

## CHAPTER 6

### CONCLUSION

After doing some experiments, some things that can be concluded are :

1. According to this project, **The best accuracy we get from this project is 88.88% in Decision Tree Algorithm with 80% train data and 20% test data, Random state = 0 and Balanced data.** The result of model performance is also very well, the precision score is 86.81%, recall gets 89.77% and F-1 Score gets 88.26%.
2. **The Decision Tree Algorithm can provide a good prediction on predicting diabetes.** Compared to XGBoost algorithm, Decision Tree works really well with balanced dataset and random state = 0 based on this project.

The Decision tree works well with balanced dataset rather than with imbalance dataset, works better with random state = 0 rather than random state = 42, and works better with 80% train data and 20% test data. We have to balanced the dataset because imbalanced dataset can cause the classifier to perform poorly on real unseen data and the severely skewed class distribution and the unequal misclassification costs. The random state = 0 works better than random state = 42 because random state = 0 more suitable for Pima Indian Diabetes Database (PIDD) than random state = 42

After doing the 5th and 6th research about comparing XGBoost and Decision Tree Algorithm without using Age attribute and balancing data, the result is XGBoost algorithm significantly has better accuracy than Decision Tree Algorithm