

**ANALISIS SENTIMEN PADA DATA ULASAN MASKAPAI
PENERBANGAN DI INDONESIA MENGGUNAKAN
ALGORITMA *NAÏVE BAYES* DENGAN SELEKSI FITUR
*INFORMATION GAIN***

TUGAS AKHIR



Ajeng Rofifa Nugrafitri Akbar

1152001020

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS BAKRIE
JAKARTA
2019**

**ANALISIS SENTIMEN PADA DATA ULASAN MASKAPAI
PENERBANGAN DI INDONESIA MENGGUNAKAN
ALGORITMA *NAÏVE BAYES* DENGAN SELEKSI FITUR
*INFORMATION GAIN***

TUGAS AKHIR

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



**Ajeng Rofifa Nugrafitri Akbar
1152001020**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS BAKRIE
JAKARTA
2019**

HALAMAN PERNYATAAN ORISINALITAS

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

Nama : Ajeng Rofifa Nugrafitri Akbar

NIM : 1152001020

Tanda Tangan : 

Tanggal : 22 Agustus 2019

HALAMAN PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama : Ajeng Rofifa Nugrafitri Akbar
NIM : 1152001020
Program Studi : Informatika
Fakultas : Teknik dan Ilmu Komputer
Judul Proposal : Analisis Sentimen Pada Data Ulasan Maskapai
Penerbangan Di Indonesia Menggunakan Algoritma
Naïve Bayes Dengan Seleksi Fitur *Information Gain*

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika Fakultas Teknik dan Ilmu Komputer, Universitas Bakrie

DEWAN PEMBAHAS

Pembimbing : Yusuf Lestanto, S.T., M.Sc.

()

Penguji : Dr. Siti Rohajawati, S.Kom, M.Kom.

()

Penguji : Guson P. Kuntarto, S.T., M.Sc.

()

Ditetapkan di : Jakarta

Tanggal : 22 Agustus 2019

UNGKAPAN TERIMA KASIH

Alhamdulillah rabbil'alamin puji dan syukur atas kehadiran Allah SWT karna berkat segala rahmat dan karunia-NYA Tugas Akhir yang berjudul “Analisis Sentimen Pada Data Ulasan Maskapai Penerbangan Di Indonesia Menggunakan Algoritma *Naïve Bayes* Dengan Seleksi Fitur *Information Gain*” dapat diselesaikan. Shalawat serta salam tak lupa pula penulis haturkan kepada junjungan 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, rintangan dan hambatan dari awal hingga akhir penyusunan. Penulis mengucapkan terimakasih sebesar-besarnya kepada dosen pembimbing Bapak Yusuf Lestanto, S.T., M.Sc. yang telah meluangkan waktu serta mengerahkan tenaga dan pikirannya 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 mengungkapkan rasa terima kasih kepada:

1. Kedua orang tua dan keluarga yang telah memberikan dukungan, doa, nasihat, motivasi, semangat serta masukan untuk penulis.
2. Bapak Hoga Saragih, S.T., M.T selaku Ketua Program Studi Informatika Universitas Bakrie dan dosen pembahas yang telah memberikan saran dan perbaikan dalam penelitian ini.
3. Bapak Guson P. Kuntarto, S.T., M.Sc. dan Ibu Dr. Siti Rohajawati, S.Kom, M.Kom selaku dosen penguji yang telah memberikan saran dan perbaikan dalam penelitian ini.
4. Salma, Dona, Finda, Oca, Alham yang senantiasa memberikan dukungan, bantuan, semangat dan doa selama proses pengerjaan skripsi ini
5. Teman-teman Informatika angkatan 2015 yang telah memberikan dukungan, semangat, motivasi serta kebersamaan dalam suka maupun duka selama 4 tahun perkuliahan yang memberikan pengalaman dan pembelajaran selama masa perkuliahan di Universitas Bakrie.

6. Kakak-kakak angkatan 2014 dan adik-adik angkatan 2016, 2017, dan 2018 Informatika Universitas Bakrie yang telah memberikan dukungan, pengalaman, motivasi dan semangatnya selama masa perkuliahan dan penyusunan Tugas Akhir ini.
7. Teman-teman Universitas Bakrie angkatan 2015 yang tidak bisa disebutkan satu persatu yang telah memberikan semangat selama penyusunan Tugas Akhir ini.
8. Seluruh Dosen Program Studi Informatika yang telah memberikan ilmu dan pembelajaran serta pengalaman yang sangat bermanfaat selama masa perkuliahan di Universitas Bakrie.
9. 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 berbagai kalangan bidang pendidikan, khususnya bidang Informatika.

Jakarta, Agustus 2019

Penulis

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI

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

Nama : Ajeng Rofifa Nugrafitri Akbar
NIM : 1152001020
Program Studi : Informatika
Fakultas : Teknik dan Ilmu Komputer

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

Analisis Sentimen Pada Data Ulasan Maskapai Penerbangan Di Indonesia Menggunakan Algoritma *NaïveBayes* Dengan Seleksi Fitur *InformationGain* beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksekutif 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 : 22 Agustus 2019

Yang menyatakan



Ajeng Rofifa Nugrafitri Akbar

**ANALISIS SENTIMEN PADA DATA ULASAN MASKAPAI
PENERBANGAN DI INDONESIA MENGGUNAKAN ALGORITMA
NAÏVE BAYES DENGAN SELEKSI FITUR *INFORMATION GAIN***

Ajeng Rofifa Nugrafitri akbar

ABSTRAK

Saat ini, transportasi udara merupakan salah satu pilihan utama masyarakat dalam melakukan perjalanan lintas kota maupun antar pulau. Tingginya minat masyarakat untuk melakukan perjalanan dengan menggunakan transportasi udara membuat maskapai penerbangan berusaha untuk memberikan pelayanan terbaik agar mendapatkan banyak penumpang. Penumpang maskapai penerbangan menggunakan web *review* perjalanan pada jejaring platform *online* untuk memilih akomodasi yang akan digunakan. Banyaknya *review* yang diberikan penumpang, membutuhkan waktu yang cukup lama untuk menyimpulkan, sehingga analisis sentimen diperlukan untuk mengetahui opini penumpang maskapai penerbangan. Pada penelitian ini, data yang diambil merupakan data *review* atau ulasan penumpang pada lima maskapai penerbangan di Indonesia dengan perjalanan domestik. Dari lima maskapai penerbangan tersebut akan diambil masing-masing 250 data dengan jumlah total 1250 data ulasan yang akan diteliti. Data tersebut kemudian dilakukan *preprocessing* dan pengklasifikasian menggunakan algoritma *Naive Bayes* dengan seleksi fitur *information gain*. Hasil pengukuran performa yang diperoleh dari klasifikasi ulasan penumpang maskapai menghasilkan akurasi sebesar 93%, *precision* 95%, *recall* 96% dan *fmeasure* 95%.

Kata kunci: Analisis sentimen, Maskapai penerbangan, *Naive bayes*, *Information gain*.

SENTIMENT ANALYSIS ON DATA REVIEW AIRLINE IN INDONESIA USING NAÏVE BAYES ALGORITHM WITH FEATURE SELECTION INFORMATION GAIN

Ajeng Rofifa Nugrafitri akbar

ABSTRACT

Nowadays, air transportation is one of the top choices for people traveling across cities and between islands. The high interest of the public to travel using air transportation makes the airlines try to provide the best service to get a lot of passengers. The travel review web is used by airline passengers on online platform networks to select the accommodation to be used. The number of reviews given by passengers requires quite a long time to be concluded. Therefore, sentiment analysis is needed to determine airline passenger opinion. In this study, the data available are data reviews or passenger reviews on five airlines in Indonesia with domestic travel. This will take 250 data from each of the five airlines for a total of 1250 reviews data that will be examined. The data is then carried out preprocessing and classification using the Naïve Bayes algorithm with information gain feature selection. The results of performance measurements obtained from airline passenger review classifications produce an accuracy of 93%, precision of 95%, recall of 96% and fmeasure of 95%.

Keyword: Sentiment Analysis, Airlines, *Naive bayes*, *Information gain*.

DAFTAR ISI

HALAMAN PERNYATAAN ORISINALITAS.....	i
HALAMAN PENGESAHAN.....	ii
UNGKAPAN TERIMA KASIH.....	iii
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI.....	v
ABSTRAK.....	vi
ABSTRACT.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR RUMUS.....	xii
DAFTAR SINGKATAN.....	xiii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	3
1.3. Batasan Masalah.....	4
1.4. Tujuan Penelitian.....	4
1.5. Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA.....	6
2.1. Penelitian Terdahulu.....	6
2.2. <i>Text Mining</i>	12
2.3. Analisis Sentimen.....	12
2.3.1. Klasifikasi Analisis Sentimen.....	13
2.4. <i>Feature Selection</i>	14
2.4.1. <i>Information Gain</i>	16
2.5. <i>Naïve Bayes Classification</i>	17
2.6. Pengukuran Performa.....	19
BAB III METODOLOGI PENELITIAN.....	21
3.1. Studi Literatur.....	22
3.2. Menentukan Rumusan Masalah.....	22
3.3. Pengumpulan Data.....	22

3.4.	<i>Preprocessing</i>	24
3.4.1.	<i>Case folding</i>	25
3.4.2.	Normalisasi	25
3.4.3.	<i>Stemming</i>	26
3.4.4.	<i>Stopword</i>	27
3.4.5.	<i>Tokenizing</i>	27
3.5.	Seleksi Fitur <i>Information Gain</i>	28
3.6.	Klasifikasi Algoritma <i>Naïve Bayes</i>	28
3.7.	Evaluasi dan Hasil Klasifikasi Algoritma	29
3.8.	Penyusunan Laporan	29
BAB IV IMPLEMENTASI DAN HASIL PENELITIAN		30
4.1.	Pengumpulan Data.....	30
4.2.	Pelabelan data.....	31
4.3.	<i>Preprocessing</i> Data	32
4.4.1.	<i>Case Folding</i>	33
4.4.2.	Normalisasi Kata.....	34
4.4.3.	<i>Stemming</i>	36
4.4.4.	<i>Stopword</i>	37
4.4.	Seleksi Fitur <i>InformationGain</i>	38
4.5.	Klasifikasi Naive Bayes	41
4.6.	Pengukuran Performa	44
BAB V KESIMPULAN DAN SARAN.....		46
5.1.	Kesimpulan	46
5.2.	Saran.....	46
DAFTAR PUSTAKA		47
LAMPIRAN		52

DAFTAR TABEL

Tabel 2.1. Rangkuman Penelitian Terdahulu	9
Tabel 2.2. <i>Confusion Matrix</i>	19
Tabel 3.1. Contoh penerapan proses <i>case folding</i>	25
Tabel 3.2. Contoh penerapan proses normalisasi kata singkatan.....	26
Tabel 3.3. Contoh penerapan proses normalisasi kata bahasa Inggris	26
Tabel 3.4. Contoh penerapan proses <i>stemming</i>	26
Tabel 3.5. Contoh penerapan proses <i>stopword</i>	27
Tabel 3.6. Contoh penerapan proses <i>tokenizing</i>	27
Tabel 4.1. Data Ulasan Maskapai.....	31
Tabel 4.2. Perbandingan data <i>actual</i> dan <i>predicted</i>	44
Tabel 4.3. <i>Confussion Matrix</i> Analisis Sentimen	44

DAFTAR GAMBAR

Gambar 2.1. Teknik Klasifikasi Sentimen	14
Gambar 3.1. Kerangka Penelitian.....	21
Gambar 3.2. Hasil Proses <i>Scraping</i> Data	23
Gambar 3.3. <i>Review</i> Negatif	23
Gambar 3.4. <i>Review</i> Positif.....	24
Gambar 3.5. Tahapan <i>Preprocessing</i>	24
Gambar 4.1. Dataset Penelitian.....	30
Gambar 4.2. <i>Class Diagram</i> Proses <i>Preprocessing</i>	32
Gambar 4.3. <i>Source Code Case Folding</i>	33
Gambar 4.4. <i>Source Code</i> Normalisasi Singkatan	34
Gambar 4.5. <i>Source Code</i> Normalisasi Bahasa Inggris	35
Gambar 4.6. <i>Source Code Stemming</i>	36
Gambar 4.7. <i>Source Code Stopword</i>	37
Gambar 4.8. <i>Class Diagram</i> proses <i>Information Gain</i>	38
Gambar 4.9. <i>Source Code</i> Menghitung Nilai <i>Entropy</i>	39
Gambar 4.10. <i>Source Code</i> Menghitung Nilai <i>Information Gain</i>	40
Gambar 4.11. Potongan Hasil Perangkingan <i>Information Gain</i>	40
Gambar 4.12. <i>Class Diagram</i> proses <i>Naïve Bayes</i>	41
Gambar 4.13. <i>Source Code</i> Perhitungan <i>Prior</i>	42
Gambar 4.14. <i>Source Code</i> Perhitungan <i>Conditional Probability</i>	43
Gambar 4.15. <i>Source Code</i> Perhitungan <i>Cmap</i>	43

DAFTAR RUMUS

Persamaan 2.1. <i>Entropy</i>	16
Persamaan 2.2. Persamaan <i>Information Gain</i>	16
Persamaan 2.3. Persamaan Umum <i>Teorema Bayes</i>	17
Persamaan 2.4. Penyederhanaan Perhitungan NBC.....	18
Persamaan 2.5. Perhitungan <i>Prior</i> NBC	18
Persamaan 2.6. Perhitungan <i>Posterior</i> NBC.....	18
Persamaan 2.7. Perhitungan Probabilitas Tertinggi NBC	18
Persamaan 2.8. Penyederhanaan Perhitungan Probabilitas Tertinggi NBC.....	18
Persamaan 2.9. <i>Accuracy</i>	20
Persamaan 2.10. <i>Precision</i>	20
Persamaan 2.11. <i>Recall</i>	20
Persamaan 2.12. <i>F-Measure</i>	20

DAFTAR SINGKATAN

AUC	<i>Area Under Curve</i>
DT	<i>Decision Trees</i>
FP	<i>False Positive</i>
FN	<i>False Negative</i>
GA	<i>Genetic Algorithm</i>
IG	<i>Information Gain</i>
KNN	<i>K-Nearest Neighbor</i>
MNB	<i>Multinomial Naive Bayes</i>
NBC	<i>Naïve Bayes Classification</i>
PCO	<i>Particle Swarm Optimization</i>
PCA	<i>Principal Component Analysis</i>
RNTN	<i>Recursive Neural Tensor Network</i>
SVM	<i>Support Vector Machine</i>
SGD	<i>Stochastic Gradient Descent</i>
TP	<i>True Positive</i>
TN	<i>True Negative</i>