Project 1 – Values Audit Essay

Introduction

This project examines the values embedded in a machine learning study on predicting obesity levels from demographic and lifestyle factors. The paper reported using a dataset of 2009 records and 16 features, while the dataset provided to me contained 2111 records and 17 features. Because of this inconsistency, an exact replication of the paper's outcomes was not possible. Instead, my notebook offers a best-effort attempt at recreating the strategies the authors described. The focus of this task is not precise accuracy but understanding how values and assumptions shape technical choices. In conducting a values audit, I show how fairness, interpretability, efficiency, and authenticity are reflected in the study's choices, and how different values can lead to different outcomes.

Decision 1: Handling Class Imbalance with SMOTE-NC

One of the first significant choices in the study was how to deal with class imbalance. The dataset contains several obesity categories, some with far more records than others. The authors used SMOTE-NC (Synthetic Minority Oversampling Technique for categorical and numeric features) to create synthetic samples for minority classes. This reflects the value of **fairness across categories**. By giving more equal weight to each obesity class in the training set, the authors aimed to ensure that the model would not neglect smaller groups such as Insufficient Weight or Obesity Type III.

The value of fairness shaped the research by allowing models to learn minority classes more effectively. Without balancing, the model might predict majority classes well but fail on smaller ones. At the same time, SMOTE creates artificial records that never truly existed in the dataset. This decision accepts that fabricating records is acceptable if it results in fairer predictions. The outcome is that performance appears stronger across categories, but results depend partly on artificial rather than fully real participants. The trade-off here was between authenticity and fairness, and the authors clearly prioritized fairness.

Decision 2: Selection of Models

The authors compared three algorithms: Logistic Regression, Random Forest, and XGBoost. Each model embodies different values. Logistic Regression is highly interpretable and mathematically straightforward. Random Forest and XGBoost are ensemble methods that are typically more powerful but less transparent. The fact that the paper evaluated all three suggests an appreciation for balancing **interpretability and performance**.

The paper concluded that Logistic Regression performed best after feature selection. This outcome highlights the authors' prioritization of transparency. Logistic Regression is easier to interpret when one wants to understand how predictors such as age, diet, and activity levels affect obesity risk. Random Forest and XGBoost may have offered stronger performance on some metrics, but they are less interpretable in a healthcare or public health setting. The authors' decision shows that in sensitive applications, explainability was considered more important than a marginal increase in accuracy.

In my replication, however, the results were different. Random Forest and XGBoost both outperformed Logistic Regression. This difference is likely due to the dataset mismatch, with 2111 rows and 17 features instead of 2009 and 16, and because I did not apply hyperparameter tuning as the authors did. More importantly, this divergence illustrates how outcomes themselves are shaped by values: my replication favored performance with tree-based models, while the original study elevated interpretability by emphasizing Logistic Regression. This shows that values guide not only which models are chosen but also which results are presented as the best.

Decision 3: Feature Selection using Recursive Feature Elimination

Another important choice was to reduce the feature set using Recursive Feature Elimination (RFE). This reflects the value of **parsimony and efficiency**. By narrowing the dataset to the most important predictors, the authors created a streamlined model that is easier to train, explain, and apply.

This value plays out in two ways. First, it reduces the risk of overfitting by removing noisy or redundant features. Second, it aligns with interpretability goals. Physicians and policymakers prefer a concise list of predictors over a long list of weak indicators. The trade-off is that parsimony can exclude subtle but real effects. For example, a less obvious lifestyle variable might still matter for obesity but could be eliminated. The choice shows a preference for **efficiency and clarity** over completeness.

Alternative Value: Authenticity over Fairness

In my implementation, I challenged the value of fairness underlying the SMOTE decision. Instead, I adopted the value of **authenticity**, avoiding the creation of artificial individuals. To do this, I trained models directly on the original imbalanced data. Logistic Regression was modified with class weights so minority classes received more attention without fabricating records. I also trained a Random Forest as a baseline without balancing.

The results differed from the SMOTE version. Accuracy dropped slightly. Random Forest achieved about 92 percent without SMOTE compared to 95 percent with it. Some Minor classes became harder to predict correctly. The benefit, however, was that every prediction was grounded in real participants from the dataset. This illustrates the trade-off: SMOTE improved fairness but introduced synthetic data, while the no-SMOTE approach preserved authenticity but reduced fairness. Both values are legitimate, and my notebook demonstrates how technical solutions embody these priorities.

Conclusion

This values audit illustrates that technical decisions in data science are always value laden. The original research reflected fairness, interpretability, and parsimony through its use of SMOTE-NC, Logistic Regression, and RFE. My alternative implementation prioritized authenticity by avoiding synthetic data, which significantly changed the outcomes. Although my replication did not exactly match the paper's results due to dataset differences, this highlights that values, assumptions, and trade-offs are embedded directly in data science systems. What counts as a good model is not only about numbers but also about which values we choose to prioritize.