

**SKRIPSI**

**IMPLEMENTASI ALGORITMA NAÏVE BAYES UNTUK  
MENGANALISIS SENTIMEN OPINI PENGGUNA TWITTER  
TERHADAP KOREAN POP (KPOP)**

***IMPLEMENTATION OF THE NAÏVE BAYES ALGORITHM TO  
ANALYZE THE SENTIMENT OF TWITTER USERS' OPINION  
ABOUT KOREAN POP (KPOP)***



Disusun oleh

**TIARA REISAONA  
19101126**

**PROGRAM STUDI S1 TEKNIK TELEKOMUNIKASI  
FAKULTAS TEKNIK TELEKOMUNIKASI DAN ELEKTRO  
INSTITUT TEKNOLOGI TELKOM PURWOKERTO**

**2023**

**SKRIPSI**

**IMPLEMENTASI ALGORITMA NAÏVE BAYES UNTUK  
MENGANALISIS SENTIMEN OPINI PENGGUNA TWITTER  
TERHADAP KOREAN POP (KPOP)**

***IMPLEMENTATION OF THE NAÏVE BAYES ALGORITHM TO  
ANALYZE THE SENTIMENT OF TWITTER USERS' OPINION  
ABOUT KOREAN POP (KPOP)***



Disusun oleh

**TIARA REISAONA  
19101126**

**PROGRAM STUDI S1 TEKNIK TELEKOMUNIKASI  
FAKULTAS TEKNIK TELEKOMUNIKASI DAN ELEKTRO  
INSTITUT TEKNOLOGI TELKOM PURWOKERTO**

**2023**

**IMPLEMENTASI ALGORITMA NAÏVE BAYES UNTUK  
MENGANALISIS SENTIMEN OPINI PENGGUNA TWITTER  
TERHADAP KOREAN POP (KPOP)**

***IMPLEMENTATION OF THE NAÏVE BAYES ALGORITHM TO  
ANALYZE THE SENTIMENT OF TWITTER USERS' OPINION  
ABOUT KOREAN POP (KPOP)***

**Skripsi ini digunakan sebagai salah satu syarat untuk memperoleh  
Gelar Sarjana Teknik (S.T.)  
Di Institut Teknologi Telkom Purwokerto  
2023**

Disusun oleh

**TIARA REISAONA  
19101126**

**DOSEN PEMBIMBING**

**Rahmat Widadi, S.Pd., M.Eng.  
Zein Hanni Pradana, S.T., M.T.**

**PROGRAM STUDI S1 TEKNIK TELEKOMUNIKASI  
FAKULTAS TEKNIK TELEKOMUNIKASI DAN ELEKTRO  
INSTITUT TEKNOLOGI TELKOM PURWOKERTO**

**2023**

**HALAMAN PENGESAHAN**  
**IMPLEMENTASI ALGORITMA NAÏVE BAYES UNTUK**  
**MENGANALISIS SENTIMEN OPINI PENGGUNA TWITTER**  
**TERHADAP KOREAN POP (KPOP)**

**IMPLEMENTATION OF THE NAÏVE BAYES ALGORITHM TO**  
**ANALYZE THE SENTIMENT OF TWITTER USERS' OPINION ABOUT**  
**KOREAN POP (KPOP)**

Disusun oleh  
**TIARA REISAONA**  
19101126

Telah dipertanggungjawabkan di hadapan Tim Penguji pada tanggal  
8 Agustus 2023

**Susunan Tim Penguji**

Pembimbing Utama : Rahmat Widadi, S.Pd., M.Eng.  
NIDN. 0631039201

(  )

Pembimbing Pendamping : Zein Hanni Pradana, S.T., M.T.  
NIDN. 0604039001

(  )

Penguji 1 : Shinta Romadhona, S.T., M.T.  
NIDN. 0611068402

(  )

Penguji 2 : Solichah Larasati, S.T., M.T.  
NIDN. 0617069301

(  )

**Mengetahui,**

Ketua Program Studi S1 Teknik Telekomunikasi  
Institut Teknologi Telkom Purwokerto

  
Prasetyo Yudiantoro, S.T., M.T.  
NIDN. 0620052201

## HALAMAN PERNYATAAN ORISINALITAS

Dengan ini saya **TIARA REISAONA**, menyatakan bahwa skripsi dengan judul “**IMPLEMENTASI ALGORITMA NAÏVE BAYES UNTUK MENGANALISIS SENTIMEN OPINI PENGGUNA TWITTER TERHADAP KOREAN POP (KPOP)**” adalah benar-benar karya saya sendiri. Saya tidak melakukan penjiplakan kecuali melalui pengutipan sesuai dengan etika keilmuan yang berlaku. Saya bersedia menanggung risiko ataupun sanksi yang dijatuhkan kepada saya apabila ditemukan pelanggaran terhadap etika keilmuan dalam skripsi saya ini.

Purwokerto, 1 Juli 2023



(Tiara Reisaona)

## PRAKATA

Puji dan syukur penulis panjatkan kehadirat Allah SWT yang telah melimpahkan kasih dan sayang-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “**Implementasi Algoritma Naïve Bayes Untuk Menganalisis Sentimen Opini Pengguna Twitter Terhadap Korean Pop (Kpop)**”.

Maksud dari penyusunan skripsi ini adalah untuk memenuhi salah satu syarat dalam menempuh ujian sarjana Teknik Telekomunikasi pada Fakultas Teknik Telekomunikasi dan Elektro Institut Teknologi Telkom Purwokerto.

Dalam penyusunan skripsi ini, banyak pihak yang sangat membantu penulis dalam berbagai hal. Oleh karena itu, penulis sampaikan rasa terima kasih yang sedalam-dalamnya kepada:

1. Kedua Orangtua penulis dan keluarga tercinta atas segala bantuan, dukungan, motivasi, serta doa yang diberikan.
2. Bapak Rahmat Widadi, S.Pd., M.Eng. selaku pembimbing I yang telah memberikan arahan dan bimbingan dalam penyusunan dan penulisan laporan ini.
3. Bapak Zein Hanni Pradana, S.T., M.T. selaku pembimbing II yang telah memberikan arahan dan bimbingan dalam penyusunan dan penulisan laporan ini.
4. Ibu Dr. Tenia Wahyuningrum, S.Kom., M.T, selaku Rektor Institut Teknologi Telkom Purwokerto
5. Bapak Prasetyo Yuliantoro, S.T.,M.T. ketua Program Studi S1 Teknik Telekomunikasi.
6. Ibu Dr. Anggun Fitria Isnawati selaku Dekan Fakultas Teknik Telekomunikasi dan Elektro.
7. Bapak M Lukman Leksono, S.Pd.,M.Pd. selaku Dosen Bahasa Indonesia yang telah membantu dalam pelabelan data sentimen.
8. Magdalena, Amalia, Vania, Intan, serta teman-teman seperjuangan angkatan 2019 yang telah memberikan semangat serta bantuan sehingga penulis dapat menyelesaikan laporan ini.

9. EXO dan NCT, *you safe my life*. Chanyeol, Jaemin *i did it*. Sampai bertemu di hari yang baik.

Akhir kata penulis menyadari bahwa tidak ada yang sempurna, untuk itu, sangat diperlukan kritik dan saran yang membangun penulis. Semoga laporan Tugas akhir ini dapat bermanfaat tidak hanya untuk penulis, namun juga untuk orang lain

Purwokerto, 8 Agustus 2023

(Tiara Reisaona)

## DAFTAR ISI

<b>HALAMAN JUDUL</b> .....	<b>I</b>
<b>HALAMAN PENGESAHAN</b> .....	<b>II</b>
<b>HALAMAN PERNYATAAN ORISINALITAS</b> .....	<b>III</b>
<b>PRAKATA</b> .....	<b>IV</b>
<b>ABSTRAK</b> .....	<b>VI</b>
<b>ABSTRACT</b> .....	<b>VII</b>
<b>DAFTAR ISI</b> .....	<b>VIII</b>
<b>DAFTAR GAMBAR</b> .....	<b>XI</b>
<b>DAFTAR TABEL</b> .....	<b>XII</b>
<b>BAB 1 PENDAHULUAN</b> .....	<b>1</b>
1.1. LATAR BELAKANG .....	1
1.2 RUMUSAN MASALAH .....	3
1.3 BATASAN MASALAH .....	3
1.4 TUJUAN .....	4
1.5 MANFAAT .....	4
1.6 SISTEMATIKA PENULISAN .....	4
<b>BAB 2 DASAR TEORI</b> .....	<b>6</b>
2.1 KAJIAN PUSTAKA .....	6
2.2 DASAR TEORI .....	9
2.2.1 KPOP .....	9
2.2.2 <i>TWITTER</i> .....	10
2.2.3 ANALISIS SENTIMEN .....	11
2.2.4 <i>DATA MINING</i> .....	12
2.2.5 <i>PRE-PROCESSING</i> .....	12
2.2.6 PELABELAN DATA .....	13
2.2.7 <i>TERM FREQUENCY - INVERSE DOCUMENT FREQUENCY (TF-IDF)</i> .....	13
2.2.8 KLASIFIKASI .....	14
2.2.9 NAÏVE BAYES .....	14
2.2.10 EVALUASI .....	16



<b>BAB 3 METODE PENELITIAN.....</b>	<b>19</b>
3.1 ALAT DAN BAHAN YANG DIGUNAKAN.....	19
3.1.1 PERANGKAT KERAS ( <i>HARDWARE</i> ).....	19
3.1.2 PERANGKAT LUNAK ( <i>SOFTWARE</i> ).....	19
3.1.3 BAHAN.....	19
3.2 ALUR PENELITIAN.....	20
3.2.1 STUDI LITERATUR.....	21
3.2.2 PENGAMBILAN DATA.....	21
3.2.3 <i>PREPROCESSING DATA</i> .....	23
3.2.4 PELABELAN DATA.....	30
3.2.5 PEMBOBOTAN <i>TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF)</i> .....	30
3.2.6 PENGOLAHAN DATA DENGAN MODEL PREDIKSI ALGORITMA NAÏVE BAYES.....	35
3.2.7 EVALUASI.....	36
<b>BAB 4 HASIL DAN PEMBAHASAN.....</b>	<b>37</b>
4.1 HASIL PENGUMPULAN DATA.....	37
4.2 HASIL <i>PREPROCESSING</i> .....	38
4.2.1 <i>CLEANING DATA</i> .....	38
4.2.2 <i>CASE FOLDING</i> .....	45
4.2.3 <i>TOKENIZE</i> .....	46
4.2.4 <i>NORMALIZATION</i> .....	47
4.2.5 <i>STOPWORD REMOVAL</i> .....	49
4.2.6 <i>STEMMING</i> .....	51
4.3 HASIL PELABELAN DATA.....	52
4.4 PEMBOBOTAN <i>TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF)</i> .....	55
4.5 PENGUJIAN KLASIFIKASI NAÏVE BAYES <i>CLASSIFIER</i> .....	56
4.6 HASIL EVALUASI KLASIFIKASI.....	57
4.6.1 SKENARIO 1 ( <i>TFIDFVECTORIZER()</i> ).....	57
4.6.2 SKENARIO 2 ( <i>TFIDFVECTORIZER(MIN_DF=2)</i> ).....	59
4.6.3 SKENARIO 3 ( <i>TFIDFVECTORIZER(MIN_DF=3)</i> ).....	62

4.6.4	SKENARIO 4 (TFIDFVECTORIZER(MIN_DF=4)) .....	64
4.6.5	SKENARIO 5 (TFIDFVECTORIZER(MIN_DF=5)) .....	67
4.7	HASIL ANALISA .....	69
<b>BAB 5 KESIMPULAN DAN SARAN .....</b>		<b>71</b>
5.1	KESIMPULAN.....	71
5.2	SARAN.....	71
<b>DAFTAR PUSTAKA .....</b>		<b>73</b>
<b>LAMPIRAN.....</b>		<b>77</b>

## DAFTAR GAMBAR

Gambar 3. 1 Diagram Alur Penelitian.....	20
Gambar 3. 2 Rangkaian Operator pengambilan data pada <i>RapidMiner</i> . ....	21
Gambar 3. 3 Rangkaian Operator Pembersihan <i>Attribute</i> .....	22
Gambar 3. 4 Diagram Preprocessing. ....	23
Gambar 3. 5 Rangkaian Operator Pembersihan Retweet.....	23
Gambar 3. 6 Rangkaian Operator Pembersihan Link. ....	24
Gambar 3. 7 Rangkaian Operator Pembersihan Hastag.....	24
Gambar 3. 8 Rangkaian Operator Pembersihan Mention. ....	25
Gambar 3. 9 Rangkaian Operator Pembersihan Simbol. ....	25
Gambar 3. 10 Rangkaian Operator Pembersihan Spasi Dan Enter tidak perlu.....	26
Gambar 3. 11 Rangkaian Operator Pembersihan Tweet Kosong dan Duplikat....	26
Gambar 4. 1 Tweet Duplikat.....	45
Gambar 4. 2 Persentase data <i>Labeling</i> .....	53
Gambar 4. 3 <i>Wordcloud</i> Sentimen.....	54
Gambar 4. 4 Visualisasi <i>Confusion Matrix</i> Skenario 1.....	57
Gambar 4. 5 Hasil Akurasi Klasifikasi dengan <i>TfidfVectorizer()</i> . ....	58
Gambar 4. 6 Visualisasi <i>Confusion Matrix</i> Skenario 2.....	60
Gambar 4. 7 Hasil Akurasi Klasifikasi dengan <i>TfidfVectorizer(min_df=2)</i> . ....	60
Gambar 4. 8 Visualisasi <i>Confusion Matrix</i> Skenario 3.....	62
Gambar 4. 9 Hasil Akurasi Klasifikasi dengan <i>TfidfVectorizer(min_df=3)</i> .....	63
Gambar 4. 10 Visualisasi <i>Confusion Matrix</i> Skenario 4.....	65
Gambar 4. 11 Hasil Akurasi Klasifikasi dengan <i>TfidfVectorizer(min_df=4)</i> . ....	65
Gambar 4. 12 Visualisasi <i>Confusion Matrix</i> Skenario 5.....	67
Gambar 4. 13 Hasil Akurasi Klasifikasi dengan <i>TfidfVectorizer(min_df=5)</i> . ....	68

## DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu .....	7
Tabel 2. 2 <i>Multiclass Confusion Matrix</i> .....	16
Tabel 3. 1 Perangkat Keras .....	19
Tabel 3. 2 Perangkat lunak.....	19
Tabel 3. 3 Sampel Hasil Pengambilan Data Tweet.....	22
Tabel 3. 4 <i>Cleaning Data</i> .....	26
Tabel 3. 5 Contoh Hasil <i>Case Folding</i> .....	27
Tabel 3. 6 Contoh Hasil <i>Tokenize</i> .....	27
Tabel 3. 7 Contoh Hasil <i>Normalization</i> .....	28
Tabel 3. 8 Contoh Hasil <i>Stopword Removal</i> .....	29
Tabel 3. 9 Contoh Hasil <i>Stemming</i> .....	30
Tabel 3. 10 Contoh Dokumen .....	30
Tabel 3. 11 Menghitung TF .....	31
Tabel 3. 12 Menghitung DF .....	31
Tabel 3. 13 Menghitung IDF.....	32
Tabel 3. 14 Menghitung TF-IDF.....	33
Tabel 3. 15 Hasil pembagian data uji dan data latih .....	36
Tabel 4. 1 Hasil Pengumpulan Data Tweet.....	37
Tabel 4. 2 Hasil <i>Cleaning Retweet</i> .....	38
Tabel 4. 3 Hasil <i>Cleaning Link</i> .....	40
Tabel 4. 4 Hasil <i>Cleaning Hastag</i> .....	41
Tabel 4. 5 Hasil <i>Cleaning Mention</i> .....	42
Tabel 4. 6 Hasil <i>Cleaning Simbol</i> .....	43
Tabel 4. 7 Hasil <i>Cleaning Hastag</i> .....	44
Tabel 4. 8 Hasil Proses <i>Case Folding</i> .....	45
Tabel 4. 9 Hasil Proses <i>Tokenize</i> .....	46
Tabel 4. 10 Hasil Proses <i>Normalization</i> .....	47
Tabel 4. 11 Hasil Proses Stop Removal.....	49
Tabel 4. 12 Hasil Proses <i>Stemming</i> .....	51
Tabel 4. 13 Hasil Pelabelan Data .....	52

Tabel 4. 14 Hasil Pembobotan TF-IDF.....	55
Tabel 4. 15 Hasil Pembagian Data <i>Training</i> Dan <i>Testing</i> .....	56
Tabel 4. 16 Hasil <i>Confusion Matrix</i> .....	58
Tabel 4. 17 Hasil <i>Confusion Matrix</i> .....	61
Tabel 4. 18 Hasil <i>Confusion Matrix</i> .....	63
Tabel 4. 19 Hasil <i>Confusion Matrix</i> .....	66
Tabel 4. 20 Hasil <i>Confusion Matrix</i> .....	68
Tabel 4. 21 Rangkuman evaluasi performa model naïve bayes .....	70