PROPOSAL RISET INFROMATIKA

PENERAPAN METODE XGBOOST DENGAN OPTIMASI SMOTE UNTUK KLASIFIKASI TINGKAT OBESITAS BERDASARKAN FAKTOR GAYA HIDUP

Disusun oleh:

Chesa Saskia Rafika / 22081010211

LATAR BELAKANG

Obesitas merupakan masalah kesehatan global yang terus meningkat dan menjadi penyebab utama penyakit tidak menular seperti diabetes dan jantung. Gaya hidup seperti pola makan, aktivitas fisik, dan kebiasaan sehari-hari berperan besar dalam menentukan tingkat obesitas. Machine learning, khususnya algoritma XGBoost, dapat digunakan untuk memprediksi tingkat obesitas secara akurat berdasarkan faktor gaya hidup. Namun, ketidakseimbangan data antar kelas (imbalanced data) dapat menurunkan performa model, sehingga diterapkan metode SMOTE untuk menyeimbangkan data dan meningkatkan akurasi klasifikasi.

RUMUSAN MASALAH

- Bagaimana penerapan metode XGBoost untuk melakukan klasifikasi tingkat obesitas berdasarkan faktor gaya hidup?
- Bagaimana pengaruh penerapan SMOTE terhadap peningkatan performa model klasifikasi obesitas?
- Seberapa baik akurasi, presisi, dan recall yang dihasilkan oleh model XGBoost dengan optimasi SMOTE dibandingkan dengan tanpa SMOTE?

TUJUAN PENELITIHAN

- Menerapkan metode XGBoost untuk klasifikasi tingkat obesitas berdasarkan data gaya hidup individu.
- Mengoptimalkan hasil klasifikasi dengan menggunakan teknik SMOTE untuk menyeimbangkan distribusi data.
- Menganalisis performa model dengan membandingkan hasil sebelum dan sesudah penerapan SMOTE.

PENELITIAN TERDAHULU

- Penelitian "Enhancing Obesity Prediction through SMOTE-based Classification Models" membandingkan Decision Tree, Random Forest, AdaBoost, dan XGBoost. XGBoost + SMOTE menghasilkan akurasi tertinggi 98,6%, menunjukkan metode ini paling efektif untuk klasifikasi obesitas berbasis gaya hidup.
- Penelitian "Optimasi AdaBoost dan XGBoost untuk Klasifikasi Obesitas Menggunakan SMOTE" Hasilnya, XGBoost menunjukkan performa terbaik dengan akurasi meningkat dari 92,4% → 94,5% setelah penerapan SMOTE

RESERCH GAP DAN NOVELTY

Penelitian mengenai klasifikasi tingkat obesitas dengan metode machine learning telah banyak dilakukan, namun sebagian besar masih berfokus pada data fisik seperti berat badan, tinggi badan, dan umur. Padahal, tingkat obesitas juga sangat dipengaruhi oleh berbagai faktor gaya hidup seperti pola makan, aktivitas fisik, serta kebiasaan harian.

Kebaruan dari penelitian ini terletak pada penerapan metode XGBoost yang dioptimasi menggunakan SMOTE untuk mengklasifikasikan tingkat obesitas berdasarkan faktor gaya hidup secara menyeluruh.

SUMBER DATA

Data yang digunakan dalam penelitian ini berasal dari dataset publik berjudul "Obesity Levels Dataset", yang tersedia di platform Kaggle. Dataset tersebut terdiri atas 2.111 baris data dan 17 atribut yang menggambarkan berbagai faktor gaya hidup, pola makan, serta kebiasaan individu yang memengaruhi tingkat obesitas.

TAHAPAN PENELITIHAN

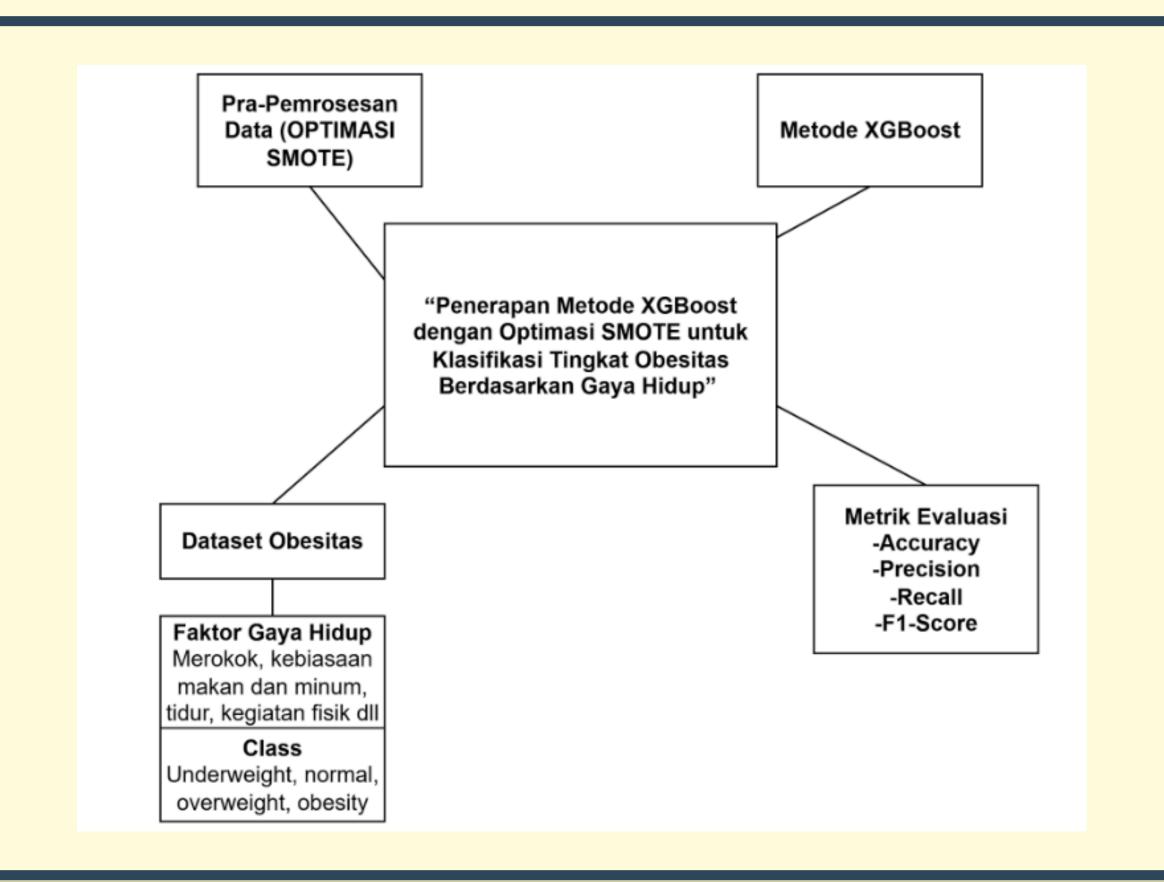
Proses penelitian disusun secara sistematis untuk memperoleh model klasifikasi obesitas yang optimal:

- 1. Pengumpulan Data
- 2. Pra-pemrosesan Data
- 3. Penyeimbangan Data (SMOTE)
- 4. Pembagian Data (Train-Test Split)
- 5. Penerapan Metode XGBoost

EVALUASI HASIL

Evaluasi model dilakukan dengan Confusion Matrix dan empat metrik utama yaitu Accuracy, Precision, Recall, F1-Score.

MIND MAP



THAILS KOU