

ABSTRAK

ANALISIS PERBANDINGAN KINERJA TEKNIK IMPUTASI UNTUK PENANGANAN *MISSING VALUE* (STUDI KASUS DATASET *DIABETES MELLITUS*)

Oleh

Vincentius Sagi Alban Anindyajati

21110037

Diabetes mellitus merupakan penyakit kronis yang prevalensinya terus meningkat di abad ke-21. Deteksi dini penyakit ini sangat penting, namun sering terkendala oleh masalah *missing value* dalam dataset medis. Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja berbagai teknik imputasi dalam menangani *missing value* pada dataset *diabetes mellitus*. Studi ini menggunakan dataset 'Pima Indians Diabetes Database' dari Kaggle, dengan fokus pada perbandingan teknik imputasi MICE dan *MissForest*, dengan KNN *Imputer* sebagai *benchmark*. Metode penelitian meliputi tahapan pengolahan data, pengembangan model menggunakan *Random Forest*, di mana optimasi hiperparameter dilakukan secara khusus pada data hasil imputasi KNN menggunakan Random Search lalu parameter terbaik tersebut kemudian diterapkan pada model yang dilatih dengan data hasil imputasi MICE dan *MissForest*. Evaluasi performa model menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa MICE dan *MissForest* secara signifikan mengungguli KNN *Imputer*. *MissForest* teridentifikasi sebagai teknik imputasi terbaik, yang memberikan peningkatan *F1-Score* (Kelas 1) sebesar +0.005829 dibandingkan KNN. Lebih penting lagi, *MissForest* menghasilkan nilai *F1-Score* tertinggi (0.807273), *Recall* tertinggi (0.828358), dan *Precision* tertinggi (0.787234) untuk kelas 'Diabetes', menunjukkan keseimbangan terbaik antara *precision* dan *recall* yang krusial untuk aplikasi medis. Temuan ini memberikan rekomendasi teknik imputasi berbasis bukti untuk meningkatkan akurasi prediksi *diabetes mellitus*, sehingga berkontribusi pada pengembangan sistem analisis data medis yang lebih andal.

Kata Kunci: *Diabetes Mellitus*, Imputasi, MICE, *MissForest*, *Random Forest*