



**ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN APLIKASI
MYTELKOMSEL MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN METODE
TF-IDF**

SKRIPSI

**FISYA ALIFIA FAWWAZI SIREGAR
NIM. 2010512060**

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA
2024**



SKRIPSI

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh
Gelar Sarjana Ilmu Komputer**

**FISYA ALIFIA FAWWAZI SIREGAR
NIM. 2010512060**

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA
2024**

PERNYATAAN ORISINALITAS

PERNYATAAN ORISINALITAS

Skripsi ini merupakan hasil karya sendiri serta semua sumber referensi yang dikutip maupun yang dirujuk telah saya nyatakan benar.

Nama : Fisyah Alifia Fawwazi Siregar

NIM : 2010512060

Tanggal : 15 Juli 2024

Apabila pada kemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan berlaku.

Jakarta, 15 Juli 2024
Yang Menyatakan,



Fisyah Alifia Fawwazi Siregar

PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Fisya Alifia Fawwazi Siregar
NIM : 2010512060
Fakultas : Ilmu Komputer
Program Studi : SI Sistem Informasi

Demi Pembangunan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Eksklusif (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

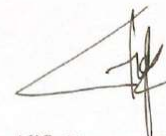
ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN APLIKASI MYTELKOMSEL MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN METODE TF-IDF

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini, Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis/peneliti dan pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sesungguhnya.

Dibuat di: Jakarta
Pada Tanggal: 15 Juli 2024

Yang Menyatakan,



Fisya Alifia Fawwazi Siregar

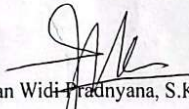
LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

Tugas akhir ini diajukan oleh:

Nama : Fisyah Alifia Fawwazi Siregar
NIM : 2010512060
Program Studi : SI Sistem Informasi
Judul : ANALISIS SENTIMEN BERBASIS ASPEK PADA
APLIKASI MYTELKOMSEL MENGGUNAKAN ALGORITMA
NAÏVE BAYES DAN METODE TF-IDF

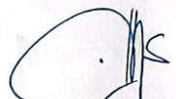
Telah berhasil dipertahankan dihadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



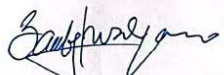
I Wayan Widi Pradnyana, S.Kom., MTI
Penguji 1



Sarika, S.Kom., M.Kom
Penguji 2




Ika Nurlaili Isnainiyah, S.Kom., M.Sc
Pembimbing 1



Bambang Tri Wahyono, S.Kom., M.Si
Pembimbing 2



Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM
Dekan



Anita Muliawati, S.Kom. MTI
Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Ujian : 27 Mei 2024

ASPECT-BASED SENTIMENT ANALYSIS OF MYTELKOMSEL APPLICATION'S REVIEW USING NAÏVE BAYES ALGORITHM AND TF-IDF METHODS

ABSTRACT

MyTelkomsel is a service in the form of an application provided by Telkomsel as one of the cellular operators that is widely used in Indonesia. Through reviews on the Google Play Store, users can find out the experiences of other users. However, the presence of a high rating with positive and negative reviews on MyTelkomsel does not guarantee that user expectations are met, so evaluation must still be carried out. This happens because it is not uncommon for users to give high ratings but with negative reviews. Apart from that, the development team can also find out which aspects have advantages and disadvantages, so they must be paid attention to as material for future evaluation. In this research, the Naïve Bayes algorithm and the TF-IDF method were used to carry out aspect-based sentiment analysis on reviews of the MyTelkomsel application, with the aim of knowing the performance of the Naïve Bayes algorithm by grouping reviews into four aspects of Nielsen's Model, namely learnability, efficiency, error, and satisfaction and two classes, namely positive and negative. Nielsen's Model is used as a benchmark to determine the strengths and weaknesses of the MyTelkomsel application. In this research, 1000 data were used in the form of user reviews taken using scrapping techniques on the Google Play Store. The results of this research obtained accuracy ranging from 79% to 93%. Apart from that, it was found that the advantages of this application were in the learnability aspect. Meanwhile, the weaknesses in this application are in the aspects of satisfaction, learnability and errors.

Keywords: *Aspect-Based Sentiment Analysis, Naïve Bayes, TF-IDF, Nielsen Model, MyTelkomsel*

ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN APLIKASI MYTELKOMSEL MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN METODE TF-IDF

ABSTRAK

MyTelkomsel merupakan sebuah layanan berupa aplikasi yang disediakan oleh Telkomsel sebagai salah satu operator seluler yang banyak digunakan di Indonesia. Melalui ulasan di Google Play Store, para pengguna dapat mengetahui pengalaman pengguna lainnya. Namun, adanya *rating* yang tinggi dengan ulasan positif dan negatif pada MyTelkomsel tidak menjamin bahwa ekspektasi pengguna terpenuhi, sehingga tetap harus dilakukan evaluasi. Hal ini terjadi karena tak jarang pengguna memberikan *rating* tinggi namun dengan ulasan yang negatif. Selain itu, tim *developer* juga dapat mengetahui aspek mana yang menjadi kelebihan dan kekurangan, sehingga harus diperhatikan untuk dijadikan bahan evaluasi kedepannya.

Pada penelitian ini, digunakan algoritma Naïve Bayes dan metode TF-IDF untuk melakukan analisis sentimen berbasis aspek pada ulasan aplikasi MyTelkomsel, dengan tujuan untuk mengetahui performansi algoritma Naïve Bayes dengan mengelompokkan ulasan ke dalam empat aspek Nielsen's Model yaitu yaitu *learnability*, *efficiency*, *error*, dan *satisfaction* serta dua kelas, yaitu positif dan negatif. Nielsen's Model digunakan sebagai tolak ukur untuk mengetahui aspek yang menjadi kelebihan dan kelemahan pada aplikasi MyTelkomsel. Dalam penelitian ini, digunakan sebanyak 1000 data berupa ulasan pengguna yang diambil menggunakan teknik *scrapping* di Google Play Store. Hasil penelitian ini mendapatkan akurasi yang berkisar antara 79% hingga 93%. Selain itu, ditemukan bahwa kelebihan pada aplikasi ini terdapat pada aspek *learnability*. Sedangkan kelemahan pada aplikasi ini terletak pada aspek *satisfaction*, *learnability*, dan *error*.

Kata Kunci: Analisis Sentimen Berbasis Aspek, Naïve Bayes, TF-IDF, Nielsen Model, MyTelkomsel

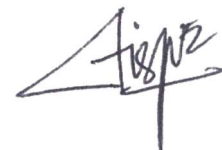
KATA PENGANTAR

Segala puji syukur dipanjatkan kepada Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan proposal penelitian yang berjudul “ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN PENGGUNA APLIKASI MYTELKOMSEL MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN METODE TF-IDF”.

Penulis menyadari bahwa penulisan penelitian ini masih jauh dari kata sempurna. Dalam perampungannya, penulis banyak mendapat dukungan dan motivasi dalam menyelesaikan penelitian ini dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih sebanyak-banyaknya kepada:

1. Allah SWT yang telah memberikan kesempatan dan kemudahan kepada penulis dalam menyelesaikan penelitian ini.
2. Bunda Ika Sari Armayani Harahap yang selalu mendoakan dan memberikan dukungan penuh kepada penulis.
3. Fidria Farzana Siregar, Fizka Fakhira Siregar, dan Filham Farizi Siregar selaku adik kandung penulis yang selalu memberikan support dan doa,
4. Mama Rohani Hsb, Nek Alom, dan Nek Ida, serta keluarga besar yang selalu mendoakan penulis
5. Muhammad Zidane Vigamada selaku partner saya yang selalu mendampingi dan membantu saya dalam menyelesaikan skripsi ini.
6. Ibu Ika Nurlaili Isnainiyah, S.Kom., M.Sc. selaku Dosen Pembimbing I dan Bapak Bambang Tri Wahyono, S.Kom., M.Si. selaku Dosen Pembimbing II yang telah membimbing penulis dalam menyusun skripsi ini.
7. Ibu Anita Muliawati, S.Kom., MTI. Selaku Ketua Jurusan Sistem Informasi
8. Mas Dennis Adiprawira selaku perwakilan pihak MyTelkomsel sebagai narasumber terkait wawancara mengenai kebutuhan analisis sentimen pada MyTelkomsel.
9. Kepada *annotator* yang telah membantu penulis dalam melakukan *labelling* data.
10. Bapak dan ibu dosen Fakultas Ilmu Komputer yang telah memberi ilmu yang bermanfaat.
11. Teman dekat dan teman-teman kuliah seperjuangan yang telah mendukung dan membantu penulis dalam menyelesaikan penelitian ini.

Jakarta, 22 April 2024



Fisya Alifia Fawwazi Siregar

DAFTAR ISI

PERNYATAAN ORISINALITAS	i
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN ...	ii
LEMBAR PENGESAHAN	iii
ABSTRACT	iv
ABSTRAK	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
DAFTAR LAMPIRAN	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	5
1.4 Manfaat Penelitian	5
1.5 Ruang Lingkup	6
1.6 Luaran	6
1.7 Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA	8
2.1 MyTelkomsel	8
2.2 Analisis Sentimen	8
2.3 Analisis Sentimen Berbasis Aspek	9
2.4 Model Nielsen	9
2.5 <i>Text Preprocessing</i>	10
2.5.1 Data Cleaning	10
2.5.2 Case Folding	10
2.5.3 Tokenizing	11
2.5.4 Stopword Removal	11
2.5.5 Stemming	11
2.6 Synthetic Minority Over-Sampling Technique (SMOTE)	12
2.7 Term Frequency-Inverse Document Frequency (TF-IDF)	13
2.8 Algoritma Naïve Bayes	13
2.9 Evaluasi	15
2.10 Word Cloud	17
2.11 Network Analysis	17

2.12	Penelitian Terkait	17
BAB III METODE PENELITIAN.....		22
3.1	Tahapan Penelitian.....	22
3.1.1	Identifikasi Masalah	22
3.1.2	Studi Literatur.....	24
3.1.3	Pengumpulan Data.....	24
3.1.4	<i>Labelling</i>	25
3.1.5	<i>Preprocessing Data</i>	26
3.1.6	Word Embedding.....	30
3.1.7	<i>Balancing Data</i> dengan SMOTE.....	30
3.1.8	Klasifikasi dengan Naïve Bayes.....	31
3.1.9	Evaluasi	32
3.1.10	Visualisasi Data	33
3.2	Alat Bantu Penelitian	33
3.3	Jadwal Penelitian	34
BAB IV HASIL DAN PEMBAHASAN.....		36
4.1	Pengumpulan Data.....	36
4.2	<i>Labelling Data</i>	36
4.3	<i>Pre-Processing Data</i>	40
4.3.1	<i>Cleaning</i>	40
4.3.2	<i>Case-folding</i>	40
4.3.3	<i>Tokenizing</i>	41
4.3.4	<i>Stopword Removal</i>	41
4.3.5	<i>Stemming</i>	42
4.4	<i>Word Embedding</i>	42
4.5	Pembagian Data	43
4.6	<i>Balancing Data</i>	44
4.7	Klasifikasi	44
4.8	Evaluasi.....	45
4.8.1	Hasil Uji Skenario 1	45
4.8.2	Hasil Uji Skenario 2	47
4.8.3	Hasil Uji Skenario 3	49
4.8.4	Hasil Uji Skenario 4	51
4.8.5	Hasil Uji Skenario 5	53
4.8.6	Hasil Uji Skenario 6	55
4.8.7	Hasil Uji Skenario 7	56
4.8.8	Hasil Uji Skenario 8	58
4.9	<i>Word Cloud</i>	60

4.9.1	<i>Word Cloud Uji Skenario 1</i>	61
4.9.2	<i>Word Cloud Uji Skenario 2</i>	61
4.9.3	<i>Word Cloud Uji Skenario 3</i>	62
4.9.4	<i>Word Cloud Uji Skenario 4</i>	63
4.9.5	<i>Word Cloud Uji Skenario 5</i>	63
4.9.6	<i>Word Cloud Uji Skenario 6</i>	64
4.9.7	<i>Word Cloud Uji Skenario 7</i>	65
4.9.8	<i>Word Cloud Uji Skenario 8</i>	65
4.10	<i>Network Analysis</i>	66
4.11	Perbandingan Hasil Skenario Pengujian Data	67
4.12	Analisis Sentimen Setiap Aspek	68
BAB V KESIMPULAN DAN SARAN		70
5.1	Kesimpulan	70
5.2	Saran	71
DAFTAR PUSTAKA		72
RIWAYAT HIDUP		76

DAFTAR GAMBAR

Gambar 1. 1 Peningkatan Jumlah Pengguna Internet di Indonesia Tahun 2023	1
Gambar 1. 2 Presentase Operator Seluler yang Digunakan di Indonesia Tahun 2023.....	2
Gambar 3. 1 Diagram Alur Penelitian	22
Gambar 3. 2 Contoh <i>Labelling</i> Dataset	26
Gambar 3. 3 Diagram Alur Preprocessing Data	27
Gambar 3. 4 Diagram Alur Kerja Naïve Bayes.....	32
Gambar 4. 1 Distribusi Ulasan Aspek Satisfaction	38
Gambar 4. 2 Distribusi Ulasan Aspek Learnability.....	38
Gambar 4. 3 Distribusi Ulasan Aspek Efficiency.....	39
Gambar 4. 4 Distribusi Ulasan Aspek Error.....	39
Gambar 4. 5 Hasil Proses Cleaning	40
Gambar 4. 6 Hasil Proses Case-Folding.....	41
Gambar 4. 7 Hasil Proses Tokenizing	41
Gambar 4. 8 Hasil Proses Stopword Removal	42
Gambar 4. 9 Hasil Proses Stemming.....	42
Gambar 4. 10 Hasil Label Encoding pada Kolom Sentimen.....	43
Gambar 4. 11 Hasil TF-IDF pada Kolom Ulasan.....	43
Gambar 4. 12 Confusion Matrix Skenario 1	45
Gambar 4. 13 Classification Report Skenario 1	46
Gambar 4. 14 Confusion Matrix Skenario 2	47
Gambar 4. 15 Classification Report Skenario 2	48
Gambar 4. 16 Confusion Matrix Skenario 3	49
Gambar 4. 17 Classification Report Skenario 3	50
Gambar 4. 18 Confusion Matrix Skenario 4	51
Gambar 4. 19 Classification Report Skenario 4	52
Gambar 4. 20 Confusion Matrix Skenario 5	53
Gambar 4. 21 Classification Report Skenario 5	54
Gambar 4. 22 Confusion Matrix Skenario 6	55
Gambar 4. 23 Classification Report Skenario 6	56
Gambar 4. 24 Confusion Matrix Skenario 7	57
Gambar 4. 25 Classification Report Skenario 7	58
Gambar 4. 26 Confusion Matrix Skenario 8	59
Gambar 4. 27 Classification Report Skenario 8	60
Gambar 4. 28 Word Cloud Skenario 1	61
Gambar 4. 29 Word Cloud Skenario 2	61
Gambar 4. 30 Word Cloud Skenario 3	62
Gambar 4. 31 Word Cloud Skenario 4	63
Gambar 4. 32 Word Cloud Skenario 5	63
Gambar 4. 33 Word Cloud Skenario 6	64
Gambar 4. 34 Word Cloud Skenario 7	65
Gambar 4. 35 Word Cloud Skenario 8	65
Gambar 4. 36 Network Analysis Sentimen Setiap Aspek.....	66
Gambar 4. 38 Akurasi Model Setiap Aspek.....	67
Gambar 4. 39 Jumlah Sentimen Positif dan Negatif Setiap Aspek	68

DAFTAR TABEL

Tabel 2. 1 <i>Confusion Matrix</i>	16
Tabel 2. 2 <i>Literature Review</i>	18
Tabel 3. 1 Tabel Identifikasi Aspek	25
Tabel 3. 2 Contoh Proses Cleaning	27
Tabel 3. 3 Contoh Proses Case Folding.....	27
Tabel 3. 4 Contoh Proses Tokenizing.....	28
Tabel 3. 5 Contoh Proses Stopword Removal.....	29
Tabel 3. 6 Contoh Proses Stemming	29
Tabel 3. 7 Skenario Tahapan Pengujian	32
Tabel 3. 8 Confusion Matrix	33
Tabel 3. 9 Jadwal Penelitian.....	34

DAFTAR LAMPIRAN

Lampiran 1. Wawancara Pra-Penelitian	76
Lampiran 2. Dokumentasi Wawancara Pra-Penelitian.....	78
Lampiran 3. Hasil Turnitin	78