

ABSTRAK

Ardis Hibatul Hakim, 10120176

PREDIKSI DIABETES MENGGUNAKAN ALGORITMA XGBOOST: STUDI KASUS PADA DATASET PASIEN DIABETES

Penulisan Ilmiah. Jurusan Sistem Informasi. Fakultas Ilmu Komputer dan
Teknologi Informasi. Universitas Gunadarma. 2023

Kata kunci : XGBoost, *machine learning*, diabetes. SMOTETomek, EDA.

(xii + 56 halaman + lampiran)

Diabetes mellitus, atau lebih dikenal sebagai diabetes, adalah penyakit kronis yang ditandai oleh peningkatan kadar gula darah dalam tubuh. Hal ini terjadi ketika tubuh tidak dapat memproduksi atau menggunakan insulin secara efektif. Salah satu cara untuk memprediksi diabetes dengan menggunakan *machine learning*. XGBoost merupakan salah satu model klasifikasi dalam *machine learning* yang sering digunakan dalam melakukan prediksi. Pada penulisan ilmiah ini, algoritma XGBoost dirancang menggunakan Python untuk memprediksi seorang terjangkit diabetes. Penelitian dilakukan menggunakan dataset dari *National Institute Of Diabetes And Digestive And Kidney Diseases* yang terdiri dari 100 ribu data pasien dengan 9 variabel. Exploratory data analysis dilakukan untuk memaksimalkan dataset yang dimiliki serta menggunakan bantuan statistik dan mempresentasikannya melalui teknik visual. Beberapa variabel dataset memuat missing values, handling duplicates, uniqueness. Serta juga dilakukan tahap pra-pemrosesan data dengan menggunakan teknik standard scalling, one hot encoding dan handling imbalance menggunakan SMOTETomek guna mencapai keseimbangan antara kelas-kelas dalam dataset. Pembagian data training dan testing dibagi dengan rasio 75% untuk data *training* dan 25% untuk data *testing* dari total data yang siap dianalisis untuk masing-masing kelas. Hasil dari analisis yang dilakukan, didapatkan bahwa metode klasifikasi XGBoost mampu mendapatkan nilai akurasi sekitar 97%, presisi sekitar 99%, dan recall sekitar 95%.

Daftar Pustaka (2013– 2023)