

CHAPTER 1

INTRODUCTION

1.1 Background

Malnutrition in Indonesia has shown a decreasing trend since 2013. Even so, this problem persists until now, noting that about 15.9% of toddlers under five are underweight, and 21.5% suffer from stunting (Indonesia's Ministry of Health, 2023). Child malnutrition significantly affects survival rates from acute and chronic diseases and has a deleterious effect on individual and communal productivity. Malnutrition encompasses undernutrition and overnutrition. The World Health Organization (WHO) explain that undernutrition leads to a higher likelihood of childhood stunting, wasting, and being underweight, while overnutrition increases the potential for overweight or obesity (WHO, 2006).

One of the primary factors contributing to this high number of malnutrition rate is the low awareness of malnutrition itself. In the 2023 Indonesian health survey report (Indonesia's Ministry of Health, 2023), it is explained that nationally, only 69.4% of the population is aware of the definition of stunting, while Java Island, the most populous region, is only 67.83%. These percentages are relatively low, especially considering that many government initiatives to combat stunting focus on increasing public awareness. This lack of awareness could hinder human development and perpetuate socioeconomic disparities across the region in Java Island.

Recognizing the strong link between malnutrition and socioeconomic status underscores the importance of understanding local conditions. However, most regions in the *status quo* mostly apply uniform solutions across diverse areas. Most programs, particularly those related to stunting prevention, focus heavily on

general awareness campaigns or socialization efforts without considering the distinct nutritional profiles of each city. This one-size-fits-all approach often fails to address the root causes of malnutrition that vary significantly across urban and rural settings, income levels, and access to health infrastructure. Without localized data and personalized strategies, these interventions risk being inefficient and unsustainable (Sentika et al., 2024).

In understanding the national nutritional state, one significant region to be researched is the cities and regencies on Java Island. Based on the BPS-Statistics Indonesia's 2024 report, Java Island has 156 million people, covering over half of the national population. As the most developed island with the biggest economy, Java Island is notoriously known as a target for pursuing a career and a better living (Putri & Judijanto, 2023). Therefore, Java has been the nation's most diversified region, roughly representing Indonesia's entire population (Mardiansjah et al., 2023). Since Java is divided into cities and regencies, clustering is one of the best ways to analyze its nutritional status (Mardiansjah et al., 2023).

Clustering plays a significant role in nutritional studies by grouping regions with similar characteristics, in this case, nutritional status (Nagari & Inayati, 2020). This method allows researchers and policymakers to identify patterns and disparities more effectively among all status that build the latent variables (Morlini, 2013), which, in this case, is a nutritional status. Regarding public health, clustering allows us to divide cities and regencies into clusters that separate them based on status such as stunting-wasting prevalence and underweight and overweight prevalence. Cities or regencies with similar nutritional profiles can be grouped

using clustering techniques, providing a data-driven foundation for targeted interventions and policy-making (Sentika et al., 2024).

Clustering can be done using parametric and non-parametric statistics. One of the most prevalent methods in parametric statistics is the Gaussian Mixture Model (GMM). The Gaussian Mixture Model (GMM) is a probabilistic clustering method that assumes data points are generated from a mixture of multiple Gaussian distributions (Giesen et al., 2024). This method is different than the other classic clustering techniques method, such as K-means, which assigns each data point to a single cluster based on distance; GMM assigns probabilities to each data, allowing for a more flexible and realistic representation of data (Wahidah & Utari, 2023).

Commonly, GMM uses the Expectation-Maximization (EM) algorithm, which is based on the maximum likelihood principle, to iteratively refine the cluster assignments, ensuring that the model converges to the best possible cluster (Raditya et al., 2022). The benefit of using the EM algorithm comes from how it guarantees monotonic improvement in likelihood with each iteration. This means it steadily approaches a better solution, making GMM and EM algorithms suitable for classifying unsupervised data with complex distributions, such as nutritional status, which can form irregular distribution. (Alqahtani & Kalantan, 2020).

The EM algorithm in this study is initialized using cluster partitions generated by model-based agglomerative hierarchical clustering (MBAHC). This method works by repeatedly merging the two clusters that cause the smallest drop in classification likelihood under a Gaussian mixture model. This method is advantageous since the underlying probabilistic model is shared by both the initialization step and the model fitting step. Hierarchical clustering is also

convenient because we need only a single run to provides the basis for initializing the EM algorithm for any number of cluster.

Nutritional status studies in Indonesia rely heavily on traditional statistical methods such as descriptive analysis or basic clustering techniques like K-means (Greve et al., 2016; Saputra & Harini, 2022). Even though these methods can provide insights, they often fail to account for the complexity of the data distribution, including overlapping distributions and non-linear relationships (Andryan et al., 2023). There is a lack of studies applying probabilistic clustering methods like the Gaussian Mixture Model (GMM) to nutritional data or the health field in Indonesia. This study aims to bridge this gap by providing a more nuanced classification with a model-based approach in GMM.

By applying GMM to the clustering of nutritional status in Java Island, this study aims to find meaningful patterns that can be used in policymaking to make an a more efficient program based on each region's characteristics. For example, GMM can be used to identify clusters of cities and regencies with high stunting or wasting prevalence and can help prioritize government programs such as food assistance, education, and healthcare. Similarly, recognizing clusters with rising trends of overweight or underweight individuals may lead to targeted awareness campaigns promoting healthier lifestyles. In this study, GMM employs a data-driven approach to inform public health decision-making, providing insights that reflect real-life situations.

1.2 Problem Identification

The problem identification of this study are:

1. How should the cities and regencies in Java Island be best classified with the Gaussian Mixture Model based on nutritional status?
2. What are the characteristics of the clusters formed, and how do they explain the nutritional situation in cities and regencies in Java Island?

1.3 Problem Limitation

This study focuses exclusively on cities and regencies within Java Island in 2023 based on Indonesia's Ministry of Health report. This study does not aim to cover the root cause (e.g., cultural practices or economic policies) but focuses on identifying patterns through clustering. Finally, the GMM assumes data follows a Gaussian distribution, which may oversimplify real-world nutritional dynamics, requiring careful validation.

1.4 Aim of Study

The aim of this study are:

1. To find the best model to classify the cities and regencies in Java Island with the Gaussian Mixture Model based on nutritional status.
2. To unveil the characteristics of the clusters formed and explain the nutritional situation in cities and regencies on Java Island.