PENGEMBANGAN APLIKASI WEB UNTUK DETEKSI PLAGIARISME KODE SUMBER JAVA BERBASIS *MACHINE LEARNING*

Tugas Akhir

Diajukan untuk Memenuhi Salah Satu Persyaratan Mencapai Derajat Sarjana Komputer



Dibuat Oleh:

STEPANUS PETRA PAMBUDI

210711441

PROGRAM STUDI INFORMATIKA

FAKULTAS TEKNOLOGI INDUSTRI

UNIVERSITAS ATMA JAYA YOGYAKARTA

2026

HALAMAN PENGESAHAN

Tugas Akhir Berjudul

PENGEMBANGAN APLIKASI WEB DETEKSI PLAGIARISME KODE SUMBER JAVA BERBASIS MACHINE LEARNING

yang disusun oleh

STEPANUS PETRA PAMBUDI

210711441

dinyatakan telah memenuhi syarat pada tanggal 28 Januari 2026

		Keterangan
Dosen Pembimbing 1	: Wilfridus Bambang Triadi H, ST., M.Cs	Telah Menyetujui
Dosen Pembimbing 2	: Vinindita Citrayasa, S.Pd., M.Hum	Telah Menyetujui
Tim Penguji		
Penguji 1	: Wilfridus Bambang Triadi H, ST., M.Cs	Telah Menyetujui
Penguji 2	: Dra. Ernawati, M.T., Ph.D.	Telah Menyetujui
Penguji 3	: Martinus Maslim, S.T., M.T., Ph.D.	Telah Menyetujui

Yogyakarta, 28 Januari 2026

Universitas Atma Jaya Yogyakarta

Teknologi Industri

Dekan

ttd.

Prof. Dr. Ir. Parama Kartika Dewa SP., S.T., M.T., IPU, ASEAN Eng.

Dokumen ini merupakan dokumen resmi UAJY yang tidak memerlukan tanda tangan karena dihasilkan secara elektronik oleh Sistem Bimbingan UAJY. UAJY bertanggung jawab penuh atas informasi yang tertera di dalam dokumen ini

PERNYATAAN ORISINALITAS & PUBLIKASI ILMIAH

Saya yang bertanda tangan di bawah ini:

Nama Lengkap : Stepanus Petra Pambudi

NPM : 210711441

Program Studi : Informatika

Fakultas : Teknologi Industri

Judul Penelitian : Pengembangan Aplikasi Web untuk Deteksi Plagiarisme Kode Sumber Java Berbasis *Machine Learning*

Menyatakan dengan ini:

1. Tugas Akhir ini adalah benar tidak merupakan salinan sebagian atau keseluruhan dari karya penelitian lain.
2. Memberikan kepada Universitas Atma Jaya Yogyakarta atas penelitian ini, berupa Hak untuk menyimpan, mengelola, mendistribusikan, dan menampilkan hasil penelitian selama tetap mencantumkan nama penulis.
3. Bersedia menanggung secara pribadi segala bentuk tuntutan hukum atas pelanggaran Hak Cipta dalam pembuatan Tugas Akhir ini.

Demikianlah pernyataan ini dibuat dan dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 11 Januari 2026

Yang menyatakan,



Stepanus Petra Pambudi

210711441

HALAMAN PERSEMBAHAN

Semua akan indah pada waktu-Nya

KATA PENGANTAR

Puji dan syukur penulis haturkan kepada Tuhan Yang Maha Esa karena berkat rahmat dan karunia-Nya penulis dapat menyelesaikan pembuatan tugas akhir "Pengembangan Aplikasi Web untuk Deteksi Plagiarisme Kode Sumber Java Berbasis *Machine Learning*" ini dengan baik.

Penulisan tugas akhir ini bertujuan untuk memenuhi salah satu syarat untuk mencapai derajat sarjana komputer dari Program Studi Informatika, Fakultas Teknologi Industri di Universitas Atma Jaya Yogyakarta.

Penulis menyadari bahwa dalam pembuatan tugas akhir ini penulis telah mendapatkan bantuan, bimbingan, dan dorongan dari banyak pihak. Untuk itu, pada kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Tuhan Yesus Kristus yang selalu membimbing dalam iman-Nya, memberikan berkat-Nya, dan menyertai penulis selalu.
2. Bapak Dr. Ir. Parama Kartika Dewa SP., S.T., M.T. IPU, selaku Dekan Fakultas Teknologi Industri, Universitas Atma Jaya Yogyakarta.
3. Bapak Paulus Mudjihartono, S.T., M.T., Ph.D, selaku Ketua Departmen Program Studi Informatika, Universitas Atma Jaya Yogyakarta.
4. Bapak Thomas Adi Purnomo Sidhi, S.T., M.T., selaku Ketua Program Studi Informatika, Universitas Atma Jaya Yogyakarta.
5. Bapak W. Bambang Triadi H. ST., M. Cs., selaku dosen pembimbing I yang telah membimbing dan memberikan masukan serta motivasi kepada penulis untuk menyelesaikan tugas akhir ini.
6. Ibu Vinindita Citrayasa, S.Pd., M.Hum., selaku dosen pembimbing II yang telah membimbing dan memberikan masukan serta motivasi kepada penulis untuk menyelesaikan tugas akhir ini.
7. Kepada kedua orang tua tercinta, yang selalu memberikan kasih sayang, dukungan, serta doa yang tiada henti. Terima kasih atas segala pengorbanan,

kesabaran, dan dorongan moral maupun material yang telah diberikan selama ini. Semangat dan ketulusan dari Ayah dan Ibu menjadi sumber motivasi utama bagi penulis dalam menyelesaikan tugas akhir ini.

Demikian laporan tugas akhir ini penulis susun dengan segala upaya dan kemampuan yang dimiliki. Penulis menyadari bahwa laporan ini masih jauh dari sempurna, sehingga kritik dan saran yang membangun sangat diharapkan demi perbaikan di masa mendatang. Semoga hasil dari tugas akhir ini dapat memberikan manfaat bagi pembaca, menjadi tambahan wawasan bagi pengembangan ilmu di bidang informatika, serta menjadi langkah awal bagi penulis untuk terus belajar dan berkarya di masa depan.

Yogyakarta, 11 Januari 2026



Stepanus Petra Pambudi

210711441

DAFTAR ISI

PENGEMBANGAN APLIKASI WEB UNTUK DETEKSI PLAGIARISME KODE SUMBER JAVA BERBASIS *MACHINE LEARNING* i
HALAMAN PENGESAHAN ii
PERNYATAAN ORISINALITAS & PUBLIKASI ILMIAH iii
HALAMAN PERSEMBAHAN iv
KATA PENGANTAR v
DAFTAR ISI vii
DAFTAR GAMBAR ix
DAFTAR TABEL x
INTISARI xi
BAB I PENDAHULUAN 12
A. Latar Belakang 12
B. Rumusan Masalah 13
C. Batasan Masalah 13
D. Tujuan Penelitian 14
E. Metode Penelitian 14
F. Sistematika Penulisan 3
BAB II TINJAUAN PUSTAKA 5
BAB III LANDASAN TEORI 12
A. Plagiarisme *Source Code* 12
B. Normalisasi dan *Preprocessing Source Code* 12
C. Representasi *Source Code* 13
D. Metode Pengukuran Kemiripan *Source Code* 14
E. Klasifikasi Biner 15
F. Evaluasi Model 15
G. *Website* 15

H. Framework Streamlit 16
I. Waterfall 16
BAB IV *DATASET* DAN PENGEMBANGAN MODEL 17
A. Deskripsi Masalah 17
B. *Dataset* 19
C. Pengembangan Model 36
BAB V IMPLEMENTASI MODEL DAN PENGUJIAN SISTEM 45
A. Implementasi Model 45
B. Implementasi Sistem 53
C. Pengujian Sistem 56
BAB VI. PENUTUP 62
A. Kesimpulan 62
B. Saran 62
DAFTAR PUSTAKA 64

DAFTAR GAMBAR

Gambar 1. 1 Model Waterfall dalam Pengembangan Sistem Deteksi Plagiarisme 2
Gambar 4. 1 Ilustrasi jenis modifikasi kode mahasiswa 17
Gambar 4. 2 Ilustrasi boilerplate vs logika inti 18
Gambar 4. 3 Diagram alur masalah deteksi plagiarism 19
Gambar 4. 4 Alur pembangkitan dataset sintetis 31
Gambar 4. 5 Distribusi label positif dan negatif 33
Gambar 4. 6 Jumlah pasangan per problem 33
Gambar 4. 7 Diagram alur preprocessing dataset 35
Gambar 4. 8 Cuplikan berkas JSONL hasil preprocessing 36
Gambar 4. 9 Diagram alur pengembangan model 38
Gambar 4. 10 Confusion matrix kandidat model pada data uji.....40
Gambar 4. 11 Kurva ROC dan nilai AUC kandidat model 41
Gambar 4. 12 Perbandingan performa dan efisiensi kandidat model 42
Gambar 4. 13 Analisis learning curve model pada berbagai ukuran dataset 43
Gambar 5. 1 function load model.....46
Gambar 5. 2 function load academic detector.....46
Gambar 5. 3 list fitur kemiripan.....47
Gambar 5. 4 function extract academic features.....49
Gambar 5. 5 function predict 51
Gambar 5. 6 Diagram Alur Implementasi Model pada Aplikasi Web.....52
Gambar 5. 7 Diagram Struktur Sistem dan Alur Data 53
Gambar 5. 8 halaman input utama dan sidebar.....54
Gambar 5. 9 Tangkapan layar tampilan hasil prediksi.....55
Gambar 5. 10 Tangkapan layar highlight baris mirip 56

DAFTAR TABEL

Tabel 2. 1 Perbandingan Penelitian.....8
Tabel 4. 1 Contoh Sample Dataset ConPlag 21
Tabel 4. 2 Perbandingan Performa Kandidat Model.....41
Tabel 5. 1 Skenario pengujian fungsional sistem 58

INTISARI

PENGEMBANGAN APLIKASI WEB UNTUK DETEKSI PLAGIARISME KODE SUMBER JAVA BERBASIS *MACHINE LEARNING*

Intisari

Stepanus Petra Pambudi
210711441

Plagiarisme pada tugas pemrograman Java di lingkungan perkuliahan tidak selalu muncul dalam bentuk penyalinan identik, melainkan sering dilakukan melalui berbagai modifikasi permukaan seperti penggantian nama variabel, perubahan komentar, pengaturan ulang struktur kode, atau *refactoring* ringan tanpa mengubah logika inti program. Kondisi ini menyebabkan metode pemeriksaan berbasis pencocokan teks sederhana menjadi kurang efektif dan berpotensi menghasilkan kesalahan deteksi. Oleh karena itu, penelitian ini dirumuskan untuk mengembangkan sebuah sistem deteksi plagiarisme *source code* Java yang mampu menangkap kesamaan logika program dan dapat digunakan sebagai alat bantu dosen ataupun asisten dosen.

Penelitian ini mengembangkan aplikasi web deteksi plagiarisme *source code* berbasis *Streamlit* yang mengintegrasikan model *machine learning* untuk klasifikasi pasangan *source code*. *Dataset* yang digunakan merupakan gabungan *dataset* ConPlag dan *dataset* sintetis yang dibangkitkan secara terkontrol. Setiap pasangan direpresentasikan menggunakan enam fitur, yaitu Kemiripan *token*, Pola penamaan variabel, kemiripan komentar, kemiripan urutan logika, kemiripan teks, dan kemiripan baris logika. Kemudian model dievaluasi menggunakan skema *group-aware split* serta *cross-validation*.

Hasil penelitian menunjukkan bahwa model SVM dengan enam fitur tersebut memberikan performa yang stabil dan efisien. Sistem yang dikembangkan mampu menghasilkan skor probabilitas, label plagiarisme, tingkat risiko, serta ringkasan fitur pendukung secara konsisten. Pengujian sistem menunjukkan bahwa aplikasi berjalan sesuai rancangan dan siap digunakan sebagai alat bantu deteksi plagiarisme *source code* Java.

Kata Kunci: deteksi plagiarisme, *source code* Java, *machine learning*, aplikasi web

Dosen Pembimbing I : W. Bambang Triadi H. ST., M. Cs.

Dosen Pembimbing II : Vinindita Citrayasa, S.Pd., M.Hum.

Jadwal Sidang Tugas Akhir : 21 Januari 2026