



**OPTIMIZED MULTI CORRELATION-BASED FEATURE SELECTION
IN SOFTWARE DEFECT PREDICTION**

Skripsi

**Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

Oleh

MUHAMMAD NABIL MUYASSAR RAHMAN

NIM 2011016210001

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
FEBRUARI 2024**



**OPTIMIZED MULTI CORRELATION-BASED FEATURE SELECTION
IN SOFTWARE DEFECT PREDICTION**

Skripsi

**Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

Oleh

MUHAMMAD NABIL MUYASSAR RAHMAN

NIM 2011016210001

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

FEBRUARI 2024

SKRIPSI

OPTIMIZED MULTI CORRELATION-BASED FEATURE SELECTION IN SOFTWARE DEFECT PREDICTION

Oleh :

Muhammad Nabil Muyassar Rahman
NIM. 2011016210001

Telah dipertahankan di depan Dosen Penguji pada tanggal 26 Januari 2024.

Susunan Dosen Penguji:

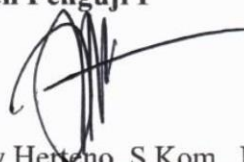
Dosen Pembimbing I



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

NIP. 198212042008011006

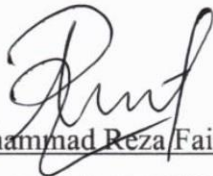
Dosen Penguji I



Rudy Herteno, S.Kom., M.Kom.

NIP. 198809252022031003

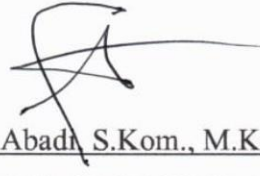
Dosen Pembimbing II



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

NIP. 197612202008121001

Dosen Penguji II



Friska Abadi, S.Kom., M.Kom.

NIP. 198809132023211010



Koordinator PS Ilmu Komputer,

Muhammad Budiman, S.T., M.Kom.

NIP. 197703252008121001

KATA PENGANTAR

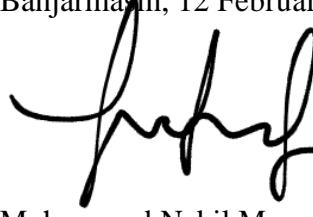
Puji syukur saya panjatkan pada Allah SWT karena berkat kasih, rahmat dan karunia-Nya penulis dapat menyelesaikan jurnal yang berjudul “*Optimized Multi Correlation-Based Feature Selection in Software Defect Prediction*” untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

Pada lembar ini penulis ingin menyampaikan ucapan terima kasih kepada semua pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan jurnal ini, adapun yang dimaksud adalah sebagai berikut:

1. Keluarga besar yang selalu memberikan bantuan, semangat, doa dan kelimpahan dukungan dalam proses penyelesaian jurnal ini.
2. Kedua orang tua yang selalu menguatkan, memberikan doa, dukungan dan optimis bahwa jurnal ini pasti dapat selesai dengan baik.
3. Kakak Muhammad Daffa Abiyyu Rahman yang selalu memberikan opini revisi terkait *review* jurnal ini.
4. Bapak Radityo Adi Nugroho, S.T., M.Kom. selaku dosen pembimbing utama dan Bapak M. Reza Faisal, S. T., M. T., Ph. D. selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian jurnal ini.
5. Bapak Irwan Budiman, S. T., M. Kom. selaku Ketua Program Studi Ilmu Komputer FMIPA ULM, serta seluruh Dosen dan Staff Prodi Ilmu Komputer yaitu Ka Azizah atas bantuan dan saran beliau jurnal ini dapat diselesaikan dengan baik.
6. Kak Ajwa Helisa dan Bang Meftah Mafazy yang telah membantu dalam pemberkasan sehingga jurnal ini bisa terselesaikan dengan baik.
7. Teman-teman dan sahabat-sahabat keluarga Ilmu Komputer yang memberikan dukungan dan selalu mengingatkan serta mendoakan dalam proses mengerjakan jurnal.
8. Serta semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian jurnal ini.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna. Meski demikian, diharapkan tulisan ini dapat memberikan manfaat bagi ilmu pengetahuan dan pembaca, serta mendapatkan berkah yang melimpah dari Tuhan Yang Maha Esa.

Banjarmasin, 12 Februari 2024

A handwritten signature in black ink, appearing to read 'Nabil', written in a cursive style.

Muhammad Nabil Muyassar Rahman

ABSTRAK

OPTIMASI PARTICLE SWARM OPTIMIZATION MENGGUNAKAN CORRELATION BASED MULTI-FILTERING PADA SOFTWARE DEFECT PREDICTION

(Oleh : Muhammad Nabil Muyassar Rahman; Pembimbing : Radityo Adi Nugroho, S.T., M.Kom. dan Mohammad Reza Faisal, S.T., M.T., Ph.D.; 2024; halaman)

Dalam prediksi cacat perangkat lunak, atribut yang berisik dan data berdimensi tinggi masih menjadi tantangan yang kritis. Makalah ini memperkenalkan pendekatan baru yang dikenal sebagai *Multi Correlation-Based Feature Selection* (MCFS), yang berusaha untuk mengatasi tantangan-tantangan ini. MCFS mengintegrasikan dua teknik seleksi fitur, yaitu *Correlation-Based Feature Selection* (CFS) dan *Correlation Matrix-Based Feature Selection* (CMFS), yang bertujuan untuk mengurangi dimensi data dan menghilangkan atribut yang berisik. Untuk mencapai hal ini, CFS dan CMFS diterapkan secara independen untuk menyaring set data, dan rata-rata tertimbang dari hasilnya dihitung untuk menentukan pemilihan fitur yang optimal. Pendekatan ini tidak hanya mengurangi dimensi data tetapi juga mengurangi dampak dari atribut yang berisik. Untuk lebih meningkatkan kinerja prediktif, makalah ini memanfaatkan algoritma *Particle Swarm Optimization* (PSO) sebagai mekanisme pemilihan fitur, khususnya menargetkan peningkatan dalam *Area Under the Curve* (AUC). Evaluasi metode yang diusulkan dilakukan pada 12 *dataset benchmark* yang bersumber dari korpus NASA MDP, yang terkenal dengan atribut yang berisik, berdimensi tinggi, dan catatan kelas yang tidak seimbang. Temuan penelitian menunjukkan bahwa MCFS mengungguli CFS dan CMFS, menghasilkan nilai AUC rata-rata 0,891, sehingga menekankan kemampuannya dalam memajukan kinerja klasifikasi dalam konteks prediksi cacat perangkat lunak menggunakan klasifikasi *K-Nearest Neighbors* (KNN).

Kata kunci: *Software Defect, Correlation-Based, Feature Selection, High Dimensional, Noisy Attribute*

ABSTRACT

OPTIMIZED MULTI CORRELATION-BASED FEATURE SELECTION IN SOFTWARE DEFECT PREDICTION

(By: Muhammad Nabil Muyassar Rahman; Supervisors: Radityo Adi Nugroho, S.T., M.Kom. and Mohammad Reza Faisal, S.T., M.T., Ph.D.; 2024; page)

In software defect prediction, noisy attributes and high-dimensional data remain to be a critical challenge. This paper introduces a novel approach known as Multi Correlation-Based Feature Selection (MCFS), which seeks to address these challenges. MCFS integrates two feature selection techniques, namely Correlation-Based Feature Selection (CFS) and Correlation Matrix-Based Feature Selection (CMFS), intending to reduce data dimensionality and eliminate noisy attributes. To accomplish this, CFS and CMFS are applied independently to filter the datasets, and a weighted average of their outcomes is computed to determine the optimal feature selection. This approach not only reduces data dimensionality but also mitigates the impact of noisy attributes. To further enhance predictive performance, this paper leverages the Particle Swarm Optimization (PSO) algorithm as a feature selection mechanism, specifically targeting improvements in the Area Under the Curve (AUC). The evaluation of the proposed method is conducted on 12 benchmark datasets sourced from the NASA MDP corpus, renowned for their noisy attributes, high dimensionality, and imbalanced class records. The research findings demonstrate that MCFS outperforms CFS and CMFS, yielding an average AUC value of 0.891, thereby emphasizing its efficacy in advancing classification performance in the context of software defect prediction using K-Nearest Neighbors (KNN) classification.

Keywords: *Software Defect, Correlation-Based, Feature Selection, High Dimensional, Noisy Attribute*

SURAT PERNYATAAN

Dengan ini saya menyatakan bahwa dalam jurnal 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.

Banjarmasin, 12 Februari 2024

Yang Menyatakan,



Muhamamd Nabil Muyassar Rahman

NIM. 2011016210001