

# Diagnosis diabetes tipe 2 dengan machine learning melalui analisis spektroskopi ATR-FTIR pada sampel saliva = Diagnosis of type 2 diabetes with machine learning via ATR-FTIR spectroscopic analysis of saliva samples

Juan Wiratama, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920548865&lokasi=lokal>

---

## Abstrak

Dataset yang digunakan pada penelitian ini didapat dari paper yang berjudul “Attenuated total reflection FTIR dataset for identification of type 2 diabetes using saliva” yang ditulis oleh Sanchez-Brito et al. pada tahun 2022. Dataset tersebut berhubungan dengan spektrum ATR-FTIR dari 1040 saliva pasien. Dataset ini kemudian digunakan pada penelitian ini untuk melatih suatu model Machine Learning menggunakan algoritma SVM dan XGBoost. Sebelum dijadikan dataset acuan untuk keperluan pelatihan model, data terlebih dahulu melalui proses pre-processing yang meliputi proses pemotongan data agar terfokus pada region Biological Fingerprint, normalisasi protein amida I, dan penurunan orde satu. Untuk keperluan cross validation, dataset terlebih dahulu dipisah menjadi data train dan data test, kemudian data train akan kembali dipisah menjadi subset train untuk tiap fold dan subset validation yang dilatih sambil melewati stratified cross validation sebanyak 10 fold. Performa model akan didapat dari hasil prediksi model terhadap subset validation yang dihasilkan di semua 10 fold, serta hasil prediksi model terhadap data test yang menunjukkan performa keseluruhan model. Didapat bahwa performa model XGBoost melampaui performa model SVM dengan nilai accuracy sebesar 91,8%; sensitivity sebesar 93,6%; dan specificity sebesar 89,9%. Performa ini berhasil mendekati performa metode diagnosis diabetes tipe 2 yang masih bersifat invasif, yaitu tes HbA1c.

.....

The dataset used in this study was obtained from the paper titled “Attenuated Total Reflection FTIR Dataset for Identification of Type 2 Diabetes Using Saliva” written by Sanchez-Brito et al. in 2022. This dataset pertains to the ATR-FTIR spectrum of saliva from 1040 patients. It was used in this research to train a machine learning model using the SVM and XGBoost algorithms. Before being used as a reference dataset for model training, the data underwent preprocessing, which included data trimming to focus on the Biological Fingerprint region, protein amide I normalization, and first-order derivative processing. For cross-validation purposes, the dataset was first split into training and testing data. The training data was further divided into train and validation subsets for each fold and trained using 10-fold stratified cross-validation. The model's performance was evaluated based on predictions on the validation subsets from all 10 folds, as well as predictions on the test data, reflecting the overall model performance. It was found that the XGBoost model outperformed the SVM model with an accuracy of 91.8%, sensitivity of 93.6%, and specificity of 89.9%. This performance approaches that of the invasive HbA1c test used for diagnosing type 2 diabetes.