

**PERBANDINGAN METODE RANDOM FOREST DAN  
LOGISTIC REGRESSION UNTUK ANALISIS SENTIMEN  
ULASAN APLIKASI TUMBUH KEMBANG ANAK  
DI PLAY STORE**

**SKRIPSI**



**Oleh:**

**MUHAMMAD ALFYANDO**

**NPM. 19081010037**

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

**2024**

**PERBANDINGAN METODE RANDOM FOREST DAN  
LOGISTIC REGRESSION UNTUK ANALISIS SENTIMEN  
ULASAN APLIKASI TUMBUH KEMBANG ANAK  
DI PLAY STORE**

**SKRIPSI**

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



Oleh:

**MUHAMMAD ALFYANDO**

**NPM. 19081010037**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"  
JAWA TIMUR  
2024**

## LEMBAR PENGESAHAN SKRIPSI

Judul : PERBANDINGAN METODE RANDOM FOREST DAN  
LOGISTIC REGRESSION UNTUK ANALISIS SENTIMEN  
ULASAN APLIKASI TUMBUH KEMBANG ANAK DI  
PLAY STORE

Oleh : Muhammad Alfyando

NPM : 19081010037

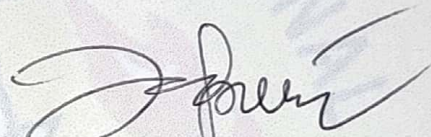
Telah Diseminarkan Dalam Ujian Skripsi Pada :  
Hari Jumat, Tanggal 5 Januari 2024

### Mengetahui

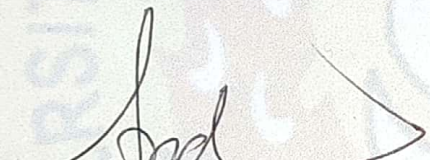
1. Dosen Pembimbing

  
Fetty Tri Anggraeny, S.Kom., M.Kom  
NIP. 19820211 202121 2 005


1. Dosen Penguji

  
Dr. Rr. Ani Dajah Rahajoe, S. T., M. Cs  
NIP. 19730512 200501 2003

2.


  
Andreas Nugroho Sihananto, S.Kom., M.Kom  
NPT. 211199 00 412271

2.

  
Fawwaz Ali Akbar, S.Kom., M.Kom  
NIP. 19920317 201803 1 002

### Menyetujui

Dekan  
Fakultas Ilmu Komputer

  
Prof. Dr. Ir. Novirina Hendrasarie, MT  
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 Program studi Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur, yang bertanda tangan di bawah ini:

Nama : Muhammad Alfyando

NPM : 19081010037

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

### **“PERBANDINGAN METODE RANDOM FOREST DAN LOGISTIC REGRESSION UNTUK ANALISIS SENTIMEN ULASAN APLIKASI TUMBUH KEMBANG ANAK DI PLAY STORE”**

Bukan merupakan plagiat dari tugas akhir atau skripsi maupun penelitian orang lain dan juga bukan merupakan produk atau *software* yang saya beli dari pihak lain. Saya juga menyatakan bahwa skripsi ini merupakan 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.

Surabaya, 2 Januari 2024

Hormat Saya,



**Muhammad Alfyando**

**NPM. 19081010037**

# **PERBANDINGAN METODE RANDOM FOREST DAN LOGISTIC REGRESSION UNTUK ANALISIS SENTIMEN ULASAN APLIKASI TUMBUH KEMBANG ANAK DI PLAY STORE**

**Nama Mahasiswa : Muhammad Alfyando**

**NPM : 19081010037**

**Program Studi : Informatika**

**Dosen Pembimbing : Fetty Tri Anggraeny, S.Kom., M.Kom.**

**Andreas Nugroho Sihananto, S. Kom., M.Kom.**

## **ABSTRAK**

Anak usia dini mempunyai peranan penting dalam pembentukan dasar perkembangan, yang melibatkan aspek-aspek seperti nilai agama, moral, sosial emosional, bahasa, kognitif, dan fisik motorik. Konsep belajar melalui bermain menjadi landasan utama dalam pendekatan pendidikan pada anak usia dini. Orang tua memiliki tanggung jawab untuk memahami pendidikan anak zaman sekarang dan memanfaatkan teknologi sebagai alat bantu. Penelitian ini mengeksplorasi aplikasi tentang tumbuh kembang anak, khususnya "Tentang Anak", "PrimaKu", dan "Teman Bumil". Fokus penelitian melibatkan analisis sentimen ulasan pengguna aplikasi menggunakan metode Random Forest dan Logistic Regression, dengan hasil yang menunjukkan performa terbaik pada masing-masing aplikasi.

Hasil penelitian mencatat bahwa pada Aplikasi "Tentang Anak," metode Logistic Regression mencapai akurasi tertinggi dengan nilai 94.34%, sedangkan pada Aplikasi "PrimaKu," metode Logistic Regression juga meraih akurasi terbaik sebesar 88.52%. Untuk Aplikasi "Teman Bumil," metode Random Forest dan Logistic Regression mencapai akurasi tertinggi masing-masing sebesar 86.94% dan 84.47%. Analisis sentimen ulasan pengguna memberikan pemahaman yang lebih baik tentang kebutuhan dan preferensi pengguna terhadap aplikasi pendidikan anak, sehingga dapat meningkatkan kualitas layanan yang disediakan oleh aplikasi tersebut.

***Kata kunci : Analisis Sentimen, Logistic Regression, Random Forest, Tumbuh Kembang Anak***

## **KATA PENGANTAR**

Dengan penuh rasa syukur, peneliti ingin mengucapkan terima kasih kepada Tuhan YME atas berkat dan rahmat-Nya yang telah memungkinkan penyelesaian skripsi berjudul "Perbandingan Metode Random Forest Dan Logistic Regression Untuk Analisis Sentimen Ulasan Aplikasi Tumbuh Kembang Anak Di Play Store".

Laporan skripsi ini disusun dalam rangka memenuhi persyaratan mata kuliah skripsi dan sebagai salah satu langkah penting dalam menyelesaikan Pendidikan tingkat Sarjana (S1) di Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur.

Peneliti sadar akan adanya keterbatasan dalam penulisan skripsi ini, dan dengan rendah hati menerima saran serta kritik yang membangun. Semoga skripsi ini dapat memberikan manfaat bagi semua pihak, baik bagi pembaca maupun untuk penelitian mendatang.

Surabaya, 2 Januari 2024

Peneliti,

**Muhammad Alfando**

**NPM. 19081010037**

## UCAPAN TERIMA KASIH

Penulisan laporan skripsi ini tidak dapat berjalan lancar tanpa bantuan, motivasi, dan dukungan dari berbagai pihak. Oleh karena itu, dengan tulus, peneliti ingin mengucapkan terima kasih kepada:

1. Tuhan YME, yang telah memberikan rahmat, hidayah, serta karunia-Nya sehingga peneliti dapat menyelesaikan penulisan skripsi ini.
2. Orang tua peneliti, yang selalu memberikan doa, dukungan moral, dan motivasi agar peneliti tetap semangat menjalani proses pengerjaan skripsi.
3. Bapak Prof. Dr. Ir. Akhmad Fauzi, M.MT., selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Ibu Dr. Ir. Novirina Hendrasarie, M.T., selaku Dekan Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur.
5. Ibu Fetty Tri Anggraeny S.Kom., M.Kom. selaku Koordinator Program Studi Informatika Fakultas Ilmu Komputer UPN “Veteran” Jawa Timur sekaligus Dosen Pembimbing I yang telah banyak memberikan arahan kepada peneliti selama proses perkuliahan dan pengerjaan skripsi.
6. Bapak Andreas Nugroho Sihananto, S.Kom., M.Kom., selaku Dosen Pembimbing II yang telah memberikan dukungan serta saran kepada peneliti, sehingga peneliti dapat menyelesaikan penyusunan skripsi ini.
7. Bapak Hendra Maulana, S.Kom., M.Kom., selaku dosen wali yang telah membantu dalam perwalian sejak awal masa perkuliahan.
8. Seluruh Bapak dan Ibu dosen, serta Staff program studi Informatika UPN “Veteran” Jawa Timur yang telah mengajar dan memberikan ilmu serta pengalaman selama masa perkuliahan.
9. Teman seperjuangan yang ada dimanapun yang selalu menghibur, mengingatkan, dan menyemangati dalam pengerjaan skripsi ini. Tak lupa juga, ucapan terima kasih kepada teman-teman informatika angkatan 2019 serta pihak-pihak lain yang terkait yang tidak disebutkan atas segala bentuk kepeduliannya dalam pengerjaan penelitian ini.

## DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI .....	i
SURAT PERNYATAAN ANTI PLAGIAT .....	ii
ABSTRAK .....	iii
KATA PENGANTAR .....	iv
UCAPAN TERIMA KASIH .....	v
DAFTAR ISI .....	vi
DAFTAR GAMBAR .....	ix
DAFTAR TABEL .....	xii
DAFTAR KODE PROGRAM .....	xiv
BAB I PENDAHULUAN .....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	4
1.3. Tujuan Penelitian .....	4
1.4. Manfaat Penelitian .....	4
1.5. Batasan Masalah .....	5
BAB II TINJAUAN PUSTAKA .....	6
2.1. Penelitian Terdahulu .....	6
2.2. Aplikasi “Tentang Anak” .....	8
2.3. Aplikasi “PrimaKu” .....	9
2.4. Aplikasi “Teman Bumil” .....	9
2.5. Analisis Sentimen .....	10
2.6. Text Mining .....	10
2.7. Text Preprocessing .....	10
2.7.1. Cleaning .....	11
2.7.2. Lowercasing .....	11
2.7.3. Normalization .....	11
2.7.4. Tokenizing .....	12
2.7.5. Stopword Removal .....	12
2.7.6. Stemming .....	12
2.8. VADER .....	12



2.9.	Synthetic Minority Over Sampling Technique (SMOTE) .....	13
2.10.	Pembobotan Teks (TF-IDF).....	14
2.11.	Random Forest .....	15
2.12.	Logistic Regression.....	16
2.13.	Confussion Matrix .....	19
<b>BAB III METODOLOGI PENELITIAN .....</b>		<b>21</b>
3.1.	Tahapan Penelitian .....	21
3.2.	Studi Literatur.....	21
3.3.	Analisa & Desain.....	22
3.4.	Akuisisi Data .....	22
3.5.	Eksplorasi Data.....	23
3.6.	Text Preprocessing .....	23
3.6.1.	Cleaning .....	23
3.6.2.	Lowercasing .....	25
3.6.3.	Normalization.....	25
3.6.4.	Tokenizing.....	26
3.6.5.	Stopword Removal.....	27
3.6.6.	Stemming .....	28
3.7.	Pelabelan Data .....	29
3.8.	Pembobotan Teks (TF-IDF) .....	31
3.9.	Handle Imbalance Data .....	33
3.10.	Klasifikasi Data.....	37
3.10.1.	Random Forest .....	37
3.10.2.	Logistic Regression.....	41
3.11.	Skema Pengujian.....	47
<b>BAB IV HASIL &amp; PEMBAHASAN.....</b>		<b>49</b>
4.1.	Hasil Akuisisi Data.....	49
4.2.	Hasil Eksplorasi Data .....	52
4.3.	Hasil Preprocessing Text.....	53
4.3.1.	Hasil Implementasi Cleaning .....	53
4.3.2.	Hasil Implementasi Lowercasing .....	54
4.3.3.	Hasil Implementasi Normalization .....	55

4.3.4.	Hasil Implementasi Tokenizing .....	56
4.3.5.	Hasil Implementasi Stopword Removal .....	56
4.3.6.	Hasil Implementasi Stemming .....	57
4.4.	Hasil Pelabelan Data .....	58
4.5.	Hasil Implementasi Pembobotan Teks (TF-IDF) .....	59
4.6.	Hasil Handle Imbalance Data.....	60
4.7.	Pembagian data.....	60
4.8.	Model.....	61
4.8.1.	Random Forest .....	61
4.8.2.	Logistic Regression.....	62
4.9.	Hasil Skema Pengujian.....	63
4.9.1.	Aplikasi “Tentan Anak” .....	65
4.9.2.	Aplikasi “PrimaKu” .....	71
4.9.3.	Aplikasi “Teman Bumil”.....	77
4.10.	Hasil dan Analisis Uji.....	84
BAB V KESIMPULAN & SARAN.....		87
5.1.	Kesimpulan.....	87
5.2.	Saran .....	88
DAFTAR PUSTAKA .....		89
BIODATA PENULIS .....		93

## DAFTAR GAMBAR

Gambar 2. 1 Preview Aplikasi “Tentang Anak” .....	8
Gambar 2. 2 Preview Aplikasi “PrimaKu” .....	9
Gambar 2. 3 Preview Aplikasi "Teman Bumil" .....	10
Gambar 2. 4 Ilustrasi Random Forest .....	16
Gambar 2. 5 Rumus Logistic Regression (David G. Kleinbaum and Mitchel Klein, Logistic Regression: A Self-Learning Text, 2nd Edition, halaman. 6).....	16
Gambar 2. 6 Shape Logistic Regression (David G. Kleinbaum and Mitchel Klein, Logistic Regression: A Self-Learning Text, 2nd Edition, halaman. 6).....	17
Gambar 3. 1 Tahapan Penelitian .....	21
Gambar 3. 2 Analisa Alur Sistem .....	22
Gambar 3. 3 Flowchart Cleaning Text .....	24
Gambar 3. 4 Flowchart Lowercasing .....	25
Gambar 3. 5 Flowchart Normalization .....	26
Gambar 3. 6 Flowchart Tokenizing .....	27
Gambar 3. 7 Flowchart Stopword Removal.....	28
Gambar 3. 8 Flowchart Stemming .....	29
Gambar 3. 9 Flowchart Pelabelan Data.....	30
Gambar 3. 10 Flowchart Pembobotan Teks .....	31
Gambar 3. 11 Flowchart Handle Imbalance Data.....	34
Gambar 3. 12 Flowchart Random Forest .....	37
Gambar 3. 13 Flowchart Logistic Regression.....	41
Gambar 4. 1 Contoh Id Aplikasi .....	49
Gambar 4. 2 Hasil Akuisisi Data Aplikasi "Tentang Anak" .....	51
Gambar 4. 3 Hasil Akuisisi Data Aplikasi "PrimaKu" .....	51
Gambar 4. 4 Hasil Akuisisi Data Aplikasi "Teman Bumil" .....	51
Gambar 4. 5 Eksplorasi Data .....	52
Gambar 4. 6 Hasil Implementasi Cleaning Text .....	54
Gambar 4. 7 Hasil Implementasi Lowercasing .....	55
Gambar 4. 8 Hasil Implementasi Normalization.....	55
Gambar 4. 9 Hasil Implementasi Tokenizing .....	56

Gambar 4. 10 Hasil Implementasi Stopword Removal.....	57
Gambar 4. 11 Hasil Implementasi Stemming .....	58
Gambar 4. 12 Hasil Implementasi Translation .....	58
Gambar 4. 13 Hasil Akurasi Skenario 1 .....	65
Gambar 4. 14 Hasil Confussion Matrix Skenario 1 .....	66
Gambar 4. 15 Hasil Akurasi Skenario 2 .....	66
Gambar 4. 16 Hasil Confussion Matrix Skenario 2 .....	67
Gambar 4. 17 Hasil Akurasi Skenario 3 .....	67
Gambar 4. 18 Hasil Confussion Matrix Skenario 2 .....	68
Gambar 4. 19 Hasil Akurasi Skenario 4 .....	68
Gambar 4. 20 Hasil Confussion Matrix Skenario 4 .....	69
Gambar 4. 21 Hasil Akurasi Skenario 5 .....	69
Gambar 4. 22 Hasil Confussion Matrix Skenario 5 .....	70
Gambar 4. 23 Hasil Akurasi Skenario 6 .....	70
Gambar 4. 24 Hasil Confussion Matrix Skenario 6 .....	71
Gambar 4. 25 Hasil Akurasi Skenario 7 .....	71
Gambar 4. 26 Hasil Confussion Matrix Skenario 7 .....	72
Gambar 4. 27 Hasil Akurasi Skenario 8 .....	72
Gambar 4. 28 Hasil Confussion Matrix Skenario 8 .....	73
Gambar 4. 29 Hasil Akurasi Skenario 9 .....	73
Gambar 4. 30 Hasil Confussion Matrix Skenario 9 .....	74
Gambar 4. 31 Hasil Akurasi Skenario 10 .....	74
Gambar 4. 32 Hasil Confussion Matrix Skenario 10 .....	75
Gambar 4. 33 Hasil Akurasi Skenario 11 .....	75
Gambar 4. 34 Hasil Confussion Matrix Skenario 11 .....	76
Gambar 4. 35 Hasil Akurasi Skenario 12 .....	76
Gambar 4. 36 Hasil Confussion Matrix Skenario 12 .....	77
Gambar 4. 37 Hasil Akurasi Skenario 13 .....	77
Gambar 4. 38 Hasil Confussion Matrix Skenario 13 .....	78
Gambar 4. 39 Hasil Akurasi Skenario 14 .....	78
Gambar 4. 40 Hasil Confussion Matrix Skenario 14 .....	79
Gambar 4. 41 Hasil Akurasi Skenario 15 .....	79

Gambar 4. 42 Hasil Confussion Matrix Skenario 15 .....	80
Gambar 4. 43 Hasil Akurasi Skenario 16 .....	80
Gambar 4. 44 Hasil Confussion Matrix Skenario 16 .....	81
Gambar 4. 45 Hasil Akurasi Skenario 17 .....	81
Gambar 4. 46 Hasil Confussion Matrix Skenario 17 .....	82
Gambar 4. 47 Hasil Akurasi Skenario 18 .....	82
Gambar 4. 48 Hasil Confussion Matrix Skenario 18 .....	83
Gambar 4. 49 Visualisasi Hasil Confussion Matrix.....	86

## DAFTAR TABEL

Tabel 2. 1 Cleaning Text.....	11
Tabel 2. 2 Lowercasing .....	11
Tabel 2. 3 Normalization.....	11
Tabel 2. 4 Tokenizing .....	12
Tabel 2. 5 Stopword Removal.....	12
Tabel 2. 6 Stemming .....	12
Tabel 2. 7 Confussion Matrix.....	19
Tabel 3. 1 Rentang Tahun Ulasan .....	23
Tabel 3. 2 Hasil Proses Cleaning Text .....	24
Tabel 3. 3 Hasil Proses Lowercasing .....	25
Tabel 3. 4 Tabel Hasil Normalization .....	26
Tabel 3. 5 Tokenizing .....	27
Tabel 3. 6 Hasil Proses Stopword Removal.....	28
Tabel 3. 7 Hasil Proses Stemming .....	29
Tabel 3. 8 Contoh Data Ulasan Untuk Pembobotan .....	31
Tabel 3. 9 Corpus Ulasan .....	32
Tabel 3. 10 Frekuensi Kemunculan Kata.....	32
Tabel 3. 11 Representasi Matriks Hasil TF-IDF .....	33
Tabel 3. 12 Contoh Data Ulasan Untuk Proses SMOTE .....	34
Tabel 3. 13 Corpus Ulasan (SMOTE).....	35
Tabel 3. 14 Frekuensi Kemunculan Kata (SMOTE).....	35
Tabel 3. 15 Hasil Pembobotan (SMOTE).....	36
Tabel 3. 16 Contoh Data Ulasan Untuk Random Forest.....	38
Tabel 3. 17 Corpus Ulasan (Random Forest).....	38
Tabel 3. 18 Frekuensi Kemunculan Kata (Random Forest).....	38
Tabel 3. 19 Hasil Pembobotan (Random Forest).....	39
Tabel 3. 20 Hasil Bootstraping.....	39
Tabel 3. 21 Contoh Data Ulasan Untuk Logistic Regression .....	42
Tabel 3. 22 Corpus Ulasan (Logistic Regression) .....	43
Tabel 3. 23 Frekuensi Kemunculan Kata (Logistic Regression) .....	43

Tabel 3. 24 Hasil Pembobotan (Logistic Regression).....	44
Tabel 3. 25 Skema Pengujian.....	47
Tabel 4. 1 Id Aplikasi.....	50
Tabel 4. 2 Hasil Pelabelan Data .....	59
Tabel 4. 3 Hasil Handle Imbalance Data .....	60
Tabel 4. 4 Variabel Pembagian Data.....	60
Tabel 4. 5 Confussion Matrix Aplikasi "Tentang Anak" .....	84
Tabel 4. 6 Confussion Matrix Aplikasi "PrimaKu" .....	84
Tabel 4. 7 Confussion Matrix Aplikasi "Teman Bumil" .....	85

## DAFTAR KODE PROGRAM

Kode Program 4. 1 Akuisisi Data .....	50
Kode Program 4. 2 Kamus Normalization .....	53
Kode Program 4. 3 Cleaning Text.....	54
Kode Program 4. 4 Lowercasing.....	54
Kode Program 4. 5 Normalization .....	55
Kode Program 4. 6 Tokenizing .....	56
Kode Program 4. 7 Stopword Removal .....	57
Kode Program 4. 8 Stemming.....	57
Kode Program 4. 9 Translation .....	58
Kode Program 4. 10 Pelabelan Data .....	59
Kode Program 4. 11 Label Encoder.....	59
Kode Program 4. 12 TF-IDF .....	60
Kode Program 4. 13 SMOTE.....	60
Kode Program 4. 14 Pembagian Data .....	61
Kode Program 4. 15 Random Forest.....	61
Kode Program 4. 16 Logistic Regression .....	63
Kode Program 4. 17 Prediksi Random Forest.....	64
Kode Program 4. 18 Prediksi Logistic Regression .....	64
Kode Program 4. 19 Import Library SKLearn .....	65
Kode Program 4. 20 Hitung Akurasi Random Forest .....	65
Kode Program 4. 21 Hitung Akurasi Logistic Regression.....	65