In addition to that, this study does not aim to provide a general solution for all hospitals since each hospital would differ in the type and amount of facility they have and the doctors who provide the diagnosis for the illness. The researchers found it too difficult to give a general solution which is why this study is only geared on developing one solution for one specific instance of this problem on one specific local hospital.

Chapter 2

REVIEW OF RELATED WORKS

2.1. Inpatient Length of Stay

One of the main concerns in healthcare today is the prediction of inpatient length of stay (Liu, et. al, 2006). Length of stay (LOS) can be defined as the term commonly used to describe the duration of a single episode of hospitalization of a patient. It is computed by subtracting the day of admission to the day of discharge. Persons who are admitted and subsequently discharged on the same day would have a length of stay of one day.

In the study of Gustafson (1968) he indicated that the physicians attending to a patient sometimes overestimates or underestimates the actual length of stay of a patient. Due to this miscalculation, hospital facilities, man power and equipment are not well allocated. Gustafson focused his study on patients who have inguinal herniotomy. He used the following statistical approaches to determine the LOS of the patients: historical mean, regression analysis, direct posterior method and Bayesian theorem. Out of the four, the Bayesian theorem was found to be most effective. In his study, Gustafson tried to classify the LOS patients from 5-16 days since these are the common LOS of patients with herniotomy.

The study of Kim et. Al (2000) used neural networks and logical analysis, both of which are data mining techniques, to estimate the LOS of the patients in the post-anesthesia unit. The results of their study showed that the neural networks approach was more efficient compared to the logical analysis. They also discussed that the prediction of LOS is dependent upon the diagnosis of the physician since it is one of the major basis for the approaches that they have used.

On the other hand, the study of Lucas, et.al (2004) which focused on patients with heart failures stated that the important factor in the prediction of patients' LOS is dependent on the facilities that a hospital possesses. An added factor to the accuracy of the prediction is the diagnosis of a hospital physician.

According to WHO, the average LOS of dengue patients is from 2-7 days. This figure though is based on the assumption that no other complications will follow the initial findings of the physician but if the patient's condition worsens the LOS may reach to up to two weeks to a month depending on the recuperation ability of the patient. Although there is no specific anti-viral medicine to cure this ailment, hospitalization of dengue patients are still encouraged. The mortality rate of patients that are not hospitalized reach to up to 50%. With proper hospital treatment it comes down to 3%. The basic treatment for the fever caused by the disease is paracetamol. Supportive treatment such as intravenous is needed to prevent the patient's condition from getting worse.

2.2. Prediction Using Data Mining

Data mining is quite visible in the fields of e-business, marketing and retail (Canlas, 2009). In the past years, data mining has also become a great tool in the healthcare sector. Healthcare industry today generates large amounts of complex data about patients, hospitals resources, disease diagnosis, electronic patient records, medical devices etc. (Desikan, et. al, 2011). Data mining is typically used for prediction and classification. In this section, the approaches used in predictions using data mining will be discussed, along with other studies which are related to them.

2.3. Neural Networks

Neural Networks (NN) consist of a structure or network of numerous interconnected units (artificial neurons). Neural Network is considered to imitate the way human brain works by acquiring knowledge from its environment through a learning process.

Kim, et al (2000) compared the efficiency of neural networks and logical regression method in predicting the length of stay of post-anesthesia patients following the general anesthesia. Their study claims that the neural network was found to be better at predicting the length of stay of a patient. They added that the drawback of this approach is that the process would need a number of samples to arrive at an accurate prediction. Also, NN's decision criteria are difficult to understand even for an expert.

2.4. Decision Trees

A decision tree is a tree-structured classifier built by partitioning the data set into homogeneous classes (Bayat, et.al, 2009). The root node is split into child nodes by selecting the variable that best classifies the samples according to a split criterion. The splitting continues on the child nodes until stopping criteria are met.

In the study of Bayat et.al, they compared the efficiency of decision trees and Bayesian network in predicting the access to the renal transplant waiting list. They have discussed that both approaches have very high predictive performance. Decision tree was able to correctly predict the access to renal transplant of 18 patients out of 20. The McNemar's test for correlation was conducted on the two approaches to determine which

approach is more efficient for the problem; it showed no significant difference between the two. This means that the two approaches have the same level of efficiency in predicting the patients who will have access to renal transplant waiting list. Moreover, the results showed that the age of the patients is an important variable in determining the accuracy of the predicting approaches.

2.5. Bayesian Networks

Bayesian networks with their associated methods are especially suited for capturing and reasoning with uncertainty (Pearl, 1988). They have been around in biomedicine and healthcare for more than a decade now and have become increasingly popular for handling the uncertain knowledge involved in establishing diagnoses of disease, in selecting optimal treatment alternatives, and predicting treatment outcome in various different areas (Lucas et. al, 2004). The Bayesian approach is also increasingly developed in areas of healthcare. Examples include the use of Bayesian networks in clinical epidemiology for the construction of disease models and within bioinformatics for the interpretation of microarray gene expression data.

Hinde et. al (2009) compared Bayesian networks against Decision Trees and Neural Networks. They showed that the Bayesian networks is not inferior to the two more popular approaches. According to the results of the study of Hinde et al (2009), it showed that the Bayesian approach has a precision value of 99.37% . And although neural network is as accurate, it is harder to train and it takes longer to train due to its complex nature. On the other hand, the results of the decision trees showed that it was more efficient than the other two approaches and its results are also easier to understand. The researchers concluded that although the Bayesian