



**PERBANDINGAN EKSTRAKSI FITUR TF-IDF DAN *FASTTEXT* PADA
ANALISIS SENTIMEN DI TWITTER TERHADAP PPKM
MENGUNAKAN METODE KLASIFIKASI *NAÏVE BAYES***

Skripsi

**Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

Oleh

Irvan Dwi Halmanto

NIM 1711016110007

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

JULI 2023



**PERBANDINGAN EKSTRAKSI FITUR TF-IDF DAN *FASTTEXT* PADA
ANALISIS SENTIMEN DI TWITTER TERHADAP PPKM
MENGUNAKAN METODE KLASIFIKASI *NAÏVE BAYES***

Skripsi

**Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

Oleh

Irvan Dwi Halmanto

NIM 1711016110007

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

JULI 2023

SKRIPSI

**PERBANDINGAN EKSTRAKSI FITUR TF-IDF DAN *FASTTEXT* PADA
ANALISIS SENTIMEN DI TWITTER TERHADAP PPKM
MENGUNAKAN METODE KLASIFIKASI *NAÏVE BAYES***

Oleh :

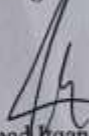
Irvan Dwi Halmanto

1711016110007

Telah dipertahankan di depan Dosen Penguji pada tanggal 13 Juli 2023

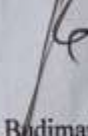
Susunan Dosen Penguji :

Pembimbing I



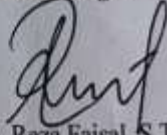
Muhammad Iqan Mazdadi, S.Kom., M.Kom.
NIP. 199006122019031013

Dosen Penguji I



Irvan Budiman, S.T., M.Kom.
NIP. 197703252008121001

Pembimbing II



M. Reza Faisal, S.T., M.T., Ph.D.
NIP. 19761220200812001

Dosen Penguji II



Muliadi, S.Kom., M.Cs.
NIP. 197804222010121002



Banjarbaru, 13 Juli 2023

Ketua Program Studi Ilmu Komputer

Irvan Budiman, S.T., M.Kom
NIP. 197703252008121001

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, 13 Juli 2023



Irvan Dwi Halmanto
NIM. 1711016110007

ABSTRAK

PERBANDINGAN EKSTRAKSI FITUR TF-IDF DAN *FASTTEXT* PADA ANALISIS SENTIMEN DI TWITTER TERHADAP PPKM MENGGUNAKAN METODE KLASIFIKASI *NAÏVE BAYES*

(Oleh: Irvan Dwi Halmanto; Pembimbing: Muhammad Itqan Mazdadi S.Kom., M.Kom dan M. Reza Faisal, S.T., M.T., Ph.D.; 2023; 61 halaman)

Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) sempat menjadi banyak perbincangan dan menimbulkan banyak pro kontra bagi masyarakat, oleh karena itu dapat dilakukan analisis sentimen yang memberikan suatu penilaian atau opini pribadi. Analisis sentimen merupakan proses mengekstrak pengetahuan atau informasi pada suatu kalimat sentimen atau opini. Pada penelitian ini melakukan perbandingan ekstraksi fitur TF-IDF dan *FastText* dengan metode klasifikasi *Naïve Bayes*. Berdasarkan hasil penelitian yang dilakukan *FastText* menghasilkan kinerja akurasi yang lebih baik dibandingkan TF-IDF, dimana akurasi terbaik dihasilkan oleh *FastText* tanpa *stopwords* yaitu sebesar 80,80%. Penelitian ini juga menjelaskan, pengaruh *stopwords* pada *FastText* yang dinilai kurang efektif karena dapat mengurangi informasi dari suatu kalimat.

Kata kunci: *Naïve Bayes*, Analisis Sentimen, Ekstraksi Fitur, TF-IDF, *FastText*.

ABSTRACT

COMPARISON OF TF-IDF AND FASTTEXT WEIGHTING MODELS IN SENTIMENT ANALYSIS ON TWITTER REGARDING PPKM USING NAÏVE BAYES CLASSIFICATION METHOD

(By: *Irvan Dwi Halmanto*; Supervisor: *Muhammad Itqan Mazdadi S.Kom., M.Kom and M. Reza Faisal, S.T., M.T., Ph.D.*; 2023; 61 pages)

The implementation of Restriction of Community Activities (PPKM) policy has sparked numerous discussions and generated mixed reactions among the public. Therefore, sentiment analysis can be conducted to provide personal evaluations or opinions. Sentiment analysis is a process of extracting knowledge or information from a sentence expressing sentiment or opinion. This study compares the feature extraction of TF-IDF and FastText using the Naïve Bayes classification method. Based on the research results, FastText outperformed TF-IDF in terms of accuracy, with the best accuracy achieved by FastText without stopwords at 80.80%. The study also explains the limited effectiveness of stopwords in FastText, as they can reduce the information contained in a sentence.

Keywords: *Naïve Bayes, Sentiment Analysis, Feature Extraction, TF-IDF, FastText.*

PRAKATA

Assalamualaikum Warahmatullahi Wabarakatuh.

Puji syukur penulis panjatkan ke hadirat Allah SWT karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “*Perbandingan Ekstraksi Fitur TF-IDF dan FastText pada Analisis Sentimen di Twitter terhadap PPKM menggunakan Metode Klasifikasi Naïve Bayes*” untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Tak lupa pula penulis panjatkan shalawat dan salam ke hadirat Rasulullah Muhammad SAW beserta para sahabat, keluarga, dan pengikut beliau hingga *yaumul qiyamah*.

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. Keluarga, terutama kedua orang tua tercinta untuk mama dan abah yang dari awal proses perkuliahan selalu memberikan bantuan, semangat, doa dan dukungan hingga sampai pada tahap proses penyelesaian skripsi ini.
2. Bapak Muhammad Itqan Mazdadi S.Kom, M.Kom selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran penyelesaian skripsi ini.
3. Bapak M. Reza Faisal, S.T., M.T., Ph.D. selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
4. Bapak Irwan Budiman, S.T., M.Kom selaku Koordinator Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
5. Seluruh Dosen dan Staff Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
6. Teman-teman yang telah bersedia membantu memecahkan kebingungan-kebingungan dan memberikan saran, dukungan, semangat, serta motivasi, selama proses penyelesaian skripsi.

7. Teman-teman keluarga Ilmu Komputer angkatan 2017 yang memberikan dukungan dan bantuan semasa kuliah dan selama proses penyelesaian skripsi.
8. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna, namun penulis mengharapkan bantuan berupa saran dan kritik yang membangun dari semua pihak demi kesempurnaan dan mutu penulisan skripsi ini.

Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, 13 Juli 2023



Irvan Dwi Halmanto

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
HALAMAN PERNYATAAN	iii
ABSTRAK	iv
<i>ABSTRACT</i>	v
PRAKATA.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	3
1.4 Tujuan	3
1.5 Manfaat Penelitian.....	3
BAB II TINJAUAN PUSTAKA.....	4
2.1 Kajian Terdahulu	4
2.2 Keaslian Penelitian	5
2.3 Landasan Teori	7
2.3.1 Analisis Sentimen	7
2.3.2 <i>Text Preprocessing</i>	7
2.3.3 TF-IDF	9

2.3.4	<i>FastText</i>	11
2.3.5	<i>K-fold Validation</i>	15
2.3.6	<i>Naïve Bayes</i>	17
2.3.7	<i>Confusion Matrix</i>	17
BAB III METODE PENELITIAN.....		19
3.1	Bahan dan Alat Penelitian	19
3.2	Variabel Penelitian.....	19
3.3	Prosedur Penelitian	20
BAB IV HASIL DAN PEMBAHASAN		23
4.1	Hasil.....	23
4.1.1	Pengumpulan Data	23
4.1.2	Pelabelan Data.....	24
4.1.3	<i>Preprocessing Data</i>	24
4.1.4	Ekstraksi Fitur	30
4.1.5	Klasifikasi	35
4.1.6	<i>Naïve Bayes</i> dengan Ekstraksi Fitur TF-IDF	36
4.1.7	<i>Naïve Bayes</i> dengan Ekstraksi Fitur <i>FastText</i>	45
4.1.8	Evaluasi.....	50
4.2	Pembahasan	53
BAB V PENUTUP.....		59
5.1	Kesimpulan.....	59
5.2	Saran	59
DAFTAR PUSTAKA		60
LAMPIRAN		

DAFTAR TABEL

Tabel 1: Tabel keaslian penelitian.....	5
Tabel 2: Penggunaan TF-IDF.....	11
Tabel 3: Proses <i>FastText Average</i>	15
Tabel 4: <i>Confusion Matrix</i>	18
Tabel 5: Tweet Positif	21
Tabel 6: Tweet Negatif.....	21
Tabel 7: Data Twitter	23
Tabel 8: Jumlah Data Perkelas	23
Tabel 9: Pelabelan Data	24
Tabel 10: Proses Data <i>Cleansing</i>	24
Tabel 11: Data <i>Cleansing</i>	25
Tabel 12: Data <i>Case Folding</i>	26
Tabel 13: Awalan (<i>Prefiks</i>).....	26
Tabel 14: Akhiran (<i>Sufiks</i>)	27
Tabel 15: Sisipan (<i>Infiks</i>)	27
Tabel 16: Gabungan Imbuhan (<i>Konfiks</i>).....	27
Tabel 17: Data <i>Stemming</i>	28
Tabel 18: Data <i>Stopwords</i>	28
Tabel 19: Data <i>Tokenization</i> dengan <i>Stopwords</i>	29
Tabel 20: Data <i>Tokenization</i> tanpa <i>stopwords</i>	29
Tabel 21: Hasil Ekstraksi Fitur TF-IDF dengan <i>stopwords</i>	31
Tabel 22: Perhitungan Ekstraksi Fitur TF-IDF dengan <i>stopwords</i>	32
Tabel 23: Hasil Ekstraksi Fitur TF-IDF tanpa <i>stopwords</i>	32
Tabel 24: Perhitungan Ekstraksi Fitur TF-IDF tanpa <i>stopwords</i>	33
Table 25: Model <i>FastText</i> dengan <i>stopwords</i>	34
Tabel 26: Model <i>FastText</i> tanpa <i>stopwords</i>	35
Tabel 27: 10 <i>K-fold cross validation</i>	36
Tabel 28: Probabilitas Prior TF-IDF dengan <i>Stopwords</i>	36
Tabel 29: Perhitungan rata-rata TF-IDF dengan <i>Stopwords</i>	37

Tabel 30: Perhitungan Standar Deviasi TF-IDF dengan <i>Stopwords</i>	37
Tabel 31: Nilai x pada data test TF-IDF dengan <i>Stopwords</i>	38
Tabel 32: Probabilitas likelihood menggunakan distribusi Gaussian TF-IDF dengan <i>Stopwords</i>	38
Tabel 33: Probabilitas Total Fitur TF-IDF dengan <i>Stopwords</i>	39
Tabel 34: Probabilitas Posterior TF-IDF dengan <i>Stopwords</i>	39
Tabel 35: Hasil kali probabilitas Posterior TF-IDF dengan <i>Stopwords</i>	40
Tabel 36: Probabilitas Prior TF-IDF tanpa <i>Stopwords</i>	41
Tabel 37: Perhitungan rata-rata TF-IDF tanpa <i>Stopwords</i>	41
Tabel 38: Perhitungan Standar Deviasi TF-IDF tanpa <i>Stopwords</i>	41
Tabel 39: Nilai x pada data test TF-IDF tanpa <i>Stopwords</i>	42
Tabel 40: Probabilitas likelihood menggunakan distribusi Gaussian TF-IDF tanpa <i>Stopwords</i>	42
Tabel 41: Probabilitas Total Fitur TF-IDF tanpa <i>Stopwords</i>	43
Tabel 42: Probabilitas Posterior TF-IDF tanpa <i>Stopwords</i>	44
Tabel 43: Hasil kali probabilitas Posterior TF-IDF tanpa <i>Stopwords</i>	45
Tabel 44: Probabilitas Prior <i>FastText</i> dengan <i>Stopwords</i>	45
Tabel 45: Perhitungan rata-rata <i>FastText</i> dengan <i>Stopwords</i>	46
Tabel 46: Perhitungan Standar Deviasi <i>FastText</i> dengan <i>Stopwords</i>	46
Tabel 47: Nilai x pada data test <i>FastText</i> dengan <i>Stopwords</i>	46
Tabel 48: Probabilitas likelihood menggunakan distribusi Gaussian <i>FastText</i> dengan <i>Stopwords</i>	46
Tabel 49: Probabilitas Total Fitur <i>FastText</i> dengan <i>Stopwords</i>	47
Tabel 50: Probabilitas Posterior <i>FastText</i> dengan <i>Stopwords</i>	47
Tabel 51: Probabilitas Prior <i>FastText</i> tanpa <i>Stopwords</i>	48
Tabel 52: Perhitungan rata-rata <i>FastText</i> tanpa <i>Stopwords</i>	48
Tabel 53: Perhitungan Standar Deviasi <i>FastText</i> tanpa <i>Stopwords</i>	48
Tabel 54: Nilai x pada data test <i>FastText</i> tanpa <i>Stopwords</i>	48
Tabel 55: Probabilitas likelihood menggunakan distribusi Gaussian <i>FastText</i> tanpa <i>Stopwords</i>	49
Tabel 56: Probabilitas Total Fitur <i>FastText</i> tanpa <i>Stopwords</i>	49
Tabel 57: Probabilitas Posterior <i>FastText</i> tanpa <i>Stopwords</i>	49

Tabel 58: <i>Naïve Bayes</i> dengan Ekstraksi Fitur TF-IDF dengan <i>stopwords</i>	50
Tabel 59: Perhitungan Akurasi Klasifikasi <i>Naïve Bayes</i> dengan Ekstraksi Fitur TF-IDF dengan <i>Stopwords</i>	50
Tabel 60: <i>Naïve Bayes</i> dengan Ekstraksi Fitur TF-IDF tanpa <i>Stopwords</i>	51
Tabel 61: Perhitungan Akurasi Klasifikasi <i>Naïve Bayes</i> Ekstraksi Fitur TF-IDF tanpa <i>Stopwords</i>	51
Tabel 62: Confusion Matrix <i>Naïve Bayes</i> dengan Ekstraksi Fitur <i>FastText</i> dengan <i>Stopwords</i>	52
Tabel 63: Perhitungan Akurasi Klasifikasi <i>Naïve Bayes</i> Ekstraksi Fitur <i>FastText</i> dengan <i>Stopwords</i>	52
Tabel 64: <i>Confusion Matrix Naïve Bayes</i> dengan Ekstraksi Fitur <i>FastText</i> tanpa <i>Stopwords</i>	52
Tabel 65: Perhitungan Akurasi Klasifikasi <i>Naïve Bayes</i> Ekstraksi Fitur <i>FastText</i> tanpa <i>Stopwords</i>	53
Tabel 66: Hasil Akurasi setiap <i>fold</i> TF-IDF dengan <i>stopwords</i> , <i>TF-IDF</i> tanpa <i>stopwords</i> , <i>FastText</i> dengan <i>stopwords</i> dan <i>FastText</i> tanpa <i>stopwords</i>	55
Tabel 67: Perbandingan Hasil Akurasi TF-IDF	56
Tabel 68: Perbandingan Hasil Akurasi <i>FastText</i>	57
Tabel 69: Perbandingan Hasil Akurasi TF-IDF dengan <i>stopwords</i> , TF-IDF tanpa <i>stopwords</i> , <i>FastText</i> dengan <i>stopwords</i> dan <i>FastText</i> tanpa <i>stopwords</i>	57

DAFTAR GAMBAR

Gambar 1. Alur TF-IDF	10
Gambar 2. Skip Gram	12
Gambar 3. Representasi 10 <i> folds cross validation</i>	16
Gambar 4. Alur Penelitian.....	20
Gambar 5. Skenario Ekstraksi Fitur	31
Gambar 6. Proses Ekstraksi <i>FastText</i> dengan <i>stopwords</i>	34
Gambar 7. Proses Ekstraksi <i>FastText</i> tanpa <i>stopwords</i>	35
Gambar 8. Grafik Perbandingan akurasi TF-IDF dengan <i>stopwords</i> dan TF-IDF tanpa <i>stopwords</i>	56
Gambar 9. Grafik Perbandingan akurasi <i>FastText</i> dengan <i>stopwords</i> dan <i>FastText</i> tanpa <i>stopwords</i>	57
Gambar 10. Grafik Perbandingan akurasi TF-IDF dengan <i>stopwords</i> , TF-IDF tanpa <i>stopwords</i> , <i>FastText</i> dengan <i>stopwords</i> dan <i>FastText</i> tanpa <i>stopwords</i>	58

DAFTAR LAMPIRAN

Lampiran 1 Kamus *Stopwords*

Lampiran 2 Kode Program