****

**PROPOSAL TUGAS AKHIR**

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

**NAMA : Kindi Setiadi**

**NIM : 41519120004**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS MERCU BUANA**

**2021**

# **LEMBAR PERSETUJUAN**

Yang bertanda tangan di bawah ini menyatakan bahwa Proposal Tugas Akhir dari mahasiswa :

Nama Mahasiswa : Kindi Setiadi

NIM : 41519120004

Fakultas : Ilmu Komputer

Program Studi : Teknik Informatika

Judul : Analisis Sentimen Pelanggan Terhadap

Layanan ShopeeFood Pada Media Sosial

Twitter Menggunakan Algoritma Naïve Bayes

dan *Support Vector Machine* (SVM)

Jakarta, 13 November 2021

Disetujui dan diterima untuk di seminarkan,

|  |  |
| --- | --- |
| Dosen Pembimbing    (Afiyati, S.Si., MT)  NIP 114690444 | Dosen Pengampu  Matakuliah Metodologi Penelitian  (Emil Robert Kaburuan, S.T., M.A., Ph.D)  NIP 1248000691 |

DAFTAR PUSTAKA

[LEMBAR PERSETUJUAN 2](#_Toc515006308)

[I. TOPIK 3](#_Toc515006309)

[II. LATAR BELAKANG 4](#_Toc515006310)

[III. RUMUSAN MASALAH 7](#_Toc515006311)

[IV. TUJUAN DAN MANFAAT 7](#_Toc515006312)

[A. Tujuan 7](#_Toc515006313)

[B. Manfaat 7](#_Toc515006314)

[V. BATASAN MASALAH 7](#_Toc515006315)

[VI. LANDASAN TEORI 8](#_Toc515006316)

[A. Penelitian Terkait 8](#_Toc515006317)

[B. Teori Pendukung 10](#_Toc515006318)

[VII. METODOLOGI 13](#_Toc515006319)

[A. Jenis Penelitian 13](#_Toc515006320)

[B. Metode Pengumpulan Data 13](#_Toc515006321)

[C. Tahap Penelitian 14](#_Toc515006322)

[VIII. DAFTAR PUSTAKA 16](#_Toc515006323)

# Topik

Topik yang diangkat pada penelitian Tugas Akhir ini adalah analisis sentimen pelanggan terhadap layanan ShopeeFood pada media sosial Twitter menggunakan algoritma naïve bayes. Adanya sentimen opini dari pelanggan mengenai layanan ShopeeFood dapat dianalisis dan dimanfaatkan untuk mendapatkan informasi yang berguna bagi pelanggan. Analisis sentimen pada penelitian ini dapat menggunakan metode *Naïve Bayes Classife*r (NBC) dan *Support Vector Machine* (SVM) yang digunakan untuk melakukan proses klasifikasi data. Algoritma pada penelitian ini diharapkan akan digunakan untuk melakukan klasifikasi sentimen berdasarkan data *Tweet* yang dapat membantu menentukan opini pelanggan yang mengandung sentimen positif, dan negatif.

# latar belakang

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 mengkonsumsi 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.

*E-commerce* mengacu pada penjualan yang dilakukan melalui media elektronik, yaitu semacam mekanisme bisnis elektronik yang berfokus pada transaksi komersial berbasis individu yang memanfaatkan internet sebagai sarana untuk memperdagangkan produk dan jasa. *E-commerce* memberikan pengembangan pendapatan dan prospek perusahaan yang sangat besar (seperti menjual barang dan memberikan layanan online) [1]. Salah satu aplikasi yang dapat membantu memenuhi kebutuhan manusia, yaitu aplikasi Shopee.

Shopee adalah salah satu pasar pertama yang diluncurkan oleh Garena, yang kini telah berganti nama menjadi SEA Group, secara langsung di bawah pengawasan PT Shopee International Indonesia, Shopee resmi diluncurkan di Indonesia pada Desember 2015. Barang-barang kecantikan, pakaian pria, pakaian wanita, handphone & aksesoris, komputer & aksesoris, peralatan rumah tangga, elektronik, makanan & minuman, dan sebagainya semuanya tersedia di *marketplace* 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 [2].

Pada April 2020, Shopee merilis ShopeeFood sebagai layanan *food delivery services*. ShopeeFood, memliki 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 [3]. Konsumen menggunakan aplikasi ini untuk memesan makanan yang aman dan higienis. Di satu sisi, para pelaku usaha tetap menjalankan usahanya dan membantu ojek online untuk menghasilkan uang [4]. Promosi merupakan hal yang penting dalam dunia bisnis karena merupakan kunci keberhasilan suatu usaha dalam memasarkan produknya. Dengan kemajuan teknologi yang cepat, media sosial menjadi bagian penting untuk memasarkan agar dapat dikenal [5].

Media sosial merupakan salah satu wadah atau media untuk berekspresi dan berpendapat atas berbagai tema di era digital saat ini. Media sosial menjadikan seperangkat teknologi komunikasi dan kolaborasi baru yang memungkinkan orang untuk terlibat dalam berbagai bentuk interaksi yang sebelumnya tidak tersedia untuk individu biasa. Salah satu media sosial yang popular saat ini adalah Twitter [6].

Twitter adalah cara yang hebat untuk bertukar ide, mendapatkan informasi, memicu pemikiran, dan dapat mengikuti apa saja yang sedang dilakukan oleh teman. 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 [7]. Pentingnya peran media sosial seperti Twitter dalam memposting *tweet*, dapat mengetahui sentimen opini dari pelanggan dengan cara analisis sentimen.

Analisis sentimen adalah metode menganalisis orang-orang yang membagikan ide-ide mereka di media sosial tentang berbagai masalah. 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 [8].

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 % [9].

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 *Naïve Bayes* menghasilkan hasil klasifikasi dengan akurasi rata-rata 85,59%, sedangkan metode SVM menghasilkan hasil dengan akurasi rata-rata 84,41%. Algoritme ﻿*Naïve 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 [10].

Dalam penelitian saat ini, penulis akan mengajukan penelitian dengan judul “Analisis Sentimen Pelanggan Terhadap Layanan ShopeeFood Pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes dan *Support Vector Machine* (SVM)”. Dalam prosesnya data yang diperoleh perlu menggunakan Twitter API, sehingga dapat menjadi bahan untuk analisis sentimen.

# rumusan masalah

Rumusan masalah yang akan dibahas pada pengerjaan Tugas Akhir ini adalah sebagai berikut :

1. Bagimana implementasi metode *Naïve Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM) pada data *tweet* pelangganterhadap layanan ShopeeFood untuk analisis sentimen ?
2. Bagaimana tingkat akurasi antara metode *Naïve Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM) ?

# tujuan dan manfaat

## Tujuan

## Adapun tujuan yang hendak dicapai dalam penyusunan Tugas Akhir adalah sebagai berikut :

1. Mengetahui analisis sentimen dengan metode *Naïve Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM) layak digunakan klasifikasi *tweet* pada media sosial Twitter.
2. Mengetahui hasil akurasi pada jenis sentimen dengan metode *Naïve Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM).

## Manfaat

Adapun manfaat dari penelitian ini adalah sebagai berikut :

1. Membantu proses klasifikasi data *tweet* dengan metode *Naïve Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM).
2. Bagi penulis dapat menambah sumber referensi bagi ilmu pengetahuan dan mengetahui informasi sentimen terhadap layanan ShopeeFood.

# batasan masalah

Adapun batasan masalah dari penelitian ini adalah sebagai berikut :

1. Penelitian ini implementasinya menggunakan algoritma naïve bayes dengan bantuan bahasa pemrograman python.
2. Data *tweet* yang diperoleh dari Twitter API dan data yang dianalisis menggunakan Bahasa Indonesia.
3. Klasifikasi sentimen *tweet* yang digunakan hanya 2 kategori, yaitu positif, dan negatif.
4. Data yang diolah hanya berupa teks.
5. Data *tweet* diambil pada bulan Januari 2021 – Desember 2021.
6. Kata kunci yang digunakan untuk pengambilan data *tweet* menggunakan ShopeeFood, shopeefood, dan hashtag #shopeefood.
7. Kamus data yang digunakan adalah kamus sentimen Sastrawi.
8. Pemanfaatan *user account* bersifat unik, *tweet* yang berulang akan dihapus.

# landasan teori

Adapun landasan teori yang menjadi dasar dari penelitian Tugas Akhir ini adalah sebagai berikut:

## Penelitian Terkait

Pada Tugas Akhir ini, penulis telah melakukan explorasi terkait dengan jurnal penelitian yang telah dilakukan sebelumnya. Berikut adalah beberapa teori penelitian terkait yang digunakan dan mendukungnya jalannya dengan Tugas Akhir ini :

Tabel 1 Penelitian Terkait

|  |  |  |  |
| --- | --- | --- | --- |
| **Topik** | **Metode** | **Hasil** | **Referensi** |
| Analisis sentimen toko online JD.ID | Metode yang digunakan Naïve Bayes Classifier (NBC) pembobotan *tf-idf* | Klasifikasi menggunakan *Naïve Bayes Classifier* (NBC) ﻿tanpa pembobotan tf-idf memiliki nilai akurasi 96,44% sedangkan dengan pembobotan tf-idf memiliki nilai | [11] |
|  |  | akurasi 98% |  |
| Analisis sentimen ulasan pembelian produk menggunakan pendekatan *Natural Language Processing* | Menggunakan algoritma KNN dan TF-IDF dengan ﻿pendekatan NLP | Teknik NLP menghasilkan akurasi 76,92%, presisi 80,00%, dan *recall* 74,07%, ﻿di bandingkan tidak menggunakan NLP hanya menghasilkan nilai akurasi sebesar 69,23%, presisi 80,00%, dan *recall* 64,52%. | [12] |
| Analisis sentimen *tweet* vaksin COVID-19 menggunakan *Recurrent Neural Network* dan *Naïve Bayes* | ﻿*Recurrent Neural Network* dan *Naïve Bayes* | Dataset untuk penelitian ini dibangun dari 5.000 tweet vaksinasi COVID-19, dengan 3800 tweet sentimen positif, 800 tweet sentimen negatif, dan 400 tweet sentimen netral. ﻿Hasil pengujian menunjukkan RNN (TF-IDF) memiliki akurasi lebih besar yaitu 97,77% dibandingkan Naïve Bayes (TF-IDF) sebesar 80% | [13] |
| Analisis sentimen tingkat kepuasan pengguna penyedia layanan telekomunikasi seluler Indonesia pada Twitter | Metode yang digunakan *Support Vector Machine* dan *Lexicon Based Features* | Metode *Support Vector Machine* dan *Lexicon Based Features* menghasilkan nilai accuracy sebesar 79%, precision sebesar 65%, *recall* sebesar 97%, dan f-measure sebesar 78% | [14] |
| Analisis sentimen *review* Shopee berbahasa Indonesia | *Improved K-Nearest Neighbor* dan *Jaro Winkler Distance Liana* | Metode *Jaro Winkler Distance* terbukti dapat meningkatkan akurasi sistem. Berdasarkan hasil pengujian, nilai k optimal ditentukan pada k=10 dengan rata-rata akurasi yang lebih baik dalam keadaan setelah restorasi kata, yaitu akurasi 0,876, presisi 0,810, *recall* 0,942, dan *f-measure* 0,882. | [15] |
| ﻿Analisis sentimen pada maskapai penerbangan di platform  Twitter menggunakan algoritma *Support Vector Machine* (SVM) | Metode pembobotan kata menggunakan metode TF-IDF dan memanfaatkan ﻿algoritma multi-class Support Vector Machine (SVM) | ﻿(Data yang digunakan dalam penelitian ini adalah opini dalam bahasa Inggris dari para pengguna Twitter terhadap maskapai penerbangan. Berdasarkan pengujian yang telah dilakukan, hasil klasifikasi terbaik diperoleh menggunakan SVM kernel RBF pada nilai parameter ??(complexity) = 10 dan ??(gamma) = 1, dengan nilai *accuracy* sebesar 84,37% dan 80,41% ketika menggunakan 10-fold cross validation. | [16] |
| ﻿Analisis sentimen Gojek pada media sosial Twitter menggunakan *Support Vector Machine* (SVM) | Metode yang digunakan *Support Vector Machine* (SVM) dan menggunakan fungsi *kernel*. ﻿Kernel yang akan digunakan adalah kernel linear dan kernel Radial Basis Function (RBF) | ﻿Hasil klasifikasi sentimen dari hasil pelabelan data secara manual menggunakan metode SVM pada Gojek menghasilkan tingkat akurasi keseluruhan terbaik sebesar 79,19% dan akurasi kappa terbaik sebesar 16,52%. Nilai akurasi keseluruhan dan kappa tersebut diperoleh dari pemodelan menggunakan kernel RBF dengan nilai Cost=1000 dan γ=0,00026 | [17] |

## Teori Pendukung

1. Python

Pada tahun 1991, Gudio van Rossum merancang bahasa pemrograman Python. Bahasa ini dimaksudkan untuk memudahkan orang lain memahami kode tersebut. Penggunaan spasi untuk membuat indentasi kode sangat penting sebagai contoh. Python bersifat portable dalam kode sumber. Berbagai platform sistem operasi seperti Windows, Mac OS, dan Linux dapat menggunakan kode sumber Python. Selain itu, Python menyediakan manajemen memori yang bertindak seperti metode pengumupulan sampah pada Java [18].

1. *Web Crawler*

*Web crawler* adalah perangkat lunak yang menggunakan pendekatan tertentu untuk mengumpulkan semua informasi di situs web secara otomatis. *Web crawler* akan mengunjungi setiap URL yang telah diberikan kepadanya satu per satu. Setiap halaman URL yang dikunjungi akan diidentifikasi siapakah ada dalam URL tersebut di dalamnya terdapat URL yang belum dikunjungi. Jika URL belum pernah dikunjungi sebelumnya, URL tersebut akan ditambahkan ke *list* URL yang akan dikunjungi. Putaran mendapatkan dan menambahkan daftar URL ini akan terus berlanjut hingga tidak ada lagi URL yang belum dikunjungi. *Crawl frontier* adalah nama yang diberikan untuk aktivitas ini.

Selama proses *crawling*, akan ditemukan situs web yang terhubung erat dengan URL. URL (Uniform Resource Locator) adalah sistem pengalamatan yang terdiri dari karakter tertentu berupa angka, huruf, dan simbol yang mengarah ke halaman web di jaringan internet. URL digunakan untuk mengarahkan orang ke halaman web di jaringan internet. Ketika URL digunakan, itu digunakan untuk mengidentifikasi alamat situs web. Ini juga terdiri dari informasi alamat situs, yang memungkinkan pengguna untuk berpindah dari satu halaman ke halaman lain dengan bantuan *hyperlink*  [19].

1. *Text Mining*

*Text mining* merupakan salah satu pendekatan yang dapat digunakan untuk melakukan klasifikasi. Text mining adalah bagian dari penambangan data yang mencari pola menarik dalam sejumlah besar data tekstual. *Text mining* mirip dengan *data mining*, dengan pengecualian bahwa metode penambangan data dimaksudkan untuk beroperasi pada data terstruktur dalam database, tetapi penambangan teks dapat bekerja pada data tidak terstruktur atau semi-terstruktur seperti dokumen teks lengkap, halaman web kode/skrip, dan lain-lain.

*Text mining* sering dibagi menjadi tiga tahap: *text pre-processing*, *feature selection*, and *text analytics*. *Tokenizing, filtering, stemming, tagging*, dan *analyzing* merupakan tahapan dari text *preprocessin*g secara umum. *Tokenizing* adalah proses pemisahan setiap kata (token) yang dimasukkan dalam dokumen input. *Filtering* adalah teknik pemilihan kata yang dibuat selama proses *tokenizing*. Ini dapat dilakukan dengan menggunakan *stop list* atau algoritma *word list*. Algoritma *stop list* akan menghilangkan kata-kata yang berlebihan seperti kata ganti, kata keterangan, konjungsi, preposisi, dan pakaian. Daripada itu, algoritma *word list* akan menyimpan kata-kata yang signifikan.

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* [20].

1. *Naïve Bayes Classifier* (NBC)

Naive Bayes Classifier adalah teknik klasifikasi Bayesian berdasarkan teorema Bayes. Teorema Bayes adalah pendekatan kategorisasi berdasarkan probabilitas dan metode statistik yang dibuat oleh ilmuwan Inggris Thomas Bayes. Ini memprediksi peluang masa depan berdasarkan pengalaman sebelumnya. Ciri utama dari Nave Bayes Classifier ini adalah asumsi (naif) yang sangat kuat dari setiap kondisi/independensi.

Menurut Olson Delen (2008), Naive Bayes menentukan probabilitas bahwa kelas penilaian benar diberikan vektor informasi objek. Pendekatan ini membuat asumsi bahwa properti objek mandiri. Probabilitas yang digunakan untuk mendapatkan estimasi akhir diturunkan sebagai jumlah dari frekuensi tabel keputusan "master".

Dibandingkan dengan model pengklasifikasi lainnya, Pengklasifikasi Naive Bayes berkinerja baik. Seperti yang ditunjukkan oleh Xhemali, Hinde Stone dalam publikasinya "Naive Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages", "Naive Bayes Classifier memiliki tingkat akurasi yang lebih tinggi daripada model classifier lainnya."

Manfaat dari strategi ini adalah hanya menggunakan sedikit data pelatihan untuk menghasilkan estimasi parameter yang diperlukan untuk klasifikasi. Karena asumsinya itu adalah variabel independen, hanya varians variabel dalam kelas yang diperlukan untuk klasifikasi, bukan keseluruhan matriks kovarians.

1. Jupyter Notebook

Jupyter Notebook adalah proyek spin-off dari proyek IPython, yang dulunya memiliki proyek Notebook IPython itu sendiri. Nama, Jupyter berasal dari bahasa pemrograman inti yang didukung : Julia, Python, dan R. Jupyter dikirimkan dengan kernel IPython, yang memungkinkan untuk menulis program dengan Python. Jupyter Notebook juga merupakan *tools* online website den sumber terbuka yang memungkinkan pembuatan dan berbagi dokumen yang mencakup kode langsung, persamaan, visualisasi, dan prosa naratif. Pembersihan dan transformasi data, simulasi numerik, pemodelan statistik, visualisasi data, dan pembelajaran mesin hanyalah beberapa aplikasi.

1. Tweepy

Tweepy adalah *community tools* dan *library* bahasa pemrograman Python untuk mengakses Twitter API. Versi terbaru dari *library* saat ini adalah 4.3.0. Library ini dapat digunakan dengan cara membuat akun developer di <https://developer.twitter.com/>,kemudian dibutuhkan *consumer key, consumer secret, access token* dan *access secret* untuk melakukan autentikasi agar dapat diakses. Selain itu, pengujian API bisa dilakukan dengan aplikasi Postman.

1. *Support Vector Machine* (SVM)

*Support Vector Machine* (SVM) adalah cara untuk mengklasifikasikan hal-hal yang pertama kali digunakan oleh Vapnik pada tahun 1998, ketika ia menemukan ide tersebut. Sebagian besar waktu, metode ini bekerja dengan menetapkan batas antara dua kelas yang sejauh mungkin dari satu sama lain . *Hyperplanes* (garis pemisah) harus dibuat di ruang input untuk mendapatkan batas maksimum antar kelas. Untuk melakukan ini, ukur margin hyperplane dan temukan titik maksimum dalam ruang. Margin adalah jarak dari hyperplane ke titik terdekat di setiap kelas. Ini adalah jarak dari hyperplane ke titik terdekat di setiap kelas. Salah satu hal yang membantu mencapai titik ini disebut "*support vector"* [21]*.*

1. *Confusion Matrix*

*Confusion matrix* adalah matrix 2x2 yang merepresentasikan hasil klasifikasi biner pada suatu dataset. Terdapat beberapa rumus umum yang dapat digunakan untuk menghitung performa klasifikasi. Hasil dari nilai *accurac*y, *precision* dan *recall* bisa ditampilkan dalam persentase [22].

1. *Accuracy* adalah jumlah proporsi prediksi yang benar. Adapun rumus perhitungan akurasi dapat dilihat dari persamaan di bawah ini.

Schematic

Description automatically generated with low confidence

1. *Precision* adalah proporsi jumlah dokumen teks yang relevan terkendali diantara semua dokumen teks yang terpilih oleh sistem. Rumus precision dapat dilihat pada persamaan di bawah ini.

Text

Description automatically generated with medium confidence

1. *Recall* adalah proporsi jumlah dokumen teks yang relevan terkendali diantara semua dokumen teks relevan yang ada pada koleksi. Rumus *recall* dapat dinyatakan dengan

Text

Description automatically generated

# metodologi

## Jenis Penelitian

Dalam penelitian ini, penulis menggunakan jenis penelitian deskriptif dengan pendekatan kuantitatif, dimana data disajikan dalam bentuk gambar, dan tabel, atau diagram. Setelah data diolah, ditunjukkan dengan hasil perhitungan numerik dan tabel atau diagram.

## Metode Pengumpulan Data

Pada penelitian ini, penulis memperoleh data dan informasi tentang objek penelitian, dengan proses pengumpulan data sebagai berikut :

1. Studi Pustaka

Studi pustaka adalah teknik untuk mengumpulkan data yang memerlukan penilaian dan referensi materi terkait menggunakan referensi tertentu. Pada langkah ini, data dikumpulkan dalam bentuk informasi dan sumber daya yang dapat digunakan untuk melakukan penelitian. Selanjutnya penulis mempelajari referensi penelitian melalui buku, jurnal, artikel, website, dan skripsi sejenis yang berhubungan tentang analisis sentimen, *text mining*, algoritma *naive bayes*, dan *support vector machine*.

1. Observasi

Observasi adalah teknik pengumpulan data yang melibatkan mempelajari hal-hal yang menjadi subjek penelitian secara langsung, pada tahap ini, penulis memanfaatkan Twitter API dengan bantuan *library* Tweepy. Jumlah data yang diambil sebanyak 1000 *tweet* dengan bantuan kata kunci ShopeeFood, shopeefood, dan *hashtag* #shopeefood yang berkaitan dengan akun @ShopeeID.

## Tahap Penelitian

Berikut ini adalah tahap penelitian yang akan dilakukan oleh penulis sebagai berikut :

Diagram

Description automatically generated

Gambar 1 Tahapan Penelitian Analisis Sentimen

1. Studi Literatur

Pada penelitian ini penulis mencari sumber – sumber yang berkaitan *text mining* dan analisis sentimen dengan menggunakan algoritma naïve bayes. Literatur yang digunakan sebagai referensi adalah dari jurnal “﻿Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes” oleh Safitri Juanita. Selain itu, penulis juga mempelajari buku “Bahasa Pemrograman Python” oleh Abdul Kadir.

1. Pengumpulan Data / *Crawling Data Tweet*

Penulis melakukan pengambilan sumber data dari kumpulan *tweet* menggunakan Twitter API dengan bantuan *library* Tweepy. Pencarian kumpulan *tweet* menggunakan kata kunci ShopeeFood, shopeefood, dan *hashtag* #shopeefood. Serta jumlah data yang diambil berkisar 1000 *tweet* berdasarkan mention akun @ShopeeID.

1. *Preprocessing*

Terlebih dahulu data yang sudah dikumpulkan dilakukan proses *preprocessing* dengan 5 langkah yaitu: (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) *Stopword removal*, Pemilihan kata kunci dari hasil token dari ﻿membuang kata-kata yang tidak memiliki arti atau tidak relevan, yaitu kata apa yang digunakan untuk merepresentasikan dokumen, disebut sebagai proses *filtering*. (4) *Tokenizing*, proses membagi urutan karakter menjadi komponen yang lebih kecil (kata/frasa) yang dikenal sebagai token. (5) *Stemming*, langkah-langkah untuk mengurangi jumlah indeks yang berbeda dalam sumber data sehingga istilah *suffix* atau *prefix* kembali ke bentuk dasarnya.

1. Klasifikasi

Langkah ini merupakan prosedur klasifikasi yang akan menggunakan *Naïve Bayes Classifier* untuk memastikan titik data mana yang termasuk dalam evaluasi positif dan negatif. Prosedur ini dilakukan setelah melakukan tahap *preprocessing*.

# daftar pustaka

[1] F. A. Widyanita, “Analisis Pengaruh Kualtitas Pelyanan E-commerce Terhadap Kepuasan Konsumen Shopee Indonesia Pada Mahasiswa Fakultas Ekonomi UII Pengguna Shopee,” vol. 151, no. 2, pp. 10–17, 2018.

[2] 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.

[3] Supangat, “Peluang Bisnis Makanan Dan Minuman Di Shopee Food Bagi UMKM,” 2021, [Online]. Available: http://repository.untag-sby.ac.id/12484/1/1211900355\_Anisa%27ul Khomariyah\_E-Bisnis\_ETS.pdf.

[4] A. D. Sofia, S. A. Rohman, and H. Chaniago, “Analisis Pengaruh Penggunaan Aplikasi Food Delivery Services Terhadap Penjualan UMKM di Masa Pandemi Covid-19 (Studi pada Kota Bandung, Indonesia),” *Int. J. Adm. Bus. Organ.*, vol. 2, no. 1, pp. 44–50, 2021.

[5] S. F. Nurjanah, R. R. Kurniati, and D. Zunaida, “Pengaruh E-commerce terhadap Keputusan Pembelian pada Belanja Online Shopee (Studi pada Konsumen Belanja Online Mahasiswa Universitas Islam Malang),” *J. Ilmu Adm. Niaga/Bisnis*, vol. 8, no. 3, pp. 154–162, 2019, [Online]. Available: http://riset.unisma.ac.id/index.php/jiagabi/article/view/3572.

[6] Y. S. Mahardika and E. Zuliarso, “Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial Twitter Menggunakan Algoritma Naives Bayes,” *Pros. SINTAK 2018*, no. 2015, pp. 409–413, 2018.

[7] 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.

[8] S. F. Pratama, R. Andrean, 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.

[9] 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.

[10] B. Laurensz and Eko Sediyono, “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.

[11] A. Sari, F. V., & Wibowo, “Analisis Sentimen Pelanggan Toko Online Jd. Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi,” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 2, no. 2, pp. 681–686, 2019.

[12] E. H. Muktafin, K. Kusrini, and E. T. Luthfi, “Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing,” *J. Eksplora Inform.*, vol. 10, no. 1, pp. 32–42, 2020, doi: 10.30864/eksplora.v10i1.390.

[13] M. Lestandy, A. Abdurrahim, and L. Syafa’ah, “Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 10, pp. 802–808, 2021.

[14] R. S. Perdana and M. A. Fauzi, “Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter dengan Metode Support Vector Machine dan Lexicon Based Features Twitter event detection View project Human Detection and Tracking View project,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1(12), no. October, pp. 1725–1732, 2017, [Online]. Available: https://www.researchgate.net/publication/320234928.

[15] L. Shanty Wato Wele Keaan, “Analisis Sentimen Review Shopee Berbahasa Indonesia Menggunakan Improved K-Nearest Neighbor dan Jaro Winkler Distance,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 7, pp. 2548–964, 2019, [Online]. Available: http://j-ptiik.ub.ac.id.

[16] 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.

[17] N. Fitriyah, B. Warsito, and D. A. I. Maruddani, “Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (Svm,” *J. Gaussian*, vol. 9, no. 3, pp. 376–390, 2020, doi: 10.14710/j.gauss.v9i3.28932.

[18] A. Kadir, *Logika Pemrograman Python*. PT Elex Media Komputindo, 2019.

[19] D. Deddy, “Implementasi Web Crawling untuk Pencarian Harga Sparepart Pada PT Asuransi Sinar Mas,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 7, no. 3, pp. 416–428, 2020, doi: 10.35957/jatisi.v7i3.505.

[20] 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.

[21] M. Athoillah, “Pengenalan Wajah Menggunakan SVM Multi Kernel dengan Pembelajaran yang Bertambah,” *J. Online Inform.*, vol. 2, no. 2, p. 84, 2018, doi: 10.15575/join.v2i2.109.

[22] L. A. Andika, P. A. N. Azizah, and R. Respatiwulan, “Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier,” *Indones. J. Appl. Stat.*, vol. 2, no. 1, p. 34, 2019, doi: 10.13057/ijas.v2i1.29998.