UNIVERSITAS
GADJAH MADA

Implementasi Balanced Random Forest Classifier Dalam Menangani Data Tidak Seimbang
Naim Nur Yasin Farhan, Danang Teguh Qoyyimi, S.Si., M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

INTISARI

IMPLEMENTASI BALANCED RANDOM FOREST CLASSIFIER DALAM MENANGANI DATA TIDAK SEIMBANG

Naim Nur Yasin Farhan
19/442601/PA/19350

Diabetes merupakan penyakit kronis yang muncul dikarenakan kebiasaan makan yang kurang baik, kurangnya aktifitas fisik, pengaruh lingkungan, kesehatan mental, atau faktor genetik. Dari faktor-faktor tersebut, dapat dilihat bahwa diabetes dapat dideteksi lebih awal agar bisa dilakukan pencegahan. Dalam memprediksi penderita diabetes, data yang umum digunakan cenderung tidak seimbang antara jumlah penderita diabetes dan yang bukan penderita diabetes. Beberapa penelitian sebelumnya sudah menunjukkan beberapa pendekatan metode pada tingkat data, algoritma, maupun gabungan keduanya. Penelitian ini memperlihatkan penerapan *machine learning* menggunakan metode *Balanced Random Forest Classifier* untuk memprediksi penderita penyakit diabetes menggunakan data *Behavioral Risk Factor Surveillance System* (BRFSS). Metode ini dipilih karena performanya yang baik dalam menangani ketidakseimbangan data yang ekstrim dengan jumlah data yang besar. Hasil yang didapatkan pada penelitian ini adalah metode *Balanced Random Forest Classifier* cukup baik mendeteksi penderita diabetes dibuktikan dengan nilai sensitifitas yang lebih baik dibandingkan metode seperti *Random Forest Classifier* dan *Random Forest Classifier* dengan modifikasi bobot kelas.

Kata Kunci: Diabetes, Ketidakseimbangan Data, Analisis Klasifikasi, *Machine Learning*, *Random Forest*, *Balanced Random Forest Classifier*

UNIVERSITAS
GADJAH MADA

Implementasi Balanced Random Forest Classifier Dalam Menangani Data Tidak Seimbang
Naim Nur Yasin Farhan, Danang Teguh Qoyyimi, S.Si., M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

BALANCED RANDOM FOREST CLASSIFIER IMPLEMENTATION ON IMBALANCED DATA CLASSIFICATION

Naim Nur Yasin Farhan
19/442601/PA/19350

Diabetes is a chronic disease that arises due to poor eating habits, lack of physical activity, environmental influences, mental health, or genetic factors. From these factors, it can be seen that diabetes can be detected early so that prevention can be carried out. When predicting diabetes sufferers, the data commonly used tends to be imbalanced between the number of diabetics and non-diabetics. Several previous studies have shown several methodological approaches at the data level, algorithm level, or a combination of both. This study shows the application of machine learning using the Balanced Random Forest Classifier method to predict diabetes patients using Behavioral Risk Factor Surveillance System (BRFSS) data. This method was chosen because of its good performance in handling extreme data imbalance with large amounts of data. The results obtained in this study indicate that the Balanced Random Forest Classifier method is quite effective in detecting diabetes patients, as shown by its higher sensitivity compared to methods such as the Random Forest Classifier and the Random Forest Classifier with class weight modification.

Keywords: Diabetes, Data Imbalance, Classification Analysis, Machine Learning, Random Forest, Balanced Random Forest Classifier