




# EVALUASI KINERJA ALGORITMA LIGHTGBM DENGAN RANDOM OVERSAMPLING DALAM KLASIFIKASI TINGKAT OBESITAS



Chesa Saskia Rafika  
(22081010211)



# LATAR BELAKANG

Obesitas merupakan salah satu masalah kesehatan global dengan prevalensi yang terus meningkat dan berisiko menyebabkan berbagai penyakit tidak menular, seperti diabetes melitus dan hipertensi. Oleh karena itu, klasifikasi tingkat obesitas menjadi penting untuk mendukung pemetaan risiko serta perencanaan intervensi kesehatan yang lebih tepat sasaran. Namun, pada praktiknya, data kesehatan sering mengalami ketidakseimbangan kelas (class imbalance), di mana jumlah data antar kategori tidak terdistribusi secara merata. Kondisi ini berpotensi menyebabkan model machine learning cenderung bias terhadap kelas mayoritas, sehingga menurunkan kemampuan model dalam mengenali kelas minoritas secara akurat.



# SUMBER DATASET

- Dataset: Dhaka\_Obesity
- Jumlah data awal: 2.182
- Data akhir setelah pra-pemrosesan: 2.157
- Jumlah fitur: 17
- Jumlah kelas obesitas: 6  
Insufficient Weight  
Normal Weight  
Overweight  
Obesity Type I  
Obesity Type II  
Obesity Type III

Tabel 2. Distribusi Kelas Tingkat Obesitas (Sebelum Pra-Pemrosesan)

No	Kelas Tingkat Obesitas	Jumlah Data
1	Normal Weight	530
2	Overweight	423
3	Insufficient Weight	378
4	Obesity Type I	317
5	Obesity Type II	283
6	Obesity Type III	251
	<b>Total</b>	<b>2.182</b>



# METODOLOGI



## Pra-Pemrosesan Data

- Pembersihan dan penyeragaman nama kolom
- Normalisasi nilai fitur kategorikal
- Penghapusan 25 data duplikat
- Dataset akhir tidak memiliki missing value

## Encoding & Pembagian Data

- Variabel target dikonversi menjadi label numerik (0–5)
- Encoding fitur:
  - Label Encoding (fitur biner)
  - Ordinal Encoding (fitur berurutan)
  - One-Hot Encoding (fitur nominal)
- Pembagian data:
  - 80% data latih (1.725 data)
  - 20% data uji (432 data)
  - Menggunakan stratified split





# METODOLOGI



## Penanganan Class Imbalance

- Menggunakan Random Oversampling (ROS)
- ROS menduplikasi sampel kelas minoritas
- Diterapkan hanya pada data latih
- Diintegrasikan dalam pipeline untuk mencegah data leakage

## Model & Optimasi

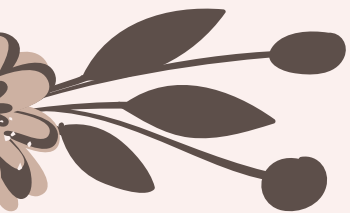
- Algoritma: Light Gradient Boosting Machine (LightGBM)
- Alasan:
  - Cocok untuk data tabular
  - Mampu menangkap hubungan non-linear
  - Efisiensi komputasi tinggi
- Optimasi hyperparameter:
  - GridSearchCV
  - Validasi silang 5-fold
  - Metrik evaluasi: Balanced Accuracy



# HASIL AKHIR MODEL LIGHTGBM

Tabel 6. Perbandingan Kinerja Model Baseline vs ROS

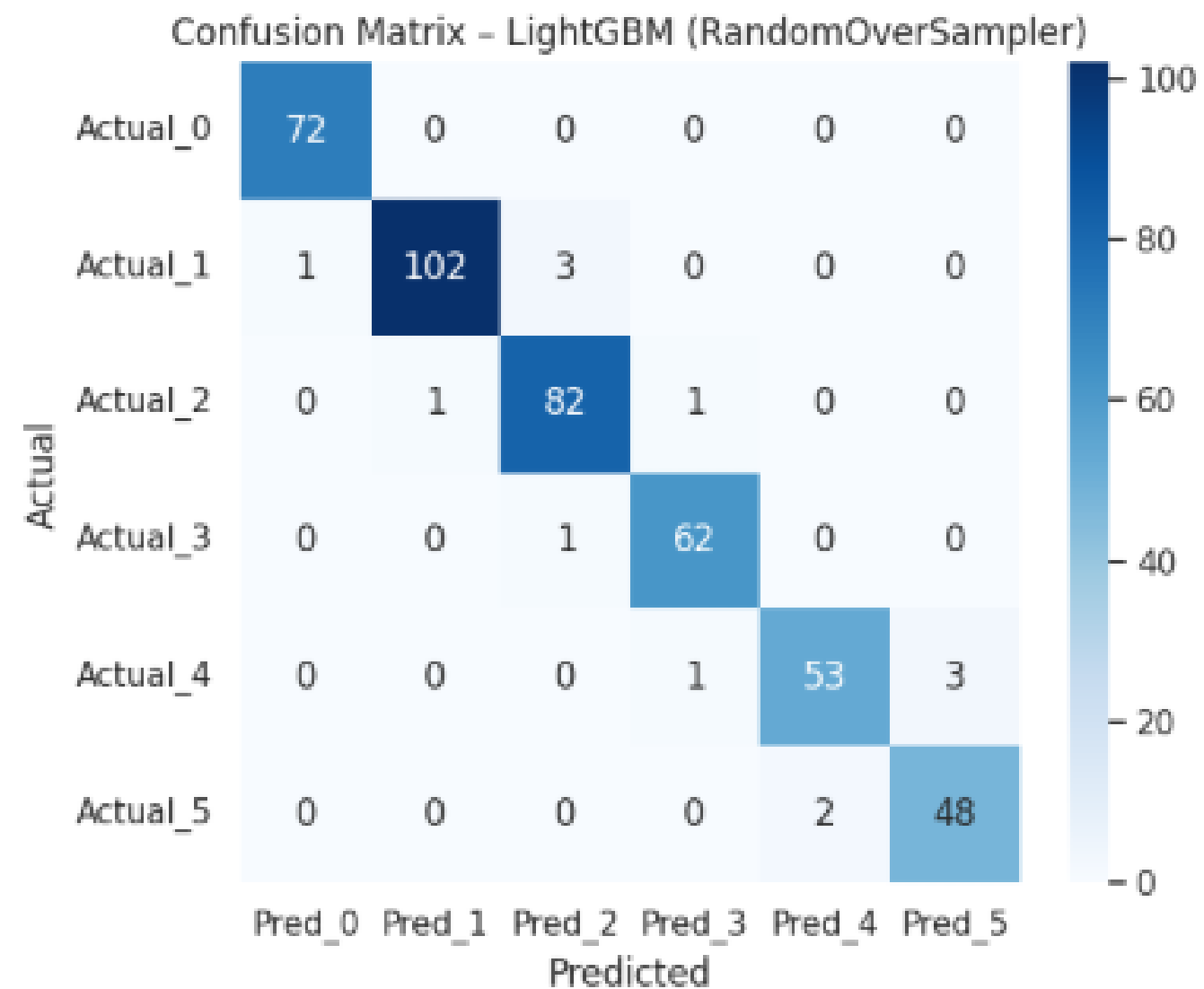
Skenario Model	Accuracy	F1-score Macro	Balanced Accuracy	Waktu Training (detik)
LightGBM Baseline	0.9583	0.9609	0.9638	67.40
LightGBM+Random OverSampler	0.9699	0.9679	0.9687	245.07



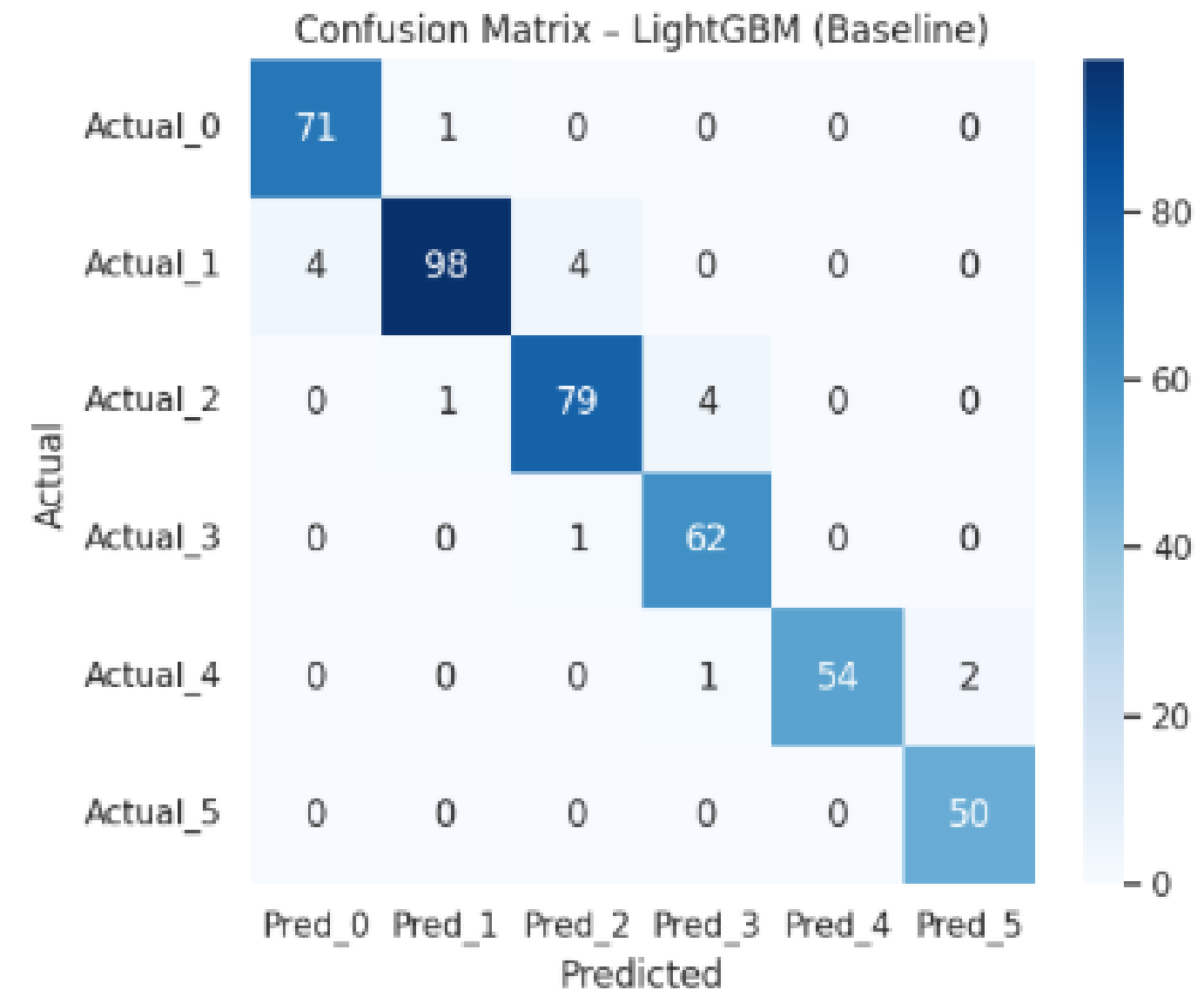
# CONFUSION MATRIX



Confusion Matrix - LightGBM (RandomOverSampler)



Confusion Matrix - LightGBM (Baseline)



# PEMBAHASAN

- Algoritma LightGBM menunjukkan kinerja sangat baik dalam klasifikasi tingkat obesitas berbasis data tabular dengan fitur numerik dan kategorikal.
- Pada skenario baseline, LightGBM telah mencapai balanced accuracy tinggi, menandakan kemampuannya dalam memodelkan hubungan non-linear antar faktor gaya hidup dan kondisi fisik.
- Penerapan Random OverSampler (ROS) pada data latih terbukti meningkatkan F1-score macro dan balanced accuracy, sehingga performa model menjadi lebih seimbang antar kelas.
- Analisis confusion matrix menunjukkan penurunan kesalahan klasifikasi pada kelas minoritas, seperti Insufficient Weight dan Obesity Type III.



# KESIMPULAN

Berdasarkan hasil penelitian, algoritma Light Gradient Boosting Machine (LightGBM) terbukti mampu mengklasifikasikan tingkat obesitas secara efektif pada data tabular dengan hubungan fitur yang bersifat non-linear. Ketidakseimbangan kelas terbukti memengaruhi keadilan performa model, khususnya pada kelas minoritas. Penerapan Random OverSampler (ROS) pada data latih mampu meningkatkan F1-score macro dan balanced accuracy, sehingga performa model menjadi lebih merata antar kelas. Meskipun waktu pelatihan meningkat, peningkatan kualitas dan keadilan prediksi menunjukkan bahwa kombinasi LightGBM dan ROS merupakan pendekatan yang efektif dan layak digunakan untuk klasifikasi tingkat obesitas pada dataset yang tidak seimbang.



✦ TERIMA KASIH ✦

