



**ANALISA SENTIMEN TERHADAP PENGGUNAAN
APLIKASI *SHOPEE FOOD* PADA TWITTER
MENGUNAKAN METODE NAÏVE BAYES DAN SUPPORT
VECTOR MACHINE (SVM)**

TUGAS AKHIR

Muhammad Yoga Pratama
41518010052

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



**ANALISA SENTIMEN TERHADAP PENGGUNAAN
APLIKASI *SHOPEE FOOD* PADA TWITTER
MENGUNAKAN METODE NAÏVE BAYES DAN SUPPORT
VECTOR MACHINE (SVM)**

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:

Muhammad Yoga Pratama

41517010052

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

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM : 41518010052

Nama : Muhammad Yoga Pratama

Judul Tugas Akhir : Analisa Sentimen Terhadap Penggunaan Aplikasi *Shopee food* Pada Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine (SVM)

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, 03 agustus 2022



UNIVERSITAS
MERCU BUANA
Muhammad Yoga Pratama

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

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

Nama Mahasiswa : Muhammad Yoga Pratama
NIM : 41518010052
Judul Tugas Akhir : Analisa Sentimen Terhadap Penggunaan Aplikasi *Shopee food* Pada Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine (SVM)

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, 03 agustus 2022



Muhammad Yoga Pratama

SURAT PERNYATAAN LUARAN TUGAS AKHIR

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

Nama Mahasiswa : Muhammad Yoga Pratama
NIM : 41518010052
Judul Tugas Akhir : Analisa Sentimen Terhadap Penggunaan Aplikasi
Shopee food Pada Twitter Menggunakan Metode
Naïve Bayes dan Support Vector Machine (SVM)

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 Internasional Tidak Bereputasi	
		Jurnal Internasional Bereputasi	
Disubmit/dipublikasikan di :	Nama Jurnal	: Jurnal Informatika: Jurnal Pengembangan IT	
	ISSN	: 2477-5126	
	Link Jurnal	https://ejournal.polteksegi.ac.id/index.php/informatika/author/submission/4143	
	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

a.n.



Leonard Goefrianto, Dr., ST, M.Sc

Jakarta, 03 Agustus 2022



Muhammad Yoga Pratama

LEMBAR PERSETUJUAN PENGUJI

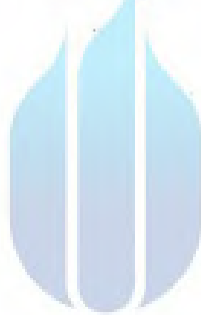
NIM : 41518010052
Nama : Muhammad Yoga Pratama
Judul Tugas Akhir : Analisa Sentimen Terhadap Penggunaan Aplikasi *Shopee food* Pada Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine (SVM)

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, 03 Agustus 2022



(Rushendra S.Kom, M.T)



UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010052
Nama : Muhammad Yoga Pratama
Judul Tugas Akhir : Analisa Sentimen Terhadap Penggunaan Aplikasi *Shopee food* Pada Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine (SVM)

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, 03 Agustus 2022



(Dr.Ir.Eliyani)
UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010052
Nama : Muhammad Yoga Pratama
Judul Tugas Akhir : Analisa Sentimen Terhadap Penggunaan Aplikasi *Shopee food* Pada Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine (SVM)

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, 03 Agustus 2022



(Anis Cherid, SE, MTI)

UNIVERSITAS
MERCU BUANA

LEMBAR PENGESAHAN

NIM : 41518010052
Nama : Muhammad Yoga Pratama
Judul Tugas Akhir : Analisa Sentimen Terhadap Penggunaan Aplikasi *Shopee food* Pada Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine (SVM)

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, 03 Agustus 2022

Menyetujui,



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

UNIVERSITAS

MERCU BUANA

Mengetahui,



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



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

KATA PENGANTAR

Terimakasih dan rasa syukur saya panjatkan kepada Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, dikarenakan Tugas Akhir yang berjudul “Sentiment Analysis of face-to-face education in Indonesia on Twitters” dapat diselesaikan dalam jangka waktu yang sudah ditentukan. Laporan Tugas Akhir ini dibuat sebagai syarat untuk LULUS sebagai sarjana Ilmu Komputer dari Universitas Mercu Buana.

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

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

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

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

Jakarta, 03 Agustus 2022

Muhammad Yoga Pratama



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. <i>SOURCE CODE</i>	244
BAB 4. <i>DATASET</i>	25
BAB 5. TAHAPAN EKSPERIMEN	26
BAB 6. HASIL SEMUA EKSPERIMEN.....	31
DAFTAR PUSTAKA	34
LAMPIRAN DOKUMEN HAKI.....	36
LAMPIRAN KORESPONDENSI	37

NASKAH JURNAL

ANALISA SENTIMEN TERHADAP PENGGUNAAN APLIKASI SHOPEE FOOD PADA TWITTER MENGGUNAKAN METODE NAÏVE BAYES DAN SUPPORT VECTOR MACHINE (SVM)

Muhammad Yoga Pratama¹, Leonard Goeiranto²

^{1,2} Faculty of Computer Science, University of Mercu Buana, Indonesia

¹141518010052@student.mercubuana.ac.id. ²leonard@mercubuana.ac.id

Abstract: In the era of the Industrial Revolution 4.0 now, many business people are utilizing technology to market their merchandise, and one of the applications that may be familiar is the Shopee application. Among them, the shopee application is not only engaged in buying and selling daily necessities on the internet. As of 2020, the app had over 50 million downloads, with various positive or negative reviews. The number of reviews or reviews on the Shopee app on Google play has a huge impact on Shopee of all types by including positive, neutral, or negative reviews that provide evaluation material to Shopee. Among them, the shopee application itself has launched the shopee food function which makes it easier for shopee users who initially only buy daily necessities online, and provide food supply services for shopee application users. Sentiment analysis is a branch of science in text mining, natural language programming, and artificial intelligence. Sentiment analysis itself or what is commonly referred to as opinion mining is part of text mining. This field studies people's opinions, emotions, evaluations, behaviors, and emotions about entities such as products, services, organizations, individuals, problems, topics, events, and their attributes. Shopee Food's current problem is that there is still a sense of dissatisfaction for customers, including shopee food does not have regular fans, and the ease of use of the application is still low. In the previous study entitled "Implementation of the Naive Bayes Algorithm for Sentiment Analysis of Shopee Reviews on the Google Play Store" it was

found that the sharing of Hold Out data with a ratio of 80:20 resulted in an algorithm accuracy of 83% and In another study entitled "Implementation of the Naïve Bayes Classifier and Association for Sentiment Analysis of Shopee E-Commerce Application Review Data on Google Play Sites" that the study used the Naïve Bayes model The classifier obtained an accuracy rate of 97.4%. Aims to increase services for shopee food application users.

Keywords: Shopee Food, Naïve Bayes, Support Vector Machine, Algorithm, Sentiment Analysis

I. PENDAHULUAN

Perkembangan pesat yang terjadi pada bidang teknologi merupakan cikal bakal pertumbuhan aplikasi mobile di Indonesia. Dimana penggunaan Smartphone di Indonesia mencapai 25% dari total penduduk sekitar 65 juta orang, sehingga banyak masyarakat memanfaatkan peluang tersebut dengan berlomba-lomba untuk menciptakan aplikasi mobile yang dapat mempermudah dalam penggunaan smartphone di kehidupan sehari-hari.

Pada akhir tahun 2019 ditemukan sebuah virus baru yang berasal dari Wuhan, China, yaitu virus Corona (COVID-19). Semakin luasnya penyebaran Virus ini termasuk di Indonesia, membuat pemerintah Indonesia mengupayakan berbagai cara untuk mencegah penyebaran Virus ini di Indonesia. Salah satu upaya yang dilakukan pemerintah Indonesia adalah dengan menetapkan PSBB atau Pembatasan Sosial Berskala Besar.

Dengan ditetapkannya PSBB, ada banyak aspek dan bidang yang terdampak dari kebijakan

pemerintah tersebut, diantaranya adalah pendidikan dan kegiatan perkantoran. Seluruh sekolah termasuk perguruan tinggi di Indonesia ditutup, mengharuskan siswa serta mahasiswa dianjurkan untuk belajar secara online dari rumah dan juga banyak kantor yang harus menerapkan kebijakan WFH (Work From Home) atau kerja dari rumah secara daring.

Di era Revolusi industry 4.0 saat ini sudah banyak pelaku bisnis yang sudah memanfaatkan teknologi dalam memasarkan barang dagangannya dan salah satu aplikasi yang mungkin sudah tidak asing lg yaitu aplikasi Shopee. Dimana diaplikasi shopee saat ini sudah tidak hanya bergerak dibidang jual beli barang kebutuhan sehari-hari secara online. Pada tahun 2020 aplikasi ini sudah didownload sebanyak 50 juta lebih dengan rating 4+ dengan berbagai ulasan baik ataupun buruk.

Banyaknya ulasan atau komentar yang ada di aplikasi Shopee pada Google play membawa berbagai macam pengaruh yang besar bagi pihak shopee dengan meliputi komentar positif, atau negative, dimana komentar tersebut menjadikan bahan evaluasi bagi pihak shopee. Dimana aplikasi shopee sendiri telah meluncurkan fitur shopee food yang dimana mempermudah pengguna shopee yang awalnya hanya berbelanja kebutuhan sehari-hari online dan menjadikan jasa penyediaan makanan bagi para pengguna aplikasi shopee.

Analisa sentimen merupakan salah satu cabang ilmu dari textmining, natural language program dan artificial intelligence. Analisis sentimen sendiri atau juga biasa disebut dengan opinion mining adalah salah satu bagian dari text mining. Bidang ini melakukan studi mengenai opini orang-orang, sentimen, evaluasi, tingkah laku dan emosi terhadap suatu entitas seperti produk, layanan, organisasi, individu, permasalahan, topik, acara dan atribut- atributnya. Permasalahan yang masih terdapat di shopee food masih ada rasa kurang memuaskan bagi para

pelanggannya, Diantaranya ialah masih kurangnya peminat tetap di shopee food, aplikasi masih kurang useable.

Pada penelitian sebelumnya yang berjudul "Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store" mendapatkan hasil bahwa pembagian data Hold Out dengan perbandingan 80:20 menghasilkan akurasi algoritma sebesar 83%, yang lebih unggul 1% dibandingkan dengan hasil akurasi rata-rata dengan menggunakan teknik pembagian data 10-Cross Fold Validation yaitu sebesar 82%[1].

Pada penelitian lainnya yang berjudul "Implementasi Naive Bayes Classifier dan Asosiasi untuk Analisis Sentimen Data Ulasan Aplikasi E-Commerce Shopee Pada Situs Google Play" bahwa penelitian tersebut menggunakan model Naive Bayes Classifier didapatkan tingkat akurasi sebesar 97,4%. Selanjutnya, pada proses asosiasi teks diperoleh informasi bahwa pengguna aplikasi Shopee Indonesia membicarakan aplikasi, ongkir, harga dan puas untuk kelas positif dan untuk kelas negatif diperoleh informasi pengguna banyak membicarakan aplikasi ongkir, buruk, susah, gambar, pengiriman, dan penjual. Hasil ulasan negatif tersebut dibuat dalam diagram fishbone untuk pemecahan masalah[2].

Dari permasalahan tersebut, penulis mencoba untuk melakukan penelitian mengenai sentimen analisa atau pendapat masyarakat terhadap Aplikasi Shopee Food untuk meningkatkan pelayanan kepada pengguna aplikasi Shopee Food.

II. PENELITIAN YANG TERKAIT

A. Penelitian Terkait

R. Ardianto, T. Rivanie, Y. Alkhalifi, F. S. Nugraha, and W. Gata[3]. Perbandingan tingkat

Muhammad Yoga Pratama
41518010052@student.mercubuana.ac.id

akurasi, presisi dan recall antara Naive Bayes dengan SVM menggunakan metode SMOTE. Perbedaan akurasi antara Naive Bayes dan Support Vector Machine dapat diamati sebesar 3,4%.

H. A. Santoso, E. H. Rachmawanto, A. Nugraha, A. A. Nugroho, D. R. I. M. Setiadi, and R. S. Basuki[4]. Hasil pengujian 30 berita, rata-rata hoax

adalah 77%, dimana semua berita terdeteksi sebagai hoax dengan persentase minimal sekitar 66% dan maksimal 91%. Hal ini menunjukkan bahwa kinerja metode yang diusulkan cukup handal untuk mendeteksi berita hoax. Sistem ini juga dilengkapi dengan proses analisis sentimen menggunakan Naïve Bayes yang dioptimalkan dengan metode PSO.

J. Samuel, G. G. M. N. Ali, M. M. Rahman, E. Esawi, and Y. Samuel[5]. Perbandingan tingkat akurasi antara Naïve bayes dengan Logistic regression. LR memiliki kinerja yang lebih baik dengan klasifikasi Tweet yang seimbang. Sedangkan Naïve Bayes memiliki kinerja yang lebih baik dengan mengklasifikasikan Tweet negatif.

M. Nur Habibi and Sunjana[6]. Berdasarkan hasil klasifikasi sentimen polarisasi politik Indonesia, pada dataset Jokowi (92,65% sentimen positif, 5,88% sentimen negatif) dan Prabowo (89,99% sentiment positif, 9,06% sentiment negatif)

B. Algoritma Klasifikasi

Text mining adalah salah satu subjek data mining, karena teks juga adalah data. Namun karena cakupannya sangat luas sehingga penggunaan istilah Text Mining dan Data Mining dibedakan. Data mining menambang pengetahuan dari kumpulan data yang banyak dan biasanya terstruktur. Perbedaannya terdapat dari tugas data miningnya, seperti klasifikasi, klasterisasi, asosiasi, estimasi atau prediksi, bukan pada karakteristik data yang diolahnya. Text mining merupakan proses menambang data yang berupa teks dimana sumber data biasanya didapatkan dari dokumen dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen[7]

Analisis Sentimen adalah kegiatan melakukan analisis terhadap pendapat, opini, sikap atau emosi seseorang mengenai suatu produk, topik atau permasalahan tertentu sehingga bisa diketahui hal tersebut masuk kedalam sentimen positif, negatif[8].

Metode Klasifikasi adalah teknik dalam data mining untuk mengelompokan data berdasarkan keterikatan data terhadap data sampel, Metode Klasifikasi juga suatu proses pengklasifikasian

dokumen-dokumen kedalam satu atau lebih kategori yang telah didefinisikan sebelumnya atau kedalam kelas-kelas dari dokumen- dokumen yang sama[9].

Algoritma Naive Bayes merupakan salah satu algoritma yang terdapat dalam teknik klasifikasi. Naive Bayes adalah klasifikasi dengan metode probabilitas dan statistik yang disajikan oleh ilmuwan Inggris Thomas Bayes, yang memprediksi peluang masa depan berdasarkan pengalaman masa lalu sehingga dikenal dengan Teorema Bayes. Teorema tersebut digabungkan dengan Naive dimana diasumsikan bahwa kondisi antar atribut adalah saling mandiri. Klasifikasi Naive Bayes mengasumsikan bahwa tidak ada fitur khusus dari kelas yang tidak memiliki hubungan dengan karakteristik kelas lain[10]. Naïve Bayes Classifier dalam melakukan klasifikasi terdapat dua proses penting yaitu learning (training) dan testing[11].

Support Vector Machine (SVM) adalah suatu teknik yang relatif baru untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi. Penelitian ini berpusat pada klasifikasi biner yang terdiri dua kelas. Kelas tersebut adalah positive atau negative . Konsep klasifikasi dengan Support Vector Machine adalah mencari hyperplane terbaik yang berfungsi sebagai pemisah dua kelas data. Algoritma Support Vector Machine bekerja dengan memisahkan data secara optimal menggunakan pengukuran jarak hyperplane dari titik terdekat daripada mencari titik maksimum untuk memaksimalkan jarak antar label kelas berdasarkan pengukuran batas keanggotaan kelas[12].

III. DATASET

Pada penelitian ini, dataset yang diperoleh merupakan data Twitter. Proses crawling data dibantu dengan menggunakan tools Rapidminer melalui API Twitter. Data yang diambil merupakan tweet yang relevan dengan topik dan studi kasus yang diambil, dalam hal ini adalah Pendidikan Tatap Muka menggunakan kata kunci yang relevan. Kata kunci yang dipakai pada penelitian ini adalah “shopee”, “shopee food”, “promo shopee food”, “shopee error” untuk studi

kasus Aplikasi Shopee Food. Pada penelitian ini data disimpan dengan format .csv dengan total record data untuk studi kasus Aplikasi Shopee Food sebanyak 2535.

C. Preprocessing Data

Preprocessing merupakan pengolahan awal data dan mempersiapkan data teks untuk dilakukan proses klasifikasi. Tahap Preprocessing dapat dilakukan langsung di aplikasi Rapid Miner maupun menggunakan tools bantuan lainnya. Seluruh proses melibatkan beberapa langkah: membersihkan teks online, penghapusan ruang spasi, memperluas singkatan, kata dasar (stemming), penghapusan kata henti (stopword removal), penanganan negasi dan terakhir seleksi fitur.

. Berikut adalah penjelasan dari lima langkah Preprocessing:

- **Cleansing Data** : Cleansing dalam penelitian ini bertujuan untuk menghapus RT, username, hashtag, angka dan URL pada tweets.

Contoh tweet : RT @Kemenag_RI: Salam #SahabatReligi, Universitas Islam Internasional Indonesia (UIII) kembali membuka peluang bagi kamu yang ingin melanjutkan

Hasil Cleansing : Salam Universitas Islam Internasional Indonesia (UIII) kembali membuka peluang bagi kamu yang ingin melanjutkan

- **Case Folding** : Case Folding dalam penelitian ini bertujuan untuk merubah Teks dari penulisan huruf besar menjadi penulisan huruf kecil.

Hasil Case Folding : salam universitas islam internasional indonesia (uiii) kembali membuka peluang bagi kamu yang ingin melanjutkan

- **Tokenization** : Tokenization dalam penelitian ini bertujuan untuk memisahkan kata kata dari dalam kalimat.

Hasil Tokenization : salam, universitas, islam, internasional, indonesia, (uiii), kembali, membuka, peluang, bagi, kamu, yang, ingin, melanjutkan,

- **Stopword Removal** : Stopword Removal dalam penelitian ini bertujuan untuk menghapus kata sambung[13].

Contoh dari Stopword Removal bahasa Indonesia adalah sebagai berikut : yang, juga, dari, dia, kami, kamu, aku, saya, ini, itu, atau, dan, pada, dengan, adalah, yaitu, ke, tak, tidak, di, pada, jika, maka, ada, pun, lain, saja, hanya, namun, seperti, kemudian, dll[14].

Hasil Stopword Removal : salam, universitas, islam, internasional, indonesia, (uiii), kembali, membuka, peluang, bagi, kamu, yang, ingin, melanjutkan,

- **Stemming** : Stemming dalam penelitian ini bertujuan untuk membentuk kata kata dasar.

Hasil Stemming : salam, universitas, islam, internasional, indonesia, kembali, membuka, peluang, bagi, kamu, yang, ingin, melanjutkan,

Remove Duplicate : Remove Duplicate dalam penelitian ini bertujuan untuk menghapus data yang memiliki duplikat.

Setelah melakukan tahap preprocessing, data yang telah dipreprocessing menjadi 1996 untuk data Aplikasi Shopee Food. Berikut merupakan hasil data setelah dilakukan tahapan preprocessing:



Retweet Count	Tweet Text
1529	
1915	a shopee mudou o frete gratis pra reais ah vai...
613	aaa cd playerrr bukak shopee tgk harga tutup
1089	aanhin ko naman tong suklay ni jungwon marami ...
185	abang shopee food nya bawa doi anjir dua gabut
...	...
58	yuk order kak
1600	yukk fee cv koin shopee langsung dm wa fastresp
1557	yukkk kak freeong shopee
1921	yung sa akin halos friends na nga isipin mo ta...
682	zonauang wts vo shopee belanja min rb

Gambar 1. Preprocessing Pendidikan Tatap Muka tanpa Labeling

D. Labeling

Pada tahap ini, data yang telah dibersihkan selama preprocessing akan ditandai sebagai kelas data. Label kelas data dibagi menjadi kelas positif dan negatif. Proses pelabelan dilakukan

menggunakan bahasa R melalui alat Google Colab. Proses tagging menggunakan bahasa R membutuhkan kamus bahasa Indonesia yang terintegrasi dengan Google Drive. Kamus Bahasa Indonesia yang digunakan terdiri dari dua kamus, kamus bahasa Indonesia kata positif dan kamus bahasa Indonesia kata negatif.

Selama fase penandaan, setiap kalimat diberi skor. Setiap kata yang terdeteksi diberi skor untuk mengevaluasi kategori sentimen. Ini memiliki nilai 1 untuk kata-kata positif dan -1 untuk kata-kata negatif. Memberi nilai 0 jika sebuah kata tidak ada dalam kamus positif atau negatif. Scoring dilakukan dengan menghitung nilai total setiap kata dalam sebuah kalimat. Jika nilainya ≥ 0 , ditandai sebagai sentimen tweet positif, sebaliknya jika nilainya < 0 , ditandai sebagai sentimen tweet negatif. Jika skornya 0, tandai sebagai tweet positif. Kelas data positif akan diberi nilai 1, sedangkan kelas data negatif akan diberi nilai 0.

Table I. Nilai Klasifikasi Otomatis

Sentimen	Positif	Negatif
Aplikasi Shopee Food	1385	610

Pada penelitian ini juga dilakukan eksperimen dengan menggunakan labeling manual terhadap dataset tersebut. Label positif merupakan tweet yang memiliki pujian sedangkan tweet negative berupa kata-kata kasar dan sarkas.

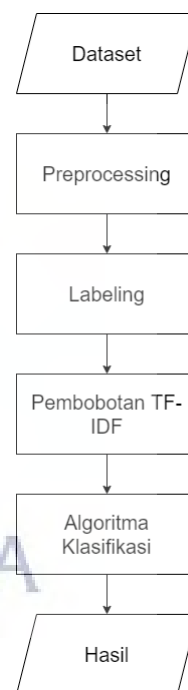
E. TF-IDF

Pembobotan TF-IDF (Term Frequency - Inverse Document Frequency) adalah proses konversi data dari data tekstual ke data numerik untuk memberikan bobot setiap kata atau fitur. TF-IDF adalah ukuran statistik yang digunakan untuk mengevaluasi pentingnya kata-kata dalam sebuah dokumen. TF adalah frekuensi sebuah kata dalam setiap dokumen yang diberikan, yang menunjukkan pentingnya kata tersebut dalam setiap dokumen. DF adalah frekuensi dokumen yang berisi kata, menunjukkan seberapa umum kata tersebut. IDF adalah kebalikan dari nilai DF. Hasil pembobotan kata dengan TF-IDF adalah hasil kali TF kali IDF. Kata-kata yang sering muncul dalam satu dokumen

memiliki bobot yang lebih besar, dan kata-kata yang muncul di beberapa dokumen memiliki bobot yang lebih kecil[15].

IV. METODE

Penelitian ini menggunakan metode kuantitatif yang dimana *Dataset* yang telah terkumpul diolah melalui Twitter API menggunakan tools *Rapid Miner*. Dalam penelitian ini, python dan r tools dibantu untuk melakukan *Preprocessing* dan labeling pada *Google Colab*. Pada penelitian ini juga dilakukan eksperimen dengan membandingkan kinerja algoritma klasifikasi Naïve Bayes dan Support Vector Machine dan juga *K-Fold Validation* untuk mengetahui algoritma mana yang memiliki tingkat akurasi terbaik. Alur penelitian diilustrasikan pada Gambar 2.



Gambar 2. Tahapan Penelitian

Tahapan pertama yang dilakukan adalah mengumpulkan *Dataset*. Selanjutnya pada tahapan berikutnya dilakukan tahapan *Preprocessing* untuk membersihkan *Dataset* mentah. Setelah data selesai dilakukan *Preprocessing*, tahapan berikutnya adalah dilakukan labeling untuk mengetahui kelas positif dan negatif. Setelah itu dilakukan tahapan pembobotan *TF-IDF* untuk memberikan bobot pada setiap kata yang terdapat di dalam dokumen. Selanjutnya, dilakukan metode klasifikasi dengan

menggunakan algoritma Naïve Bayes, Support Vector Machine dan juga *K-Fold Cross Validation*.

Sebelum tahap klasifikasi dilakukan, data terlebih dahulu dibagi menjadi data latih dan data uji. Data latih tersebut digunakan untuk proses pembelajaran pada algoritma Naïve Bayes dan Support Vector Machine. Langkah selanjutnya adalah menggunakan data uji untuk proses pengujian model klasifikasi Naïve Bayes dan Support Vector Machine. Pada penelitian ini digunakan algoritma Support Vector Machine karena cocok untuk klasifikasi teks dan cara kerja algoritma yang dapat mengatasi data outlier. Sedangkan algoritma Naïve Bayes cocok untuk data teks karena Naïve Bayes menggunakan metode probabilitas, peluang kata muncul.

V. HASIL SEMUA EKSPERIMEN

Pada bagian ini akan menampilkan hasil semua eksperimen yang telah dilakukan. Pada penelitian ini melakukan eksperimen pada studi kasus yaitu Pendidikan Tatap Muka. Eksperimen dilakukan dengan menggunakan dua algoritma yaitu Support Vector Machine dan Naïve Bayes. Terdapat tiga skenario yang menjadi pembagian persentase data dan juga *10-Fold Cross Validation*. Pembagian presentasi data merupakan pemisahan data *training* dan testing berdasarkan persentasenya, contoh 90%:10% berarti 90% merupakan data *training* dan 10% merupakan data testing. Pada penelitian ini pembagian persentase dibagi menjadi tiga skenario eksperimen yaitu eksperimen pertama menggunakan pembagian data 90%:10%, eksperimen kedua menggunakan pembagian data 80%:20% dan eksperimen yang ketiga menggunakan pembagian data 70%:30%. Penelitian ini juga membandingkan nilai akurasi yang dilakukan dengan menggunakan metode labeling manual dan metode labeling otomatis.

Hasil dari kinerja model dengan menggunakan metode labeling otomatis dan pembagian data atau percentage *Split* 90%:10% bahwa studi kasus Aplikasi Shopee Food memiliki nilai akurasi 89% untuk algoritma SVM dan 80% untuk algoritma Naïve Bayes.

Table II. SVM Aplikasi Shopee Food *Split* 90:10 Labeling Otomatis

	Precision	Recall	Accuracy
0	92%	64%	76%
1	88%	98%	93%

Table III. Naïve Bayes Aplikasi Shopee Food *Split* 90:10 Labeling Otomatis

	Precision	Recall	Accuracy
0	71%	69%	70%
1	84%	86%	85%

Hasil dari kinerja model dengan menggunakan metode labeling otomatis dan pembagian data atau percentage *Split* 80%:20% bahwa studi kasus Aplikasi Shopee Food memiliki nilai akurasi 89% untuk algoritma SVM dan 84% untuk algoritma Naïve Bayes.

Table IV.

	Precision	Recall	Accuracy
0	94%	65%	77%
1	88%	98%	93%

Table V. Naïve Bayes Aplikasi Shopee Food *Split* 80:20 Labeling Otomatis

	Precision	Recall	Accuracy
0	77%	72%	75%
1	88%	90%	89%

Hasil dari kinerja model dengan menggunakan metode labeling otomatis dan pembagian data atau percentage *Split* 70%:30% bahwa untuk studi kasus Aplikasi Shopee Food memiliki nilai akurasi 86% untuk algoritma SVM dan 84% untuk algoritma Naïve Bayes.

Table VI. SVM Aplikasi Shopee Food *Split* 70:30 Labeling Otomatis

	Precision	Recall	Accuracy
0	96%	59%	73%
1	84%	99%	91%

Table VII. Naïve Bayes Aplikasi Shopee Food *Split* 70:30 Labeling Otomatis

	Precision	Recall	Accuracy
0	74%	71%	73%
1	88%	89%	89%

Dari tabel berikut disimpulkan bahwa eksperimen dengan menggunakan labeling otomatis dalam menentukan kelas data. Model percentage split 80:20 memiliki nilai akurasi terbaik pada data yang diuji dengan nilai akurasi 89% dan 84% pada

eksperimen labeling otomatis memiliki hasil yang lebih baik karena memberikan nilai evaluasi yang mendekati keseimbangan setelah proses penyetalan. 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.

VI. KESIMPULAN

Dari table berikut disimpulkan bahwa eksperimen dengan menggunakan labeling otomatis lebih representative dalam menentukan kelas data. Model percentage split 80:20 memiliki nilai akurasi terbaik pada data yang diuji pada aplikasi shopee food dengan nilai akurasi 89% pada eksperimen labeling otomatis sehingga bisa disimpulkan model percentage split 80:20 dengan menggunakan algoritma Support Vector Machine merupakan skenario eksperimen yang paling baik pada penelitian ini.

DAFTAR PUSTAKA

- [1] J. W. Iskandar and Y. Nataliani, "Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 6, pp. 1120–1126, 2021, doi: 10.29207/resti.v5i6.3588.
- [2] N. A. C. A, D. H. Citra, and W. Purnama, "The Implementation of Naïve Bayes Algorithm for Sentiment Analysis of Shopee Reviews on Google Play Store Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store," vol. 2, no. April, pp. 47–54, 2022.
- [3] R. Ardianto, T. Rivanie, Y. Alkhalifi, F. S. Nugraha, and W. Gata, "Sentiment Analysis on E-Sports for Education Curriculum Using Naive Bayes and Support Vector Machine," *J. Ilmu Komput. dan Inf.*, vol. 13, no. 2, pp. 109–122, 2020, doi: 10.21609/jiki.v13i2.885.
- [4] H. A. Santoso, E. H. Rachmawanto, A. Nugraha, A. A. Nugroho, D. R. I. M. Setiadi, and R. S. Basuki, "Hoax classification and sentiment analysis of Indonesian news using Naive Bayes optimization," *Telkomnika (Telecommunication Comput. Electron. Control.)*, vol. 18, no. 2, pp. 799–806, 2020, doi: 10.12928/TELKOMNIKA.V18I2.14744.
- [5] J. Samuel, G. G. M. N. Ali, M. M. Rahman, E. Esawi, and Y. Samuel, "COVID-19 public sentiment insights and machine learning for tweets classification," *Inf.*, vol. 11, no. 6, pp. 1–22, 2020, doi: 10.3390/info11060314.
- [6] M. Nur Habibi and Sunjana, "Analysis of Indonesia Politics Polarization before 2019 President Election Using Sentiment Analysis and Social Network Analysis," *Int. J. Mod. Educ. Comput. Sci.*, vol. 11, no. 11, pp. 22–30, 2019, doi: 10.5815/ijmecs.2019.11.04.
- [7] W. Yulita *et al.*, "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier," *Jdmsi*, vol. 2, no. 2, pp. 1–9, 2021.
- [8] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," *Integer J.*, vol. 2, no. 1, pp. 32–41, 2017, [Online]. Available: <https://t.co/jrvaMsgBdH>
- [9] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018, doi: 10.33096/ilkom.v10i2.303.160-165.
- [10] S. Dwiasnati and Y. Devianto, "Utilization of Prediction Data for Prospective Decision Customers Insurance Using the Classification Method of C.45 and Naive Bayes Algorithms," *J. Phys. Conf. Ser.*, vol. 1179, no. 1, 2019, doi: 10.1088/1742-6596/1179/1/012023.
- [11] M. Rasyida, "Naïve Bayes Classification untuk Penentuan Status Penduduk Miskin," *J. Inform. Kaputama(JIK)*, vol. 4, no. 2, 2020.
- [12] D. A. Kristiyanti, Normah, and A. H. Umam, "Prediction of Indonesia presidential election results for the 2019-2024 period using twitter sentiment analysis," *Proc. 2019 5th Int. Conf. New Media Stud. CONMEDIA 2019*, pp. 36–42, 2019, doi: 10.1109/CONMEDIA46929.2019.8981823.

- [13] R. Tineges, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)," *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 650, 2020, doi: 10.30865/mib.v4i3.2181.
- [14] F. Syadid, "Analisis Sentimen Komentar Netizen Terhadap Calon Presiden Indonesia 2019 Dari Twitter Menggunakan Algoritma Term Frequency-Invers Document Frequency (Tf-Idf) Dan Metode Multi Layer Perceptron (Mlp) Neural Network," *Skripsi Univ. Islam Negeri Syarif Hidayatullah Jakarta*, p. 72, 2019.
- [15] Rian Hidayat, "Uin Syarif Hidayatullah Jakarta Uin Syarif Hidayatullah Jakarta," no. 95, pp. 1–28, 2017, [Online]. Available: http://repository.uinjkt.ac.id/dspace/bitstream/123456789/33026/1/NITA_FITRIANI-FKIK.pdf



KERTAS KERJA

Ringkasan

Kertas kerja ini berisi tentang kelengkapan material dari artikel jurnal dengan judul “Analisa Sentimen Mengenai Aplikasi *Shopee food* Pada Twitter Menggunakan Metode Naive Bayes Dan Support Vector Machine”. Seluruh hasil penelitian Tugas Akhir yang tidak dimasukkan kedalam artikel jurnal. Pada kertas kerja ini disajikan: literatur review, analisis perancangan, *source code*, *dataset* yang digunakan, tahapan eksperimen dan seluruh hasil eksperimen.

- Bagian 1: Literature Review menjabarkan mengenai beberapa jurnal yang terkait dengan penelitian.
- Bagian 2: Analisis dan Perancangan menjabarkan analisis masalah dan analisis model.
- Bagian 3: *Source code* menjabarkan kumpulan kode di setiap prosesnya mulai dari Membaca data, Preprocessing, *Labeling*, Implementasi model, sampai pengujian dan evaluasi.
- Bagian 4: *Dataset* melampirkan pengambilan *dataset*.
- Bagian 5: Tahapan eksperimen berisi pengumpulan data, Preprocessing, Lebeling, Implementasi model, dan evaluasi model.
- Bagian 6: Hasil implementasi eksperimen secara keseluruhan yang mencakup jawaban dari skenario percobaan pengujian yang sudah disebutkan pada bagian 5. Bagian ini berguna untuk menemukan model Analisa sentimen terbaik yang dilihat dari beberapa pengujian berdasarkan Opini masyarakat yang positif dan negatif. Kemudian, hasil yang didapat digunakan berupa akurasi, presisi dan recal Penggunaan Aplikasi *Shopee food* di Indonesia untuk bahan evaluasi atas kebijakan yang diterapkan.