



**ANALISIS SENTIMEN TERHADAP *ULASAN* KEPUASAN PELANGGAN  
PADA MARKETPLACE TOKOPEDIA DI JEJARING SOSIAL *TWITTER*  
MENGUNAKAN ALGORITMA *NAIVE BAYES***

**TUGAS AKHIR**

**Muhammad Abi Nurhakim**

**1710511083**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA**

**FAKULTAS ILMU KOMPUTER**

**PROGRAM STUDI INFORMATIKA**

**2022**



**ANALISIS SENTIMEN TERHADAP *ULASAN* KEPUASAN PELANGGAN  
PADAMARKETPLACE TOKOPEDIA DI JEJARING SOSIAL *TWITTER*  
MENGUNAKAN ALGORITMA *NAIVE BAYES***

**TUGAS AKHIR**

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

**Muhammad Abi Nurhakim  
1710511083**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA  
FAKULTAS ILMU KOMPUTER  
PROGRAM STUDI INFORMATIKA  
2022**

## PERNYATAAN ORISINALITAS

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

Nama : Muhammad Abi Nurhakim

NIM : 1710511083

Tanggal : 15 Desember 2021

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

Jakarta, 15 Desember 2021  
Yang Menyatakan,



(Muhammad Abi Nurhakim)

## **PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS**

---

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

Nama : Muhammad Abi Nurhakim  
NIM : 1710511083  
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 Royalty Free Right*) atas karya ilmiah saya yang berjudul: **ANALISIS SENTIMEN TERHADAP ULASAN KEPUASAN PELANGGAN PADA MARKETPLACE TOKOPEDIA DI JEJARING SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAIVE BAYES**

Beserta perangkat yang ada (jika digunakan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Tugas Akhir 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 : 15 Desember 2021

Yang Menyatakan,



(Muhammad Abi Nurhakim)

## LEMBAR PENGESAHAN

Dengan ini dinyatakan bahwa Tugas Akhir berikut:

Nama : Muhammad Abi Nurhakim  
NIM : 1710511083  
Program Studi : Informatika  
Judul : Analisis Sentimen Terhadap Ulasan Kepuasan Pelanggan Pada Marketplace Tokopedia Di Jejaring Sosial Twitter Menggunakan Algoritma Naïve Bayes

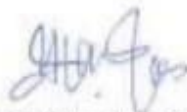
Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



DR. Didit Widiyanto, S.Kom., M.Si.  
Penguji I



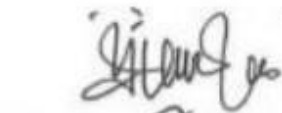
Iin Ernawati, S.Kom., M.Si.  
Penguji II



Yuni Widiastiwi, S.Kom., M.Si.  
Pembimbing I



Nurul Chamidah, S.Kom., M.Kom.  
Pembimbing II



Yuni Widiastiwi, S.Kom., M.Si.  
Ketua Program Studi

Ditetapkan di : Jakarta  
Tanggal Ujian : 15 Desember 2021



# ANALISIS SENTIMEN TERHADAP ULASAN KEPUASAN PELANGGAN PADA MARKETPLACE TOKOPEDIA DI JEJARING SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES

Muhammad Abi Nurhakim

## Abstrak

Ulasan konsumen tentang kepuasan berbelanja di marketplace merupakan informasi yang bernilai dan dapat diolah dengan baik. Informasi ini dapat dimanfaatkan untuk mendapatkan evaluasi produk dan layanan baik oleh konsumen maupun pihak penjual atau marketplace dengan melakukan analisis ulasan pengalaman berbelanja tersebut melalui jejaring sosial *twitter* guna mendapatkan informasi yang diperlukan. Kegiatan analisis ulasan tidak dapat cukup dengan melihat jumlah *rating* saja, namun diperlukan melihat seluruh isi ulasan untuk dapat mengetahui arti yang lebih luas dan umum dari ulasan tersebut. Apabila dalam jumlah sedikit dapat dilakukan secara manual, namun untuk melihat ulasan dalam jumlah banyak, diperlukan sistem agar dapat menganalisis secara lebih efektif dan memudahkan dalam memahami maksud ulasan. Pada penelitian ini yang menggunakan metode klasifikasi Naïve Bayes yang dibagi ke dalam 2 kelas yaitu positif dan negatif. Hasil klasifikasi menggunakan metode algoritma *Naïve Bayes* didapatkan hasil akurasi sebesar 79,02%, *precision* sebesar 80,30%, *recall* sebesar 77,94% dan *specificity* sebesar 80,15%.

**Kata Kunci:** *Twitter*, Ulasan, Klasifikasi, *Naïve Bayes*

# **SENTIMENT ANALYSIS OF CUSTOMER SATISFACTION REVIEWS ON THE TOKOPEDIA MARKETPLACE ON TWITTER SOCIAL NETWORK USING NAVE BAYES ALGORITHM**

**Muhammad Abi Nurhakim**

## **Abstract**

Consumer reviews about shopping satisfaction in the marketplace are valuable information and can be processed properly. This information can be used to obtain product and service evaluations by both consumers and sellers or marketplaces by analyzing the shopping experience reviews through the Twitter social network in order to obtain the necessary information. Review analysis activities cannot be enough just to look at the number of ratings, but it is necessary to look at the entire content of the review to be able to find out the broader and general meaning of the review. If a small amount can be done manually, but to see a large number of reviews, a system is needed to be able to analyze more effectively and make it easier to understand the purpose of the review. In this study using the Naïve Bayes classification method which is divided into 2 classes, namely positive and negative. The results of the classification using the Naïve Bayes algorithm obtained an accuracy of 79.02%, precision of 80.30%, recall of 77.94% and specificity of 80.15%.

**Keywords:** *Twitter, Reviews, Classification, Nave Bayes*

## KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala karunia-Nya, sehingga penulis mampu menyelesaikan tugas akhir berjudul “**Analisis Sentimen Terhadap Ulasan Pelanggan Pada Marketplace Tokopedia Di Jejaring Sosial Twitter Menggunakan Naïve Bayes**”. Penulis berterima kasih kepada:

1. Ibu Yuni Widiastiwi, S.Kom, M.Si. selaku dosen pembimbing 1 yang telah memberikan saran, bimbingan dan nasihatnya selama penelitian berlangsung dan penulisan tugas akhir ini.
2. Ibu Nurul Chamidah, S.Kom, M.Kom. selaku dosen pembimbing 2 yang telah memberikan saran, bimbingan dan nasihatnya selama penelitian berlangsung dan penulisan tugas akhir ini.
3. Ibu Dr. Ermatita, M.Kom. selaku Dekan Fakultas Ilmu Komputer.
4. Ibu Yuni Widiastiwi, S.Kom, M.Si. selaku Ketua Program Studi Informatika dan Dosen Pembimbing Akademik.
5. Bapak/Ibu dosen Informatika Universitas Pembangunan Nasional Veteran Jakarta terima kasih atas ilmu-ilmu yang telah diberikan dan bermanfaat.
6. Kedua orang tua tercinta, Mohammad Rizal dan Susi R Puspitadewi yang senantiasa selalu memberikan semangat dan kasih sayang serta rela memberikan waktunya untuk mendengar keluh kesah penulis.
7. Teman-teman, saudara, dan pihak-pihak lain yang rasanya tak mungkin saya sebutkan satu per satu, terima kasih atas saran dan dukungan yang telah diberikan.

Akhir kata, semoga tugas akhir ini dapat bermanfaat bagi para pembacanya.

Jakarta, 15 Desember 2021

Penulis



## DAFTAR ISI

|  |           |
|--|-----------|
| HALAMAN JUDUL .....  | i         |
| PERNYATAAN ORISINALITAS .....  | ii        |
| PERNYATAAN PERSETUJUN PUBLIKASI TUGAS AKHIR .....                              | iii       |
| LEMBAR PENGESAHAN .....  | iv        |
| ABSTRAK .....  | v         |
| ABSTRACT .....   | vi        |
| KATA PENGANTAR .....   | vii       |
| DAFTAR ISI .....   | viii      |
| DAFTAR GAMBAR .....  | xi        |
| DAFTAR TABEL .....   | xii       |
| DAFTAR SIMBOL .....  | xiii      |
| <b>BAB I PENDAHULUAN</b> .....   | <b>1</b>  |
| <b>1.1 Latar Belakang</b> .....  | <b>1</b>  |
| <b>1.2 Rumusan Masalah</b> .....   | <b>3</b>  |
| <b>1.3 Tujuan Penelitian</b> .....   | <b>3</b>  |
| <b>1.4 Manfaat Penelitian</b> .....  | <b>4</b>  |
| <b>1.5 Ruang Lingkup</b> .....   | <b>4</b>  |
| <b>1.6 Luaran Yang Diharapkan</b> .....  | <b>4</b>  |
| <b>1.7 Sistematika Penulisan</b> .....   | <b>4</b>  |
| <b>BAB II TINJAUAN PUSTAKA</b> .....   | <b>6</b>  |
| <b>2.1 Twitter</b> .....   | <b>6</b>  |
| <b>2.2 Ulasan</b> .....  | <b>6</b>  |
| <b>2.3 Tokopedia</b> .....   | <b>7</b>  |
| <b>2.4 Analisis Sentimen</b> .....   | <b>7</b>  |
| <b>2.5 Twitter Crawling</b> .....  | <b>7</b>  |
| <b>2.6 Text Analysis</b> .....   | <b>8</b>  |
| <b>2.6.1 Text Mining</b> .....   | <b>8</b>  |
| <b>2.7 Pre-processing Data</b> .....   | <b>9</b>  |
| <b>2.8 Pembobotan TF-IDF (Term Frequency-Inverse Document Frequency)</b> ..... | <b>11</b> |
| <b>2.9 Metode Algoritma Naive Bayes</b> .....                                  | <b>12</b> |

|  |                                      |           |
|--|--------------------------------------|-----------|
| 2.10                                       | Evaluasi .....                       | 14        |
| 2.11                                       | Python .....                         | 16        |
| 2.12                                       | Review Riset yang Relevan .....      | 16        |
| <b>BAB III METODOLOGI PENELITIAN .....</b> |                                      | <b>17</b> |
| 3.1  | Identifikasi Masalah .....           | 18        |
| 3.2  | Studi Literatur .....                | 18        |
| 3.3  | Pengumpulan Data .....               | 18        |
| 3.3.1                                      | Pelabelan Data .....                 | 18        |
| 3.4  | <i>Pre-processing Data</i> .....     | 19        |
| 3.4.1                                      | Pembersihan Data .....               | 20        |
| 3.4.2                                      | <i>Case Folding</i> .....            | 20        |
| 3.4.3                                      | <i>Stopword Removal</i> .....        | 20        |
| 3.4.4                                      | <i>Stemming</i> .....                | 20        |
| 3.4.5                                      | <i>Tokenisasi</i> .....              | 20        |
| 3.5  | Pembobotan TF-IDF .....              | 21        |
| 3.6  | Klasifikasi <i>Naïve Bayes</i> ..... | 21        |
| 3.7  | Analisa pengujian dan Evaluasi ..... | 21        |
| 3.8  | Visualisasi .....                    | 22        |
| 3.9  | Alat Yang Digunakan .....            | 22        |
| 3.10                                       | Jadwal Penelitian .....              | 23        |
| <b>BAB IV HASIL DAN PEMBAHASAN .....</b>   |                                      | <b>24</b> |
| 4.1  | Data .....                           | 24        |
| 4.1.1                                      | Pelabelan Data .....                 | 25        |
| 4.2  | <i>Pre-processing Data</i> .....     | 28        |
| 4.2.1                                      | Pembersihan Data .....               | 28        |
| 4.2.2                                      | <i>Case Folding</i> .....            | 29        |
| 4.2.3                                      | <i>Stopword Removal</i> .....        | 30        |
| 4.2.4                                      | <i>Stemming</i> .....                | 30        |
| 4.2.5                                      | <i>Tokenisasi</i> .....              | 31        |

|                                       |  |    |
|---------------------------------------|--|----|
| 4.3                                   | Pembobotan TF-IDF .....                          | 31 |
| 4.4                                   | Klasifikasi Naïve Bayes .....                    | 34 |
| 4.4.1                                 | Tahap Data Training .....                        | 35 |
| 4.4.2                                 | Tahap Data Testing .....                         | 41 |
| 4.5                                   | Evaluasi .....                                   | 46 |
| 4.6                                   | Visualisasi .....                                | 47 |
| 4.6.1                                 | Wordcloud Sentimen Positif .....                 | 47 |
| 4.6.2                                 | Frekuensi Kemunculan Kata Sentimen Positif ..... | 48 |
| 4.6.3                                 | Wordcloud Sentimen Negatif .....                 | 49 |
| 4.6.4                                 | Frekuensi Kemunculan Kata Sentimen Negatif ..... | 50 |
| <b>BAB V PENUTUP</b> .....            |  | 52 |
| 5.1                                   | Kesimpulan .....                                 | 52 |
| 5.2                                   | Saran .....                                      | 53 |
| <b>DAFTAR PUSTAKA</b> .....           |  | 54 |
| <b>RIWAYAT HIDUP</b> .....            |  |    |
| <b>LAMPIRAN</b> .....                 |  |    |
| Lampiran 1 .....                      |  |    |
| <i>Similarity Index</i> Skripsi ..... |  |    |
| Lampiran 1 (Lanjutan) .....           |  |    |
| Lampiran 1 (Lanjutan) .....           |  |    |
| Lampiran 1 (Lanjutan) .....           |  |    |
| Lampiran 1 (Lanjutan) .....           |  |    |
| Lampiran 1 (Lanjutan) .....           |  |    |
| Lampiran 1 (Lanjutan) .....           |  |    |
| Lampiran 1 (Lanjutan) .....           |  |    |
| Lampiran 2 .....                      |  |    |
| <i>Source Code</i> .....              |  |    |
| Lampiran 2 (lanjutan) .....           |  |    |




## DAFTAR GAMBAR

|  |           |
|--|-----------|
| <b>Gambar 1. Tahapan pre-processing data .....</b>                             | <b>10</b> |
| <b>Gambar 2. Metodologi Penelitian.....</b>                                    | <b>17</b> |
| <b>Gambar 3. Hasil tampilan dari crawling data tweet berbentuk format xlsx</b> | <b>25</b> |
| <b>Gambar 4. Visualisasi Hasil Pelabelan Data Tweet.....</b>                   | <b>28</b> |
| <b>Gambar 5. Wordcloud Sentimen Positif.....</b>                               | <b>48</b> |
| <b>Gambar 6. Hasil Frekuensi Kemunculan Kata Sentiment Positif.....</b>        | <b>49</b> |
| <b>Gambar 7. Wordcloud Sentimen Negatif.....</b>                               | <b>50</b> |
| <b>Gambar 8. Hasil Frekuensi Kemunculan Kata Sentiment Negatif.....</b>        | <b>51</b> |

## DAFTAR TABEL

|   |    |
|---|----|
| Tabel 2.1. Confusion Matrix .....                     | 15 |
| Tabel 3.1. Interpretasi Koefisien Cohen's Kappa ..... | 19 |
| Tabel 3.2. Jadwal Penelitian.....                     | 23 |
| Tabel 4.1. Contoh Pelabelan Data tweet .....          | 25 |
| Tabel 4.2. Hasil Pelabelan Data Tweet.....            | 26 |
| Tabel 4.3. Tabel Pembersihan Data.....                | 29 |
| Tabel 4.4. Tabel Case Folding.....                    | 29 |
| Tabel 4.5. Tabel Stopword Removal .....               | 30 |
| Tabel 4.6. Tabel Stemming.....                        | 31 |
| Tabel 4.7. Tabel Tokenisasi.....                      | 31 |
| Tabel 4.8. Sampel Dokumen (Tweet).....                | 32 |
| Tabel 4.9. Perhitungan Pembobotan (TF-IDF).....       | 33 |
| Tabel 4.10. Hasil Pembagian Data.....                 | 35 |
| Tabel 4.11. Sampel Data Training.....                 | 35 |
| Tabel 4.12. Sampel Data Testing .....                 | 41 |
| Tabel 4.13. Hasil Pre-processing Data Testing.....    | 41 |
| Tabel 4.14. Hasil Pembobotan Data Testing.....        | 42 |
| Tabel 4.15. Hasil Confusion Matrix.....               | 46 |

## DAFTAR SIMBOL

| Simbol Flowchart |  |            |  |
|------------------|--|------------|--|
| No               | Simbol   | Nama       | Keterangan   |
| 1                |   | Terminator | Simbol ini digambarkan sebagai lingkaran yang dimana kegiatan dimulai atau berakhir.   |
| 2                |   | Flow       | Simbol ini digambarkan arus panah yang menunjukkan suatu hubungan yang menyambungkan simbol ke simbol lainnya ke dalam proses kegiatan tersebut. |
| 3                |  | Process    | Simbol ini digambarkan kotak persegi panjang yang menunjukkan suatu proses yang dijalankan.  |