



**PENERAPAN ALGORITMA *RANDOM FOREST* DENGAN
SELEKSI FITUR ABC PADA *SOFTWARE DEFECT PREDICTION***

SKRIPSI

**Untuk memenuhi persyaratan
Dalam menyelesaikan sarjana Strata-1 Ilmu Komputer**

Oleh

**ZAINA FADIA LAILA HIDAYATI
NIM 1911016220025**

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
DESEMBER 2023**



**PENERAPAN ALGORITMA *RANDOM FOREST* DENGAN SELEKSI
FITUR ABC PADA *SOFTWARE DEFECT PREDICTION***

SKRIPSI

**Untuk memenuhi persyaratan
Dalam menyelesaikan sarjana Strata-1 Ilmu Komputer**

**Oleh
ZAINA FADIA LAILA HIDAYATI
NIM 1911016220025**

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU**

DESEMBER 2023

SKRIPSI

PENERAPAN ALGORITMA RANDOM FOREST DENGAN SELEKSI FITUR ABC PADA SOFTWARE DEFECT PREDICTION

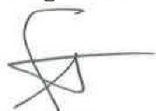
Oleh :

ZAINA FADIA LAILA HIDAYATI
NIM. 1911016220025

telah dipertahankan di depan Dosen Penguji pada tanggal 22 Desember 2023

Susunan Penguji :

Pembimbing Utama



Friska Abadi, S.Kom., M.Kom
NIP. 198809132023211010

Dosen Penguji 1



Mohammad Reza Faisal, S.T., M.T., Ph.D
NIP. 197612202008121001

Pembimbing Pendamping



Muliadi, S.Kom., M.Sc
NIP. 197804222010121002

Dosen Penguji II



Dwi Kartini, S.Kom., M.Kom
NIP. 198704212012122003



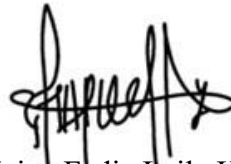
Yogi Pratomo, S.T., M.Kom

Yogi Pratomo, S.T., M.Kom
NIP. 197508252008121001

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, 1 Desember 2023



Zaina Fadia Laila Hidayati

NIM. 1911016220025

ABSTRAK

PENERAPAN ALGORITMA *RANDOM FOREST* DENGAN SELEKSI FITUR ABC PADA *SOFTWARE DEFECT PREDICTION*

(Oleh: Zaina Fadia Laila Hidayati; Pembimbing: Friska Abadi, S.Kom., M.Kom. dan Muliadi, S.Kom, M.Sc.; 2023; 49 Halaman)

Cacat yang mungkin timbul pada perangkat lunak saat proses pengembangan dapat mempengaruhi kualitas perangkat lunak. Oleh karena itu, perlu upaya untuk mencegah adanya kecacatan yang besar pada perangkat lunak yakni dengan prediksi cacat *software*. Metode klasifikasi digunakan untuk memprediksi cacat *software* agar dapat meminimalisir kecacatan. Namun, dataset yang digunakan pada proses klasifikasi kemungkinan terdapat fitur yang kurang relevan atau jumlah fiturnya terlalu banyak. Hal ini dapat diatasi dengan melakukan seleksi fitur pada dataset. Pada penelitian ini, algoritma *Random Forest* diterapkan untuk proses klasifikasi dan algoritma *Artificial Bee Colony* (ABC) sebagai metode seleksi fitur. Tujuan dari penelitian ini adalah untuk mengetahui akurasi yang dihasilkan oleh *Random Forest* dengan seleksi fitur ABC. Dari hasil penelitian yang dilakukan pada 3 dataset Relink, tanpa seleksi fitur didapatkan rata-rata akurasi sebesar 73,37%. Dan setelah diterapkan seleksi fitur ABC, rata-rata akurasi terbaik yang dihasilkan adalah sebesar 82,03%.

Kata Kunci: Prediksi Cacat Software, Seleksi Fitur, *Random Forest*, *Artificial Bee Colony*

ABSTRACT

APPLICATION OF THE RANDOM FOREST ALGORITHM WITH ABC FEATURES SELECTION IN DEFECT PREDICTION SOFTWARE

(By: Zaina Fadia Laila Hidayati; Supervisors: Friska Abadi, S.Kom., M.Kom; Muliadi, S.Kom, M.Sc.; 2023; 49 Pages)

Defects that may arise in software during the development process can affect the quality of the software. The classification method is used to predict software defects to minimize defects. However, the dataset used in the classification process may contain less relevant or have too many features. This can be overcome by selecting features in the dataset. In this research, the Random Forest algorithm is applied for the classification process, and the Artificial Bee Colony (ABC) algorithm as a feature selection method. The research aims is to determine the accuracy produced by Random Forest with ABC feature selection. From the results of research conducted on 3 Relink datasets, without feature selection, an average accuracy of 73.37% was obtained. After implementing ABC feature selection, the average accuracy increased to 82,03%

Keywords: *Artificial Bee Colony, Feature Selection, Random Forest, Software Defect Prediction*

PRAKATA

Puji dan Syukur penulis panjatkan kepada Tuhan yang maha Esa, karena atas nikmat dan kasih sayang-Nya penulis dapat menyelesaikan skripsi yang berjudul ***PENERAPAN ALGORITMA RANDOM FOREST DENGAN SELEKSI FITUR ABC PADA SOFTWARE DEFECT PREDICTION*** untuk memenuhi persyaratan dalam menyelesaikan Pendidikan S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

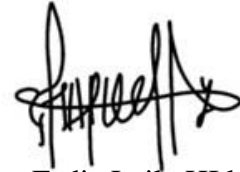
Pada lembar ini penulis bermaksud menyampaikan ucapan terimakasih kepada pihak-pihak yang telah mendukung dalam penyusunan skripsi ini, Adapun yang dimaksud adalah sebagai berikut.

1. Keluarga yang senantiasa mendukung, memberikan semangat, bantuan, dan doa dalam proses penyelesaian skripsi ini.
2. Bapak Friska Abadi, S.Kom., M.Kom. selaku dosen pembimbing utama dan yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
3. Bapak Muliadi, S.Kom., M.Sc. selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
4. Bapak Irwan Budiman S.T., M.Kom. selaku Ketua Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
5. Seluruh Dosen dan staff Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
6. Teman-teman dan sahabat-sahabat keluarga Ilmu Komputer angkatan 2019 yang memberikan dukungan dan selalu mengingatkan serta mendoakan dalam proses mengerjakan skripsi.
7. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna, namun penulis mengharapkan bantuan serupa berupa saran dan kritik

yang membangun dari semua pihak demi kesempurnaan dan mutu penulisan skripsi ini. Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, 1 Desember 2023



Zaina Fadia Laila Hidayati

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	i
PERNYATAAN.....	iii
ABSTRAK	iv
<i>ABSTRACT</i>	v
PRAKATA.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xii
BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian	3
1.5 Batasan Masalah	3
BAB II TINJAUAN PUSTAKA	
2.1 Kajian Terdahulu	4
2.2 Landasan Teori.....	7
2.2.1 Prediksi Cacat Software.....	7
2.2.2 Dataset Relink.....	7
2.2.3 Z-Score Normalization.....	9
2.2.4 Random Forest.....	9
2.2.5 Artificial Bee Colony.....	10
2.2.6 Split Data	13

2.2.7 Cross Validation.....	13
2.2.8 Akurasi.....	13
BAB III METODE PENELITIAN	
3.1 Alat Penelitian.....	15
3.2 Bahan Penelitian	15
3.3 Prosedur Penelitian	15
BAB IV HASIL DAN PEMBAHASAN	
4.1 Hasil.....	18
4.1.1 Pengumpulan Dataset.....	18
4.1.2 Normalisasi Z-Scores.....	20
4.1.3 Split data	21
4.1.4 Seleksi Fitur ABC.....	22
4.1.5 Klasifikasi Random Forest.....	31
4.2 Pembahasan.....	44
BAB V PENUTUP	
5.1 Kesimpulan	49
5.2 Saran	49
DAFTAR PUSTAKA	50
LAMPIRAN.....	53

DAFTAR TABEL

Tabel 1. Keaslian Penelitian.....	6
Tabel 2. Perencanaan Penelitian	6
Tabel 3. Spesifikasi Dataset Relink	8
Tabel 4. Jumlah data tiap kelas pada dataset Relink.....	18
Tabel 5. Dataset Apache	19
Tabel 6. Dataset Safe.....	19
Tabel 7. Dataset Zxing	19
Tabel 8. Dataset Apache sudah dinormalisasi.....	20
Tabel 9. Dataset Safe sudah dinormalisasi.....	20
Tabel 10. Dataset Zxing sudah dinormalisasi	21
Tabel 11. Hasil Pembagian Dataset	22
Tabel 12. Inisialisasi Posisi Seleksi Fitur ABC	22
Tabel 13. Posisi Awal Setelah Dikonversi Biner	23
Tabel 14. Subset fitur N ke-1	23
Tabel 15. Nilai fitness awal.....	24
Tabel 16. Perbedaan posisi awal dan fase employee	25
Tabel 17. Nilai <i>fitness</i> dataset Apache pada Iterasi 10	26
Tabel 18. Nilai <i>fitness</i> dataset Safe pada iterasi 10.....	26
Tabel 19. Nilai <i>fitness</i> dataset Zxing pada iterasi 10	27
Tabel 20. Nilai <i>fitness</i> dataset Apache pada iterasi 50.....	27
Tabel 21. Nilai <i>fitness</i> dataset Safe pada iterasi 50.....	27
Tabel 22. Nilai <i>fitness</i> dataset Zxing pada iterasi 50	28
Tabel 23. Nilai <i>fitness</i> dataset Apache pada iterasi 100.....	28
Tabel 24. Nilai <i>fitness</i> dataset Safe pada iterasi 100.....	29
Tabel 25. Nilai <i>fitness</i> dataset Zxing pada iterasi 100	29
Tabel 26. Data Fitur Terpilih	30
Tabel 27. Parameter Random Forest.....	31
Tabel 28. Data Bootstrapping	32
Tabel 29. Perhitungan Gini Impurity pada root node	33

Tabel 30. Perhitungan Gini Impurity pada Node 1.1	34
Tabel 31. Perhitunga gini impurity node 1.1.1	35
Tabel 32. Perhitungan gini impurity node 1.2	36
Tabel 33. Perhitungan Gini Impurity node 1.2.1	37
Tabel 34. Perhitungan gini impurity node 1.2.1.1	38
Tabel 35. Perhitungan gini impurity node 1.2.1.1.1	39
Tabel 36. Hasil Evaluasi Dataset Relink.....	40
Tabel 37. Akurasi Dataset Apache.....	41
Tabel 38. Akurasi Dataset Safe.....	42
Tabel 39. Akurasi Dataset Zxing	43
Tabel 40. Hasil Akurasi Dataset Relink.....	47

DAFTAR GAMBAR

Gambar 1. Struktur Umum Random Forest	10
Gambar 2. Alur Artificial Bee Colony	12
Gambar 3. Pembagian Data	13
Gambar 4. Diagram Alur Penelitian.....	16
Gambar 5. Jumlah data pada tiap kelas dataset relink	18
Gambar 6. Root Node Pohon Keputusan Random Forest	33
Gambar 7. Pohon Keputusan node 1.1	34
Gambar 8. Pohon Keputusan node 1.1.1	36
Gambar 9. Pohon Keputusan node 1.2.....	37
Gambar 10. Pohon Keputusan node 1.2.1	38
Gambar 11. Pohon Keputusan node 1.2.1.1	39
Gambar 12. Pohon Keputusan pertama random forest	40
Gambar 14. Nilai Akurasi Dataset Relink.....	41
Gambar 15. Nilai Akurasi dataset Apache.....	42
Gambar 16. Nilai Akurasi Dataset Safe	43
Gambar 17. Nilai Akurasi Dataset Zxing.....	44
Gambar 18. Fitur Terpilih	46
Gambar 19. Grafik Akurasi Dataset Relink	48