



**ANALISIS SENTIMEN ULASAN *CASH ON DELIVERY* DI  
SHOPEE MENGGUNAKAN METODE *NAIVE BAYES***

**SKRIPSI**

**MHARTIAN JORDAN HUTASOIT**

**NIM. 1710511048**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA**

**FAKULTAS ILMU KOMPUTER**

**PROGRAM STUDI S1 INFORMATIKA**

**2024**



**ANALISIS SENTIMEN ULASAN *CASH ON DELIVERY* DI  
SHOPEE MENGGUNAKAN *NAIVE BAYES***

**SKRIPSI**

**Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Komputer Pada  
Fakultas Ilmu Komputer**

**MHARTIAN JORDAN HUTASOIT**

**NIM. 1710511048**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA**

**FAKULTAS ILMU KOMPUTER**

**PROGRAM STUDI S1 INFORMATIKA**

**2024**

## PERNYATAAN ORISINALITAS

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

Nama : Mhartian Jordan Hutasoit

NIM : 1710511048

Tanggal : 31 Juli 2024

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

Jakarta, 31 Juli 2024

Yang menyatakan



(Mhartian Jordan Hutasoit)

## **PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

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

Nama : Mhartian Jordan Hutasoit

NIM : 1710511048

Fakultas : Ilmu Komputer

Program Studi : Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non Eksklusif (Non-Exclusive Royalti Free Right) atas karya ilmiah saya yang berjudul:

### **ANALISIS SENTIMEN ULASAN *CASH ON DELIVERY* DI SHOPEE MENGUNAKAN METODE *NAIVE BAYES***

Dengan hak bebas royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan 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.

Dibuat di : Jakarta

Pada Tanggal : 31 Juli 2024



(Mhartian Jordan Hutasoit)

# LEMBAR PENGESAHAN


## LEMBAR PENGESAHAN SKRIPSI

Dengan ini dinyatakan bahwa Tugas Akhir/Skripsi berikut :

Nama : Mhartian Jordan Hutasoit  
NIM : 1710511048  
Program Studi : S1 Informatika  
Judul Tugas Akhir : Analisis Sentimen Ulasan *Cash On Delivery* di Shopeee Menggunakan *Naive Bayes*

Telah berhasil dipertahankan dihadapan Tim Penguji dan diterima sebagai bagian dari persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jakarta

  
(Dr. Widya Cholij, M.I.T)  
Penguji I

  
(Muhammad Adrezo, S.Kom., M.Sc)  
Penguji II

  
(Bayu Hananto, S.Kom., M.Kom.)  
Pembimbing

  
(Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM)  
Dekan Fakultas Ilmu Komputer

  
(Dr. Widya Cholij, M.I.T)  
Ketua Program Studi

Disahkan di : Jakarta  
Tanggal : 30 Juli 2024



## KATA PENGANTAR

Puji dan syukur kepada Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi. Skripsi ini disusun untuk memenuhi persyaratan untuk memperoleh gelar Sarjana Komputer pada Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jakarta.

Dalam penyusunan skripsi ini tidak terlepas dari dukungan, bantuan, dan bimbingan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih sebesar-besarnya kepada:

- 1 Bapak Bayu Hananto, M.Kom. selaku dosen pembimbing dari pihak jurusan, yang telah memberikan arahan dan bimbingan dalam penyusunan skripsi ini.
- 2 Ibu Dr. Ermatita, M.Kom., selaku Dekan Fakultas Ilmu Komputer, yang telah memberikan kesempatan dan fasilitas untuk melaksanakan penelitian ini.
- 3 Dr. Widya Cholil, M.I.T selaku Ketua Program Studi Sarjana Jurusan S1 Informatika.
- 4 Orang tua dan keluarga yang selalu memberikan doa dan dukungan.
- 5 Sahabat dan rekan-rekan mahasiswa di Fakultas Ilmu Komputer, yang selalu memberi dukungan dan bantuan dalam berbagai hal.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna, Oleh karena itu, kritik maupun saran yang diharapkan untuk penyempurnaan skripsi ini di masa mendatang. Akhir kata penulis berharap ini dapat bermanfaat bagi pihak yang membacanya.

Jakarta, 31 Juli 2024



Mhartian Jordan Hutasoit

# ANALISIS SENTIMEN ULASAN *CASH ON DELIVERY* DI SHOPEE MENGGUNAKAN METODE *NAIVE BAYES*

Mhartian Jordan Hutasoit

## ABSTRAK

Penulis melakukan penelitian untuk menganalisis sentimen ulasan *Cash On Delivery*(COD) di Shopee menggunakan algoritma *Naive Bayes*. Metode ini digunakan karena kemampuannya dalam menangani masalah klasifikasi teks secara efektif. Dalam penelitian menggunakan data yang berasal dari *google play store* dengan teknik *scraping*. Ulasan-ulasan tersebut akan dimasukkan ke dalam dua kategori sentimen yang berbeda, yaitu sentimen positif dan negatif. Proses analisis dimulai pengumpulan data ulasan, diikuti oleh tahap *text-preprocessing* yang meliputi *case folding*, *tokenization*, *normalization*, *stopword removal*, dan *stemming*. Tahap selanjutnya membagi data menjadi data latih sebanyak 80% dan data uji sebanyak 20%. Data latih akan melewati tahapan metode TF-IDF(*Term Frequency-Invers Document Frequency*) untuk melakukan pembobotan pada data dan data uji akan dilakukan pengujian data latih menggunakan klasifikasi algoritma *Naive Bayes*.

Hasil penelitian menunjukkan bahwa model dengan memanfaatkan algoritma *Naive Bayes* mampu mendapatkan tingkat *accuracy* sebesar 66%, dengan *precision* tertinggi pada kategori sentimen positif sebesar 67%, *recall* sebesar 30% dan F-1 *score* 41% . Model ini juga menunjukkan performa yang baik dalam mengidentifikasi ulasan negatif dengan tingkat *precision* sebesar 66% dengan *recall* sebesar 90% dan F-1 *Score* sebesar 76%. Penelitian ini memberikan pemahaman mengenai persepsi pelanggan terhadap layanan COD mereka. Dengan memanfaatkan hasil analisis sentimen ini, Shopee dapat melakukan perbaikan dan penyesuaian layanan untuk meningkatkan kepuasan pelanggan. Peneliti ini juga menunjukkan bahwa algoritma *Naive Bayes* adalah alat yang efektif dan efisien untuk analisis sentimen dalam *E-commerce*

**Kata Kunci:** *Cash On Delivery*, Analisis Sentimen, *Naive Bayes*, *google play store*, Shopee, TF-IDF, *E-commerce*

# ***Sentiment Analysis of Cash on Delivery Reviews on Shopee Using the Naive Bayes Method***

*Mhartian Jordan Hutasoit*

## ***ABSTRACT***

*The author conducted research to analyze the sentiment of Cash On Delivery (COD) reviews on Shopee using the Naive Bayes algorithm. This method was chosen due to its effectiveness in handling text classification problems. The research utilized data from the Google Play Store collected through scraping techniques. These reviews were categorized into two sentiment categories: positive and negative. The analysis process began with data collection, followed by text preprocessing stages including case folding, tokenization, normalization, stopword removal, and stemming. The next stage involved splitting the data into 80% training data and 20% test data. The training data underwent TF-IDF (Term Frequency-Inverse Document Frequency) weighting, and the test data was evaluated using the Naive Bayes classification algorithm.*

*The research results showed that the model utilizing the Naive Bayes algorithm achieved an accuracy rate of 66%, with the highest precision in the positive sentiment category at 67%, a recall of 30%, and an F-1 score of 41%. The model also demonstrated good performance in identifying negative reviews with a precision rate of 66%, a recall of 90%, and an F-1 score of 76%. This research provides insights into customer perceptions of their COD services. By leveraging the results of this sentiment analysis, Shopee can make improvements and adjustments to enhance customer satisfaction. This study also demonstrates that the Naive Bayes algorithm is an effective and efficient tool for sentiment analysis in e-commerce.*

*Keywords: Cash On Delivery, Sentiment Analysis, Naive Bayes, Google Play Store, Shopee, TF-IDF, E-commerce*



## DAFTAR ISI

COVER.....	I
PERNYATAAN ORISINALITAS.....	II
PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS .....	I
LEMBAR PENGESAHAN .....	IV
KATA PENGANTAR .....	V
ABSTRAK.....	VI
<i>ABSTRACT</i> .....	VII
DAFTAR ISI.....	VIII
DAFTAR GAMBAR.....	XI
DAFTAR TABEL.....	XII
BAB I PENDAHULUAN .....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	2
1.3. Batasan Masalah.....	2
1.4. Tujuan Penelitian .....	2
1.5. Manfaat Penelitian .....	3
1.6. Luaran yang Diharapkan .....	3
1.7. Sistematika Penulisan.....	3
BAB II LANDASAN TEORI.....	5
2.1. Analisis Sentimen.....	5
2.2. <i>E-commerce</i> .....	5
2.3. Shopee .....	6
2.4. <i>Cash On Delivery</i> .....	6
2.5. <i>Data Mining</i> .....	6
2.6. <i>Text Mining</i> .....	7
2.6.1. Pengumpulan Data .....	7
2.6.2. <i>Text Preprocessing</i> .....	8
2.7. TF-IDF .....	9
2.8. Algoritma <i>Naive Bayes</i> .....	10
2.9. Evaluasi Model.....	11
2.10. Penelitian Relevan .....	13

BAB III METODOLOGI PENELITIAN .....	16
3.1. Tahapan Penelitian .....	16
3.1.1. Pengumpulan Data .....	16
3.1.2. Pelabelan Data .....	17
3.1.3. <i>Preprocessing</i> Data .....	18
3.1.4. Pembagian Data .....	22
3.1.5. TF-IDF .....	22
3.1.6. Klasifikasi Algoritma <i>Naive Bayes</i> .....	22
3.1.7. Evaluasi dan Hasil Model .....	22
3.1.8. Hasil Visualisasi .....	23
3.2. Alat Bantuan Penelitian.....	23
3.3. Jadwal Penelitian.....	23
BAB IV HASIL DAN PEMBAHASAN .....	25
4.1. Pengumpulan Data .....	25
4.2. Pelabelan Data.....	26
4.3. <i>Preprocessing</i> Data .....	27
4.3.1. <i>Case Folding</i> .....	27
4.3.2. <i>Cleaning</i> .....	28
4.3.3. <i>Tokenization</i> .....	29
4.3.4. <i>Normalization</i> .....	30
4.3.5. <i>Stopword Removal</i> .....	31
4.3.6. <i>Stemming</i> .....	32
4.4. Pembagian Data .....	33
4.5. Pembobotan TF-IDF .....	33
4.6. Proses Klasifikasi .....	37
4.6.1. Data Latih .....	38
4.4.2. Data Uji.....	40
4.5. Evaluasi Dan Hasil Analisis .....	43
4.6. Hasil Visualisasi <i>Word Cloud</i> .....	46
BAB V KESIMPULAN DAN SARAN .....	47
5.1. Kesimpulan .....	47
5.2. Saran.....	47
DAFTAR PUSTAKA .....	49
RIWAYAT HIDUP .....	52

LAMPIRAN.....	53
HASIL PLAGIARISME.....	59

## DAFTAR GAMBAR

Gambar 3. 1. Alur Penelitian .....	16
Gambar 4. 1. Hasil Scraping data ulasan .....	25
Gambar 4. 2. WordCloud Data Ulasan .....	46

## DAFTAR TABEL

Tabel 2. 1. Confusion Matrix untuk dua kelas .....	12
Tabel 3. 1. Contoh hasil pelabelan data ulasan .....	17
Tabel 3. 2. Contoh Case Folding data ulasan.....	18
Tabel 3. 3. Contoh Cleaning data ulasan.....	19
Tabel 3. 4. Contoh Tokenization data ulasan .....	19
Tabel 3. 5. Contoh Normalization data ulasan .....	20
Tabel 3. 6. Contoh Stopword Removal data ulasan. ....	21
Tabel 3. 7. Contoh Stemming data ulasan .....	21
Tabel 3. 8. Jadwal Penelitian.....	24
Tabel 4. 1. Hasil pelabelan data ulasan .....	26
Tabel 4. 2. Hasil Case Folding data ulasan .....	27
Tabel 4. 3. Hasil Cleaning data ulasan.....	28
Tabel 4. 4. Hasil Tokenization data ulasan.....	29
Tabel 4. 5. Hasil Normalization data ulasan .....	30
Tabel 4. 6. Hasil Stopword Removal data ulasan .....	31
Tabel 4. 7. Hasil Stemming data ulasan .....	32
Tabel 4. 8. Hasil pembagian data ulasan.....	33
Tabel 4. 9. Hasil Preprocessing sampel data ulasan.....	34
Tabel 4. 10. Hasil perhitungan TF-IDF pada sampel data .....	35
Tabel 4. 11. Hasil probabilitas data latih kelas positif dan negatif .....	39
Tabel 4. 12. Sampel data uji.....	40
Tabel 4. 13. Hasil probabilitas data uji kelas positif dan negatif .....	41
Tabel 4. 14. Confusion Matrix evaluasi hasil data uji.....	43
Tabel 4. 15. Hasil accuracy, recall, precision dan F-1 score .....	45