ON THE PREDICTION ACCURACY OF THE MID-TERM ELECTRICITY PRICES USING OPTIMIZED SUPPORT VECTOR MACHINE

by

Abdussalam Taher Mohamed

Submitted in partial fulfillment of the requirements
for the degree of Master of Applied Science

at

Dalhousie University
Halifax, Nova Scotia
July 2016

To my beloved parents and family, and all people I love for their encouragement, support, and love.

Table of Contents

List of Tables

List of Figures

Abstract

In the modern electricity market, it is crucial to have precise electricity price forecasting. However, few studies have focused on this area. Mid-term electricity price forecasting (MTEPF) has numerous applications, such as scheduling future power plant maintenance, risk management, planning future contracts, purchasing raw materials, and determining market pricing. To forecast electricity prices, some factors are especially significant, such as choosing the most useful price features that influence the market price, and choosing the proper prediction model that is able to predict price behavior using historical data. In forecast modeling, feature selection techniques are an important step in data pre-processing prior to creating the prediction model. Selecting the most relevant input features increases the prediction accuracy and minimizes the data and training time. In this research, various feature selection techniques are compared and analyzed. The techniques are then used as filters prior to electricity price forecasting and their influence on prediction accuracy and mean absolute percentage error (MAPE) of each selected subset is compared. The proposed SVM method and other forecasting methods are evaluated using data from the New England ISO, which is published on their official website. Optimization of SVM parameters and kernels has also been proposed in this thesis to further improve the prediction accuracy obtained by the presented SVM model. The results obtained in this research indicate that with the same input data, the optimized SVM model achieved the highest prediction accuracy. Furthermore, our research findings show that using the SVM Regression model and an optimization of its parameters can improve overall system prediction accuracy compared with other forecasting models investigated in this thesis.

List of Abbreviations and Symbols Used

MTEPF	Medium Term Electricity Price Forecast
MCP	Market Clearing Price
SVM	Support Vector Machine
SVR	Support Vector Regression
SRM	Structure Risk Minimization
ERM	Empirical Risk Minimization
$\Re$	Real Values
ε	Insensitive Loss Function
Φ	Non-linear Transformation
w	Weight Vector
b	Bias Term
$\|\|w\|\|^2$	Norm of Weight Vector
ξ	Slack Variable
C	Penalty Parameter
α	Lagrangian Multiplier
SMO	Sequential Minimal Optimization
RBF	Radial Basis Function
k	Kernel Mapping Function
γ	Free parameter of the Gaussian radial basis function
ISO-NE	Independent System Operator New England
DT	Day of the week
NGPrice	Daily Natural Gas Price
FuelPrice	Daily Generators Fuel Price
CrudeOilPrice	Daily Crude Oil Price

Energy	Daily demand for ISO-NE in GWh
PkDEMAND	Daily Peak Demand in MW
PkHr	The hour in which the daily peak demand occurred
PkDB	Dry Bulb temperature at the hour of the daily peak.
PkDP	Dew Point temperature at the hour of the daily peak.
PkWTHI	Temperature-Humidity Index at the time of the peak
MinDEMAND	Actual minimum hourly demand in MW
NEL	System's monthly actual net energy for load in GWh
SYSPeak	Actual system peak load for the day in MW
SYSPkHr	Hour of the system peak load
Net_ int	Total net interchange for the day in GWh
ANN	Artificial Neural Networks
MLP	Multilayer Perceptron
BP	Back Propagation
CNN3L	Cascade- Forward Backprop - Three Layers
LR3L	Layer Recurrent Network - Three Layers
LeastMedSq	Least Median Squared
RMSE	Relative Mean Square Error
MAPE	Mean Absolute Percentage Error
r	Correlation Coefficient
QP	Quadratic Programming
ISDA	Iterative Single Data Algorithm
L1QP	Soft-margin Minimization by Quadratic Programming

Acknowledgements

First and above all, I praise Allah for providing me this opportunity and granting me the capability to proceed successfully. I am heartily thankful to the Libyan-North American Scholarship Program and my supervisor, Dr. ME El Hawary, whose encouragement, guidance and support enabled me to develop an understanding of this subject. I have learned precious lessons from his personality, vision, and professionalism.

I am also grateful to my supervisory committee, Dr. W. Phillips and Dr. Gu, for reviewing my work, and I thank FGS and the Department of Electrical and Computer Engineering.

Chapter 1

Introduction

1.1 Overview

Price prediction plays an important role in the scheduling and administration of electricity markets. Recently, many researchers have focused on short-term electricity price forecasting [1, 2], while very little has been done to investigate mid-term electricity price forecasting, which can range from a few weeks up to one year. Researching mid-term price forecasting is necessary for many aspects of mid-term planning in electricity markets, such as maintenance scheduling, power generation dispatching, future contracting and investments [3, 4]. Mid-term electricity price forecasting is a complex task due to the length of the prediction period and the unsteady nature of electricity prices [5].

In addition, limited explanatory data currently exist for use in price forecasting. Unlike short-term forecasting, only a few attributes of historical data are available for mid-term forecasting. Many of the forecasting engines need a large set of data for training and testing, so some methods are not applicable for mid-term forecasting. In this research, we focus on using the machine learning technique (SVM) to predict electricity prices. The mid-term electricity price forecasting task is addressed, and different issues related to forecasting are considered. As well, based on the ISO-NE publicly available data, a mid-term horizon of price forecast is developed. The contributions of this research are developing mid-term electricity price forecasts for electricity markets using publicly available data, and addressing and resolving various issues associated with the electricity price forecasting problem, such as selecting input features and parameter optimization. Feature selection (pre-processing) is the process by which the best subset of attributes in the data set is automatically searched and selected. The selection of the relevant features to a target variable creates a new subset of features based on the importance of these features. Feature selection is an important step for creating and training the forecasting model with the most relevant features, which can significantly improve forecasting accuracy and minimize training time [6]. Feature selection is also important due to correlations in some features, which can affect the accuracy of the forecasting model. This technique can be used to build new features that are independent [7]. Having superfluous features can be tricky for modeling algorithms, and keeping unrelated features in the dataset can result in overfitting. In fact, overfitting of the training data can manipulate the prediction modeling and affect the accuracy. Using feature selection pre-processing on the input dataset could benefit the prediction model by decreasing dataset size, decreasing training time, and improving the prediction accuracy simply by including only the most informative features. Feature selection has already proven to be a productive research area in data mining [8] and machine learning [9] and has been applied to many other fields. Some feature selection techniques are proposed as a pre-processing tool [10, 11, 12, 13, 14].

In this research, we also conduct a study of feature pre-processing techniques and their performances in electricity price forecasting. Irrelevant correlated features are observed in most electricity price data, so effective feature pre-processing technique should be included in the forecasting model to enhance accuracy.

1.2 Thesis Objective

This thesis comprehensively studies SVM Regression to perform MTEPF and explore the models performance. Different SVM kernels are introduced and used with the model, and their performances are explored and compared. Furthermore, three benchmark models - Least Median Squared, Neural Networks (NN), and Radial Basis Function (RBF) are demonstrated in this work. Each is individually trained and tested, and the results are compared with the performance of the proposed SVM Regression model. Case study data from New England ISO are used to train and test the performance of the understudy models on real daily data.

This thesis also studies effective input feature selection. SVM Regression parameter optimization solvers are applied to select the most important informative factors to reduce the training data and improve forecasting accuracy, and different optimization solvers are explored to select the best SVM parameters.

1.3 Structure of the Thesis

Chapter 2 presents an overview of electricity price forecasting.

Chapter 3 presents a literature review of related forecasting methodologies.

Chapter 4 presents a theoretical background of SVM Regression and its parameters. It also describes the mapping kernels, quadratic programming optimization solvers and performance measures used in this research to evaluate the forecasting accuracy.

Chapter 5 presents an overview of the New England power market, data collection, and pre-processing technique, and also presents software used in this research for modeling and data analysis. Chapter 6 presents a study of implementing the process of selecting explanatory input features to increase prediction accuracy.

Chapter 7 presents the SVM approach, parameter and kernel optimization, comprehensive comparative price analysis, and a discussion of the results achieved.

Finally, Chapter 8 presents a conclusion of this research and suggests directions for possible future research work.

Chapter 2

Electricity Market Price Forecasting

2.1 Brief Description of Power Trading

Electricity is by its nature difficult to store and has to be available on demand. Consequently, unlike other products, it is not possible under normal operating conditions to keep it in stock, ration it, or have customers queue for it. Furthermore, demand and supply vary continuously. There is, therefore, a physical requirement for a controlling agency, i.e., a transmission system operator, to coordinate the dispatch of generating units to meet the expected demand of the system across the transmission grid. Markets may operate outside national boundaries. In economic terms, elec-

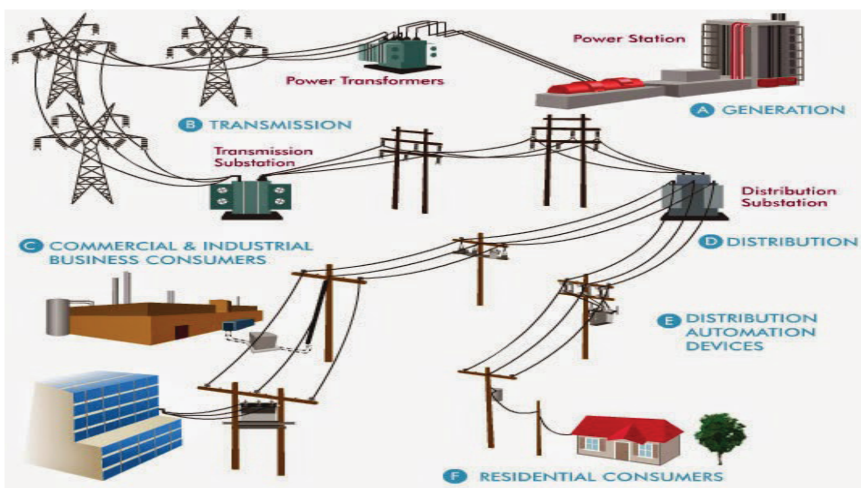


Figure 2.1: General Structure of Power System [15]

tricity (both power and energy) is a commodity capable of being bought, sold and traded. An electricity market is a system allowing purchases through bids to buy, sales through offers to sell, and short-term trades that are generally in the form of financial or obligation swaps. Supply and demand values are used by traders to set the market clearing price. Long-term trades are considered contracts between bids and offers. Electricity markets are categorized into two major classifications: regulated and deregulated electricity trading systems.

In a regulated system, the individual structure administrates generation, transmission, and distribution. The electricity price in a regulated system is determined by a local utility company to cover the generation and transmission costs, and to generate sufficient revenue for the company to expand in the future. On the other hand, in a deregulated system, many structures are granted to do electricity supply and distribution. The electricity price in deregulated systems is determined by the market supply and demand correlation.

An independent system operator (ISO) manages dealings in the electricity market. By managing the bid and offer dealings, ISOs obtain market trading knowledge, which is then used to maintain the system balance. In any electricity market, power and energy are considered commodities. Power is the rate of the transferred electric energy and is measured in megawatts (MW). Electrical Energy is the energy generated by flows of electric charges and is measured in megawatt-hours (MWh). Transmission congestion and electricity derivative markets are developed in major electricity operators trading by virtue of the reorganization of electric power systems. This reorganization progress has been often developed in parallel with the reorganization of natural gas markets.

2.2 Electricity Market Clearing Price

Modern deregulated electricity markets are organized as day-ahead and real-time venues. In day-ahead electricity markets, demand bids and generation offers determine the electricity price of the next operating day. In real-time electricity markets, five-minute-interval electricity prices are calculated based on the grid operating condition. The electricity market clearing price (MCP) commonly denotes the day-ahead electricity market price. Electricity market clearing prices are presented when the electric market is in an equilibrium state (clear of shortages and surpluses). Figure 2.3 shows how the electricity MCP is determined. When the electricity MCP exists,

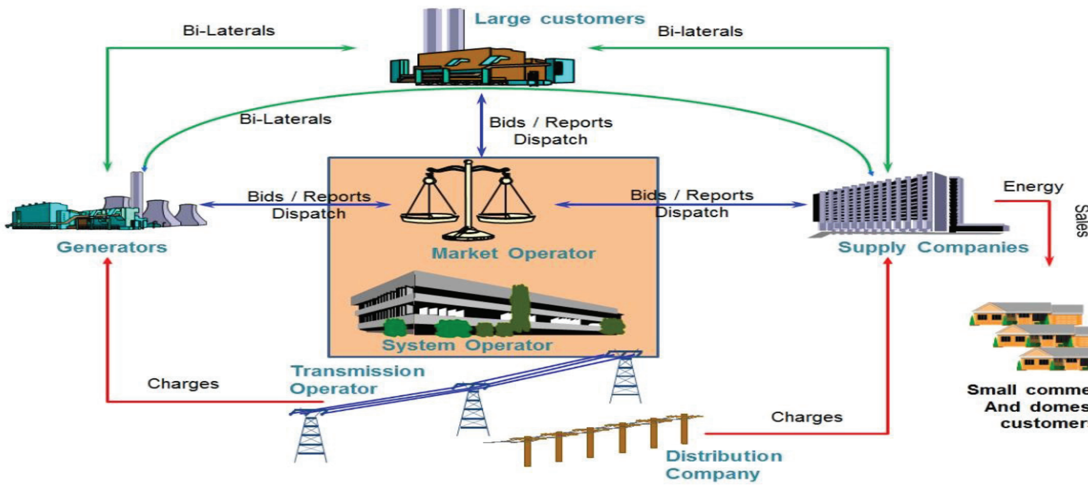


Figure 2.2: General Structure of Electricity Trading [16]

all supply offers equal or below the MCP will be picked up. To maintain market equality and avoid corruption, all picked-up supply offers will be paid the same (MCP), regardless of their offer.

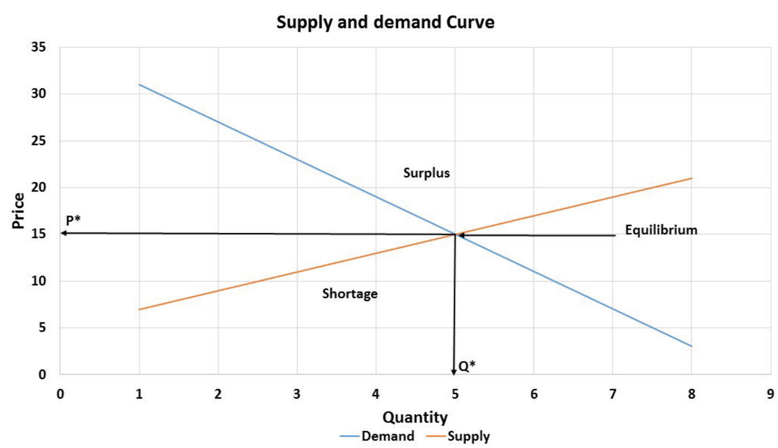


Figure 2.3: Demand and Supply Curve

The electricity MCP is determined the same way in a deregulated market. Short-term horizon forecasts of the electricity demand and reserve energy are issued by an ISO. These forecasts are updated in real time as information changes. When this forecast is available to the public, both electricity suppliers and large-size consumers will decide their participation (supplying or purchasing) and at what price. Both the suppliers and consumers can send offers to the ISO until the deadline. The ISO then accepts the lowest offered price and goes up to the higher prices until consumer demand has been met. All approved suppliers are paid based on the last accepted offer (MCP). The process of MCP determination motivates electricity suppliers to offer low prices in order to participate in the market.

2.3 Electricity Market Price Forecasting

Predicting electricity MCP is the process of forecasting a future horizon electricity price based on a given forecast of related predictors, such as demand, weather, and fuel prices. A precise forecast model significantly helps market participators to plan their bidding and purchasing strategies so as to increase their revenues.

Because of the uncertainty of some related predictors to MCP forecasting, electricity MCP is not easy to predict and is a difficult task. Some forecasting predictors have direct and linear relationships with the MCP, while some predictors have a complex relationship (nonlinear) with the MCP. Forecasting the electricity MCP can be classified into three forecast horizon types: short-term, mid-term, and long-term.

Short-term forecasting, usually referred to as day-ahead MCP forecasting, is a commonly used technique to predict electricity prices. Short-term MCP forecasting tools can help electricity suppliers develop their bidding and better manage their energy resources [17]. Electricity consumers also need short-term forecasting to establish their purchasing strategy.

Mid-term forecasting is used to predict electricity MCP on a time horizon of up to one year. It is applied in planning future contracts and in decision-making. Forecasting electricity prices on a longer time horizon is much more difficult than on a shorter one due to the unstable nature of electricity prices. There are three major difficulties involved in mid-term forecasting. Firstly, mid-term forecasting is not connected to the instant past trend and cannot employ it. Secondly, predicting peak prices (spikes) is difficult due to the unavailability of instant past data. Thirdly, sourcing long historical data to train the model is challenging due to its unavailability. Given these three main difficulties, the mid-term forecasting modeling must have a robust adaptability in the training period in order to precisely predict future prices.

The long-term MCP forecasting is dominated by economic development and political decisions. It is established annually for the determination of long-term contract/decision, network growth and development plans.

2.4 The Necessity and Importance of MTEPF

In any electricity market, there is a need for forecasting mid-term electricity prices. Such forecasting is very important for planning new contracts, arranging maintenance, dispatching power generation, purchasing fuels, and reducing financial risks. Some of MTEPF's applications are mentioned in [18, 19].

Chapter 3

Electricity Price Forecasting and Related Methodologies

3.1 Literature Review

Unlike some other electrical power topics such as electric load forecasting, electricity price forecasting is still a relatively new topic of research [20]. Electricity prices are highly unstable in nature and can experience unexpected high or low price spikes [21]. Because of this volatility, the demand and supply trends need to be monitored and balanced in real-time. Many variables may have an effect on electricity price volatility, such as sudden generation or transmission system failure, generation availability, unexpected weather changes, and instability in fuel prices. These variables can change from one market to another. Hence, it is very important to understand and analyze the relationship between forecasting input variables and electricity price, a topic which has not yet been widely explored by researchers.

A study by Dan Werner [22] shows, in the New England electricity market, the importance of ramping costs on the electricity price instability. In the absence of cost-effective storage, ramping costs are major contributors to price volatility in the electricity market. Flexible production can function like storage in guaranteeing price stability. Electricity price forecasting methods can be classified as traditional and computational intelligence methods. The traditional methods were built from the statistical methods. Currently, short-term electricity price forecasting utilizes many statistical techniques, including regression methods like dynamic regression [23], time series [24, 25], auto-regressive integrated moving average (ARIMA) [26], stochastic model [27], and wavelet transformation [28]. Early in 1993, Mbamalu and El-Hawary in [29] presented a suboptimal seasonal autoregressive model using an iteratively reweighted least squares algorithm which was used for load forecasting. Subsequently, Artificial Neural Networks (ANN) was applied to predict electricity prices due to its easy running and flexibility in processing non-linear relationships. Various ANN methods were developed in the literature of electricity price forecasting [30, 31, 32, 33]. ANN is utilized by many deregulated electric markets to predict the electricity price. More recently, a new learning method called Support Vector Machine (SVM) has emerged in electricity price forecasting [34, 35, 36], SVM performs structural risk minimization. Some research papers [34, 37], it was concluded that SVM achieves better performance compared to ANN on electricity demand forecasting. Many algorithms are used with SVM to optimize its parameters in order to improve prediction accuracy. Among these algorithms are Sequential Minimal Optimization (SMO) [38], Iterative Single Data Algorithm (ISDA) [39], soft-margin minimization by quadratic programming (L1QP) [40], and genetic algorithms [41]. A kernel-based learning method called Least Squares Support Vector Machine (LS-SVM) was proposed by Suykens and Vandewalle [42]. In classical SVMs, the solution is found by solving a set of quadratic programming (QP) problems, while with LS-SVM, a set of linear equations are solved to find a solution. An example of LS-SVM application with fuzzy logic is presented in [43]. Hybrid models have recently been developed to forecast electricity prices. These models are built using different forecasting methods and achieve better performance due to the overall combination of methods [44, 20].

3.2 Electricity Price Forecasting Models

There are two general classes of forecasting techniques: quantitative approaches and qualitative approaches. Quantitative approaches are built on algorithms of varying complexity, while qualitative approaches are built on educated guessing. In this research, quantitative approaches will be discussed.

The two main types of quantitative approaches are statistical approaches and computational intelligence approaches. A third lesser-used type time-series approaches makes predictions based on past patterns in the data.

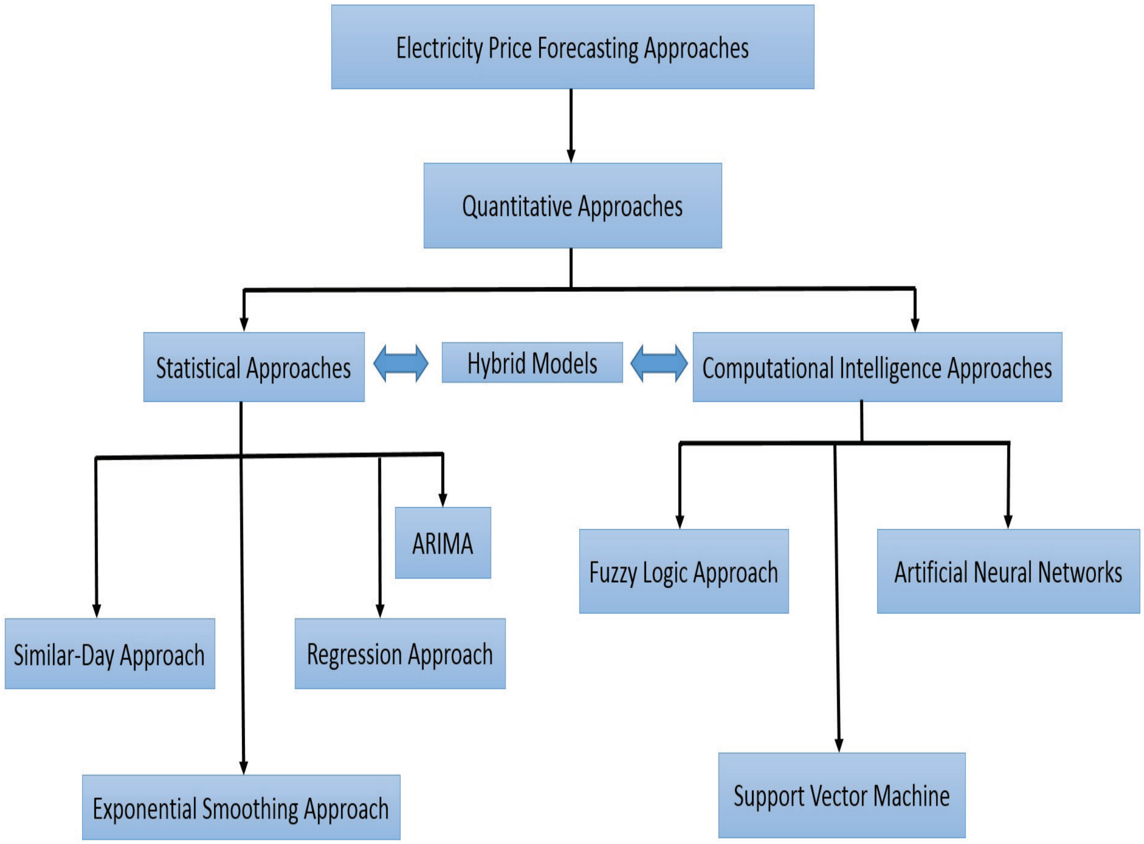


Figure 3.1: Electricity Price Forecasting Approaches

3.2.1 Statistical Approaches

Statistical approaches predict current prices by using a mathematical relationship of past prices and past related factors, such as supply, demand, fuel prices, and weather conditions. Statistical approaches have a limited capability to model the nonlinearity of electricity prices and related variables. Thus, they are rarely used to forecast electricity prices.

3.2.1.1 Similar-Day Approach

This approach is based on searching previous data for days with similar features to the target day, and considering these past values as forecasts of future prices. Similar features could include characteristics like day, time, weather or demand records. A linear regression forecast model may be used to combine similar days.

3.2.1.2 Exponential Smoothing Approach

This approach was widely used for load forecasting in the past. An exponentially weighted average of previous observations is used to construct the predicted price value.

$$\hat{Y}_t = \quad P_t = \alpha Y_t + (1-\alpha) P_{t-1} \tag{3.1}$$

P_t is the weighted average smooth value of the previous observations where the weights exponentially depend on the parameter $\alpha \in (0, 1)$.
Some advanced models of exponential smoothing have been developed to adapt time series with seasonal and trend effects.

3.2.1.3 Regression Approach

Regression is a popular and commonly used statistical method. Multiple regression is used to fit a relationship between a dependent variable (Target) and independent variables (Predictors). Multiple regression is fitted so that the sum of squared errors is minimized.

$$P_t = b_0 + b_1 X_t^{(1)} + ... + b_k X_t^{(k)} + \varepsilon_t \tag{3.2}$$

The regression approach assumes that the relationship between the dependent variable (P_t) and each of the independent variables (X_i) is linear, so it is recommended to check scatter plots of (P, X_i) to see if any relationship plot shows nonlinearity. Some data transformation techniques may be used to attain linearity. In general, linear regression is easy to use and interpret and thus should be tried first to solve a forecast problem. However, if linear regression is unable to fit a good model, then nonlinear models can be used.

3.2.1.4 Regression Trees Approach

The common regression tree building approach allows input variables to be a mixture of continuous and categorical variables. A decision tree is created when each decision node in the tree contains a test of some input variable's value. The terminal nodes of the tree have the predicted output variable values. A regression tree is designed to estimate real-valued functions instead of being used for classification methods.

Alternating Model Trees [45] is a common method for dealing with regression problems that need interpretable modeling. Model trees are decision trees with multiple linear regression models at the leaf nodes. The alternating model trees approach was proposed to solve regression problems, as the model performs very well in classification. Alternating model trees for regression have splitter and prediction nodes, and simple linear regression functions are used at the prediction nodes. In a standard decision tree, an internal node known as ”splitter node” divides the data based on a selected variable-importance test. Additive regression using forward stage-wise modeling is also applied to grow the tree. The model grows an alternating model tree by minimizing squared error.

3.2.1.5 AutoRegressive Moving Average Approach (ARMA)

ARMA is a standard time-series approach that considers and studies the random nature and time correlations of the occurrence. The price value (P_t) is expressed linearly in the $ARMA(p,q)$ model with (p) autoregressive terms and (q) moving average terms. This model was introduced by Box and Jenkins book in 1971. The equation below shows a model constructed from AutoRegressive and Moving-Average models.

$$P_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i P_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{3.3}$$

Where c is a constant, ε_t is a white noise, and φ_i and θ_i are model parameters.

The ARMA approach assumes that the price time series is non-stationary. If it is stationary, then some transformations such as differencing can be done to the series to make it non-stationary. Another approach called AutoRegressive Integrated Moving Average (ARIMA) was introduced by the same authors in 1976. It includes some minor differencing in the formulation. Overall, the AutoRegressive approach extends the basis of all time-series models.

3.2.2 Computational Intelligence Approaches

Computational Intelligence Approaches are a naturally inspired group of techniques. These techniques are developed to powerfully handle the problems that traditional approaches could not solve. Computational Intelligence techniques are able to adapt the problems complexity and non-linearity by learning and evaluation.
Computational Intelligence Approaches includes Fuzzy Logic, Artificial Neural Networks (ANN), Support Vector Machine (SVM) and evolutionary computation.

3.2.2.1 Fuzzy Logic Approach

Fuzzy Logic is a type of reasoning that identifies more than simple true and false values. With fuzzy logic, propositions can be characterized with degrees of truthfulness and falsehood. All propositions have a degree of truth that lies in the closed interval [0, 1].
Let X be an universe set. A fuzzy subset of X is then considered the couple (X, μ_A), where $\mu : X \rightarrow [0, 1]$. That is, using this notation, we consider $A = (X, \mu_A)$ and the function μ_A is named the membership function. A fuzzy inference system utilizes fuzzy logic to represent inputs to outputs. The integration of the two approaches can greatly reduce the difficulties that each approach has when working independently. A Neuro-Fuzzy model utilizes the useful properties of both approaches, so a neural network can be trained to learn the nonlinear behavior of a complex system. Afterward, the fuzzy logic can use the NN learned knowledge to generate rules and membership functions. Nima Amjady [46] proposed a new fuzzy neural network method for short-term (day-ahead) price forecasting of electricity markets. The fuzzy neural network was developed to have inter-layer and feed-forward architecture with a new hyper-cubic training technique. A combination of fuzzy logic and an efficient learning algorithm was presented to create a proper model for the non-stationary behavior and outliers of the price series. The proposed method was tested on the Spanish electricity market and it was claimed that the method provided more accurate results than the other price forecasting techniques.

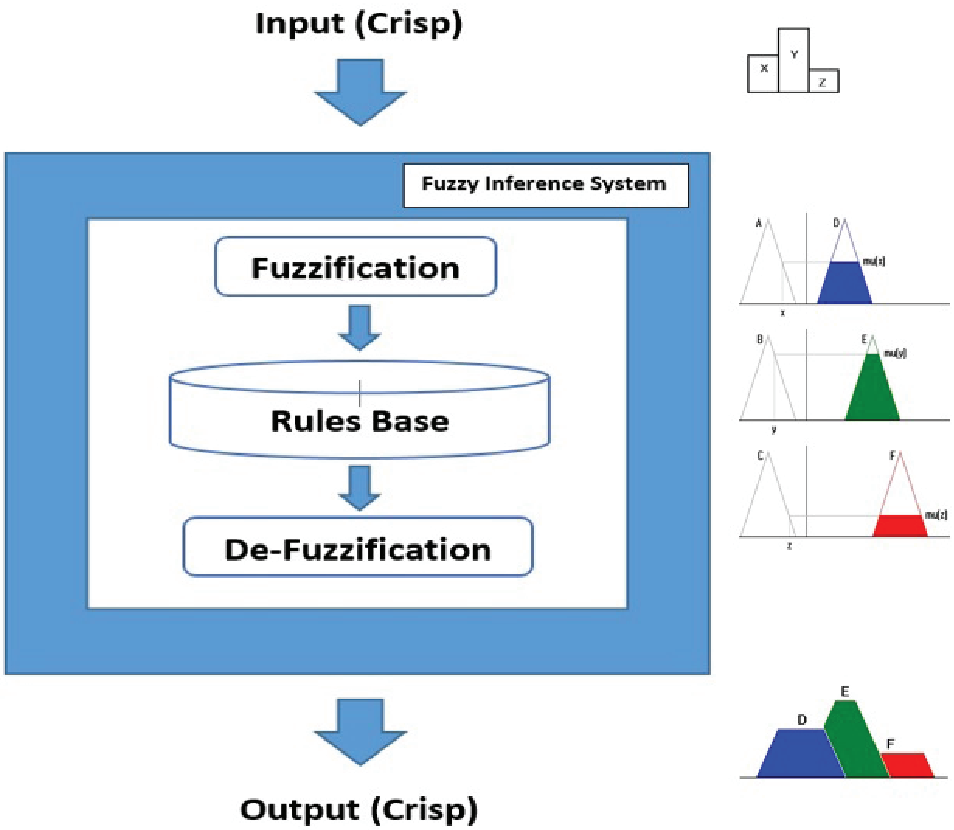


Figure 3.2: Fuzzy Inference System

3.2.2.2 Artificial Neural Networks (ANN)

Artificial Neural Network (ANN), a popular method for training and solving non-linear problems, has been widely used in forecasting. ANN has many models inspired by biological neural networks and, based on this principle, has an artificial neuron. Neural Networks are characteristically structured in layers which are interconnected by nodes containing an activation function. The input layer connects the inputs to the network via weighted connections. The artificial neuron model is processed in the hidden layers which are then connected to an output layer.

A typical ANN is shown in Figure 3.2. Individual inputs $X_1, X_2, ..., X_n$ are multiplied by connection weights $w_{1,1}, ..., w_{1,n}$. The summation equation at the hidden layer can

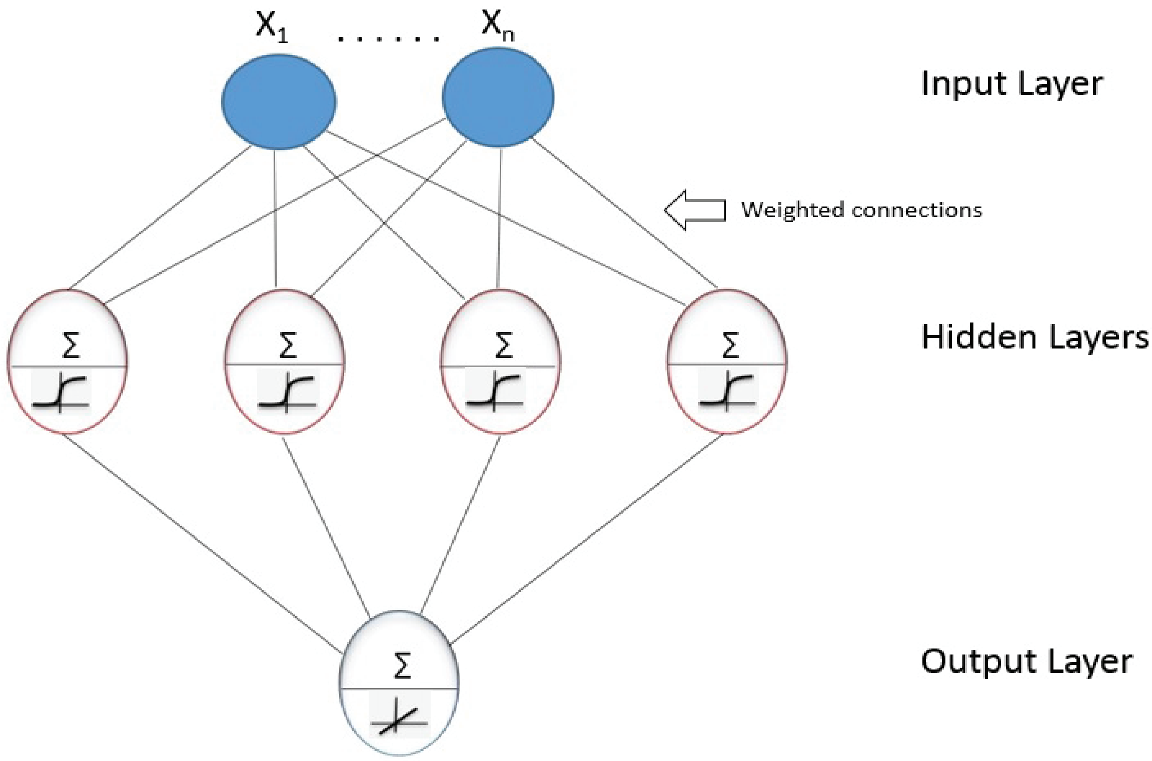


Figure 3.3: Artificial Neuron Configuration

be written as:

$$S_j(n) = \sum_{i=1}^{n} w_{ji}(n)x_i(n) \tag{3.4}$$

Where i denotes the input layer, j denotes the hidden layer, and n is the number of inputs.

Afterward, a transfer function is applied to the summation of the weighted inputs with bias b. The output of neuron y can be written as:

$$y_j = \varphi_j(s_j) + b_j \tag{3.5}$$

Where φ_j denotes the transfer function of the hidden layers.

The transfer function could be any one of the following:

- Linear Function: $f(x) = x$
- Log-sigmoid Function: $f(x) = \frac{1}{1+\exp(-x)}$
- Hyperbolic Tangent Function: $f(x) = \frac{\exp(2x)-1}{\exp(2x)+1}$

The most commonly used transfer function is the sigmoid function. The ANN has a method of learning algorithm which adjusts the connection weights according to the inputs changes. Neural Networks use many different learning algorithms, including backward propagation, feedforward propagation, etc. In each ANN cycle, a supervised process learning occurs. The ANN does a random guess of the connection weights when it is initially obtained with inputs, and then calculates the error between the answer and the actual value before adjusting the connection weights to minimize the error. In function approximation, the most widely used artificial neural networks architecture is the one with a backpropagation learning algorithm. In the backpropagation technique, a gradient descent is done within the solutions vector space to reach a global minima. The global minima is the theoretical solution with the lowest possible error. In some cases, the solution space is irregular, which causes the neural network to settle down in a local minima. The learning rules in the neural networks have mathematical terms to analyze the network performance and control the speed and momentum of the learning. The speed (β-coefficient) is the rate of convergence toward the global minima. Momentum helps neural networks to overcome local minima in the error surface and settle down at or around the global minima. To use the neural network in function approximation, three preparation steps are required: training, validation, and testing. The biggest portion of the dataset is used for the training step. In this step, the independent variables (predictors) and the dependent variable (target) are fed into the neural network and the approximation problem is solved by minimizing the error. The error is defined as the difference between the actual and desired outputs. Many training algorithms can be used to train the data, including the Levenberg-Marquardt algorithm, the Bayesian Regularization algorithm, and the BFGS Quasi-Newton algorithm. In non-linear regression problems, the fastest training function is generally Levenberg-Marquardt, which is also the default training function for a Feedforward neural network. The training algorithm estimates the neuron biases and layer connection weights. The network training will stop when its performance meets the defined goal within a specified tolerance. The validation step is used to minimize data overfitting. At the outset of the training, the validation error typically decreases. However, when the network starts to overfit the data, the error on the validation set starts to increase. The training will stop if the validation error keeps increasing for a specific number of iterations. The last stage is the testing step, where the neural network is exposed to the sample test data. Based on network performance and forecasting accuracy in the test step, more training and optimization may be required to enhance the performance and accuracy. Deepak Singhal and K. S. Swarup in [31] presented a three-layer back propagation (BP) ANN approach in a deregulated market to map complex interdependencies between electricity prices and other historical factors. They developed a day-ahead forecast of market behaviors in order to predict future electricity prices and quantities. Paras Mandal and co-authors in [32] applied a method of using a neural network model based on the similar-day technique to forecast day-ahead electricity price in the PJM market. Factors influencing electricity price were discussed and applied to train and forecast future electricity prices.

J. George and co-authors in [30] presented a new methodology for midterm energy forecasting, proposing an optimized adaptive ANN model. The proposed neural network transforms the input predictors to differences or relative differences to predict future values. The ANN parameters are optimized to enhance forecasting accuracy and neural network performance. A radial basis function (RBF) network is a type of feedforward network that utilizes radial basis functions as neuron activation functions. The K-means clustering technique is used to train the hidden layers and a supervised learning technique is used to train the output layer. The output of the RBF network is a direct combination of the inputs radial basis functions and neuron parameters.

Eibe Frank [47] discussed using supervised training of Gaussian RBF networks in WEKA software. He also discussed utilizing a RBFRegressor package, learning center locations with global and local variance parameters, and learning attribute weights.

In this research, several neural networks architectures will be employed to train and forecast the electricity price to compare their performance with the proposed SVM model.

3.2.2.3 Support Vector Machine (SVM)

A Support Vector Machine (SVM) is a modern classification and regression tool based on structural risk minimization (SRM). The SVM extended version Support Vector Regression (SVR) was developed to model time series prediction problems. SVM utilizes Vapnik's ϵ-insensitive loss function to fit nonlinear problems. Using SVM, the input data is mapped into a higher dimensional space by a nonlinear mapping technique called the kernel trick. Then a linear function is used to solve the problem in the feature space [48]. The kernel trick is employed in the SVM to do the inner product of original space vectors by performing similarity transformation in terms of the original space in the feature space. Different kernels are used with SVM to map the input data. When training an SVM model, the parameters (e.g., cost of error (C), ϵ-insensitive, kernel type and kernel parameters) need to be carefully chosen or optimized in order to increase the model accuracy. In general, some of the SVM advantages are that it can sidestep data overfitting and local minimum, minimize unpredictable data error, and needs less parameter setting compared to other models. SVM is a very robust prediction model that is widely used in classification and regression problems.

D.C. Sansom, T. Downs, and T.K. Saha [49] published one of the first papers on the applications of SVM in electricity price forecasting. In their work, they presented a comparison of a multilayer perceptron and an SVM with the same inputs and concluded that the SVM produces more consistent forecasts and requires less time for optimal training.

Dongxiao Niu and co-authors in [50] proposed a system to forecast short-term power load by using SVM with feature selection ant colony optimization, after which a fuzzy-rough method is applied to select the optimal feature subset. The results were compared with single SVM and BP neural networks.

Sapankevych, I. Nicholas, and Ravi Sankar in [51] presented a survey of time series predictions using support vector machines. The authors mentioned that the estimation process of complex applications require advanced time series prediction algorithms. The paper provided a motivation for using SVM to model nonlinear, non-stationary and not defined a prior. They claim in the survey that SVM has better performance compared to other nonlinear techniques.

The Support Vector Machine is applied in the proposed model in this research and will be discussed in detail in Chapter 4.

Chapter 4

Support Vector Machine and Related Methodologies

4.1 Introduction

Support Vector Machines (SVMs) were introduced by Vladimir Vapnik and first presented at the Computational Learning Theory (COLT) conference in 1992. All SVM features, however, were already present in machine learning since the 1960s. In 1992, the SVM features were put together to model a basic Support Vector Machine and maximal margin classifier [48]. The soft margin version was first presented in 1995 [52, 48], The soft margin allows training data points errors while the SVM fitting a model. Soft margin has a small penalty parameter to reduce model complexity and increase model generalization.
SVM is developed for solving problems of nonlinear classification and regression, while Support Vector Regression (SVR) is developed specifically to solve regression problems. SVMs implement the structural risk minimization (SRM) principle, which has proven to be more efficient than the empirical risk minimization (ERM) principle used in neural network models [53].

4.2 SVM Model

In regression problems, the training data has the form $\{(x_1, y_1), ...(x_i, y_i)\} \subset X \times \Re$, where each X denotes input space of the sample. The objective of SVR is to determine a function $f(x)$ that tolerates ε deviation from the targets y_i for the entire training data. The generic SVR estimating function takes the form:

$$f(x) = (w * \Phi(x)) + b \qquad w \in X, b \in \Re \tag{4.1}$$

Φ denotes a nonlinear transformation from n to high-dimensional space.

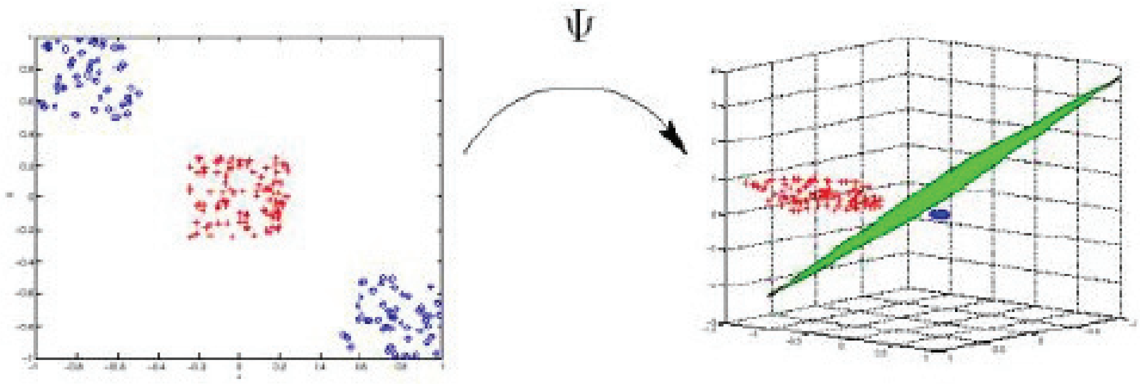


Figure 4.1: Non-linear High Dimensional Mapping [52]

The input data x is initially presented in a feature space of m-dimensional using specified nonlinear representation. This feature space then creates a linear model. Using mathematical representation, the linear model (in the feature space) $f(x, w)$ is given by:

$$f(x, w) = (w^T \Phi(x)) + b \tag{4.2}$$

Where $\Phi(x)$ is a mapping function, w represents the weight vector and b is a bias term.

Parameters w and b can be determined by minimizing the expression of the error function:

$$R_{reg}(f) = C(\sum_{i=1}^{n} |f(x_i, w) - y_i|_\varepsilon) + \frac{1}{2}||w||^2 \tag{4.3}$$

Where, $|f(x, w) - y|_\varepsilon$

$$= \begin{cases} 0, & \text{if } |f(x, w) - y| < \varepsilon. \\ |f(x, w) - y| - \varepsilon, & \text{if } |f(x, w) - y| > \varepsilon. \end{cases} \tag{4.4}$$

Where $||w||^2$ is the norm of the weight vector, the term ε represents the Vapnik's ε-insensitive loss function, and C is the penalty parameter.

If the predicted value is within the loss function zone, the approximation error will be zero. For predicted values that are outside the loss function zone, the loss is equal

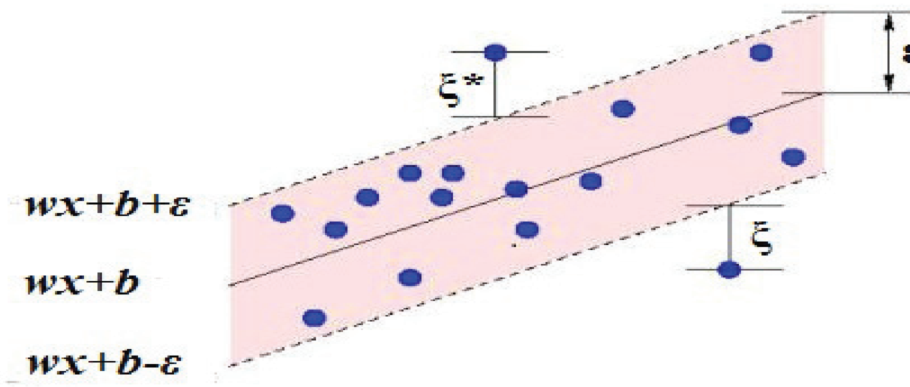


Figure 4.2: Epsilon Band with Slack Variables [52]

to the distance between the data point and the margin of ε-insensitive.
SVM Regression performs linear regression in the high-dimension feature space using ε-insensitive loss and, at the same time, tries to reduce model complexity by minimizing $||w||^2$. This can be described by introducing (non-negative) slack variables ξ_i,ξ_i^* $i = 1, ..n$ to measure the deviation of training samples outside the ε-insensitive zone. Thus SVM regression is formulated as the minimization of the following functional:

$$R_{reg}(f) = C\sum_{i=1}^{n}(\xi_i + \xi_i^*) + \frac{1}{2}||w||^2 \tag{4.5}$$

Subject to the constraints:

$$y_i - f(x_i, w) \le \varepsilon + \xi_i^*.$$
$$f(x_i, w) - y_i \le \varepsilon + \xi_i.$$

4.3 SVM Parameters

The parameters of the optimal hyperplane $f(x, w)$ can be determined by using Lagrangian multipliers;

$$w = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\Phi(x_i) \tag{4.6}$$

Kernel functions allow the dot product to be executed in a feature space of high dimension using a space data input of low-dimensional without identifying the transformation Φ.

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) k(x_i, x_j) + b \tag{4.7}$$

The Support Vector Regression (SVR) performance depends on choosing parameters, such as cost of error (C), the width of the loss function (ε), and which kernel function is used for data mapping.

SVR utilizes optimization algorithms in feature space to perform regression computation. SVR uses a quadratic formulation as an optimization algorithm. So, to improve the accuracy of the original SVM, some optimization algorithms such as Sequential Minimal Optimization (SMO), Iterative Single Data Algorithm (ISDA), and Soft-margin Minimization by Quadratic Programming (L1QP) will be utilized to improve the training process of SVM. In this electricity price forecasting research, the program WEKA 3.7 [54] is used for SVM modeling, the SVM parameters are trained from the data using Sequential Minimal Optimization (SMO) algorithm, and the RBF kernel is used for SVM modeling.

4.4 Mapping Kernels

Kernel data mapping is a technique used to change the feature representation from input space to feature space. The data is then mapped in a higher dimension, where linear relations exist among the data. A linear model can be applied to do regression computations in this space. Each kernel has an associated feature mapping ϕ, where this feature mapping ϕ takes the input $x \in X$ (input space) and maps it to F (feature space). Kernel $k(x, x^{'})$ takes two inputs and give their similarity in feature space. Kernel can be constructed by defining a mapping ϕ, such that $\forall\ x, x^{'} \in X$.

$$k(x, x^{'}) = \langle \phi(x), \phi(x^{'}) \rangle \tag{4.8}$$

4.4.1 Linear Kernel

Training an SVM with a linear kernel is not only faster than with another kernel, it also performs better in problems that are linearly separable. The linear kernel outperforms when there are a lot of variables and fewer observations. This is because mapping the data to a feature space does not really improve the performance. An SVM linear kernel requires a lower parameter setting and only the (C) regularization parameter needs to be optimized.

$$k(x_i, x_j) = x_i^T x_j \tag{4.9}$$

4.4.2 Polynomial Kernel

A polynomial kernel is a mapping function that is generally utilized by SVM and other techniques to model nonlinear mapping. This type of kernel performs similarity comparisons of the input data over the mapped data in higher dimension space, which gives a better understanding of non-linear problems.
In regression analysis, the polynomial kernel looks also to the interaction of input samples to determine their similarity. The polynomial linear regression is modeled with a degree of polynomial to find the relationship between dependent and independent variables. It is a duplicate to the higher dimension space of a polynomial kernel.

$$k(x_i, x_j) = (\gamma x_i^T x_j + c)^d \quad , \quad \gamma > 0. \tag{4.10}$$

Polynomial kernel regulating parameters are the slope γ, the constant term C and the polynomial degree d.

4.4.3 Sigmoid Kernel

The Sigmoid Kernel is a kernel utilizing a sigmoid function to map the data to a higher dimension. It is also called a Hyperbolic Tangent Kernel. The sigmoid kernel was developed from a bipolar sigmoid function. In neural networks, the bipolar sigmoid function is frequently used as a neuron's activation function. SVM model with a sigmoid kernel functions similar to a two-layer perceptron neural network.

$$k(x_i, x_j) = \tanh(\gamma x_i^T x_j + c) \quad , \quad \gamma > 0. \tag{4.11}$$

Where γ is the slope and C is the intercept constant .

4.4.4 Gaussian RBF Kernel

The Gaussian kernel is the most common kernel of the Radial Basis Function kernels.

$$k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad , \quad \gamma > 0. \tag{4.12}$$

When training SVM with the Radial Basis Function (RBF) kernel, two regulating parameters must be considered: C and γ. The parameter γ plays an important role in the performance of the kernel, and should be carefully tuned or optimized to the problem. If γ is overestimated, the kernel exponential function will act as linear and the feature space will lose its non-linear power. If γ is underestimated, the exponential function will lack regularization.

The penalty parameter C, popular in all SVM kernels, controls trade-offs between the margin and the size of the slack variables. A small C will lead to higher training errors, whereas a large C will lead to a performance similar to that of a hard-margin SVM.

4.5 SVM Optimization Solvers

Using some numerical Quadratic Programming (QP) optimization tools to solve the dual optimization problem is a traditional method. The numerical QP method is slow and impractical for handling huge data sets, generally because it is time-wasting and also needs a considerable amount of memory. Several algorithms have been developed to solve the QP problem with this limitation. The technique used by most of these algorithms is reducing the size of the QP problem by iteratively dealing with smaller QP subproblems.

4.5.1 Sequential Minimal Optimization (SMO)

The Sequential Minimal Optimization (SMO) algorithm proposed in (Platt, 1999a) sidesteps dealing with numerical QP tools by systematically solving a large number of small optimization sub- sets that includes only two Lagrange multipliers at a time. This is the smallest number of multipliers to deal with because the linear equivalence constraint $\sum_{i=1}^{n} \alpha_i y_i = 0$ in the original QP problem.

The main SMO-developed idea is to keep improving the overall objective function by fine-tuning two Lagrange multipliers at one time. The SMO algorithm analytically solves a QP subproblem for the two selected Lagrange multipliers and consequently modifies the SVM model. The SMO method's size is linear in the quantity of training data. The time complexity is described as being between linear and quadratic regarding the size of the training set for various datasets.

Alex J. Smola and Bernhard Schlkopf in [48] modified the original SMO algorithm to solve regression problems. Shevade et al. in [55] presented more modifications to SMO for better performance.

4.5.2 Iterative Single Data Algorithm (ISDA)

The Iterative Single Data Algorithm (ISDA) was developed to sidestep the use of typical QP solvers. The most important feature of ISDA is that it deals with one data point at a time to develop the objective function. The Kernel AdaTron (KA) is the old version of ISDA for SVM, which uses kernel functions to map data into SVM feature space and performs AdaTron learning in the feature space. The first addition of the Kernel AdaTron algorithm for regression was presented in [56], where the KA algorithm is built on a gradient ascent method. KA is a straightforward algorithm, it uses exponential method to quickly converge the quadratic problem and find the optimal solution. The development of ISDA is based on the notion that the equality constraint of the SVM optimization problem can be removed when a positive definite kernel is used.

The straightforwardness of ISDA is the result of the circumstance that, in the training step, the Lagrangians constraints do not need to be achieved. Therefore, the ISDA is an ideal technique for solving outsized SVM problems containing large training data sets. The ISDA will solve the QP faster compared to other typical algorithms, while performing the same generalization results.

4.5.3 L1 Soft-margin Minimization by Quadratic Programming (L1QP)

In the soft margin SVM, a linear approximation function is considered in the feature space. By minimizing the approximation error for the training data and the response, the approximation function can be determined.
This can be reached by minimizing:

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{M}\zeta_i \tag{4.13}$$

Subject to the constraints:

$$y_i(w^T\phi(x_i) + b) \geq 1 - \zeta_i \qquad \text{for} \qquad i = 1, \dots, M \tag{4.14}$$

Where w is the weight vector, ζ_i is the positive slack variable, x_i is training data, y_i is the response variable, C is the margin parameters, and M is the number of the training data.
L1 soft-margin SVM is widely used for function approximation. L1 SVM solutions depend on the margin C parameter, which is used to enforce the constraints [29]. If C is extremely large or greater than the biggest Lagrange multiplier (α_i) calculated, the margin is essentially hard. If C is smaller than the biggest original Lagrange multiplier (α_i), the margin is soft.

4.6 Performance Evaluation

To evaluate model forecasting results, specific performance measures are used. The Relative Mean Square Error (RMSE) for the predicted period is calculated by (4.15), and for each model the Mean Absolute Percentage Error (MAPE) is also calculated. MAPE is often utilized to compare forecast performance; it is calculated by (4.16) and has valuable statistical properties which are often useful for the purpose of reporting. MAPE is also the most commonly used evaluation measure, being expressed in generic percentage terms that are understandable to users [57].

Correlation coefficient (r) is a coefficient that measures the strength of the statistical relationship (degree of linear dependence) between variables. In other words, the coefficient shows how close two variables lie along a line. In this research, it is used to measure the correlation coefficient between the observed values (Y_i) and the fitted values ($\hat{Y}_i$). r is determined by Eq.(4.17).Given n historical observations of electricity price (y_i) and the corresponding forecasted price ($\hat{y}_i$) for $i = 1, \ldots, n$, the equations are defined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \tag{4.15}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} |\frac{y_i - \hat{y}_i}{y_i} \times 100| \tag{4.16}$$

$$r = \frac{cov(X_i, Y)}{\sqrt{var(X_i) var(Y))}} \tag{4.17}$$

Where cov designates the covariance and var the variance.

Chapter 5

Data Collection and Pre-Processing

5.1 Introduction

Data collection and pre-processing are significant steps in regression analysis and forecasting. This research focuses on utilizing SVM to predict mid-term electricity prices. Prediction accuracy is mainly dependent on the selection of the model input data. An input data collection process using feature selection techniques has been developed in this research to help forecasting models obtain the best prediction accuracy. The process will be discussed in detail in the next chapter. Data pre-processing is an important standard technique which is used in machine learning and data analysis to obtain optimal results. Data pre-processing involves initial preparation of data, which includes various tasks such as data normalizing, data cleaning, and data reduction. In the context of electricity price forecasting, the most frequently used data pre-processing actions are outliers (abnormal prices) detection and manipulation, normalization, and data transformation [58]. In this research, outlier operations are not performed because abnormal values explain the real nature of the data, and manipulating them may result in losing their informative feature. In the present work, only data normalization and reduction are applied, as normalization has been found to enhance prediction accuracy [58].

ISO-NE Electric Power Market data is used as an example in this research.

5.2 Electric Power Market - New England (ISO-NE)

Recently, electricity systems located near to each other have been integrated as a result of high demand in some areas that had been causing frequent sudden shortage and/or surpluses. Forming a larger integrated system helps keep the demand and supply stable, increases system reliability, and delivers improved service quality.

The ISO-NE electricity market [59] agrees to participate with any market that meets its competency and thus works as an open market. Electricity suppliers of ISO-NE and Maritime Canada have transmission access and wholesale markets for both systems.

The New England wholesale electricity market operates much like a commodity market. Buyers and sellers of electricity gather (electronically) and bid on sales and purchases of electricity. The price of electricity is set in the marketplace according to demand (how much electricity parties want to buy) and supply (how much electricity is for sale). The electricity that is available for sale in the New England wholesale marketplace is usually generated primarily by natural gas and nuclear generation. The pricing in the New England wholesale marketplace largely determines the pricing that Maritime Electric pays for the electricity it needs to purchase outside of electricity supplied by long-term contracts.

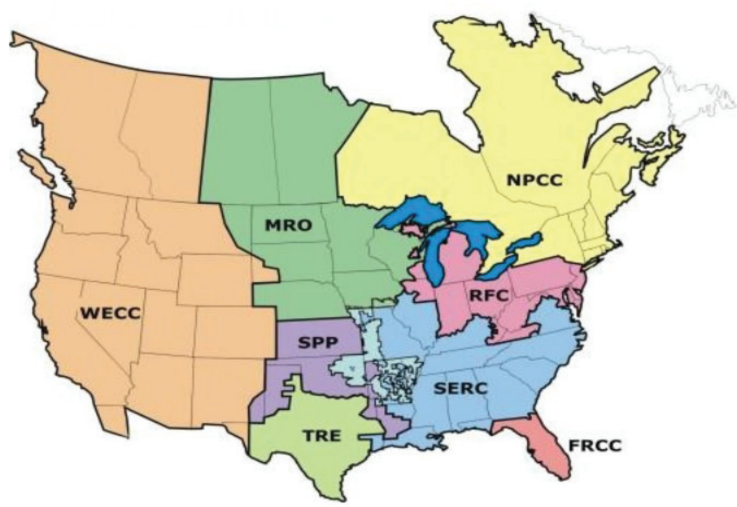


Figure 5.1: North America Electric Reliability Corporation Regions

ISO-NE grid market stats [59]:

- 350 generators.
- 31,000 MW of generating capacity.
- 600 MW of active demand resources.
- 1,700 MW of energy efficiency with capacity supply obligations.

- $7.8 billion in transmission investment since 2002.
- Over 400 buyers and sellers in wholesale electricity marketplace; $7.2 billion traded in 2015.

5.3 Data Collection

Electricity price can be predicted by processing models that evaluate a set of variables such as fuel prices, generation capacity, electricity supply, demand, and weather. In any typical electricity price forecasting model, all available variables that affect the electricity price should be included. Practically and in real life, however, only variables with high informative and feature importance should be included to reduce the data size, speed up the forecast process, and increase the forecast accuracy. Based on the factors mentioned, a data and features selection process is proposed and presented using different attributes and search methods in order to select the optimal input features that will be taken into account to do forecasting in this research.

In selecting the best influencing features, the ISO-NE actual daily data is split into two parts: training dataset and testing dataset. The testing dataset will be used as out-of-sample data to evaluate the model forecasting. The training dataset consists of data points from March 2003 to March 2009, while the testing dataset consists of data points from March 2009 to September 2009.

Once the proposed forecasting models are trained using historical data, future electricity prices are forecasted using the proposed trained model. The forecasting work in this research is based on the assumption that all forecasting input data have already been precisely predicted.

5.4 Input Data Pre-Processing

The ISO-NE daily selected forecasting input features dataset is initially filtered for outliers or missing values. In this research, electricity price outliers filtering is not performed because abnormal values explain the real nature of the data, and manipulating them may result in losing the informative feature. An interquartile range filter

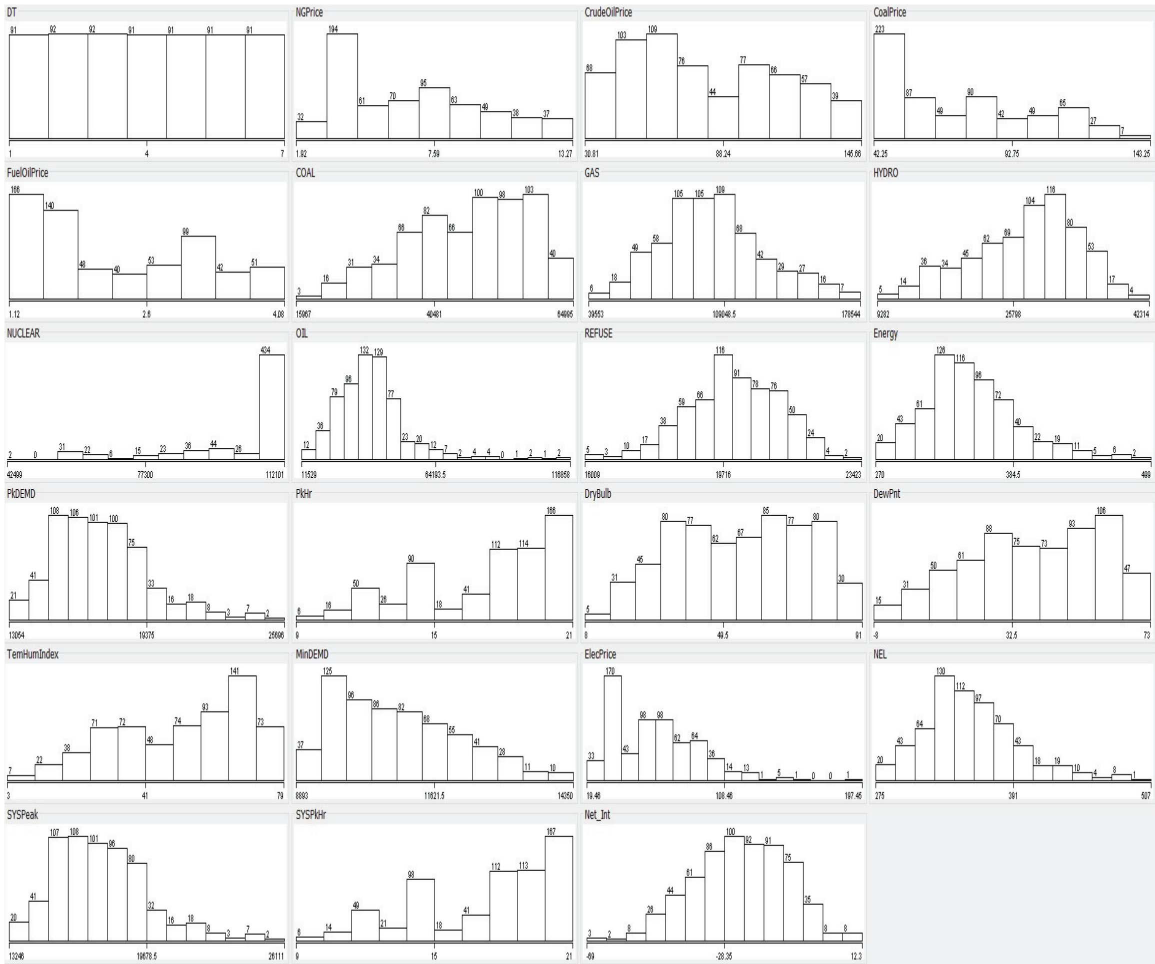


Figure 5.2: Attributes Statistical Distribution

is used to detect outliers and extreme values in the input dataset based on interquartile ranges.

Outliers:

$$Q3 + OF * IQR < x \leq Q3 + EVF * IQR.$$

or

$$Q1 - EVF * IQR \leq x < Q1 - OF * IQR.$$

Extreme values:

$$x > Q3 + EVF * IQR.$$

or

$$x < Q1 - EVF * IQR.$$

Where Q1 is 25% quartile, Q3 is 75% quartile, IQR is Interquartile Range (difference between Q1 and Q3),OF is Outliers Factor, and EVF is Extreme Value Factor. In

Figure 5.3: Outliers and Extreme Value Filtering

forecasting, the most recommended data, the pre-processing practice is normalization. Normalization converts each data in each variable into a value between 0 and 1 with respect to the variable's maximum and minimum values.

$$z_i = \frac{x_i - min(x)}{max(x) - min(x)} \tag{5.1}$$

Where $x = (x_1, \ldots, x_n)$ and z_i is the i^{th} normalized data.

Normalizing the input data enhances the forecasting accuracy so that better results can be obtained. Normalization can eliminate the problem whereby bigger values may terminate lower values. It is advisable to normalize the training data and the testing data together to avoid mismatching the data parameters.

De-normalization is applied when the forecasting of the electricity price is completed. It converts the forecast ratio values to real values. De-normalization can be defined as:

$$x_i = z_i \times (max(x) - min(x)) + min(x) \tag{5.2}$$

5.5 Use of Software

All graphs in this thesis were generated using MATLAB [60]. IBM SPSS Statistics [61] was used to perform input feature correlations analysis. Figures of neural networks architectures were generated using NN toolbox in MATLAB. Data pre-processing, price forecasting training and testing in the neural network method were performed using MATLAB script with NN toolbox. SVR modeling, data pre-processing and forecasting results were performed using WEKA [54]. SVR optimization solvers training and forecasting results were performed using MATLAB.

Chapter 6

Effective Input Features Selection for Electricity Price Forecasting

In forecast modeling, choosing feature selection techniques is a crucial step in data pre-processing prior to creating the prediction model. Selecting the most significant input features is important for increasing the prediction accuracy and minimizing the data and training time. In this chapter, some feature selection techniques are compared and analyzed and then used as a filter prior to electricity price forecasting. The influence of feature selection techniques on prediction accuracy and mean absolute percentage error (MAPE) of each selected subset is also compared.

6.1 Introduction

Feature selection is a process in which the best subset of attributes in the dataset is automatically searched and selected. Selecting the relevant features of a target variable and creating a new subset of features based on the importance of these features is very important. Feature selection (pre-processing) is an important step in creating and training the forecasting model with the most relevant features, which can significantly improve forecasting accuracy and minimize training time [62]. Feature selection is also important due to the correlation of some features which can affect the accuracy of the forecasting model. Feature selection techniques can be used to build new features that are independent [63].

The presence of excessive training features can misguide the forecasting modeling. Hence, including uncorrelated features in the training data could affect the forecasting accuracy due to redundancy. Minimizing the modeling time, reducing the data size, and enhancing forecasting accuracy are the main advantages of implementing features selection prior to modeling a forecast.

Feature selection has been a productive research area in data mining [8] and machine learning [9] and has been applied to numerous other fields. Some feature selection techniques are proposed as pre-processing tools [10, 11, 12, 13, 14].

In this thesis, a study of feature pre-processing techniques is conducted and their performances in electricity price forecasting measured. Irrelevant correlated features are observed in most electricity price data, so effective feature pre-processing is essential for an element to be included in the forecasting model to enhance higher accuracy.

6.2 Correlation Criteria

To understand the relationship between any two variables, a covariance formula is used to show the direction of the relationship and its relative strength. The correlation coefficient calculation takes the variable covariance and divides it by the product of the standard deviation of the two variables. This will ensure correlation between values of -1 and +1. A correlation of +1 can be interpreted as both variables having a perfect positive correlation relationship, and a -1 indicates they have a perfectly negative correlation. The Pearson correlation coefficient is commonly used as a measure of the linear dependencies degree between two variables. It is defined as

$$R_i = \frac{cov(x_i, y)}{\sqrt{var(x_i)var(y)}} \tag{6.1}$$

Where *cov* designates the covariance and *var* designates the variance.

The estimated equation of the Pearson correlation is given by:

$$R_i = \frac{\sum_{k=1}^{m}(x_{k,i} - \bar{x}_i)(y_k - \bar{y})}{\sqrt{\sum_{k=1}^{m}(x_{k,i} - \bar{x}_i)^2 \sum_{k=1}^{m}(y_k - \bar{y})^2}} \tag{6.2}$$

Figure 6.1 shows the variables correlation coefficients with respect to electricity price. The correlation table illustrates how good the variable correlation factor is and in which direction it points, and also illustrates how significant the correlation is by

		ElecPrice
ElecPrice	Pearson Correlation	1
	Sig. (2-tailed)	
	N	821
DT	Pearson Correlation	-.080*
	Sig. (2-tailed)	.022
	N	821
NGPrice	Pearson Correlation	.898**
	Sig. (2-tailed)	.000
	N	821
CrudeOilPrice	Pearson Correlation	.675**
	Sig. (2-tailed)	.000
	N	821
CoalPrice	Pearson Correlation	.746**
	Sig. (2-tailed)	.000
	N	821
FuelOilPrice	Pearson Correlation	.770**
	Sig. (2-tailed)	.000
	N	821
COAL	Pearson Correlation	.481**
	Sig. (2-tailed)	.000
	N	821
GAS	Pearson Correlation	.181**
	Sig. (2-tailed)	.000
	N	821
HYDRO	Pearson Correlation	.074*
	Sig. (2-tailed)	.035
	N	821
NUCLEAR	Pearson Correlation	-.139**
	Sig. (2-tailed)	.000
	N	821
OIL	Pearson Correlation	.453**
	Sig. (2-tailed)	.000
	N	821
REFUSE	Pearson Correlation	.499**
	Sig. (2-tailed)	.000
	N	821

		ElecPrice
Energy	Pearson Correlation	.434**
	Sig. (2-tailed)	.000
	N	821
PkDEMD	Pearson Correlation	.392**
	Sig. (2-tailed)	.000
	N	821
PkHr	Pearson Correlation	-.066
	Sig. (2-tailed)	.059
	N	821
Dry Bulb	Pearson Correlation	.007
	Sig. (2-tailed)	.849
	N	821
DewPnt	Pearson Correlation	-.085*
	Sig. (2-tailed)	.015
	N	821
TemHumIndex	Pearson Correlation	-.042
	Sig. (2-tailed)	.235
	N	821
MinDEMD	Pearson Correlation	.418**
	Sig. (2-tailed)	.000
	N	821
NEL	Pearson Correlation	.429**
	Sig. (2-tailed)	.000
	N	821
SYSPeak	Pearson Correlation	.388**
	Sig. (2-tailed)	.000
	N	821
SYSPkHr	Pearson Correlation	-.053
	Sig. (2-tailed)	.126
	N	821
Net_Int	Pearson Correlation	-.197**
	Sig. (2-tailed)	.000
	N	821

Figure 6.1: Variable Correlation Coefficients

showing the p-value of each variable test.

6.3 Feature Selection

Feature pre-processing techniques employ different search strategies and evaluation criteria. In this section, we will discuss these strategies and criteria. The feature selection process is separated into two parts:

- **Attribute Evaluator**: a technique that evaluates the merit of features subsets by considering the degree of duplication between features and feature individual predictability. The evaluator searches for features which have high correlation with the target and low intercorrelation between features.
- **Search Method**: a technique that examines the space of feature subsets by using different search techniques.

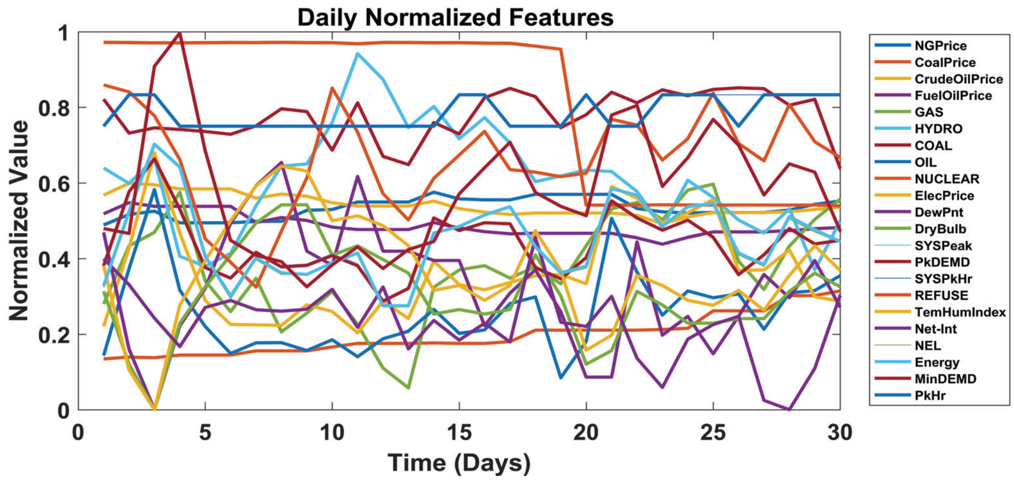


Figure 6.2: All Available Input Features

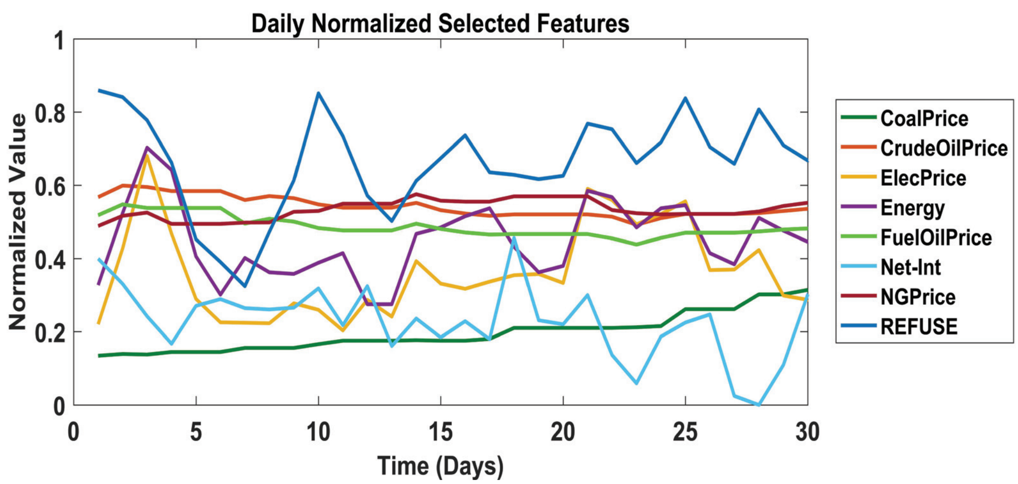


Figure 6.3: Selected Input Features

6.3.1 Features Pre-Processing Techniques

Some Attribute Evaluator methods are [54]:

1. **CfsSubsetEval**: Assesses the value of features combinations by recognizing each feature ability to predict the target, and data duplication between features.
2. **Principal Components**: This method is utilized in combination with a Ranker. It examines the components and performs a transformation. By considering the variance percentage in the input data, the dataset size can be reduced.
3. **ReliefFAttributeEval**: This method is continually read instance and considers the feature value of the nearest instance to assess the value of the feature.

4. **WrapperSubsetEval**: This method utilizes machine learning programming to assess the feature subsets. Cross-validation is used in conjunction to estimate the accuracy of a subset.

The most popular search methods are [54]:

1. **Best-First Search**: It utilizes greedy hill-climbing algorithm [8] to examine feature subsets. As all likely feature including or excluding at a point, Best-first may search forward, backward, or in both directions.
2. **Greedy-Step Wise Search**: It executes a greedy search of feature subsets. It may search forward or backward. Search stops when the including or excluding of any feature brings a decrease in accuracy estimation [64].
3. **Exhaustive Search**: It executes backtracking algorithms through the space of feature subsets. It is used to reduce the features space and resulting the best subset available.
4. **Genetic Search**: It utilizes genetic algorithm [65] to search potential informative features combinations. Genetic algorithm is able of search large spaces, it performs stochastic search method. GA is different from other local search methods, it implements global search to find the global optimum.
5. **Random Search**: It executes a random examination in the feature subsets. Random search results in the best subset with fewer features than the initial or random start points [62].
6. **Ranker**: It is used in combination with attribute evaluators. The Rank is recognized by each features subset evaluations.

From the above-mentioned feature selection techniques, different users may choose different methods based on their purpose and requirements.

6.4 Developed Methodology

This section studies the effect of feature selection on mid-term electricity price forecasting. Generally, there are three main steps involved in building a prediction model: data pre-processing, feature selection, and model selection [66].

The data pre-processing purpose involves the initial preparation of data, which includes tasks such as data normalizing, data cleaning, and data reduction. In the context of electricity price forecasting, the most important data pre-processing actions are normalization and feature selection [66], so they are used in the present work to enhance prediction accuracy.

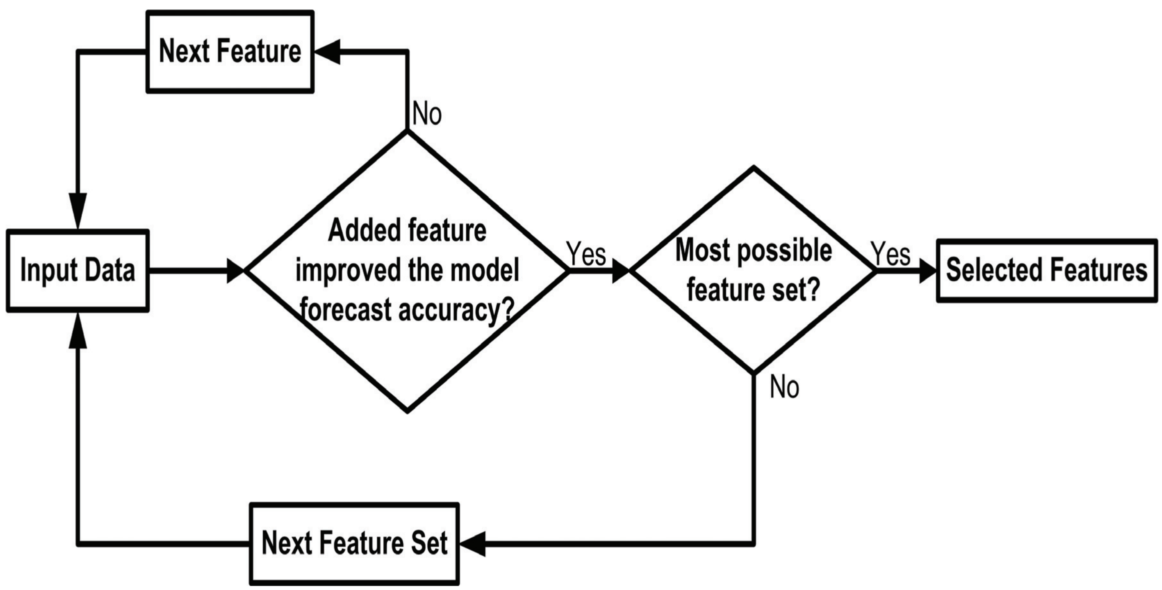


Figure 6.4: Input Data Selection Process

Feature selection concentrates on finding the most important and explanatory predictors and using them as inputs in the forecasting model. Given the limited available data for mid-term forecasting, developing appropriate input feature selection methodologies is not only complex but important. In general, the two main feature selection methodologies are filter and wrapper [66]. In the filter method, features are chosen based on their importance to the target variable, regardless of the forecasting model used. Wrapper methods, on the other hand, search for an optimal feature subset obtained from the full set, with the relevance of the features being evaluated by the accuracy of the final predictions.

Feature pre-processing is introduced and illustrated in Figure 6.4. The features pre-processing procedure includes the following steps:

1. **Input set generation**: Initial input data is generated; the feature set can be subset of the original feature set or the most influential features.
2. **Feature set evaluation**: After adding a new feature to the input feature set, it is evaluated and compared with the previous best one according to a certain evaluation criterion. If the new feature set turns out to be better, it replaces the previous best feature set. Different evaluation methods can be employed to select the best feature set, such as Attribute Evaluator and search method.
3. **Stopping criterion**: The process of feature input set generation and evaluation is repeated until a given mean absolute error percentage is satisfied.
4. **Results**: The selected optimum feature set is validated by testing the data sets.

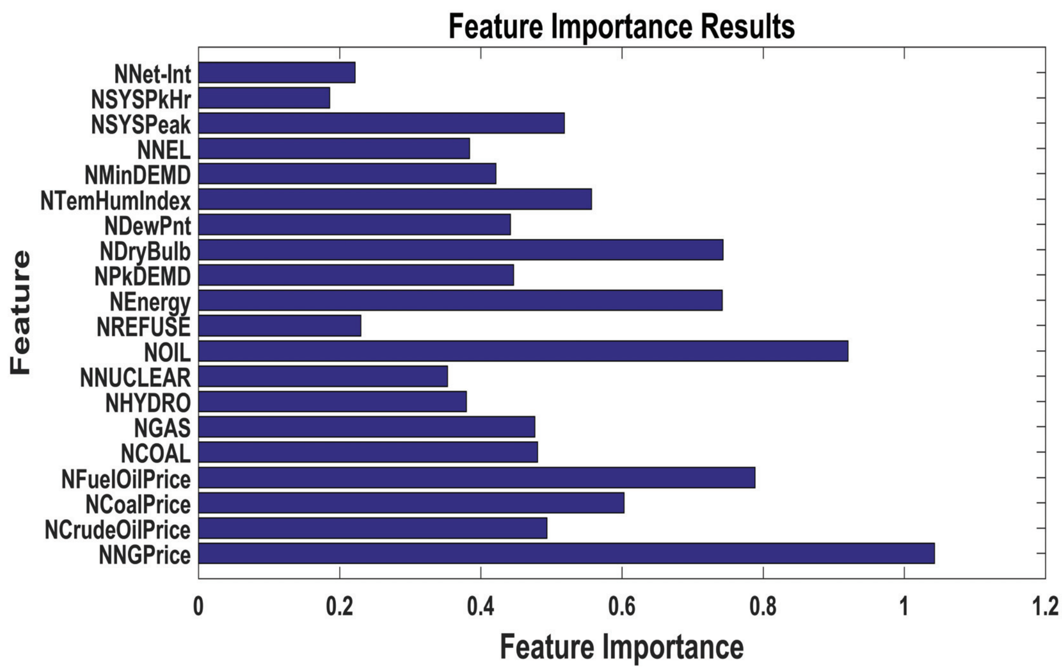


Figure 6.5: Features Importance

According to ISO-NE market characteristics, the most influential explanatory features of the electricity price are determined using the regression tree in Figure 6.5. It is obvious that the fuel prices are the most important features and that the natural gas fuel price is the most influential one among them, as the ISO-NE generates 45% of its power from natural gas.

6.5 Numerical Results and Discussion

The dataset was used to train, test and analysis the effectiveness of several feature selection techniques in electricity price forecasting. The study will be carried with real data from New England ISO. ISO-NE publishes daily historical data of New England regional demand, production, weather and electricity prices on its website [59]. The data used in this research was split into training period (Jan 01, 2008 to Jan 07, 2010) and testing period (Jan 08, 2010 to March 31, 2010). Different feature selection methods will first be applied on the ISO-NE dataset to generate feature sets. These feature sets will then be employed to train and test the SVM prediction model separately. The prediction accuracy of the model represents the quality of the feature set.

Table 6.1: Feature Selection Techniques

Attribute Evaluator	Search Method	Selected Features No.
CFS	**BF Forward**	**5**
CFS	**BF Backward**	**8**
CFS	**BF Bi-directional**	**5**
CFS	**Exhaustive**	**5**
CFS	**Genetic**	**7**
CFS	**GreedyStepwise**	**5**
CFS	**PSO**	**7**
CFS	**Random**	**5**
CFS	**Rank Search**	**5**
ReliefF	**Ranker**	**8**

The New England real daily data includes 20 features. We initially train the SVM model with all available features and then apply different feature selection techniques to generate feature sets, as shown in Table 6.1. The SVM model is trained separately with each selected feature set. Some other data set combinations are selected to train and test the SVM model, as follows:

- **Group 1**: All features.
- **Group 2**: Natural gas price only.
- **Group 3**: All fuel prices.
- **Group 4**: All fuel prices and supply data.
- **Group 5**: All fuel prices, supply and demand data.

- **Group 6**: All fuel prices and weather data.

Some search methods have the same selected feature set, so one of the same feature set will be applied to train and test the SVM model, as follows:

- BestFirst_F, BestFirst_Bi, Exhaustive, GreedyStepwise, Random, and Rank Search: Selected features $\{NGPrice, COAL, OIL, REFUSE, Net\_Int\}$.
- Genetic, PSO: Selected features $\{NGPrice, CrudeOilPrice, CoalPrice, FuelOilPrice, REFUSE, Energy, Net\_Int\}$.

To evaluate the results, some performance measures are used. Relative Mean Square Error (RMSE) for the predicted period is calculated by Eq.(4.15). For each model, the Mean Absolute Percentage Error (MAPE) is also calculated. It is used to compare the performance of different models Eq.(4.16). The correlation coefficient (r) measures the strength of the statistical relationship (degree of linear dependence) between variables. In other words, the coefficient shows how close two variables lie along a line. In this research, it is used to measure the correlation coefficient between the observed values (Y_i) and the fitted values ($\hat{Y}_i$). r is determined by Eq.(4.17).

Table 6.2: Forecast Accuracy of the Feature Sets

FS	**MAPE**	**RMSE**	**r**	**MAE**	**T.Time**
Group1	**6.42**	**5.47**	**0.97**	**3.66**	**61.51**
BF Forward	**8.08**	**6.79**	**0.95**	**4.64**	**3.13**
BF Backward	**7.49**	**6.64**	**0.96**	**4.28**	**5.94**
Genetic PSO	**7.37**	**6.31**	**0.96**	**4.25**	**5.22**
Ranker	**6.77**	**5.87**	**0.97**	**3.89**	**6.18**
Group2	**9.77**	**10.20**	**0.90**	**6.13**	**0.51**
Group3	**8.85**	**9.02**	**0.92**	**5.45**	**1.82**
Group4	**6.46**	**5.13**	**0.97**	**3.63**	**10.78**
Group5	**6.17**	**5.29**	**0.97**	**3.46**	**28.83**
Group6	**8.36**	**7.75**	**0.94**	**4.96**	**4.7**

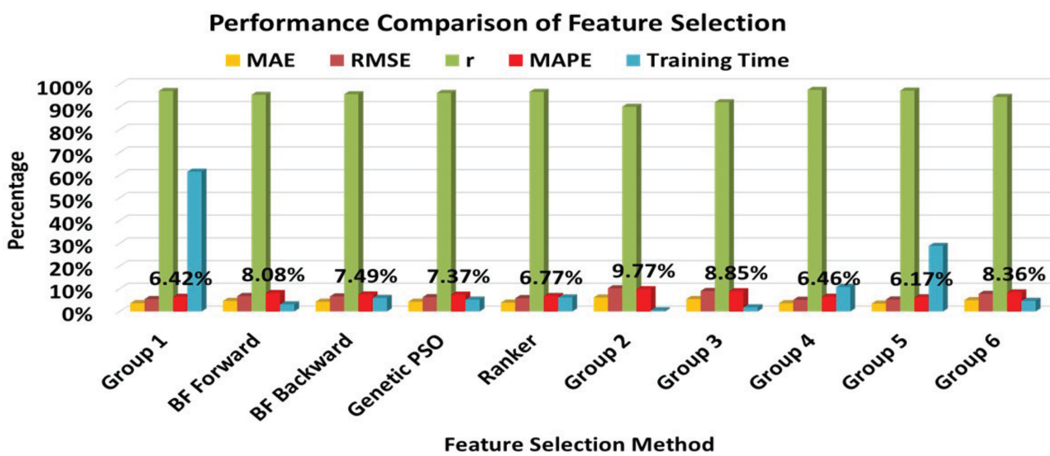


Figure 6.6: FS Performance Comparison Chart

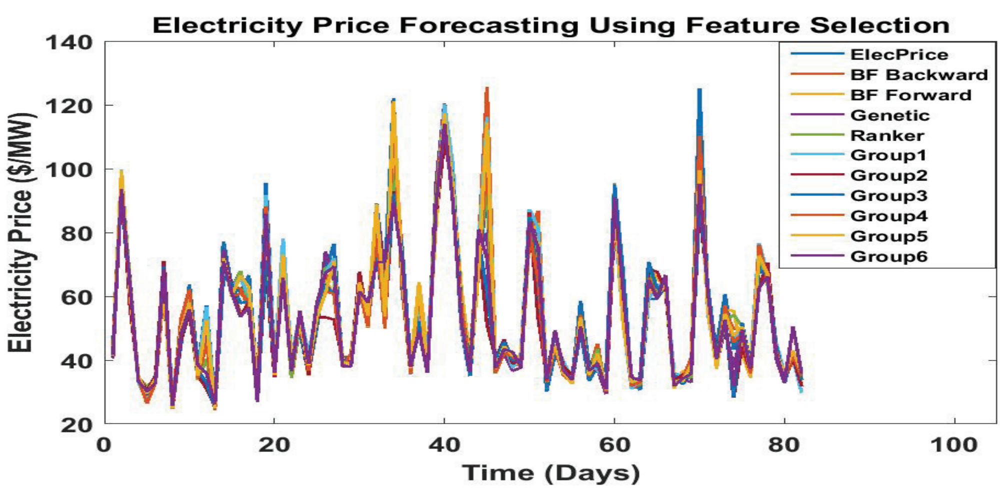


Figure 6.7: Electricity Price Forecasting Using Different FS

Table 6.2 shows a performance comparison between different feature selection methods used in electricity price forecasting using SVM modeling.

From the results, it can be observed that the Group 5 feature set has the lowest percentage error (MAPE) and the highest correlation coefficient (r). Hence, the Group5 feature set has the lowest prediction error for electricity price and is the best model for fitting the non-linearity of the data. In Figure 6.7, we can see that

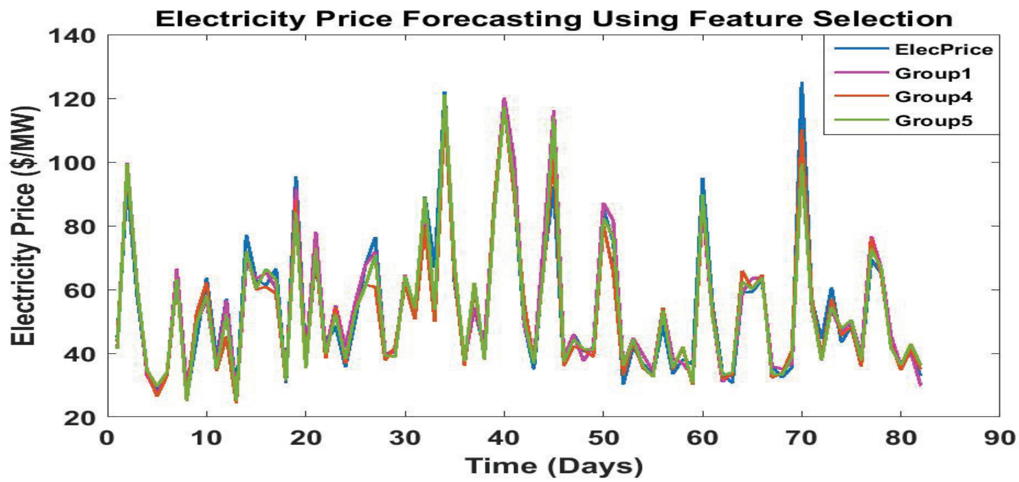


Figure 6.8: Best Performed Feature Sets

the feature set price forecasting has captured price variations trends as well as price magnitudes.

In Figure 6.8, the most accurate price forecasting model using feature selection is shown. It is clear that Group5 FS has the lowest MAPE and thus the most accurate prediction FS.

6.6 Summary

In this chapter, the problem of feature selection for electricity price forecasting was presented. In the proposed feature selection methodology, based on the market properties, the input features are selected from a large set of inputs using search methods. From all factors that may influence the electricity price, those which can better address the price variations and are predictable should be selected. Real daily data for New England ISO are used with Support Vector Machine modeling to predict the daily electricity price using different feature selection methods. For this purpose, the program WEKA is used for data pre-processing and for performing feature selection search methods and SVM modeling with a Radial Basis Function kernel.

The performance of the proposed feature selection technique is compared with some feature search methods. The obtained results indicate the superiority of the features selection in generating high-quality prediction. From the results, we can clearly see that the Group5 feature set, which consists of fuel prices, demand and supply data, has the best-predicted price.

Chapter 7

Mid-Term Electricity Price Forecasting Using SVM Approach

In the modern electricity market, it is important to have precise electricity price forecasting in a mid-term time horizon, but few studies have focused on the mid-term forecasting of electricity prices. The mid-term time horizon of electricity price forecasting has many applications, such as future maintenance scheduling of power plants, planning future contracts, purchasing raw materials, and determining market pricing. Factors that are important for forecasting electricity prices include choosing the most useful price features that influence the market price, and choosing the proper prediction model that is able to predict price variation behavior using historical data. The proposed SVM method, along with a few other methods, is evaluated using data from the New England ISO that are published on their official website.

7.1 Introduction

Price prediction plays an important role in the scheduling and administration of electricity markets. Recently, many studies were presented focusing on short-term electricity price forecasting [67, 68]. In contrast, there is very little research focusing on mid-term electricity price forecasting. Mid-term forecasting can range from a few weeks to one year. Researching mid-term price forecasts is necessary for mid-term planning in electricity markets that involves scheduling maintenance, dispatching power generation, and future contracting and investment [69, 70]. Mid-term electricity price forecasting is a complex task due to the long prediction period and the unsteady nature of electricity prices. In addition, limited explanatory data exists for use in such price forecasting. In mid-term forecasting historical data, only a few attributes are available. Moreover, many forecasting engines need large sets of data for training and testing, so some methods are not applicable for mid-term forecasting. This chapter focuses on using the data mining technique (SVM) to predict electricity prices. The mid-term electricity price forecasting task is addressed, and different issues related to the forecasting are considered. In addition, based on the NE-ISO publicly available data, a year-ahead price forecast is developed. The contributions of this research are developing mid-term electricity price forecasts for electricity markets using publicly available data and addressing and resolving issues associated with mid-term price forecasting problem, such as selecting proper input features.

7.2 Developed Electricity Price Forecasting Methodology

This section explores mid-term electricity price forecasting. It is assumed that the daily average fuel price, demand data, and weather data are available, and that the objective is to forecast the daily average electricity price. Generally, there are three main steps involved in building a prediction model: data pre-processing, feature selection, and model selection [24].

Data pre-processing involves the initial preparation of data, which includes tasks such as data normalizing, data cleaning, and data reduction. In the context of electricity price forecasting, the most frequently used data pre-processing actions are outliers (abnormal prices) detection and manipulation, normalization, and data transformation [24]. In this research, outliers operation of electricity prices is not performed, because abnormal values explain the real nature of the data, and manipulating them may result in losing informative features. In the present work, only data normalization and reduction are applied, as normalization has been found to enhance prediction accuracy [24].

Feature selection concentrates on finding the most important and explanatory predictors in order to use them as inputs in the forecasting model. Given the limited amount of available data for mid-term forecasting, developing appropriate input feature selection methodologies has become increasingly complex. Feature selection developed

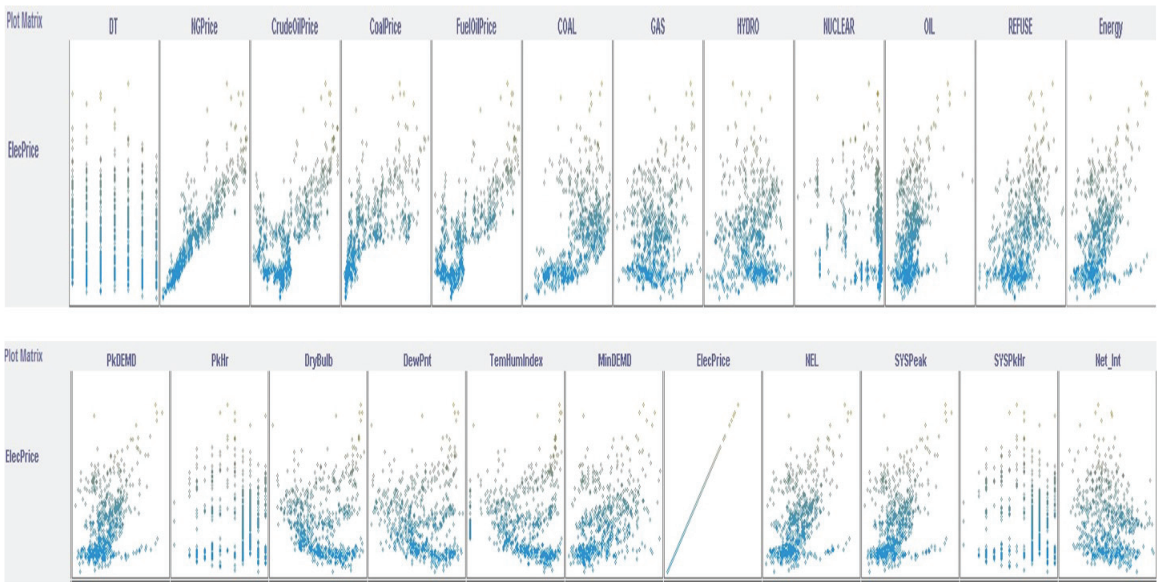


Figure 7.1: Features Scatter Plots

methodology and techniques were discussed in the previous chapter.

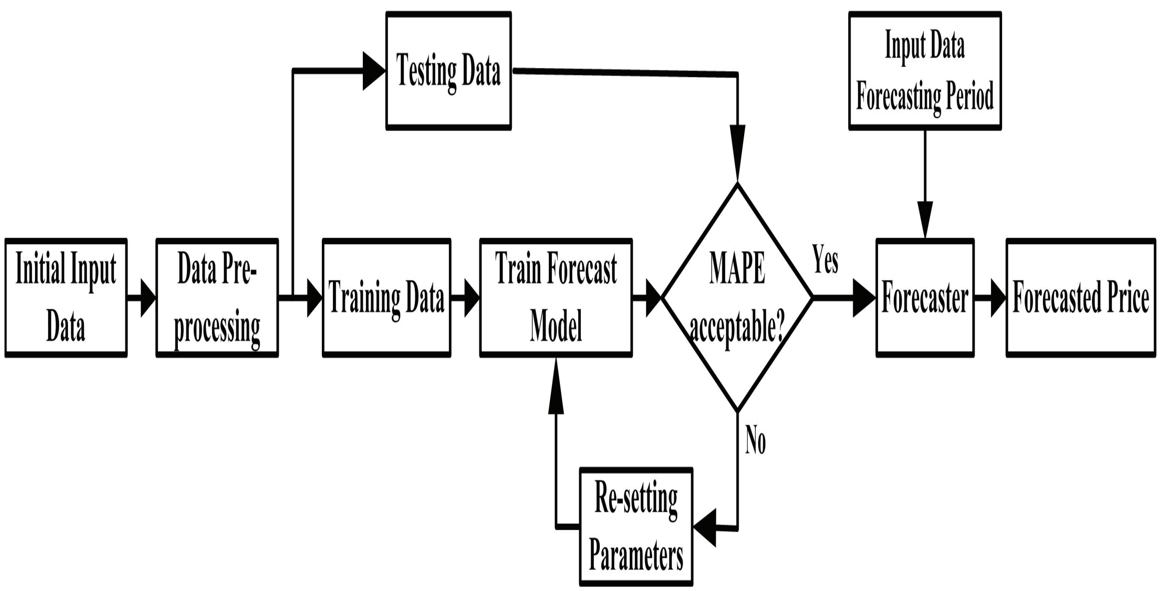


Figure 7.2: Electricity Price Forecaster

A design of the proposed forecasting method is shown in Figure 7.2. The forecaster is trained using training data, after which its performance is evaluated by the out-of-sample test set. The validation period has two applications: to identify optimum input features, and to find the optimum SVM parameters.

7.3 SVM Optimization

The SVM and mapping kernels have training parameters that can affect the performance of training and forecasting. These parameters can be manually adjusted or fine-tuned to improve their outcome. As shown and discussed in Chapter 4, SVM training improvement can be achieved by adjusting the slack variables penalty weight (C), and searching for the best trade-off between generalizing the model and outliers. Large C values will greatly penalize outliers and result in a strong hyperplane that avoids outliers and potentially loses model generalization. Furthermore, small C values will penalize outliers and result in soft-margin SVM behavior. Penalty weight (C) is separate from kernel choice and is a parameter of the minimization problem.
Other than for the linear kernel, other kernels might have tunable parameters. For instance, the RBF kernels gamma (γ) parameter is optimized jointly with (C) parameter. The RBF kernel is widely used for its flexibility in fitting data, along with some other popular kernels such as polynomials.
In this research, three QP optimization solvers discussed in section 4.5 are examined to train the SVM. Their results will be discussed in the next section.

7.4 Numerical Results and Discussion

The relationships between electricity price and fuel costs, load demand, weather conditions and economic/demographic factors are very important for prediction accuracy. These features should be considered when determining the data to be used for electricity price forecasting. In this study, the data set consists of calendar days, daily average (natural gas, crude oil, fuel oil) prices, daily peak load values, daily average (Dew Point, Dry Bulb, Humidity Index), and Import/Export power. Daily data from 2003 to 2009 is gathered from the New England ISO [54]. As fuel prices are the most influential factors in pricing electricity, they are included in the data set of this study. Daily normalized electricity and fuel prices are given in Figure 7.2.
The normalized daily load peak demand curve is shown in Figure 7.3. Due to the lack of information, economical and demographical factors are not included in this

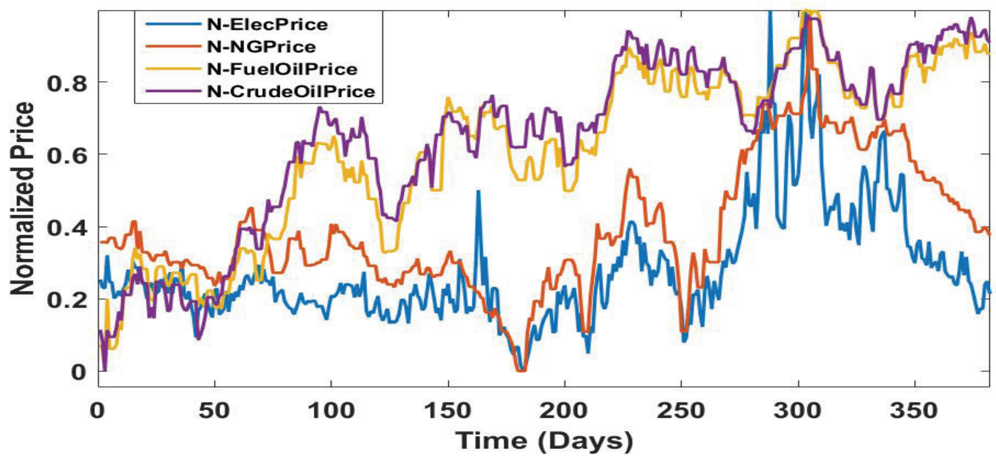


Figure 7.3: Normalized Average Daily Electricity and Fuels Prices

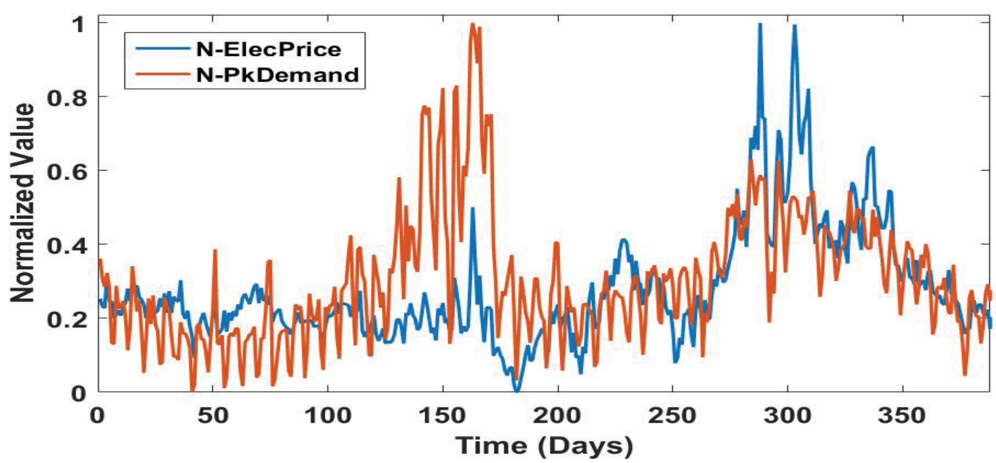


Figure 7.4: Normalized Average Daily Electricity Price and Peak Demand

study. The data set is constructed from fifteen inputs, as follows:

- One input that indicates the weekday.
- Three inputs that indicate fuel prices.
- Seven inputs that indicate electrical loads.
- Three inputs that indicate weather conditions.
- One input that indicates Import/Export power.

Normalizing the input data enhances the forecasting accuracy so that better results can be obtained. Normalization can eliminate the problem whereby larger values may terminate lower values. It is beneficial to normalize training data and testing data together to avoid mismatching the data parameters.

As discussed in section 4.4.4, the radial basis function kernel is used in SVM modeling. When pre-processing the data to build an SVM model, 90% of the data is used for training and 10% for testing the model. All of the input data are normalized, and in order to enhance the forecasting accuracy and achieve better results, the WEKA built-in Sequential Minimal Optimization (SMO) algorithm is used to train the SVM. Different SVM parameters optimization approaches are detailed in [71, 72, 73, 74]. Almost all types of neural networks were trained and tested on an out-of-sample dataset, but only the out-performed network types are included in this research. The best among them is included in the comparison to the SVM performance. Cascade-forward backprop 3 layers (CNN3L) performed the best with MAPE= 10.30%. Figure 7.8 illustrates the neural network models' performance.

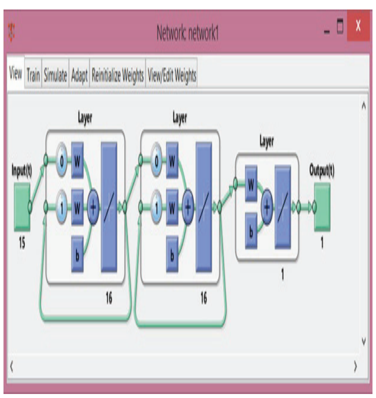


Figure 7.5: Layer Recurrent Neural Network

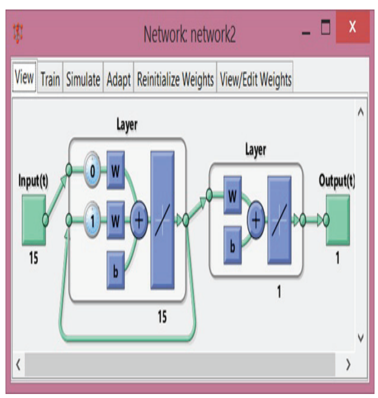


Figure 7.6: Elman Neural Network

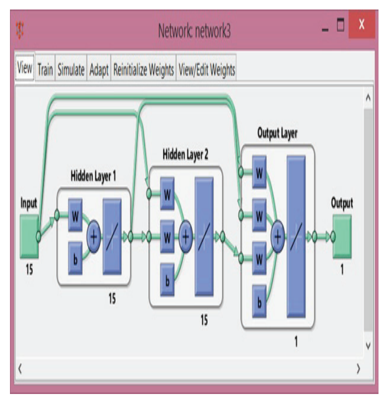


Figure 7.7: Cascade-forward Neural Network

Table 7.1: Comparison of Neural Networks Forecasting Performances

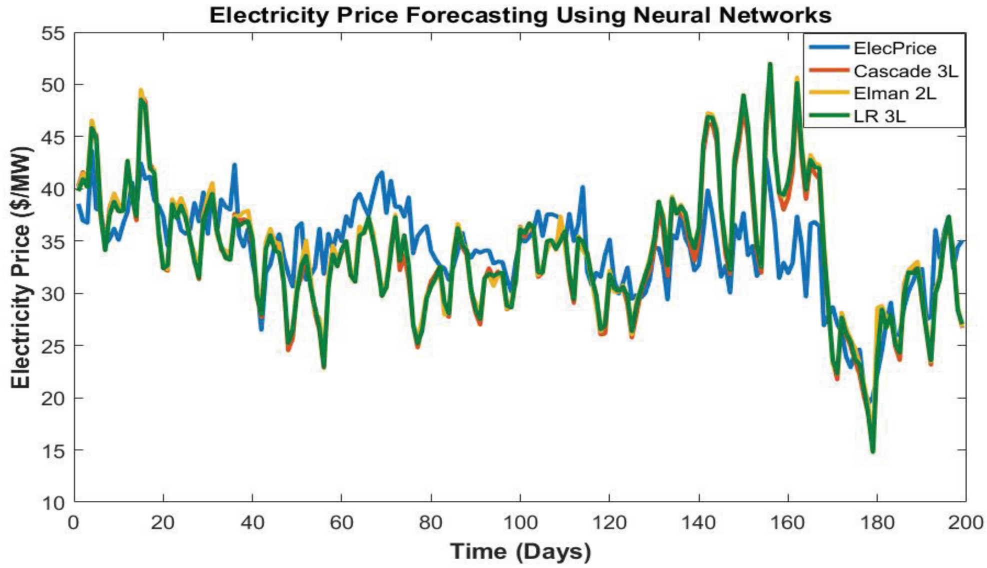


Figure 7.8: The Best Neural Network Models' Performance

Network Type	T.Time	r%	MSE	MAPE	MAE	RMSE
Layer Recurrent 3L	11	65.70	21.4295	10.3738	3.5786	4.6292
Elman 2L	7	64.6	21.8646	10.6529	3.6576	4.676
Cascade-forward 3L	2	66.8	20.5656	10.3046	3.544	4.5349

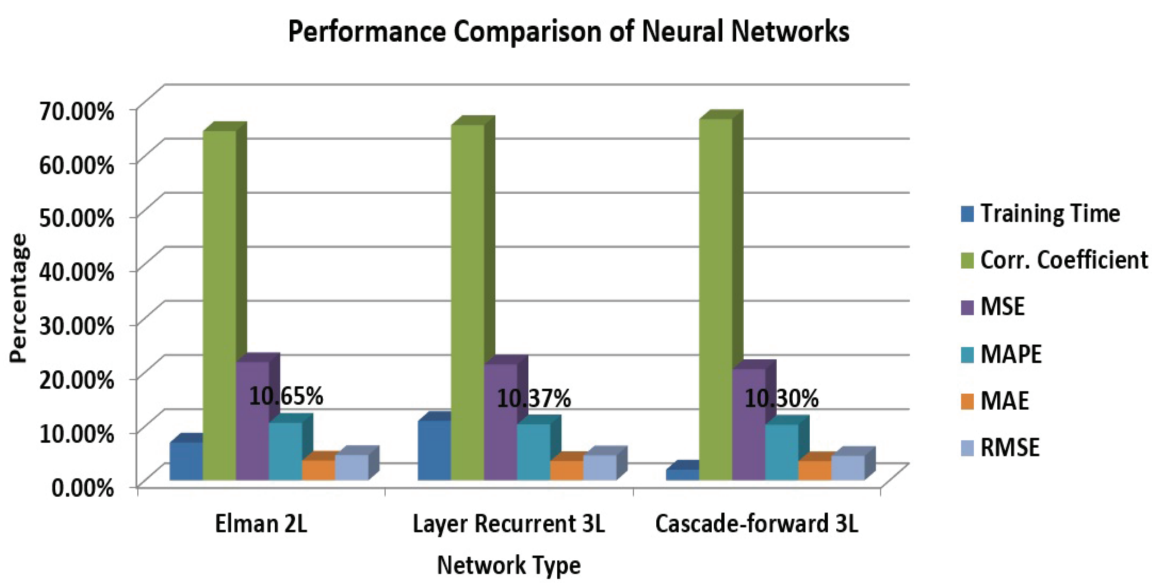


Figure 7.9: Neural Networks Performance Comparison Chart

Other models, such as Least Median Squared (LeastMedSq) and Radial Basis Function (RBF) are trained and tested with the same data to compare the performance of the proposed SVM Regression model.

To evaluate the results, some performance measures are used. Relative Mean Square Error (RMSE) for the predicted period is calculated by Eq.(4.15). For each model, the Mean Absolute Percentage Error (MAPE) is also calculated and is used to compare the performance of different models Eq.(4.16). MAPE, which is most commonly used to evaluate forecasts, has valuable statistical properties which are often useful for purpose of reporting. It is expressed in generic percentage terms that are understandable to users [57]. Finally, the correlation coefficient (r) is a coefficient that measures the strength of the statistical relationship (degree of linear dependence) between variables. In other words, the coefficient shows how close two variables lie along a line. In this research, it is used to measure the correlation coefficient between the observed values (Y_i) and the fitted values ($\hat{Y}_i$). r is determined by Eq.(4.17).

Table 7.2: Comparison of Model Forecasting Performance

Model	**T.Time**	**r%**	**MSE**	**MAPE**	**MAE**	**RMSE**
LeastMedSq	22.4	68	15.0163	8.7951	3.03	3.8751
Cascade-forward NN	2	66.8	20.5656	10.3046	3.544	4.5349
RBFRegressor	2.91	68.72	29.5276	11.5224	3.9088	5.4339
SVM	13.4	75	11.7441	7.4857	2.5683	3.427

Table 7.2 presents a comparison of the optimum forecasting results of different models. The SVM model has the lowest percentage error (MAPE), the lowest regression squared error (RMSE), and the highest correlation coefficient (r). These factors indicate that SVM has the lowest prediction error for electricity price and is the best model fitting the non-linearity of the data.

Figure 7.13 shows that SVM captures price variation trends as well as price magnitudes, but Figure 7.12 shows that SVM and the other models are not capable of

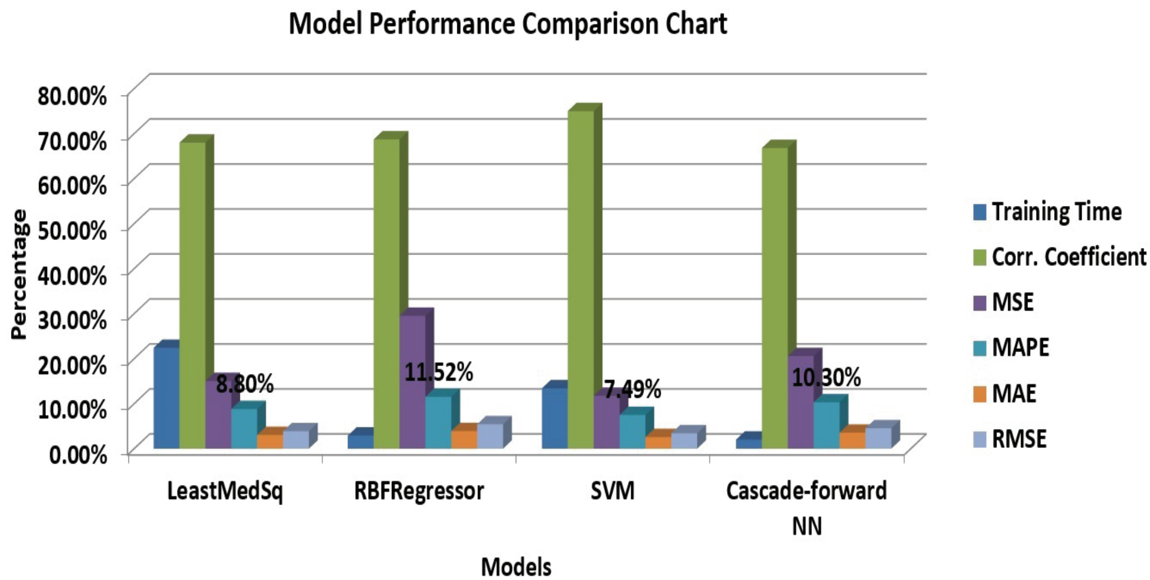


Figure 7.10: Model Performance Comparison Chart

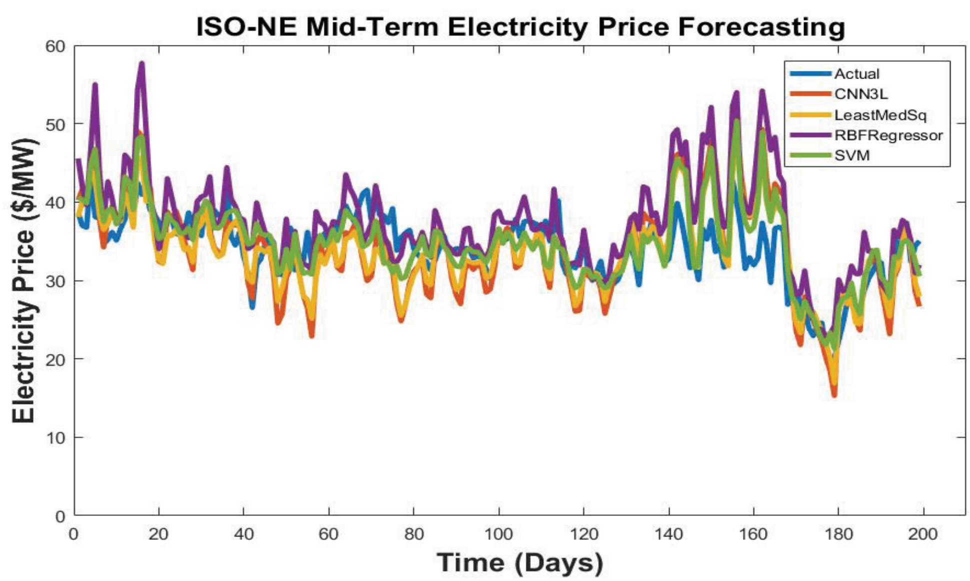


Figure 7.11: Different Model Performances Against Actual Price

following price spikes as sharply as they occur in the real world. However, SVM can estimate the variations in time, so it makes the most accurate predictions.

7.4.1 SVMs Parameter and Kernel Optimization

The existence of parameters in SVMs and kernels can affect training and regression results, but these parameters can be fine-tuned to improve SVM performance. The

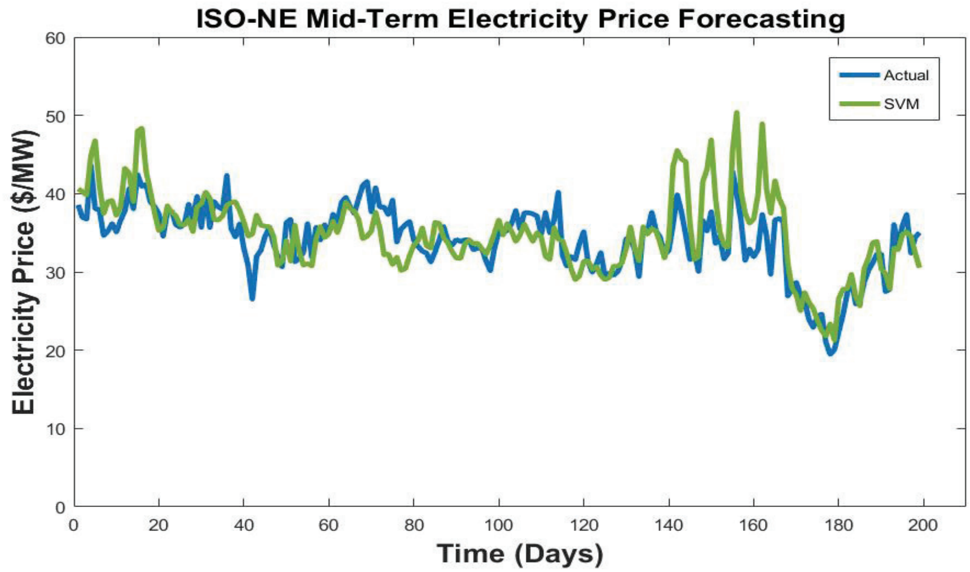


Figure 7.12: SVM Model Performance Against Actual Price

simplest approach in SVM regression improvement is to adjust the slack variables penalty weight (C) and watch for the best trade-off between allowing outlier errors and generalizing the model. Large values of C will penalize outlier instances and, as a result, the hyperplane will strongly avoid regression errors. Eventually, a $C = \infty$ generates a hard-margin SVM behavior, which sacrifices model generalization.

Optimization of SVM parameters using search techniques is beneficial for getting better regression results. Kernel generalization power can be understood by optimizing kernel hyper-parameters, as generalization may not show if default or random parameters were used. In this section, different SVM QP solving algorithms, several kernels and an approach to tuning the SVM performance are presented. Three QP optimization solvers mentioned in section 4.5 are used and compared to model the SVM by solving the QP problem for the two Lagrange multipliers. A linear kernel is used to map the data with all solvers in this comparison.

Table 7.3: Performance Comparison of SVM QP Optimization Solver

Network Type	T.Time	r%	MSE	MAPE	MAE	RMSE
SMO	3	70.7	13.5834	8.3369	2.8739	3.6856
ISDA	4	71.8	12.7863	8.0851	2.7925	3.5758
L1QP	660	71.2	13.3209	8.2369	2.8422	3.6498

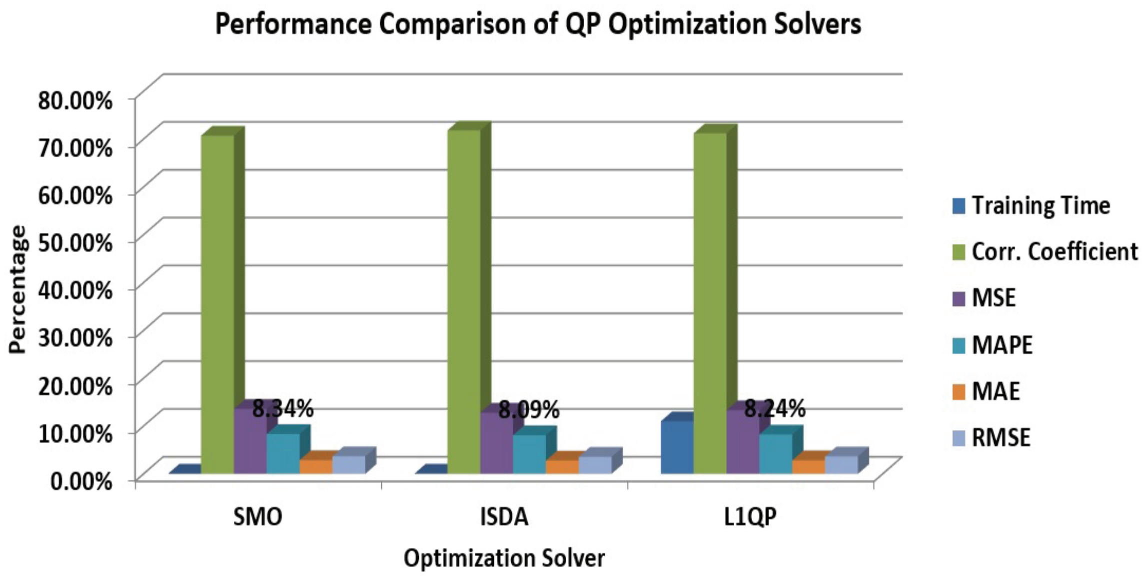


Figure 7.13: Performance Comparison of SVM QP Solvers

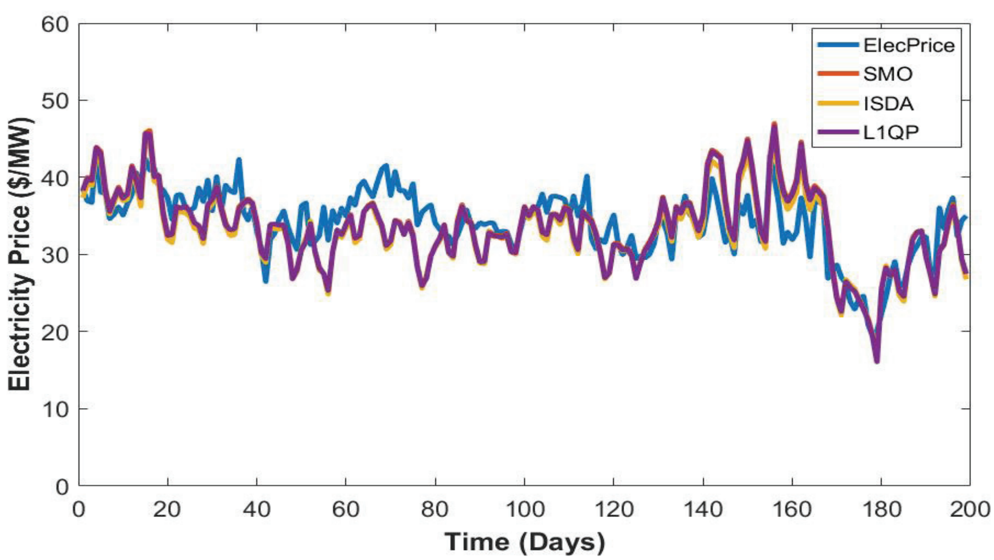


Figure 7.14: Electricity Price Forecasting Using Different SVM Optimization Solvers

Table 7.3 represents a forecasting performance comparison of three QP optimization solvers. From the table, it is clear that ISDA outperforms the other solvers for this research dataset, as it has the largest correlation coefficient (r) and the lowest mean absolute percentage error (MAPE). CVParameterSelection [75] is a meta-classifier in

```
=== Classifier model (full training set) ===

Cross-validated Parameter selection.
Classifier: weka.classifiers.functions.SMOreg
Cross-validation Parameter: '-C' ranged from 1.0 to 10.0
with 5.0 steps
Classifier Options: -C 10 -N 0 -I
"weka.classifiers.functions.supportVector.RegSMOImproved -T
0.001 -V -P 1.0E-12 -L 0.001 -W 1" -K
"weka.classifiers.functions.supportVector.RBFKernel -G
0.01 -C 250007"
```

Figure 7.15: Complexity Parameter Optimization Using Cross-validated Parameter selection

WEKA. It can optimize over a random number of parameters and performs parameter selection by cross-validation. Furthermore, it can be utilized to optimize the complexity parameter (C) of SVM and the gamma parameter (γ) of the RBF kernel. To analyze the kernels performance on an optimization basis, each kernel is tested

```
=== Classifier model (full training set) ===

Cross-validated Parameter selection.
Classifier: weka.classifiers.functions.LibSVM
Cross-validation Parameter: '-G' ranged from 0.01 to 0.1
with 10.0 steps
Classifier Options: -G 0.09000000000000002 -S 3 -K 2 -D 3 -R
0.0 -N 0.5 -M 40.0 -C 10.0 -E 0.001 -P 0.1 -Z -model "C:
\\Program Files\\Weka-3-7" -seed 1
```

Figure 7.16: RBF Gamma Parameter Optimization Using Cross-validated Parameter selection

and its parameters fine-tuned independently. In [76], the authors pointed out that there is no single best kernel which has optimal performance in all problems.

Table 7.4: Performances Comparison of SVM Kernels

Kernel Type	T.Time	r%	MSE	MAPE	MAE	RMSE
Linear	10.55	71.23	12.382	8.0942	2.7747	3.5188
Polynomial	11.98	71.24	14.6266	8.6867	2.9978	3.8245
RBF	13.4	75	11.7441	7.4857	2.5683	3.427

When assessing the results kernel-wise in Table 7.4, the RBF kernel stands out with the larger correlation coefficient and the least percentage error. In building an

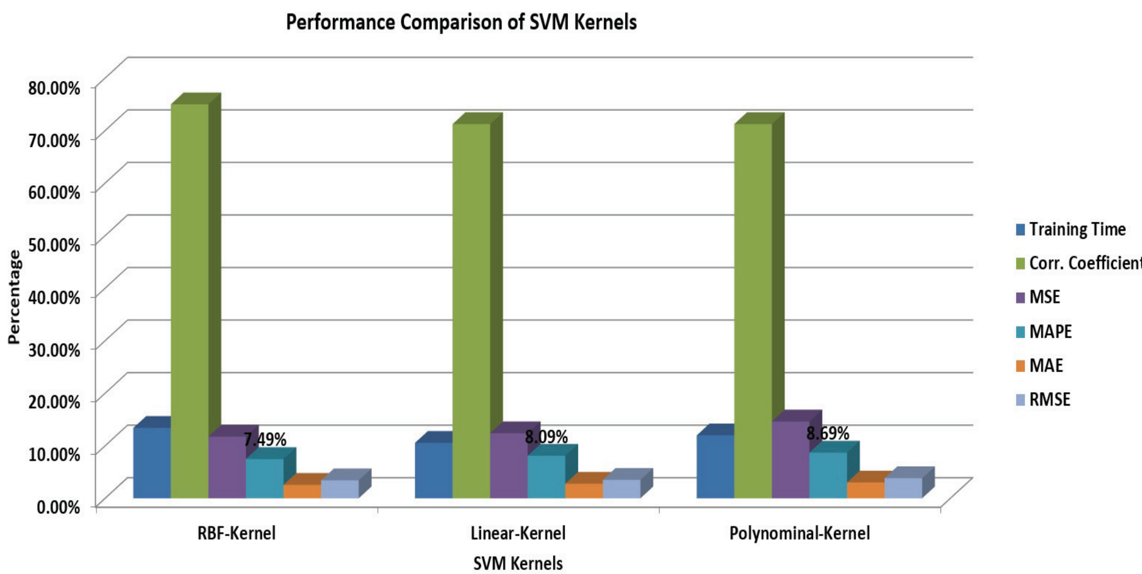


Figure 7.17: Forecasting Performance Comparison of SVM Kernels

SVM for regression, some learning parameters can be employed. The most important parameters are the ε-insensitive loss function and the penalty parameter C, which defines the adjustments between margin and slack variables. The parameter C has wide-ranging values and the best performance can be assessed using a cross-validation technique to find the optimal C value. Both parameters are selected by the user or optimized using an optimization technique prior to training the SVM.

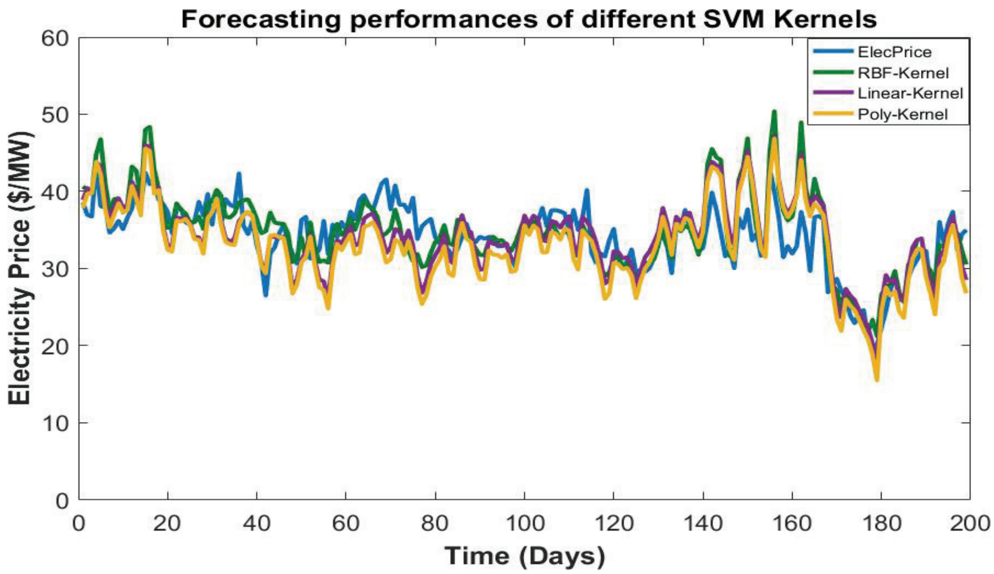


Figure 7.18: Electricity Price Forecasting Using Different SVM Kernels

7.5 Summary

In this chapter, the problem of mid-term electricity price forecasting was presented. In the proposed forecasting methodology, based on the market properties, the input features are selected from a large set of inputs using the search method. From all factors that may influence the electricity price, those which can better address the price variations and are predictable should be selected. The daily electricity price for New England ISO is predicted with Support Vector Machines using real fuel prices, daily demand values, and weather conditions. For this purpose, the program WEKA is used for SVM modeling. As well, the Radial Basis Function kernel is used for mapping nonlinear input values to a higher dimension feature space. The SVM meta-parameters are determined from the training data using a built-in SMO algorithm. This approach saves the trial and error time of selecting these parameters and enhances prediction accuracy. The performance of the proposed SVM technique is compared with some benchmark techniques and the obtained results indicate the superiority of the SVM in generating high-quality prediction intervals. The results show that SVM has the best-predicted price, thus indicating the models strong capability in predicting price variation trends (i.e., sign and magnitude of price variations). This capability shows to what extent a model can identify the relation between input data and the value to be predicted. From the SVM parameters and kernel optimization results, we can conclude that selecting the correct kernel, adjusting kernel parameters, adjusting the model complexity penalty, and selecting the optimal subset of features can lead to significant improvements in forecasting performance. The inclusion of input features that can explain the economic conditions of the market will increase the model's accuracy to a great extent.

Chapter 8

Conclusion

8.1 Conclusion

In this research, different techniques were applied to perform MTEPF. Computational methods, which better understand price variations than traditional methods, were used for forecasting in deregulated markets. Given the lack of comprehensive literature available on SVM to do regression and perform MTEPF, machine learning techniques like SVM had not previously been thoroughly considered in electricity price forecasting. Therefore, this research work applied SVM Regression and optimized its parameters to improve the prediction accuracy of future electricity prices. This thesis evaluated the importance of feature selection in forecasting and also evaluated the performance of SVM Regression in electricity price forecasting. Real daily data from New England ISO were used to experiment and explore the applicability of the developed feature selection and SVM Regression forecasting. Various feature selection techniques were explored and experimented on the dataset, and a developed feature selection method was discussed. To compare the performance of the SVM Regression model, various forecasting methods were applied in this research, including Neural Network Architectures, Median Least Squared, and Radial Basis Function. A statistical analysis of electricity price forecasting was done separately for each forecasting method. After finding the results for all methods, a comparison analysis was made. The performance of each method was evaluated on the basis of mean absolute percentage error ($MAPE$), root mean squared error ($RMSE$), correlation coefficient (r) and other evaluating measures. The results show that with the same dataset, the SVM Regression forecasting model achieved the highest forecasting accuracy among all applied forecasting models.

Various SVM optimization solvers were utilized in this research to train and optimize the SVM parameters to fasten the quadratic programming solution and increase the model forecasting accuracy. Various kernels were used with the SVM Regression model and their performances were compared.

8.2 Summary of Contributions

The work presented in this thesis contributes to improving the quality of feature selection through building a developed method for input dataset for electricity price forecasting. Three primary concerns of dataset are data size, selection of the most important variables, and improving the quality of the forecasting. A developed feature selection method is a good approach for addressing these concerns. Further techniques are applied to perform feature selection. This thesis also presented a step towards forecasting mid-term electricity prices using SVM Regression. The relative success of the SVM Regression model is due to its flexibility and efficiency. In this thesis, we have shown how feature selection and the SVM Regression model can be combined to achieve the highest accuracy in electricity price forecasting. These techniques can easily be generalized to other aspects of forecasting models.

8.2.1 List of Publications

The work of this thesis is based on the following publications:

1. A. Mohamed and M. E. El-Hawary. "Mid-Term Electricity Price Forecasting Using SVM," In IEEE Proc. The 29^{th} Annual IEEE Canadian Conference on Electrical and Computer Engineering, 15-18 MAY, 2016.
2. A. Mohamed and M. E. El-Hawary. "Effective Input Features Selection for Electricity Price Forecasting," In IEEE Proc. The 29^{th} Annual IEEE Canadian Conference on Electrical and Computer Engineering, 15-18 MAY, 2016.
3. A. Mohamed and M. E. El-Hawary. "On Optimization of SVMs Kernels and Parameters for Electricity Price Forecasting,"Manuscript submitted for publication in IEEE Electrical Power and Energy Conference 2016.

8.3 Scope of Future Work

SVM Regression performance relies on the tuning of its parameters, the kernels used, and feature selection of the training data. Although this thesis presented a detailed performance evaluation of feature selection techniques and proposed SVM Regression to predict mid-term electricity prices, future work can focus on improving forecasting accuracy by:

- Studying the prediction accuracy of electricity peak prices (spikes) and how to improve the predicting of these abnormal behaviors.
- Examining other kernels and optimization solvers for SVM regression training and evaluating their out-of-sample performance.

Bibliography

[1] Catalo, JPSa, et al. "Short-term electricity prices forecasting in a competitive market: a neural network approach." Electric Power Systems Research 77.10 (2007): 1297-1304.

[2] Shafie-Khah, M., M. Parsa Moghaddam, and M. K. Sheikh-El-Eslami. "Price forecasting of day-ahead electricity markets using a hybrid forecast method." Energy Conversion and Management 52.5 (2011): 2165-2169.

[3] A. Mohamed and M. E. El-Hawary. "Mid-Term Electricity Price Forecasting Using SVM," In IEEE Proc. The 29^{th} Annual IEEE Canadian Conference on Electrical and Computer Engineering, 15-18 MAY, 2016.

[4] A. Mohamed and M. E. El-Hawary. "Effective Input Features Selection for Electricity Price Forecasting," In IEEE Proc. The 29^{th} Annual IEEE Canadian Conference on Electrical and Computer Engineering, 15-18 MAY, 2016.

[5] A. Mohamed and M. E. El-Hawary. "On Optimization of SVMs Kernels and Parameters for Electricity Price Forecasting,"Manuscript submitted for publication in IEEE Electrical Power and Energy Conference 2016.

[6] Shafie-Khah, M., M. Parsa Moghaddam, and M. K. Sheikh-El-Eslami. "Price forecasting of day-ahead electricity markets using a hybrid forecast method." Energy Conversion and Management 52.5 (2011): 2165-2169.

[7] Lei Wu; Shahidehpour, M., "A Hybrid Model for Day-Ahead Price Forecasting," in Power Systems, IEEE Transactions on , vol.25, no.3, pp.1519-1530, Aug. 2010 doi: 10.1109/TPWRS.2009.2039948

[8] Suganthi, L., and Anand A. Samuel. "Energy models for demand forecasting A review." Renewable and sustainable energy reviews 16.2 (2012): 1223-1240.

[9] Escribano, Alvaro, J. Ignacio Pea, and Pablo Villaplana. "Modelling Electricity Prices: International Evidence*." Oxford bulletin of economics and statistics 73.5 (2011): 622-650.

[10] Karegowda, Asha Gowda, A. S. Manjunath, and M. A. Jayaram. "Comparative study of attribute selection using gain ratio and correlation based feature selection." International Journal of Information Technology and Knowledge Management 2.2 (2010): 271-277.

[11] Nguyen, Minh Hoai, and Fernando De la Torre. "Optimal feature selection for support vector machines." Pattern recognition 43.3 (2010): 584-591.

[12] Cai, Deng, Chiyuan Zhang, and Xiaofei He. "Unsupervised feature selection for multi-cluster data." Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2010.

[13] Song, Qinbao, Jingjie Ni, and Guangtao Wang. "A fast clustering-based feature subset selection algorithm for high-dimensional data." Knowledge and Data Engineering, IEEE Transactions on 25.1 (2013): 1-14.

[14] Bermejo, Pablo, et al. "Fast wrapper feature subset selection in high-dimensional datasets by means of filter re-ranking." Knowledge-Based Systems 25.1 (2012): 35-44.

[15] "http://bang2good.blogspot.ca/2015/04/introduction-to-power-system.html" Introduction to power system, Mawsoo3a.

[16] LinkedIn. (2014, July 17). White Paper - A Smart Electric Market Means a Smart Market Infrastructure,[Article]. Retrieved from https://www.linkedin.com/pulse/20140717202051-75805467-white-paper-a-smart-electric-market-means-a-smart-market-infrastructure.

[17] Saini, Lalit M., Sanjeev K. Aggarwal, and Ajit Kumar. "Parameter optimisation using genetic algorithm for support vector machine-based price-forecasting model in National electricity market." Generation, Transmission & Distribution, IET 4.1 (2010): 36-49.

[18] Hogarth, Robin M., and Spyros Makridakis. "Forecasting and planning: An evaluation." Management Science 27.2 (1981): 115-138.

[19] Nakano, Mikihisa. "Collaborative forecasting and planning in supply chains: The impact on performance in Japanese manufacturers." International Journal of Physical Distribution & Logistics Management 39.2 (2009): 84-105.

[20] Lei Wu; Shahidehpour, M., "A Hybrid Model for Day-Ahead Price Forecasting," in Power Systems, IEEE Transactions on , vol.25, no.3, pp.1519-1530, Aug. 2010 doi: 10.1109/TPWRS.2009.2039948

[21] Escribano, Alvaro, J. Ignacio Pea, and Pablo Villaplana. "Modelling Electricity Prices: International Evidence*." Oxford bulletin of economics and statistics 73.5 (2011): 622-650.

[22] Werner, Dan. "Electricity Market Price Volatility: The Importance of Ramping Costs." 2014 Annual Meeting, July 27-29, 2014, Minneapolis, Minnesota. No. 169619. Agricultural and Applied Economics Association, 2014.

[23] Bompard, E., et al. "Dynamic price forecast in a competitive electricity market." Generation, Transmission & Distribution, IET 1.5 (2007): 776-783.

[24] Box, George EP, et al. Time series analysis: forecasting and control. John Wiley & Sons, 2015.

[25] Martinez Alvarez, Francisco, et al. "Energy time series forecasting based on pattern sequence similarity." Knowledge and Data Engineering, IEEE Transactions on 23.8 (2011): 1230-1243.

[26] Lee, Yi-Shian, and Lee-Ing Tong. "Forecasting time series using a methodology based on autoregressive integrated moving average and genetic programming." Knowledge-Based Systems 24.1 (2011): 66-72.

[27] Kian, A.; Keyhani, A., "Stochastic price modeling of electricity in deregulated energy markets," in System Sciences, 2001. Proceedings of the 34th Annual Hawaii International Conference on , vol., no., pp.7 pp.-, 6-6 Jan. 2001

[28] Pindoriya, N. M., S. N. Singh, and S. K. Singh. "An adaptive wavelet neural network-based energy price forecasting in electricity markets." Power Systems, IEEE Transactions on 23.3 (2008): 1423-1432.

[29] Mbamalu, G. A. N., and M. E. El-Hawary. "Load forecasting via suboptimal seasonal autoregressive models and iteratively reweighted least squares estimation." Power Systems, IEEE Transactions on 8.1 (1993): 343-348.

[30] Tsekouras, George J., Nikos D. Hatziargyriou, and Evangelos N. Dialynas. "An optimized adaptive neural network for annual midterm energy forecasting." Power Systems, IEEE Transactions on 21.1 (2006): 385-391.

[31] Singhal, Deepak, and K. S. Swarup. "Electricity price forecasting using artificial neural networks." International Journal of Electrical Power & Energy Systems 33.3 (2011): 550-555.

[32] Mandal, P.; Senjyu, T.; Urasaki, N.; Funabashi, T.; Srivastava, A.K., "A Novel Approach to Forecast Electricity Price for PJM Using Neural Network and Similar Days Method," in Power Systems, IEEE Transactions on , vol.22, no.4, pp.2058-2065, Nov. 2007

[33] Mandal, P.; Senjyu, T.; Funabashi, T., "Neural network models to predict short-term electricity prices and loads," in Industrial Technology, 2005. ICIT 2005. IEEE International Conference on , vol., no., pp.164-169, 14-17 Dec. 2005

[34] Aggarwal, Sanjeev Kumar, Lalit Mohan Saini, and Ashwani Kumar. "Electricity price forecasting in deregulated markets: A review and evaluation." International Journal of Electrical Power & Energy Systems 31.1 (2009): 13-22.

[35] Sansom, Damien C., Tom Downs, and Tapan K. Saha. "Evaluation of support vector machine based forecasting tool in electricity price forecasting for Australian national electricity market participants." Journal of Electrical and Electronics Engineering, Australia 22.3 (2003): 227-234.

[36] LI, Yuan-cheng, Ting-jian FANG, and Er-keng YU. "Study of support vector machines for short-term load forecasting [J]." Proceedings of the Csee 6 (2003): 10.

[37] Xinxing Pan and B. Lee, "A comparison of support vector machines and artificial neural networks for mid-term load forecasting," Industrial Technology (ICIT), 2012 IEEE International Conference on, Athens, 2012, pp. 95-101.

[38] S. K. Shevade, S. S. Keerthi, C. Bhattacharyya and K. R. K. Murthy, "Improvements to the SMO algorithm for SVM regression," in IEEE Transactions on Neural Networks, vol. 11, no. 5, pp. 1188-1193, Sep 2000.

[39] Kecman, Vojislav, T-M. Huang, and Michael Vogt. "Iterative single data algorithm for training kernel machines from huge data sets: Theory and performance." Support vector machines: theory and applications. Springer Berlin Heidelberg, 2005. 255-274.

[40] Wu, Qiang, and Ding-Xuan Zhou. "SVM soft margin classifiers: linear programming versus quadratic programming." Neural computation 17.5 (2005): 1160-1187.

[41] Saini, Lalit M., Sanjeev K. Aggarwal, and Ajit Kumar. "Parameter optimisation using genetic algorithm for support vector machine-based price-forecasting model in National electricity market." Generation, Transmission & Distribution, IET 4.1 (2010): 36-49.

[42] Suykens, Johan AK, and Joos Vandewalle. "Least squares support vector machine classifiers." Neural processing letters 9.3 (1999): 293-300.

[43] Zhang, Qian, et al. "A fuzzy group forecasting model based on least squares support vector machine (LS-SVM) for short-term wind power." Energies 5.9 (2012): 3329-3346.

[44] Shafie-Khah, M., M. Parsa Moghaddam, and M. K. Sheikh-El-Eslami. "Price forecasting of day-ahead electricity markets using a hybrid forecast method." Energy Conversion and Management 52.5 (2011): 2165-2169.

[45] Frank, Eibe and Mayo, Michael and Kramer, Stefan. "Alternating Model Trees." Proceedings of the 30th Annual ACM Symposium on Applied Computing SAC '15 (2015).

[46] Amjady, Nima. "Day-ahead price forecasting of electricity markets by a new fuzzy neural network." Power Systems, IEEE Transactions on 21.2 (2006): 887-896.

[47] E. Frank. "Fully supervised training of Gaussian radial basis function networks in WEKA." Technical Report 04/14, Department of Computer Science, University of Waikato, (2014).

[48] Smola, Alex J., and Bernhard Schlkopf. "A tutorial on support vector regression." Statistics and computing 14.3 (2004): 199-222.

[49] Niu, Dongxiao, Yongli Wang, and Desheng Dash Wu. "Power load forecasting using support vector machine and ant colony optimization." Expert Systems with Applications 37.3 (2010): 2531-2539.

[50] Sansom, Damien C., Tom Downs, and Tapan K. Saha. "Evaluation of support vector machine based forecasting tool in electricity price forecasting for Australian national electricity market participants." Journal of Electrical and Electronics Engineering, Australia 22.3 (2003): 227-234.

[51] Sapankevych, Nicholas I., and Ravi Sankar. "Time series prediction using support vector machines: a survey." Computational Intelligence Magazine, IEEE 4.2 (2009): 24-38.

[52] "http://kernelsvm.tripod.com/" Support Vector Machine Regression.

[53] Jain, Amit, and B. Satish. "Clustering based short term load forecasting using support vector machines." PowerTech, 2009 IEEE Bucharest. IEEE, 2009.

[54] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.

[55] Shevade, Shirish Krishnaj, et al. "Improvements to the SMO algorithm for SVM regression." Neural Networks, IEEE Transactions on 11.5 (2000): 1188-1193.

[56] Veropoulos, Konstantinos. "Machine learning approaches to medical decision making." Diss. University of Bristol, 2001.

[57] Swanson, David A., Jeff Tayman, and T. M. Bryan. "MAPE-R: a rescaled measure of accuracy for cross-sectional subnational population forecasts." Journal of Population Research 28.2-3 (2011): 225-243.

[58] Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. "Time series analysis: forecasting and control. 2008." (1970).

[59] "http://www.iso-ne.com/markets-operations/iso-express" New England Independent System Operator.

[60] The MathWorks, Inc. MATLAB. Version R2015a. URL: $http : //www.mathworks.com/support/?s_tid = gn_supp$

[61] IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

[62] Liu, Huan, and Lei Yu. "Toward integrating feature selection algorithms for classification and clustering." Knowledge and Data Engineering, IEEE Transactions on 17.4 (2005): 491-502.

[63] Crone, Sven F., and Nikolaos Kourentzes. "Feature selection for time series prediction A combined filter and wrapper approach for neural networks." Neurocomputing 73.10 (2010): 1923-1936.

[64] Witten, Ian H., and Eibe Frank. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2005.

[65] Goldberg, David E., Bradley Korb, and Kalyanmoy Deb. "Messy genetic algorithms: Motivation, analysis, and first results." Complex systems 3.5 (1989): 493-530.

[66] Kantardzic, Mehmed. Data mining: concepts, models, methods, and algorithms. John Wiley & Sons, 2011.

[67] Catalo, JPSa, et al. "Short-term electricity prices forecasting in a competitive market: a neural network approach." Electric Power Systems Research 77.10 (2007): 1297-1304.

[68] Shafie-Khah, M., M. Parsa Moghaddam, and M. K. Sheikh-El-Eslami. "Price forecasting of day-ahead electricity markets using a hybrid forecast method." Energy Conversion and Management 52.5 (2011): 2165-2169.

[69] Tanlapco, Eva, Jacques Lawarre, and Chen-Ching Liu. "Hedging with futures contracts in a deregulated electricity industry." Power Systems, IEEE Transactions on 17.3 (2002): 577-582.

[70] Conejo, Antonio J., et al. "Optimal involvement in futures markets of a power producer." Power Systems, IEEE Transactions on 23.2 (2008): 703-711.

[71] Deng, Naiyang, Yingjie Tian, and Chunhua Zhang. "Support vector machines: optimization based theory, algorithms, and extensions." CRC Press, 2012.

[72] Chang, Chih-Chung, and Chih-Jen Lin. "LIBSVM: A library for support vector machines." ACM Transactions on Intelligent Systems and Technology (TIST) 2.3 (2011): 27.

[73] Aydin, Ilhan, Mehmet Karakose, and Erhan Akin. "A multi- objective artificial immune algorithm for parameter optimization in support vector machine." Applied Soft Computing 11.1 (2011): 120-129.

[74] Zhang, XiaoLi, XueFeng Chen, and ZhengJia He. "An ACO- based algorithm for parameter optimization of support vector machines." Expert Systems with Applications 37.9 (2010): 6618-6628.

[75] Kohavi, Ron. Wrappers for performance enhancement and oblivious decision graphs. Diss. stanford university, 1995.

[76] A. Ali and A. Abraham. "An empirical comparison of kernel selection for support vector machines." In Proceedings of the second international conference on hybrid intelligent systems: design, management and applications. The Netherlands: IOS Press, pages 32130, 2002.

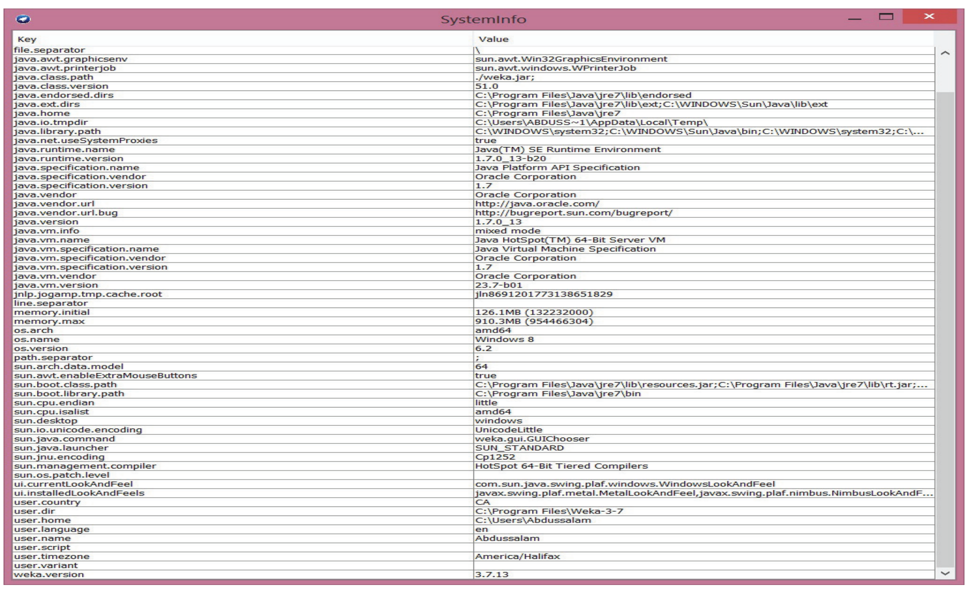
Appendices

Appendix A

WEKA Data Mining Software

Waikato Environment for Knowledge Analysis (WEKA) is an open source software. It was written in Java, it operates on different operating systems. It is developed by Waikato University in New Zealand. It has many algorithms to perform data preprocessing, feature selection, classification and regression.

WEKA 3.7.13 developer version is used in this research.


SystemInfo

Key	Value
file.separator	\
java.awt.graphicsenv	sun.awt.Win32GraphicsEnvironment
java.awt.printerjob	sun.awt.windows.WPrinterJob
java.class.path	./weka.jar;
java.class.version	51.0
java.endorsed.dirs	C:\Program Files\Java\jre7\lib\endorsed
java.ext.dirs	C:\Program Files\Java\jre7\lib\ext;C:\WINDOWS\Sun\Java\lib\ext
java.home	C:\Program Files\Java\jre7
java.io.tmpdir	C:\Users\ABDUSS~1\AppData\Local\Temp\
java.library.path	C:\WINDOWS\system32;C:\WINDOWS\Sun\Java\bin;C:\WINDOWS\system32;C:\...
java.net.useSystemProxies	true
java.runtime.name	Java(TM) SE Runtime Environment
java.runtime.version	1.7.0_13-b20
java.specification.name	Java Platform API Specification
java.specification.vendor	Oracle Corporation
java.specification.version	1.7
java.vendor	Oracle Corporation
java.vendor.url	http://java.oracle.com/
java.vendor.url.bug	http://bugreport.sun.com/bugreport/
java.version	1.7.0_13
java.vm.info	mixed mode
java.vm.name	Java HotSpot(TM) 64-Bit Server VM
java.vm.specification.name	Java Virtual Machine Specification
java.vm.specification.vendor	Oracle Corporation
java.vm.specification.version	1.7
java.vm.vendor	Oracle Corporation
java.vm.version	23.7-b01
jnlp.jogamp.tmp.cache.root	jln869120177313865l829
line.separator	
memory.initial	126.1MB (132232000)
memory.max	910.3MB (954466304)
os.arch	amd64
os.name	Windows 8
os.version	6.2
path.separator	;
sun.arch.data.model	64
sun.awt.enableExtraMouseButtons	true
sun.boot.class.path	C:\Program Files\Java\jre7\lib\resources.jar;C:\Program Files\Java\jre7\lib\rt.jar;...
sun.boot.library.path	C:\Program Files\Java\jre7\bin
sun.cpu.endian	little
sun.cpu.isalist	amd64
sun.desktop	windows
sun.io.unicode.encoding	UnicodeLittle
sun.java.command	weka.gui.GUIChooser
sun.java.launcher	SUN_STANDARD
sun.jnu.encoding	Cp1252
sun.management.compiler	HotSpot 64-Bit Tiered Compilers
sun.os.patch.level	
ui.currentLookAndFeel	com.sun.java.swing.plaf.windows.WindowsLookAndFeel
ui.installedLookAndFeels	javax.swing.plaf.metal.MetalLookAndFeel,javax.swing.plaf.nimbus.NimbusLookAndF...
user.country	CA
user.dir	C:\Program Files\Weka-3-7
user.home	C:\Users\Abdussalam
user.language	en
user.name	Abdussalam
user.script	
user.timezone	America/Halifax
user.variant	
weka.version	3.7.13

Figure A.1: WEKA System Information

WEKA graphical user interfaces (GUI) chooser in Figure A.2 contains three applications: Explorer, Experimenter and Knowledge Flow. Explorer is the simplest way to explore WEKA.

Many sources are used in WEKA to import and store data. WEKA uses ARFF file format, and it allows other format such as URLs, and CSV.

Figure A.2: WEKA Graphical User Interfaces

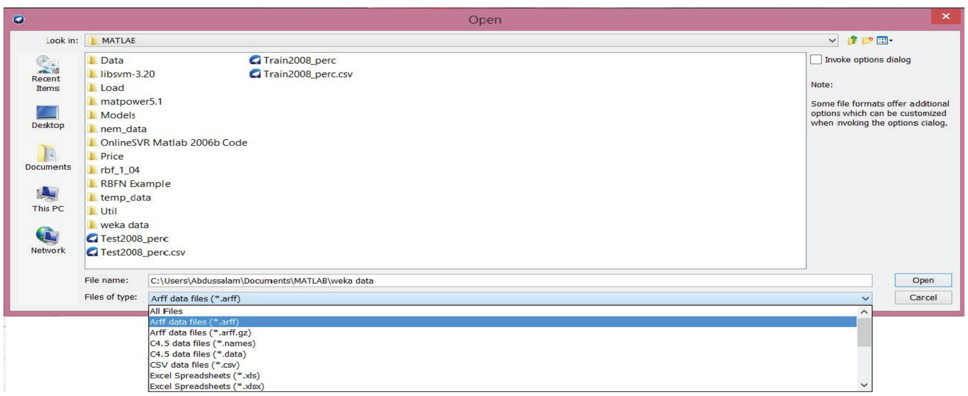


Figure A.3: WEKA Data Sources

WEKA Explore window in Figure A.4 contains panels for different data mining tasks such as preprocess where filters are used to transform data by resample, normalizeetc.

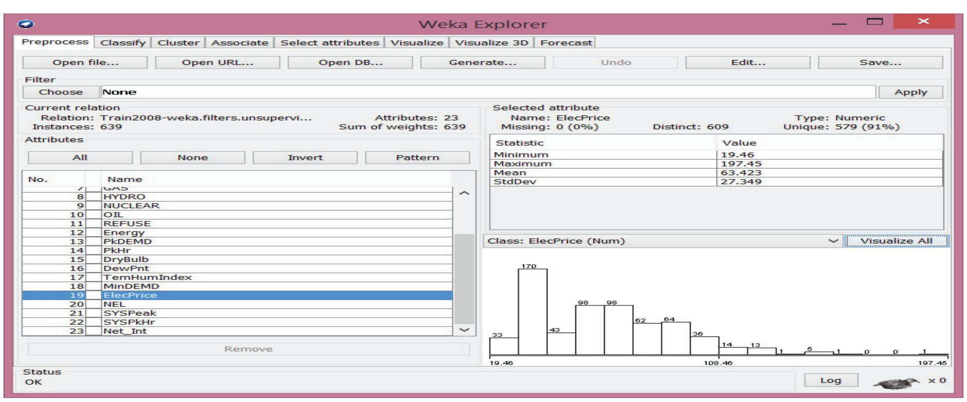


Figure A.4: WEKA Explore GUI

Classify panel in Figure A.5 consists of regression and classification techniques.

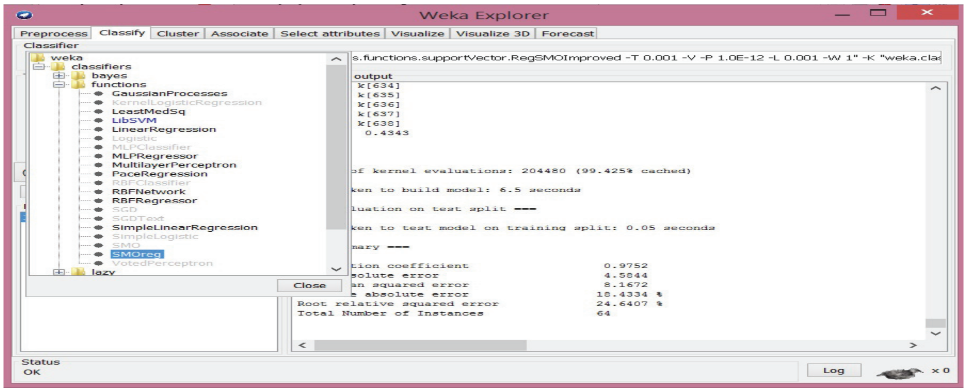


Figure A.5: WEKA Regression Techniques

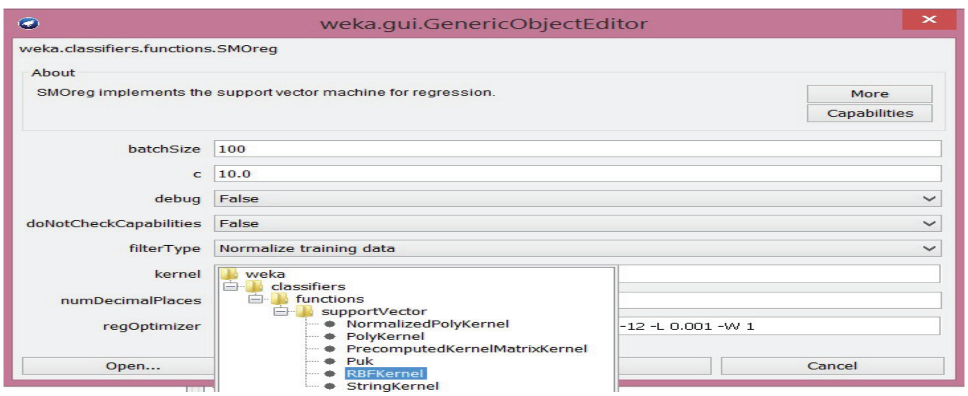


Figure A.6: WEKA SVM Kernels

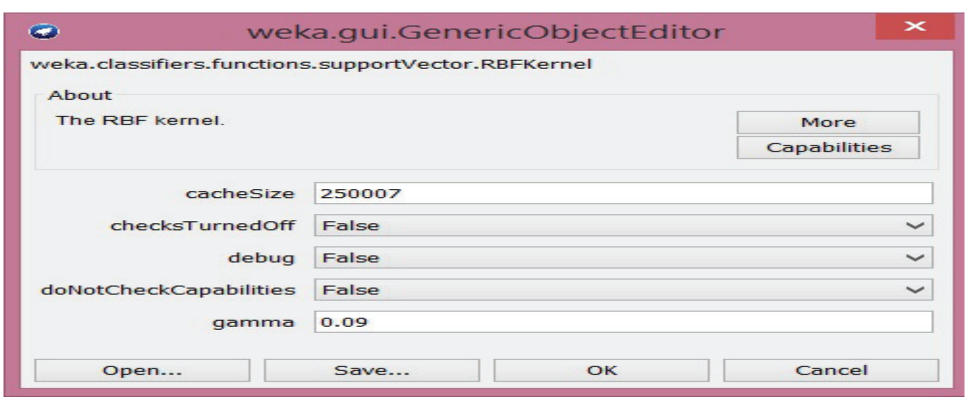


Figure A.7: RBF Kernel Parameters

Select attribute panel in Figure A.8 provides access to many search methods to perform feature selection by measuring the value of features and select the most important features in a dataset.

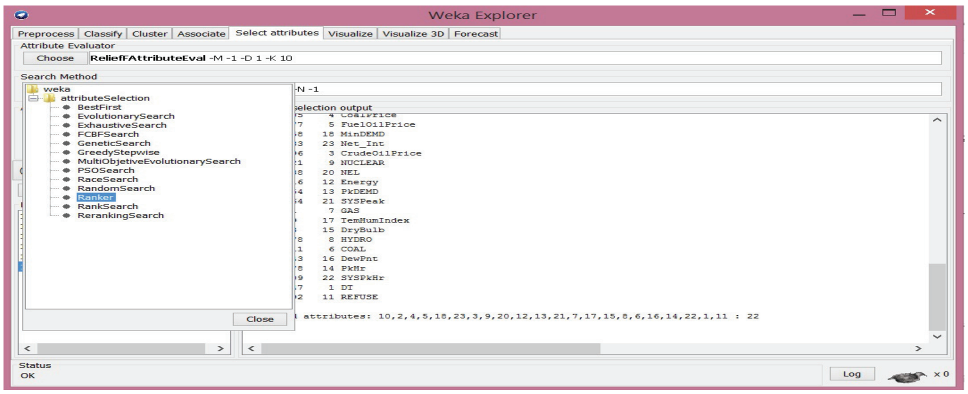


Figure A.8: WEKA Attributes Selection Methods

Visualization panel is used to generate Scatter plots, ROC curves, Trees, and graphs.

Figure A.9: WEKA Visualization Panel