CHAPTER I
INTRODUCTION

1.1 Background

Adequate housing quality is one of the important elements in improving community welfare. A house not only functions as a place to live, but also serves as a primary support for creating healthy, safe, and productive living conditions [1]. However, in several regions, including Jombang Regency, there are still many families living in houses with physical conditions that do not meet proper housing standards. These problems include damage to the foundation, walls, floors, and roofs, as well as limited sanitation facilities and restricted access to basic services. Inadequate housing conditions can negatively affect health, reduce productivity, and increase the economic vulnerability of the community [2].

As a response to these problems, the government has implemented The RTLH Assistance Program, which is intended to help low-income families obtain proper and decent housing [3]. Although the objective of this program is highly strategic, the process of determining the priority of assistance recipients remains a significant challenge. Priority assessments are generally carried out manually through field surveys and direct verification by officers, making the process vulnerable to subjectivity and inconsistencies in decision-making [4]. In addition, the large amount of data and the diversity of variables that must be analyzed make the classification process of prospective assistance recipients inefficient when relying solely on traditional methods [5].

RTLH data itself includes various important aspects that influence housing feasibility, including the physical condition of the building, the availability of sanitation facilities, access to clean water, and the socio-economic condition of the family [6]. The complexity of the relationships among these variables requires an analytical approach that is capable of processing information comprehensively and producing more accurate classification results [7]. Manual systems are not capable of accommodating the dynamics of variables effectively, especially when the relationships between features are nonlinear or influence one another. As a result, the determination of assistance recipient priorities may become less accurate and not properly targeted [8].

Previous studies have widely applied various classification algorithms for determining social assistance recipients. Julian Bramanda (2025), in a study entitled "Classification of Communities Receiving Government Social Assistance Using the C4.5 Algorithm Method," successfully implemented the C4.5 Decision Tree algorithm to identify assistance recipients based on attributes such as income, home ownership, and health conditions. However, the study revealed that the C4.5 algorithm has limitations in handling numerical data optimally and is prone to overfitting, especially when the dataset has high complexity and interdependent variables [9]. On the other hand, Damuri et al. (2021), in the journal entitled "Implementation of Data Mining Using the Naïve Bayes Algorithm for Classification of Basic Food Assistance Eligibility," showed that the Naïve Bayes algorithm was able to achieve an accuracy of 86% with a precision of 88%. Nevertheless, this approach assumes independence between features, an assumption that is often inconsistent with the reality of complex and interconnected socio-economic data [10].

These two algorithms are also not specifically designed to handle class imbalance, which commonly occurs in social assistance data, and tend to be less efficient when dealing with large volumes of data. Therefore, to overcome these limitations, this study proposes the use of the Light Gradient Boosting Machine (LightGBM) algorithm. LightGBM offers high computational efficiency through histogram-based learning and leaf-wise growth techniques, allowing faster processing of large-scale data with lower memory consumption. In addition, LightGBM is capable of handling nonlinear relationships and complex interactions between variables, as well as providing internal mechanisms for handling imbalanced data [11].

To further optimize the model performance, Bayesian Optimization was also applied to perform hyperparameter tuning in a more efficient and effective manner compared to grid search or random search methods. With this combination, the study is expected to achieve not only higher classification accuracy, but also a more robust model that is ready to be implemented on real-scale social assistance program data [12]. By combining the advantages of LightGBM and Bayesian Optimization, the developed classification model is expected to provide more accurate, stable, and efficient results in determining the priority level of RTLH assistance recipients [13].

The implementation of this model not only helps improve objectivity in the process of determining assistance recipients, but also supports the efforts of the Jombang Regency Government in achieving more targeted housing assistance distribution. Through this data-driven approach, the research is expected to provide a meaningful contribution to improving the effectiveness of housing assistance programs and strengthening efforts toward equitable community welfare.

1.2 Problem Statement

The research problems addressed in this study are formulated as follows:

1. How can the LightGBM algorithm be implemented to classify the priority of RTLH assistance recipients on data with diverse variables and complex relationships between features?
2. How can the Bayesian Optimization method be used to determine the optimal hyperparameter combination for LightGBM in order to obtain the best classification performance?

1.3 Research Objectives

The objectives of this undergraduate thesis entitled “Classification of Priority Recipients for RTLH Assistance Using LightGBM with Bayesian Optimization in Jombang Regency” are as follows:

1. To implement the LightGBM algorithm for classifying the priority of RTLH assistance recipients in order to handle the diversity of variables and the complexity of relationships between features in household data in Jombang Regency.
2. To optimize the performance of the LightGBM model through the implementation of the Bayesian Optimization method in determining the most optimal hyperparameter combination for the classification performance of RTLH assistance recipient priorities.

1.4 Research Benefits

Several benefits that can be obtained from this undergraduate thesis entitled "Classification of Priority Recipients for RTLH Assistance Using LightGBM with Bayesian Optimization in Jombang Regency" are as follows:

1. This study provides an understanding of the implementation of LightGBM in the RTLH classification system. It explains how LightGBM is used to process variables related to housing physical conditions and socio-economic characteristics in order to produce more accurate classifications of RTLH assistance priorities
2. The results of this research can serve as a reference in selecting appropriate classification methods for determining assistance priorities by showing the advantages of LightGBM optimized with Bayesian Optimization.
3. The classification model developed in this research can support the improvement of data-based RTLH assistance distribution. It can help related institutions improve the verification process and determine assistance recipients in a more objective and targeted manner.
4. This research encourages the use of artificial intelligence in the fields of housing and social welfare. It also provides opportunities for the development of AI-based decision support systems for social assistance programs, particularly through the implementation of LightGBM and hyperparameter optimization.

1.5 Scope and Limitations

The limitations of this undergraduate thesis entitled "Classification of Priority Recipients for RTLH Assistance Using LightGBM with Bayesian Optimization in Jombang Regency" are as follows:

1. The dataset used in this study consists of RTLH data from Jombang Regency obtained from the related institution for the period of May 31, 2024, to October 1, 2025.
2. The variables analyzed in this study include socio-economic characteristics and the physical condition of housing available in the RTLH dataset, including age, education, occupation, gender, income, housing status,

housing assets, land status, land assets, electricity access, assistance status, area classification, room function, foundation, columns, frame structure, ceiling, beams, *sloof*, windows, ventilation, floor materials and conditions, wall materials and conditions, roof materials and conditions, house area, land area, number of occupants, water sources, distance to waste disposal sites, and toilet facilities (bathroom, toilet seat, and type of waste disposal system).