KOMPARASI TEKNIK IMPUTASI UNTUK NILAI NOL DAN PARADIGMA SELEKSI FITUR (*GA* DAN *RFE-SHAP*) UNTUK KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN *RANDOM FOREST*

Skripsi

Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer

Oleh
MUHAMMAD HENRY ADITYA
NIM 2211016210016

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

KOMPARASI TEKNIK IMPUTASI UNTUK NILAI NOL DAN PARADIGMA SELEKSI FITUR (*GA* DAN *RFE-SHAP*) UNTUK KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN *RANDOM FOREST*

Skripsi

Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer

Oleh

MUHAMMAD HENRY ADITYA

NIM 2211016210016

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

FEBRUARI 2026

SKRIPSI

KOMPARASI TEKNIK IMPUTASI UNTUK NILAI NOL DAN PARADIGMA SELEKSI FITUR (*GA* DAN *RFE-SHAP*) UNTUK KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN *RANDOM FOREST*

Oleh:

MUHAMMAD HENRY ADITYA

NIM 2211016210016

Telah dipertahankan di depan Dosen Penguji pada tanggal 25 Februari 2026.

Susunan Dosen Penguji:

Pembimbing I

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

Dosen Penguji I

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

Pembimbing II

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

Dosen Penguji II

Fatma Indriani, S.T., M.I.T., Ph.D.
NIP. 198404202008122004

Banjarbaru, 26 Februari 2026

Koordinator Program Studi Ilmu Komputer

Dwi Kartini, 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, 25 Februari 2026

Yang Menyatakan,

Muhammad Henry Aditya

NIM. 2211016210016

ABSTRAK

KOMPARASI TEKNIK IMPUTASI UNTUK NILAI NOL DAN PARADIGMA SELEKSI FITUR (GA DAN RFE-SHAP) UNTUK KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN RANDOM FOREST

(Oleh: Muhammad Henry Aditya; Pembimbing: Muhammad Itqan Mazdadi, S.Kom., M.Kom. dan Muliadi, S.Kom., M.Cs.; 2026; 78 halaman)

Diabetes salah satu penyakit kronis yang menjadi tantangan kesehatan global dan memerlukan deteksi dini yang akurat serta andal. Pemanfaatan *machine learning*, khususnya menggunakan *dataset Pima Indians Diabetes*, telah banyak dilakukan untuk mendukung klasifikasi penyakit ini. Namun, *dataset* tersebut memiliki permasalahan kualitas data berupa nilai nol yang tidak valid secara medis. Selain itu, perbedaan teknik imputasi dan paradigma seleksi fitur berpotensi menghasilkan variasi performa sekaligus memengaruhi keterbacaan model. Penelitian ini bertujuan mengevaluasi secara komparatif pengaruh kombinasi teknik imputasi (*mean*, KNN, dan MICE) dan seleksi fitur dengan paradigma berbeda, yaitu *Genetic Algorithm* (GA) dan RFE-SHAP, terhadap performa dan keterbacaan model *Random Forest*. Hasil penelitian menunjukkan bahwa kombinasi MICE–GA menghasilkan performa paling tinggi dan efisien pada akurasi (0,7727), *precision* (0,6610), dan F1-*score* (0,6903), mean–RFE-SHAP unggul pada *recall* (0,7407) dengan jumlah fitur lebih sedikit, sedangkan *mean*–GA memperoleh AUC tertinggi (0,8406), di mana seluruh kombinasi mampu menghasilkan AUC > 0,8 dan RFE-SHAP mengidentifikasi *Glucose* sebagai fitur paling dominan diikuti BMI, *Age*, dan Insulin berdasarkan nilai kepentingan fitur (SHAP *importance*) berada pada rentang 0,0740 hingga 0,1733 serta arah kontribusi (*directional contribution*) pada rentang –0,0028 hingga +0,0031.

Kata Kunci: Diabetes, Imputasi Data, Seleksi Fitur, Random Forest, RFE-SHAP

ABSTRACT

COMPARISON OF IMPUTATION TECHNIQUES FOR ZERO VALUES AND FEATURE SELECTION PARADIGM (GA AND RFE-SHAP) FOR DIABETES DISEASE CLASSIFICATION USING RANDOM FOREST

(By: Muhammad Henry Aditya; Supervisor: Muhammad Itqan Mazdaadi, S.Kom., M.Kom. & Muliadi, S.Kom., M.Cs.; 2026; 78 pages)

Diabetes is a chronic disease that poses a global health challenge and requires accurate and reliable early detection. The use of machine learning, particularly using the Pima Indians Diabetes dataset, has been widely used to support the classification of this disease. However, this dataset has data quality issues in the form of medically invalid null values. Furthermore, differences in imputation techniques and feature selection paradigms have the potential to produce performance variations and affect model readability. This study aims to comparatively evaluate the effect of a combination of imputation techniques (mean, KNN, and MICE) and feature selection with different paradigms, namely the Genetic Algorithm (GA) and RFE-SHAP, on the performance and readability of the Random Forest model. The results showed that the combination of MICE–GA produced the highest and most efficient performance in accuracy (0,7727), precision (0,6610), and F1-score (0,6903), mean–RFE-SHAP excelled in recall (0,7407) with fewer features, while mean–GA obtained the highest AUC (0,8406), where all combinations were able to produce AUC > 0.8 and RFE-SHAP identified Glucose as the most dominant feature followed by BMI, Age, and Insulin based on feature importance values (SHAP importance) in the range of 0,0740 to 0,1733 and directional contribution in the range of –0,0028 to +0,0031.

***Keywords**: Diabetes, Data Imputation, Feature Selection, Random Forest, RFE-SHAP*

PRAKATA

Puji syukur penulis haturkan kepada Tuhan Yang Maha Esa, Allah SWT, karena atas berkat, rahmat, dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul "Komparasi Teknik Imputasi untuk Nilai Nol dan Paradigma Seleksi Fitur (*GA* dan *RFE-SHAP*) untuk Klasifikasi Penyakit Diabetes Menggunakan *Random Forest*".

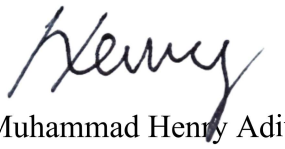
Melalui lembar ini, penulis ingin menyampaikan ucapan terima kasih sebesar-besarnya kepada berbagai pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi sehingga pada akhirnya skripsi ini dapat diselesaikan, adapun pihak-pihak yang dimaksud adalah sebagai berikut.

1. Keluarga besar, terutama Abah, Mama, Wina, Kai, Nini, Tante Ati, Bapak H. M. Isra Ismail, S.H., M.H., Ibu Hj. Hairiyati, S.H. dan sekeluarga, Ibu Ir. Hj. Ernawati dan sekeluarga, Ibu Hairiyah dan sekeluarga, Bapak Dr.Eng Ir. Irfan Prasetia, S.T., M.T. dan sekeluarga, yang selalu memberikan doa, semangat, dukungan, hingga kepercayaan, sehingga penulis termotivasi untuk mengejar pendidikan setinggi-tingginya serta menjadi bahan bakar semangat dalam menyelesaikan skripsi ini sebaik mungkin.
2. Bapak Muhammad Itqan Mazdadi, S.Kom., M.Kom. selaku dosen pembimbing utama serta Bapak Muliadi, S.Kom., M.Cs. selaku dosen pembimbing pendamping dan dosen pembimbing akademik yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
3. Ibu Dwi Kartini, S.Kom., M.Kom. selaku Koordinator Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
4. Seluruh dosen dan staf Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
5. Teman-teman seperjuangan, keluarga besar NVIDIA'22, yang telah menjadi bagian atas perjuangan kuliah selama ini.

6. 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 dan penulis mengharapkan bantuan berupa kritik dan saran 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 dari Tuhan Yang Maha Esa, Allah SWT.

Banjarbaru, 25 Februari 2026



Muhammad Henry Aditya

DAFTAR ISI

DAFTAR TABEL

Tabel **Halaman**

DAFTAR GAMBAR

Gambar **Halaman**

DAFTAR LAMPIRAN

Lampiran