

**ANALISIS ULASAN DATA LAYANAN TELEKOMUNIKASI
INDONESIA PADA *TWITTER* MENGGUNAKAN METODE
NAIVE BAYES DENGAN *FEATURE SELECTION***

TUGAS AKHIR



RIFINDA FITRIANI

1152001010

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS BAKRIE
JAKARTA
2020**

**ANALISIS ULASAN DATA LAYANAN TELEKOMUNIKASI
INDONESIA PADA *TWITTER* MENGGUNAKAN METODE
NAIVE BAYES DENGAN *FEATURE SELECTION***

TUGAS AKHIR

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana
Komputer**



RIFINDA FITRIANI

1152001010


**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS BAKRIE
JAKARTA
2020**

HALAMAN PERNYATAAN ORISINALITAS

Tugas Akhir adalah hasil karya saya sendiri, dan semua sumber baik dikutip maupun dirujuk telah saya nyatakan dengan benar.

Nama : Rifinda Fitriani

NIM : 1152001010

Tanda Tangan : 

Tanggal : 10 Maret 2020

HALAMAN PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama : Rifinda Fitriani
NIM : 1152001010
Program Studi : Informatika
Fakultas : Teknik dan Ilmu Komputer
Judul Proposal : Analisis Ulasan Data Layanan Telekomunikasi Indonesia
Pada *Twitter* Menggunakan Metode *Naive Bayes* dengan
Feature Selection

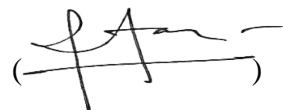
Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian dari persyaratan untuk melakukan penelitian pada Program Studi Informatika Fakultas Teknik dan Ilmu Komputer Universitas Bakrie

DEWAN PENGUJI

Pembimbing : Prof. Dr. Hoga Saragih, ST.,MT.,



Penguji : Refyul Rey Fatri, M.Sc.



Penguji : Sigit Wijayanto, M.Sc.



Ditetapkan di : Jakarta

Tanggal : 10 Maret 2020

UCAPAN TERIMA KASIH

Alhamdulillahahirabbil'aalamiin, puji dan syukur kehadiran Allah SWT karena berkat segala rahmat dan karunia-Nya Tugas Akhir yang berjudul “Analisis Ulasan Data Layanan Telekomunikasi Indonesia Pada *Twitter* Menggunakan Metode *Naive Bayes* dengan *Feature Selection*” dapat diselesaikan. Shalawat serta salam tak lupa pula peneliti hanturkan kepada Nabi Besar Muhammad SAW yang telah membimbing umatnya ke masa yang terang benderang penuh dengan cahaya iman.

Penyusunan Tugas Akhir ini tidak lepas dari berbagai kesulitan dan hambatan dari awal hingga akhir penyusunan. Penulis menyampaikan terima kasih yang sebesar-besarnya kepada dosen pembimbing Bapak Prof. Dr. Hoga Saragih, ST, MT. yang telah meluangkan waktu serta mengerahkan tenaga dan pikiran untuk membantu proses pengerjaan dan penyusunan Tugas Akhir ini. Begitu banyak pihak yang turut membantu dan memberikan dukungan, masukan, nasihat serta doa selama penyusunan Tugas Akhir ini. Oleh karena itu, dengan segala hormat dan kerendahan hati, penulis ingin mengungkapkan rasa terima kasih kepada:

1. Kedua orang tua dan keluarga yang telah membeikan dukungan do'a, nasihat, motivasi, dan semangat serta selalu menjadi pengingat selama ini.
2. Bapak Prof. Hoga Saragih S.T.,M.T. selaku Ketua Program Studi Informatika, dan sekaligus sebagai dosen pembimbing yang telah membimbing dan memberikan saran dan perbaikan dalam penelitian ini.
3. Bapak Refyul Rey Fatri, M.Sc. dan Bapak Sigit Wijayanto, M.Sc. selaku dosen penguji yang memberikan saran dan perbaikan dalam penelitian ini.
4. Salma, Ofi, Dona yang senantiasa memberikan dukungan, bantuan, semangat dan doa selama proses pengerjaan skripsi ini.
5. Teman – teman Informatika 2015 : Ahmad Gadran, Ahmad Muklis, Alifian Azmi, Alhamsya, Ayu Hapsari, Aziz Sentosa Setiana, Meivy Miranda, Nadika Safira, Ocarina Cloramidina, Rani Septhia Dewi,

Wahyu Widodo

6. Teman – teman Informatika angkatan 2015 yang tidak bisa disebutkan satu persatu telah memberikan dukungan, dan semangat serta kebersamaan dalam suka maupun duka selama 4 tahun perkuliahan yang memberikan dan pembelajaran selama masa perkuliahan di Universitas Bakrie.
7. Kakak – kakak angkatan 2014 dan adik – adik 2016 dan 2017 Informatika Universitas Bakrie yang telah memberikan dukungan, pengalaman, motivasi dan semangatnya selama masa perkuliahan dan penyusunan Tugas Akhir ini.
8. Teman – teman Universitas Bakrie angkatan 2015 yang tidak bisa disebutkan satu persatu yang telah memberikan semangat selama penyusunan Tugas Akhir ini.
9. Seluruh Dosen Program Studi Informatika yang telah memberikan ilmu dan pembelajaran serta pengalaman yang sangat bermanfaat selama masa perkuliahan di Universitas Bakrie.
10. Seluruh pihak Universitas Bakrie yang terlibat langsung maupun tidak yang telah memberikan pengalaman, bantuan, motivasi, dan fasilitas yang sangat membantu selama masa perkuliahan dan penyusunan Tugas Akhir ini.

Semoga Allah SWT selalu memberikan kebaikan dan keberkahan kepada kita semua. Semoga Tugas Akhir ini dapat memberikan manfaat dan berguna bagi semua kalangan bidang pendidikan, khususnya bidang Informatika.

Jakarta, Maret 2020

Penulis

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI

Sebagai *civitas* akademik Universitas Bakrie, saya yang bertanda tangan di bawah ini:

Nama : Rifinda Fitriani
NIM : 1152001010
Program Studi : Informatika
Fakultas : Teknik dan Ilmu Komputer

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bakrie **Hak Bebas Royalti Noneksklusif** (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul:

**Analisis Ulasan Data Layanan Telekomunikasi Indonesia Pada *Twitter*
Menggunakan Metode *Naive Bayes* dengan *Feature Selection***

Dengan Hak Bebas Royalti Noneksklusif ini Universitas Bakrie berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta untuk kepentingan akademis.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta
Pada tanggal : 10 Maret 2020

Yang menyatakan



Rifinda Fitriani

**ANALISIS ULASAN DATA LAYANAN TELEKOMUNIKASI
INDONESIA PADA *TWITTER* MENGGUNAKAN METODE *NAIVE
BAYES* DENGAN *FEATURE SELECTION***

Rifinda Fitriani

ABSTRAK

Twitter adalah salah satu layanan media sosial dan *microblogging* online yang paling umum digunakan dengan banyaknya pengguna dapat berbagi pikiran dan pendapat mereka tentang berbagai aspek. Hal ini karena *twitter* dianggap sebagai sumber yang kaya informasi untuk pengambilan keputusan dan analisis sentimen. Analisis sentimen ini bertujuan untuk mengatasi *opinion mining* yang berasal dari *tweets* untuk menentukan positif dan negatif pada *Twitter* agar dapat mengetahui kualitas provider telekomunikasi yang berada di Indonesia. Dalam penelitian ini, data yang diambil merupakan *tweet* ulasan pelanggan masing-masing pengguna provider yang berada di *twitter*. Terdapat tiga provider yang dijadikan kategori dalam melakukan analisis sentimen. Dari tiga provider tersebut akan diambil masing-masing 300 data dengan total jumlah data sebanyak 900 data ulasan yang akan dilakukan penelitian. Data tersebut kemudian dilakukan *preprocessing* dan pengklasifikasian menggunakan algoritma *Naive Bayes* dengan perbandingan seleksi fitur antara *Information Gain* dan *Term Frequency Invers Document Frequency*. Menghasilkan pengukuran performa diperoleh masing-masing akurasi sebesar 77,50% untuk TF-IDF dan *Naive Bayes*, serta akurasi yang diperoleh untuk IG dan *Naive Bayes* sebesar 87,50%.

Kata Kunci : *Twitter, Naive Bayes Classifier (NBC), Analisis Sentimen, TF-IDF, Information Gain, Seleksi Fitur.*

**ANALISIS ULASAN DATA LAYANAN TELEKOMUNIKASI
INDONESIA PADA *TWITTER* MENGGUNAKAN METODE *NAIVE
BAYES* DENGAN *FEATURE SELECTION***

Rifinda Fitriani

ABSTRACT

Twitter is one of the most commonly used online social media and microblogging services with many users sharing their thoughts and opinions on various aspects. This is because Twitter is considered as a rich source of information for decision making and sentiment analysis. This sentiment analysis aims to overcome the opinion mining that comes from tweets to determine the positive and negative of Twitter in order to find out the quality of telecommunications providers in Indonesia. In this study, the data taken is a customer review tweet of each provider user who is on Twitter. There are three providers that are categorized as sentiment analysis. Of the three providers, 300 data will be taken each with a total amount of 900 data review data that will be conducted research. The data is then carried out preprocessing and classification using the Naive Bayes algorithm with a feature selection comparison between Information Gain and Term Frequency Invers Document Frequency. The resulting performance measurements obtained an accuracy of 77.50% for TF-IDF and Naive Bayes, as well as accuracy obtained for IG and Naive Bayes of 87.50%.

Keyword : Twitter, Naive Bayes Classifier (NBC), Sentiment Analysis, TF-IDF, Information Gain.

DAFTAR ISI

HALAMAN PERNYATAAN ORISINALITAS	i
HALAMAN PENGESAHAN	ii
UCAPAN TERIMA KASIH	iii
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI.....	v
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	x
DAFTAR TABEL	xi
DAFTAR RUMUS	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Batasan Masalah.....	4
1.5 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Penelitian Terkait	6
2.2 Analisis Sentimen	12
2.2.1. Pengertian Analisis Sentimen.....	12
2.2.2. Klasifikasi Analisis Sentimen	12
2.3 Twitter	13
2.4 Text Mining	14
2.5 Naive Bayes Classifier (NBC).....	14
2.6 Pemilihan Fitur (Feature Selection).....	16
2.6.1 Information Gain	18
2.6.2 Term Frequency Invers Dokument Frequency (TF-IDF)	18
2.7 Evaluasi dan validasi Klasifikasi.....	20
BAB III METODOLOGI PENELITIAN	22
3.1 Studi Literatur	23
3.2 Menentukan Rumusan Masalah	23

3.3	Pengumpulan Data	24
3.4	<i>Preprocessing</i>	24
3.4.1.	<i>Case Folding</i>	24
3.4.2.	Normalisasi	25
3.4.3.	<i>Stemming</i>	26
3.4.1.	<i>Stopword Removal</i>	26
3.4.2.	<i>Tokenizing</i>	28
3.5	Seleksi Fitur (<i>feature selection</i>)	29
3.6	Klasifikasi <i>Naive Bayes Classifier</i> (NBC)	29
3.7	Analisis Hasil dan Pembahasan Klasifikasi	30
3.8	Penyusunan Laporan	30
3.9	Jadwal Penelitian	30
BAB IV IMPLEMENTASI DAN PEMBAHASAN		32
4.1	Pengumpulan Data	32
4.2	Dataset Ulasan Provider	32
4.3	Hasil <i>Preprocessing Data</i>	33
4.3.1.	<i>Case Folding</i>	34
4.3.2.	Normalisasi	34
4.3.3.	<i>Stemming</i>	37
4.3.4.	<i>Stopword Removal</i>	37
4.4	Hasil Seleksi Fitur (<i>Feature Selection</i>)	38
4.4.1.	Hasil <i>Information Gain</i>	39
4.4.2.	Hasil <i>Term Frequency Invers Dokument Frequency</i> (TF-IDF)..	41
4.5	Hasil Klasifikasi <i>Naive Bayes</i> (NBC)	42
4.6	Pengukuran Performa	45
BAB V KESIMPULAN DAN SARAN		48
5.1	Kesimpulan	48
5.2	Saran	49
DAFTAR PUSTAKA		50
LAMPIRAN		53

DAFTAR TABEL

Table 2.1 Penelitian Terkait	10
Table 2.2 <i>Predicted Class</i>	20
Table 3.1 Proses penerapan <i>Case Folding</i>	25
Table 3.2 Proses penerapan Normalisasi	25
Table 3.3 Proses penerapan <i>Stemming</i>	26
Table 3.4 Proses Penerapan <i>Stopword Removal</i>	27
Table 3.5 Proses Penerapan <i>Tokenizing</i>	28
Table 3.6 Rencana Kegiatan Penelitian	31
Table 4.1 Data Ulasan <i>Provider</i>	33
Table 4.2 Perbandingan data <i>Predicted</i> dan <i>Actual</i>	46
Table 4.3 <i>Confussion Matrix</i> Sentimen	46

DAFTAR GAMBAR

Gambar 3.1 Kerangka Penelitian	22
Gambar 3.2 <i>Flowchart</i> Proses <i>Stopword Removal</i>	27
Gambar 3.3 <i>Flowchart</i> Tahapan <i>Preprocessing</i>	29
Gambar 4.1 <i>Source Code Case Folding</i>	34
Gambar 4.2 <i>Source Code</i> Normalisasi Singkatan	35
Gambar 4.3 <i>Source Code</i> Normalisasi Bahasa Inggris	36
Gambar 4.4 <i>Source Code Stemming</i>	37
Gambar 4.5 <i>Source Code Stopword Removal</i>	38
Gambar 4.6 Hasil Perhitungan Frekuensi Kata	39
Gambar 4.7 <i>Source Code</i> Perhitungan Nilai <i>Entropy</i>	40
Gambar 4.8 <i>Source Code</i> Perhitungan Nilai <i>Information Gain</i>	40
Gambar 4.9 Hasil Perangkingan <i>Information Gain</i>	41
Gambar 4.10 <i>Source Code</i> Perhitungan TF-IDF	41
Gambar 4.11 Potongan Hasil TF-IDF	42
Gambar 4.12 Pembagian Data	42
Gambar 4.13 <i>Class Diagram Naive Bayes</i>	43
Gambar 4.14 <i>Source Code</i> Perhitungan <i>Prior</i>	44
Gambar 4.15 <i>Source Code</i> Perhitungan <i>Conditional Probability</i>	44
Gambar 4.16 <i>Source Code</i> Perhitungan <i>Cmap</i>	45
Gambar 4.17 <i>Source Code</i> Penentuan Kelas Sentimen	45

DAFTAR RUMUS

Persamaan 2.1	14
Persamaan 2.2	15
Persamaan 2.3	15
Persamaan 2.4	15
Persamaan 2.5	16
Persamaan 2.6	16
Persamaan 2.7	18
Persamaan 2.8	18
Persamaan 2.9	19
Persamaan 2.10	19
Persamaan 2.11	19
Persamaan 2.12	21
Persamaan 2.13	21
Persamaan 2.14	21
Persamaan 2.15	21