

**PERBANDINGAN ANALISIS *WORDNET* DAN *K-NEAREST NEIGHBOR*
PADA ULASAN APLIKASI SIREKAP 2024**

SKRIPSI

Diajukan Oleh :

Emyzar Hafliida Tanjung

2011102441240



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR
SAMARINDA
JULI 2024**

**PERBANDINGAN ANALISIS *WORDNET* DAN *K-NEAREST NEIGHBOR*
PADA ULASAN APLIKASI SIREKAP 2024**

SKRIPSI

Diajukan Sebagai Salah Satu Persyaratan Untuk Memenuhi Gelar Sarjana Fakultas Sains dan
Teknologi Universitas Muhammadiyah Kalimantan Timur

Diajukan Oleh :

Emyzar Hafliida Tanjung

2011102441240



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS MUHAMMADIYAH KALIMANTAN TIMUR**

SAMARINDA

JULI 2024

LEMBAR PERSETUJUAN

**PERBANDINGAN ANALISIS *WORDNET* DAN *K-NEAREST NEIGHBOR*
PADA ULASAN APLIKASI SIREKAP 2024**

Diajukan Oleh :

EMYZAR HAFLIDA TANJUNG

2011102441240

Disetujui untuk diujikan

Pada tanggal ~~30~~ 30 Juni 2024

Pembimbing



Naufal Azmi Verdikha, S.Kom., M.Eng.

NIDN. 1114048801

Mengetahui,

Koordinator Tugas Akhir/Skripsi/Tesis/Disertasi



Abdul Rahim, S.Kom., M.Cs.

NIDN. 111024103

LEMBAR PENGESAHAN

PERBANDINGAN ANALISIS *WORDNET* DAN *K-NEAREST NEIGHBOR* PADA ULASAN APLIKASI SIREKAP 2024

SKRIPSI



Diajukan Oleh :

EMYZAR HAFLIDA TANJUNG

2011102441240

Diseminarkan dan diujikan

Pada tanggal 11 Juli 2024

Penguji I	Penguji II
 <u>Taghfirul Azhima Yoga Siswa, S.Kom., M.Kom</u> NIDN. 1118038805	 <u>Nurfal Azmi Verdikha, S.Kom., M.Eng.</u> NIDN. 1114048801

Mengetahui,

Ketua

Program Studi Teknik Informatika



Arbansyah, S.Kom., M.TI
NIDN. 1118019203

PERNYATAAN KEASLIAN PENELITIAN

Saya yang bertanda tangan di bawah ini:

Nama : Emyzar Hafliida Tanjung

NIM : 2011102441240

Program Studi : S1 Teknik Informatika

Judul Penelitian : “Perbandingan Analisis *WordNet* dan *K-Nearest Neighbor* Pada Ulasan Aplikasi Sirekap 2024”.

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 risiko atau sanksi yang dijatuhkan kepada saya apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam tugas skripsi saya ini, atau klaim dari pihak lain terhadap keaslian karya saya ini

Samarinda, 30 Juni 2024

Yang membuat pernyataan



Emyzar Hafliida Tanjung

2011102441240

ABSTRAK

Perkembangan model klasifikasi telah mencakup bidang klasifikasi sentimen yang bersumber dari data teks. Penelitian mengenai *Machine Learning* dan *Natural Language Processing* (NLP) yang menggunakan data ulasan terhadap suatu aplikasi sudah pernah dilakukan sebelumnya. Bagi pengguna, ulasan aplikasi sering digunakan sebagai sumber informasi untuk mengetahui lebih lanjut tentang aplikasi tersebut. Tujuan penelitian ini untuk mengetahui hasil perbandingan analisis klasifikasi *WordNet* dan *K-Nearest Neighbor* menggunakan ekstraksi fitur TF-IDF dalam mendapatkan hasil dari nilai evaluasi *F1-Score* pada ulasan Aplikasi “Sirekap 2024”. Penelitian ini menggunakan sebanyak 8358 data ulasan, dengan teknik pengambilan data melalui proses *Scraping*. Hasil penelitian berdasarkan uji coba yang telah dilakukan, metode *K-Nearest Neighbor* lebih baik dalam mengklasifikasikan dibanding metode *WordNet*, dengan hasil perbandingan *K-Nearest Neighbor* sebesar 32% sedangkan *WordNet* sebesar 17,50%.

Kata kunci: Sirekap 2024, *K-Nearest Neighbor*, *WordNet*, Perbandingan.

ABSTRACT

The development of classification models has covered the field of sentiment classification sourced from text data. Research on machine learning and natural language processing (NLP) using review data on an application has been done before. For users, application reviews are often used as a source of information to find out more about the application. The purpose of this study is to determine the comparison results of WordNet and K-Nearest Neighbor classification analysis using TF-IDF feature extraction in obtaining the results of the F1-Score evaluation value on the “Sirekap 2024” application review. This research uses a total of 8358 review data, with data retrieval techniques through the Scraping process. The results of research based on trials that have been carried out, the K-Nearest Neighbor method is better at classifying than the WordNet method, with the results of the K-Nearest Neighbor comparison of 32% while WordNet is 17.50%.

Keywords: Sirekap 2024, K-Nearest Neighbor, WordNet, Comparison.

PRAKATA

Alhamdulillah Rabbil'alamin. Puji dan syukur atas kehadiran Allah subhanahu wa ta'ala, atas segala rahmat dan karunia-Nya, yang telah memberikan kekuatan, semangat, dan kesabaran, sehingga membantu penulis dalam menyelesaikan karya skripsi ini.

Dengan segala kekurangan dalam penulisan skripsi ini sangat disadari oleh penulis. Penulis menyadari bahwa selama proses penulisan skripsi banyak pihak yang telah memberikan dorongan, bantuan, bimbingan, motivasi, masukan dan do'a yang tiada hentinya diterima oleh penulis sehingga mampu menyelesaikan skripsi ini. Oleh karena itu penulis ingin mengucapkan terimakasih tak terhingga kepada:

1. Orang tua tercinta Ayahanda Ghazali Tanjung dan Ibunda Paridah yang tiada henti selalu memberikan cinta, kasih sayang, dukungan, nasihat, do'a, serta kesabarannya dalam menghadapi setiap langkah hidup penulis.
2. Bapak Arbansyah, S.Kom., M.TI selaku Ketua Program Studi S1 Teknik Informatika.
3. Bapak Naufal Azmi Verdikha, S.Kom., M.Eng selaku dosen pembimbing yang telah dengan sabar memberikan bimbingan, masukan dan saran-saran yang sangat membantu dalam menyempurnakan skripsi ini.
4. Kepada saudara saudari saya kak Zahra Khalida Tanjung, Maulida Ashfiya Tanjung dan adik saya Muhammad Zulfi Naufal Tanjung, yang selalu memberikan dukungan penuh kepada penulis.
5. Ucapan terima kasih juga penulis sampaikan kepada diri sendiri atas dedikasi dan kerja keras yang telah diberikan selama proses penulisan skripsi ini. Terima kasih telah tetap kuat dan semangat meskipun menghadapi banyak tantangan. Perjalanan ini telah mengajarkan banyak hal berharga yang akan selalu diingat.
6. Rekan-rekan seperjuangan prodi Teknik Informatika Universitas Muhammadiyah Kalimantan Timur yang tidak dapat disebutkan satu persatu.
7. Teruntuk seseorang dimasa depan saya yang masih dipersiapkan oleh Allah, terimakasih sudah mendo'akan saya untuk tidak berpacaran hingga saat ini. Semoga kelak kita dipertemukan diwaktu yang terbaik, ketika kita sudah menjadi orang yang sukses dan bisa membahagiakan orang tersayang, sehingga skripsi ini bisa terbentuk karena sembari mewujudkan mimpi dan menunggu mu tanpa berpacaran.

Samarinda, 30 Juni 2024
Penyusun,



Emyzar Haflida Tanjung

DAFTAR ISI

Halaman

Lembar Persetujuan.....	ii
Lembar Pengesahan.....	iii
Pernyataan Keaslian Penelitian	iv
Abstrak	v
<i>Abstract</i>	vi
Prakata	vii
Daftar Isi.....	viii
Daftar Tabel.....	ix
Daftar Gambar	x
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Masalah.....	3
BAB II METODE PENELITIAN	4
2.1 Objek Penelitian	4
2.2 Alat dan Bahan	4
2.3 Prosedur Penelitian.....	5
BAB III HASIL DAN PEMBAHASAN.....	12
3.1 Hasil.....	12
3.2 Pembahasan	21
BAB IV KESIMPULAN DAN SARAN.....	23
4.1 Kesimpulan.....	23
4.2 Saran.....	23
DAFTAR RUJUKAN	24
LAMPIRAN	26
RIWAYAT HIDUP	39

DAFTAR TABEL

Tabel	Halaman
Tabel 2. 1 Library Python.....	4
Tabel 2. 2 Kriteria Indeks.....	7
Tabel 2. 3 Penjelasan nilai konfusi matriks.....	8
Tabel 3. 1 Frekuensi Masing-Masing Rating	13
Tabel 3. 2 Hasil <i>F1 Score</i> Setiap <i>Fold</i>	20
Tabel 3. 3 Hasil Perbandingan 2 Metode	21

DAFTAR GAMBAR

Gambar	Halaman
Gambar 2. 1 Diagram Alur Penelitian.....	5
Gambar 2. 2 Dasar <i>Confusion Matrix</i>	8
Gambar 2. 3 <i>Confusion matrix 5x5</i>	8
Gambar 2. 4 <i>Cross Validation k = 10</i>	11
Gambar 3. 1 Hasil <i>Scraping Data</i>	12
Gambar 3. 2 Hasil Distribusi <i>Rating</i>	12
Gambar 3. 3 Hasil <i>Wordcloud rating 1</i>	13
Gambar 3. 5 Hasil <i>Add Id</i>	14
Gambar 3. 6 Sebelum dan Setelah <i>Lowercase</i>	14
Gambar 3. 7 Hasil Sebelum dan Setelah <i>Remove Char</i>	15
Gambar 3. 8 Hasil Sebelum dan Setelah <i>Spellchecker</i>	15
Gambar 3. 9 Hasil Sebelum dan Setelah <i>Stemming</i>	15
Gambar 3. 10 Hasil Sebelum dan Setelah <i>Translated</i>	16
Gambar 3. 11 Hasil Hitung Fungsi <i>Score</i>	16
Gambar 3. 12 Hasil <i>Confusion Matrix Wordnet</i>	17
Gambar 3. 13 Hasil Ekstraksi Fitur TF-IDF	18
Gambar 3. 14 Hasil Nilai <i>K-Fold 1</i>	19
Gambar 3. 15 Hasil Nilai <i>K-Fold 10</i>	19
Gambar 3. 16 Hasil <i>F1-Score Macro</i> Setiap <i>Fold</i>	20
Gambar 3. 17 Hasil <i>Confusion Matrix Overall</i>	20
Gambar 3. 18 Hasil Perbandingan 2 Metode.....	21

DAFTAR LAMPIRAN

Lampiran	Halaman
Lampiran 1 Pengambilan Data	26
Lampiran 2 Analisis Data	26
Lampiran 3 Pra Proses.....	28
Lampiran 4 <i>WordNet</i>	30
Lampiran 5 Klasifikasi	31
Lampiran 6 Kartu Kendali Bimbingan	35
Lampiran 7 Hasil Uji Turnitin.....	36
Lampiran 8 Surat Ijin Penelitian	38