



**KOMPARASI ANALISIS KLASIFIKASI SENTIMEN ULASAN DI
GOOGLE PLAY STORE BERBASIS *WORD CLOUD* DENGAN
ALGORITMA NAÏVE BAYES DAN *K-NEAREST NEIGHBOR* (K-NN)
(STUDI KASUS : APLIKASI TOKOPEDIA)**

SKRIPSI

**Untuk memenuhi persyaratan
dalam menyelesaikan program sarjana Strata-1 Matematika**

**Oleh:
TRI DONI ARSIWENDO PURBA
NIM 1911011210019**

**PROGRAM STUDI MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
SEPTEMBER 2024**

HALAMAN PENGESAHAN

SKRIPSI

**Komparasi Analisis Klasifikasi Sentimen Ulasan Di Google Play Store Berbasis Word
Cloud Dengan Algoritma Naïve Bayes dan K-Nearest Neighbor (KNN)
(Studi Kasus : Aplikasi Tokopedia)**

Oleh:
Tri Doni Arsiwendo Purba
1911011210019



telah dipertahankan di depan Dosen Penguji pada tanggal *02 Oktober 2024*
Susunan Dosen Penguji:

Pembimbing I

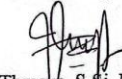


Akhmad Yusuf, S.Si, M.Kom
NIP198004022005011001

Dosen Penguji:

1. Oni Soesanto, S.Si, M.Si ()
2. Dr. Muhammad Ahsar K, S.Si,
M.Sc ()


Pembimbing II



Thresya, S.Si, M.Si
NIP 197205042000122002



Disetujui pada tanggal *3 Oktober 2024*
Studi Matematika FMIPA ULM


Umah Hijriati, S.Si, M.Si
NIP 197911222008012013

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis di acu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, 01 Oktober 2024



Tri Doni Arsiwendo Purba

1911011210019

ABSTRAK

KOMPARASI ANALISIS KLASIFIKASI SENTIMEN ULASAN DI GOOGLE PLAY STORE BERBASIS WORD CLOUD DENGAN ALGORITMA NAIVE BAYES DAN K-NEAREST NEIGHBOR (K-NN) (STUDI KASUS : APLIKASI TOKOPEDIA), (Oleh : Tri Doni Arsiwendo Purba ; Pembimbing : Akhmad Yusuf, S.Si., M.Kom, dan Thresye, S.Si., M.Si., 2024; 82 halaman)

Penelitian ini bertujuan untuk mendokumentasikan sentimen konsumen dan mengidentifikasi sentimen yang paling sering muncul pada ulasan aplikasi Tokopedia di *Google Play Store*. Analisis sentimen dilakukan dengan menggunakan *Word Cloud*, *Naïve Bayes*, dan *K-Nearest Neighbor* (KNN). Penelitian ini menggunakan metode studi literatur untuk mengumpulkan referensi pendukung yang berkaitan dengan analisis sentimen, algoritma *Naïve Bayes*, dan KNN, serta metode studi kasus untuk mengumpulkan data ulasan pengguna aplikasi Tokopedia di *Google Play Store*. Hasil penelitian ini memberikan wawasan tentang sentimen konsumen terhadap aplikasi Tokopedia dan membantu dalam memahami faktor-faktor yang memengaruhi kepuasan pengguna. *Word Cloud* menampilkan kata-kata kunci yang paling umum muncul dalam ulasan, seperti "tokopedia", " kirim", dan "barang". Klasifikasi sentimen dilakukan dengan menggunakan *Naïve Bayes* dan KNN, dengan hasil menunjukkan bahwa *Naïve Bayes* (skenario III) memberikan akurasi terbaik yaitu 86,50%, presisi 87,83%, *recall* 97,65%, dan *F-1 score* 92,48%.

Kata kunci: Analisi Sentimen, *Word Cloud*, Klasifikasi, *Naïve Bayes*, *K-Nearest Neighbor* (KNN)

ABSTRACT

A COMPARISON OF SENTIMENT CLASSIFICATION ANALYSIS OF REVIEWS ON THE GOOGLE PLAY STORE BASED ON WORD CLOUD WITH NAIVE BAYES AND K-NEAREST NEIGHBOR (K-NN) ALGORITHMS (CASE STUDY: TOKOPEDIA APPLICATION), (By: Tri Doni Arsiwendo Purba; Supervisor: Akhmad Yusuf, S.Si., M.Kom, and Thresye, S.Si., M.Si., 2024; 82 pages)

This research aims to document consumer sentiment and identify the most frequent sentiments expressed in reviews of the Tokopedia app on the Google Play Store. Sentiment analysis was conducted using Word Cloud, Naïve Bayes, and K-Nearest Neighbor (KNN). This research employed a literature review method to gather supporting references related to sentiment analysis, the Naïve Bayes algorithm, and KNN, as well as a case study method to collect data on user reviews of the Tokopedia app on the Google Play Store. The findings of this research provide insights into consumer sentiment towards the Tokopedia app and help understand the factors influencing user satisfaction. The Word Cloud displays the most common keywords appearing in reviews, such as "tokopedia," " kirim," and "barang." Sentiment classification was performed using Naïve Bayes and KNN, with results showing that Naïve Bayes (scenario III) achieved the best accuracy at 86.50%, precision at 87.83%, recall at 97.65%, and F-1 score at 92.48%.

Keywords: Sentiment Analysis, Word Cloud, Classification, Naïve Bayes, K-Nearest Neighbor (KNN)

PRAKATA

Puji syukur kepada Tuhan Yesus Kristus, Sang Sumber Hikmat dan Kebenaran, yang senantiasa menyertai dan menuntun penulis dalam menyelesaikan skripsi ini. Melalui kasih karunia-Nya, penulis dapat menyelesaikan skripsi yang berjudul "KOMPARASI ANALISIS KLASIFIKASI SENTIMEN ULASAN DI GOOGLE PLAY STORE BERBASIS WORD CLOUD DENGAN ALGORITMA NAIVE BAYES DAN K-NEAREST NEIGHBOR (K-NN) (STUDI KASUS: APLIKASI TOKOPEDIA)". Penyusunan skripsi ini bertujuan untuk memenuhi salah satu persyaratan dalam rangka menyelesaikan program sarjana strata-1 Matematika di Program Studi MAtematika Fakultas MIPA Universitas Lambung Mangkurat Banjarbaru.

Penulis menyadari sepenuhnya bahwa dalam penyusunan skripsi ini tidak terlepas dari bantuan, dukungan, kerjasama, maupun bimbingan dari berbagai pihak. Oleh karena itu, dalam kesempatan ini penulis ingin mengucapkan terimakasih yang sebesar-besarnya kepada :

1. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lambung Mangkurat Banjarbaru.
2. Ketua Program Studi Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lambung Mangkurat Banjarbaru.
3. Bapak Dr. Pardi Affandi, S.Si, M.Sc selaku dosen pembimbing akademik yang telah memberikan bimbingan, arahan, dan nasihat selama masa perkuliahan.
4. Bapak Akhmad Yusuf, S.Si, M.Kom dan Ibu Thresye, S.Si, M.Si selaku dosen pembimbing yang telah bersedia meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan, arahan, dan nasihat hingga akhirnya penulis dapat menyelesaikan skripsi ini.
5. Dosen penguji Bapak Oni Soesanto, S.Si, M.Si dan Bapak Dr. Muhammad Ahsar Karim, S.Si, M.Sc yang telah memberikan kritik dan saran demi kesempurnaan skripsi ini.
6. Segenap dosen-dosen yang telah membekali ilmu yang bermanfaat kepada penulis selama perkuliahan.

7. Orang tua, abang, kakak, dan adik penulis yang telah bekerja keras untuk melengkapi semua kebutuhan penulis selama masa pendidikan dan mempercayai, mendukung sepenuh hati penulis dalam menyelesaikan pendidikan.
8. Saulina Simanullang S.Tr.Kom selaku teman dekat saya yang telah banyak membantu dan kebersamai saya dengan meluangkan waktu, tenaga dan pikiran selama perkuliahan
9. Seluruh rekan mahasiswa Matematika FMIPA ULM banjarbaru dan semua pihak yang turut membantu hingga selesainya skripsi ini yang tidak dapat penulis sebutkan satu-persatu, terimakasih.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna, namun penulis berharap skripsi ini dapat bermanfaat bagi pembaca dan menjadi sumbangsih bagi pengembangan ilmu pengetahuan. Semoga Tuhan Yesus memberkati.

Banjarbaru, Oktober 2024

Tri Doni Arsiwendo Purba

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
PERNYATAAN	iii
ABSTRAK	iv
ABSTRACT	v
PRAKATA	vi
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	Xi
DAFTAR LAMPIRAN	xii
ARTI LAMBANG DAN SINGKATAN	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan Penelitian	3
1.3 Sistematika Penulisan	3
BAB II TINJAUAN PUSTAKA	5
2.1 <i>Google Play Store</i> dan Ulasan Pelanggan	5
2.2 <i>Word Cloud</i>	5
2.3 <i>Web Scraping</i>	5
2.4 <i>Preprocessing</i>	6
2.5 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	7
2.6 Analisis Sentimen	8
2.7 Klasifikasi	10
2.8 Naïve Bayes	10
2.9 <i>K-Nearest Neighbor (KNN)</i>	11
2.10 <i>Confusion Matrix</i>	12
BAB III METODE PENELITIAN	14
3.1 Sumber Data	14

3.2 Pembagian Data	14
3.3 Langkah-langkah Penelitian	14
3.4 <i>Flowchart</i> Penelitian	16
BAB IV HASIL DAN PEMBAHASAN	17
4.1 Pengumpulan Data	17
4.2 Pelabelan Data	20
4.3 Normalisasi Data Text	21
4.4 <i>Preprocessing</i> Data	22
4.4.1 <i>Cleaning</i> dan <i>Case Folding</i>	23
4.4.2 <i>Stopword Removal</i>	23
4.4.3 <i>Tokenizing</i>	24
4.4.4 <i>Stemming</i>	24
4.5 <i>Word Cloud</i>	25
4.6 Pembagian Data	26
4.7 Pembobotan Data	26
4.8 Proses Manual Klasifikasi Naïve Bayes	31
4.9 Implementasi Klasifikasi Naïve Bayes dalam Program	35
4.10 Proses Manual Klasifikasi KNN dengan <i>Cosine</i> <i>Simalirity</i>	36
4.11 Implementasi Klasifikasi KNN dengan <i>Cosine</i> <i>Simalirity</i> dalam Program	41
4.12 Evaluasi Model	42
4.12.1 Evaluasi Model Naïve Bayes	42
4.12.2 Evaluasi Model KNN dengan <i>Cosine</i> <i>Simalirity</i>	44
4.13 Intrepretasi Hasil	45
BAB V PENUTUP	47
5.1 Kesimpulan	47
5.2 Saran	48
DAFTAR PUSTAKA	49
LAMPIRAN	51

DAFTAR TABEL

Tabel	Halaman
2.1 Ilustrasi Confusion <i>matrix</i>	13
4.1 Hasil Data yang diambil dan akan digunakan, lalu disimpan ke dalam file format csv	19
4.2 Hasil Pelabelan Data Ulasan	20
4.3 Hasil Proses Normalisasi data text	22
4.4 Hasil <i>Preprocessing Cleaning</i> dan <i>Case folding</i>	23
4.5 Hasil <i>Preprocessing Stopword Removal</i>	24
4.6 Hasil <i>Preprocessing Tokenizing</i>	24
4.7 Hasil <i>Preprocessing Stemming</i>	24
4.8 10 Kata paling banyak muncul	25
4.9 Skenario split data	26
4.10 Daset Ulasan	26
4.11 Dataset ulasan setelah <i>preprocessing</i>	27
4.12 Hasil perhitungan TF dan df_t	28
4.13 Hasil perhitungan IDF_t	29
4.14 Hasil perhitungan TF-IDF	31
4.15 Pembagian dataset latihan dan uji setelah <i>preprocessing</i>	32
4.16 Frekuensi kemunculan <i>term</i> pada setiap kelas	33
4.17 Probabilitas setiap <i>term</i> pada setiap kelas	34
4.18 Hasil Klasifikasi dengan metode Naïve Bayes	35
4.19 Hasil perkalian skalar TF-IDF data latihan dan data uji	37
4.20 Hasil kuadrat, total kuadrat, dan akar kuadrat dari total kuadrat tiap TF-IDF setiap data latihan D dan data uji Q	39
4.21 Hasil <i>Cosine Simalirity</i>	40
4.22 Hasil jarak data uji yang telah diurutkan dari yang terbesar ke terkecil	40
4.23 Confusion Matrix dan K terbaik pada skenario masing-masing	41
4.24 Hasil Evaluasi Naïve Bayes dan KNN	46

DAFTAR GAMBAR

Gambar	Halaman
Gambar 1. Alur Penelitian	16
Gambar 2. Pengambilan Data dengan <i>Web Scrapping</i>	17
Gambar 3. Hasil pengambilan data dengan teknik <i>Web Scrapping</i>	18
Gambar 4. Menampilkan Kolom <i>Dataframe Content</i>	18
Gambar 5. Jumlah Dataset Positif dan Negatif	21
Gambar 6. Alur <i>Preprocessing</i> Data	22
Gambar 7. Hasil <i>Word Cloud</i>	25

DAFTAR LAMPIRAN

Lampiran	Halaman
Lampiran 1. Dataset Ulasan Pengguna Aplikasi Tokopedia di <i>Google Play Store</i>	51
Lampiran 2. Dataset yang telah dilabelkan dan dinormalisasi	52
Lampiran 3. Dataset Hasil <i>Preprocessing</i>	53
Lampiran 4. Codingan Tahapan pengambilan Dataset Ulasan	54
Lampiran 5. Codingan Tahapan <i>Preprocessing</i>	55
Lampiran 6. Codingan Pembobotan Kata (TF-IDF)	59
Lampiran 7. Codingan Naïve Bayes untuk mencari <i>Accuracy</i> dan <i>Confusion Matrix</i> pada masing-masing skenario	60
Lampiran 8. Codingan Mencari nilai <i>Cosine Similarity</i>	63
Lampiran 9. Codingan Metode KNN untuk mencari <i>Accuracy</i> , K terbaik, dan <i>Confusion Matrix</i> pada masing-masing skenario	63

ARTI LAMBANG DAN SINGKATAN

TF	: jumlah keberadaan <i>term</i> dalam suatu dokumen
$tf_{(t,d)}$: nilai <i>term</i> frekuensi <i>term</i> (t) pada dokumen (d)
$IDF_{(t)}$: Nilai inverse document frequency pada <i>term</i> (t)
df_t	: Jumlah dokumen yang mengandung <i>term</i> (<i>document frequency</i>)
N	: Jumlah keseluruhan dokumen
$W_{(t,d)}$: hasil perkalian TF dan $IDF_{(t)}$
$P(V_j X_i)$: probabilitas kelas V_j diberikan kata X_i (<i>posterior probability</i>)
$P(V_j)$: probabilitas prior kelas V_j
$P(X_i V_j)$: probabilitas kata X_i untuk kelas V_j (<i>conditional probability</i>)
V_j	: kelas klasifikasi $j = 0$ (kelas negatif), $j = 1$ (kelas positif)
N_j	: jumlah dokumen pada kelas j
w_{ij}	: bobot <i>term</i> i pada kelas j
$\sum_{w' \in T} w_{ij}'$: jumlah bobot <i>term</i> pada dokumen kelas j
B	: jumlah seluruh <i>term</i> pada dokumen
Cos_{ij}	: Nilai kemiripan data uji dan data latih
d_{ik}	: panjang vektor d_i data latih pada data ke- k
d_{jk}	: panjang vektor d_j data uji pada data ke- k
k	: data ke- k
TP	: <i>True Positive</i>
TN	: <i>True Negative</i>
FP	: <i>False Positive</i>
FN	: <i>False Negative</i>
KNN	: <i>K-Nearest Neighbor</i>