CHAPTER V
CONCLUSION AND SUGGESTIONS

This chapter presents the conclusions derived from the conducted research along with recommendations for future studies. The conclusions encompass the performance of the top-performing model in predicting movie admissions, as well as the contributions of pre-release metadata to the prediction outcomes. Meanwhile, the future recommendations are directed toward developing more complex models, incorporating additional features, and exploring alternative methods to enhance prediction accuracy in the future..

5.1 Conclusion

Based on the conducted research, it can be concluded that the movie admissions prediction model based on pre-release metadata was successfully developed and yielded satisfactory results. In accordance with the established problem statements, the conclusions are formulated as follows:

1. The movie admissions prediction model for the Indonesian film industry was successfully developed using the XGBoost, LightGBM, and CatBoost algorithms through several primary stages, namely data preprocessing, feature engineering, train-test splitting, and model training. During the preprocessing phase, missing value handling and a logarithmic transformation on the target variable were executed to mitigate data skewness. Subsequently, feature engineering was performed utilizing a Bayesian smoothing approach to construct popularity features. The three algorithms were then trained on the processed data and evaluated using RMSE, MAE, MAPE, and R^2 metrics. Furthermore, hyperparameter tuning was conducted via Random Search and Bayesian Optimization to enhance model performance. The results demonstrate that all three models are capable of effectively learning the relationship between pre-release metadata and movie admissions, thereby validating their utility as predictive models within the context of the film industry.
2. Based on the experimental results and model evaluation, CatBoost emerged as the top-performing algorithm compared to XGBoost and LightGBM. This

superiority is demonstrated by its achieving the lowest RMSE of 0.7698 and the highest R^2 value of 0.8729 in the optimal scenario. Although XGBoost exhibited performance improvements following hyperparameter tuning and LightGBM maintained stable performance, CatBoost consistently outperformed them, delivering more accurate and robust predictions on this research dataset.

3. Based on feature importance and SHAP analyses of the optimal CatBoost model, individual popularity-based features exert the most dominant influence on movie admissions predictions, with Lead_Actor_Popularity being the highest contributor. SHAP values indicate that higher individual popularity scores positively impact admissions, meaning that well-known individuals involved in a film significantly increase its likelihood of attracting larger audiences. Conversely, genre_popularity exhibits a relatively minor influence, indicating that genre is not a primary driver of admissions within this model. These findings demonstrate that individual-based factors play a more critical role than categorical factors in the film industry, serving as a vital consideration for movie production and marketing strategies.

5.2 Suggestions

Based on the findings of this research, several recommendations can be put forward for both future studies and the further development of the implemented system, as follows :

1. Future research is recommended to expand the dataset and incorporate external features, such as production budget, marketing strategies, and social media data, alongside advanced feature engineering techniques to enrich data representation and enhance prediction accuracy.
2. From a modeling perspective, future studies could explore more complex algorithms or ensemble approaches, alongside more adaptive hyperparameter tuning methods, to achieve more optimal and stable model performance.