



**ANALISIS SENTIMEN *REVIEW* APLIKASI BERITA *ONLINE* PADA
GOOGLE PLAY MENGGUNAKAN METODE NAÏVE BAYES
STUDI KASUS: TRIBUNNEWS.COM**

SKRIPSI

ADITYA SAMIAJI

1810511052

PROGRAM STUDI SARJANA INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

JAKARTA

2022



**ANALISIS SENTIMEN *REVIEW* APLIKASI BERITA *ONLINE* PADA
GOOGLE PLAY MENGGUNAKAN METODE NAÏVE BAYES**

STUDI KASUS: TRIBUNNEWS.COM

SKRIPSI

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

ADITYA SAMIAJI

1810511052

PROGRAM STUDI SARJANA INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

JAKARTA

2022

PERNYATAAN ORISINALITAS

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

Nama : Aditya Samiaji

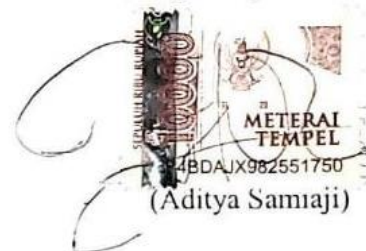
NIM : 1810511052

Tanggal : 19 Juli 2022

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

Jakarta, 19 Juli 2022

Yang Menyatakan,


METERAI
TEMPEL
4BDAJX982551750
(Aditya Samiaji)

**PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK
KEPENTINGAN AKADEMIS**

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

Nama : Aditya Samiaji

NIM : 1810511052

Fakultas : Ilmu Komputer

Program Studi : S1 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:

**Analisis Sentimen *Review* Aplikasi Berita *Online* Pada Google Play
Menggunakan Metode Naïve Bayes Studi Kasus: Tribunnews.com**

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 sesungguhnya.

Dibuat di : Jakarta

Pada Tanggal : 19 Juli 2022

Yang Menyatakan,



(Aditya Samiaji)

LEMBAR PENGESAHAN

Dengan ini dinyatakan bahwa Tugas Akhir sebagai berikut:

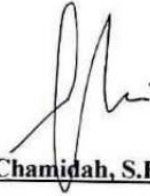
Nama : Aditya Samiaji
NIM : 1810511052
Program Studi : S1 Informatika
Judul Proposal : Analisis Sentimen *Review* Aplikasi Berita *Online* Pada
Google Play Menggunakan Metode Naïve Bayes Studi
Kasus: Tribunnews.com

Telah berhasil dipertahankan dihadapan Tim Penguji dan diterima sebagai bagian dari persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi S-1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



Henki Bayu Seta, S.Kom, MTI.

Penguji I



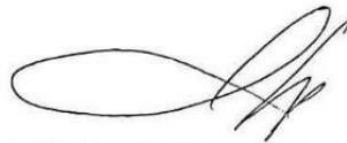
Nurul Chamidah, S.Kom., M.Kom

Penguji II



Bayu Hananto, S.Kom., M.Kom.

Pembimbing I



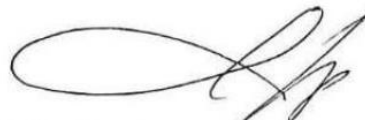
Desta Sandya Prasvita, S.Komp., M.Kom.

Pembimbing II



Dr. Ermatita, M.Kom.

Dekan



Desta Sandya Prasvita, S.Komp., M.Kom.

Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Pengesahan : 19 Juli 2022



**ANALISIS SENTIMEN *REVIEW* APLIKASI BERITA *ONLINE* PADA
GOOGLE PLAY MENGGUNAKAN METODE NAÏVE BAYES STUDI
KASUS: TRIBUNNEWS.COM**

Aditya Samiaji

ABSTRAK

Aplikasi Tribunnews merupakan sebuah aplikasi berita online yang sudah diunduh lebih dari 500 ribu pengguna. Pesatnya perkembangan teknologi seluler genggam membawa dampak yang sangat besar bagi umat manusia dalam penggunaan smartphone dan internet. Banyaknya orang yang mencari berita secara online membuat orang mengunduh aplikasi berita online. Berita sudah menjadi kebutuhan informasi harian bagi masyarakat untuk mengetahui perkembangan berbagai kejadian. Tribunnews.com termasuk salah satu portal berita online yang memiliki banyak pengunduh. Sehingga penulis melakukan penelitian analisis sentimen dengan melakukan klasifikasi sentimen ulasan menggunakan metode Naïve Bayes. Hasil pengujian beberapa model menggunakan seleksi fitur dan tidak menggunakan seleksi fitur didapat akurasi tertinggi yaitu model dengan seleksi fitur dan di *oversampling undersampling* memiliki *accuracy* sebesar 0.96; nilai *precision* sebesar 0,97; nilai *recall* sebesar 0,98; nilai *specificity* sebesar 0.88 dan AUC sebesar 0.93.

Kata Kunci: Analisis Sentimen, Tribunnews.com, Naïve Bayes, Klasifikasi, Seleksi Fitur.

**SENTIMENT ANALYSIS OF ONLINE NEWS APPLICATIONS ON
GOOGLE PLAY USING NAÏVE BAYES METHOD CASE STUDY:
TRIBUNNEWS.COM**

Aditya Samiaji

ABSTRACT

The Tribunnews application is an online news application that has been downloaded by more than 500 thousand users. The development of handheld cellular technology has had a huge impact on mankind in the use of smartphones and the internet. The number of people who search for news in online makes people download online news applications. News has become a daily information need for the public to know the development of various events. Tribunnews.com is one of the online news portals that has many downloaders. So that the author conducts sentiment analysis research by classifying sentiment using the Naïve Bayes method. The test results of several models using feature selection and not using feature selection, the highest accuracy is the model with feature selection and oversampled undersampled has an accuracy value of 0.96; precision value of 0.97; recall value of 0.98; specificity value of 0.88; and an AUC value of 0.93.

Keywords: Sentiment Analysis, Tribunnews.com, Naïve Bayes, Classifying, Feature Selection.

KATA PENGANTAR

Puji Syukur saya ucapkan kepada Allah SWT, atas karunia yang telah diberikan, sehingga penulisan Skripsi dapat selesai dikerjakan secara tepat waktu. Penelitian ini disusun sebagai persyaratan untuk menyelesaikan studi Sarjana Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta. Skripsi ini berjudul “Analisis Sentimen *Review* Aplikasi Berita *Online* Pada *Google Play* Menggunakan Metode Naïve Bayes Studi Kasus: *Tribunnews.com*”.

Penulis ingin mengungkapkan banyak terima kasih kepada pihak yang telah membantu dalam segi apapun, yang penulis tujukan kepada:

1. Ibu Dr. Ermatita, M.Kom. selaku Dekan Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.
2. Bapak Desta Sandya Prasvita, S.Komp., M.Kom. selaku Ketua Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta dan selaku dosen pembimbing kedua penulis.
3. Bapak Bayu Hananto, S.Kom, M.Kom. selaku dosen pembimbing kesatu proposal tugas akhir yang telah bersedia meluangkan waktu dan memberikan arahan.
4. Orang tua serta keluarga yang tak pernah berhenti mendoakan serta memberikan dukungan dalam berbagai hal.
5. Dina Ismayanti yang senantiasa memberikan dukungan kepada penulis selama mengerjakan penelitian.
6. Teman-teman seperjuangan Program Studi Informatika Angkatan 2018 yang telah memberikan semangat untuk menyelesaikan skripsi ini.

Penulis menyadari masih banyaknya kekurangannya dan masih jauh dari kata sempurna dalam menyusun skripsi ini. Akhir kata penulis ucapkan terimakasih.

Jakarta, 24 Juni 2022

Penulis

DAFTAR ISI

PERNYATAAN ORISINALITAS	iii
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS	iv
LEMBAR PENGESAHAN	v
ABSTRAK.....	vi
ABSTRACT.....	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xiv
DAFTAR SIMBOL.....	xv
DAFTAR LAMPIRAN.....	xv
BAB 1 PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Tujuan Penelitian.....	3
1.4. Ruang Lingkup	3
1.5. Manfaat Penelitian.....	3
1.6. Luaran yang Diharapkan	4
1.7. Sistematika Penulisan.....	4
BAB 2 TINJAUAN PUSTAKA	5
2.1. Analisis Sentimen.....	5
2.2. Berita	5
2.3. Tribunnews.com	6

2.4.	<i>Data Mining</i>	7
2.5.	<i>Text Mining</i>	7
2.6.	<i>Fleiss Kappa</i>	7
2.7.	<i>Text Pre-Processing</i>	9
2.8.	Pembobotan Kata (<i>Term Weighting</i>).....	10
2.8.1.	TF-IDF (<i>Term Frequency Inverse Document Frequency</i>).....	10
2.9.	Seleksi Fitur.....	11
2.9.1.	<i>Information Gain</i>	11
2.10.	K-Fold Cross Validation	12
2.11.	Klasifikasi.....	12
2.11.1.	<i>Naïve Bayes Classifier</i>	13
2.12.	Evaluasi Model.....	14
2.13.	Penelitian Terdahulu.....	15
BAB 3 METODOLOGI PENELITIAN		19
3.1.	Tahap Penelitian	19
3.2.	Uraian Tahapan Penelitian	20
3.2.1.	Pengumpulan Data.....	20
3.2.2.	Pelabelan Data	20
3.2.3.	Text Preprocessing.....	21
3.2.4.	<i>Term Weighting</i>	22
3.2.5.	Seleksi Fitur	23
3.2.6.	<i>K-Fold Cross Validation</i>	23
3.2.7.	Proses Klasifikasi Menggunakan <i>Naïve Bayes Classifier</i>	24
3.2.8.	Evaluasi Model	25
3.2.9.	Analisis Hasil.....	25
3.2.10.	Visualisasi.....	25

3.3. Alat Bantu Penelitian.....	26
3.3.1. Spesifikasi Perangkat Keras.....	26
3.3.2. Spesifikasi Perangkat Lunak.....	26
BAB 4 HASIL DAN PEMBAHASAN	27
4.1. Pengumpulan Data	27
4.2. Pelabelan Data	27
4.3. Praproses Data	31
4.3.1. <i>Case Folding</i>	31
4.3.2. <i>Data Cleaning</i>	32
4.3.3. Normalisasi Data.....	33
4.3.4. <i>Stopword Removal</i>	34
4.3.5. <i>Stemming</i>	35
4.3.6. <i>Tokenizing</i>	36
4.4. <i>Term Weighting</i>	36
4.5. Seleksi Fitur.....	38
4.6. Klasifikasi Naïve Bayes	44
4.6.1. Pembagian Data	44
4.6.2. Pemodelan Data	46
4.7. Evaluasi dan Analisis Hasil.....	49
4.7.1. Evaluasi Model Naïve Bayes.....	49
4.7.2. Evaluasi Model Naïve Bayes-IG	52
4.7.3. Analisis Hasil Klasifikasi.....	56
4.8. Visualisasi	57
4.8.1. Hasil Penelitian	57
4.8.2. <i>Wordcloud</i> Sentimen Positif	58
4.8.3. <i>Wordcloud</i> Sentimen Negatif.....	59

BAB 5 PENUTUP	61
5.1. Kesimpulan.....	61
5.2. Saran.....	62
DAFTAR PUSTAKA	63
RIWAYAT HIDUP.....	67
LAMPIRAN.....	68

DAFTAR TABEL

Tabel 2.1 Skala Nilai Kappa.....	8
Tabel 2.2 Review Penelitian Terdahulu	15
Tabel 3.1 Proses Pelebelan Data	21
Tabel 4.1 Contoh Data Ulasan	27
Tabel 4.2 Contoh Pelabelan Data Ulasan	28
Tabel 4.3 Gabungan Hasil Penilaian Anator	29
Tabel 4.4 Case Folding.....	31
Tabel 4.5 Data Cleaning	32
Tabel 4.6 Normalisasi Data	33
Tabel 4.7 Stopword Removal.....	34
Tabel 4.8 Stemming	35
Tabel 4.9 Tokenizing.....	36
Tabel 4.10 Penghitungan TF-IDF	37
Tabel 4.11 Bobot Data Untuk Seleksi Fitur	39
Tabel 4.12 Contoh Hasil Penghitungan Information Gain.....	41
Tabel 4.13 Hasil Pengurutan Fitur	42
Tabel 4.14 Contoh Pemilihan Fitur	43
Tabel 4.15 Jumlah fitur terseleksi	44
Tabel 4.16 Pembagian Data K-Fold CV.....	45
Tabel 4.17 Jumlah Pembagian Data	45
Tabel 4.18 Data TF-IDF.....	46
Tabel 4.19 Sampel Data	48
Tabel 4.20 Hasil Pencarian K-Fold Model Naïve Bayes	50
Tabel 4.21 Confusion Matrix Model Naïve Bayes.....	50
Tabel 4.22 Penerapan Resampling pada Data Latih.....	51
Tabel 4.23 Confusion Matrix Model Naïve Bayes Dengan Resample	52
Tabel 4.24 Tabel Hasil Pengujian Model Naïve Bayes-IG	53
Tabel 4.25 Tabel Model Naïve Bayes-IG SMOTE Random Undersampler.....	54
Tabel 4.26 Confusion Matrix Model Naïve Bayes-IG tanpa sampling.....	54
Tabel 4.27 Confusion Matrix Model Naïve Bayes-IG Dengan Resample.....	55
Tabel 4.28 Performa Setiap Model.....	56

DAFTAR GAMBAR





Gambar 3.1 Alur Penelitian	19
Gambar 3.2 Tahapan Preprocessing Data.....	22
Gambar 3.3 Tahapan Seleksi Fitur Information Gain	23
Gambar 4.1 Grafik Performa Model	58
Gambar 4.2 WordCloud Sentimen Positif.....	59
Gambar 4.3 WordCloud Sentimen Negatif	60

DAFTAR LAMPIRAN

Lampiran 1. Kamus Normalisasi Slangword	68
Lampiran 2. Daftar Stopwords	83
Lampiran 3. Similarity Index Turnitin	88
Lampiran 4. Source Code.....	100
Lampiran 5. Dataset	103

DAFTAR SIMBOL

Diagram Flowchart

No	Gambar	Nama	Keterangan
1.		<i>Input/Output</i>	Memiliki fungsi sebagai input atau hasil pada suatu langkah.
2.		<i>Flow line</i>	Digunakan untuk menentukan arah dalam alur suatu program.
3.		<i>Proses</i>	Sebagai penjelasan proses pengolahan data pada sebuah program.
4.		<i>Terminator</i>	Memiliki fungsi dalam mengawali dan mengakhiri alur.