

**EKSPERIMEN ALGORITMA RANDOM FOREST UNTUK PREDIKSI
BUG JUST-IN-TIME PADA PERANGKAT LUNAK OPEN-SOURCE
MENGUNAKAN DATASET APACHEJIT**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi *Teknik Komputer*



disusun oleh

MUHAMMAD ROFIQ FAISAL

22.83.0769

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**EKSPERIMEN ALGORITMA RANDOM FOREST UNTUK PREDIKSI
BUG JUST-IN-TIME PADA PERANGKAT LUNAK OPEN-SOURCE
MENGUNAKAN DATASET APACHEJIT**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Teknik Komputer



disusun oleh

MUHAMMAD ROFIQ FAISAL

22.83.0769

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**EKSPERIMEN ALGORITMA RANDOM FOREST UNTUK PREDIKSI
BUG JUST-IN-TIME PADA PERANGKAT LUNAK OPEN-SOURCE
MENGUNAKAN DATASET APACHEJIT**

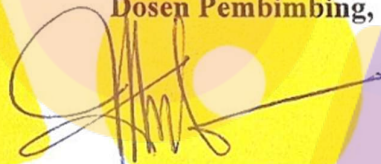
yang disusun dan diajukan oleh

MUHAMMAD ROFIQ FAISAL

22.83.0769

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 21 Januari 2026

Dosen Pembimbing,



Melwin Syafrizal, S.Kom., M.Eng., Ph.D.
NIK. 190302105

HALAMAN PENGESAHAN
SKRIPSI
EKSPERIMEN ALGORITMA RANDOM FOREST UNTUK PREDIKSI
BUG JUST-IN-TIME PADA PERANGKAT LUNAK OPEN-SOURCE
MENGGUNAKAN DATASET APACHEJIT

yang disusun dan diajukan oleh

Muhammad Rofiq Faisal

22.83.0769

Telah dipertahankan di depan Dewan Penguji
pada tanggal 21 Januari 2026
Susunan Dewan Penguji

Nama Penguji

Dony Ariyus, S.S., M.Kom.
NIK. 1903022128

Wahid Miftahul Ashari, S.Kom., M.T.
NIK. 190302452

Melwin Syafrizal, S.Kom., M.Eng., Ph.D.
NIK. 190302105

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 21 Januari 2026

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : MUHAMMAD ROFIQ FAISAL
NIM : 22.83.0769

Menyatakan bahwa Skripsi dengan judul berikut:

EKSPERIMEN ALGORITMA RANDOM FOREST UNTUK PREDIKSI BUG
JUST-IN-TIME PADA PERANGKAT LUNAK OPEN-SOURCE
MENGUNAKAN DATASET APACHEJIT

Dosen Pembimbing: Melwin Syafrizal S.Kom., M.Eng., Ph.D.

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, 21 Januari 2026

Yang Menyatakan,



Muhammad Rofiq Faisal.

KATA PENGANTAR

Segala puji dan syukur penulis ucapkan kehadirat Allah SWT, yang telah melimpahkan rahmat dan karunia-Nya, serta memberikan kekuatan kepada penulis baik bersifat lahir maupun batin. Shalawat serta salam tak lupa penulis ucapkan kepada junjungan yakni Nabi Muhammad SAW, karena berkat beliau kita dapat berkembang dari zaman jahiliyah hingga pada zaman yang penuh dengan ilmu pengetahuan dan kemajuan ini.

Dalam proses penyusunan skripsi ini, penulis menyadari bahwa penyelesaian skripsi tidak terlepas dari bantuan, bimbingan, dan dukungan berbagai pihak. Oleh karena itu, pada kesempatan ini penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Melwin Syafrizal S.Kom., M.Eng., Ph.D. selaku Dosen Pembimbing, yang telah memberikan bimbingan, arahan, dan masukan selama proses penyusunan skripsi ini.
2. Bapak Dony Ariyus, S.S., M.Kom. dan Bapak Wahid Miftahul Ashari, S.Kom., M.T. selaku Dosen Penguji, atas saran, kritik, dan masukan yang diberikan pada saat ujian pendadaran sehingga skripsi ini dapat disempurnakan.
3. Orang tua dan keluarga penulis, yang telah memberikan doa, dukungan moral, dan material selama proses perkuliahan hingga penyusunan skripsi ini.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan. Oleh karena itu, penulis mengharapkan kritik dan saran yang bersifat membangun sebagai bahan perbaikan di masa yang akan datang. Semoga skripsi ini dapat memberikan manfaat bagi pembaca dan pihak-pihak yang berkepentingan.

Yogyakarta, 21 Januari 2026

Penulis

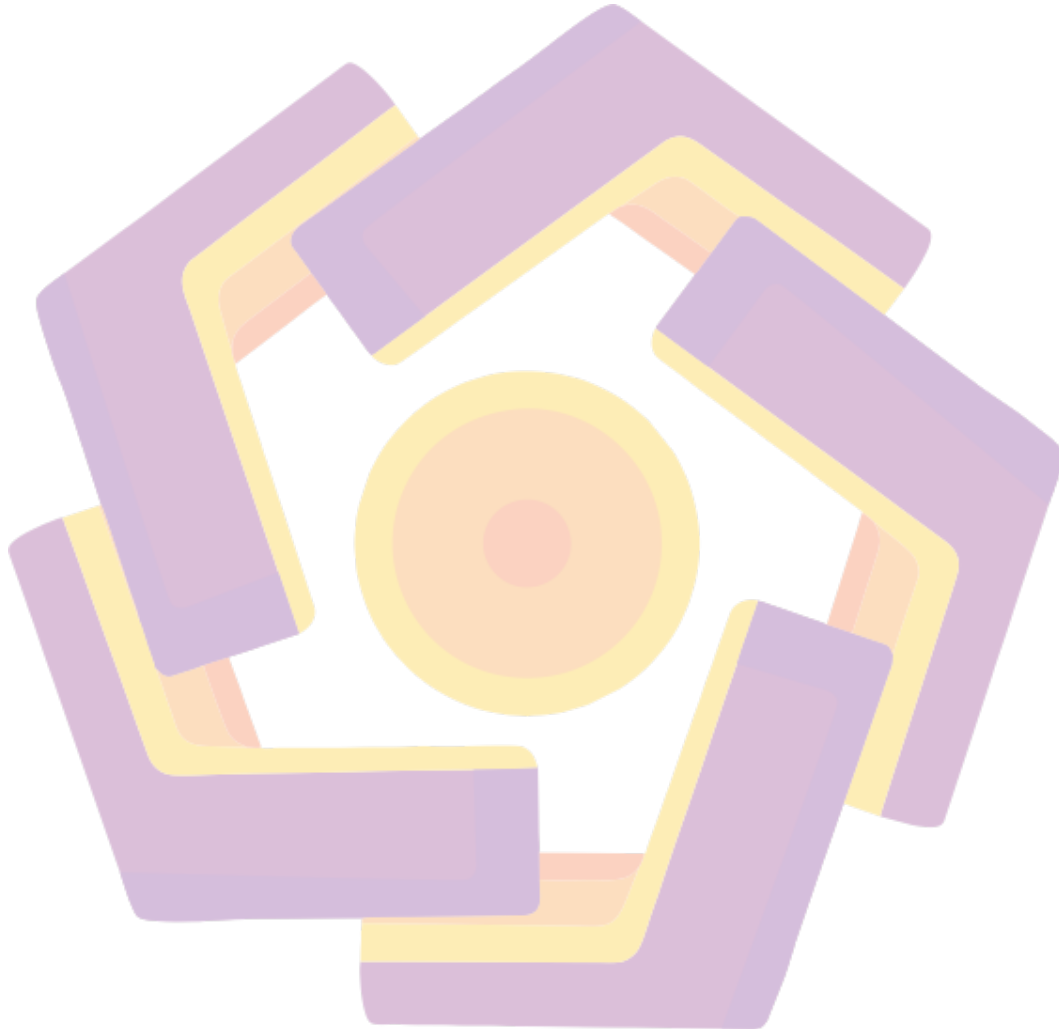
DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
KATA PENGANTAR	v
DAFTAR ISI.....	vi
DAFTAR TABEL.....	viii
DAFTAR GAMBAR.....	ix
DAFTAR LAMBANG DAN SINGKATAN	x
DAFTAR ISTILAH.....	xi
INTISARI	xii
<i>ABSTRACT</i>	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA	7
2.1 Studi Literatur	7

2.2	Dasar Teori.....	16
2.3	Penelitian Terdahulu dan Analisis	23
BAB III METODE PENELITIAN		26
3.1	Objek Penelitian.....	26
3.2	Analisis Masalah dan Solusi	28
3.3	Alur Penelitian	29
3.4	Alat dan Bahan.....	34
BAB IV HASIL DAN PEMBAHASAN		35
4.1	Implementasi Sistem.....	35
4.2	Hasil Evaluasi Model.....	37
4.3	Visualisasi dan Analisis Hasil.....	40
4.4	Hasil Pembahasan	45
BAB V PENUTUP		45
5.1	Kesimpulan	47
5.2	Saran	48
DAFTAR PUSTAKA		49
LAMPIRAN.....		51

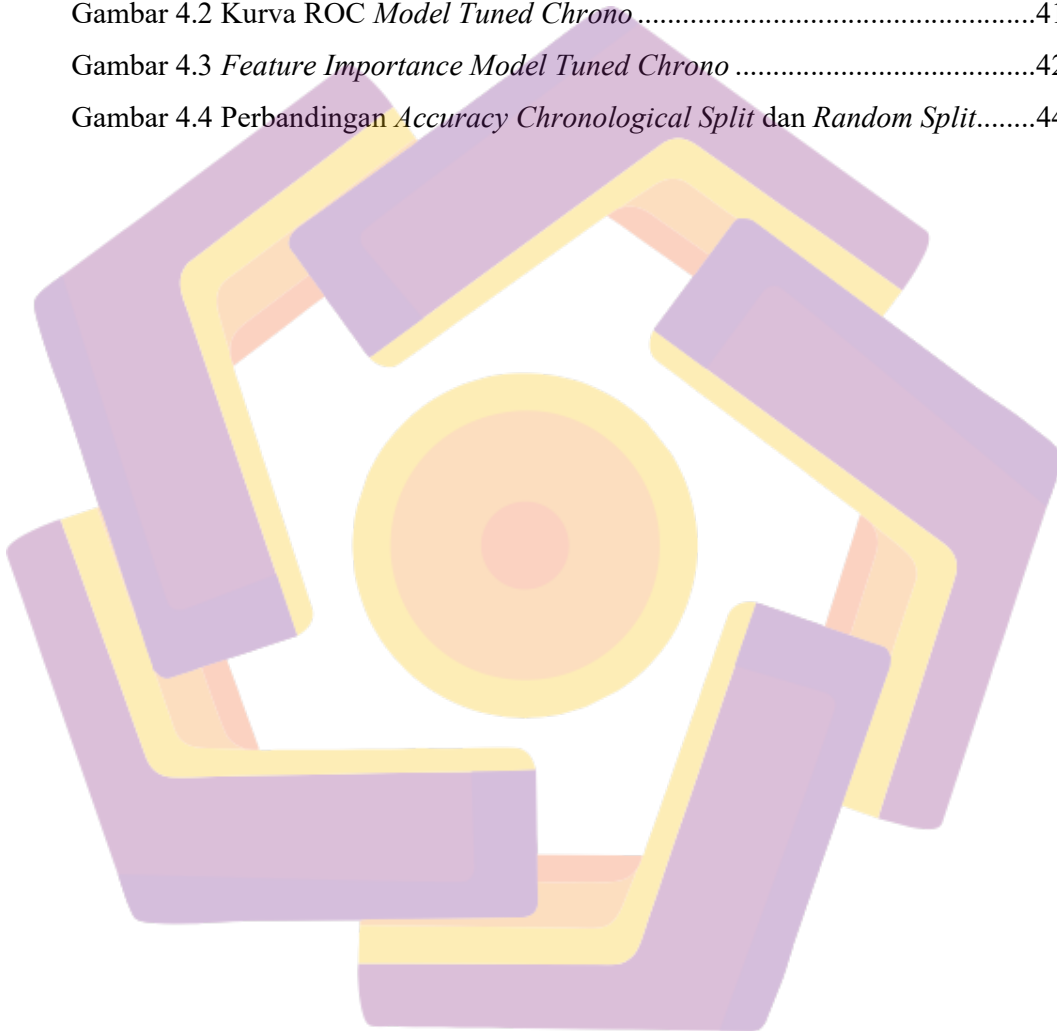
DAFTAR TABEL

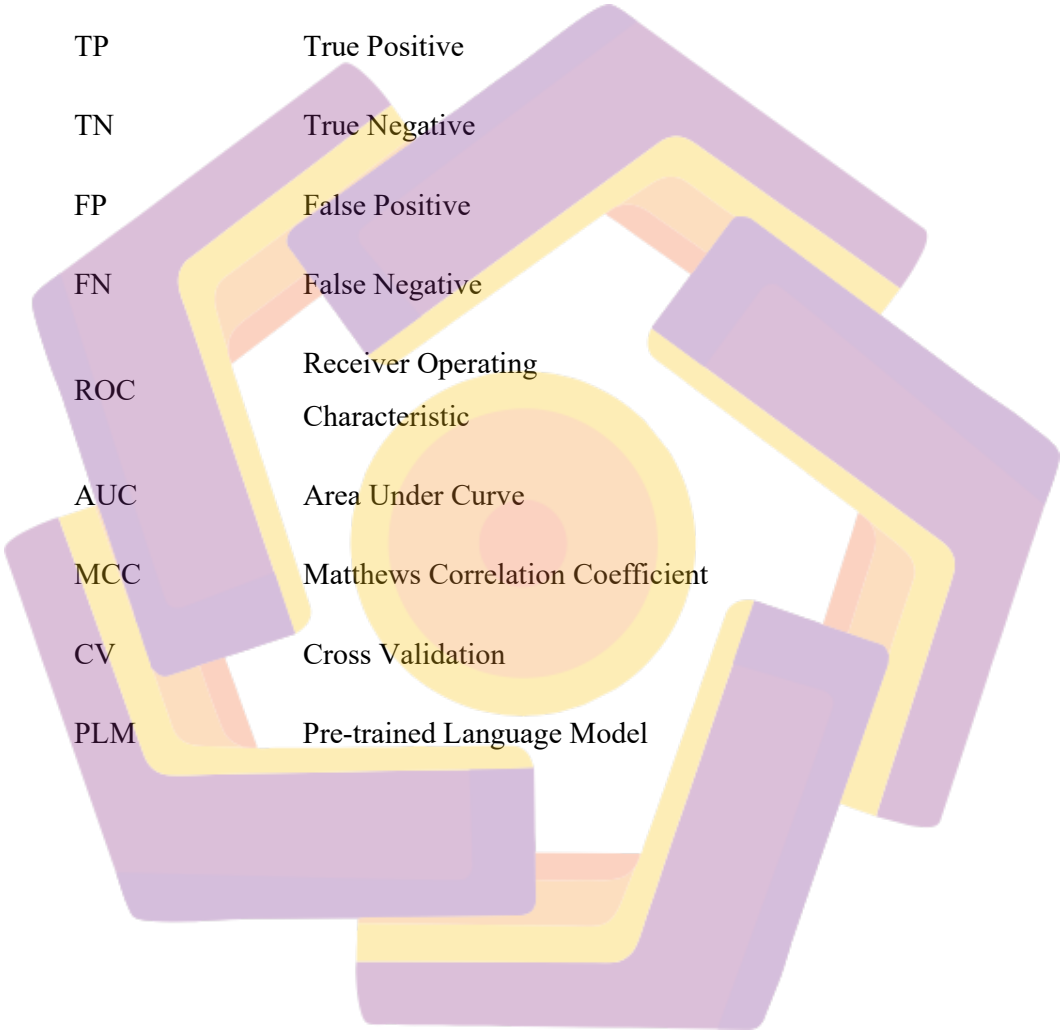
Tabel 2.1. Keaslian Penelitian	10
Tabel 3.1. Struktur Dataset ApacheJIT	27
Tabel 4.1 Hasil Evaluasi Model <i>Random Forest</i> pada Dua Skema Validasi.....	37



DAFTAR GAMBAR

Gambar 2.1. Perbandingan Validasi <i>Random Split</i> dan <i>Chronological Split</i>	22
Gambar 3.1 Alur Penelitian	30
Gambar 4.1 <i>Confusion matrix</i> Model Tuned Chrono	40
Gambar 4.2 Kurva ROC <i>Model Tuned Chrono</i>	41
Gambar 4.3 <i>Feature Importance</i> Model Tuned Chrono	42
Gambar 4.4 Perbandingan <i>Accuracy Chronological Split</i> dan <i>Random Split</i>	44



DAFTAR LAMBANG DAN SINGKATAN

SDP	Software Defect Prediction
JIT-DP	Just-in-Time Defect Prediction
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
ROC	Receiver Operating Characteristic
AUC	Area Under Curve
MCC	Matthews Correlation Coefficient
CV	Cross Validation
PLM	Pre-trained Language Model

DAFTAR ISTILAH

<i>Commit</i>	Satuan perubahan kode yang direkam ke dalam repositori versi
<i>Bug</i>	Kesalahan pada perangkat lunak yang menyebabkan perilaku tidak sesuai
<i>Clean Commit</i>	Commit yang tidak memperkenalkan cacat
<i>Buggy Commit</i>	Commit yang memperkenalkan cacat
<i>Chronological Split</i>	Metode pembagian data berdasarkan urutan waktu
<i>Random Split</i>	Metode pembagian data secara acak
<i>Temporal Leakage</i>	Kebocoran informasi masa depan ke data pelatihan
<i>Concept Drift</i>	Perubahan pola data seiring waktu
<i>Feature Importance</i>	Ukuran kontribusi fitur terhadap keputusan model
<i>Confusion Matrix</i>	Tabel evaluasi hasil klasifikasi
<i>Recall</i>	Proporsi bug yang berhasil terdeteksi oleh model
<i>Precision</i>	Proporsi prediksi bug yang benar
<i>ROC Curve</i>	Kurva performa klasifikasi berbasis threshold
<i>Dataset ApacheJIT</i>	Dataset commit-level untuk JIT-DP dari proyek Apache

INTISARI

Kualitas perangkat lunak dipengaruhi oleh kemampuan sistem dalam mengidentifikasi kesalahan lebih awal selama proses pengembangan. *Just-in-Time Defect Prediction* (JIT-DP) adalah metode yang memperkirakan kemungkinan terjadinya *bug* saat *commit* dilakukan dengan memanfaatkan metrik perubahan kode dan aktivitas pengembang. Tujuan dari penelitian ini adalah untuk mengembangkan dan menilai model prediksi *bug* berbasis algoritma *Random Forest* menggunakan Dataset ApacheJIT yang mencakup metrik tingkat *commit* dari berbagai proyek perangkat lunak *open-source*.

Langkah-langkah dalam penelitian ini mencakup pra-pemrosesan data, pembagian data dengan dua metode validasi (*Random Split* dan *Chronological Split*), pelatihan model, dan penilaian performa menggunakan metrik *Recall*, *Precision*, *F1-score*, ROC-AUC, dan MCC. Hasil dari eksperimen menunjukkan bahwa *Random Forest* dapat memberikan performa prediksi yang konsisten pada kedua metode validasi, meskipun *Random Split* menunjukkan performa yang lebih tinggi secara statistik. Namun, metode *Chronological Split* dianggap lebih representatif karena mencerminkan sifat waktu dalam alur *commit* dan mengurangi risiko kebocoran data. Analisis *Feature Importance* menunjukkan bahwa ukuran perubahan kode, terutama jumlah baris yang ditambahkan dan dihapus, adalah indikator utama dalam mempengaruhi terjadinya *commit* cacat.

Temuan ini memperkuat hasil penelitian sebelumnya yang menekankan pentingnya metrik perubahan dalam metode JIT-DP. Secara keseluruhan, penelitian ini menunjukkan bahwa algoritma *Random Forest* efektif digunakan dalam tugas prediksi *bug* berbasis *commit*, dengan efisiensi komputasi yang baik dan kemampuan interpretasi yang cukup. Diharapkan, hasil penelitian ini dapat menjadi landasan untuk mengembangkan alat bantu kualitas perangkat lunak yang dapat memberi peringatan dini kepada pengembang tentang risiko cacat pada tahap *commit*.

Kata Kunci: Prediksi Bug, Just-In-Time, Random Forest, ApacheJIT, Scikit-learn

ABSTRACT

Software quality is affected by how well a system can spot problems early during the software development stages. Just-in-Time Defect Prediction (JIT-DP) predicts the chances of issues when a change is made by analyzing code update metrics and developer activities. This research focuses on creating and testing a defect prediction model for commits, using the Random Forest algorithm and the ApacheJIT dataset, which consists of commit data from several open-source software projects.

The approach taken in this research includes preparing the data, dividing the dataset using two validation methods—Random Split and Chronological Split—training the model, and assessing its performance. To evaluate how well the model performs, metrics such as Recall, Precision, F1-score, ROC-AUC, and Matthews Correlation Coefficient (MCC) are used, ensuring a thorough analysis of its predicting abilities in various validation contexts.

Findings reveal that the Random Forest model performs consistently with both validation techniques, but the Random Split method yields better statistical results. On the other hand, the Chronological Split method better reflects real-life situations by maintaining time order and minimizing data leaks. An analysis of Feature Importance shows that the size of code changes, especially the number of lines added and removed, is the key factor in predicting faulty commits.

Keyword: *Bug Prediction, Just-In-Time, Random Forest, ApacheJIT, Scikit-learn.*