

## CHAPTER 1

### THE PROBLEM

This study is dealing with the prediction of Dengue Haemorrhagic Fever in Cimahi, West Java, Indonesia. Before it is discussed in great detail, the Rationale, Theoretical Framework, Conceptual Framework/Paradigm, Statement of the Problem, Hypotheses, Assumption, Scope and Delimitation, Importance of the study, and Definition of terms are discussed briefly in chapter 1.

#### 1.1 Rationale

Dengue haemorrhagic fever (DHF) is one of the endemic diseases in various parts of Indonesia. The number of DHF cases tends to increase and spread wider. It is stimulated by many factors such as: agent or vector, living behavior, and immune of human [3,4,7]. The growing up of agent or vector is influenced by weather and living behavioral too [3,7]. Dengue fever is very infectious, its spread cannot be limited by the concept of territory. Up to this moment, no medicine nor vaccine has been found to overcome the disease. The disease is still potential to cause outbreaks and usually occurs with a grace period of 3 to 5 years [7].

Cimahi, one of the city in West Java, had high incidence of DHF. It ranked first from year 2005 to 2007 and second in years 2008 and 2009. The Health Department of Cimahi so far has recorded the incidence number of DHF in each district of Cimahi. But, the data only is only used to obtain the SMP (*Sebelum Masa Penularan*). It is a very limited analysis because no observation of the factors that can cause the disease is ever taken.

In response to those problems, this study analyzes the relationship among the factors influencing the incidence of this disease and generate patterns of the relationship among the factors to predict the DHF conditions in the future. This study is expected to provide optimal patterns or models and accurate predictions so that the Health Department of Cimahi can take into action earlier in the future to prevent DHF spreading in this area.

## 1.2 Theoretical Framework

This study attempts to predict the DHF conditions in Cimahi for each district in the future. Firstly, the prediction system will search the optimal input for fuzzy logic includes the type and parameter of membership function, number of linguistic value and rules using genetic algorithm and fuzzy grid tables. Whereas, generating the input data for prediction is using the forecast method.

In this study, fuzzy logic is used because fuzzy logic can classify the data with fuzzy or blur, and not crisp so if DHF condition is red, it may not be fully red, it has membership values in range 0-1. Genetic algorithm is used because fuzzy logic system need optimal input to obtain optimal result if there is no information from an expert. Last, fuzzy grid tables are used to generate rules that will be the input to the fuzzy system.

## 1.3 Conceptual Framework/Paradigm

This sub chapter will discuss research variables and the relationship with the conceptual research. There are at least eight research variables applied in this study namely :

**Table 1-1 Research Variables**

No	Relationship with Conceptual Research	Variable
1	Class, output, or consequent for fuzzy system	Incidence Rate (IR)
2	The factors that stimulated DHF incidences (Antecedent for fuzzy rule)	Weather data, consist of : temperature, rainfall, solar radiation, relative humidity, and wind velocity
4	The factors that stimulated DHF incidences	Living Behavior, consist of : Using clean water, washing hands, using latrines, fighting larvae, vegetables and fruits consumption, physical activities, and no smoking.
5	Input for GA	Population size (pop size), Probability crossover (pc)
6	Output of GA / Input for fuzzy	Number of linguistic value , Type and parameter of membership function, Rules
7	GA evaluation variable / Optimal input for fuzzy system	Fitness value
8	Comparison of class target and class output	Training, testing or prediction accuracy

## 1.4 Statement of the Problem

Statement of the problems of this study are :

1. What are the correlations of IR and weather data ?

2. What are the correlations of IR and living behavioral ?
3. What is the optimal input to fuzzy system which influence the accuracy of prediction result ?
4. What is the optimal combination parameter of GA to predict DHF condition in the future based on learning data ?

### 1.5 Hypothesis

The hypothesis of this study are :

1. There are correlation of IR and weather data, because weather data is one of factors that influenced the growth of agent/vectors.
2. There are correlation of IR and living behavioral data, such as if the human has good living behavioral on any weather conditions then the DHF condition is lower.
3. The optimal input to fuzzy system may result in an accurate system to predict the spread of DHF.
4. There is the optimal combination parameter of GA to predict DHF condition in the future based on learning data.
5. The system can obtain the prediction accuracy more than 80%.

### 1.6 Assumption

The assumptions of this study are :

1. Living behavioral data is only available for two years, 2008 and 2009, but the data used for this study is the data from 2005 to 2009. So to bridge the gap, regression is used for each living behavioral data of each district because it is assumed that there is a trend of up or down every year.
2. Data for training and testing is labeled based on IR
3. Fitness value and accuracy of testing and prediction by comparing class target and class output
4. Number of linguistic value of DHF condition is 2 based on surveillance information and the rule of the Indonesia Health Departement
5. Minimum number of linguistic values of input variable is 2 and maximum is 3.
6. The membership function type that observe in this study only 3 type namely : Trapezium, Phi, and Triangle.

### 1.7 Scope and Delimitation

The scopes and delimitations of this study are :

1. This study is conducted based on the data taken from the Health Department of Cimahi
2. This study used weather and living behavioral data from 2005 to 2009 in each district of Cimahi
3. This study used ANN method to forecast the weather data but this study did not discuss in detail about ANN.

### 1.8 Importance of the Study

The importance of this study is the government especially the Health Department will be able to predict DHF conditions in each district earlier so that they can take more comprehensive actions to prevent the outbreak and reduce the number of DHF incidences (IR).

### 1.9 Definition of Terms

ANN	Artificial Neural Network, one of soft computing algorithm
Fitness value	Maximum accuracy of best individu/solution of GA
<i>SMP (Sebelum Masa Penularan)</i>	The month or time when a district or area has the lowest incidence so that prevention can be carried out and the month is the calculation result of the last 5 years data
IR (Incidence Rate)	The rate of DHF spreading based on standard of Indonesia Health Departement
Population size (pop size)	Number of candidate solution
Probability crossover (pc)	Probability of successfully recombination