CHAPTER 1
INTRODUCTION

This chapter presents the fundamental background of the research on obesity level classification. It covers the research background, research questions, research objectives, research significance, and research limitations, which serve as the foundation for conducting this study.

1.1. Research Background

Obesity has become an increasingly significant public health concern due to its association with a higher risk of non communicable diseases, including cardiovascular disease, diabetes, and cancer. It also contributes substantially to increased mortality rates and escalating healthcare expenditures worldwide. However, obesity should not be viewed merely as a matter of excessive body weight. Rather, it is a multifactorial condition influenced by the complex interaction of biological, behavioral, sociocultural, environmental, economic, and technological factors. Since obesity is determined by numerous interrelated factors, approaches relying solely on a single indicator or individual behavioral factor are insufficient to explain variations in nutritional status across a population [1].

The urgency of obesity prevention and control in Indonesia is particularly evident in urban areas. A study based on the Indonesian Basic Health Research (Riskesdas) conducted in 2007 and 2018 reported that the prevalence of obesity among the urban population increased substantially from 23.0% in 2007 to 50.1% in 2018. This increase was associated with changes in dietary patterns and lifestyle, particularly the growing consumption of high calorie foods and declining levels of physical activity. These findings highlight the need for more targeted obesity prevention strategies among urban communities [2]. The rapid increase in obesity prevalence emphasizes the importance of early identification of individuals at risk and more comprehensive nutritional status assessment. In this context, data driven predictive approaches have become increasingly relevant because they are capable of processing multiple variables simultaneously. Conventional analytical methods

often have limitations in capturing the complex and interactive relationships among obesity related factors. Therefore, machine learning has emerged as a promising approach for providing more accurate predictions in obesity risk assessment [3].

Obesity level classification based only on Body Mass Index (BMI) has several limitations because BMI does not consider other factors, such as dietary habits and physical conditions, that may influence obesity levels. Therefore, a machine learning approach using lifestyle and physical condition variables is needed to provide a more comprehensive estimation [4]. In predictive modeling, tree boosting algorithms are widely used because they can build strong predictive models. XGBoost and LightGBM are decision tree based ensemble methods that are suitable for handling large scale data. Previous studies have shown that XGBoost generally performs better in accuracy and sensitivity, LightGBM performs better in specificity, indicating differences in performance across evaluation metrics. Therefore, comparing XGBoost and LightGBM is relevant to identify which model more stable performance across all obesity classes [5].

Multiclass obesity classification frequently encounters the problem of class imbalance, where the distribution of classes is uneven. In such situations, the majority class tends to dominate the training process, causing the model to achieve high overall accuracy while exhibiting biased predictions toward the majority class. Under imbalanced data distributions, XGBoost and LightGBM have been shown to produce different levels of performance across evaluation metrics such as accuracy, sensitivity, and specificity. Consequently, comparing both algorithms under identical experimental conditions is essential for assessing the consistency of their predictive performance across all classes [5]. In obesity classification studies, addressing class imbalance is commonly considered an integral part of the experimental design. Previous research has demonstrated that the application of data balancing techniques, such as the Synthetic Minority Over sampling Technique (SMOTE), together with hyperparameter optimization, can significantly improve model performance, with XGBoost achieving an accuracy of 0.945 accompanied by high precision, recall, and F1 score values. These findings indicate that the model training strategy, including the application of data balancing techniques and hyperparameter tuning, can substantially influence classification performance.

Therefore, such approaches should be systematically evaluated through controlled experiments [6]. Nevertheless, resampling strategies are not always the most appropriate solution for class imbalance problems. In medical datasets, synthetic oversampling techniques such as SMOTE and Adaptive Synthetic Sampling (ADASYN) may alter the underlying structure of the training data by generating artificial samples, potentially affecting the predictive characteristics of the resulting classification model [7].

The effectiveness of resampling techniques depends on the classification algorithm, the evaluation metrics used, and the degree of class imbalance in the dataset. Resampling does not always improve model performance, and its impact may differ depending on the selected method and imbalance level. Therefore, its benefits and limitations need to be evaluated according to the dataset and research objective [8]. Random Oversampling (ROS) is one of the simplest oversampling techniques, which works by randomly duplicating minority class samples in the training data to balance class distribution before model construction. Although ROS is simple and easy to implement, its effectiveness should still be empirically validated through a fair comparison between models trained without data balancing and models trained using ROS [9].

Evaluation metrics play a crucial role in imbalanced multiclass classification problems. Previous studies have highlighted that overall accuracy can be dominated by the majority class and therefore may not accurately reflect the model's overall classification performance. In contrast, macro averaged metrics, such as the macro F1 score (F1 macro), assign equal importance to every class, making them more appropriate for evaluating performance across all classes. Consequently, a model's capability to distinguish between closely related obesity categories should be demonstrated through confusion matrix analysis and class wise evaluation metrics rather than relying solely on overall accuracy [10]. Based on this background, this study focuses on obesity classification using lifestyle and physical condition factors from the Dhaka Obesity Dataset. It compares the performance of XGBoost and LightGBM with Random Oversampling applied to training data. The analysis evaluates model robustness through different train test split ratios and comparable hyperparameter configurations, while identifying the optimal model using an

evaluation framework that emphasizes balanced performance across all classes.

1.2. Research Questions

Based on the research background described above, the research questions addressed in this study are as follows:

1. How are the XGBoost and LightGBM models implemented with the application of Random Oversampling (ROS) on the training data for obesity level classification?
2. How do different train test split ratios and hyperparameter configurations affect the performance of the XGBoost + ROS and LightGBM + ROS models?
3. How does the performance of models without ROS compare with those incorporating ROS, and which model achieves the most optimal performance for obesity level classification?

1.3. Research Objectives

Based on the research questions presented above, the objectives of this study are as follows:

1. To implement XGBoost and LightGBM models with the application of Random Oversampling (ROS) on the training data for obesity level classification using the Dhaka Obesity Dataset.
2. To analyze the effects of different train test split ratios and sequential manual hyperparameter tuning on the performance of the XGBoost + ROS and LightGBM + ROS models for obesity level classification.
3. To analyze the impact of applying Random Oversampling (ROS) compared with a baseline scenario without ROS on the performance of XGBoost and LightGBM, and to determine the model that achieves the most optimal performance based on accuracy, macro precision, macro recall, macro F1 score, and balanced accuracy.

1.4. Research Significance

This study is expected to contribute to the advancement of machine learning applications in the healthcare domain, particularly in obesity level classification using tabular data. From an academic perspective, this research contributes to the

existing body of knowledge by providing a comparative analysis of the XGBoost and LightGBM algorithms with the application of Random Oversampling (ROS) on the training data to address class imbalance. Furthermore, this study provides insights into the use of evaluation metrics, including accuracy, macro precision, macro recall, macro F1 score, balanced accuracy, confusion matrix, and train test gap, to assess classification performance more comprehensively. From a practical perspective, the findings of this study may serve as a reference for selecting the most appropriate model for obesity level classification and support the development of more reliable and systematically evaluated data driven prediction systems.

1.5. Scope and Limitations of the Study

To ensure that this study remains focused within the intended scope, the following limitations are defined:

1. This study uses the secondary Dhaka Obesity Dataset from Mendeley Data, consisting of 2,182 respondents and 17 attributes related to lifestyle, physical condition, and obesity level.
2. The classification target uses the Obesity_Level variable, which is adjusted based on the Asia Pacific BMI classification adopted by the Indonesian Ministry of Health into five classes: Underweight, Normal, Overweight, Obesitas_I, and Obesitas_II.
3. The dataset is divided using stratified train test split with ratios of 60:40, 70:30, and 80:20. Hyperparameter testing is conducted manually on learning_rate, max_depth, n_estimators, subsample, and colsample_bytree.
4. The compared models are limited to XGBoost and LightGBM, both without ROS and with Random Oversampling applied to the training data.
5. Model evaluation uses accuracy, macro precision, macro recall, macro F1 score, balanced accuracy, confusion matrix, classification report, and train test gap.
6. This study does not discuss feature interpretability in depth, such as feature importance or SHAP. The system implementation is used only as a simple demonstration of the best model.