

**PERBANDINGAN EKSTRAKSI FITUR *TF-IDF* DAN
WORD2VEC PADA ANALISIS SENTIMEN KUISIONER
EVALUASI DOSEN MENGGUNAKAN *LOGISTIC*
*REGRESSION***

TUGAS AKHIR

Oleh:

ARYO PRATAMA GINANTAKA

NIM.1910817310012



PROGRAM STUDI TEKNOLOGI INFORMASI

FAKULTAS TEKNIK

UNIVERSITAS LAMBUNG MANGKURAT

BANJARMASIN

2024

**PERBANDINGAN EKSTRAKSI FITUR *TF-IDF* DAN
WORD2VEC PADA ANALISIS SENTIMEN KUISIONER
EVALUASI DOSEN MENGGUNAKAN *LOGISTIC*
*REGRESSION***

TUGAS AKHIR

Diajukan untuk Memenuhi Salah Salah Satu Syarat
Sarjana Strata-1 Teknologi Informasi

Oleh:

ARYO PRATAMA GINANTAKA

NIM.1910817310012



PROGRAM STUDI TEKNOLOGI INFORMASI

FAKULTAS TEKNIK

UNIVERSITAS LAMBUNG MANGKURAT

BANJARMASIN

2024

LEMBAR PERNYATAAN

Saya yang bertanda tangan di bawah ini,

Nama : Aryo Pratama Ginantaka

NIM : 1910817310012

Fakultas : Teknik

Program Studi : Teknologi Informasi

Judul Tugas Akhir : Perbandingan Ekstraksi Fitur TF-IDF dan Word2Vec Pada Analisis Sentimen Kuisisioner Evaluasi Dosen Menggunakan Logistic Regression

Dosen Pembimbing : Nurul Fathanah Mustamin, S.Pd., M.T

Dengan ini saya menyatakan bahwa dalam Tugas Akhir 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 rujukan.

Banjarmasin, Desember 2023

Aryo Pratama Ginantaka
NIM.1910817310012

LEMBAR PENGESAHAN

LEMBAR PENGESAHAN SKRIPSI PROGRAM STUDI S-1 TEKNOLOGI INFORMASI

Perbandingan Ekstraksi Fitur *TF-IDF* dan
Word2Vec Pada Analisis Sentimen Kuisioner
Evaluasi Dosen Menggunakan *Logistic Regression*

oleh

Aryo Pratama Ginantaka (1910817310012)

Telah dipertahankan di depan Tim Penguji pada 3 Januari 2024 dan
dinyatakan

L U L U S

Komite Penguji :

Ketua : Muti'a Maulida, S.Kom., M.T.I
NIP 198810272019032013

Anggota 1 : Andreyan Rizky Baskara, S.Kom., M.Kom
NIP 199307032019031011

Anggota 2 : Dr. Ir. Yuslena Sari, S.Kom., M.Kom
NIP 198411202015042002

Pembimbing Utama : Nurul Fathanah Mustamin, S.Pd., MT
NIP 199110252019032018

Banjarbaru, 12 JAN 2024
diketahui dan disahkan oleh:

Wakil Dekan Bidang
Akademik



Andreyan Rizky Baskara, S.T., M.T.
NIP 199307032019031011

Koordinator Program Studi
S-1 Teknologi Informasi.

Andreyan Rizky Baskara, S.Kom., M.Kom
NIP 199307032019031011

PERSETUJUAN TUGAS AKHIR

PERSETUJUAN TUGAS AKHIR

PERBANDINGAN EKSTRAKSI FITUR TF-IDF DAN WORD2VEC PADA
ANALISIS SENTIMEN KUISIONER EVALUASI DOSEN MENGGUNAKAN
LOGISTIC REGRESSION

OLEH

ARYO PRATAMA GINANTAKA

NIM. 1910817310012

Telah diperiksa dan terpenuhi semua persyaratan akademik, administrasi dan disetujui
untuk dipertahankan di hadapan dewan penguji

Banjarmasin, 15 Desember 2023

Pembimbing utama,



Nurul Fathanah Mustamin, S.Pd., M.T.

NIP. 199110252019032018

ABSTRAK

Dosen memegang peranan krusial dalam membentuk lingkungan belajar. Oleh karena itu, evaluasi dosen menjadi aspek penting dalam upaya meningkatkan kualitas kampus. Data dari kolom kritik dan saran pada kuisioner evaluasi dosen dapat dilakukan analisis sentimen untuk mengetahui teks tersebut memiliki nilai positif atau negatif. Analisis sentimen adalah teknik dalam pemrosesan bahasa alami yang bertujuan untuk mengidentifikasi dan mengekstrak sentimen atau opini dari teks. Sentimen termasuk jenis data teks tidak terstruktur, sehingga untuk dapat diproses, teks tersebut harus diubah menjadi data terstruktur dengan proses ekstraksi fitur. Ekstraksi fitur yang umum digunakan adalah TF-IDF dan Word2Vec. Penelitian ini bertujuan untuk membandingkan metode ekstraksi fitur, yaitu TF-IDF dan Word2Vec menggunakan metode Logistic Regression. Dataset yang digunakan adalah kolom kritik dan saran dari kuisioner evaluasi dosen sebanyak 2000 data yang terbagi menjadi kelas positif dan negatif. Penelitian ini melakukan uji coba dengan rasio pembagian data 70%, 80%, 90%, 100% sebagai data latih dan 30%, 20%, 10%, 20 data uji. Pada data latih diterapkan SMOTE untuk menangani ketidakseimbangan kelas. Kemudian dilakukan klasifikasi dengan metode Logistic Regression. Hasil dari penelitian ini yaitu ekstraksi fitur TF-IDF tanpa menggunakan stopword menunjukkan kinerja yang sangat baik dengan accuracy 98%, 99%, 98% dan 90%. Sedangkan ekstraksi fitur Word2Vec tanpa stopword mendapatkan nilai akurasi yang lebih rendah yaitu 87%, 89%, 90%, dan 90%.

Kata Kunci : Analisis Sentimen, Evaluasi Dosen, TF-IDF, Word2Vec, *Logistic Regression*

ABSTRACT

Lecturers play a crucial role in shaping the learning environment. Therefore, evaluating lecturers is an essential aspect of efforts to improve campus quality. Data from the critique and suggestion columns in lecturer evaluation questionnaires can undergo sentiment analysis to determine whether the text has a positive or negative value. Sentiment analysis is a natural language processing technique aimed at identifying and extracting sentiment or opinions from text. Sentiment is a type of unstructured text data, so to process it, the text must be transformed into structured data through the feature extraction process. Commonly used feature extraction methods include TF-IDF and Word2Vec. This study aims to compare feature extraction methods, namely TF-IDF and Word2Vec, using the Logistic Regression method. The dataset used consists of 2000 data points from the critique and suggestion columns in lecturer evaluation questionnaires, divided into positive and negative classes. The study conducts experiments with data splitting ratios of 70%, 80%, 90%, 100% for training data and 30%, 20%, 10%, 20 data points for testing. The training data is subjected to SMOTE to address class imbalance. Classification is then performed using the Logistic Regression method. The results of this study show that TF-IDF feature extraction without using stop words exhibits excellent performance with accuracies of 98%, 99%, 98%, and 90%. In contrast, Word2Vec feature extraction without stop words yields lower accuracy values of 87%, 89%, 90%, and 90%.

Keywords: *Sentiment Analysis, Lecturer Evaluation, TF-IDF, Word2Vec, Logistic Regression*

KATA PENGANTAR

Dengan mengucapkan puji dan syukur atas kehadiran Tuhan Yang Maha Esa. Yang telah melimpahkan anugerah-Nya sehingga penulis dapat menyelesaikan Tugas Akhir ini dengan judul “Perbandingan Ekstraksi Fitur TF-IDF dan Word2Vec Pada Analisis Sentimen Kuisisioner Evaluasi Dosen Menggunakan Logistic Regression” dengan baik. Tugas akhir ini merupakan salah satu syarat kelulusan untuk memperoleh gelar Sarjana Strata-1 pada Program Studi Teknologi Informasi Fakultas Teknik Universitas Lambung Mangkurat (ULM) Banjarmasin.

Selama berlangsungnya penulisan ini, penulis mendapatkan banyak arahan bantuan, binaan, saran dan kritik sehingga penyusunan laporan ini telah sampai ke tahap penyelesaian. Penulis sangat bersyukur dan mengucapkan banyak terima kasih kepada semua pihak yang telah membantu, terutama kepada:

1. Orang tua dan keluarga di rumah yang telah memberikan motivasi dan turut membantu demi kelancaran penyelesaian Tugas Akhir ini.
2. Dosen Pembimbing Utama, Nurul Fathanah Mustamin, S.Pd., M.T, yang telah memberikan waktu, bimbingan dan arahan dalam proses penyelesaian dan penulisan laporan Tugas Akhir ini.
3. Dosen-dosen beserta staf di program Studi Teknologi Informasi yang telah mengarahkan dan teman-teman yang membantu dalam proses penyelesaian Tugas Akhir ini.
4. Teman seperjuangan dari masa perkuliahan sampai Tugas Akhir ini dan seluruh teman-teman angkatan 2019 Program Studi Teknologi Informasi.

Akhir kata, penulis mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah turut serta membantu dalam penyelesaian program dan penulisan laporan ini yang tidak bisa disebutkan satu per satu. Harapan yang paling besar dari penyusunan laporan ini adalah agar bermanfaat untuk banyak orang, baik untuk pribadi, teman-teman, serta pembaca. Penulis juga mengharapkan

saran dan kritik demi perbaikan dan penyempurnaan laporan ini. Semoga laporan ini dapat bermanfaat bagi para pembaca dan semua pihak yang membutuhkan.

Banjarmasin, Desember 2023
Penulis,

Aryo Pratama Ginantaka

DAFTAR ISI

HALAMAN SAMPUL LUAR.....	i
HALAMAN SAMPUL DALAM.....	ii
LEMBAR PERNYATAAN	iii
LEMBAR PENGESAHAN	iv
PERSETUJUAN TUGAS AKHIR	v
ABSTRAK	vi
<i>ABSTRACT</i>	vii
KATA PENGANTAR.....	viii
DAFTAR ISI	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN	xv
DAFTAR RIWAYAT HIDUP.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian.....	4
BAB II TINJAUAN PUSTAKA	5
2.1 Landasan Teori	5
2.1.1 Kuisisioner Evaluasi Dosen	5
2.1.2 Analisis Sentimen	5
2.1.3 Data Mining.....	6
2.1.4 Text Preprocessing.....	6
2.1.5 TF-IDF.....	7
2.1.6 Word2Vec	8
2.1.7 <i>Synthetic Minority Over-sampling Technique (SMOTE)</i>	9
2.1.8 Logistic Regression	10
2.1.9 Confusion Matrix	12

2.2 Penelitian Terkait.....	14
2.2.1 Analisis Perbandingan Pembobotan TF-IDF Dan TF-RF Pada Trending Topic Di Twitter Dengan Menggunakan Klasifikasi K-Nearest Neighbor	14
2.2.2 Analisis Sentimen Twitter Bahasa Indonesia dengan Word2Vec	14
2.2.3 Perbandingan Model Pembobotan TF-IDF Dan Word2Vec Pada Klasifikasi Analisis Sentimen Perpindahan Ibukota Indonesia Dengan Metode <i>Xtreme Gradient Boosting</i>	15
2.2.4 Analisis Sentimen Pengguna Twitter Terhadap Kenaikan Harga Bahan Bakar Minyak (BBM) Menggunakan Metode <i>Logistic Regression</i>	15
2.2.5 Analisis Sentimen Evaluasi Pembelajaran Tatap Muka 100 Persen pada Pengguna Twitter menggunakan Metode <i>Logistic Regression</i>	16
2.3 Kerangka Pemikiran	18
BAB III METODOLOGI PENELITIAN.....	19
3.1 Alat dan Bahan	19
3.2 Alur Penelitian	20
3.2.1 Identifikasi Masalah	20
3.2.2 Studi Literatur.....	20
3.2.3 Pengumpulan Data	20
3.2.4 Melakukan Eksperimen.....	21
3.2.5 Evaluasi	29
3.2.6 Implementasi	29
BAB IV HASIL DAN PEMBAHASAN.....	30
4.1 Hasil.....	30
4.1.1 Pengumpulan Data	30
4.1.2 Pelabelan Data	30
4.1.3 Preprocessing Data	33
4.1.4 Pembagian Data.....	36
4.1.5 TF-IDF.....	36
4.1.6 Word2Vec	38
4.1.7 <i>Synthetic Minority Over-sampling Technique (SMOTE)</i>	39
4.1.8 Logistic Regression	41
4.1.9 Evaluasi	46
4.1.10 Implementasi Sistem	53
4.2 Pembahasan	54

BAB V KESIMPULAN	56
5.1 Kesimpulan.....	56
5.2 Saran	56
DAFTAR PUSTAKA	57

DAFTAR TABEL

Tabel 1.1 Penelitian Poornima dan Sathiya.....	3
Tabel 1.2 Penelitian Aloysius.....	3
Tabel 2.1 Penelitian Terkait.....	17
Tabel 3.1 Contoh Komentar.....	22
Tabel 3.2 Hasil Cleaning.....	23
Tabel 3.3 Hasil Case Folding.....	24
Tabel 3.4 Hasil Tokenization.....	24
Tabel 3.5 Hasil Stopword.....	25
Tabel 3.6 Hasil Stemming.....	25
Tabel 3.7 Contoh Pengujian Confusion Matrix.....	27
Tabel 4.1 Acuan Pelabelan.....	32
Tabel 4.2 Hasil Cleaning.....	33
Tabel 4.3 Hasil Case Folding.....	34
Tabel 4.4 Hasil Tokenization.....	34
Tabel 4.5 Hasil Stopword.....	35
Tabel 4.6 Hasil Stemming.....	36
Tabel 4.7 Hasil TF-IDF.....	37
Tabel 4.8 Hasil SMOTE.....	40
Tabel 4.9 Parameter Logistic Regression.....	41
Tabel 4.10 Hasil Prediksi Logistic Regression dengan TF-IDF DS.....	42
Tabel 4.11 Hasil Prediksi Logistic Regression + TF-IDF TS.....	43
Tabel 4.12 Hasil Prediksi Logistic Regression + W2V DS.....	44
Tabel 4.13 Hasil Prediksi Logistic Regression + W2V TS.....	45
Tabel 4.14 Confusion Matix TF-IDF dengan Stopword.....	46
Tabel 4.15 Classification Report TF-IDF dengan Stopword.....	47
Tabel 4.16 Confusion Matrix TF-IDF tanpa Stopword.....	48
Tabel 4.17 Classification Report TF-IDF tanpa Stopword.....	48
Tabel 4.18 Confusion Matrix W2V dengan Stopword.....	49
Tabel 4.19 Classification Report W2V dengan Stopword.....	50
Tabel 4.20 Confusion Matrix W2V tanpa Stopword.....	50
Tabel 4.21 Classification Report W2V tanpa Stopword.....	51

DAFTAR GAMBAR

Gambar 2.1 Arsitektur CBOW.....	8
Gambar 2.2 Arsitektur Skip Gram	9
Gambar 2.3 Fungsi Logistik	12
Gambar 2.4 Confusion Matrix [18].....	12
Gambar 2.5 Kerangka Pemikiran.....	18
Gambar 3.1 Alur Penelitian.....	20
Gambar 3.2 Tahap Penelitian	21
Gambar 4.1 Data Kuisisioner Evaluasi Dosen	30
Gambar 4.2 Implementasi Pelabelan Otomatis.....	31
Gambar 4.3 Hasil Pelabelan Data Otomatis.....	31
Gambar 4.4 Hasil Pemilihan Data.....	32
Gambar 4.5 Jumlah Data Perkelas	32
Gambar 4.6 Hasil Prediksi Logistic Regression + TF-IDF DS.....	42
Gambar 4.7 Hasil Prediksi Logistic Regression + TF-IDF TS	43
Gambar 4.8 Hasil Prediksi Logistic Regression + W2V DS.....	44
Gambar 4.9 Hasil Prediksi Logistic Regression + W2V TS	45
Gambar 4.10 Halaman Input Teks	53
Gambar 4.11 Halaman Hasil Prediksi	53
Gambar 4.12 Perbandingan Akurasi Model.....	54

DAFTAR LAMPIRAN

Lampiran 1. Source Code Google Collaboratory.....	60
Lampiran 2. Source Code Implementasi Sistem.....	63
Lampiran 3. Hasil Prediksi Logistic Regression.....	68
Lampiran 4. Lembar Konsultasi.....	74
Lampiran 5. Surat Permohonan Data	75