



**Analisis Sentimen Pelanggan Terhadap Layanan ShopeeFood
Pada Media Sosial Twitter Menggunakan Algoritma *Naïve Bayes*
dan *Support Vector Machine* (SVM)**

TUGAS AKHIR

Kindi Setiadi
41519120004

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



**Analisis Sentimen Pelanggan Terhadap Layanan ShopeeFood
Pada Media Sosial Twitter Menggunakan Algoritma *Naïve Bayes*
dan *Support Vector Machine* (SVM)**

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:
Kindi Setiadi
41519120004

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

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM : 41519120004

Nama : Kindi Setiadi

Judul Tugas Akhir : Analisis Sentimen Pelanggan Terhadap Layanan
ShopeeFood Pada Media Sosial Twitter Menggunakan
Algoritma *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, 27 Juli 2022



Kindi Setiadi

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

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

Nama Mahasiswa : Kindi Setiadi
NIM : 41519120004
Judul Tugas Akhir : Analisis Sentimen Pelanggan Terhadap Layanan
ShopeeFood Pada Media Sosial Twitter
Menggunakan Algoritma *Naive 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, 27 Juli 2022



Kindi Setiadi

SURAT PERNYATAAN LUARAN TUGAS AKHIR

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

Nama Mahasiswa : Kindi Setiadi
NIM : 41519120004
Judul Tugas Akhir : Analisis Sentimen Pelanggan Terhadap Layanan
ShopeeFood Pada Media Sosial Twitter
Menggunakan Algoritma *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		Diajukan	✓
		Jurnal Nasional Terakreditasi	✓		
		Jurnal International Tidak Bereputasi		Diterima	
		Jurnal International Bereputasi			
Disubmit/dipublikasikan di :	Nama Jurnal	: KOMPUTA: Jurnal Ilmiah Komputer dan Informatika			
	ISSN	: 2089-9033			
	Link Jurnal	: https://ojs.unikom.ac.id/index.php/komputa			
	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, 27 Juli 2022



Kindi Setiadi

LEMBAR PERSETUJUAN

Nama Mahasiswa : Kindi Setiadi
NIM : 41519120004
Judul Tugas Akhir : Analisis Sentimen Pelanggan Terhadap Layanan
ShopeeFood Pada Media Sosial Twitter
Menggunakan Algoritma *Naïve Bayes* dan *Support
Vector Machine* (SVM)

Tugas Akhir ini telah diperiksa dan disetujui

Jakarta, 7 Juli 2022

Menyetujui,



(Afiyati, S.Si., MT)
Dosen Pembimbing

LEMBAR PERSETUJUAN PENGUJI

NIM : 41519120004
Nama : Kindi Setiadi
Judul Tugas Akhir : Analisis Sentimen Pelanggan Terhadap Layanan
ShopeeFood Pada Media Sosial Twitter
Menggunakan Algoritma *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, 27 Juli 2022



(Puji Rahayu, Dr, MT)

LEMBAR PERSETUJUAN PENGUJI

NIM : 41519120004
Nama : Kindi Setiadi
Judul Tugas Akhir : Analisis Sentimen Pelanggan Terhadap Layanan
ShopeeFood Pada Media Sosial Twitter
Menggunakan Algoritma *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, 27 Juli 2022



(Wawan Sunawan, S.Kom, MT)

LEMBAR PERSETUJUAN PENGUJI

NIM : 41519120004
Nama : Kindi Setiadi
Judul Tugas Akhir : Analisis Sentimen Pelanggan Terhadap Layanan
ShopeeFood Pada Media Sosial Twitter
Menggunakan Algoritma *Naive 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, 27 Juli 2022



(Sabar Rudiarto, M.Kom)

LEMBAR PENGESAHAN

NIM : 41519120004
Nama : Kindi Setiadi
Judul Tugas Akhir : Analisis Sentimen Pelanggan Terhadap Layanan
ShopeeFood Pada Media Sosial Twitter Menggunakan
Algoritma *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, 22 Agustus 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 kepada Allah SWT yang telah memberikan nikmat dan rahmatnya sehingga penulis dapat menyelesaikan laporan tugas akhir ini yang berjudul; “Analisis Sentimen Pelanggan Terhadap Layanan ShopeeFood Pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes dan Support Vector Machine (SVM)” dengan lancar. Tugas akhir ini disusun untuk memenuhi syarat memperoleh gelar sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana. Penulis menyadari bahwa 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.Si., MT selaku dosen pembimbing tugas akhir yang telah memberikan masukan saat bimbingan dalam penyusunan tugas akhir ini hingga selesai.
2. Bapak Ir. Emil R. Kaburuan, Ph.D., IPM selaku Kepala Prodi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.
3. 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.
4. Orang tua yang selalu memberikan dukungan secara penuh dan doa sehingga dapat menyelesaikan semua laporan tugas akhir dengan lancar.
5. Teman – teman dan rekan – rekan kerja yang selalu memberikan dukungan dan memotivasi dalam melakukan penulisan tugas akhir ini agar dapat terselesaikan dengan baik.

Akhir kata, penulis berharap tugas akhir ini dapat bermanfaat atau mungkin bisa dikembangkan menjadi hal yang lebih baik lagi.

Jakarta, 7 Juli 2022

Kindi Setiadi

DAFTAR ISI

HALAMAN SAMPUL.....	i
HALAMAN JUDUL	i
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR... iii	iii
SURAT PERNYATAAN LUARAN TUGAS AKHIR..... iv	iv
LEMBAR PERSETUJUAN	v
LEMBAR PERSETUJUAN PENGUJI..... vi	vi
LEMBAR PENGESAHAN	ix
ABSTRAK	x
ABSTRACT	xi
KATA PENGANTAR.....	xii
DAFTAR ISI.....	xiii
NASKAH JURNAL	1
KERTAS KERJA.....	11
BAB 1. LITERATUR REVIEW	12
BAB 2. ANALISIS DAN PERANCANGAN.....	20
BAB 3. SOURCE CODE	28
BAB 4. DATASET.....	33
BAB 5. TAHAPAN EKSPERIMEN.....	35
BAB 6. HASIL SEMUA EKSPERIMEN.....	43
DAFTAR PUSTAKA	49
LAMPIRAN DOKUMEN HAKI.....	53
LAMPIRAN KORESPONDENSI	54
CURRICULUM VITAE	56

NASKAH JURNAL

Analisis Sentimen Pelanggan Terhadap Layanan ShopeeFood Pada Media Sosial Twitter Menggunakan Algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM)Kindi Setiadi^{1, 2}

^{1,2} Teknik Informatika – Universitas Mercu Buana
Jl. Meruya Selatan No. 1 Kembangan, Kota Jakarta Barat, 11650
E-mail : 41519120004@student.mercubuana.ac.id^{1, 2}

Abstrak

Banyaknya pengguna internet dan teknologi yang semakin canggih saat ini, kita dapat dengan cepat menemukan apa yang kita inginkan melalui E-Commerce dan dapat digunakan sebagai peluang bisnis dengan membuka toko online. Salah satunya adalah Shopee yang telah merilis layanan ShopeeFood sebagai layanan pesan antar makanan secara online yang dapat memudahkan masyarakat. ShopeeFood memiliki kurang lebih dari 500 mitra merchant di Jabodetabek dalam setahun. Semakin banyaknya minat masyarakat terhadap layanan tersebut. Dalam hal ini masyarakat banyak memberikan pendapat mereka menggunakan media sosial terutama twitter. Tujuan penelitian ini adalah dapat menganalisis opini masyarakat yang nantinya dapat diklasifikasikan oleh metode naïve bayes dan support vector machine, kemudian hasil klasifikasi terdapat 2 kelas negatif dan positif. Metode penelitian yang digunakan adalah pengumpulan data, pembersihan data, dan klasifikasi. Adapun hasil performa analisis sentimen menggunakan algoritma naïve bayes dengan tingkat akurasi 90.43%, presisi 99.74%, *recall* 76.65%, dan *f1-score* 86.68%. Sedangkan algoritma support vector machine memiliki tingkat akurasi 99.27%, presisi 99.60%, *recall* 98.60%, dan *f1-score* 99.10%. Hasil perbandingan dari kedua algoritma tersebut dapat disimpulkan bahwa support vector machine memiliki akurasi yang lebih baik sebesar 99.27% dibandingkan algoritma naïve bayes sebesar 90.43%. Sehingga akurasi pada algoritma support vector machine dapat dimanfaatkan sebagai model yang dapat digunakan sebagai prediksi sebuah kalimat.

Kata kunci : shopeefood, analisis sentimen, naïve bayes, support vector machine

Abstract

Most people the internet and technologies more advanced now, we can quickly find what we want from e-commerce and can be used as online shop. One of them is Shopee which has release ShopeeFood service as service food delivery. ShopeeFood has more than 500 merchant partners in Jabodetabek in a year. Research purposes is can analyze opinion people and can be classified to naïve bayes and support vector machine and then the result has two class negative and positive. The research method used is data collection, data cleansing, and classification. As for the performance results of sentiment analysis using algorithm naïve bayes with level accuracy 90.43%, precision 99.74%, recall 76.65%, dan f1-score 86.68%. While the algorithm support vector machine have level accuracy 99.27%, presisi 99.60%, recall 98.60%, dan f1-score 99.10%. The results of the comparison of the two algorithms can be concluded that the support vector machine has a better accuracy of 99.27% than the naïve bayes algorithm of 90.43%. So that the accuracy of the support vector machine algorithm can be used as a model that can be used as a prediction of a sentence.

Keywords : shopeefood, sentiment analysis, naïve bayes, support vector machine

1. PENDAHULUAN

Di masa yang sangat berkembang ini, penggunaan internet tidak hanya untuk berkomunikasi tetapi juga untuk berbisnis, memesan transportasi, membeli makanan dan minuman, bekerja, bahkan membeli barang. Keinginan konsumen untuk membuat penilaian yang cepat akan memiliki pengaruh langsung atau tidak langsung terhadap perilaku konsumen saat membeli dan

mengonsumsi barang yang dibeli secara online. Karena banyaknya orang yang menggunakan internet, kita dapat dengan cepat menemukan apa yang kita inginkan melalui *E-Commerce*, yang juga dapat digunakan sebagai peluang bisnis dengan membuka toko online. Salah satu aplikasi *E-Commerce* yang dapat membantu memenuhi kebutuhan manusia, yaitu aplikasi Shopee.

Pada tahun 2017, Shopee melayani hingga 244,8 juta pesanan, dengan pertumbuhan 604,5 juta pesanan pada tahun 2018. Pada tahun 2018, sebanyak 206,9 juta pesanan dilakukan, naik menjadi 98,3 juta pada tahun 2017. Pembeli aktif Shopee berjumlah 49,9 juta pada tahun 2018, naik dari 21,7 juta pada tahun 2017, menunjukkan bahwa penjual aktif telah melonjak [1]. Pada April 2020, Shopee merilis ShopeeFood sebagai layanan *food delivery services*. ShopeeFood, memiliki fungsi tidak jauh beda seperti halnya Gofood dan Grabfood, yang merupakan layanan pesan antar makanan dan minuman. Dengan menambahkan fitur yang terdapat di ShopeeFood seperti pengiriman gratis, cash back, promosi, diskon 10%, 15%, 50%, 60%, dan 75% saat menggunakan Shopee Pay, atau diskon 100% untuk pelanggan baru, ShopeeFood telah menjadi sangat populer di kalangan konsumen[2].

Pada saat ini Twitter semakin populer sebagai alat komunikasi yang ampuh di kalangan pengguna internet. Berdasarkan konferensi resmi pengembang Twitter Chirp 2010, Twitter memiliki 106 juta akun dan 180 juta pengunjung unik bulanan pada April 2010. Jumlah pengguna Twitter diperkirakan meningkat 300.000 per hari[3]. Pentingnya peran media sosial seperti Twitter dalam memposting *tweet*, dapat mengetahui sentimen opini dari pelanggan dengan cara analisis sentimen. Analisis sentimen telah digunakan untuk menentukan reaksi pelanggan terhadap suatu produk, preferensi politik, dan pemikiran tentang film yang baru dirilis, serta proyeksi keuntungan film, dan sebagainya[4].

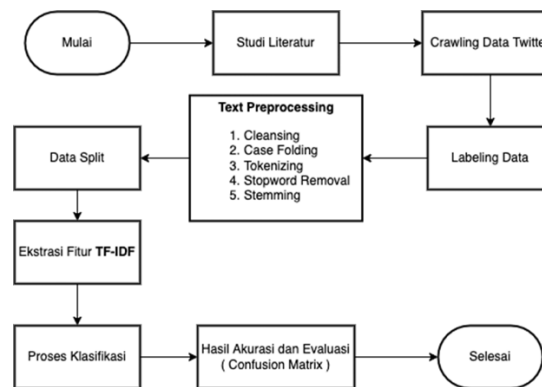
Berdasarkan penelitian sebelumnya dengan judul “Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes”. Hasil penelitian tersebut memiliki pola persepsi negatif pada dataset tweet pemilu 2019 adalah 52%, jauh lebih besar dari pola persepsi positif 18%, dan pandangan netral memiliki nilai 31% lebih tinggi daripada persepsi positif. Akurasi dataset training adalah 81%, sedangkan akurasi dataset testing adalah 76%. Rata-rata nilai presisi sentimen positif sebesar 86,65%, sentimen negatif sebesar 77,15%, dan sentimen netral sebesar 80,95 %, sedangkan nilai rata-rata recall pada sentimen positif sebesar 36,8%, sentimen negatif sebesar 93,2 %, dan sentimen netral sebesar 86,8 %[5].

Kemudian penelitian lainnya dengan judul “Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19”. Dalam penelitian ini, melakukan perbandingan hasil analisis sentimen dari vaksin Sinovac dan vaksin merah-putih. Sumber data yang digunakan dari media sosial Twitter. Data yang digunakan sebanyak 845 tweet, dengan menggunakan dua kata kunci, yaitu “vaksinmerahputih” dan “vaksinsinovac”. Setelah itu, data dipisahkan menjadi 253 set pelatihan dan 592 set tes. Algoritma *Support Vector Machine* dan *Naive Bayes* digunakan untuk klasifikasi. Pendekatan Naive Bayes menghasilkan hasil klasifikasi dengan akurasi rata-rata 85,59%, sedangkan metode SVM menghasilkan hasil dengan akurasi rata-rata 84,41%. Algoritma *Naive Bayes* mengembalikan 66% sentimen positif dan 34% sentimen negatif untuk istilah "vaccinovac", sedangkan "vaccinmerahputih" mengembalikan 89% sentimen positif dan 11% sentimen negatif. Pendekatan SVM mengembalikan 96% sentimen positif dan 4% sentimen negatif untuk istilah "vaccinovac", sedangkan "vaksinmerahputih" mengembalikan 98% sentimen positif dan 2% negatif[6].

Penelitian sebelumnya menunjukkan bahwa *machine learning* dapat digunakan untuk melakukan analisis sentimen di media sosial untuk menentukan bagaimana perasaan orang tentang masalah yang menjadi bahan perdebatan. Berdasarkan hal tersebut dan konteks masalah, maka dilakukan penelitian untuk analisis sentimen pelanggan terhadap layanan ShopeeFood di media sosial Twitter. Adapun tujuan penelitian ini adalah mengetahui implementasi analisis sentimen dengan metode *Naive Bayes* dan *Support Vector Machine* dan mengetahui tingkat performa hasil akurasi pada kedua algoritma tersebut.

2. METODOLOGI

Tahapan penelitian digunakan untuk menunjukkan bagaimana menggambarkan dan menyelesaikan masalah penelitian melalui banyak tahapan. Berikut ini disajikan rancangan atau tahapan - tahapan penelitian serta penjelasan yang disampaikan oleh penulis :



Gambar 1. Flowchart Tahap Penelitian

2.1 Studi Literatur

Pada penelitian ini penulis mencari sumber – sumber yang berkaitan *text mining* dan analisis sentimen dengan menggunakan algoritma *naïve bayes* dan *support vector machine*[7]. Selain itu, penulis juga mempelajari tata cara penulisan pada buku bahasa pemrograman Python [8].

2.2 Pengumpulan Data / Crawling Data & Pelabelan Data

Penulis melakukan pengambilan sumber data dari kumpulan *tweet* menggunakan Twitter API dengan bantuan *library* Tweepy. Pencarian kumpulan data *tweet* menggunakan kata kunci '("Shopee Food" OR "ShopeeFood") lang:id -has:links', artinya akan mencocokkan frasa persis di isi dalam *tweet* dan data *tweet* hanya Bahasa Indonesia serta tidak mengandung url / link . Serta jumlah data yang diambil berkisar 2000 data *tweet* dari tanggal 02-06-2022 s/d 17-06-2022, kemudian data disimpan dalam format csv.

Pada tahap selanjutnya, penulis melakukan pelabelan data dengan bantuan rekan tanpa melibatkan pakar. Pelabelan data menggunakan Google Spreadsheet dan penentuan label berdasarkan opini rekan yang kalimat tersebut mendekati negatif dan positif. Hasil data yang sudah dilabeli sebanyak 1498 negatif dan 504 positif.

2.3 Text Preprocessing

Setelah proses pengambilan data selesai, kemudian pada dataset dilakukan tahapan *text preprocessing* yang merupakan langkah awal dalam pembersihan data sesuai dengan kebutuhannya agar *text mining* dapat mengolahnya lebih lanjut. Algoritma *text mining* sering beroperasi dengan cara yang sama seperti teknik *data mining*. *Text mining* dapat digunakan untuk mengatasi berbagai masalah, termasuk *analysis*, *classification*, *clustering*, atau *prediction* dan *information retrieval*[9].

1. *Cleansing*, teknik menghilangkan kata-kata yang tidak perlu dari sebuah teks untuk mengurangi *noise* seperti Kata, tanda baca, link url, huruf, serta emoji.
2. *Case folding*, Semua karakter huruf atau teks diubah dari huruf kapital menjadi huruf kecil.
3. *Tokenizing*, proses membagi urutan karakter menjadi komponen yang lebih kecil (kata/frasa) yang dikenal sebagai token.
4. *Stopword removal*, membuang kata-kata yang tidak memiliki arti atau tidak relevan, yaitu kata apa yang digunakan untuk merepresentasikan dokumen.
5. *Stemming*, langkah-langkah untuk mengurangi jumlah indeks yang berbeda dalam sumber data sehingga istilah *suffix* atau *prefix* kembali ke bentuk dasarnya.

2.4 Data Split

Kumpulan data untuk penelitian ini akan dibagi menjadi data latih dan data uji. Pada penelitian ini diterapkan teknik *train test split* menggunakan modul *scikit learn library* dengan bahasa pemrograman Python. Jumlah data yang digunakan dalam teknik sebesar 80% data latih dan 20% data uji.

2.5 Ekstraksi Fitur TF-IDF

Setelah melakukan tahapan *text preprocessing* dan *splitting data*, kemudian penulis melakukan tahap selanjutnya yaitu menggunakan *Term Frequency - Inverse Document Frequency* (TF-IDF). TF-IDF merupakan proses yang memberikan bobot pada setiap kata untuk menentukan seberapa

dekat istilah terhubung ke dokumen[10]. Tujuan dari TF adalah untuk menampilkan frekuensi suatu istilah dalam *tweet*. IDF berusaha untuk menentukan seberapa sering sebuah kata muncul di semua *tweet* [11].

2.6 Naïve Bayes

Strategi ini menggunakan teorema Bayes, yang didirikan Thomas Bayes pada abad ke-18. *Naïve Bayes Classification* (NB) adalah klasifikasi statistik yang dapat digunakan untuk meramalkan kemungkinan keanggotaan kelas. Menurut Wu dan Kumar, algoritma *Naïve Bayes* adalah teknik klasifikasi yang terkenal dan salah satu dari 10 algoritma teratas untuk *data mining*. *Naïve Bayes* menggunakan teori probabilitas, cabang matematika, untuk menentukan kemungkinan klasifikasi yang mungkin dengan menganalisis frekuensi setiap klasifikasi dalam data pelatihan[12].

$$P(H|X) = \frac{P(X|H)}{P(X)} \cdot P(H) \quad (1)$$

Dimana:

x = Data dengan class yang belum diketahui

H= Hipotesis data merupakan suatu class spesifik

P(H|X) = Probabilitas hipotesis H berdasar kondisi X (*posteriori probabilitas*)

P(H) = Probabilitas hipotesis H (*prior probabilitas*)

P(X|H) = Probabilitas X berdasarkan kondisi pada hipotesis H

P(X) = Probabilitas X

2.7 Support Vector Machine (SVM)

Teknik *Support Vector Machine* (SVM) digunakan untuk mengenali pola dalam data pelatihan berlabel dan kemudian membuat model pembelajaran mesin selama proses klasifikasi. Model yang dihasilkan kemudian diterapkan pada data *tweet* tambahan yang tidak berlabel untuk menghasilkan perkiraan kelas sentimen, apakah negatif, atau positif[13].

2.8 Evaluasi

Evaluasi dilakukan menggunakan instrumen bernama *confusion matrix* untuk mengukur nilai *accuracy*, *precision*, *recall*, dan *f-score* dengan tujuan menguji performa model. Rumus untuk menghitung *confusion matrix* adalah Persamaan (2) sampai Persamaan (4) Parameter TP (*True Positive*) menunjukkan jumlah prediksi positif dari kelas positif nyata. FP (*False Positive*) menunjukkan jumlah prediksi positif dari kelas yang sebenarnya negatif. TN (*True Negative*) menunjukkan jumlah prediksi negatif yang dihasilkan oleh kelas negatif nyata. FN (*False Negative*) menunjukkan jumlah prediksi negatif dari kelas positif benar[14].

$$Accuracy (A) = \frac{(TP + TN)}{(TP + FP + FN + TN)} \quad (2)$$

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

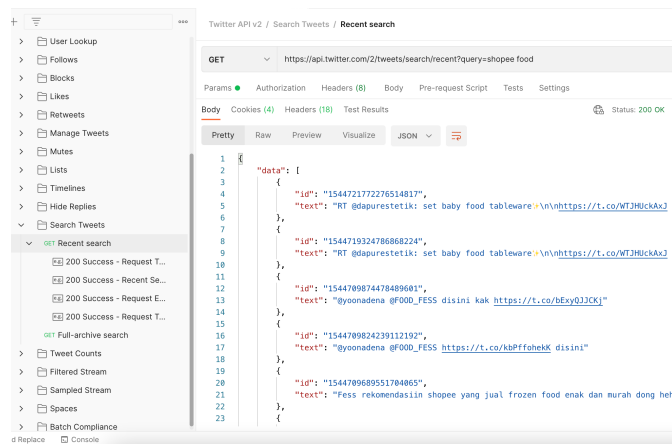
$$Precision (P) = \frac{(TP)}{(TP + FN)} \quad (4)$$

3. HASIL DAN PEMBAHASAN

Berikut ini adalah hasil – hasil penelitian serta pembahasan mengenai tahapan dalam melakukan analisis sentiment dari mulai pengumpulan data, *labeling data*, *text preprocessing*, dan penerapan algoritma.

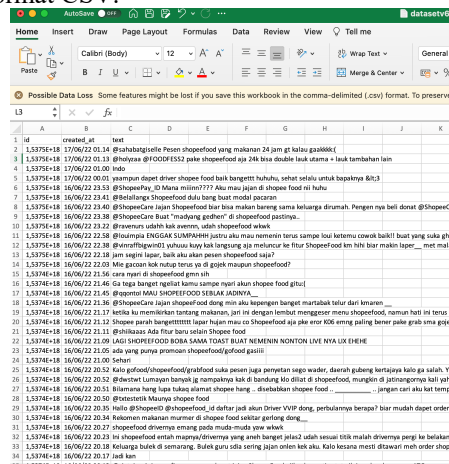
3.1 Pengumpulan Data & Labeling Data

Tahap awal yang dilakukan oleh penulis pada penelitian ini adalah melakukan *testing* pada Twitter API V2 dalam bentuk *collection*. *Collection* dapat digunakan dengan perangkat lunak Postman. Setelah itu, *collection* dapat di *import* dan diuji coba. Berikut contoh *testing* pada “Search Tweets” pada Postman.



Gambar 2. Result Twitter API

Pada Gambar 2 terdapat hasil berupa format json yang isi datanya berupa id dan *text*. Tahap selanjutnya adalah melakukan pengambilan data dengan bantuan *library* Tweepy dan Python. Data yang diambil sejumlah 2000 data dari tanggal 02-06-2022 s/d 17-06-2022 berupa id, *created_at* dan *text* yang disimpan dalam format CSV.



Gambar 3. Hasil Crawling Data

Dapat dilihat Gambar 3 data pada kolom *text* masih belum bersih, contohnya masih terdapat *special character*. Setelah itu, pada tahap berikutnya penulis menentukan label secara manual pada setiap *tweet*, yang berjumlah 2000 data dan kelas label terdiri negatif, dan positif seperti pada Gambar 4.

Id	Created_at	Text	Label
15369620424	15/06/22 13:4	SHOPEE FOOD KENAPA SIH CC	negatif
15369612575	15/06/22 13:3	mitra shopee food menghindari	negatif
15369610917	15/06/22 13:3	shopee food kok lemot sih	negatif
15369590647	15/06/22 13:2	@worksfess Apa ini salah satu	negatif
15369580886	15/06/22 13:2	@inYugi sepemahamanku tuh n	positif
15369536582	15/06/22 13:0	@dailydinbo Trauma sama pem	negatif
15369474368	15/06/22 12:4	dari kurir menjadi driver shopee lebih santai, tapi tetap masih be	positif
15369432599	15/06/22 12:2	Mesan di Shopeefood, yg nganti	negatif
15369400596	15/06/22 12:1	Hello marrybrown bangi kenap	negatif
15369383032	15/06/22 12:0	Kalo bener ada album fisiknya €	negatif
15369373183	15/06/22 12:0	@Yolo9918 @ayun_khurrotul @c	negatif
15369357774	15/06/22 11:5	-1fess gess Shopeefood tuh fil	negatif
15369341096	15/06/22 11:5	@susipudijastuti Kayaknya yg d	negatif

Gambar 4. Hasil Labeling Data Manual

3.2 Text Preprocessing

Dalam tahap selanjutnya, penulis mencoba *import* data dalam bentuk csv yang diubah menjadi *dataframe* dengan tujuan untuk dilakukan pembersihan data.

```
data = pd.read_csv("ds.csv")
# overview dataset
data.head(20)

Out[35]:
```

	id	created_at	text
0	1537488780144010000	17/06/22 01:14	@sahabatgiselle Pesen shopeefood yang makanan ...
1	1537488657913400000	17/06/22 01:13	@holyyaa @FOODFESS2 pake shopeefood aja 24k bi...
2	1537488458813480000	17/06/22 01:00	Indo\n\nPesawat siapin budget : 1 - 1.5jt pp\n...
3	1537488425862400000	17/06/22 00:01	yaampun dapet driver shopee food baik bangettt...
4	1537476595473900000	16/06/22 23:53	@ShopeePay_ID Mana miinn???? Aku mau jajan di...
5	1537475626949890000	16/06/22 23:41	@Belalanga Shopeefood dlu bang buat modif pa...
6	1537475139630760000	16/06/22 23:40	@ShopeeCare Jajan Shopeefood biar bisa makan b...
7	1537474607741900000	16/06/22 23:38	@ShopeeCare Buat "madyang gadhan" di Shopeefoo...
8	1537470638844590000	16/06/22 23:32	@wenerus udahh kak evenm, udah Shopeefood sekek
9	1537464743044320000	16/06/22 22:58	@tumpia ENGGAH SUMPAN#H jastu aku mau nemen...
10	1537459822461990000	16/06/22 22:38	@vinnBiginD1 juhau kuyj kak langsung aja...
11	1537454532010390000	16/06/22 22:18	jam ngini laeet, baik aku akan pesen Shopeefo...
12	1537450846964220000	16/06/22 22:03	Mie gacoo kok mup tena ya di ggek maupn...
13	1537449162718370000	16/06/22 21:56	cara nyari di Shopeefood gmn sh
14	1537446324547540000	16/06/22 21:46	Ga tega banget ngeliat kamu sampai nyari akun s...
15	1537446344549250000	16/06/22 21:45	@qantari MAU SHOPEEFOOD SEBLAK JADINYA...
16	1537443984980390000	16/06/22 21:36	@ShopeeCare jajan Shopeefood dong min aku kepe...
17	1537439179512290000	16/06/22 21:17	katika lu memblinon tartang makanan, jeri ml...
18	1537437890715580000	16/06/22 21:12	Shopee parnh bangetttttt laper huan mtu oo...

Gambar 5. Bentuk *Dataframe*

Pada Gambar 5 terlihat bentuk dari *dataframe*, setelah itu, tahapan yang dilakukan oleh penulis adalah melakukan proses *text preprocessing* yang memiliki tahapan berikut.

1. *Cleansing*, merupakan tahapan untuk menghapus *emoticon*, url / links, *mentions*, *hashtags*, tanda baca, dan angka pada data [15]. Berikut hasil *cleansing* pada Gambar 6.

	text	cleansing
0	@sahabatgiselle Pesen shopeefood yang makanan ...	pesen yang makanan jam gt kalau gaakkkk
1	@holyyaa @FOODFESS2 pake shopeefood aja 24k bi...	pake aja k bisa double lauk utama lauk tambaha...
2	Indo\n\nPesawat siapin budget : 1 - 1.5jt pp\n...	indo pesawat siapin budget jt pp kereta k pp t...
3	yaampun dapet driver shopee food baik bangettt...	yaampun dapet driver baik bangettt huhuhu seha...
4	@ShopeePay_ID Mana miinn???? Aku mau jajan di...	mana miinn aku mau jajan di nii huhu

Gambar 6. Hasil *Cleansing*

2. *Case folding*, tahapan ini peneliti melakukan proses mengecilkan semua huruf dan menghapus kalimat terkait shopee, shopeefood, shopeefood_id, dan food. Berikut Gambar 7 hasil dari *casefolding*.

Out[29]:

	text	casefolding
1796	lewat shopee food, itupun aku yang membayarnya...	lewat itupun aku yang membayarnya sembari menu...
1797	@rlgxxg gk muluk muluk pengen sushi di shopeefo...	gk muluk muluk pengen sushi di ada promo disko...
1798	Shopeefood kenapa yaa????	kenapa yaa
1799	gw bener bener baru buka shopeefood mau jajan ...	gw bener bener baru buka mau jajan trus tbtb m...
1800	@jpb_menfess biasanya aku kirimn buatan aku s...	biasanya aku kirimn buatan aku sendiri teh ja...

Gambar 7. Hasil *Casefolding*

3. *Tokenizing*, tahapan ini peneliti melakukan pemecahan kalimat menjadi potongan kata menggunakan fungsi bawaan Python yang disebut *split()*. Berikut Gambar 8 hasil dari *tokenizing*.

Out[31]:

	text	tokenizing
0	@sahabatgiselle Pesen shopeefood yang makanan ...	[pesen, yang, makanan, jam, gt, kalau, gaakkkk]
1	@holyyaa @FOODFESS2 pake shopeefood aja 24k bi...	[pake, aja, k, bisa, double, lauk, utama, lauk...
2	Indo\n\nPesawat siapin budget : 1 - 1.5jt pp\n...	[indo, pesawat, siapin, budget, jt, pp, kereta...
3	yaampun dapet driver shopee food baik bangettt...	[yaampun, dapet, driver, baik, bangettt, huhuh...
4	@ShopeePay_ID Mana miinn???? Aku mau jajan di...	[mana, miinn, aku, mau, jajan, di, nii, huhu]

Gambar 8. Hasil *Tokenizing*

4. *Stopword Removal*, *stopwords* adalah pemilihan kata kunci dari hasil token dan membuang kata-kata yang tidak memiliki arti atau tidak relevan, yaitu kata yang digunakan untuk merepresentasikan dokumen, yang disebut sebagai proses *filtering*. Contoh dari *stopword*

Tabel 1. Perhitungan TF-IDF

No	Term	TF			df	idf $\log\frac{n}{df}$	W = df * idf		
		D1	D2	D3			D1	D2	D3
1	shopeefood	1	1	1	3	0	0	0	0
2	emang	1	0	0	1	0,477	0,477	0	0
3	pelit	1	0	0	1	0,477	0,477	0	0
4	barget	1	0	0	1	0,477	0,477	0	0
5	semenjak	0	1	0	1	0,477	0	0,477	0
6	lau	0	1	0	1	0,477	0	0,477	0
7	murah	0	1	1	2	0,176	0	0,176	0,176
8	order	0	1	0	1	0,477	0	0,477	0
9	situ	0	1	0	1	0,477	0	0,477	0
10	beli	0	0	1	1	0,477	0	0	0,477
11	bjaa	0	0	1	1	0,477	0	0	0,477

Dari Tabel 1 di atas dapat dilihat simulasi dari operasi perhitungan TF-IDF dari dokumen/teks masukan. Ekstraksi ini bertujuan untuk mengubah dokumen teks dalam bentuk numerik agar mudah dijadikan sebagai vektor inputan saat pelatihan data dan pengklasifikasian data menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine*.

3.4 Klasifikasi

Algoritma dalam penelitian ini menggunakan dua algoritma yaitu, *Naïve Bayes* dan *Support Vector Machine* (SVM). Dataset yang digunakan sebanyak 1236 data yang telah dibersihkan, terdiri dari 734 negatif dan 502 positif dan pengujian algoritma dilakukan sebanyak 2 kelas menggunakan data latih 80% dan data uji 20%. Model yang digunakan *Naïve Bayes* adalah *Multinomial Naive Bayes*, sedangkan SVM model yang digunakan adalah C-Support Vector Classification atau SVC dengan kernel rbf.

Berikut performa dari pengujian algoritma *Naïve Bayes* pada dan Gambar 12 dapat dideskripsikan bahwa klasifikasi algoritma *Naïve Bayes* memiliki tingkat akurasi 90.43%, presisi 99.74%, *recall* 76.65%, dan *f1-score* 86.68%.

```

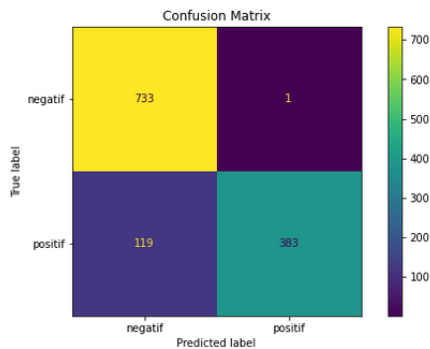
Training time: 0.011940s; Prediction time: 0.008975s
precision    recall  f1-score   support

   0         0.86    1.00    0.93     732
   1         1.00    0.77    0.87     501

 accuracy          0.90     1233
 macro avg         0.93    0.88    0.90     1233
 weighted avg      0.92    0.90    0.90     1233

 accuracy score = 90.43%
 Precision = 99.74%
 Recall: 76.65%
 F1-Score: 86.68%
 731 1 117 384
    
```

Gambar 12. Akurasi MultinomialNB



Gambar 13. Confusion Matrix Naïve Bayes

Pada Gambar 13 menunjukkan bahwa *confusion matrix* dengan ukuran 2 x 2 secara benar sebanyak 733 data sebagai negatif dan 383 sebagai data positif. Selain itu model salah dalam memprediksi 1 data ke dalam kelas positif yang seharusnya negatif (*False Negatif*), serta model menyatakan 119 data yang salah dalam kelas negatif yang seharusnya positif (*False Positif*).

Selain itu, performa pada algoritma *Support Vector Machine* (SVM) pada Gambar 14 dapat dideskripsikan bahwa algoritma *Support Vector Machine* (SVM) memiliki tingkat akurasi 99.27%, presisi 99.60%, *recall* 98.60%, dan *f1-score* 99.10%.

```

Training time: 2.224051s; Prediction time: 3.003963s
precision    recall  f1-score   support

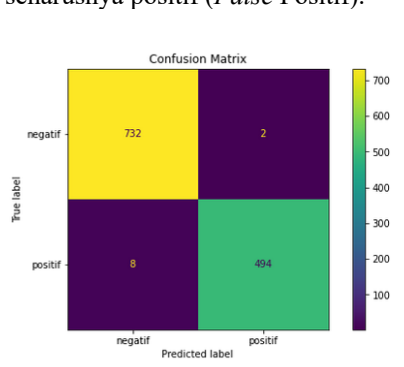
   0         0.99    1.00    0.99     732
   1         1.00    0.99    0.99     501

 accuracy          0.99    1233
 macro avg         0.99    0.99    0.99    1233
weighted avg         0.99    0.99    0.99    1233

accuracy score = 99.27%
Precision = 99.60%
Recall: 98.60%
F1-Score: 99.10%
730 2 7 494
    
```

Gambar 14. Akurasi Support Vector Classification

Berdasarkan hasil evaluasi model pada Gambar 15 menyatakan bahwa secara benar sebanyak 731 data sebagai negatif dan 494 sebagai data positif. Selain itu model salah dalam memprediksi 2 data ke dalam kelas positif yang seharusnya negatif (*False Negatif*), serta model menyatakan 8 data yang salah dalam kelas negatif yang seharusnya positif (*False Positif*).



Gambar 15. Confusion Matrix Support Vector Machine

Berdasarkan perbandingan yang telah dilakukan bahwa algoritma *Support Vector Machine* (SVM) memiliki akurasi yang lebih baik sebesar 99.27% dibandingkan algoritma *Naïve Bayes* sebesar 90.43%. Sehingga akurasi pada algoritma *Support Vector Machine* (SVM) dapat dimanfaatkan sebagai model yang dapat digunakan sebagai prediksi sebuah kalimat.

4. PENUTUP

Berdasarkan hasil dan pembahasan yang telah dijelaskan, maka dapat disimpulkan bahwa algoritma *naïve bayes* memiliki performa *accuracy* 90.27%, *precision* 99.74%, *recall* 76.20%, dan *f1-score* 86.39%. Proses klasifikasi dengan waktu *training* 0.014990s dan waktu *prediction* 0.010943s. Sedangkan algoritma *support vector machine* memiliki performa *accuracy* 99.27%, *precision* 99.60%, *recall* 98.60%, dan *f1-score* 99.10%. Proses klasifikasi dengan waktu *training* 2.252885s dan waktu *prediction* 2.972732s.

Sehingga dapat dibuktikan bahwa hasil algoritma *support vector machine* lebih akurat dibandingkan *naïve bayes* dan model pada algoritma *support vector machine* dapat digunakan sebagai proses analisis sentimen data Twitter.

DAFTAR PUSTAKA

- [1] E. S. Sulistiyawati and A. Widayani, “Marketplace Shopee Sebagai Media Promosi Penjualan UMKM di Kota Blitar,” *J. Pemasar. Kompetitif*, vol. 4, no. 1, p. 133, 2020, doi: 10.32493/jpkpk.v4i1.7087.
- [2] Supangat, “Peluang Bisnis Makanan Dan Minuman Di Shopee Food Bagi UMKM,” 2021, [Online]. Available: http://repository.untag-sby.ac.id/12484/1/1211900355_Anisa%27ulKhomariyah_E-Bisnis_ETS.pdf.
- [3] G. A. Buntoro, “Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter,” *INTEGER J. Inf. Technol.*, vol. 1, no. 1, pp. 32–41, 2017, [Online]. Available: https://www.researchgate.net/profile/Ghulam_Buntoro/publication/316617194_Analisis_Sentimen_Calon_Gubernur_DKI_Jakarta_2017_Di_Twitter/links/5907eee44585152d2e9ff992/Analisis-Sentimen-Calon-Gubernur-DKI-Jakarta-2017-Di-Twitter.pdf.
- [4] S. F. Pratama, R. Andrian, and A. Nugroho, “Analisis Sentimen Twitter Debat Calon

- Presiden Indonesia Menggunakan Metode Fined-Grained Sentiment Analysis,” *JOINTECS (Journal Inf. Technol. Comput. Sci.*, vol. 4, no. 2, p. 39, 2019, doi: 10.31328/jointecs.v4i2.1004.
- [5] S. Juanita, “Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes,” *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 552, 2020, doi: 10.30865/mib.v4i3.2140.
- [6] 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, May 2021, doi: 10.22146/jnteti.v10i2.1421.
- [7] S. Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti *et al.*, “Analisis Sentimen Pemindahan Ibu Kota Negara dengan Feature Selection Algoritma Naive Bayes dan Support Vector Machine,” *Masa Berlaku Mulai*, vol. 1, no. 3, pp. 504–512, 2017.
- [8] A. Kadir, *Logika Pemrograman Python*. PT Elex Media Komputindo, 2019.
- [9] E. E. Pratama and R. L. Atmi, “A Text Mining Implementation Based on Twitter Data to,” *J. Comput. Soc.*, vol. 1, no. 1, pp. 91–100, 2020.
- [10] R. Asmara, M. F. Ardiansyah, and M. Anshori, “Analisa Sentiment Masyarakat terhadap Pemilu 2019 berdasarkan Opini di Twitter menggunakan Metode Naive Bayes Classifier,” *INOVTEK Polbeng - Seri Inform.*, vol. 5, no. 2, p. 193, 2020, doi: 10.35314/isi.v5i2.1095.
- [11] 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, Jul. 2020, doi: 10.30865/mib.v4i3.2181.
- [12] A. Mustopa, Hermanto, Anna, E. B. Pratama, A. Hendini, and D. Risdiansyah, “Analysis of User Reviews for the PeduliLindungi Application on Google Play Using the Support Vector Machine and Naive Bayes Algorithm Based on Particle Swarm Optimization,” in *2020 Fifth International Conference on Informatics and Computing (ICIC)*, Nov. 2020, vol. 2, pp. 1–7, doi: 10.1109/ICIC50835.2020.9288655.
- [13] H. C. Husada and A. S. Paramita, “Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM),” *Teknika*, vol. 10, no. 1, pp. 18–26, 2021, doi: 10.34148/teknika.v10i1.311.
- [14] F. D. Ananda and Y. Pristyanto, “Analisis Sentimen Pengguna Twitter Terhadap Layanan Internet Provider Menggunakan Algoritma Support Vector Machine,” *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 20, no. 2, pp. 407–416, May 2021, doi: 10.30812/matrik.v20i2.1130.
- [15] D. Darwis, N. Siskawati, and Z. Abidin, “Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional,” *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.

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 disajikan: literature review, dataset yang digunakan, source code, dan hasil eksperimen secara keseluruhan. Pada bagian literature review berisi artikel – artikel jurnal yang memuat tentang analisis sentimen menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM) sebagai bahan acuan untuk menyusun kerangka dalam kegiatan penelitian. Bagian analisis dan perancangan, penulis menjelaskan tahapan – tahapan dalam melakukan pengambilan data dari media social twitter dan memberikan sampel dalam *text preprocessing* dan perhitungan manual TF-IDF. Setelah itu bagian *source code* merupakan ringkasan – ringkasan tahapan analisis sentiment dengan bantuan Bahasa pemrograman Python serta menggunakan *library* pendukung. Dataset merupakan isi kumpulan data opini netizen tentang Shopee Food untuk dilakukan pemberian label manual dan diolah menjadi dataframe. Kemudian bagian tahapan eksperimen merupakan proses pembagian data latih dan uji, setelah itu dilakukan pembobotan dan pembuatan model. Tahapan hasil semua eksperimen membandingkan akurasi pada masing – masing algoritma dan uji coba klasifikasi dengan streamlit.