CHAPTER V
CONCLUSIONS AND SUGGESTIONS

5.1. Conclusion

From the results of this research on predicting stress levels based on sleep conditions using the Random Forest and XGBoost methods, the following conclusions can be derived:

1. The Random Forest and XGBoost models were successfully developed to predict stress levels based on eight physiological sleep parameters, namely Snoring Rate, Respiration Rate, Body Temperature, Limb Movement, Blood Oxygen, Rapid Eye Movement, Sleeping Hours, and Heart Rate. The model development process began with data collection using the SaYoPillow dataset obtained from Kaggle, consisting of 5,670 data points distributed evenly across five stress level classes. The dataset was then subjected to preprocessing using Min-Max Scaler normalization prior to training with both algorithms. To enhance model performance, hyperparameter optimization was performed using Bayesian Optimization, which successfully identified the optimal parameter combinations for each model. The trained models were subsequently deployed into a Streamlit-based web application, enabling direct use by end users.
2. According on the performance comparison results of the four tested models, Random Forest with Bayesian Optimization showed the best overall performance with an accuracy of 0.9841, precision of 0.9844, recall of 0.9841, and F1-Score of 0.9841. Meanwhile, default XGBoost and XGBoost with Bayesian Optimization produced identical performance with an accuracy of 0.9762, precision of 0.9774, recall of 0.9762, and F1-Score of 0.9763. The application of Bayesian Optimization was proven to provide a more significant performance improvement on Random Forest compared to XGBoost, with an accuracy increase of 0.0079. From the results of the test scenarios, it was found that physiological parameters of sleep are very strong predictors for determining stress levels, as evidenced by the high performance of both models above 97% in all test scenarios. The only factor that showed a significant impact on performance was the composition of the dataset split, where Random Forest achieved its best performance at a 70:30 ratio with an accuracy of 0.9894.

5.2. Suggestion

In accordance with the results obtained, several recommendations can be made for subsequent research:

1. The SaYoPillow dataset used in this study is simulated, resulting in highly consistent and clean patterns. Future research is recommended to use real-world clinical data collected directly from wearable devices or sleep monitors to better represent real-world conditions and address various challenges, such as noise and class imbalance.
2. This study only used eight physiological sleep parameters. Future research could consider adding other features such as psychological data, daily activity, or user demographics to improve the model's predictive capabilities more comprehensively.
3. Further research could expand the comparison by including other algorithms such as LightGBM, CatBoost, or deep learning approaches such as LSTM that are able to capture temporal patterns from sequential sleep data.
4. The application developed in this study is still based on a local website. Future research could develop a mobile-based application or integrate it directly with wearable devices to enable real-time predictions and greater public access.