



***COMPARISON BETWEEN K-FOLD CROSS VALIDATION AND
PERCENTAGE SPLIT IN DECISION TREE
ALGORITHMS FOR ANEMIA CLASSIFICATION***

Skripsi

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

**Oleh
NANDA PUTRI RAHMAWATI
NIM 2011016320021**

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JUNI 2026**



***COMPARISON BETWEEN K-FOLD CROSS VALIDATION AND
PERCENTAGE SPLIT IN DECISION TREE
ALGORITHMS FOR ANEMIA CLASSIFICATION***

Skripsi

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

**Oleh
NANDA PUTRI RAHMAWATI
NIM 2011016320021**

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JUNI 2026**

SKRIPSI

COMPARISON BETWEEN K-FOLD CROSS VALIDATION AND PERCENTAGE SPLIT IN DECISION TREE ALGORITHMS FOR ANEMIA CLASSIFICATION

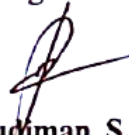
Oleh :

NANDA PUTRI RAHMAWATI
NIM. 2011016320021

telah dipertahankan di depan Dosen Penguji pada tanggal 30 April 2026

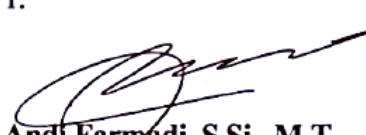
Susunan Penguji :

Pembimbing Utama

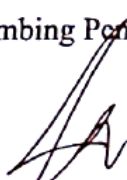

Irwan Budiman, S.T., M.Kom.
NIP. 197703252008121001

Penguji

1.

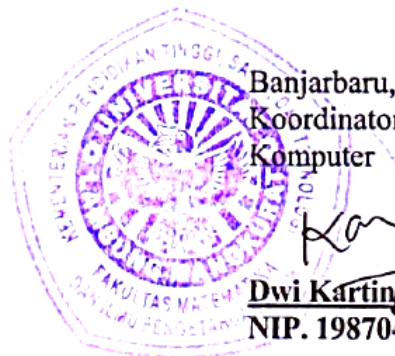

Andi Farmadi, S.Si., M.T.
NIP. 198212042008011006

Pembimbing Pendamping


**Muhammad Itqan Mazdadi, S.Kom.,
M.Kom.**
NIP. 199006122019031013

2.


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



Banjarbaru, 11 Juni 2026
Koordinator Program Studi Ilmu
Komputer


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

KATA PENGANTAR

Puji syukur saya panjatkan pada Allah SWT karena berkat kasih, rahmat dan karunia-Nya penulis dapat menyelesaikan jurnal yang berjudul “*Comparison Between K-Fold Cross Validation And Percentage Split In Decision Tree Algorithms For Anemia Classifications*” 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. Allah SWT, yang telah memberikan rahmat dan karunianya selama masa perkuliahan dan pengerjaan tugas akhir ini. Karunia berupa kemudahan dan kelancaran dalam pengerjaan jurnal serta diberikannya rahmat seperti kesehatan, keuangan, waktu, serta yang terpenting iman dan islam yang Alhamdulillah menjadi salah satu faktor penting dalam terselesaikannya tugas akhir in.
2. Kedua orang tua saya, yang selalu membantu dalam jerih payahnya, keringatnya, doa dan harapannya, semangat yang selalu diberikan yang memotivasi saya untuk dapat menyelesaikan laporan tugas akhir ini sesegera mungkin. Teruntuk Ibunda tercinta, yang tak pernah lelah untuk memberikan dukungan dan doanya, yang demikian membuat saya semangat untuk menyelesaikan pendidikan perguruan tinggi saya. Tak lupa juga kepada Ayahanda, Sabirin, yang tak pernah mengeluh membiayai biaya pendidikan saya, yang demikian membuat saya semangat untuk menyelesaikan pendidikan perguruan tinggi saya.
3. Diri saya sendiri, yang mampu menyelesaikan tugas akhir ini. Terimakasih karena selama pengerjaan jurnal tidak pernah menyerah, selalu mencari solusi dan mampu untuk keluar ketika dihadapkan oleh masalah.
4. Bapak Irwan Budiman, S.T., M.Kom. selaku dosen pembimbing utama dan Bapak Muhammad Itqan Mazdadi, S.Kom., M.Kom. selaku dosen

pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian jurnal ini.

5. Ibu Dwi Kartini, S.Kom., M.Kom. selaku Ketua Program Studi Ilmu Komputer FMIPA ULM, serta seluruh Dosen dan staff.
6. Teman-teman dan sahabat-sahabat keluarga Ilmu Komputer yang memberikan dukungan dalam proses mengerjakan jurnal.
7. 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.

Banjarbaru, 23 Juni 2026



Nanda Putri Rahmawati

ABSTRAK

PERBANDINGAN ANTARA K-FOLD CROSS VALIDATION DAN PERCENTAGE SPLIT PADA ALGORITMA DECISION TREE UNTUK KLASIFIKASI ANEMIA

(Oleh: Nanda Putri Rahmawati; Pembimbing: Irwan Budiman, S.T., M.Kom. dan Muhammad Itqan Mazdadi, S.Kom., M.Kom.: 2026: halaman)

Anemia merupakan tantangan kesehatan global yang signifikan yang ditandai dengan defisit patologis dalam konsentrasi hemoglobin, yang sering kali menyebabkan ketidakstabilan fisiologis. Diagnosis klinis yang akurat biasanya bergantung pada tes darah lengkap (complete blood count/CBC), yang memberikan parameter hematologi penting untuk keperluan klasifikasi. Meskipun model machine learning telah menunjukkan efikasi yang tinggi dalam mendiagnosis anemia, penelitian yang ada sering kali bergantung pada strategi partisi data statis yang mungkin mengabaikan keandalan evaluasi dan stabilitas performa. Penelitian ini mengatasi kesenjangan tersebut dengan mengalihkan fokus dari perbandingan arsitektur ke ketahanan validasi, secara khusus mengevaluasi performa algoritma C4.5 di berbagai teknik pembagian data. Penelitian ini memanfaatkan dataset yang terdiri dari 1.281 rekam klinis dengan 14 fitur numerik dan 9 label jenis anemia. Untuk menilai stabilitas, dua strategi partisi yang berbeda diimplementasikan, yaitu Percentage Split statis (berkisar dari rasio 60:40 hingga 90:10) dan K-Fold Cross Validation iteratif (dengan nilai K sebesar 3, 5, 7, 10, dan 15). Hasil eksperimen menunjukkan bahwa algoritma C4.5 mencapai performa puncaknya menggunakan Percentage Split 90:10, dengan mencapai rata-rata akurasi sebesar 99,46%, presisi 98,32%, dan recall 99,28%. Sebagai perbandingan, pendekatan K-Fold (K=10) menghasilkan akurasi yang sedikit lebih rendah namun lebih stabil, yaitu sebesar 99,19% dengan standar deviasi yang berkurang secara signifikan ($\pm 0,09$), yang menyoroti keandalannya untuk aplikasi klinis. Meskipun percentage split dengan rasio tinggi memaksimalkan paparan pelatihan dan potensi prediktif, metode K-Fold memberikan tolok ukur yang lebih objektif dan dapat digeneralisasi dengan memperhitungkan seluruh distribusi data. Penelitian ini lebih lanjut mengidentifikasi tantangan dalam mengklasifikasikan kelas minoritas, seperti Leukemia dengan trombositopenia, akibat kelangkaan data bawaan. Pada akhirnya, penelitian ini menegaskan bahwa algoritma C4.5, apabila dipasangkan dengan protokol partisi yang optimal, tetap menjadi solusi yang tangguh dan sangat dapat diinterpretasikan untuk skrining anemia klinis, mengungguli beberapa arsitektur modern yang kompleks.

Kata kunci: Anemia; Decision Tree; K-Fold Cross Validation; Percentage Split; Algoritma C4.5

ABSTRACT

COMPARISON BETWEEN K-FOLD CROSS VALIDATION AND PERCENTAGE SPLIT IN DECISION TREE ALGORITHMS FOR ANEMIA CLASSIFICATIONS

(By : Nanda Putri Rahmawati; Pembimbing : Irwan Budiman, S.T., M.Kom. and Muhammad Itqan Mazdadi, S.Kom., M.Kom.: 2026: page)

Anemia is a significant global health challenge characterized by a pathological deficit in hemoglobin concentration, often leading to physiological instability. Accurate clinical diagnosis typically relies on complete blood count (CBC) tests, which provide critical hematological parameters for classification. While machine learning models have demonstrated high efficacy in diagnosing anemia, existing research often relies on static data partitioning strategies that may overlook evaluation reliability and performance stability. This study addresses this gap by shifting the focus from architectural benchmarking to validation robustness, specifically evaluating the C4.5 algorithm's performance across different data-splitting techniques. The research utilizes a dataset comprising 1,281 clinical records with 14 numerical features and 9 anemia-type labels. To assess stability, two distinct partitioning strategies were implemented: a static Percentage Split (ranging from 60:40 to 90:10) and iterative K-Fold Cross Validation (with K values of 3, 5, 7, 10, and 15). Experimental results demonstrate that the C4.5 algorithm achieved its peak performance using the 90:10 Percentage Split, reaching an average accuracy of 99.46%, a precision of 98.32%, and a recall of 99.28%. In comparison, the K-Fold (K=10) approach yielded a slightly lower but more stable accuracy of 99.19% with a significantly reduced standard deviation (± 0.09), highlighting its reliability for clinical applications. While the high-ratio percentage split maximizes training exposure and predictive potential, the K-Fold method provides a more objective, generalizable benchmark by accounting for the entire data distribution. The study further identifies challenges in classifying minority classes, such as Leukemia with thrombocytopenia, due to inherent data scarcity. Ultimately, this research confirms that the C4.5 algorithm, when paired with an optimal partitioning protocol, remains a robust and highly interpretable solution for clinical anemia screening, outperforming several complex modern architectures.

Keywords: Anemia; Decision Tree; K-Fold Cross Validation; Percentage Split; C4.5 Algorithm

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.

Banjarbaru, 23 Juni 2026

Yang Menyatakan,



Nanda Putri Rahmawati

NIM. 2011016320021