



**PERBANDINGAN EKSTRAKSI FITUR *FASTTEXT* DAN *TF-IDF-ICF*
PADA ANALISIS SENTIMEN DI *TWITTER* TERHADAP ISU
PENUNDAAN PEMILU 2024 MENGGUNAKAN METODE *RANDOM
FOREST***

Skripsi

Untuk Memenuhi Persyaratan

Dalam Menyelesaikan Strata-1 Ilmu Komputer

Oleh

MUHAMMAD RIFQI

NIM 1711016310019

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU**

JUNI 2024



**PERBANDINGAN EKSTRAKSI FITUR *FASTTEXT* DAN *TF-IDF-ICF*
PADA ANALISIS SENTIMEN DI *TWITTER* TERHADAP ISU
PENUNDAAN PEMILU 2024 MENGGUNAKAN METODE *RANDOM
FOREST***

Skripsi

Untuk Memenuhi Persyaratan

Dalam Menyelesaikan Strata-1 Ilmu Komputer

Oleh

Muhammad Rifqi

NIM 1711016310019

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU**

JUNI 2024

SKRIPSI

PERBANDINGAN EKSTRAKSI FITUR FASTTEXT DAN TF-IDF-ICF PADA
ANALISIS SENTIMEN DI TWITTER TERHADAP ISU PENUNDAAN PEMILU 2024
MENGUNAKAN METODE RANDOM FOREST

Oleh :

MUHAMMAD RIFQI

NIM. 1711016310019

Telah dipertahankan di depan Dosen Penguji pada

Susunan Dosen Penguji :

Pembimbing I



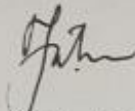
Triando Hamonangan Saragih S.Kom., M.Kom
NIP. 199308242019031012

Dosen Penguji I



Radityo Adi Nugroho, S.T, M.Kom.
NIP. 198212042008011006

Pembimbing II



Fatma Indriani, S.T, M.I.T, Ph.D
NIP. 198404202008122004

Dosen Penguji II



Friska Abadi, S.Kom, M.Kom
NIP. 198809132023211010



Bayu Baru,
Koordinator Program Studi Ilmu Komputer

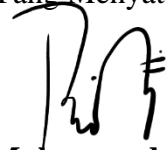
Bayu Baru, S.T., M.Kom
NIP. 19703252008121001

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 diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Banjarbaru, 24 Juni 2024

Yang Menyatakan,



Muhammad Rifqi
NIM. 1711016310019

ABSTRAK

PERBANDINGAN EKSTRAKSI FITUR *FASTTEXT* DAN *TF-IDF-ICF* PADA ANALISIS SENTIMEN DI *TWITTER* TERHADAP ISU PENUNDAAN PEMILU 2024 MENGGUNAKAN METODE *RANDOM FOREST*

(Oleh : Muhammad Rifqi; Pembimbing: Triando Hamonangan Saragih S.Kom., M.Kom. dan Fatma Indriani, S.T., M.I.T., Ph.D.; 2024; 65 halaman)

Isu penundaan pemilu pada tahun 2024 sempat menjadi banyak perbincangan dan menimbulkan banyak perdebatan bagi masyarakat, oleh karena itu dapat dilakukan analisis sentimen yang memberikan suatu penilaian atau opini pribadi. Analisis sentimen merupakan proses mengekstrak untuk mendapatkan pengetahuan atau suatu informasi pada suatu kalimat sentimen atau sebuah opini. Pada penelitian ini melakukan perbandingan antara metode ekstraksi fitur *FastText* dan *TF-IDF-ICF* dengan metode klasifikasi *Random Forest*. Berdasarkan hasil penelitian yang dilakukan *TF-IDF-ICF* mendapatkan hasil kinerja akurasi yang lebih baik dibandingkan dengan *FastText*, dimana akurasi terbaik didapatkan dengan *TF-IDF-ICF* yang dilakukan proses *stemming* dan *stopwords* yaitu sebesar 79,60%. Sedangkan untuk *FastText* yang dilakukan proses *stemming* dan *stopwords* mendapatkan akurasi 73,33%.

Kata kunci: Pemilu, Analisis Sentimen, Ekstraksi Fitur, *Random Forest*, *FastText*, *TF-IDF-ICF*.

ABSTRACT

COMPARISON OF FEATURE EXTRACTION USING FASTTEXT AND TF-IDF-ICF FOR SENTIMENT ANALYSIS ON TWITTER REGARDING THE ELECTION DELAY 2024 ISSUE USING THE RANDOM FOREST METHOD

(By: Muhammad Rifqi; Advisor: Triando Hamonangan Saragih, S.Kom., M.Kom. and Fatma Indriani, S.T., M.I.T., Ph.D.; 2024; 65 Pages)

The issue of election postponement in 2024 has sparked much discussion and debate among the public. Therefore, sentiment analysis can be conducted to provide an assessment or personal opinion. Sentiment analysis is the process of extracting knowledge or information from a sentence that expresses a sentiment or opinion. This study compares the feature extraction methods FastText and TF-IDF-ICF with the Random Forest classification method. Based on the research results, TF-IDF-ICF achieved better accuracy performance compared to FastText, with the best accuracy obtained using TF-IDF-ICF combined with stemming and stopword removal, reaching 79.60%. Meanwhile, FastText with stemming and stopword removal achieved an accuracy of 73.33%.

Keywords: *Election, Sentiment Analysis, Feature Extraction, Random Forest, FastText, TF-IDF-ICF.*

PRAKATA

Puji syukur penulis panjatkan kepada Allah SWT. karena atas berkat, rahmat, dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul **PERBANDINGAN EKSTRAKSI FITUR *FASTTEXT* DAN *TF-IDF-ICF* PADA ANALISIS SENTIMEN DI *TWITTER* TERHADAP ISU PENUNDAAN PEMILU 2024 MENGGUNAKAN METODE *RANDOM FOREST*** untuk memenuhi syarat dalam menyelesaikan Pendidikan program S-1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Tak lupa penulis panjatkan sholawat dan salam ke hadirat Rasulullah Muhammad SAW. beserta para sahabat, keluarga, dan pengikut beliau hingga *yaumul qiama*.

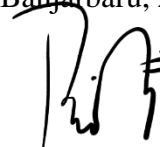
Pada lembar ini penulis ingin menyampaikan ucapan terimakasih kepada pihak-pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut:

1. Allah SWT. karena atas limpahan karunia-Nya dan bantuan-Nya lah penulis dapat menyelesaikan skripsi.
2. Keluarga terutama kedua orang tua dan saudara Saya yang selalu memberikan do'a, semangat, bantuan, dan dukungan dalam proses penyelesaian skripsi.
3. Bapak Triando Hamonangan Saragih, S.Kom., M.Kom. dan Ibu Fatma Indriani, S.T, M.I.T., Ph.D selaku Dosen Pembimbing Akademik.
4. Bapak Triando Hamonangan Saragih, S.Kom., M.Kom. selaku Dosen Pembimbing Utama dan Ibu Fatma Indriani, S.T, M.I.T., Ph.D. selaku Dosen Pembimbing Pendamping yang telah memberikan waktu dan bimbingannya kepada penulis dalam pengerjaan skripsi.
5. Bapak Irwan Budiman, S.T., M.Kom. selaku Koordinator Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
6. Seluruh Dosen dan Staf Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama berkuliah di Ilmu Komputer yang sangat bermanfaat.

7. Teman-teman keluarga Ilmu Komputer Angkatan 2017 yang memberikan bantuan dan dukungan yang senantiasa menjadi alat pacu penulis untuk selalu berusaha dan semangat untuk menyelesaikan skripsi.
8. Teman-teman yang telah bersedia membantu memecahkan kebingungan-kebingungan dan memberikan saran, dukungan, semangat, serta motivasi, selama proses penyelesaian skripsi.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan skripsi ini masih jauh dari kesempurnaan, maka penulis mengharapkan bantuan berupa kritik dan saran yang membangun dari semua pihak sangat diharapkan demi penyempurnaan dan mutu penulisan skripsi ini. Semoga tulisan ini dapat bermanfaat bagi semua pihak untuk dapat menambah pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, 24 Juni 2024



Muhammad Rifqi

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
PERNYATAAN	iii
ABSTRAK	iv
<i>ABSTRACT</i>	v
PRAKATA	vi
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
BAB I	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
BAB II TINJAUAN PUSTAKA	5
2.1 Kajian Terdahulu	5
2.2 Keaslian Penelitian	6
2.3 Landasan Teori	9
2.3.1 <i>Twitter</i>	9
2.3.2 Pemilihan Umum	9
2.3.3 <i>Text Preprocessing</i>	10
2.3.4 <i>TF-IDF-ICF</i>	12
2.3.5 <i>FastText</i>	12
2.3.6 <i>Random Forest</i>	15
2.3.7 <i>K-Fold Cross Validation</i>	16
2.3.8 <i>Confusion Matrix</i>	17

BAB III METODE PENELITIAN -----	20
3.1 Alat Penelitian-----	20
3.2 Bahan Penelitian -----	20
3.3 Variabel Penelitian -----	20
3.4 Prosedur Penelitian -----	20
BAB IV HASIL DAN PEMBAHASAN -----	25
4.1 Hasil -----	25
4.1.1 Pengumpulan Data -----	25
4.1.2 Pelabelan Data -----	27
4.1.3 Preprocessing Data-----	28
4.1.4 Ekstraksi Fitur -----	42
4.1.5 Pembagian Data -----	52
4.1.6 Klasifikasi <i>Random Forest</i> -----	53
4.1.6.1 <i>Random Forest</i> dengan Ekstraksi Fitur <i>TF-IDF-ICF</i> -----	53
4.1.6.2 <i>Random Forest</i> dengan Ekstraksi Fitur <i>FastText</i> -----	56
4.2 Pembahasan-----	59
BAB V PENUTUP -----	63
5.1 Kesimpulan -----	63
5.2 Saran-----	63
DAFTAR PUSTAKA -----	64
LAMPIRAN -----	67

DAFTAR TABEL

Tabel 1: Keaslian Penelitian.....	6
Tabel 2: Penelitian yang Diusulkan	8
Tabel 3: Proses <i>FastText Average</i>	15
Tabel 4: <i>FastText Average</i>	15
Tabel 5: Tabel <i>Confusion Matrix</i>	18
Tabel 6. <i>Confusion Matrix Multiclass</i>	18
Tabel 7: Komentar Positif	22
Tabel 8: Komentar Negatif.....	22
Tabel 9: Komentar Netral.....	22
Tabel 10: Contoh Data <i>Twitter / X</i>	25
Tabel 11: Jumlah data perkelas	27
Tabel 12: Pelabelan Data	27
Tabel 13: Contoh Proses <i>Cleansing Data</i>	29
Tabel 14: Data <i>Cleaning</i>	29
Tabel 15: Data <i>Case Folding</i>	31
Tabel 16: kamus <i>lemmatization</i>	33
Tabel 17: data <i>lemmatization</i>	33
Tabel 18: awalan(<i>Prefiks</i>)	35
Tabel 19: Akhiran (<i>Sufiks</i>)	35
Tabel 20: Sisipan (<i>Infiks</i>)	36
Tabel 21: Gabungan Imbuhan (<i>Konfiks</i>)	36
Tabel 22: Data <i>Stemming</i>	36
Tabel 23: Data <i>Stopwords</i>	38
Tabel 24: Data <i>tokenization</i> menggunakan <i>stemming</i> dan <i>stopwords</i>	39
Tabel 25: Data <i>tokenization</i> tanpa <i>stemming</i> dan <i>stopwords</i>	41
Tabel 26: contoh hasil ekstraksi <i>TF-IDF</i> dengan <i>stemming</i> dan <i>stopwords</i>	43
Tabel 27: contoh hasil ekstraksi <i>TF-IDF</i> tanpa <i>stemming</i> dan <i>stopwords</i>	43
Tabel 28: Contoh perhitungan ekstraksi fitur <i>TF-IDF</i> dengan <i>stemming</i> dan <i>stopwords</i>	44
Tabel 29: Contoh perhitungan untuk data dengan <i>stemming</i> dan <i>stopwords</i> pada isu penundaan pemilu	44
Tabel 30: Contoh perhitungan ekstraksi fitur <i>TF-IDF</i> tanpa <i>stemming</i> dan <i>stopwords</i>	45
Tabel 31: Contoh perhitungan untuk data tanpa <i>stemming</i> dan <i>stopwords</i> pada isu penundaan pemilu.....	45
Tabel 32: contoh perhitungan <i>ICF</i> dengan <i>stemming</i> dan <i>stopwords</i>	46
Tabel 33: contoh perhitungan <i>ICF</i> tanpa <i>stemming</i> dan <i>stopwords</i>	46

Tabel 34: Contoh perhitungan ekstraksi fitur <i>TF-IDF-ICF</i> dengan <i>stemming</i> dan <i>stopwords</i> pada data isu penundaan pemilu	46
Tabel 35: Contoh perhitungan ekstraksi fitur <i>TF-IDF-ICF</i> dengan <i>stemming</i> dan <i>stopwords</i> pada data isu penundaan pemilu	47
Tabel 36: contoh hasil ekstraksi <i>TF-IDF-ICF</i> dengan <i>stemming</i> dan <i>stopwords</i> .	48
Tabel 37: contoh hasil ekstraksi <i>TF-IDF-ICF</i> tanpa <i>stemming</i> dan <i>stopwords</i>	48
Tabel 38: Contoh model <i>FastText</i> dengan <i>stemming</i> dan <i>stopwords</i>	49
Tabel 39: Contoh model <i>FastText</i> tanpa <i>stemming</i> dan <i>stopwords</i>	49
Tabel 40: Contoh vector perkata dengan <i>stemming</i> dan <i>stopwords</i>	50
Tabel 41: Contoh vector perkata tanpa <i>stemming</i> dan <i>stopwords</i>	50
Tabel 42: Contoh word average dengan <i>stemming</i> dan <i>stopwords</i>	51
Tabel 43: Contoh word average tanpa <i>stemming</i> dan <i>stopwords</i>	52
Tabel 44: 10 <i>K-fold cross validation</i>	52
Tabel 45: Hasil Akurasi <i>Fold</i> 1-10 pada Ekstraksi Fitur <i>TF-IDF-ICF</i> dengan <i>Stemming</i> dan <i>Stopwords</i>	55
Tabel 46: Hasil Akurasi <i>Fold</i> 1-10 pada Ekstraksi Fitur <i>TF-IDF-ICF</i> tanpa <i>Stemming</i> dan <i>Stopwords</i>	56
Tabel 47: Hasil Akurasi <i>Fold</i> 1-10 pada Ekstraksi Fitur <i>FastText</i> dengan <i>Stemming</i> dan <i>Stopwords</i>	58
Tabel 48: Hasil Akurasi <i>Fold</i> 1-10 pada Ekstraksi Fitur <i>FastText</i> tanpa <i>Stemming</i> dan <i>Stopwords</i>	59
Tabel 49: Perbandingan Hasil Akurasi Nilai <i>Fold</i> 1-10.....	62

DAFTAR GAMBAR

Gambar 1. <i>Skip Gram</i>	13
Gambar 2. Visualisasi Metode <i>Random Forest</i>	16
Gambar 3. <i>K-Fold Cross Validation (K-10)</i>	17
Gambar 4. Prosedur Penelitian.....	21
Gambar 5. <i>Confusion Matrix TF-IDF-ICF</i> dengan <i>Stemming</i> dan <i>Stopwords</i>	54
Gambar 6. <i>Confusion Matrix TF-IDF-ICF</i> tanpa <i>Stemming</i> dan <i>Stopwords</i>	56
Gambar 7. <i>Confusion Matrix FastText</i> dengan <i>Stemming</i> dan <i>Stopwords</i>	58
Gambar 8. <i>Confusion Matrix FastText</i> tanpa <i>Stemming</i> dan <i>Stopwords</i>	59
Gambar 9. Grafik Perbandingan akurasi <i>TF-IDF-ICF</i> dan <i>FastText</i>	62

DAFTAR LAMPIRAN

Lampiran 1. Source code preprocessing data.....	68
Lampiran 2. Source code Ekstraksi Fitur <i>TF-IDF ICF</i>	70
Lampiran 3. Source code Ekstraksi Fitur <i>TF-IDF ICF</i> tanpa <i>stemming</i> dan <i>stopwords</i>	71
Lampiran 4. Source code Ekstraksi Fitur <i>FastText</i>	72
Lampiran 5. Source code <i>Random Forest</i> dengan <i>TF-IDF-ICF</i>	76
Lampiran 6. Source code <i>Random Forest</i> dengan <i>FastText</i>	78
Lampiran 7. Source code <i>Random Forest</i> dengan <i>TF-IDF-ICF</i> tanpa <i>stemming</i> dan <i>stopwords</i>	81
Lampiran 8. Source code <i>Random Forest</i> dengan <i>FastText</i> tanpa <i>stemming</i> dan <i>stopwords</i>	83
Lampiran 9. Kamus <i>Stopword</i>	85
Lampiran 10. Kamus <i>Lemmatization</i>	91