

**DETEKSI ULASAN PADA E-COMMERCE AMAZON
MENGUNAKAN ALGORITMA K-NEAREST NEIGHBOR DAN
PARTICLE SWARM OPTIMIZATION**

Tesis
untuk memenuhi sebagian persyaratan
mencapai derajat Sarjana S-2 Program Studi
Magister Sistem Informasi



Nitami Lestari Putri
30000320420049

SEKOLAH PASCASARJANA
UNIVERSITAS DIPONEGORO
SEMARANG

2023

HALAMAN PENGESAHAN

TESIS

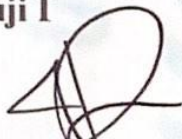
DETEKSI ULASAN PADA E-COMMERCE AMAZON MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR DAN PARTICLE SWARM OPTIMIZATION

Oleh:
Nitami Lestari Putri
30000320420049

Telah diujikan dan dinyatakan lulus ujian tesis pada tanggal 22 Desember 2023 oleh tim penguji Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro.

Semarang, 27 Desember 2023
Mengetahui,

Penguji I



Prof. Dr. Rahmat Gernowo, M.Si
NIP. 196511231994031003

Penguji II



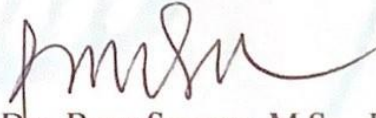
Prof. Dr. Ir. R. Rizal Isnanto, S.T.,
M.M., M.T., IPU, ASEAN Eng.
NIP. 197007272000121001

Pembimbing I



Dr. Budi Warsito, S.Si., M.Si
NIP. 197508241999031003

Pembimbing II



Drs. Bayu Surarso, M.Sc., Ph.D
NIP. 196311051988031001

Mengetahui :
Dekan Sekolah Pascasarjana
Universitas Diponegoro



Dr. R.B. Sularso, S.H., M.Hum
NIP. 196701011991031005

Ketua Program Studi
Magister Sistem Informasi



Drs. Bayu Surarso, M.Sc., Ph.D
NIP. 196311051988031001

**PERNYATAAN PERSETUJUAN
PUBLIKASI TESIS UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Diponegoro, saya yang bertanda tangan di bawah ini:

Nama	:	Nitami Lestari Putri
NIM	:	30000320420049
Program Studi	:	Magister Sistem Informasi
Program	:	Sekolah Pascasarjana
Jenis Karya	:	Tesis

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Diponegoro Hak Bebas Royalti Noneksklusif atas karya ilmiah saya yang berjudul :

**DETEKSI ULASAN PADA E-COMMERCE AMAZON MENGGUNAKAN
ALGORITMA K-NEAREST NEIGHBOR DAN PARTICLE SWARM
OPTIMIZATION**

beserta perangkat yang ada. Dengan Hak bebas Royalti Noneksklusif ini Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*) merawat, dan mempublikasikan tesis saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta.

SEKOLAH PASCASARJANA

Dibuat di : Semarang
Pada tanggal : 8 Desember 2023
Yang menyatakan



Nitami Lestari Putri
NIM. 30000320420049

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam tesis ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar akademik di suatu perguruan tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Semarang, 27 Agustus 2023



Nitami Lestari Putri

SEKOLAH PASCASARJANA

KATA PENGANTAR

Assalamualaikum Warahmatullahi Wabarakatuh.

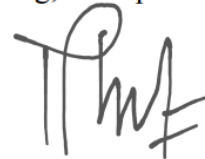
Puji Syukur saya panjatkan ke hadirat Allah SWT karena atas berkat rahmat dan karunia-Nya saya dapat menyelesaikan tesis yang berjudul “Deteksi Ulasan Pada *E-Commerce* Amazon Menggunakan Algoritma *K-Nearest Neighbor* dan *Particle Swarm Optimization*” untuk memenuhi syarat dalam menyelesaikan pendidikan Magister Sistem Informasi, Sekolah Pascasarjana, Universitas Diponegoro. Tidak lupa pula saya panjatkan sholawat dan salam ke hadirat Rasullulah Nabi Muhammad SAW beserta para sahabat, keluarga, dan pengikut beliau hingga akhir jaman.

Pada lembar ini penulis ingin menyampaikan ucapan terima kasih kepada pihak-pihak yang sangat mendukung saya dalam pembuatan dan penyusunan tesis ini, Adapun yang dimaksud adalah sebagai berikut:

1. Bapak Dr. R. B. Sularto, S.H., M.Hum yang merupakan selaku Dekan Sekolah Pascasarjana Universitas Diponegoro.
2. Bapak Drs. Bayu Surarso, M.SC., Ph.D Selaku Ketua Program Studi Magister Sistem Informasi, Sekolah Pascasarjana, Universitas Diponegoro dan juga sebagai Dosen Pembimbing II dalam penyelesaian tesis ini.
3. Bapak Dr. Budi Warsito, S.Si., M.Si Selaku Dosen Pembimbing Utama yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian tesis ini.

Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapatkan keridhaan Allah SWT.

Semarang, 12 September 2023



Nitami Lestari Putri, S.Kom.

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN TESIS	ii
PERNYATAAN PERSETUJUAN	iii
PUBLIKASI TESIS UNTUK KEPENTINGAN AKADEMIS.....	iii
PERNYATAAN.....	iv
KATA PENGANTAR	v
DAFTAR ISI.....	vi
DAFTAR GAMBAR	ix
DAFTAR TABEL	x
DAFTAR ARTI LAMBANG DAN SINGKATAN	xi
ABSTRAK	xiii
ABSTRACT.....	xiv
BAB I PENDAHULUAN	15
1.1 Latar Belakang	15
1.2 Tujuan Penelitian	18
1.3 Manfaat Penelitian	18
BAB II LANDASAN TEORI.....	19
2.1 Tinjauan Pustaka	19
2.2 Dasar Teori.....	22
2.2.1 Ulasan Palsu	22
2.2.2 <i>E-Commerce</i> Amazon	23
2.2.3 Prapengolahan Teks	24
2.2.4 Analisis Sentimen.....	26
2.2.5 Analisis Sentimen Berbasis Leksikon.....	26
2.2.6 Analisis Sentimen <i>Textblob</i>	28
2.2.7 <i>CountVectorizer</i>	29
2.2.8 <i>K-Nearest Neighbor</i>	30
2.2.9 <i>Particle Swarm Optimization</i>	32
2.2.10 <i>Cross validation</i>	33

2.2.11	<i>Confusion Matrix</i>	34
BAB III METODE PENELITIAN.....		36
3.1	Bahan dan Alat Penelitian.....	36
3.1.1	Bahan Penelitian.....	36
3.1.2	Alat Penelitian.....	36
3.2	Prosedur Penelitian.....	36
3.2.1	Studi Literatur	38
3.2.2	Pengumpulan Dataset Ulasan Produk Amazon.....	38
3.2.3	Prapengolahan Teks	38
3.2.4	Pelabelan Sentimen Menggunakan <i>Textblob Library</i>	40
3.2.5	<i>CountVectorizer</i> dan <i>TF-IDF Transformers</i>	40
3.2.6	Split Data.....	41
3.2.7	Klasifikasi Dengan 5 Bagian Validasi Silang	41
3.2.8	Evaluasi <i>Confusion Matrix</i>	41
3.2.9	Hasil Optimasi Akurasi Menggunakan PSO.....	41
3.3	Kerangka Sistem Informasi.....	42
BAB IV HASIL PENELITIAN DAN PEMBAHASAN		45
4.1	Hasil Penelitian	45
4.1.1	Pengumpulan Data	45
4.1.2	Prapengolahan Teks	48
4.1.3	Analisis Sentimen.....	53
4.1.4	Ekstraksi Fitur	56
4.1.4	Split Data.....	59
4.1.5	Klasifikasi Dengan 5 Bagian Validasi Silang	60
4.1.6	Implementasi Sistem	65
4.2	Pembahasan.....	72
4.2.1	Hasil Analisis Sentimen	72
4.2.2	Hasil Klasifikasi <i>K-Nearest Neighbor</i>	76
4.2.3	Hasil <i>Particle Swarm Optimization</i>	81
BAB V KESIMPULAN DAN SARAN.....		82
5.1	Kesimpulan	82
5.2	Saran.....	82

DAFTAR PUSTAKA	83
LAMPIRAN	88



SEKOLAH PASCASARJANA

DAFTAR GAMBAR

Gambar 2.1 Skema KNN 2 dimensi menggunakan $k=1$ dan $k=3$	32
Gambar 2.2 Ilustrasi prosedur dari <i>k-fold cross validation</i>	34
Gambar 2.3 <i>Confusion matrix</i>	35
Gambar 3.1 Diagram alur proses penelitian.....	37
Gambar 3.2 Kerangka Sistem Informasi.....	42
Gambar 4. 1 <i>Flowchart</i> kombinasi algoritma PSO dan KNN	62
Gambar 4.2 Halaman dasbor.....	66
Gambar 4.3 Tampilan sub-halaman Amazon <i>fake product reviews</i> dataset	66
Gambar 4. 4 Tampilan sub-halaman <i>pre-processing</i>	67
Gambar 4. 5 Tampilan sub-halaman <i>Pre-processed</i> data	67
Gambar 4. 6 Tampilan sub-halaman <i>lexicon-based sentiment analysis</i>	68
Gambar 4. 7 Tampilan sub-halaman data <i>sentiment</i>	68
Gambar 4. 8 Tampilan sub-halaman <i>Countvectorizer</i> + TF-IDF Transformer	69
Gambar 4. 9 Hasil pembobotan kata	69
Gambar 4. 10 Tampilan sub-halaman <i>K-Nearest Neighbor</i>	70
Gambar 4. 11 Tampilan sub-halaman <i>K-Nearest Neighbor</i> (lanjutan).....	71
Gambar 4. 12 Tampilan sub-halaman <i>Particle Swarm Optimization</i>	72
Gambar 4. 13 Distribusi sentimen pada <i>dataset</i> ulasan palsu produk Amazon	72
Gambar 4. 14 Visualisasi diagram batang.....	73
Gambar 4. 15 Visualisasi diagram batang.....	75
Gambar 4. 16 Perbandingan akurasi KNN.....	76
Gambar 4. 17 <i>Confusion Matrix</i>	77
Gambar 4. 18 Perbandingan label sentimen.....	80

DAFTAR TABEL

Tabel 2.1 Skema matriks <i>CountVectorizer</i>	30
Tabel 4.1 Daftar fitur pada dataset Amazon	45
Tabel 4.3 Daftar fitur beserta kategori dan jumlah ulasan	46
Tabel 4.4 Ulasan teks	47
Tabel 4.5 <i>Case folding</i>	48
Tabel 4.6 Penghapusan tanda baca.....	49
Tabel 4.7 <i>Stemming</i>	50
Tabel 4.8 <i>Lemmatization</i>	50
Tabel 4.9 Tokenisasi	51
Tabel 4.10 Penghapusan <i>stopwords</i>	52
Tabel 4.11 Skor polaritas	53
Tabel 4.12 Total jumlah label sentimen.....	56
Tabel 4.13 20 kata yang paling sering muncul menggunakan <i>countvectorizer</i>	57
Tabel 4.14 Hasil penerapan TF-IDF <i>Transformers</i>	58
Tabel 4.15 <i>K-Fold Cross Validation</i>	60
Tabel 4.16 Perbandingan hasil akurasi, presisi, <i>recall</i> , dan F1-score pada KNN.....	61
Tabel 4.17 Hasil akurasi KNN-PSO	64

SEKOLAH PASCASARJANA

DAFTAR ARTI LAMBANG DAN SINGKATAN

DAFTAR ARTI LAMBANG

Lambang	Arti Lambang
$\bar{P}$	Nilai polaritas
$\bar{S}$	Nilai subjektivitas
d	Jarak <i>Euclidean</i>
Σ	Mewakili jumlah dari serangkaian bilangan atau variabel
w	Nilai bobot
V	Kecepatan partikel
$pbest$	Posisi terbaik yang pernah dicapai oleh partikel
$gbest$	Posisi terbaik yang pernah dicapai oleh seluruh populasi
tf	<i>Term Frequency</i>
idf	<i>Inverse Document Frequency</i>
C	<i>Count</i>
t	<i>Term Document</i>
d	<i>Document</i>
$\log$	Logaritma bilangan
TP	<i>True positive</i>
TN	<i>True negative</i>
FP	<i>False positive</i>
FN	<i>False negative</i>

DAFTAR SINGKATAN

Singkatan	Kepanjangan Singkatan
TP	<i>Topic Distribution</i>
LDA	<i>Latent Dirichlet Allocation</i>
CD	<i>Cluster Distribution</i>
AHC	<i>Agglomerative Hierarchical Clustering</i>
VP	<i>Verified Purchase</i>
CNN	<i>Convolutional Neural Network</i>
LSTM	<i>Long Short Term-Memory</i>
KNN	<i>K-Nearest Neighbor</i>
KL	<i>Kullback-Leibler</i>
UTSJ	<i>Unsurvised Topic Sentiment Joint</i>
POS	<i>Part of Speech</i>
JST	<i>Joint Sentiment/Topic</i>
SJTSM	<i>Sentence Joint Topic Sentiment Model</i>
SVM	<i>Support Vector Machine</i>
NLP	<i>Natural Language Processing</i>

Singkatan	Kepanjangan Singkatan
API	<i>Application Programming Interface</i>
CV	<i>Count Vectorizer</i>
PSO	<i>Particle Swarm Optimization</i>
NLTK	<i>Natural Language Toolkit</i>
TF-IDF	<i>Term Frequency – Inverse Document Frequency</i>
lb	<i>lower bound</i>
ub	<i>upper bound</i>



SEKOLAH PASCASARJANA

**DETEKSI ULASAN PADA *E-COMMERCE* AMAZON MENGGUNAKAN
ALGORITMA *K-NEAREST NEIGHBOR* DAN *PARTICLE SWARM
OPTIMIZATION***

ABSTRAK

Ulasan *online* menjadi faktor penting yang mendorong konsumen untuk membeli barang di *e-commerce*. Dalam *e-commerce*, ulasan pelanggan sebelumnya dapat membantu pembeli membuat keputusan yang lebih baik dengan memberikan informasi tentang kualitas produk, kekuatan dan kelemahan, perilaku penjual, harga, dan waktu pengiriman. Namun, keberadaan ulasan palsu menimbulkan tantangan dalam menilai sentimen yang diungkapkan oleh pelanggan asli secara benar. Dalam penelitian ini, berfokus pada analisis sentimen dan bertujuan untuk mengeksplorasi peran sentimen dalam ulasan produk Amazon. Penelitian ini menggunakan *K-Nearest Neighbor* (KNN) dengan menghitung optimasi menggunakan *Particle Swarm Optimization* (PSO) untuk memperoleh hasil kinerja terbaik dalam analisis sentimen dalam ulasan produk Amazon. Dalam mengekstrak skor polaritas dari ulasan, penelitian ini menggunakan pendekatan analisis sentimen berbasis leksikon yaitu *Textblob Library* dan menetapkan label sentimen dari ulasan produk. Hasil dari pemodelan yang diusulkan mencapai tingkat akurasi sebesar 83.07% pada nilai $k=15$. Kemudian, nilai akurasi terbaik yang diperoleh KNN dioptimasi menggunakan PSO dan memperoleh nilai k optimal pada $k=18$ dengan akurasi sebesar 83.28%. Hasil dari penelitian ini dapat membantu konsumen dalam membuat Keputusan pembelian dan membantu penjual dalam meningkatkan nilai produk dan layanan mereka berdasarkan *feedback* yang diberikan oleh pelanggan.

Kata Kunci : Ulasan produk Amazon, Analisis sentimen, *Textblob Library*, *K-Nearest Neighbor*, *Particle Swarm Optimization*

REVIEW DETECTION ON AMAZON E-COMMERCE USING K-NEAREST NEIGHBOR ALGORITHM AND PARTICLE SWARM OPTIMIZATION

ABSTRACT

Online reviews are a significant influence encouraging consumers to buy things in e-commerce. In e-commerce, previous customer reviews can help buyers make better decisions by providing information about product quality, strengths and weaknesses, seller behaviour, prices, and delivery times. However, the existence of fake reviews poses a challenge in accurately assessing the sentiments of real customers. In this study, we focus on sentiment analysis and aim to explore the role of sentiment in Amazon product reviews. This study uses K-Nearest Neighbor (KNN) by calculating optimization using Particle Swarm Optimization (PSO) to obtain the best performance results in sentiment analysis in Amazon product reviews. In exploring polarity scores from reviews, this study uses a lexicon-based sentiment analysis approach, namely the Textblob Library that attaches sentiment labels from product reviews. The results of the proposed modeling reach an accuracy level of 83.07% at $k=15$. Then, the best accuracy value obtained by KNN was optimized using PSO and obtained the optimal k value at $k=18$ with an accuracy of 83.28%. The findings of this study will help buyers decide what to buy, and vendors will be able to raise the value of their goods and services by utilizing customer input.

Keywords : Amazon's product reviews, Sentiment analysis, Textblob Library, K-Nearest Neighbor, Particle Swarm Optimization

SEKOLAH PASCASARJANA