

CHAPTER 6

CONCLUSION

Based on the results obtained from the case and the dataset used in this project, it can be concluded that:

1. Extreme gradient boosting and Logistics regression can be applied to predict loan approvals for this project. The best accuracy obtained from several experiments is quite good. By using Extreme gradient boosting acquisition of 82% using dataset 60:40, and Logistics Regression of 69% for dataset ratio 70:30.
2. For the case in this project, the XGBoost algorithm is more suitable than logistic regression, from the accuracy results it can be seen that XGboost has 82% and Logistics Regression has 69%, this is because XGBoost uses a tree-based approach which is suitable for the data set in this project. And XGBoost has gradient and hessian as the error calculation while logistic regression only uses gradient for error calculation. Then XGBoost can choose the best variable using the gain calculation function, while logistic regression cannot choose a variable because this model uses all existing variables. Xgboost can have better accuracy because apart from being able to choose the best few variables, this model has its weighting for the variables used.
3. Based on the selection of features that use correlation, we can see that the feature that most influences the outcome of loan status is the credit history, and the one that does not affect the applicant income. Meanwhile, according to the statistics in the EDA process, the features that influence the outcome of the loan status decision are marriage, education, credit history, property area, and those that do not affect are gender, dependents, self-employed.

For further research, it is expected to be able to explore the XGBoost algorithm with the same data in order to produce better accuracy values and compare it with other tree-based algorithms.