CHAPTER V
CONCLUSION AND RECOMENDATION

5.1 Conclusion

The implementation of Content-Based Filtering combined with HDBSCAN successfully produced a developer recommendation system that is relevant and effective based on activity log data. The optimal HDBSCAN configuration with min_cluster_size = 2 and min_samples = 2 achieved the highest Silhouette Score of 0.559 and the lowest noise percentage of 18.71%.

Under this configuration, the system reached its best performance with Hit Ratio = 0.88, Recall = 0.80, and MRR = 0.563. The cosine similarity threshold testing used as a filtering mechanism for the training data showed that using the entire dataset without filtering (100% threshold, 4,505 data entries) produced the highest Hit Ratio and Recall values. This demonstrates that data completeness in historical logs significantly contributes to the coverage and accuracy of the recommendation results.

Compared to pure Content-Based Filtering (CBF), the hybrid CBF + HDBSCAN approach outperformed it across nearly all evaluation metrics, proving that clustering developer expertise patterns helps overcome the limitations of standard CBF. Therefore, the proposed system is suitable for implementation to support decision-making in developer task assignment within software development projects.

5.2 Recomendation

From this study, several recommendations can be proposed for future research developments, as follows:

1. This study uses the TF-IDF method for text representation, which focuses on word frequency without considering semantic context. Future research can implement word embedding-based models such as Word2Vec, FastText, or BERT to enable the system to understand word meaning in a contextual manner and improve recommendation accuracy.
2. This study focuses on developer activity log content. Future studies may incorporate additional factors such as project history, task difficulty level, task completion performance, or the level of collaboration between developers to produce more comprehensive recommendations.
3. Dimensionality reduction in Content-Based Filtering. Since pure CBF has already shown relatively good performance, future research can explore the application of

dimensionality reduction techniques such as PCA, SVD, or UMAP on TF-IDF representations before similarity computation. This aims to evaluate whether dimensionality reduction can improve recommendation accuracy while also reducing computational time, enabling a more structured comparison between standard CBF, CBF with dimensionality reduction, and CBF combined with HDBSCAN.

Overall, these suggestions are expected to provide direction for further improvement of the proposed system, particularly in enhancing recommendation accuracy, computational efficiency, and contextual understanding of developer activity data. By exploring more advanced text representation techniques and incorporating additional contextual features, future research can develop a more robust and adaptive recommendation system that better reflects real-world software development environments.