



**SENTIMEN ANALISIS TERHADAP LAYANAN FIRST MEDIA
MENGUNAKAN METODE *NAÏVE BAYES***

SKRIPSI

RIZKI INDAH PRATIWI

1710511020

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2021



**SENTIMEN ANALISIS TERHADAP LAYANAN FIRST MEDIA
MENGUNAKAN METODE *NAÏVE BAYES***

SKRIPSI

**Diajukan Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana
Komputer**

RIZKI INDAH PRATIWI

1710511020

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2021

PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya sendiri, dan sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Rizki Indah Pratiwi

NIM : 1710511020

Tanggal : 2 Juli 2021

Bilamana dikemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 2 Juli 2021

Yang menyatakan,



Rizki Indah Pratiwi

PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

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

Nama : Rizki Indah Pratiwi

NIM : 1710511020

Fakultas : Ilmu Komputer

Program Studi : Informatika

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

**Sentimen Analisis Terhadap Layanan First Media Menggunakan Metode
*Naïve Bayes***

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

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada tanggal : 2 Juli 2021

Yang menyatakan,



(Rizki Indah Pratiwi)

LEMBAR PENGESAHAN

Dengan ini dinyatakan bahwa Skripsi berikut:

Nama : Rizki Indah Pratiwi
NIM : 1710511020
Program Studi : Informatika
Judul Skripsi : Sentimen Analisis Terhadap Layanan First Media
Menggunakan Metode *Naïve Bayes*

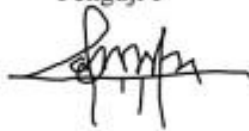
Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



Iin Ernawati, S.Kom., M.Si.
Penguji I



Ing. Artambo B. Pangaribuan., BSc.
Penguji II



Dr. Ermatita, M.Kom.
Pembimbing I



Nurul Chamidah, S.Kom, M.Kom.
Pembimbing II



Dr. Ermatita, M.Kom.
Dekan



Yuni Widiastiwi, S.Kom., M.Si.
Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Ujian : 13 Juli 2021



SENTIMEN ANALISIS TERHADAP LAYANAN FIRST MEDIA MENGUNAKAN METODE *NAÏVE BAYES*

Rizki Indah Pratiwi

Abstrak

Twitter menjadi salah satu media sosial yang sering dimanfaatkan konsumen untuk menyampaikan keluhan ataupun opini terkait layanan yang diberikan oleh suatu perusahaan. Dimana keluhan atau opini yang disampaikan melalui *tweet* tersebut dapat diolah untuk mengetahui sentimen yang terkandung dalam *tweet* tersebut. Oleh karena itu penelitian ini dilaksanakan untuk melakukan proses klasifikasi data *tweet* yang didalamnya mengandung sentimen positif maupun negatif mengenai layanan dari salah satu perusahaan yaitu First Media. Metode yang digunakan untuk proses klasifikasi pada penelitian ini yaitu dengan menggunakan metode *Naïve Bayes*. Data *tweet* diambil dengan menggunakan *Twitter API (Application Programming Interface)* yang tersedia pada *platform Twitter*. Dimana data *tweet* akan diberi label positif dan negatif untuk memudahkan data tersebut diproses. *Tweet* yang diambil adalah *tweet* yang menyertakan *@FirstMediaCares* sebagai kata kunci. Selanjutnya data *tweet* yang terkumpul dibagi kedalam data latih sebesar 70% dan data uji sebesar 30%. Hasil evaluasi yang berfungsi untuk menguji performa dari metode *Naïve Bayes* dalam mengklasifikasikan *tweet* kedalam kategori positif dan negatif menunjukkan nilai akurasi yang didapat yaitu sebesar 89.47%.

Kata kunci: *Twitter, tweet, Naïve Bayes, analisis sentimen, evaluasi*

SENTIMEN ANALISIS TERHADAP LAYANAN FIRST MEDIA MENGUNAKAN METODE NAÏVE BAYES

Rizki Indah Pratiwi

Abstract

Twitter is one of the social media that is often used by consumers to submit complaints or opinions regarding the services provided by a company. Where complaints or opinions submitted through the tweet can be processed to find out the sentiments contained in the tweet. Therefore, this research was conducted to carry out the process of classifying tweet data in which it contains positive or negative sentiments regarding the services of one company, First Media. The method used for the classification process in this study is the Naïve Bayes method. Tweet data is retrieved using the Twitter API (Application Programming Interface) available on the Twitter platform. Where the tweet data will be labeled positive and negative to make it easier for the data to be processed. Retrieved tweets are tweets that include @FirstMediaCares as a keyword. Furthermore, the tweet data collected is divided into training data by 70% and test data by 30%. The results of the evaluation that serves to test the performance of the Naïve Bayes method in classifying tweets into positive and negative categories show the accuracy value obtained is 89.47%.

Keywords: *Twitter, tweet, Naïve Bayes, sentiment analysis, evaluation*

KATA PENGANTAR

Puji serta syukur penulis panjatkan kehadirat Allah SWT atas segala limpahan rahmat, karunia dan ridha-Nya, sehingga Skripsi dengan judul “Sentimen Analisis Terhadap Layanan First Media Menggunakan Metode *Naïve Bayes*” berhasil diselesaikan. Tak lupa, penulis ingin mengucapkan banyak terima kasih kepada:

1. Kedua orang tua penulis, Tohirin (bapak) dan Almh. Kartimah (ibu), serta keluarga yang selalu memberikan doa dan motivasi sehingga penulis dapat menyelesaikan skripsi ini.
2. Ibu Dr. Ermatita, M. Kom. dan Ibu Nurul Chamidah, S.Kom, M.Kom., selaku dosen pembimbing yang telah membimbing serta memberikan saran yang bermanfaat.
3. Ibu Dr. Ermatita, M. Kom. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta.
4. Ibu Yuni Widiastiwi, S.Kom., M.Si. selaku Ketua Program Studi Informatika Universitas Pembangunan Nasional Veteran Jakarta.
5. Bapak Bayu Hananto, S.Kom., M.Kom selaku dosen pembimbing akademik.
6. Bapak/Ibu dosen Informatika Universitas Pembangunan Nasional Veteran Jakarta yang telah memberikan ilmu yang bermanfaat.
7. Teman-teman Informatika angkatan 2017, kakak tingkat, dan seluruh rekan mahasiswa yang tidak dapat disebutkan namanya satu-persatu oleh penulis. Terima kasih atas segala motivasi, saran yang diberikan serta pengalaman suka dan duka semasa kuliah yang dilewati bersama-sama.
8. Semua teman-teman yang tidak bisa penulis sebutkan satu-persatu. Terima kasih atas segala dukungan, bantuan, dan doa yang telah diberikan.

Akhir kata penulis ucapkan terima kasih dan semoga skripsi ini dapat bermanfaat bagi pembaca.

Jakarta, 2 Juli 2021

Penulis

DAFTAR ISI

SENTIMEN ANALISIS TERHADAP LAYANAN FIRST MEDIA
MENGUNAKAN METODE *NAÏVE BAYES*

HALAMAN JUDUL	i
PERNYATAAN ORISINALITAS.....	ii
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI.....	iii
LEMBAR PENGESAHAN	iv
Abstrak.....	v
Abstract.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xiii
DAFTAR SIMBOL	xiv
DAFTAR LAMPIRAN	xv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	4
1.5 Batasan Penelitian	4

1.6	Luaran Yang Diharapkan	5
1.7	Sistematika penulisan.....	5
BAB 2 TINJAUAN PUSTAKA.....		7
2.1	<i>Twitter</i>	7
2.2	Sentimen Analisis.....	8
2.3	<i>Naïve Bayes</i>	8
2.4	Evaluasi.....	11
2.5	Penelitian Terdahulu	12
BAB 3 METODOLOGI PENELITIAN		13
3.1	Kerangka Pikir	13
3.1.1	Penentuan Topik Permasalahan	14
3.1.2	Identifikasi Masalah	14
3.1.3	Studi Literatur	14
3.1.4	Pengumpulan Data	14
3.1.5	Pelabelan Data.....	15
3.1.6	Praproses Data.....	15
3.1.7	Pembobotan Kata	16
3.1.8	Pembagian Data	17
3.1.9	<i>Training</i>	17
3.1.10	<i>Testing</i>	18
3.1.11	Evaluasi.....	18
3.1.12	Analisa Hasil	18
3.2	Perangkat Penelitian.....	19
3.2.1	Perangkat Lunak (<i>software</i>)	19

3.2.2	Perangkat Keras (<i>hardware</i>)	19
3.3	Jadwal Penelitian.....	19
BAB 4 HASIL DAN PEMBAHASAN		21
4.1	Data	21
4.2	Pelabelan Data.....	22
4.3	Praproses Data.....	26
4.4	Pembobotan Kata (TF-IDF)	31
4.5	Pembagian Data	34
4.6	Proses <i>Training</i>	35
4.7	Proses <i>Testing</i>	42
4.8	Evaluasi.....	47
4.9	Analisa Hasil	50
BAB 5 PENUTUP.....		59
5.1	Kesimpulan	59
5.2	Saran.....	59
DAFTAR PUSTAKA		61
RIWAYAT HIDUP		66
LAMPIRAN.....		67

DAFTAR TABEL

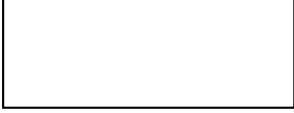
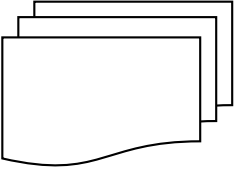




<i>Tabel 2.1 Confussion Matrix</i>	11
<i>Tabel 3.1 Jadwal Penelitian</i>	20
<i>Tabel 4.1 Contoh Proses Pelabelan Data Tweet</i>	22
<i>Tabel 4.2 Rincian Hasil dari Proses Pelabelan Data</i>	23
<i>Tabel 4.3 Kategori Nilai Kappa</i>	25
<i>Tabel 4.4 Hasil Label Akhir</i>	25
<i>Tabel 4.5 Tahap Cleaning Data Tweet</i>	26
<i>Tabel 4.6 Tahap Case Folding pada Data Tweet</i>	27
<i>Tabel 4.7 Tahap Tokenizing Data Tweet</i>	28
<i>Tabel 4.8 Tahap Stopword Removal pada Data Tweet</i>	29
<i>Tabel 4.9 Tahap Stemming Data Tweet</i>	31
<i>Tabel 4.10 Dokumen (tweet)</i>	32
<i>Tabel 4.11 Perhitungan TF-IDF</i>	33
<i>Tabel 4.12 Rincian Pembagian (split) Data</i>	35
<i>Tabel 4.13 Rincian Data Training</i>	35
<i>Tabel 4.14 Sampel Data Training</i>	36
<i>Tabel 4.15 Hasil Bobot Trainig Set</i>	36
<i>Tabel 4.16 Rincian Data Testing</i>	42
<i>Tabel 4.17 Sampel Data Testing</i>	43
<i>Tabel 4.18 Data setelah Melalui Tahapan Praproses</i>	43
<i>Tabel 4.19 Bobot Testing Set</i>	43

<i>Tabel 4.20 Confusion Matrix</i>	<i>48</i>
<i>Tabel 4.21 Perbandingan Hasil Confusion Matrix.....</i>	<i>51</i>
<i>Tabel 4.22 Penerapan Metode Undersampling</i>	<i>51</i>
<i>Tabel 4.23 Confusion Matrix untuk Hasil Metode Undersampling</i>	<i>52</i>

DAFTAR GAMBAR

<i>Gambar 3.1 Kerangka Pikir.....</i>	13
<i>Gambar 4.1 Hasil Crawling yang Telah Disimpan dalam Bentuk File .csv.....</i>	22
<i>Gambar 4.2 Alur Pembobotan TF-IDF</i>	32
<i>Gambar 4.3 Wordcloud Sentimen Positif.....</i>	54
<i>Gambar 4.4 Wordcloud Sentimen Negatif</i>	55
<i>Gambar 4.5 Kata yang Sering Muncul pada Sentimen Positif.....</i>	56
<i>Gambar 4.6 Kata yang Sering Muncul pada Sentimen Negatif.....</i>	57
<i>Gambar 4.7 Keterangan Waktu Terjadinya Gangguan</i>	58

DAFTAR SIMBOL

Simbol	Nama Simbol	Keterangan
	Simbol Proses	Menunjukkan suatu proses yang dilakukan
	Simbol <i>Multiple</i> Dokumen	Menunjukkan dokumen yang digunakan lebih dari satu
	Simbol Arus	Berfungsi untuk menghubungkan antar simbol dan menunjukkan alur proses
	Simbol Titik Terminal	Menunjukkan suatu titik permulaan (<i>start</i>) atau akhir (<i>stop</i>) dari suatu proses
	Simbol Data	Simbol sebagai masukan atau keluaran data untuk suatu proses
	Simbol Dokumen	Dokumen yang dibutuhkan dalam suatu proses

DAFTAR LAMPIRAN

Lampiran 1 <i>Similarity Index</i> Skripsi	68
Lampiran 2 <i>Source Code</i> Proses Klasifikasi <i>Tweet</i> (Klasifikasi <i>Tweet_NB.ipynb</i>)...	79