

CHAPTER 6

CONCLUSION

This chapter is to answer the questions asked at the beginning of the project, and also gives the conclusion of the observed results. The algorithms managed to do very well in terms of predicting actual illness given the common symptoms, especially for XGBoost which consistently performs better and has a more accurate prediction compared to random forest and TensorFlow's multilayer perceptrons. And because of it the best algorithm for this sort of scenario, in which the features are abundant, and the dataset is small is XGBoost. And as seen in the data reduction scenario, even with better data (more agreeable entries) the result will be worse compared to more data even with less agreeable entries. To improve the result of the prediction of all algorithms it is better to have a larger number of data for the algorithms to learn from, also make sure the data is balanced and if possible, give more features so one illness is more distinguishable from the others. Because generally algorithms take numbers as label, be sure to encode the labels to become numbers first, so it will be acceptable for the algorithms.

For future research, it can be suggested that larger symptoms and more illnesses may also be added so it can cover not only just oral symptoms but also other symptoms in other areas of the body. The amount of data can also be increased for better predictions. And finally, for a more robust method of evaluation, K-Fold cross validation is recommended, as it is used to flag problems like overfitting or selection bias and gives insights on how the model will generalize to an independent dataset.