

ABSTRAK

Laksono, Bagus Dwi. 2026. Klasifikasi Risiko Penyakit *Gestasional Diabetes Mellitus* pada Ibu Hamil Menggunakan Metode *Naive Bayes*. Tugas Akhir. Program Sarjana. Program Studi Teknik Informatika. Universitas Muhammadiyah Jember.

Pembimbing: Dudi Irawan, S.T., M.Kom.; Ari Eko Wardoyo, S.T., M.Kom.

Gestasional Diabetes Mellitus (GDM) merupakan kondisi gangguan metabolisme glukosa yang terjadi selama kehamilan dan diperkirakan berdampak pada sekitar 14% kehamilan di seluruh dunia. Deteksi dini DMG menjadi tantangan tersendiri karena proses diagnosis konvensional masih bergantung pada pemeriksaan laboratorium seperti Tes Toleransi Glukosa Oral (TTGO) yang memerlukan waktu, biaya, dan ketersediaan fasilitas kesehatan yang memadai. Penelitian ini bertujuan untuk membangun model klasifikasi risiko GDM pada ibu hamil menggunakan algoritma *Naive Bayes* yang dikombinasikan dengan teknik *Random Over Sampling* (ROS) untuk mengatasi permasalahan ketidakseimbangan kelas pada data. Metodologi penelitian menggunakan kerangka kerja CRISP-DM yang mencakup tahap *business understanding*, *data understanding*, *data preparation*, *modelling*, *evaluation*, dan *deployment*. *Dataset* yang digunakan adalah *Gestational Diabetes Mellitus* (GDM Data Set) yang diperoleh dari Kaggle dengan total 3.525 data dari populasi wilayah India, menggunakan 15 fitur klinis seperti usia, jumlah kehamilan, BMI, HDL, riwayat keluarga, OGTT, hemoglobin, dan tekanan darah. Proses *preprocessing* meliputi *data cleaning*, imputasi *median* untuk *missing value*, penanganan *outlier* menggunakan metode IQR, *encoding*, dan normalisasi menggunakan *StandardScaler*. Pembagian data dilakukan dengan rasio 80:20 dan ROS diterapkan hanya pada data latih untuk menghindari *data leakage*. Hasil evaluasi menggunakan *Confusion Matrix* pada 705 data uji menunjukkan bahwa model *Gaussian Naive Bayes* dengan ROS menghasilkan akurasi sebesar 94,47%, *precision* 92,42%, *recall* 93,43%, dan *F1-Score* 92,92%. Nilai *recall* yang tinggi menunjukkan kemampuan model dalam meminimalkan kesalahan deteksi pada ibu hamil yang berisiko GDM. Model kemudian diimplementasikan dalam bentuk aplikasi web berbasis *Streamlit* sebagai alat bantu skrining awal bagi tenaga kesehatan.

Kata Kunci: CRISP-DM, *Gestasional Diabetes Mellitus*, Klasifikasi, *Naive Bayes*, *Random Over Sampling*, *Streamlit*

ABSTRACT

Laksono, Bagus Dwi. 2026. Classification of Gestational Diabetes Mellitus Risk in Pregnant Women Using the Naive Bayes Method. Undergraduate Thesis. Bachelor's Program in Informatics Engineering. University of Muhammadiyah Jember.

Advisors: (1) Dudi Irawan, S.T., M.Kom.; (2) Ari Eko Wardoyo, S.T., M.Kom.

Gestational Diabetes Mellitus (GDM) is a condition of impaired glucose metabolism occurring during pregnancy, estimated to affect approximately 14% of pregnancies worldwide. Early detection of GDM remains challenging as conventional diagnosis still relies on laboratory examinations such as the Oral Glucose Tolerance Test (OGTT), which requires considerable time, cost, and adequate healthcare facilities. This study aims to develop a GDM risk classification model for pregnant women using the Naive Bayes algorithm combined with the Random Over Sampling (ROS) technique to address class imbalance in the dataset. The research methodology follows the CRISP-DM framework, covering the stages of business understanding, data understanding, data preparation, modelling, evaluation, and deployment. The dataset used is the Gestational Diabetes Mellitus (GDM Data Set) obtained from Kaggle, consisting of 3,525 records from an Indian population, with 15 clinical features including age, number of pregnancies, BMI, HDL, family history, OGTT results, hemoglobin levels, and blood pressure. Preprocessing steps included data cleaning, median imputation for missing values, outlier handling using the IQR method, encoding, and normalization using StandardScaler. Data was split at an 80:20 ratio, and ROS was applied exclusively to the training data to prevent data leakage. Evaluation results using the Confusion Matrix on 705 test samples showed that the Gaussian Naive Bayes model with ROS achieved an accuracy of 94.47%, precision of 92.42%, recall of 93.43%, and F1-Score of 92.92%. The high recall value demonstrates the model's capability in minimizing false negatives in detecting at-risk pregnant women. The model was subsequently deployed as a Streamlit-based web application to serve as an early screening tool for healthcare professionals.

Keywords: CRISP-DM, Gestasional Diabetes Mellitus, Classification, Naive Bayes, Random Over Sampling, Streamlit