

**PERBANDINGAN KINERJA ALGORITMA  
MULTINOMIAL DAN COMPLEMENT NAÏVE BAYES  
UNTUK ANALISIS SENTIMEN ULASAN APLIKASI  
TRAVELOKA DI GOOGLE PLAY STORE**

**SKRIPSI**



Oleh :

**MOCHAMMAD ARYA SALSABILA**

**NPM. 19081010001**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"  
JAWA TIMUR**

**2023**

**PERBANDINGAN KINERJA ALGORITMA  
MULTINOMIAL DAN COMPLEMENT NAÏVE BAYES  
UNTUK ANALISIS SENTIMEN ULASAN APLIKASI  
TRAVELOKA DI GOOGLE PLAY STORE**

**SKRIPSI**

Diajukan Untuk Memenuhi Sebagai Persyaratan Dalam Menempuh Gelar Sarjana  
Komputer Program Studi Informatika



Oleh :

**MOCHAMMAD ARYA SALSABILA**

**NPM. 19081010001**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"  
JAWA TIMUR  
2023**

## LEMBAR PENGESAHAN SKRIPSI

**Judul : PERBANDINGAN KINERJA ALGORITMA MULTINOMIAL  
DAN COMPLEMENT NAÏVE BAYES UNTUK ANALISIS  
SENTIMEN ULASAN APLIKASI TRAVELOKA DI GOOGLE  
PLAY STORE**

**Oleh : MOCHAMMAD ARYA SALSABILA**

**NPM : 19081010001**

**Telah Diseminarkan Dalam Ujian Skripsi Pada :**

**Hari Senin, Tanggal 22 Mei 2023**

**Mengetahui**

**Dosen Pembimbing**

**Dosen Penguji**

1.

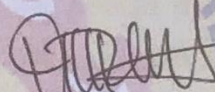


**Eva Yulia Puspaningrum, S.Kom.,**

**M.Kom.**

**NIP. 19890705 2021212 002**

1.

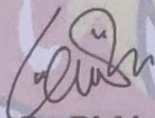


**Henni Endah Wahanani, S.T.,**

**M.Kom.**

**NIP. 19780922 2021212 005**

2.



**Agung Mustika Rizki, S.Kom.,**

**M.Kom.**

**NIP. 19930725 202203 1008**

2.



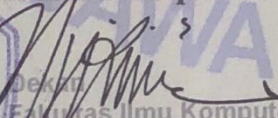
**Chrystia Ali Putra, S.Kom., M.T.**

**NIP. 19861008 2021211 001**

**Menyetujui**

**Dekan**

**Fakultas Ilmu Komputer**

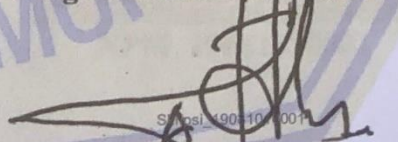


**Dr. H. Novrina Hendrasarie, M.T.**

**NIP. 19681126 199403 2 001**

**Koordinator**

**Program Studi Informatika**



**Fetty Tri Anggraeny, S.Kom.,**

**M.Kom.**

**NIP. 19820211 2021212 005**



## SURAT PERNYATAAN ANTI PLAGIAT

Saya mahasiswa Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur, yang bertanda tangan di bawah ini:

Nama : Mochammad Arya Salsabila

NPM : 19081010001

Dengan ini menyatakan bahwa judul skripsi atau tugas akhir yang saya ajukan dan kerjakan, yang berjudul:

**“PERBANDINGAN KINERJA ALGORITMA MULTINOMIAL DAN  
COMPLEMENT NAÏVE BAYES UNTUK ANALISIS SENTIMEN  
ULASAN APLIKASI TRAVELOKA DI GOOGLE PLAY STORE”**

Bukan merupakan plagiat dari skripsi atau tugas akhir maupun penelitian orang lain dan juga bukan merupakan produk atau software yang saya beli dari pihak lain. Saya juga menyatakan bahwa skripsi ini adalah pekerjaan saya sendiri, kecuali yang dinyatakan dalam daftar pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional “Veteran” Jawa Timur maupun institusi pendidikan lainnya.

Jika ternyata kemudian hari pernyataan ini terbukti tidak benar, maka saya siap menerima segala konsekuensinya.

Suarabaya, 22 Mei 2023

Hormat Saya,



**Mochammad Arya Salsabila**  
**NPM. 19081010001**

# PERBANDINGAN KINERJA ALGORITMA MULTINOMIAL DAN COMPLEMENT NAÏVE BAYES UNTUK ANALISIS SENTIMEN ULASAN APLIKASI TRAVELOKA DI GOOGLE PLAY STORE

Nama Mahasiswa : Mochammad Arya Salsabila

NPM : 19081010001

Program Studi : Informatika

Dosen Pembimbing : Eva Yulia Puspaningrum, S.Kom., M.Kom.

Agung Mustika Rizki, S.Kom., M.Kom.

## ABSTRAK

Traveloka sebagai aplikasi perjalanan *online* terpopuler di Indonesia tentu saja menerima berbagai kritik dan saran dari para penggunanya. Hal ini terlihat dari banyaknya ulasan pengguna aplikasi Traveloka di Google Play Store. Penelitian ini mengimplementasikan algoritma *Multinomial Naïve Bayes* dan *Complement Naïve Bayes* untuk melakukan analisis sentimen pada 121.378 ulasan pengguna aplikasi Traveloka. Tujuan dari penelitian ini adalah untuk mengetahui kinerja dari kedua algoritma dalam mengklasifikasikan ulasan ke dalam kategori sentimen positif atau negatif. *Multinomial Naïve Bayes* bekerja dengan cara menghitung probabilitas kemunculan kata dalam dokumen, sedangkan *Complement Naïve Bayes* bekerja dengan cara menghitung probabilitas kemunculan kata diseluruh kelas kecuali kelas yang sedang diamati. Hasil pengujian menunjukkan bahwa rasio data *train* dan *test* 70%:30% memberikan kinerja terbaik untuk kedua algoritma yang diuji, dan *Complement Naïve Bayes* mengungguli *Multinomial Naïve Bayes* dalam hal *accuracy*, dengan perolehan nilai *accuracy* berturut-turut sebesar 86,46% dan 84,75%. Selain itu, penggunaan parameter *norm True* pada *Complement Naïve Bayes* memberikan kinerja yang lebih baik daripada *norm False*, dengan perolehan nilai *accuracy* masing-masing sebesar 88,22% dan 86,46%.

**Kata Kunci** : Analisis Sentimen, *Multinomial Naïve Bayes*, *Complement Naïve Bayes*, Traveloka

## **KATA PENGANTAR**

Puji syukur penulis panjatkan kehadirat Allah Subhaanahu Wa Ta'aalaa, atas segala berkat dan rahmat-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "Perbandingan Kinerja Algoritma Multinomial Dan Complement Naïve Bayes Untuk Analisis Sentimen Ulasan Aplikasi Traveloka Di Google Play Store".

Laporan skripsi ini dibuat dalam rangka untuk memenuhi mata kuliah skripsi dan sebagai salah satu persyaratan untuk menyelesaikan pendidikan Strata Satu (S1) di Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur.

Penulis menyadari bahwa masih banyak kekurangan dalam penulisan skripsi ini. Oleh karena itu, penulis berharap adanya saran dan kritik yang membangun. Semoga skripsi ini dapat memberikan manfaat bagi semua pihak, baik bagi pembaca maupun penelitian selanjutnya.

Suarabaya, 22 Mei 2023

Penulis

## UCAPAN TERIMA KASIH

Penulisan laporan skripsi ini tentu tidak terlepas dari bantuan, motivasi, dan dukungan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih kepada:

1. Bapak Prof. Dr. Ir. Akhmad Fauzi, M.MT., selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Ibu Dr. Ir. Novirina Hendrasarie, M.T., selaku Dekan Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Fetty Tri Anggraeny, S.Kom., M.Kom., selaku Koordinator Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Ibu Eva Yulia Puspaningrum, S.Kom., M.Kom., selaku Dosen Wali sekaligus Dosen Pembimbing I yang dengan sabar dan ikhlas meluangkan waktu, tenaga, serta pikiran untuk membimbing dan mengarahkan penulis selama masa perkuliahan dan proses penyelesaian skripsi ini.
5. Bapak Agung Mustika Rizki, S.Kom., M.Kom., selaku Dosen Pembimbing II yang telah memberikan arahan, dukungan, serta saran kepada penulis sehingga penulis dapat menyelesaikan penyusunan skripsi ini.
6. Seluruh Dosen Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur yang telah memberikan ilmunya kepada penulis selama perkuliahan.
7. Kedua orang tua penulis, Bapak Samiarso Eddy Prasetyo, S.T., dan Ibu Erna Wirdaningsih, S.E., serta adik yang selalu memberikan motivasi dan doa untuk penulis.
8. Dimas Seno Herlambang, Nadhif Mahardika Awandi, Fahri Izzuddin Zulkarnaen, Puteri Aulia Fahlia, dan teman-teman seperjuangan dari program studi Informatika angkatan 2019 yang telah memberikan bantuan dan dukungan selama penulis menempuh perkuliahan dan menyelesaikan skripsi.

Penulis hanya bisa berharap, semoga Allah Subhaanahu Wa Ta’aalaa senantiasa memberikan perlindungan dan balasan yang lebih di kemudian hari.

## DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI .....	ii
SURAT PERNYATAAN ANTI PLAGIAT .....	iii
ABSTRAK .....	iv
KATA PENGANTAR .....	v
UCAPAN TERIMA KASIH.....	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR .....	x
DAFTAR TABEL.....	xii
DAFTAR PSEUDOCODE .....	xiii
DAFTAR KODE PROGRAM.....	xiv
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Tujuan .....	4
1.4 Manfaat .....	4
1.5 Batasan Masalah.....	4
BAB II TINJAUAN PUSTAKA.....	6
2.1 Penelitian Terdahulu .....	6
2.2 Traveloka.....	7
2.3 Analisis Sentimen .....	7
2.4 Text Preprocessing .....	8
2.4.1 Case Folding .....	8
2.4.2 Cleansing.....	8
2.4.3 Tokenizing.....	9
2.4.4 Normalization.....	9
2.4.5 Stopword Removal.....	10
2.4.6 Stemming .....	10
2.5 Valence Aware Dictionary and Sentiment Reasoner .....	11
2.6 Term Frequency-Inverse Document Frequency.....	11
2.7 Naïve Bayes .....	12
2.7.1 Multinomial Naïve Bayes .....	14
2.7.2 Complement Naïve Bayes.....	14



2.8 Confusion Matrix .....	15
<b>BAB III METODOLOGI PENELITIAN.....</b>	<b>18</b>
3.1 Tahapan Penelitian .....	18
3.2 Studi Literatur .....	18
3.3 Pengumpulan Data .....	19
3.4 Pengolahan Data.....	20
3.4.1 Text Preprocessing .....	21
3.4.2 Labeling VADER.....	26
3.5 Visualisasi Data.....	26
3.6 Pembagian Data .....	27
3.7 Ekstraksi Fitur .....	27
3.8 Klasifikasi .....	29
3.8.1 Multinomial Naïve Bayes .....	30
3.8.2 Complement Naïve Bayes.....	33
3.9 Evaluasi .....	36
3.10 Skenario Pengujian.....	36
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>37</b>
4.1 Lingkungan Penelitian .....	37
4.1.1 Spesifikasi Perangkat Keras .....	37
4.1.2 Spesifikasi Perangkat Lunak.....	37
4.2 Pengumpulan Data .....	38
4.3 Pengolahan Data.....	41
4.3.1 Text Preprocessing .....	41
4.3.2 Labeling VADER.....	50
4.4 Visualisasi Data.....	53
4.5 Pembagian Data .....	58
4.6 Ekstraksi Fitur .....	60
4.7 Klasifikasi .....	62
4.7.1 Multinomial Naïve Bayes .....	62
4.7.2 Complement Naïve Bayes.....	65
4.8 Evaluasi .....	68
4.9 Hasil Pengujian .....	69
4.9.1 Skenario 1 .....	70
4.9.2 Skenario 2 .....	71

4.9.3 Skenario 3 .....	72
4.9.4 Skenario 4 .....	74
4.9.5 Skenario 5 .....	75
4.9.6 Skenario 6 .....	76
4.10 Perbandingan Hasil Pengujian .....	78
BAB V PENUTUP.....	80
5.1 Kesimpulan .....	80
5.2 Saran.....	81
DAFTAR PUSTAKA .....	82
BIODATA PENULIS .....	86

## DAFTAR GAMBAR

Gambar 3.1 Tahapan Penelitian .....	18
Gambar 3.2 Pengumpulan Data .....	19
Gambar 3.3 Scraping Ulasan Pengguna.....	20
Gambar 3.4 Pengolahan Data.....	21
Gambar 3.5 Case Folding.....	22
Gambar 3.6 Cleansing .....	22
Gambar 3.7 Tokenizing.....	23
Gambar 3.8 Normalization.....	23
Gambar 3.9 Stopword Removal.....	24
Gambar 3.10 Stemming .....	25
Gambar 3.11 Pembagian Data.....	27
Gambar 3.12 Ekstraksi Fitur .....	28
Gambar 3.13 Klasifikasi.....	30
Gambar 4.1 Proses Scraping Ulasan .....	40
Gambar 4.2 Ulasan Hasil Scraping .....	40
Gambar 4.3 Ulasan Hasil Case Folding .....	42
Gambar 4.4 Ulasan Hasil Cleansing .....	43
Gambar 4.5 Ulasan Hasil Tokenizing .....	44
Gambar 4.6 Ulasan Hasil Normalization .....	45
Gambar 4.7 Ulasan Hasil Stopword Removal .....	47
Gambar 4.8 Ulasan Hasil Stemming.....	49
Gambar 4.9 Ulasan Hasil Text Preprocessing.....	50
Gambar 4.10 Menerjemahkan Ulasan ke Bahasa Inggris .....	51
Gambar 4.11 Ulasan Hasil Pengolahan Data .....	53
Gambar 4.12 Diagram Batang Distribusi Sentimen Per Tahun .....	54
Gambar 4.13 Wordcloud Sentimen Negatif Tahun 2018.....	56
Gambar 4.14 Wordcloud Sentimen Positif Tahun 2018 .....	56
Gambar 4.15 Wordcloud Sentimen Negatif Tahun 2019.....	56
Gambar 4.16 Wordcloud Sentimen Negatif Tahun 2020.....	57
Gambar 4.17 Wordcloud Sentimen Positif Tahun 2019 .....	56
Gambar 4.18 Wordcloud Sentimen Positif Tahun 2020 .....	57

Gambar 4.19 Wordcloud Sentimen Negatif Tahun 2021.....	57
Gambar 4.20 Wordcloud Sentimen Positif Tahun 2021 .....	57
Gambar 4.21 Wordcloud Sentimen Negatif Tahun 2022.....	57
Gambar 4.22 Wordcloud Sentimen Positif Tahun 2022 .....	57
Gambar 4.23 Hasil Label Encoding pada Kolom Label .....	58
Gambar 4.24 Jumlah Data Train dan Data Test.....	59
Gambar 4.25 Document-Term Matrix Hasil TF-IDF .....	61
Gambar 4.26 Subset Document-Term Matrix Hasil TF-IDF.....	62
Gambar 4.27 Confusion Matrix Skenario 1 .....	70
Gambar 4.28 Classification Report Skenario 1.....	70
Gambar 4.29 Confusion Matrix Skenario 2 .....	71
Gambar 4.30 Classification Report Skenario 2.....	72
Gambar 4.31 Confusion Matrix Skenario 3 .....	73
Gambar 4.32 Classification Report Skenario 3.....	73
Gambar 4.33 Confusion Matrix Skenario 4 .....	74
Gambar 4.34 Classification Report Skenario 4.....	74
Gambar 4.35 Confusion Matrix Skenario 5 .....	75
Gambar 4.36 Classification Report Skenario 5.....	76
Gambar 4.37 Confusion Matrix Skenario 6 .....	77
Gambar 4.38 Classification Report Skenario 6.....	77

## DAFTAR TABEL

Tabel 2.1 Data Teks Sebelum dan Sesudah Case Folding .....	8
Tabel 2.2 Data Teks Sebelum dan Sesudah Cleansing .....	9
Tabel 2.3 Data Teks Sebelum dan Sesudah Tokenizing .....	9
Tabel 2.4 Data Teks Sebelum dan Sesudah Normalization .....	9
Tabel 2.5 Data Teks Sebelum dan Sesudah Stopword Removal .....	10
Tabel 2.6 Data Teks Sebelum dan Sesudah Stemming.....	10
Tabel 2.7 Confusion Matrix untuk Klasifikasi Dua Kelas.....	16
Tabel 3.1 Contoh Hasil Text Preprocessing.....	25
Tabel 3.2 Contoh Koleksi Dokumen.....	28
Tabel 3.3 Contoh Document-Term Matrix Hasil TF-IDF .....	29
Tabel 3.4 Contoh Perhitungan Feature Probability .....	32
Tabel 3.5 Contoh Vektor Bobot TF-IDF untuk Perhitungan MNB.....	32
Tabel 3.6 Contoh Perhitungan Complementary Feature Probability .....	34
Tabel 3.7 Contoh Vektor Bobot TF-IDF untuk Perhitungan CNB.....	35
Tabel 3.8 Skenario Pengujian .....	36
Tabel 4.1 Spesifikasi Perangkat Keras yang Digunakan .....	37
Tabel 4.2 Spesifikasi Perangkat Lunak yang Digunakan.....	37
Tabel 4.3 Isi File Normalization Resource .....	44
Tabel 4.4 Daftar Kata Stopword pada Sastrawi .....	46
Tabel 4.5 Isi File Stemming Resource .....	48
Tabel 4.6 Jumlah Sentimen Berdasarkan Tahun.....	54
Tabel 4.7 Ringkasan Hasil Pengujian untuk Semua Skenario .....	78

## DAFTAR PSEUDOCODE

Pseudocode 4.1 Pengumpulan Data .....	38
Pseudocode 4.2 Fungsi Case Folding .....	41
Pseudocode 4.3 Fungsi Cleansing.....	42
Pseudocode 4.4 Fungsi Tokenizing.....	43
Pseudocode 4.5 Fungsi Normalization.....	45
Pseudocode 4.6 Fungsi Stopword Removal.....	46
Pseudocode 4.7 Fungsi Stemming .....	47
Pseudocode 4.8 Fungsi Apply Stemming .....	48
Pseudocode 4.9 Labeling Otomatis dengan VADER .....	51
Pseudocode 4.10 Fungsi Plot Sentiment Distribution.....	53
Pseudocode 4.11 Fungsi Plot Sentiment Wordcloud .....	55
Pseudocode 4.12 Label Encoding .....	58
Pseudocode 4.13 Pembagian Data Train dan Data Test .....	59
Pseudocode 4.14 TF-IDF .....	60
Pseudocode 4.15 Multinomial Naïve Bayes .....	63
Pseudocode 4.16 Complement Naïve Bayes.....	65
Pseudocode 4.17 Evaluasi Model .....	68

## DAFTAR KODE PROGRAM

Kode Program 4.1 Pengumpulan Data.....	39
Kode Program 4.2 Menghapus Ulasan dengan Nilai NaN .....	41
Kode Program 4.3 Mengaktifkan Progress Bar dengan Tqdm .....	41
Kode Program 4.4 Fungsi Case Folding .....	42
Kode Program 4.5 Fungsi Cleansing .....	42
Kode Program 4.6 Fungsi Tokenizing .....	43
Kode Program 4.7 Fungsi Normalization .....	45
Kode Program 4.8 Fungsi Stopword Removal .....	46
Kode Program 4.9 Fungsi Stemming .....	47
Kode Program 4.10 Fungsi Apply Stemming .....	49
Kode Program 4.11 Memisahkan Data Hasil Normalization.....	51
Kode Program 4.12 Labeling Otomatis dengan Vader .....	52
Kode Program 4. 13 Menggabungkan Hasil Pengolahan Data.....	52
Kode Program 4.14 Fungsi Plot Sentiment Distribution .....	54
Kode Program 4.15 Fungsi Plot Sentiment Wordcloud.....	55
Kode Program 4.16 Label Encoding.....	58
Kode Program 4.17 Pembagian Data Train dan Data Test .....	59
Kode Program 4.18 Ekstraksi Fitur dengan TF-IDF.....	60
Kode Program 4.19 Menampilkan Document-Term Matrix.....	61
Kode Program 4.20 Menyimpan Hasil Esktraksi Fitur .....	62
Kode Program 4.21 Class Multinomial Naïve Bayes .....	64
Kode Program 4. 22 Class Complement Naïve Bayes.....	66
Kode Program 4.23 Klasifikasi .....	68
Kode Program 4.24 Fungsi Evaluate Model .....	69
Kode Program 4.25 Evaluasi.....	69