

BAB II

LANDASAN TEORI

2.1 Penelitian terdahulu

Tabel 2. 1 Penelitian Terdahulu

Tabel Penelitian Terdahulu 1	
Judul	Sentiment Analysis of User Satisfaction Towards Sales Promotion of Gojek Application Service Using Support Vector Machine (SVM)[14]
Nama Penulis	Haryani, Calandra Alencia,Widjaja, Andree E. Hery, Hery,Ferdinand, Ferry Vincenttius
Jurnal	Ultima InfoSys : Jurnal Ilmu Sistem Informasi
Tahun	2023
Hasil	Hasil penelitian ini akan membantu Gojek dalam memahami kepuasan pelanggan dari sudut pandang promosi atau penawaran dan membantu perusahaan menerapkan strategi promosi yang lebih baik agar tetap kompetitif terhadap pesaing lainnya.
Kesimpulan	Berdasarkan analisis sentimen yang dilakukan pada penjualan promosi layanan aplikasi Gojek, disimpulkan masih banyaknya keluhan yang menandakan kepuasan pelanggan masih perlu optimal. Dari data yang diperoleh, perusahaan perlu meninjau kembali kebijakan, proses penggunaan, dan perolehan promosi atau diskon yang ditawarkan kepada pelanggan serta mempertimbangkan strategi penjualan promosi yang lebih baik untuk layanan seperti GoFood, GoRide, dan GoCar untuk masa depan.
Tabel Penelitian Terdahulu 2	
Judul	Aspect-Based Sentiment Analysis on Application Review using CNN (Case Study : Peduli Lindungi Application)[15]
Nama Penulis	Aritonang, Putri Arta; Johan; Monika Evelin; Prasetiawan, Iwan
Jurnal	Ultima Infosys : Jurnal Ilmu Sistem Informasi
Tahun	2022

Hasil	Temuan riset menunjukkan bahwa model CNN berhasil mencapai hasil yang sangat baik dalam klasifikasi aspek dengan skor F1 mencapai 92,23%, sementara dalam klasifikasi sentimen, skornya mencapai 95,13%.
Kesimpulan	Penelitian ini menunjukkan hasil yang baik dari Berbasis Aspek Analisis Sentimen menggunakan model CNN pada klasifikasi aspek dan sentimen data review. Temuan riset menunjukkan bahwa model CNN berhasil mencapai hasil yang sangat baik dalam klasifikasi aspek dengan skor F1 mencapai 92,23%, sementara dalam klasifikasi sentimen, skornya mencapai 95,13%..
Tabel Penelitian Terdahulu 3	
Judul	Sentimental Analysis on Web Scraping Using Machine Learning Method[16]
Nama Penulis	Sahu, Saurabh; Divya, Km; Rastogi, Dr. Neeta; Yadav, Puneet Kumar; Perwej, Dr. Yusuf
Jurnal	Journal of Information and Computational Science
Tahun	2022
Hasil	Tujuan dari penelitian ini yaitu untuk mengetahui pemrosesan dalam NLP.
Kesimpulan	Untuk menetapkan peringkat sentimen tertimbang pada entitas, topik, tema, dan kategori yang disertakan dalam kalimat atau frasa, sistem analisis sentimen untuk analisis teks mengintegrasikan pemrosesan bahasa alami (NLP) dengan pendekatan machine learning.
Tabel Penelitian Terdahulu 4	
Judul	Pemanfaatan Big Data dalam Peningkatan Efektivitas Strategi Komunikasi Marketing Terpadu pada Perusahaan E-Commerce[12]
Nama Penulis	Syahdina Damayari Syira, Achmad Fauzi, Choiroel Woestho, Laurencia Vilani, Prado Dian Firmansyah, Demas Rizky Pratama, Atun Dwi Apriliana, Naufal Shafly Abdul Ghaffar, Dhea Amelia Putri
Jurnal	Jurnal Ekonomi Manajemen Sistem Informasi
Tahun	2023
Hasil	Penelitian ini bertujuan untuk memahami peran Big Data dalam strategi pemasaran komunikasi pada platform E-commerce.

Kesimpulan	Hasil studi menunjukkan bahwa Big Data memiliki kemampuan untuk secara akurat memprediksi harga dan volume pasokan yang akan disiapkan oleh perusahaan dalam jangka waktu yang signifikan.
------------	--

Tabel Penelitian Terdahulu 5

Judul	Analisa Konten Media Sosial E-Commerce Pada Instagram Menggunakan Metode Sentimen Analysis Dan Lda-Based Topic Modeling (Studi Kasus: Shopee Indonesia) Analysis of Content Social Media E-Commerce in Instagram Using Sentiment Analysis and Lda Based Topic Modeling (Study Case : Shopee Indonesia)[17]
Nama Penulis	Irawan Noor Kabiru, Puspita Kencana Sari, S.kom, MTI
Jurnal	eProceedings of Management
Tahun	2019
Hasil	Penelitian ini dilakukan dengan tujuan mengetahui sentimen pengguna shopee pada social media.
Kesimpulan	Hasil penelitian menunjukkan bahwa komentar-komentar di Instagram Shopee selama event 11:11 cenderung bersentimen negatif. Dengan menerapkan metode Naïve Bayes pada analisis sentimen dan pemodelan topik, dapat disimpulkan bahwa mayoritas komentar memiliki sentimen negatif.

Tabel Penelitian Terdahulu 6

Judul	Sentimen Analisis Aplikasi E-Commerce Berdasarkan Ulasan Pengguna Menggunakan Algoritma Stochastic Gradient Descent[18]
Nama Penulis	Wilya Kurnia
Jurnal	Jurnal Teknologi dan Sistem Informasi
Tahun	2023
Hasil	Penelitian ini bertujuan untuk meramalkan label sentimen dari ulasan yang diberikan oleh pengguna aplikasi Tokopedia dan Shopee selama bulan Maret hingga Mei 2021.
Kesimpulan	Menurut data penilaian dari Maret hingga Mei 2021, sekitar 66% dari sepuluh ribu ulasan

	pengguna Tokopedia menyatakan kepuasan yang tinggi terhadap platform e-commerce tersebut.
Tabel Penelitian Terdahulu 7	
Judul	Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based Dan Support Vector Machine[19]
Nama Penulis	Rachmad Mahendrajaya, Ghulam Asrofi Buntoro, Moh. Bhanu Setyawan
Jurnal	Journal of Komputek
Tahun	2019
Hasil	Mengumpulkan komentar pengguna tentang layanan Go-Pay dari Twitter untuk analisis sentimen menggunakan metode Lexicon Based. Data terdiri dari seribu dua ratus sepuluh entri, terbatas pada rentang waktu antara 24 Juli 2019 dan 30 Juli 2019. Tujuannya adalah untuk mengklasifikasikan sentimen dari komentar tersebut..
Kesimpulan	Kita meninjau sentimen terhadap Go-Pay melalui analisis ulasan pengguna di Twitter. Dari pendekatan berbasis lexicon, ditemukan bahwa mayoritas komentar pengguna cenderung positif, dengan sembilan ratus dua puluh tiga ulasan mendukung hal tersebut. Di sisi lain, ada dua ratus delapan puluh tujuh ulasan yang bersifat negatif.
Tabel Penelitian Terdahulu 8	
Judul	ANALISIS SENTIMEN MASYARAKAT TERHADAP VAKSINASI COVID-19 PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)[20]
Nama Penulis	Herwinskyah, Arita Witanti
Jurnal	Jurnal Sistem Informasi dan Informatika
Tahun	2022
Hasil	Analisis sentimen memiliki suatu tujuan untuk memeriksa dokumen atau teks yang mengandung komentar untuk mengidentifikasi sikap yang terkandung di dalamnya, sehingga dapat diklasifikasikan sebagai positif, negatif, atau netral.

Kesimpulan	Hasil penelitian menggunakan empat ribu tujuh ratus delapan data indeks tweet yang telah diolah menunjukkan bahwa 83,6% dari respon yang dianalisis bersifat positif, sedangkan 16,4% sisanya bersifat negative.
------------	--

Tabel Penelitian Terdahulu 9

Judul	Komparasi Metode KNN dan Naive Bayes Terhadap Analisis Sentimen Pengguna Aplikasi Shopee[21]
Nama Penulis	Salman Alfaris, Kusnawi
Jurnal	Indonesian Journal of Computer Science
Tahun	2023
Hasil	Penelitian ini bertujuan untuk mengevaluasi apakah Naïve Bayes memiliki tingkat akurasi yang lebih tinggi daripada K-Nearest Neighbors dalam konteks yang sama.
Kesimpulan	Menurut penelitian ini, K-Nearest Neighbors dan Naïve Bayes Classifier dapat diterapkan untuk menganalisis sentimen dari ulasan pengguna aplikasi Shopee di Playstore dengan tingkat akurasi mencapai 70,00% menggunakan metode K-Nearest Neighbor.

Tabel Penelitian Terdahulu 10

Judul	Perbandingan Algoritma Klasifikasi Support Vector Machine, Random Forest dan Logistic Regression Pada Ulasan Shopee[22]
Nama Penulis	Saepudin, Asep, Faqih, Ahmad , Dwilestari, Gifthera
Jurnal	Jurnal TEKNO KOMPAK
Tahun	2024
Hasil	Hasil evaluasi algoritma Support Vector Machine mendapatkan akurasi 91%, lalu Random Forest mendapatkan 94% akurasi dan akurasi Logistic Regression sebesar 86%
Kesimpulan	Dari hasil perbandingan algoritma dipastikan bahwa Random Forest memiliki akurasi tertinggi diikuti oleh Support Vector Machine dan Logistic Regression.

Berdasarkan tabel 2.1 menunjukkan penelitian penelitian terdahulu yang sudah dilakukan depan topik yang sesuai dengan Sentimen Analisis. Berdasarkan

penelitian terdahulu, dilakukan penelitian bahwa kepuasan pelanggan terhadap aplikasi gojek masih kurang optimal ditandakan dengan masih banyaknya keluhan. Dari data yang diperoleh Perusahaan gojek perlu meninjau kebijakan yang ada dan layanan untuk masa depan Perusahaan[14]. Penelitian terdahulu selanjutnya menunjukkan bahwa aplikasi peduli lindungi memiliki pelayanan yang baik ditunjukan dengan skor akurasi yang bagus[15]. Berdasarkan penelitian selanjutnya menunjukkan sentiment analysis dengan cara mengambil data menggunakan scrapping dan mengetahui pemrosesan dalam NLP (Natural Language processing)[16].

Dilakukan penelitian bahwa *big data* memiliki peran penting dalam meningkatkan efisiensi dan kecepatan bisnis *e-commerce* dengan memungkinkan analisis data yang akurat, aksesibilitas yang lebih baik, serta penerapan aplikasi analisis yang memberikan pengalaman pelanggan yang lebih baik[12]. Pertumbuhan pesat *e-commerce* di Indonesia, yang menempati pada tahun 2018, mereka adalah yang paling cepat di dunia, dan pada tahun 2019 mereka menempati peringkat ketujuh secara global, mencerminkan kebutuhan akan platform perdagangan elektronik yang lebih efisien dan fleksibel bagi konsumen modern. Selanjutnya, penelitian terdahulu ini menganalisis sentimen pengguna Instagram terhadap Shopee selama periode penjualan 11:11 dengan menggunakan teknik analisis sentimen *Naïve Bayes* dan pemodelan topik untuk mengidentifikasi topik-topik yang sering dibahas dalam komentar pengguna. Metodologi penelitian melibatkan pengumpulan data komentar dari Instagram, Shopee, *preprocessing data*, analisis sentimen, dan pemodelan topik untuk mengekstrak pengetahuan dari data tidak terstruktur tersebut[17].

Pada penelitian ini *e-commerce*, terutama melalui platform seperti Tokopedia dan Shopee, telah mempercepat pertumbuhan ekonomi dengan memungkinkan transaksi jual-beli secara elektronik, didukung oleh ulasan pengguna yang memberikan pandangan penting terhadap kualitas layanan dan produk. Penelitian menggunakan metode *machine learning*, untuk menganalisis sentimen ulasan pengguna pada platform Google Play Store, memberikan wawasan

yang berharga bagi tokopedia dan shopee dalam memahami persepsi pengguna terhadap aplikasi mereka[18]. Selain dari *e-commerce*, Perkembangan teknologi komunikasi menuju digitalisasi telah mendorong penggunaan dompet *digital* seperti Go-Pay di Indonesia, yang tercermin dari meningkatnya transaksi *online*, dengan Twitter menjadi platform untuk menganalisis sentimen pengguna terhadap layanan tersebut menggunakan metode *Lexicon Based* dan *Support Vector Machine*[19].

Berdasarkan penelitian menggunakan metode analisis sentimen dengan algoritma Mesin Vector Dukungan untuk menggali tanggapan masyarakat Indonesia terhadap kebijakan vaksinasi COVID-19 di platform media sosial Twitter dengan fokus pada pembobotan data tweet dan klasifikasi sentimen positif atau negatif[20]. Selanjutnya dalam penelitian lain juga ada penelitian Penelitian yang menggunakan algoritma KNN dan Naïve Bayes untuk analisis sentimen ulasan produk di Shopee, menunjukkan bahwa KNN memiliki akurasi lebih tinggi daripada Naïve Bayes, dengan nilai akurasi masing-masing 73,33% dan 70,48%[23]. Selanjutnya dalam penelitian lain juga ada penelitian yang menggunakan algoritma SVM, Random Forest dan logistic regression untuk analisis sentiment ulasan pada produk di shopee dari Hasil evaluasi algoritma Support Vector Machine mendapatkan akurasi 91%, lalu Random Forest mendapatkan 94% akurasi dan akurasi Logistic Regression sebesar 86% [22].

Berdasarkan beberapa penelitian terdahulu, terdapat research gap dalam penelitian ini yang berkaitan dengan pemilihan algoritma untuk melakukan sentimen analisis[14][15][16][17][19][20][23][22]. Alur penelitian yang digunakan mengacu pada metode CRISP-DM[14][15][16][18][19][20][23][22].Kebaruan dari penelitian ini adalah e-commerce yang dipilih, algoritma yang dibandingkan serta rekomendasi e-commerce yang diberikan. Selain itu pada penelitian ini rekomendasi yang diberikan ada 2 yaitu algoritma yang memiliki hasil terbaik dan rekomendasi e-commerce.

2.2 Sentimen Analisis

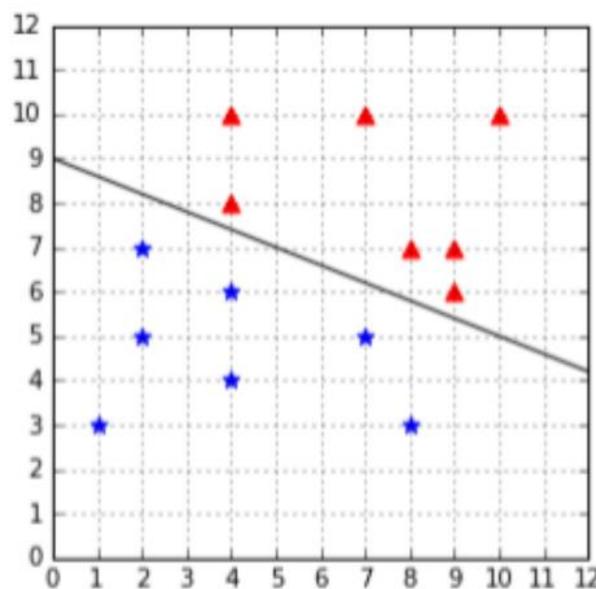
Analisis sentimen merupakan aspek ilmiah dalam memisahkan teks berdasarkan opini yang terkandung di dalamnya [17]. Analisis sentimen melibatkan

pengidentifikasi orientasi emosional dalam suatu teks berdasarkan konteksnya, mengklasifikasikan apakah teks tersebut bersifat positif, negatif, atau netral[17]. Dalam hal ini dalam melakukan sentimen analisis berupa review dari pengguna. Pada situs Google Play Store dilengkapi dengan fitur review agar pengguna bisa mengirimkan ulasan berupa kritik, pujian, saran, atau penilaian lain terhadap suatu aplikasi[18]. Ulasan dari pengguna ini sering digunakan sebagai alat tolak ukur yang efektif dan efisien dalam menemukan informasi terhadap suatu produk atau jasa[18].

2.3 Teori Algoritma yang digunakan

2.3.1 Support Vector Machine

Support Vector Machine adalah sebuah metode yang cukup baru yang digunakan untuk membuat prediksi, baik untuk mengelompokkan data (klasifikasi) maupun memperkirakan nilai (regresi). Algoritma SVM asli diciptakan oleh Vladimir Vapnik dan turunan standar saat ini (margin lunak) diusulkan oleh Corinna Cortes dan Vapnik Vladimir[19]. SVM adalah algoritma klasifikasi yang populer dan dapat digunakan untuk mengklasifikasikan data ke dalam dua atau lebih kelas[22]. Berikut merupakan gambaran hyperplane:



Gambar 2. 1 Hyperplane memisahkan dua kelas

Berdasarkan Gambar 2.1 Dua titik yang menjadi patokan hyperplane disebut dengan support vector. Dapat dilihat bahwa memiliki dua kelompok data yang disebut klasifikasi, kemudian tugas SVM adalah membagi dua kelompok ini sebaik mungkin atau menentukan hyperplane terbaik, pembagian dimana garis batasnya dapat memisahkan dua kelompok dengan jarak terjauh antara titik terluar di masing masing kelompok dengan garis pembatas itu sendiri[24].

2.3.2 Naïve Bayes

Algoritma Naive Bayes adalah salah satu algoritma klasifikasi yang populer dalam bidang machine learning. Algoritma ini didasarkan pada Teorema Bayes dengan asumsi naif (naive assumption), yaitu anggapan bahwa semua atribut dalam dataset adalah independen secara kondisional terhadap kelasnya[23]. Berikut merupakan rumus dari algoritma naïve bayes:

$$P(A|B) = \frac{P(A) \times P(B|A)}{P(B)}$$

Gambar 2. 2 Rumus Naive Bayes

Berdasarkan Gambar 2.2 rumus tersebut dapat dijelaskan bahwa Dimana B adalah data yang akan dicari class-nya, sedangkan A merupakan hipotesis class. $P(A|B)$ merupakan peluang terjadinya A terhadap kondisi B, $P(A)$ merupakan peluang terjadinya A, dan $P(B)$ adalah peluang terjadinya B.

2.3.3 K-Nearest Neighbor (K-NN)

Algoritma KNN menjadi salah satu pilihan utama dalam machine learning karena kemudahan penerapannya untuk menangani masalah yang kompleks. KNN mampu mengatasi ketidakseimbangan kelas dalam dataset. Karena algoritma ini menggunakan mayoritas tetangga terdekat untuk pengklasifikasian, ia dapat memberikan hasil yang baik bahkan ketika ada kelas minoritas yang kurang representatif[23]. Berikut merupakan rumus KNN:

$$D(x_i, y_i) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Keterangan :

D (xi, yi) =jarak

Xi = data latih

Yi = data uji

i = variabel data

n = Dimensi data

Gambar 2.3 Rumus KNN

Berdasarkan Gambar 2.3 rumus tersebut dapat dijelaskan Langkah Langkah dalam algoritma K-NN sebagai berikut:

1. Menentukan nilai parameter K.
2. Menghitung jarak antara data yang akan diprediksi (data uji) dengan contoh data pelatihan. Jika data memiliki atribut numerik, maka menggunakan metrik jarak tertentu
3. Urutkan jarak secara menurun (descending).
4. Pilih K jarak terdekat dari data uji berdasarkan nilai parameter K
5. Tentukan kelas terbanyak dari K dan diklasifikasikan ke dalam kelas tersebut.

2.4 Google Collabs dan Web Scrapping

Pengambilan data dari Google Play Store menggunakan algoritma Python dengan alat bantu Google Colab yang dikembangkan oleh Google[21]. Data ulasan tersebut disimpan dalam format CSV (Comma Separated Values) agar dapat digunakan dalam tahap preprocessing Data selanjutnya[21]. Selain itu dapat menggunakan *web scrapping* sebagai perangkat lunak yang mensimulasikan pengalaman manusia di *Web* interaksi digunakan dalam proses *web scrapping*, yang mencakup ekstraksi konten sistematis dan penggabungan dari *Internet*[16]. Yang membedakan dalam hal ini adalah *web scrapping* digunakan untuk menyelesaikan operasi ini secara otomatis dengan software aplikasi yang ada atau extension chrome tanpa menggunakan skrip[16].