

**KOMPARASI ALGORITMA NAÏVE BAYES DAN SVM PADA
ANALISIS SENTIMEN ULASAN PRODUK DI MARKETPLACE**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

NURUL ZALZA BILAL JANNAH

20.11.3682

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

**KOMPARASI ALGORITMA NAÏVE BAYES DAN SVM PADA
ANALISIS SENTIMEN ULASAN PRODUK DI MARKETPLACE**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

NURUL ZALZA BILAL JANNAH

20.11.3682

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

HALAMAN PERSETUJUAN

SKRIPSI

**KOMPARASI ALGORITMA NAÏVE BAYES DAN SVM PADA ANALISIS
SENTIMEN ULASAN PRODUK DI MARKETPLACE**

yang disusun dan diajukan oleh
NURUL ZALZA BILAL JANNAH
20.11.3682

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 16 Februari 2024

Dosen Pembimbing,



Kusnawi, S.Kom., M.Eng.
NIK. 190302112

HALAMAN PENGESAHAN
SKRIPSI
KOMPARASI ALGORITMA NAÏVE BAYES DAN SVM PADA ANALISIS
SENTIMEN ULASAN PRODUK DI MARKETPLACE

yang disusun dan diajukan oleh

NURUL ZALZA BILAL JANNAH

20.11.3682

Telah dipertahankan di depan Dewan Penguji
pada tanggal 16 Februari 2024.

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Ike Verawati, M.Kom
NIK. 190302237

Rumini, M.Kom
NIK. 190302246

Kusnawi, S.Kom, M.Eng
NIK. 190302112



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 16 Februari 2024

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Nurul Zalza Bilal Jannah
NIM : 20.11.3682

Menyatakan bahwa Skripsi dengan judul berikut:

KOMPARASI ALGORITMA NAÏVE BAYES DAN SVM PADA ANALISIS SENTIMEN ULASAN PRODUK DI MARKETPLACE

Dosen Pembimbing : Kusnawi, S.Kom., M.Eng.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 16 Februari 2024

Yang Menyatakan,



Nurul Zalza Bilal Jannah

HALAMAN PERSEMBAHAN

Alhamdulillah segala puji syukur atas rahmat dan karunia Allah SWT yang telah memberikan petunjuk dan kemudahan bagi penulis dalam menyelesaikan skripsi ini dengan sebaik-baiknya. Dengan ini penulis mempersembahkan skripsi ini kepada semua pihak yang telah membantu dan memberikan motivasi serta dukungan yang besar sehingga skripsi ini bisa diselesaikan dengan sebaik-baiknya.

Melalui skripsi ini penulis mempersembahkan terima kasih kepada:

1. Dengan penuh rasa syukur kepada Allah SWT yang telah memberikan rahmat dan karuniaNya sehingga penulis memiliki kemampuan untuk menyelesaikan skripsi ini.
2. Kepada orang tua penulis Bapak Trimanto dan Ibu Noor Subai'ah yang tak hentinya memberikan doa, cinta kasih, dan dukungannya untuk penulis.
3. Saudara penulis yaitu Reynand, Ajeng, dan Ririn Nurdianti yang sudah menjadi penguat, pendengar, dan pendukung penulis selama ini.
4. Sahabat "Till Jannah" yaitu Ririn, Rafli, Sindhi yang telah menemani penulis dalam dunia perkuliahan, saling mendukung satu sama lain, dan berprogress bersama.
5. Teman perjalanan penulis dalam menyusun skripsi ini, Redi Husein. Terima kasih sudah memberikan semangat dan mendengarkan keluh kesah penulis.
6. Teman-teman kontrak Umi Qhobibah (Alti, Salsa, Fiola) beserta Zahra, Ayu, dan mas Fahry dalam memberikan dukungan kepada penulis untuk menyelesaikan skripsi ini.
7. Teman-teman pengurus HMIF periode 2022/2023 yang telah berjuang dalam mencapai kesuksesan dan memberi dukungan satu sama lain dalam masa sulit selama mengurus HMIF.
8. Seluruh anggota *group* NCT terutama Haechan yang telah memberikan pengaruh positif dan motivasi kepada penulis secara tidak langsung melalui karyanya.
9. *Last but not least*, untuk diri sendiri, terima kasih karena telah berjuang dan menyelesaikan skripsi ini dengan tepat waktu.

KATA PENGANTAR

Segala Puji serta Syukur penulis panjatkan kepada Allah SWT karena berkat Pertolongan serta Kemudahan dari-Nya penulis mampu menyelesaikan skripsi dengan judul **“Komparasi Algoritma Naive Bayes dan SVM pada Analisis Sentimen Ulasan Produk di Marketplace”** dengan tepat waktu. Skripsi ini ditulis dalam rangka memenuhi syarat kelulusan untuk mencapai Gelar Sarjana pada Fakultas Ilmu Komputer Program Studi Informatika Universitas AMIKOM Yogyakarta.

Dalam menyelesaikan penulisan skripsi ini, penulis banyak memperoleh bantuan baik, pengajaran dan arahan dari berbagai pihak secara langsung maupun tidak langsung. Dengan ini penulis menyampaikan terimakasih kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M. selaku rektor Universitas Amikom Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom., Ph. D. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Ibu Windha Mega Pradnya D, M.Kom. selaku Ketua Program Studi SI Informatika Universitas Amikom Yogyakarta.
4. Bapak Kusnawi, S.Kom., M.Eng. selaku dosen pembimbing yang selalu memberikan arahan dan saran dalam proses penulisan skripsi ini.
5. Bapak Trimanto dan Ibu Noor Subai'ah selaku orang tua penulis yang memberikan dukungan moril dan materil, doa dan motivasi kepada penulis.

Penulis menyadari bahwa skripsi ini masih banyak kekurangan karena keterbatasan pengetahuan yang dimiliki oleh penulis. Oleh karena itu kritik dan saran sangat diharapkan bagi perbaikan penelitian selanjutnya. Harapan penulis, semoga skripsi ini bermanfaat untuk para pembaca serta peneliti selanjutnya.

Yogyakarta, 30 Januari 2024

Nurul Zalza Bilal Jannah

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	iv
HALAMAN PERSEMBAHAN.....	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	i
DAFTAR GAMBAR.....	ii
DAFTAR ISTILAH.....	iii
INTISARI.....	iv
<i>ABSTRACT</i>	v
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA.....	6
2.1 Studi Literatur.....	6
2.2 Dasar Teori.....	15
2.2.1 Analisis Sentimen.....	15
2.2.2 <i>Marketplace</i>	15
2.2.3 Produk Perawatan Kecantikan.....	15
2.2.4 Data Mining.....	16
2.2.5 <i>Text Mining</i>	16
2.2.6 Python.....	17
2.2.7 Google Colab.....	17
2.2.8 <i>Preprocessing Text</i>	17
2.2.9 Algoritma <i>Naïve Bayes</i>	18

2.2.10	<i>Support Vector Machine (SVM)</i>	19
2.2.11	<i>K-Fold Cross Validation</i>	20
2.2.12	<i>Confusion Matrix</i>	20
2.2.13	<i>Lexicon Based</i>	21
2.2.14	<i>Exploratory Data Analysis</i>	22
BAB III METODE PENELITIAN		23
3.1	<i>Dataset</i>	23
3.2	<i>Alur Penelitian</i>	23
3.2.1	<i>Scraping Data</i>	24
3.2.2	<i>Cleaning Data</i>	24
3.2.3	<i>Translate</i>	25
3.2.4	<i>Labeling Data</i>	25
3.2.5	<i>Preprocessing Data</i>	25
3.2.6	<i>Wordcloud</i>	26
3.2.7	<i>K-Fold Cross Validation</i>	26
3.2.8	<i>Ekstraksi Fitur</i>	27
3.2.9	<i>Pemodelan Data</i>	27
3.2.10	<i>Komparasi Kinerja Algoritma</i>	28
3.3	<i>Alat dan Bahan</i>	28
3.3.1	<i>Alat Penelitian</i>	28
3.3.2	<i>Bahan Penelitian</i>	29
BAB IV HASIL DAN PEMBAHASAN		31
4.1	<i>Scraping Data</i>	31
4.2	<i>Cleaning Data</i>	31
4.3	<i>Translate Data</i>	32
4.4	<i>Labeling Data</i>	34
4.5	<i>Preprocessing Data</i>	35
4.5.1	<i>Case Folding</i>	35
4.5.2	<i>Cleansing</i>	36

4.5.3	<i>Word Normalization</i>	37
4.5.4	<i>Stopword Removal</i>	38
4.5.5	<i>Stemming</i>	39
4.5.6	<i>Tokenizing</i>	40
4.6	<i>Wordcloud</i>	41
4.7	<i>K-Fold Cross Validation</i>	42
4.8	TF-IDF	43
4.9	Pemodelan Data	45
4.9.1	Pelatihan	45
4.9.1.1	<i>Naïve Bayes</i>	45
4.9.1.2	SVM.....	45
4.9.2	Evaluasi.....	46
4.9.2.1	<i>Naïve Bayes</i>	46
4.9.2.2	SVM.....	46
4.9.3	Hasil	47
4.9.3.1	<i>Naïve Bayes</i>	47
4.9.3.2	SVM	48
4.10	Komparasi Kinerja Algoritma.....	48
BAB V PENUTUP.....		51
5.1	Kesimpulan	51
5.2	Saran	51
REFERENSI.....		53

DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian.....	9
Tabel 2. 2 Ilustrasi <i>Confusion Matrix</i>	20
Tabel 3. 1 Contoh Ulasan.....	24
Tabel 3. 2 Contoh <i>Dataset</i>	29
Tabel 4. 1 <i>Cleaning Data</i>	31
Tabel 4. 2 <i>Translate Data</i>	33
Tabel 4. 3 <i>Labeling Data</i>	34
Tabel 4. 4 <i>Case Folding</i>	35
Tabel 4. 5 <i>Cleansing</i>	36
Tabel 4. 6 <i>Word Normalization</i>	37
Tabel 4. 7 <i>Stopword Removal</i>	38
Tabel 4. 8 <i>Stemming</i>	39
Tabel 4. 9 <i>Tokenizing</i>	40
Tabel 4. 10 Hasil <i>10-Fold Cross Validation</i>	43
Tabel 4. 11 Hasil <i>Confusion Matrix NB</i>	46
Tabel 4. 12 Hasil <i>Confusion Matrix SVM</i>	47

DAFTAR GAMBAR

Gambar 2. 1 <i>Hyperplane SVM</i>	19
Gambar 2. 2 <i>K-Fold Cross Validation</i>	20
Gambar 3. 1 Alur Penelitian	23
Gambar 3. 2 <i>Preprocessing Data</i>	25
Gambar 3. 3 Pemodelan Data	27
Gambar 4. 1 <i>Cleaning Data Manual</i>	32
Gambar 4. 2 <i>Query Translate Data</i>	33
Gambar 4. 3 <i>Query Labeling Data</i>	34
Gambar 4. 4 <i>Wordcloud Data Positif</i>	41
Gambar 4. 5 <i>Wordcloud Data Negatif</i>	42
Gambar 4. 6 <i>Query K-Fold Cross Validation</i>	42
Gambar 4. 7 <i>Query TF-IDF</i>	44
Gambar 4. 8 <i>Heatmap TF-IDF</i>	44
Gambar 4. 9 Pelatihan <i>Naïve Bayes</i>	45
Gambar 4. 10 Pelatihan SVM	46
Gambar 4. 11 Perbandingan Akurasi NB dan SVM Pada Setiap <i>Fold</i>	49
Gambar 4. 12 Perbandingan Rata-Rata Akurasi NB dan SVM	49

DAFTAR ISTILAH

Algoritma	metode yang direncanakan secara sistematis untuk menyelesaikan suatu masalah
Dataset	kumpulan data terstruktur
Data Training	bagian dari dataset yang digunakan untuk melatih model
Data Testing	bagian dari dataset yang digunakan untuk menguji model
Heatmap	representasi grafik dari data dalam bentuk matriks yang divisualisasikan menggunakan warna
Fold	lipatan atau bagian dari dataset yang digunakan selama iterasi dalam proses cross validation
Library	modul untuk mempercepat pengembangan perangkat lunak
Platform	rencana kerja atau media untuk menjalankan software

INTISARI

Pada saat ini semakin banyak orang beralih untuk berbelanja secara *online* di *marketplace* yang ada seperti Shopee. *Marketplace* memberikan berbagai keuntungan dan kerugian kepada *customer* seperti biaya lebih murah dan barang yang dikirim tidak sesuai dengan pesanan. Ulasan produk dari *customer* sangat berpengaruh kepada tingkat penjualan pelaku bisnis sehingga dilakukan analisis sentimen. Pentingnya melakukan analisis sentimen ulasan produk di *marketplace* untuk menambah gambaran tentang bagaimana produk tersebut diterima oleh pengguna.

Penelitian ini menggunakan algoritma *Naïve Bayes* dan SVM untuk analisis sentimen *dataset review* produk perawatan kecantikan didapat dari hasil *scraping* Shopee. *Dataset* dibersihkan dengan mengeliminasi ulasan yang tidak sesuai dengan tujuan penelitian. Data diolah dengan diberi label sentimen positif atau negatif menggunakan *vader lexicon*. Selanjutnya melalui tahapan *preprocessing* (*case folding, cleansing, word normalization, stopword removal, stemming, dan tokenizing*). Kemudian dibagi menggunakan metode *K-Fold Cross Validation* dilanjutkan dengan TF-IDF dan pemodelan data. Algoritma *Naïve Bayes* memperoleh nilai akurasi paling tinggi sebesar 85,53% pada *fold* 2 dan nilai akurasi paling rendah sebesar 77,16% pada *fold* 3. Sedangkan algoritma SVM memperoleh nilai akurasi paling tinggi sebesar 88,58% pada *fold* 2 dan nilai akurasi paling rendah sebesar 82,99% pada *fold* 7. Dengan ini dinyatakan bahwa SVM dapat bekerja lebih baik untuk analisis sentimen ulasan produk perawatan kecantikan pada *marketplace* Shopee karena memperoleh nilai rata-rata akurasi lebih tinggi sebesar 86,14% dibandingkan dengan algoritma *Naïve Bayes*.

Kata kunci: Analisis Sentimen, Ulasan Produk, *Naïve Bayes*, SVM, *Marketplace*.

ABSTRACT

At this time more and more people are switching to shopping online in existing marketplaces such as Shopee. Marketplaces provide various advantages and disadvantages to customers such as lower costs and goods sent not according to orders. Product reviews from customers greatly affect the sales level of business people so that sentiment analysis is carried out. The importance of conducting sentiment analysis of product reviews in the marketplace is to add an overview of how the product is received by users.

This research uses Naïve Bayes and SVM algorithms for sentiment analysis of beauty care product review datasets obtained from Shopee scraping results. The dataset is cleaned by eliminating reviews that do not match the research objectives. The data is processed by labeling positive or negative sentiments using the vader lexicon. Furthermore, it goes through preprocessing stages (case folding, cleansing, word normalization, stopword removal, stemming, and tokenizing). Then divided using the K-Fold Cross Validation method followed by TF-IDF and data modeling. The Naïve Bayes algorithm obtained the highest accuracy value of 85.53% on fold 2 and the lowest accuracy value of 77.16% on fold 3. While the SVM algorithm obtained the highest accuracy value of 88.58% on fold 2 and the lowest accuracy value of 82.99% on fold 7. With this it is stated that SVM can work better for sentiment analysis of beauty care product reviews on the Shopee marketplace because it gets a higher average accuracy value of 86.14% compared to the Naïve Bayes algorithm.

Keyword: *Sentiment Analysis, Review Products, Naïve Bayes, SVM, Marketplaces*