


Krisna Novianto

ARTIKEL JURNAL: Menggali Sentimen Konsumen terhadap produk iPhone Bekas di Tokopedia Pendekatan Naïve Bayes d...

 Quick Submit

 Quick Submit

 Universitas Bhayangkara Jakarta Raya

Document Details

Submission ID

trn:oid::1:3296734269

Submission Date

Jul 16, 2025, 2:40 PM GMT+7

Download Date

Jul 16, 2025, 2:59 PM GMT+7

File Name

dan_Support_Vector_Machine_-_202110715003_KRISNA_NOVIANTO_1.doc

File Size

2.5 MB

13 Pages

3,325 Words

21,352 Characters




12% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- ▶ Bibliography
- ▶ Quoted Text
- ▶ Cited Text
- ▶ Small Matches (less than 8 words)
- ▶ Internet sources

Top Sources

- 0%  Internet sources
- 11%  Publications
- 6%  Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Top Sources

- 0% Internet sources
- 11% Publications
- 6% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Publication	Ridho Sholehurrohman, Ivani Valentine, Kurnia Muludi. "Penilaian Kualitas Udara...	1%
2	Student papers	Universitas Negeri Manado	<1%
3	Student papers	Universitas Diponegoro	<1%
4	Publication	M Febrin As Shidiq, Debby Alita. "ANALISIS SENTIMEN MASYARAKAT TERHADAP ...	<1%
5	Student papers	University of Central Florida	<1%
6	Publication	Octavia Salwa Dzaky Fadhillah, Jajam Haerul Jaman, Carudin Carudin. "PERBANDI...	<1%
7	Publication	Rahmad Firdaus, Rafiq Al Hariri, Hasanatul Fu'adah Amran. "Sentimen Analisis M...	<1%
8	Student papers	Institut Bisnis dan Teknologi Indonesia (INSTIKI)	<1%
9	Publication	Sitti Nurlaili Qofifa, Rizka Ardiansyah, Yuri Yudhaswana Joeffie, Wirdayanti, Nouv...	<1%
10	Publication	Mohammad Bayu Anggara. "Comparison of Naïve Bayes and SVM Methods in Sen...	<1%
11	Publication	Atiqah Hanifah Shalihah, Gigih Forda Nama, Puput Budi Wintoro, Yessi Mulyani. "...	<1%

12	Publication	Muhammad Rizky Julianto, Yuma Akbar, Tri Wahyudi. "Analisis Sentimen Respon ..."	<1%
13	Student papers	Universitas Prima Indonesia	<1%
14	Student papers	Universitas Sumatera Utara	<1%
15	Student papers	Universitas Muhammadiyah Purwokerto	<1%
16	Publication	Alfun Roehatul Jannah, Ria Kristi, Muhammad Habibi. "Metode Latent Dirichlet A..."	<1%
17	Student papers	Universitas Negeri Jakarta	<1%
18	Publication	Herlawati Herlawati, Rahmadya Trias Handayanto, Prima Dina Atika, Fata Nidaul ...	<1%
19	Publication	Tamora Nonia Wijaya, Rini Indriati, Muhammad Najibulloh Muzaki. "ANALISIS SE..."	<1%
20	Publication	Aulia Farhani, Sutisna. "Analisis Sentimen Terhadap Kendaraan Listrik di Indones..."	<1%
21	Publication	Fajar Romadoni, Yuyun Umaidah, Betha Nurina Sari. "Text Mining Untuk Analisis ..."	<1%
22	Publication	Marwa Sulehu, Markani Markani. "Integrasi Pemrosesan Bahasa Alami dalam Bot..."	<1%
23	Publication	Muhamad Fajar Yudhistira Herjanto, Carudin Carudin. "ANALISIS SENTIMEN ULAS..."	<1%
24	Publication	Muhammad Alif Maghriby, Herry Irawan. "Analisis Persepsi Publik Mengenai Res..."	<1%
25	Publication	Rakhmi Khalida, Siti Setiawati. "Analisis Sentimen Sistem E-Tilang Menggunakan ..."	<1%

1. Pendahuluan

iPhone merupakan salah satu *smartphone* yang diminati oleh masyarakat Indonesia. Namun, harga seri terbaru iPhone yang mencapai puluhan juta rupiah sering dianggap terlalu tinggi oleh sebagian konsumen yang berasal dari kalangan menengah ke bawah (Hakim & Harahap, 2024). Bahkan oleh konsumen luar negeri, menilai iPhone sebagai produk *overpriced*. Berdasarkan laporan dari Blibli Friends, perbandingan harga antara iPhone dan Android cukup mencolok. Di Amerika Serikat, harga *smartphone* Android berada di kisaran US\$100 atau sekitar Rp1,4 juta rupiah, sedangkan iPhone terbaru dijual mulai dari US\$999 atau sekitar Rp14 juta. Selain itu, merek *Apple* sering diasosiasikan dengan status sosial dan gaya hidup modern, sehingga meningkatkan daya tariknya di kalangan masyarakat.

Selain terkenal karena spesifikasi dan performanya yang unggul, produk Apple juga sering diasosiasikan dengan status sosial dan gaya hidup modern, yang semakin menambah daya tarik di mata konsumen (N'da et al., 2023). Sebagai alternatif, banyak masyarakat memilih untuk membeli produk iPhone *second* yang masih dalam kondisi baik. Pilihan ini memungkinkan mereka tetap bisa menikmati teknologi dan fitur canggih dari produk iPhone tanpa harus mengeluarkan biaya sebesar saat membeli unit baru.

Tokopedia merupakan salah satu *marketplace* terpercaya dan terkemuka peringkat kedua di Indonesia dan mencatat total kunjungan global sebanyak 58 juta berdasarkan data dari SimilarWeb pada Februari tahun 2025. Tokopedia menyediakan berbagai macam produk, salah satunya adalah produk elektronik yaitu iPhone *second* (Nasikah & Fuadi, 2022). Pembelian secara daring melalui Tokopedia dapat memberikan kemudahan bagi masyarakat untuk berbelanja tanpa harus mengunjungi toko fisik, sehingga lebih hemat waktu dan efisien. Namun tantangan tetap ada, terutama dalam membangun kepercayaan konsumen kepada penjual terhadap produk *second* seperti iPhone. Kekhawatiran terhadap kualitas dan keaslian barang yang dijual di Tokopedia (Salsabilaa et al., 2024).

Tokopedia menyediakan fitur ulasan yang menjadi salah satu elemen penting dalam membangun kepercayaan pembeli terhadap penjual. Penilaian ini membantu calon pembeli memahami reputasi penjual dan kualitas produk yang ditawarkan konsumen (Hidayansyah, 2024). Namun jumlah ulasan yang begitu banyak di fitur ulasan membuat proses membaca dan menganalisis secara manual menjadi tidak efisien dan memakan banyak waktu. Oleh sebab itu, dibutuhkan metode yang lebih efisien untuk mengelola ulasan dalam jumlah besar, salah satunya melalui penerapan teknik analisis sentimen (Handayanto et al., 2021).

Analisis sentimen adalah metode yang digunakan untuk mengolah data teks, seperti ulasan, guna memahami opini atau tanggapan konsumen yang terkandung apakah bersifat positif atau negatif, sehingga dapat dijadikan dasar untuk mengambil keputusan yang lebih tepat terkait suatu produk, layanan, atau topik tertentu. (Herlawati et al., 2024).

Beberapa penelitian sebelumnya telah menyatakan bahwa pendekatan analisis sentimen efektif digunakan untuk menangkap persepsi konsumen terhadap suatu produk. Dalam salah satu penelitian yang membahas ulasan produk masker *Top Brand* di platform

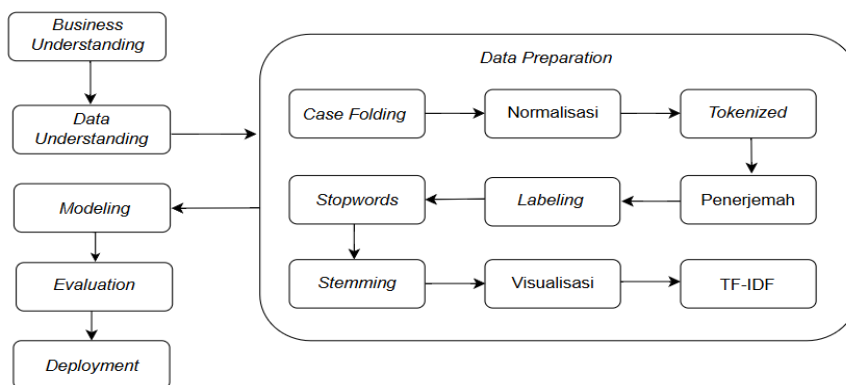
Tokopedia dengan menerapkan algoritma *Naïve Bayes*, ditemukan bahwa sebagian besar ulasan memiliki kecenderungan positif. Hasil penelitian tersebut juga memperlihatkan bahwa algoritma *Naïve Bayes* memiliki kemampuan klasifikasi sentimen yang cukup baik, dengan akurasi mencapai 88%.

Penelitian lainnya yang mengkaji analisis sentimen terhadap produk *skincare* dari merek *Skintific* menggunakan data komentar dari media sosial, yang mencakup tanggapan pengguna dan identitas akun mereka. Penelitian ini menerapkan algoritma *Support Vector Machine* (SVM) dengan kernel linear untuk mengklasifikasikan opini pengguna. Hasil penelitian menunjukkan bahwa metode SVM tersebut cukup efektif dalam membedakan sentimen positif dan negatif pada ulasan berbahasa Indonesia, dengan tingkat akurasi mencapai 94%.

Penelitian ini berfokus pada penerapan dua metode klasifikasi, yaitu *Naïve Bayes* dan *Support Vector Machine* (SVM), dalam menganalisis sentimen konsumen terhadap ulasan iPhone *second* yang dijual di Tokopedia. Tujuan dari studi ini adalah untuk menilai dan membandingkan tingkat akurasi kedua algoritma tersebut dalam mengidentifikasi opini pengguna, baik yang bersifat positif maupun negatif. Melalui perbandingan ini, diharapkan dapat diperoleh gambaran yang lebih menyeluruh mengenai efektivitas masing-masing metode dalam memahami preferensi serta kekhawatiran pelanggan terhadap produk iPhone *second* di platform *e-commerce*.

2. Metode Penelitian

Penelitian ini dilakukan menggunakan teknik CRISP-DM (*Cross-Industry Standard Process for Data Mining*) sebagai kerangka kerja dalam pelaksanaan penelitian ini. Gambar 1 merupakan kerangka penelitian yang menjelaskan langkah-langkah dalam penelitian ini.



Sumber: Hasil Penelitian (2025)

Gambar 1. Kerangka Penelitian

2.1 Machine Learning

Machine Learning merupakan cabang dari kecerdasan buatan yang berfokus pada pengembangan algoritma dan model yang memungkinkan komputer untuk belajar dan membuat prediksi atau keputusan tanpa perlu diprogram secara eksplisit. (Roihan et al., 2020).

2.2 Analisis Sentimen

Analisis sentimen merupakan teknik yang digunakan untuk memahami opini, perasaan, atau penilaian yang diberikan oleh pengguna internet melalui media sosial (Amelia & Yustiana, 2024). Hal ini sangat relevan dalam dunia bisnis, di mana analisis sentimen diterapkan untuk menganalisis ulasan produk, layanan pelanggan, dan tanggapan konsumen terhadap suatu produk atau merek.

2.3 Web Scraping

Web Scraping merupakan metode pengumpulan data secara otomatis dari situs web menggunakan teknik-teknik tertentu untuk mengekstrak informasi yang tersedia di halaman web (Henrys, 2021). Teknik *Web Scraping* dapat memanfaatkan bahasa pemrograman seperti Python, yang sangat populer karena memiliki pustaka-pustaka seperti *BeautifulSoup*, *Scrapy*, dan *Selenium*.

2.4 Naïve Bayes

Naïve Bayes merupakan algoritma pembelajaran mesin yang mengandalkan *Teorema Bayes* untuk memprediksi kategori dari suatu data. Algoritma ini beroperasi dengan asumsi bahwa setiap fitur dalam data saling independen, yang membuat perhitungan probabilitas menjadi lebih efisien dan lebih sederhana. Berikut perhitungan menggunakan *Teorema Bayes* dan distribusi *multinomial*, yang dijelaskan melalui rumus berikut ini (Dania Siregar et al., 2023).

$$P(C_k|x) = \frac{P(C_k) \prod_{i=1}^n P(x_i|C_k)}{P(x)} \quad (1)$$

Keterangan:

$P(C_k|x)$ = Probabilitas kelas C_k diberikan fitur X (posterior).

$P(C_k)$ = Kelas tertentu yang ingin diidentifikasi dalam data X .

$P(x_i|C_k)$ = Kelas yang dimaksud untuk dianalisis.

$P(x)$ = Probabilitas total dari semua kemungkinan fitur

2.5 Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan algoritma pembelajaran mesin yang beroperasi dengan cara mencari *hyperplane* terbaik yang dapat memisahkan dua kelas data dengan margin terbesar di dalam ruang fitur (Oktavia et al., 2023). Dalam SVM, titik data yang paling krusial adalah *support vectors*, yaitu titik-titik yang paling dekat dengan *hyperplane* pemisah (Widyastuti et al., 2024). *Support Vectore Machine*, *hyperplane* pemisah dijelaskan melalui persamaan berikut:

$$w_0 + w_1x_1 + w_2x_2 > 0 \quad (2)$$

Oleh karena itu, setiap titik yang terletak di atas *hyperplane* pemisah memenuhi kondisi:

$$w_0 + w_1x_1 + w_2x_2 < 0 \quad (3)$$

17 Sedangkan setiap titik yang terletak di bawah hyperplane pemisah memenuhi kondisi:

$$w_0 + w_1x_1 + w_2x_2 > 0 \quad (4)$$

3. Hasil dan Pembahasan

3.1 Business Understanding

11 Tren pembelian iPhone bekas secara online menunjukkan peningkatan, khususnya melalui platform Tokopedia. Dari hasil survei awal, ditemukan bahwa 58,1% responden memilih Tokopedia karena dinilai memiliki tingkat kepercayaan yang tinggi. Fakta ini memperlihatkan bahwa rasa percaya terhadap platform sangat memengaruhi keputusan konsumen dalam membeli perangkat elektronik bekas secara daring. Penelitian ini berupaya mengevaluasi persepsi konsumen terhadap produk dan layanan melalui analisis sentimen pada ulasan pengguna. Proses klasifikasi dilakukan dengan pendekatan *machine learning* menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM) yang dikenal memiliki performa baik dalam pengolahan teks. Temuan dari penelitian ini diharapkan dapat memberikan wawasan kepada pelaku bisnis dan platform *e-commerce* dalam memahami pola sentimen konsumen, merancang strategi pemasaran yang lebih terarah, serta memberikan dukungan pengambilan keputusan bagi calon pembeli berdasarkan kecenderungan ulasan yang tersedia.

3.2 Data Understanding

Data ulasan pelanggan diperoleh dari empat toko yang menjual iPhone *second* di Tokopedia, yakni *Boboy Store*, *JoyTaroGadgetOfficial*, *Maxsense Store*, dan *Yunishop00*. Pengambilan data dilakukan secara otomatis menggunakan metode *web scraping* dengan bantuan library *Selenium*, *BeautifulSoup*, dan *Pandas*.

Ulasan yang dianalisis berasal dari pembeli yang memposting ulasan dalam rentang waktu Januari 2023 hingga Januari 2025, dengan total 1.863 baris data dalam format CSV. Gambar 2 menampilkan dataset dari salah satu toko hasil teknik *web scraping*, yaitu Toko Boboy Store.

	Username	Product	Tanggal	Rating	Ulasan
0	Ranita	Iphone 13 128Gb / 256Gb / 512Gb Bekas Fullset	1 hari lalu	5	toko sangat bagus udah biasa langganan disini
1	Alexander	Iphone 14 Pro 128Gb / 256Gb / 512Gb / 1TB Seco...	1 hari lalu	5	barang bagus, imei tembus, hape mulus. semoga ...
2	Erdy	Iphone 12 Pro 128gb / 256gb / 512gb Bekas Fullset	1 hari lalu	5	Sangat memuaskan barang dari bosku yg 1 ini ,...
3	D***g	Iphone 13 Mini 128Gb / 256Gb / 512Gb Second Im...	2 hari lalu	5	Makasih banyak bosku, sudah sampai barangnya a...
4	upi	Iphone 13 128Gb / 256Gb / 512Gb Bekas Fullset	2 hari lalu	5	Sudah yg ketiga kalinya beli disini, hasilnya ...

Sumber: Hasil Penelitian (2025)

Gambar 2. Hasil *Scraping* Ulasan pada Toko Boboystore

Proses penyaringan dilakukan dengan Python untuk memastikan hanya ulasan produk iPhone *second* yang digunakan dalam analisis. Hasil penyaringan ditampilkan pada Gambar 3.

	Username	Product	Tanggal	Rating	Ulasan
13	MUSTARING	Apple Charger Original 20 W Usb C Power Adapto...	3 minggu lalu	5	Barang sangat baik
17	Andhika	Ipad Air 2 16GB 32GB 64GB 128GB Wifi Only Wifi...	3 minggu lalu	5	Admin ramah, toko yang terjamin
18	dicky	iPad 8 7 6 5 128GB 32GB WiFi Only Wifi Cell Fu...	3 minggu lalu	5	alus madep pokona mah
32	Bagas	iPad Pro M1 2021 11 & 12.9 inch WiFi Cellular ...	1 bulan lalu	5	Mantap banget sih masih aman sampe sekarang so...
39	Adi	SEIN Samsung Galaxy S10 Plus 12/1TB 8/512GB 8/...	1 bulan lalu	5	recommended seller

Sumber: Hasil Penelitian (2025)

Gambar 3. Hasil Produk Bukan iPhone *Second*

Data dari keempat toko yang telah disaring digabungkan menjadi satu dataset menggunakan Python, kemudian disimpan dalam format CSV. Gambar 4 menampilkan hasil akhir proses tersebut.

	Username	Product	Tanggal	Rating	Ulasan
0	N***a	iphone 11 64gb 128gb second fullset mulus terawat	Hari ini	5	BH suntikan, baterai sudah pernah diganti... Tid...
1	Irfan	iPhone 11 Second IMEI Terdaftar BC (BeaCukai)...	1 hari lalu	5	Barang sesuai pesanan, mulus banget, IMEI terd...
2	Muhammad	iPhone XR 64gb 128gb second fullset ex inter	2 hari lalu	5	baterai cepat panas. boros pemakaian, sehari s...
3	A***t	iphone 13 pro max 256gb second fullset mulus e...	4 hari lalu	5	Seller Fast Respon, Produk sesuai deskripsi, o...
4	W***y	iPhone 15 pro 256gb second fullset mulus terawat	5 hari lalu	5	Kamera jernih bgt aku pake sampai sekarang ser...
...
1858	D***i	IPHONE 11 GREEN 128GB MULUS LIKE NEW 100% ...	Lebih dari 1 tahun lalu	5	pengiriman cepat sampai, toko bertanggung jawa...
1859	A***a	IPHONE 11 PRO MAX SILVER 256GB 100% ORIGINAL [...]	Lebih dari 1 tahun lalu	5	Barangnya original 📶📶 Sinyal tidak terblokir ...
1860	Yassina	IPHONE 12 PRO MAX PACIFIC BLUE 256GB MULUS LIK...	Lebih dari 1 tahun lalu	5	Barang mulus, bh 97, pengirimannya cepat, pac...
1861	eka	IPHONE 12 PRO MAX PACIFIC BLUE 512GB ORIGINAL ...	Lebih dari 1 tahun lalu	5	barang ok, msh bagus dan mulus kaya baru, toko...
1862	Hermawan	IPHONE 13 PRO SIERRA BLUE 128GB MULUS LIKE NEW...	Lebih dari 1 tahun lalu	5	Barangnya original 🙌, Sinyal aman tidak berma...

1863 rows x 5 columns

Sumber: Hasil Penelitian (2025)

Gambar 4. Hasil Penggabungan Dataset

3.3 Data Preparation

Tahap ini meliputi proses awal dalam menyiapkan data ulasan iPhone *second* di Tokopedia sebelum memasuki tahap pemodelan.

1). Case Folding

Pada tahap *case folding*, data dibersihkan untuk meningkatkan kualitasnya. Langkah ini mencakup penghapusan emoji, angka, spasi ganda, karakter tidak relevan, serta spasi di sekitar tanda baca. Nilai null dan kolom yang tidak relevan dihapus, sementara kolom *Rating* dan *Ulasan* dipertahankan karena dianggap penting untuk analisis sentimen.

2). Tokenized

Pada tahap tokenisasi, teks dipecah menjadi unit-unit kecil seperti kata atau frasa untuk memudahkan pemrosesan lanjutan. Perbandingan sebelum dan sesudah tokenisasi ditampilkan pada Gambar 5.

23

22

Ulasan	Tokenized
bh disuntik tidak ril btra i cpt panas ceritany...	[bh, disuntik, tidak, ril, btra i, cpt, panas, ...
sesuai pesanan mulus bget imei terdaftar siny...	[sesuai, pesanan, mulus, bget, imei, terdaftar...
baterai cept panas bagian microphone nya berke...	[baterai, cept, panas, bagian, microphone, nya...
seller fast respon produk sesuai deskripsi ove...	[seller, fast, respon, produk, sesuai, deskrip...
kamera jernih bgt aku pake sampai sekarang ser...	[kamera, jernih, bgt, aku, pake, sampai, sekar...

Sumber: Hasil Penelitian (2025)

Gambar 5. Perbedaan Teks antara Kolom Ulasan dan Kolom *Tokenized*

3). Normalisasi

Tahap normalisasi dilakukan setelah tokenisasi, dengan tujuan menyempurnakan data melalui koreksi kata tidak baku, singkatan, dan kesalahan penulisan. Tabel 1 menampilkan perbandingan kata sebelum dan sesudah normalisasi.

Tabel 1. Perbandingan Kata Sebelum dan Setelah Normalisasi

Sebelum	Sesudah	Sebelum	Sesudah
amboxing	unboxing	belanj	belanja
batre	baterai	cc	coba-coba
bh	battery health	cpt	cepat

Sumber: Hasil Penelitian (2025)

4). Normalisasi

Tahap penerjemahan dilakukan untuk mempermudah proses pelabelan sentimen dengan mengonversi teks ulasan ke dalam bahasa Inggris. Langkah ini juga memungkinkan pemanfaatan alat analisis yang lebih optimal karena banyak di antaranya dirancang khusus untuk bahasa Inggris. Gambar 6 menunjukkan perbandingan antara kolom *Ulasan* asli dan kolom *English_Review* hasil terjemahan.

Ulasan	English_Review
baterai health suntikkan tidak asli baterai ce...	the health battery injection is not authentic,...
sesuai pesanan mulus banget imei terdaftar sin...	according to the order is very smooth imei is ...
baterai cepat panas bagian microphone nya berk...	the battery is hot, the microphone is constrai...
seller cepat respon produk sesuai deskripsi ke...	seller fast product response according to the ...

Sumber: Hasil Penelitian (2025)

Gambar 6. Hasil Penerjemahan ke Kolom *English_Review*

5). Labelling

Pada tahap pelabelan, ulasan berbahasa Inggris diklasifikasikan ke dalam dua kategori sentimen: positif dan negatif. Proses ini menggunakan pustaka VADER (*Valence Aware Dictionary and Sentiment Reasoner*), yang menentukan polaritas berdasarkan skor komposit. Ulasan dengan skor ≥ 0 dilabeli sebagai positif, sedangkan skor < 0 dianggap negatif. Untuk meningkatkan akurasi analisis sesuai konteks, peneliti menambahkan kosakata khusus (*custom lexicon*) ke dalam VADER. Daftar kata negatif tambahan ditampilkan pada Tabel 2.

Tabel 2 Daftar Kata-Kata dengan Sentiment Negatif

6
6

Kata	Skor
<i>blisters</i>	-3.0
<i>hot_battery</i>	-3.0
<i>injection</i>	-3.0

Sumber: Hasil Penelitian (2025)

Selanjutnya dilakukan pengecekan manual untuk memastikan kesesuaian hasil pelabelan, karena tidak semua ulasan iPhone *second* di Tokopedia terklasifikasi dengan tepat secara otomatis. Gambar 7 menunjukkan hasil akhir pelabelan berdasarkan polaritas sentimen masing-masing ulasan.

Rating	Ulasan	English_Review	Label
1	baterai health suntikkan tidak asli baterai ce...	the health battery injection is not authentic,...	negatif
5	sesuai pesanan mulus banget imei terdaftar sin...	according to the order is very smooth imei is ...	positif
2	baterai cepat panas bagian microphone nya berk...	the battery is hot, the microphone is constrai...	negatif
5	seller cepat respon produk sesuai deskripsi ke...	seller fast product response according to the ...	positif
5	kamera jernih banget aku pakai sampai sekarang...	i really use clear camera until now i often us...	positif

Sumber: Hasil Penelitian (2025)

Gambar 7. Hasil dari Proses Pelabelan

6). Stopwords

Kata-kata yang tidak memiliki nilai informasi penting dihapus menggunakan proses *stopwords removal*. Hasil implementasinya ditampilkan pada Tabel 3.

Tabel 3. Hasil dari Proses Pelabelan

Sebelum Stopwords	Sesudah Stopwords
<i>minus</i> nya hanya sinyal bulanan imei tidak terdaftar bea cukai tapi semoga aja bertahan lama <i>thanks</i>	<i>minus</i> hanya sinyal bulanan imei tidak terdaftar semoga bertahan lama.

Sumber: Hasil Penelitian (2025)

7). Stemming

Proses *stemming* dilakukan untuk mengubah setiap kata pada kolom ulasan ke bentuk dasarnya. Hasil dari proses ini ditampilkan pada Tabel 9.

Tabel 4. Hasil dari Proses *Stemming*

Sebelum Stemming	Sesudah Stemming
<i>minus</i> nya hanya sinyal bulanan imei tidak terdaftar bea cukai tapi semoga aja bertahan lama <i>thanks</i> seller.	<i>minus</i> hanya sinyal bulan imei tidak daftar bea cukai semoga tahan lama.

Sumber: Hasil Penelitian (2025)

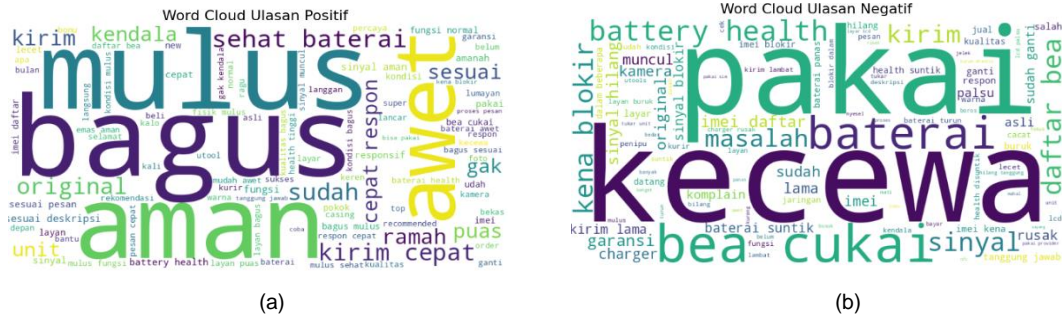
8). Visualisasi

Tahap selanjutnya adalah visualisasi data menggunakan *Word Cloud* untuk menggambarkan kata-kata yang paling sering muncul dalam ulasan iPhone *second* di Tokopedia, berdasarkan klasifikasi sentimen positif dan negatif. Visualisasi ini bertujuan untuk mengidentifikasi pola kepuasan dan ketidakpuasan konsumen, sehingga dapat memberikan

15

13

gambaran umum bagi calon pembeli mengenai aspek-aspek yang sering menjadi perhatian. Ukuran kata yang lebih besar menunjukkan frekuensi kemunculan yang lebih tinggi dalam dataset. Gambar 8 menampilkan perbandingan *Word Cloud* antara sentimen positif dan negatif.



Sumber: Hasil Penelitian (2025)

Gambar 8. *Word Cloud* dengan (a) Sentimen Positif (b) *Word Cloud* dengan Sentimen Negatif

9). **Term Frequency (TF) dan Inverse Document Frequency (IDF)**

Sebelum dilakukan transformasi menggunakan TF-IDF, data ulasan iPhone *second* dibagi menjadi tiga subset: data latih (70%), serta data validasi dan data uji (masing-masing dari sisa 30%). Pembagian ini bertujuan untuk mengevaluasi performa model secara menyeluruh. Proses TF-IDF dilakukan pada data latih untuk membangun pemahaman awal model terhadap kata-kata dan bobot pentingnya dalam teks. Tabel 5 menampilkan hasil perhitungan TF-IDF pada dua contoh ulasan.

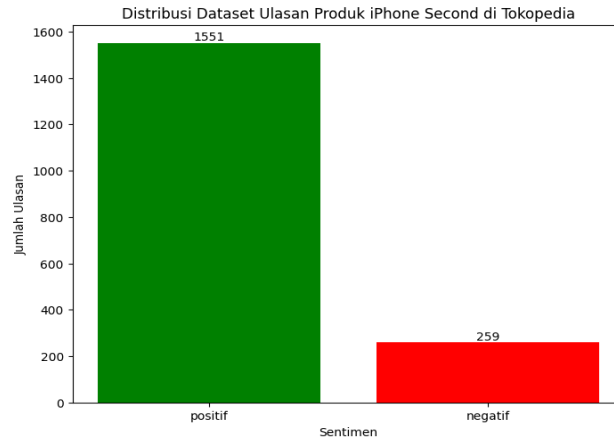
Tabel 5. Hasil perhitungan TF-IDF

Kata	TF		TF Normalisasi		DF	IDF	TF-IDF	
	U1	U2	U1	U2			U1	U2
sesuai	1	0	1/6	0	1	$\log(2/1) = 0,301$	$,301 \times (1/6) = 0,050$	0
pesan	1	0	1/6	0	1	0,301	0,050	0
mulus	1	0	1/6	0	1	0,301	0,050	0
imei	1	1	1/6	1/7	2	$\log(2/2) = 0$	0	0
daftar	1	0	1/6	0	1	0,301	0,050	0

Sumber: Hasil Penelitian (2025)

3.4 **Modeling**

Sebelum membangun model *Machine Learning*, perlu diawali dengan analisis distribusi sentimen pada dataset ulasan iPhone *second* di Tokopedia yang telah melalui proses TF-IDF.



Sumber: Hasil Penelitian (2025)

Gambar 9. Distribusi Sentimen Data Ulasan iPhone *Second* di Tokopedia

Gambar 9 menunjukkan distribusi kelas yang tidak seimbang, di mana ulasan negatif jumlahnya lebih sedikit, sehingga dapat memengaruhi kemampuan model dalam mengenali kelas tersebut. Untuk mengatasi hal ini, digunakan teknik *oversampling* dengan metode SMOTE. Selanjutnya, model klasifikasi sentimen dibangun menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine (SVM)*. Kedua algoritma dilatih pada data yang sama dan dibandingkan performanya untuk menentukan model terbaik.

1). *Naïve Bayes*

Model klasifikasi sentimen menggunakan algoritma *Multinomial Naïve Bayes* berhasil mencapai akurasi sebesar 93% dalam mengelompokkan ulasan iPhone *second* di Tokopedia ke dalam kelas positif dan negatif. Hasil ini menunjukkan bahwa algoritma tersebut mampu mengenali pola sentimen dengan cukup baik, meskipun masih memiliki keterbatasan dalam menangani konteks yang lebih kompleks pada data teks.

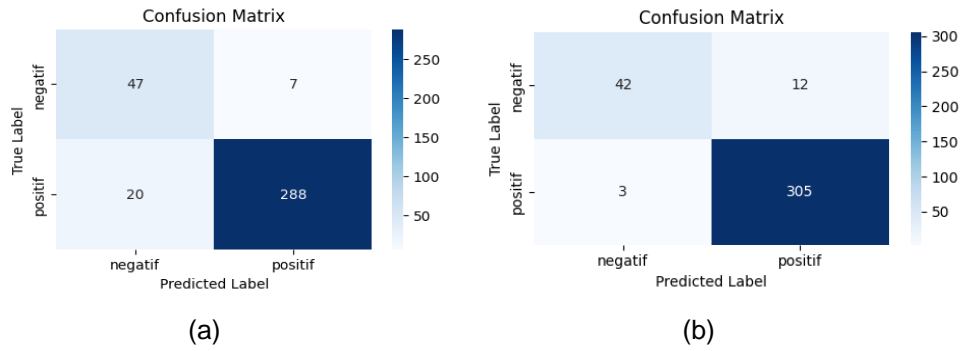
2). *Support Vector Machine (SVM)*

Model klasifikasi sentimen menggunakan algoritma *Support Vector Machine (SVM)* dengan threshold 0.2 menghasilkan akurasi sebesar 96%. Angka ini menunjukkan bahwa SVM mampu mengidentifikasi sentimen ulasan iPhone *second* di Tokopedia dengan tingkat ketepatan yang tinggi. Kinerja yang unggul ini mencerminkan kemampuan SVM dalam memisahkan kelas positif dan negatif secara lebih optimal dibandingkan model *Multinomial Naïve Bayes*, terutama dalam menangani dimensi fitur yang tinggi dari representasi teks.

3.5 *Evaluation*

Berdasarkan hasil evaluasi, algoritma *Naïve Bayes* dan *Support Vector Machine (SVM)* sama-sama menunjukkan performa yang cukup baik dalam klasifikasi sentimen ulasan iPhone *second* di Tokopedia. Meskipun demikian, masing-masing memiliki karakteristik dan keunggulan tersendiri yang memengaruhi pemilihan model sesuai dengan kebutuhan analisis.

Gambar 10 menunjukkan *confusion matrix* dari kedua model, yang merepresentasikan akurasi klasifikasi ulasan ke dalam sentimen positif dan negatif.



Sumber: Hasil Penelitian (2025)

Gambar 10 Confusion Matrix dengan model (a) Naïve Bayes, (b) Confusion Matrix dengan model Support Vector Machine (SVM)

Model *Naïve Bayes* menghasilkan 27 kesalahan klasifikasi, terdiri dari 19 ulasan positif yang keliru diklasifikasikan sebagai negatif dan 8 ulasan negatif yang diprediksi positif. Sementara itu, model SVM hanya mencatat 16 kesalahan, dengan 2 pada ulasan positif dan 14 pada ulasan negatif. Hasil ini menunjukkan bahwa SVM memiliki tingkat akurasi yang lebih tinggi dan lebih andal, khususnya dalam mengenali sentimen positif. Dengan jumlah kesalahan yang lebih rendah, SVM terbukti lebih stabil dalam menghadapi ketidakseimbangan data. Ringkasan evaluasi dari kedua algoritma disajikan pada Tabel 6.

Tabel 6. Perbandingan evaluasi kedua model yaitu Naïve Bayes dan SVM.

Algoritma	Accuracy	Precision	Recall	F1-Score
<i>Naïve Bayes</i>	0.93	0.84	0.90	0.87
<i>Support Vector Machine</i>	0.96	0.95	0.88	0.91

Sumber: Hasil Penelitian (2025)

Support Vector Machine (SVM) dipilih sebagai algoritma terbaik karena menunjukkan performa yang lebih unggul dalam mengklasifikasikan sentimen positif dan negatif pada ulasan iPhone *second* di Tokopedia. SVM mencapai akurasi sebesar 95%, lebih tinggi dibandingkan *Naïve Bayes* yang mencatatkan akurasi 92%. Keunggulan ini semakin terlihat pada hasil pengujian model, di mana SVM hanya menghasilkan 9 kesalahan dari total 362 ulasan, sehingga mencapai akurasi 97% (Gambar 11a). Sebaliknya, *Naïve Bayes* melakukan 37 kesalahan klasifikasi pada data yang sama, dengan akurasi sebesar 90% (Gambar 11b). Temuan ini menunjukkan bahwa model *Support Vector Machine* lebih konsisten dan akurat dalam memproses data ulasan yang tidak seimbang.

Evaluasi Model Support Vector Machine dengan Data Uji
 Akurasi : 0.9751
 Precision : 0.9748
 Recall : 0.9751
 F1-Score : 0.9748
 Prediksi Salah: 9 data

(a)

Evaluasi Model Naive Bayes dengan Data Uji
 Akurasi : 0.8978
 Precision : 0.9285
 Recall : 0.8978
 F1-Score : 0.9061
 Prediksi Salah: 37 data

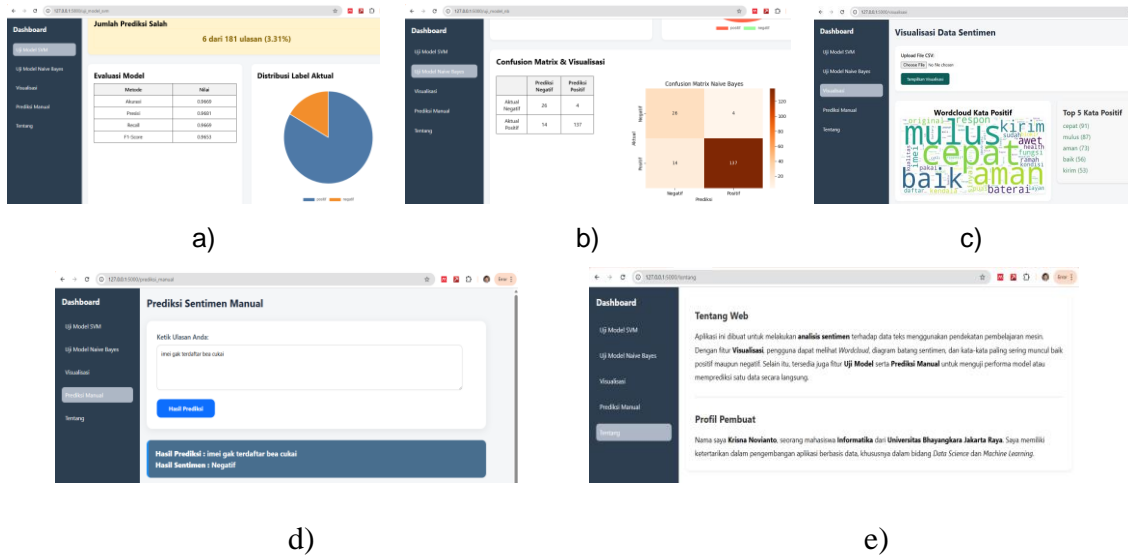
(b)

Sumber: Hasil Penelitian (2025)

Gambar 11. Hasil Evaluasi (a) Algoritma SVM dengan Data Validasi, (b) Evaluasi Algoritma Naïve Bayes dengan Data Validasi

3.6 Deployment

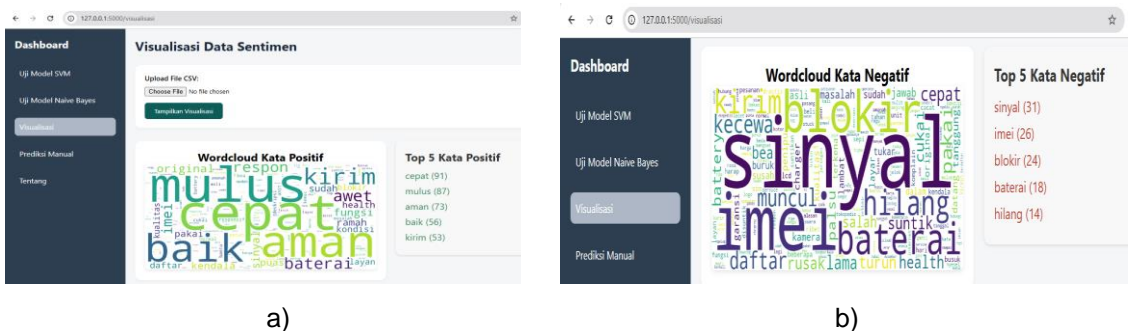
Pada tahap *deployment*, model *Support Vector Machine* diimplementasikan di dalam aplikasi *web* bernama analisis sentimen iPhone *second* di Tokopedia menggunakan *framework Flask*. File TF-IDF yang dihasilkan dari pemrosesan data latih ulasan produk iPhone *second* juga digunakan dalam sistem ini. *Web* ini menyediakan empat fitur yaitu: uji model SVM, uji model *Naïve Bayes*, visualisasi, prediksi manual, dan tentang *web*. Gambar 13 merupakan seluruh tampilan fitur dari *web*.



Sumber: Hasil Penelitian (2025)

Gambar 12. Tampilan *web* sentimen analisis produk iPhone *second* di Tokopedia pada (a) fitur uji model SVM, (b) Fitur uji model *Naïve Bayes*, (c) Fitur Visualisasi (d) Fitur prediksi manual, (e) Fitur tentang

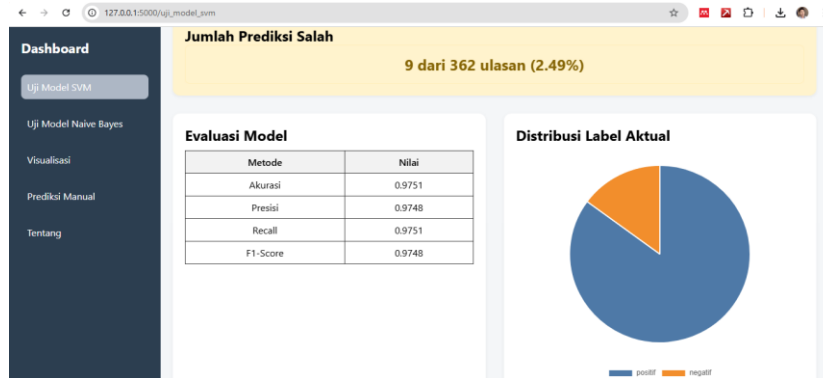
Pada fitur prediksi manual, pengguna mengetik langsung kalimat ulasan yang berkaitan dengan produk iPhone *second* di Tokopedia. Kalimat ulasan yang diketik pengguna adalah: *“imei gak terdaftar di bea cukai”*, menghasilkan bahwa kalimat tersebut memiliki sentimen “negatif”, dapat dilihat pada Gambar 13 (b). Sementara itu, dilakukan pengunggah data latih ke fitur visualisasi sebagaimana pada Gambar 14 (a) dan (b), menghasilkan tampilan *word cloud* dan lima daftar kata yang sering muncul pada masing-masing sentimen.



Sumber: Hasil Penelitian (2025)

Gambar 13. Hasil Visualisasi dari data latih (a) sentimen positif (b) sentimen Negatif

Pada fitur uji model SVM, dilakukan pengujian terhadap 362 ulasan. Hasilnya, model salah memprediksi 9 ulasan dan mencapai akurasi sebesar 97%. Capaian ini menunjukkan bahwa *Support Vector Machine* merupakan model dengan performa terbaik. Rincian hasil ditampilkan pada Gambar 15.



Sumber: Hasil Penelitian (2025)

Gambar 14. Hasil Prediksi Model SVM terhadap data latih

4. Kesimpulan

Penelitian ini menunjukkan bahwa calon pembeli iPhone *second* di Tokopedia berpotensi menghadapi berbagai risiko, seperti kondisi produk yang tidak sesuai deskripsi, kerusakan pada komponen penting, serta isu legalitas seperti IMEI yang tidak terdaftar. Informasi tersebut umumnya baru terungkap melalui ulasan pengguna. Analisis sentimen terhadap ulasan memberikan gambaran yang bermanfaat bagi konsumen. Ulasan positif cenderung menyoroti kondisi fisik yang baik dan pengiriman cepat, sedangkan ulasan negatif berkaitan dengan masalah teknis dan keaslian perangkat. Fitur *word cloud* dan prediksi otomatis membantu menyederhanakan informasi tersebut agar lebih mudah dipahami. Dari dua algoritma yang digunakan, *Support Vector Machine* menunjukkan performa paling unggul dengan akurasi 96%, dibandingkan *Naïve Bayes* yang mencapai 93%. Hasil ini menjadikan SVM sebagai algoritma yang direkomendasikan untuk klasifikasi sentimen dalam konteks ini.

Daftar Pustaka

- Amelia, E. E., & Yustiana, I. (2024). Analisis Sentimen Pada Ulasan Produk Pakaian UNIQLO Dengan Algoritma Naive Bayes. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 8(1), 668–674. <https://doi.org/10.36040/jati.v8i1.8697>
- Dania Siregar, Faroh Ladayya, Naufal Zhafran Albaqi, & Bintang Mahesa Wardana. (2023). Penerapan Metode Support Vector Machines (SVM) dan Metode Naïve Bayes Classifier (NBC) dalam Analisis Sentimen Publik terhadap Konsep Child-free di Media Sosial Twitter. *Jurnal Statistika dan Aplikasinya*, 7(1), 93–104. <https://doi.org/10.21009/jsa.07109>

- Hakim, I., & Harahap, L. S. (2024). Analisis Sentimen Terhadap Iphone 16 Pada Data Twitter Menggunakan Orange Data Mining. *Jurnal Multidisiplin Saintek*, 5(3). <https://ejournal.warunayama.org/kohesi%0AKohesi>:
- Handayanto, R. T., Herlawati, Atika, P. D., Khasanah, F. N., Yusuf, A. Y. P., & Septia, D. Y. (2021). Analisis Sentimen Pada Situs Google Review dengan Naïve Bayes dan Support Vector Machine. *Jurnal Komtika (Komputasi dan Informatika)*, 5(2), 153–163. <https://doi.org/10.31603/komtika.v5i2.6280>
- Henrys, K. (2021). *Importance of web scraping in e-commerce and e-marketing*. January, 1–10. <https://essay.utwente.nl/100867/>
- Herlawati, Srisulistiowati, D. B., Agustin, S. C., Syafina, P. H., Rachmatin, N., Setiawati, S., & *. (2024). *Metode Naïve Bayes dan Support Vector Machine untuk Mengolah Sentimen Ulasan dan Komentar di Platform Digital*. 5(2), 197–212.
- Hidayansyah, R. (2024). Pengaruh Ulasan Dan Penilaian Konsumen Terhadap Keputusan Pembelian Pada Marketplace Tokopedia Di Indonesia. *Digital Repository UNILA*. <http://digilib.unila.ac.id/id/eprint/79168>
- N'da, K., Ge, J., Ren, S. J., & Wang, J. (2023). *What matters for international consumers ' choice preferences for smartphones : Evidence from a cross-border ecommerce platform*. 1–20. <https://doi.org/10.1371/journal.pone.0285551>
- Nasikah, D., & Fuadi, S. (2022). Pengaruh Persepsi Keamanan, Kemudahan Bertransaksi, Kepercayaan Konsumen, Kualitas Produk Dan Harga Terhadap Keputusan Pembelian Pada Marketplace Tokopedia. *Jurnal Manajemen*, 16(1), 162–173.
- Oktavia, D., Ramadhan, Y. R., & Minarto. (2023). Analisis Sentimen Terhadap Penerapan Sistem E-Tilang Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (SVM). *Kajian Ilmiah Informatika dan Komputer*, 4(1), 407–417. <https://doi.org/10.30865/klik.v4i1.1040>
- Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang. (*Indonesian Journal on Computer and Information Technology*, 5(April), 75–82.
- Salsabilaa, N. A., Sa'adah, U., & Fauzi, F. (2024). Analisis Sentimen Pada Ulasan Aplikasi Tokopedia Menggunakan Klasifikasi Naïve Bayes. *PRISMA, Prosiding Seminar Nasional Matematika*, 44–51. <https://proceeding.unnes.ac.id/prisma> ISSN
- Widyastuti, A., Astuti, Y., Wisesa, B. A., & Hengki. (2024). Analisis Sentimen pada Ulasan Produk dengan SVM dan Word2Vec Sentiment Analysis on Product Reviews with SVM and Word2Vec. *JIKO (Jurnal Informatika dan Komputer)*, 8(1), 173–185. <https://doi.org/10.26798/jiko.v8i1.1498>