

**ANALISIS SENTIMEN MENGGUNAKAN ALGORITMA
NAIVE BAYES PADA ULASAN PRODUK *E-COMMERCE* (Studi
Kasus: Tokopedia)**

SKRIPSI

**OLEH:
MOCHAMMAD ALDI YUSUF
NIM. 19610009**



**PROGAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM MAULANA MALIK IBRAHIM MALANG
2025**

**ANALISIS SENTIMEN MENGGUNAKAN ALGORITMA
NAIVE BAYES PADA ULASAN PRODUK *E-COMMERCE* (Studi
Kasus: Tokopedia)**

SKRIPSI

**Diajukan Kepada
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
untuk Memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Matematika (S.Mat)**

**Oleh
MOHAMMAD ALDI YUSUF
NIM. 19610009**

**PROGAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM MAULANA MALIK IBRAHIM MALANG
2025**

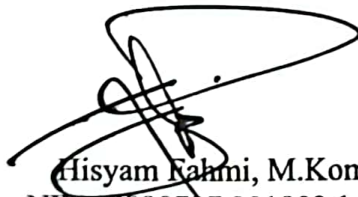
**ANALISIS SENTIMEN MENGGUNAKAN ALGORITMA
NAIVE BAYES PADA ULASAN PRODUK *E-COMMERCE* (Studi
Kasus: Tokopedia)**

SKRIPSI

Oleh
Mochammad Aldi Yusuf
NIM. 19610009

Telah Disetujui Untuk Diuji
Malang, 13 Juni 2025

Dosen Pembimbing I


Hisyam Fahmi, M.Kom.
NIP. 19890727 201903 1 018

Dosen Pembimbing II


Mohammad Nafie Jauhari, M.Si.
NIPPPK. 19870218 202321 1 018

Mengetahui,
Ketua Progam Studi Matematika



Dr. Elly Susanti, M.Sc
NIP. 19741129 200012 2 005

**ANALISIS SENTIMEN MENGGUNAKAN ALGORITMA
NAIVE BAYES PADA ULASAN PRODUK *E-COMMERCE* (Studi
Kasus: Tokopedia)**

SKRIPSI

Oleh
Mochammad Aldi Yusuf
NIM. 19610009

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima sebagai Salah Satu Persyaratan
untuk Memperoleh Gelar Sarjana Matematika (S.Mat)

Tanggal 17 Juni 2025

Ketua Penguji	:	Dr. Mohammad Jamhuri, M. Si.	 
Anggota Penguji 1	:	Muhammad Khudzaifah, M. Si.	 
Anggota Penguji 2	:	Hisyam Fahmi, M.Kom.	 
Anggota Penguji 3	:	Muhammad Nafie Jauhari, M.Si.	 

Mengetahui,
Ketua Program Studi Matematika


Dr. Elly Susanti, M.Sc
NIP. 19741129 200012 2 005

HALAMAN PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Nama : Mochammad Aldi Yusuf
NIM : 19610009
Progam Studi : Matematika
Fakultas : Sains dan Teknologi
Judul Skripsi : Analisis Sentimen Menggunakan Algoritma Naïve Bayes
Pada Ulasan Produk E-Commerce (Studi Kasus:
Tokopedia)

Menyatakan dengan sebenar-benarnya Selain mencantumkan nama penulis asli dalam daftar pustaka, saya menyatakan dengan tegas bahwa skripsi ini sepenuhnya merupakan hasil karya orisinal saya dan tidak merupakan gabungan dari informasi, karya, atau gagasan pihak lain. Saya siap bertanggung jawab atas segala konsekuensi yang timbul jika di kemudian hari terbukti atau dapat dibuktikan bahwa tesis ini mengandung unsur jiplakan.

Malang, 17 Juni 2025

Yang Membuat Pernyataan,



Mochammad Aldi Yusuf
NIM. 19610009

MOTTO

“Sak sibuk-sibuk e penggawean, ojo lali karo SHOLAT e”
(Almaghfurlah RKH. Muhammad Badruddin Anwar)

“Semua Orang Memiliki Gilirannya Masing- Masing”
(Gold D Roger - One Piece)

“Mengapa Aku Harus Merubah Sikapku Hanya Karena Kamu Orang Penting”
(Monkey D Luffy – One Piece)

PERSEMBAHAN

Dengan ucapan Bismillahirrahmanirrahim, segala puji bagi Allah SWT dan dengan penuh keikhlasan, penulis mempersembahkan tulisan ini kepada: Sembah sujud serta rasa syukur kepada Allah SWT, atas limpahan Rahman, Rahim-Nya yang telah memberikan kekuatan, semangat dan hidayah kepada penulis sehingga skripsi ini dapat terselesaikan dengan lancar dan baik.

Kedua orang tua penulis Bapak tersayang Suseno dan Ibu tercinta Sulistyoningsih yang telah membantu, mensupport, dan mendoakan penulis di sepanjang waktu.

Karya sederhana ini saya persembahkan dengan tulus kepada Bapak dan Ibu.

Karya ini juga saya persembahkan kepada seluruh saudara, teman, dan kerabat yang telah membantu serta memberikan dukungan hingga karya ini dapat diselesaikan.

Para dosen pembimbing, para dosen yang telah membagi ilmunya sehingga penulis dapat menyelesaikan karya ini.

Dan terakhir kepada Mochammad Aldi Yusuf, sang penulis yang terus bertahan dan terus melangkah walaupun kadang terhenti, terima kasih dan tetaplah terus berjuang.

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Puji Syukur kehadiran Allah Swt, atas segala limpahan rahmat dan karunia-Nya serta hidayah-Nya sehingga penulis dapat menyelesaikan draf skripsi yang berjudul “Analisis Sentimen Menggunakan Algoritma *Naive Bayes* Pada Ulasan Produk *E-commerce* (Studi Kasus: Tokopedia)”. Shalawat serta salam mudah-mudahan senantiasa tetap tercurahkan kepada Nabi Muhammad SAW, yang telah membawa dari jalan gelap gulita yakni era jahiliyah menuju jalan yang terang benderang, yakni *ad-dinul Islam* (agama islam).

Skripsi ini dibuat oleh peneliti untuk memenuhi salah satu syarat untuk memperoleh gelar sarjana matematika dari Progam Studi Matematika Fakultas Sains dan Teknologi Universitas Islam Maulana Malik Ibrahim Malang. Peneliti mengucapkan terima kasih kepada:

1. Prof. Dr. H. M. Zainuddin, MA. selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Sri Harini, M.Si. selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Elly Susanti, M.Sc. selaku ketua Progam Studi Matematika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Hisyam Fahmi, M.Kom. selaku dosen pembimbing I yang telah memberikan berbagai pengetahuan, motivasi, nasihat, serta arahan kepada peneliti.
5. M. Nafie Jauhari, M.Si. selaku dosen pembimbing II yang telah memberikan berbagai pengetahuan, motivasi, nasihat, serta arahan kepada peneliti.
6. Dr. Mohammad Jamhuri, M. Si. selaku ketua penguji yang telah memberikan berbagai pengetahuan, motivasi, nasihat, serta arahan kepada peneliti.
7. Muhammad Khudzaifah, M. Si. selaku penguji I yang telah memberikan berbagai pengetahuan. Motivasi, nasihat, serta arahan kepada peneliti.

8. Seluruh dosen Progam Studi Matematika yang telah membimbing sejak awal semester hingga terbentuknya skripsi.
9. Orang tua dan seluruh keluarga yang senantiasa mendoakan, memberi nasihat, dukungan, semangat, kasih sayang sehingga penulis dapat menyelesaikan tugas akhir.
10. Teman-teman alumni yang selalu memberikan dorongan berupa ucapan dan bantuan dalam berbagai keadaan.

Semoga Allah Swt senantiasa memberikan balasan atas segala kebaikan yang ditujukan kepada peneliti. Harapan kedepannya, baik akademisi dan pembaca akan menganggap skripsi ini berharga dan dapat menambah pengetahuan.

Wassalamu 'alaikum Warahmatullahi Wabarakatuh

Malang, 17 Juni 2025

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN.....	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERNYATAAN KEASLIAN TULISAN.....	v
MOTTO	vi
PERSEMBAHAN	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN	xiv
ABSTRAK	xv
ABSTRACT.....	xvi
مستخلص البحث	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	5
1.3 Tujuan Penelitian	5
1.4 Manfaat Penelitian	5
1.5 Batasan Masalah	6
1.6 Definisi Istilah	6
BAB II KAJIAN PUSTAKA	8
2.1 <i>Data Mining</i>	8
2.2 <i>Text Mining</i>	9
2.2.1 Analisis Sentimen	10
2.2.2 <i>Data Cleansing</i>	10
2.2.3 <i>Text Preprocessing</i>	11
2.2.4 Pembobotan Kata (<i>Term Weigthing</i>)	12
2.3 Klasifikasi	13
2.3.1 <i>Naive Bayes Classifier</i>	14
2.4 <i>K-Fold Cross Validation</i>	17
2.5 <i>Online Review</i>	17
2.6 Analisis Sentimen Menurut Pandangan Islam	18
BAB III METODE PENELITIAN	21
3.1 Jenis Penelitian	21
3.2 Data dan Sumber Data	21
3.3 Teknik Analisis Data	22
3.3.1 Analisis Sentimen	22
3.3.2 <i>Confusion Matrix</i>	24
BAB IV HASIL DAN PEMBAHASAN.....	26
4.1 Data Penelitian.....	26
4.2 <i>Data Cleansing</i>	28
4.3 <i>Text Preprocessing</i>	30
4.3.1 <i>Casefolding</i>	30
4.3.2 <i>Cleansing dan Normalization</i>	31

4.3.3	<i>Stopword Removal</i>	32
4.3.4	<i>Tokenizing</i>	33
4.3.5	<i>Stemming</i>	34
4.4	<i>Term Weigthing (TF-IDF)</i>	36
4.4.1	Perhitungan <i>Term Frequency (TF)</i>	36
4.4.2	Perhitungan <i>Inverse Document Frequency (IDF)</i>	37
4.4.3	Perhitungan <i>TF-IDF</i>	38
4.5	<i>Training Data</i>	38
4.5.1	<i>Multinomial Naive Bayes</i>	39
4.5.2	<i>K-Fold Cross Validation</i>	40
4.6	<i>Testing Data</i>	41
4.6.1	<i>Confusion Matrix</i>	42
4.7	Kajian Keislaman	47
BAB V PENUTUP		49
5.1	Kesimpulan	49
5.2	Saran	50
DAFTAR PUSTAKA		51
LAMPIRAN		54
RIWAYAT HIDUP		55

DAFTAR TABEL

Tabel 3.1	Variabel dan Definisi Variabel.....	22
Tabel 3.2	Konsep <i>Confusion Matrix</i>	24
Tabel 4.1	Data Sampel Ulasan Produk.....	27
Tabel 4.2	<i>Pseudocode Data Cleansing</i>	29
Tabel 4.3	<i>Pseudocode Casefolding</i>	30
Tabel 4.4	<i>Pseudocode Cleansing dan Normalization</i>	31
Tabel 4.5	<i>Pseudocode Stopword Removal</i>	33
Tabel 4.6	<i>Pseudocode Tokenizing</i>	33
Tabel 4.7	<i>Pseudocode Stemming</i>	34
Tabel 4.8	Hasil <i>Text Preprocessing</i>	35
Tabel 4.9	Contoh Perhitungan <i>Term Frequency (TF)</i>	37
Tabel 4.10	Hasil Perhitungan TF-IDF.....	38
Tabel 4.11	Sampel Dokumen <i>Training</i>	39
Tabel 4.12	Hasil <i>K-Fold Cross Validation</i>	41
Tabel 4.13	Nilai Akurasi <i>Testing Naive Bayes</i>	46
Tabel 4.14	Nilai Akurasi <i>Testing Random Forest</i>	46

DAFTAR GAMBAR

Gambar 4.1 Presentase Data Ulasan Produk.....	26
Gambar 4.2 Hasil <i>Confusion Matrix Naive Bayes</i>	42
Gambar 4.3 Hasil <i>Confusion Matrix Random Forest</i>	43

DAFTAR LAMPIRAN

Lampiran 1 Perhitungan Manual TF-IDF	54
Lampiran 2 Perhitungan Manual Naive Bayes	54

ABSTRAK

Yusuf, Mochammad Aldi. 2025. **Analisis Sentimen Menggunakan Algoritma Naive Bayes Pada Ulasan Produk E-Commerce (Studi Kasus: Tokopedia)**. Skripsi. Progam Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Hisyam Fahmi, M.Kom. (II) Mohammad Nafie Jauhari, M.Si.

Kata Kunci: Analisis Sentimen, Tokopedia, *Naive Bayes*, *Random Forest*, TF-IDF, *e-commerce*, ulasan produk

Tokopedia merupakan salah satu e-commerce di Indonesia yang menyediakan peluang layanan bisnis dengan memanfaatkan jaringan internet. Tokopedia menyediakan berbagai macam kategori produk yang sangat menarik, namun tidak semua kategori produk memiliki nilai yang baik dari sudut pandang konsumen. Setiap konsumen dapat meninggalkan ulasan atau komentar terhadap produk yang dibeli. Meningkatnya ulasan produk yang diberikan oleh konsumen telah menciptakan tantangan baru bagi perusahaan dan membutuhkan pendekatan analisis data berupa *text mining* yang dapat melihat ulasan atau pendapat yang dikeluhkan oleh konsumen. Pada penelitian ini, pendekatan analisis data yang dilakukan adalah analisis sentimen menggunakan algoritma *Naive Bayes* dan *Random Forest*. Validasi model dilakukan menggunakan *K-Fold Cross Validation* dengan nilai $k = 5$ serta menggunakan teknik *Laplacian Correction* untuk model *Naive Bayes*. Hasil akurasi tertinggi dari *K-Fold Cross Validation* digunakan untuk model terbaik dalam pengujian data testing. Berdasarkan hasil pengujian pada penelitian ini, algoritma *Naive Bayes* memiliki nilai akurasi lebih tinggi dibandingkan dengan *Random Forest* dengan nilai akurasi sebesar 92,49%.

ABSTRACT

Yusuf, Mochammad Aldi. 2025. **Sentiment Analysis Using Naive Bayes Algorithm on E-Commerce Product Reviews (Case Study: Tokopedia)**. Undergraduate Thesis. Mathematics Departement, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Advisors: (I) Hisyam Fahmi, M.Kom. (II) Mohammad Nafie Jauhari, M.Si.

Kata Kunci: Sentiment Analysis, tokopedia, Naive Bayes, Random Forest, TF-IDF, e-commerce, Product Reviews

Tokopedia is one of the leading e-commerce platforms in Indonesia that provides business opportunities through internet-based services. Although it offers a wide range of attractive product categories, not all products receive favorable responses from consumers. Customers are able to leave reviews or comments on products they have purchased. The growing volume of user reviews poses a challenge for companies and necessitates a data analysis approach such as text mining to extract consumer feedback and complaints. In this study, sentiment analysis was conducted using the Naive Bayes and Random Forest algorithms. Model validation was carried out using K-Fold Cross Validation with $k = 5$, and Laplacian Correction was applied to the Naive Bayes model. The highest accuracy score from the K-Fold Cross Validation was used to select the best model for testing. Based on the test results, the Naive Bayes algorithm achieved a higher accuracy than the Random Forest algorithm, with an accuracy rate of 92.49%.

مستخلص البحث

يوسف، محمد أدي. ٢٠٢٥. تحليل المشاعر باستخدام خوارزمية بايز الساذجة في مراجعات منتجات التجارة الإلكترونية (دراسة حالة: توكويديا). البحث اجلامعي. قسم الرياضيات، كلية العلوم والتكنولوجيا، جامعة مولان مالك إبراهيم الإسلامية الحكومية مالنح. املشرف: ١. هشام فهمي، ماجستير في علوم الحاسوب ٢. محمد نافع جوهري، ماجستير في العلوم

الكلمات المفتاحية: تحليل المشاعر، توكويديا، بايز الساذج، الغابة العشوائية، TF-IDF، التجارة الإلكترونية، مراجعات المنتجات

توكويديا هي إحدى منصات التجارة الإلكترونية في إندونيسيا التي توفر فرص خدمات الأعمال من خلال الاستفادة من شبكة الإنترنت. توفر توكويديا مجموعة متنوعة من فئات المنتجات الجذابة للغاية، ولكن ليست كل فئات المنتجات ذات قيمة جيدة من وجهة نظر المستهلك. يمكن لكل مستهلك ترك مراجعة أو تعليق على المنتج الذي تم شراؤه. أدى العدد المتزايد من مراجعات المنتجات التي يقدمها المستهلكون إلى خلق تحديات جديدة للشركات ويتطلب نهجًا لتحليل البيانات في شكل استخراج النصوص الذي يمكنه رؤية المراجعات أو الآراء التي اشتكى منها المستهلكون. في هذه الدراسة، فإن نهج تحليل البيانات المستخدم هو تحليل المشاعر باستخدام خوارزميات Naive Bayes و Random Forest. يتم التحقق من صحة النموذج باستخدام التحقق المتقاطع K-Fold بقيمة $k = 5$ واستخدام تقنية التصحيح لابلان لنموذج Naive Bayes. يتم استخدام أعلى نتائج الدقة من التحقق المتقاطع K-Fold لأفضل نموذج في اختبار البيانات. بناءً على نتائج الاختبار في هذه الدراسة، تتمتع خوارزمية Naive Bayes بقيمة دقة أعلى مقارنة بالغابة العشوائية بقيمة دقة تبلغ 92.49٪.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Tokopedia merupakan salah satu *e-commerce* di Indonesia yang menyediakan peluang layanan bisnis atau jual beli dengan memanfaatkan jaringan internet. *E-commerce* sendiri adalah bentuk dari digitalisasi ekonomi yang memudahkan perusahaan atau manusia dalam melakukan kegiatan jual beli yang memaksimalkan penggunaan internet (Nur'aini, 2023). Tokopedia.com merupakan sebuah perusahaan *e-commerce* yang didirikan oleh William Tanuwijaya dan Leontinus Alpha Edison pada 6 Februari 2009 dan resmi diperkenalkan kepada public pada bulan Agustus di tahun yang sama (Novendra, 2022). Tokopedia merupakan *e-commerce* yang sangat diminati oleh Masyarakat Indonesia. Berdasarkan data iPrice Group (2024, Januari 19) kuartal II (Q2) 2022, Tokopedia secara konsisten menduduki peringkat teratas sebagai platform *e-commerce* paling populer di Indonesia, dengan rata-rata kunjungan bulanan mencapai angka 158.346.667 (iPrice, 2022). Hal ini dikarenakan Tokopedia menyediakan berbagai macam kategori produk yang sangat menarik seperti computer, laptop, otomotif, makanan, minuman, mainan, perlengkapan rumah tangga, perlengkapan bayi dan lain-lain. Meskipun kategori produk yang ditawarkan oleh Tokopedia sangat banyak dan bervariasi namun tidak semua produk memiliki nilai yang baik dari sudut pandang konsumen atau pelanggan.

Konsumen atau pelanggan dapat memberikan penilaian berupa ulasan produk, rating maupun *emoticon* sebagai nilai kepuasan yang diperoleh dari pembelian salah satu produk yang ditawarkan. Penilaian berupa ulasan produk yang

diberikan dapat menjadi sebuah informasi bagi calon pelanggan yang akan melakukan pembelian. Menurut (Prasetyo & Adriyanto, 2022), ulasan produk merupakan informasi yang sangat penting bagi calon pelanggan dalam pengambilan keputusan, hal ini disebabkan oleh fakta bahwa ulasan pelanggan berperan penting dalam membantu konsumen untuk menemukan produk atau layanan yang paling sesuai dengan kebutuhan mereka, serta memberikan gambaran yang lebih jelas tentang kualitas produk tersebut berdasarkan pengalaman langsung konsumen lainnya. Ulasan produk terkadang berisi tentang pendapat konsumen yang puas maupun kurang puas terhadap produk yang telah dibeli (Fahmi & Sari, 2021).

Seiring dengan meningkatnya jumlah pengguna aktif dan transaksi yang terjadi di Tokopedia, jumlah ulasan produk yang terkumpul juga mengalami pertumbuhan yang eksponensial. Meningkatnya jumlah ulasan produk telah menciptakan tantangan bagi perusahaan dan produsen dalam mengelola dan menganalisis data ulasan produk, karena dapat berdampak negatif pada pengambilan keputusan bisnis, seperti pengembangan produk baru atau perbaikan layanan pelanggan. Oleh karena itu, perlu dilakukan upaya untuk mempercepat dan meningkatkan akurasi proses analisis data. Teknik pendekatan atau metode analisis yang dapat dipakai adalah *text mining*. *Text mining* dapat dimanfaatkan untuk melihat ulasan atau pendapat apa saja yang dikeluhkan dan sering dibahas oleh konsumen. Salah satu metode analisa *text mining* yang dapat diaplikasikan dalam sebuah perusahaan atau produsen yang mengeluarkan produk atau jasa yaitu analisis sentimen. Teknik ini diaplikasikan karena dapat mengelompokkan pendapat (*feedback*) dari positif, negatif, dan netral sehingga dapat mempercepat

tugas perusahaan atau produsen dalam meninjau kembali kelebihan dan kekurangan produk yang telah mereka jual.

Terdapat penelitian sebelumnya mengenai analisis sentiment yang dilakukan oleh (Wahyu Sejati et al., 2023). Penelitian ini menganalisis sentimen ulasan konsumen terhadap produk di Shopee dengan menggunakan metode *Naive Bayes*. Hasil penelitian menunjukkan bahwa tingkat akurasi pada pengujian dengan 5 kelas, menggunakan 80% data latih dan 20% data uji, adalah sebesar 52,66%. Sementara itu, pada pengujian dengan 3 kelas, yang menggunakan 90% data latih dan 10% data uji, akurasi tertinggi yang diperoleh mencapai 77,78%. Adapun penelitian lain yang dilakukan oleh (Indrayuni, 2019) yang membahas tentang analisis sentimen terhadap produk kecantikan menggunakan metode *Naive Bayes*, algoritma yang digunakan mampu menghasilkan tingkat akurasi yang tinggi hingga 90,50% dengan nilai AUC sebesar 0.715 dalam klasifikasi sentimen review produk kosmetik berbahasa Indonesia dengan menambahkan penerapan generate N-gram pada tahap *preprocessing*. Adapula penelitian lain yang dilakukan oleh (Stephenie et al., 2020) yang membahas tentang analisis sentimen terhadap ulasan produk di platform *e-commerce* Tokopedia menggunakan model klasifikasi *random forest*, dengan parameter optimal berupa *mtry* sebanyak 73 dan *ntree* sebanyak 50 serta menggunakan TF-IDF dan pelabelan sentimen menggunakan pendekatan *lexicon-based*. Model pada penelitian ini menggunakan *K-Fold Cross Validation* dengan nilai $k = 10$ dan *confusion matrix*, dengan hasil akurasi tertinggi mencapai 97,38%

Penerapan analisis sentimen menggunakan algoritma *Multinomial Naive Bayes* akan dilakukan pada penelitian ini untuk mengevaluasi efektifitas nilai akurasi metode *Naive Bayes* dalam menganalisis sentimen konsumen terhadap

produk atau jasa, serta mengansumsikan kata-kata dalam sebuah dokumen atau teks yang dianggap mengikuti distribusi multinomial yaitu frekuensi kemunculan kata-kata dihitung dan digunakan sebagai fitur untuk klasifikasi (Abbas et al., 2019). Serta mengidentifikasi aspek-aspek apa saja yang harus diperhatikan oleh pihak perusahaan dalam meningkatkan produk atau layanan yang mereka berikan serta factor-faktor apa saja yang dapat mempengaruhi keputusan pemilihan oleh konsumen. Hal ini telah diterangkan dalam firman Allah Swt surat An-Nisa' [4] ayat 29 yang telah ditashih oleh Ibnu Hibban yang berbunyi:

يَا أَيُّهَا الَّذِينَ آمَنُوا لَا تَأْكُلُوا أَمْوَالَكُمْ بَيْنَكُمْ بِالْبَاطِلِ إِلَّا أَنْ تَكُونَ تِجَارَةً عَنْ تَرَاضٍ مِّنْكُمْ
وَلَا تَقْتُلُوا أَنْفُسَكُمْ إِنَّ اللَّهَ كَانَ بِكُمْ رَحِيمًا

“Wahai orang-orang yang beriman, janganlah kamu memakan harta sesamamu dengan cara yang batil (tidak benar), kecuali berupa perniagaan atas dasar suka sama suka diantara kamu. Janganlah kamu membunuh dirimu. Sesungguhnya Allah Swt adalah Maha Penyayang kepadamu.”

Ayat di atas menerangkan tentang hakikat jual beli yang harus dilakukan dengan adanya suka sama suka (ridho) antara penjual dan pembeli dengan tujuan agar terhindar dari perkara batil seperti terjadinya riba, sehingga ketika didalam transaksi ada perkara batil didalamnya maka transaksi tersebut merupakan tindakan batil dan tidak dapat dijadikan landasan dalam dunia bisnis. Namun ridho tidak dapat diketahui oleh manusia karena ridho tersembunyi dalam hati dan hanya dapat diketahui oleh Allah Swt.

Berdasarkan pemaparan latar belakang dan tafsir diatas, penelitian ini bertujuan untuk mengevaluasi efektivitas metode klasifikasi *Naive Bayes* dalam menganalisis sentimen ulasan pelanggan atau konsumen terhadap produk

berdasarkan ulasan mereka. Melalui analisis ini, diharapkan dapat mengidentifikasi aspek-aspek produk yang perlu ditingkatkan serta mengukur tingkat kepuasan baik dari penjual maupun pembeli. Sehingga dirumuskan penelitian yang berjudul “Analisis Sentimen Menggunakan Algoritma *Naive Bayes* pada Ulasan Produk *E-Commerce* (Studi Kasus: Tokopedia)”.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang sebelumnya, maka diperoleh rumusan masalah sebagai berikut:

1. Bagaimana hasil analisis sentimen terhadap ulasan produk di Tokopedia menggunakan metode *Naive Bayes*?
2. Bagaimana evaluasi performa analisis sentiment terhadap ulasan produk di Tokopedia menggunakan metode *Naive Bayes*?

1.3 Tujuan Penelitian

Berdasarkan uraian rumusan masalah sebelumnya, sehingga didapatkan tujuan masalah sebagai berikut:

1. Untuk menilai seberapa baik algoritma *Naive Bayes* dalam melakukan analisis sentimen terhadap ulasan produk di Tokopedia.
2. Untuk mengevaluasi akurasi dan efektivitas metode *Naive Bayes* dalam analisis sentimen terhadap ulasan produk Tokopedia.

1.4 Manfaat Penelitian

Adapun manfaat yang dapat diperoleh dalam penelitian yaitu:

1. Bagi Penulis

Menambah wawasan dan mengimplementasikan ilmu yang didapatkan dari analisis sentimen menggunakan metode *Naive Bayes* pada data ulasan produk e-commerce (studi kasus: Tokopedia).

2. Bagi Progam Studi

Menjadi sumber rujukan pembelajaran bagi mahasiswa mengenai topik analisis sentimen.

3. Bagi Instansi

Membantu perusahaan dalam mengidentifikasi pandangan pengguna terhadap produk yang ditawarkan melalui ulasan baik maupun buruk, sehingga dapat digunakan sebagai acuan untuk meningkatkan kualitas produk dan memperbaiki kelemahan menuju perbaikan yang lebih baik.

1.5 Batasan Masalah

Adapun batasan masalah pada penelitian ini yaitu:

1. Data yang digunakan terdiri dari review konsumen dan sentimen yang berjumlah 5.400 data.
2. Review konsumen yang digunakan untuk klasifikasi adalah review konsumen yang berbahasa Indonesia.

1.6 Definisi Istilah

Berikut adalah beberapa istilah yang digunakan dalam penelitian ini:

<i>Preprocessing</i>	: Proses pengolahan data yang bertujuan untuk mengonversi data mentah menjadi pola yang lebih mudah dimengerti oleh sistem.
<i>Data Training</i>	: Data yang digunakan untuk melatih algoritma program agar dapat menghasilkan model yang tepat.
<i>Data Testing</i>	: Data yang digunakan untuk menguji kinerja model dan menilai tingkat keakuratannya.
<i>Data Cleansing</i>	: Proses menemukan nilai yang hilang, membersihkan <i>noise values</i> , mengidentifikasi <i>outlier</i> , duplikasi dalam data dan memperbaiki data yang tidak konsisten.
<i>Casefolding</i>	: Mengubah semua teks menjadi huruf kecil (lowwercease) atau huruf besar (uppercase).
<i>Cleansing</i>	: Proses menghapus komponen-komponen yang tidak penting atau komponen yang tidak ada kaitannya dengan dokumen.
<i>Normalization</i>	: Proses mengubah kata yang tidak baku menjadi baku, <i>noise</i> , dan kata yang salah eja.
<i>Stopword</i>	: Proses menghapus kata-kata yang umum dan tidak memiliki makna.
<i>Tokenizing</i>	: Proses memecah teks menjadi beberapa kata-kata penyusunnya atau token.
<i>Stemming</i>	: Proses mengubah kata-kata menjadi kata dasarnya.

BAB II

KAJIAN PUSTAKA

2.1 *Data Mining*

Data mining merupakan proses menemukan sebuah informasi baru pada sebuah *database*. *Data mining* adalah suatu pendekatan yang diterapkan untuk mengeksplorasi dan mengidentifikasi informasi atau pengetahuan baru dari kumpulan data yang rumit, dengan memanfaatkan berbagai teknik analisis data, termasuk metode statistik, matematika, kecerdasan buatan, dan *machine learning* (Nimri et al., 2022). Menurut (Kamber & Han, 2006), *data mining* merupakan proses mengidentifikasi suatu informasi untuk menemukan sebuah pengetahuan baru seperti asosiasi, anomali, pola, perubahan, dan struktur signifikan dalam data besar yang tersimpan dalam *database*. *Data mining* juga merupakan bagian dari *Knowledge Discovery in Database* (KDD).

Menurut (Kamber & Han, 2006), *Knowledge Discovery in Database* (KDD) memiliki beberapa langkah sebagai berikut:

1. *Data cleaning* merupakan tahapan yang bertujuan untuk mengidentifikasi dan memperbaiki kesalahan atau ketidakkonsistenan dalam data.
2. *Data integration* merupakan usaha untuk menyatukan informasi dari berbagai sumber agar membentuk satu kumpulan data yang terorganisir.
3. *Data selection* merupakan tahapan dalam data mining yang berfokus pada pengambilan subset data penting dari set data yang lebih besar.
4. *Data transformation* merupakan tahapan yang melibatkan modifikasi data agar sesuai dengan struktur data yang optimal untuk keperluan analisis.

5. *Data mining* merupakan tahapan yang bertujuan untuk mengidentifikasi pola, tren, atau informasi penting dari data dalam jumlah besar.
6. *Pattern evaluation* merupakan tahapan yang bertujuan untuk menentukan seberapa efektif pola atau model yang telah diidentifikasi dalam mempresentasikan data.
7. *Knowledge representation* adalah proses yang bertujuan untuk menyimpan dan mengorganisir informasi yang diperoleh dari data mining dan digunakan untuk pengambilan keputusan dan analisis lebih lanjut.

2.2 Text Mining

Text mining juga disebut sebagai *text data mining*. Karena, dirancang untuk menemukan atau memperoleh pengetahuan implisit yang tersembunyi dalam teks yang tidak terstruktur (Khotimah et al., 2019). Menurut (Sun et al., 2018), *text mining* atau *text analytics* merupakan proses pengolahan data teks yang bertujuan untuk menganalisis data teks semi-terstruktur atau tidak terstruktur. Hal yang membedakan antara *text mining* dan *data mining* terletak pada data yang digunakan.

Menurut (Khotimah et al., 2019), *text mining* dan data mining memiliki perbedaan yang terletak pada sumber data yang digunakan. *Text mining* menggunakan data teks yang tidak terstruktur dan tidak berasal dari *database*, sedangkan data mining menggunakan data yang terstruktur dan berasal dari *database*. Namun memiliki kesamaan yang terletak pada ukuran data yang digunakan. Data yang digunakan memiliki ukuran cukup besar dan berdimensi cukup tinggi yang memiliki struktur kompleks dan terus berubah.

2.2.1 Analisis Sentimen

Analisis sentimen secara umum adalah mengenali opini seseorang, yang meliputi penilaian, evaluasi, emosi, kepercayaan, atau kepuasan yang diekspresikan dalam sebuah teks. Menurut (Akbari et al., 2017), analisis sentimen merupakan proses mengolah data tekstual secara otomatis untuk mendapatkan suatu informasi. Proses ini sangat penting dalam konteks bisnis karena, dapat membantu perusahaan memahami persepsi konsumen terhadap produk atau layanan mereka. Analisis sentimen tidak hanya berfungsi untuk mengevaluasi opini publik, tetapi juga dapat menjadi alat strategis dalam pengambilan keputusan yang lebih baik berdasarkan ulasan yang diberikan konsumen lain.

Penerapan analisis sentimen sangat luas, mencakup berbagai sektor seperti jasa kesehatan, jasa keuangan, produk konsumen, dan peristiwa social dan politik. Analisis sentimen bukan cuma berfungsi untuk mengidentifikasi opini, tapi juga untuk mengukur intensitas opini tersebut, baik yang bersifat positif, negatif, atau netral. Karena, opini yang terungkap dalam teks dapat mencerminkan sikap dan perilaku manusia (Liu, 2020). Dengan memanfaatkan hasil analisis sentiment, perusahaan atau organisasi dapat memperoleh informasi yang lebih mendalam tentang persepsi public, sehingga dapat membuat keputusan yang lebih akurat dan strategis untuk meningkatkan layanan dan produk yang diberikan.

2.2.2 Data Cleansing

Data mentah yang diperoleh dari berbagai sumber sering kali masih mengandung masalah seperti data yang tidak lengkap, terduplikat, *noise values*, *outlier*, hingga ketidakkonsistenan. Menurut (Alasadi & Bhaya, 2017), data

cleansing atau *data cleaning* menjadi langkah awal yang sangat krusial dalam preprocessing data untuk menemukan nilai yang hilang, membersihkan *noise values*, mengidentifikasi *outlier*, dan memperbaiki data yang tidak konsisten. Hal ini dapat mempengaruhi hasil *output data mining* atau *text mining* menjadi tidak dapat diandalkan dan buruk.

Data cleansing atau *data cleaning* merupakan proses penting sebelum melakukan proses *preprocessing text* yang bertujuan untuk mengidentifikasi serta memperbaiki kesalahan dan ketidakkonsistenan dalam dataset. Teknik-teknik seperti deteksi *outlier*, penghapusan data duplikasi, serta normalisasi data telah banyak dikembangkan untuk proses ini. Banyak algoritma dan alat otomatisasi guna mempercepat dan meningkatkan akurasi dalam proses *data cleansing* atau *data cleaning* (Singh & Gaur, 2019).

2.2.3 Text Preprocessing

Text preprocessing merupakan proses menyiapkan dataset sebelum melakukan proses analisis data yang bertujuan untuk menyaring data penting dari beberapa kumpulan data, dapat juga diartikan untuk mentransformasikan data teks yang tidak terstruktur menjadi data yang terstruktur (Surbakti et al., 2021).

Secara umum langkah-langkah yang dilakukan dalam text preprocessing adalah sebagai berikut:

1. *Casefolding*, merupakan tahapan yang dilakukan untuk mengganti seluruh karakter dalam teks menjadi huruf kecil (*lowercase*) atau huruf besar (*uppercase*) dengan tujuan meningkatkan efisiensi dalam proses pencarian informasi.

2. *Spelling Normalization*, adalah tahapan mengubah beberapa kata yang salah eja dengan tujuan untuk menghindari perluasan jumlah perhitungan dimensi kata.
3. *Tokenizing*, merupakan tahap membagi rangkaian karakter lengkap menjadi segmen kata individual. Proses ini bertujuan untuk memecah teks menjadi token atau kata-kata penyusunnya.
4. *Filtering*, merupakan langkah untuk menghilangkan kata-kata yang tidak relevan. Algoritma yang umum digunakan yaitu *stoplist* (untuk menghapus kata yang kurang signifikan) dan *wordlist* (untuk menyimpan kata yang penting). Tujuan dari langkah ini yaitu untuk memisahkan kata-kata yang sering muncul atau tidak memiliki makna, sehingga hanya kata-kata yang bermakna dan relevan yang akan dilanjutkan ke langkah berikutnya.
5. *Stemming*, adalah proses mengubah kalimat menjadi kata dasar yang bertujuan untuk mempersiapkan teks agar dapat dieksekusi oleh komputer.

2.2.4 Pembobotan Kata (*Term Weighthing*)

Term weighting menggunakan TF-IDF merupakan salah satu pendekatan pembobotan kata yang bertujuan untuk memberikan bobot hubungan suatu kata (*term*) terhadap sebuah dokumen (Nurjannah & Fitri Astuti, 2013). *Term* merupakan salah satu unit dalam sebuah dokumen yang digunakan untuk memetakan atau memahami isi dari sebuah dokumen.

TF-IDF adalah metode pembobotan *term* yang sering dijadikan acuan untuk membandingkan metode pembobotan lainnya. Dalam metode ini perhitungan *term t* dalam suatu dokumen dilakukan dengan menggabungkan nilai *Term Frequency* (frekuensi kemunculan *term* dalam dokumen) dengan *Inverse*

Document Frequency (kebalikan dari frekuensi kemunculan term tersebut di seluruh dokumen) (Nurjannah & Fitri Astuti, 2013).

Pada umumnya persamaan TF-IDF sebagai berikut:

$$w_{ij} = tf_{ij} \times idf_j \quad (2.1)$$

$$w_{ij} = tf_{ij} \times \log \frac{D}{df_j}$$

Keterangan:

w_{ij} : Adalah bobot term t_j terhadap dokumen d_i

tf_{ij} : Adalah jumlah kemunculan term t_j dalam dokumen d_i

D : Adalah jumlah semua dokumen dalam koleksi

df_j : Adalah jumlah dokumen yang mengandung term t_j

2.3 Klasifikasi

Klasifikasi merupakan salah satu teknik dalam *data mining* yang bertujuan untuk memetakan data ke dalam kelompok atau kelas yang telah ditentukan. Menurut (Setio et al., 2020), klasifikasi adalah metode *supervised learning* yang membutuhkan data training berlabel untuk menghasilkan sebuah aturan yang mengkategorikan data uji ke dalam kelompok atau kelas yang telah ditetapkan. Beberapa contoh teknik klasifikasi adalah *Naive Bayes classifier*, *Decision Tree*, *Rule-based classifier*, *Neural network*, dan *Support Vector Machine*. Metode klasifikasi menggunakan algoritma pembelajaran untuk menentukan model yang menghasilkan hubungan optimal antara kumpulan atribut dan label kelas dari data *input*.

2.3.1 Naive Bayes Classifier

Metode *Naive Bayes classifier* merupakan teknik klasifikasi probabilistik yang relatif sederhana, yang beroperasi dengan menghitung serangkaian probabilitas berdasarkan frekuensi kombinasi nilai dalam dataset yang di analisis (Rezki et al., 2020). Teknik klasifikasi ini digunakan untuk memprediksi kemungkinan di masa mendatang berdasarkan informasi yang sudah ada sebelumnya, yang dikenal dengan istilah probabilitas bersyarat (Fauzan & Hikmah, 2022). Probabilitas bersyarat adalah perhitungan yang menunjukkan kemungkinan terjadinya suatu peristiwa H , dengan asumsi bahwa peristiwa lain yaitu, peristiwa X , telah terjadi. Hal ini dinyatakan dengan simbol $P(H|X)$, yang mengaitkan probabilitas H dan X (Suhartono et al., 2018). Metode *Naive Bayes* didasarkan pada Teorema *Bayes*, yang diperkenalkan oleh ilmuwan asal inggris, Thomas Bayes. Teorema *Bayes* dapat dinyatakan dalam bentuk Persamaan 2.2 sebagai berikut:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (2.2)$$

Keterangan:

- X : Merupakan data dengan kelas belum diketahui
- H : Hipotesis data X termasuk dalam suatu kelas spesifik
- $P(H|X)$: Probabilitas hipotesis H berdasarkan kondisi X
- $P(H)$: Probabilitas hipotesis H
- $P(X|H)$: Probabilitas X berdasarkan kondisi
- $P(X)$: Probabilitas dari X

Multinomial Naive Bayes adalah algoritma *Naive Bayes* yang dirancang untuk memproses data dalam bentuk teks atau dokumen. Menurut (Abbas et al., 2019)(Stephenie et al., 2020) metode ini efektif digunakan dalam klasifikasi teks seperti analisis sentimen. Model ini bekerja dengan menghitung probabilitas dari suatu dokumen termasuk ke dalam kelas tertentu berdasarkan frekuensi kata yang muncul, dan kemudian memilih kelas dengan probabilitas tertinggi. Berikut Persamaan 2.3 digunakan untuk mencari nilai probabilitas tertinggi di setiap kategori *VMap* (Jurafsky & Martin, 2009) :

$$VMap = \arg \max P(Vj | a1, a2, \dots, an) \quad (2.3)$$

Selanjutnya, dengan menggunakan Persamaan Teorema Bayes, maka persamaannya menjadi seperti yang tercantum pada Persamaan 2.4 berikut:

$$VMap = \arg \max(Vj \in V) \frac{P(a1, a2, \dots, an | P(Vj))}{P(a1, a2, \dots, an)} \quad (2.4)$$

Karena nilai $P(a1, a2, a3)$ tetap untuk semua kategori (Vj), maka Persamaan 2.4 berubah menjadi Persamaan 2.5 berikut:

$$VMap = \arg \max(Vj \in V) P(a1, a2, \dots, an | P(Vj)) \quad (2.5)$$

Persamaan 2.5 dapat disederhanakan menjadi Persamaan 2.6 berikut:

$$VMap = \arg \max(Vj \in V) P(Vj) \prod_{k=1}^n P(Wk | Vj) \quad (2.6)$$

Keterangan:

$VMap$: Kelas dengan probabilitas tertinggi

$\arg \max$: Fungsi yang memilih nilai dengan hasil tertinggi dari sebuah ekspresi probabilistik

$P(Vj)$: Probabilitas awal (prior) untuk kelas Vj

$P(Wk | Vj)$: Probabilitas kata Wk muncul dalam kelas Vj

V_j : Kategori j

W_k : Bobot kata

$\prod_{k=1}^n P(W_k|V_j)$: Perkalian dari semua probabilitas kata

$P(V_j)$ atau probabilitas awal kelas (*prior*) dapat dihitung menggunakan Persamaan 2.7 berikut.

$$P(V_j) = \frac{|doc\ j|}{|sampel|} \quad (2.7)$$

Keterangan:

$|doc\ j|$: Total dokumen dari tiap kategori

$|sampel|$: Total dokumen seluruh kategori

Sedangkan $P(W_k|V_j)$ atau *likelihood* dengan *laplacian smoothing* (*add-one smoothing*) dapat dihitung menggunakan Persamaan 2.8 berikut.

$$P(W_k|V_j) = \frac{nk + 1}{n + |kosakata|} \quad (2.8)$$

Keterangan:

nk : Jumlah nilai TF-IDF dari kata W_k pada kelas V_j

n : Total nilai TF-IDF semua kata pada kelas V_j

Metode *Naive Bayes* memiliki karakteristik yang cepat dan tingkat akurasi yang tinggi saat diterapkan pada dataset yang luas dan bervariasi (Fauzan & Hikmah, 2022). Karena dataset yang digunakan cukup besar, diperlukan metode yang memiliki kinerja cepat serta akurasi yang tinggi. Nilai akurasi yang dihasilkan oleh metode *Naive Bayes* akan meningkat jika dikolaborasikan dengan metode pembobotan kata TF-IDF (Syarifuddin, 2020).

2.4 K-Fold Cross Validation

Menurut (Supartini et al., 2017), *Cross Validation* (CV) adalah suatu metode yang digunakan untuk menilai akurasi model yang dibangun berdasarkan dataset tertentu. Dalam CV, Solusi diperoleh dengan membagi data menjadi beberapa bagian (*fold*), di mana setiap bagian digunakan sebagai data uji pada setiap iterasi. Pada proses ini, data akan dibagi menjadi data *training* dan data *testing* di setiap uji coba.

Proses *Cross Validation* (CV) adalah metode yang membagi data yang tersedia menjadi K bagian atau *fold* secara acak. Setiap *fold* secara bergantian berfungsi sebagai data testing pada setiap iterasi CV. Sebanyak K percobaan dilakukan, di mana dalam setiap percobaan, satu fold ke- K digunakan sebagai data testing, sedangkan *fold* lainnya digunakan sebagai data *training*. Jumlah percobaan yang dilakukan akan sesuai dengan jumlah *fold* yang telah dibuat.

2.5 Online Review

Menurut *Oxford Dictionaries Online*, kata online berarti “*controlled by or connected to a computer*”, yang dalam bahasa Indonesia berarti “dikontrol oleh atau terhubung ke komputer”. Kata review dalam bahasa Indonesia berarti “ulasan”. Dalam Kamus Besar Bahasa Indonesia, ulasan diberi makna sebagai kupasan, tafsiran, serta komentar.

Menurut (Syakira & Moeliono, 2019), *online review* merupakan umpan balik berupa pendapat atau tinjauan dari konsumen yang didasarkan pada pengalaman mereka menggunakan suatu produk, dan hal ini dapat berpengaruh

pada calon pembeli baru dalam meningkatkan ketertarikan mereka untuk membeli produk tersebut.

2.6 Analisis Sentimen Menurut Pandangan Islam

Jual beli atau “البُيْع” merupakan bentuk jama’ dari lafadz “يَبِعُ” yang memiliki arti dalam bahasa Indonesia, menukar sesuatu dengan sesuatu yang lain. Hal ini berlaku pula bagi sesuatu yang bukan harta seperti khamr. Menurut syara’ jual beli memiliki arti berpindahnya kepemilikan hak suatu barang atau kemanfaatan suatu barang kepada orang lain dengan cara barter atau dengan mengganti kemanfaatan suatu barang dengan barang atau harga. Berdasarkan keterangan kitab fathul qorib, jual beli memiliki beberapa syarat sah yaitu penjual dan pembeli, adanya barang, dan adanya ijab dan qobul.

Jual beli harus dilakukan secara suka sama suka (saling ridho) antara penjual dan pembeli pada saat melakukan transaksi. Hal ini telah diterangkan dalam firman Allah Swt surat An-Nisa’ [4] ayat 29 dan Hadist Rasulullah SAW yang telah ditashih oleh Ibnu Hibban yang berbunyi:

يَا أَيُّهَا الَّذِينَ آمَنُوا لَا تَأْكُلُوا أَمْوَالَكُمْ بَيْنَكُمْ بِالْبَاطِلِ إِلَّا أَنْ تَكُونَ تِجَارَةً عَنْ تَرَاضٍ مِّنْكُمْ
وَلَا تَقْتُلُوا أَنْفُسَكُمْ إِنَّ اللَّهَ كَانَ بِكُمْ رَحِيمًا

“Wahai orang-orang yang beriman, janganlah kamu memakan harta sesamamu dengan cara yang batil (tidak benar), kecuali berupa perniagaan atas dasar suka sama suka diantara kamu. Janganlah kamu membunuh dirimu. Sesungguhnya Allah Swt adalah Maha Penyayang kepadamu.”

وَلَقَوْلِهِ - صلى الله عليه وسلم : أَمَّا الْبَيْعُ أَنْ تَرْضَى (ص - 325 - كتاب مغني المحتاج الى معرفة

معاني الفاظ المنهاج - كتاب البيع - المكتبة الشاملة)

“*Sesungguhnya jual beli (harus) atas dasar saling ridho (suka sama suka).*”

قال بعضهم : وَحَقِيقَةُ التَّرَاضِي لَا يَعْلَمُهَا إِلَّا اللَّهُ وَالْمَرَادُهَا هُنَا أَمَارَتُهُ كَالِإِيجَابِ وَالْقَبُولِ وَكَالتَعَاطَى عِنْدَ

الْقَائِلِ بِهِ. فَذَلِكَ عَلَى أَنَّ مُجَرَّدَ التَّرَاضِي هُوَ الْمَنَاطُ. وَلَا بُدَّ مِنَ الدَّلَالَةِ عَلَيْهِ بِلَفْظٍ أَوْ إِشَارَةٍ أَوْ كِتَابَةٍ،

بِأَيِّ لَفْظٍ وَقَعَ وَعَلَى أَيِّ صِفَةٍ كَانَ ، وَبِأَيِّ إِشَارَةٍ مُفِيدَةٍ حَصَلَ. (ص - 126 - كتب التفسير

الوسيط لطنطاوي - سورة النساء الآيات الى - المكتبة الشاملة)

Menurut beberapa ulama' ridho (suka sama suka) merupakan sifat tersembunyi yang hanya diketahui oleh Allah SWT. Orang lain tidak dapat mengetahui ridho kecuali dengan adanya tanda-tanda (qorinah) seperti, ucapan yang keluar dari keduanya, isyarah maupun ungkapan berupa tulisan.

Berdasarkan firman Allah Swt surat An-Nisa' [4] ayat 29, hadits Rasullullah SAW yang telah distashih Ibnu Hibban dan kutipan dari kitab Mughnil Muhtaj, dapat disimpulkan bahwa jual beli itu harus dilakukan dengan adanya saling ridho (suka sama suka) tanpa adanya perkara batil didalamnya ketika didalam transaksi ada perkara batil didalamnya maka transaksi tersebut merupakan tindakan batil dan dapat dijadikan landasan dalam dunia bisnis. Hal ini (saling ridho) juga berlaku terhadap rekan kerja kita apabila kita menjalankan bisnis secara kelompok. Namun ridho tidak dapat diketahui oleh manusia dikarenakan ridho memiliki sifat tersembunyi dalam hati dan hanya dapat diketahui oleh Allah Swt. Manusia dapat mengetahui ridho dengan adanya tanda-tanda pada saat melakukan transaksi seperti

pembeli meninggalkan teks ulasan pada toko online penjual berupa ulasan positif maupun negatif yang dapat digunakan sebagai informasi untuk meningkatkan kualitas produk dan jasa yang mereka berikan. Bagi pembeli yang akan membeli produk atau jasa yang ditawarkan mereka dapat melihat ulasan yang diberikan oleh pembeli lain. Ulasan tersebut bisa digunakan untuk informasi dalam proses pengambilan keputusan.

BAB III

METODE PENELITIAN

3.1 Jenis Penelitian

Jenis penelitian yang diterapkan pada penelitian ini adalah penelitian deskriptif dengan metode kuantitatif. Penelitian deskriptif (*Descriptive research*) merupakan penelitian yang dilakukan pada beberapa masalah berbentuk fakta-fakta terkini dan bersumber dari sebuah populasi, serta bertujuan untuk memberikan gambaran yang sistematis, faktual, dan akurat mengenai karakteristik, sifat-sifat, atau fenomena dalam suatu populasi atau sampel (Syahrizal & Jailani, 2023). Data yang dikumpulkan menggunakan instrumen penelitian dengan pendekatan penelitian kuantitatif yang berlandaskan ideologi positivis dan digunakan untuk mempelajari suatu populasi atau sampel dengan tujuan untuk mengkarakterisasi dan mengevaluasi hipotesis yang dipilih (Sugiyono, 2013).

3.2 Data dan Sumber Data

Dalam penelitian ini data yang digunakan adalah data sekunder yang berupa data ulasan produk konsumen Tokopedia berbahasa Indonesia sebanyak 5.400 ulasan produk dari 29 kategori produk dan setiap ulasan produk memiliki satu label sentimen. Data ulasan produk diperoleh dari sumber *opensource* Kaggle *datasets* (<https://www.kaggle.com/datasets/jocelyndumlaou/prdect-id-indonesian-emotion-classification>). Variabel yang digunakan pada penelitian ini terdiri dari 1 variabel *independent* dan 1 variabel *dependent* yang akan disajikan dalam Tabel 3.1:

Tabel 3.1 Variabel dan Definisi Variabel

Variabel	Definisi Operasional Variabel
Ulasan Produk(Variabel <i>Independent</i>)	Komentar konsumen terkait produk Tokopedia.com
Sentimen(Variabel <i>Dependent</i>)	Perasaan emosional yang terkandung dalam isi komentar

3.3 Teknik Analisis Data

3.3.1 Analisis Sentimen

Berikut adalah langkah-langkah yang dilakukan dalam penelitian ini:

1. Data *cleansing*

Tahap awal dari proses analisis sentimen adalah data *cleansing* yang bertujuan untuk menemukan nilai yang hilang, membersihkan *noise values*, mengidentifikasi *outlier*, duplikasi dalam data dan memperbaiki data yang tidak konsisten.

2. Proses *Text Preprocessing*

Pada proses *text preprocessing* memiliki beberapa tahapan sebagai berikut:

- a. *Casefolding* merupakan tahap awal *text preprocessing* yang bertujuan untuk mengubah semua teks yang terdapat pada ulasan produk menjadi huruf kecil (*lowercase*).
- b. *Cleansing* dan *normalization*, *cleansing* dilakukan bertujuan untuk menghapus komponen-komponen yang tidak penting dan komponen-komponen yang tidak ada kaitannya dengan dokumen seperti tanda baca, angka, karakter khusus, spasi berlebih dan memperbaiki kata yang memiliki huruf *double*. Sedangkan tahap *normalization* merupakan tahapan yang dilakukan untuk mengubah seluruh kata yang memiliki kesalahan dalam ejaan, *noise*, atau kata yang tidak baku menjadi baku.

- c. *Stopword removal* merupakan tahapan dalam *text preprocessing* yang bertujuan untuk menghapus atau menghilangkan kata-kata umum yang sering muncul dan tidak memiliki makna tertentu supaya jumlah kata yang tersimpan dalam daftar token tidak terlalu banyak.
- d. *Tokenizing* merupakan tahapan yang bertujuan untuk memecah teks menjadi token atau kata-kata penyusunnya.
- e. Setelah melakukan tokenisasi akan dilakukan proses pembobotan kata menggunakan TF-IDF dengan menggunakan persamaan 2.1 yang bertujuan untuk mengidentifikasi kata-kata yang memiliki makna penting dalam sebuah dokumen dan dapat meningkatkan akurasi dalam proses evaluasi model.

3. Proses *Training*

Pada proses *training* terdapat beberapa tahapan sebagai berikut:

- a. Sebanyak 80% dari data ulasan produk digunakan sebagai data *training* yang berjumlah 4.314 data.
- b. Membangun model dengan menggunakan metode *multinomial naive bayes*.
- c. Evaluasi performa model pada data training menggunakan *K-fold cross validation* yang bertujuan mengambil nilai akurasi terbaik untuk digunakan pada evaluasi model pada data testing. Penelitian ini menggunakan nilai *K* yang digunakan adalah 5 yang berarti proses evaluasi model dilakukan sebanyak lima kali.

4. Proses *Testing*

Pada proses *testing* terdapat beberapa tahapan sebagai berikut:

- a. Sebanyak 20% dari data ulasan produk digunakan sebagai data testing yang berjumlah 1.079 data.
- b. Model terbaik dari proses *K-fold cross validation* digunakan untuk mengklasifikasi data uji dan dievaluasi menggunakan *Confusion Matrix* yang bertujuan untuk mengetahui nilai akurasi, presisi, *recall* dan *f-measure*.

3.3.2 Confusion Matrix

Confusion matrix merupakan tahap pengujian atau menghitung nilai performa yang bertujuan untuk mengetahui nilai akurasi, presisi, *recall*, dan *f-measure*. Menurut (Anam et al., 2024) *Confusion Matrix* merupakan sebuah tabel yang di dalamnya berisi banyak data uji yang diprediksi benar dan salah oleh model klasifikasi dan nilai *true false* dijadikan pedoman dalam menghitung nilai performa. Konsep dari *Confusion Matrix* dapat dilihat dalam Tabel 3.2 berikut.

Tabel 3.2 Konsep *Confusion Matrix*

Prediksi Aktual	Kelas Prediksi	
	<i>Negative</i>	<i>Postive</i>
Kelas Negatif	<i>True Negative (TN)</i>	<i>False Positive (FP)</i>
Kelas Positif	<i>False Negative (FN)</i>	<i>True Positive (TP)</i>

Berikut persamaan untuk melakukan perhitungan nilai performa berdasarkan konsep *Confusion Matrix* pada Tabel 3.2 di atas.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \quad (3.1)$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (3.2)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (3.3)$$

$$f - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\% \quad (3.4)$$

Keterangan:

TP: Jumlah data yang memiliki nilai positif dan diklasifikasikan sebagai data positif.

FP: Jumlah data yang memiliki nilai negatif tetapi diklasifikasikan sebagai data positif.

FN: Jumlah data yang memiliki nilai positif tetapi diklasifikasikan sebagai data negatif.

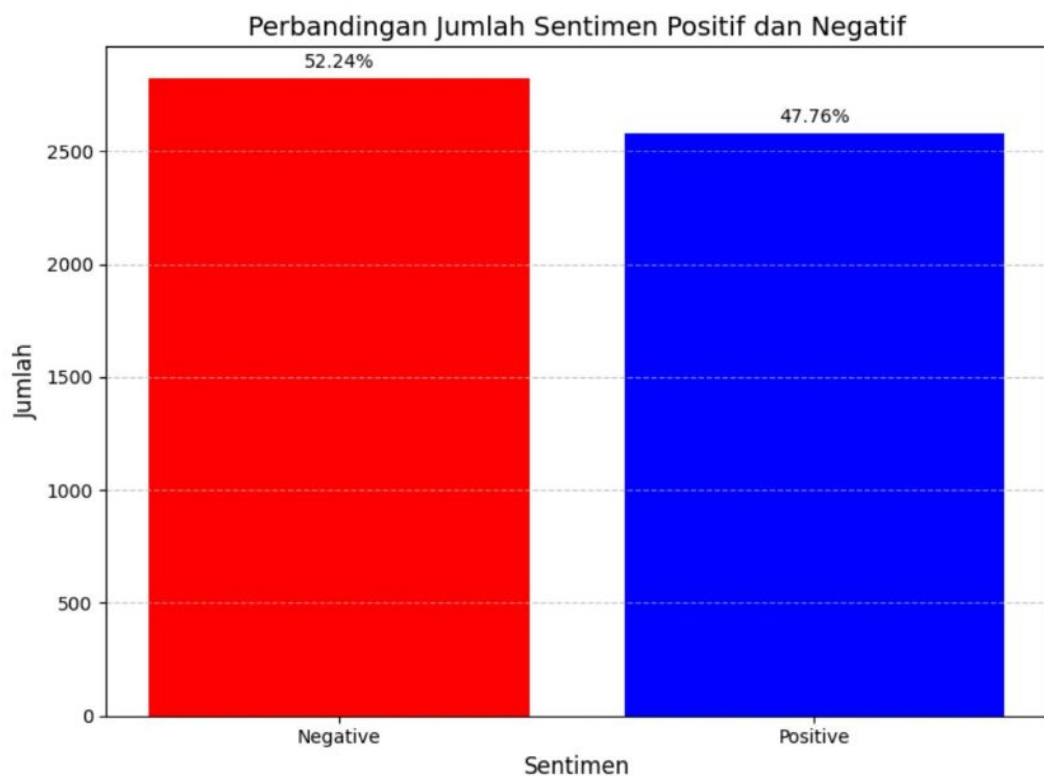
TN: Jumlah data yang memiliki nilai negatif dan diklasifikasikan sebagai data negatif.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Data Penelitian

Data penelitian berupa data ulasan produk konsumen Tokopedia berbahasa Indonesia sebanyak 5.400 ulasan produk dan setiap ulasan produk memiliki satu label sentimen. Ulasan produk yang berlabel sentimen positif sebanyak 2.579 ulasan produk sedangkan ulasan produk yang berlabel sentimen negatif sebanyak 2.821 ulasan produk. Visualisasi perbandingan sentiment positif dan negatif dapat dilihat pada Gambar 4.1.



Gambar 4.1 Presentase Data Ulasan Produk

Berdasarkan Gambar 4.1 ulasan produk positif memiliki presentase sebesar 47,76 % dan ulasan produk negatif memiliki presentase sebesar 52,24% dari 5.400

data ulasan produk yang digunakan. Berikut sampel data ulasan produk yang disajikan dalam Tabel 4.1 berikut.

Tabel 4.1 Data Sampel Ulasan Produk

No	Customer Review	Sentiment
1.	Barang sudah diterima nih kak, makasih yaa	Positive
2.	Gampang dibawa-bawa, terlalu imut ukurannya	Positive
3.	LANGGANAN ??????????	Positive
4.	bagus, pengiriman cepet banget bakal jadi langganan nih	Positive
5.	Kartu bekerja dengan baik begitupun sellernya yang sigap menjawab semua pertanyaan..	Positive
6.	Pengirimannya cepat, besok langsung sampai. Nggak expect apa-apa sebenarnya, tapi ternyata bagus. Bahannya lembut banget, gelap oke buat nutup mata + ternyata ada gelnya juga, kirain gaada hehe. Overall worth it siih first impression-nya bagus ????	Positive
7.	plastik yang uk. 50x75 itu tebal banget bagus.. yang uk 40x60 agak tipis.. pengiriman cepat lumayan. thanks	Positive
8.	Produk nya OK..kaca bersih, karet nya OK.. Kain pembersih nya OK jg Bersihin kaca walupun di awal sudah dipasang	Positive
9.	bahan nya bagus banget	Positive
10.	barang sudah sampai dengan baik.. packing rapih.. dan .. terima kasih	Positive
11.	Semua kelipet yg pakai mata ayam , print an juga gak rapi , kecewa sih	Negative
12.	Barang nya kotor banyak bintik hitam dan bintik putih mungkin jamur.	Negative
13.	packing berantakan ada yg bocir tumpah semua	Negative
14.	Sangat kecewa, extra box tidak dikirim sama sekali. Saya beli 4 sepatu dan 4 extra box tapi tidak ada extraboxnya. Saya hanya terima 4 sepatu di dalam 1 box besar.	Negative
15.	Harga di struk yang dimasukkan di kotaknya beda dengan harga yang ditampilkan di tokopedia. Kok bisa harganya lebih mahal ya? Di struk beda jauh harganya	Negative
16.	pesen playdough 3, dikirim cuma 1, pesen stamp tp ga dikirim..padahal byr full..ga ada cfm sama sekali..sangat ga recommended tokonya (maaf sy ksh bintang 1, lain kali hrs lebih amanah kl jual)	Negative
17.	barang tidak sesuai gambar....bukan original. kaitan plastik semua...di gambar kaitan stainless/besi	Negative
18.	barang tidak sesuai orderan	Negative
19.	gak sesuai ekspektasi. klo buat simpan barang2 yg agak berat kurang recommended. dan prosesnya lama, klo prosesnya gak sampe 2 hari mungkin sampe lebih cepat kali ya.	Negative

Lanjutan Tabel 4.1 Data Sampel Ulasan Produk

20.	sedih bgt belom ada 2bln waxingnya udah rusak dan gabisa dipake lg, padahal baru dipake 2x. dr awal dtg tutupnya penyok dan ga terlalu kedap udara jd pinggiran tutupnya rada basah. trus selama 1bln lebih disimpan di suhu ruang trus tutupnya selalu rapet, tp pas mau dipake lg taunya udah banyak semutnya:(pas dicoba pake spatula teksturnya jg mirip madu jd rada cair dan gamau lengket lg:(katanya kalo cair buang bagian atasnya aja, ini mah ampe bawah2 semua teksturnya kayak madu:(berarti dibuang semua dong? sedih bgt mana masih banyak lg sayang bgt:(	Negative
-----	--	----------

4.2 Data Cleansing

Data cleansing merupakan tahap awal dalam penelitian ini, yang bertujuan untuk menemukan data yang terduplikasi atau kembar. Dataset mentah yang digunakan dalam penelitian ini berjumlah 5.400 ulasan produk dan setiap ulasan memiliki satu label sentimen. Pada tahap ini, dilakukan pemeriksaan terhadap keberadaan data yang terduplikasi dengan bantuan sistem yang dibuat dengan *google collaboratory* dan bahasa pemrograman Python.

Proses *cleansing* ini, menggunakan library pandas dan beberapa fungsi bahasa pemrograman Python seperti `uplicated`, `drop_duplicates`, dan `reset_index` untuk membaca file csv, menemukan data yang terduplikasi, menghapus data yang terduplikasi dan mereset indexnya. Setelah melakukan pemeriksaan ini, ditemukan 7 data ulasan produk yang terduplikasi sehingga data ulasan yang terduplikasi harus dihapus untuk menjaga integritas dataset. Setelah dilakukan penghapusan terhadap data yang terduplikasi, jumlah data ulasan produk yang tersisa menjadi 5.393 data sehingga dataset yang digunakan menjadi lebih bersih dan representatif. Implementasi *data cleaning* yang disajikan dalam bentuk pseudocode dapat dilihat pada Tabel 4.2.

Tabel 4.2 Pseudocode Data Cleansing

```

BEGIN
  Input
    File dataset berformat CSV bernama 'PRDECT-ID
Dataset
  .csv'
  Process
    Baca dataset dari file CSV
    data ← BacaCSV('PRDECT-ID Dataset.csv')
    Hitung jumlah total data
    jumlah_data ← HitungJumlahBaris(data)
    Cetak('Jumlah Data:', jumlah_data)
    Deteksi data duplikat
    Kembar ← TandaiDuplikat(data)
    jumlah_kembar ← HitungJumlahDuplikat(data)
    Cetak('Jumlah Data Kembar:', jumlah_kembar)
    Hapus data yang duplikat
    HapusBarisDuplikat(data)
    Hitung jumlah data setelah duplikat dihapus
    hasil ← HitungJumlahBaris(data)
    Cetak('Jumlah data setelah dihapus:', hasil)
    Reset indeks dataset
    reset ← ResetIndeks(data)
    Hitung jumlah data setelah indeks direset
    jumlah_reset ← HitungJumlahBaris(reset)
    Cetak('Jumlah data cleansing:', jumlah_reset)
  Output
    Informasi jumlah data awal, jumlah data duplikat,
jumlah
    data setelah pembersihan, dan data yang telah
direset
END

```

4.3 Text Preprocessing

Setelah melalui melalui proses pembersihan data, langkah berikutnya sangat penting untuk mengkonversi ulasan produk yang masih dalam bentuk tidak terorganisir menjadi format data terorganisir yang siap untuk diuji.

4.3.1 Casefolding

Casefolding merupakan tahap pertama dalam tahap *text preprocessing* yang bertujuan untuk mengubah seluruh karakter dalam teks menjadi huruf kecil dan meminimalisir kesalahan saat proses pencarian kata. Tahap ini mengimplementasikan fungsi `.str.lower()` pada *library* *pandas*, yang secara efisien mengubah seluruh karakter atau huruf pada kolom yang dituju menjadi huruf kecil. Implementasi tahap casefolding yang disajikan dalam bentuk pseudocode pada Tabel 4.3.

Tabel 4.3 *Pseudocode Casefolding*

```

BEGIN
  Input
    Kolom 'Customer Review' dari dataset (cleaning)
  Process
    FUNCTION casefolding (text)
      text ← ubah semua huruf dalam teks menjadi huruf
      kecil
    RETURN text

    cleaning['Customer Review'] ← cleaning['Customer
    Review'].apply(casefolding)
  Output
    Tampilkan lima entri pertama dari cleaning['Customer
    Review']
END

```

4.3.2 *Cleansing dan Normalization*

Tahap yang kedua adalah *cleansing* dan *normalization*. *Cleansing* sendiri dilakukan bertujuan untuk menghapus komponen-komponen yang tidak penting dan komponen-komponen yang tidak ada kaitannya dengan dokumen seperti tanda baca, nomor, karakter khusus, spasi ekstra dan memperbaiki kata yang memiliki huruf double. Sedangkan tahap *normalization* merupakan tahapan dalam preprocessing teks yang dilakukan untuk memperbaiki seluruh kata yang salah eja, menghilangkan *noise*, atau mengganti kata tidak baku menjadi bentuk yang sesuai dengan kaidah bahasa. Implementasi tahap *cleansing* dan *normalization* yang disajikan dalam bentuk pseudocode pada Tabel 4.4.

Tabel 4.4 *Pseudocode Cleansing dan Normalization*

```

BEGIN
  Input
    Kolom 'Customer Review' dari dataset cleaning
  Process
    Definisikan kamus normalisasi
      norm ← pasangan kata tidak baku dan padanannya
    FUNCTION normalisasi(teks)
      Untuk setiap pasangan (kata_singkat, kata_baku)
dalam
      norm:
        Ganti kata_singkat di dalam teks dengan kata_baku
        Menggunakan ekspresi regular
      RETURN teks

    FUNCTION perlu_normalisasi(teks)
      IF teks mengandung:
        karakter khusus selain huruf dan angka
        angka
        spasi berlebih dari satu karakter

```

Lanjutan Tabel 4.4 *Pseudocode Cleansing dan Normalization*

```

ELSE:
    RETURN True
ELSE:
    RETURN False

FUNCTION normalisasi_teks(teks)
    Hapus karakter khusus (selain huruf dan angka)
    Hapus angka
    Ubah spasi ganda menjadi satu spasi
    Hapus spasi diawal dan akhir
    Ubah huruf berulang menjadi satu huruf saja
    RETURN teks
Terapkan fungsi normalisasi() pada seluruh entri kolom
'Customer Review':
    Cleaning['Customer Review'] ← cleaning['Customer
    Review'].apply(normalisasi)
Tandai teks yang masih perlu dinormalisasi:
    mask ← cleaning['Customer Review'].apply(perlu_norma
    lisasi)
Untuk entri yang bertanda mask = True, lakukan
pembersihan lanjutan:
    cleaning[mask, 'Customer Review'] ← cleaning[mask,
    'Customer Review'].apply(normalisasi_teks)
Output
    Tampilkan lima entri pertama dari kolom 'Customer
    Review' setelah normalisasi
END

```

4.3.3 Stopword Removal

Tahap yang ketiga adalah *stopword removal*, yang digunakan untuk menghilangkan kata-kata umum yang sering muncul namun tidak memiliki makna spesifik, agar jumlah kata dalam daftar token tetap efisien. Proses

implementasi *stopword removal* yang disajikan dalam bentuk pseudocode pada Tabel 4.5.

Tabel 4.5 Pseudocode Stopword Removal

```
BEGIN
  Input
    Kolom 'Customer Review' dari dataset cleaning
  Process
    Definisikan list stopwords tambahan (more_stop)
    Gabungkan stopwords tambahan dengan stopwords default
    dari Sastrawi
    FUNCTION stopwords(teks)
      Hapus stopwords menggunakan StopWordRemover.remove
      (teks)
      Hapus stopwords tambahan menggunakan regex
    END FUNCTION
    cleaning['Customer Review'] ← cleaning['Customer
    Review'].apply(lambda ulasan: stopwords(ulasan))
  Output
    Tampilkan lima entri pertama dari cleaning['Customer
    Review']
END
```

4.3.4 Tokenizing

Tokenizing merupakan sebuah tahapan yang berfungsi untuk memecah kalimat menjadi satuan kata. Implementasi tahap *tokenizing* yang disajikan dalam bentuk pseudocode pada Tabel 4.6.

Tabel 4.6 Pseudocode Tokenizing

```
BEGIN
  Input
    Kolom 'Customer Review' dari dataset cleaning
  Process
```

Lanjutan Tabel 4.6 Pseudocode Tokenizing

```

FUNCTION tokenisasi(ulasan)
    tokens ← pisahkan teks menjadi kata-kata berdasarkan
    spasi
    RETURN tokens
END FUNCTION

Tokenisasi ← cleaning['Customer Review'].apply(lambda
ulasan:tokenisasi(ulasan))

Output
    Tampilkan lima entri pertama dari Tokenisasi

END

```

4.3.5 Stemming

Tahap akhir dari *preprocessing* teks adalah *stemming*, tahap *stemming* bertujuan untuk mengonversi seluruh kata dalam dokumen ke bentuk dasarnya dengan menghilangkan seluruh imbuhan kata. Implementasi tahap *stemming* yang disajikan dalam bentuk pseudocode pada Tabel 4.7.

Tabel 4.7 Pseudocode Stemming

```

BEGIN
    Input
        List token (hasil tokenisasi) dari kolom 'Customer
        Review'
    Process
        Inisialisasi stemmer
            Buat objek Stemmer dari pustaka Sastrawi
        FUNCTION stemming(daftar_kata)
            Buat list kosong untuk menampung kata hasil stemming
            (akar_kata)
            Untuk setiap kata dalam daftar_kata:
                Ubah kata menjadi bentuk dasar menggunakan stemmer
                Tambahkan ke dalam list akar_kata
            Gabungkan semua akar_kata menjadi satu string dengan

```

Lanjutan Tabel 4.7 Pseudocode Stemming

```

    spasi sebagai pemisah
    Cetak string hasil stemming untuk verifikasi
    RETURN string hasil stemming
Terapkan fungsi stemming pada setiap entri dalam
Variabel hasil tokenisasi:
    hasil_stemming ← Tokenisasi.apply(stemming)
Simpan hasil_stemming ke dalam file CSV:
    SimpanCSV(hasil_stemming, 'Cleaningdata.csv')
Baca kembali file csv hasil stemming:
    data_clean ← BacaCSV('Cleaningdata.csv', encoding =
        'latin1')
Tampilkan lima entri pertama dari data_clean untuk
Verifikasi
Output
    File 'Cleaningdata.csv' yang berisi hasil stemming
    Lima entri pertama dari file tersebut
END

```

Tabel 4.8 berikut, merupakan hasil dari *text preprocessing* seperti *casefolding*, *cleansing* dan *normalization*, *stopword removal*, *tokenizing*, dan *stemming* yang dilakukan

Tabel 4.8 Hasil Text Preprocessing

<i>Preprocessing</i>	Sebelum	Sesudah
<i>Casefolding</i>	Pengirimannya cepat, besok langsung sampai. Nggak expect apa-apa sebenarnya, tapi ternyata bagus. Bahannya lembut banget, gelap oke buat nutup mata + ternyata ada gelnya juga, kirain gaada hehe. Overall worth it siih first impression-nya bagus ????	pengirimannya cepat, besok langsung sampai. nggak expect apa-apa sebenarnya, tapi ternyata bagus. bahannya lembut banget, gelap oke buat nutup mata + ternyata ada gelnya juga, kirain gaada hehe. overall worth it siih first impression-nya bagus ????
<i>Cleansing dan normalization</i>	pengirimannya cepat, besok langsung sampai. nggak expect apa-apa sebenarnya, tapi ternyata bagus. bahannya lembut banget, gelap oke buat nutup mata +	pengirimannya cepat besok langsung sampai tidak expect apa apa sebenarnya tapi ternyata bagus bahannya lembut banget gelap oke buat nutup mata

Lanjutan Tabel 4.8 Hasil Text Preprocessing

	ternyata ada gelnya juga, kirain gaada hehe. overall worth it siih first impression-nya bagus ????	ternyata ada gelnya juga kirain tidak ada hehe overall worth it siih first impression nya bagus
<i>Stopword</i>	pengirimannya cepat besok langsung sampai tidak expect apa apa sebenarnya tapi ternyata bagus bahannya lembut banget gelap oke buat tutup mata ternyata ada gelnya juga kirain tidak ada hehe overall worth it siih first impression nya bagus	Pengiriman cepat besok langsung sampai tidak apa apa sebenar tapi bagus bahan lembut banget gelap oke buat tutup mata ada gel juga kirain tidak ada bagus
<i>Tokenizing</i>	Pengiriman cepat besok langsung sampai tidak apa apa sebenar tapi bagus bahan lembut banget gelap oke buat tutup mata ada gel juga kirain tidak ada bagus	“pengiriman”, “cepat”, “besok”, “langsung”, “sampai”, “tidak”, “apa”, “apa”, “sebenar”, “tapi”, “bagus”, “bahan”, “lembut”, “banget”, “gelap”, “oke”, “buat”, “tutup”, “mata”, “ada”, “gel”, “juga”, “kirain”, “tidak”, “ada”, “bagus”
<i>Stemming</i>	“pengiriman”, “cepat”, “besok”, “langsung”, “sampai”, “tidak”, “apa”, “apa”, “sebenar”, “tapi”, “bagus”, “bahan”, “lembut”, “banget”, “gelap”, “oke”, “buat”, “tutup”, “mata”, “ada”, “gel”, “juga”, “kirain”, “tidak”, “ada”, “bagus”	“kirin”, “cepat”, “besok”, “langsung”, “sampai”, “tidak”, “apa”, “apa”, “benar”, “tapi”, “bagus”, “bahan”, “lembut”, “banget”, “gelap”, “oke”, “buat”, “tutup”, “mata”, “ada”, “gel”, “juga”, “kira”, “tidak”, “ada”, “bagus”

4.4 Term Weigthing (TF-IDF)

Sesudah tahap *text preprocessing* sudah selesai dilakukan, akan dihasilkan sekumpulan kata yang selanjutnya akan menjadi *term* untuk langkah selanjutnya, yaitu *term weigthing* atau pembobotan kata dengan mengimplementasikan *Term Frequency-Inverse Document Frequency* atau TF-IDF. *Term weigthing* menggunakan TF-IDF dapat dilakukan menggunakan Persamaan 2.1.

4.4.1 Perhitungan Term Frequency (TF)

Term Frequency (TF) merupakan ukuran frekuensi kemunculan suatu kata dalam sebuah dokumen. Nilai *Term Frequency* didapatkan dengan membagi jumlah term atau kata “*t*” dalam dokumen “*d*” dengan total term dalam dokumen

“*d*”. Berikut contoh perhitungan manual TF kata “barang” pada 20 dokumen sampel yang akan disajikan dalam Tabel 4.9.

Tabel 4.9 Contoh Perhitungan *Term Frequency* (TF)

Dokumen	Jumlah Kata	Kata	Frekuensi Kemunculan Kata	Nilai TF
D1	6	barang	1	1/6
D2	5		0	0/5
D3	1		0	0/1
D4	8		0	0/8
D5	9		0	0/9
D6	8		0	0/8
D7	10		0	0/10
D8	9		0	0/9
D9	3		0	0/3
D10	7		1	1/7
D11	9		0	0/9
D12	9		1	1/9
D13	5		0	0/5
D14	17		0	0/17
D15	15		0	0/15
D16	21		0	0/21
D17	12		1	1/12
D18	3		1	1/3
D19	22		1	1/22
D20	71		0	0/71

4.4.2 Perhitungan *Inverse Document Frequency* (IDF)

Inverse Document Frequency (IDF) berfungsi untuk menilai seberapa penting suatu term dalam keseluruhan kumpulan dokumen. Berikut contoh perhitungan manual nilai IDF untuk kata “barang” pada 20 dokumen sampel.

$$IDF(\text{barang}) = \log \frac{D}{df_{\text{barang}}}$$

$$IDF(\text{barang}) = \log \frac{20}{6}$$

$$IDF(\text{barang}) = 0,522879$$

4.4.3 Perhitungan TF-IDF

Perhitungan TF-IDF dilakukan menggunakan Persamaan 2.1 dengan cara mengalikan nilai TF dan IDF. Berikut perhitungan bobot TF-IDF untuk 7 kata pada setiap dokumen sampel yang disajikan pada Tabel 4.10.

Tabel 4.10 Hasil Perhitungan TF-IDF

Dokumen	Kata						
	barang	terima	kakak	kasih	iya	gampang	bawa
D1	0,08714	0,23299	0,21683	0,13731	0,21683	0	0
D2	0	0	0	0	0	0,26020	0,26020
D3	0	0	0	0	0	0	0
D4	0	0	0	0	0	0	0
D5	0	0	0	0	0	0	0
D6	0	0	0	0	0	0	0
D7	0	0	0	0	0	0	0
D8	0	0	0	0	0	0	0
D9	0	0	0	0	0	0	0
D10	0,0747	0,09985	0	0,1177	0	0	0
D11	0	0	0	0	0	0	0
D12	0,0581	0	0	0	0	0	0
D13	0	0	0	0	0	0	0
D14	0	0,04112	0	0	0	0	0
D15	0	0	0	0	0	0	0
D16	0	0	0	0,03923	0	0	0
D17	0,04357	0	0	0	0	0	0
D18	0,17429	0	0	0	0	0	0
D19	0,02377	0	0	0	0	0	0
D20	0	0	0	0	0	0	0

4.5 Training Data

Proses *training* dilakukan menggunakan data sebanyak 80% dari data ulasan produk sebagai data *training*, yang secara keseluruhan berjumlah 4.314 data. Untuk keperluan contoh perhitungan, digunakan 20 data dari Tabel 4.1 yang diklasifikasikan secara manual sesuai kelasnya dan telah melalui proses *preprocessing*, seperti yang ditampilkan pada Tabel 4.11.

Tabel 4.11 Sampel Dokumen *Training*

Dokumen	Teks Hasil <i>Preprocessing</i>	Kelas
D1	barang terima kakak terima kasih iya	P
D2	gampang bawa terlalu imut ukur	P
D3	langgan	P
D4	bagus kirim cepat banget bakal jadi langgan nih	P
D5	kartu kerja baik begitu seller sigap jawab semua tanya	P
D6	kirim cepat besok langsung apa apa benar nyata	P
D7	plastik ukuran tebal banget bagus ukuran tipis kirim cepat lumayan	P
D8	produk kaca bersih karet kain bersih bersih kaca walau	P
D9	bahan bagus banget	P
D10	barang sudah sampai baik rapih terima kasih	P
D11	semua lipat pakai mata ayam print tidak rapi kecewa	N
D12	barang kotor banyak bintik hitam bintik putih mungkin jamur	N
D13	packing rantak bocor tumpah semua	N
D14	sangat kecewa ekstra kotak kirim sama sekali beli sepatu ekstra kotak ekstra kotak terima sepatu kotak besar	N
D15	harga struk masuk kotak beda harga tampil tokopedia harga lebih mahal struk beda jauh harga	N
D16	kirim cuma tidak kirim padahal bayar penuh sama sekali tidak saran toko maaf kasih bintang kali harus lebih amanah kalau jual	N
D17	barang sesuai gambar bukan original kait plastik semua gambar kait stenless besi	N
D18	barang sesuai order	N
D19	tidak sesuai ekspektasi kalau buat simpan barang agak berat kurang rekomendasi proses lama kalau proses tidak sampai hari mungkin sampai lebih cepat	N
D20	sedih belum ada bulan rusak tidak bisa pakai lagi baru pakai awal datang tutup penyok tidak terlalu kedap udara pinggir tutup agak basah selama bulan lebih simpan suhu ruang terus tutup rapat mau pakai lagi tahu sudah banyak semut coba pakai spatula tekstur juga mirip madu jadi agak cair tidak mau lengket lagi kata kalau cair buang bagi atas saja ini sampai bawah semua	N

4.5.1 Multinomial Naive Bayes

Berdasarkan dokumen *training* yang disajikan dalam Tabel 4.11, model *prior probabilities* dibangun menggunakan Persamaan 2.7, seperti dijelaskan berikut ini.

$$P(V_j|positif) = \frac{|doc\ j|}{|sampel|} = \frac{10}{20} = \frac{1}{2} = 0,5$$

$$P(V_j|negatif) = \frac{|doc\ j|}{|sampel|} = \frac{10}{20} = \frac{1}{2} = 0,5$$

Dalam data *training*, terdapat 20 dokumen, di mana masing-masing kelas positif dan negatif muncul sebanyak 10 kali. Jadi peluang suatu dokumen termasuk dalam kelas positif adalah 10 dari 20, atau $\frac{1}{2}$ begitu juga untuk kelas negatif. Selanjutnya, dilakukan perhitungan probabilitas bersyarat (*conditional probabilities*) berdasarkan Persamaan 2.8. Sebagai contoh kasus, perhitungan *conditional probabilities* untuk kata “barang” pada kelas positif dan negatif.

$$\begin{aligned} P(W_{barang}|positif) &= \frac{nk + 1}{n + |kosakata|} = \frac{0,1618434 + 1}{10,251089 + 153} = \frac{1,1618434}{163,251089} \\ &= 0,0071169 \end{aligned}$$

$$\begin{aligned} P(W_{barang}|negatif) &= \frac{nk + 1}{n + |kosakata|} = \frac{0,299731 + 1}{9,8579459 + 153} \\ &= \frac{1,299731}{162,8579459} = 0,0079808 \end{aligned}$$

4.5.2 K-Fold Cross Validation

K-fold Cross Validation merupakan teknik validasi model yang digunakan untuk mengukur performa generalisasi model terhadap data yang belum pernah dilihat sebelumnya. Tahap ini, evaluasi model dilakukan pada data *training* yang berjumlah 4.314 data dengan nilai $K = 5$, artinya data dibagi menjadi lima bagian, masing-masing berisi kurang lebih 863 data. Tujuan proses ini adalah untuk memperoleh nilai akurasi terbaik pada data *training* yang nantinya akan digunakan dalam evaluasi terhadap data *testing*.

Model yang digunakan untuk proses evaluasi ini adalah *Multinomial Naive Bayes* dengan parameter $\alpha = 1$, yang berarti model menggunakan teknik *Laplacian Correction* untuk menghindari pembagian nol akibat tidak adanya kemunculan fitur tertentu pada kelas tertentu. Setiap iterasi evaluasi menghasilkan nilai akurasi validasi yang kemudian dibandingkan. Model dengan akurasi tertinggi dari seluruh iterasi dipilih sebagai model terbaik dan digunakan untuk proses prediksi pada data testing. Implementasi teknik ini dilakukan menggunakan pustaka *scikit-learn* dengan *StratifiedKFold*.

Berikut hasil dari *5-Fold Cross Validation* terhadap data *training* yang disajikan pada Tabel 4.12.

Tabel 4.12 Hasil *K-Fold Cross Validation*

Iterasi ke-	Akurasi Validasi
1	0,9096
2	0,9224
3	0,9282
4	0,9224
5	0,9304
Rata-Rata Akurasi	0,9226

Berdasarkan Tabel 4.12 dapat dilihat bahwa proses *K-Fold Cross Validation* menghasilkan nilai akurasi validasi yang bervariasi antara 0,9096 hingga 0,9304. Akurasi tertinggi yang diperoleh terdapat pada iterasi ke-5 dengan nilai 0,9304, yang nantinya akan digunakan sebagai model terbaik untuk tahap evaluasi pada data *testing*.

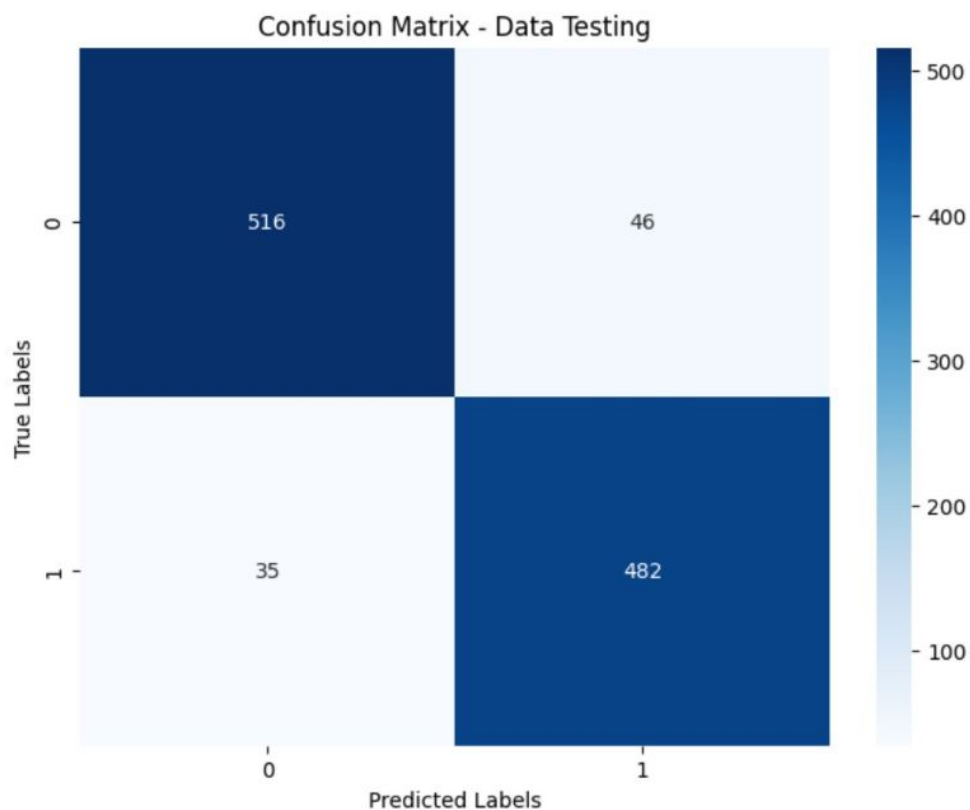
4.6 Testing Data

Pada tahap *testing*, data yang digunakan adalah data *testing* sebanyak 20% dari total data ulasan produk yaitu 1.079 data. Model yang digunakan untuk mengklasifikasikan data *testing* adalah model terbaik yang diperoleh dari proses *K-*

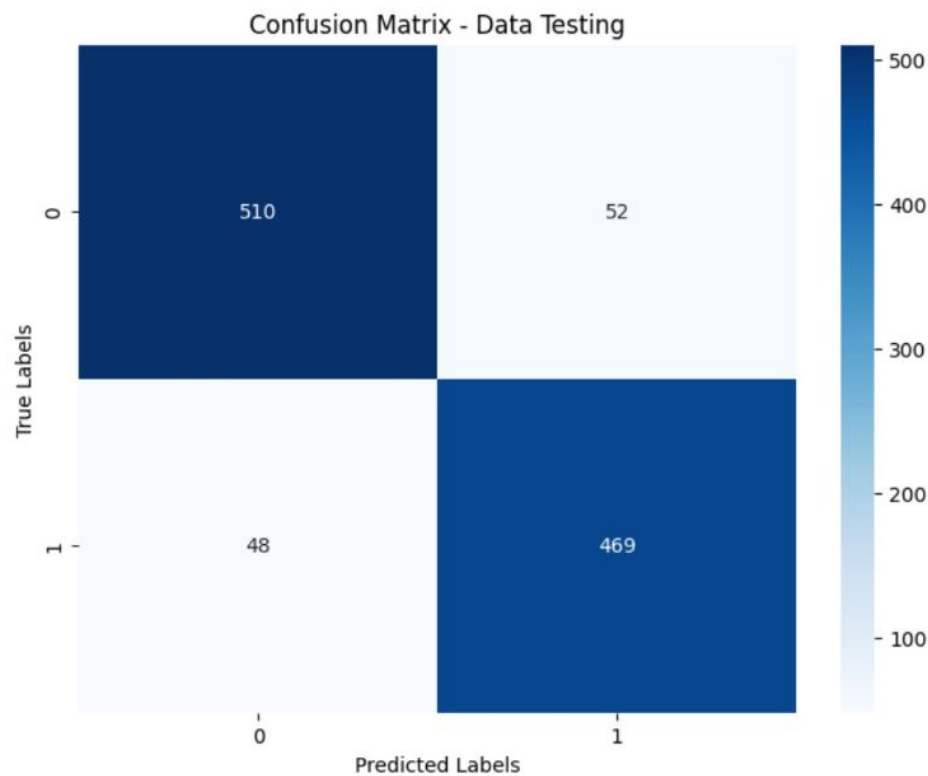
fold Cross Validation. Evaluasi performa model dilakukan dengan menggunakan Confusion Matrix, yang bertujuan untuk mengukur tingkat akurasi, presisi, *recall*, dan *f-measure*. Hasil dari testing data ini memberikan gambaran tentang seberapa baik model mampu mengenali pola dan melakukan klasifikasi secara akurat terhadap data baru.

4.6.1 Confusion Matrix

Proses *testing* data dilakukan guna mengevaluasi dan mengukur nilai performa sistem yang telah dibangun. Hasil dari proses *testing* data yang dilakukan adalah berupa nilai *Confusion Matrix*. Berikut merupakan hasil matrix dan nilai akurasi, presisi, *recall*, dan *f-measure*.



Gambar 4.2 Hasil *Confusion Matrix Naive Bayes*



Gambar 4.3 Hasil *Confusion Matrix Random Forest*

Gambar 4.2 menunjukkan *Confusion Matrix Naive Bayes* yang digunakan sebagai dasar evaluasi performa sistem klasifikasi yang telah dibangun. Berdasarkan Gambar 4.2, diperoleh nilai *True Positive* (TP) sebanyak 482, yang menunjukkan jumlah data kelas 1 (positif) yang berhasil diklasifikasikan dengan benar. Nilai *True Negative* (TN) adalah 516 yaitu jumlah data kelas 0 (negatif) yang berhasil diklasifikasikan dengan benar. Sementara itu, terdapat *False Positive* (FP) sebanyak 46, yang berarti data kelas 0 (negatif) salah diprediksi sebagai kelas 1 (positif). Adapun *False Negative* (FN) sebanyak 35, yaitu data kelas 1 (positif) yang salah diklasifikasikan sebagai kelas 0 (negatif). Nilai ini akan digunakan untuk menghitung matrik evaluasi seperti akurasi, presisi, *recall*, dan *f-measure*.

Berdasarkan Gambar 4.3, diperoleh nilai *True Positive* (TP) sebanyak 469, yang menunjukkan jumlah data kelas 1 (positif) yang berhasil diklasifikasikan dengan benar. Nilai *True Negative* (TN) adalah 510 yaitu jumlah data kelas 0 (negatif) yang berhasil diklasifikasikan dengan benar. Sementara itu, terdapat *False Positive* (FP) sebanyak 52, yang berarti data kelas 0 (negatif) salah diprediksi sebagai kelas 1 (positif). Adapun *False Negative* (FN) sebanyak 48, yaitu data kelas 1 (positif) yang salah diklasifikasikan sebagai kelas 0 (negatif). Nilai ini akan digunakan untuk menghitung matrik evaluasi seperti akurasi, presisi, *recall*, dan *f-measure*.

Berikut ini merupakan proses perhitungan nilai akurasi, presisi, *recall*, dan *f-measure* pada kelas positif untuk Gambar 4.2

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$Akurasi = \frac{482 + 516}{482 + 516 + 46 + 35} \times 100\%$$

$$Akurasi = \frac{998}{1079} \times 100\%$$

$$Akurasi = 92,5\%$$

$$Presisi = \frac{TP}{TP + FP} \times 100\%$$

$$Presisi = \frac{482}{482 + 46} \times 100\%$$

$$Presisi = \frac{482}{528} \times 100\%$$

$$Presisi = 91,29\%$$

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

$$Recall = \frac{482}{482 + 35} \times 100\%$$

$$Recall = \frac{482}{517} \times 100\%$$

$$Recall = 93,21\%$$

$$f - measure = \frac{2 \times presisi \times recall}{presisi \times recall} \times 100\%$$

$$f - measure = \frac{2 \times 0,9129 \times 0,9321}{0,9129 \times 0,9321} \times 100\%$$

$$f - measure = \frac{1,7027}{1,8450} \times 100\%$$

$$f - measure = 92,28\%$$

Berikut ini merupakan proses perhitungan nilai akurasi, presisi, *recall*, dan *f-measure* pada kelas positif untuk Gambar 4.3

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$Akurasi = \frac{469 + 510}{469 + 510 + 52 + 48} \times 100\%$$

$$Akurasi = \frac{979}{1079} \times 100\%$$

$$Akurasi = 90,02\%$$

$$Presisi = \frac{TP}{TP + FP} \times 100\%$$

$$Presisi = \frac{469}{469 + 52} \times 100\%$$

$$Presisi = \frac{469}{521} \times 100\%$$

$$Presisi = 90,02\%$$

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

$$Recall = \frac{469}{469 + 48} \times 100\%$$

$$Recall = \frac{469}{517} \times 100\%$$

$$Recall = 90,71\%$$

$$f - measure = \frac{2 \times presisi \times recall}{presisi \times recall} \times 100\%$$

$$f - measure = \frac{2 \times 0.9002 \times 0.9071}{0.9002 \times 0.9071} \times 100\%$$

$$f - measure = \frac{1.6348}{1.8073} \times 100\%$$

$$f - measure = 90,45\%$$

Tabel 4.13 Nilai Akurasi *Testing Naive Bayes*

Akurasi	0,9249			
	precision	recall	f1-score	support
0 (negatif)	0,94	0,92	0,93	562
1 (positif)	0,91	0,93	0,92	517
<i>accuracy</i>			0,92	1079
<i>macro avg</i>	0,92	0,93	0,92	1079
<i>Weighted avg</i>	0,93	0,92	0,92	1079

Tabel 4.14 Nilai Akurasi *Testing Random Forest*

Akurasi	0,9073			
	precision	Recall	f1-score	support
0 (negatif)	0,91	0,91	0,91	562
1 (positif)	0,90	0,91	0,90	517
<i>accuracy</i>			0,91	1079
<i>macro avg</i>	0,91	0,91	0,91	1079
<i>Weighted avg</i>	0,91	0,91	0,91	1079

Berdasarkan hasil pengujian pada Tabel 4.13 dan Tabel 4.14, model *Naive Bayes* menunjukkan performa yang lebih baik dibandingkan dengan *Random Forest* dalam mengklasifikasikan sentimen. *Naive Bayes* mencapai akurasi

sebesar 0,9249 atau sekitar 92,49%, sedangkan *Random Forest* hanya memperoleh akurasi sebesar 0,9073 atau 90,73%. Selain itu, nilai *precision*, *recall*, dan *f1-score* dari model *Naive Bayes* secara konsisten lebih tinggi, baik untuk kelas negatif maupun positif. Untuk kelas negatif, *Naive Bayes* memiliki *f1-score* sebesar 0,93, lebih tinggi dibandingkan *Random Forest* yang hanya mencapai 0,91. Hal serupa juga terjadi pada kelas positif, di mana *f1-score Naive Bayes* sebesar 0,92, mengungguli *Random Forest* yang hanya 0,90. Rata-rata makro dan rata-rata tertimbang (*weighted average*) pada seluruh metrik evaluasi juga menunjukkan keunggulan *Naive Bayes*. Dengan demikian, dapat disimpulkan bahwa model *Naive Bayes* lebih efektif dalam melakukan klasifikasi data pada pengujian ini dibandingkan model *Random Forest*.

4.7 Kajian Keislaman

Berdasarkan data testing berupa ulasan pelanggan sebanyak 1.079 data, model klasifikasi sentimen dibangun untuk mendeteksi apakah suatu ulasan bersifat positif atau negatif. Hasil evaluasi model ditunjukkan melalui confusion matrix, yang mencerminkan seberapa akurat model membaca tanda-tanda lahiriah berupa tulisan pelanggan tersebut. Hal ini sesuai dengan keterangan dalam Kitab *At-Tafsir Al-Wasith* karya Syaikh Tantawi pada penafsiran surat An-Nisa [4] ayat 29.

قال بعضهم : وَحَقِيقَةُ التَّرَاضِي لَا يَعْلَمُهَا إِلَّا اللَّهُ وَالْمَرَاذُهَا هُنَا أَمَارَتُهُ كَالِإِيجَابِ وَالْقَبُولِ وَكَالتَعَاطَى عِنْدَ الْقَائِلِ بِهِ. فَذَلِكَ عَلَى أَنَّ مُجَرَّدَ التَّرَاضِي هُوَ الْمَنَاطُ. وَلَا بُدَّ مِنَ الدَّلَالَةِ عَلَيْهِ بِلَفْظٍ أَوْ إِشَارَةٍ أَوْ كِتَابَةٍ، بِأَيِّ لَفْظٍ وَقَعَ وَعَلَى أَيِّ صِفَةٍ كَانَ ، وَبِأَيِّ إِشَارَةٍ مُفِيدَةٍ حَصَلَ. (ص - 126 - كتب التفسير

الوسيط لطنطاوي - سورة النساء الآيات الى - المكتبة الشاملة)

Menurut beberapa ulama hakikat *at-taradi* (ridho) sejatinya tidak diketahui kecuali oleh Allah Swt. Namun dalam praktik muamalah, ridho dapat diketahui dengan adanya tanda-tanda (*Qorinah*) yang tampak seperti ijab qobul, isyarat atau ucapan berbentuk tulisan seperti ulasan pelanggan. Ucapan berbentuk tulisan disini menjadi bentuk konkret dari kesepakatan batin yang dapat dinilai dan dijadikan dasar hukum. Hal ini menunjukkan bahwa dalam interaksi sosial dan ekonomi, ekspresi melalui tulisan tangan dapat mewakili maksud dan perasaan seseorang.

Sebagaimana ucapan berbentuk tulisan merupakan salah satu indikator dalam mengidentifikasi ridho, maka dalam penelitian ini, ucapan berupa tulisan atau ulasan pelanggan merupakan manifestasi dari pengalaman dan keridhoan mereka terhadap transaksi. Model klasifikasi yang mampu membaca dan memahami makna dari ulasan pelanggan tersebut secara akurat memiliki nilai teknis, namun juga mencerminkan nilai metodologis yang selaras dengan Islam, yaitu membaca niat dan keridhoan melalui bentuk nyata berupa ulasan pelanggan yang tertulis. Dengan demikian indikator ucapan berupa tulisan dapat dijadikan prinsip universal yang berlaku dalam dunia data dan teknologi modern yaitu representasi sah dari makna batin yang dapat diolah, ditafsirkan dan dipertanggungjawabkan.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil analisis confusion matrix, model *Naive Bayes* menunjukkan performa klasifikasi yang lebih baik dibandingkan *Random Forest*. Model *Naive Bayes* menghasilkan *True Positive* (TP) sebanyak 482 dari 517 data positif dan *True Negative* (TN) sebanyak 516 dari 562 data negatif, dengan jumlah *False Positive* (FP) sebanyak 46 dan *False Negative* (FN) sebanyak 35. Sedangkan model *Random Forest* menghasilkan TP sebanyak 469 dari 517 data positif dan TN sebanyak 510 dari 562 data negatif, dengan jumlah FP sebanyak 52 dan FN sebanyak 48.

Selain itu, berdasarkan metrik evaluasi, model *Naive Bayes* juga unggul dalam semua aspek pengukuran performa. *Naive Bayes* mencatatkan akurasi sebesar 0,9249 atau 92,49%, lebih tinggi dibandingkan *Random Forest* yang hanya memperoleh akurasi sebesar 0,9073 atau 90,73%. Selain itu, nilai *precision*, *recall*, dan *f1-score* dari model *Naive Bayes* secara konsisten lebih tinggi, baik untuk kelas negatif maupun positif. Untuk kelas negatif, *Naive Bayes* memiliki *f1-score* sebesar 0,93, lebih tinggi dibandingkan *Random Forest* yang hanya mencapai 0,91. Hal serupa juga terjadi pada kelas positif, di mana *f1-score* *Naive Bayes* sebesar 0,92, mengungguli *Random Forest* yang hanya 0,90. Rata-rata makro dan rata-rata tertimbang (*weighted average*) pada seluruh metrik evaluasi juga menunjukkan keunggulan *Naive Bayes*. Dengan demikian, dapat disimpulkan bahwa model *Naive Bayes* lebih efektif dalam melakukan klasifikasi data pada pengujian ini dibandingkan model *Random Forest*.

5.2 Saran

Hasil penelitian ini menunjukkan bahwa masih diperlukan upaya lebih lanjut untuk meningkatkan kinerja sistem. Oleh karena itu, penulis menyarankan beberapa hal untuk pengembangan penelitian selanjutnya.

1. Menerapkan pendekatan *Aspect-Based Sentiment Analysis* untuk analisis mendalam dan dapat mengidentifikasi aspek-aspek spesifik dari produk yang perlu diperbaiki.
2. Menambahkan kelas sentimen netral supaya memberikan gambaran yang lebih realistis terhadap persepsi pengguna.
3. Menggunakan data ulasan yang lebih luas dan bervariasi. Hal ini bertujuan untuk meningkatkan kemampuan generalisasi model terhadap berbagai jenis data.
4. Menggunakan representasi kata yang lebih canggih seperti, *Word2Vec*, *Glove*, atau *FastText* yang dapat menangkap konteks dan makna kata secara lebih baik dibandingkan TF-IDF.

DAFTAR PUSTAKA

- Abbas, M., Ali, K., Jamali, A., Ali Memon, K., & Aleem Jamali, A. (2019). Multinomial Naive Bayes Classification Model for Sentiment Analysis. *IJCSNS International Journal of Computer Science and Network Security*, 19(3), 62–67. <https://doi.org/10.13140/RG.2.2.30021.40169>
- Akbari, M. I. H. A. D., Novianty, A., & Setianingsih, C. (2017). Analisis Sentimen Menggunakan Metode Learning Vector Quantization. *E-Proceeding of Engineering*, 4(2), 2283. https://openlibrary.telkomuniversity.ac.id/pustaka/files/135356/jurnal_eproc/analisis-sentimen-menggunakan-metode-learning-vector-quantization.pdf
- Alasadi, S. A., & Bhaya, W. S. (2017). Review of data preprocessing techniques in data mining. In *Journal of Engineering and Applied Sciences* (Vol. 12, Issue 16, pp. 4102–4107). <https://doi.org/10.3923/jeasci.2017.4102.4107>
- Anam, K., Rinaldi, A. R., & Fathurrohman, F. (2024). Komparasi Algoritma Machine Learning Dalam Klasifikasi Loyalitas Nasabah Bank Berbasis Particle Swarm Optimization. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(4), 8212–8218. <https://doi.org/10.36040/jati.v8i4.10941>
- Fahmi, H., & Sari, W. P. (2021). *Opinion Mining Analysis on Online Product Reviews Using Naïve Bayes and Feature Selection*. IEEE. <https://doi.org/10.1109/ICIMTech53080.2021.9535081>
- Fauzan, A. C., & Hikmah, K. (2022). Implementasi Algoritma Naive Bayes Dalam Analisis Polarisasi Opini Masyarakat Terkait Vaksin Covid-19. *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab*, 7(2), 122–128. <https://doi.org/10.36341/rabit.v7i2.2403>
- Indrayuni, E. (2019). Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes. *Jurnal Khatulistiwa Informatika*, 7(1), 29–36. <https://doi.org/10.31294/jki.v7i1.1>
- Jurafsky, D., & Martin, J. H. (2009). Speech and Language Processing. In *Upper Saddle River, N.J.* (Pearson in). Pearson Education.
- Kamber, M., & Han, J. (2006). Data Mining: Concepts and Techniques : Concepts and Techniques. In *The Fundamentals of Political Science Research*.
- Khotimah, N., Darsyah, M. Y., & Nur, I. M. (2019). Analisis Sentimen Terhadap Review E-Commerce Dengan Metode Stochastic Gradient Descent. *Skripsi*, 1–11.
- Liu, B. (2020). Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Second Edition. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Second Edition*, May, 1–432. <https://doi.org/10.1017/9781108639286>
- Nimri, R., Phillip, M., & Kovatchev, B. (2022). Decision Support Systems and Closed-Loop. *Diabetes Technology and Therapeutics*, 24(S1), S58–S75. <https://doi.org/10.1089/dia.2022.2504>

- Novendra, C. S. (2022). Analysis of William Tanuwijaya Leadership Style as a CEO of Tokopedia. *Researchgate.Net*, June.
- Nur'aini, S. (2023). Peningkatan Perekonomian Indonesia Melalui E-Commerce. *Jurnal Andromeda*, 1(1).
- Nurjannah, M., & Fitri Astuti, I. (2013). Penerapan Algoritma Term Frequency-Inverse Document (TF-IDF) Untuk Text Mining. *Jurnal Informatika Mulawarman*, 8(3), 110–113.
- Prasetyo, A., & Adriyanto, A. T. (2022). “the Effect of Celebrity Endorsment Towards Purchasing Intention Through Brand Perception As Intervening Variable on Shopee Consumers During the Covid-19 Pandemic.” *International Journal of Economics, Business and Accounting Research (IJEBAR)*, 6(1), 489. <https://doi.org/10.29040/ijebar.v6i1.4472>
- Rezki, M., Kholifah, D. N., Faisal, M., Priyono, P., & Suryadithia, R. (2020). Analisis Review Pengguna Google Meet dan Zoom Cloud Meeting Menggunakan Algoritma Naïve Bayes. *Jurnal Infortech*, 2(2), 264–270. <https://doi.org/10.31294/infortech.v2i2.9286>
- Setio, P. B. N., Saputro, D. R. S., & Bowo Winarno. (2020). Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5. *PRISMA, Prosiding Seminar Nasional Matematika*, 3, 64–71.
- Singh, N., & Gaur, B. (2019). Data Preprocessing: A Step-by-Step Guide for Clean and Usable Data. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 10(2), 1148–1153. <https://doi.org/10.61841/turcomat.v10i2.14384>
- Stephenie, Warsito, B., & Prahutama, A. (2020). Sentiment Analysis on Tokopedia Product Online Reviews Using Random Forest Method. *E3S Web of Conferences*, 202, 1–10. <https://doi.org/10.1051/e3sconf/202020216006>
- Sugiyono. (2013). *Metodologi Penelitian Kuantitatif, Kualitatif dan R & D*. Alfa Beta.
- Suhartono, Kurniawan, F., & Imran, B. (2018). Identification of virtual plants using bayesian networks based on parametric L-system. *International Journal of Advances in Intelligent Informatics*, 4(1), 40–52. <https://doi.org/10.26555/ijain.v4i1.157>
- Sun, W., Cai, Z., Li, Y., Liu, F., Fang, S., & Wang, G. (2018). Data processing and text mining technologies on electronic medical records: A review. *Journal of Healthcare Engineering*, 2018. <https://doi.org/10.1155/2018/4302425>
- Supartini, I. A. M., Sukarsa, I. K. G., & Srinadi, I. G. A. M. (2017). Analisis Diskriminan Pada Klasifikasi Desa Di Kabupaten Tabanan Menggunakan Metode K-Fold Cross Validation. *E-Jurnal Matematika*, 6(2), 106. <https://doi.org/10.24843/mtk.2017.v06.i02.p154>
- Surbakti, A. Q., Hayami, R., & Al Amien, J. (2021). Analisa Tanggapan Terhadap Psbb Di Indonesia Dengan Algoritma Decision Tree Pada Twitter. *Jurnal CoSciTech (Computer Science and Information Technology)*, 2(2), 91–97.

<https://doi.org/10.37859/coscitech.v2i2.2851>

- Syahrizal, H., & Jailani, M. S. (2023). Jenis-Jenis Penelitian Dalam Penelitian Kuantitatif dan Kualitatif. *Jurnal QOSIM Jurnal Pendidikan Sosial & Humaniora*, 1(1), 13–23. <https://doi.org/10.61104/jq.v1i1.49>
- Syakira, S. N., & Moeliono, N. N. K. (2019). Pengaruh online consumer review oleh beauty vlogger suhay salim terhadap proses keputusan pembelian produk freeman feeling beautiful facial masks. *EProceedings of Management*, 6(2), 4412–4420.
- Syarifuddinn, M. (2020). Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naïve Bayes Dan Knn. *INTI Nusa Mandiri*, 15(1), 23–28. <https://doi.org/10.33480/inti.v15i1.1347>
- Wahyu Sejati, Ankur Singh Bist, & Amirsyah Tambunan. (2023). Pengembangan Analisis Sentimen dalam Rekayasa Software Engineering menggunakan tinjauan literatur sistematis. *Jurnal MENTARI: Manajemen, Pendidikan Dan Teknologi Informasi*, 2(1), 95–103. <https://doi.org/10.33050/mentari.v2i1.377>

LAMPIRAN

Lampiran 1 Perhitungan Manual TF-IDF

<https://bit.ly/4dVwVAI>

Lampiran 2 Perhitungan Manual Naive Bayes

<https://bit.ly/3ZrOeTD>

RIWAYAT HIDUP



Mochammad Aldi Yusuf, lahir di Malang, 15 Januari 2001. Bertempat tinggal di Perum. Asrikaton Indah Blok H 9 No 4, Kecamatan Pakis, Kabupaten Malang. Penulis merupakan anak pertama dari tiga bersaudara. Putra dari Bapak Suseno dan Ibu Sulistyoningsih. Penulis sudah melakukan perjalanan dalam pendidikan formal dan agama dari TK Siti Khadijah dan lulus pada tahun 2007. Kemudian melanjutkan perjalanannya menuju pendidikan dasar di SDN Pandanwangi 5, Kota Malang dan lulus tahun 2013, lalu melanjutkan perjalanan menempuh pendidikan agama di Pondok Pesantren An- Nur 2 Al Murtadlo, Bululawang, Malang selama 6 tahun dan pendidikan formal di SMP An-Nur, Bululawang, Malang dan lulus tahun 2016 serta SMA An- Nur, Bululawang, Malang dan lulus tahun 2019. Selanjutnya penulis menempuh pendidikan tinggi di Universitas Islam Negeri Malang melalui jalur SNMPTN pada tahun 2019. Selama menempuh pendidikan tinggi, penulis turut berkontribusi aktif dalam beberapa kegiatan eksternal kampus. Beberapa pengalaman yang pernah dilalui oleh penulis yaitu sebagai Ketua Koordinator IKSAN Kampus (Ikatan Santri Alumni An Nur regional Kampus) UIN selama satu periode dan anggota aktif IKSAN KAMPUS. Penulis memiliki ketertarikan terhadap bidang data analisis. Penulis menerima kritik dan saran atas penelitian ini. Kritik dan saran berkaitan penelitian ini dapat dikirim pada email penulis yaitu aldiyusuf981@gmail.com.



BUKTI KONSULTASI SKRIPSI

Nama : Mochammad Aldi Yusuf
NIM : 19610009
Fakultas / Jurusan : Sains dan Teknologi / Matematika
Judul Skripsi : Analisis Sentimen Menggunakan Algoritma Naive Bayes
Pada Ulasan Produk E-Commerce (Studi Kasus: Tokopedia)
Pembimbing I : Hisyam Fahmi, M. Kom.
Pembimbing II : Muhammad Nafie Jauhari, M.Si.

No	Tanggal	Hal	Tanda Tangan
1.	23 April 2024	Konsultasi Topik dan Data	1.
2.	27 Agustus 2024	Konsultasi Bab I, II, dan III	2.
3.	2 Oktober 2024	Konsultasi Bab I, II, dan III	3.
4.	9 Oktober 2024	Konsultasi Bab I, II, dan III	4.
5.	22 Oktober 2024	Konsultasi Bab I, II, dan III	5.
6.	15 Oktober 2024	Konsultasi Kajian Agama Bab I dan II	6.
7.	16 Oktober 2024	ACC Kajian Agama Bab I dan II	7.
8.	1 November 2024	ACC BAB I, II, dan III	8.
9.	1 November 2024	ACC Seminar Proposal	9.
10.	20 April 2025	Konsultasi Revisi Seminar Proposal	10.
11.	28 April 2025	Konsultasi Bab IV dan V	11.
12.	16 Mei 2025	Konsultasi Bab IV dan V	12.
13.	20 Mei 2025	Konsultasi Bab IV dan V	13.
14.	21 Mei 2025	Konsultasi Bab IV dan V	14.
15.	23 Mei 2025	ACC Bab IV dan V	15.



KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI
MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Gajayana No.50 Dinoyo Malang Telp. / Fax. (0341) 558933

16.	22 Mei 2025	Konsultasi Kajian Agama Bab IV	15.
17.	23 Mei 2025	ACC Kajian Agama Bab IV	17.
18.	23 Mei 2025	ACC Seminar Hasil	18.
19.	4 Juni 2025	Konsultasi Revisi Seminar Hasil	19.
20.	5 Juni 2025	Konsultasi Revisi Seminar Hasil	20.
21.	13 Juni 2025	ACC Sidang Skripsi	21.
22.	17 Juni 2025	ACC Keseluruhan	22.

Malang, 17 Juni 2025

Mengetahui,

Ketua Program Studi Matematika



Dr. Elly Susanti, M.Sc.

NIP. 19741129 200012 2 005