Document Validation Report

for Benchmarking

Correspondence: Yes
Percentages: 60.0%

Conclusion:

The repository contains Python code that (1) builds synthetic 2-D regression benchmarks and (2) trains/ evaluates models on a real Moscow apartment-price data set using R^2 and MAPE, which matches the documentation's description of synthetic data and the 26-feature housing data set. The code also includes a GA+VAE pipeline that can be interpreted as the proposed benchmark-generation method. However, several documented details are missing or diverge: the healthcare (diabetes) classification data set is not handled anywhere; the train/test split is 67/33 instead of the stated 80/20; only regression objectives are implemented (no classification); and the documentation's outlier policy, feature counts, and exact evaluation protocol are only partially reproduced. Therefore correspondence is partial, not full.