CHAPTER I
INTRODUCTION

1.1. Background of Study

The rapid advancement of information technology has led to an increasing number of tasks in software development, making the task management process more complex. In a dynamic and collaborative work environment, task assignment management plays an important role in determining team effectiveness and the success of software development. Each company may have numerous tasks that need to be allocated to the most suitable workers based on their competencies, experience, and work history[1] .

However, in practice, the task assignment process is still largely carried out manually by project managers or team leaders. Determining the most appropriate software developer for a particular task is not always straightforward, as expertise and experience are dynamic factors that evolve over time and across projects. A mismatch between task characteristics and a software developer's competencies can result in delays in task completion, increased workload, and reduced software product quality. Furthermore, manual assignment processes are prone to subjective bias and inefficiency, especially when dealing with a large number of workers and tasks[2].

Several studies have shown that human factors, such as individual experience and expertise, have a significant influence on the effectiveness of task recommendation systems[1]. If task assignment is carried out solely based on intuition or static profile data (for example, the area of expertise listed in a developer's profile), the results are often suboptimal because they do not specifically reflect the developer's actual experience. Therefore, a more adaptive approach is required by utilizing historical data of developer activities, such as commit logs and previous task completion records[3]. This history-based approach enables the system to identify work patterns and software developer preferences based on real activities.

Currently, many recommendation systems employ the Content-Based Filtering method, which utilizes content information to generate recommendations for new inputs. For instance, modern approaches that combine keyword extraction and machine learning techniques, such as the integration of word embedding and Fuzzy logic, are capable of improving recommendation accuracy by up to 30% compared to traditional Content-Based Filtering methods that rely solely on Cosine Similarity [4].

In another study, a recommndation system was developed by utilizing textual data through the Content-Based Filtering method combined with the DBSCAN algorithm [5]. This approach aims to improve recommendation quality by grouping data based on content similarity. DBSCAN is used to automatically form clusters without requiring the number of clusters to be predefined, making it more flexible in handling large-scale data with uneven distributions. After the clustering process is completed, each task or document belongs to a cluster representing similar topics. When a user submits new data or requests, the system analyzes the content to determine the closest cluster and then provides recommendations from the cluster that is most relevant to the characteristics of the input.

In line with this concept, the present study also applies a Content-Based Filtering approach but employs a different clustering method. While previous studies combined Content-Based Filtering with DBSCAN, this research integrates Content-Based Filtering with HDBSCAN in order to improve flexibility and accuracy in the clustering process.

This study performs clustering using HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise). HDBSCAN is capable of identifying clusters with varying density levels without requiring the number of clusters to be specified in advance[6]. In addition, this algorithm only requires the minimum cluster size parameter, making it more adaptive in handling data with uneven distributions [8]. The clustering results produce groups of tasks that are closely related based on their content similarity.

HDBSCAN was selected because of its advantages in identifying clusters with varying densities without the need to determine the number of clusters beforehand. One of the main benefits of HDBSCAN is its ability to generate clusters with different density levels, unlike standard DBSCAN. In other words, HDBSCAN can identify both dense and more loosely connected task groups. Furthermore, the algorithm only requires the specification of the desired minimum cluster size parameter [7].

After the clusters are formed, Content-Based Filtering plays a role in the recommendation process. When a new task is submitted, the system analyzes the task description to determine the nearest cluster and subsequently selects developers who have previously handled tasks within the same cluster. This approach does not require preference data from other users, as in Collaborative Filtering, making it suitable for work environments with limited interaction data.

Therefore, the proposed method is expected to assist in determining the most suitable developers in a more objective and efficient manner based on content analysis and historical developer activity data.

1.2. Problem Formulation

Based on the background that has been described, the research problems discussed in this study are formulated as follows:

1. How can Content-Based Filtering and HDBSCAN be implemented for developer recommendations in a web-based task management application?
2. How effective are the results of applying Content-Based Filtering and HDBSCAN for developer recommendations in task management?

1.3. Research Objectives

Based on the formulated research problems, the objectives of this study are as follows:

1. To implement the Content-Based Filtering and HDBSCAN methods in providing developer recommendations for a web-based task management application at PT. Tunas Kreasi Digital.
2. To analyze the results of applying the Content-Based Filtering and HDBSCAN methods in generating developer recommendations for the task management application at PT. Tunas Kreasi Digital.

1.4. Significance Study

Based on the research objectives above, the expected benefits of this study are as follows:

1. This research contributes to the development of knowledge in the field of recommendation systems, particularly in the context of task assignment for workers (developers).
2. This study may serve as a reference or evaluation material for future research, especially studies related to task assignment recommendations for developers.

1.5. Scope and Limitation

In this study, several limitations are applied in order to maintain the scope of the research so that it remains focused and well-defined, as follows:

1. The developer log data are obtained from GitHub commit logs and daily logs from PT. Tunas Kreasi Digital.
2. The research is conducted only within the scope of PT. Tunas Kreasi Digital.
3. The data used in this study consist of log data recorded from 2024 to 2025.

These limitations are established to ensure that the research remains within a manageable scope and allows for more focused analysis and testing in accordance with the main objectives of the study.