



**ANALISIS SENTIMEN BLIBLI MENGGUNAKAN METODE *SUPPORT
VECTOR MACHINE* UNTUK PREDIKSI KEPUASAN PELANGGAN**

TUGAS AKHIR



Oleh:

Dion Ardyansyah Putro Widodo

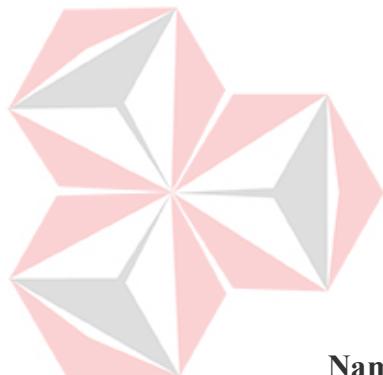
18410100112

**FAKULTAS TEKNOLOGI DAN INFORMATIKA
UNIVERSITAS DINAMIKA
2025**

**ANALISIS SENTIMEN BLIBLI MENGGUNAKAN METODE *SUPPORT
VECTOR MACHINE* UNTUK PREDIKSI KEPUASAN PELANGGAN**

TUGAS AKHIR

**Diajukan sebagai salah satu syarat untuk menyelesaikan
Program Sarjana**



**UNIVERSITAS
Dinamika**

Oleh:

Nama	: Dion Ardyansyah Putro Widodo
NIM	: 18410100112
Program Studi	: S1 Sistem Informasi

FAKULTAS TEKNOLOGI DAN INFORMATIKA

UNIVERSITAS DINAMIKA

2025

Tugas Akhir

ANALISIS SENTIMEN BLIBLI MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE* UNTUK PREDIKSI KEPUASAN PELANGGAN

Dipersiapkan dan disusun oleh
Dion Ardyansyah Putro Widodo
NIM: 18410100112

Telah diperiksa, dibahas dan disetujui oleh Dewan Pembahasan
Pada: Rabu , 13 Agustus 2025

Susunan Dewan Pembahasan

Pembimbing

- I. Arifin Puji Widodo, S.E., MSA
NIDN 0721026801
- II. I Gusti Ngurah Alit Widana Putra, S.T., M.Eng.
NIDN 0805058602

Digitally signed by
e124a5-9267-4b99-86a0-
a9fe0044e78
Date: 2025.08.14 10:49:41
+07'00'

Pembahasan

- Vivine Nurcahyawati, M.Kom
NIDN 0723018101

Digitally signed
by Vivine
Nurcahyawati
Date: 2025.08.14
12:12:09 +07'00'

Tugas Akhir ini telah diterima sebagai salah satu persyaratan

Untuk memperoleh gelar Sarjana:

Digitally signed by
Julianto
Date: 2025.08.18
23:16:09 +07'00'
Julianto Lemantara, S.Kom., M.Eng.



NIDN 0722108601

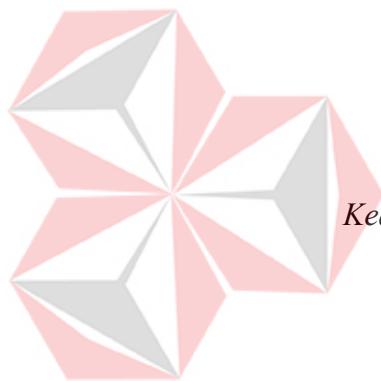
Dekan Fakultas Teknologi dan Informatika

UNIVERSITAS DINAMIKA



Pantang menyerah, karena setiap tantangan adalah peluang untuk berkembang

UNIVERSITAS
Dinamika



*Tugas Akhir ini saya persembahkan kepada
Kedua orang tua saya yang telah berjuang sampai saat ini*

UNIVERSITAS
Dinamika

SURAT PERNYATAAN
PERSETUJUAN PUBLIKASI DAN KEASLIAN KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama : Dion Ardyansyah Putro Widodo
NIM : 18410100112
Program Studi : S1 Sistem Informasi
Fakultas : Fakultas Teknologi dan Informatika
Jenis Karya : Tugas Akhir
Judul Karya : **ANALISIS SENTIMEN BLIBLI MENGGUNAKAN
METODE SUPPORT VECTOR MACHINE UNTUK
PREDIKSI KEPUASAN PELANGGAN**

Menyatakan dengan sesungguhnya bahwa:

1. Demi pengembangan Ilmu Pengetahuan, Teknologi, dan Seni, saya setuju memberikan kepada Universitas Dinamika Hak Bebas Royalti Non-Eksklusif (*Non-Exclusive Royalti Free Right*) atas seluruh isi/Sebagian karya ilmiah saya tersebut di atas untuk disimpan, dialihmediakan dan dikelola dalam bentuk pangkalan data (*database*) untuk selanjutnya didistribusikan atau dipublikasikan demi kepentingan akademis dengan tetap mencantumkan nama saya sebagai penulis atau pencipta sebagai pemilik Hak Cipta.
2. Karya tersebut di atas adalah karya asli saya, bukan plagiat baik Sebagian maupun keseluruhan, kutipan karya atau pendapat orang lain yang ada dalam karya ilmiah ini adalah semata hanya rujukan yang dicantumkan dalam Daftar Pustaka saya.
3. Apabila dikemudian hari ditemukan dan terbukti terhadap tindakan plagiat pada karya ilmiah ini, maka saya bersedia untuk menerima pencabutan terhadap gelar kesarjanaan yang telah diberikan kepada saya.

Demikian surat pernyataan ini saya buat dengan sebenarnya.

Surabaya, 13 Agustus 2025

Yang menyatakan



Dion Ardyansyah Putro Widodo

NIM: 18410100112

ABSTRAK

Blibli, sebagai salah satu e-commerce terkemuka di Indonesia, mengalami penurunan *Top Brand Index* pada 2024, sehingga diperlukan analisis sentimen untuk memahami persepsi pengguna. Penelitian ini bertujuan menganalisis sentimen ulasan pengguna terhadap aplikasi Blibli di Google Play Store menggunakan metode *Support Vector Machine* (SVM) dengan pelabelan *lexicon-based* dan pembobotan TF-IDF, guna mengetahui kecenderungan sentimen positif atau negatif serta mengidentifikasi pola dan tren opini yang dapat menjadi dasar perbaikan layanan dan strategi peningkatan kepuasan pelanggan. Berdasarkan 2000 ulasan, diperoleh 1.576 data setelah *preprocessing*, yang dibagi menjadi 70% pelatihan dan 30% pengujian. Pelabelan dilakukan dengan pendekatan *lexicon-based* dan vektorisasi TF-IDF. Hasil menunjukkan SVM mengklasifikasikan sentimen dengan distribusi 70,61% positif dan 29,39% negatif dari 473 data uji, dengan akurasi 86,68%. Performa pada ulasan positif unggul secara signifikan (*precision* 90%, *recall* 91%) dibandingkan ulasan negatif (*precision* 78%, *recall* 77%), menunjukkan model lebih efektif dalam mendeteksi sentimen positif. Rata-rata *F1-score* 84% mengonfirmasi konsistensi dan keandalan kinerja model. Visualisasi *wordcloud* memperlihatkan kata dominan pada tiap sentimen, mendukung interpretasi hasil analisis.

Kata Kunci: *Sentimen, Blibli, SVM, Google Play Store, Lexicon-Based*



UNIVERSITAS
Dinamika

KATA PENGANTAR

Puji syukur kehadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir ini dengan judul "Analisis Sentimen Blibli Menggunakan Metode *Support Vector Machine* Untuk Prediksi Kepuasan Pelanggan.". Tugas Akhir ini tidak akan terselesaikan tanpa dukungan dan bantuan dari berbagai pihak. Oleh karena itu, pada kesempatan ini, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Orang tua dan keluarga, yang selalu memberikan dukungan moral dan material selama masa studi hingga penyelesaian Tugas Akhir ini.
2. Bapak Arifin Puji Widodo, S.E., MSA selaku dosen pembimbing pertama, yang telah memberikan bimbingan, arahan, dan dukungan dalam penyusunan Tugas Akhir ini.
3. Bapak I Gusti Ngurah Alit Widana Putra, S.T., M.Eng. selaku dosen pembimbing kedua sekaligus dosen wali, atas kesabaran, masukan dan saran yang sangat berharga.
4. Ibu Vivine Nurcahyawati, M.Kom selaku dosen pembahas, atas segala masukan dan saran yang sangat berharga.
5. Semua pihak yang tidak dapat disebutkan satu per satu, namun telah memberikan kontribusi dalam penyelesaian Tugas Akhir ini.

Penulis menyadari bahwa Tugas Akhir ini masih jauh dari sempurna. Oleh karena itu, penulis terbuka untuk menerima kritik dan saran yang membangun demi perbaikan di masa yang akan datang. Semoga Tugas Akhir ini dapat memberikan manfaat bagi semua pihak.

Surabaya, 13 Agustus 2025

Dion Ardyansyah Putro Widodo

DAFTAR ISI

	Halaman
ABSTRAK	vii
KATA PENGANTAR.....	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL.....	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan.....	3
1.5 Manfaat	4
BAB II LANDASAN TEORI	5
2.1 Penelitian Terdahulu.....	5
2.2 <i>E-Commerce</i>	6
2.3 Analisis Sentimen.....	6
2.4 Google Play Store	7
2.5 <i>Text Mining</i>	7
2.6 <i>Text Preprocessing</i>	8
2.7 <i>Lexicon Based</i>	8
2.8 TF-IDF	9
2.9 <i>Support Vector Machine (SVM)</i>	9
2.10 <i>Confusion Matrix</i>	10
BAB III METODOLOGI PENELITIAN.....	12
3.1 Metode Penelitian.....	12
3.1.1 Tahap Awal	12
3.1.2 Analisis Data	15
3.1.3 Tahap Akhir	30
BAB IV HASIL DAN PEMBAHASAN.....	31



UNIVERSITAS
Dinamika

4.1 <i>Data Scraping</i>	31
4.2 <i>Preprocessing Data</i>	31
4.3 Pembagian Data	35
4.4 Pelabelan Data Latih dengan <i>Lexicon Based</i>	35
4.5 Klasifikasi <i>Support Vector Machine</i>	36
4.6 Validasi dan Evaluasi Data.....	37
4.7 Visualisasi Data.....	38
4.8 Evaluasi.....	40
BAB V PENUTUP	41
5.1 Kesimpulan	41
5.2 Saran.....	42
DAFTAR PUSTAKA	43
LAMPIRAN	45



UNIVERSITAS
Dinamika

DAFTAR GAMBAR

	Halaman
Gambar 3.1 <i>Flow Diagram</i> Metode Penelitian.....	12
Gambar 3.2 <i>Flow Diagram</i> Tahap Awal	12
Gambar 3.3 Contoh Ulasan Pelanggan	13
Gambar 3.4 <i>Flow Diagram</i> Analisis Data.....	15
Gambar 3.5 <i>Flow Diagram</i> Tahap Akhir	30
Gambar 4.1 Data ulasan hasil <i>Scraping</i>	31
Gambar 4.2 Proses hapus data duplikat	32
Gambar 4.3 Data hasil <i>cleaning</i>	32
Gambar 4.4 Hasil proses <i>Case Folding</i>	33
Gambar 4.5 Hasil proses normalisasi kata	33
Gambar 4.6 Hasil <i>Stopword Removal</i>	34
Gambar 4.7 Hasil <i>Stemming Data</i>	34
Gambar 4.8 Hasil proses <i>Tokenize</i>	35
Gambar 4.9 Hasil pelabelan data	36
Gambar 4.10 Prediksi Sentimen SVM	37
Gambar 4.11 Hasil validasi Model SVM	38
Gambar 4.12 Hasil <i>wordcloud</i> negatif	39
Gambar 4.13 Hasil <i>wordcloud</i> positif	39



DAFTAR TABEL

	Halaman
Tabel 1.1 Top Brand Index.....	2
Tabel 2.1 Penelitian Terdahulu.....	5
Tabel 2.2 <i>Confusion Matrix</i>	10
Tabel 3.1 Contoh data hasil <i>scraping</i>	17
Tabel 3.2 Contoh data hasil <i>Cleaning</i>	19
Tabel 3.3 Contoh data hasil <i>Case Folding</i>	20
Tabel 3.4 Contoh data hasil Normalisasi Kata.....	20
Tabel 3.5 Contoh data hasil <i>Stopword Removal</i>	21
Tabel 3.6 Contoh data hasil <i>Stemming</i>	22
Tabel 3.7 Contoh data hasil <i>Tokenize</i>	23



DAFTAR LAMPIRAN

	Halaman
L1. 1 Proses <i>data scraping</i>	45
L1. 2 Proses baca data.....	46
L1. 3 Proses <i>cleaning</i>	47
L1. 4 Proses <i>case folding</i>	48
L1. 5 Membuat fungsi penggantian kata baku.....	48
L1. 6 Proses normalisasi kata.....	49
L1. 7 Membuat fungsi <i>stopword removal</i>	49
L1. 8 Proses <i>Stopword removal</i>	49
L1. 9 Proses <i>Stemming data</i>	50
L1. 10 Proses <i>Tokenize</i>	50
L1. 11 Proses Pembagian Data.....	51
L1. 12 Proses Pelabelan Data Latih dengan <i>Lexicon</i>	52
L1. 13 Proses <i>TF-IDF</i>	53
L1. 14 Proses pelatihan dan prediksi dengan model yang telah dilatih	53
L1. 15 Proses <i>confusion matrix</i> dan <i>classification report</i>	54
L1. 16 Proses pembuatan <i>wordcloud</i>	54

BAB I

PENDAHULUAN

1.1 Latar Belakang

Google Play Store merupakan platform distribusi aplikasi yang populer di seluruh dunia, memungkinkan pengguna untuk mengunduh berbagai aplikasi, termasuk aplikasi *e-commerce* seperti Blibli. Melalui Google Play Store, pengguna dapat memberikan pendapat mereka berupa komentar dan penilaian untuk setiap aplikasi yang diunduh. Ulasan ini bervariasi berdasarkan pengalaman masing-masing pengguna dan dapat berisi opini, rekomendasi, atau saran terkait aplikasi tersebut.

Blibli, sebagai salah satu aplikasi *e-commerce* terkemuka di Indonesia, tidak jarang menerima ulasan dari pengguna mengenai layanan, produk, serta fitur yang ditawarkannya. Aspek-aspek yang sering disoroti dalam ulasan mencakup kemudahan penggunaan aplikasi, kualitas produk, dan layanan pelanggan. Ulasan-ulasan tersebut memberikan gambaran mengenai sejauh mana pengguna merasa puas atau tidak puas terhadap aplikasi Blibli..

Menurut (Ahn & Lee, 2024), Ulasan daring, yang bersumber dari pengalaman pengguna, dinilai lebih autentik dan meyakinkan dibanding iklan tradisional, serta berperan strategis dalam membangun kepercayaan dan memengaruhi perilaku beli konsumen. Berdasarkan pada tabel 1.1 *Top Brand Index*, dapat dilihat bahwa Blibli mengalami perubahan dari tahun 2022 ke 2024, di mana terdapat peningkatan dari 10,10% pada tahun 2022 menjadi 10,60% pada tahun 2023. Namun, pada tahun 2024 terjadi penurunan yang cukup tajam hingga mencapai 6,60%. Hal ini dapat mempengaruhi keputusan pelanggan dalam menggunakan Blibli. Oleh karena itu, dibutuhkan analisis sentimen lanjutan guna mendapatkan pemahaman yang lebih komprehensif mengenai tanggapan pelanggan terhadap pengalaman berbelanja online di platform ini.

Tabel 1.1 Top Brand Index

No.	Nama Brand	TBI 2022	TBI 2023	TBI 2024
1.	Shopee.co.id	43.70%	45.80%	50.50%
2.	Lazada.co.id	14.70%	15.10%	13.30%
3.	Tokopedia.com	14.90%	11.30%	10.40%
4.	Bukalapak.com	8.10%	4.70%	7.60%
5.	Blibli.com	10.10%	10.60%	6.60%

Sumber: <https://www.topbrand-award.com/>

Mengacu pada permasalahan yang telah dijelaskan, peneliti melakukan studi berjudul “Analisis Sentimen Blibli Menggunakan Metode *Support Vector Machine* untuk Prediksi Kepuasan Pelanggan” guna memperoleh wawasan yang lebih mendalam mengenai respons pengguna terhadap pengalaman berbelanja online di platform *e-commerce* Blibli. Platform *e-commerce* Blibli juga menyediakan aplikasi mobile di Google Play Store, yang dimanfaatkan oleh banyak pengguna untuk melakukan pembelian produk dan layanan. Seperti halnya Twitter, aplikasi Blibli di Google Play Store juga berfungsi sebagai wadah bagi pelanggan untuk menyampaikan komentar dan ulasan terkait layanan yang disediakan oleh Blibli. Dengan demikian, analisis opini terhadap data dari Google Play Store dapat menghasilkan wawasan yang lebih mendalam mengenai tingkat kepuasan pelanggan terhadap layanan yang disediakan oleh Blibli. Metode *Support Vector Machine* juga dapat diterapkan dalam analisis sentimen terhadap data dari Google Play Store untuk mengklasifikasikan sentimen dalam bentuk data teks.

Analisis sentimen atau *opinion mining* merupakan proses otomatis untuk memahami, mengekstraksi, dan mengolah data teks guna memperoleh informasi sentimen yang terdapat dalam kalimat opini. Selanjutnya, dilakukan klasifikasi polaritas teks dalam kalimat atau dokumen ke dalam dua kemungkinan kategori, yaitu positif dan negatif (Atmadja, 2022). Analisis sentimen dapat diterapkan menggunakan berbagai pendekatan, salah satunya yaitu dengan metode *Support Vector Machine*. Metode ini menggunakan teori probabilitas untuk memprediksi kategori atau label yang tepat untuk setiap dokumen atau teks berdasarkan kemunculan kata atau fitur dalam dokumen.

Penelitian ini bertujuan menganalisis sentimen ulasan pengguna terhadap aplikasi Blibli di Google Play Store menggunakan metode *Support Vector Machine* (SVM) dengan pelabelan *lexicon-based* dan pembobotan TF-IDF, guna mengetahui

kecenderungan sentimen positif atau negatif serta mengidentifikasi pola dan tren opini yang dapat menjadi dasar perbaikan layanan dan strategi peningkatan kepuasan pelanggan.

1.2 Rumusan Masalah

Mengacu pada latar belakang yang telah dipaparkan, dapat dirumuskan permasalahan utama yaitu bagaimana menerapkan analisis sentimen terhadap ulasan pelanggan Blibli yang berasal dari data Google Play Store menggunakan algoritma *Support Vector Machine*, sehingga dapat diperoleh persentase ulasan yang termasuk dalam kategori sentimen positif dan negatif.

1.3 Batasan Masalah

Mengacu pada rumusan masalah yang telah disusun sebelumnya, maka ruang lingkup permasalahan dalam penelitian ini dibatasi sebagai berikut.:

1. Analisis sentimen hanya dilakukan pada data Google Play Store Blibli yang berbahasa Indonesia.
2. Analisis sentimen meliputi pengelompokan negatif dan positif.
3. Metode yang diterapkan untuk melakukan analisis sentimen adalah algoritma *Support Vector Machine* (SVM).
4. Data yang dianalisis hanya mencakup ulasan dan penilaian pengguna terhadap aplikasi Blibli di Google Play Store selama periode tahun 2024.

1.4 Tujuan

Tujuan penelitian ini adalah menganalisis sentimen ulasan pengguna terhadap aplikasi Blibli di Google Play Store menggunakan metode *Support Vector Machine* (SVM) dengan pelabelan *lexicon-based* dan vektorisasi TF-IDF, guna mengetahui kecenderungan sentimen positif atau negatif serta mengidentifikasi pola dan tren opini yang dapat menjadi dasar perbaikan layanan dan strategi peningkatan kepuasan pelanggan.

1.5 Manfaat

Mengacu pada tujuan yang telah dipaparkan sebelumnya, maka manfaat dari penelitian ini adalah sebagai berikut:

1. Meningkatkan pemahaman tentang sentimen pelanggan terhadap layanan Blibli berdasarkan ulasan dan rating pada aplikasi di Google Play Store.
2. Memberikan informasi yang berguna bagi Blibli dalam pengambilan keputusan bisnis yang lebih baik, terutama dalam hal pengembangan dan peningkatan layanan.
3. Membuktikan tingkat efektivitas dan akurasi metode Support Vector Machine dalam menganalisis sentimen pada data ulasan aplikasi Blibli di Google Play Store.
4. Memberikan acuan bagi peneliti berikutnya dalam melakukan studi serupa serta turut memperluas wawasan dalam kajian analisis sentimen pada layanan e-commerce.



UNIVERSITAS
Dinamika

BAB II

LANDASAN TEORI

1.1 Penelitian Terdahulu

Penelitian ini merujuk pada sejumlah studi sebelumnya sebagai landasan untuk memperkuat teori yang digunakan dalam analisis penelitian yang dilakukan. Rangkuman penelitian terdahulu disajikan pada Tabel 2.1 berikut.

Tabel 2.1 Penelitian Terdahulu

Judul	Penulis	Hasil Penelitian	Akurasi
Penerapan Metode <i>Support Vector Machine</i> Untuk Analisis Sentimen Pengguna Twitter	Zidna Alhaq, Ali Mustopa, Sri Mulyatun, Joko Dwi Santoso	Berdasarkan 300 data tweet klasifikasi sentimen menggunakan SVM dengan TF-IDF dan <i>Cosine Similarity</i> mencapai akurasi sebesar 93% dan Evaluasi SVM dengan <i>K-Fold Cross Validation</i> mencapai akurasi sebesar 93% dan rata-rata sebesar 83,5%	SVM dengan TF-IDF dan <i>Cosine Similarity</i> 93% SVM dengan <i>K-Fold Cross Validation</i> 93% dan rata-rata sebesar 83,5%
Penerapan Algoritma SVM Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia	Dedi Darwis, Eka Shintya Pratiwi, A. Ferico Octaviansyah Pasaribu	Berdasarkan hasil data crawling sebanyak 2000 data dan dilakukan <i>Preprocessing</i> sehingga tersisa 1890 data yang dapat diolah. Klasifikasi sentimen menggunakan SVM dengan TF-IDF mencapai akurasi sebesar 82%	Akurasi 82% <i>precision</i> 90%, <i>recall</i> 88% <i>f1-score</i> 89%.
Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma <i>Support Vector Machine</i> (SVM)	Irma Kumala Yasin Mustofa, Abraham Salihi	Berdasarkan 3000 data ulasan klasifikasi sentimen menggunakan algoritma <i>Support Vector Machine</i> menghasilkan akurasi sebesar 98% dan <i>f1-score</i> sebesar 0.98 atau sebesar 98%	Akurasi 98% <i>f1-score</i> 98%

1.2 *E-Commerce*

E-commerce, atau perdagangan elektronik, merupakan aktivitas transaksi bisnis yang dilakukan secara daring melalui jaringan internet. Melalui platform ini, konsumen dapat membeli produk maupun layanan secara langsung melalui situs web atau aplikasi yang disediakan oleh penyedia *e-commerce*.

Menurut (Wulandary et al., 2023), *E-commerce* menawarkan sejumlah keunggulan yang tidak dimiliki oleh transaksi konvensional. Proses pembelian dapat dilakukan dengan cepat dan praktis, tanpa mengharuskan pelanggan mendatangi toko secara langsung. Dari sisi pembayaran, konsumen diberi keleluasaan untuk memilih metode yang sesuai, seperti transfer bank, kartu kredit, hingga *cash on delivery* (COD) yang memudahkan proses transaksi. Selain itu, platform *e-commerce* juga memungkinkan pembeli untuk mengakses informasi produk secara lengkap serta memberikan umpan balik atau ulasan langsung kepada penjual maupun penyedia layanan. Ragam produk yang ditawarkan pun sangat bervariasi, sehingga konsumen dapat membandingkan harga dan kualitas dari berbagai vendor sebelum memutuskan pembelian. Keunggulan lain yang ditawarkan adalah efisiensi waktu, karena seluruh proses mulai dari pencarian produk, pembayaran, hingga pengiriman dapat dilakukan secara daring. Tidak hanya itu, layanan komunikasi pelanggan yang tersedia, seperti fitur *live chat*, pusat layanan (*call center*), maupun pesan langsung, turut mempermudah interaksi antara penjual dan pembeli.

1.3 Analisis Sentimen

Analisis sentimen atau *opinion mining* merupakan studi yang berfokus pada bagaimana mengidentifikasi dan menyelesaikan permasalahan yang berkaitan dengan opini, sikap, dan emosi masyarakat terhadap suatu entitas yang dapat merepresentasikan individu. (Utami & Artana, 2022).

Analisis sentimen bertujuan untuk mengelompokkan opini atau komentar yang berasal dari media sosial ke dalam kategori positif, negatif, atau netral, sehingga dapat memberikan gambaran umum terhadap persepsi publik. Informasi ini bermanfaat sebagai bahan evaluasi dan pengambilan keputusan, misalnya dalam

perbaikan kualitas layanan maupun pengembangan produk.(Aryanti & Mahendra, 2023).

1.4 Google Play Store

Google Play Store adalah platform utama untuk distribusi aplikasi pada perangkat Android yang menawarkan ribuan pilihan aplikasi yang dapat diunduh oleh pengguna. Kepopulerannya terus bertumbuh sejalan dengan meningkatnya jumlah pengguna smartphone, menjadikannya sebagai opsi favorit dibandingkan layanan distribusi aplikasi lainnya. Data tahun 2022 menunjukkan bahwa sekitar 96% aplikasi di Google Play Store tersedia secara gratis, sehingga semakin menarik minat masyarakat. Melalui platform ini, pengguna memiliki akses untuk mengunduh aplikasi sekaligus meninjau ulasan, peringkat, dan jumlah unduhan sebagai acuan sebelum memutuskan untuk memasang aplikasi tersebut. (Rahma, Intan Larasati Anisa, Shafira Nurin Qolbi, 2022).

Dalam analisis sentimen pada data dari Google Play Store, kita dapat memanfaatkan informasi yang tersedia pada setiap ulasan pengguna terhadap aplikasi atau konten digital yang diunduh. Ulasan pengguna pada Google Play Store biasanya terdiri dari rating bintang dan komentar atau ulasan tertulis. Rating bintang menunjukkan seberapa baik pengguna menilai aplikasi atau konten digital tersebut, sedangkan komentar atau ulasan tertulis memberikan informasi lebih rinci mengenai kelebihan, kekurangan, serta pengalaman pengguna dalam menggunakan aplikasi atau konten digital tersebut.

1.5 Text Mining

Menurut Firdaus & Firdaus, (2021), *Text Mining* adalah proses menemukan pengetahuan dalam bentuk teks dari sebuah database, yang juga dikenal sebagai aktivitas pencarian atau penggalian informasi dari data berbasis teks. Proses ini mencakup pencarian pengetahuan baru yang sebelumnya belum diketahui, sehingga dapat dipahami dan memiliki nilai praktis. *Text Mining* bertujuan mengungkap pola atau informasi berharga dari kumpulan data teks atau korpus besar yang bersifat tidak terstruktur, yaitu koleksi teks yang merepresentasikan penggunaan bahasa secara lengkap dalam bentuk tulisan.

1.6 *Text Preprocessing*

Pemrosesan teks adalah proses untuk mengekstraksi, mengelola, dan menyusun informasi dengan menganalisis keterkaitan serta pola-pola yang terdapat dalam data teks yang bersifat semi-terstruktur maupun tidak terstruktur. (Pudjiantoro et al., 2021).

Preprocessing dilakukan dengan tujuan mengubah data mentah, yang diperoleh dari proses pengambilan data, menjadi format yang lebih terstruktur dan efisien untuk mendukung analisis sentimen. (Saraswati & Riminarsih, 2020).

Berikut ini adalah tahapan yang ada pada *preprocessing* :

1. *Cleaning* adalah tahap pembersihan teks dengan cara menghapus tanda baca, angka, simbol, tautan URL, serta nama pengguna yang terdapat dalam data teks. (Khairunnisa & Faraby, 2021).
2. *Case Folding* adalah proses mengonversi seluruh huruf kapital pada ulasan dalam data latih dan data uji menjadi huruf kecil. (Pudjiantoro et al., 2021).
3. *Tokenize* merupakan proses memecah teks menjadi kata-kata individual dengan menggunakan spasi dan tanda baca sebagai pemisah. (Arsi & Waluyo, 2021).
4. Normalisasi Kata adalah proses mengubah kata tidak baku menjadi bentuk baku serta mengganti akronim atau singkatan dengan kata lengkap yang sesuai. (Ditami et al., 2022).
5. *Stopword Removal* adalah proses menghapus kata-kata yang tidak relevan dengan topik dalam dokumen, karena keberadaan kata-kata tersebut tidak berpengaruh signifikan terhadap akurasi dalam proses klasifikasi sentimen. (Pudjiantoro et al., 2021).
6. *Stemming* adalah proses mengevaluasi kata-kata yang telah dikonversi ke dalam huruf kecil, dengan tujuan menyeragamkan bentuk kata sehingga jumlah variasi kata dalam data latih dapat dikurangi.

1.7 *Lexicon Based*

Lexicon Based merupakan metode analisis sentimen yang mengandalkan pendekatan berbasis kamus. Dalam metode ini, kamus *lexicon* menetapkan bobot pada kata-kata berdasarkan daftar yang telah diklasifikasikan sebelumnya, sehingga memungkinkan proses pengelompokan kalimat secara otomatis sesuai dengan

sentimennya. Oleh karena itu, pendekatan ini sering digunakan dalam proses pelabelan data latih sebelum model dilatih lebih lanjut. (Machová et al., 2020). Salah satu sumber *lexicon* dalam Bahasa Indonesia yang banyak digunakan adalah InSet (*Indonesian Sentiment Lexicon*), yang dapat diakses melalui repositori GitHub: <https://github.com/fajri91/InSet> (Fajri Koto, 2016).

1.8 TF-IDF

TF-IDF (*Term Frequency–Inverse Document Frequency*) adalah metode representasi teks ke dalam bentuk numerik yang memadukan informasi tentang frekuensi kemunculan kata dalam suatu dokumen (*term frequency*) dengan tingkat kelangkaannya di seluruh kumpulan dokumen (*inverse document frequency*) (Fahlevvi, 2022). Proses ini bertujuan untuk mengubah ulasan yang awalnya berupa kalimat menjadi vektor numerik yang dapat diproses oleh algoritma klasifikasi. Melalui TF-IDF, setiap kata direpresentasikan dalam bentuk nilai numerik yang mencerminkan tingkat relevansinya terhadap dokumen, bukan berdasarkan kamus atau daftar kata yang telah ditetapkan sebelumnya. Representasi vektor ini kemudian menjadi masukan bagi model klasifikasi untuk mempelajari pola dan keterkaitan antar kata dalam menentukan kategori sentimen.

1.9 *Support Vector Machine* (SVM)

Support Vector Machine (SVM) merupakan salah satu metode dalam *machine learning* yang memanfaatkan hipotesis berupa fungsi linear pada ruang fitur berdimensi tinggi (Isnain et al., 2021). SVM dilatih menggunakan algoritma pembelajaran yang berlandaskan pada teori optimasi. Metode ini pertama kali diperkenalkan oleh Vapnik pada tahun 1992 dan dikenal sebagai sistem pembelajaran yang mampu melakukan generalisasi dengan baik dalam proses klasifikasi pola data. SVM menerapkan konsep hyperplane sebagai alat pemisah antar kelas data. Akurasi model yang dihasilkan sangat bergantung pada pemilihan fungsi kernel serta parameter yang digunakan selama proses pelatihan. SVM banyak dimanfaatkan di berbagai bidang, termasuk klasifikasi teks, analisis sentimen, dan pengenalan pola.

SVM memiliki keunggulan dalam memetakan data input non-linear ke ruang berdimensi tinggi, sehingga memungkinkan pemisahan data secara efektif berkat penggunaan fungsi kernel yang dimilikinya (Irfani et al., 2020). Kernel linear adalah salah satu jenis kernel yang digunakan dalam algoritma SVM. Fungsinya untuk menghitung tingkat kesamaan antara dua vektor fitur melalui operasi *dot product* (perkalian dalam). Kernel ini ideal untuk digunakan pada data yang dapat dipisahkan secara linier di ruang fitur, namun kurang efektif jika diterapkan pada data yang memiliki pola non-linier. Adapun rumus fungsi kernel linear adalah sebagai berikut:

$$K(x_i, x_j) = x_i \cdot x_j$$

Di sini, x_i dan x_j adalah dua vektor fitur yang mewakili dua sampel data, dan · merupakan operasi dot product atau perkalian titik antara kedua vektor.

1.10 Confusion Matrix

Confusion Matrix digunakan untuk melakukan proses evaluasi performa sistem yang menguji hasil akhir klasifikasi dengan mengukur nilai performa dari suatu sistem yang telah dibuat atau dikembangkan (Sulaiman et al., 2021). Tabel 2.2 menyajikan *confusion matrix* yang terdiri dari empat kombinasi antara nilai prediksi dan nilai aktual yang berbeda.

Tabel 2.2 *Confusion Matrix*

Classification	Predicted Positive	Predicted Negative
Actual Positive Case	True Positive (TP)	False Negative (FN)
Actual Negative Case	False Positive (FP)	True Negative (TN)

Untuk menghitung akurasi prediksi kelas *positive* dan *negative* menggunakan rumus persamaan sebagai berikut:

$$Akurasi = \frac{TN+TP}{FN+FP+TN+TP}.$$

Untuk menghitung *Recall* pada prediksi kelas positif atau negatif digunakan persamaan:

$$Recall = \frac{TP}{TP + FN}$$

Recall menunjukkan proporsi data positif yang berhasil terdeteksi dengan benar oleh model dari seluruh data positif yang sebenarnya.

Untuk menghitung *Precision* pada prediksi kelas positif atau negatif digunakan persamaan:

$$Precision = \frac{TP}{TP + FP}$$

Precision menunjukkan proporsi prediksi positif yang benar-benar positif dibandingkan dengan seluruh prediksi positif yang dihasilkan model.

Untuk menghitung *F1-Score* digunakan persamaan:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

F1-Score merupakan rata-rata harmonis antara *Precision* dan *Recall* yang berguna untuk menilai keseimbangan performa model pada kedua metrik tersebut.

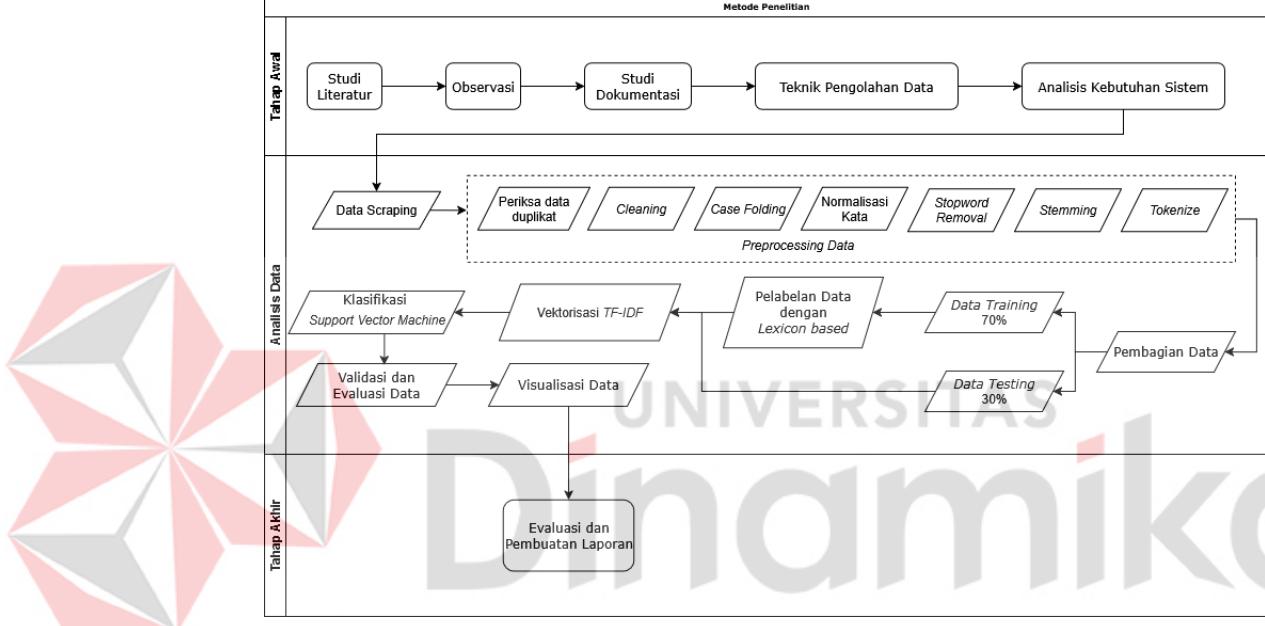


BAB III

METODOLOGI PENELITIAN

3.1 Metode Penelitian

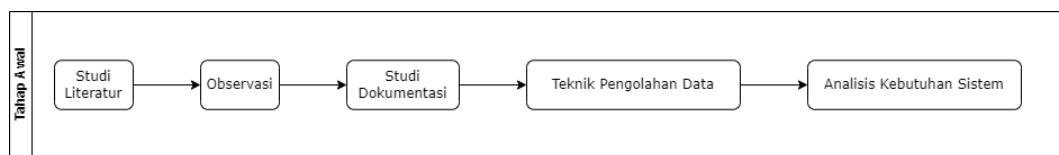
Pada bagian ini dijelaskan mengenai langkah-langkah yang dilakukan dalam menyelesaikan penelitian. Penelitian ini terdiri dari tiga tahapan utama, yaitu Tahap Awal, Analisis Data, dan Tahap Akhir, yang ditampilkan pada Gambar 3.1.



Gambar 3.1 *Flow Diagram* Metode Penelitian

3.1.1 Tahap Awal

Tahap awal merupakan langkah yang dilakukan sebelum memasuki proses analisis data dalam penelitian ini. Alur proses pada tahap awal ditampilkan pada gambar 3.2.



Gambar 3.2 *Flow Diagram* Tahap Awal

A. Studi Literatur

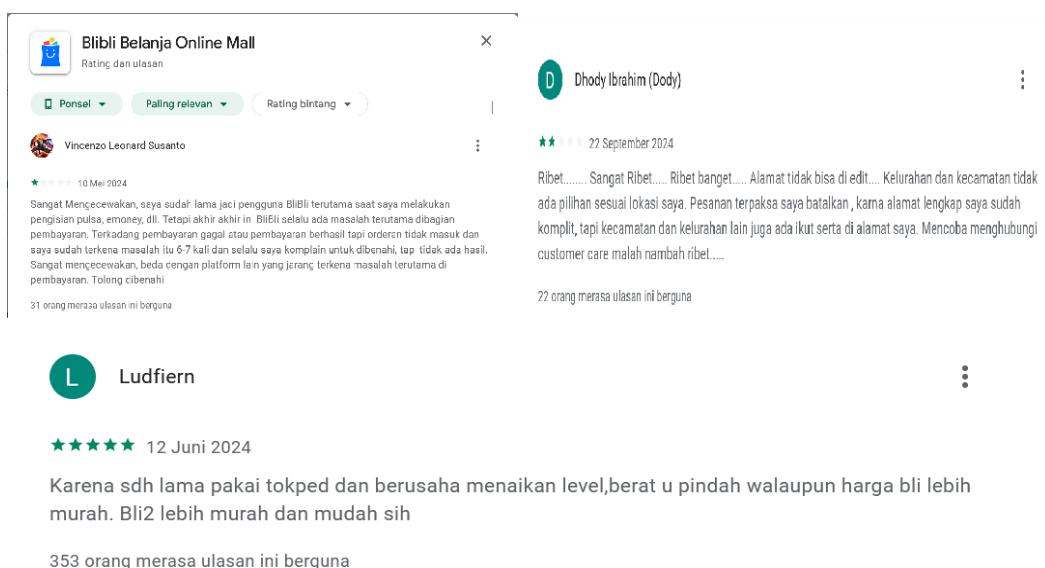
Studi literatur merupakan tahap pertama dari metode penelitian pada penelitian kali ini. Pada tahap ini akan dilakukan dengan membaca beberapa referensi terkait analisis sentimen, metode *Support Vector Machine*, dan aplikasinya pada data dari Google Play Store mengenai aplikasi Blibli. Referensi yang digunakan terdiri dari jurnal ilmiah, dan artikel.

B. Observasi

Observasi dilakukan dengan mengamati perilaku pengguna di platform Google Play Store mengenai aplikasi Blibli, khususnya pada penulisan ulasan produk dan tingkat kepuasan pelanggan yang tertera di dalamnya. Observasi dilakukan secara langsung dengan memantau ulasan produk di platform Google Play Store.

C. Studi Dokumentasi

Studi dokumentasi dilakukan dengan cara mengumpulkan data dari platform Google Play Store, khususnya berupa ulasan produk. Ulasan tersebut kemudian diolah dan dianalisis guna memperoleh informasi mengenai sentimen pelanggan. Gambar 3.3 menunjukkan contoh ulasan terkait aplikasi Blibli yang terdapat di Google Play Store.



Gambar 3.3 Contoh Ulasan Pelanggan

D. Teknik Pengolahan Data

Teknik Pengolahan Data dalam penelitian ini merupakan tahapan untuk mengubah data mentah menjadi data yang siap dianalisis. Proses ini diawali dengan pengumpulan data dari Google Play Store menggunakan *library google-play-scrapers*, yang menghasilkan 2000 entri berisi nama pengguna, rating, dan ulasan berbahasa Indonesia. Selanjutnya dilakukan pembersihan data dengan menghapus entri yang duplikat, kosong, maupun tidak relevan serta menghilangkan karakter yang tidak diperlukan.

Tahapan berikutnya adalah *preprocessing data*, yang meliputi *case folding*, *tokenizing*, *stopword removal*, dan *stemming* agar ulasan menjadi lebih terstruktur. Selanjutnya, data dibagi menjadi dua bagian, yaitu data latih dan data uji. Data latih kemudian diberi label sentimen positif atau negatif menggunakan pendekatan *lexicon-based* dengan memanfaatkan kamus sentimen InSet. Setelah pelabelan, teks diubah menjadi representasi numerik melalui metode TF-IDF sehingga dapat diolah oleh algoritma klasifikasi.

E. Analisis Kebutuhan Sistem

Analisis kebutuhan sistem dilakukan untuk menentukan kebutuhan yang diperlukan dalam penelitian, meliputi perangkat lunak, bahasa pemrograman *Python* beserta *library* yang digunakan, serta spesifikasi minimum perangkat keras yang mendukung proses penelitian.

1. Kebutuhan Perangkat Lunak (*Software*)

Kebutuhan perangkat lunak mencakup aplikasi atau program yang digunakan selama proses penelitian. Adapun perangkat lunak yang digunakan dalam penelitian ini adalah sebagai berikut:

- a. *Google Play Scraper* berfungsi untuk melakukan pengambilan data (*scraping*) dari Google Play Store.
- b. *Google Colaboratory* digunakan sebagai platform dalam menjalankan proses mulai dari scraping hingga visualisasi data.
- c. *Library pandas* dimanfaatkan untuk mengimpor dan mengekspor data dalam format .csv..

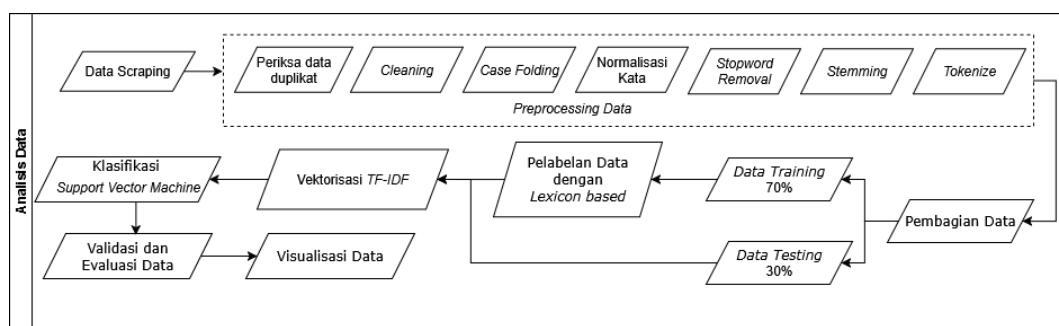
- d. Library *sastrawi* digunakan dalam tahap stemming pada proses pra-pemrosesan teks.
 - e. *Library nltk* berperan dalam proses tokenisasi dan penghapusan kata-kata umum (stopwords).
 - f. *Library sklearn* digunakan untuk melakukan klasifikasi dengan algoritma *Support Vector Machine* serta membuat confusion matrix.
 - g. *Library matplotlib* digunakan dalam pembuatan visualisasi data seperti *pie chart*.
 - h. *Library wordcloud* digunakan untuk menampilkan visualisasi data dalam bentuk *word cloud*.

2. Kebutuhan Perangkat Keras (*Hardware*)

Kebutuhan perangkat keras merupakan peralatan fisik yang mendukung dalam proses penggunaan perangkat lunak dalam penelitian ini. Perangkat keras yang digunakan dalam penelitian ini adalah sebagai berikut :

- a. Komputer yang digunakan dalam penelitian ini dilengkapi dengan prosesor Intel(R) Pentium(R) Silver N6000.
 - b. Komputer yang digunakan dalam penelitian ini memiliki kapasitas RAM (*Random Access Memory*) sebesar 4GB.

3.1.2 Analisis Data



Gambar 3.4 *Flow Diagram* Analisis Data

Tahap analisis data bertujuan untuk mengenali serta memahami kebutuhan yang diperlukan dalam proses analisis, yang selanjutnya akan dilanjutkan dengan pemrosesan data tersebut..

A. Data Scraping

Pada tahap ini, proses pengambilan data dilakukan menggunakan bahasa pemrograman Python dengan bantuan *library* Google Play scraper. ID aplikasi blibli dari Google Play Store yang sudah diketahui dimasukkan ke *code* untuk mendapatkan data-data dari Google Play Store, kemudian melakukan *import library* pandas. *Library* Pandas digunakan untuk mengubah data *scraping* Google Play Store menjadi format *Comma-separated values* atau (.csv). Berikut ini pseudocode dan contoh data hasil yang berhasil di-*scraping*.

Pseudocode untuk Mengambil Data Ulasan dari Google Play Store

1. Install Library google-play-scraper

- Gunakan perintah `pip install google-play-scraper` untuk menginstal library yang dibutuhkan.

2. Import Library yang Diperlukan

- Import `reviews` dan `Sort` dari library `google_play_scraper` untuk mengambil data ulasan aplikasi dari Google Play Store.
- Import `pandas` untuk pengolahan data, jika diperlukan pada langkah selanjutnya.
- Import `datetime` untuk bekerja dengan tanggal jika diperlukan dalam filtering waktu (tidak digunakan dalam kode ini, namun diimpor untuk kemungkinan penggunaan).

3. Tentukan ID Aplikasi

- Tentukan variabel `app_id` dengan ID aplikasi yang akan diambil datanya (misalnya, "blibli.mobile.commerce").

4. Definisikan Fungsi `get_reviews`

- Buat fungsi `get_reviews` yang menerima parameter berikut:
 - `app_id`: ID aplikasi
 - `lang`: Bahasa yang diinginkan (default "id" untuk Bahasa Indonesia)
 - `count`: Jumlah ulasan yang ingin diambil (default 2000)
 - `sort`: Metode pengurutan ulasan (default `Sort.NEWEST`)
 - `filter_score_with`: Filter ulasan berdasarkan skor tertentu (default 'None')

- `filter_device_with`: Filter ulasan berdasarkan jenis perangkat (default 'None')
- `continuation_token`: Token untuk melanjutkan pengambilan ulasan berikutnya (default 'None')
- Di dalam fungsi:
 - Gunakan fungsi `reviews` dari library `google_play_scraper` untuk mengambil data ulasan berdasarkan parameter yang diberikan.
 - Return hasil pengambilan data `result` dan `continuation_token`.
 - Jika terjadi kesalahan, tangkap exception dan cetak pesan error, lalu kembalikan 'None, None'.

5. Panggil Fungsi `get_reviews`

- Panggil fungsi `get_reviews` dengan `app_id` yang sudah ditentukan.
- Simpan hasil ke dalam variabel `reviews` dan `continuation_token`.

6. Cek Hasil Pengambilan Data

- Jika `reviews` tidak kosong:
 - Cetak jumlah ulasan yang berhasil diambil.
 - Jika terdapat ulasan, cetak satu contoh ulasan untuk pengecekan.
- Jika `reviews` kosong, cetak pesan "Tidak dapat mengambil ulasan."

Tabel 3.1 Contoh data hasil *scraping*

Username	Rating	Review Text
Mustafa mus	5	sangat puas
Wiji Utami	5	Like
Ezra Putra	5	Okee
Heru budi utomo Utomo	2	Blibli, kl kalian tdk ada perubahan, 2thn lg nasibmu akan seperti elevania.
Cak Bogang8383	5	Oke

B. *Preprocessing Data*

Preprocessing data adalah rangkaian langkah untuk membersihkan dan menyiapkan data mentah agar layak digunakan dalam analisis dan pemodelan, sehingga kualitas dan akurasi hasil meningkat. Tahapan ini mencakup pemeriksaan duplikat, pembersihan karakter, penyeragaman huruf, normalisasi kata, penghapusan *stopword*, *stemming*, dan tokenisasi, yang dilakukan secara berurutan agar data siap untuk pembobotan dan klasifikasi.

1. Periksa data duplikat

Pada tahap ini bertujuan mendeteksi dan menghapus data yang sama atau berulang, karena duplikasi dapat menimbulkan bias pada pelatihan model dan menurunkan kemampuan generalisasinya. Berikut ini pseudocode untuk proses periksa data duplikat.

Preprocessing Data Ulasan

- Import library yang diperlukan:
 - Import 'pandas' untuk memanipulasi data.
 - Import 'matplotlib.pyplot' untuk visualisasi data.
 - Import 'Counter' dari 'collections' untuk menghitung frekuensi kata.
 - Import 're' dan 'string' untuk penghapusan simbol, angka, emoji, dan karakter khusus.
 - Import 'nltk' untuk stopword removal dan stemming.
 - Import 'Sastrawi' untuk stemming bahasa Indonesia.
- Baca data ulasan yang sudah di-scraping:
 - Gunakan 'pandas' untuk membaca file CSV ('hasil_scraping_ulasan_app_blibli.csv').
 - Tampilkan informasi umum data dan beberapa baris pertama.
- Buat DataFrame dengan kolom yang relevan:
 - Pilih kolom 'Date', 'Username', 'Rating', dan 'Review Text'.
 - Tampilkan beberapa baris pertama.
- Hapus Data Duplikat:
 - Hapus data yang memiliki nilai duplikat pada kolom 'Review Text'.
 - Tampilkan informasi data setelah penghapusan duplikat.

2. *Cleaning*

Pada tahap *cleaning*, dilakukan proses penghapusan tanda baca dan karakter khusus dari teks guna mempermudah analisis dalam pemrosesan bahasa alami (*Natural Language Processing/NLP*). Langkah ini bertujuan untuk menghapus elemen non-alfabet yang tidak memberikan informasi penting bagi analisis, seperti tanda seru (!), tanda tanya (?), maupun simbol-simbol khusus seperti (@, #). Proses penghapusan dilakukan menggunakan *library re* di Google Colab dengan bahasa

pemrograman Python. Berikut ini pseudocode dan contoh data hasil yang telah melewati tahapan *cleaning*.

Proses Cleaning:

- Definisikan fungsi-fungsi pembersihan data:
 - `remove_URL`: Menghapus URL dalam teks.
 - `remove_html`: Menghapus tag HTML.
 - `remove_symbols`: Menghapus simbol.
 - `remove_numbers`: Menghapus angka.
 - `remove_emoji`: Menghapus emoji.
 - `hapus_katadouble`: Menghapus karakter berulang.
- Terapkan fungsi-fungsi tersebut pada kolom 'Review Text' dan simpan hasilnya ke kolom baru 'cleaning'.

Tabel 3.2 Contoh data hasil *Cleaning*

Username	Rating	Review Text	Cleaning
Mustafa mus	5	sangat puas	sangat puas
Wiji Utami	5	Like	Like
Ezra Putra	5	Okee	Okee
Heru budi utomo Utomo	2	Blibli, kl kalian tdk ada perubahan, 2thn lg nasibmu akan seperti elevania.	Blibli kl kalian tdk ada perubahan thn lg nasibmu akan seperti elevania
Cak Bogang8383	5	Oke	Oke

3. Case Folding

Tahap ini melibatkan normalisasi teks dalam pemrosesan bahasa alami (*Natural Language Processing/NLP*) dengan cara mengonversi seluruh karakter dalam teks menjadi huruf kecil. Tujuannya adalah untuk mencegah perbedaan antara huruf kapital dan huruf kecil memengaruhi hasil analisis teks. Proses ini dilakukan menggunakan library *re* pada Google Colab dengan bahasa pemrograman Python. Di bawah ini disajikan pseudocode beserta contoh data yang telah melalui tahap case folding..

Proses Case Folding:

- Definisikan fungsi `case_folding` untuk mengubah teks menjadi huruf kecil.
- Terapkan fungsi pada kolom 'cleaning' dan simpan hasilnya pada kolom 'case_folding'.

Tabel 3.3 Contoh data hasil *Case Folding*

Username	Rating	Review Text	Cleaning	Case Folding
Mustafa mus	5	sangat puas	sangat puas	sangat puas
Wiji Utami	5	Like	Like	like
Ezra Putra	5	Okee	Okee	okee
Heru budi utomo Utomo	2	Blibli, kl kalian tdk ada perubahan, 2thn lg nasibmu akan seperti elevania.	Blibli kl kalian tdk ada perubahan thn lg nasibmu akan seperti elevania	blibli kl kalian tdk ada perubahan thn lg nasibmu akan seperti elevania
Cak Bogang8383	5	Oke	Oke	oke

4. Normalisasi Kata

Dalam tahap normalisasi kata pada pemrosesan bahasa alami (*Natural Language Processing/NLP*), kata-kata tidak baku atau slang diubah menjadi bentuk baku yang sesuai dengan aturan bahasa yang berlaku. Tujuan utamanya adalah untuk memastikan konsistensi kata agar lebih mudah dianalisis. Proses ini meliputi penggantian kata-kata informal seperti "gak" menjadi "tidak", atau "yg" menjadi "yang". Normalisasi ini dilakukan menggunakan *library nltk* di Google Colab dengan bahasa pemrograman Python. Berikut ini adalah pseudocode dan contoh data yang telah melalui tahapan normalisasi kata.

Normalisasi Kata:

- Baca file kamus tidak baku ('kamuskatabaku1.xlsx').
- Definisikan fungsi 'replace_taboo_words' untuk mengganti kata-kata tidak baku.
- Terapkan fungsi normalisasi pada kolom 'case_folding' dan simpan hasilnya ke kolom 'normalisasi'.

Tabel 3.4 Contoh data hasil Normalisasi Kata

Review Text	Cleaning	Case Folding	Normalisasi Kata
sangat puas	sangat puas	sangat puas	sangat puas
Like	Like	like	suka
Okee	Okee	okee	oke
Blibli, kl kalian tdk ada perubahan, 2thn lg nasibmu akan seperti elevania.	Blibli kl kalian tdk ada perubahan thn lg nasibmu akan seperti elevania	blibli kl kalian tdk ada perubahan thn lg nasibmu akan seperti elevania	blibli kalau kalian tidak ada perubahan tahun lagi nasibmu akan seperti elevania
Oke	Oke	oke	oke

5. Stopword Removal

Tahap ini merupakan bagian dari pemrosesan bahasa alami (*Natural Language Processing/NLP*) yang bertujuan menghapus kata-kata yang dianggap tidak memiliki peran penting dalam pemahaman isi teks. Kata-kata tersebut dikenal sebagai *stopwords*, seperti kata hubung (misalnya "dan", "atau"), kata ganti umum (seperti "saya", "kamu"), dan kata-kata lain yang sering muncul tetapi tidak memberikan kontribusi berarti dalam analisis. Proses ini dilakukan menggunakan library *nltk* di Google Colab dengan bahasa pemrograman Python. Berikut ini pseudocode dan contoh data hasil yang telah melewati tahapan *stopword removal*.

Penghapusan Stopword:

- Unduh daftar stopwords untuk bahasa Indonesia.
- Tambahkan kata spesifik yang ingin dihapus.
- Definisikan fungsi 'remove_stopwords' untuk menghapus stopwords.
- Terapkan fungsi pada kolom 'normalisasi' dan simpan hasilnya di kolom 'stopword_removal'.

Tabel 3.5 Contoh data hasil *Stopword Removal*

<i>Cleaning</i>	<i>Case Folding</i>	<i>Normalisasi Kata</i>	<i>Stopword Removal</i>
sangat puas	sangat puas	sangat puas	puas
Like	like	suka	suka
Okee	okee	oke	oke
Blibli kl kalian tdk ada perubahan thn lg	blibli kl kalian tdk ada perubahan thn lg	blibli kalau kalian tidak ada perubahan tahun lagi	perubahan
nasibmu akan seperti elevania	nasibmu akan seperti elevania	nasibmu akan seperti elevania	nasibmu
Oke	oke	oke	elevania

6. Stemming

Tahapan ini merupakan bagian dari proses dalam pemrosesan bahasa alami (*Natural Language Processing/NLP*) yang berfungsi untuk menghapus imbuhan atau bentuk turunan kata, sehingga diperoleh bentuk dasar atau akar katanya. Tujuannya adalah untuk menyederhanakan teks dan menghasilkan representasi data yang lebih seragam. Proses ini dilakukan menggunakan library *Sastrawi* pada platform Google Colab dengan bantuan bahasa pemrograman Python. Berikut ini pseudocode dan contoh data hasil yang telah melewati tahapan *stemming*.

Stemming Data:

- Buat objek stemmer menggunakan Sastrawi.
- Definisikan fungsi `stem_text` untuk mengubah setiap kata menjadi bentuk dasarnya.
- Terapkan stemming pada kolom `stopword_removal` dan simpan hasilnya di kolom `stemming_data`.

Simpan Hasil Preprocessing:

- Simpan DataFrame hasil preprocessing ke dalam file CSV ('hasil_stemming_data.csv').

Tabel 3.6 Contoh data hasil *Stemming*

Case Folding	Normalisasi Kata	Stopword Removal	Stemming
sangat puas	sangat puas	puas	puas
like	suka	suka	suka
okee	oke	oke	oke
blibli kl kalian tdk ada	blibli kalau kalian tidak ada	perubahan	ubah nasib
perubahan thn lg nasibmu	perubahan tahun lagi nasibmu	nasibmu	elevania
akan seperti elevania	akan seperti elevania	elevania	
oke	oke	oke	oke

7. *Tokenize*

Pada tahap ini, dilakukan proses dalam pemrosesan bahasa alami (*Natural Language Processing/NLP*) yang bertujuan membagi teks menjadi bagian-bagian kecil yang disebut token. Token dapat berupa kata, frasa, simbol, atau elemen lain yang memiliki arti khusus. Proses ini dijalankan menggunakan library *nltk* di Google Colab dengan bahasa pemrograman Python. Berikut ini pseudocode dan contoh data hasil yang telah melewati tahapan *tokenize*.

Tokenisasi:

- Definisikan fungsi `tokenize` untuk membagi teks menjadi kata-kata (token).
- Terapkan fungsi pada kolom `stemming_data` dan simpan hasilnya di kolom `tokenize`.
- Simpan DataFrame hasil preprocessing ke dalam file CSV ('hasil_preprocessing_data.csv').

Tabel 3.7 Contoh data hasil *Tokenize*

Case Folding	Normalisasi Kata	Stopword Removal	Stemming	Tokenize
sangat puas	sangat puas	puas	puas	['puas']
like	suka	suka	suka	['suka']
okee	oke	oke	oke	['oke']
blibli kl kalian tdk	blibli kalau kalian tidak	perubahan	ubah nasib	['ubah',
ada perubahan thn lg	ada perubahan tahun lagi	nasibmu	elevania	'nasib',
nasibmu akan seperti	nasibmu akan seperti	elevania		'elevania']
elevania	elevania		oke	
oke	oke	oke	oke	['oke']

C. Pembagian Data

Langkah ini mencakup pemisahan dataset menjadi dua bagian, yakni data pelatihan (*training*) dan data pengujian (*testing*). Data pelatihan dimanfaatkan untuk membangun model dengan algoritma *Support Vector Machine*, sementara data pengujian digunakan untuk mengevaluasi performa model setelah proses pelatihan selesai. Tahapan ini dilakukan menggunakan *library sklearn* di Google Colab dengan bahasa pemrograman Python. Berikut ini pseudocode pada tahapan pembagian data.

```
# Split data latih dan uji
- Load Dataset
  - Baca file CSV 'hasil_preprocessing_data.csv' ke dalam dataframe 'data'
  - Hapus baris yang memiliki nilai kosong (NA)

- Split Dataset
  - Tentukan kolom fitur yang akan digunakan, yaitu 'stemming_data'
  - Bagi data menjadi:
    - 'X_train_raw' (70% dari data untuk pelatihan)
    - 'X_test_raw' (30% dari data untuk pengujian)
    - dengan parameter 'random_state=42' untuk memastikan hasil pembagian konsisten

- Create DataFrames
  - Buat dataframe 'train_df' dari 'X_train_raw'
  - Buat dataframe 'test_df' dari 'X_test_raw'

- Display Dataset Sizes
  - Cetak jumlah data latih dan data uji
```

- Visualize Data Distribution

- Buat diagram batang (bar chart) dengan dua batang:
- Batang pertama untuk jumlah data latih
- Batang kedua untuk jumlah data uji
- Tambahkan label jumlah data di atas masing-masing batang
- Beri judul pada grafik: "Jumlah Data Latih dan Uji"
- Tampilkan grafik.

D. Pelabelan Data Latih dengan *Lexicon Based*

Proses pelabelan data dilakukan untuk mengidentifikasi dan mengklasifikasikan sentimen yang terkandung dalam sebuah teks ke dalam kategori sentimen tertentu. Dalam penelitian ini, pelabelan dilakukan menggunakan pendekatan *lexicon-based* dengan memanfaatkan lexicon InSet, yaitu kumpulan kata yang telah dilengkapi dengan label sentimen spesifik. Pelabelan dilakukan dengan mencocokkan kata-kata dalam teks terhadap daftar yang ada di lexicon tersebut, lalu memberikan skor sentimen berdasarkan kecocokan kata-kata yang ditemukan. Berikut ini pseudocode dan contoh data hasil yang telah melewati tahapan pelabelan data dengan *Lexicon Based*.

Pseudocode: Labeling Data Training & Testing with Lexicon-Based Approach

- Load Lexicon

- BACA file "positive.tsv" dengan pemisah tab, skip baris pertama,
- simpan kolom sebagai ["word", "weight"] → positive_lexicon_df
- BACA file "negative.tsv" dengan pemisah tab, skip baris pertama,
- simpan kolom sebagai ["word", "weight"] → negative_lexicon_df

- Konversi kolom bobot menjadi numerik

- UBAH positive_lexicon_df["weight"] menjadi numerik
- UBAH negative_lexicon_df["weight"] menjadi numerik

- Ubah lexicon menjadi dictionary

- positive_lexicon ← dictionary dari (word, weight) di positive_lexicon_df
- negative_lexicon ← dictionary dari (word, weight) di negative_lexicon_df

- Definisikan fungsi untuk menentukan sentimen (binary)
 - FUNGSI `determine_sentiment_binary(tokens):`
 - JIKA `tokens` adalah string berbentuk list:
 - UBAH string menjadi list menggunakan `ast.literal_eval`
 - JIKA gagal, hilangkan tanda kurung & tanda kutip, lalu split
 - JIKA `tokens` adalah list:
 - GABUNG semua elemen jadi string
 - JIKA `tokens` adalah string:
 - UBAHKAN ke huruf kecil lalu split menjadi list kata
 - `score ← 0`
 - UNTUK setiap word dalam `tokens`:
 - JIKA word ada di `positive_lexicon`:
 - TAMBAHKAN bobot positif ke `score`
 - ELSE JIKA word ada di `negative_lexicon`:
 - `neg_weight ← bobot dari negative_lexicon`
 - TAMBAHKAN `neg_weight` jika negatif, selain itu kurangi `score` dengan bobot
 - KEMBALIKAN "Positif" jika `score > 0`, selain itu "Negatif"
 - Label Data Training
 - Terapkan fungsi `'determine_sentiment'` pada setiap baris `'train_df['tokenize']'` → simpan ke kolom 'Sentiment'
 - Terapkan fungsi `'replace_none'` untuk mengganti nilai 'None' di kolom 'Sentiment'
 - Simpan dataframe `'train_df'` yang sudah dilabeli ke file `'train_data_labelled.csv'`
 - Label Data Testing (untuk evaluasi)
 - Ulangi langkah yang sama pada dataframe `'test_df'` menggunakan `lexicon`
 - Simpan dataframe `'test_df'` yang sudah dilabeli ke file `'test_data_labelled.csv'`
 - Visualize Sentiment Distribution (Training Data)

- Gunakan `sns.countplot` untuk menampilkan distribusi jumlah data berlabel "Positif" dan "Negatif"
- Tambahkan judul grafik: "Distribusi Sentimen pada Data Latih"
- Tampilkan grafik.

E. Vektorisasi TF-IDF

Setelah pelabelan data, dilakukan vektorisasi TF-IDF untuk mengubah teks hasil *preprocessing* menjadi bentuk numerik yang siap digunakan pada tahap klasifikasi. Metode ini menghitung bobot kata berdasarkan frekuensi kemunculannya di suatu dokumen (*Term Frequency*) dan tingkat kelangkaannya di seluruh dokumen (*Inverse Document Frequency*). Hasil pembobotan menghasilkan matriks numerik yang menjadi input bagi algoritma SVM pada proses klasifikasi. Berikut ini pseudocode pada tahapan vektorisasi TF-IDF.

```
# Pseudocode vektorisasi TF-IDF
- Prepare Training and Testing Data
  - Ambil fitur teks dari `train_df['tokenize']` → simpan ke `X_train`
  - Ambil label sentimen dari `train_df['Sentiment']` → simpan ke `y_train`
  - Ambil fitur teks dari `test_df['tokenize']` → simpan ke `X_test`
  - Ambil label sentimen dari `test_df['Sentiment']` → simpan ke `y_test`
- Convert Text to Numerical Features (TF-IDF)
  - Inisialisasi `TfidfVectorizer`
  - Fit dan transform `X_train` menjadi matriks fitur `X_train_vec`
  - Transform `X_test` menggunakan vectorizer yang sama menjadi `X_test_vec`
```

F. Klasifikasi *Support Vector Machine*

Setelah melalui tahapan pembagian data, penerapan algoritma *Support Vector Machine* (SVM) dilakukan dengan memanfaatkan *library* scikit-learn (sklearn) yang berbasis Python. Pada tahap ini, data latih yang telah dibagi digunakan untuk melatih model SVM. Penelitian ini menerapkan kernel linear karena sifatnya yang

sederhana serta kemampuan untuk memisahkan data secara linier sesuai dengan karakteristik dataset yang digunakan..

SVM dengan kernel linear beroperasi dengan menemukan sebuah hyperplane yang dapat memisahkan dua kelas dalam ruang fitur. Algoritma ini bertujuan untuk memperluas margin antara kedua kelas tersebut dengan menentukan pemisah linier yang paling optimal. Dalam prosesnya, SVM menghitung nilai bobot dan bias yang dibutuhkan untuk menetapkan posisi hyperplane. Hyperplane inilah yang kemudian digunakan dalam proses klasifikasi data baru.

Model ini dilatih menggunakan data pelatihan untuk menghitung parameter bobot dan bias, sehingga mampu membedakan data ke dalam dua kelas (positif dan negatif) dengan margin pemisah yang paling luas. Setelah proses pelatihan selesai, model SVM dapat digunakan untuk mengklasifikasikan data pengujian berdasarkan letak data terhadap hyperplane yang telah ditentukan. Berikut ini pseudocode pada tahapan klasifikasi SVM.

Pseudocode: Klasifikasi Sentimen dengan Support Vector Machine (SVM)

- Train SVM Model
 - Inisialisasi model SVM dengan kernel linear (`^SVC(kernel="linear")`)
 - Latih model menggunakan `'X_train_vec'` dan `'y_train'`
- Prediction
 - Gunakan model SVM untuk memprediksi sentimen pada `'X_test_vec'`
 - Simpan hasil prediksi ke `'y_pred'`
- Display Prediction Summary
 - Hitung total data yang diprediksi (`total = len(y_pred)`)
 - Hitung jumlah prediksi "Positif" (`positif`)
 - Hitung jumlah prediksi "Negatif" (`negatif`)
 - Hitung persentase:
 - `'persen_positif = (positif / total) * 100'`
 - `'persen_negatif = (negatif / total) * 100'`
 - Cetak total data, persentase positif, dan persentase negatif.

G. Validasi dan Evaluasi Data

Hasil pengujian klasifikasi *Support Vector Machine* akan divalidasi dan dievaluasi untuk menentukan tingkat akurasi. Proses validasi dilakukan dengan membagi data menjadi dua set, yaitu data latih (*training set*) dan data uji (*test set*). Data latih digunakan untuk melatih model, sedangkan data uji digunakan untuk menguji performa model. Evaluasi akurasi algoritma dilakukan menggunakan *Confusion Matrix*, yang membantu untuk mengukur sejauh mana model dapat memprediksi kelas dengan benar, serta menghitung metrik evaluasi seperti *precision*, *recall*, dan *F1-score*. Berikut ini pseudocode pada tahapan validasi dan evaluasi data.

Pseudocode: Evaluasi Model SVM

- Hitung Confusion Matrix
 - Gunakan fungsi `confusion_matrix(y_test, y_pred)` untuk menghitung matriks kebingungan
 - Cetak hasil confusion matrix
- Visualisasi Confusion Matrix
 - Buat plot dengan ukuran 6x6
 - Gunakan `sns.heatmap` untuk memvisualisasikan confusion matrix dengan:
 - `annot=True` (menampilkan nilai dalam kotak)
 - `fmt="d"` (format integer)
 - `cmap="Blues"` (warna biru)
 - `xticklabels` dan `yticklabels` diisi dengan label `[Positif, Negatif]`
 - Beri judul pada grafik: '"Confusion Matrix SVM"'
 - Tampilkan plot
- Hitung Akurasi Model
 - Gunakan `accuracy_score(y_test, y_pred)` untuk menghitung akurasi
 - Cetak akurasi dalam persen dengan dua angka desimal
- Tampilkan Classification Report
 - Gunakan `classification_report(y_test, y_pred)` untuk mendapatkan metrik evaluasi:
 - Precision

- Recall
- F1-Score
- Cetak laporan klasifikasi tersebut.

H. Visualisasi Data

Visualisasi data dilakukan untuk menyajikan hasil akhir dalam bentuk *WordCloud* agar mempermudah peneliti dan pembaca dalam memahami informasi yang diperoleh. Kata-kata yang paling sering muncul dalam teks akan divisualisasikan dalam bentuk *WordCloud*, dengan ukuran huruf mencerminkan frekuensi kemunculannya. Proses ini menggunakan *library WordCloud* pada bahasa pemrograman Python. Berikut ini pseudocode pada tahapan visualisasi data.

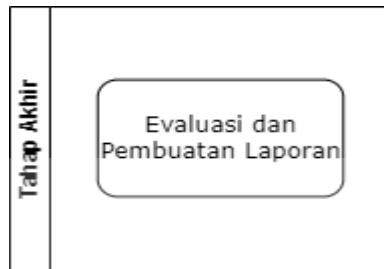
Pseudocode: Visualisasi Data dengan Wordcloud Berdasarkan Sentimen

- Persiapan DataFrame untuk hasil prediksi
 - Untuk setiap data pada X_test:
 - Jika tipe data berupa list token, gabungkan menjadi string.
 - Bersihkan tanda kutip tunggal ', ganda ", atau backtick ` dari setiap token.
 - Jika bukan list, konversi ke string lalu bersihkan tanda kutip/backtick.
 - Simpan hasil pembersihan ini pada kolom tokenize.
 - Simpan label prediksi (y_pred) pada kolom Sentiment_Pred.
- Pisahkan dan gabungkan token berdasarkan sentimen
 - Ambil semua teks dari baris dengan prediksi Negatif, gabungkan menjadi satu string sent_neg_pred.
 - Ambil semua teks dari baris dengan prediksi Positif, gabungkan menjadi satu string sent_pos_pred.
- Definisikan fungsi create_wordcloud
 - Input: text (string), title (judul).
 - Buat wordcloud berwarna latar hitam, ukuran 800x400 piksel.
 - Tampilkan wordcloud menggunakan matplotlib tanpa sumbu koordinat.
 - Beri judul sesuai parameter title.
- Visualisasi wordcloud untuk hasil prediksi data uji

- Panggil `create_wordcloud(sent_neg_pred, "Wordcloud Prediksi Negatif (Data Uji)").`
- Panggil `create_wordcloud(sent_pos_pred, "Wordcloud Prediksi Positif (Data Uji)").`

3.1.3 Tahap Akhir

Pada tahap akhir penelitian ini, terdapat rangkuman serta rekomendasi bagi pembaca dan peneliti guna meningkatkan kualitas penelitian mendatang atau penelitian lainnya.



Gambar 3.5 *Flow Diagram* Tahap Akhir

A. Evaluasi

Evaluasi pada tahap akhir dalam penelitian ini mencakup analisis menyeluruh terhadap hasil dari tahapan sebelumnya, termasuk validasi dan evaluasi model *Support Vector Machine* yang digunakan. Evaluasi ini dilakukan untuk memastikan bahwa kualitas dan akurasi dari analisis sentimen yang diterapkan benar-benar sesuai dengan tujuan penelitian, yaitu memahami opini publik tentang kepuasan pelanggan terhadap aplikasi Blibli berdasarkan ulasan di Google Play Store.

B. Pembuatan Laporan

Penyusunan laporan merupakan tahap akhir dari penelitian ini yang bertujuan untuk mencatat secara lengkap seluruh rangkaian proses penelitian, mulai dari tahap perencanaan hingga analisis data. Laporan ini berisi hasil analisis sentimen menggunakan metode *Support Vector Machine*, kesimpulan mengenai opini publik terkait tingkat kepuasan pelanggan terhadap aplikasi Blibli, serta rekomendasi yang diharapkan dapat menjadi acuan bagi penelitian sejenis di masa yang akan datang.

BAB IV

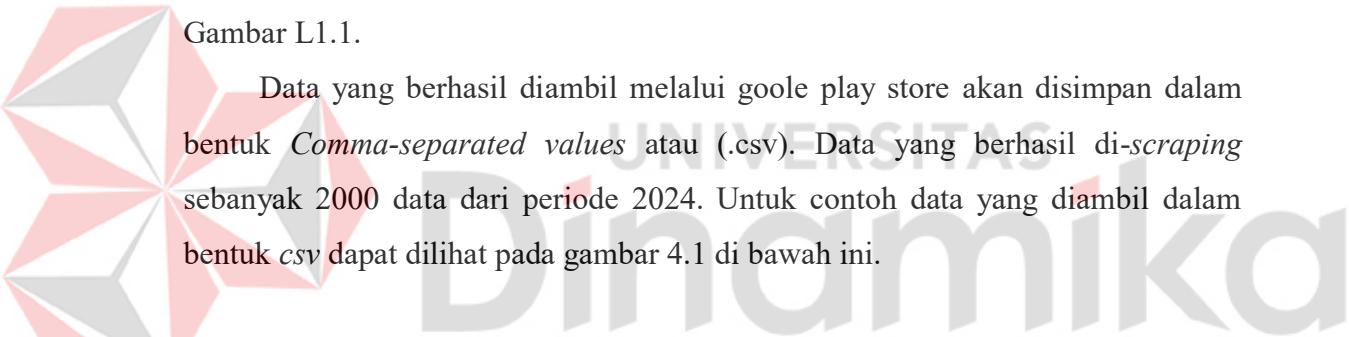
HASIL DAN PEMBAHASAN

Bagian ini menguraikan hasil serta pembahasan dari penelitian yang telah dilaksanakan, sesuai dengan tahapan-tahapan yang dirancang berdasarkan metode yang digunakan dalam studi ini.

4.1 Data Scraping

Data Scraping adalah tahap pengambilan data dengan menggunakan bahasa pemrograman python dan dengan bantuan library google_play_scraper. ID aplikasi blibli dari Google Play Store yang sudah diketahui dimasukkan ke code untuk mendapatkan data-data dari Google Play Store, kemudian memasukkan bahasa, negara, dan jumlah data yang dibutuhkan. Penulisan kodennya dapat dilihat pada Gambar L1.1.

Data yang berhasil diambil melalui google play store akan disimpan dalam bentuk *Comma-separated values* atau (.csv). Data yang berhasil di-scraping sebanyak 2000 data dari periode 2024. Untuk contoh data yang diambil dalam bentuk csv dapat dilihat pada gambar 4.1 di bawah ini.



A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Review ID	Username	Rating	Review Text	Date													
2	8ba7104f-7fb7-4db0-a806-6248a419056e	Mustafa mus,5,sangat puas	2024-11-09 18:07:36															
3	e75f70d1-ba34-4cc9-9d0a-d48477fe9e91	Wiji Utami,5,Like	2024-11-09 17:07:10															
4	c43f3aee4-4564-4fe5-9997-5fd7d8a00053	Era Putra,5,okeee	2024-11-09 14:59:18															
5	538500c4-4741-4f52-9c88-f650c4486cba	Tyas Aulia,1,Aplikasi buruk....jangan pake aplikasi ini...payleter di tolak padahal di lazada dan shopee bisa semua payleternya	2024-11-09 13:50:57															
6	08bd2025-603c-4453-944b-19cc41b1922	Cak Bogang8383,5,oke	2024-11-09 12:58:29															
7	91fd7d64-8329-4ecb-8c2e-853e2988da30	Heru budi utomo,2,"Blibli, kli kalian tdk ada perubahan, 2thn lg nasibmu akan seperti elevania."	2024-11-09 11:34:18															
8	4bf7c8d1-a937-41ff-aa6c-80473d4b804f	itsmail Real,5,mantul lah ini. walaupun ongkirnya masih belum gratis.	2024-11-09 10:47:30															
9	f9b682e9-0da3-4181-896a-360f1ae5e289	Bayu Sarasito,5,sangat memuaskan	2024-11-09 09:55:20															
10	dbc5b5e8-fecc-4560-9933-0ff9b9156241	Abdul Hayat,5,Kreennn...sangat Bagusssss & Membantu	2024-11-09 07:34:04															
11	5793cc22-49cc-4c98-a21f-0db25816409f	Master Kombak,4,Bagus	2024-11-09 06:59:36															
12	a7b79e78-3f9a-404c-bd18-67b1008b3898	Doju Store,5,Semoga terus berkembang menjadi lebih baik lagi aplikasinya	2024-11-09 06:09:09															
13	3600a5cb-41e2-4d68-a5cc-8667342bda16	karsono wa,1,"Tampilan aplikasinya membingungkan, kebanyakan animasi² yang justru mengganggu pada saat proses memilih barang"	2024-11-09 05:37:37															
14	d89079be-6210-43e2-ba83-695051df28b2	Taufik Tri Risqj,1,Sangat buruk. Aplikasi tidak optimal dan kuota promo tidak jelas. Padahal sudah beli voucher promo taunya gabisa digunakan. Buruk dan tidak ri																
15	ec77775e-76c3-4490-ad4e-5afe032dafb6	masasep gituloh,5,mantab betul	2024-11-09 04:54:09															
16	ea3e2401-b0b4-48c2-b1b8-d2746b8216a7	Firman Medianaaa,5,Good	2024-11-09 03:04:24															
17	140e8dd0-eb3c-4b4d-8635-7c08508e71de	Aan W.,1,"Pengguna paylater saya ditolak, dan sekarang nomor saya di teror promosi pinjol. Tolong profesional sedikit, kalau ditolak ya hapus data saya. Bukan																
18	b490a010-8523-4bdc-9289-1972787b833	Suharto Ps,5,cepat,2024-11-09 01:13:21																
19	ab544fe5-3035-4b89-8c6d-20d330b5eab5	fardia kusti,1,"ulasan masa de cek, tolong namanya belanja di blibli selalu mengecewakan gamau ngaku kalau barang yang dikirim gak sesuai produk bagus cuma s																
20	45b999e4-76d8-4425-9a0d-ab0fd0f993867	Benny Ardiansah,5,mantaps	2024-11-09 00:25:08															
21	bf133596-169c-48d2-8524-c3de8b54582	Najeges,5,keren,2024-11-08 22:31:20																
22	c6e0238a-30ce-478e-81d3-219d8396f393	Amir Hamzah,5,mantap	2024-11-08 22:14:13															

Gambar 4.1 Data ulasan hasil Scraping

4.2 Preprocessing Data

Preprocessing data adalah serangkaian langkah untuk membersihkan dan mempersiapkan data mentah agar lebih mudah dan akurat diolah dalam proses analisis atau pemodelan. Dalam proses ini dibagi dengan 7 tahap. Pada tahap

preprocessing data menggunakan *code editor Google Colaboratory* dengan bahasa *python*, sebelum melakukan *preprocessing data* periksa data yang sudah melalui *data scraping*. Penulisan kodenya dapat dilihat pada gambar L1.2.

Setelah membaca data maka akan dilakukan proses penghapusan data duplikat karena data duplikat dapat mempengaruhi hasil analisis. Penulisan kodenya dapat dilihat pada gambar 4.2 di bawah ini.

```
▶ df.drop_duplicates(subset=['Review Text'], keep='first', inplace=True)
df.info()

→ <class 'pandas.core.frame.DataFrame'>
Index: 1604 entries, 0 to 1999
Data columns (total 4 columns):
 #   Column      Non-Null Count  Dtype  
---  --          --          --      
 0   Date        1604 non-null   object 
 1   Username    1604 non-null   object 
 2   Rating      1604 non-null   int64  
 3   Review Text 1604 non-null   object 
dtypes: int64(1), object(3)
memory usage: 62.7+ KB
```

Gambar 4.2 Proses hapus data duplikat

Pada tahap berikutnya yaitu *cleaning*. Tahap ini merupakan penghapusan *url*, *HTML*, symbol, angka, emoji, karakter berulang, dan tanda baca pada ulasan. Bentuk penulisan kode lengkapnya dapat dilihat pada gambar L1.3 dan hasilnya pada gambar 4.3

	Date	Username	Rating	Review Text	cleaning
0	2024-11-09 18:07:36	Mustafa mus	5	sangat puas	sangat puas
1	2024-11-09 17:07:10	Wiji Utami	5	Like	Like
2	2024-11-09 14:59:18	Ezra Putra	5	okee	okee
3	2024-11-09 13:50:57	Tyas Aulia	1	Aplikasi buruk...jangan pake aplikasi ini...p...	Aplikasi buruk jangan pake aplikasi ini paylet...
4	2024-11-09 12:58:29	Cak Bogang8383	5	oke	oke
...
1993	2024-06-11 15:51:28	Ali Yusup	5	mudah dan praktis	mudah dan praktis
1995	2024-06-11 14:16:52	fajar nugroho	1	Tampilan baru malah bikin ruwet buka aplikasi ...	Tampilan baru malah bikin ruwet buka aplikasi ...
1996	2024-06-11 14:13:16	Abdul Rajak Rumbati	1	Semoga membantu karena saya ini pemula	Semoga membantu karena saya ini pemula
1998	2024-06-11 10:56:49	ELFRY KUHUELA	5	Pelayanan terbaik.. tak di ragukan lagi. Pesan...	Pelayanan terbaik tak di ragukan lagi Pesan d...
1999	2024-06-11 09:10:53	Nur Wening	1	Jelek. Cancel pesanan malah saldo masuk ke bil...	Jelek Cancel pesanan malah saldo masuk ke bilp...

Gambar 4.3 Data hasil *cleaning*

Tahap berikutnya merupakan *Case Folding*, pada tahap ini dilakukan perubahan setiap ulasan yang memiliki huruf besar atau *uppercase* diubah menjadi huruf kecil atau *lowercase*. Bentuk penulisan kodennya dapat dilihat pada gambar L1.4 dan hasilnya pada gambar 4.4

	Date	Username	Rating	Review Text	cleaning	case_folding
0	2024-11-09 18:07:36	Mustafa mus	5	sangat puas	sangat puas	sangat puas
1	2024-11-09 17:07:10	Wiji Utami	5	Like	Like	like
2	2024-11-09 14:59:18	Ezra Putra	5	okee	okee	okee
3	2024-11-09 13:50:57	Tyas Aulia	1	Aplikasi buruk...jangan pake aplikasi ini...p...	Aplikasi buruk jangan pake aplikasi ini paylet...	aplikasi buruk jangan pake aplikasi ini paylet...
4	2024-11-09 12:58:29	Cak Bogang8383	5	oke	oke	oke
...
1993	2024-06-11 15:51:28	Ali Yusup	5	mudah dan praktis	mudah dan praktis	mudah dan praktis
1995	2024-06-11 14:16:52	fajar nugroho	1	Tampilan baru malah bikin ruwet buka aplikasi ...	Tampilan baru malah bikin ruwet buka aplikasi ...	tampilan baru malah bikin ruwet buka aplikasi ...
1996	2024-06-11 14:13:16	Abdul Rajak Rumbati	1	Semoga membantu karena saya ini pemula	Semoga membantu karena saya ini pemula	semoga membantu karena saya ini pemula
1998	2024-06-11 10:56:49	ELFRY KUHUELA	5	Pelayanan terbaik.. tak di ragukan lagi. Pesan...	Pelayanan terbaik tak di ragukan lagi Pesan d...	pelayanan terbaik tak di ragukan lagi pesan d...
1999	2024-06-11 09:10:53	Nur Wening	1	Jelek. Cancel pesanan malah saldo masuk ke bil...	Jelek Cancel pesanan malah saldo masuk ke bil...	jelek cancel pesanan malah saldo masuk ke bil...

1604 rows x 6 columns

Gambar 4.4 Hasil proses *Case Folding*

Tahap berikutnya yaitu normalisasi kata, pada tahap ini dilakukan perubahan kalimat tidak baku menjadi kalimat baku dengan membuat fungsi penggantian kata baku seperti pada gambar L1.5 lalu memasukkan kamus kata baku dan proses normalisasi kata seperti pada gambar L1.6 dan hasilnya pada gambar 4.5

	Date	Username	Rating	Review Text	cleaning	case_folding	normalisasi
0	2024-11-09 18:07:36	Mustafa mus	5	sangat puas	sangat puas	sangat puas	sangat puas
1	2024-11-09 17:07:10	Wiji Utami	5	Like	Like	like	suka
2	2024-11-09 14:59:18	Ezra Putra	5	okee	okee	okee	oke
3	2024-11-09 13:50:57	Tyas Aulia	1	Aplikasi buruk...jangan pake aplikasi ini...p...	Aplikasi buruk jangan pake aplikasi ini paylet...	aplikasi buruk jangan pake aplikasi ini payle...	aplikasi buruk jangan pakai aplikasi ini payle...
4	2024-11-09 12:58:29	Cak Bogang8383	5	oke	oke	oke	oke
...
1993	2024-06-11 15:51:28	Ali Yusup	5	mudah dan praktis	mudah dan praktis	mudah dan praktis	mudah dan praktis
1995	2024-06-11 14:16:52	fajar nugroho	1	Tampilan baru malah bikin ruwet buka aplikasi ...	Tampilan baru malah bikin ruwet buka aplikasi ...	tampilan baru malah bikin ruwet buka aplikasi ...	tampilan baru malah bikin ruwet buka aplikasi ...
1996	2024-06-11 14:13:16	Abdul Rajak Rumbati	1	Semoga membantu karena saya ini pemula			
1998	2024-06-11 10:56:49	ELFRY KUHUELA	5	Pelayanan terbaik.. tak di ragukan lagi. Pesan...	Pelayanan terbaik tak di ragukan lagi Pesan d...	pelayanan terbaik tak di ragukan lagi pesan d...	pelayanan terbaik tak di ragukan lagi pesan d...
1999	2024-06-11 09:10:53	Nur Wening	1	Jelek. Cancel pesanan malah saldo masuk ke bil...	Jelek Cancel pesanan malah saldo masuk ke bil...	jelek cancel pesanan malah saldo masuk ke bil...	jelek cancel pesanan malah saldo masuk ke bil...

1604 rows x 7 columns

Gambar 4.5 Hasil proses normalisasi kata

Tahap berikutnya yaitu *Stopword Removal*, pada tahap ini dilakukan proses penghapusan kata-kata umum yang tidak memiliki makna penting dalam analisis

seperti "yang", "dan", "di", "ke", dan sebagainya. Proses ini diawali dengan membuat fungsi *stopword removal* seperti pada Gambar L1.7, kemudian memasukkan daftar kata *stopword* ke dalam fungsi tersebut seperti yang ditunjukkan pada Gambar L1.8, dan hasil dari penghapusan *stopword* dapat dilihat pada Gambar 4.6.

	Date	Username	Rating	Review Text	cleaning	case_folding	normalisasi	stopword_removal
0	2024-11-09 18:07:36	Mustafa mus	5	sangat puas	sangat puas	sangat puas	sangat puas	puas
1	2024-11-09 17:07:10	Wiji Utami	5	Like	Like	like	suka	suka
2	2024-11-09 14:59:18	Ezra Putra	5	okee	oke	oke	oke	oke
3	2024-11-09 13:50:57	Tyas Aulia	1	Aplikasi buruk...jangan pake aplikasi ini...p...	Aplikasi buruk jangan pake aplikasi ini paylet...	aplikasi buruk jangan pake aplikasi ini paylet...	aplikasi buruk jangan pakai aplikasi ini payle...	buruk pakai payleter tolak lazada shope paylet...
4	2024-11-09 12:58:29	Cak Bogang8383	5	oke	oke	oke	oke	oke
...
1993	2024-06-11 15:51:28	Ali Yusup	5	mudah dan praktis	mudah dan praktis	mudah dan praktis	mudah dan praktis	mudah praktis
1995	2024-06-11 14:16:52	fajar nugroho	1	Tampilan baru malah bikin ruwet buka aplikasi ...	Tampilan baru malah bikin ruwet buka aplikasi ...	tampilan baru malah bikin ruwet buka aplikasi ...	tampilan baru malah bikin ruwet buka aplikasi ...	tampilan bikin ruwet buka bertele langsung menu
1996	2024-06-11 14:13:16	Abdul Rajak Rumbati	1	Semoga membantu karena saya ini pemula	semoga membantu pemula			
1998	2024-06-11 10:56:49	ELFRY KUHUEL A	5	Pelayanan terbaik.. tak di ragukan lagi. Pesan...	Pelayanan terbaik tak di ragukan lagi Pesan dg...	pelayanan terbaik tak di ragukan lagi pesan dg...	pelayanan terbaik tak di ragukan lagi pesan de...	pelayanan terbaik rugukan pesan nominal tujuan...
1999	2024-06-11 09:10:53	Nur Wening	1	Jelek. Cancel pesanan malah saldo masuk ke bil...	Jelek Cancel pesanan malah saldo masuk ke bil...	jelek cancel pesanan malah saldo masuk ke bil...	jelek cancel pesanan malah saldo masuk ke bil...	jelek cancel pesanan saldo masuk bilpay pindah...

Gambar 4.6 Hasil Stopword Removal

Tahap berikutnya yaitu *Stemming Data*. Pada tahap ini dilakukan proses pengubahan kata-kata dengan menghilangkan imbuhan pada setiap kata sehingga menjadi kata dasar. Proses ini dilakukan dengan memanfaatkan *library* Sastrawi. Bentuk penulisan kodennya dapat dilihat pada gambar L1.9 dan hasilnya pada gambar 4.7

	Date	Username	Rating	Review Text	cleaning	case_folding	normalisasi	stopword_removal	stemming_data
0	2024-11-09 18:07:36	Mustafa mus	5	sangat puas	sangat puas	sangat puas	sangat puas	puas	puas
1	2024-11-09 17:07:10	Wiji Utami	5	Like	Like	like	suka	suka	suka
2	2024-11-09 14:59:18	Ezra Putra	5	okee	oke	oke	oke	oke	oke
3	2024-11-09 13:50:57	Tyas Aulia	1	Aplikasi buruk...jangan pake aplikasi ini...p...	Aplikasi buruk jangan pake aplikasi ini paylet...	aplikasi buruk jangan pake aplikasi ini paylet...	aplikasi buruk jangan pakai aplikasi ini payle...	buruk pakai payleter tolak lazada shope paylet...	buruk pakai payleter tolak lazada shope paylet...
4	2024-11-09 12:58:29	Cak Bogang8383	5	oke	oke	oke	oke	oke	oke
...
1993	2024-06-11 15:51:28	Ali Yusup	5	mudah dan praktis	mudah dan praktis	mudah dan praktis	mudah dan praktis	mudah praktis	mudah praktis
1995	2024-06-11 14:16:52	fajar nugroho	1	Tampilan baru malah bikin ruwet buka aplikasi ...	Tampilan baru malah bikin ruwet buka aplikasi ...	tampilan baru malah bikin ruwet buka aplikasi ...	tampilan baru malah bikin ruwet buka aplikasi ...	tampilan bikin ruwet buka bertele langsung menu	tampil bikin ruwet buka bertele langsung menu
1996	2024-06-11 14:13:16	Abdul Rajak Rumbati	1	Semoga membantu karena saya ini pemula	semoga membantu pemula	moga bantu mula			
1998	2024-06-11 10:56:49	ELFRY KUHUEL A	5	Pelayanan terbaik.. tak di ragukan lagi. Pesan...	Pelayanan terbaik tak di ragukan lagi Pesan dg...	pelayanan terbaik tak di ragukan lagi pesan dg...	pelayanan terbaik tak di ragukan lagi pesan de...	pelayanan terbaik rugukan pesan nominal tujuan...	layan baik rugu pesan nominal tuju degan aman ...
1999	2024-06-11 09:10:53	Nur Wening	1	Jelek. Cancel pesanan malah saldo masuk ke bil...	Jelek Cancel pesanan malah saldo masuk ke bil...	jelek cancel pesanan malah saldo masuk ke bil...	jelek cancel pesanan saldo masuk bilpay pindah...	jelek cancel pesan saldo masuk bilpay pindah ...	jelek cancel pesan saldo masuk bilpay pindah s...

Gambar 4.7 Hasil Stemming Data

Tahap akhir dari *preprocessing data* yaitu *Tokenize*, pada tahap ini dilakukan proses pemecahan kalimat menjadi potongan-potongan kata (token) agar memudahkan dalam proses analisis selanjutnya. Bentuk penulisan kodennya dapat dilihat pada Gambar L1.10 dan hasil pemisahan kata tersebut dapat dilihat pada Gambar 4.8.

	Date	Username	Rating	Review Text	cleaning	case_folding	normalisasi	stopword_removal	stemming_data	tokenize
0	2024-11-09 18:07:36	Mustafa mus	5	sangat puas	sangat puas	sangat puas	sangat puas	puas	puas	[puas]
1	2024-11-09 17:07:10	Wiji Utami	5	Like	Like	like	suka	suka	suka	[suka]
2	2024-11-09 14:59:18	Ezra Putra	5	oke	oke	oke	oke	oke	oke	[oke]
3	2024-11-09 13:50:57	Tyas Aulia	1	Aplikasi buruk....jangan pake aplikasi ini...p...	Applikasi buruk jangan pake aplikasi ini payleter...	aplikasi buruk jangan pake aplikasi ini payleter...	aplikasi buruk jangan pake aplikasi ini payleter...	buruk pakai payleter tolak lazada shope...	buruk pakai payleter tolak lazada shope payleter...	[buruk, pakai, payleter, tolak, lazada, shope...]
4	2024-11-09 12:58:29	Cak Bogang8383	5	oke	oke	oke	oke	oke	oke	[oke]
...
1993	2024-06-11 15:51:28	Ali Yusup	5	mudah dan praktis	mudah dan praktis	mudah dan praktis	mudah dan praktis	mudah praktis	mudah praktis	[mudah, praktis]
1995	2024-06-11 14:16:52	fajar nugroho	1	Tampilan baru malah bikin ruwet buka aplikasi ...	Tampilan baru malah bikin ruwet buka aplikasi ...	tampilan baru malah bikin ruwet buka aplikasi ...	tampilan baru malah bikin ruwet buka aplikasi ...	tampilan bikin ruwet buka bertele langsung menu	tampilan bikin ruwet buka tele langsung menu	[tampil, bikin, ruwet, buka, tele, langsung, m...]
1996	2024-06-11 14:13:16	Abdul Rajak Rumbati	1	Semoga membantu karena saya ini pemula	Semoga membantu karena saya ini pemula	semoga membantu karena saya ini pemula	semoga membantu karena saya ini pemula	semoga membantu pemula	moga bantu mula	[moga, bantu, mula]
1998	2024-06-11 10:56:49	ELFRY KUHUELA	5	Pelayanan terbaik.. tak di ragukan lagi.. Pesan...	Pelayanan terbaik tak di ragukan lagi.. Pesan dg...	pelayanan terbaik tak di ragukan lagi.. Pesan dg...	pelayanan terbaik tak di ragukan lagi.. Pesan dg...	pelayanan terbaik ragukan pesan nominal tujuan...	layan baik rugu pesan nominal tuju... dengan aman ...	[layan, baik, rugu, pesan, nominal, tuju, dega...]

Gambar 4.8 Hasil proses *Tokenize*

4.3 Pembagian Data

Setelah dilakukan *Preprocessing data* menghasilkan 1576 data yang tersisa untuk dilakukan pembagian data, dalam penelitian ini data dibagi dengan *test_size=0.3* atau rasio 70:30 di mana 70% data digunakan untuk pelatihan dan 30% sisanya untuk pengujian atau sebanyak 1103 data pelatihan dan 473 data pengujian. Pembagian ini dilakukan secara acak namun konsisten dengan menetapkan parameter *random_state* untuk menjaga reproduksibilitas hasil. Untuk penulisan kodennya dapat dilihat pada gambar L1.11.

4.4 Pelabelan Data Latih dengan *Lexicon Based*

Setelah melalui tahapan pembagian data, tahap selanjutnya adalah pelabelan data menggunakan pendekatan lexicon based. Pada tahap ini, data pelatihan diberi label berdasarkan kamus sentimen yang telah disusun sebelumnya, yang berisi daftar kata positif dan negatif.

Proses pelabelan dilakukan dengan mencocokkan kata-kata dalam setiap ulasan dengan kata-kata yang terdapat dalam kamus *lexicon*. Jika jumlah kata positif lebih banyak dari kata negatif dalam sebuah ulasan, maka ulasan tersebut diberi label "Positif". Sebaliknya, jika jumlah kata negatif lebih dominan, maka ulasan diberi label "Negatif". Proses pelabelan ini dapat dilihat pada gambar L1.12 yang menampilkan fungsi *lexicon based* dan penghitungan skor sentimen pada setiap data pelatihan. Hasil dari proses pelabelan ditampilkan pada gambar 4.9 yang menunjukkan data ulasan beserta label sentimen masing-masing.

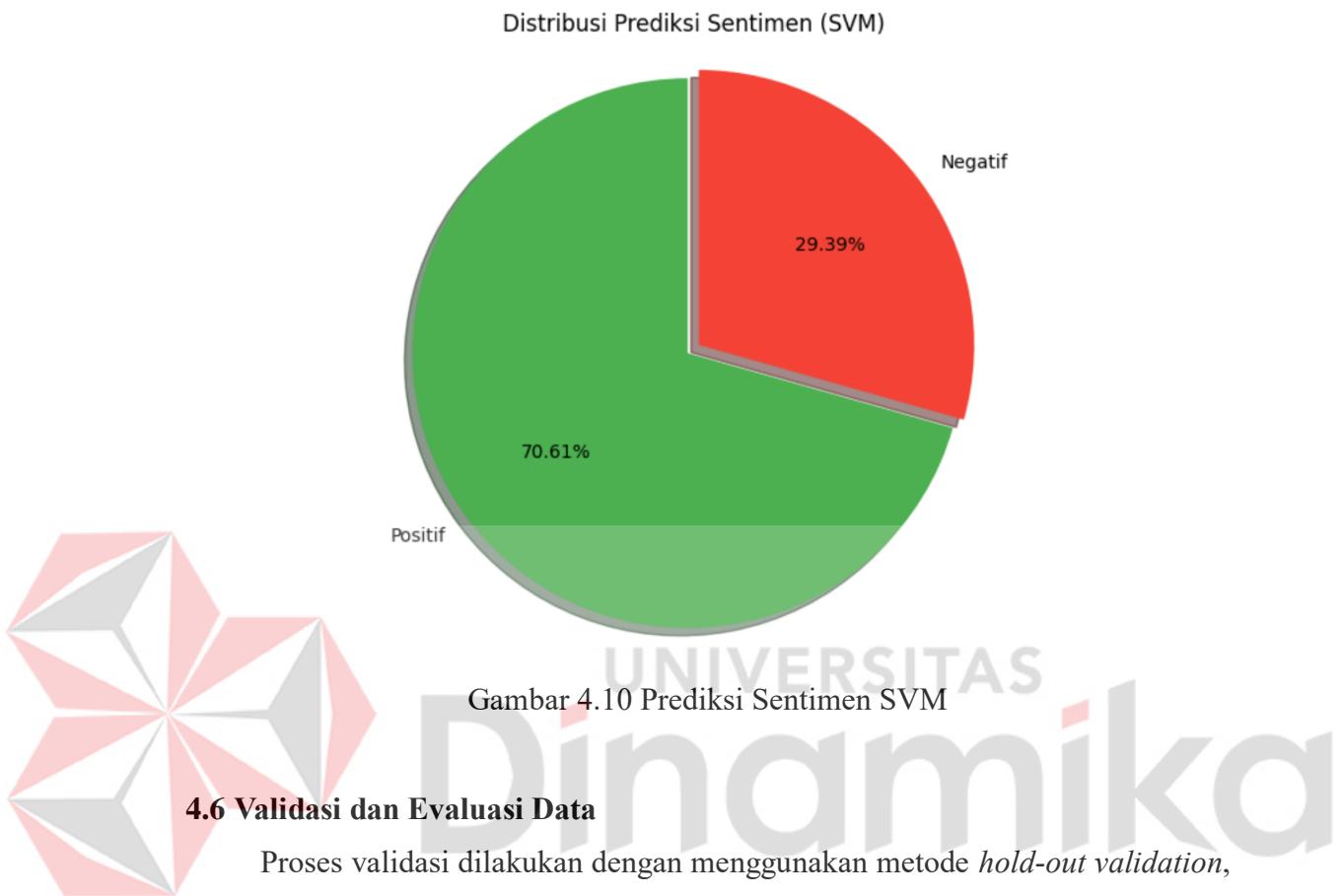
	A	B	C	D	E	F	G	H	I	J	K
1	tokenize, sentiment										
2	['transaksi', 'mudah', 'respon', 'komplain', 'cepat'],	Positif									
3	['first', 'time'],	Negatif									
4	['moga', 'kirim', 'sesuai', 'pesan'],	Positif									
5	['mntap'],	Negatif									
6	['belanja', 'pesan', 'paket', 'posisi', 'kirim', 'ubah', 'ubah', 'mentangmentang'],	Positif									
7	['amanah', 'muas'],	Positif									
8	['produk', 'original', 'ragu', 'belanja', 'mewah'],	Positif									
9	['bantu', 'situasi', 'apa', 'jaga', 'reting', 'bayar'],	Positif									
10	['limit', 'haikin', 'min'],	Negatif									
11	['after', 'sales', 'keren', 'hadap', 'customer', 'proses', 'cepat', 'ongkir', 'mantap', 'banding', 'hijau', 'oyen'],	Positif									
12	['tahun', 'pakai', 'aman', 'lancar', 'jaya', 'kendala'],	Positif									
13	['sukses'],	Positif									

Gambar 4.9 Hasil pelabelan data

4.5 Klasifikasi *Support Vector Machine*

Setelah data ulasan dibagi menjadi data latih yang diberi label dengan *lexicon* dan data uji maka dilakukan klasifikasi menggunakan algoritma *Support Vector Machine* (SVM). Sebelum pelatihan model data teks terlebih dahulu dirubah ke dalam bentuk numerik menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*). Proses ini dilakukan dengan *TfidfVectorizer* dari *library scikit-learn* yang menghasilkan matriks vektor berdasarkan frekuensi dan bobot kata dalam setiap dokumen. Hasil dari vektorisasi ini, yaitu *X_train_vec* dan *X_test_vec*, menjadi input bagi algoritma *Support Vector Machine* (SVM), seperti yang ditunjukkan pada Gambar L1.13. Setelah melalui proses vektorisasi model SVM kemudian proses pelatihan model dilakukan dengan kernel linier (*kernel='linier'*) dan dilakukan prediksi dengan data uji menghasilkan Persentase Sentimen Positif 70.61% dan Persentase Sentimen Negatif 29.39% dari total data

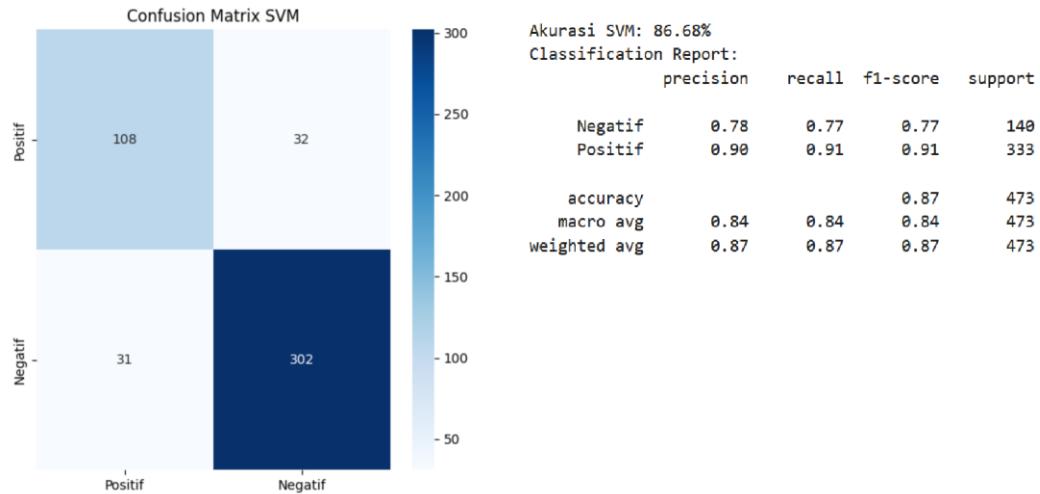
uji yang diprediksi 473 dapat dilihat pada gambar 4.10. Proses pelatihan dan prediksi model ini ditunjukkan pada gambar L1.14.



4.6 Validasi dan Evaluasi Data

Proses validasi dilakukan dengan menggunakan metode *hold-out validation*, dengan membagi dataset menjadi dua bagian, yaitu 70% data latih yang dilabeli dengan *lexicon* dan 30% data uji menggunakan fungsi *train_test_split*, setelah model dilatih menggunakan data latih, model dilakukan validasi dengan data uji untuk mengukur efektivitasnya dalam mengklasifikasikan sentimen.

Berdasarkan hasil validasi, model SVM menghasilkan prediksi dengan distribusi sentimen sebagai berikut: Persentase sentimen positif sebesar 70,61% dan persentase sentimen negatif sebesar 29,39% dari total 473 data uji yang diprediksi. Hasil validasi dapat ditunjukkan pada *confusion matrix* dan *classification report* pada Gambar 4.25 dan untuk penulisan kodenya dapat dilihat pada gambar L1.15.



Gambar 4.11 Hasil validasi Model SVM

Hasil evaluasi terhadap model SVM ditunjukkan melalui beberapa metrik evaluasi, yaitu *accuracy*, *precision*, *recall*, dan *f1-score*, serta *confusion matrix* sebagai gambaran visual performa model dalam memprediksi label sentimen. Model SVM memberikan hasil sebagai berikut:

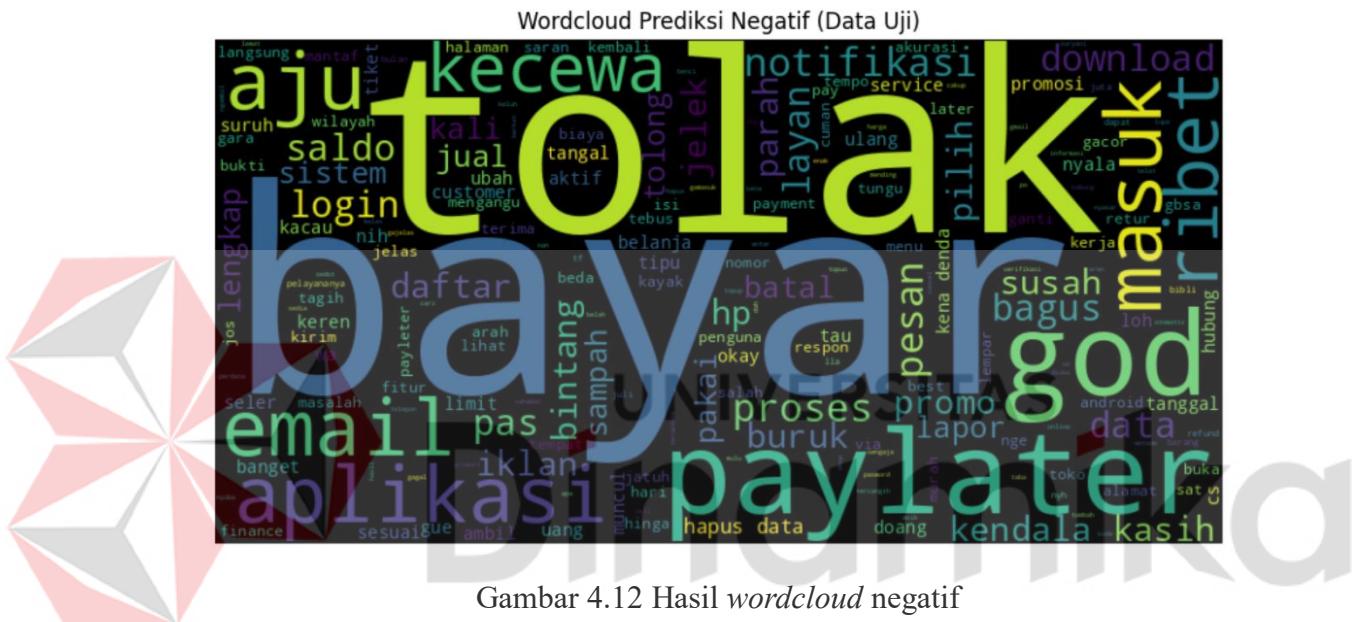
- Akurasi model sebesar 86.68%, artinya sekitar 86% data uji berhasil diklasifikasikan dengan benar.
- Kelas Positif memiliki *precision* 0,90, *recall* 0,91, dan *F1-score* 0,91, menunjukkan performa yang sangat baik dalam mendeteksi ulasan positif.
- Kelas Negatif memiliki *precision* 0,78, *recall* 0,77, dan *F1-score* 0,77, menandakan model masih kurang optimal dalam mengenali ulasan negatif.
- *Macro Average F1-score* sebesar 0,84 dan *Weighted Average F1-score* sebesar 0,87 mengindikasikan bahwa secara keseluruhan model cukup baik, meskipun terdapat ketidakseimbangan performa antar kelas.

4.7 Visualisasi Data

Tahap visualisasi data dilakukan dengan cara menvisualkan data dengan *wordcloud*. Visualisasi *wordcloud* digunakan untuk mengidentifikasi kata-kata yang paling sering muncul dalam ulasan yang telah dianalisis sebelumnya. Ukuran huruf yang bervariasi dalam *wordcloud* merepresentasikan frekuensi kemunculan setiap kata, sehingga memberikan gambaran visual mengenai dominasi istilah dalam data. Langkah awal dalam melakukan visualisasi data diperlukan data

lengkap dari *hasil_labelling_data* dan data dipisahkan berdasarkan kolom *sentiment* menjadi dua bagian yaitu sentimen positif dan sentimen negatif. Proses pembuatan *wordcloud* dapat dilihat pada gambar L1.16.

Pada *wordcloud* negatif terdapat beberapa kata negatif yang sering muncul seperti “ribet”, “tolak”, dan “kecewa” sedangkan untuk *wordcloud* positif terdapat “bagus”, “mantap”, dan “banget”. Gambar 4.12 di bawah ini merupakan hasil *wordcloud* negatif, sedangkan gambar 4.13 merupakan hasil *wordcloud* positif.



Gambar 4.12 Hasil *wordcloud* negatif



Gambar 4.13 Hasil *wordcloud* positif

4.8 Evaluasi

Berdasarkan seluruh tahapan yang telah dilaksanakan dalam penelitian ini menunjukkan bahwa analisis sentimen terhadap ulasan aplikasi Blibli di Google Play Store telah dilakukan secara terstruktur dan komprehensif. Setiap proses mulai dari pengambilan data melalui *data scraping*, tahap *preprocessing data*, pembagian dataset, pelabelan sentimen, pelatihan model klasifikasi menggunakan algoritma *Support Vector Machine* (SVM), validasi serta evaluasi performa model, hingga tahap visualisasi data telah diterapkan dengan metodologi yang sesuai dan didukung oleh teknik yang relevan untuk mencapai hasil yang optimal.

Proses *preprocessing data* yang mencakup normalisasi kata, *stopword removal*, *stemming*, hingga *tokenize* berhasil menghasilkan data teks yang bersih dan siap dianalisis. Penggunaan pendekatan *lexicon-based* dalam pelabelan data latih menunjukkan efektivitas dalam mengklasifikasikan data secara awal sebelum proses klasifikasi menggunakan model pembelajaran mesin.

Algoritma *Support Vector Machine* (SVM) menunjukkan kinerja yang baik dalam klasifikasi sentimen dengan distribusi 70,61% positif dan 29,39% negatif dari 473 data uji. Model mencapai akurasi 86,68% dengan *precision* 90% dan *recall* 91% pada kelas positif. Namun, performa pada kelas negatif lebih rendah (*precision* 78% dan *recall* 77%), yang menunjukkan kesulitan dalam mengenali ulasan negatif. Hal ini disebabkan oleh kendala pada tahap *preprocessing*, seperti ketidak sempurnaan dalam pembersihan dan normalisasi teks, yang memengaruhi kualitas fitur dan menurunkan kemampuan model dalam membedakan sentimen negatif.

Visualisasi data menggunakan wordcloud memperkuat pemahaman terhadap dominasi kata dalam masing-masing sentimen, memberikan gambaran intuitif mengenai persepsi pengguna terhadap aplikasi Blibli. Kata-kata seperti "bagus", "mantap", dan "banget" muncul dominan pada ulasan positif, sementara "ribet", "tolak", dan "kecewa" mendominasi ulasan negatif. Secara keseluruhan, penelitian ini berhasil menjawab tujuan awal dengan menunjukkan bahwa model SVM dapat mengklasifikasikan sentimen dengan cukup baik.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa analisis sentimen terhadap ulasan aplikasi Blibli di Google Play Store menggunakan metode *Support Vector Machine* (SVM) berhasil dilaksanakan dengan tahapan yang sistematis, mulai dari pengambilan data, *preprocessing*, pelabelan data latih menggunakan pendekatan *lexicon-based*, pelatihan model, hingga evaluasi hasil. Model SVM yang diterapkan mampu mengklasifikasikan sentimen dengan akurasi sebesar 86.68%, dengan performa yang lebih baik pada ulasan positif dibandingkan ulasan negatif. *Precision* dan *recall* pada kelas positif masing-masing mencapai 90% dan 91%, menunjukkan kemampuan model yang tinggi dalam mendeteksi ulasan positif. Sebaliknya, kelas negatif memiliki *precision* 78% dan *recall* 77%, menandakan bahwa model masih kesulitan mengenali ulasan negatif. Nilai *Macro Average F1-score* sebesar 84% dan *Weighted Average F1-score* sebesar 87% menunjukkan bahwa model bekerja dengan cukup baik meskipun terdapat ketidakseimbangan performa antar kelas.

Visualisasi wordcloud memperlihatkan kata-kata dominan pada masing-masing sentimen, di mana "*bagus*", "*mantap*", dan "*promo*" banyak muncul pada ulasan positif, sedangkan "*ribet*", "*susah*", dan "*buruk*" mendominasi ulasan negatif. Temuan ini mendukung hasil evaluasi bahwa model SVM efektif memprediksi tingkat kepuasan pelanggan sekaligus memberikan wawasan mengenai aspek yang perlu dipertahankan maupun diperbaiki.

Selain itu, analisis terhadap ulasan negatif mengungkap beberapa permasalahan utama yang perlu menjadi perhatian pengembang aplikasi Blibli. Kata-kata seperti "*bayar*", "*tolak*", dan "*paylater*" menunjukkan adanya kendala pada proses pembayaran, termasuk transaksi yang gagal atau metode pembayaran yang ditolak. Kemunculan kata "*aplikasi*", "*login*", dan "*data*" mengindikasikan adanya permasalahan teknis yang mengganggu pengalaman pengguna, seperti kesulitan masuk atau bug sistem. Kata "*ribet*" dan "*pesan*" menggambarkan bahwa alur pembelian masih dianggap kurang praktis, sementara kata "*mahal*"

memunculkan persepsi harga yang tinggi. Selain itu, kata "*layan*" dan "*jelek*" mencerminkan ketidakpuasan terhadap layanan pelanggan.

5.2 Saran

Berdasarkan hasil penelitian yang telah diperoleh, terdapat beberapa saran yang dapat diberikan sebagai bahan pertimbangan, baik bagi pihak pengembang aplikasi Blibli maupun bagi penelitian selanjutnya.

1. Saran Praktis (untuk Pengembang Aplikasi Blibli)
 - a. Memperbaiki sistem pembayaran, khususnya pada metode seperti PayLater, guna meminimalkan penolakan transaksi yang dapat mengganggu pengalaman pengguna.
 - b. Meningkatkan stabilitas dan performa aplikasi dengan melakukan perbaikan pada masalah login serta bug yang ditemukan.
 - c. Menyederhanakan proses pemesanan agar lebih praktis, cepat, dan mudah dipahami oleh pengguna.
 - d. Meningkatkan kualitas layanan pelanggan melalui pelatihan staf dan percepatan waktu respon terhadap keluhan maupun pertanyaan pengguna.
2. Saran Akademis (untuk Penelitian Selanjutnya)
 - a. Memperluas dataset untuk meningkatkan akurasi, generalisasi model, dan keragaman kosakata *lexicon*.
 - b. Mengembangkan *lexicon* dengan menambah kata positif-negatif melalui sinonim, penyesuaian konteks e-commerce, dan pembobotan kata.
 - c. Mengeksplorasi algoritma *machine learning* lain, seperti *Random Forest*, *Naive Bayes*, LSTM, dan BERT, untuk dibandingkan kinerjanya dengan SVM pada penelitian ini.
 - d. Menerapkan klasifikasi sentimen multikategori, termasuk netral dan emosi spesifik, untuk analisis yang lebih komprehensif.

DAFTAR PUSTAKA

- Ahn, Y., & Lee, J. (2024). The Impact of Online Reviews on Consumers' Purchase Intentions: Examining the Social Influence of Online Reviews, Group Similarity, and Self-Construal. *Journal of Theoretical and Applied Electronic Commerce Research*, 19(2), 1060–1078. <https://doi.org/10.3390/jtaer19020055>
- Arsi, P., & Waluyo, R. (2021). Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 8(1), 147. <https://doi.org/10.25126/jtiik.0813944>
- Aryanti, P. A. N., & Mahendra, I. B. M. (2023). Analisis Sentimen Opini Berbahasa Indonesia Pada Sosial Media Menggunakan TF-IDF dan Support Vector Machine. *Jurnal Elektronik Ilmu Komputer Udayana*, 12(1), 45–52. <https://ojs.unud.ac.id/index.php/jlk/article/view/JLK.2023.v12.i01.p06/48126>
- Ditami, G. R., Ripanti, E. F., & Sujaini, H. (2022). Implementasi Support Vector Machine untuk Analisis Sentimen Terhadap Pengaruh Program Promosi Event Belanja pada Marketplace. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 8(3), 508. <https://doi.org/10.26418/jp.v8i3.56478>
- Fahlevvi, M. R. (2022). Analisis Sentimen Terhadap Ulasan Aplikasi Pejabat Pengelola Informasi Dan Dokumentasi Kementerian Dalam Negeri Republik Indonesia Di Google Playstore Menggunakan Metode Support Vector Machine. *Jurnal Teknologi Dan Komunikasi Pemerintahan*, 4(1), 1–13. <https://doi.org/10.33701/jtkp.v4i1.2701>
- Firdaus, A., & Firdaus, W. I. (2021). Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan). *Jurnal JUPITER*, 13(1), 66.
- Irfani, F. F., Triyanto, M., & Hartanto, A. D. (2020). Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma Support Vector Machine. *Jurnal Bisnis, Manajemen Dan Informatika Analisis*, 16, 258–266.
- Isnain, A. R., Sakti, A. I., Alita, D., & Marga, N. S. (2021). Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma

- SVM. *JDMSI*, 2(1), 31–37.
- Khairunnisa, S., & Faraby, S. Al. (2021). *Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi. 5(April), 406–414. <https://doi.org/10.30865/mib.v5i2.2835>*
- Machová, K., Mikula, M., Gao, X., & Mach, M. (2020). Lexicon-based sentiment analysis using particle swarm optimization. *Electronics (Switzerland)*, 9(8), 1–22. <https://doi.org/10.3390/electronics9081317>
- Pudjiantoro, T. H., Umbara, F. R., & Trihatmoko, B. (2021). Analisis Sentimen Terhadap E-commerce Pada Media Sosial Twitter Menggunakan Metode Naïve bayes. *Snia.Unjani.Ac.Id*, 1–7.
- Rahma, Intan Larasati Anisa, Shafira Nurin Qolbi, A. S. Z. (2022). *EXPLORATORY DATA ANALYSIS Analysis of Google Playstore Datasets Using ANALISIS DATASET GOOGLE PLAYSTORE MENGGUNAKAN METODE EXPLORATORY DATA ANALYSIS Analysis of Google Playstore Datasets Using Exploratory Data Analysis Methods. December.*
- Saraswati, M., & Riminarsih, D. (2020). Analisis Sentimen Terhadap Pelayanan Krl Commuterline Berdasarkan Data Twitter Menggunakan Algortima Bernoulli Naive Bayes. *Jurnal Ilmiah Informatika Komputer*, 25(3), 225–238. <https://doi.org/10.35760/ik.2020.v25i3.3256>
- Sulaiman, F. I., Winarno, W. W., & Kurniawan, M. P. (2021). *Perancangan Aplikasi Klasifikasi Sentimen Berbasis Web Terhadap Mata Uang Kripto.*
- Utami, N. W., & Artana, M. (2022). *Text Mining Dalam Analisis Sentimen Pembelajaran Daring Di Masa Pandemi Covid 19 Menggunakan Algoritma K-Nearest Neighbor. 4(1), 59–64.*
- Wulandary, E., Mas'ud, M., Arifin, & Ashoer, M. (2023). Pengaruh Kualitas Layanan Era Digitalisasi Media Pemasaran Online E-Commerce dengan Menggunakan E-Service Quality Untuk Mengetahui Kepuasan Pelanggan. *Center of Economic Students Journal*, 6(1), 44–59. <https://doi.org/10.56750/csej.v6i1.568>