CHAPTER 3
RESEARCH METHODOLOGY

1. Literature Study

This process consists of collecting and reading journals that are related to respiratory infection disease and journals related to the accuracy of using basic Bayes Theorem to predict probabilities. All the journals in this literature study are used as references for the making of this journal and programs.

2. Collecting and Augmenting Data

This project uses data collected from various sources which then will be accumulated and analysed to prevent duplicate data. The data will then be used as a basis to create a program using Bayes theorem to predict how many people could be infected after certain amount of time and then the data will be analysed to be used as a preventative way to reduce risk of another outbreak. Naïve Bayes theorem can also be used to generate missing data or increase the amount of data by adding data which has been slightly modified from the existing data.

3. Implementation with Programs.

Making code programs for analysing the data and then finding the basic reproduction number of the virus. Then using the analysed data to create a program to predict and then visualized the prediction rate of said virus.

4. Testing

Testing the data by splitting the existing data to 2 group, train and test. This project uses 1755 data accumulated from various sources on the Hong Kong 2003 SARS outbreak with a time-frame of 15 Feb 2003 – 31 May 2003 as the last day of the onset of cases. Half of the existing data will be used as training data for the system to analyse and half of the data will be used as test data, where the theorem will be applied and further analysed to compare.

5. Analyse

Comparing the results from the program with the data from the Hong Kong 2003 SARS outbreak to make the conclusion on what can be used and when to do it to prevent another outbreak of an epidemic.