



SKRIPSI

IMPLEMENTASI ENSEMBLE LEARNING UNTUK ANALISIS SENTIMEN PADA APLIKASI REKU

RATIH AISYAH

NPM 20082010190

DOSEN PEMBIMBING

Reisa Permatasari, S.T., M.Kom

Dhian Satria Yudha Kartika, S.Kom., M.Kom

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SISTEM INFORMASI
SURABAYA
2024**



SKRIPSI

IMPLEMENTASI ENSEMBLE LEARNING UNTUK ANALISIS SENTIMEN PADA APLIKASI REKU

RATIH AISYAH

NPM 20082010190

DOSEN PEMBIMBING

Reisa Permatasari, S.T., M.Kom

Dhian Satria Yudha Kartika, S.Kom., M.Kom

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SISTEM INFORMASI
SURABAYA
2024**

Halaman ini sengaja dikosongkan

LEMBAR PENGESAHAN

IMPLEMENTASI ENSEMBLE LEARNING UNTUK ANALISIS SENTIMEN PADA APLIKASI REKU

Oleh :

RATIH AISYAH
NPM. 20082010190

Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Skripsi Prodi Sistem Informasi
Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 11
Juni 2025

Menyetujui

Reisa Permatasari, S.T., M.Kom
NIP. 19920514 2022032 007

(Pembimbing I)

Dhian Satria Yudha Kartika, S.Kom., M.Kom
NPT. 2011986 052224 9

(Pembimbing II)

Nur Cahyo Wibowo, S.Kom., M.Kom
NIP. 19790317 2021211 002

(Penguji I)

Dr. Eng. Agussalim, M.T
NIP. 19850811 2019031 005

(Penguji II)

Iqbal Ramadhani Mukhlis, S.Kom., M.Kom
NIP. 19930305 2024061 002

(Penguji III)

Mengetahui,
Dekan Fakultas Ilmu Komputer


Prof. Dr. Ir. Novarina Hendrasarie, MT
NIP. 19681126 199403 2 001

Halaman ini sengaja dikosongkan

LEMBAR PERSETUJUAN

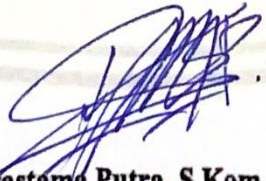
IMPLEMENTASI ENSEMBLE LEARNING UNTUK ANALISIS SENTIMEN PADA APLIKASI REKU

Oleh :

RATH AISYAH
NPM. 20082010190

Menyetujui,

Koordinator Program Studi Sistem Informasi
Fakultas Ilmu Komputer


Agung Brastama Putra, S.Kom, M.Kom.
NIP. 19851124 2021211 003

Halaman ini sengaja dikosongkan

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini :

Nama : Ratih Aisyah
NPM : 20082010190
Program : Sarjana (S1)
Program Studi : Sistem Informasi
Fakultas : Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis di sitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila di kemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya



ABSTRAK

Nama Mahasiswa / NPM : Ratih Aisyah / 20082010190
Judul Skripsi : Implementasi Ensemble Learning Untuk Analisis Sentimen Pada Aplikasi Reku
Dosen Pembimbing : 1. Reisa Permatasari, S.T., M.Kom.
2. Dhian Satria Yuda Kartika, S.Kom., M.Kom.

Popularitas cryptocurrency di Indonesia mengalami pertumbuhan signifikan, tercermin dari peningkatan jumlah investor kripto dan nilai transaksi yang terus membaik. Seiring meningkatnya minat masyarakat terhadap aplikasi finansial seperti Reku, analisis sentimen menjadi alat penting untuk memahami opini dan kepuasan pengguna. Namun, pendekatan konvensional dalam analisis sentimen memiliki keterbatasan dalam mengolah kompleksitas bahasa dan makna dalam ulasan pengguna.

Penelitian ini menerapkan pendekatan ensemble learning menggunakan kombinasi tiga arsitektur klasifikasi, yaitu Convolutional Neural Network (CNN), Support Vector Machine (SVM) dengan Word2Vec embedding, dan model transformer IndoBERT. Evaluasi dilakukan terhadap performa masing-masing model serta hasil gabungannya dalam ensemble menggunakan metode Soft Voting dan Weighted Soft Voting. Penilaian dilakukan dengan metrik akurasi, presisi, recall, dan F1-score untuk mengukur pengaruh penggabungan model terhadap peningkatan akurasi analisis sentimen. Hasil eksperimen menunjukkan bahwa pendekatan ensemble learning dengan metode Weighted Soft Voting memberikan akurasi terbaik, yaitu 91%, mengungguli performa masing-masing model tunggal. Sebagai implementasi, model optimal dikembangkan menjadi aplikasi berbasis web menggunakan Streamlit, yang dapat digunakan sebagai alat bantu dalam mengevaluasi opini pengguna terhadap aplikasi Reku. Aplikasi ini mampu menghasilkan akurasi sebesar 90% dalam mengklasifikasikan sentimen pengguna.

Kata Kunci : *Ensemble Learning*, Reku, Analisis Sentimen, CNN, SVM, IndoBERT, *Weighted Soft Voting*, *Word2Vec*, Streamlit.

ABSTRACT

Student Name / NPM : Ratih Aisyah / 20082010190

Thesis Title : Implementation Of Ensemble Learning for Sentiment Analysis on Reku Application

Advisors : 1. Reisa Permatasari, S.T., M.Kom.
2. Dhian Satria Yuda Kartika, S.Kom., M.Kom.

The popularity of cryptocurrency in Indonesia has experienced significant growth, as reflected in the increasing number of crypto investors and improving transaction volumes. With the growing public interest in financial applications like Reku, sentiment analysis has become an essential tool to understand user opinions and satisfaction. However, conventional sentiment analysis methods have limitations in processing the complexity of language and meaning in user reviews.

This study applies an ensemble learning approach by combining three classification architectures: Convolutional Neural Network (CNN), Support Vector Machine (SVM) with Word2Vec embedding, and the transformer-based model IndoBERT. The performance of each individual model as well as their combined ensemble results is evaluated using soft voting and weighted soft voting methods. The evaluation is carried out using accuracy, precision, recall, and F1-score metrics to measure the impact of model combination on improving sentiment analysis accuracy. Experimental results show that the ensemble learning approach with the weighted soft voting method achieves the best accuracy of 91%, outperforming each individual model. As an implementation, the optimal model is developed into a web-based application using Streamlit, which can be used as a tool to evaluate user opinions on the Reku application. This application is capable of achieving an accuracy of 90% in classifying user sentiments.

Keyword : Ensemble Learning, Reku, Sentiment Analysis, CNN, SVM, IndoBERT, Weighted Soft Voting, Word2Vec, Streamlit.

KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul "**Implementasi Ensemble Learning Untuk Analisis Sentimen Pada Aplikasi**" sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, **Universitas Pembangunan Nasional Veteran Jawa Timur**.

Dalam penyusunan skripsi ini, penulis mendapat banyak dukungan, bimbingan, serta bantuan dari berbagai pihak. Oleh karena itu, dengan segala hormat dan rasa terima kasih yang mendalam, penulis ingin menyampaikan kepada:

1. Ibu Prof. Dr. Ir. Novirina Hendrasari, M.T., selaku Dekan Fakultas Ilmu Komputer yang telah memfasilitasi serta dukungan dalam menyelesaikan studi di fakultas ini.
2. Bapak Agung Brastama Putra, S.Kom., M.Kom., selaku Koordinator Program Studi Sistem Informasi yang telah memfasilitasi serta dukungan dalam menyelesaikan studi di fakultas ini.
3. Bapak Dhian Satria Yudha Kartika, S.Kom., M.Kom., selaku dosen wali dan dosen pembimbing kedua, yang telah memberikan arahan, bimbingan, dan motivasi dalam penyusunan skripsi ini.
4. Ibu Reisa Permatasari, S.T., M.Kom selaku dosen pembimbing pertama yang telah memberikan arahan, bimbingan, dan motivasi dalam penyusunan skripsi ini.
5. Orang tua dan keluarga besar, yang selalu memberikan doa, dukungan, semangat, serta kasih sayang yang tiada henti.
6. Dosen penguji seminar proposal dan seminar hasil, yang telah memberikan masukan, kritik, dan saran yang sangat berharga dalam meningkatkan kualitas penelitian ini.
7. Seluruh dosen Program Studi Sistem Informasi yang telah memberikan ilmu, pengalaman, dan wawasan yang sangat bermanfaat selama masa perkuliahan.

8. Teman-teman saya, Fia, Vita, Nisa, Sesiil, Ita, Kalel Fuzz, dan Bisquad yang selalu memberikan semangat dan dukungan dalam menyelesaikan skripsi ini.
9. Seluruh pihak yang turut memberikan dukungan dan bantuan, baik secara langsung maupun tidak langsung, yang tidak dapat saya sebutkan satu per satu dalam penyusunan skripsi ini.

Penulis menyadari bahwa penelitian ini, mengenai implementasi ensemble learning untuk analisis sentimen pada aplikasi Reku, memiliki ruang untuk pengembangan lebih lanjut. Oleh karena itu, saran dan kritik konstruktif sangat diharapkan demi penyempurnaan di masa mendatang. Harapan dari penelitian ini adalah dapat memberikan kontribusi signifikan dalam peningkatan kualitas analisis sentimen pada platform finansial seperti Reku, serta menjadi referensi berharga bagi penelitian selanjutnya di bidang klasifikasi teks dan ensemble learning.

Akhir kata, penulis berharap skripsi ini dapat memberikan wawasan yang bermanfaat bagi pembaca serta mendukung kemajuan ilmu pengetahuan dan teknologi di masa depan.

Surabaya, 17 Juni 2025

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN	iii
LEMBAR PERSETUJUAN	v
SURAT PERNYATAAN BEBAS PLAGIASI	vii
ABSTRAK	viii
ABSTRACT	ix
KATA PENGANTAR.....	x
DAFTAR ISI.....	xii
DAFTAR GAMBAR	xv
DAFTAR TABEL	xvii
DAFTAR LAMPIRAN	xviii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitian.....	6
1.5 Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA	9
2.1 Dasar Teori	9
2.1.1 Analisis Sentimen	9
2.1.2 Aplikasi Reku.....	10
2.1.3 Algoritma Deep Learning	10
2.1.4 <i>Ensemble learning</i>	14
2.1.5 Python	16
2.1.6 Scraping.....	17
2.1.7 Text Pre-Processing.....	18
2.1.8 Lexicon.....	18
2.1.9 Synthetic Minority Over Sampling Technique (SMOTE)	19
2.1.10 BERT Embedding	19
2.1.10 Word2Vec	20
2.1.12 <i>Confusion Matrix</i>	21
2.2 Penelitian Terdahulu.....	22
BAB III DESAIN DAN IMPLEMENTASI SISTEM	29

3.1	Studi Literatur	30
3.2	Analisis Kebutuhan	30
3.3	Pengumpulan Data	31
3.4	Data <i>Cleansing</i>	32
3.5	<i>Exploratory Data Analysis</i> (EDA)	32
3.6	Data <i>Preprocessing</i>	33
3.7	<i>Labeling</i> Data	36
3.8	<i>Word Embedding</i>	36
3.9	Data <i>Splitting</i> and <i>Handling Imbalance Data</i>	37
3.10	Klasifikasi Model	37
3.10.1	<i>Convolutional Neural Network</i> (CNN)	37
3.10.2	<i>Support Vector Machine</i> (SVM)	38
3.10.3	IndoBERT	38
3.10.4	<i>Ensemble Learning</i>	38
3.11	Model Evaluation	39
3.12	Pengembangan Sistem	39
3.13	Uji Validasi Sistem	40
BAB IV PENGUJIAN DAN ANALISA		41
4.1	Metode Pengujian	41
4.1.1	Pengumpulan Data	41
4.1.2	Data <i>Cleansing</i>	43
4.1.3	<i>Exploratory Data Analysis</i> (EDA)	45
4.1.4	Data <i>Preprocessing</i>	49
4.1.5	Labelisasi Data	53
4.1.6	Ekstraksi Fitur (Word2Vec)	55
4.1.7	Data <i>Splitting</i>	56
4.1.8	SMOTE	57
4.1.9	Klasifikasi Model	58
4.1.10	Model Evaluation	62
4.2	Analisis Skenario Pengujian	67
4.2.1	Pengaruh Arsitektur Model	67
4.2.2	Pengaruh Data dan Distribusi Kelas	67
4.2.3	Pengaruh Teknik <i>Ensemble Learning</i>	68
4.2.4	Kesimpulan Analisis Skenario Pengujian	69

4.3	Pengembangan Sistem	70
4.4	Uji Validasi Sistem.....	72
BAB V PENUTUP		75
5.1	Kesimpulan	75
5.2	Saran.....	76
DAFTAR PUSTAKA.....		77
LAMPIRAN.....		83

DAFTAR GAMBAR

Gambar 1. 1 Data Cryptocurrency di Indonesia.....	2
Gambar 1. 2 Perbandingan Traffic Share Reku dengan Aplikasi Lainnya	2
Gambar 2. 1 Aplikasi Reku	10
Gambar 2. 2 Perbandingan Performa <i>Machine Learning</i> dan <i>Deep Learning</i>	11
Gambar 2. 3 Plot SVM.....	13
Gambar 2. 4 Tahapan CNN pada klasifikasi retina.....	14
Gambar 2. 5 Soft Voting.....	16
Gambar 2. 6 Alur <i>Scraping</i>	17
Gambar 2. 7 Alur <i>Pre-Processing</i>	18
Gambar 2. 8 Cara Kerja Algoritma BERT	20
Gambar 3. 1 Tahapan Penelitian	29
Gambar 3. 2 Ulasan Pengguna	32
Gambar 3. 3 File CSV Hasil <i>Stemming</i>	35
Gambar 3. 4 <i>Flowchart Labeling Data</i>	36
Gambar 4. 1 Source code scraping data dari Google Play Store	42
Gambar 4. 2 Source code tanggal akhir scraping data Reku	42
Gambar 4. 3 <i>Source Code Cleansing Data</i>	44
Gambar 4. 4 Source Code Histogram	45
Gambar 4. 5 Visualisasi Histogram.....	46
Gambar 4. 6 <i>Source Code Word Cloud</i>	46
Gambar 4. 7 Visualisasi Word Cloud	47
Gambar 4. 8 Source Code <i>Latent Dirichlet Allocation</i>	47
Gambar 4. 9 Source Code Tokenisasi	49
Gambar 4. 10 <i>Source Code Stopword Removal</i>	51
Gambar 4. 11 <i>Source Code Stemming</i>	52
Gambar 4. 12 <i>Source Code Lexicon</i>	54
Gambar 4. 13 <i>Source Code Word2Vec</i>	55
Gambar 4. 14 <i>Source Code Data Splitting</i>	57
Gambar 4. 15 <i>Source Code SMOTE</i>	58
Gambar 4. 16 <i>Source Code Convolutional Neural Network</i>	59

Gambar 4. 17 <i>Source Code SVM Polinomial</i>	59
Gambar 4. 18 <i>Source Code IndoBERT</i>	60
Gambar 4. 19 <i>Source Code Soft Voting</i>	61
Gambar 4. 20 <i>Source Code Weighted Soft Voting</i>	62
Gambar 4. 21 <i>Confusion Matrix Model CNN</i>	63
Gambar 4. 22 <i>Metrik Evaluasi CNN</i>	63
Gambar 4. 23 <i>Confusion Matrix Model SVM</i>	64
Gambar 4. 24 <i>Metrik Evaluasi SVM</i>	64
Gambar 4. 25 <i>Confusion Matrix Model IndoBERT</i>	65
Gambar 4. 26 <i>Metrik Evaluasi IndoBERT</i>	65
Gambar 4. 27 <i>Confusion Matrix Model Soft Voting</i>	66
Gambar 4. 28 <i>Confusion Matrix Model Weighted Soft Voting</i>	66
Gambar 4. 29 <i>Classification Report Model Soft Voting</i>	68
Gambar 4. 30 <i>Classification Report Weighted Soft Voting</i>	68
Gambar 4. 31 <i>ROC AUC Model Weighted Soft Voting</i>	69
Gambar 4. 32 <i>Tampilan Fitur Upload File Excel yang Berisi Data Review</i>	71
Gambar 4. 33 <i>Tampilan Fitur Tabel Keseluruhan Data</i>	71
Gambar 4. 34 <i>Tampilan Fitur Tabel Hasil Analisis Sentimen</i>	72

DAFTAR TABEL

Tabel 2. 1 Komponen Confusion Matrix.....	21
Tabel 2. 2 Penelitian Terdahulu.....	22
Tabel 3. 1 Proses <i>Case Folding</i>	33
Tabel 3. 2 Hasil Alur Proses Tokenisasi.....	33
Tabel 3. 3 Hasil Alur Proses Stopword Removal.....	34
Tabel 3. 4 Hasil Alur Proses Stemming	35
Tabel 4. 1 <i>Sample Scraping Data</i>	43
Tabel 4. 2 <i>Sample Cleansing Data</i>	45
Tabel 4. 3 Hasil Tokenisasi Data.....	50
Tabel 4. 4 Hasil <i>Stopword Removal</i>	51
Tabel 4. 5 Hasil <i>Stemming</i>	53
Tabel 4. 6 Hasil Labelisasi	54
Tabel 4. 7 Hasil Klasifikasi Secara Keseluruhan	70
Tabel 4. 8 <i>Confusion Matrix</i> Data Validasi <i>Soft Voting</i>	72
Tabel 4. 9 Confusion Matrix Data Validasi Weighted Soft Voting.....	73
Tabel 4. 10 Perbandingan Sentimen Hasil Prediksi Sistem	74

DAFTAR LAMPIRAN

Lampiran 1 Kode Program Bagian Preprocessing	83
Lampiran 2 Kode Program Pengembangan Sistem Bagian 1	84
Lampiran 3 Kode Program Pengembangan Sistem Bagian 2	85
Lampiran 4 Kode Program Pengembangan Sistem Bagian 3	86
Lampiran 5 Kode Program Pengembangan Sistem Bagian 4	87
Lampiran 6 Kode Program Pengembangan Sistem Bagian 5	88