

**FORECASTING MODEL OF ENERGY CONSUMPTION USING  
LEAST SQUARES SUPPORT VECTOR MACHINES**

By

ARVINDER SINGH A/L SARJIT SINGH

FINAL PROJECT REPORT

Submitted to the Department of Electrical & Electronic Engineering  
in Partial Fulfillment of the Requirements  
for the Degree  
Bachelor of Engineering (Hons)  
(Electrical & Electronic Engineering)

Universiti Teknologi PETRONAS

Bandar Seri Iskandar

31750 Tronoh

Perak Darul Ridzuan

© Copyright 2012

by

ARVINDER SINGH A/L SARJIT SINGH, 2012

## **CERTIFICATION OF APPROVAL**

### **FORECASTING MODEL OF ENERGY CONSUMPTION USING LEAST SQUARES SUPPORT VECTOR MACHINES**

by

Arvinder Singh a/l Sarjit Singh

A project dissertation submitted to the  
Department of Electrical & Electronic Engineering  
Universiti Teknologi PETRONAS  
in partial fulfilment of the requirement for the  
Bachelor of Engineering (Hons)  
(Electrical & Electronic Engineering)

Approved:

---

Dr, ROSDIAZLI IBRAHIM  
Project Supervisor

UNIVERSITI TEKNOLOGI PETRONAS

TRONOH, PERAK

September 2012

## **CERTIFICATION OF ORIGINALITY**

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

---

ARVINDER SINGH A/L SARJIT SINGH

## **ABSTRACT**

In the oil and gas industry, accurate measurement of gas is a very important aspect for the gas transmission operation. The outgoing gas flow during the transmission operation is monitored and maintained by a metering system. The metering system must be ensured reliable and dependable at all cost to maintain the billing integrity between distributors and customers. The main concern is products sold and returned as money worth product to seller and buyer. An existing system in Transmission Operation Division (TOD), PETRONAS Gas Berhad (PGB), Gurun is held responsible to calculate the energy consumption from the sales gas produced. The system consists of a turbine meter, measuring equipment which are pressure transmitter and temperature transmitter, gas chromatography and flow computer. However, the system is a standalone system that does not have any reference system to verify the integrity of it. Customers are billed according to the amount of energy consumption calculated and any error in calculation will cause loss of profit to the company and affect PETRONAS's business credibility. As a solution, a Least Squares Support Vector Machines (LS-SVM) prediction model of energy consumption is proposed as a verification system of the outgoing gas flow. The model will predict the energy consumption and compare it with the results of the existing metering system to ensure the reliability and accuracy of the system. The billing integrity between PETRONAS and the customers could be maintained and in the future if the project is expanded, it will have the potential of saving of millions of dollars to Malaysian oil and gas companies.

## **ACKNOWLEDGEMENT**

The author would like to address his greatest gratitude to the almighty for giving the strength to actually undertake this project and to complete it within the specified time.

The utmost gratitude goes to the author's supervisor, Dr. Rosdiazli Ibrahim for his continuous guidance and support throughout the project. The author would like to extend his gratitude to Mr. EleangeshwaranPathmanathan who was there throughout the critical part of the project and tirelessly guided the author despite his many other obligations. The author would like to thank UniversitiTeknologi PETRONAS (UTP) for giving him a chance to undertake research work as a life experience that is hard to get by. This learning experience was indeed a hands on experience. Lastly, the author would like to thank his friends and family for their tireless support throughout the whole project which had managed to bring to the success of this Final Year Project.

## TABLE OF CONTENTS

<b>ABSTRACT</b>	.	.	.	.	.	.	.	.	iv
<b>LIST OF FIGURES</b>	.	.	.	.	.	.	.	.	viii
<b>LIST OF TABLES</b>	.	.	.	.	.	.	.	.	ix
<b>CHAPTER 1:</b>	<b>INTRODUCTION</b>	.	.	.	.	.	.	.	<b>1</b>
	1.1. Chapter Overview	.	.	.	.	.	.	.	1
	1.2. Background Study	.	.	.	.	.	.	.	1
	1.3. Problem Statement	.	.	.	.	.	.	.	2
	1.4. Project Objectives	.	.	.	.	.	.	.	3
	1.5. Scope of Study	.	.	.	.	.	.	.	4
	1.6. Relevance of Project	.	.	.	.	.	.	.	4
	1.7. Feasibility of Project	.	.	.	.	.	.	.	5
<b>CHAPTER 2:</b>	<b>LITERATURE REVIEW</b>	.	.	.	.	.	.	.	<b>6</b>
	2.1. Chapter Overview	.	.	.	.	.	.	.	6
	2.2. Natural Gas	.	.	.	.	.	.	.	6
	2.3. Volume and Energy Calculation	.	.	.	.	.	.	.	7
	2.4. Forecasting models of gas consumption	.	.	.	.	.	.	.	8
	2.5. Artificial Neural Networks (ANN)	.	.	.	.	.	.	.	8
	2.6. Support Vector Machines (SVM)	.	.	.	.	.	.	.	9
	2.7. LS-SVM	.	.	.	.	.	.	.	10
	2.8. LS-SVM and ANN comparison	.	.	.	.	.	.	.	13

<b>CHAPTER 3:</b>	<b>METHODOLOGY</b>	<b>14</b>
	3.1 Chapter Overview	14
	3.1 Research Methodology and activities	14
	3.2 Gantt charts	17
	3.3 Key Milestones	20
	3.4 Tools and software used	21
<b>CHAPTER 4:</b>	<b>RESULTS AND DISCUSSION</b>	<b>22</b>
	4.1 Chapter overview	22
	4.2 Data preprocessing	22
	4.3 Selection of hidden neurons	23
	4.4 Selection of data division	24
	4.5 BPNN model development	26
	4.6 LS-SVM model development	27
	4.3 Comparison in performance	30
<b>CHAPTER 5:</b>	<b>CONCLUSION AND RECCOMENDATIONS</b>	<b>34</b>
	5.1 Chapter overview	34
	5.2 Conclusion	34
	5.3 Reccomendations	35
<b>REFERENCES</b>		<b>36</b>

## LIST OF FIGURES

Figure 1	Gas Metering Station	3
Figure 2	Natural Gas Vehicle	6
Figure 3	Typical BPNN network	9
Figure 4	Separating Hyper plane	10
Figure 5	Project Flow Activities	15
Figure 6	Chart results of different number of hidden neurons	24
Figure 7	Chart results of different data division	25
Figure 8	Energy Consumption training data for BPNN	26
Figure 9	Energy Consumption testing data for BPNN	27
Figure 10	LS-SVM Flowchart	28
Figure 11	Energy Consumption training data for LS-SVM	29
Figure 12	Energy Consumption testing data for LS-SVM	29
Figure 13	Chart result comparison between LS-SVM and BPNN	31



## LIST OF TABLES

Table 1	Comparison between LS-SVM and ANN	13
Table 2	Gantt chart for FYP 1	18
Table 3	Gantt chart for FYP 2	19
Table 4	FYP 1 milestones	20
Table 5	FYP 2 milestones	20
Table 6	Results of different number of hidden neurons	23
Table 7	Results of different data division	24
Table 8	Results for BPNN model	26
Table 9	Results for LS-SVM model	28
Table 10	Hyper parameter optimization steps for LS-SVM	30
Table 11	Result comparison between LS-SVM and BPNN	31
Table 12	1 <sup>st</sup> Gen and 2 <sup>nd</sup> Gen Comparison	31

# CHAPTER 1

## INTRODUCTION

### 1.1 Chapter Overview

In this chapter of the report, the background of the project which includes the basic information of the existing metering and gas distribution system will be included and discussed thoroughly. Then the problem statement which gives the insight of the current problem at the gas metering system will be discussed. This is indeed important due to the fact that we must understand the root problem properly.

After this is done, we will move on to the development of the LS-SVM energy forecasting model which will be assumed as the solution to that problem. Other than that, the objectives and the scope of study of the project will also be included in this chapter of the report. Last but not least will be the relevance of the projects and also the summary of this chapter.

### 1.2 Background Study

When we get involved in the oil and gas industry especially in gas transmission where every cent matters, we need to implement accurate measurement techniques. Natural gas can be transported using various techniques and in this project we would be dealing with pipelines as a medium for gas transportation to the customers. Once gas has been transported to the customer, the amount sent needs to be calculated of what we can say in technical terms is that it must be metered. This is done by employing a metering system that manages and monitors outgoing gas along the pipelines and these stations must be placed at suitable locations in order to effectively measure and monitor the gas flow at specific time and place.

Since natural gas is indeed a very valuable commodity, a reliable and dependable metering system needs to be employed no matter what because this will affect the profitability of a gas transmission facility.

PETRONAS Gas Berhad(PGB) has an existing metering system which is the Nur Metering Station located at Kulim . This station is responsible to calculate the energy consumption of the sales gas that is being produced. The metering system within the station consists of a turbine meter, measuring equipment such as pressure transmitter and temperature transmitter, gas chromatography and a flow computer. All these equipment will take form of the system and will calculate the energy consumption based on the multiple inputs. The inputs are basically pressure, temperature, gross volume, calorific value and lastly specific gravity.

The Nur Metering station needs to be reliable and accurate because it will affect the billing integrity between PETRONAS Gas Berhad(PGB) and its customer. So that is why we need to have an energy forecasting system to be developed which will have the function of a verification system for the billing. The model is to predict the outgoing gas flow from the facility based on the given inputs and to compare it with the results from the existing metering system. So to meet the objectives, the LS-SVM forecasting model is used.

### **1.3 Problem Statement**

The current metering station uses the American Gas Association (AGA) equations to calculate the energy consumption of the sales gas being transferred. In order to do the calculations, there are multiple inputs that are used to gain the energy consumption. These inputs are the pressure (P), temperature (T), gross volume ( $V_g$ ), calorific value (CV) and specific gravity (sg).

Under normal operating conditions, we do have to realise the fact that the pressure and temperature does vary within a certain span. This will cause change that will result in a difference between the calculated energy consumption and the actual energy consumption. When these two values do not tally, a critical error in billing of the transmitted gas to the customers will happen.

Since the current metering system that has been deployed is a standalone system, it does not have any other way to compare and verify the reliability and accuracy of the calculated energy consumption and the sales gas that has been transmitted. Since the

current billing is done based on the amount of energy consumption calculated, any error in calculation will give a different reading at the customer's side. As results, the integrity of the metering system employed by PETRONAS will be questioned and this not will only cause loss of profit to PETRONAS but also affect PETRONAS's business credibility.



Figure 1: Gas Metering System

In order to verify the equipment used for billing purposes, there is a dire need to conduct periodic validation. This periodic validation is set to be done twice yearly or yearly basis based on the contract that has been signed between PETRONAS Gas Berhad (PGB) and the customer. So in the intervals between the two different validations, there might be changes in the readings from the devices such as drift in readings and other issues such as flow computer giving anonymous readings. That is why there is a need for a possible solution to be developed in order to verify the billing that is done by the existing metering system.

#### **1.4 Project Objectives**

This project is collaboration between University Technology of PETRONAS (UTP), Group Technology Solution (GTS) and Transmission Operation Division (TOD), PETRONAS Gas Berhad (PGB), Gurun. The objectives of this project are listed as below:

- To develop a LS-SVM model of energy consumption that is based on a few process variables such as temperature, pressure, flowrate and characteristics of natural gas.
- To develop a LS-SVM model of energy consumption as a verification tool of the outgoing gas flow.
- To develop a LS-SVM prediction model that will produce accurate readings that is less than one percent error of energy prediction.

### **1.5 Scope of Study**

In this study, the LS-SVM model of energy consumption will be used as a verification tool to compare the billing integrity of the existing metering system that has been commissioned by TOD, PGB, Gurun . There will be 5 inputs which are pressure, temperature, gross volume, calorific value and specific gravity while the output is energy. The model is to predict the energy consumption of the system based on all the 5 given inputs and the predicted energy will be compared with the results of the actual energy.

So in this project, the LS-SVM algorithm will be used to develop a forecasting model of energy consumption. This algorithm is MATLAB based and from the inputs and outputs relation, the model can be successfully derived. The new outputs will then be predicted based on the trained data earlier.

### **1.6 Relevance of Project**

This project is proposed with the main aim of ensuring the reliability and accuracy of the existing metering facility in PGB, Gurun. This is done in line of maintaining billing integrity between PGB and the customers which is important because gas transmission is the core business of PGB. If it is done correctly, PETRONAS will get to maintain its business credibility and a major loss in company's profit can be

prevented certainly. This will also give the customers more confidence in trust when dealing with PETRONAS.

### **1.7 Feasibility of Project**

This particular project that is based on the actual situation in TOD, PGB, Gurun and a preliminary investigation has been done by an engineer from PGB. The idea of the investigation is to find the most suitable and promising method to be proposed. Based on the literature review, LS-SVM is found suitable for the project.

The project is scheduled to be completed within two semesters. The project can be completed within the scope and the timeline as long as the individual responsible for the project gives full commitment and strive to get the most optimum results and present a successful project at the end.

## CHAPTER 2

### LITERATURE REVIEW

#### 2.1 Chapter overview

This chapter outlines the theories and also the literature review of this project which covers natural gas, transportation of natural gas, volume and energy calculations of natural gas, forecasting models of gas consumption, artificial neural networks (BPNN), Support Vector Machines(SVM) and also Least Squares Support Vector Machines( LS-SVM).

These were taken from books, journals, conference papers and many other publications that are related to the topic in order to gain a better understanding of the project.

#### 2.2 Natural Gas

Natural gas is a mix of various hydrocarbons that consist of methane, ethane, propane and butane. There are also traces of hydrogen sulfide and nitrogen in natural gas.[1] Methane has the highest composition in natural gas which is 82% meanwhile ethane, propane and other heavier hydrocarbons are only about 19%. The small quantities of hydrogen sulfide and nitrogen are negligible.[2]Natural gas has been accepted widely as an alternative source of fuel and has been promoted as a clean source of fuel. It is used as energy source for vehicles and even power plants that use gas turbines. [3]



Figure 2: Natural Gas Vehicle (NGV)

### 2.3 Volume and Energy Calculation

Natural gas is usually being sent in bulk using pipelines. The volume of the gas sent is calculated from the metered volume readings, pressure and temperature measurement and gas composition data. Usually when we want find the energy consumption of the sales gas; we often calculate using the equation below which is based on ISO 6979 and the AGA Report No. 8 that describes the compressibility factor standard and analysis. This equation is for calculating hourly volume of gas (standard cubic meter) for each hour

$$V_h = \sum_h \left( \frac{V_{g(meas)} \times P_{meas} \times T_{base} \times Z_{base}}{P_{base} \times T_{meas} \times Z_{meas}} \right)$$

As we know, the energy measurement is one of the important part in the gas industry where every cent counts. The amount of energy contained in the natural gas is determined on far greater effort and is more complex than the billing of electrical energy. This is because this involves fluid mechanics and also calorific properties of natural gases have been taken into account[4]. The equation that is usually used for hourly calculation of energy flow (in GJ) is as the one below.

$$E_h = \frac{(V_h \times CV)}{1000}$$

$E_h$  =Energy consumption per hour (GJ)

$V_h$  =Volume of gas per hour (Sm<sup>3</sup>)

$CV$  =Calorific value (mJ/Sm<sup>3</sup>)

$V_{g(meas)}$  =Measured gas volume (m<sup>3</sup>)

$P_{meas}$  =Measured pressure (kPa)



$P_{base}$  =Base pressure (kPa) (101.325 kPa)

$T_{meas}$  =Measured temperature (K)

$T_{base}$  =Base temperature (K) (288.15K)

$Z_{meas}$  =Compressibility factor at  $P_{meas}$  and  $T_{meas}$  calculated by reference to AGA Report No. 8

$Z_{base}$  =Compressibility factor at  $P_{base}$  and  $T_{base}$  calculated by reference to AGA Report No. 8

## 2.4 Forecasting models of gas consumption

Natural gas distributors do need to predict the amount of energy consumed because meters at the customer side are rarely read. Estimates are usually based on the previous month's consumption patterns[5]. There are many methods to forecast the natural gas consumption such as Neural Networks, LS-SVM, genetic methods and also grey models[6]. In this paper we will be focusing on LS-SVM and neural networks will be made as a benchmark.

## 2.5 Artificial Neural Network(ANN)

Artificial Neural Networks (ANN) is regarded as one of the modern method of forecasting and is highly favorable because it has a clear model, an easy implementation and has a certainly good performance[7]. It has the ability to approach non-linear systems.

However in highly complex system, its ability in approximation will be limited by the algorithm of the neurons[8]. This ANN actually imitates the way the human brain works and is a great tool in classification, pattern recognition and forecasting. The typical ANN is basically made of combination of layers of neurons and the most commonly used ANN for forecasting is multi layers perceptron (MLP) [9].

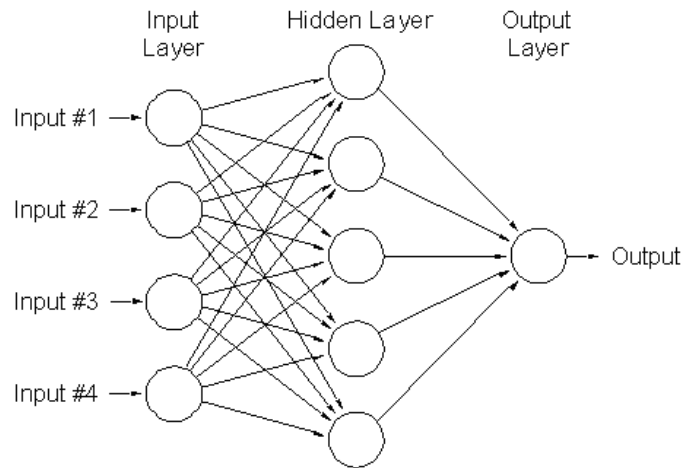


Figure 3: Typical BP-NN network.

The ANN used in this project as a benchmark is the Back –Propagation Neural Network. An ANN using this method will conduct supervised learning and the system is supplied with the set of patterns for it to learn and also the desired system response. This method has many advantages and one of them is when it has been trained with the initial set of patterns, their recognition of similar patterns will be done quickly[10].

The Back Propagation Neural Networks (BPNN) have many limitations and this covers a range from dangers of over fitting right up till difficulty of obtaining stable solutions. This somehow proves that BPNN unreliable even though all the parameters have been determined earlier beforehand.[11]. This is why in the later part; LS-SVM will be used to improve the performance of ANN.

## 2.6 Support Vector Machines(SVM)

All this while support vector machines (SVM) have been used a method to solve many pattern recognition problems because it is known to be a very powerful classifier .[12]SVM was first proposed by Vapnik et al. and since then it has been related deeply to the field of statistical learning. This algorithm has been used in many areas in the present day such as in the energy demand forecasting [13][14]

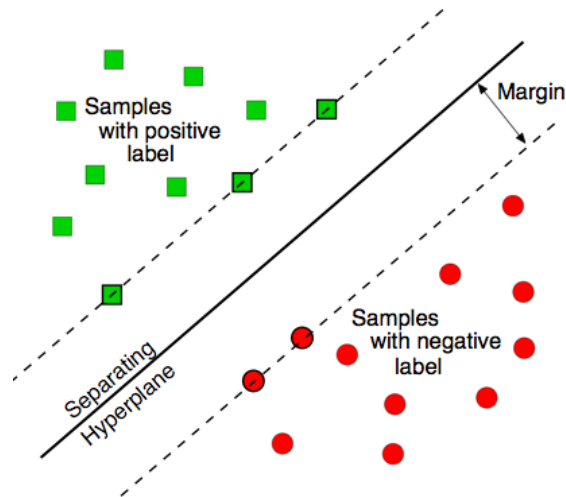


Figure 4: Separating Hyperplane

The difference between BPNN and SVM is that SVM aims in minimizing an upper bound of the generalization error through maximizing the margin between the separating hyper plane and the data on the structure based on structure risk minimization principle. By doing so SVM is able to overcome certain weak points of BPNN such as local minimal solutions and the selection of the number of hidden units.[15] In the case of non-linear problems, SVM uses kernel tricks to produce non-linear boundaries where the kernel is to maps low dimensional data at a high dimensional space via a mapping function and also constructs a separating hyper plane to maximize the margin.[12][16] Radial basis function (RBF) is the most common kernel and the kernel parameters must be determined before training.[11]

## 2.7 Least Squares Support Vector Machines(LS-SVM)

SVM has been applied for some time successfully but there were a few problems that occurred. These problems happen when the number of training vector increases. When the number of training vectors increase, the storage (memory) requirement will increase pretty much rapidly.[17] A perfect description would be when one tries to use large data sets to model for SVM, problems start to occur. A simple explanation behind it would be in order to solve quadratic optimization problems, SVM needs resources that are at least a square or what we can say the power of two to the number of training samples. This will cause the training time and memory that is needed to increase by quite a lot when the data set increases in size[18]

These issues were pretty much a problem and has to be dealt with. So in order to solve this issue, LS-SVM was made a solution. LS-SVM does it by solving the quadratic problems by converting the inequality constraint in SVM to equality ones. By doing so we don't need to solve QP problems and can focus on solving linear systems only.[14] This will directly enhance the performance of the model.

LS-SVM is said to have a smaller calculation complexity and is capable of performing faster calculations when it is compared to SVM. LS-SVM can be used to solve pattern recognition problems because it is created based on SVM. The formulation for LS-SVM starts when we take a training set and it is to be estimated based on non-linear function as in where  $\phi(\cdot): R^n \rightarrow R^{n_h}$  is a mapping function to a high dimensional and potentially infinite dimensional feature space. Next, the optimization is formulated in primal weight space;

$$\min \Phi(w, e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^l e_i^2$$

Subject to;

$$y_i = w^T \phi(x_i) + b + e_i \quad i = 1, \dots, l$$

The optimization formulation in Eq. 13 is ridge regression cost function formulated in feature space. By constructing the Lagrangian of the problem, the dual problem is derived;

$$L(w, b, e, \alpha) = \Phi(w, e) - \sum_{i=1}^l \alpha_i \{w^T \phi(x_k) + b + e_i - y_i\}$$

The conditions for optimality are given by as below;

$$\begin{aligned}
\frac{\partial L}{\partial w} = 0 &\rightarrow w = \sum_{i=1}^l \alpha_i \phi(x_i) \\
\frac{\partial L}{\partial b} = 0 &\rightarrow \sum_{i=1}^l \alpha_i = 0 \\
\frac{\partial L}{\partial e_i} = 0 &\rightarrow \alpha_i = \gamma e_i \quad i = 1, \dots, l \\
\frac{\partial L}{\partial \alpha_i} = 0 &\rightarrow w^T \phi(x_i) + b + e_i - y_i \quad i = 1, \dots, l
\end{aligned}$$

When we eliminate the variables  $w$  and  $e$  and solve in  $\alpha$  and  $b$  we get the following solution in dual space

$$\begin{bmatrix} 0 & \mathbf{1}_v^T \\ \mathbf{1}_v & \Omega + \mathbf{1}/\gamma \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}$$

Where  $y = [y_1; \dots; y_l]$ ,  $\mathbf{1}_v = [1; \dots; 1]$ , and  $\alpha = [\alpha_1; \dots; \alpha_l]$  The “kernel trick” is applied here as shown below;

$$\begin{aligned}
\Omega_{ij} &= \phi(x_i)^T \phi(x_j) \\
&= K(x_i, x_j) \quad i, j = 1, \dots, l
\end{aligned}$$

So the resulting LS-SVM model is;

$$f(x) = \sum_{i=1}^l \alpha_i K(x_i, x_j) + b$$

In this paper, RBF kernel function is used;

$$K(x_i, x_j) = \exp\left(-\frac{|x - x_i|^2}{\sigma}\right)$$

The LS-SVM formulation can be used to handle large datasets with no dimensionality problem. Kernel functions are used to replace the high order mapping function. In this report, the RBF kernel will be used exclusively for all computations involving kernel

operations. In the case of RBF kernel, only two hyper parameters need to be tuned ( $\gamma$ ,  $\sigma$ ), which is lesser than standard SVM.[11]

## 2.8 Comparison between LS-SVM and ANN

The ANN method chosen as mentioned earlier in this report is the Back Propagation Neural Network (BP-NN) and is used as a benchmark algorithm to be tested against the LS-SVM algorithm. So a few papers were reviewed and a comparison was made based the train and testing error difference. The results are shown as per in the table below;

No	Brief Title	BPNN error %		LS-SVM error %		Ref
		Train	Test	Train	Test	
1	CO2 Emissions Model	7.29	6.8	1.24	5.26	[11]
2	Self-Adaptive	9.39		2.78		[20]
3	Elec Load Forecast	14.186		5.691		[21]
4	Hard Landing	4.2		3.7		[22]
5	Building cool (MAERR)	4.18		1.65		[23]

Table 1: Comparison between LS-SVM and ANN

So based on the table of comparison, it can be said that LS-SVM should be tried for this data set to prove that it can produce a better performing model .

## **CHAPTER 3**

### **METHODOLOGY**

#### **3.1 Overview**

This chapter of this report outlines clearly the methodology or the guidelines of solving the issued discussed in the problem statement in the earlier part of this report.

This chapter is to have 4 main parts which are the Research Methodology and Project Activities, the Gantt charts for both FYP 1 and FYP 2, the key milestones of the project which marks the important times of the project and the software and tools used in order to complete this project.

#### **3.2 Research Methodology and Project Activities**

This part of the report basically contains the research methodology and project activities. The research methodology is basically divided to five different stages which are the preliminary research work by going through publications, the development of the Neural Network model using MATLAB, the development of the LS-SVM model using MATLAB, the testing and validation works of both models and finally is the part where will have the results and discussion and not to forget most importantly is the final documentation.

The Project Activities on the other hand is the activities that were discussed within the research methodology i.e. what are we doing at that point of the research methodology. The project activities are represented in the form of a diagram as per in the next page. So after the diagram we will be looking at the first part which is the preliminary research work.

The project work flowchart is as per in the following page

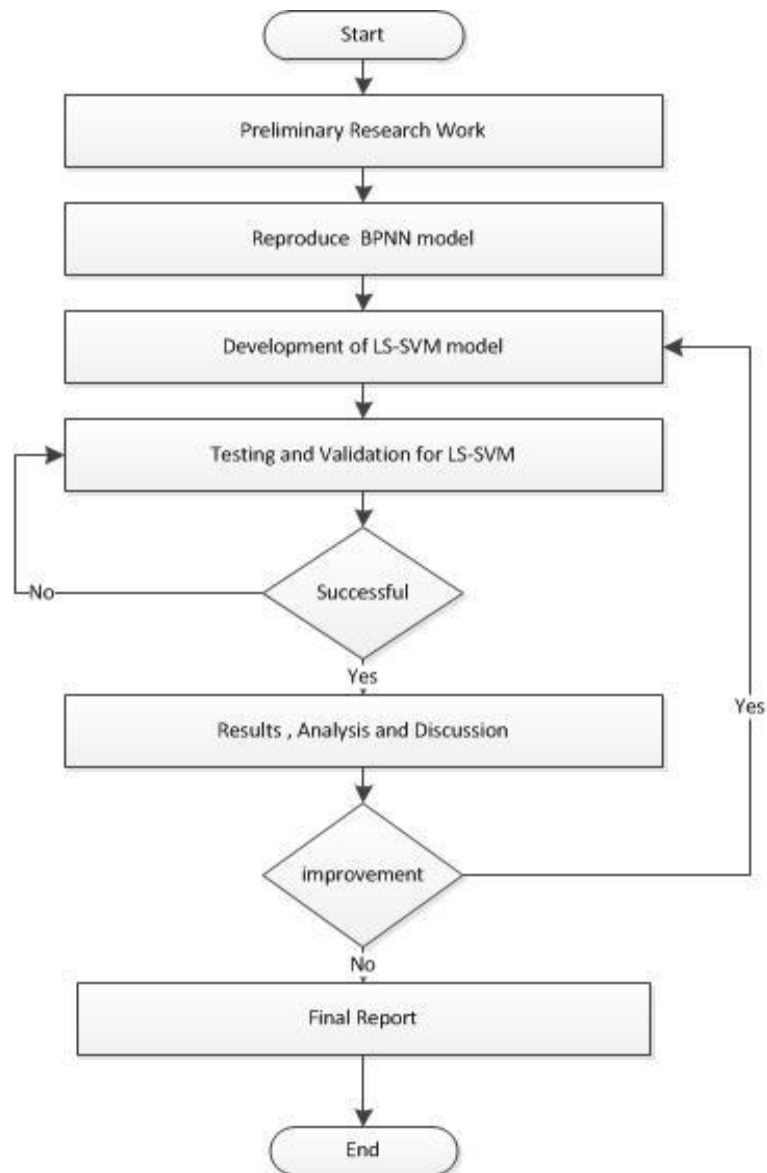


Figure 5: Project activities flow

### **Preliminary research work**

Before the proper work in the project can be carried out, a proper understanding of the topic must be gained and this is to be done by doing preliminary research work. Preliminary research work basically consists of going through journals, research papers, books and other publications that are available both online and in the Information Resource Center (IRC). This is done in order to gain a better understanding of the project. The later parts would be learning how to use both the



BPNN and the LS-SVM algorithm properly for MATLAB modeling and understanding what results we should be getting.

### **Development of Neural Network model**

In this part of the project we would be talking of the development of the Neural Network Model or to be specific it is the development of the Back-Propagation Neural Network (BP-NN). By using a few chosen values like number of hidden neurons and the data division, we will try to get the best parameters to be used in order to develop the best performing model.

After that, the BP-NN MATLAB model will then be used as a benchmark for the LS-SVM model. This BP-NN model will be remodeled several times to the point where we will get the most desired results that will be used as a comparison in the results and discussion part.

### **Development of LS-SVM model**

The development of the LS-SVM model starts right after the finalized BP-NN model has been obtained. This LS-SVM model will be developed using the LS-SVM toolbox which runs on MATLAB and from here we will get the prediction model of energy consumption of sales gas. Many trials are done in order to get the best model in terms of comparable results to the BP-NN model's results.

### **Testing and Validation**

This part of the project is only started when the BP-NN and the LS-SVM models are fully developed. The testing and validation of the models are done when the data set is divided in the algorithm itself into 3 main parts which are the training data, testing data and also the validation data in the MATLAB software itself. This part of the project flow is much related to the development of the BP- NN and the LS-SVM model because it is done at around the same time.

If the results are not satisfactory, the model will have to be retrained in order for us to get a better performing model. If all the results are satisfactory, then only we can move to the next part of this report which is the analysis, discussion and also the final documentation.

## **Results, analysis and discussion**

Once the testing and validation are successful; the results can be analyzed and then later be discussed for further evaluation. So in this part of the project, the results will be discussed thoroughly. This will cover from the BP-NN benchmark algorithm results right until the LS-SVM results and its comparison. Then when this is finished, the final documentation can be done.

## **Final Documentation**

Once the results have been discussed thoroughly, it is now the part where it will be compiled in the form of the final documentation which is the dissertation and also possible the technical paper.

### **3.3 Gantt charts**

In this part, the Gantt charts for both FYP 1 and FYP 2 will be displayed in the next few pages where it will show the projects timeline. The starting weeks and agendas and also the finishing weeks and its agendas will be shown in the form of tables. The schedules of the activities are not finalized and can change from time to time depending on the progress achieved on the project.

So far as in this project, everything is up to date and is following the progress as per in the Gantt chart

**FYP 1**

No.	Detail/ Week	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Selection and Confirmation of Project Title														
2	Preliminary Research Work on Related Topics														
3	Submission of Extended Proposal						●								
4	Study on MATLAB Software for Model Development														
5	Proposal Defence														
6	BPNN Model Development in MATLAB Software														
7	Submission of Interim Draft Report													●	
8	Submission of Interim Final Report														●

Table 2: Ganttchart for FYP 1

## FYP 2

No.	Detail/ Week	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	LS-SVM Model Development	■	■	■	■	■	■								
2	Testing and Validation work							■	■						
3	Submission of Progress Report								●						
4	Results, Analysis and Discussion							■	■	■					
5	Pre-EDX										●				
6	Submission of Draft Report											●			
7	Submission of Dissertation (softbound)												●		
8	Submission of Technical Paper												●		
9	Oral Presentation													●	
10	Submission of Project Dissertation(softbound)														●

Table 3: Ganttchart for FYP 2

### 3.4 Key milestones

The key milestones are used in order to mark important dates in the whole work progress and also the end stage of the project. This is indeed very important due to the fact that we need to monitor the progress very closely to make sure that the project is right on schedule.

The key milestones are shown as per in the two tables below which are table 3.4 and table 3.5

FYP 1															
No.	Detail/ Week	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Completion of preliminary research work														
2	Completion of Interim Report														

Table 4: FYP 1 Milestones

SECOND SEMESTER															
No.	Detail/ Week	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Completion of model development														
2	Completion of testing and validation work														
3	Completion of results analysis and discussion														
4	Completion of final documentation														

Table 5: FYP 2 milestones

The key milestones that were aimed as per in the first semester that covered the preliminary research work and also the completion of the preliminary report were achieved successfully.

As for the second semester, everything managed to be achieved right up to the submission of this final report and the technical paper. The other work can be done in time as long as the progress is continuously monitored and the project is made on schedule

### **3.5 Tools and Software used**

The software used in this project would be MATLAB 2010a and Microsoft Office 2010. MATLAB is used basically to run both the BP-NN and LS-SVM toolbox while Microsoft Office 2010 is used for the preparation of the reports. Since this is a simulation project, there are no physical tools that are being used.

## CHAPTER 4

### RESULTS AND DISCUSSION

#### 4.1 Chapter Overview

This chapter basically covers the results that we have gained up till recently. This results that contain issues regarding to the project will be further discussed in this part of the report. Since this project is to develop a LS-SVM energy consumption model and have the BPNN as a benchmark, we would be getting results from both algorithms and performance of both algorithms will be discussed and a comparison will be made.

In this chapter, we will cover sub-chapters that will bring to the development of the final LS-SVM model. This consists of data preprocessing, selection of number of hidden neurons for BPNN algorithm, selection of data division, developing the model using BPNN algorithm ,developing the model using LS-SVM algorithm and finally the comparison in terms of performance.

#### 4.2 Data Preprocessing

For this project, the data that we would be using is obtained from the Nur Metering station as it has been explained in the introduction part of this project. This metering station is held responsible to meter outgoing gas flow from PETRONAS Gas Berhad (PGB). So there are 5 main input variables and 1 output that consist of 3446 data points that have been take in hourly basis at the Nur Metering station from Dec 2010 to April 2011. These data points have been shifted by one day for forecasting purposes.

For this project, before we can actually do the modeling,we need to remove all of the bad inputs and non-values. After that is done then only we can use both data sets for modeling and algorithm testing purposes. So the first thing we would do now is the process of selection of the number of hidden neurons for the BPNN algorithm.

### 4.3 Selection of hidden neurons

Since we would be using the BP-NN model as a benchmark, there is a need to set up the basic parameters that will come up with the best results. This can be done by first getting the optimum number of hidden neurons for the algorithm to work the best with this particular data set.

Before we can determine the best number of hidden neurons, there would be a need to set the data division as a constant variable. So in this case is we will set that 70% of samples will be used for training purposes, 10% for validation, and 20% for testing. The total amount of data points that will be used are 3446 data points which is the whole data set for this purpose.

Once all the constants are set, then only different number of hidden neurons can be tested. The numbers of hidden neurons that are tested 10, 15, 20, and 25 with a constant data division of 0.7:0.1:0.2. The results that were collected are presented in the table below

Hidden neurons	Train error %	Test error %	Time (s)
10	1.996	2.336	9.484
15	2.222	2.116	5.214
20	1.935	1.740	17.74
25	2.015	1.847	9.386

Table 6: Results of different number of hidden neurons

By analyzing the results in the table, it is obvious that the time taken for the whole process does not differ much because it is only a matter of a few seconds. Due to this, the time taken won't be taken into consideration for analysis. There would only be focus on the train error percentage and test error percentage. For a better analysis, the data is presented in the form of chart as follows.



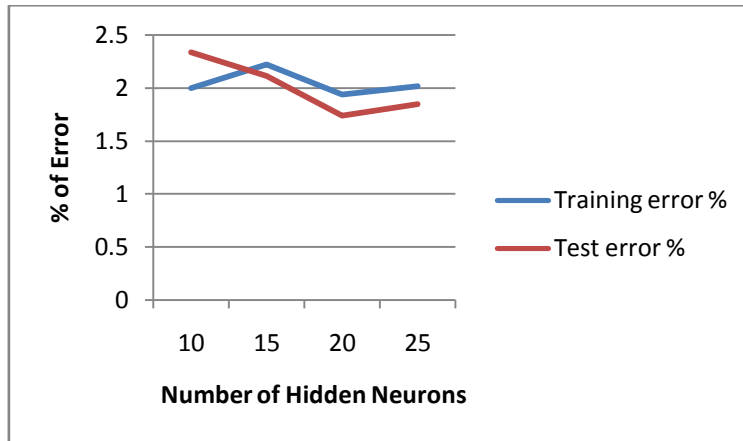


Figure 6: Chart results of different number of hidden neurons

So from what we can see from the results of the graph, the best number of hidden neurons to be chosen for this data set is 20 hidden neurons. This is because it has the lowest training and testing error percentage.

#### 4.4 Selection of data division

The numbers of hidden neurons have already been chosen in the previous section and it is 20 hidden neurons. In this part of the report, we will be having the number of hidden neurons constant and the data division varied. From this we will get the best possible combination of data division for this data set.

The data division will be arranged according to training data:validationdata:testing data with a few choices as per the table below. The performance criteria will be the training error percentage and also the testing error percentage. Results as per table.

Option	Data Division ratio			Train error %	Test error %
	Train	Valid	Test		
1	0.3	0.1	0.6	1.900	2.094
2	0.4	0.1	0.5	2.309	2.267
3	0.5	0.1	0.4	2.264	2.244
4	0.6	0.1	0.3	2.224	2.030
5	0.7	0.1	0.2	1.935	1.740

Table 7: Results of different data division

By looking at the results of the table we cannot see much difference but it is very obvious that option 6 has the best results in terms of train error percentage and also tests error percentage. Even though option 2 performs slightly better than option 6 in terms of train error, but its test error is still higher. This can be seen well in the form of the graph as per below.

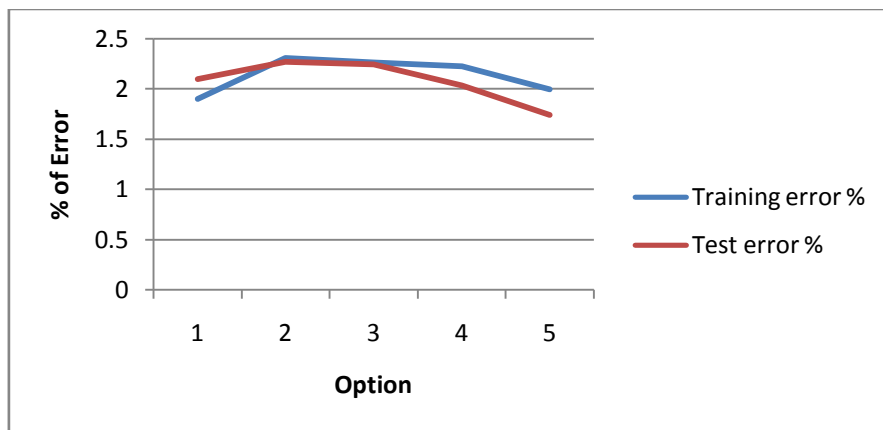


Figure 7: Chart results of different data division

So from the graph above it is very obvious that the best performing option is option 5 with the data division of 0.7:0.1:0.2 where training : validation: testing . From section 4.3 and 4.4, we can conclude that the parameters to be used for the BP-NN algorithm are;

- Data division = 0.7:0.1: 0.2 (train:validation:test)
- Number of layers = 2
- Number of hidden neurons = 20
- Input layer transfer function = tan-sigmoid ( set in algorithm)
- Hidden layer transfer function = linear ( set in algorithm)
- Training algorithm = Lavenberg-Marquat ( set in algorithm)

## 4.5 BP-NN model development

In this part of the report, we would be using all, the data points which are all the 3446 points. The performance will be measured and will be presented in the table form. This will be done by developing the forecasting model of gas consumption using the BPNN- algorithm in MATLAB.

So after running the BP-NN model, we will get the results which are presented in the form of table that consists of different performance measuring parameters as per in the table below. The computational time of the model will be included too in this report as a performance measure.

	Accuracy	RMSE	R	Time
Training data	98.07	0.7048	0.8618	11.3 sec
Testing data	98.26	0.6575	0.8776	

Table 8: Results for BPNN model

It can be considered that the performance of the BP-NN model is very good due to its ability to produce training and testing error less than 2%. The table above only shows the values of the performance indicators. The figures in the next page will show the distribution of the real and predicted s in the form of graph that were obtained during the development and testing of the model.

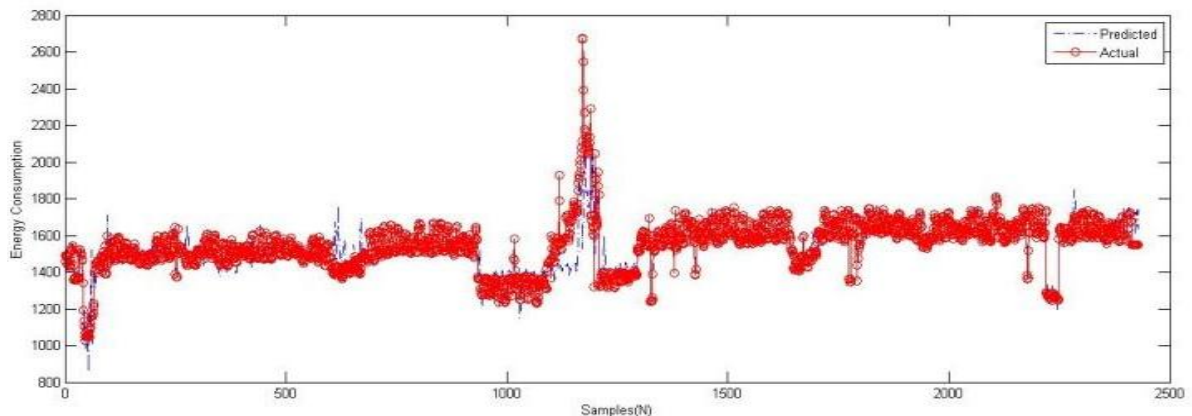


Figure 8: Energy Consumption training data for BPNN

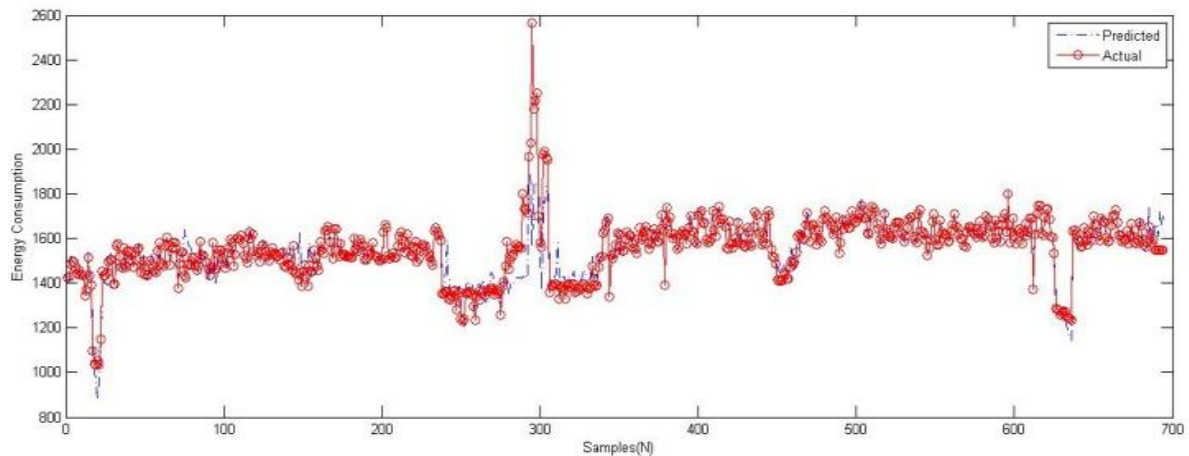


Figure 9: Energy Consumption testing data for BPNN

So from the actual and predicted s for both the training and testing data, we can visually say that there is not much of error.

#### 4.6 LS-SVM model development

Once we have already done the BP-NN model, it is now time to do the modeling using the LS-SVM algorithm. The amount of data points will be same which 3446 is also and the data division used is 0.8:0.2 which is cross validation: testing. The performance will be measured and will be presented in the table form. This will be done by developing the forecasting model of gas consumption using the LS-SVM algorithm in MATLAB. A clearer picture of how the LS-SVM algorithm works will be shown in the following flowchart.

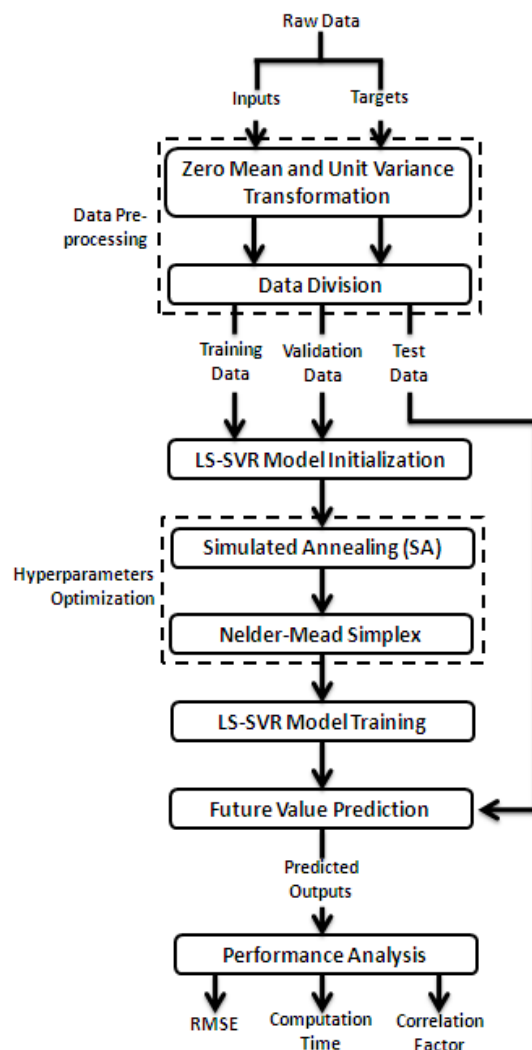


Figure 10: LS-SVM Flowchart

So after running the LS-SVM model, we will get the results which are presented in the form of table that consists of different performance measuring parameters as per in the table below. The computational time of the model will be included too in this report as a performance measure.

	Accuracy	RMSE	R	Time
Training data	98.73	0.5044	0.9297	934.3 sec
Testing data	98.59	0.5732	0.9085	

Table 9: Results for LS-SVM model

It can be considered that the performance of the LS-SVM model is very good due to its ability to produce training and testing error less than 2%. The table above only shows the values of the performance indicators. The figures below will show the

distribution of the real and predicted s in the form of graph that was obtained during the development and testing of the model. After that, there will be shown the hyper parameter optimization steps for energy consumption for the LS-SVM model in the form of a table and this will be followed by an analysis of the results for this part of the report.

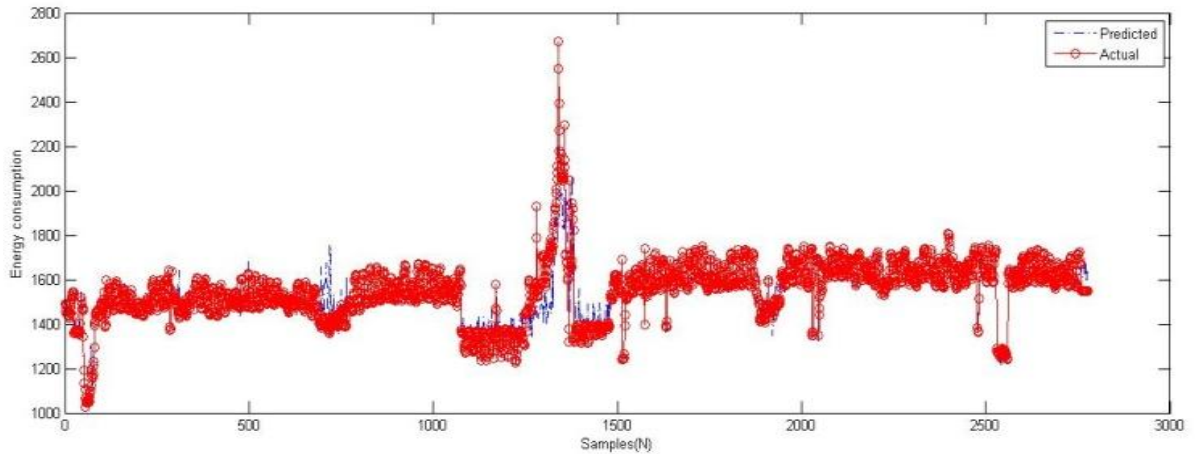


Figure 11: Energy Consumption training data for LS-SVM

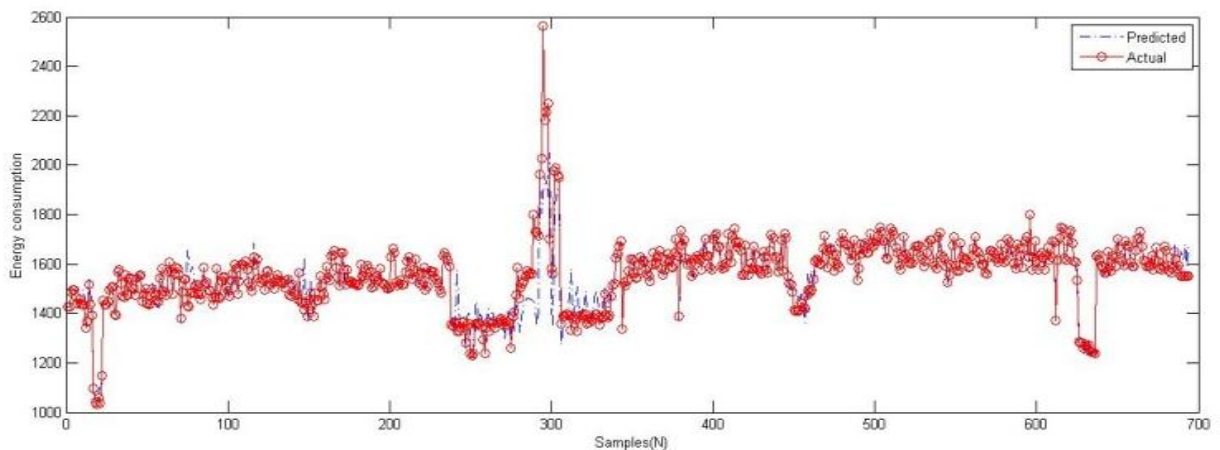


Figure 12: Energy Consumption testing data for LS-SVM

So from the actual and predicted s for both the training and testing data, we can visually say that there is not much of error.

Iteration	Function evaluation	min f(x)	log(y)	log( $\sigma^2$ )	Procedure
1	3	4.242217e+003	3.2340	-1.0868	Initial
2	7	4.242217e+003	3.2340	-1.0868	Shrink
3	9	4.242217e+003	3.2340	-1.0868	Contract inside
4	10	4.242217e+003	3.2340	-1.0868	Reflect
5	12	4.242217e+003	3.2340	-1.0868	Contract inside
6	14	4.229819e+003	2.8965	-1.1618	reflect
7	16	4.229819e+003	2.8965	-1.1618	Contract outside
8	17	4.229819e+003	2.8965	-1.1618	Reflect
9	18	4.229819e+003	2.8965	-1.1618	Reflect
10	22	4.228449e+003	2.8168	-1.2462	Shrink
11	24	4.228449e+003	2.8168	-1.2462	Contract inside
12	28	4.228074e+003	2.7723	-1.2228	Shrink
13	30	4.228074e+003	2.7723	-1.2228	Contract inside
14	34	4.228072e+003	2.7679	-1.2362	Shrink
15	36	4.228035e+003	2.7823	-1.2320	Reflect
16	40	4.228034e+003	2.7773	-1.2274	Shrink
17	42	4.228027e+003	2.7845	-1.2252	Reflect
18	46	4.228025e+003	2.7834	-1.2286	Shrink
19	48	4.228022e+003	2.7870	-1.2276	Reflect
20	50	4.228022e+003	2.7870	-1.2276	Contract inside

TABLE 10: Hyper parameter optimization steps for LS-SVM

So by looking at the figures that covered the real and predicted  $s$  and also the hyper parameter optimization steps, we can say that the values of the hyper parameters selected by the LS-SVM algorithm greatly affect the performance of the LS-SVM model. The initial values that were set for this case are basically;

$$\gamma = 25.3809$$

$$\sigma = 0.5808$$

These two values were computed using simulated annealing (SA) and was further optimized using the Nedler-Mead simplex algorithm. So this LS-SVM can be said to use the Nedler Mead- Simulated Annealing algorithm or in short NM-CSA. In this case the objective is to minimize the model error (MSE) which is based on a few limitations which are the maximum number of optimization steps, maximum number of function evaluations, stopping criterion based on the relative change in value of the function in each step and stopping criterion based on the change in the minimizer in each step.

From the simplex that was shown in the earlier table we managed to further optimize the values. That is where we managed to get optimized hyper parameters which are as below;

$$\gamma = 16.2326$$

$$\sigma = 0.5413$$

#### 4.7 Comparison in performance between LS-SVM and BP-NN

Both the models for BP-NN and LS-SVM have been developed and the results have already been presented in the earlier sections. One of the results is the energy consumption in the graphs that shows highly non-linear distribution. From this it shows that the model is capable to map nonlinearities with precision and also accuracy. The performance difference between the BP-NN and LS-SVM model in terms of accuracy is mapped in the table and chart below.

Algorithm	Train error %	Test error %	Time taken
BP-NN	1.93	1.74	11.3
LS-SVM	1.27	1.41	934.2

Table 11: Result comparison between LS-SVM and BPNN

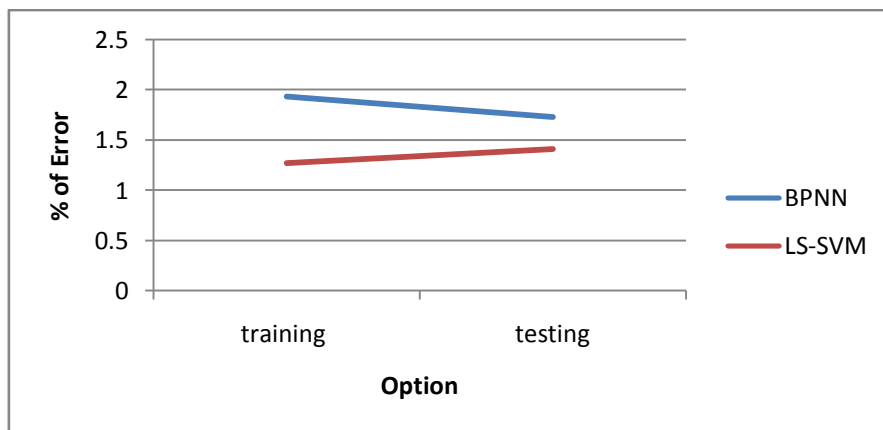


Figure 12: Chart result comparison between LS-SVM and BPNN

As we can see from both the table and graph, the error produced from the LS-SVM and BP-NN are both less than 2% which means that it has a very high accuracy and can be used in the real world applications.



Since this project is focusing on LS-SVM , it can be said that LS-SVM is a far better performing algorithm for this data set as the percentage error for both training and testing data is very much less. The time taken to perform the computations is very much longer. This can be overcome in the future by using a better PC to perform the computations.

The earlier comparison does not provide enough evidence that LS-SVM is a better performing algorithm. In order to provide further justification that LS-SVM is a better performing algorithm compared to BP-NN for this data set, the results gained in this project is to be compared with the previous student that has been using the same dataset but focusing on various different type of neural networks. The best results from the previous FYP will be made as the first generation results and the current project would be dubbed as the second generation results. The main criterion that will be used would be the percentage training and also the percentage of testing error. The data would be presented in the form of tables below for better analysis.

	Train error %	Test error %
1 <sup>st</sup> Generation (NN)	2.7	2.8
2 <sup>nd</sup> Generation (BP-NN)	1.93	1.74
2 <sup>nd</sup> Generation (LS-SVM)	1.27	1.41

Table 12: 1<sup>st</sup> Gen and 2<sup>nd</sup> Gen Comparison

So by looking at the table above , it is obvious that the 2<sup>nd</sup> generation neural networks (BPNN) performs far much better than the best results that were managed to be gained from the 1<sup>st</sup> generation neural networks(NN) in terms of both the percentage of training and testing error. This shows that there is significant improvement done to the student's previous work.

The accuracy of the neural networks was felt not satisfactory enough and was then further improved by implementing the LS-SVM algorithm to produce a better performing model and the results were proven much better for the LS-SVM model because there was a reduction of both training and testing error. So to be said , the results of this project is indeed good because it managed to meet the aim of proving

that the LS-SVM algorithm is indeed a better performing algorithm for this data set which is to make the forecasting model of energy consumption.

## CHAPTER 5

### CONCLUSION AND RECCOMENDATIONS

#### 5.0 Chapter Overview

This chapter of the report will be discussion the conclusion and the recommendations that are to be put forward based on the research, findings and also the research that has already been done. The recommendations are put forward as suggestions for further development of this project

#### 5.1 Conclusion

Developing an accurate forecasting model of energy consumption is indeed very important in order to get a good prediction which is not only accurate but also very much precise. This in whole will help PETRONAS Gas Berhad (PGB) to maintain its billing integrity with its clients and losses in terms of error in measurement can be prevented

There were two forecasting models of energy consumption that were developed using two separate algorithms which are BP-NN and LS-SVM. This modeling was done on MATLAB . Both the BP-NN and LS-SVM models have been successfully constructed and have been tested with the same environment i.e. data division and etc.

As a result, models with very high accuracy(less than 2% error) have been obtained for both LS-SVM and BP-NN. LS-SVM in whole was a better performing model with both the training error percentage and also testing error percentage was lower than the BP-NN. However, the computational time for the LS-SVM model was very much longer and this can be improved later on by using better hardware to process the data sets.

So as a conclusion, it is safe to say that all the objectives of this project have been achieved and the LS-SVM can be deployed on site. It can also be applied to similar plants elsewhere.

## **5.2 Recommendation**

There are a few recommendations that are to be put forward for this project in this report so that it can be further improved. The current project only used data from one metering station for now. What is recommended is to use more data sets from different metering station so that the reliability can be increased.

Another problem that was faced was the computational time for the LS-SVM is too long when compared to the BP-NN model. What is recommended is to use computer with better hardware to compute it or to use alternative hybrid LS-SVM algorithms that might perform better in terms of computational time and also accuracy. So further studies on possible algorithms is encouraged for further improvements on this project

## REFERENCES

1. *LNG: An eco-friendly cryogenic fuel for sustainable development.* **Satish Kumar a, Hyouk-Tae Kwon b, Kwang-Ho Choi b, Wonsub Lim a, Jae Hyun Cho a, Kyungjae Tak a.** Seoul : s.n., 2011.
2. *Introduction to LNG.* **Foss, Michelle Michot.** 2007.
3. **El-Sherif, A.S.** *Effects of natural gas composition on the nitrogen oxide, flame structure and burning velocity under laminar premixed flame conditions.* Cairo : Elsevier Science Ltd, 1998.
4. *Thermodynamic research improves energy measurement of natural gas.* **Jaeschke, M.** Dorsten, Germany : Thermochemica Acta, 2002.
5. *Prediction of Natural Gas Consumption.* **R. L. Zhang, D. J. Walton and W. D. Hoskins.** Winnipeg, Manitoba, Canada : IEEE, 1993.
6. *The Forecasting of China Natural Gas Consumption Based on Genetic.* **Chen Rui, Wang Jian, Wang Li.** Henan : IEEE, 2009.
7. *Several-Hours-Ahead Electricity Price and Load.* **Paras Mandal, Tomonobu Senjyu and Toshihisa Funabashi.** s.l. : IEEE.
8. *A NOVEL CHAOTIC NEURAL NETWORKS AND APPLICATION.* **WEI-FENG SHI, SHI-LONG XUE.** Shanghai, China : Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, 2005.
9. *Comparison of direct and iterative artificial neural network forecast approaches.* Zonguldak, Turkey : ScienceDirect, 2009.
10. *Artificial Neural Network for Forecasting Daily Loads of Canadian Electric Utility.* **B.S. Kermanshah, i C.H. Poskar, G. Swift.** Winnipeg, Manitoba, Canada : IEEE, 1993.
11. *Development of CO<sub>2</sub> Emission Model of an Acid-Gas Incinerator using Nelder-Mead LS-SVR.* **Elangeshwaran Pathmanathan, Rosdiazli Ibrahim, Vijanth Sagayan Asirvadam.** Tronoh, Perak, Malaysia : s.n.

12. *Model selection for the LS-SVM.Application to handwriting recognition.* **Mathias M.Adankon\*, MohamedCheriet.** Montréal,Canada : ScienceDirect, 2008.
13. *A normal least squares support vector machine (NLS-SVM) and its learning algorithm.* Shanghai,China : ScienceDirect, 2008.
14. *Applying LS-SVM to Predict Primary Energy Consumption.* **Wang Yi, Li Ying.** Beijing,China : s.n.
15. *Pruning LS-SVM Based Battery Model for Electric Vehicles.* **Xiao Lei, C.C. Chan, Kaipei Liu, Li Ma.** s.l. : School of Electrical Engineering, Wuhan University.
16. *REGRESSION MODELS USING PATTERN SEARCH ASSISTED LEAST SQUARES SUPPORT VECTOR MACHINES.* Pune , India : Institution of Chemical Engineers, 2005.
17. *A PARALLEL SVM TRAINING ALGORITHM ON LARGE-SCALE CLASSIFICATION PROBLEMS.* **JIAN-PEI ZHANG, ZHONG-WEI LI, JING YANG.** Guangzhou : s.n., 2005.
18. *Sequential Bootstrapped Support Vector Machines-A SVM Accelerator.* **Xuchun Li, Yan Zhu, Eric Sung.** Montreal : Proceedings of International Joint Conference on Neural Networks, 2005.
19. *On-line least squares support vector machine algorithm in gas prediction.* **ZHAO Xiao-hu, WANG Gang, ZHAO Ke-ke, TAN De-jian.** Jiangsu 221008, China : ScienceDirect, 2009.