

**ANALISIS SENTIMEN KEPUASAN PENGGUNA APLIKASI  
OYO MENGGUNAKAN MULTINOMIAL NAIVE BAYES  
DENGAN SELEKSI FITUR CHI-SQUARE**

**SKRIPSI**

**Diajukan untuk memenuhi salah satu syarat kelulusan  
di Program Studi Sains Data**



**Disusun Oleh:**

**Rizky Aziz**

**20083010007**

**PROGRAM STUDI SAINS DATA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAWA TIMUR  
SURABAYA  
2024**

**LEMBAR PENGESAHAN**

**ANALISIS SENTIMEN KEPUASAN PENGGUNA APLIKASI OYO  
MENGUNAKAN MULTINOMIAL NAIVE BAYES DENGAN SELEKSI  
FITUR CHI-SQUARE**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Sains Data  
pada : Selasa, 14 Mei 2024

**Program Studi S-1 Sains Data  
Fakultas Ilmu Komputer  
Universitas Pembangunan Nasional Veteran Jawa Timur  
Surabaya**

Oleh :

Rizky Aziz

NPM. 20083010007

Disetujui oleh Tim Penguji Skripsi :

Penguji 1

Penguji 2

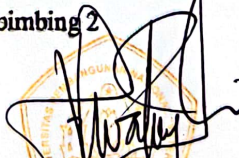
  
**Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.**  
NIP. 198303102021211006

  
**Kartika Maulida Hindrayani S.Kom., M.Kom**  
NIP. 199209092022032009

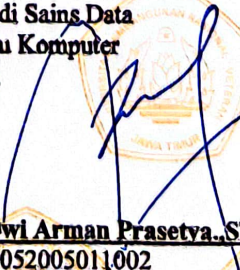
Pembimbing 1

Pembimbing 2

  
**Tresna Maulana Fabrudin, S.ST., M.T**  
NIP. 199305012022031007

  
**Wahyu Syaifullah Jahharis Saputra, S.Kom., M.Kom**  
NIP. 198608252021211003

  
**Fakultas Ilmu Komputer  
Dekan,**

**Mengetahui,  
Program Studi Sains Data  
Fakultas Ilmu Komputer  
Koordinator,**  


**Prof. Dr. Ir. Novirina Hendrasarie, MT**  
NIP. 196811261994032001

**Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU.**  
NIP. 198012052005011002

Surabaya, Mei, 2024

## SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Rizky Aziz  
NPM : 20083010007  
Program Studi : Sains Data

Menyatakan bahwa judul Skripsi / Tugas Akhir sebagai berikut:

**Analisis Sentimen Kepuasan Pengguna Aplikasi OYO  
Menggunakan Multinomial Naive Bayes dengan Seleksi Fitur  
Chi-square**

Bukan merupakan plagiat dari Skripsi/ Tugas Akhir/ Penelitian orang lain dan juga bukan merupakan produk/ *software*/ hasil karya yang saya beli dari orang lain

Saya juga menyatakan bahwa Skripsi/ Tugas Akhir 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 di institusi pendidikan lain.

Jika ternyata dikemudian hari pernyataan ini terbukti tidak benar, maka Saya bertanggung jawab penuh dan siap menerima segala konsekuensi, termasuk pembatalan ijazah dikemudian hari

Surabaya, 14 Mei 2024

Hormat Saya



Rizky Aziz  
NPM. 20083010007

## **ABSTRAK**

# **ANALISIS SENTIMEN KEPUASAN PENGGUNA APLIKASI OYO MENGGUNAKAN MULTINOMIAL NAIVE BAYES DENGAN SELEKSI FITUR CHI-SQUARE**

**Nama Mahasiswa / NPM** : Rizky Aziz / 20083010007  
**Program Studi** : Sains Data, FASILKOM, UPN Veteran Jatim  
**Dosen Pembimbing 1** : Tresna Maulana Fahrudin, S.ST., M.T  
**Dosen Pembimbing 2** : Wahyu Syaifullah Jauharis Saputra., S.Kom., M.Kom

### **Abstrak**

Perkembangan teknologi informasi dan aplikasi seluler telah membawa perubahan signifikan dalam industri perhotelan. Namun, pengguna sering menghadapi tantangan dalam memilih akomodasi yang sesuai. OYO, sebagai penyedia layanan akomodasi terkemuka, berkomitmen untuk memberikan pengalaman menginap yang unggul. Oleh karena itu, penelitian ini bertujuan menganalisis sentimen ulasan pengguna aplikasi OYO di Google Playstore menggunakan Multinomial Naïve Bayes dan seleksi fitur Chi-Square untuk mengeksplorasi hubungan antara faktor-faktor yang mempengaruhi kepuasan pengguna. Langkah penelitian meliputi pengumpulan data, pra-pemrosesan, pelabelan, dan pembagian data, dilanjutkan dengan pembobotan TF-IDF dan seleksi fitur Chi-Square. Hasil menunjukkan akurasi model meningkat dari 71% menjadi 79% setelah penerapan seleksi fitur Chi-Square, menunjukkan peningkatan evaluasi model klasifikasi sebesar 7%-10%. Penelitian ini memberikan dasar untuk pengembangan lebih lanjut dalam meningkatkan kualitas layanan OYO dan memahami pengalaman serta preferensi pengguna dengan lebih baik.

**Kata kunci:** Analisis Sentimen, Aplikasi OYO, Google Playstore, Multinomial Naïve Bayes, Seleksi Fitur Chi-Square.

## ***ABSTRACT***

### **SENTIMENT ANALYSIS OF USER SATISFACTION WITH THE OYO APPLICATION USING MULTINOMIAL NAIVE BAYES WITH CHI- SQUARE FEATURE SELECTION**

**Student Name / NPM** : Rizky Aziz / 20083010007  
**Study Program** : Sains Data, FASILKOM,UPN Veteran Jatim  
**Advisor 1** : Tresna Maulana Fahrudin, S.ST., M.T  
**Advisor 2** : Wahyu Syaifullah Jauharis Saputra., S.Kom., M.Kom

#### ***Abstract***

The advancement of information technology and mobile applications has significantly transformed the hospitality industry. However, users often face challenges in choosing accommodations that suit their needs. OYO, as a leading accommodation service provider, is committed to offering superior stay experiences. Therefore, this study aims to analyze user reviews of the OYO app on Google Playstore using Multinomial Naïve Bayes and Chi-Square feature selection to explore the factors affecting user satisfaction. The research steps include data collection, preprocessing, labeling, and data splitting, followed by TF-IDF weighting and Chi-Square feature selection. The results show that the model accuracy increased from 71% to 79% after applying Chi-Square feature selection, indicating an improvement in model evaluation by 7%-10%. This study provides a foundation for further development in enhancing the quality of OYO's services and better understanding user experiences and preferences.

***Keywords: Sentiment Analysis, OYO App, Google Playstore, Multinomial Naïve Bayes, Chi-Square Feature Selection***

## KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kepada Allah SWT, sumber segala rahmat dan karunia-Nya, serta sholawat dan salam yang tak terhingga penulis sampaikan kepada junjungan kita, Nabi Besar Muhammad SAW. Berkat limpahan kasih sayang-Nya, penulis berhasil menyelesaikan skripsi ini yang berjudul "Analisis Sentimen Kepuasan Pengguna Oyo Menggunakan Multinomial Naive Bayes Dengan Seleksi Fitur Chi-Square", sebagai salah satu persyaratan untuk menyelesaikan Program Studi S1/Sains Data di Universitas Pembangunan Nasional "Veteran" Jawa Timur.

Pencapaian penulisan skripsi ini tidak terlepas dari dukungan dan bantuan yang diberikan oleh berbagai pihak. Oleh karena itu, penulis ingin mengungkapkan rasa terima kasih yang tulus kepada semua yang telah membantu dalam kelancaran penulisan tugas akhir ini. Dengan kesempatan ini, penulis ingin menyampaikan penghargaan dan ucapan terima kasih kepada:

1. Kedua orang tua dan keluarga yang selalu mendoakan dan memberikan motivasi;
2. Prof. Dr. Ir. Akhmad Fauzi, M.MT., IPU selaku Rektor Universitas Pembangunan Nasional "Veteran" Jawa Timur;
3. Ibu Dr. Ir. Ni Ketut Sari, MT selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional "Veteran" Jawa Timur;
4. Bapak Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU selaku Koordinator Program Studi Sains Data Universitas Pembangunan Nasional "Veteran" Jawa Timur;
5. Tresna Maulana Fahrudin, S.ST., M.T Selaku Dosen pembimbing 1;
6. Wahyu Syaifullah JS., S.Kom., M.Kom Selaku Dosen Pembimbing 2;
7. Bapak dan Ibu dosen Program Studi Sains Data UPN "Vetran" Jawa Timur yang sudah berkenan untuk memberikan waktu dan berkontribusi dalam penelitian ini.
8. Teman-teman program studi Sains Data yang tidak dapat disebutkan satu-persatu yang telah membantu saya dalam menyelesaikan skripsi ini;

Semoga Tuhan memberkati usaha ini dengan hasil yang terbaik. Saya mengakui bahwa skripsi ini memiliki kekurangan tertentu yang perlu diperbaiki. Semoga hasil penelitian ini memberikan manfaat yang besar bagi semua, terutama dalam kemajuan ilmu pengetahuan di bidang Sains Data.

Surabaya, 3 Mei 2024

Penulis

## DAFTAR ISI

LEMBAR PENGESAHAN.....	ii
ABSTRAK .....	iv
<i>ABSTRACT</i> .....	v
KATA PENGANTAR .....	vi
DAFTAR ISI .....	viii
DAFTAR GAMBAR .....	x
DAFTAR TABEL .....	xii
BAB I PENDAHULUAN .....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	4
1.3. Batasan Masalah.....	4
1.4. Tujuan Penelitian.....	5
1.5. Manfaat Penelitian.....	5
BAB II TINJAUAN PUSTAKA .....	7
2.1. Dasar Teori.....	7
2.1.1. Analisis Sentimen.....	7
2.1.2. OYO.....	8
2.1.3. <i>Google Playstore</i> .....	8
2.1.4. <i>Web Crawling</i> .....	8
2.1.5. <i>Text preprocessing</i> .....	9
2.1.7. TF-IDF ( <i>Term Frequency – Inverse Document Frequency</i> ).....	11
2.2. Penelitian Terdahulu.....	16
BAB III METODOLOGI PENELITIAN.....	25
3.1. Variabel Penelitian dan Sumber Data.....	25
3.2. Langkah Analisis.....	27
3.2.1 Studi Literatur .....	27
3.2.2 <i>Crawling</i> Ulasan OYO pada Google Playstore.....	28
3.2.3 <i>Prepocosing</i> .....	28
3.2.4 <i>Labeling</i> .....	31
3.2.5 Pembobotan TF-IDF.....	32
3.2.6 Seleksi Fitur Chi-Square .....	34



3.2.7	Multinomial Naïve Bayes .....	36
3.2.8	Evaluasi Model.....	38
3.3.	Diagram Alir Penelitian.....	38
3.4.	Jadwal Penelitian.....	40
BAB IV HASIL DAN PEMBAHASAN .....		41
4.1.	Pengumpulan Data .....	41
4.2.	<i>Preprocessing</i> Data.....	42
4.2.1	<i>Clean Text</i> .....	43
4.2.2	<i>Case Folding</i> .....	44
4.2.3	<i>Tokenize</i> .....	45
4.2.4	<i>Remove Slang</i> .....	46
4.2.5	<i>Stopword</i> .....	47
4.2.6	<i>Stemming</i> .....	48
4.3.	<i>Labeling</i> .....	49
4.4.	Pembobotan TF-IDF.....	49
4.5.	Seleksi Fitur Chi-Square.....	51
4.5.1	Seleksi Fitur (Split data 70:30).....	52
4.5.2	Seleksi Fitur (Split data 80:20).....	53
4.5.3	Seleksi Fitur (Split data 90:10).....	53
4.6.	Model Klasifikasi Multinomial Naïve Bayes .....	57
4.7.	Evaluasi Model.....	59
4.7.1	Klasifikasi Multinomial Naïve Bayes Classifier tanpa Seleksi Fitur Chi-Square .....	60
4.7.2	Klasifikasi Multinomial Naïve Bayes Classifier dengan Seleksi Fitur Chi-Square .....	63
4.8.	Visualisasi Data.....	70
BAB V PENUTUP .....		73
5.1	Kesimpulan .....	73
5.2	Saran .....	74
DAFTAR PUSTAKA .....		75
LAMPIRAN .....		78
BIODATA PENULIS .....		92

## DAFTAR GAMBAR

Gambar 3. 1. <i>Flowchart Preprocessing</i> .....	28
Gambar 3. 2. Flowchat Diagram Alir .....	39
Gambar 3. 3. Diagram Sistem penelitian .....	40
Gambar 4. 1. Hasil <i>Scraping/Crawling</i> Data .....	42
Gambar 4. 2. <i>Import</i> Dataset OYO .....	43
Gambar 4. 3. Hasil <i>Cleaning Text</i> .....	44
Gambar 4. 4. Hasil <i>Case Folding</i> .....	45
Gambar 4. 5. Hasil <i>Tokenize</i> .....	45
Gambar 4. 6. Kamus <i>Slang Word</i> .....	46
Gambar 4. 7. Hasil <i>Romove Slang</i> .....	47
Gambar 4. 8. Hasil <i>Stopword</i> .....	48
Gambar 4. 9. Hasil <i>Stemming</i> .....	48
Gambar 4. 10. Hasil Proses <i>Labeling</i> .....	49
Gambar 4. 11. Hasil <i>Labeling</i> Data .....	49
Gambar 4. 12. Jumlah Fitur Perkata .....	50
Gambar 4. 13. Hasil Dari Pembobotan TF-IDF .....	51
Gambar 4. 14. Hasil Nilai K terbaik pada Seleksi Fitur Chi -Square (Split Data 70:30) .....	52
Gambar 4. 15. Hasil Nilai K terbaik pada Seleksi Fitur Chi -Square (Split Data 80:20) .....	53
Gambar 4. 16. Hasil Nilai K terbaik pada Seleksi Fitur Chi -Square (Split Data 90:10) .....	53
Gambar 4. 17. Hasil pengurutan lima fitur teratas dan terbawah .....	54
Gambar 4. 18. Hasil Seleksi Fitur (Split Data 70:30) .....	55
Gambar 4. 19. Hasil Seleksi Fitur (Split Data 80:20) .....	56
Gambar 4. 20. Hasil Seleksi Fitur (Split Data 90:10) .....	57
Gambar 4. 21. <i>Classification Report</i> sekenario pengujian ke-1 .....	60
Gambar 4. 22. <i>Classification Report</i> Sekenario pengujian ke-2 .....	61
Gambar 4. 23. <i>Classification Report</i> Sekenario pengujian ke-3 .....	62
Gambar 4. 24. <i>Classification Report</i> Sekenario Pengujian ke-4 .....	64

Gambar 4. 25. <i>Classification Report</i> Sekenario Pengujian ke-5 .....	65
Gambar 4. 26. <i>Classification Report</i> Sekenario Pengujian ke-6 .....	66
Gambar 4. 27. Visualisasi Ulasan Positif Dan Negatif .....	71
Gambar 4. 28. Visualisasi Pie Chart .....	71

## DAFTAR TABEL

Tabel 2. 2. <i>Confusion Matrix</i> .....	15
Tabel 2. 1. Tabel studi literatur yang relevan dalam Skripsi.....	16
Tabel 3. 1. Tabel Struktur Data Kolom Awal .....	25
Tabel 3. 2. Tabel Struktur Data Lanjutan.....	26
Tabel 3. 3. <i>Clean Text</i> .....	29
Tabel 3. 4. <i>Case Folding</i> .....	29
Tabel 3. 5. <i>Tokenize</i> .....	30
Tabel 3. 6. <i>Remove Slang</i> .....	30
Tabel 3. 7. <i>Tokenize</i> .....	31
Tabel 3. 8. <i>Stemming</i> .....	31
Tabel 3. 9. Contoh Labelisasi data.....	32
Tabel 3. 10. Contoh Data Penelitian .....	33
Tabel 3. 11. TF-IDF Pada Setiap Term Didokumen.....	34
Tabel 3. 12. Contoh Data Penelitian .....	35
Tabel 3. 13. Frekuensi Kemuncilan Kata.....	35
Tabel 3. 14. Hasil Seleksi Fitur .....	36
Tabel 3. 15. Contoh data Penelitian.....	36
Tabel 3. 16. Probabilitas Kata Tiap Sentimen .....	37
Tabel 3. 17. Jadwal Kegiatan .....	40
Tabel 4. 1. Data Latih Dan Data Uji.....	60
Tabel 4. 2. <i>Confusion Matrix</i> Skema satu.....	61
Tabel 4. 3. <i>Confusion Matrix</i> Skema dua.....	62
Tabel 4. 4. <i>Confusion Matrix</i> Skema Tiga .....	63
Tabel 4. 5. Jumlah Fitur Pada Setiap Data Latih Dan Data Uji.....	64
Tabel 4. 6. <i>Confusion Matrix</i> Skema empat.....	64
Tabel 4. 7. <i>Confusion Matrix</i> Skema Lima .....	65
Tabel 4. 8. <i>Confusion Matrix</i> Skema Enam .....	66
Tabel 4. 9. Peforma Model.....	67
Tabel 4. 10. Prediksi sentimen untuk data teks baru .....	69
Tabel 4. 11. Visualisasi Kata Positif.....	72
Tabel 4. 12. Visualisasi Kata Negatif .....	72

## DAFTAR LAMPIRAN

Lampiran 1 . Hasil uji plagiasi .....	78
Lampiran 2 . Data penelitian .....	79
Lampiran 3 . <i>Source Code</i> yang digunakan untuk analisis .....	80