	F1-Score	Akurasi
50% - 50%	83,87	99,69	91,1	83,66
60% - 40%	84,22	99,68	91,3	84,01
70% - 30%	83,94	99,57	91,09	83,66
80% - 20%	83,84	99,74	91,1	83,69
90% - 10%	82,17	99,74	90,11	82,07
Rata-rata	84,48	99,76	90,94	83,42

Berdasarkan tabel XIV di atas, pengujian terhadap algoritma *naïve bayes* skema 1 mendapatkan hasil rata-rata akurasi sebesar 83,42%. Akurasi terendah terdapat pada pengujian menggunakan proporsi data 90% - 10% dengan nilai akurasi 82,07%. Akurasi terbaik didapatkan melalui pengujian menggunakan proporsi dataset 60% - 40%. Nilai *precision* terbaik didapat dengan pengujian pada proporsi dataset 80% - 20% dan 90% - 10% dengan nilai 99,74%.

Pengujian kedua untuk algoritma *naïve bayes* juga dilakukan menggunakan *GridSearchCV* menggunakan detail parameter seperti pada tabel 6.7. Parameter "*alpha*" terbaik yang didapatkan adalah "0.5". Detail model *Naïve Bayes* yang digunakan pada skema pertama digambarkan pada tabel XV di bawah ini.

TABEL XV. NAÏVE BAYES SKEMA 2

Parameter	Nilai
<i>alpha</i>	0.5
<i>fit_prior</i>	True
<i>cross validation</i>	30

TABEL XVI.

AKURASI KESELURUHAN NAÏVE BAYES SKEMA 2

Proporsi Data	Precision	Recall	F1-Score	Akurasi
50% - 50%	83,88	99,79	91,15	83,75
60% - 40%	84,23	99,74	91,33	84,06
70% - 30%	83,95	99,66	91,13	83,73
80% - 20%	83,82	99,61	91,04	83,59
90% - 10%	82,14	99,47	89,98	81,86
Rata-rata	83,60	99,65	90,93	83,40

Pada pengujian *naïve bayes* skema ke-2, hasil akurasi rata-rata yang dihasilkan lebih rendah daripada *naïve bayes* skema pertama yaitu sebesar 83,40%. Akurasi terbaik pada skema kedua juga didapatkan melalui pengujian menggunakan proporsi dataset 60% - 40%. Nilai *precision* terbaik didapat dengan pengujian pada proporsi dataset 50% - 50% dengan nilai 99,79%, *recall* dan *f1-score* terbaik didapatkan pada pengujian menggunakan dataset dengan proporsi pembagian 60% data latih dan 40% data uji sebesar 84,23%.

Berdasarkan 2 skema pengujian algoritma *naïve bayes* yang telah dilakukan didapatkan hasil bahwa skema 1 memiliki rata-rata yang lebih tinggi pada *recall*, dan *f1-score* dibanding skema 2 yaitu 99,65% *recall*, dan 90,93% *f1-score*. Sedangkan *naïve bayes* skema 1 lebih baik pada *precision* dan akurasi dibanding *naïve bayes* skema 2 dengan nilai rata-rata *precision* 84,48% dan rata-rata akurasi 83,42% untuk skema 1.

TABEL XVII. RATA-RATA KESELURUHAN

Skema	Precision	Recall	F1-Score	Akurasi
SVM Skema 1	85,42	96,72	90,71	83,37
SVM Skema 2	84,48	99,76	91,45	84,04
Naïve Bayes Skema 1	84,48	99,76	90,94	83,42
Naïve Bayes Skema 2	83,60	99,65	90,93	83,40

Pada perbandingan algoritma *SVM* dan *naïve bayes* berdasarkan proporsi data latih dan data uji pada tabel 6.17 di atas, didapatkan algoritma *SVM* menjadi algoritma yang lebih baik dalam analisa sentimen dengan nilai *precision* tertinggi yaitu 85,42% pada *SVM* skema 1, *f1-score* tertinggi dengan nilai 91,45% pada *SVM* skema 2, dan akurasi tertinggi pada *SVM* skema 2 dengan nilai 84,04%. Sedangkan rata-rata *recall* tertinggi dengan nilai 99,76% didapatkan oleh *SVM* skema 2 dan *naïve bayes* skema 1.

V. KESIMPULAN

Berdasarkan pengujian algoritma *SVM* dan *Naïve Bayes* dalam melakukan analisa sentimen mengenai vaksin *booster*, didapatkan hasil berupa perbandingan algoritma yang menunjukkan bahwa algoritma *SVM* menjadi algoritma yang lebih baik dalam melakukan analisa sentimen dibandingkan *naïve bayes*. *SVM* menjadi algoritma yang lebih baik dengan nilai rata-rata *precision* 85,42% pada *SVM* skema 1, rata-rata *recall* 99,76%, rata-rata *f1-score* 91,45%, dan rata-rata akurasi 84,04% pada *SVM* skema 2. Hal ini juga menunjukkan bahwa varian *SVM* skema 1 menjadi algoritma yang lebih baik untuk mendapatkan nilai *precision* yang lebih tinggi dengan konfigurasi parameter "*C* = 10.0", "*gamma* 0,1", dan "*kernel* = *rbf*", sedangkan varian *SVM* skema 2 dengan konfigurasi parameter "*C* = 1.0", "*gamma* = 1", dan "*kernel* = *linear*" adalah algoritma yang lebih baik untuk mendapatkan nilai *recall*, *f1-score*, dan akurasi yang lebih tinggi. Jadi, dapat disimpulkan bahwa *SVM* menjadi algoritma yang lebih baik dalam melakukan analisa sentimen mengenai vaksin *booster* karena dapat memberikan hasil yang lebih tinggi pada *precision*, *recall*, *f1-score*, dan memberikan akurasi prediksi yang lebih baik.

UCAPAN TERIMA KASIH

Terima kasih penulis sampaikan kepada berbagai pihak atas bantuan, arahan, dan bimbingan yang diberikan dalam penelitian.

DAFTAR PUSTAKA

- [1] A. Hasan, S. Moin, A. Karim, and S. Shamshirband, "Machine Learning-Based Sentiment Analysis for Twitter Accounts," *Math. Comput. Appl.*, vol. 23, no. 1, p. 11, 2018, doi: 10.3390/mca23010011.
- [2] M. I. Aditama, R. I. Pratama, K. H. U. Wiwaha, and N. A. Rakhmawati, "Analisis Klasifikasi Sentimen Pengguna Media Sosial Twitter Terhadap Pengadaan Vaksin COVID-19," *J. Inf. Eng. Educ. Technol.*, vol. 4, no. 2, pp. 90–92, 2020.
- [3] S. Yousefinaghani, R. Dara, S. Mubareka, A. Papadopoulos, and S. Sharif, "An analysis of COVID-19 vaccine sentiments and opinions on Twitter," *Int. J. Infect. Dis.*, vol. 108, pp. 256–262, 2021, doi: 10.1016/j.ijid.2021.05.059.
- [4] R. Wagh and P. Punde, "Survey on Sentiment Analysis using Twitter Dataset Rasika," *2018 Second Int. Conf. Electron. Commun. Aerosp. Technol.*, pp. 208–211, 2018.
- [5] B. Laurensz and Eko Sedyono, "Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 10, no. 2, pp. 118–123, 2021, doi: 10.22146/jnteti.v10i2.1421.
- [6] H. P. P. Zuriel and A. Fahrurrozi, "Implementasi Algoritma Klasifikasi Support Vector Machine Untuk Analisa Sentimen Pengguna Twitter Terhadap Kebijakan Psbb," *J. Ilm. Inform. Komput.*, vol. 26, no. 2, pp. 149–162, 2021, doi: 10.35760/ik.2021.v26i2.4289.
- [7] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, "Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm," *Procedia Comput. Sci.*, vol. 161, pp. 765–772, 2019, doi: 10.1016/j.procs.2019.11.181.
- [8] A. Bayhaqy, S. Sfenianto, K. Nainggolan, and E. R. Kaburuan, "Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes," *2018 Int. Conf. Orange Technol. ICOT 2018*, no. October, pp. 1–6, 2018, doi: 10.1109/ICOT.2018.8705796.
- [9] N. Hayatin, G. I. Marthasari, and L. Nuraini, "Optimization of Sentiment Analysis for Indonesian Presidential Election using Naïve Bayes and Particle Swarm Optimization," *J. Online Inform.*, vol. 5, no. 1, pp. 81–88, 2020, doi: 10.15575/join.v5i1.558.
- [10] A. P. Natasuwarna, "Seleksi Fitur Support Vector Machine pada Analisis Sentimen Keberlanjutan Pembelajaran Daring," *Techno.Com*, vol. 19, no. 4, pp. 437–448, 2020, doi: 10.33633/tc.v19i4.4044.
- [11] N. Ruhjana, "Analisis Sentimen Terhadap Penerapan Sistem Plat Nomor Ganjil / Genap Pada Twitter Dengan Metode Klasifikasi Naïve Bayes," *J. IKRA-ITH Inform.*, vol. 3, no. 1, pp. 94–99, 2019.
- [12] D. A. Nurdeni, I. Budi, and A. B. Santoso, "Sentiment Analysis on Covid19 Vaccines in Indonesia: From the Perspective of Sinovac and Pfizer," *3rd 2021 East Indones. Conf. Comput. Inf. Technol. EIConCIT 2021*, pp. 122–127, 2021, doi: 10.1109/EIConCIT50028.2021.9431852.
- [13] V. K. S. Que, A. Iriani, and H. D. Purnomo, "Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 9, no. 2, pp. 162–170, 2020, doi: 10.22146/jnteti.v9i2.102.
- [14] H. Tuhuteru and A. Iriani, "Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naïve Bayes Classifier," *J. Inform. J. Pengemb. IT*, vol. 3, no. 3, pp. 394–401, 2018, doi: 10.30591/jpit.v3i3.977.
- [15] K. Aurangzeb, N. Ayub, and M. Alhussein, "Aspect Based Multi-Labeling Using SVM Based Ensembler," *IEEE Access*, vol. 9, pp. 26026–26040, 2021, doi: 10.1109/ACCESS.2021.3055768.
- [16] N. S. Mohd Nafis and S. Awang, "An Enhanced Hybrid Feature Selection Technique Using Term Frequency-Inverse Document Frequency and Support Vector Machine-Recursive Feature Elimination for Sentiment Classification," *IEEE Access*, vol. 9, no. 1, pp. 52177–52192, 2021, doi: 10.1109/ACCESS.2021.3069001.
- [17] A. Bijaksana, P. Negara, H. Muhandi, and I. M. Putri, "Analisis Sentimen Maskapai Penerbangan Menggunakan Metode Naive Bayes Dan Seleksi Fitur Information Gain Sentiment Analysis on Airlines Using Naïve Bayes Method and Feature Selection Information Gain," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 3, pp. 599–606, 2020, doi: 10.25126/jtiik.202071947.
- [18] P. Shah, P. Swaminarayan, and M. Patel, "Sentiment analysis on film review in Gujarati language using machine learning," *Int. J. Electr. Comput. Eng.*, vol. 12, no. 1, pp. 1030–1039, 2022, doi: 10.11591/ijece.v12i1.pp1030-1039.
- [19] S. Khomsah, "Naive Bayes Classifier Optimization on Sentiment Analysis of Hotel Reviews," *J. Penelit. Pos dan Inform.*, vol. 10, no. 2, p. 157, Dec. 2020, doi: 10.17933/jppi.2020.100206.
- [20] A. Idrus, H. Brawijaya, and Maruloh, "Sentiment Analysis of State Officials News on Online Media Based on Public Opinion Using Naive Bayes Classifier Algorithm and Particle Swarm Optimization," *2018 6th Int. Conf. Cyber IT Serv. Manag. CITSM 2018*, pp. 1–7, 2019, doi: 10.1109/CITSM.2018.8674331.
- [21] 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.
- [22] M. Lestandy, A. Abdurrahim, and L. Syafa'ah, "Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naive Bayes," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 802–808, 2021.
- [23] A. Nugroho and R. Ali Fahmi, "On-Time Flight Departure Prediction System Using Naive Bayes Classification Method (Case Study: XYZ Airline)," *Int. J. Comput. Trends Technol.*, vol. 54, no. 1, pp. 4–10, 2017, doi: 10.14445/22312803/ijctt-v54p102.
- [24] Styawati, Andi Nurkholis, Zaenal Abidin, and Heni Sulistiani, "Optimasi Parameter Support Vector Machine Berbasis Algoritma Firefly Pada Data Opini Film," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 5, pp. 904–910, 2021, doi: 10.29207/resti.v5i5.3380.
- [25] U. Enri, "Optimasi Parameter Support Vector Machines untuk Prediksi Nilai Tukar Rupiah Terhadap Dollar Amerika Serikat," *J. gerbang*, vol. 8, no. 1, pp. 65–72, 2018.
- [26] J. Abellán and J. G. Castellano, "Improving the Naive Bayes classifier via a quick variable selection method using maximum of entropy," *Entropy*, vol. 19, no. 6, pp. 1–17, 2017, doi: 10.3390/e19060247.

KERTAS KERJA

Ringkasan

Kertas kerja ini merupakan material kelengkapan artikel jurnal dengan judul “PERBANDINGAN ALGORITMA *SVM* DAN *NAÏVE BAYES* DALAM ANALISA SENTIMEN TERHADAP PENGGUNA *TWITTER* MENGENAI VAKSIN *BOOSTER*”. Kertas kerja berisi semua material hasil penelitian Tugas Akhir yang tidak dimuat/atau disertakan di artikel jurnal. Di dalam kertas kerja ini disajikan: literature review, dataset yang digunakan, source code, dan hasil eksperimen secara keseluruhan.

1. Bagian 1, berisi *Literature Review* yang menjabarkan mengenai beberapa artikel yang berkaitan dengan penelitian.
2. Bagian 2, Analisis dan Perancangan yang menjabarkan analisis masalah dan perancangan model.
3. Bagian 3, *Source Code* berisi mengenai *script* bahasa pemrograman yang digunakan untuk mengumpulkan data, membaca data, *pre-processing* data, *feature selection*, implementasi algoritma, pelatihan dan pengujian, dan evaluasi.
4. Bagian 4, *Dataset* berisi penjelasan mengenai data yang digunakan untuk penelitian ini.
5. Bagian 5, Tahapan Eksperimen berisi tahapan yang dilakukan dalam penelitian ini mulai dari pengumpulan data, *pre-processing*, *feature selection*, implementasi algoritma, dan evaluasi.
6. Bagian 6, Hasil Semua Eksperimen berisi penjelasan eksperimen secara keseluruhan dari semua percobaan yang dilakukan melalui tahapan yang terdapat pada bagian 5. Bagian ini menjelaskan mengenai percobaan untuk menemukan implementasi algoritma terbaik dengan melakukan percobaan menggunakan beberapa skema algoritma yang digunakan.