

**PERBANDINGAN METODE K – NEAREST NEIGHBOR (KNN) DAN
NAIVE BAYES TERHADAP ANALISIS SENTIMEN PADA PENGGUNA
E-WALLET APLIKASI DANA MENGGUNAKAN FITUR EKSTRAKSI
TF-IDF**

SKRIPSI

Diajukan Oleh :

Muhammad Rayhan Elfansyah

2011102441075



PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR

JULI 2024

**PERBANDINGAN METODE K – NEAREST NEIGHBOR (KNN) DAN
NAIVE BAYES TERHADAP ANALISIS SENTIMEN PADA PENGGUNA
E-WALLET APLIKASI DANA MENGGUNAKAN FITUR EKSTRAKSI
TF-IDF**

SKRIPSI

Diajukan Sebagai Salah Satu Persyaratan Untuk Memperoleh Gelar Sarjana Teknik
Informatika Fakultas Sains Dan Teknologi Universitas Muhammadiyah Kalimantan Timur

Diajukan Oleh :

Muhammad Rayhan Elfansyah

2011102441075



PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR

JULI 2024

HALAMAN PERSETUJUAN

**PERBANDINGAN METODE K – NEAREST NEIGHBOR (KNN) DAN
NAIVE BAYES TERHADAP ANALISIS SENTIMENT PADA
PENGGUNA E-WALLET APLIKASI DANA MENGGUNAKAN FITUR
EKSTRAKSI TF-IDF**

SKRIPSI

Diajukan Oleh :

Muhammad Rayhan Elfansyah

2011102441075

Disetujui untuk diujikan

Pada tanggal 27 Juni 2024

Pembimbing



Rudiman, S.Kom, M.Sc

NIDN 1105068202

Mengetahui,

Koordinator Tugas Akhir/Skripsi/Tesis/Disertasi



Abdul Rahim, S.Kom., M.Cs

NIDN 0009047901

HALAMAN PENGESAHAN

**PERBANDINGAN METODE K – NEAREST NEIGHBOR (KNN) DAN
NAIVE BAYES TERHADAP ANALISIS SENTIMENT PADA
PENGGUNA E-WALLET APLIKASI DANA MENGGUNAKAN FITUR
EKSTRAKSI TF-IDF**

Tugas akhir/skripsi



Diajukan oleh:

Muhammad Rayhan Elfansyah

2011102441075

Diseminarkan dan Diujikan

Pada Tanggal 17 Juli 2024

Penguji I	Penguji II
 <u>Fendy Yulianto, S.Kom., M.Kom.</u> NIDN 1102079402	 <u>Rudiman, S.Kom, M.Sc</u> NIDN 1105068202

Mengetahui,

Ketua

Program Studi/Teknik Informatika



Arbansyah, S.Kom, M.TI

NIDN. 1118019203

PERNYATAAN KEASLIAN PENELITIAN

Saya yang bertanda tangan di bawah ini:

Nama : Muhammad Rayhan Elfansyah
NIM : 2011102441075
Program Studi : S1 Teknik Informatika
Judul Penelitian : Perbandingan Metode K-Nearest Neighbor (KNN) dan Naive Bayes Terhadap Analisis Sentiment Pada Pengguna Aplikasi DANA Menggunakan Fitur Ekstraksi TF-IDF

Menyatakan bahwa Skripsi yang saya tulis ini benar-benar hasil karya saya sendiri, dan bukan merupakan hasil plagiasi/falsifikasi/fabrikasi baik sebagian atau seluruhnya.

Atas pernyataan ini, saya siap menanggung resiko atau sanksi yang dijatuhkan kepada saya apa bila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam skripsi saya ini, atau klaim dari pihak lain terhadap keaslian karya saya ini.

Samarinda, 30 Juni 2024

Yang membuat pernyataan



Muhammad Rayhan Elfansyah

NIM : 2011102441075

ABSTRAK

Penelitian ini membandingkan akurasi metode K-Nearest Neighbor (KNN) dan Naive Bayes dalam klasifikasi sentimen pengguna aplikasi *E-Wallet* DANA menggunakan fitur ekstraksi *Term Frequency-Inverse Document Frequency* (TF-IDF). Data ulasan pengguna dikumpulkan melalui teknik web scraping dan dilabeli oleh ahli bahasa dan model *lexicon*. Setelah melalui proses *pre-processing* seperti *case folding*, *cleaning*, *tokenizing*, *stopword removal*, dan *stemming*, data diklasifikasikan menggunakan metode KNN dan Naive Bayes. Hasil penelitian menunjukkan bahwa pelabelan data oleh ahli bahasa signifikan dalam meningkatkan akurasi kedua metode klasifikasi tersebut. Selain itu, penggunaan TF-IDF sebagai metode pembobotan kata terbukti efektif dalam meningkatkan performa model klasifikasi sentimen. Analisis sentimen terhadap ulasan pengguna aplikasi DANA mengungkapkan berbagai keluhan dan masalah yang dihadapi oleh pengguna, memberikan informasi yang dapat digunakan untuk memperbaiki fitur dan layanan yang diberikan sehingga dapat meningkatkan kepuasan pengguna. Penelitian ini juga memberikan perbandingan antara metode KNN dan Naive Bayes yang dapat menjadi referensi bagi peneliti lain dalam memilih metode yang sesuai untuk analisis sentimen pada dataset serupa.

Kata kunci: Analisis Sentimen, Aplikasi DANA, K-Nearest Neighbor, Naive Bayes

ABTRACT

This research compares the accuracy of the K-Nearest Neighbor (KNN) and Naive Bayes methods in classifying user sentiment towards the DANA e-wallet application using Term Frequency-Inverse Document Frequency (TF-IDF) feature extraction. User review data was collected through web scraping techniques and labeled by linguists and lexicon models. After undergoing pre-processing steps such as case folding, cleaning, tokenizing, stopword removal, and stemming, the data was classified using the KNN and Naive Bayes methods. The research results indicate that data labeling by linguists significantly improves the accuracy of both classification methods. Additionally, using TF-IDF as a word weighting method proves effective in enhancing the performance of sentiment classification models. Sentiment analysis of user reviews of the DANA application reveals various complaints and issues faced by users, providing information that can be used to improve the features and services offered, thereby increasing user satisfaction. This research also provides a comparison between the KNN and Naive Bayes methods, which can serve as a reference for other researchers in selecting appropriate methods for sentiment analysis on similar datasets.

Keywords: Sentiment Analysis, DANA Application, K-Nearest Neighbor, Naive Bayes

PRAKATA

Puji dan syukur saya panjatkan kepada Allah SWT. Atas ridho-Nya saya dapat menyelesaikan penyusunan skripsi ini. Adapun judul skripsi yang saya ajukan adalah “Perbandingan Metode K-Nearest Neighbor (KNN) Dan Naive Bayes Terhadap Analisis Sentimen Pada Pengguna E-Wallet Aplikasi DANA Menggunakan Fitur Ekstraksi TF-IDF ”

Skripsi ini diajukan untuk memenuhi syarat kelulusan mata kuliah Skripsi di Fakultas Sains dan Teknologi, Universitas Muhammadiyah Kalimantan Timur. Tidak dapat disangka bahwa butuh usaha yang keras dalam penyelesaian pengerjaan skripsi ini. Namun, karya ini tidak akan selesai tanpa orang-orang tercinta di sekeliling saya yang mendukung dan membantu. Terima kasih saya sampaikan kepada:

- 1) Kedua orang tua penulis, yaitu superhero dan panutan penulis Bapak Fadliansyah, terimakasih selalu berjuang dalam mengupayakan yang terbaik untuk kehidupan penulis, beliau memang tidak sempat merasakan pendidikan sampai bangku perkuliahan, namun beliau mampu mendidik penulis, memotivasi, memberikan dukungan hingga penulis mampu menyelesaikan studinya sampai sarjana.
- 2) Pintu surga penulis, Ibunda Fatmawati, yang tidak henti-hentinya memberikan kasih sayang dengan penuh cinta dan selalu memberikan dukungan serta doa yang teramat tulus sehingga penulis mampu menyelesaikan studinya sampai sarjana.
- 3) Bapak Dr. Muhammad Musiyam, M.T. Selaku Rektor Univeritas Muhammadiyah Kalimantan Timur yang telah memberi kesempatan pada peneliti untuk menempuh pendidikan di Universitas Muhammadiyah Kalimantan Timur.

- 4) Bapak Arbansyah, Selaku Ketua Program Studi Teknik Informatika dan sebagai dosen pembimbing akademik penulis di Fakultas Sains dan Teknologi Universitas Muhammadiyah Kalimantan Timur.
- 5) Bapak Rudiman, M.Sc. Selaku Dosen Pembimbing penulis yang telah membagi ilmu, bimbingan dan arahan selama proses bimbingan.
- 6) Seorang wanita yang bernama Nadhira Sintha Aulia yang selalu ada disaat senang dan susah, terimakasih atas kesabaran, cinta, dan pengertian yang tiada batas, Kehadiranmu adalah anugerah terindah yang menguatkan langkahku setiap hari.
- 7) Sahabat penulis, Khoirul Huda, Ikhsan Nuttakwa, Ipan Hasmadi, Farhat Jundullah, Zulkarnaen, dan anggota kos lainnya terimakasih atas dukungan dan semangat yang diberikan kepada penulis selama ini.
- 8) Teman satu bimbingan dan teman kuliah penulis yang bernama Rendy Nurdiansyah, Anggiq Karisma, Ilham, Takhta Perlawanan, dll yang sudah membantu saya dan mensupport saya, bersama-sama menyelesaikan skripsi ini sampai akhir.

Semoga segala kebaikan dan pertolongan semuanya mendapat berkah dari Allah SWT. Dan akhirnya penulis menyadari bahwa skripsi ini tidak luput dari berbagai kekurangan, karena keterbatasan ilmu yang saya miliki. Untuk itu dengan kerendahan hati saya mengharapkan saran dan kritik yang sifatnya membangun dari semua pihak demi membangun laporan penelitian ini

Samarinda, 30 Juni 2024



Muhammad Rayhan Elfansyah

DAFTAR ISI

	Halaman
HALAMAN PERSETUJUAN	i
HALAMAN PENGESAHAN.....	ii
PERNYATAAN KEASLIAN PENELITIAN	iii
ABSTRAK	iv
<i>ABTRACT</i>	v
PRAKATA	vi
DAFTAR ISI.....	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xii
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah.....	4
1.3. Tujuan Penelitian.....	4
1.4. Manfaat Penelitian.....	5
BAB II METODOLOGI PENELITIAN.....	6
2.1. Obyek Penelitian.....	6
2.2. Alat Dan Bahan	6
2.3. Prosedur Penelitian	7
2.3.1. Pengumpulan Data (<i>Scrapping</i>)	8
2.3.2. <i>Labelling Data</i>	9
2.3.3. <i>Pre-Processing</i>	10
2.3.4. Pembobotan Kata (TF-IDF)	10
2.3.5. Split Data.....	11
2.3.6. Klasifikasi.....	12
2.3.7. K-Nearest Neighbor (KNN).....	12
2.3.8. Naive Bayes.....	13
2.3.9. Evaluasi	14
2.3.10. Jadwal Penelitian.....	15
BAB III HASIL DAN PEMBAHASAN.....	17
3.1. Hasil Pengumpulan Data	17

3.2.	Hasil <i>Labelling</i> Data.....	18
3.3.	Hasil <i>Pre-Processing</i>	20
3.3.1.	<i>Case Folding</i>	20
3.3.2.	<i>Cleaning</i>	21
3.3.3.	<i>Tokenizing</i>	22
3.3.4.	<i>Stopword Removal</i>	23
3.3.5.	<i>Stemming</i>	24
3.4.	Hasil Pembobotan Kata (TF-IDF).....	25
3.5.	Hasil <i>Visualiasi</i>	28
3.6.	Hasil Evaluasi.....	29
3.6.1.	Metode K-Nearest Neighbor (KNN).....	31
3.6.2.	Metode Naive Bayes	36
BAB IV PENUTUP		39
4.1.	Simpulan.....	39
4.2.	Implikasi	40
DAFTAR RUJUKAN		42
LAMPIRAN		44
RIWAYAT HIDUP		59

DAFTAR TABEL

Tabel	Halaman
2.1 Jadwal Penelitian.....	15
3.1 Hasil Pengumpulan Data.....	17
3.2 Hasil Labelling Data Expert.....	18
3.3 Hasil Labelling Data Lexicon.....	19
3.4 Hasil Case Folding.....	21
3.5 Hasil Cleaning.....	22
3.6 Hasil Tokenizing.....	23
3.7 Hasil Stopword Removal.....	24
3.8 Hasil Stemming.....	24
3.9 Pembagian Data KNN.....	30
3.10 Pembagian Data Naive Bayes.....	30
3.11 Hasil Data Uji KNN Expert.....	35
3.12 Hasil Data Uji KNN Lexicon.....	35
3.13 Hasil Data Uji Naive Bayes Expert.....	37
3.14 Hasil Data Uji Naive Bayes Lexicon.....	37

DAFTAR GAMBAR

Gambar	Halaman
2.1 Alur Penelitian.....	7
2.2 Hasil Pengumpulan Data	8
2.3 Pelabelan Data Expert	9
2.4 Pelabelan Data Lexicon.....	9
3.1 Pembobotan Kata (TF-IDF)	26
3.2 Skoring TF.....	27
3.3 WordCloud Positif.....	28
3.4 WordCloud Negatif.....	28
3.5 WordCloud Netral	29
3.6 Hasil Akurasi Nilai K.....	32
3.7 Grafik Akurasi Nilai K.....	33
3.8 Hasil Klasifikasi KNN Expert.....	33
3.9 Hasil Klasifikasi KNN Lexicon	34
3.10 Hasil Confusion Matriks KNN Expert	34
3.11 Hasil Confusion Matriks KNN Lexicon.....	34
3.12 Hasil Klasifikasi Naive Bayes Expert	36
3.13 Hasil Klasifikasi Naive Bayes Lexicon.....	36
3.14 Hasil Confusion Matriks Naive Bayes Expert.....	36
3.15 Hasil Confusion Matriks Naive Bayes Lexicon	37

DAFTAR LAMPIRAN

Lampiran	Halaman
L1 CV Expert Labelling Data	44
L2 Sertifikat Expert	46
L3 Surat Ijin Penelitian.....	48
L4 Code Pengumpulan Data.....	49
L5 Install Library	49
L6 Membaca Dataset.....	50
L7 Cek Informasi Dataset.....	50
L8 Cek Jumlah Sentimen Positif, Negatif, Dan Netral	50
L9 Membaca Dataset Kolom Review dan Sentimen.....	50
L10 Pre-Processing Data	50
L11 Pembobotan Kata (TF-IDF).....	52
L12 Hasil Akurasi Nilai K.....	52
L13 Grafik Akurasi Nilai K.....	53
L14 Split Data dan Evaluasi Confusion Matrix Metode KNN.....	53
L15 Split Data dan Evaluasi Confusion Matrix Metode Naive Bayes	54
L16 Visualisasi Wordcloud Positif, Negatif, Dan Netral	54
L17 Kartu Kendali Bimbingan	55