



**IMPLEMENTASI TEKNIK SELEKSI FITUR MENGGUNAKAN *FILTER*  
*FEATURE RANKING* (FFR) UNTUK PREDIKSI CACAT PERANGKAT  
LUNAK**

**Skripsi**

Untuk Memenuhi Persyaratan  
Dalam Menyelesaikan Sarjana Strata-1 Ilmu Komputer

Oleh  
**ABDUL ZULFIKAR YAHYA**  
**NIM. 1811016310024**

**PROGRAM STUDI S1-ILMU KOMPUTER**  
**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**  
**UNIVERSITAS LAMBUNG MANGKURAT**  
**BANJARBARU**  
**MEI 2025**



**IMPLEMENTASI TEKNIK SELEKSI FITUR MENGGUNAKAN FILTER  
FEATURE RANKING (FFR) UNTUK PREDIKSI CACAT PERANGKAT  
LUNAK**

**Skripsi**

Untuk Memenuhi Persyaratan  
Dalam Menyelesaikan Sarjana Strata-1 Ilmu Komputer

Oleh  
**ABDUL ZULFIKAR YAHYA**  
**NIM. 1811016310024**

**PROGRAM STUDI S1 ILMU KOMPUTER**  
**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**  
**UNIVERSITAS LAMBUNG MANGKURAT**  
**BANJARBARU**  
**MEI 2025**

# HALAMAN PENGESAHAN

## SKRIPSI

### IMPLEMENTASI TEKNIK SELEKSI FITUR MENGGUNAKAN FILTER FEATURE RANKING (FFR) UNTUK PREDIKSI CACAT PERANGKAT LUNAK

Oleh:  
**ABDUL ZULFIKAR YAHYA**  
NIM. 1811016310024

Telah dipertahankan di depan Dosen Penguji pada hari Kamis, tanggal 15 Mei 2025

Susunan Dosen Penguji:

**Pembimbing I**



Rudy Herteno, S.Kom., M.Kom.  
NIP. 198809252022031003

**Dosen Penguji I**



Triando Hamonangan Saragih, S.Kom., M.Kom.  
NIP. 199308242019031012

**Pembimbing II**



Radityo Adi Nugroho, S.T., M.Kom.  
NIP. 198212042008011006

**Dosen Penguji II**



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

Banjarbaru, Kamis/15 Mei 2025  
Koordinator Program Studi Ilmu Komputer



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

## **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, Mei 2025

Yang Menyatakan



**Abdul Zulfikar Yahya**

**NIM. 1811016310024**

## ABSTRAK

### IMPLEMENTASI TEKNIK SELEKSI FITUR MENGGUNAKAN FILTER FEATURE RANKING (FFR) UNTUK PREDIKSI CACAT PERANGKAT LUNAK

(Oleh: Abdul Zulfikar Yahya; Pembimbing: Rudy Herteno, S.Kom., M.Kom. dan Radityo Adi Nugroho, S.T., M.Kom; 2025; 72 halaman)

Software Defect Prediction (SDP) merupakan aktivitas penting dalam fase pengujian Software Development Life Cycle (SDLC) untuk mengidentifikasi modul yang rawan cacat, sehingga sumber daya pengujian dapat digunakan secara efisien. Salah satu tantangan dalam SDP adalah pemilihan fitur yang relevan untuk meningkatkan kinerja model prediksi. Filter feature ranking adalah metode seleksi fitur yang mengevaluasi fitur berdasarkan kriteria statistik tanpa melibatkan algoritma pembelajaran mesin tertentu. Penelitian ini bertujuan untuk mengetahui kinerja metode filter feature ranking menggunakan tiga teknik seleksi fitur Information Gain (IG), Relief (Rel), dan Chi-Square (CS). serta kombinasi di antaranya pada dua algoritma klasifikasi, yaitu Naïve Bayes dan Random Forest. Eksperimen dilakukan pada tiga dataset publik: ReLink, AEEEM, dan SOFTLAB. Evaluasi kinerja dilakukan menggunakan metrik Area Under the ROC Curve (AUC). Hasil penelitian menunjukkan bahwa penerapan filter feature ranking pada dataset AEEEM dan SOFTLAB dapat menghasilkan nilai AUC sebesar 0.70 hingga 0.87 pada kedua model klasifikasi, menunjukkan bahwa filter feature ranking mampu dalam meningkatkan kinerja prediksi cacat perangkat lunak. Namun, dataset ReLink menunjukkan performa yang lebih rendah, dengan nilai AUC hanya berkisar antara 0.60 hingga 0.70, sehingga kurang cocok untuk diterapkan dengan metode ini.

**Kata Kunci:** Software Defect Prediction, Information Gain, Relief, Chi-Square, Filter Feature Ranking.

## **ABSTRACT**

### **IMPLEMENTATION OF FEATURE SELECTION TECHNIQUE USING FILTER FEATURE RANKING (FFR) FOR SOFTWARE DEFECT PREDICTION**

*(By: Abdul Zulfikar Yahya; Advisors: Rudy Herteno, S.Kom., M.Kom. and Radityo Adi Nugroho, S.T., M.Kom; 2025; 72 pages)*

*Software Defect Prediction (SDP) is an important activity in the testing phase of the Software Development Life Cycle (SDLC) to identify defect-prone modules, so that testing resources can be used efficiently. One of the challenges in SDP is the selection of relevant features to improve the performance of the prediction model. Filter feature ranking is a feature selection method that evaluates features based on statistical criteria without involving specific machine learning algorithms. This study aims to determine the performance of the filter feature ranking method using three feature selection techniques Information Gain (IG), Relief (Rel), and Chi-Square (CS), as well as a combination of them on two classification algorithms, namely Naïve Bayes and Random Forest. Experiments were conducted on three public datasets: ReLink, AEEEM, and SOFTLAB. Performance evaluation was performed using the Area Under the ROC Curve (AUC) metric. The results show that applying the feature ranking filter to the AEEEM and SOFTLAB datasets can produce AUC values of 0.70 to 0.87 in both classification models, indicating that the feature ranking filter is capable of improving the performance of software defect prediction. However, the ReLink dataset showed lower performance, with AUC values only ranging from 0.60 to 0.70, making it less suitable for application with this method.*

**Keywords:** *Software Defect Prediction, Information Gain, Relief, Chi-Square, Filter Feature Ranking.*

## PRAKATA

Puji syukur penulis panjatkan ke Tuhan Yang Maha Esa karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “Implementasi Teknik Seleksi Fitur Menggunakan Filter Feature Ranking (FFR) Untuk Prediksi Cacat Perangkat Lunak” untuk memenuhi syarat dalam menyelesaikan pendidikan program S1-Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

Pada lembar ini penulis ingin menyampaikan ucapan terimakasih kepada pihak-pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut:

1. Keluarga yang selalu memberikan bantuan, semangat, doa dan dukungan dalam proses penyelesaian skripsi ini.
2. Bapak Rudy Herteno, S.Kom., M.Kom selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
3. Bapak Radityo Adi Nugroho, S.T., M. Kom selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
4. Ibu Dwi Kartini, S.Kom., M.Kom selaku Koordinator Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
5. Bapak Triando Hamonangan Saragih, S.Kom., M.Kom dan Bapak Friska Abadi, S.Kom., M.Kom selaku dosen penguji yang telah memberikan masukan-masukkan dalam skripsi ini.
6. Seluruh dosen dan staff Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
7. Teman-teman dan sahabat-sahabat keluarga Ilmu Komputer angkatan 2018 yang memberikan dukungan dan selalu mengingatkan serta mendoakan dalam proses mengerjakan skripsi.

8. 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, 15 Mei 2025



Abdul Zulfikar Yahya

## DAFTAR ISI

Skripsi .....	i
HALAMAN PENGESAHAN.....	ii
PERNYATAAN.....	iii
ABSTRAK.....	iv
<i>ABSTRACT</i> .....	v
PRAKATA.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Masalah.....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Kajian Terdahulu.....	5
2.2 Landasan Teori.....	6
2.2.1 Prediksi Cacat Perangkat Lunak.....	6
2.2.2 Dataset ReLink.....	7
2.2.3 Dataset AEEEM.....	8
2.2.4 Dataset SOFTLAB.....	9
2.2.5 Seleksi Fitur.....	10
2.2.6 <i>Information Gain</i> .....	11

2.2.7 <i>Relief</i> .....	12
2.2.8 <i>Chi-Square</i> .....	13
2.2.9 <i>Naïve Bayes</i> .....	14
2.2.10 <i>Random Forest</i> .....	15
2.2.11 <i>Normalization Min-Max</i> .....	16
2.2.12 <i>Area Under the ROC Curve</i> .....	16
2.2.13 Keaslian Penelitian .....	18
<b>BAB III METODE PENELITIAN</b> .....	19
3.1 Alat Penelitian .....	19
3.2 Prosedur Penelitian .....	19
<b>BAB IV HASIL DAN PEMBAHASAN</b> .....	22
4.1 Hasil.....	22
4.1.1 Pengumpulan Data.....	22
4.1.2 Pembobotan .....	26
4.1.3 Filter Feature Ranking .....	29
4.1.4 Klasifikasi.....	34
4.1.5 Evaluasi .....	36
a. AUC Pada Dataset ReLink.....	36
b. AUC Pada Dataset AEEEM .....	41
c. AUC Pada Dataset SOFTLAB .....	56
4.2 Pembahasan .....	62
<b>BAB V PENUTUP</b> .....	67
5.1 Kesimpulan.....	67
5.2 Saran .....	67
<b>DAFTAR PUSTAKA</b> .....	68

## DAFTAR TABEL

Tabel 1. Spesifikasi dataset ReLink.....	8
Tabel 2. Spesifikasi dataset AEEEM.....	8
Tabel 3. Spesifikasi dataset SOFTLAB.....	9
Tabel 4. Keakuratan Hasil Klasifikasi berdasarkan nilai AUC.....	17
Tabel 5. Keaslian Penelitian.....	18
Tabel 6. Tabel Perancangan Penelitian.....	18
Tabel 7. Hasil Klasifikasi berdasarkan nilai AUC.....	21
Tabel 8. Daftar Metrik ReLink.....	23
Tabel 9. Spesifikasi dataset ReLink.....	23
Tabel 10. Daftar Metrik AEEEM.....	24
Tabel 11. Spesifikasi dataset AEEEM.....	25
Tabel 12. Daftar Metrik SOFTLAB.....	25
Tabel 13. Spesifikasi dataset SOFTLAB.....	26
Tabel 14. Hasil dari label encoding.....	26
Tabel 15. Contoh hasil feature ranking dari dataset ReLink - apache.....	27
Tabel 16. Contoh hasil yang sudah di normalisasi.....	28
Tabel 17. Contoh hasil Information Gain dataset ReLink – apache.....	29
Tabel 18. Contoh hasil Relief dataset ReLink – apache.....	30
Tabel 19. Contoh hasil Chi-Square dataset ReLink – apache.....	31
Tabel 20. Contoh hasil IG+Rel dataset ReLink – apache.....	31
Tabel 21. Contoh hasil IG+CS dataset ReLink – apache.....	32
Tabel 22. Contoh hasil Rel+CS dataset ReLink – apache.....	33
Tabel 23. Contoh hasil IG+Rel+CS dataset ReLink – apache.....	33
Tabel 24. Contoh hasil Accuracy Naive Bayes dataset ReLink - apache.....	34
Tabel 25. Contoh hasil Accuracy Random Forest dataset ReLink - apache.....	35
Tabel 26. Hasil AUC Naive Bayes dataset ReLink - apache.....	36
Tabel 27. Hasil AUC Random Forest dataset ReLink - apache.....	37
Tabel 28. Hasil AUC Naive Bayes dataset ReLink - safe.....	38
Tabel 29. Hasil AUC Random Forest dataset ReLink - safe.....	38
Tabel 30. Hasil AUC Naive Bayes dataset ReLink - zxing.....	39

Tabel 31. Hasil AUC Random Forest dataset ReLink - zxing.....	40
Tabel 32. Hasil AUC Naive Bayes dataset AEEEM - EQ.....	41
Tabel 33. Hasil AUC Random Forest dataset AEEEM - EQ.....	42
Tabel 34. Hasil AUC Naive Bayes dataset AEEEM - JDT .....	44
Tabel 35. Hasil AUC Random Forest dataset AEEEM - JDT .....	45
Tabel 36. Hasil AUC Naive Bayes dataset AEEEM - LC .....	47
Tabel 37. Hasil AUC Random Forest dataset AEEEM - LC.....	48
Tabel 38. Hasil AUC Naive Bayes dataset AEEEM - ML .....	50
Tabel 39. Hasil AUC Random Forest dataset AEEEM - ML .....	52
Tabel 40. Hasil AUC Naive Bayes dataset AEEEM - PDE.....	53
Tabel 41. Hasil AUC Random Forest dataset AEEEM - PDE .....	55
Tabel 42. Hasil AUC Naive Bayes dataset SOFTLAB - ar1 .....	56
Tabel 43. Hasil AUC Random Forest dataset SOFTLAB - ar1 .....	57
Tabel 44. Hasil AUC Naive Bayes dataset SOFTLAB - ar4.....	58
Tabel 45. Hasil AUC Random Forest dataset SOFTLAB - ar4.....	59
Tabel 46. Hasil AUC Naive Bayes dataset SOFTLAB - ar5 .....	59
Tabel 47. Hasil AUC Random Forest dataset SOFTLAB - ar5.....	60
Tabel 48. Hasil AUC Naive Bayes dataset SOFTLAB - ar6 .....	61
Tabel 49. Hasil AUC Random Forest dataset SOFTLAB - ar6.....	61
Tabel 50. Rerata AUC Naive Bayes dataset ReLink .....	63
Tabel 51. Rerata AUC Naive Bayes dataset AEEEM.....	63
Tabel 52. Rerata AUC Naive Bayes dataset SOFTLAB.....	64
Tabel 53. Rerata AUC Random Forest dataset ReLink .....	65
Tabel 54. Rerata AUC Random Forest dataset AEEEM .....	65
Tabel 55. Rerata AUC Random Forest dataset SOFTLAB .....	65

## DAFTAR GAMBAR

Gambar 1. Model Pencegahan Cacat Perangkat Lunak .....	7
Gambar 2. Skema Kinerja Algoritma Random Forest .....	15
Gambar 3. Grafik ROC (Diskrit dan Kontinu).....	17
Gambar 4. Alur Penelitian.....	19
Gambar 5. AUC Naive Bayes dengan Feature Ranking.....	64
Gambar 6. AUC Random Forest dengan Feature Ranking.....	66