

**ANALISIS SENTIMEN PADA ULASAN APLIKASI  
TOKOPEDIA DI *GOOGLE PLAY STORE* DENGAN  
ALGORITMA *SUPPORT VECTOR MACHINE*  
DAN *NAÏVE BAYES***

**SKRIPSI**

**Oleh:**

**Ahmad Syahrul Firmansyah**

**201910225391**



**PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS BHAYANGKARA JAKARTA RAYA  
2023**

## LEMBAR PERSETUJUAN PEMBIMBING

Judul Skripsi : Analisis Sentimen Pada Ulasan Aplikasi  
Tokopedia Di Google Play Store Dengan  
Algoritma *Support Vector Machine* Dan *Naïve Bayes*

Nama Mahasiswa : Ahmad Syahrul Firmansyah

Nomor Pokok Mahasiswa : 201910225391

Program Studi / Fakultas : Informatika / Ilmu Komputer

Tanggal Lulus Ujian Skripsi : 17-07-2023



Pembimbing I

Herlawati, S.Si., M.M., M.Kom.

NIDN: 0311097302

Pembimbing II

Rafika Sari, S.Si., M.Si.

NIDN: 0329098902

## LEMBAR PENGESAHAN

Judul Tugas akhir : Analisis Sentimen Pada Aplikasi  
Tokopedia Di Ulasan Google Play store  
Dengan Algoritma *Support Vector  
Machine* dan *Naive Bayes*.

Nama Mahasiswa : Ahmad Syahrul Firmansyah

Nomor Pokok Mahasiswa : 201910225391

Program Studi/Fakultas : Informatika / Ilmu Komputer

Tanggal Lulus Ujian Tugas akhir : 17/07/2023

Bekasi, 01/08/2023

MENGESAHKAN,

Ketua Tim Penguji : Wowon Priatna, ST, M.Ti

NIDN : 0429118007

Penguji I : Sri Rejeki, S.Kom, M.M.

NIDN : 0320116602

Penguji II : Herlawati, S.Si., M.M., M.Kom.

NIDN : 0311097302

MENGETAHUI,

Ketua Prodi  
Informatika

Dekan  
Fakultas Ilmu Komputer

  
Ahmad Fathurrozi, S.E., M.M.S.I  
NIP. 2012486

  
Dr. Dra. Tyastuti Sri Lestari, M.M  
NIP. 1408206



UNIVERSITAS BHAYANGKARA JAKARTA RAYA  
FAKULTAS ILMU KOMPUTER  
PROGRAM STUDI INFORMATIKA

## LEMBAR PERNYATAAN BUKAN PLAGIASI

Yang bertanda tangan dibawah ini :

Nama : Ahmad Syahrul Firmansyah  
NPM : 201910225391  
Program Studi : Informatika  
Fakultas : Ilmu Komputer  
Judul Tugas Akhir : Analisis Sentimen Pada Ulasan Aplikasi Tokopedia Di  
Google Play Store Dengan Algoritma *Support Vector  
Machine* Dan *Naïve Bayes*

Dengan ini menyatakan bahwa hasil penulisan skripsi yang telah saya buat ini merupakan **hasil karya saya sendiri dan benar keasliannya**. Apabila dikemudian hari penulisan skripsi ini merupakan plagiat atau penjiplakan terhadap karya orang lain, maka saya bersedia mempertanggungjawabkan sekaligus bersedia menerima sanksi berdasarkan tata tertib di Universitas Bhayangkara Jakarta Raya.

Demikian pernyataan ini saya buat dalam keadaan sadar dan tidak dipaksakan dari pihak manapun.

Bekasi, 01 Juli 2023

Penulis



Ahmad Syahrul Firmansyah  
NPM. 201910225391

## ABSTRAK

**Ahmad Syahrul Firmansyah. 201910225391.** Analisis Sentimen Pada Ulasan Aplikasi Tokopedia Di Google Play Store Dengan Algoritma Support Vector Machine Dan Naïve Bayes.

Perkembangan teknologi yang pesat membawa pengaruh yang besar bagi manusia. Salah satu pengaruh dari teknologi tersebut merupakan jual beli secara *online* melalui media internet. Pesatnya perkembangan jual beli *online* memunculkan banyaknya marketplace di Indonesia salah satunya, yaitu Tokopedia. Tokopedia telah diunduh sebanyak 100 juta lebih dan memiliki 6 juta ulasan pada *Google Play Store*. Pada penelitian ini akan menguji hasil penerapan *Support Vector Machine* dan *Naïve Bayes* dalam menganalisis data sentiment ulasan pengguna Tokopedia pada *Google Play Store*. Data yang digunakan adalah data ulasan pengguna Tokopedia dalam *Google Play Store* sebanyak 20.000 data berbahasa Indonesia. Metode *Support Vector Machine* dan *Naïve Bayes* digunakan untuk mengklasifikasikan sentimen positif dan negatif dari ulasan pengguna. Hasil dari penerapan metode *Support Vector Machine* pada klasifikasi ulasan pengguna Tokopedia dengan data uji 90:10 menghasilkan tingkat *accuracy* 89%, *Recall* 90%, *Precision* 90%, *F-Measure* 90%, sedangkan penerapan metode *Naïve Bayes* pada klasifikasi ulasan pengguna Tokopedia dengan data uji 90:10 menghasilkan tingkat *accuracy* 70%, *Recall* 66%, *Precision* 93%, *F-Measure* 77%. Performa dari algoritma *Support Vector Machine* menghasilkan tingkat *accuracy* lebih tinggi dari algoritma *Naïve Bayes*.

**Kata Kunci :** Analisis Sentimen, *Support Vector Machine*, *Naïve Bayes*, Ulasan, Tokopedia.

## ***ABSTRACT***

**Ahmad Syahrul Firmansyah. 201910225391. Sentiment Analysis Of Tokopedia App Reviews On Google Play Store With Support Vector Machine And Naïve Bayes Algorithms**

*The rapid development of technology has a great influence on humans. One of the influences of this technology is buying and selling online through internet media. The rapid development of buying and selling online has led to many marketplaces in Indonesia, one of which is Tokopedia. Tokopedia has been downloaded more than 100 million and has 6 million reviews on the Google Play Store. This research will examine the results of implementing the Support Vector Machine in analyzing sentiment data from Tokopedia user reviews on the Google Play Store. The data used is Tokopedia user review data in the Google Play Store of 20,000 data. The data used are reviews in Indonesian. The Support Vector Machine and Naïve Bayes methods are used to classify positive and negative sentiments from user reviews. The ability of the Support Vector Machine to determine the best hyperplane makes this algorithm have a high level of generality while the advantage of Naïve Bayes is that it only requires a small amount of training data to determine the estimated parameters needed in the classification process. The results of applying the Support Vector Machine method to the classification of Tokopedia user reviews with 90:10 test data produce an accuracy rate of 89%, Recall 90%, Precision 90%, F-Measure 90%, while the application of the Naïve Bayes method to the classification of Tokopedia user reviews with data the 90:10 test produced an accuracy rate of 70%, Recall 66%, Precision 93%, F-Measure 77%. The performance of the Support Vector Machine algorithm produces a higher level of accuracy than the Naïve Bayes algorithm.*

**Keywords:** *Sentiment Analysis, Support Vector Machine, Naïve Bayes, Review, Tokopedia*

## LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIK

Sebagai sivitas akademik Universitas Bhayangkara Jakarta Raya, saya yang bertanda tangan di bawah ini :

Nama : Ahmad Syahrul Firmansyah  
NPM : 201910225391  
Program Studi : Informatika  
Fakultas : Ilmu Komputer  
Jenis Karya : Skripsi

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

Analisis Sentimen Pada Ulasan Aplikasi Tokopedia Di *Google Play Store* Dengan Algoritma *Support Vector Machine* Dan *Naïve Bayes*. beserta perangkat yang ada (bila diperlukan). Dengan hak bebas royalti non-eksklusif ini, Universitas Bhayangkara Jakarta Raya berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya dan mempublikasikannya di Internet atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik hak cipta.

Segala bentuk tuntutan hukum yang timbul atas pelanggaran hak cipta dalam karya ilmiah ini menjadi tanggung jawab saya pribadi

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Bekasi  
Pada tanggal : 17 Juli 2023  
Yang Menyatakan



Ahmad Syahrul Firmansyah  
NPM. 201910225391

## KATA PENGANTAR

Puji syukur kami panjatkan kehadiran Allah SWT yang telah memberikan rahmat serta karunia-Nya kepada kami sehingga ini kami menyampaikan terimakasih kepada semua pihak terkait yang telah membantu terselesaikannya penelitian ini.

1. Bapak Irjen Pol. (Purn) Prof Dr. Drs. Bambang Karsono, SH., M.M. Selaku Rektor Universitas Bhayangkara Jakarta Raya.
2. Ibu Dr. Dra Tyastuti Sri Lestari, M.M. Selaku Dekan Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya.
3. Bapak Ahmad Fathurrozi, S.E., M.M.S.I. Selaku Ketua Program Studi Informatika Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya.
4. Ibu Herlawati, S.Si., M.M., M.Kom. Selaku Dosen Pembimbing I Skripsi yang memberikan masukan serta saran dan motivasi dalam mengerjakan tugas akhir ini.
5. Ibu Rafika Sari, S.Si., M.Si. Selaku Dosen Pembimbing II Skripsi yang memberikan masukan serta saran dan motivasi dalam mengerjakan tugas akhir ini.
6. Kedua orang tua yang selalu mendoakan dan teman-teman serta sahabat yang selalu mendukung.

Saya menyadari bahwa tugas akhir ini masih jauh dari kesempurnaan. Oleh karena itu, segala saran dan kritik yang membangun dari semua pihak diharapkan demi perbaikan pada laporan selanjutnya. Semoga penelitian ini bermanfaat khususnya bagi penyusun dan bagi pembaca lain pada umumnya.

Bekasi, 19 Maret 2023



Ahmad Syahrul Firmansyah

## DAFTAR ISI

|                                                          |           |
|----------------------------------------------------------|-----------|
| LEMBAR PERSETUJUAN PEMBIMBING .....                      | ii        |
| LEMBAR PENGESAHAN .....                                  | iii       |
| LEMBAR PERNYATAAN BUKAN PLAGIASI .....                   | iv        |
| ABSTRAK .....                                            | v         |
| ABSTRACT .....                                           | vi        |
| LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI .....            | vii       |
| KATA PENGANTAR.....                                      | viii      |
| DAFTAR ISI.....                                          | ix        |
| DAFTAR TABEL .....                                       | xii       |
| DAFTAR GAMBAR.....                                       | xiii      |
| DAFTAR LAMPIRAN .....                                    | xiv       |
| <b>BAB I    PENDAHULUAN.....</b>                         | <b>1</b>  |
| 1.1 Latar Belakang.....                                  | 1         |
| 1.2 Identifikasi Masalah .....                           | 6         |
| 1.3 Rumusan Masalah .....                                | 6         |
| 1.4 Tujuan Penelitian.....                               | 7         |
| 1.5 Manfaat Penelitian.....                              | 7         |
| 1.6 Batasan Masalah.....                                 | 7         |
| 1.7 Sistematika Penulisan .....                          | 8         |
| <b>BAB II   TINJAUAN PUSTAKA.....</b>                    | <b>10</b> |
| 2.1 Kajian Literatur.....                                | 10        |
| 2.2 Text Mining .....                                    | 14        |
| 2.2.1 <i>Knowledge Discovery in Database (KDD)</i> ..... | 14        |
| 2.2.2 <i>Machine Learning</i> .....                      | 21        |

|                |                                             |           |
|----------------|---------------------------------------------|-----------|
| 2.2.3          | <i>Data Mining</i> .....                    | 21        |
| 2.3            | Support Vector Machine (SVM) .....          | 22        |
| 2.4            | Naïve Bayes .....                           | 23        |
| 2.5            | Analisis Sentimen .....                     | 25        |
| 2.5.1          | Tokopedia .....                             | 26        |
| 2.5.2          | <i>Google Play Store</i> .....              | 27        |
| 2.6            | Bahasa Pemrograman .....                    | 28        |
| <b>BAB III</b> | <b>METODOLOGI PENELITIAN</b> .....          | <b>29</b> |
| 3.1            | Objek Penelitian .....                      | 29        |
| 3.2            | Kerangka Penelitian.....                    | 30        |
| 3.3            | Metode Pengumpulan Data .....               | 31        |
| 3.4            | Metode Analisis Data .....                  | 32        |
| 3.5            | Analisis Kebutuhan Sistem.....              | 33        |
| 3.6            | Data Penelitian.....                        | 34        |
| <b>BAB IV</b>  | <b>HASIL DAN PEMBAHASAN</b> .....           | <b>35</b> |
| 4.1            | Data Selection.....                         | 35        |
| 4.2            | Pelabelan Data .....                        | 37        |
| 4.3            | Preprocessing .....                         | 38        |
| 4.3.1          | <i>Cleansing</i> .....                      | 38        |
| 4.3.2          | <i>Case Folding</i> .....                   | 39        |
| 4.3.3          | <i>Tokenazing</i> .....                     | 39        |
| 4.3.4          | <i>Normalization</i> .....                  | 41        |
| 4.3.5          | <i>Stopword Removal</i> .....               | 42        |
| 4.3.6          | <i>Stemming</i> .....                       | 43        |
| 4.4            | Transformation .....                        | 44        |
| 4.4.1          | <i>Transformation Secara Otomatis</i> ..... | 44        |
| 4.4.2          | <i>Transformation Secara Manual</i> .....   | 45        |

|                                                                 |           |
|-----------------------------------------------------------------|-----------|
| 4.5 Text Mining .....                                           | 48        |
| 4.5.1 Hitungan SVM Dan <i>Naïve Bayes</i> Secara Otomatis ..... | 49        |
| 4.5.2 Hitungan SVM Dan <i>Naïve Bayes</i> Secara Manual .....   | 49        |
| 4.6 Evaluation.....                                             | 59        |
| <b>BAB V PENUTUP.....</b>                                       | <b>66</b> |
| 5.1 Kesimpulan.....                                             | 66        |
| 5.2 Saran .....                                                 | 66        |
| <b>DAFTAR PUSTAKA.....</b>                                      | <b>68</b> |
| <b>LAMPIRAN.....</b>                                            | <b>75</b> |



## DAFTAR TABEL

|                                                                                                |    |
|------------------------------------------------------------------------------------------------|----|
| Tabel 2. 1 Penelitian Terkait .....                                                            | 11 |
| Tabel 2.2 <i>Confusion Matrix</i> .....                                                        | 19 |
| Tabel 4.1 Tahapan metode <i>Scraping</i> .....                                                 | 35 |
| Tabel 4.2 Tahapan Penghapusan Kolom Yang Di Anggap Tidak Relevan...                            | 37 |
| Tabel 4.3 Tahapan Pelabelan Data Ulasan .....                                                  | 37 |
| Tabel 4.4 Penerapan <i>Cleansing</i> Pada Ulasan Pengguna Tokopedia .....                      | 39 |
| Tabel 4.5 Penerapan <i>Case Folding</i> .....                                                  | 39 |
| Tabel 4.6 Penerapan <i>Tokenazing</i> .....                                                    | 40 |
| Tabel 4.7 Penerapan <i>Normalization</i> .....                                                 | 41 |
| Tabel 4.8 Penerapan <i>Stopword Removal</i> .....                                              | 42 |
| Tabel 4.9 Penerapan <i>Stemming</i> .....                                                      | 43 |
| Tabel 4.10 Tahapan <i>Transformation</i> .....                                                 | 44 |
| Tabel 4. 11 Ulasan Yang Digunakan Untuk Perhitungan TF-IDF.....                                | 45 |
| Tabel 4. 12 Ulasan Yang Digunakan Setelah <i>Preprocessing</i> .....                           | 45 |
| Tabel 4. 13 Hasil Normalisasi Kata .....                                                       | 46 |
| Tabel 4. 14 Hasil <i>Document Frequency</i> (DF) .....                                         | 46 |
| Tabel 4. 15 Hasil <i>Inverse Document Frequency</i> (IDF) .....                                | 47 |
| Tabel 4. 16 Hasil Dari TF-IDF .....                                                            | 48 |
| Tabel 4.17 Tahapan Text Mining Menggunakan Support Vector Machine .                            | 49 |
| Tabel 4.18 Tahapan <i>Text Mining</i> Menggunakan <i>Naïve Bayes</i> .....                     | 49 |
| Tabel 4. 19 Hasil TF-IDF 3 Ulasan.....                                                         | 50 |
| Tabel 4. 20 Data <i>Traning</i> Sebelum <i>Preprocessing</i> .....                             | 52 |
| Tabel 4. 21 Data <i>Traning</i> Sesudah <i>Preprocessing</i> .....                             | 52 |
| Tabel 4. 22 Hasil Probabilitas Ulasan 1 .....                                                  | 56 |
| Tabel 4. 23 Hasil Probabilitas Ulasan 2 .....                                                  | 56 |
| Tabel 4. 24 Data <i>Testing</i> Sebelum <i>Preprocessing</i> .....                             | 56 |
| Tabel 4. 25 Data <i>Testing</i> Sesudah <i>Preprocessing</i> .....                             | 57 |
| Tabel 4.26 Hasil Probabilitas Data <i>Testing</i> .....                                        | 58 |
| Tabel 4.27 Hasil Perbandingan model <i>Support Vector Machine</i> dan <i>Naïve Bayes</i> ..... | 63 |

## DAFTAR GAMBAR

|                                                                                        |    |
|----------------------------------------------------------------------------------------|----|
| Gambar 1.1 Ulasan Pengguna Aplikasi Tokopedia .....                                    | 3  |
| Gambar 1.2 Hasil Jawaban Kuesioner .....                                               | 5  |
| Gambar 2.1 Proses <i>Knowledge Discovery in Database</i> .....                         | 15 |
| Gambar 2.2 <i>Support Vector Machine</i> .....                                         | 23 |
| Gambar 3.1 Lokasi Kuesioner Penelitian.....                                            | 29 |
| Gambar 3. 2 Kerangka Penelitian .....                                                  | 31 |
| Gambar 3. 3 Data Hasil Scraping Ulasan Aplikasi Tokopedia.....                         | 34 |
| Gambar 4.1 Hasil Metode <i>Scraping</i> .....                                          | 36 |
| Gambar 4.2 Hasil Penghapusan Kolom Yang Di Anggap Tidak Relevan .....                  | 37 |
| Gambar 4.3 Hasil Pelabelan Data Ulasan .....                                           | 38 |
| Gambar 4.4 <i>Confusion Matrix</i> Menggunakan SVM Dengan Skenario 90:10 .....         | 59 |
| Gambar 4.5 <i>Confusion Matrix</i> Menggunakan <i>Naïve Bayes</i> Skenario 90:10 ..... | 59 |
| Gambar 4.6 Ulasan Positif Yang Di Visualisasikan Dengan <i>Wordcloud</i> .....         | 64 |
| Gambar 4.7 Ulasan Negatif Yang Di Visualisasikan Dengan <i>Wordcloud</i> .....         | 65 |

## DAFTAR LAMPIRAN

1. Hasil Plagiarisme Checker Originality
2. Biodata Mahasiswa
3. Kartu Konsultasi Tugas Akhir Pembimbing 1
4. Kartu Konsultasi Tugas Akhir Pembimbing 2
5. *Data Set* Penelitian

