CHAPTER 6
CONCLUSION

Based on the result, XGBoost can be used on diabetes prediction, and it is able to achieve accuracy above 90% which is good. So based on the result, both models are usable and good for diabetes prediction. From Table 5.1 to 5.5 and Figure 5.1 shows that with 70% ratio of training set is the best ratio for this dataset, because ANN 1, ANN 3, ANN 4, XGBoost 1 and XGBoost 2 achieve the highest score compared with other ratio. ANN 3 is able to achieve better accuracy scores at 80% of training data and achieve the same accuracy score at 70% of training data as the XGBoost models. Some of the ANN models are able to achieve same result as the XGBoost models , as we can see with 80%:20% ratio, ANN 4 have the same accuracy score as the XGBoost models but it has better recall and f1 score. With 60%:40%, ANN 1, ANN 2, ANN 3, and XGBoost 1 able to achieve the same accuracy scores, but ANN 3 have the highest recall but lower precision than ANN 1, ANN 2, and XGBoost 1, as for XGBoost 2 is second lowest accuracy score. And the last with 50% of training data XGBoost 2 have the best overall performance. XGBoost with hyperparameters tuning (XGBoost 2) able to achieve the same result or better performance than XGBoost with default parameters (XGBoost 1), the results shows that with 75% or 50% of training data XGBoost 2 outperforms XGBoost 1 but with 60% of training data XGBoost 1 have better performance than XGBoost 2. And with 80% ANN 3 able to outperform the XGBoost models.

So, with lower ratio of training set XGBoost are better but with higher ratio of training set ANN 3 able to achieve better result than the XGBoost models. The ANN models were built with 3, 4, 5, 6 hidden layers respectively, but the model with 5 hidden layers give the best performances over all ANN models, the last model's performances began decreasing or have big gap than the previous with 5 hidden layers. by adding more hidden layers we also can improve the ANNs performances but it's only to certain extend, after that the performance began to drop. We can improve the XGBoost models by adding more data to the training set and hyperparameters tuning to get the optimal parameters.