



***SENTIMENT ANALYSIS ULASAN PENGGUNA APLIKASI PADA PT.  
XYZ DALAM BIDANG FARMASI MENGGUNAKAN ALGORITMA  
NAÏVE BAYES CLASSIFIER***

**SKRIPSI**

**Tia Arlin Dita**

**1810512088**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA  
FAKULTAS ILMU KOMPUTER  
PROGRAM STUDI SISTEM INFORMASI  
2022**



***SENTIMENT ANALYSIS ULASAN PENGGUNA APLIKASI PADA PT.  
XYZ DALAM BIDANG FARMASI MENGGUNAKAN ALGORITMA  
NAÏVE BAYES CLASSIFIER***

**SKRIPSI**

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

**Tia Arlin Dita**

**1810512088**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA**

**FAKULTAS ILMU KOMPUTER**

**PROGRAM STUDI SISTEM INFORMASI**

**2022**

## PERNYATAAN ORISINALITAS

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

Nama : Tia Arlin Dita

NIM : 1810512088

Tanggal : 17 Juli 2022

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

Jakarta, 17 Juli 2022

Yang menyatakan,



(Tia Arlin Dita)

## LEMBAR PENGESAHAN

Dengan ini dinyatakan bahwa Tugas Akhir berikut :

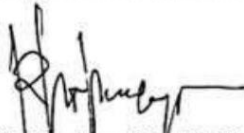
Nama : Tia Arlin Dita

NIM : 1810512088

Program Studi : S1 - Sistem Informasi

Judul Tugas Akhir : *Sentiment Analysis* Ulasan Pengguna Aplikasi Pada PT. XYZ Dalam Bidang Farmasi Menggunakan Algoritma *Naive Bayes Classifier*

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



Rudhy Ho Parabaya, SE., MMSI

Penguji I



Rio Wirawan., S.Kom., MMSI

Penguji II



Ati Zaidiah, S.Kom., MTL

Pembimbing I



Ika Nurlaili Isnainiyah, S.Kom., M.Sc.

Pembimbing II



Dr. Ermatita M. Kom.

Dekan



Helena Nurramdhani Irmada, S.Pd., M.Kom.

Ketua Program Studi

Ditetapkan di : Jakarta  
Tanggal Ujian : 11 Juli 2022



**PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK  
KEPENTINGAN AKADEMIS**

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

Nama : Tia Arlin Dita  
NIM : 1810512088  
Fakultas : Ilmu Komputer  
Program Studi : S1 - Sistem Informasi

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:

***SENTIMENT ANALYSIS* ULASAN PENGGUNA APLIKASI  
PADA PT. XYZ DALAM BIDANG FARMASI  
MENGUNAKAN ALGORITMA *NAÏVE BAYES*  
*CLASSIFIER***

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 : 17 Juli 2022  
Yang menyatakan,



(Tia Arlin Dita)

**SENTIMENT ANALYSIS ULASAN PENGGUNA APLIKASI PADA PT.  
XYZ DALAM BIDANG FARMASI MENGGUNAKAN ALGORITMA  
NAÏVE BAYES CLASSIFIER**

**Tia Arlin Dita**

**ABSTRAK**

Teknologi informasi telah memberikan dampak positif pada kehidupan masyarakat. Penerapan aplikasi pada suatu perusahaan merupakan kebutuhan yang tidak bisa dihindari. PT. XYZ yang bergerak dibidang farmasi adalah perusahaan industri farmasi di Indonesia telah menerapkan aplikasi yang dijalankan pada google play store. Aplikasi pada google play store dapat diakses secara luas oleh masyarakat. Untuk mengetahui penerimaan masyarakat terhadap aplikasi tersebut, perlu dilakukan analisis terhadap opini masyarakat pada aplikasi tersebut. Ulasan pengguna aplikasi pada google play store dapat digunakan untuk menggali informasi opini masyarakat. Dalam penelitian ini akan dilakukan klasifikasi dengan menggunakan metode naïve bayes untuk mengetahui opini masyarakat terhadap aplikasi PT. XYZ yang bergerak dibidang farmasi. Tahapan dalam melakukan *Sentiment Analysis* ini adalah mengumpulkan ulasan yang akan dilakukan *Pre-Processing* setelah itu dilakukan klasifikasi dengan menggunakan metode naïve bayes. Hasil dari penelitian ini didapatkan data ulasan yang didapatkan adalah 900 yang terdiri dari 450 data positif dan 450 data negatif. Hasil yang didapatkan dari proses klasifikasi menggunakan algoritma naïve bayes classifier dengan pembagian data latih sebesar 80% serta data uji 20% mendapatkan hasil yang baik. Hasil ini ditunjukkan oleh uji evaluasi menggunakan confusion matrix sehingga didapatkan nilai akurasi sebesar 90%

**Kata kunci:** *Sentiment Analysis, google play store, naïve bayes classifier, klasifikasi*

# **SENTIMENT ANALYSIS USER REVIEWS APPLICATIONS AT PT. XYZ IN PHARMACEUTICAL FIELD USING NAVE BAYES CLASSIFIER ALGORITHM**

**Tia Arlin Dita**

## **ABSTRACT**

*Information technology has had a positive impact on people's lives. The application of applications in a company is a necessity that cannot be avoided. PT. XYZ, which is engaged in pharmaceuticals, is a pharmaceutical industry company in Indonesia that has implemented an application that runs on the Google Play Store. Applications on the Google Play Store can be accessed widely by the public. To determine public acceptance of the application, it is necessary to analyze public opinion on the application. Application user reviews on the Google Play Store can be used to explore public opinion information. In this study, classification will be carried out using the nave Bayes method to find out public opinion on the application of PT. XYZ which is engaged in pharmacy. The stage in conducting this Sentiment Analysis is to collect reviews that will be carried out Pre-Processing after that the classification is carried out using the Nave Bayes method. The results of this study obtained that the review data obtained were 900 consisting of 450 positive data and 450 negative data. The results obtained from the classification process using the naive bayes classifier algorithm with the distribution of training data by 80% and 20% test data get good results. These results are shown by an evaluation test using a confusion matrix so that an accuracy value of 90% is obtained.*

**Keywords:** *Sentiment Analysis, google play store, nave bayes classifier, classification*

## KATA PENGANTAR

Dengan Memanjatkan puja dan puji syukur atas kehadiran Allah SWT yang telah melimpahkan rahmat, taufik, serta hidayah-nya sehingga penulis dapat menyelesaikan skripsi ini dengan judul **“SENTIMENT ANALYSIS ULASAN PENGGUNA APLIKASI PADA PT. XYZ DALAM BIDANG FARMASI MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER”**, sebagai salah satu syarat untuk menyelesaikan Program Sarjana (S1) Jurusan Sistem Informasi Fakultas Ilmu Komputer UPN Veteran Jakarta.

Penulis menyadari bahwa skripsi ini tidak mungkin terselesaikan tanpa adanya dukungan, bantuan, bimbingan, serta nasehat dari berbagai pihak selama penyusunan skripsi ini. Pada kesempatan ini penulis menyampaikan terima kasih setulus-tulusnya kepada:

1. Ibu Dr. Ermatita, M.Kom selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta.
2. Ibu Helena Nurramdhani Irmada, S.Pd.,M selaku Ketua program studi S1 Sistem Informasi Universitas Pembangunan Nasional Veteran Jakarta.
3. Ibu Ati Zaidiah,S.Kom.,MTI., selaku Dosen Pembimbing 1.
4. Ibu Ika Nurlaili Isnainiyah, S.Kom., M.Sc., selaku Dosen Pembimbing 2.
5. Orang tua serta keluarga penulis yang selalu mendoakan serta memberikan dukungan moril.
6. Seluruh Sahabat yang selalu membantu dan mendoakan penulis.
7. Seluruh pihak yang terlibat dalam kelancaran penyusunan skripsi ini.yang tidak bisa disebutkan satu-persatu.

Peneliti menyadari bahwa masih terdapat banyak kesalahan serta kekurangan dalam penyusunan Tugas Akhir ini, oleh karena itu penulis menerima kritik serta saran yang membangun. Penulis juga berharap agar penelitian yang telah dilakukan ini dapat bermanfaat bagi penulis dan pihak lain yang terkait.

Jakarta, 23 Mei 2022

Penulis



## DAFTAR ISI

PERNYATAAN ORISINALITAS .....	ii
LEMBAR PENGESAHAN .....	iii
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS .....	iv
ABSTRAK .....	v
<i>ABSTRACT</i> .....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI .....	viii
DAFTAR TABEL .....	xi
DAFTAR GAMBAR .....	xii
DAFTAR RUMUS .....	xiii
LAMPIRAN .....	xiv
<b>BAB 1</b> PENDAHULUAN .....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	2
1.3. Tujuan Penelitian .....	3
1.4. Manfaat Penelitian .....	3
1.5. Ruang Lingkup Penelitian .....	3
1.6. Sistematika Penulisan .....	3
<b>BAB 2</b> LANDASAN TEORI .....	5
2.1. Google Play .....	5
2.2. Apotek .....	5
2.3. <i>Sentiment Analysis</i> .....	6
2.4. <i>Text Mining</i> .....	6
2.5. <i>Pre-Processing</i> .....	7
2.5.1. <i>Data Cleansing</i> .....	7
2.5.2. <i>Case Folding</i> .....	7
2.5.3. <i>Tokenization</i> .....	7
2.5.4. <i>Filtering</i> .....	7
2.5.5. <i>Steaming</i> .....	8
2.6. Pembobotan TF-IDF .....	8

2.7.	Klasifikasi .....	8
2.8.	<i>Naïve Bayes Classifier</i> .....	8
2.9.	Penelitian Sebelumnya .....	10
<b>BAB 3 METODOLOGI PENELITIAN</b> .....		12
3.1.	Tahapan Penelitian .....	12
3.1.1.	Penentuan Topik .....	13
3.1.2.	Pengambilan Data .....	13
3.1.3.	<i>Pre-Processing</i> .....	13
3.1.4.	Pelabelan Data .....	13
3.1.5.	Pembobotan TF-IDF .....	14
3.1.6.	<i>Naïve Bayes Classifier</i> .....	14
3.1.7.	Pembagian Data Latih dan Data Uji .....	15
3.1.8.	Pembentukan Model .....	15
3.1.9.	Evaluasi .....	15
3.1.10	Hasil dan Visualisasi .....	16
3.2	Metode Analisis Data .....	16
3.3	Alat dan Bahan Penelitian .....	17
3.4.	Jadwal Penelitian .....	17
<b>BAB 4 HASIL DAN PEMBAHASAN</b> .....		19
4.1.	Pengumpulan Data .....	19
4.2.	<i>Pre-Processing</i> .....	20
4.2.1.	<i>Data Cleansing</i> .....	20
4.2.2.	<i>Case Folding</i> .....	21
4.2.3.	<i>Filtering</i> .....	22
4.2.4.	<i>Tokenization</i> .....	23
4.2.5.	<i>Steaming</i> .....	24
4.3.	Pelabelan Data .....	25
4.4.	Pembobotan TF-IDF .....	26
4.5.	<i>Naïve bayes classifier</i> .....	28
4.5.1.	Pembagian Data Latih dan Data Uji .....	29
4.5.2.	Pembentukan Model <i>Naïve Bayes</i> .....	29
4.6.	Evaluasi Algoritma <i>Naïve Bayes</i> .....	30
4.6.1.	<i>Confusion Matrix</i> .....	30
4.6.2.	<i>Accuracy</i> .....	32

<b>4.7. Visualisasi Hasil</b> .....	33
<b>4.7.1. Visualisasi Hasil Positif</b> .....	33
<b>4.7.2. Visualisasi Hasil Negatif</b> .....	34
<b>BAB 5 PENUTUP</b> .....	36
<b>5.1. Kesimpulan</b> .....	36
<b>5.2. Saran</b> .....	36
<b>DAFTAR PUSTAKA</b> .....	37
<b>RIWAYAT HIDUP</b> .....	39
<b>LAMPIRAN</b> .....	40

## DAFTAR TABEL

Tabel 1. Penelitian Sebelumnya .....	10
Tabel 2. Confusion Matrix.....	15
Tabel 3. Jadwal Penelitian.....	18
Tabel 4. Data Cleansing.....	20
Tabel 5. Case Folding.....	21
Tabel 6. Filtering .....	22
Tabel 7. Tokenization.....	23
Tabel 8. Steaming.....	24
Tabel 9. Contoh TF-IDF Manual .....	27
Tabel 10. Pembagian Data Latih dan Data Uji .....	29
Tabel 11. Confusion Matrix 80:20.....	31
Tabel 12. Confusion Matrix 70:30.....	31
Tabel 13. Accuracy 80:20 .....	32
Tabel 14. Accuracy 70:30 .....	32

## DAFTAR GAMBAR

gambar 2. Tahapan Penelitian.....	12
gambar 3. Hasil scraping .....	19
gambar 4. Pelabelan Data .....	26
gambar 5. Hasil Pelabelan Data .....	26
gambar 6. Visualisasi Positif .....	33
gambar 7. Visualisasi Negatif.....	34

## DAFTAR RUMUS

Rumus 2.1. Skor TF-IDF .....	8
Rumus 2.2. <i>Naïve Bayes Classifier</i> .....	9
Rumus 3.1 Pembobotan TF-IDF .....	14
Rumus 3.2 Teorema Bayes.....	14
Rumus 3.3 <i>Accuracy</i> .....	16

## LAMPIRAN

<b>Lampiran 1.</b> Pengambilan Data .....	40
<b>Lampiran 2.</b> Import dan Pendefinisian Data.....	41
<b>Lampiran 3.</b> Data Cleansing .....	41
<b>Lampiran 4.</b> Case Folding .....	43
<b>Lampiran 5.</b> Filtering.....	44
<b>Lampiran 6.</b> Tokenization .....	44
<b>Lampiran 7.</b> Steaming.....	45
<b>Lampiran 8.</b> Pelabelan data .....	45
<b>Lampiran 9.</b> Pembobotan TF-IDF .....	45
<b>Lampiran 10.</b> Pembagian Data Latih dan Data Uji.....	46
<b>Lampiran 11.</b> Confusion Matrix .....	46
<b>Lampiran 12.</b> Accuracy.....	47
<b>Lampiran 13.</b> Visualisasi data .....	48
<b>Lampiran 14.</b> Hasil Uji Turnitin .....	49