



**PREDIKSI CACAT *SOFTWARE* MENGGUNAKAN SELEKSI FITUR
FP-GROWTH PADA KNN YANG DIOPTIMASI DENGAN
ALGORITMA GENETIKA**

Skripsi

Untuk memenuhi tugas melakukan
penelitian dalam rangka penyusunan skripsi

Oleh

LUKMANUL HAKIM

NIM 1811016210026

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

JUNI 2023



**PREDIKSI CACAT *SOFTWARE* MENGGUNAKAN SELEKSI FITUR
FP-GROWTH PADA KNN YANG DIOPTIMASI DENGAN
ALGORITMA GENETIKA**

Skripsi

Untuk memenuhi tugas melakukan
penelitian dalam rangka penyusunan skripsi

Oleh

LUKMANUL HAKIM

NIM 1811016210026

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

JUNI 2023

SKRIPSI

PREDIKSI CACAT *SOFTWARE* MENGGUNAKAN SELEKSI FITUR *FP-GROWTH* PADA KNN YANG DIOPTIMASI DENGAN ALGORITMA GENETIKA

Oleh:

LUKMANUL HAKIM

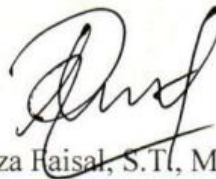
1811016210026

Telah dipertahankan di depan Dosen Penguji pada tanggal 21 Juni 2023

Susunan Dosen Penguji:

Pembimbing I

Dosen Penguji I



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

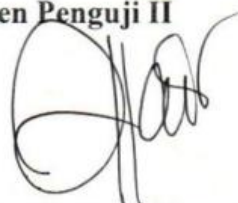

M. Reza Faisal, S.T., M.T., Ph.D.

NIP. 198212042008011006

NIP. 197612202008121001

Pembimbing II

Dosen Penguji II



Andi Farmadi, S.Si., M.T.

Dodon Turianto Nugrahadi S.Kom., M.Eng.

NIP. 197307252008011006

NIP. 198001122009121002



Banjarbaru, 26 Juni 2023

Departemen Studi Ilmu Komputer

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

NIP. 198212042008011006

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjaan 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, 25 Juni 2023

Yang menyatakan,



Lukmanul Hakim

NIM. 1811016210026

ABSTRAK

PREDIKSI CACAT *SOFTWARE* MENGGUNAKAN SELEKSI FITUR *FP-GROWTH* PADA KNN YANG DIOPTIMASI DENGAN ALGORITMA GENETIKA

(Oleh: Lukmanul Hakim; Pembimbing: Radityo Adi Nugroho, S.T., M.Kom. dan Farmadi, S.Si., M.T.; 2023; 73 halaman)

Prediksi cacat *software* adalah suatu proses membangun model *Machine Learning* untuk memprediksi kemungkinan terjadinya cacat atau kesalahan dalam *software*. Namun, pengembangan model prediksi cacat *software* dengan menggunakan metode *Machine Learning* sangatlah kompleks dan sulit, terutama ketika mencari atribut pengukuran kompleksitas asli yang relevan dan dapat diandalkan. Terlalu banyak atribut pengukuran kompleksitas asli dapat mengarah ke *Curse of Dimensionality* yang membuat model *Machine Learning* dapat mengalami *overfitting* pada data latih, sehingga menghasilkan model yang buruk pada data uji. Pada penelitian ini untuk menangani permasalahan tersebut, dilakukan pendekatan seleksi fitur *association rules* dengan algoritma *Fp-Growth* untuk mengatasi kompleksitas atribut pengukuran dan meningkatkan akurasi prediksi cacat *software*, serta algoritma klasifikasi *K-Nearest Neighbor* (KNN) sebagai metode klasifikasi yang dioptimasi dengan algoritma genetika. Tujuan dari penelitian ini adalah untuk mengetahui kinerja yang dihasilkan oleh seleksi fitur *FP-Growth* pada klasifikasi KNN yang dioptimasi algoritma genetika menggunakan 12 dataset NASA MDP. Hasil penelitian menunjukkan pendekatan klasifikasi KNN, seleksi fitur *FP-Growth* dengan klasifikasi KNN, klasifikasi KNN yang dioptimasi dengan algoritma genetika dan seleksi fitur *FP-Growth* pada klasifikasi KNN yang dioptimasi dengan algoritma genetika memiliki nilai rata-rata AUC secara berurutan yaitu 0,739, 0,796, 0,888 dan 0,880.

Kata Kunci: Prediksi Cacat *Software*, *Associaion Rules*, *FP-Growth*, Algoritma Genetika, KNN

ABSTRACT

SOFTWARE DEFECT PREDICTION USING FP-GROWTH FEATURE SELECTION ON KNN OPTIMIZED WITH GENETIC ALGORITHM

(By: Lukmanul Hakim; Advisors: Radityo Adi Nugroho, S.T., M.Kom. and Farmadi, S.Si., M.T.; 2023; 73 pages)

Software defect prediction is the process of building Machine Learning models to predict the likelihood of defects or errors in software. However, developing software defect prediction models using Machine Learning methods is complex and difficult, especially when looking for relevant and reliable native complexity measurement attributes. Too many original complexity measurement attributes can lead to the Curse of Dimensionality which makes Machine Learning models can experience overfitting on training data, resulting in poor models on test data. In this research, to deal with these problems, an association rules feature selection approach is carried out with the Fp-Growth algorithm to overcome the complexity of measurement attributes and improve the accuracy of software defect prediction, as well as the K-Nearest Neighbor (KNN) classification algorithm as a classification method optimized with genetic algorithms. The purpose of this study is to determine the performance generated by FP-Growth feature selection on KNN classification optimized by genetic algorithm using 12 NASA MDP datasets. The results showed that the KNN classification approach, FP-Growth feature selection with KNN classification, KNN classification optimized with genetic algorithms and FP-Growth feature selection on KNN classification optimized with genetic algorithms have average AUC values of 0.739, 0.796, 0.888 and 0.880, respectively.

Keywords: Software Defect Prediction, Associoaion Rules, FP-Growth, Genetic Algorithm, KNN

PRAKATA

Puji syukur penulis panjatkan ke Tuhan Yang Maha Esa karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul **“PREDIKSI CACAT *SOFTWARE* MENGGUNAKAN SELEKSI FITUR *FP-GROWTH* PADA KNN YANG DIOPTIMASI DENGAN ALGORITMA GENETIKA”** 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 Radityo Adi Nugroho, S.T., M.Kom. selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
3. Bapak Andi Farmadi, S.Si., M.T. 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. Seluruh staf keluarga PTIK ULM yang memberikan dukungan, bimbingan dan doa dalam pembuatan skripsi
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, 25 Juni 2023



Lukmanul Hakim

DAFTAR ISI

SKRIPSI	ii
PERNYATAAN	iii
ABSTRAK	iv
ABSTRACT	v
PRAKATA	vi
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xii
BAB I	13
1.1 Latar Belakang	13
1.2 Rumusan Masalah	14
1.3 Batasan Masalah	15
1.4 Tujuan Penelitian	15
1.5 Manfaat Penelitian	15
BAB II	16
2.1 Kajian Terdahulu	16
2.2 Machine Learning	19
2.3 Prediksi Cacat Software	20
2.4 Dataset NASA MDP	20
2.5 Associatin Rules	23
2.5.1 FP-Growth	24
2.6 Algoritma Genetika	27
2.7 K-Nearest Neighbors Algorithm	30
2.8 Cross Validation	31
2.9 Confusion Matrix	33
2.10 Area Under Curve (AUC)	35
BAB III	36
3.1 Alat Penelitian	36
3.2 Bahan Penelitian	36
3.3 Variabel Penelitian	36

3.4	Prosedur Penelitian.....	37
BAB IV	40
4.1	Hasil.....	40
4.1.1	Pengumpulan Data.....	40
4.1.2	Seleksi Fitur.....	40
4.1.3	Tahap <i>Modeling</i>	51
4.2	Pembahasan.....	71
BAB V	73
5.1	Kesimpulan.....	73
5.2	Saran.....	73
DAFTAR PUSTAKA	74
LAMPIRAN	79

DAFTAR TABEL

Tabel 1. kajian terdahulu.....	18
Tabel 2. Penelitian yang Diusulkan	19
Tabel 3. Spesifikasi Atribut Dataset NASA MDP	21
Tabel 4. Kriteria Dataset	22
Tabel 5. Lanjutan	23
Tabel 6. <i>Confusion Matrix</i>	33
Tabel 7. Hasil <i>Weight by Correlation</i> pada MC2	41
Tabel 8. Lanjutan	42
Tabel 9. Hasil perhitungan <i>size, frequency, score</i> dan <i>support</i> pada MC1	43
Tabel 10. Kelompok <i>non-Defect</i> pada MC2	44
Tabel 11. Lanjutan	45
Tabel 12. Kelompok <i>Defect</i> MC2	45
Tabel 13. Lanjutan	46
Tabel 14. Hasil diskritisasi MC2.....	47
Tabel 15. Lanjutan	48
Tabel 16. Hasil seleksi fitur	48
Tabel 17. Lanjutan	49
Tabel 18. Lanjutan	50
Tabel 19. Hasil Algoritma Genetika	52
Tabel 20. Lanjutan	53
Tabel 21. Hasil AUC CM1	55
Tabel 22. Hasil AUC JM1.....	56
Tabel 23. Lanjutan.....	57
Tabel 24. Hasil AUC KC1	58
Tabel 25. Hasil AUC KC3	59
Tabel 26. Hasil AUC MC1	60
Tabel 27. Lanjutan	61
Tabel 28. Hasil AUC MC2	62
Tabel 29. Hasil AUC MW1	63
Tabel 30. Hasil AUC PC1	64

Tabel 31. Lanjutan	65
Tabel 32. Hasil AUC PC2.....	66
Tabel 33. Hasil AUC PC3.....	67
Tabel 34. Hasil AUC PC4.....	68
Tabel 35. Lanjutan	69
Tabel 36. Hasil AUC PC5.....	70
Tabel 37. Perbandingan AUC tertinggi setiap pendekatan	72

DAFTAR GAMBAR

Gambar 1. Rumus untuk <i>support</i> , <i>confidence</i> dan <i>lift</i> dengan aturan $X \rightarrow Y$...	24
Gambar 2. <i>Flowchart FP-Growth</i>	25
Gambar 3. Alur Algoritma Genetika.....	30
Gambar 4. Pembagian Data pada <i>10-Fold Cross Validation</i>	32
Gambar 5. Alur Penelitian.....	37
Gambar 6. Grafik AUC CM1.....	56
Gambar 7. Grafik AUC JM1.....	57
Gambar 8. Grafik AUC KC1.....	58
Gambar 9. Grafik AUC KC3.....	60
Gambar 10. Grafik AUC MC1.....	61
Gambar 11. Grafik AUC MC2.....	62
Gambar 12. Grafik AUC MW1.....	64
Gambar 13. Grafik AUC PC1.....	65
Gambar 14. Grafik AUC PC2.....	66
Gambar 15. Grafik AUC PC3.....	68
Gambar 16. Grafik AUC PC4.....	69
Gambar 17. Grafik AUC PC5.....	70
Gambar 18. Grafik Perbandingan AUC tertinggi setiap pendekatan.....	71