

CHAPTER 6

CONCLUSION

Several conclusions are gained in this research about predicting heart disease with the XGBoost algorithm:

1. The XGBoost without feature selection has better overall performance evaluation scores than AdaBoost without feature selection with higher precision, roc area, and accuracy. The XGBoost with Chi Square feature selection also has better overall performance evaluation scores than AdaBoost with Forward Feature Selection but with lower recall on the XGBoost with Chi Square feature selection. This shows XGBoost is better in predicting heart disease than AdaBoost with both of models using optimal parameters.
2. The XGBoost with Chi Square feature selection has the best performance with 11 feature selected (“thalach”, “ca”, “thal”, “oldpeak”, “exang”, “chol”, “age”, “cp”, “sex”, “restecg”, and “slope”) than Mutual Information, ANOVA, Forward Feature Selection, Backward Feature Selection, Recursive Feature Elimination, and Feature Importance. The XGBoost with Chi Square feature selection is compared with XGBoost without feature selection shows that XGBoost with Chi Square feature selection has better overall performance. This also shows that feature selection can make the XGBoost improve its performance further with fewer features.

With these conclusions, the importance of the recall evaluation score may be considered in future research especially in the medical field as it is affected by a False Negative which indicates the model predicted a negative on a supposedly positive target. The XGBoost hyperparameter may be tuned further with different potential search spaces that may further increase performance.