

# DIABETES PATIENT CLASSIFICATION USING K-NEAREST NEIGHBOR ALGORITHM WITH BALANCING

Banu Nai Ngalis, Hari

Informatika

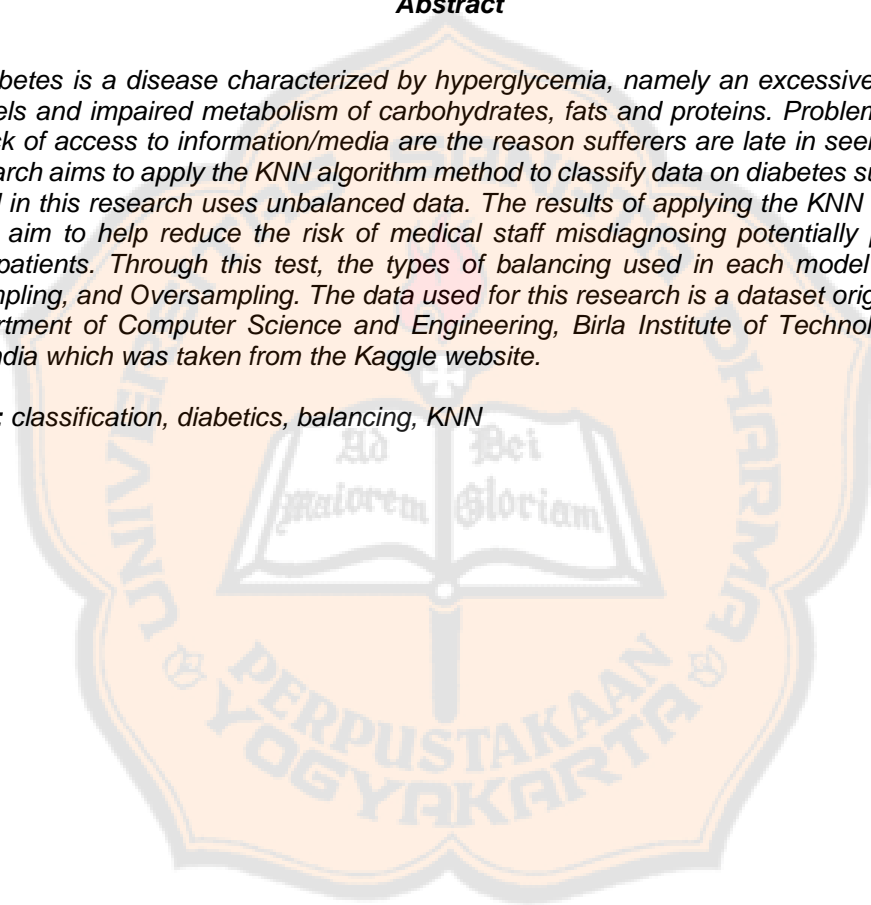
Universitas Sanata Dharma Yogyakarta

e-mail: banunaingalis77@gmail.com

## Abstract

*Diabetes is a disease characterized by hyperglycemia, namely an excessive increase in sugar levels and impaired metabolism of carbohydrates, fats and proteins. Problems that arise due to lack of access to information/media are the reason sufferers are late in seeing a doctor. This research aims to apply the KNN algorithm method to classify data on diabetes sufferers. The data used in this research uses unbalanced data. The results of applying the KNN method with balancing aim to help reduce the risk of medical staff misdiagnosing potentially positive and negative patients. Through this test, the types of balancing used in each model are Smote, Undersampling, and Oversampling. The data used for this research is a dataset originating from the Department of Computer Science and Engineering, Birla Institute of Technology, Mesra, Ranchi, India which was taken from the Kaggle website.*

**Keyword:** classification, diabetics, balancing, KNN



# KLASIFIKASI PASIEN DIABETES MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR DENGAN BALANCING

Banu Nai Ngalis, Hari

Informatika

Universitas Sanata Dharma Yogyakarta

e-mail: banunaingalis77@gmail.com

## Abstrak

*Diabetes adalah salah satu penyakit yang ditandai dengan terjadinya hiperglikemia yaitu meningkatnya kadar gula secara berlebihan dan gangguan metabolisme karbohidrat, lemak, dan protein. Permasalahan yang timbul karena kurangnya akses informasi/media menjadi alasan penderita terlambat untuk memeriksakan diri ke dokter. Penelitian ini bertujuan menerapkan metode algoritma KNN untuk mengklasifikasikan data penderita diabetes. Data yang digunakan pada penelitian ini menggunakan data yang tidak seimbang. Hasil dari penerapan metode KNN dengan balancing bertujuan untuk membantu mengurangi resiko dari kesalahan diagnosa petugas medis kepada pasien yang berpotensi positif dan negatif. Melalui pengujian ini bentuk macam balancing yang digunakan pada model masing-masing adalah Smote, Undersampling, dan Oversampling. Data yang digunakan untuk penelitian ini adalah dataset yang berasal dari bagian Department of Computer Science and Engineering, Birla Institute of Technology, Mesra, Ranchi, India yang diambil dari website Kaggle.*

**Kata kunci:** *klasifikasi, penderita diabetes, balancing, KNN.*