

**KLASIFIKASI SENTIMEN ULASAN PENGGUNA ADOBE
SCAN PADA SITUS GOOGLE PLAY STORE DENGAN
METODE RANDOM FOREST**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Ilmu Komputer



disusun oleh

DESVINA KUSUMANINGRUM

21.11.4193

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**KLASIFIKASI SENTIMEN ULASAN PENGGUNA ADOBE
SCAN PADA SITUS GOOGLE PLAY STORE DENGAN
METODE RANDOM FOREST**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Ilmu Komputer



disusun oleh

DESVINA KUSUMANINGRUM

21.11.4193

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**KLASIFIKASI SENTIMEN ULASAN PENGGUNA ADOBE SCAN PADA
SITUS GOOGLE PLAY STORE DENGAN METODE RANDOM FOREST**

yang disusun dan diajukan oleh

Desvina Kusumaningrum

21.11.4193

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 12 Maret 2025

Dosen Pembimbing,



I Made Artha Agastya, S.T., M.Eng., Ph.D.

NIK. 190302352

HALAMAN PENGESAHAN

SKRIPSI

**KLASIFIKASI SENTIMEN ULASAN PENGGUNA ADOBE SCAN PADA
SITUS GOOGLE PLAY STORE DENGAN METODE RANDOM FOREST**

yang disusun dan diajukan oleh

Desvina Kusumaningrum

21.11.4193

Telah dipertahankan di depan Dewan Penguji
pada tanggal 12 Maret 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Windha Mega Pradnya D, S.Kom., M.Kom.
NIK. 190302185



Dwi Nurani, S.Kom., M.Kom.
NIK. 190302236



I Made Artha Agastya, S.T., M.Eng., Ph.D.
NIK. 190302352



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 12 Maret 2025

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Desvina Kusumaningrum
NIM : 21.11.4193

Menyatakan bahwa Skripsi dengan judul berikut:

Klasifikasi Sentimen Ulasan Pengguna Adobe Scan Pada Situs Google Play Store Dengan Metode Random Forest

Dosen Pembimbing : I Made Artha Agastya, S.T., M.Eng., Ph.D.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 12 Maret 2025

Yang Menyatakan,



10000
METERAI
TEMPEL
2568BAMX236984398

Desvina Kusumaningrum

HALAMAN PERSEMBAHAN

Dengan rasa syukur yang mendalam karena telah diberi kesempatan untuk menyelesaikan skripsi ini. Sebagai wujud terima kasih atas motivasi yang diberikan, penulis mempersembahkan skripsi ini kepada:

1. Keluarga kecil dan kedua orang tua tercinta yang senantiasa memberikan doa, kasih sayang, serta dukungan tiada henti.
2. Bapak I Made Artha Agastya, S.T., M.Eng., Ph.D., selaku dosen pembimbing yang telah membantu dan membimbing saya selama penyusunan skripsi ini.
3. Pasangan saya yang selalu menjadi motivasi dan dukungan dalam berbagai hal, baik dalam keadaan senang maupun sulit.
4. Seluruh teman dekat yang telah memberikan semangat dan masukan untuk skripsi saya.

KATA PENGANTAR

Puji syukur kehadiran Tuhan Yang Maha Esa atas rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi ini dengan judul “Klasifikasi Sentimen Ulasan Pengguna Adobe Scan Pada Situs Google Play Store Dengan Metode Random Forest” sebagai salah satu syarat dalam menyelesaikan studi program sarjana jurusan Informatika.

Dalam penyusunan skripsi ini, penulis menyadari bahwa pencapaian ini tidak terlepas dari dukungan, bimbingan, serta bantuan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Keluarga kecil dan kedua orang tua, yang selalu memberikan doa, kasih sayang, serta dukungan moral dan material selama masa perkuliahan hingga selesainya skripsi ini.
2. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas AMIKOM Yogyakarta.
3. Bapak Hanif Al Fatta, S.Kom., M.Kom., Ph.D., selaku Dekan Fakultas Ilmu Komputer.
4. Bapak I Made Artha Agastya, S.T., M.Eng., Ph.D., selaku dosen pembimbing yang selalu memberi ilmu dan arahnya yang sangat berharga dalam penyusunan skripsi ini.
5. Ibu Windha Mega Pradnya Dhuhiya, S.Kom., M.Kom. dan Ibu Dwi Nurani, S.Kom., M.Kom., selaku dosen penguji yang telah memberikan masukan dan arahnya dalam penyempurnaan skripsi ini.
6. Orang-orang terdekat, yang selalu memberikan dukungan, semangat, serta motivasi dalam menyelesaikan skripsi ini.

Yogyakarta, 12 Maret 2025

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xiii
DAFTAR LAMBANG DAN SINGKATAN	xiv
DAFTAR ISTILAH	xv
INTISARI	xvi
ABSTRACT.....	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	3
BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5
2.2 Dasar Teori.....	15

2.2.1	Data Mining	15
2.2.2	Analisis Sentimen	15
2.2.3	Google Play Store	15
2.2.4	Adobe Scan	15
2.2.5	Preprocessing Data.....	16
2.2.6	Pembobotan TF-IDF	16
2.2.7	Oversampling	17
2.2.8	Undersampling	18
2.2.9	Random Forest	18
2.2.10	Confusion Matrix	19
BAB III METODE PENELITIAN		21
3.1	Objek Penelitian.....	21
3.2	Alur Penelitian	21
3.2.1	Pengumpulan Data	22
3.2.2	Preprocessing Data.....	22
3.2.3	Pelabelan Data	22
3.2.4	Penyeimbangan Jumlah Data Positif	23
3.2.5	Pembobotan TF-IDF	23
3.2.6	Pembagian Data	23
3.2.7	Teknik Sampling.....	23
3.2.8	Klasifikasi Model.....	24
3.2.9	Evaluasi.....	24
3.3	Alat dan Bahan.....	24
3.3.1	Data Penelitian	24
3.3.2	Alat/Instruments.....	24

BAB IV HASIL DAN PEMBAHASAN	25
4.1. Pengumpulan Data	25
4.2. Preprocessing Data.....	26
4.2.1 Cleaning Data.....	26
4.2.2 Tokenizing	27
4.2.3 Filtering.....	27
4.2.4 Stemming	28
4.3. Pelabelan Data	29
4.4. Penyeimbangan Jumlah Data Positif	30
4.5. Pembobotan TF-IDF	31
4.6. Pembagian Data	33
4.6.1 Skenario 1: Data Asli 2298 Kelas Positif.....	34
4.6.2 Skenario 2: Data 1000 Kelas Positif	35
4.6.3 Skenario 3: Data 400 Kelas Positif	35
4.7. Klasifikasi dan Evaluasi Skenario 1: Data Asli 2298 Kelas Positif.....	36
4.7.1 Model Tanpa Sampling.....	36
4.7.2 Model Oversampling SMOTE.....	37
4.7.3 Model Oversampling RandomOverSampler.....	37
4.7.4 Model Undersampling RandomUnderSampler.....	38
4.7.5 Model Undersampling ClusterCentroids	39
4.8. Klasifikasi dan Evaluasi Skenario 2: Data 1000 Kelas Positif	40
4.8.1 Model Tanpa Sampling.....	40
4.8.2 Model Oversampling SMOTE.....	40
4.8.3 Model Oversampling RandomOverSampler.....	41
4.8.4 Model Undersampling RandomUnderSampler.....	42

4.8.5 Model Undersampling ClusterCentroids	43
4.9. Klasifikasi dan Evaluasi Skenario 3: Data 400 Kelas Positif	43
4.9.1 Model Tanpa Sampling	44
4.9.2 Model Oversampling SMOTE	44
4.9.3 Model Oversampling RandomOverSampler	45
4.9.4 Model Undersampling RandomUnderSampler	46
4.9.5 Model Undersampling ClusterCentroids	46
4.10 Pembahasan	47
BAB V PENUTUP	51
5.1 Kesimpulan	51
3.2 Saran	51
REFERENSI	52

DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian	7
Tabel 2. 2 <i>Confusion Matrix</i>	19
Tabel 4. 1 Pemilihan Kolom	26
Tabel 4. 2 <i>Cleaning Data</i>	26
Tabel 4. 3 <i>Tokenizing</i>	27
Tabel 4. 4 <i>Filtering</i>	27
Tabel 4. 5 Penghilangan token kurang dari 4 karakter	28
Tabel 4. 6 <i>Stemming</i>	28
Tabel 4. 7 Hasil Pelabelan Data	29
Tabel 4. 8 Jumlah Distribusi Kelas	29
Tabel 4. 9 Pengurangan Data Positif Menjadi 1000	30
Tabel 4. 10 Pengurangan Data Positif Menjadi 400	31
Tabel 4. 11 Hasil Pembagian Data Skenario 1.....	34
Tabel 4. 12 Hasil Pembagian Data Skenario 2.....	35
Tabel 4. 13 Hasil Pembagian Data Skenario 3.....	35
Tabel 4. 14 Hasil Evaluasi Skenario 1 Tanpa Sampling.....	36
Tabel 4. 15 Jumlah Data Skenario 1 <i>SMOTE</i>	37
Tabel 4. 16 Hasil Evaluasi Skenario 1 <i>SMOTE</i>	37
Tabel 4. 17 Jumlah Data Skenario 1 <i>RandomOverSampler</i>	38
Tabel 4. 18 Hasil Evaluasi Skenario 1 <i>RandomOverSampler</i>	38
Tabel 4. 19 Jumlah Data Skenario 1 <i>RandomUnderSampler</i>	38
Tabel 4. 20 Hasil Evaluasi Skenario 1 <i>RandomUnderSampler</i>	39
Tabel 4. 21 Jumlah Data Skenario 1 <i>ClusterCentroids</i>	39
Tabel 4. 22 Hasil Evaluasi Skenario 1 <i>ClusterCentroids</i>	39
Tabel 4. 23 Hasil Evaluasi Skenario 2 Tanpa Sampling.....	40
Tabel 4. 24 Jumlah Data Skenario 2 <i>SMOTE</i>	41
Tabel 4. 25 Hasil Evaluasi Skenario 2 <i>SMOTE</i>	41
Tabel 4. 26 Jumlah Data Skenario 2 <i>RandomOverSampler</i>	41
Tabel 4. 27 Hasil Evaluasi Skenario 2 <i>RandomOverSampler</i>	42

Tabel 4. 28 Jumlah Data Skenario 2 <i>RandomUnderSampler</i>	42
Tabel 4. 29 Hasil Evaluasi Skenario 2 <i>RandomUnderSampler</i>	42
Tabel 4. 30 Jumlah Data Skenario 2 <i>ClusterCentroids</i>	43
Tabel 4. 31 Hasil Evaluasi Skenario 2 <i>ClusterCentroids</i>	43
Tabel 4. 32 Hasil Evaluasi Skenario 3 Tanpa Sampling.....	44
Tabel 4. 33 Jumlah Data Skenario 3 <i>SMOTE</i>	44
Tabel 4. 34 Hasil Evaluasi Skenario 3 <i>SMOTE</i>	44
Tabel 4. 35 Jumlah Data Skenario 3 <i>RandomOverSampler</i>	45
Tabel 4. 36 Hasil Evaluasi Skenario 3 <i>RandomOverSampler</i>	45
Tabel 4. 37 Jumlah Data Skenario 3 <i>RandomUnderSampler</i>	46
Tabel 4. 38 Hasil Evaluasi Skenario 3 <i>RandomUnderSampler</i>	46
Tabel 4. 39 Jumlah Data Skenario 3 <i>ClusterCentroids</i>	46
Tabel 4. 40 Hasil Evaluasi Skenario 3 <i>ClusterCentroids</i>	47
Tabel 4. 41 Perbandingan Performa Model Kelas Positif.....	48
Tabel 4. 42 Perbandingan Performa Model Kelas Negatif	48

DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian	21
Gambar 4. 1 Jumlah Distribusi Kelas Dalam Persen	30
Gambar 4. 2 <i>TF-IDF</i> Skenario 1: 2298 kelas positif	31
Gambar 4. 3 <i>TF-IDF</i> Skenario 2: 1000 kelas positif	32
Gambar 4. 4 <i>TF-IDF</i> Skenario 3: 400 kelas positif	33
Gambar 4. 5 Distribusi Data Skenario 1 Dalam Persen	34
Gambar 4. 6 Distribusi Data Skenario 2 Dalam Persen	35
Gambar 4. 7 Distribusi Data Skenario 3 Dalam Persen	36
Gambar 4. 8 Visualisasi Perbandingan Kelas Positif	49
Gambar 4. 9 Visualisasi Perbandingan Kelas Negatif	50



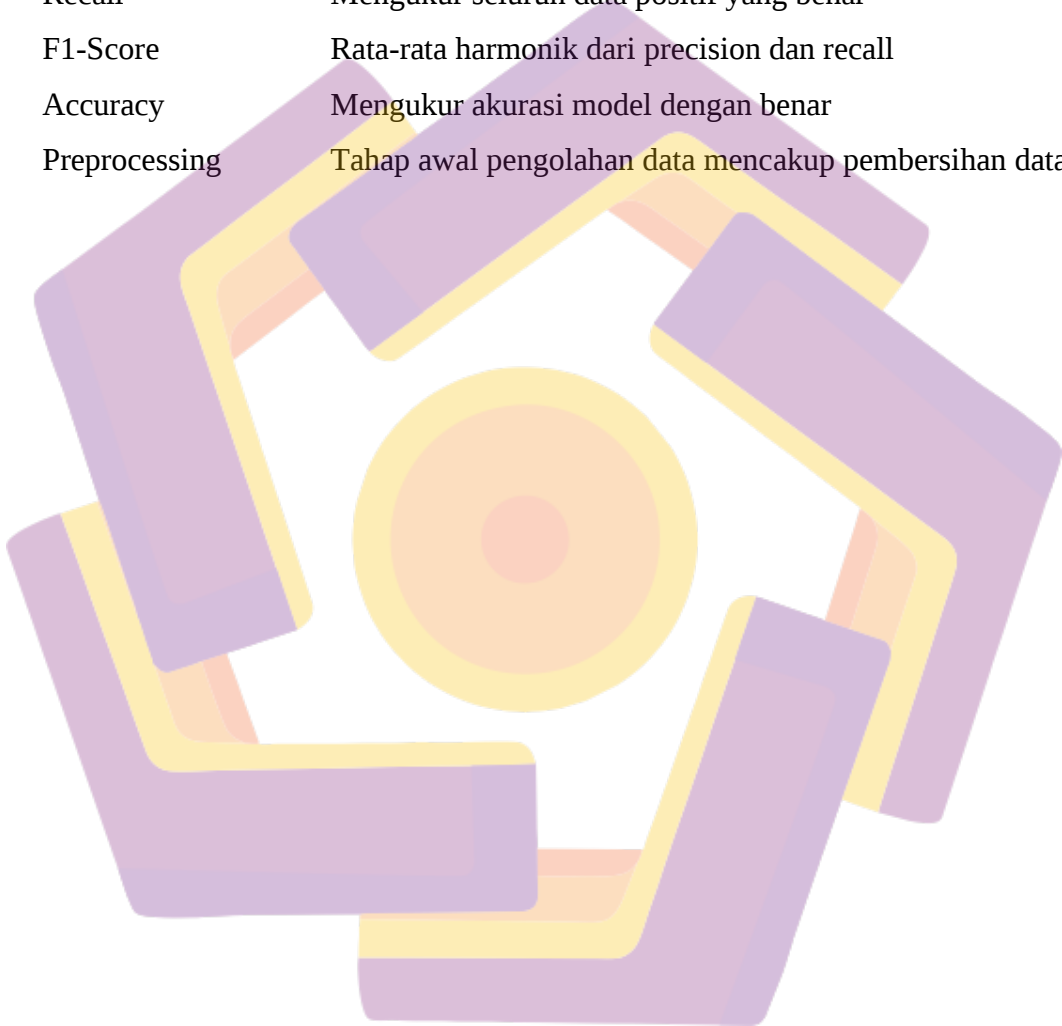
DAFTAR LAMBANG DAN SINGKATAN



OCR	Optical Character Recognition
TF	Term Frequency
IDF	Inverse Document Frequency
TF-IDF	Term Frequency - Inverse Document Frequency
SMOTE	Synthetic Minority Over-sampling Technique
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
CART	Classification and Regression Tree
P_i	Probabilitas kemunculan term dalam dokumen
i	Jumlah dokumen

DAFTAR ISTILAH

Dataset	Sekumpulan data
Confusion Matrix	Matriks berisi TP, TN, FP, FN
Precision	Mengukur ketepatan model data positif
Recall	Mengukur seluruh data positif yang benar
F1-Score	Rata-rata harmonik dari precision dan recall
Accuracy	Mengukur akurasi model dengan benar
Preprocessing	Tahap awal pengolahan data mencakup pembersihan data



INTISARI

Perkembangan teknologi pemindai dokumen semakin pesat dengan kehadiran aplikasi berbasis mobile seperti Adobe Scan. Dengan teknologi OCR dan integrasi kecerdasan buatan, Adobe Scan menjadi solusi bagi pengguna dalam mendigitalkan dokumen secara praktis. Dengan lebih dari 100 juta unduhan dan rating tinggi di Google Play Store, ulasan pengguna menjadi sumber informasi penting untuk mengevaluasi dan memahami kebutuhan pengguna.

Penelitian ini dilakukan untuk mengklasifikasikan sentimen menjadi positif dan negatif. Data yang digunakan berasal dari ulasan pengguna aplikasi Adobe Scan di Google Play Store. Untuk mengatasi ketidakseimbangan data, diterapkan berbagai teknik *sampling*, yaitu tanpa *sampling*, *oversampling* (*SMOTE* dan *RandomOverSampler*), serta *undersampling* (*RandomUnderSampler* dan *ClusterCentroids*). Model klasifikasi menggunakan algoritma *Random Forest* dan dievaluasi menggunakan *confusion matrix*.

Hasil penelitian menunjukkan bahwa algoritma *Random Forest* mampu mengklasifikasikan sentimen ulasan dengan baik. Pada kelas positif, *precision* tertinggi sebesar 97% pada skenario 1 dengan teknik *RandomUnderSampler*, *recall* tertinggi sebesar 100% pada skenario 1 dan 2 dengan teknik tanpa *sampling*, serta *f1-score* tertinggi sebesar 96% pada skenario 1 dengan teknik tanpa *sampling*, *SMOTE*, dan *RandomOverSampler*. Untuk kelas negatif, *precision* tertinggi sebesar 95% pada skenario 3 dengan teknik tanpa *sampling*, *recall* tertinggi sebesar 79% pada skenario 2 dengan *RandomUnderSampler*, dan *f1-score* tertinggi sebesar 74% pada skenario 3 dengan teknik *RandomOverSampler*. Penelitian ini diharapkan bermanfaat bagi pengembang aplikasi Adobe Scan untuk meningkatkan kualitas aplikasi.

Kata kunci: sentimen, Adobe Scan, *Random Forest*, Google Play Store, *sampling*.

ABSTRACT

The development of document scanner technology is increasingly rapid with the presence of mobile-based applications such as Adobe Scan. With OCR technology and artificial intelligence integration, Adobe Scan is a solution for users to digitize documents practically. With more than 100 million downloads and high ratings on the Google Play Store, user reviews are an important source of information to evaluate and understand user needs.

This study was conducted to classify sentiment into positive and negative. The data used came from user reviews of the Adobe Scan application on the Google Play Store. To overcome data imbalance, various sampling techniques were applied, namely without sampling, oversampling (SMOTE and RandomOverSampler), and undersampling (RandomUnderSampler and ClusterCentroids). The classification model uses the Random Forest algorithm and is evaluated using a confusion matrix.

The results of the study show that the Random Forest algorithm is able to classify review sentiment well. In the positive class, the highest precision is 97% in scenario 1 with the RandomUnderSampler technique, the highest recall is 100% in scenarios 1 and 2 with the technique without sampling, and the highest f1-score is 96% in scenario 1 with the technique without sampling, SMOTE, and RandomOverSampler. For the negative class, the highest precision is 95% in scenario 3 with the technique without sampling, the highest recall is 79% in scenario 2 with RandomUnderSampler, and the highest f1-score is 74% in scenario 3 with the RandomOverSampler technique. This research is expected to be useful for Adobe Scan application developers to improve the quality of the application.

Keyword: sentiment, Adobe Scan, Random Forest, Google Play Store, sampling.