



**VISUALISASI DATA ANALISIS SENTIMEN TERHADAP APLIKASI *E-COMMERCE* ZALORA BERDASARKAN ULASAN DI *GOOGLE PLAY STORE* MENGGUNAKAN METODE *NAIVE BAYES CLASSIFIER***

**SKRIPSI**

**RADITYA RAFI HARIMASYA**

**2010512112**

**PROGRAM STUDI S1 SISTEM INFORMASI**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAKARTA**

**2024**



**VISUALISASI DATA ANALISIS SENTIMEN TERHADAP APLIKASI *E-COMMERCE* ZALORA BERDASARKAN ULASAN DI *GOOGLE PLAY STORE* MENGGUNAKAN METODE *NAIVE BAYES CLASSIFIER***

**SKRIPSI**

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh  
Gelar Sarjana Komputer**

**RADITYA RAFI HARIMASYA**

**2010512112**

**PROGRAM STUDI S1 SISTEM INFORMASI**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAKARTA**

**2024**

## PERNYATAAN ORISINALITAS

Tugas akhir ini merupakan hasil karya sendiri dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan benar.

Nama : Raditya Rafi Harimasya

NIM : 2010512112

Program Studi : S1 - Sistem Informasi

Judul : VISUALISASI DATA ANALISIS SENTIMEN TERHADAP APLIKASI *E-COMMERCE*  
ZALORA BERDASARKAN ULASAN DI *GOOGLE PLAY STORE* MENGGUNAKAN  
METODE *NAIVE BAYES CLASSIFIER*

Bilamana dikemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 15 Juni 2024



Raditya Rafi Harimasya

**PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK  
KEPENTINGAN AKADEMIS**

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Raditya Rafi Harimasya

NIM : 2010512112

Fakultas : Fakultas Ilmu Komputer

Program Studi : S1 Sistem Informasi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty Free Right*) atas skripsi saya yang berjudul **"VISUALISASI DATA ANALISIS SENTIMEN TERHADAP APLIKASI E-COMMERCE ZALORA BERDASARKAN ULASAN DI GOOGLE PLAY STORE MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER"**.

Dengan Hak Bebas Royalti ini, Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/memformatkan, mengelola dalam bentuk pangkalan data (basis data), merawat, dan mempublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta, dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Ditetapkan di : Jakarta

Tanggal Persetujuan : 15 Juni 2024

Yang Menyatakan,



(Raditya Rafi Harimasya)

## LEMBAR PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Raditya Rafi Harimasya

NIM : 2010512112

Program Studi : S1 Sistem Informasi

Judul Skripsi/TA : Visualisasi Data Analisis Sentimen Terhadap Aplikasi *E-commerce* Zalora Berdasarkan Ulasan di *Google Play Store* Menggunakan Metode *Naive Bayes Classifier*

Telah berhasil dipertahankan di hadapan tim penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar sarjana pada program studi S1 Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



**Ati Zaidiah, S.Kom., MTL.**  
Dosen Pembimbing I



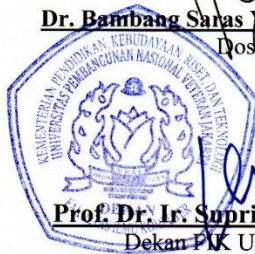
**Nindy Irzavika, S.SI., M.T.**  
Dosen Pembimbing II



**Dr. Bambang Saras Yulistiawan, S.T., M.Kom.**  
Dosen Penguji I



**Iin Ernawati, S.Kom., M.Si.**  
Dosen Penguji II



**Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM**  
Dekan FK UPN Veteran Jakarta



**Anita Muliawati, S.Kom., MTL.**  
Ketua Program Studi S1 Sistem Informasi

Ditetapkan di : Jakarta  
Tanggal Ujian : 31 Mei 2024

# VISUALISASI DATA ANALISIS SENTIMEN TERHADAP APLIKASI *E-COMMERCE* ZALORA BERDASARKAN ULASAN DI *GOOGLE PLAY STORE* MENGGUNAKAN METODE *NAIVE BAYES CLASSIFIER*

Raditya

## ABSTRAK

Perdagangan elektronik, atau *e-commerce*, adalah aktivitas jual-beli melalui internet yang meliputi pembelian, penjualan barang, layanan, dan informasi secara *online*. Pertumbuhan *e-commerce* dipicu oleh kemajuan teknologi dan internet, yang memunculkan berbagai perusahaan *e-commerce*, termasuk ZALORA, sebuah platform *e-commerce* yang berfokus pada *fashion*, menghadapi tantangan dalam menjaga peringkatnya yang menurun di pasar yang kompetitif. Evaluasi dilakukan melalui analisis sentimen pengguna dengan metode *Naïve Bayes Classifier* terhadap ulasan di *Google Play Store*. Data dikumpulkan dari Desember 2022 hingga Oktober 2023. Data melalui tahapan *preprocessing*, pembobotan kata menggunakan *TF-IDF*, dan pembagian data. *Naïve Bayes* menghasilkan *accuracy* 86%, *precision* 95%, *recall* 88%, dan *F1-Score* 91% dengan rasio data latih dan uji 70:30. Selain hasil klasifikasi, terdapat hasil visualisasi *word cloud* beserta grafik frekuensi kata terbanyak yang digunakan untuk melakukan identifikasi sentimen positif dan negatif. Luaran lainnya yaitu dilakukan pembuatan sistem *website* sederhana menggunakan *Framework Streamlit* untuk melakukan visualisasi hasil.

**Kata Kunci :** Analisis Sentimen, *Google Play Store*, *Naïve Bayes classifier*, Zalora

*DATA VISUALIZATION OF SENTIMENT ANALYSIS ON ZALORA E-COMMERCE  
APPLICATION BASED ON REVIEWS ON GOOGLE PLAY STORE USING NAIVE  
BAYES CLASSIFIER*

**Raditya Rafi Harimasya**

***ABSTRACT***

*Electronic commerce, or e-commerce, is the activity of buying and selling goods and services over the internet, encompassing online purchases, sales, services, and information. The growth of e-commerce is driven by technological advancements and the internet, leading to the emergence of various e-commerce companies, including ZALORA, an e-commerce platform focusing on fashion, which faces challenges in maintaining its declining ranking in the competitive market. Evaluation is conducted through user sentiment analysis using the Naïve Bayes Classifier method on reviews from the Google Play Store. Data were collected from December 2022 to October 2023 and underwent preprocessing, word weighting using TF-IDF, and data splitting. Naïve Bayes achieved an accuracy of 86%, precision of 95%, recall of 88%, and F1-Score of 91% with an 70:20 ratio of training and testing data. In addition to classification results, there are visualization outputs such as word clouds and graphs depicting the frequency of most-used words for identifying positive and negative sentiments. Another output is the development of a simple website system using the Streamlit Framework for visualization purposes.*

***Keywords:*** *Sentiment Analysis, Google Play Store, Naïve Bayes classifier, Zalora*

## KATA PENGANTAR

Puji dan syukur dipanjatkan kepada Tuhan Yang Maha Esa atas rahmat dan karunianya sehingga penulis dapat menyelesaikan penyusunan tugas akhir yang berjudul “VISUALISASI DATA ANALISIS SENTIMEN TERHADAP APLIKASI E-COMMERCE ZALORA BERDASARKAN ULASAN DI GOOGLE PLAY STORE MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER” dengan baik. Pada kesempatan ini, penulis ingin menyampaikan terima kasih kepada seluruh pihak yang telah ikut mendukung penulis dalam menyelesaikan skripsi ini hingga selesai, terutama kepada:

1. Kedua orang tua peneliti yang selalu mendukung seluruh kegiatan yang dilakukan peneliti.
2. Ibu Ati Zaidiah, S.Kom., MTI. selaku dosen pembimbing 1 yang telah memberikan saran dan bimbingan yang sangat bermanfaat dari awal penyusunan skripsi hingga selesai.
3. Ibu Nindy Irzavika, S.SI., M.T. selaku dosen pembimbing 2 yang juga telah memberikan saran dan bimbingan yang sangat bermanfaat dari awal penyusunan skripsi hingga selesai.
4. Ibu Anita Muliawati, S.Kom., MTI. selaku Kepala Jurusan Sistem Informasi di Universitas Pembangunan Nasional Veteran Jakarta.
5. Vincentius Ludwig dan Hanifa Widya, selaku teman dekat peneliti yang senantiasa memberikan dukungan serta saran selama penyusunan skripsi ini.
6. Seluruh pihak yang terlibat baik secara langsung maupun tidak langsung yang membantu dalam penyusunan skripsi ini.

Peneliti menyadari bahwa dalam proses penulisan tugas akhir ini masih terdapat beberapa kekurangan baik dari materi hingga Teknik penulisan. Oleh karena itu, peneliti sangat mengharapkan masukan dan kritik yang konstruktif sebagai bahan perbaikan di masa mendatang agar memberikan bermanfaat yang signifikan bagi peneliti sendiri serta semua pihak yang memiliki kepentingan terkait.



Jakarta, 22 April 2024

A handwritten signature in black ink, consisting of a large, stylized initial 'P' followed by a cursive name and a long horizontal flourish.

Peneliti

## DAFTAR ISI

<b>PERNYATAAN ORISINALITAS.....</b>	<b>ii</b>
<b>PERNYATAAN PERSETUJUAN PUBLIKASI.....</b>	<b>iii</b>
<b>ABSTRAK.....</b>	<b>iv</b>
<b>ABSTRACT.....</b>	<b>v</b>
<b>KATA PENGANTAR.....</b>	<b>vi</b>
<b>LEMBAR PENGESAHAN.....</b>	<b>viii</b>
<b>DAFTAR SIMBOL.....</b>	<b>xiii</b>
<b>DAFTAR GAMBAR.....</b>	<b>xv</b>
<b>DAFTAR TABEL.....</b>	<b>xvii</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	7
1.3. Tujuan Penelitian.....	7
1.4. Manfaat Penelitian.....	7
1.5. Batasan Penelitian.....	8
1.6. Sistematika Penulisan.....	8
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>10</b>
2.1. <i>E-Commerce</i> .....	10
2.2. ZALORA.....	10
2.3. <i>Google Play Store</i> .....	11
2.4. <i>Python</i> .....	11
2.5. <i>Text Mining</i> .....	11



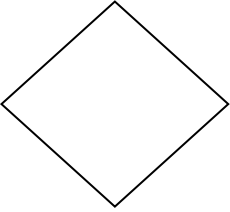
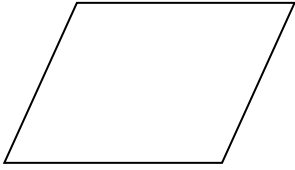

2.6. Analisis Sentimen.....	12
2.7. <i>Text Preprocessing</i> .....	13
2.7.1. <i>Data Cleansing</i> .....	13
2.7.2. <i>Case Folding</i> .....	13
2.7.3. <i>Spelling Normalization</i> .....	14
2.7.4. <i>Stemming</i> .....	14
2.7.5. <i>Stopword Removal / Filtering</i> .....	14
2.7.6. <i>Tokenizing</i> .....	14
2.8. <i>Term Weighting</i> .....	15
A. <i>Term Frequency (TF)</i> .....	15
B. <i>Inverse Document Frequency (IDF)</i> .....	15
2.9. <i>Naïve Bayes Classifier</i> .....	16
2.10. <i>Confusion matrix</i> .....	17
2.11. Evaluasi Performa Model.....	18
2.11.1. <i>Accuracy (Akurasi)</i> .....	18
2.11.2. <i>Precision (Presisi)</i> .....	18
2.11.3. <i>Recall</i> .....	19
2.11.3. <i>F1-Score</i> .....	19
2.12. <i>Word Cloud</i> .....	19
2.13. <i>Streamlit</i> .....	20
2.14. <i>Blackbox Testing</i> .....	20
2.15. Kajian Literatur.....	20
<b>BAB III METODE PENELITIAN.....</b>	<b>24</b>
3.1. Lokasi dan Waktu Penelitian.....	24

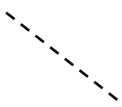
3.2.	Alur Penelitian.....	24
3.3.	Identifikasi Masalah.....	25
3.4.	Studi Pustaka.....	25
3.4.1.	<i>Data Scraping</i> (Pengumpulan Data).....	25
3.4.2.	<i>Data Labelling</i> (Pelabelan Data).....	26
3.4.3.	<i>Data Preprocessing</i> .....	26
3.4.4.	<i>Term Weighting</i> (Pembobotan).....	30
3.4.5.	<i>Splitting Data</i> .....	30
3.4.6.	<i>Model Training</i> .....	31
3.4.7.	Evaluasi Performa Model.....	31
3.4.8.	Visualisasi <i>Wordcloud</i> .....	31
3.4.9.	Klasifikasi dan Visualiasi Sentimen Terhadap Sebuah Aspek.....	32
3.4.10.	Perancangan Sistem.....	32
3.4.11.	Implementasi dan Pengujian Sistem.....	32
3.5.	Alat Bantu Penelitian.....	33
3.6.	Waktu dan Tempat Penelitian.....	33
3.7.	Jadwal Penelitian.....	33
<b>BAB IV HASIL DAN PEMBAHASAN.....</b>		<b>36</b>
4.1.	Pengumpulan Data.....	36
4.2.	Data Labeling.....	37
4.3.	Preprocessing.....	38
4.3.1.	Data Cleansing (Pembersihan Data).....	38
4.3.2.	Case Folding.....	39
4.3.3.	<i>Spelling Normalization</i> .....	40

4.3.4.	<i>Stemming</i> .....	40
4.3.5.	<i>Stopword Removal</i> .....	41
4.3.6.	<i>Tokenizing</i> .....	42
4.4.	<i>Term Weighting</i> (Pembobotan).....	42
4.5.	<i>Splitting Data</i> (Pembagian Data).....	44
4.6.	Klasifikasi Naïve Bayes Classifier.....	45
4.7.	Evaluasi Model Naïve Bayes Classifier.....	45
4.8.	Analisis Hasil Klasifikasi dan Visualisasi.....	47
4.8.1.	Visualisasi <i>Wordcloud</i> Sentimen Keseluruhan.....	47
4.8.2.	Visualisasi <i>Wordcloud</i> Sentimen Positif.....	48
4.8.3.	Visualisasi <i>Wordcloud</i> Sentimen Negatif.....	50
4.8.4.	Klasifikasi dan Visualisasi Sentimen Terhadap Berbagai Aspek.....	51
4.9.	Perancangan Sistem.....	58
4.10.	Implementasi dan Pengujian Sistem.....	58
4.10.1.	Tampilan Halaman <i>Home</i> .....	59
4.10.2.	Tampilan Halaman Informasi Data.....	60
4.10.3.	Tampilan Halaman <i>Naïve Bayes Classifier</i> .....	60
4.10.4.	Tampilan Halaman <i>Wordcloud</i> .....	62
4.10.5.	Tampilan Halaman Sentimen Berdasarkan Aspek.....	63
<b>BAB V PENUTUP</b> .....		<b>67</b>
5.1.	Kesimpulan.....	67
5.2.	Saran.....	68
<b>DAFTAR PUSTAKA</b> .....		<b>69</b>
<b>DAFTAR RIWAYAT HIDUP</b> .....		<b>71</b>

<b>LAMPIRAN.....</b>	<b>73</b>
----------------------	-----------

## DAFTAR SIMBOL

No	Simbol	Nama	Keterangan
1		<i>Terminator</i>	Simbol yang berfungsi untuk menggambarkan kegiatan awal atau akhir sebuah alur proses.
2		<i>Process</i>	Simbol yang berfungsi untuk menggambarkan suatu proses yang berjalan pada sebuah alur proses
3		<i>Decision</i>	Simbol yang berfungsi untuk menggambarkan suatu pengambilan keputusan terhadap langkah selanjutnya pada alur proses.
4		<i>Data</i>	Simbol yang berfungsi untuk menggambarkan suatu input atau output dari sebuah data dalam suatu alur proses
5		<i>Flow Line</i>	Simbol yang berfungsi untuk menghubungkan satu simbol dengan simbol lainnya yang

			memiliki keterkaitan dalam suatu alur proses.
6		<i>Dotted Line</i>	Simbol yang berfungsi menggambarkan alur proses yang menunjukkan hubungan antara suatu langkah yang dapat bersifat opsional atau bergantung pada kondisi tertentu.



## DAFTAR GAMBAR

Gambar 1. 1. Pengunjung bulanan aplikasi <i>e-commerce fashion</i> indonesia kuartal II 2021 (databoks, 2021).....	2
Gambar 1. 2. List pengunjung bulanan <i>e-commerce</i> kuartal I 2022 di indonesia (databoks, 2022).....	3
Gambar 1. 3. Perbandingan pengunjung e-commerce dalam kurun waktu 2019 – 2022 di Indonesia (iprice, 2022).....	4
Gambar 3. 1. Diagram Alur Penelitian.....	24
Gambar 3. 2. Contoh Visualisasi Wordcloud.....	37
Gambar 4. 2. Hasil Pelabelan Data ulasan Zalora.....	38
Gambar 4. 3. Hasil Pembersihan Data ( <i>Cleaning</i> ) Terhadap Data Ulasan.....	39
Gambar 4. 4. Hasil <i>Confusion Matrix</i> dengan Rasio data 70 : 30.....	46
Gambar 4. 5. <i>Word Cloud</i> Label Positif dan Negatif.....	47
Gambar 4. 6. Grafik <i>Word Cloud</i> Label Positif dan Negatif.....	48
Gambar 4. 7. <i>Word Cloud</i> Label Positif.....	49
Gambar 4. 8. Grafik <i>Word Cloud</i> Label Positif.....	49
Gambar 4. 9. <i>Word Cloud</i> Label Negatif.....	50
Gambar 4. 10. Grafik <i>Word Cloud</i> Label Negatif.....	51
Gambar 4. 11. Visualisasi Kata Terbanyak Terhadap Aspek Barang.....	52
Gambar 4. 12. Visualisasi Kata Terbanyak Terhadap Aspek Layanan.....	54
Gambar 4. 13. Visualisasi Kata Terbanyak Terhadap Aspek Pengiriman.....	55
Gambar 4. 14. Visualisasi Kata Terbanyak Terhadap Aspek Pembayaran.....	56
Gambar 4. 15. Visualisasi Kata Terbanyak Terhadap Aspek Fitur.....	57
Gambar 4. 16. Tampilan Halaman <i>Home</i> .....	58
Gambar 4. 17. Tampilan Halaman Informasi Data.....	60

Gambar 4. 18. Tampilan Halaman Naive Bayes Classifier.....	61
Gambar 4. 19. Tampilan Halaman <i>Wordcloud</i> .....	62
Gambar 4. 20. Tampilan Halaman Sentimen Berdasarkan Aspek.....	64

## DAFTAR TABEL

Tabel 2. 1. Confusion Matrix.....	18
Tabel 2. 2. Penelitian Terdahulu.....	20
Tabel 3. 1. Pelabelan Berdasarkan <i>Score</i> .....	26
Tabel 3. 2. Sebelum dan Sesudah Cleaning.....	27
Tabel 3. 3. Sebelum dan Sesudah <i>Case Folding</i> .....	27
Tabel 3. 4. Sebelum dan Sesudah Normalization.....	28
Tabel 3. 5. Sebelum dan Sesudah Stemming.....	28
Tabel 3. 6. Sebelum dan Sesudah Stopword Removal (Filtering) .....	29
Tabel 3. 7. Sebelum dan Sesudah Tokenizing .....	30
Tabel 3. 6. Confusion Matrix.....	31
Tabel 3. 7. Jadwal Penelitian.....	34
Tabel 4. 1. Tahapan <i>Preprocessing Case Folding</i> .....	39
Tabel 4. 2. Tahapan <i>Preprocessing Spelling Normalization</i> .....	40
Tabel 4. 3. Tahapan <i>Preprocessing Stemming</i> .....	41
Tabel 4. 4. Tahapan <i>Preprocessing Stopword Removal</i> .....	41
Tabel 4. 5. Tahapan <i>Preprocessing Tokenizing</i> .....	42
Tabel 4. 6. Data Sampel Perhitungan <i>TF-IDF</i> Pada Data Sampel.....	42
Tabel 4. 7. Hasil Perhitungan Pembobotan <i>TF-IDF</i> Pada Data Sampel.....	43
Tabel 4. 8. Pembagian Data dengan Rasio 70:30.....	44
Tabel 4. 9. Pembagian Data dengan Rasio 80:20.....	44
Tabel 4. 10. Pembagian Data dengan Rasio 90:10.....	45
Tabel 4. 11. Black-Box Testing Test Case Terhadap Sistem Website Sederhana.....	65