

**ANALISIS SENTIMEN MENGGUNAKAN CLASSIFIER SVM
KERNEL RBF DAN PENDEKATAN ENSEMBLE LEARNING
DENGAN RANDOM FOREST**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

RIRIN HARTI NINGSIH

20.11.3628

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

**ANALISIS SENTIMEN MENGGUNAKAN CLASSIFIER SVM
KERNEL RBF DAN PENDEKATAN ENSEMBLE LEARNING
DENGAN RANDOM FOREST**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

RIRIN HARTI NINGSIH

20.11.3628

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

202

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN MENGGUNAKAN CLASSIFIER SVM
KERNEL RBF DAN PENDEKATAN ENSEMBLE LEARNING
DENGAN RANDOM FOREST**

yang disusun dan diajukan oleh

Ririn Harti Ningsih

20.11.3628

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 19 Maret 2024

Dosen Pembimbing,



Hastari Utama, M.Cs

NIK. 190302230

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN MENGGUNAKAN CLASSIFIER SVM
KERNEL RBF DAN PENDEKATAN ENSEMBLE LEARNING
DENGAN RANDOM FOREST**

yang disusun dan diajukan oleh

Ririn Harti Ningsih

20.11.3628

Telah dipertahankan di depan Dewan Penguji
pada tanggal 19 Maret 2024

Susunan Dewan Penguji

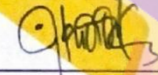
Nama Penguji

Anna Baita, M.Kom
NIK. 190302290

Ali Mustopa, M.Kom
NIK. 190302192

Hastari Utama, M.Cs
NIK. 190302230

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 19 Maret 2024

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 : RIRIN HARTI NINGSIH
NIM : 20.11.3628

Menyatakan bahwa Skripsi dengan judul berikut:

**ANALISIS SENTIMEN MENGGUNAKAN CLASSIFIER SVM KERNEL RBF
DAN PENDEKATAN ENSEMBLE LEARNING DENGAN RANDOM FOREST**

Dosen Pembimbing : Hastari Utama, M.Cs

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, 19 Maret 2024

Yang Menyatakan,



Ririn Harti Ningsih

HALAMAN PERSEMBAHAN

Alhamdulillah segala puji Syukur atas rahmat dan karunia Allah SWT yang telah memberikan petunjuk dan kemudahan bagi penulis dalam menyelesaikan skripsi ini dengan sebaik-baiknya. Dengan ini penulis mempersembahkan skripsi ini kepada semua pihak yang telah membantu dengan memberikan motivasi serta dukungan yang besar sehingga skripsi ini terselesaikan dengan sebaik-baiknya.

Melalui skripsi ini penulis mempersembahkan terima kasih kepada :

1. Dengan penuh rasa Syukur kepada Allah SWT yang telah memberikan rahmat dan karuniaNya sehingga penulis memiliki kemampuan untuk menyelesaikan skripsi ini.
2. Kepada keluarga penulis Bapak M.Syafiri , Ibu Sri Harti, Kakak Okta Yelmita serta Ade Susilo,dan keponakan penulis Maryam Nafisah dan Emir faqih yang tak hentinya memberikan doa, cinta kasih dan dukungannya untuk penulis.
3. Sahabat “Till Jannah” yaitu Nurul Zalza, Sindhi, Rafli, dan Ririn Nurdiyanti yang telah menemani penulis selama perkuliahan ,saling mendukung serta tempat berkeluh kesah.
4. Sahabat penulis semasa sekolah yaitu Nessa Tuzahara dan Mitha Rahmasarita yang menemani penulis, memberikan dukungan serta mendengarkan keluh kesah penulis.
5. Teman-Teman SMA yaitu Regina Melin, Regina Dita, dan Fania Auliaa Kahar.Terima kasih selalu dari masa Les hingga sekarang yang menemani penulis .
6. Rifky Akbar Maulana yang memberikan dukungan serta kasih sayang dan menemani penulis dalam menyelesaikan skripsi ini.
7. Terakhir,untuk diri sendiri terima kasih telah kuat berjuang dalam menyelesaikan skripsi ini dengan tepat waktu.

KATA PENGANTAR

Alhamdulillah rabbil 'alamin, segala puji dan Syukur penulis panjatkan atas kehadiran Allah subhanahu wa ta 'ala segala rahmat dan karunia-Nya sehingga penulis mampu menyelesaikan skripsi dengan judul “ANALISIS SENTIMEN MENGGUNAKAN CLASSIFIER SVM KERNEL RBF DAN PENDEKATAN ENSAMBLE LEARNING DENGAN RANDOM FOREST” dengan tepat waktu. Skripsi ini ditulis untuk memenuhi syarat kelulusan untuk mencapai Gelar Sarjana pada Fakultas Ilmu Komputer Program Studi Informatika Universitas Amikom Yogyakarta.

Dalam menyelesaikan penulisan skripsi ini, penulis banyak memperoleh bantuan baik, pengajaran dan arahan dari berbagai pihak secara langsung maupun tidak langsung. Dengan ini penulis menyampaikan terimakasih kepada:

1. Bapak prof. Dr. M. Suyanto, M.M. selaku rector Universitas Amikom Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom., Ph.D. selaku Dekan fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Ibu Windha Mega Pradnya D, M.Kom. selaku ketua program Studi S1 Informatika Universitas Amikom Yogyakarta.
4. Bapak Hastari Utama, M.Cs selaku dosen pembimbing saya yang telah memberikan waktu, saran, kritik, motivasi dan bimbingan dalam skripsi ini sehingga dapat menyelesaikan skripsi selesai dengan baik.
5. Bapak M. Syafiri dan Ibu Sri Harti selaku orang tua penulis yang memberikan dukungan, doa dan motivasi kepada penulis.

Penulis menyadari bahwa skripsi ini masih banyak kekurangan karena keterbatasan pengetahuan yang dimiliki oleh penulis. Oleh karena itu kritik dan saran sangat diharapkan bagi perbaikan penelitian selanjutnya.

Yogyakarta, 28 Februari 2024

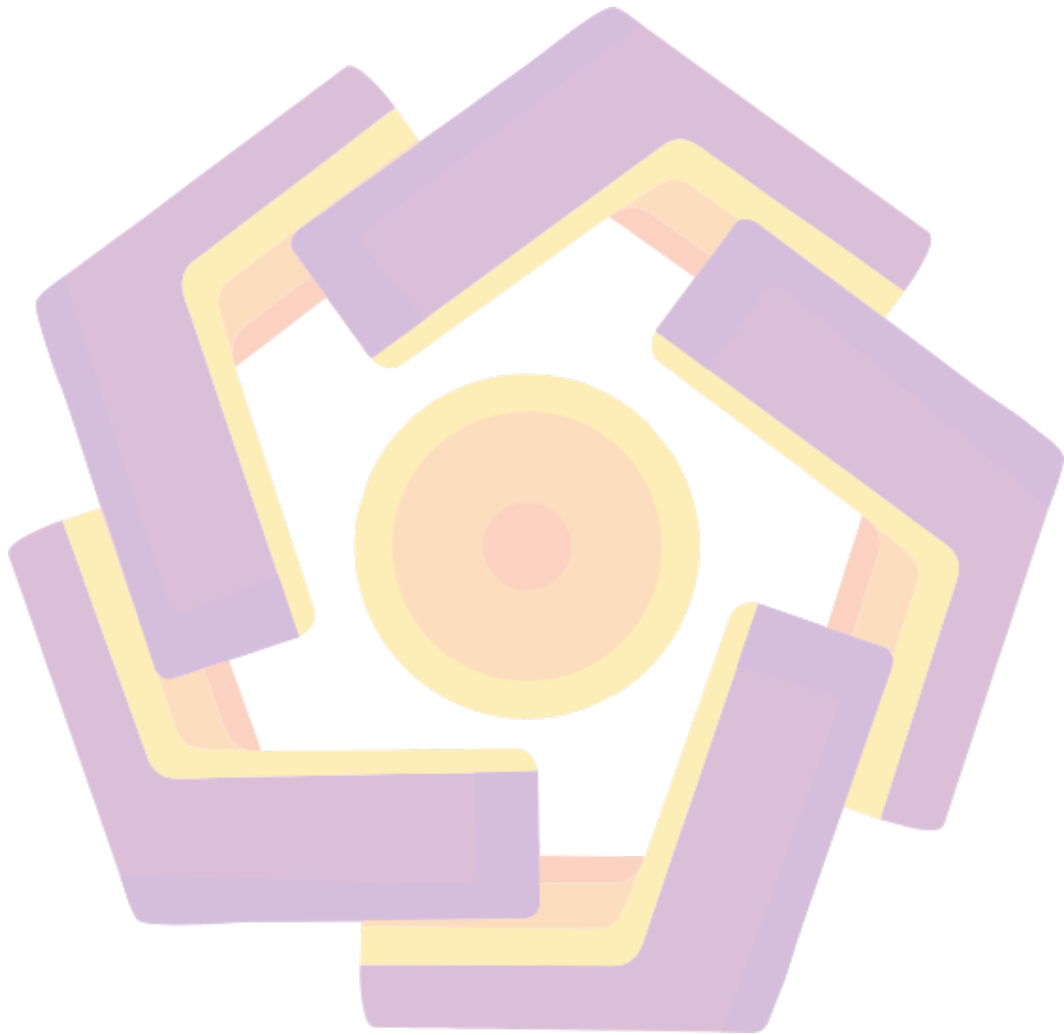
Ririn Harti Ningsih

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	x
DAFTAR GAMBAR	xi
DAFTAR ISTILAH	xiii
INTISARI	xiv
ABSTRACT	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	7
2.1 Studi Literatur	7
2.1 Dasar Teori	14
2.2.1 Analisis Sentimen	14
2.2.2 Dana	14
2.2.1 <i>Text Mining</i>	14
2.2.2 <i>Support Vector Machine</i>	15
2.2.3 <i>Adaptive Boosting (Adabost)</i>	16
2.2.4 <i>Bagging</i>	16
2.2.5 <i>Random Forest</i>	17

BAB III METODE PENELITIAN	19
3.1 <i>Dataset</i>	19
3.2 Alur Penelitian	19
3.2.1 <i>Scraping Data</i>	20
3.2.2 <i>Labelling Data</i>	21
3.2.3 <i>Preprocessing Data</i>	22
3.2.4 <i>Wordcloud</i>	23
3.2.5 <i>Pemodelan Data</i>	23
3.3 Alat dan Bahan.....	28
3.3.1 Alat Penelitian	28
3.3.2 Bahan Penelitian	29
BAB IV HASIL DAN PEMBAHASAN	31
4.1 Pengambilan Data	31
4.2 <i>Preprocessing Data</i>	34
4.2.1 <i>Case Folding</i>	34
4.2.2 <i>Data Cleaning</i>	35
4.2.3 <i>Tokenizing</i>	36
4.2.4 <i>Stopword Removal</i>	37
4.2.5 <i>Stemming</i>	38
4.2.6 <i>Visualisasi Wordcloud</i>	38
4.3 Ekstrasi Fitur Menggunakan TF-IDF.....	40
4.3.1 TF-IDF Pada <i>K-fold</i>	40
4.3.2 TF-IDF Pada Data <i>Training</i> 80% dan Data <i>Testing</i> 20%	41
4.4 Evaluasi	42
4.4.1 <i>Confusion Matrix</i> pada <i>K-fold</i>	42
4.4.2 <i>Confusion Matrix</i> pada Data <i>Training</i> 80% dan Data <i>Testing</i> 20%	54
4.5 Hasil	64
4.5.1 <i>Modelling</i> dengan <i>K-fold</i>	64
4.5.2 <i>Modelling</i> dengan Data <i>Training</i> 80% dan Data <i>Testing</i> 20%.....	66

BAB V PENUTUP	69
5.1 Kesimpulan	69
5.2 Saran	70
REFERENSI.....	71



DAFTAR TABEL


Tabel 2. 1 Keaslian Penelitian	9
Tabel 3. 1 Hasil Labeling data.....	21
Tabel 3. 2 Data Traning dan Data Testing	28
Tabel 3. 3 Contoh Dataset.....	29
Tabel 4. 1 Data hasil Scrapping.....	31
Tabel 4. 2 Code Labelling	32
Tabel 4. 3 Label positif ,Negatif, dan netral.....	32
Tabel 4. 4 Jumlah Dataset Berdasarkan Sentiment	33
Tabel 4. 5 Case Folding	34
Tabel 4. 6 Data Cleaning.....	35
Tabel 4. 7 Tokenizing	36
Tabel 4. 8 Stopword Removal.....	37
Tabel 4. 9 Hasil Stemming.....	38
Tabel 4. 10 Hasil Akurasi Pemodelan Pembagian Data Rasio 80:20	67

DAFTAR GAMBAR

Gambar 2. 1 Cara kerja Random Forest.....	17
Gambar 3. 1 Grafik Jumlah Dataset.....	19
Gambar 3. 2 Alur Penelitian.....	20
Gambar 3. 3 Ilustrasi Support Vector Machine.....	24
Gambar 3. 4 Alur kerja algoritma SVM	25
Gambar 3. 5 Alur kerja SVM Rbf dan Random forest	26
Gambar 4. 1 Grafik Sentimen.....	33
Gambar 4. 2 Code dari Case Folding.....	34
Gambar 4. 3 code pada Data cleaning	35
Gambar 4. 4 Code pada Tokenizing	36
Gambar 4. 5 Code dari <i>Stopword Removal</i>	37
Gambar 4. 6 Code Stemming Data pada Colab.....	38
Gambar 4. 7 Wordcloud Positif.....	39
Gambar 4. 8 Wordcloud Negatif	39
Gambar 4. 9 Wordcloud Netral	40
Gambar 4. 10 Split data pada K-Fold	41
Gambar 4. 11 TF-IDF pada K-Fold.....	41
Gambar 4. 12 TF-IDF pada Data Train dan Data Test	41
Gambar 4. 13 SMOTE Data pada modelling Data Train dan Test.....	42
Gambar 4. 14 Split Data untuk Modelling Data Train dan Testing.....	42
Gambar 4. 15 Code untuk Confusion Matrix pada SVM RBF	43
Gambar 4. 16 Confusion Matrix 5 Fold pada SVM RBF	44
Gambar 4. 17 Hasil 5-Fold SVM RBF	45
Gambar 4. 18 Code untuk Confusion Matrix SVM Linear.....	45
Gambar 4. 19 Confusion Matrix 5 Fold pada SVM Linear	46
Gambar 4. 20 Hasil 5-Fold SVM Linear.....	47
Gambar 4. 21 Code Confusion Matrix pada SVM Linear	47
Gambar 4. 22 Confusion Matrix 5 Fold pada SVM polynomial	48
Gambar 4. 23 Hasil 5-Fold SVM Polynomial.....	49
Gambar 4. 24 Code Confusion Matrix pada Metode Adaboost	49
Gambar 4. 25 Confusion Matrix 5 Fold pada Adaboost	50
Gambar 4. 26 Hasil 5 fold Adaboost	50
Gambar 4. 27 Code Confusion Matrix pada Metode Bagging.....	51
Gambar 4. 28 Confusion Matrix 5 Fold pada Bagging.....	52
Gambar 4. 29 Hasil 5-Fold Bagging.....	52
Gambar 4. 30 Code Confusion Matrix pada Metode Random Forest	53
Gambar 4. 31 Confusion Matrix 5 Fold pada Random Forest	54

Gambar 4. 32 Hasil 5-fold Random Forest	54
Gambar 4. 33 Code Confusion Matrix pada SVM RBF	55
Gambar 4. 34 Confusion Matrix pada SVM RBF	56
Gambar 4. 35 Hasil Confusion Matrix Metode SVM RBF.....	56
Gambar 4. 36 Code Confusion Matrix pada SVM Linear	57
Gambar 4. 37 Confusion Matrix pada SVM Linear	58
Gambar 4. 38 Hasil Confusion Matrix Metode SVM Linear.....	58
Gambar 4. 39 Code Confusion Matrix pada SVM Polynomial	59
Gambar 4. 40 Confusion Matrix pada SVM Polynomial.....	59
Gambar 4. 41 Hasil Confusion Matrix SVM Polynomial.....	60
Gambar 4. 42 Code Confusion Matrix Adaboost.....	60
Gambar 4. 43 Confusion Matrix pada Adaboost.....	61
Gambar 4. 44 Hasil Confusion Matrix Adaboost.....	61
Gambar 4. 45 Code Confusion Matrix pada Bagging	62
Gambar 4. 46 Confusion Matrix pada Bagging.....	62
Gambar 4. 47 Hasil Confusion Matrix Bagging.....	63
Gambar 4. 48 Code Confusion Matrix Random Forest	63
Gambar 4. 49 Confusion Matrix pada Random Forest.....	64
Gambar 4. 50 Hasil Confusion Matrix Random Forest	64
Gambar 4. 51 Perbandingan dari Setiap Fold dengan metode SVM Rbf,Linear, Polynomial,Adaboost dan Random Forest.....	65
Gambar 4. 52 Hasil Akurasi Fold dari Setiap Metode dengan SVM RBF,Linear,Polynomial,Adaboost,Bagging, dan Random Forest.	65
Gambar 4. 53 Grafik Akurasi dengan Pemodelan Pembagian Data Rasio 80:20 .	66

DAFTAR ISTILAH



Algoritma	metode yang direncanakan secara sistematis untuk menyelesaikan masalah
Data Training	sampel yang digunakan untuk melatih model
Data Testing	sampel yang digunakan untuk menguji model
Fold	lipatan atau bagian dari dataset yang digunakan selama iterasi dalam proses cross validation
Cross-Validation	teknik untuk mengevaluasi kinerja model
Akurasi	Tingkat ketepatan prediksi model dengan nilai sebenarnya

INTISARI

Banyaknya Perusahaan di Indonesia yang bergerak pada bidang fintech. Salah satu aplikasi yang memiliki banyak peminat di Indonesia adalah Aplikasi Dana. Tetapi, dibalik keberhasilan aplikasi Dana sendiri tentu saja memiliki Tingkat kepuasan berbeda pada pengguna dan pengalaman yang berbeda. Sehingga ulasan yang ada pada Google Playstore memiliki ulasan positif, negative, dan netral. Ulasan tersebut juga merupakan hal penting untuk perkembangan aplikasi kedepannya. Penelitian ini menggunakan algoritma SVM Rbf dan *Random Forest* untuk analisis sentiment data ulasan Aplikasi Dana yang didapat dari hasil scraping *Google Play Store*. Dataset tersebut dilakukan labeling untuk mempermudah preprocessing dataset dengan *case folding*, *data cleaning*, *tokenizing*, *stop Removal*, *stemming*, dan SMOTE. Kemudian menggunakan teknik *split data* dengan menggunakan dua cara untuk melihat perbandingan pada metode *K-Fold Cross Validation* algoritma SVM Rbf mendapatkan nilai rata-rata akurasi fold 74% sedangkan untuk *split data* rasio 80:20 mendapatkan akurasi 94% sedangkan, algoritma *Random Forest* pada metode *K-Fold Cross Validation* mendapatkan nilai rata-rata akurasi 75% dan pada *split data* rasio 80:20 mendapatkan akurasi 95%. Dengan ini menyatakan bahwa *Random Forest* berkerja lebih baik untuk analisis sentiment ulasan Dana pada *Google Play Store* dikarenakan pada penelitian dengan menggunakan metode *split data* yang berbeda nilai rata-rata akurasi *Random Forest* lebih tinggi di bandingkan oleh algoritma SVM Rbf.

Kata kunci: Analisis Sentimen, Ulasan Dana, SVM RBF, Random Forest

ABSTRACT

There are many companies in Indonesia that operate in the fintech sector. One application that has a lot of interest is Dana Application. However, behind the success of the Dana application itself, it has different experiences for each user and also different levels of satisfaction. Users' review in Google Playstore comes in positive, negative and neutral outcome. Thus, these reviews are important for future development of application. This research employs the SVM Rbf and Random Forest algorithms for sentiment analysis of user reviews on the Dana application obtained from scraping the Google Play Store. The dataset is labeled to facilitate preprocessing with case-folding, data cleaning, tokenizing, stop removal, stemming, and SMOTE. Subsequently, the data is split using two methods to compare the K-Fold Cross Validation method. For the SVM Rbf algorithm, the average accuracy fold is 74% in the K-Fold Cross Validation method, while for the 80:20 data split ratio, the accuracy is 94%. The Random Forest algorithm, in the K-Fold Cross Validation method, achieves an average accuracy of 75%, and with an 80:20 data split ratio, the accuracy is 95%. This indicates that Random Forest performs better for sentiment analysis of Dana reviews on the Google Play Store. This conclusion is drawn from the research using different data split methods, where the average accuracy of Random Forest is higher compared to the SVM Rbf algorithm.

Keyword: Sentiment Analysis ,Dana Reviews , SVM RBF ,Random Forest