

**OPTIMASI METODE KLASIFIKASI MENGGUNAKAN
FASTTEXT DAN GRID SEARCH PADA ANALISIS
SENTIMEN ULASAN APLIKASI SEABANK**

SKRIPSI

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



disusun oleh

BAGAS RESTYA ERMAWAN

21.11.4182

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

**OPTIMASI METODE KLASIFIKASI MENGGUNAKAN
FASTTEXT DAN GRID SEARCH PADA ANALISIS
SENTIMEN ULASAN APLIKASI SEABANK**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

BAGAS RESTYA ERMAWAN

21.11.4182

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

HALAMAN PERSETUJUAN

SKRIPSI

**OPTIMASI METODE KLASIFIKASI MENGGUNAKAN
FASTTEXT DAN GRID SEARCH PADA ANALISIS
SENTIMEN ULASAN APLIKASI SEABANK**

yang disusun dan diajukan oleh

Bagas Restya Ermawan

21.11.4182

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 19 Desember 2024

Dosen Pembimbing,

Nuri Cahyono, M.Kom

NIK. 190302278

HALAMAN PENGESAHAN

SKRIPSI

**OPTIMASI METODE KLASIFIKASI MENGGUNAKAN
FASTTEXT DAN GRID SEARCH PADA ANALISIS
SENTIMEN ULASAN APLIKASI SEABANK**

yang disusun dan diajukan oleh

Bagas Restya Ermawan

21.11.4182

Telah dipertahankan di depan Dewan Penguji
pada tanggal 19 Desember 2024

Susunan Dewan Penguji

Nama Penguji

Ali Mustopa, S.Kom., M.Kom.
NIK. 190302192

Agung Nugroho, M.Kom.
NIK. 190302242

Nuri Cahyono, M.Kom.
NIK. 190302378

Tanda Tangan



Skrripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 19 Desember 2024

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : **Bagas Restya Ermawan**
NIM : **21.11.4182**

Menyatakan bahwa Skripsi dengan judul berikut:

Optimasi Metode Klasifikasi Menggunakan Fasttext Dan Grid Search Pada Analisis Sentimen Ulasan Aplikasi Seabank

Dosen Pembimbing : **Nuri Cahyono, M.Kom**

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 Desember 2024

Yang Menyatakan,



Bagas Restya Ermawan

HALAMAN PERSEMBAHAN

Terwujudnya Skripsi ini tidak akan selesai tanpa bantuan dari berbagai pihak yang terus mendorong peneliti untuk menyelesaikannya. Oleh karena itu, peneliti mengucapkan terima kasih kepada:

1. Nuri Cahyono, M.kom selaku Dosen Pembimbing yang selalu memberikan nasihat, bimbingan, serta waktu yang telah diluangkan dalam memberikan arahan dan koreksi yang sangat berharga selama proses penyusunan Skripsi ini. Bimbingan dan dukungan yang diberikan sangat membantu peneliti dalam menyelesaikan penelitian ini dengan lebih baik.
2. Keluarga Peneliti, bapak dan ibu tersayang yang telah memberikan dukungan, baik moril maupun materil dan senantiasa memberikan doa yang tiada henti-hentinya kepada peneliti. Tidak lupa, Kakak peneliti yang selalu menghibur dan memberikan dukungan penuh kepada peneliti.
3. Anisha Putri Octaviyanti, yang selalu memberikan dukungan, semangat, doa, dan motivasi yang tiada henti selama penyusunan skripsi ini. Kehadirannya selalu menjadi sumber inspirasi dan kebahagiaan yang membuat peneliti mampu menyelesaikan penelitian ini dengan baik.
4. Teman-teman RAM yaitu Aldino Marsel Pratama dan Mahendra Bayu Prayoga yang telah berjuang bersama dalam masa studi dan penyusunan Skripsi. Hiburan, diskusi, kerjasama, dan bantuan yang diberikan sangat berarti bagi peneliti.
5. *Last but not least*, peneliti juga ingin menyampaikan terima kasih kepada diri saya sendiri, terima kasih telah bertahan di setiap tantangan, menjaga semangat di tengah kesulitan, dan tidak menyerah meski rintangan terasa berat.

KATA PENGANTAR

Puji syukur saya panjatkan kepada Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya yang melimpah, sehingga saya dapat menyelesaikan skripsi ini dengan baik. Skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan pendidikan Strata 1 di Program Studi Informatika, Universitas Amikom Yogyakarta.

Selama proses penyusunan skripsi ini, saya menerima banyak dukungan, arahan, dan bimbingan dari berbagai pihak yang telah berperan penting. Dengan rasa hormat dan terima kasih yang mendalam, saya sampaikan apresiasi yang tulus kepada bapak, ibu, serta kakak tercinta yang senantiasa memberikan doa, dukungan moral, serta kasih sayang yang tiada henti. Ucapan terima kasih yang sebesar-besarnya juga saya haturkan kepada Bapak Nuri Cahyono, M.Kom, selaku dosen pembimbing, atas bimbingan, ilmu, dan arahan berharga yang beliau berikan sepanjang proses penulisan skripsi ini.

Tidak lupa, saya mengucapkan terima kasih kepada pacar dan teman-teman yang telah memberikan semangat, dukungan, dan kebersamaan yang berarti selama proses penyusunan skripsi ini. Setiap bantuan, baik besar maupun kecil, sangat berharga dan membantu saya dalam menyelesaikan tahap demi tahap penelitian ini.

Saya menyadari bahwa skripsi ini masih memiliki kekurangan dan jauh dari sempurna. Oleh karena itu, saya dengan tangan terbuka menerima kritik dan saran yang membangun guna perbaikan di masa yang akan datang. Semoga skripsi ini dapat memberikan manfaat bagi para pembaca, berkontribusi terhadap pengembangan ilmu pengetahuan, dan memberi dampak positif bagi masyarakat. Dengan segala kerendahan hati, saya ucapkan terima kasih.

Yogyakarta, 19 Desember 2024

Penulis

DAFTAR ISI

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

2.2	Dasar Teori.....	12
2.2.1	Analisis Sentimen	12
2.2.2	Aplikasi <i>Seabank</i>	12
2.2.3	<i>Support Vector Machine</i> (SVM).....	13
2.2.4	<i>K-Nearest Neighbors</i> (KNN).....	16
2.2.5	<i>Fasttext Word Embedding</i>	18
2.2.6	<i>Grid Search Cross Validation</i>	20
2.2.7	<i>Confusion Matrix</i>	22
BAB III METODE PENELITIAN		25
3.1	Alur Penelitian	25
3.1.1	<i>Data Gathering</i>	26
3.1.2	<i>Text Preprocessing</i>	26
3.1.4	<i>Labelling</i>	31
3.1.5	Ekstraksi Fitur (<i>Fasttext Word Embedding</i>).....	31
3.1.6	SMOTE	31
3.1.7	<i>Split Data</i>	32
3.1.8	<i>K-Nearest Neighbors</i> (KNN).....	32
3.1.9	<i>Support Vector Machine</i> (SVM).....	32
3.1.10	<i>Grid Search</i>	33
3.1.11	<i>Confusion Matrix</i>	33
3.2	Alat dan Bahan.....	33
BAB IV HASIL DAN PEMBAHASAN		35
4.1	<i>Data Gathering</i>	35
4.2	<i>Text Preprocessing</i>	36
4.2.1	<i>Case Folding</i>	36

4.2.2	<i>Text Cleaning</i>	37
4.2.3	<i>Slang Word Normalization</i>	37
4.2.4	<i>Stopwords Removal</i>	38
4.2.5	<i>Tokenizing</i>	38
4.2.6	<i>Stemming</i>	38
4.3	<i>Labelling</i>	39
4.4	<i>Fasttext Word Embedding</i>	40
4.5	SMOTE.....	41
4.6	<i>Split Data</i>	41
4.7	Hasil <i>Text Mining</i>	42
4.7.1	Hasil Klasifikasi Algoritma KNN.....	44
4.7.2	Hasil Klasifikasi Algoritma KNN dengan <i>Grid Search</i>	47
4.7.3	Hasil Klasifikasi Algoritma SVM.....	53
4.7.4	Hasil Klasifikasi Algoritma SVM dengan <i>Grid Search</i>	55
4.7.5	Hasil Perbandingan Metode Klasifikasi.....	59
BAB V PENUTUP		61
5.1	Kesimpulan	61
5.2	Saran	61
REFERENSI		63
LAMPIRAN.....		67

DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	7
Tabel 2.2. Confusion Matrix	23
Tabel 3.1 Contoh Tahap <i>Case Folding</i>	26
Tabel 3.2 Contoh Tahap <i>Cleaning</i>	27
Tabel 3.3 Contoh Tahap <i>Stang Words Normalization</i>	28
Tabel 3.4 Contoh Proses <i>Stopword Removal</i>	29
Tabel 3.5 Contoh Tahap <i>Tokenizing</i>	29
Tabel 3.6 Contoh Proses <i>Stemming</i>	30
Tabel 4.1 Hasil <i>Grid Search KNN</i>	47
Tabel 4.2 Hasil <i>Grid Search SVM</i>	56
Tabel 4.3 Tabel perbandingan hasil metode klasifikasi	59

DAFTAR GAMBAR

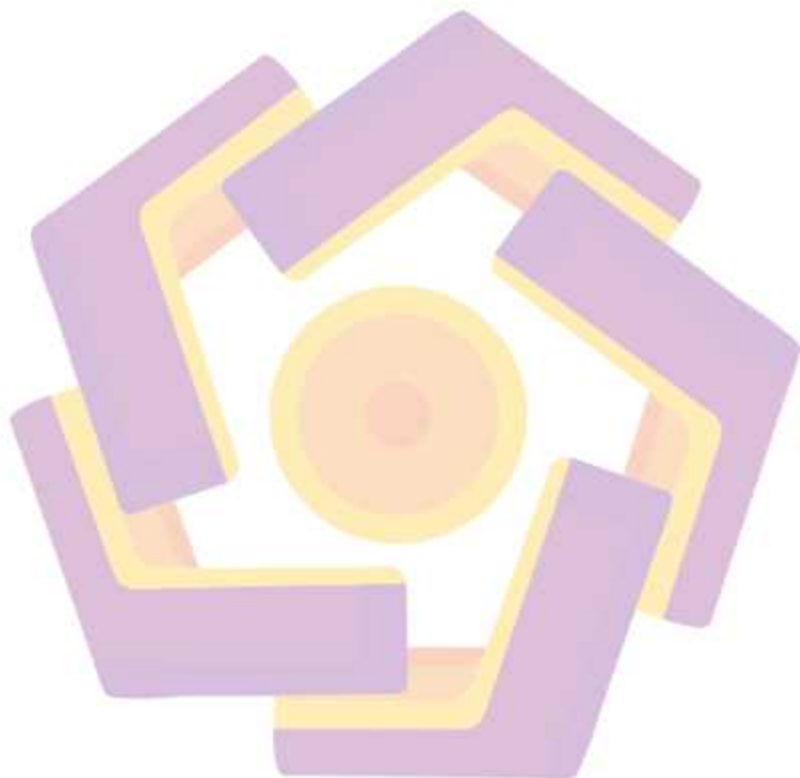
Gambar 2.1. Logo <i>Seabank</i>	12
Gambar 2.2. <i>Hyperplane</i> pada metode SVM.....	14
Gambar 2.3. Parameter <i>C</i> di metode SVM.....	15
Gambar 2.4. perbandingan besaran nilai <i>C</i>	16
Gambar 2.5. Algoritma <i>K-Nearest Neighbors (KNN)</i>	18
Gambar 2.6. Arsitektur model <i>Skip-gram</i>	19
Gambar 2.7. Cara kerja <i>Fasttext</i>	20
Gambar 2.8. Perbedaan <i>Grid Search</i> dan <i>Random Search</i>	21
Gambar 3.1 Alur Penelitian.....	25
Gambar 4.1. Data ulasan <i>Seabank</i>	35
Gambar 4.2. Data ulasan <i>Seabank</i> setelah pemilihan fitur.....	36
Gambar 4.3. Hasil <i>Case Folding</i>	36
Gambar 4.4. Hasil <i>Text Cleaning</i>	37
Gambar 4.5. Hasil <i>Slang Word Normalization</i>	37
Gambar 4.6. Hasil <i>Stopwords Removal</i>	38
Gambar 4.7. Hasil <i>Tokenizing</i>	38
Gambar 4.8. Hasil <i>Stemming</i>	39
Gambar 4.9. Hasil <i>Labelling</i>	39
Gambar 4.10. Diagram Persebaran Data.....	40
Gambar 4.11. Hasil <i>Fasttext Word Embedding</i>	40
Gambar 4.12. Persebaran data setelah SMOTE.....	41
Gambar 4.13. Hasil klasifikasi model KNN menggunakan TF-IDF.....	42
Gambar 4.14. <i>Confusion matrix</i> model KNN menggunakan TF-IDF.....	43

Gambar 4.15. Hasil klasifikasi model KNN menggunakan <i>fasttext</i>	44
Gambar 4.16. <i>Confusion matrix</i> model KNN menggunakan <i>fasttext</i>	46
Gambar 4.17. Hasil klasifikasi model KNN dengan <i>Grid Search</i>	48
Gambar 4.18. <i>Confusion matrix</i> model KNN dengan <i>Grid Search</i>	49
Gambar 4.19. Hasil klasifikasi model SVM dengan TF-IDF	51
Gambar 4.20. <i>Confusion matrix</i> model SVM dengan TF-IDF	52
Gambar 4.21. Hasil klasifikasi model SVM menggunakan <i>Fasttext</i>	53
Gambar 4.22. <i>Confusion matrix</i> model SVM menggunakan <i>Fasttext</i>	54
Gambar 4.23. Hasil klasifikasi model SVM dengan <i>Grid Search</i>	56
Gambar 4.24. <i>Confusion matrix</i> model SVM dengan <i>Grid Search</i>	58



DAFTAR LAMPIRAN

Lampiran 1 Kode <i>Scraping Data</i>	67
Lampiran 2 Kode <i>Text Mining</i>	67



DAFTAR LAMBANG DAN SINGKATAN



ξ	<i>Slack variable</i>
Σ	<i>Sigma</i>
NLP	<i>Natural Language Processing</i>
KNN	<i>K-Nearest Neighbors</i>
SVM	<i>Support Vector Machines</i>
DCT	<i>Decision Tree</i>
RF	<i>Random Forest</i>
LR	<i>Logistic Regression</i>
PSO	<i>Particle Swarm Optimization</i>
RBF	<i>Radius Basis Function</i>
VADER	<i>Valence Aware Dictionary for Sentiment Reasoning</i>
CV	<i>Cross Validation</i>
URL	<i>Uniform Resource Locators</i>
SMOTE	<i>Synthetic Minority Over-sampling Technique</i>

DAFTAR ISTILAH

<i>Hyperparameter</i>	Parameter model sebelum pelatihan.
<i>Tuning</i>	Mengoptimalkan hyperparameter
<i>Grid Search</i>	Optimasi dengan mencoba semua kombinasi parameter
<i>Fasttext</i>	Algoritma <i>embedding</i> kata dari Facebook AI
Klasifikasi	Mengelompokkan data ke kelas tertentu
Parameter	Nilai dalam model yang diatur saat pelatihan
Sentimen	Analisis emosi dalam teks
<i>Fintech</i>	Teknologi untuk layanan keuangan
<i>roBERTa</i>	Model NLP lanjutan dari BERT
<i>Naïve Bayes</i>	Algoritma klasifikasi probabilistik
<i>Smartphone</i>	Ponsel pintar
<i>E-commerce</i>	Perdagangan online
<i>Supervised Learning</i>	Pembelajaran dengan data berlabel
<i>Hyperplane</i>	Garis pemisah dalam SVM
<i>Margin</i>	Jarak antara data dan <i>hyperplane</i>
Regresi	Model untuk hubungan antar variabel
Vektor	Representasi nilai dengan arah
<i>Support Vectors</i>	Titik data penentu <i>margin</i>
<i>Kernel</i>	Fungsi pemetaan data di SVM
<i>Linear</i>	Hubungan langsung antar variabel
<i>Polynomial</i>	Hubungan dengan pangkat variabel
<i>Euclidean</i>	Jarak lurus antar dua titik
<i>Manhattan</i>	Jarak berdasarkan pergerakan kotak
<i>Cosine</i>	Kesamaan sudut antar vektor
<i>Correlation</i>	Ukuran hubungan antar variabel
<i>Hamming</i>	Jarak posisi berbeda dalam <i>string</i>
<i>Library</i>	Kumpulan kode siap pakai
<i>Word Embedding</i>	Representasi kata sebagai vektor

<i>Word2vec</i>	Algoritma untuk <i>embedding</i> kata
<i>Sufiks</i>	Akhiran kata
<i>Prefiks</i>	Awalan kata
<i>n-gram</i>	Urutan n kata berturut-turut
<i>Skip-gram</i>	Model prediksi kata sekitar
<i>Data mining</i>	Menemukan pola dalam data
<i>Confusion matrix</i>	Tabel evaluasi model
<i>Preprocessing</i>	Pemrosesan awal data
<i>Lowercase</i>	Mengubah teks jadi huruf kecil
<i>Slang</i>	Bahasa informal
<i>Machine learning</i>	Mesin belajar dari data
<i>Sastrawi</i>	<i>Library stemming</i> Bahasa Indonesia
<i>Bayesian optimization</i>	Optimasi menggunakan <i>Bayesian</i>
<i>Genetic algorithm</i>	Optimasi berbasis seleksi alam.

INTISARI

Kemajuan teknologi digital telah membawa perubahan besar dalam sektor perbankan, dengan hadirnya aplikasi perbankan digital seperti *SeaBank*. Namun, beberapa pengguna melaporkan masalah serius seperti kehilangan saldo atau keterlambatan transfer, yang memerlukan analisis mendalam terhadap persepsi pengguna melalui ulasan di *Google Play Store*. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna aplikasi *SeaBank* menggunakan metode klasifikasi *Support Vector Machine* (SVM) dan *K-Nearest Neighbors* (KNN) yang dioptimasi dengan *Grid Search* dan ekstraksi fitur *FastText*. Data yang digunakan sebanyak 15.000 ulasan dari periode 27 Juli 2024 hingga 26 September 2024, yang melalui tahapan *preprocessing*, *labelling manual* menjadi tiga kelas (positif, negatif, netral), dan penyeimbangan data menggunakan SMOTE. Hasil penelitian menunjukkan bahwa ekstraksi fitur *Fasttext* dan optimasi *Grid Search* berhasil meningkatkan performa kedua model, dengan akurasi KNN meningkat dari 87% menjadi 91% dan SVM dari 75% menjadi 92%. Model SVM dengan parameter optimal ($C: 100$, $kernel: 'rbf'$, $gamma: 'scale'$) menunjukkan performa terbaik dengan *precision* 0.93, *recall* 0.92, dan *F1-score* 0.92. Hasil penelitian ini dapat dimanfaatkan oleh tim pengembang *SeaBank* untuk meningkatkan kualitas layanan berdasarkan umpan balik pengguna, serta dapat dikembangkan lebih lanjut dengan menggunakan metode klasifikasi lain seperti *Random Forest* atau teknik ekstraksi fitur seperti BERT.

Kata Kunci: Analisis Sentimen, *SeaBank*, *Support Vector Machine*, *K-Nearest Neighbors*, *Grid Search*

ABSTRACT

Digital technology advancement has brought significant changes in the banking sector, with the emergence of digital banking applications like SeaBank. However, several users reported serious issues such as balance loss or transfer delays, requiring an in-depth analysis of user perception through Google Play Store reviews. This research aims to analyze user review sentiments of the SeaBank application using Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) classification methods optimized with Grid Search and FastText feature extraction. The study utilized 15,000 reviews collected from July 27, 2024, to September 26, 2024, which underwent preprocessing stages, manual labeling into three classes (positive, negative, neutral), and data balancing using SMOTE. The results showed that Grid Search optimization successfully improved both models' performance, with KNN accuracy increasing from 87% to 91% and SVM from 75% to 92%. The SVM model with optimal parameters (C: 100, kernel: 'rbf', gamma: 'scale') demonstrated the best performance with 0.93 precision, 0.92 recall, and 0.92 F1-score. These findings can be utilized by the SeaBank development team to improve service quality based on user feedback, and can be further developed using other classification methods such as Random Forest or feature extraction techniques like BERT.

Keywords: *sentiment analysis, SeaBank, Support Vector Machine, K-Nearest Neighbors, Grid Search*