

**TUGAS AKHIR**

**PENGARUH EKSTRAKSI FITUR TERHADAP ANALISIS SENTIMENT  
PADA DATA REVIEW PELAYANAN INDIHOME BERBASIS *NAÏVE BAYES***



**SALSABILA MAZYA PERMATANING TYAS**

**1810651044**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS MUHAMMADIYAH JEMBER**

**2021**

**TUGAS AKHIR**

**PENGARUH EKSTRAKSI FITUR TERHADAP ANALISIS SENTIMENT  
PADA DATA REVIEW PELAYANAN INDIHOME BERBASIS *NAÏVE BAYES***

Disusun Untuk Melengkapi Dan Memenuhi Syarat Kelulusan

Guna Meraih Sarjana Komputer

Teknik Informatika Universitas Muhammadiyah Jember



**SALSABILA MAZYA PERMATANING TYAS**

**1810651044**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS MUHAMMADIYAH JEMBER**

**2021**

**HALAMAN PERSETUJUAN SIDANG TUGAS AKHIR**

**PENGARUH EKSTRAKSI FITUR TERHADAP ANALISIS SENTIMENT  
PADA DATA REVIEW PELAYANAN INDIHOME BERBASIS *NAÏVE BAYES***

Oleh:

**Salsabila Mazya Permataning Tyas**

**1810651044**

Telah disetujui bahwa Laporan Tugas Akhir ini untuk diajukan pada sidang Tugas Akhir sebagai salah satu syarat kelulusan dan mendapatkan gelar Sarjana Komputer (S.Kom)

di

Universitas Muhammadiyah Jember

**Disetujui oleh,**

Pembimbing I

**Dr. Bagus Setya Rintyarna, ST, M.Kom**  
NIDN. 0729017904

Pembimbing II

**Wiwik Suharsti, S.Kom, M.Kom**  
NIDN. 0006097801

**HALAMAN PENGESAHAN**  
**PENGARUH EKSTRAKSI FITUR TERHADAP ANALISIS SENTIMENT**  
**PADA DATA REVIEW PELAYANAN INDIHOME BERBASIS *NAÏVE BAYES***

Oleh:

**Salsabila Mazya Permataning Tyas**

**1810651044**

Telah mempertanggung jawabkan Laporan Tugas Akhirnya pada sidang Tugas Akhir tanggal 28 Desember 2021 sebagai salah satu syarat kelulusan dan mendapatkan gelar Sarjana Komputer

(S.Kom)

di

Universitas Muhammadiyah Jember

**Disetujui oleh,**

**Penguji I**

  
Miftahur Rahman, S.Kom, M.Kom  
NIDN. 0724039201

**Pembimbing I**

  
Dr. Bagus Setya Rintyarna, ST, M.Kom  
NIDN/ 0729017904

**Penguji II**

  
Agung Nilogiti, S.T., M.Kom  
NIDN. 0030037701

**Pembimbing II**

  
Wiwik Suharso, S.Kom, M.Kom  
NIDN. 0006097601


**Mengesahkan,**

**Dekan Fakultas Teknik**

  
Dr. Ir. Nanang Saiful Rizal, ST., MT., IPM  
NPK. 1978040510308366

**Mengetahui,**

**Ketua Program Studi Teknik Informatika**

  
Ari Eko Wardono, S.Kom, M.Kom  
NIP. 197502214 200501 1 001

## PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Salsabila Mazya Permataning Tyas

NIM : 1810651044

Program Studi : Teknik Informatika

Perguruan Tinggi : Universitas Muhammadiyah Jember

Menyatakan bahwa Tugas Akhir yang berjudul “PENGARUH EKSTRAKSI FITUR TERHADAP ANALISIS SENTIMENT PADA DATA *REVIEW* PELAYANAN INDIHOME BERBASIS *NAÏVE BAYES*” adalah ASLI dan BELUM PERNAH dibuat orang lain, kecuali yang diacu dalam Daftar Pustaka pada Tugas Akhir ini.

Demikian Surat Pernyataan ini dibuat dengan sebenar-benarnya dan apabila pernyataan ini tidak sesuai, penulis bersedia mendapat sanksi dari akademik.

Jember, 28 Desember 2021



Salsabila Mazya Permataning Tyas

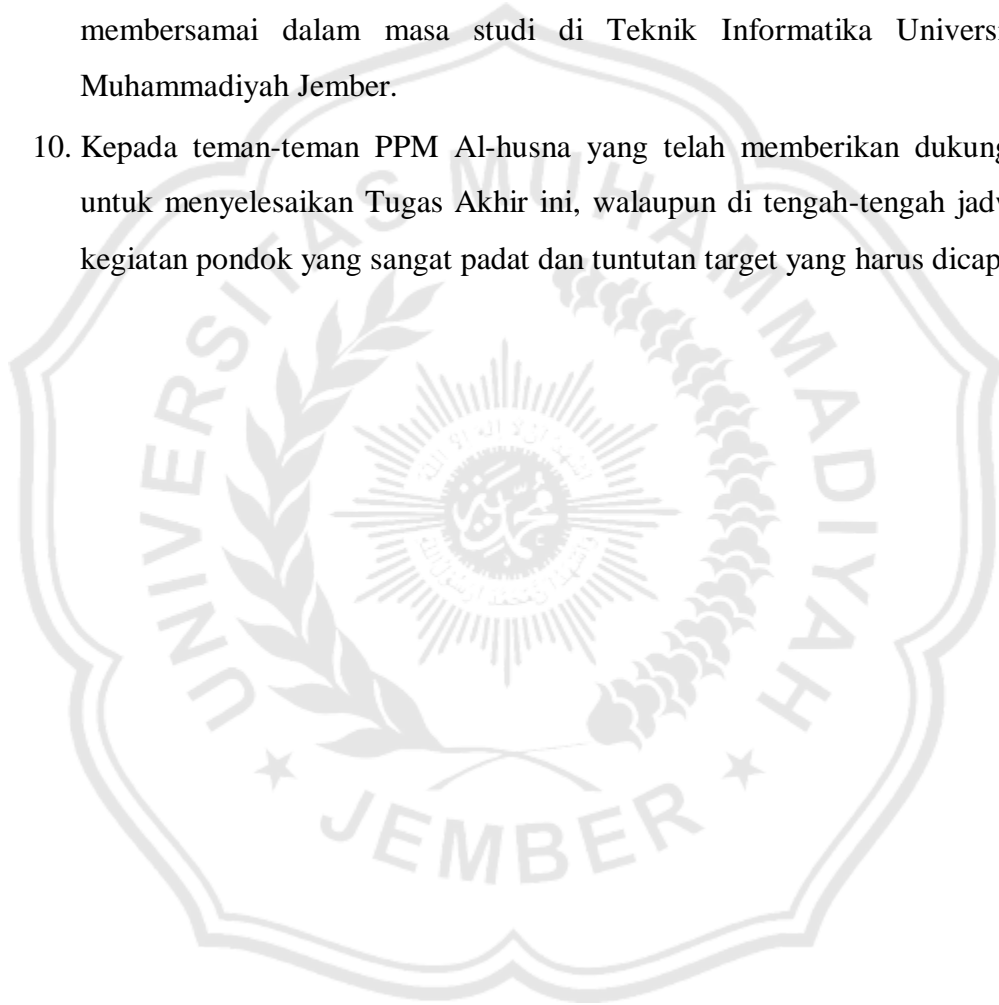
## KATA PENGANTAR

Alhamdulillah segala puji bagi Allah yang senantiasa memeberikan nikmat yang sangat luar biasa dalam hidup ini kepada penulis berupa nikmat untuk selalu bersujud dan meminta ampunan kepada-Nya. Serta berkat rahmat-Nya penulis diberikan kemudahan dalam menyelesaikan studi di kampus Universitas Muhammadiyah Jember.

Atas segala upaya, bimbingan, dan arahan dari semua pihak, penulis mengucapkan banyak terimakasih. Dalam kesempatan ini penulis mengucapkan terima kasih sebanyak-banyaknya kepada:

1. Allah SWT atas segala rahmat, nikmat, dan hidayah-Nya, terima kasih atas kehendak-Mu telah memberikan gelar Sarjana Komputer.
2. Bapak Ari Eko Wardoyo, M.Kom. selaku Ketua Program Studi Teknik Informatika Universitas Muhammadiyah Jember.
3. Bapak Dr. Bagus Setya Rintyarna, ST, M.Kom. selaku dosen pembimbing I dan Bapak Wiwik Suharso, S.Kom, M.Kom. selaku dosen pembimbing II yang telah meluangkan waktu untuk memberi arahan dan membimbing penulis dalam menyelesaikan tugas akhir ini.
4. Bapak Miftahur Rahman, S.Kom., M.Kom. selaku dosen penguji I dan Bapak Agung Nilogiri, S.T., M.Kom. selaku dosen penguji II yang telah memberikan saran dan masukan dalam penyelesaian tugas akhir ini.
5. Kepada Bapak dan Ibu Dosen Program Studi Teknik Informatika Universitas Muhammadiyah Jember yang telah memberikan ilmunya kepada penulis.
6. Kepada orang tua, adik, serta keluarga yang selalu mendoakan, memberikan kasih sayang, nasihat, dan dukungannya sehingga penulis dapat menyelesaikan tugas akhir ini.
7. Kepada M. Derryl Qinanda yang selalu memberikan dukungan dan semangat, serta sabar mendengarkan setiap keluh kesah dalam menempuh masa studi di Teknik Informatika Universitas Muhammadiyah Jember.

8. Kepada Amalia Nur Azizah teman seperjuangan yang selalu menguatkan satu sama lain dalam mencapai target kuliah maupun target pondok.
9. Kepada Ayu, Indah, Tedi, Ajeng, Veren, dan Amada teman seperjuangan belajar bersama mulai dari awal kuliah sampai akhir. Serta untuk teman-teman mahasiswa Teknik Informatika Angkatan 2018 yang sudah kebersamai dalam masa studi di Teknik Informatika Universitas Muhammadiyah Jember.
10. Kepada teman-teman PPM Al-husna yang telah memberikan dukungan untuk menyelesaikan Tugas Akhir ini, walaupun di tengah-tengah jadwal kegiatan pondok yang sangat padat dan tuntutan target yang harus dicapai.



## **Motto**

*“If you’re insecure, guess what? The rest of the world is, too. Don’t overestimate the competition and underestimate yourself. You are better than you think”*

Timothy Ferriss





## DAFTAR ISI

DAFTAR ISI.....	i
DAFTAR GAMBAR.....	iv
DAFTAR TABEL .....	v
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
<b>1.1 Latar Belakang .....</b>	<b>1</b>
<b>1.2 Rumusan Masalah .....</b>	<b>4</b>
<b>1.3 Tujuan.....</b>	<b>4</b>
<b>1.4 Manfaat.....</b>	<b>4</b>
<b>1.5 Batasan Masalah.....</b>	<b>4</b>
<b>BAB II TINJAUAN PUSTAKA .....</b>	<b>6</b>
<b>2.1 <i>Text Mining</i> .....</b>	<b>6</b>
2.1.1 <i>Text Preprocessing</i> .....	6
a. <i>Case Folding</i> .....	7
b. <i>Data Cleansing</i> .....	7
c. <i>Tokenization</i> .....	7
d. <i>Stopword Removal</i> .....	7
e. <i>Stemming</i> .....	7
<b>2.2 Analisis Sentimen .....</b>	<b>7</b>
<b>2.3 Twitter .....</b>	<b>8</b>
<b>2.4 <i>Undersampling</i> .....</b>	<b>8</b>
<b>2.5 <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i> .....</b>	<b>8</b>
<b>2.6 <i>Word2Vec</i> .....</b>	<b>9</b>
<b>2.7 <i>Naïve Bayes Classification</i>.....</b>	<b>11</b>
<b>2.8 <i>K-Fold Cross Validation</i> .....</b>	<b>12</b>
<b>2.9 <i>Python</i>.....</b>	<b>13</b>
<b>2.10 Penelitian Terdahulu .....</b>	<b>13</b>

2.10.1	Word2Vec Convolutional Neural Networks For Classification Of News Articles And Tweets (Beakcheol Jang, Inhwan Kim, Jong Wook Kim, 2019)	14
2.10.2	Enhance The power of Sentiment Analysis (Zhang & Desouza, 2014)	14
2.10.3	Sentiment Analysis of Government Policy on Corona Case Using Naïve Bayes Algorithm (Auliya Rahman Isnain, Nurman Satya, Debby Alita, 2021)	14
2.10.4	Word2Vec Model for Sentiment Analysis of Product Reviews in Indonesian Language (M. Ali Fauzi, 2019)	15
<b>BAB III METODOLOGI PENELITIAN</b>		<b>16</b>
<b>3.1</b>	<b>Tahapan Penelitian</b>	<b>16</b>
3.1.1	Pengumpulan Data	17
3.1.2	Pelabelan Data	17
3.1.3	<i>Balancing Data</i>	18
3.1.4	<i>Preprocessing</i>	18
3.1.5	Ekstraksi Fitur	24
3.1.6	<i>Cross Fold Validation</i>	32
3.1.7	Implementasi Metode Klasifikasi	36
<b>BAB IV IMPLEMENTASI DAN HASIL</b>		<b>39</b>
<b>4.1</b>	<b>Implementasi Sentiment Analysis</b>	<b>40</b>
4.1.1	<i>Crawling</i> Data Twitter	40
4.1.2	Labelisasi	41
4.1.3	<i>Balancing Data</i>	42
4.1.4	<i>Preprocessing Text</i>	43
4.1.5	Ekstraksi Fitur Pembobotan Kata	46
4.1.6	<i>K-Fold Cross Validation</i>	49
<b>4.2</b>	<b>Hasil Analisis</b>	<b>59</b>
<b>BAB V KESIMPULAN DAN SARAN</b>		<b>61</b>



## DAFTAR GAMBAR

Gambar 2.1 Arsitektur Word2Vec CBOW dan Skip-gram.....	10
Gambar 2.2 <i>K-fold cross vold validation</i> .....	13
Gambar 3.1 Flowchart Penelitian .....	10
Gambar 4.1 API Twitter pada Portal Twitter Developer .....	40
Gambar 4.2 Dataset berformat csv .....	41
Gambar 4.3 Data yang sudah diberi label negatif .....	41
Gambar 4.4 Data yang sudah diberi label positif .....	42
Gambar 4.5 Data yang sudah seimbang .....	42
Gambar 4.6 Grafik Tingkat Akurasi TF-IDF K = 2 .....	50
Gambar 4.7 Grafik Tingkat Akurasi TF-IDF K = 5 .....	51
Gambar 4.8 Grafik Tingkat Akurasi TF-IDF K = 7 .....	52
Gambar 4.9 Grafik Tingkat Akurasi TF-IDF K = 10 .....	53
Gambar 4.10 Grafik Tingkat Akurasi Word2Vec K = 2.....	54
Gambar 4.11 Grafik Tingkat Akurasi Word2Vec K = 5.....	55
Gambar 4.12 Grafik Tingkat Akurasi Word2Vec K = 7.....	56
Gambar 4.13 Grafik Tingkat Akurasi Word2Vec K = 10.....	57

## DAFTAR TABEL

Tabel 3. 1 Contoh potongan data <i>tweet</i> .....	18
Tabel 3. 2 Contoh pelabelan data .....	19
Tabel 3. 3 Tahap <i>Cleansing</i> .....	20
Tabel 3. 4 Tahap <i>Case Folding</i> .....	21
Tabel 3. 5 Tahap <i>Stopword Removal</i> .....	23
Tabel 3. 6 Tahap <i>Stemming</i> .....	24
Tabel 3. 7 Tahap <i>Tokenize</i> .....	25
Tabel 3. 8 Perhitungan jumlah <i>term</i> dalam setiap dokumen .....	27
Tabel 3. 9 Perhitungan TF.....	27
Tabel 3. 10 Perhitungan TF-IDF .....	28
Tabel 3. 11 Dokumen Teks .....	29
Tabel 3. 12 <i>One hot vector encoding</i> .....	30
Tabel 3. 13 Hasil <i>y_pred</i> .....	30
Tabel 3. 14 Contoh data bobot TF-IDF.....	32
Tabel 3. 15 Frekuensi jumlah kelas .....	39
Tabel 3. 16 <i>Mean</i> atau nilai rata-rata .....	40
Tabel 3. 17 Standar deviasi .....	37
Tabel 3. 18 Gaussian setiap dokumen.....	62
Tabel 3. 19 Hasil prediksi sentimen.....	62
Tabel 4. 1 Hasil Labelisasi .....	44
Tabel 4. 2 Data Hasil <i>Balancing</i> .....	45
Tabel 4. 3 Potongan Tahap <i>Preprocessing Text</i> .....	45
Tabel 4. 4 Potongan Pembobotan Kata TF-IDF.....	48
Tabel 4. 5 Potongan Pembobotan Kata <i>Word2Vec</i> .....	51
Tabel 4. 6 Perbandingan Perhitungan Akurasi <i>k-fold cross validation</i> TF-IDF dan hasil pengujian.....	60
Tabel 4. 7 Perbandingan Perhitungan Akurasi <i>k-fold cross validation</i> <i>Word2Vec</i> dan hasil pengujian .....	61
Tabel 4. 8 Hasil Perhitungan <i>k-fold cross validation</i> TF-IDF dan <i>Word2Vec</i> .....	62