SHORT TERM ELECTRICITY PRICE FORECASTING USING BIOLOGICALLY INSPIRED SUPPORT VECTOR MACHINE

MOHAMAD SHAFIQ BIN ROSLAN

BACHELOR OF ELECTRICAL ENGINEERING WITH HONOURS

(INDUSTRIAL POWER)

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

SUPERVISOR DECLARATION

"I hereby declare that I have read through this report entitled "Short Term Electricity Price Forecasting using Biologically Inspired Support Vector Machine" and found that it complies the partial fulfillment for awarding the degree of Bachelor of Electrical Engineering"

Signature	:	…………………………………
Supervisor's Name	:	DR. INTAN AZMIRA BINTI WAN ABDUL RAZAK
Date	:	6 JUNE 2018

SHORT TERM ELECTRICITY PRICE FORECASTING USING BIOLOGICALLY INSPIRED SUPPORT VECTOR MACHINE

MOHAMAD SHAFIQ BIN ROSLAN

A report submitted in partial fulfillment of the requirement for the degree of

Bachelor of Electrical Engineering (Power Industrial)

Faculty of Electrical Engineering

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2018

I declare that this report entitled Short Term Electricity Price Forecasting using Biologically Inspired Support Vector Machine" is the result of my own research except as cited in the references. The report has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature	:	…………………………………
Name	:	MOHAMAD SHAFIQ BIN ROSLAN
Date	:	6 JUNE 2018

To my beloved mother and father

ACKNOWLEDGEMENT

In preparing this report, I used to be in contact with some people, academicians, and practitioners. They have contributed towards my understanding and thought. In particular, I wish to express my sincere appreciation to my main project supervisor, Dr. Intan Azmira Binti Wan Abdul Razak, for guidance, encouragement, critics, and friendship. I am also very thankful to my friends who are in about the same case study for the help and give some suggestions. Without their continued support and interest, this project would not have been same as presented here.

I would like to express my gratitude to Universiti Teknikal Malaysia Melaka (UTeM) for giving me the chance to pursue my Bachelor in Electrical Engineering. My fellow postgraduate students should also be recognized for their support. My honest appreciation also extends to all my colleagues and others who have provided assistance on various occasions. Their views and tips are beneficial certainly. Unluckily, it is not viable to listing all of them in this limited space.

Last but not least, I am grateful to all my family members for all their moral support. I also would like to thank all those involved directly or indirectly in carrying out this project. Without their aide, I might not be able to complete this project at any

ABSTRACT

In deregulated electricity markets, consumers have the ability to choose their electricity supplier. Power producers and customers use short term price forecasts to manage and plan for bidding approaches, and hence increase the utility's profit and energy efficiency. Thus, it allows for greater price flexibility and increased competition between electric providers. This project proposes a prediction model for short-term electricity price forecasting using Support Vector Machine (SVM) and Bacterial Foraging Optimization Algorithm (BFOA). Least Square Support Vector Machine (LSSVM) which is an algorithm that is improved from the SVM is used in this project. Bacterial Foraging Optimization Algorithm (BFOA) is used to optimize the parameters of the LSSVM model which is gamma (γ) and sigma (σ). Furthermore, BFOA optimizes number of features to be fed into LSSVM. The parameters of BFOA are varied to find the best LSSVM-BFOA configuration. The result showed that LSSVM-BFOA able to predict electricity price with good Mean Absolute Percentage Error (MAPE). The model is examined in the Ontario, electricity market in Canada.

ABSTRAK

Dalam pasaran elektrik yang tidak dapat dikawal, pengguna mempunyai keupayaan untuk memilih pembekal elektrik mereka. Pengeluar dan pelanggan kuasa menggunakan ramalan harga jangka pendek untuk mengurus dan merancang pendekatan bidaan, dan dengan itu meningkatkan keuntungan dan kecekapan tenaga utility. Oleh itu, ia membolehkan fleksibiliti harga yang lebih tinggi dan peningkatan persaingan antara pembekal elektrik. Projek ini mencadangkan model ramalan untuk ramalan harga elektrik jangka pendek menggunakan Mesin Vektor Sokongan Biologi (SVM) dan Algoritma Pengoptimuman Pengekalan Bacterial (BFOA). Mesin Vector Sokongan Sisi Kecil (LSSVM) yang merupakan algoritma yang diperbaiki daripada SVM digunakan dalam projek ini. Algoritma Pengoptimuman Pengekalan Bacterial (BFOA) digunakan untuk mengoptimumkan parameter model LSSVM iaitu gamma (γ) dan sigma (σ). Selain itu, BFOA mengoptimumkan jumlah ciri yang akan dimasukkan ke dalam LSSVM. Parameter BFOA adalah berbeza untuk mencari konfigurasi LSSVM-BFOA terbaik. Hasilnya menunjukkan bahawa LSSVM-BFOA dapat meramalkan harga elektrik dengan Ralat Peratusan Maksimum Mutlak yang baik (MAPE). Model ini diperiksa di Ontario, pasaran elektrik di Kanada.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

MCP	-	Market Clearing Price
LSSVM	-	Least Square Support Vector Machine
BFOA	-	Bacterial Foraging Optimization Algorithm
NN	-	Neural Network
MAPE	-	Mean Absolute Percentage Error
ARIMA	-	Autoregressive Integrated Moving Average
ANN	-	Artificial Neural Network
SVM	-	Support Vector Machine
KKT	-	Karush- Kuhn-Tucker
QP	-	Quadratic Programming
SSE	-	Sum Square Error
GA	-	Genetic Algorithm
PSO	-	Particle Swarm Optimization
HOEP	-	Hourly Ontario Electricity Price
S	-	Number of bacteria
Nc	-	Number of Chemotactic
Ns	-	Number of steps during swim
Nre	-	Number of reproduction steps
Ned	-	Number of elimination-dispersal
Ped	-	Number of Probability elimination-dispersal

MAE	-	Mean absolute error
R	-	Regression
PDP	-	Pre-Dispatch Prices
MARS	-	Multivariate adaptive regression splines

CHAPTER 1

INTRODUCTION

1.1 Project Introduction

The world population is increasing day by day and somehow affected the usage of electricity. Furthermore, with current technologies, electricity becomes the source to deliver the power. Electricity needed to run everything in daily lives. When the usage is increasing, the price of electricity for sure will increase as well. With this problem, electricity price forecasting is performed. Electricity Price Forecasting (EPF) plays an important role in the wholesale electricity market. Forecasting electricity price essential for better bidding which is important for both consumer and suppliers. The consumer can maximize their usage and minimize their expense. Energy supplier aim is sell all their generated power and maximize their income [1].

A power market company can forecast the fluctuation of wholesale prices with an affordable level of accuracy. This may regulate its bidding strategy and its own particular generation or utilization plan in order to maximize the profits in day-ahead mercantilism cut back the risk. This also helps the customer to control their usage of the electricity in their daily activities.

Somehow, as compared to predicting the load or demand, forecasting electricity price is more challenging. This is because of fluctuation of price where unexpected spikes might occur for any point of series. There are others factors that affecting the unpredictability in value such as the weather condition, an imbalance between supply and demand and unexpected disturbance at generation and transmission sites. Time-Series (TS), Neural Network (NN) and Support Vector Machine (SVM) are some of the methods that the previous researchers used to forecast electricity price.

1.2 Motivation

Nowadays, fluctuation of electricity price is very common due to various factors. Demand is the most important factor that affects the electricity price. Demand increase because of consumer behavior toward electricity, holidays and peak hours during weekdays. Another important factor is fuel cost that affected by gas prices or coal prices. Besides, weather condition, power plant, transmission, and distribution system for maintenance as known as the factor affecting electricity price forecasting. Both of the suppliers and consumers have the same goal which is to gain the profit. The consumers want to schedule their energy consumption for the next days and maximize their usage, while the suppliers aim to sell all generated energy and maximize their profit [1].

1.3 Problem Statement

There are a lot of methods used for electricity price forecasting. This method can be divided into two categories which are simpler approaches and complex approaches. From previous researches, most of the methods can predict electricity well during normal condition. However, when the spike occurrences happened, the forecast error become large. Some researchers reported least Square Support Vector Machine (LSSVM) can deal with the spikes occurrences [2]. Hence, Least Square Support Vector Machine (LSSVM) is chosen as forecast model.

One of the factors of that cause a bad result of forecast price is the improper selection of features. Due to that, it leads to low accuracy and efficiency of price forecast. Too many feature selection cause the time taken to carry out the analysis become longer. While, insufficient features may also lead to high forecast error. So, the selecting on the features based on the correlation result should be chosen correctly.

Some researched reported that stand-alone forecast model cannot produce a good accuracy. Therefore, this project proposed to combine the optimizer to the forecast model. This optimizer works to select the significant inputs or optimize parameter of forecast model [3]. The selection of parameter is important for optimization to improve the accuracy of optimization.

1.4 Objectives

The objectives of this project areas below:

i. To analyze the correlation between forecast input and target price by using correlation analysis.
ii. To develop Least Square Support Vector Machine (LSSVM) as main forecast engine and optimize LSSVM parameter and features using Bacterial Foraging Optimization Algorithm (BFOA).

1.5 Scope

Scopes of this project are:

i. This project uses LSSVM as main forecast engine.
ii. BFOA is proposed to optimize LSSVM parameter and features of forecasting.
iii. Programming language of MATLAB is written.
iv. The data are taken from Ontario, Canada.
v. Objective function of Mean Absolute Percentage Error (MAPE) is used to observe the forecast error.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Electricity price forecasting has become one of the serious tasks in the operation of the electrical power system.

This participant consists of consumers and suppliers. Both suppliers and consumers need an accurate electricity price prediction to make their profit. The consumer's wish is to maximize power usage and minimize their price while the suppliers desire to sell all their produced power and boost their income.

Based on the previous forecast researchers, electricity price is hard to forecast due to some factors. Some of the factors affected this forecast are load behavior, unstable climate and fuel cost [4].

Many methods have been used to forecast electricity price during normal condition as it works well. However, during the occurrences of a spike, the forecast error becomes high. In this project, Least Square Support Vector Machine (LSSVM) was chosen as a main method and Bacterial Foreign Optimization Algorithm (BFOA) was chosen as an optimization technique.

2.2 Methods of Electricity Price Forecasting

Some of the popular methods of forecast are Time-Series, Neural Network, and Support Vector Machine. The theory and basic principle of each of the method will be explained in the next section.

2.2. 1 Time-Series

A time series is a consecutive arrangement of information data, measured typically over amount of times. A time series can be constant or discrete. In a constant time series information are measured at each example of time, while a discrete time series contains perceptions measured at discrete purposes of time [5].

Strong explanatory model which is Time-Series has been used for forecasting by simulating the variation of historical price [6]. There are many techniques that have been used for time series and one of those techniques is Autoregressive Integrated Moving Average (ARIMA). From previous researches, times series forecasting by ARIMA has become one of the favorite methods [7].

ARIMA models are essentially created to estimate the comparing subordinate variable. This ARIMA model can be partitioned into two kinds of figure which are post sample period forecast and sample period forecast. The sample period forecast is utilized to create certainty interim in the model and the post sample period figure is utilized to produce honest to goodness gauges for arranging and different purpose [8].

Rafał Weron & Misiorek (2008): The authors examine two markets using a wide range of advanced Time Series models: For the Nordpool market, they report an MAPE of only 3.2%. The other market they examine is the Californian market where a MAPE of 12.96% is the best result for a model in which spikes have been pre-processed in a way that they are dampened [20]. Dawit Hailu Mazengia (2008): In a recent study of the Ontario Electricity market in 2007, the authors report a MAPE is 17.85%. The forecasting was done by suing multiple linear regression (MLR) and included exogenous variables for technologies, market power and network congestion using a rolling window approach [21].

2.2.2 Neural Network

Neural Network is a mathematical model that emulates the functional architecture of the human brain [9]. NN works like a human brain. Thus, NN will learn and identify the design of input data. After that, NN will generate the result by applying the knowledge that has been stored in their brain [5].

Neural Network (ANN) is a frequent forecasting technique of electricity price. However, an artificial neural network has intrinsic restitution, and for instance, when the data is out of training sample, the error is extremely large. Thus, it has limited generalization capability and uncontrolled convergence [10].

In the application of electricity price forecasting, NN is analyzed in two cases. In the primary case, we consider just the cost of electric vitality as information and the system preparing is done exclusively in light of valuable information. At that point in the second case, both electric vitality utilization and costs are considered as sources of inputs [11].

Hamdireza Zareipour developed Artificial Neural Network (ANN) for forecasting short-term electricity market price and its application to operation planning of demand-side Bulk Electricity Market Customer (BECMs) in Ontario market [24]. The authors reported that the MAPE from the model is 18.30% based on electricity price in 2004 which come out better that the previous researches.

2.2.3 Support Vector Machine

SVM is a model that uses to analyze data and recognize the pattern for estimation and classification [12]. There are many advantages of SVM compared to the other methods.

SVM can deal with many problems such as local minima, high structural input data, and over-fitting. However, SVM has a disadvantage which is bind with the burden of large set input of data. In order to overcome this problem, Least Square Support Vector Machine (LSSVM) is suggested to lower the burden of SVM [4].

LSSVM is the least squares formulation of a standard SVM. [13]. LSSVM solves a system of linear equations that namely as Karush- Kuhn-Tucker (KKT), instead of a quadratic programming (QP) [14].

In other words, KKT is more easy to use than QP. Even though SVM is improved, LSSVM still retains the concept of old SVM which great in observation capability. LSSVM also reduce the sum square errors (SSEs) of practical statistics sets while synchronously lowering error.

2.3 Optimization Algorithm

In this project, optimization is used to be hybrid with the forecast engine. Optimization can be defined as a minimization or maximization problem. From previous researchers, parameters of forecast engine can be optimized or important inputs can be chosen by optimization algorithm [3]. There are many methods of optimization algorithm such as Genetic Algorithm (GA) and Particle Swarm Optimization (PSO).

2.3.1 Genetic Algorithm

GA is an algorithm based on the survival of the fittest and the natural evolution process via reproduction. The objective function is often referred to as fitness function. Three main operations in GA are selection, crossover and mutation.

There are four core elements that influencing the performance of GA which are population size, number of generations, crossover rate and mutation rate. Furthermore, the algorithm is usually terminated when the generation reaches its maximum value or an acceptable fitness value is obtained.

2.3.2 Particle Swarm Optimization

PSO was introduced by James Kennedy and Rusell Eberhart. PSO mimics the behavior of a group of migrating birds or fish trying to reach unknown destination in the search space changing its velocity. Each individually in a group will randomly move around to find food and announce the sources of the food to its neighbors so the neighbors will approach the same location [2].

A lot of researchers are interested in using PSO since then. The field of swarm intelligence is now become an exciting and constantly evolving the subject of research. Almost every area in design, optimization or scheduling applications, and computational intelligence has used PSO.

2.3.3 Bacterial Foraging Optimization Algorithm (BFOA)

The BFOA was found to search food quicker than other optimization methods [16]. Moreover, the BFOA shows a great performance rather than other meta-heuristic optimization approaches [17]. The BFOA was found to perform better in terms of fast convergence, simplicity in programming, accuracy, and flexibility.

In this project, BFOA is modelled to be hybrid with the LSSVM engine. BFOA is a foraging activity that ingesting and locating nutrient or food that present in human intestines based on bacteria named E.coli.

2.4 Summary and Discussion of the Review

Least Square Support Vector Machine (LSSVM) shows better MAPE than other methods at Ontario, Canada. LSSVM produce 13.0871% MAPE compared to Time-Series and Neural-Network which are 17.63% and 18.80% respectively. Table 2.1 below shows the result of previous MAPE.

Table 2.1: Previous Result of MAPE

References	Method	Test data	MAPE (%)
[2]	Time Series (ARIMA)	2006	17.6300
[2]	Neural Network	2006	18.8000
[2]	LSSVM	2010	13.0871

Based on the result, it can be concluded that LSSVM is a good method for conducting the electricity price forecasting. Each of the methods has advantages and disadvantages that affect the result of MAPE. LSSVM is a technique that not requires big set of data to attain the connection between input and output. LSSVM is good to deal with immense dimensional input and also can step down the over-fitting. Moreover, LSSVM minimizes the sum square errors (SSEs) of training data set while simultaneously lowering margin error [4].

Compare to LSSVM, Time-Series method more suitable for linear problem while price sequence is a non-linear pattern. Meanwhile, NN has problem with over-fitting and under-fitting. This can cause generalization problem as the model cannot figure out the connection between input and output. Besides, training process of NN require extra time.

2.5 Conclusion

Based on the previous research, Least Square Support Vector Machine (LSSVM) shows a better result on performance and accuracy than the other two methods which are Time-Series and Neural Network. So, LSSVM is chosen to be the main forecast engine for this project. In addition, this project proposed to add the optimizer Bacterial Foraging Optimization Algorithm (BFOA) is selected as the optimization technique to improve the performance and accuracy of Support Vector Machine.

CHAPTER 3

METHODOLOGY

3.1 Introduction

This chapter highlights the fundamental of main forecast engine Support Vector Machine and the optimizer which is Bacterial Foraging Algorithm respectively. The overall project activities were stated accordingly in section 3.2 which can be referred to the flowchart.

3.2 Project Flow

This subtopic shows overall project flow of the proposed model. Besides, the flowchart of LSSVM and BFOA will be explained.

3.2.1 Flowchart of Overall Proposed Model

From Figure 3.1, the project flow is started with correlation analysis on past demand and price with the target price. Correlation analysis is operated to examine the correlation between two variables. The correlation analysis is performed in MATLAB and then the correlation analysis is used to pre-select input features. The input features with high correlation coefficient will be selected as input and then will be optimized by BFOA.

After that, the input will be fed to the LSSVM. Selection of input feature is important because it will affect the output which is MAPE. Based on the previous literature review, many types of research use demand and price are commonly used as input [4].

The parameter of LSSVM which are gamma (γ) and sigma (σ) and the selected features will be optimized by Bacterial Foraging Optimization Algorithm (BFOA). The proses of BFOA can be seen in section 3.2.3 while the flowchart (BFOA process) can be seen at Figure 3.2 and Figure 3.3. Then, with the optimized features and LSSVM parameter, LSSVM training is carried out to train and construct the model. The training is followed by LSSVM testing that used to estimate the model performance [18].

From Figure 3.1, Mean absolute percentage error (MAPE) is used as an objective function for this project. MAPE which is also known as mean absolute percentage deviation (MAPD) is a measure of prediction accuracy of a forecasting method in statistics. It has been adopted as the accuracy criteria to assess and compare the performance of a model. From the LSSVM testing, Mean Absolute Percentage Error (MAPE) is calculated as below [19].

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{z_i - z'i}{z_i} \times 100\% \tag{3.1}$$

where z_i is the actual value and z'_i is forecasted value. Meanwhile, N is the number of observations used for analysis.

The final MAPE is the lowest MAPE obtained from the testing process. If the MAPE need to be improved, the process of features and parameter selection by BFOA will be repeated. The process continues until the lowest value of MAPE can be obtained by the hybrid method.

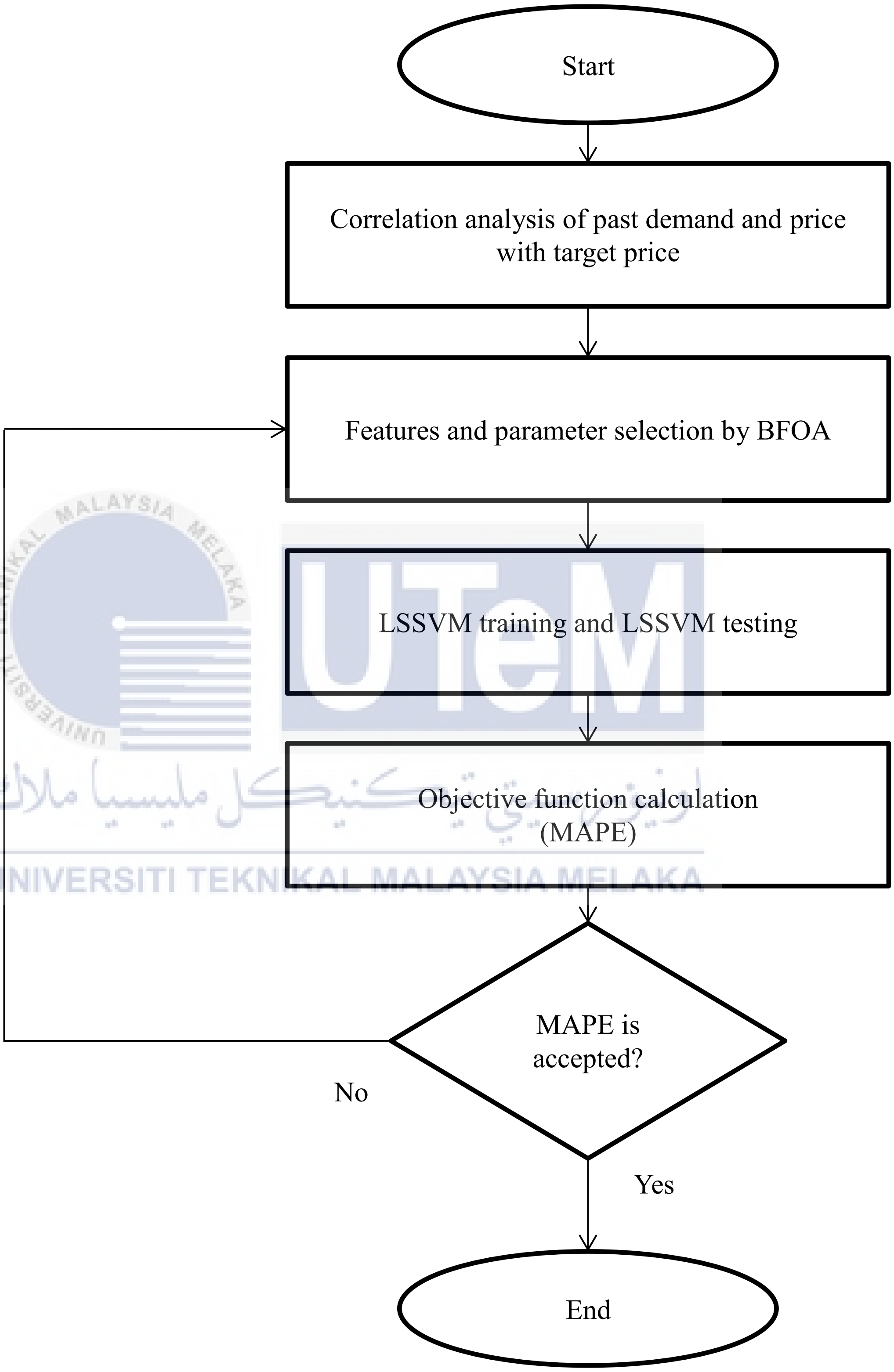

Figure 3.1 Overall Project Flow

3.2.2 Fundamental of BFOA

Bacteria Foraging Optimization Algorithm (BFOA) is a new method of optimization algorithm that inspires by nature phenomenon proposed by Passino (2002). The application of group foraging strategy of a swarm of *E.coli* bacteria in multi-optimal function optimization is the main idea of this new algorithm [20]. To gain the maximum energy obtained per unit time, the bacteria search for its nutrient. Each bacterium also communicates with others by sending signals. Mimic the movement of chemotactic of virtual bacteria in the problem search space is the main idea of BFOA. There is six parameter of BFOA which are number of bacteria (S), number of chemotactic steps (N_c), number of steps during swimming (N_s), number of reproduction steps (N_{re}), number of elimination-dispersal steps (*Ned*) and probability of elimination-dispersal (P_{ed}). There are four main steps in BFOA that are explained as below [21],[23].

i) **Chemotaxis**: This is a process of mimicking the movement of.an E.coli cell via swimming and tumbling by means of flagella. Biologically, an E.coli bacterium can swim for a timeframe in a similar heading or it might tumble and it interchange between these two movements. The $\theta^i(j,k,l)$ shows the *i*-th bacterium at *j*-th chemotactic, *k*-th reproductive and *l*-th elimination-dispersal step. The tumble (run length unit) is specified from the size of the step taken in random direction that represented by C(i). The movement of the bacterium can be represented by:

$$\theta^i(j+1,k,l) = \theta^i(j,k,l) + C(i)\frac{\Delta(i)}{\sqrt{\Delta^T(i)\Delta(i)}} \tag{3.2}$$

where Δ shows a vector in the random direction.

ii) **Swarming:** The behaviour of E.coli has been observed for several motile species of bacteria, where intricate and stable spatio-temporal patterns (swarms) are created in semisolid nutrient medium. When placed amidst a semisolid matrix with a single nutrient chemo-effecter, a group of E.coli cells arranges themselves by moving up the nutrient gradient in a traveling ring. The cells when stimulated by a high level of succinate, release an attractant aspartate which helps them to form into groups. Thus, the high bacterial density is formed in concentric patterns of swarms. The following function shows the cell-to-cell signaling in E. coli swarm;

$$\begin{aligned} J_{cc}\big(\theta, P(j,k,l)\big) &= \sum_{i=1}^{S} J_{cc}\left(\theta, \theta^{i}(j,k,l)\right) \\ &= \sum_{i=1}^{S}[-d_{attract}\exp(-w_{attract}\sum_{m=1}^{P}(\theta_m - \theta_m^i)^2)] \\ &+ \sum_{i=1}^{S}[h_{repell}\exp(-w_{repell}\sum_{m=1}^{P}(\theta_m - \theta_m^i)^2)] \end{aligned} \tag{3.3}$$

where $J_{cc}\big(\theta, P(j,k,l)\big)$ is the objective function value to be added to the actual objective function (to be minimized) to present a time-varying objective function. S is the sum of bacteria, p is the number of variables to be optimized, which are present in each bacterium and $\theta = [\theta_1, \theta_2, .., \theta_p]^T$ is a point in the p-dimensional search domain. The coefficients of d_{attract}, w_{attract}, h_{repell}, w_{repell} should be chosen properly.

iii) **Reproduction**: The healthier bacteria will split into two bacteria at the same location while the others die. This makes the swarm size keep constant.

iv) **Elimination and Dispersal**: Sudden changes in the local environment where a bacterium population lives may happen because of various reasons. For example, a group of bacteria in a place may be killed by a significant local rise in temperature. Events can take place in term of all the bacteria are killed in a region or a group is scattered into another place.

3.2.3 Process of Bacterial Foraging Optimization Algorithm

Figure 3.2 and Figure 3.3 show the process of BFOA which is elaborated as below [23].

Step 1: Declaration of BFOA parameter which are S, N_c, N_s, N_{re}, N_{ed} and P_{ed}. At the same time, BFOA initialize the parameters of LSSVM (γ and σ) and features selected from correlation analysis.

Step 2: Elimination-Dispersal loop: $l = l+1$

Step 3: Reproduction loop: $k = k+1$

Step 4: Chemotaxis loop: $j=j+1$

(i) For $i = 1, 2,\ldots,S$ take a chemotactic step for bacterium 'i' as follows:

(ii) Compute cost $J(i,j,k,l)$

Let $J(i,j,k,l)=J(i,j,k,l)+J_{cc}(\theta^i(j,k,l),P(j,k,l))$

(This adds on the cell to cell attraction effect to the nutrient concentration)

(iii) To save the value, let $J_{last} = J(i,j,k,l)$, from a run as the better cost might be found there

(iv) Tumble: Generate a random vector $\Delta(i) \in R^p$ with each element $\Delta_m(i)$, m = 1, 2… p, a random number on [-1, 1] where R is a real number.

(v) Move: let

$$\theta^i(j+1,k,l) = \theta^i(j,k,l) + C(i)\frac{\Delta(i)}{\sqrt{\Delta^T(i)\Delta(i)}}$$

This results in a step of size C(i) in a direction of the tumble for bacterium i

(vi) Compute $J(i,j+1.k,l)$

The next step can be carried out if the cost function or loss is minimum else move to step (ii).

(vii) Swim

(a) Let m=0 (counter for swim length)

(b) While m<N_s

Let m = m+1

If $J(i,j+q,k,l) < J_{last}$ (if there is improvement), let $J_{last} = J(i,j+1,k,l)$

and let $\theta^i(j+1,k,l) = \theta^i(j,k,l) + C(i)\frac{\Delta(i)}{\sqrt{\Delta^T(i)\Delta(i)}}$ and use this

$\theta^i(j+1,k,l)$ to compute the new $J(i,j+1,k,l)$.

Else, let m=N_s. End of while statement

(viii) Go to next bacterium ($i+1$) if $i \neq S$

Step 5: If j<N_C move to step 3. The bacteria continue its chemotaxis as it still alive.

Step 6: Reproduction

a) For the given k and l, and for each i=1,2,…,S, let

$$J^i_{health} = \sum_{j=l}^{Nc+1} J\,(i,j,k,l)$$

be the health of bacterium i. J_{health} is sort in ascending values which means the lower health shows the higher value.

b) The highest J_{health} values make the S_r bacterium die and the best values make other S_r bacteria to split.

Step 7: If k < N_{re}, move to step 3. This shows that the numbers of specified reproduction steps are still not enough as specified.

Step 8: Elimination-Dispersal

For i=1,2,……,S with probability P_{ed}, eliminate and disperse each bacterium. Bacteria are eliminated and disperse one to a random location on the optimization domain.

Step 9: If l < N_{ed}, move to step 2, otherwise end.

All the steps of BFOA are compressed in Figure 3.2 and Figure 3.3.

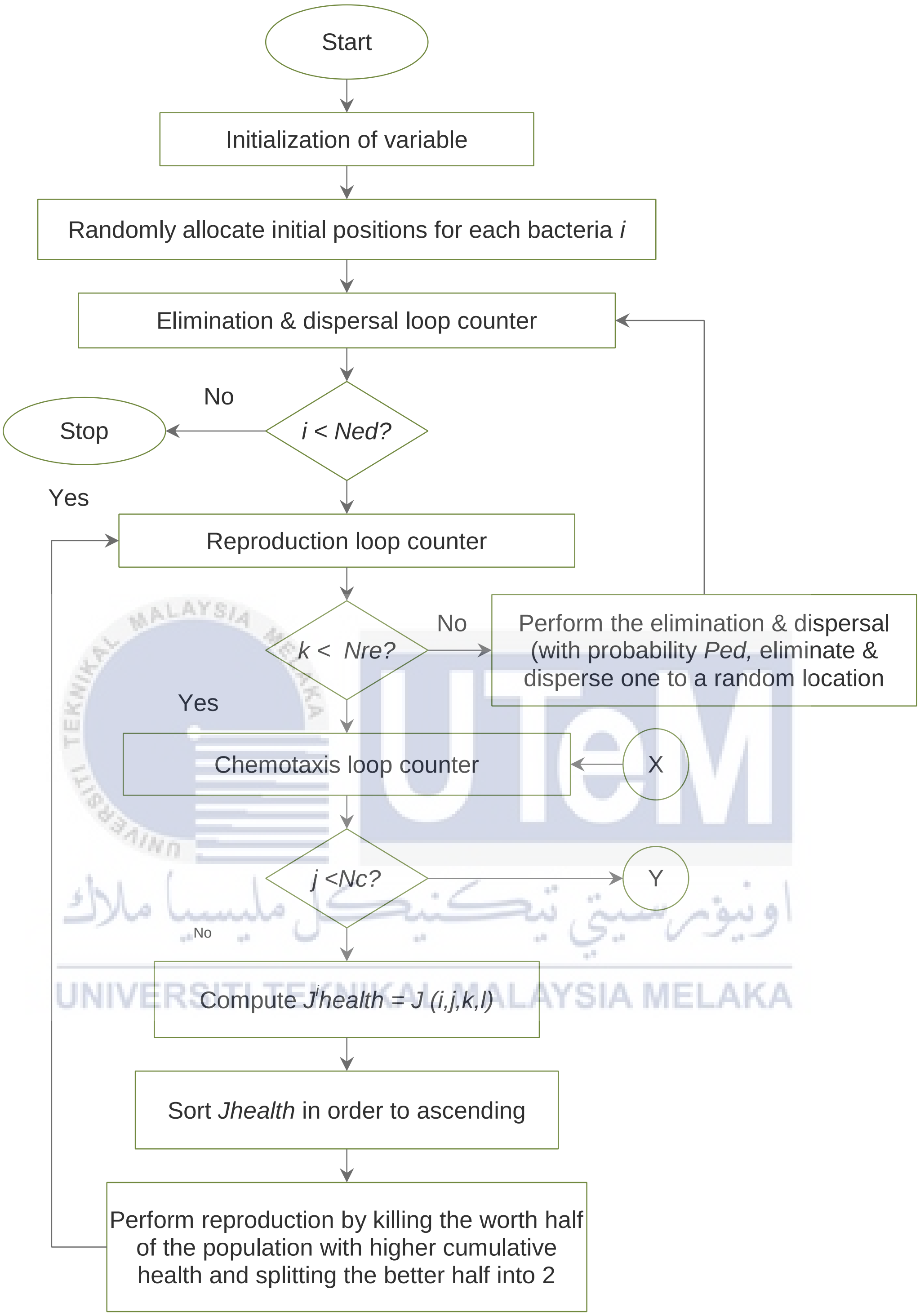

Figure 3.2 Bacterial Foraging Optimization Algorithm Proses (A) [3]

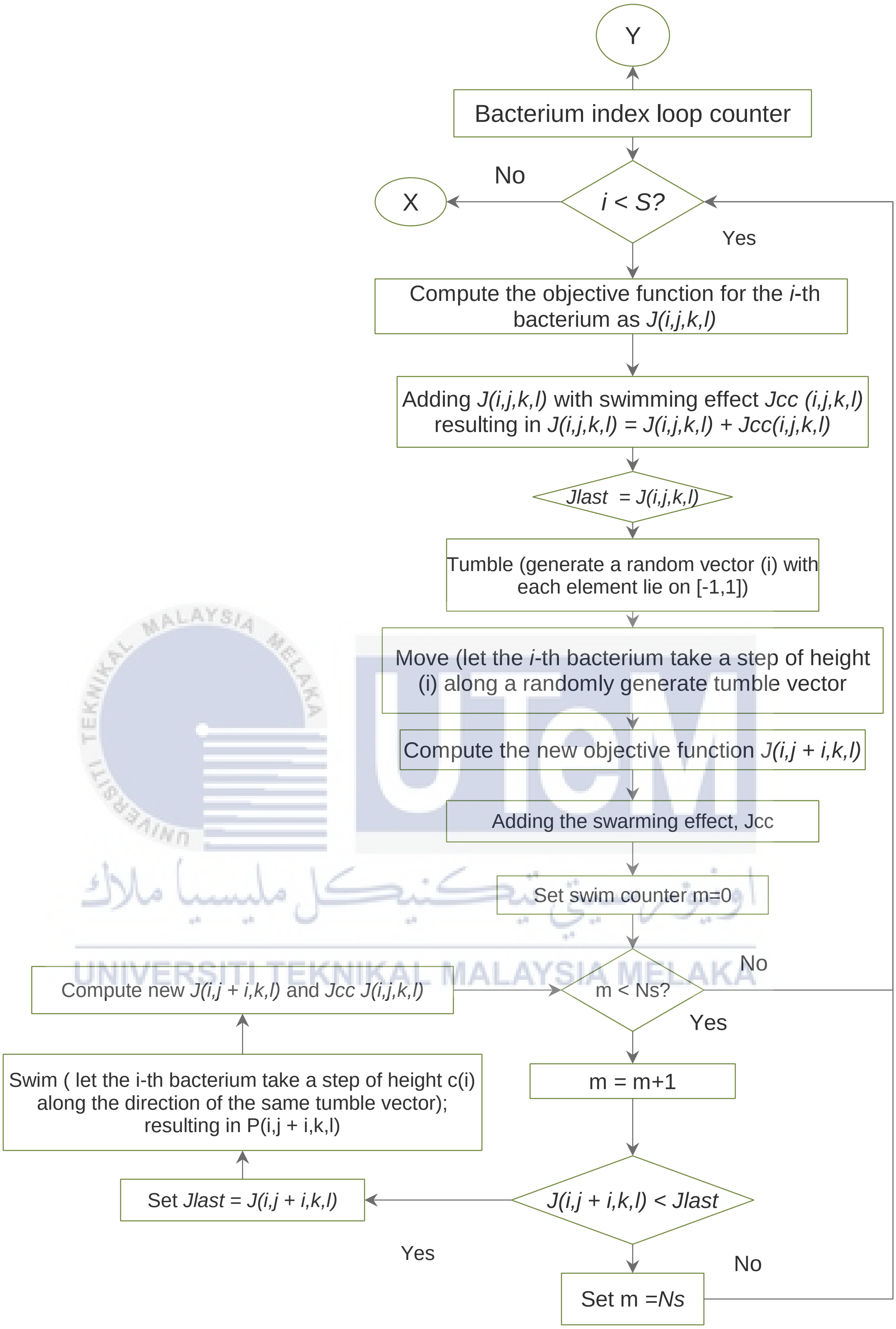

Figure 3.3 Bacterial Foraging Optimization Algorithm Proses (B) [3]

CHAPTER 4

EXPECTED RESULT

4.1 Introduction

This chapter shows the preliminary result of the project which is correlation analysis. The correlation between features and target prices are observed. The features with higher correlation coefficient will be selected as the input for forecasting. Then, selected features and LSSVM parameter will be optimized by BFOA. The optimized parameter and features will be fed into LSSVM for training and testing.

4.2 Correlation Analysis

The correlation analysis is carried out to spot the correlations between features and target price. The correlation coefficient is a measure of linear association between two variables. This type of analysis is used to verify if there a possible relation between variables. The correlation result can be either positive or negative based on the values calculated. Values of the correlation coefficient are always between -1 and +1.

The data are publicly available at http://www.ieso.ca/ for Hourly Ontario Electricity Price (HOEP) and Market Demands is used for this project. The demand and price throughout years 2006 and 2007 are taken for correlation analysis. The date selected starting is starting from January to December for both years 2006 and 2007. The correlation result is recorded as Table 4.1 and Table 4.2. The notation (d-n) indicates the number of day before the targeted day.

Table 4.1: Correlation coefficient of past demand and past price for years 2006 and 2007

Correlation / Features	Past Demand 2006	Past Price 2006	Past Demand 2007	Past Price 2007
d-1	0.9980	0.1674	0.9588	-0.3383
d-2	0.9886	0.6612	0.9421	-0.3906
d-3	0.9682	0.1751	0.9427	-0.3039
d-4	0.9791	0.5041	0.9664	-0.1695
d-5	0.8942	0.2530	0.9477	-0.1685
d-6	0.9623	0.2217	0.8597	-0.0416
d-7	0.9812	0.4683	0.9142	0.3684

The correlation coefficient of past price and past demand with the targeted price are illustrated as in Figure 4.1 until Figure 4.4.

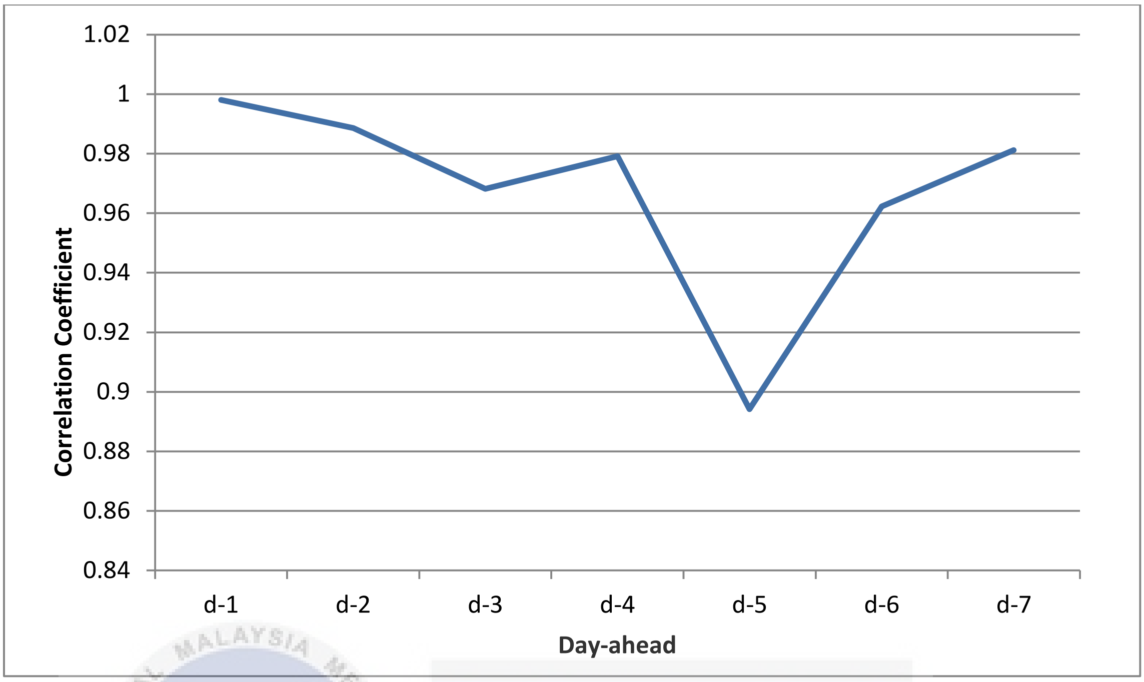

Figure 4.1 Correlation coefficient of past demand with target price 2006

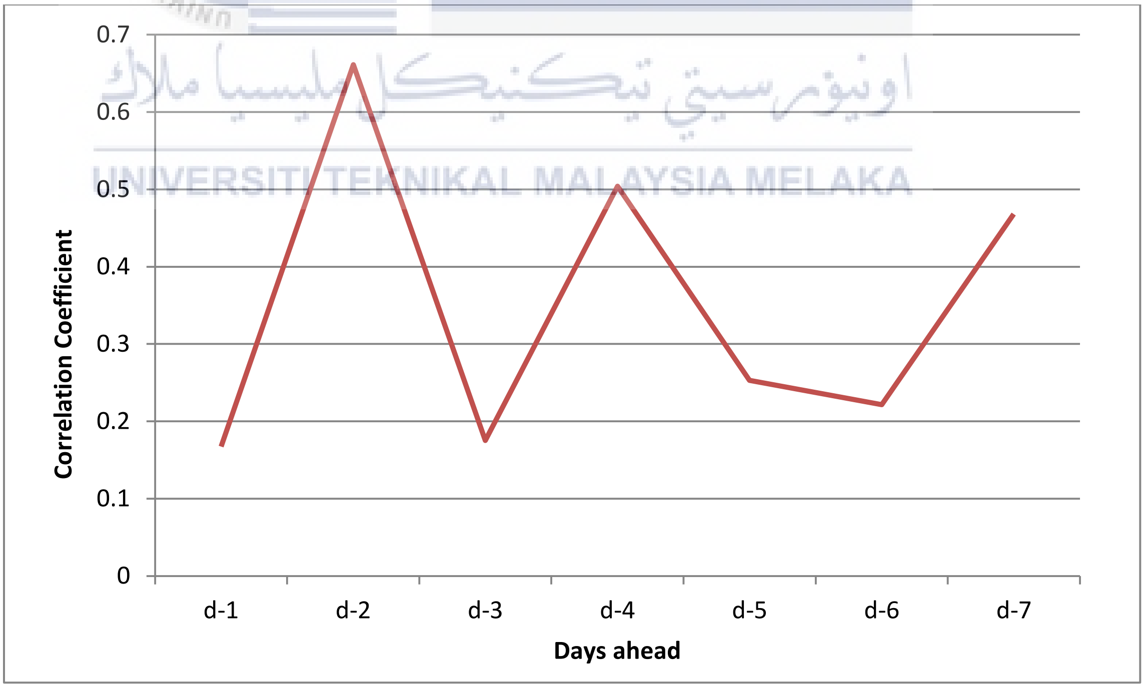

Figure 4.2 Correlation coefficient of past price with target price 2006

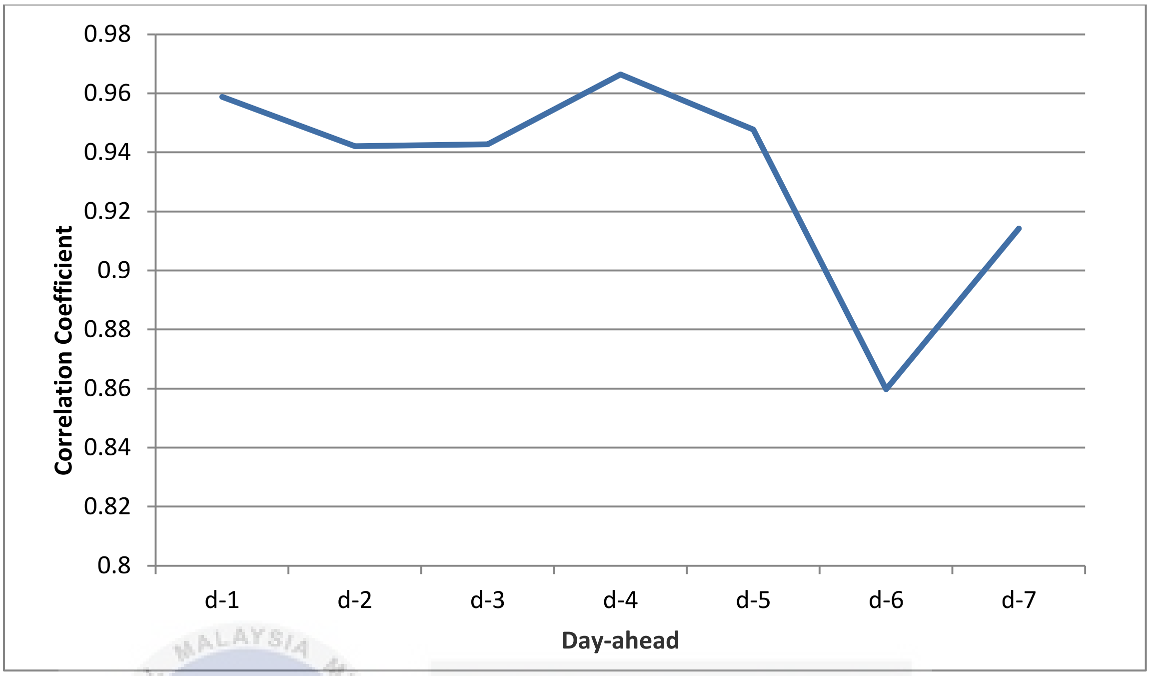

Figure 4.3 Correlation coefficient of past demand with target price 2007

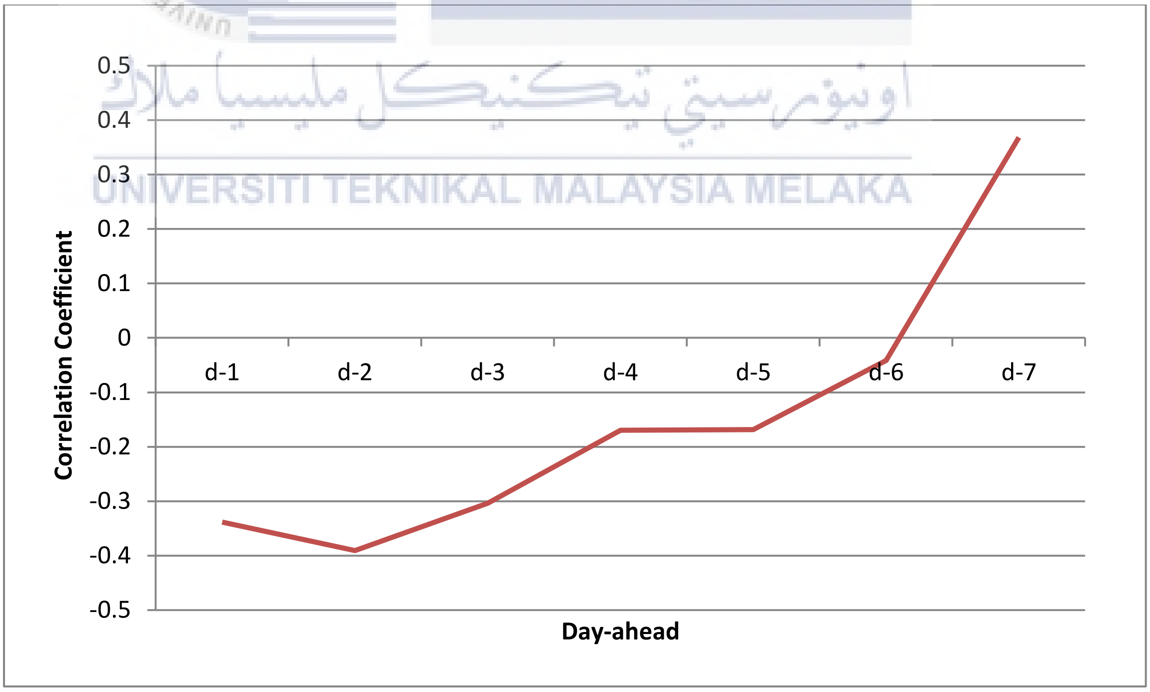

Figure 4.4 Correlation coefficient of past price with target price 2007

Based on the previous study, high correlation can be a review as correlation coefficient with the range [0.5-1.0], while medium correlation is referred for correlation within the range [0.3-0.49] [4]. So, the features with high correlation are selected as input features in this project. Meanwhile, past seven days demand are selected to reduce computational burden to the data.

From Figure 4.1 until Figure 4.4, the correlation of Hourly Demand for 2006 and 2007 are highly correlated. Meanwhile, the correlation of HOEP unstable as some of the days are highly correlated and some of the days are low correlated. It can be concluded that correlation coefficient of Hourly Demand is better than correlation coefficient of HOEP.

From table 4.1, the correlation coefficient of Hourly Demand is taken as input features. Each days contain 24 hours, it make the total input features is 168. These input features are used to be fed into LSSVM network which is training and testing.

4.3 Training and Testing Period for Forecast

As comparison with previous researches, six forecast models are developed to represent the whole year of 2004. Year 2004 was selected to be fair comparison with other forecast models [3]. Each model is trained with ten weeks data prior to the forecasting week as shown in Table.

Table 4.2: Training and testing period in 2004

Training Period	**Testing Period**
8th March - 23th April	26th April – 2nd May
15th March - 2nd May	3rd May – 9th May
7th June – 25th July	26th July – 1st August
14th June – 1st August	2nd August – 8th August
25th October – 12th December	13th December – 19th December
1st November – 19th December	20th December- 26th December

4.4 Performance of LSSVM-BFOA

There are several considerations that can be manipulated to see performance of LSSVM-BFOA. Changing the value of each parameter is one the important consideration to see the performance. This is to observe the effect of parameter on value of MAPE. Then, the best parameter will give out lowest value of MAPE is taken for the next analysis. The performance of the model is tested for six weeks of year 2004.

There are 4 parameter that has to be varied which are number of bacteria in population, number of chemotactic, number of steps during swimming and number of reproduction steps. The starting BFOA parameter setting for every week is set as table below.

Table 4.3: Early parameter setting

Type of Parameter	**Parameter Value**
S	2
Nc	10
Ns	2
Nre	1

4.4.1 Tuning BFOA parameter for week 1

4.4.1.1 Varying Number of Bacteria, *S* parameter

For varying *S* parameter, *Nc*, *Ns* and *Nre* is fixed as in Table 4.3. Table 4.4 shows that the value of MAPE is inconsistent with the increasing number of bacteria. However the time taken for the simulation to run is increasing as the number of bacteria increase. For week 1, number of bacteria that selected is 6.

Table 4.4: Varying *S* parameter in week 1

S	**MAPE (%)**	**Time Taken (Hours)**
2	18.8865	0.0068
4	18.3346	0.0133
6	17.9972	0.0201
8	18.1733	0.0283
10	18.4398	0.0324
12	18.2225	0.0409
14	18.0336	0.0462
16	18.2504	0.0514
18	18.6143	0.0590
20	18.2288	0.0657

4.4.1.2 Varying Number of Chemotactic Steps, N_c Parameter

The best S parameter which is 6 taken as fix value for varying *Nc* parameter without changing the value of N_s and *Nre*. Table 4.5 shows that the value of *Nc* = 20 give better MAPE than others value of *Nc*. The time taken for *Nc* is increase as the number of *Nc* increase.

Table 4.5: Varying *Nc* Parameter in week 1

Nc	**MAPE (%)**	**Time Taken (Hours)**
10	17.9972	0.0201
15	18.0694	0.0299
20	17.7664	0.0391

4.4.1.3 Varying Number of Steps during Swimming, *Ns* Parameter

Nc parameter from previous analysis which is 20 is taken as fix value parameter of *Nc* in varying Ns parameter. The value of MAPE decrease from *Ns* = 2 to *Ns* = 6, however rise back at *Ns* = 8. Parameter *Ns* = 6 is selected as best value that give better MAPE. The time taken for simulation is increase as number of Ns increase.

Table 4.6: Varying *Ns* Parameter in week 1

Ns	**MAPE (%)**	**Time Taken (Hours)**
2	17.9972	0.0201
4	17.8006	0.0446
6	17.5739	0.0502
8	17.7452	0.0789

4.4.1.4 Varying Number of Reproduction, *Nre* Paramter

In this experiment, parameter value of *S*=6, *Nc*=20 and *Ns*=6 is carry forward in varying Number of reproduction *Nre*. The value of MAPE gives a better result as the number of *Nre* increase. The time taken for varying *Nre* is increase as *Nre* parameter increase.

Table 4.7: Varying *Nre* Parameter in week 1

Nre	**MAPE (%)**	**Time Taken (Hours)**
1	17.5739	0.0502
2	17.0824	0.0951
3	16.9319	0.1416
4	16.8936	0.2157

4.4.2 Tuning BFOA parameter for week 2

4.4.2.1 Varying Number of Bacteria, *S* parameter

Based on the table below, the value of MAPE decrease from *S*=2 until *S*=10, however the MAPE rise back after *S*=10. For week 2, *S*=10 show the better result than other value of *S*. the time taken for simulation increase as the number of bacteria, *S* increase.

Table 4.8: Varying *S* parameter in week 2

S	**MAPE (%)**	**Time Taken (Hours)**
2	20.6877	0.0072
4	20.5129	0.0176
6	20.3094	0.0224
8	20.2460	0.0308
10	19.7848	0.0350
12	21.0155	0.0399
14	20.6737	0.0475
16	20.6728	0.0545
18	20.1397	0.0604
20	20.7374	0.0653

4.4.2.2 Varying Number of Chemotactic Steps, *Nc* Parameter

The best *S* parameter from Table 4.8 is taken as fix value to varying the *Nc* parameter. Based on the table 4.9, *Nc*=10 give a better result of MAPE than *Nc*=15 and *Nc*=20. The MAPE increase as the *Nc* parameter increase which not good.

Table 4.9: Varying N_c Parameter in week 2

Nc	**MAPE (%)**	**Time Taken (Hours)**
10	19.7848	0.0350
15	20.6849	0.0475
20	19.9987	0.0631

4.4.2.3 Varying Number of Steps during Swimming, *Ns* Parameter

The parameter *S*=10 and *Nc*=10 is taken to varying the *Ns* parameter. The value of MAPE at the beginning is 19.7848% when *Ns* parameter equal to 2. However the value of MAPE increase to 20.0391% as *Ns*=4. Although MAPE value is higher when *Ns*=6 compared when *Ns*=2, the MAPE value decrease again when the *Ns* value rose to 6. *Ns*=6 is taken to varying the next parameter *Nre*.

Table 4.10: Varying *Ns* Parameter in week 2

Ns	**MAPE (%)**	**Time Taken (Hours)**
2	19.7848	0.0350
4	20.0391	0.0370
6	19.7011	0.0400
8	19.9000	0.0436

4.4.2.4 Varying Number of Reproduction, *Nre* Parameter

The control parameter which is S=10, *Nc*=10 and Ns=6 is use to tuning the *Nre* parameter. From table below it show that as value of *Nre* increase, the MAPE decrease. The *Nre*=4 is taken as best value that give better MAPE result, meanwhile the take taken for simulation of *Nre* increase as the value of *Nre* increase.

Table 4.11: Varying *Nre* Parameter in week 2

Nre	**MAPE (%)**	**Time Taken (Hours)**
1	19.7011	0.0400
2	18.7147	0.0816
3	18.5509	0.1087
4	18.4120	0.1481

4.4.3 Tuning BFOA parameter for week 3

4.4.3.1 Varying Number of Bacteria, *S* parameter

For week 3 the value of S parameter that give better MAPE is *S*=8. The parameter *S*=8 give the result of MAPE which is16.2351%. Other values of *S* show inconsistent MAPE as the value of *S* parameter increase. However, the time taken is rise as the *S* parameter increase.

Table 4.12: Varying *S* parameter in week 3

S	**MAPE (%)**	**Time Taken (Hours)**
2	16.7070	0.0117
4	16.8664	0.0185
6	16.7691	0.0291
8	16.2351	0.0407
10	16.4321	0.0529
12	16.7577	0.0639
14	16.2398	0.0710
16	16.5262	0.0850
18	16.4324	0.0884
20	16.2465	0.1113

4.4.3.2 Varying Number of Chemotactic Steps, *Nc* Parameter

The best *S* parameter which is 8 is taken as fix value for tuning *Nc* parameter without any change in *Ns* and *Nre* parameter. From table 4.13 its show that the value of MAPE is increase as the value of *Nc* increase. However the different of MAPE is small which is 0.0807% when *Nc*=10 and *Nc*=20.

Table 4.13: Varying N_c Parameter in week 3

Nc	**MAPE (%)**	**Time Taken (Hours)**
10	16.2351	0.0407
15	16.2407	0.0661
20	16.3158	0.0901

4.4.3.3 Varying Number of Steps during Swimming, *Ns* Parameter

After getting the value of *S*=8 and *Nc*=10, the value is used to vary the *Ns* parameter. As the value of *Ns* parameter increase from 2 to 6, the value of MAPE is increase. However as the *Ns* reached 8 the MAPE give 16.0445% which is lower than others MAPE. *Ns*=8 is taken as control parameter to vary the *Nre* parameter.

Table 4.14: Varying *Ns* Parameter in week 3

Ns	**MAPE (%)**	**Time Taken (Hours)**
2	16.2351	0.0407
4	16.3717	0.0500
6	16.5335	0.0568
8	16.0445	0.0558

4.4.3.4 Varying Number of Reproduction, *Nre* Paramter

From previous analysis, *S*=8 , *Nc*=10 and *Ns*=8 parameter is carry forward in this experiment. Base on the table below, as the value of *Nre* parameter increase, the value of MAPE is inconsistent. The MAPE show a decreasing percent as the value of *Nre* increase from 1 to 2, however the MAPE increase back as the value of *Nre* increase from 2 to 3. In conclusion, *Nre*=2 give the lower MAPE which is 15.4365% than other value of *Nre*.

Table 4.15: Varying *Nre* Parameter in week 3

Nre	**MAPE (%)**	**Time Taken (Hours)**
1	16.0445	0.0558
2	15.4365	0.1061
3	16.2327	0.1545
4	15.8171	0.1551

4.4.4 Tuning BFOA parameter for week 4

4.4.4.1 Varying Number of Bacteria, *S* parameter

Based on the table below, parameter *S*=20 give the lowest MAPE result compared to other *S* parameter. Beside the inconsistent of value of MAPE, the time taken for the simulation is increasing as the value of *S* increase.

Table 4.16: Varying *S* parameter in week 4

S	**MAPE (%)**	**Time Taken (Hours)**
2	24.0729	0.0067
4	25.1498	0.0131
6	23.6497	0.0196
8	26.3277	0.0270
10	27.6011	0.0326
12	24.0761	0.0394
14	25.7058	0.0454
16	24.0599	0.0517
18	24.4549	0.0584
20	23.6296	0.0675

4.4.4.2 Varying Number of Chemotactic Steps, *Nc* Parameter

The value of *S*=20 from previous analysis is used to vary the *Nc* parameter. In this analysis, the value of MAPE increase as *Nc* parameter increase from 10 to 15. However, the MAPE decrease again when the *Nc* rise from 15 to 20. The parameter *Nc*=20 give the lowest MAPE which is 23.1660%.

Table 4.17: Varying N_c Parameter in week 4

Nc	**MAPE (%)**	**Time Taken (Hours)**
10	23.6296	0.0675
15	23.8355	0.0989
20	23.1660	0.1732

4.4.4.3 Varying Number of Steps during Swimming, *Ns* Parameter

From table below, the value of MAPE is decreasing as the parameter *Ns* increasing from 2 to 6. After *Ns*=6 the value of MAPE rose again at *Ns*=8. Furthermore, the time taken also is inconsistent as the value of *Ns* increase. In this analysis, *Ns*=6 is selected cause give good MAPE=21.9138% with time taken 0.1636 hours.

Table 4.18: Varying *Ns* Parameter in week 4

Ns	**MAPE (%)**	**Time Taken (Hours)**
2	23.1660	0.1732
4	22.8541	0.2005
6	21.9138	0.1636
8	22.6457	0.1724

4.4.4.4 Varying Number of Reproduction, *Nre* Paramter

The value of *S*=20, *Nc*=20 and *Ns*= 6 is taken as control parameter to vary the *Nre* parameter. From table 4.19 the value of MAPE become lower as the value of *Nre* is rising. The *Nre*=4 is chosen that give better value MAPE which is 21.0689% for week 4.

Table 4.19: Varying *Nre* Parameter in week 4

Nre	**MAPE (%)**	**Time Taken (Hours)**
1	21.9138	0.1636
2	21.7292	0.3019
3	21.4930	0.6570
4	21.0689	0.9277

4.4.5 Tuning BFOA parameter for week 5

4.4.5.1 Varying Number of Bacteria, *S* parameter

Based on the table below, it shows that the value of MAPE is not stable as the value of *S* parameter increasing. For week 5, the value of *S*=8 is suitable that give better MAPE compared to others value which is 18.1338%. The time increase simultaneously as the value of S increases.

Table 4.20: Varying *S* parameter in week 5

S	**MAPE (%)**	**Time Taken (Hours)**
2	18.1754	0.0094
4	18.1720	0.0174
6	18.1741	0.0266
8	18.1338	0.0325
10	18.1740	0.0446
12	18.1674	0.0550
14	18.1650	0.0631
16	18.1752	0.0710
18	18.1711	0.0818
20	18.1742	0.0914

4.4.5.2 Varying Number of Chemotactic Steps, *Nc* Parameter

The best *S* parameter which is 8 is taken as fix value for varying the *Nc* parameter. Based on table 4.21 the value of MAPE is increase from 18.1338% to 18.1709% as the value of *Nc* parameter increase. The value of *Nc*=10 then is chosen as control parameter for next analysis.

Table 4.21: Varying N_c Parameter in week 5

Nc	**MAPE (%)**	**Time Taken (Hours)**
10	18.1338	0.0325
15	18.1363	0.0526
20	18.1709	0.0721

4.4.5.3 Varying Number of Steps during Swimming, *Ns* Parameter

After getting the value of *S*=8 and *Nc*=10, the Ns parameter can be varied. The value of MAPE is increasing as the value increase from *Ns*=2 to *Ns*=6. However, the value of MAPE decrease again when the *Ns* equal to 8. The *Ns*=2 is chosen as control parameter with the different 0.0031% of MAPE compared to *Ns*=8.

Table 4.22: Varying *Ns* Parameter in week 5

Ns	**MAPE (%)**	**Time Taken (Hours)**
2	18.1338	0.0325
4	18.1492	0.0326
6	18.1719	0.0418
8	18.1369	0.0425

4.4.5.4 Varying Number of Reproduction, *Nre* Paramter

For this analysis, the control parameter is *S*=8, *Nc*=10 and *Ns*=2. From table below it shows that, although the MAPE value is higher when *Nre*=2 compared to *Nre*=1, the MAPE value decrease again when the value of *Ns* rose higher. The value of MAPE is 18.1335% when *Nre*=4.

Table 4.23: Varying *Nre* Parameter in week 5

Nre	**MAPE (%)**	**Time Taken (Hours)**
1	18.1338	0.0325
2	18.1524	0.0662
3	18.1345	0.0966
4	18.1335	0.1323

4.4.6 Tuning BFOA parameter for week 6

4.4.6.1 Varying Number of Bacteria, *S* parameter

From table below, it show that *S*=14 give the better MAPE result compared to the others value *S*. Furthermore, the value of MAPE is inconsistent as the value of *S* rise higher. The time taken for this simulation is increasing as the value of *S* increase.

Table 4.24: Varying *S* parameter in week 6

S	**MAPE (%)**	**Time Taken (Hours)**
2	27.2156	0.0060
4	27.1964	0.0130
6	27.2173	0.0183
8	27.1957	0.0250
10	27.1850	0.0315
12	27.1312	0.0368
14	27.0930	0.0442
16	27.2305	0.0502
18	27.1866	0.0572
20	27.1603	0.0619

4.4.6.2 Varying Number of Chemotactic Steps, *Nc* Parameter

The value of *S*=14 is taken to vary the *Nc* parameter without changing the early control parameter *Ns* and *Nre*. Based on table below, it shows that the value of MAPE is increasing as the value of *Nc* is increasing. The value of *Nc*=10 is chosen cause give a better MAPE which is 27.0930%.

Table 4.25: Varying N_c Parameter in week 6

Nc	**MAPE (%)**	**Time Taken (Hours)**
10	27.0930	0.0442
15	27.1679	0.0659
20	27.2035	0.0873

4.4.6.3 Varying Number of Steps during Swimming, *Ns* Parameter

S and *Nc* parameter from previous analysis which are 14 and 10 is taken as fix value parameter in varying the *Ns* parameter. The value of MAPE is increase as the value of *Ns* increase based on the table 4.26. From the table below, it shows that the *Ns*=2 give the better result of MAPE which is 27.0930% compared to the other value of *Ns*.

Table 4.26: Varying *Ns* Parameter in week 6

Ns	**MAPE (%)**	**Time Taken (Hours)**
2	27.0930	0.0442
4	27.1037	0.0497
6	27.1427	0.0526
8	27.1598	0.0474

4.4.6.4 Varying Number of Reproduction, *Nre* Paramter

The control parameter for varying *Nre* parameter are *S*=14, *Nc*=10 and *Ns*=2. Based on the table below, the value of MAPE is inconsistent as the value of *Nre* increasing. The *Nre*=1 the only parameter that give a better MAPE with 27.0930%. The time taken is increase with the increasing value of *Nre*.

Table 4.27: Varying *Nre* Parameter in week 6

Nre	**MAPE (%)**	**Time Taken (Hours)**
1	27.0930	0.0442
2	27.2130	0.0819
3	27.1714	0.1236
4	27.1381	0.1680

4.4.7 Selected Value of LSSVM-BFOA Parameter

Table below show the value of all parameter selected representing for each week. *S* is number of bacteria, *Nc* is chemotactic steps, Ns is number of swim, *Nre* is number of reproduction steps, *Ned* is number of elimination-dispersal and *ped* is probability of elimination-dispersal. This parameter must be chosen properly. In this experiment, parameter of *S*, *Nc*, *Ns* and *Nre* all are case dependent, because each week represents different season. However, Ns must be smaller than *Nc*. As in general, rising the value of *S*, *Nc*, *Ns*, and *Nre* may increase computational complexity, simulation time but perhaps leading for a better optimization progress where bacteria can find more searching space.

Table 4.28: Selected value of LSSVM-BFOA parameter

Week	***S***	***Nc***	***Ns***	***Nre***	***Ned***	***Ped***
1	6	20	6	4		
2	10	10	6	4		
3	8	10	8	2	2	0.25
4	20	20	6	4		
5	8	10	2	4		
6	14	10	2	1		

4.5 Result of MAPE for Every Test Week

Table below shows the result for every test week that includes the gamma, sigma, total features selected, regression, MAE and MAPE. Based on the result, gamma and sigma are case dependent because every week has its own behavior. The value of gamma and sigma for week 1, 2 and 4 are high compare to week 3, 5 and 6. At the beginning of the testing, the value of features is 168. From the table below, the total features selected after the testing is almost half of the 168. Meanwhile, the regression for all test weeks show the value that higher than 0.5 except for week 6 which is lower than 0.5. When the value of regression is high, it shows that the price of that week is less volatile. It can be concluded that week 1 until week 5 regression is less volatility than week 6 that consider as high volatile. Week 6 has low regression because that week known as high demand winter. Lastly, the result showed that MAE and MAPE are correlated. When the value of MAE high, the value of MAPE also high. From table below, the highest value of MAE is gained by week 6 and the lowest MAE is week 1. Meanwhile, the highest value of MAPE is obtained by week 6 and the lowest MAPE is week 3.

Table 4.29: Result of every test week

Week	**Gamma**	**Sigma**	**Total Features selected**	**Regression (r)**	**MAE**	**MAPE (%)**
1	1.9531	11.2423	82	0.7059	7.3868	16.8936
2	1.2369	7.6661	70	0.7626	8.2635	18.4120
3	0.7948	3.6890	81	0.7939	7.8550	15.4365
4	1.2724	7.5913	77	0.8511	7.8821	21.0689
5	0.9004	0.0790	79	0.7025	11.1150	18.1335
6	0.5854	0.7651	84	0.4240	14.6255	27.0930

4.6 Actual versus Prediction Graph

4.6.1 Graph of week 1

Week 1 testing is said to be a period where the Ontario market demand reached its low spring low point. To forecast the electricity prices for weeks 1 of testing period (April 26th – May 2nd), a series of training data from March 8th – April 23th is used for model development. Based on the developed models the electricity price for week 1 is forecasted and the results are shown in the figure below. The weekly percentage error for this forecast is calculated to be 16.8936% which is a considerably average error. On April 29th the graph showed the highest volatility price value, so that predicted point unable to follow the actual point.

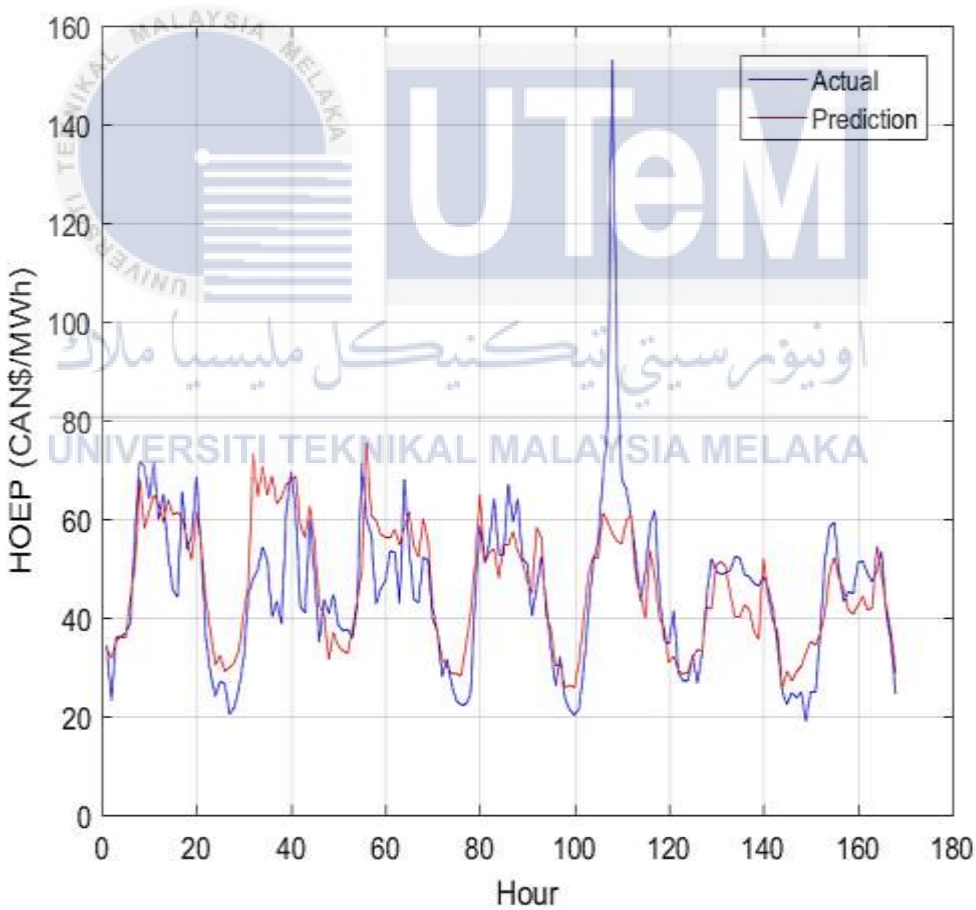

Figure 4.5: Week 1 actual vs prediction graph

4.6.2 Graph of week 2

Week 2 testing is also said to be a period where the Ontario market demand reached its low spring low point. To forecast the electricity prices for weeks 2 of testing period (May 3rd – 9th), a series of training data from March 15th – May 2nd is used for model development. Based on the developed models the electricity price for week 2 is forecasted and the results are shown in the figure below. Week 2 graph show less stable in term of spikes as some of the prediction line can't follow the actual line at the graph. The value of MAPE of week 2 also more high than MAPE of week 1.The weekly percentage error for this forecast is calculated to be 18.4120% which is a considerably high error.

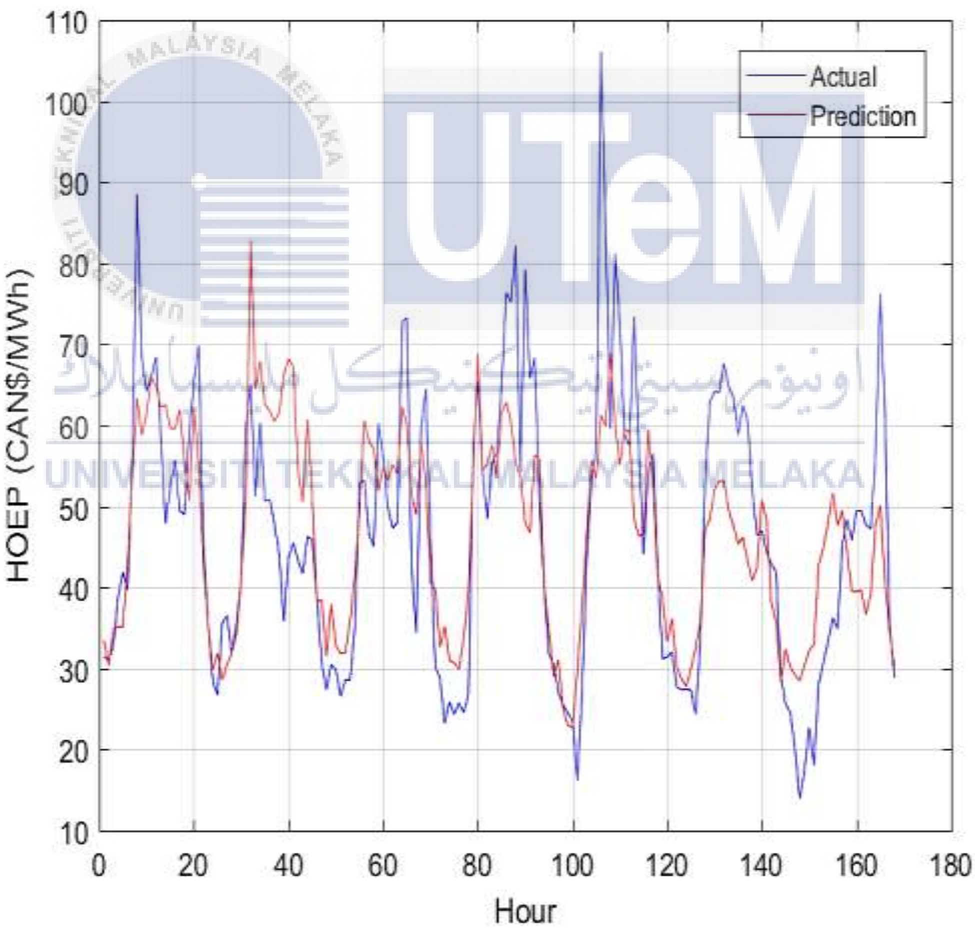

Figure 4.6 Week 2 actual vs prediction graph

4.6.3 Graph of week 3

Week 3 testing is said to be a period where the Ontario market is summer peak demand point. To forecast the electricity prices for weeks 3 of testing period (July 26th – August 1st), a series of training data from Jun 7th – July 25th is used for model development. Based on the developed models the electricity price for week 3 is forecasted and the results are shown in the figure below. Figure below showed that the graph almost stable cause the spikes at the graph are low. When the graph is almost stable will resulting lower MAPE. It shown in table 4.29 that week 3 has the lowest MAPE compared to the other weeks. The weekly percentage error for this forecast is calculated to be 15.4365% which is a considerably average error.

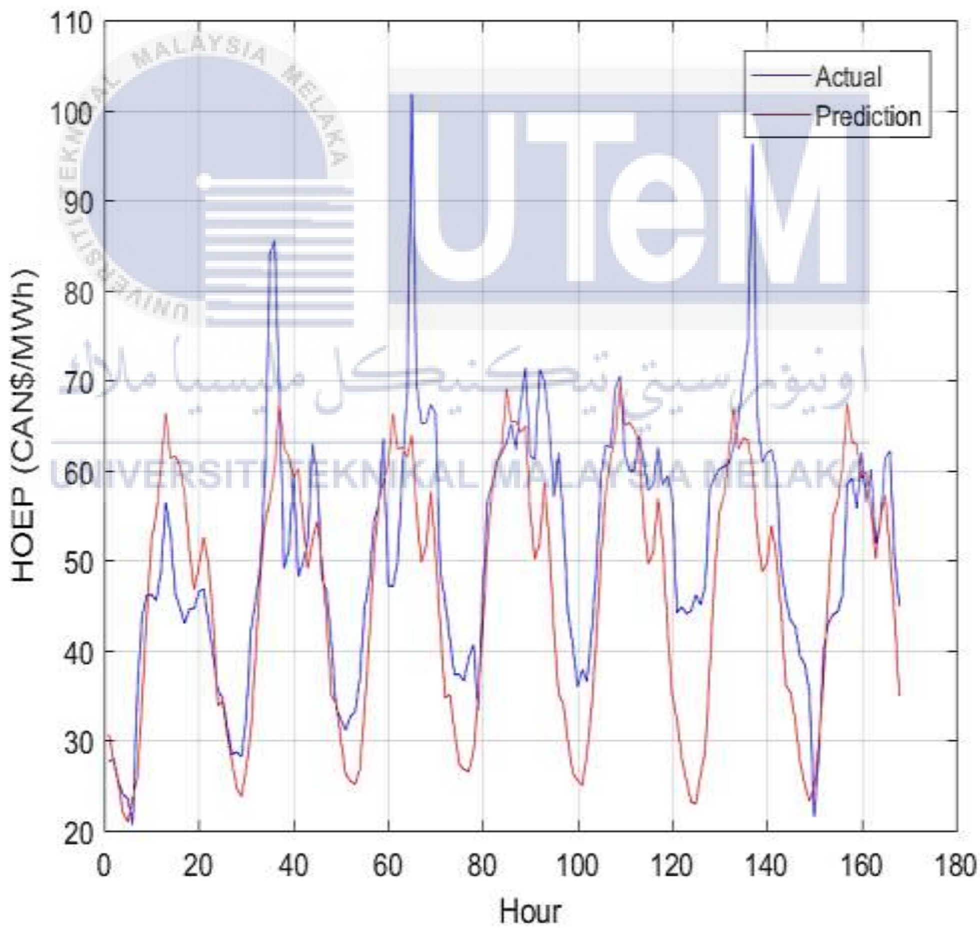

Figure 4.7: Week 3 actual vs prediction graph

4.6.4 Graph of week 4

Week 1 testing is said to be a period where the Ontario market is summer peak demand point. To forecast the electricity prices for weeks 4 of testing period (August 2nd – 8th), a series of training data from Jun 14th – August 1st is used for model development. Based on the developed models the electricity price for week 4 is forecasted and the results are shown in the figure below. From the graph, it shows that every day of actual line and prediction line are not match to each other resulting to high MAPE. The prices for the first 3 days are high volatile that cause spikes in the graph. The weekly percentage error for this forecast is calculated to be 21.0689% which is a considerably high error.

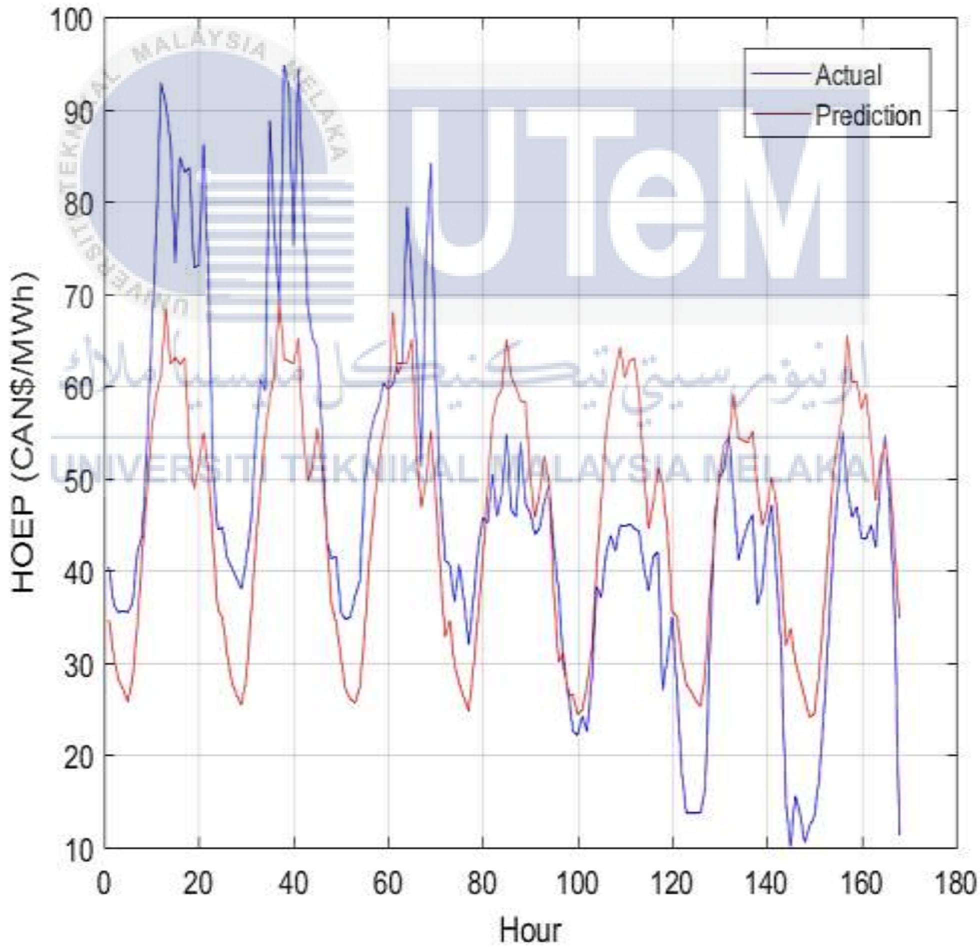

Figure 4.8: Week 4 actual vs prediction graph

4.6.5 Graph of week 5

Week 5 testing is said to be a period where the Ontario market reached its high demand winter point. To forecast the electricity prices for weeks 5 of testing period (December 13th – 19th), a series of training data from October 25th – December 12th is used for model development. Based on the developed models the electricity price for week 5 is forecasted and the results are shown in the figure below. The prediction line for every day in this testing week almost can’t match up to the actual line. It shows that the price of the week testing is quite high that resulting to high value of MAPE. The weekly percentage error for this forecast is calculated to be 18.1335% which is a considerably high error.

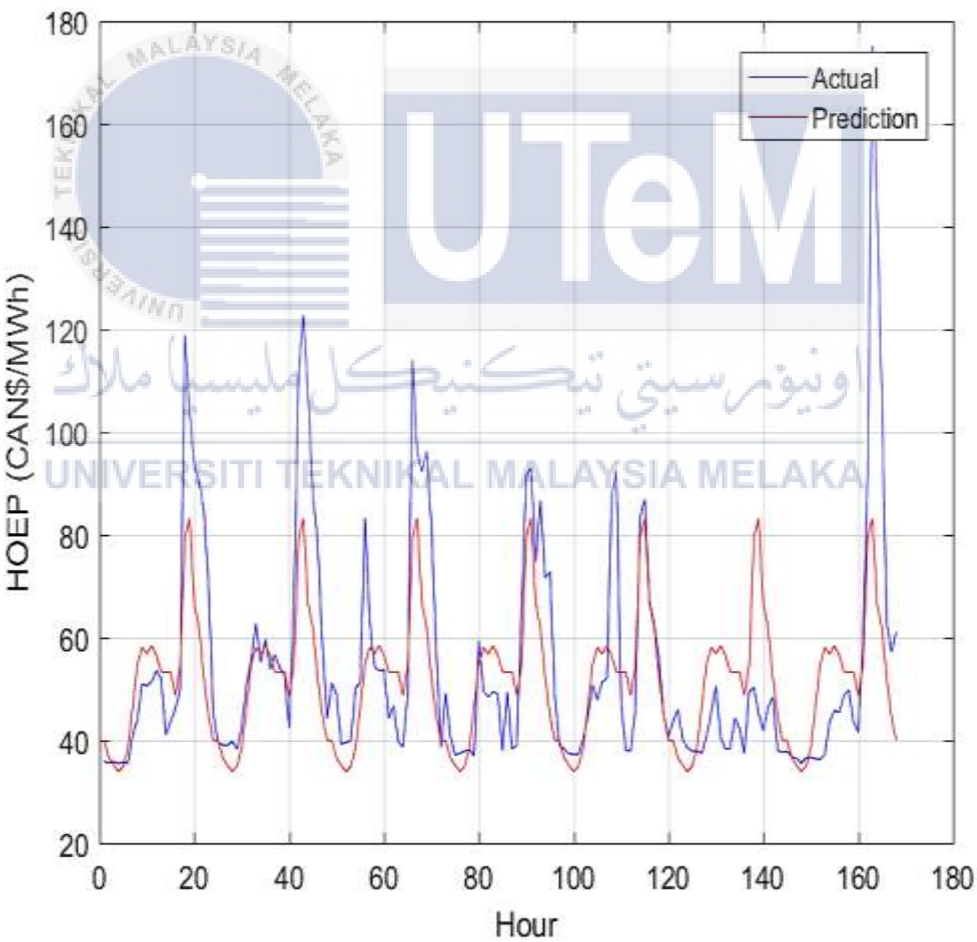

Figure 4.9: Week 5 actual vs prediction graph

4.6.5 Graph of week 6

Week 6 also testing is said to be a period where the Ontario market high demand winter point. To forecast the electricity prices for weeks 6 of testing period (December 20th – 26th), a series of training data from October 11th – December 19th is used for model development. Based on the developed models the electricity price for week 6 is forecasted and the results are shown in the figure below. Week 6 shows it has the worst performance based on the MAPE. Besides, the graph of actual HOEP has too many high spikes and the highest volatility. In the winter, the demand to high was causing the high value of electricity price in the market. The weekly percentage error for this forecast is calculated to be 27.0930% which is a considerably high error.

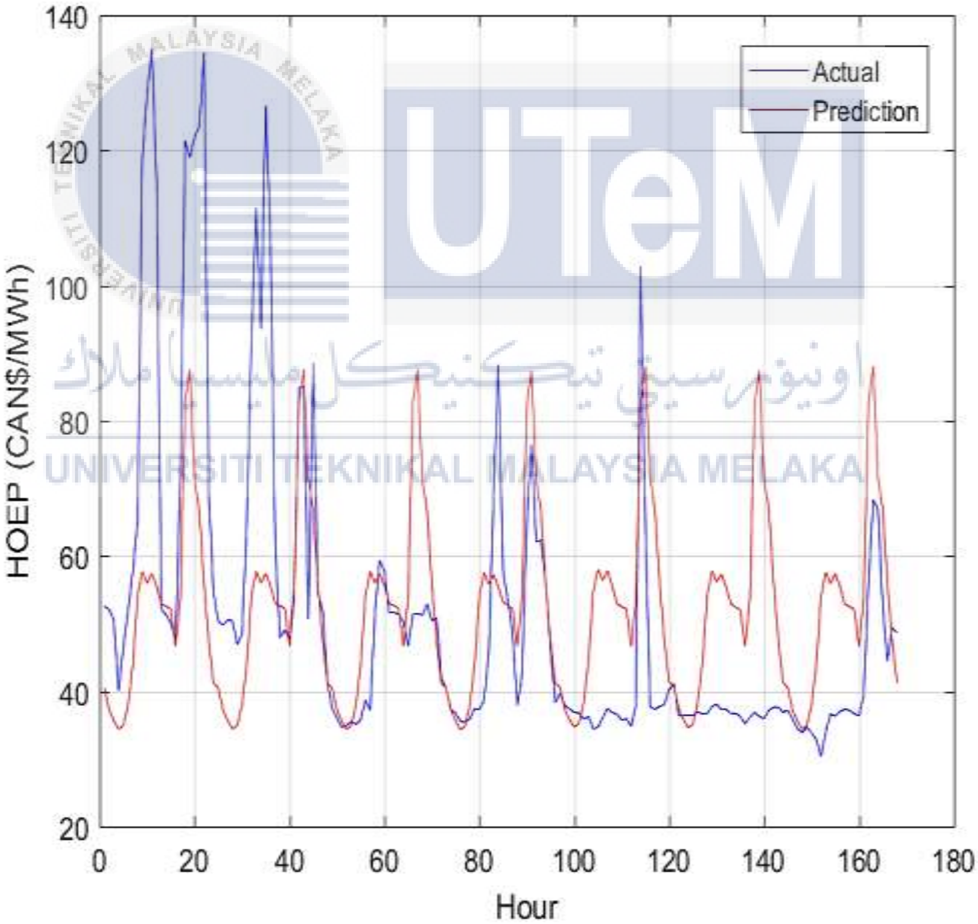

Figure 4.10: Week 6 actual vs prediction graph

4.7 Comparison of MAPE with Other Existing Methods

Based on the result, BFOA+LSSVM can be compared with two other models. First model is known as Pre-Dispatch Prices (PDP). The PDP is a technique that generated based on the most recent available market information to give the market participant with an estimate of the real time HOEP [24]. Second model is Multivariate adaptive regression splines (MARS). MARS technique is a non-parametric regression that used for application of forecasting and data mining [26]. From table below the proposed method surpasses other methods with the average MAPE of 19.51% for those six weeks of testing. Meanwhile the previous research that produces the best MAPE was 20.40% (Hamid Zareipour, C.A.Cañizares, Kanishka Bhattacharya, James A. Thomson, 2006). Furthermore, the proposed method performs the best MAPE for Week 1, 2 and 5 as compared with other model. In addition, the MAPE during Week 4 and 6 shows the worst performance for the proposed method LSSVM + BFOA. It can be concluded that the worst performance caused by the fluctuation of price during Week 4 and 6.

Table 4.30 Comparison proposed model vs other model

Method			**Week 1**	**Week 2**	**Week 3**	**Week 4**	**Week 5**	**Week 6**	**Average**
[25]	(2006)	PDP	39.7	30.3	36.9	31.6	60.2	37.3	40
[26]	(2006)	MARS	25.8	22.6	12.5	14.9	21.7	25	20.4
Proposed Method LSSVM+BFOA			16.89	18.41	15.43	21.06	18.13	27.09	19.51

4.8 Conclusion on chapter 4

Correlation analysis is performed to observe the significant features for electricity price forecasting. From the preliminary result, it shows a higher correlation of demand with the targeted price as compared to the past price with targeted price. So, only the past seven days demand will be used as features for training and forecasting processing.

All the 168 inputs are to be optimized by BFOA and to be fed into LSSVM network. On the others hand, BFOA optimizes parameters of LSSVM which are gamma and sigma. BFOA parameter also must be tuning to select the best value for each week because each bacterium has their own behavior that suitable for each week of training and testing. The best value of each parameter then can produce the best MAPE for the proposed method.

In conclusion, the LSSVM+BFOA are a simpler forecast electricity model compared to the other methods that has been mentioned in table 4.30. LSSVM+BFOA give the average MAPE is 19.51% which is lowest compared to other existing methods.

CHAPTER 5

CONCLUSION AND RECOMMENDATION

Electricity price forecasting is an essential task in power system operation and planning. Short term forecast model would be useful for both producer and consumer in developing bidding strategies or negotiation skills either in the pool market or through bilateral contracts. An accurate forecast model enables the power producer at generation sites to review and change the bids of supply and price prior to the dispatch day or hour. Hence, the output from the generators can be managed based on the price forecast to gain a maximum profit. Meanwhile, consumers can use the developed model to manage and maximize their consumption or hedge themselves against price spike occurrences.

Based on the problem statement, Least Square Support Vector Machine (LSSVM) is a good forecast model compared to the other stand-alone forecast model. Bacterial Foraging Optimization Algorithm (BFOA) that acts as optimizer for LSSVM is good choices to forecast electricity as BFOA optimize features and parameters selection. In term of accuracy, some methods may be better that LSSVM+BFOA due to the complex occurrences compare this method because this method is simpler.

From the configuration of LSSVM+BFOA shows the value of parameter for each week is different due to random initialization of bacteria. The value of parameter that selected gives the better MAPE result compared to the other value. However, the fluctuations in electricity price affect the result of MAPE for certain weeks that cause a bad performance.

As recommendations, the volatility analysis must be implemented to observe the volatility of price series in a year. This is to improve the analysis in the area of price spike that occurs. In addition, to improve the value of MAPE, LSSVM-BFOA can be applied for very short term period such as an hour ahead price forecasting as the price of electricity is changing hourly. In future, the application of the developed models tested on Ontario electricity market in year 2004 can be applied in Malaysia when the deregulated electricity market exists.

REFERENCES

[1] a. Ikonomopoulos, D. Bargiotas, M. Alamaniotis, L. H. Tsoukalas, and a. Alamaniotis, "Day-ahead Electricity Price Forecasting using Optimized Multiple-Regression of Relevance Vector Machines," *8th Mediterr. Conf. Power Gener. Transm. Distrib. Energy Convers. (MEDPOWER 2012)*, pp. 51–51, 2012.

[2] I. Azmira, K. Siah, and T. Khawa, "Feature Selection and Parameter Optimization With Ga-Lssvm in Electricity Price Forecasting," *J. Electr. Eng.*, vol. 15, no. 1, pp. 159–166, 2014.

[3] I. A. Wan Abdul Razak, I. Zainal Abidin, Y. Keem Siah, A. A. Zainul Abidin, T. K. Abdul Rahman, and A. Ahmad, "A Novel Method of BfOA-LSSVM for Electricity Price Forecasting," *ARPN J. Eng. Appl. Sci.*, vol. 11, no. 8, pp. 4961–4968, 2016.

[4] I. A. W. A. Razak, I. Z. Abidin, K. S. Yap, A. A. Z. Abidin, T. K. A. Rahman, and M. N. M. Nasir, "A novel hybrid method of LSSVM-GA with multiple stage optimization for electricity price forecasting," *PECON 2016 - 2016 IEEE 6th Int. Conf. Power Energy, Conf. Proceeding*, pp. 390–395, 2017.

[5] R. K. R. A. Agrawal, "An Introductory Study on Time Series Modeling and Forecasting," *arXiv Prepr. arXiv1302.6613*, vol. 1302.6613, pp. 1–68, 2013.

[6] J. Zhang, J. Han, R. Wang, and G. Hou, "Day-ahead electricity price forecasting based on rolling time series and least square-support vector machine model," *2011 Chinese Control Decis. Conf.*, pp. 1065–1070, 2011.

[7] E. G. Kardakos, M. C. Alexiadis, S. I. Vagropoulos, C. K. Simoglou, P. N. Biskas, and a G. Bakirtzis, "Application of time series and artificial neural network models in short-term forecasting of PV power generation," *Power Eng. Conf. (UPEC), 2013 48th Int. Univ.*, pp. 1–6, 2013.

[8] S. Katara, A. Faisal, and G. M. Engmann, "A Time Series Analysis of

Electricity Demand in Tamale , Ghana," vol. 4, no. 6, pp. 269–275, 2014.

[9] E. N. Chogumaira, S. Member, T. Hiyama, and S. Member, "Training Artificial Neural Networks for Short- term Electricity Price Forecasting," no. 1, pp. 1–4, 2009.

[10] A. R.-K. and B. M. Ali Shiriz, Mohammad Afsharx and C. of E. zSNL/CIPCE, School of Electrical and Computer Engineering, "Electricity Price Forecasting Using Support Vector Machines by Considering Oil and Natural Gas Price Impacts," pp. 2–6, 2015.

[11] A. E. K, "Forecasting the Electricity Price in Iran Power Market : A Comparison between Neural Networks and Time Series Methods," 2014.

[12] T. Papadimitriou, P. Gogas, and E. Stathakis, "Forecasting energy markets using support vector machines," *Energy Econ.*, vol. 44, pp. 135–142, 2014.

[13] X. Yan and N. A. Chowdhury, "Mid-term electricity market clearing price forecasting using multiple least squares support vector machines," *IET Gener. Transm. Distrib.*, vol. 8, no. 9, pp. 1572–1582, 2014.

[14] G. Xie, S. Wang, Y. Zhao, and K. K. Lai, "Hybrid approaches based on LSSVR model for container throughput forecasting: A comparative study," *Appl. Soft Comput. J.*, vol. 13, no. 5, pp. 2232–2241, 2013.

[15] W. J. S. Gomes, A. T. Beck, R. H. Lopez, and L. F. F. Miguel, "A probabilistic metric for comparing metaheuristic optimization algorithms," *Struct. Saf.*, vol. 70, pp. 59–70, 2018.

[16] V. Sharma, "A Review of Bacterial Foraging Optimization and Its Applications," *Int. J. Comput. Appl.*, no. Ncfaaiia, pp. 9–12, 2012.

[17] Y. T. Amghar and H. Fizazi, "A Hybrid Bacterial Foraging Optimization Algorithm and," vol. 13, no. 2, 2017.

[18] P. Samui and D. P. Kothari, "Utilization of a least square support vector machine (LSSVM) for slope stability analysis," *Sci. Iran.*, vol. 18, no. 1, pp. 53–58, 2011.

[19] L. Xie and H. Zheng, “Electricity price forecasting by clustering-least squares support vector machine,” *2013 6th Int. Congr. Image Signal Process.*, no. 3, pp. 1357–1361, 2013.

[20] S. Vani, R. Kopperundevi, T. Thilagavathi, R. K. A. Presath, and R. Rajathy, “OPF-based Market clearing procedure using bacterial foraging algorithm and auction-based Market clearing procedures in a pool based electricity market - A comparison,” *2012 IEEE Students’ Conf. Electr. Electron. Comput. Sci. Innov. Humanit. SCEECS 2012*, no. 5, pp. 5–9, 2012.

[21] S. Das, A. Biswas, S. Dasgupta, and A. Abraham, “Bacterial Foraging Optimization Algorithm : Theoretical Foundations , Analysis , and Applications,” *Found. Comput. Intell.*, vol. 3, pp. 23–55, 2009.

[22] M. A. Muñoz, S. K. Halgamuge, W. Alfonso, and E. F. Caicedo, “Simplifying the bacteria foraging optimization algorithm,” *2010 IEEE World Congr. Comput. Intell. WCCI 2010 - 2010 IEEE Congr. Evol. Comput. CEC 2010*, no. 1, pp. 18–23, 2010.

[23] S. Das, A. Biswas, S. Dasgupta, and A. Abraham, “Chapter 5 Bacterial Foraging Optimization Algorithm,” *www.Softcomputing.Net*, pp. 62–73.

[24] Zareipour, Hamidreza. “Price Forecasting and Optimal Operation of Wholesale Customers in a Competitive Electricity Market.” (2006).

[25] Zareipour, Hamid et al. “Forecasting the hourly Ontario energy price by multivariate adaptive regression splines.” 2006 IEEE Power Engineering Society General Meeting (2006): 7 pp.-.

[26] Zareipour, Hamid et al. “Application of Public-Domain Market Information to Forecast Ontario's Wholesale Electricity Prices.” IEEE Transactions on Power Systems 21 (2006): 1707-1717.

[20] A. M. RAFAL WERON, "Short-term electricity price forecasting with time series models: A review and evaluation.," Complex Electricity Markets, vol. 10, pp. 231-254, 2006.

[21] ANGELICA GIANFREDA, L. G. Forecasting Italian Electricity Zonal Prices with Exogenous Variables, University of Verona, pp. 2036-4679, January 2011.