CHAPTER 1
INTRODUCTION

1.1. Background

Vast majority of people ignore or are oblivious of their oral health, even though issues that arise from oral health can be severe if not discovered quickly enough can lead to a multitude of problems. One of the issues that also contribute to this is that the range and capability of seeing dentists are limited in some areas, not to mention that the cost of checking and the time for queuing just for a check-up can be overwhelming, especially for people at the lower economic level of society. In this project, the main idea is to predict the possibility of illness based on symptoms from the mouth area, meaning there are: gums, teeth, tongue, lips, etc. Based on the research done, at times something that seems to be trivial symptoms can lead to major issues, such as diabetes, bulimia, oral cancer, or kidney failure.

It has to be made clear that this tool does not provide medical advice. This tool is not intended to be a substitute for professional medical advice, diagnosis, or treatment. The data for the symptoms is gathered through thorough research of journals and also trusted medical websites (ex. WebMD) that cover the diseases that will be given as a result of the algorithms' prediction, therefore several features will then be encoded are used to classify each disease, and then it will be given labels. Three algorithms are used to compare the performance. Extreme Gradient Boosting (XGBoost) will be used mainly as the prediction algorithm, but there are a couple of algorithms that will be used as performance comparisons. Comparisons will be done using a rather basic random forest algorithm and also a more sophisticated multilayer perceptron using TensorFlow.

It is found with distinctive features of illness and algorithms to classify diseases based on symptoms given, classification will give accurate illness based on the symptoms given as an input.

1.2. Problem Formulation

1. How accurate is machine learning's prediction to actual symptoms?
2. What algorithm will be better in solving the problem?
3. How to improve the results?
4. Why is encoding necessary for the dataset?

1.3. Scope

This project aims to deal with major illnesses that can be classified based on multiple distinctive oral symptoms in the area of the mouth. Those illnesses in question are diabetes, bulimia, oral cancer, and kidney failure. Therefore, this study does not cover mild symptoms which by this project's definition, symptoms given are only singular symptoms or a couple of symptoms and they do not occur in multiple areas of the mouth.

1.4. Objective

There are two main objectives for this project. The first is to achieve a satisfying result in illness prediction, the criteria for that is to get an accurate prediction for each illness. The second is to compare the performance of several algorithms used to predict illness.