

**PERBANDINGAN DAN OPTIMASI ALGORITMA MACHINE
LEARNING MENGGUNAKAN BAYESIAN OPTIMIZATION
UNTUK DETEKSI ULASAN PALSU**

SKRIPSI

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



disusun oleh

NAJIB SAUQI RUBBAYANI

21.11.4406

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**PERBANDINGAN DAN OPTIMASI ALGORITMA MACHINE
LEARNING MENGGUNAKAN BAYESIAN OPTIMIZATION
UNTUK DETEKSI ULASAN PALSU**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh
NAJIB SAUQI RUBBAYANI
21.11.4406

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

HALAMAN PERSETUJUAN

SKRIPSI

**PERBANDINGAN DAN OPTIMASI ALGORITMA MACHINE
LEARNING MENGGUNAKAN BAYESIAN OPTIMIZATION UNTUK
DETEKSI ULASAN PALSU**

yang disusun dan diajukan oleh

Najib Sauqi Rubhayani

21.11.4406

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 18 Juni 2025

Dosen Pembimbing,



Windha Mega Pradnya Dhuwita, S.Kom., M.Kom.

NIK. 190302185

HALAMAN PENGESAHAN

SKRIPSI

**PERBANDINGAN DAN OPTIMASI ALGORITMA MACHINE
LEARNING MENGGUNAKAN BAYESIAN OPTIMIZATION UNTUK
DETEKSI ULASAN PALSU**

yang disusun dan diajukan oleh

Najib Sauqi Rubbayani

21.11.4406

Telah dipertahankan di depan Dewan Penguji
pada tanggal 18 Juni 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Theophilus Bayu Sasongko, S.Kom., M.Eng.
NIK. 190302375

Ike Verawati, S.Kom., M.Kom.
NIK. 190302237

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

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 18 Juni 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Najib Sauqi Rubbayani

NIM : 21.11.4406

Menyatakan bahwa Skripsi dengan judul berikut:

Perbandingan dan Optimasi Algoritma Machine Learning Menggunakan Bayesian Optimization untuk Deteksi Ulasan Palsu

Dosen Pembimbing : Windha Mega Pradnya Dhuhiita, S.Kom., 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, 18 Juni 2025

Yang Menyatakan,

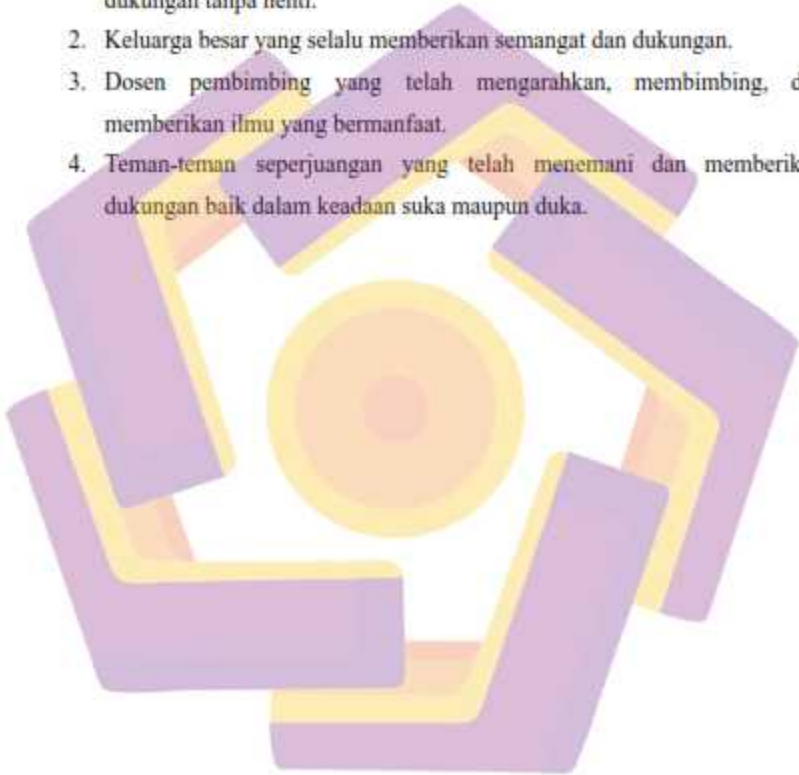


Najib Sauqi Rubbayani

HALAMAN PERSEMBAHAN

Dengan segala kerendahan hati dan rasa syukur kepada Tuhan Yang Maha Esa, skripsi ini saya persembahkan untuk:

1. Kedua orang tua tercinta, yang selalu memberikan cinta, doa, dan dukungan tanpa henti.
2. Keluarga besar yang selalu memberikan semangat dan dukungan.
3. Dosen pembimbing yang telah mengarahkan, membimbing, dan memberikan ilmu yang bermanfaat.
4. Teman-teman seperjuangan yang telah menemani dan memberikan dukungan baik dalam keadaan suka maupun duka.



KATA PENGANTAR

Puji dan syukur penulis panjatkan atas kehadiran Tuhan Yang Maha Esa karena atas rahmat dan karunianya, penulis dapat menyelesaikan skripsi yang berjudul “Perbandingan dan Optimasi Algoritma Machine Learning Menggunakan Bayesian Optimization untuk Deteksi Ulasan Palsu” dengan baik.

Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

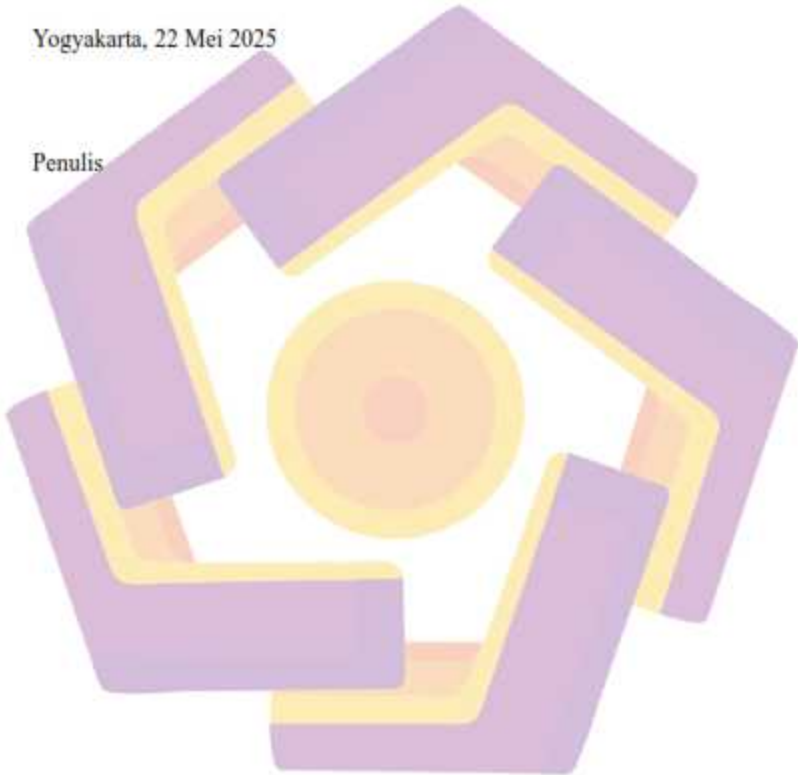
Pada proses penyusunan skripsi ini, penulis mendapatkan banyak bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas Amikom Yogyakarta.
2. Ibu Eli Pujastuti, M.Kom. selaku Ketua Program Studi Informatika Universitas Amikom Yogyakarta.
3. Ibu Windha Mega Pradnya Dhuhiha, S.Kom., M.Kom., selaku Dosen Pembimbing Skripsi yang telah memberikan arahan dan masukan yang sangat berarti kepada penulis dalam penyusunan skripsi ini.
4. Bapak Tristanto Ari Aji, M.Kom., selaku dosen wali yang senantiasa memberikan arahan dan dukungan selama masa studi di Universitas Amikom Yogyakarta.
5. Kedua orang tua dan keluarga tercinta yang selalu memberikan motivasi, doa, dan kasih sayangnya kepada penulis.
6. Rekan-rekan dan sahabat yang telah membagikan ilmu, pengalaman, serta inspirasi selama menjalani perjalanan akademik di Universitas Amikom Yogyakarta.

Penulis menyadari bahwa skripsi ini jauh dari kata sempurna. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi penyempurnaan di masa mendatang. Semoga skripsi ini dapat memberikan kontribusi positif bagi dunia akademik, khususnya di bidang klasifikasi ulasan palsu dan penerapan machine learning.

Yogyakarta, 22 Mei 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.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN.....	xii
DAFTAR LAMBANG DAN SINGKATAN.....	xiii
DAFTAR ISTILAH.....	xiv
INTISARI.....	xvii
ABSTRACT.....	xviii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Studi Literatur.....	5
2.1 Dasar Teori.....	21
2.2.1 Ulasan Palsu.....	21
2.2.2 Machine Learning.....	21
2.2.3 Klasifikasi.....	22
2.2.4 Data Preprocessing.....	22
2.2.5 Term Frequency-Inverse Document Frequency (TF-IDF).....	22
2.2.6 N-Gram.....	23
2.2.7 Random Forest.....	24
2.2.8 Support Vector Machine (SVM).....	24
2.2.9 Naive Bayes.....	26
2.2.10 XGBoost.....	26
2.2.11 Bayesian Optimization (BO).....	27

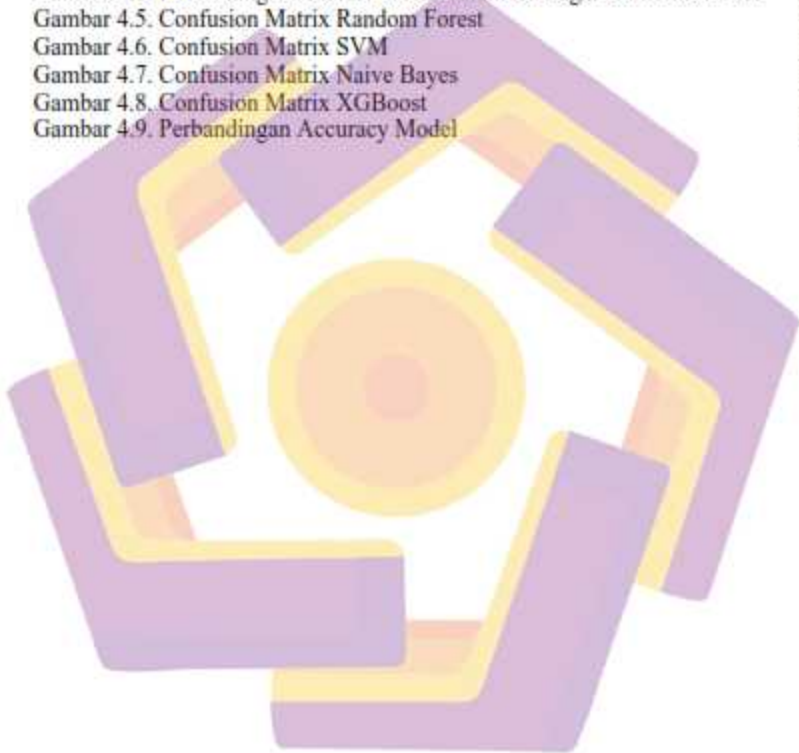
2.2.12 Confusion Matrix.....	29
2.2.13 Waktu Inferensi.....	30
BAB III METODE PENELITIAN.....	32
3.1 Objek Penelitian.....	32
3.2 Alur Penelitian.....	33
3.2.1 Memuat Dataset.....	34
3.2.2 Data Preprocessing.....	34
3.2.3 Data Splitting.....	34
3.2.4 TF-IDF.....	35
3.2.5 Training Tanpa Hyperparameter Tuning.....	35
3.2.6 Training Dengan Hyperparameter Tuning BO.....	36
3.2.7 Evaluasi.....	38
3.3 Alat dan Bahan.....	38
3.3.1 Data Penelitian.....	38
3.3.2 Alat/Instrumen.....	39
BAB IV HASIL DAN PEMBAHASAN.....	41
4.1 Dataset.....	41
4.2 Data Preprocessing.....	42
4.2.1 Cleaning.....	42
4.2.2 One-hot Encoding.....	43
4.2.3 Label Encoding.....	45
4.3 Data Splitting.....	45
4.4 TF-IDF.....	46
4.5 Training Tanpa Hyperparameter Tuning.....	47
4.6 Training Dengan Hyperparameter Tuning BO.....	49
4.7 Evaluasi.....	51
4.8 Perbandingan Dengan Penelitian Sebelumnya.....	56
BAB V PENUTUP.....	59
5.1 Kesimpulan.....	59
5.2 Saran.....	59
REFERENSI.....	60
LAMPIRAN.....	63

DAFTAR TABEL

Tabel 2.1. Keaslian Penelitian	9
Tabel 2.2. Confusion Matrix	29
Tabel 3.1. Hyperparameter Algoritma Random Forest dan Nilainya	36
Tabel 3.2. Hyperparameter Algoritma SVM dan Nilainya	37
Tabel 3.3. Hyperparameter Algoritma Naive Bayes dan Nilainya	37
Tabel 3.4. Hyperparameter Algoritma XGBoost dan Nilainya	37
Tabel 3.5. Deskripsi Dataset	39
Tabel 4.1. Cleaning	42
Tabel 4.2. Sebelum One-hot Encoding	43
Tabel 4.3. Sesudah One-hot Encoding	43
Tabel 4.4. Label Encoding	45
Tabel 4.5. Jumlah Fitur TF-IDF	46
Tabel 4.6. Data TF-IDF Pelatihan	46
Tabel 4.7. Data TF-IDF Uji	47
Tabel 4.8. Perbandingan Kinerja Model dalam Berbagai Kombinasi Fitur	48
Tabel 4.9. Performa Model Random Forest Dengan dan Tanpa max_depth	49
Tabel 4.10. Kombinasi Nilai Hyperparameter Terbaik	50
Tabel 4.11. Performa Model Setelah Hyperparameter Tuning	51
Tabel 4.12. Perbandingan Hasil Evaluasi Model	54
Tabel 4.13. Perbandingan Penelitian	57

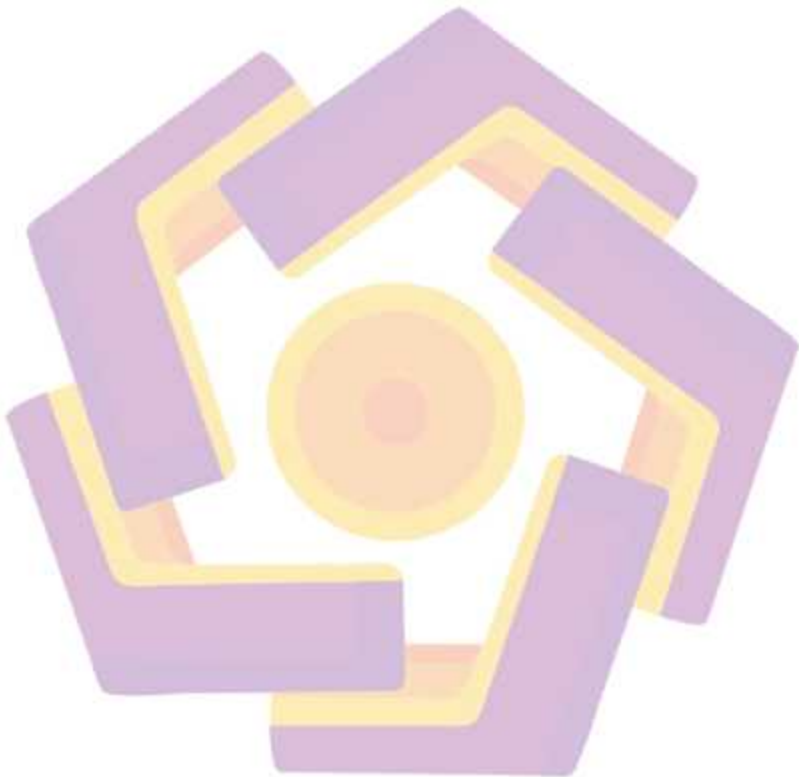
DAFTAR GAMBAR

Gambar 2.1. Alur Kerja Random Forest	24
Gambar 3.1. Alur Penelitian	33
Gambar 4.1. Dataset	41
Gambar 4.2. Distribusi Dataset Ulasan Palsu	42
Gambar 4.3. Distribusi Data Train dan Test	46
Gambar 4.4. Perbandingan Akurasi Model dalam Berbagai Kombinasi Fitur	49
Gambar 4.5. Confusion Matrix Random Forest	52
Gambar 4.6. Confusion Matrix SVM	52
Gambar 4.7. Confusion Matrix Naive Bayes	53
Gambar 4.8. Confusion Matrix XGBoost	53
Gambar 4.9. Perbandingan Accuracy Model	54

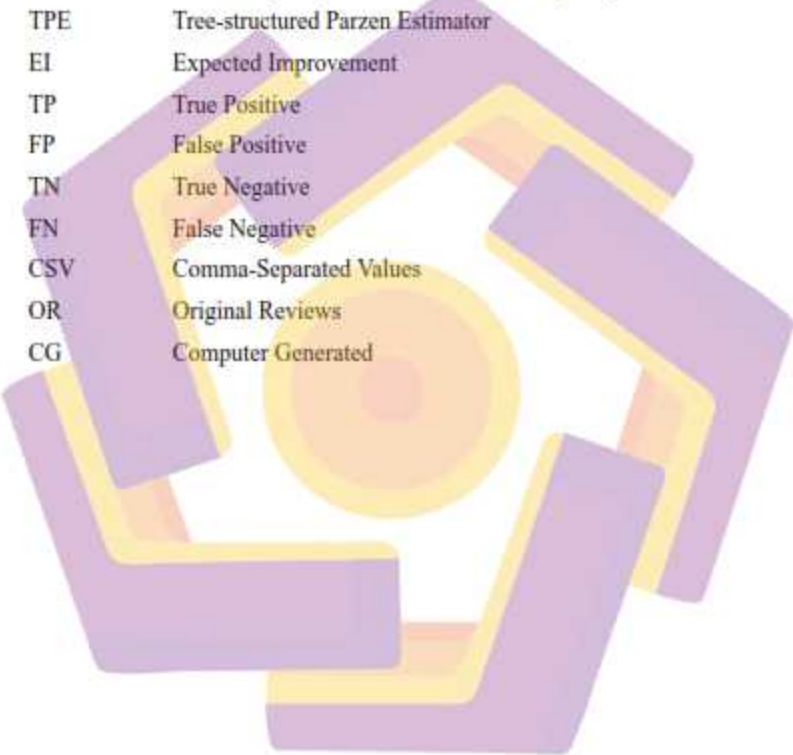


DAFTAR LAMPIRAN

Lampiran 1. Source Code	63
Lampiran 2. Exported Model	63
Lampiran 3. Web App Sederhana Hasil Deployment Model Terbaik (SVM)	64

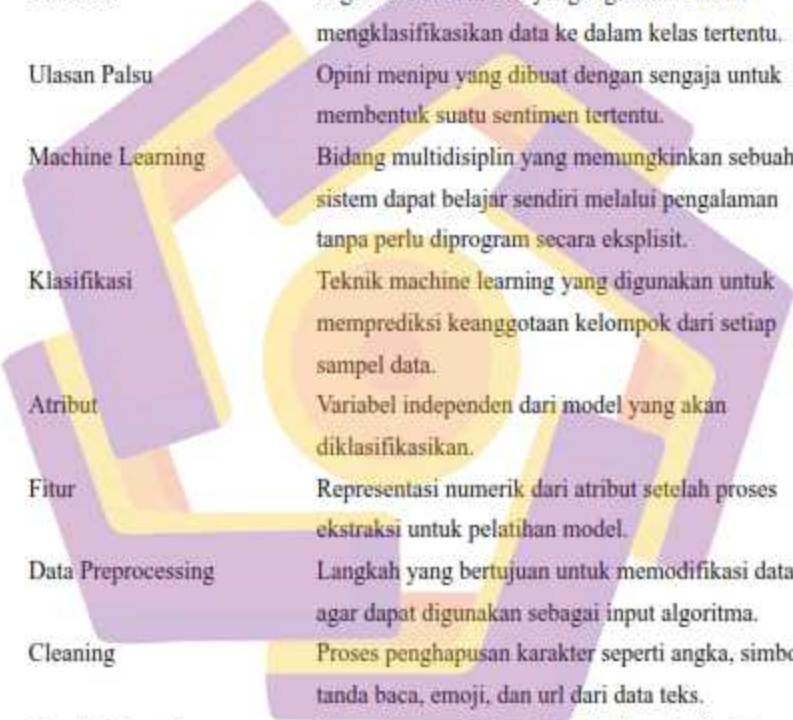


DAFTAR LAMBANG DAN SINGKATAN



BO	Bayesian Optimization
SVM	Support Vector Machines
XGBoost	eXtreme Gradient Boosting
TF-IDF	Term Frequency-Inverse Document Frequency
TPE	Tree-structured Parzen Estimator
EI	Expected Improvement
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative
CSV	Comma-Separated Values
OR	Original Reviews
CG	Computer Generated

DAFTAR ISTILAH



E-Commerce	Platform perdagangan online.
Hyperparameter	Parameter yang diatur sebelum pelatihan model.
Hyperparameter Tuning	Proses mencari kombinasi hyperparameter yang memberikan kinerja terbaik pada model.
Classifier	Algoritma atau model yang digunakan untuk mengklasifikasikan data ke dalam kelas tertentu.
Ulasan Palsu	Opini menipu yang dibuat dengan sengaja untuk membentuk suatu sentimen tertentu.
Machine Learning	Bidang multidisiplin yang memungkinkan sebuah sistem dapat belajar sendiri melalui pengalaman tanpa perlu diprogram secara eksplisit.
Klasifikasi	Teknik machine learning yang digunakan untuk memprediksi keanggotaan kelompok dari setiap sampel data.
Atribut	Variabel independen dari model yang akan diklasifikasikan.
Fitur	Representasi numerik dari atribut setelah proses ekstraksi untuk pelatihan model.
Data Preprocessing	Langkah yang bertujuan untuk memodifikasi data agar dapat digunakan sebagai input algoritma.
Cleaning	Proses penghapusan karakter seperti angka, simbol, tanda baca, emoji, dan url dari data teks.
One-hot Encoding	Proses konversi data kategorikal menjadi vektor biner dengan panjang sesuai dengan jumlah kategori.
Label Encoding	Proses konversi data kategorikal ke dalam format numerik.
Data Splitting	Pemisahan data menjadi subset pelatihan dan pengujian.

TF-IDF	Teknik pemberian bobot fitur teks berdasarkan frekuensi kata dan tingkat kepentingannya.
Unigram	Fitur yang merepresentasikan satu kata tunggal.
Bigram	Fitur yang merepresentasikan dua kata yang berdekatan.
Random Forest	Algoritma yang bekerja dengan cara membangun kumpulan pohon yang berasal dari berbagai sampel dalam dataset.
SVM	Algoritma yang bekerja dengan mencari hyperplane optimal untuk memisahkan kelas.
Naive Bayes	Algoritma yang didasarkan pada teorema Bayes yang mengasumsikan bahwa semua atribut bersifat independen.
XGBoost	Algoritma boosting yang merupakan penyempurnaan dari algoritma Gradient Boosting.
Bayesian Optimization	Metode hyperparameter tuning berbasis model probabilistik yang membentuk model pengganti untuk memperkirakan fungsi objektif.
TPE	Salah satu model pengganti (surrogate model) dalam Bayesian Optimization.
Expected Improvement	Fungsi akuisisi yang digunakan untuk memilih titik berikutnya yang selain mengeksplorasi wilayah nilai optimal saat ini, juga menghitung peningkatan yang diharapkan dari titik baru.
Evaluasi	Proses yang dilakukan untuk memastikan bahwa model memiliki kinerja yang optimal.
Confusion Matrix	Metode yang digunakan untuk mengevaluasi kinerja model klasifikasi dengan cara membandingkan hasil prediksi model dengan nilai aktual.
Label	Kategori atau output yang ingin diprediksi oleh model.

Kelas	Merujuk pada data kategoris yang merepresentasikan label yang ada pada objek setelah proses klasifikasi.
Accuracy	Perbandingan antara jumlah prediksi yang benar dengan total jumlah data.
Precision	Proporsi instance positif yang berhasil diprediksi dengan benar dibandingkan dengan total instance yang diprediksi sebagai positif.
Recall	Proporsi kasus positif yang berhasil diprediksi dengan benar dibandingkan dengan total jumlah kasus positif.
F1-Score	Rata-rata harmonik dari precision dan recall.
Corpus	Kumpulan teks yang digunakan sebagai sumber data.
Waktu Inferensi	Lama waktu yang dibutuhkan model untuk membuat prediksi terhadap data baru.

INTISARI

Peningkatan penggunaan *e-commerce* di Indonesia memunculkan tantangan baru, salah satunya yaitu penyebaran ulasan palsu yang dapat mempengaruhi persepsi konsumen serta reputasi dan profitabilitas penjual. Ulasan palsu dapat meningkatkan rating produk secara tidak wajar dan menyesatkan pembeli. Oleh karena itu, diperlukan solusi berbasis teknologi untuk mendeteksinya. Salah satu metode yang dapat digunakan adalah *machine learning*. Meskipun begitu, kinerja model *machine learning* sangat dipengaruhi oleh penyetelan *hyperparameter* (*hyperparameter tuning*). Untuk menghindari penyetelan *hyperparameter* secara manual yang memakan waktu dan rawan kesalahan, ada berbagai metode yang dapat digunakan, salah satunya *Bayesian Optimization* (BO). Penelitian ini mengusulkan penggunaan algoritma *machine learning* seperti *Random Forest*, *SVM*, *Naive Bayes*, dan *XGBoost* untuk klasifikasi ulasan palsu, serta *hyperparameter tuning* BO untuk meningkatkan performa model. Dataset yang digunakan terdiri dari 40432 data berlabel, dan proses penelitian meliputi *preprocessing* (*cleaning*, *one-hot encoding*, *label encoding*), ekstraksi fitur menggunakan TF-IDF dengan *unigram* dan *bigram*, *data splitting* (80:20), pelatihan model tanpa dan dengan *hyperparameter tuning*, dan evaluasi. Hasil dari penelitian menunjukkan bahwa BO mampu meningkatkan kinerja model *machine learning* yang diusulkan, dengan peningkatan *accuracy* sebesar 2,1% pada *Random Forest*, 0,27% pada *SVM*, 0,23% pada *Naive Bayes*, dan 3,15% pada *XGBoost*. Dibandingkan dengan model *fakeRoBERTa* dari penelitian sebelumnya yang memiliki *accuracy* 96,64%, model terbaik dari penelitian ini, yaitu *SVM* dengan BO (95,31%), memiliki waktu inferensi rata-rata 10,12 detik lebih cepat. Keunggulan ini penting dalam implementasi nyata yang memerlukan respons cepat. Hasil penelitian ini diharapkan dapat memberikan solusi yang efektif dalam mendeteksi ulasan palsu secara otomatis khususnya pada platform *e-commerce*, serta dapat dijadikan sebagai referensi oleh para peneliti di bidang kecerdasan buatan pada penelitian mendatang.

Kata kunci: ulasan palsu, *machine learning*, klasifikasi, *hyperparameter tuning*, *Bayesian optimization*.

ABSTRACT

The increasing use of e-commerce in Indonesia has led to new challenges, one of which is the spread of fake reviews that can influence consumer perceptions as well as the reputation and profitability of sellers. Fake reviews can artificially inflate product ratings and mislead buyers. Therefore, a technology-based solution is needed to detect them. One method that can be used is machine learning. However, the performance of machine learning models is highly influenced by hyperparameter tuning. To avoid manual hyperparameter tuning, which is time-consuming and prone to error, various methods can be applied, one of which is Bayesian Optimization (BO). This study proposes the use of machine learning algorithms such as Random Forest, SVM, Naive Bayes, and XGBoost for fake review classification, along with BO-based hyperparameter tuning to improve model performance. The dataset used consists of 40432 labeled data points, and the research process includes preprocessing (cleaning, one-hot encoding, label encoding), feature extraction using TF-IDF with unigram and bigram, data splitting (80:20), model training with and without hyperparameter tuning, and evaluation. The results of the study show that BO can enhance the performance of the proposed machine learning models, with accuracy improvements of 2.1% for Random Forest, 0.27% for SVM, 0.23% for Naive Bayes, and 3.15% for XGBoost. Compared to the fakeRoBERTa model from previous research, which achieved 96.64% accuracy, the best-performing model in this study, SVM with BO (95.31%), has an average inference time that is 10.12 seconds faster. This advantage is crucial for real-world implementations requiring quick responses. The results of this research are expected to provide an effective automatic solution for detecting fake reviews on e-commerce platforms and serve as a reference for future studies in the field of artificial intelligence.

Keyword: fake reviews, machine learning, classification, hyperparameter tuning, Bayesian optimization.