การคำนวณสินค้าคงเหลือขั้นต่ำสำหรับอุตสาหกรรมเครื่องประดับด้วยการพยากรณ์การบริโภค
โดยใช้โครงข่ายประสาทเทียม

นางนฤมล ยอดเพชร์

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์และสารสนเทศ ภาควิชาคณิตศาสตร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2553

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

SAFETY STOCK CALCULATION FOR JEWELRY INDUSTRIES BASED ON CONSUMPTION FORECAST BY ARTIFICIAL NEURAL NETWORK

Mrs. Naroumon Yordphet

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Computer Science and Information
Department of Mathematics
Faculty of Science
Chulalongkorn University
Academic Year 2010
Copyright of Chulalongkorn University

Thesis Title	SAFETY STOCK CALCULATION FOR JEWELRY INDUSTRIES BASED ON CONSUMPTION FORECAST BY ARTIFICIAL NEURAL NETWORK
By	Mrs. Naroumon Yordphet
Field of Study	Computer Science and Information
Thesis Advisor	Siripun Sanguansintukul, Ph.D.

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree

.... Dean of the Faculty of Science

(Professor Supot Hannongbua, Dr.rer.nat.)

THESIS COMMITTEE

.... Chairman

(Professor Chidchanok Lursinsap, Ph.D.)

.... Thesis Advisor

(Siripun Sanguansintukul, Ph.D.)

.... External Examiner

(Worasit Choochaiwattana, Ph.D.)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

นฤมล ยอดเพชร์ : การคำนวณสินค้าคงเหลือขั้นต่ำสำหรับอุตสาหกรรมเครื่องประดับด้วยการพยากรณ์การบริโภคโดยใช้โครงข่ายประสาทเทียม. (SAFETY STOCK CALCULATION FOR JEWELRY INDUSTRIES BASED ON CONSUMPTION FORECAST BY ARTIFICIAL NEURAL NETWORK) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: อ.ดร.สิริพันธุ์ สงวนสินธุกุล, 69 หน้า.

วิทยานิพนธ์นำเสนอ การคำนวณสินค้าคงเหลือขั้นต่ำบนพื้นฐานของการบริโภคของชิ้นส่วนในอุตสาหกรรมเครื่องประดับซึ่งถูกตรวจสอบโดยใช้ความสามารถในการพยากรณ์ของโครงข่ายประสาทเทียม โดยปกติธุรกิจเครื่องประดับเกี่ยวข้องกับสมัยนิยม ดังนั้นการเปลี่ยนแปลงอย่างรวดเร็วทำให้การพยากรณ์มีความซับซ้อนมากขึ้น อุปสงค์มีการเปลี่ยนแปลงขึ้นลงตามความต้องการของลูกค้า และการแข่งขันที่รุนแรงในตลาดการค้า ปัจจัยเช่น การส่งมอบสินค้าที่ล่าช้าและคุณภาพของส่วนประกอบของสินค้าที่ไม่ได้มาตรฐานอาจเป็นสาเหตุให้ส่วนประกอบของสินค้าขาดแคลน เพื่อป้องกันการขาดแคลนของส่วนประกอบของสินค้า สนับสนุนการจัดการอุปทาน และการสร้างความพึงพอใจให้กับลูกค้า โครงข่ายประสาทเทียมจึงถูกนำมาใช้ประโยชน์ในการพยาการณ์การบริโภค สินค้าคงเหลือขั้นต่ำเป็นวิธีทางหนึ่งที่ช่วยในการป้องการขาดแคลน สินค้าคงเหลือขั้นต่ำถูกสร้างขึ้นเมื่อบริษัทมีคำสั่งซื้อส่วนประกอบเพื่อการผลิต หรือ มีคำสั่งซื้อมากกว่าความต้องการ ในทางตรงกันข้าม การมีสินค้าคงเหลือขั้นต่ำเป็นสิ่งสำคัญต่อการลดค่าใช้จ่าย และเป็นสิ่งที่อยู่ในความสนใจของบริษัท ในทางปฏิบัติ อุตสาหกรรมได้ทำการคำนวณสินค้าคงเหลือขั้นต่ำด้วยประสบการณ์ของตนเอง โดยขาดองค์ความรู้ในการสนับสนุนการคำนวณสินค้าคงเหลือขั้นต่ำ และนำไปสู่การคงเหลือของส่วนประกอบของสินค้าในสินค้าคงคลัง ดังนั้น สินค้าคงเหลือขั้นต่ำเป็นสิ่งที่สำคัญในการป้องกันการผันผวนที่ไม่อาจคาดเดาได้ของอุปสงค์และอุปทาน วิทยาพนธ์นี้เน้นการเพิ่มความถูกต้องของการพยากรณ์ ในอุตสาหกรรมเครื่องประดับ โครงข่ายประสาทเทียมถูกนำมาใช้เป็นเครื่องมือในการพยากรณ์ โดยมีสมมุติฐานที่ว่าความถูกต้องของการบริโภคนำไปสู่การคำนวณสินค้าคงเหลือขั้นต่ำที่แม่นยำ

ภาควิชา คณิตศาสตร์ ลายมือชื่อนิสิต Nam/ Yordpet.

สาขาวิชาวิทยาการคอมพิวเตอร์และสารสนเทศ ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก Siripun Sanguansintukul

ปีการศึกษา 2553

5173618423: MAJOR COMPUTER SCIENCE AND INFORMATION
KEYWORDS :SAFETY STOCK / FORECAST / ARTIFICIAL NEURAL NETWORK

NAROUMON YORDPHET: SAFETY STOCK CALCULATION FOR JEWELRY INDUSTRIES BASED ON CONSUMPTION FORECAST BY ARTIFICIAL NEURAL NETWORK. ADVISOR: SIRIPUN SANGUANSINTUKUL, Ph.D., 69 pp.

This thesis presents the safety stock calculation based on consumption of components in the jewelry business was investigated using the forecasting capability of an Artificial Neural Network (ANN). Generally, this business also has links with fashion; therefore, rapid change in the fashion industry makes the forecasting situation more complicated. The demand fluctuates with customer requirements and high competition in the marketplace. Factors such as late delivery and bad component quality can cause shortages of components. To prevent shortages, provide support to supply management and enhance customer satisfaction, an ANN is utilized for consumption forecast. Safety stock is the way to protect against shortages. Safety stock is created when a company either orders before an order is needed or orders more than the expected demand. In the contrast, keeping less safety stock is important for decreasing cost and increasing interests for a firm. In practice, business has calculated safety stock based on their experiences. non theoretical support, and leads for retaining component in the warehouse. Hence, safety stock is essential to protect against the fluctuation of unexpected demand and supply. The central focus of this paper is to enhance the forecast on safety stock accuracy in the jewelry business. The artificial neural network is employed as a tool for the prediction. It is assumed that the accurate consumption leads to the accuracy in safety stock calculation.

Department : Mathematics Student's Signature Naroumon Yordphet
Field of Study : Computer Science and Information Advisor's Signature Siripun Sanguansintukul
Academic Year : 2010

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor Siripun Sanguansintukul, Ph.D. for all her great encouragement, suggestion and assistance. Without her support and attention, this thesis would have never been completed.

I am also grateful to the thesis committee, Professor Chidchanok Lursinsap, Ph.D. and Worasit Choochaiwattana, Ph.D. for their constructive criticisms and invaluable advises.

Finally, I also express the deeply thanks to my family and relative for their encouragements since the beginning of my study and practically, when I feel break down.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

CONTENTS

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

LIST OF TABLES

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Motivation and Problem Description

A product cannot be assembled unless the required materials are available. Material shortage has several root causes including late supplier delivery and unexpected customer demand. Safety stock is one of the approaches used to protect against shortages. Safety stock is created when a company either orders before an order is needed or orders more than the expected demand. In contrast, keeping less safety stock is important for decreasing cost and increasing interests for a firm.

Safety stock is excess inventory that a company holds to guards against uncertainty in demand, lead time and supply, and can be used to improve customer service and reduce stock outages [1].

Keeping more safety stock helps to serve customer satisfaction but increases the holding cost. Due to the jewelry business related to fashion, new products are launched every season. Therefore, the products have a short time life cycle. It is very important to have products on hand to support the market and guarantee customer satisfaction. In practice, business has calculated safety stock based on their experiences, non theoretical support, and leads for retaining component in the warehouse. Hence, safety stock is essential to protect against the fluctuation of unexpected demand and supply. The demand side is affected by customer demand such as consumable behavior. Supply side is affected by supplier demand such as supplier stock shortages.

There are several approaches to reduce safety stock such as better collaboration with suppliers, using better technology, and enhancing the forecast accuracy. The central focus of this paper is to enhance the forecast on safety stock accuracy in the jewelry business. The ANN is employed as a tool for the prediction. It is assumed that the accurate consumption leads to the accuracy in safety stock calculation [2].

The central focus of this thesis is to enhance the forecast on safety stock accuracy in the jewelry business. The artificial neural network is employed as a tool for the prediction. It is assumed that the accurate consumption leads to the accuracy in safety stock calculation.

1.2 The Objectives of the Research

1) To enhance the accuracy of safety stock calculation based on consumption using the artificial neural network in a Jewelry business environment.

2) To assist Jewelry business to get the close prediction in the safety stock.

1.3 Literature review

In the recent survey, there are related works corresponding to safety stock forecast. Some of them are based on artificial intelligence and expert systems as follows:

L. Zhang et al. [1] provided a model on forecasting safety stock of enterprise resource planning (ERP) based on BP neural network. Their experiment consists of 8 input attributes:

1) Actual safety stock.

2) Inventory turnover.

3) Customer satisfaction.

4) Number of customers.

5) Lead time.

6) Product quality and reliability.

7) Delivery reliability.

8) Supply reliability.

Along with 1 output attribute as shown in Figure 1.1.

There are 15 enterprises as the sample in study. They choose the first 10 enterprises as training samples and the other 5 enterprises as inspection samples. They used the learning algorithm of BP neural network to train the training samples to train the network and get a well trained BP network. Then they input the inspection samples to inspect the BP neural network model, and compare the outputs of the model with the actual safety stock, which are the results of the experts' estimation. The forecasting relative errors are controlled within 20% and the result output shown all the relative errors are below 15%. It proves that the results for the forecasting are precise. They can adjust the weights and thresholds by training the network time after time to improve precision and reduce the relative error. By this experiment, they can get simulated outputs which are quite accurate and close to the desired outputs.

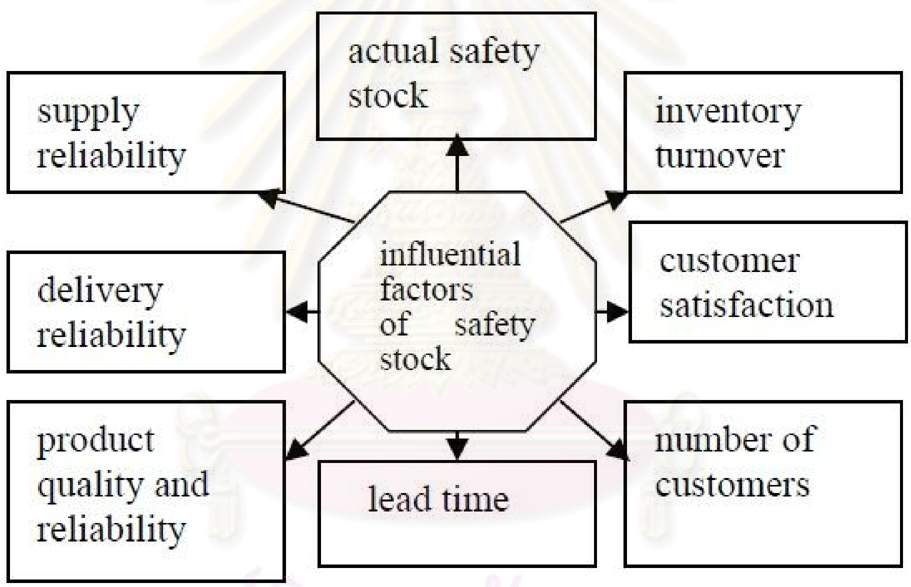


Figure 1.1: Influential factors of safety stock.

David C. Heath et al. [3] offered modeling of the evolution of demand forecast with application to safety stock analysis in production/distribution systems with seasonal stochastic demand. They combined the Martingale model of Forecast Evolution (MMFE) with a linear programming model of production and distribution planning implemented in a rolling horizon fashion.

The performance attributes of the production/distribution system are cost and customer service. Major cost categories are production costs (proportional to standard hours of production), transportation cost (proportional to volume-miles), and inventory holding costs (including financing cost and a premium for overflow inventory). Customer service is measured by a weighted average of monthly fill rates across product lines and regions. The preliminary simulations shown that annual cost could be reduced by several million dollars annually if the safety stock factor were reduced and that, provided the new Statistical Method of forecasting was implemented and used in a timely manner to plan production, there would be little adverse impact on customer service by reducing safety stock. Theoretical research is focusing on the form of the optimal inventory policy under non-trivial demand models and on state-space reduction techniques to implement the MMFE in computational dynamic programming.

Jukka Korpela et al. [4] proposed to adjust the safety stock level with the Analysis Hierarchy Process (AHP) which a theory of measurement for dealing with tangible and intangible criteria that has been applied to numerous areas. The AHP method is used for adjusting this basis level by taking into account the risks related to such as production, logistics or the overall operating environment. By using the AHP, the risk factors are mapped in the various areas, their importances are determined, and their impact on safety stock requirement is analyzed. Furthermore, strategic viewpoints such as the strategies of the supplier company and the customers can be included in the process of determining the safety stock level. As the AHP is a flexible decision support tool, the safety stock requirements can easily be adjusted whenever there are changes in the risk or strategic factors. E.g. a strike treat by the logistics service providers would trigger a safety stock requirement analysis.

Zhang Lianfu et al. [5] escalated the idea of the impact factors of the safety stock by a random variable, and analyzed them under random demand, random lead time, random requirement and certain lead time. They found the effect of increases in customer satisfaction with the increase of safety stock due to the shortage prevention. The establishment of safety stock level depends on the level of inventory demand, lead time changes as well as service level.

Qing-Yuan Li and Su-Jian Li [6] highlighted the safety stock management by VMI concept. The target is to cut down the storage cost and holding cost of the product and more efficient and more profitable in supply chain. The concept of VMI is focusing the management of stock based on the effective forecasting. The demand is not certain then people always put the uncertain factors in safety stock. The dynamic difference between forecast data and real demand data is applied in this paper to control the dynamic of safety stock. They used MEANS for applying the probability distribution to the research. The model of safety stock under VMI in this paper can show as below;

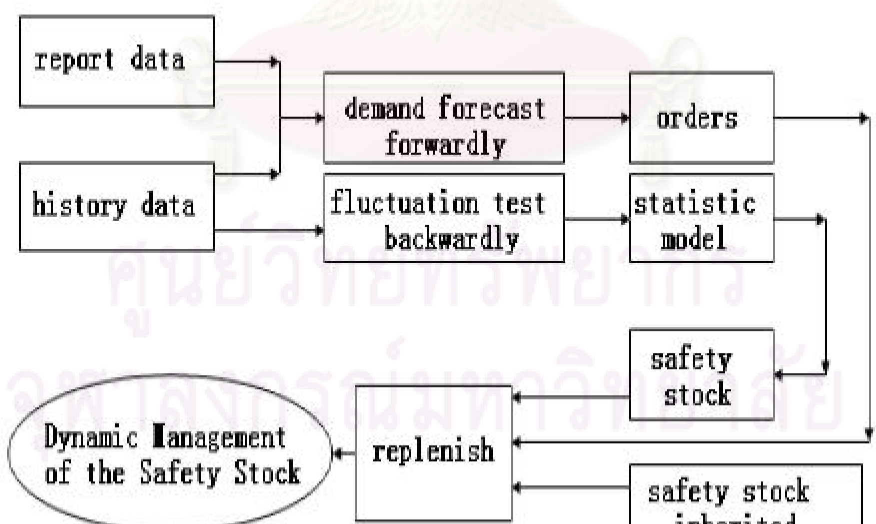


Figure 1.2: A total inventory solution under VMI

The solution covers both demand and safety stock which shown the final distribution of customer and lead to accurate safety stock of the inventory.

Zhao Ping and Liu Jingjing [7] introduced the artificial neural network is predicted the safety stock of the selling product. They claimed that the traditional method is limited to calculate the safety stock because it focus only demand and lead time. Although, in the real satiation, they are various factors is concerned to consider on calculation such as supply reliability, product cost, shortage cost, expectation of the service standard of the products. In this paper, they selected the 9 input layer nodes to be the factor to consider the safety stock as the following;

1) Lead time.

2) Fluctuation of lead time.

3) Supply reliability.

4) Product average demand.

5) Fluctuations in demand.

6) Product cost.

7) Shortage cost.

8) Product replacement rate.

9) Customer Service Level.

Their ANN architecture of this paper sets the maximum training times is 1000, the minimum expected error is 0.0001, and the learning rate is 0.1. and also they normalized the data set due to uniform measurement, accelerating neural network leaning and training speed.

Houmin Yan et al [8] proposed the models to redesign the manufacturing and distribution process to reduce the safety stock by specified service level. They suggested two models as the following;

1) To study product family contains one product and focus on resequencing of operations and shown in the figure 1.3

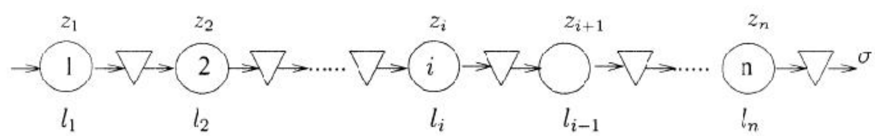


Figure 1.3: One-product model

2) To study product family contains two products and focus on merging of similar operations for postponing the point of differentiation and shown in the figure 1.4

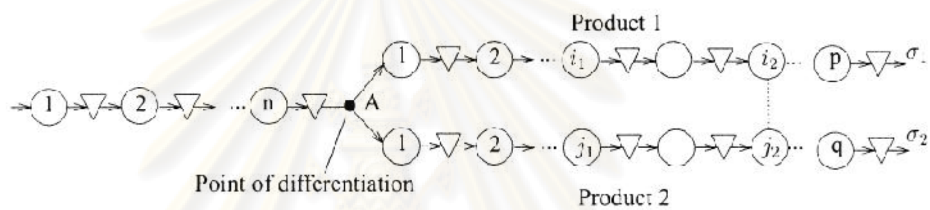


Figure 1.4: Two-product model

In the study, there are some limitations in their model in the term of uncertainties. Randomness in demand is the only uncertainty indicated in this paper. Other uncertainties, such as stochastic lead time, may affect in the safety stock reduction but they are not considered in this paper.

1.4 The Scope of study

1. Data is obtained from one of the Jewelry business in Thailand.
2. Eight variables are affected to calculate the consumption prediction by WEKA software based on the artificial neural network.
3. To take the consumption forecast to calculate the safety stock

CHAPTER II

BACKGROUND AND THEORETICAL FOUNDATION

2.1 Industry Background and Data Selection

2.1.1 Industry Background

Regarding to the fashion business, many things can change rapidly such as the market trends, customers' expectations and company objectives. Fashion has a short time horizon to the customer satisfaction and defense from the competition.

For this particular company, there are more than ten thousand Stock Keeping Unit (SKU) to support customers and more than 50,000 materials to produce these SKUs and keep in the warehouse. Jewelry product is consisting of the materials such as metal parts, crystal, plastic parts and packaging. There are suppliers both overseas and local which supply material to company. Company has to deliver the product to the main warehouse. After that, the product to each jewelry branch. Ordering is the short time from customer then safety stock acts as a buffer, its existence can enable enterprises to maintain a certain level of customer service. But how to set up a reasonable amount of safety stock. If the quantity of safety stock is set too low, then the shortage probability will increase, and rise out of stock cost, if which is set too high, will take up the capital, consume inventory management costs.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

In the company practice, safety stock is calculated using the business experience based on demand, lead time and scarp rate change.

2.1.2 Data Selection

One of the materials has been chosen because of its high overstock and high cost of holding. The data set is the actual consumption from the year 2008 until the mid of 2010.

Due to the complication in jewelry making, there are no standard inputs in the predicting process.

Thus, the experimental inputs are influenced by the safety stock research paper [1] and reviews of company performance in this situation. Here, the concerned variables include purchasing history, scrap rate, balance inventory, exchange rate, service level, minimum order, price of component and lead time. The description of each parameter is the following;

- X1 = Purchasing history is the volume of component over time bought from the suppliers.
- X2 = Scrap rate is the number of component which cannot be assembled due to the wrong specification.
- X3 = Balance inventory is the quantity left of stock in the warehouse at the end of month.
- X4 = Exchange rate is the rate of price which one country's currency is exchanged to the other currency.
- X5 = Service level is the performance measurement of a service.
- X6 = Minimum order is the lowest amount acceptable to a vendor.
- X7 = Price of a component is the selling price of component and
- X8 = Lead time is the number of days from the order date until components arrive at the factory.

The output is the component consumption, which will be further used to forecast safety stock.

2.2 Background on Data Mining

2.2.1 Data mining Concepts

Data mining is a technology that blends traditional data analysis methods with sophisticated algorithms for processing large volume of data [9]. It has also opened up exciting opportunities for exploring and analyzing new types of data and for analyzing old types of data in new ways. There are three well known data mining technique for predictive modeling, cluster analysis and association analysis.

- Predictive modeling related to develop classification models which capable of predicting the value of a class label (target variable) as a function of other variables (explanatory variables). The model is learnt from historical observations, where the class label of each sample is known, when constructed, a classification model is predicting the class label of new samples whose class is unknown, in the medical fields for forecasting whether a patient has a given disease based on the results of medical tests.
- Association analysis or pattern discovery concerned to discover patterns which described strong correlations of the features in the data or association of the features that happen frequently in the data. The discovered patterns are shown in the form of association rule, the benefit of analysis can use in market basket analysis, for example, the task of finding items are frequently purchased together based on point-of-sale data collection at cash registers.
- Cluster analysis involved to partition a data set into groups of closely related data in term of the observations belonging to different clusters. This analysis can be used, for instance, to find customer segments in a similar of purchasing behavior or to category of documents pertaining to related topics.

Data mining is a step of knowledge discovery in databases or KDD process to convert raw data into useful knowledge. The KDD process consists of a series of transformation steps:

- Data preprocessing: to transform the raw source data into an appropriate form for the subsequent analysis.
- Actual data mining: to transform the prepared data into patterns or models such as classification models, clustering models, association patterns, etc.

- Postprocessing of data mining results: to assess validity and usefulness of the extracted patterns and models, and presents interesting knowledge to the final users – business analysts, scientists, planners, etc. – by using appropriate visual metaphors or integrating knowledge into decision support systems.

2.2.2 Normalization

Normalization is the data transformation [10]; it may improve the accuracy and efficiency of mining algorithms related with neural networks, nearest neighbor and clustering classifiers. This method provided better results if the data to be analyzed have been normalized, that is, scaled to specific ranges such as [0.0, 1.0]. If using the neural network back propagation algorithm for classification mining, normalizing the input values for each attribute measured in the training samples will help speed up the learning phase. For distance-based methods, normalization is benefit for prevent attributes with initially large ranges from outweighing attributes with initially smaller ranges.

The process that represents transforming dissimilarity index from its value into a range of 0 and 1 is called *normalization* [11]. There are several ways to normalize an index as follows:

1. Min-max Normalization performs a linear transformation on the original data. Suppose that min_A and max_A are the minimum and maximum values of an attribute A. Min-max normalization maps a value name as v of A to *v'* in the range [new_min_A, new_max_A] by the equation below;

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

$$v' = ((v - min_A) / (max_A - min_A)) * (new\_max_A - new\_min_A) + new\_min_A \qquad (1)$$

2. z-score normalization, the value of attribute A is normalized based on the mean and standard deviation of A. The value v of A is normalized to v' by computing:

$$v' = ((v - \bar{A}) / \sigma_A) \qquad (2)$$

Where $\bar{A}$ and $\boldsymbol{\sigma}_A$ are the mean and the standard deviation of attribute A, respectively. This method of normalization is useful when the actual minimum and maximum of attribute A are unknown.

3. Decimal scaling normalizes by moving the decimal point of values of attribute A. The number of decimal points moved depends on the maximum absolute value of A. The value v of A is normalized to v' by computing:

$$v' = (v/10^j) \qquad (3)$$

Where j is the smallest integer such that Max (|v |) <1.

Normalization can change the original data and it is necessary to save the normalization parameters (the mean and the standard deviation if using the z-score normalization and the minimum and the maximum values if using the min-max normalization) so that future data can be normalized in the same manner.

2.2.3 Artificial Neural Network (ANN)

The artificial neural network (ANN) [12] is a mathematical structure designed to mimic the information processing functions of a network of neurons in the brain. ANNs that respond to inputs through modifiable weights, thresholds, and mathematical transfer functions that process information through many interconnected units are highly parallel systems. Each unit processes the pattern of activity it receives from other units, and then broadcasts its response to still other units. ANNs are particularly well suited for problems in which large datasets contain complicated nonlinear relations among many different inputs.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ANNs may not be well described by a set of known processes or simple mathematical formulae that are able to find and identify complex patterns in datasets.

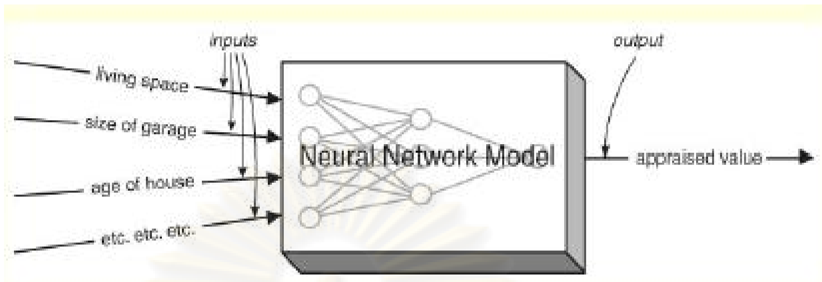


Figure 2.1: Classification; Neural Network [13]

In Figure 2.1 illustrated the artificial neural network received data in term of the number, after that, it proceed in the internal process and then transmitted to be the output

The types of Neural Network are as the following;

2.2.3.1 F. Rosenblatt was first introduced the Perceptron in 1958 [14]. It accepts only binary input and output value (0 or 1) that is a very simple neural net type with two neuron layers. The net is able to solve basic logical operations like AND or OR and the learning process is supervised. It is also used for pattern classification purposes. Perceptron can not solve more complicated logical operations (like the XOR problem).

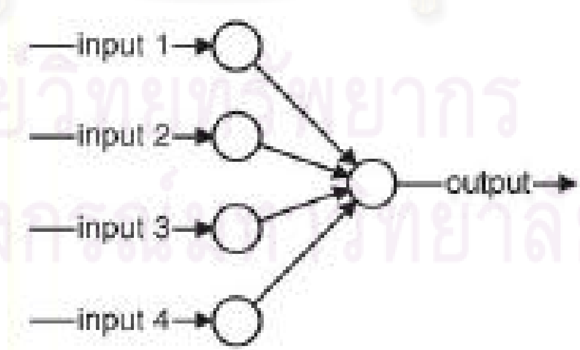


Figure 2.2: Perceptron [13]

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Figure 2.2 show the perceptron is the neural net which there is 2 layers such as Input layer and output layer. This one is similar to Logistic regression.

2.2.3.2 M. Minsky and Papert were first introduced the Multilayer Perceptron in 1969 [12]. It has one or more hidden neuron layers between its input and output layers and is an extended Perceptron. Multilayer Perceptron is able to solve every logical operation, including the XOR problem, due to its extended structure.

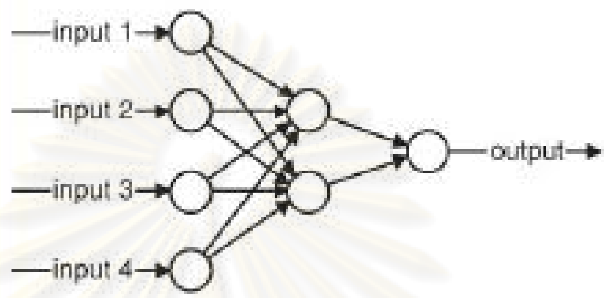


Figure 2.3: The MultiLayer Perceptron one or more hidden neuron layers [12]

Figure 2.3 reveals the multilayer perceptron which the neural net, there are 3 layers: input layer, hidden layer, and output layer.

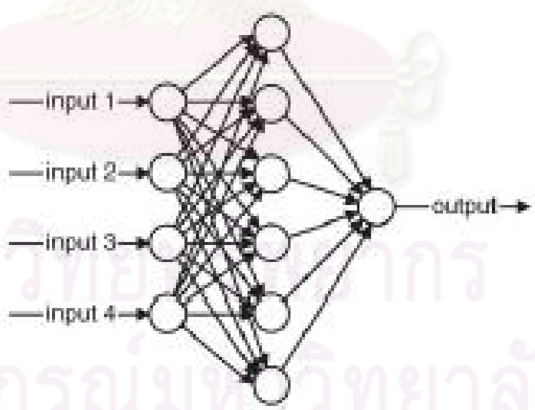


Figure 2.4: The MultiLayer Perceptron which multiple hidden neuron layers [12]

Figure 2.4 explains the MLP which add the number of node in the hidden node layer. Then, MLP can well learn nonlinear function in the other hand; it may have the problem in term of homogenous.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

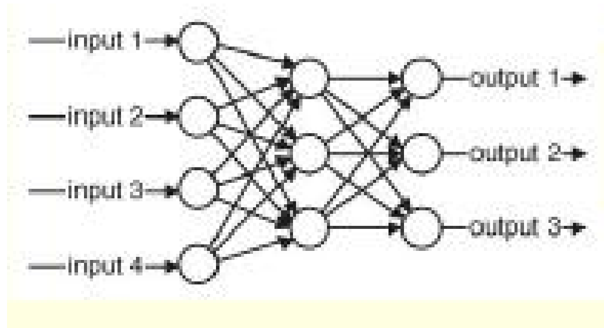


Figure 2.5: Exemplify the MLP allowed multiple output node [12]

A simple neurons mathematical model was presented by American psychologist W McCulloch and mathematician W Pitts in 1943, namely MultiLayer Perceptron model [13], which created the theoretical research about the artificial neural network model. The feedback interlinkage network and defined the energy function were proposed by American physicist J. J. Hopfield [14] in 1980s, it is the function about the neuron state and connection weights, which can be used to solve optimization problems and associative memory. In 1986, D. E. Rumelhartand J. L. McClelland brought forward the back-propagation algorithm of multilayer feedforward network, called BPnetwork or BP algorithm. This algorithm is used to solve the problems and it is the most widely used neural network algorithm presently (ref) that Perceptron cannot settle.

BP neural network can be divided into input layer, hidden layer, and output layer, and it belongs to the learning algorithm with mentor [ref], it is consisting of neurons and the connection between neurons. BP neural network is composed of positive propagation and the back propagation. In the positive propagation phase, the state of every layer neurons will only affect the neurons state in the next layer; if the expected output cannot be gotten in the output layer, the network enters into the error's back propagation phase. The error signal of back propagation network changes the networks connecting of all layers, to find out the best weight set and realize the correct network output. The output of the input layer neuron is equivalent to the input values.

After the publication of "Learning Internal Representations by Error Propagation" in 1986 (Though backpropagation itself dates from 1969), the rediscovery of the backpropagation algorithm [8] was probably the main reason behind the repopularisation of neural networks. The original network utilized multiple layers of weight-sum units of the type

$$f = g(w'x + b) \tag{4}$$

Where g was a sigmoid function or logistic function such as used in logistic regression.

A form of stochastic Gradient descent was done Training. The employment of the chain rule of differentiation in deriving the appropriate parameter seems to 'backpropagate errors', hence the nomenclature that updates results in an algorithm. However, it is essentially a form of gradient descent. Steepest gradient descent methods cannot be relied upon to give the solution without a good starting point, and determining the optimal parameters in a model of this type is not trivial. In recent times, networks with the same architecture as the backpropagation network are referred to as Multi-Layer Perceptrons. This name does not impose any limitations on the type of algorithm used for learning.

The backpropagation network was much controversy about whether such learning could be implemented in the brain or not, partly and there generated much enthusiasm at the time because a mechanism for reverse signaling was not obvious at the time, but most importantly because there was no plausible source for the 'teaching' or 'target' signal.

The algorithm of backpropagation learning can divide into two phases: propagation and weight update.

Phase 1: Propagation

Each propagation is involved the following steps:

1. A training pattern’s input through the neural network of forward propagation in order to generate the propagation's output activations.
2. The propagation’s output activations of back propagation through the neural network using the training pattern's target in order to generate the deltas of all output and hidden neurons.

Phase 2: Weight update

For each weight-synapse:

1. Multiply its output delta and input activation to get the gradient's weight.
2. Bring the weight of the gradient in the opposite direction by subtracting a ratio of it from the weight.

This ratio influence the speed and quality of learning is called learning rate. The sign of the gradient's weight indicates where the error is increasing; this is why the weight must be updated in the opposite direction. Repeat the phase 1 and 2 until the network performance is good enough.

2.3 Safety Stock Concept

Safety stock is acted as a buffer, its existence [7] can enable enterprises to maintain a certain customer service level. Customers are not the only source which can create a condition where demand exceeds the forecast. Suppliers may not ship the quantity ordered, thus creating an unplanned shortage. Incoming inspection may refute some or all of a receipt. Methods of destruct testing may render sample quantities unusable. The materials may be retained for periodic testing when samples have a shelf life. Loss may occur while handling processes, or in storage. Design problems or infant mortality may cause failures in testing. R&D usage may consume materials for new development of product. Service part requirements siphon off pars for the installed customer base. Random unplanned usage may occur if a given part is a designated alternate for another part which stocks-out.

Safety stock level [15] is a decision of management policy (business level). Companies have three techniques that use commonly at a part level:

- Statistical: safety stock is calculated based on the historical deviation of actual from planned usage via statistical methods
- Fixed: safety stock is set at a fixed quantity which is not interested the usage rate.
- Time Period: safety stock is set based on covering all of the some portion number of future periods' demand.

The part level selection technique is based primarily on the type of part (demand- vs. independent-demand), type of demand to which the part is subject (consistent or sporadic), and degree of variability to which the part is subject (low vs. high). There are many factors which affect the choice such as replenishment lead time, target customer service level, availability of history, part cost, number of time periods to protect from stock-out, impact on production or service of a stock out, and MRP or inventory subsystem designs. In addition, companies also over plan at the level of Master Schedule in order to create parts set of products in case the level of product is more than forecasted or providing the flexibility of option mix.

The companies' s problem face with the safety stock, however, is not the using technique. . The problem is determining how much to carry of which parts at any given time, based on the customer commitments, product life cycle, and competitive environment. The selected technique becomes an element within the integral strategy of company. Table 2.1 provides a comparison of safety stock techniques based on various company conditions.

CONSIDERATION	STATISTICAL	FIXED	TIME PERIOD
Phase-In Part	Ineffective for a new part unless new part picks up history from superseded part	Set high enough to prevent stock-out	Set demand (Forecasted Usage) high enough to prevent stock-out
Phase-Out Part	Set Customer Service Percent to 0	Set Fixed Safety Stock to 0	Set Time Period demand to 0 or a low quantity
Demand Profile (dependent, independent, or some combination)	Any demand profile, though erratic demand profiles will be smoothed	Steady or decreasing demand profile, since technique does not adjust to changing conditions	Any demand profile, since technique is based on demand in future periods
Variability Profile (demand, lead time, yield, and so on)	Any variability profile, though peaks will be smoothed	Any variability profile, with the Fixed Quantity set high enough to cover one or more peak conditions	Any variability profile, with the Time Period demand set high enough to cover one or more peak conditions
Typical Usage	Independent-demand parts where coverage is based primarily on part Safety Stock Customer Service Percent	Dependent-demand parts where safety stock coverage is set to zero or inexpensive independent-demand parts where excess coverage to provide a 100 percent Customer Service Level is desired	Dependent-demand parts where coverage is based primarily on days or weeks supply for a product

Table2.1: Comparison of Safety Stock Techniques [15]

2.4 Means Absolute Deviation (MAD)

Mean absolute deviation [16] is the approximated value of the standard deviation, which is the sum of the absolute deviations between the forecast and actual usage divided by the number of periods. The number of periods included in the average is important to the result. The mean absolute deviation's calculation is shown;

$$\text{MAD} = \sum |\text{actual - forecast}| / n \qquad (4)$$

One of the advantages of MAD is based on absolute values. Consequently, the errors of opposite signs do not cancel each other out when they are added. There are sum the errors, which is not interested the sign and obtain a measure of the average error. In case of comparison, we can select the method, which has the lowest MAD.

2.5 Tracking Signal

One problem is to identify whether the difference is caused by random variation or is due to a bias in the forecast, when there is a difference between forecast and actual values. Forecast bias is a persistent tendency for a forecast to be higher or lower the actual value of the data. We cannot do anything with random variation, but bias can be corrected.

Tracking signal is using to control the forecast bias. The tool used to monitor the quality of forecast is a tracking signal. It is calculation the ratio of the algebraic sum of the forecast errors divided by MAD:

$$\text{Tracking Signal} = \text{algebraic sum of forecast errors / MAD} \qquad (5)$$

Or

$$\text{Tracking Signal} = \sum (\text{actual} - \text{forecast}) / \text{MAD} \qquad (6)$$

The forecast errors are summed over time, which can indicate whether there is a bias in the forecast. In order to monitor the accuracy of forecast, the tracking signal values are compared against predetermined limits. The limits can range from $\pm$ 3 to $\pm$ 8 which are usually based on experience and judgment. The normally limits in the business use $\pm$ 4, which can be compared to 3 standard deviations limits. The forecast should be reviewed when the error fall outside these limits.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

CHAPTER III

METHODOLOGY

This chapter describes the experimental procedure which can be divided in to these following steps:

3.1 Data preparation

3.2 Training the network

3.3 Safety stock calculation

3.4 Evaluation

A framework is designed for the whole forecasting process. The framework in the experiment study is illustrated in Figure 3.1

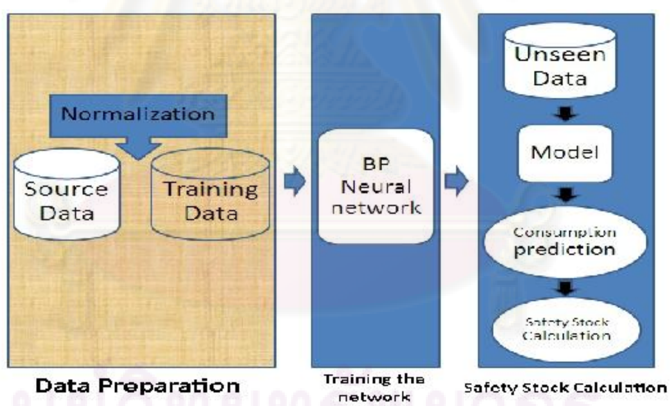


Figure 3.1: Experimental Framework [17]

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

3.1 Data preparation

The input variables mentioned in Chapter 2 will be further processed. Different input measurements are to be rescaled so that their variables reflect their importances. Standardizing either input or target variable tends to make the training process better behaved by improving the numerical condition of the optimization problem and ensuring the various default values involved in initialization and termination are appropriate.

Min-max normalization performs a linear transformation on the original data [11]. Suppose that min_A and max_A are the minimum and maximum values of an attribute A. Min-max normalization maps a value named as v of A to v' in the range [new_min_A, new_max_A] by computing;

$$v' = ((v - min_A) / (max_A - min_A)) * (new\_max_A - new\_min_A) + new\_min_A$$

Min-max normalization preserves the relationships among the original data values. It will encounter an “out-of-bounds” error if a future input case for normalization falls outside of the original data range for A.

In this thesis, data set are firstly normalized by min-max normalization into the range of [0,1] that will advantage for the weight balance process.

3.2 Training the networks

The ANN has proved to be an effective technique and is utilized in many areas such as engineering [18], medicine [19], and business [20].

The ANN can be considered as the information processing system that has certain performance characteristics in common with biological neural networks.

Figure 3.2 shows a typical architecture of a back propagation network. A back propagation neural network is a multi-layer neural network of error back propagation [1]. There are 3 layers such as input units, hidden units and output units. Each layer is composed of neurons, which are interconnected with each other by weights. In each neuron, a specific mathematical function called activation function accepts input from previous layer and generates output for the next layer. In the experiment, the utilized activation function is the hyperbolic tangent sigmoid transfer function [21].

The multilayer perceptron is trained under supervision using the back-propagation algorithm [1] using a well-known machine learning: WEKA [22].

The outputs from training the network are the predicted consumption values. These values will be furthered utilized for safety stock forecast.

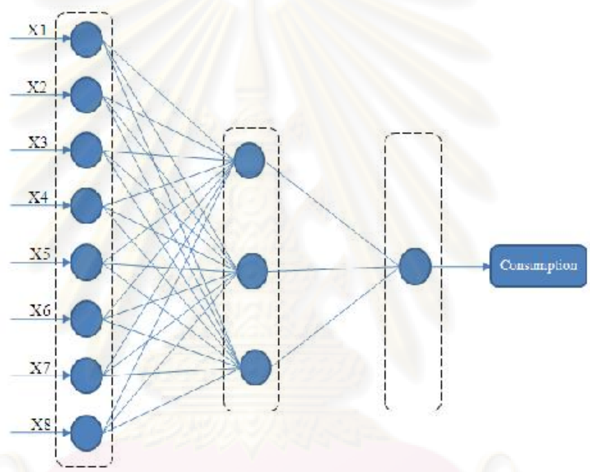


Figure 3.2: Back propagation ANNs Architecture

3.3 Safety Stock Calculation

Two purposes of safety stock;

3.3.1 on the customer side: to prevent and reduce the impact from the customers' unexpected demand.

3.3.2 on the supplier side: to prevent supplier stock shortage.

The safety stock can be given as [23];

Safety stock = standard deviation of error * service factor

$$S = \boldsymbol{\sigma} * \mathrm{Cdf}\,(P) \qquad (7)$$

Where

S is the required safety stock

$\boldsymbol{\sigma}$ is the standard deviation of the prediction error which is calculated from consumption during 6 periods [23].

Cdf (P) is referred as safety factor. It is the function of service level (P). Here, the service level has chosen to be 90% [24]. Thus, the safety factor is equal to 1.28.

Service level measures the performance. In this case, service level related to percentage of customers that do not experience an out-of-stock situation. We have to convert the service level into an error level so called the service factor and use the cumulative normal distribution to be a service factor [23].

The company preferred the service level of component to be 90% because they found that the risk to keep the over stock is 33.65 percent of component stock. On the other hand, service level at 95% and 99% have the risk in keeping the over stock to be 43.26 and 60.57 percent of component stock, respectively. However, this particular company has component stock budget 35 percent of inventory keeping. Thus, 90% of service level is chosen.

3.4 Evaluation

In business, one of the most widely used measurements for forecast accuracy is the mean absolute deviation (MAD) [16].

MAD is the average of the sum of the absolute errors;

$$\mathrm{MAD} = \frac{\sum|actual - predicted|}{n} \qquad (8)$$

The benefit of MAD is based on absolute values. The errors of opposite signs do not cancel each other out when they are added. If we are comparing different forecasting

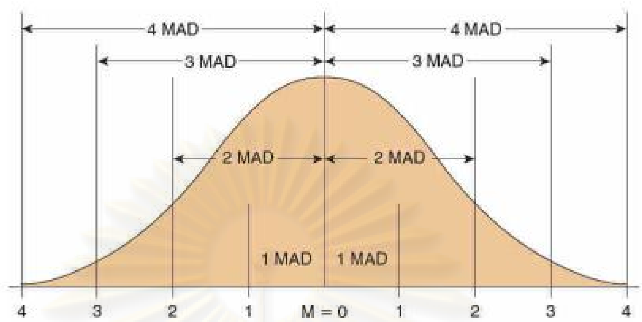


Figure 3.3: The distribution of MAD

Figure 3.3 shows the distribution of MAD which implies the bias in the tracking signal [20]. To monitor the forecast accuracy, the value of the tracking signal is compared against predetermined limits. This limits can range from ± 3 to ± 8. Generally, the acceptable limits for tracking signal are within ± 4, which correspond roughly to three standard deviations. If the errors fall outside these limits, the prediction should be reviewed [25]. The forecasting model is performing well, when the tracking signal should be around zero. The tracking signal indicates the direction of the forecasting error; the positive tracking signal indicates that the forecast is too high, while the negative tracking signal means that the forecast is too low [26].

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

CHAPTER IV

EXPERIMENTAL RESULTS

This chapter describes the experimental results of the proposed method consisting of four sections. The first section describes the experimental processing results. The second section is the evaluations and discussions.

4.1 The Experimental Processing Results

In this experiment, the WEKA software [22] was employed for the network training. The data set consists of 372 records. The data set is split into the training and test sets with the ratio 70:30. Therefore, there are 260 and 112 records for the training set and the test sets, respectively.

There are a number of parameters that need to be analyzed. Thus, many experiments are run to find the optimal result. The ranges of examined parameters are: the learning rate is between 0.1 to 0.3, the hidden nodes are between 2 to 10 nodes, and the momentum value is from 0.1 to 0.3. All combinations of these parameters are investigated.

Then, those parameters with the least mean absolute error are selected. According to the experiment results, the best architecture of the network is 8-3-1; there are 8 input nodes, 3 hidden nodes and 1 output node. The optimal parameters from the experiments are 0.1 for the learning rate, and 0.2 for the momentum. The number of epochs is set to 50000. The mean absolute error (MAE) is 0.0030. The correlation between the target and the predicted value is 0.9737 as shown in Figure 4.1. Figure 4.1 shows the linear relationship plot between the actual consumption and predicted consumption from the network. It can be easily seen that the majority of points are in the diagonal line. This indicates that the predicted values are close or the same as the target values. It is confirmed by the experimental results that the correlation between target and predicted value is 0.9737, which is very close to 1. With the assumption that the accurate consumption leads to the precise estimation in safety stock. Therefore, the predicted consumption values from the network are then used to calculate the safety stock.

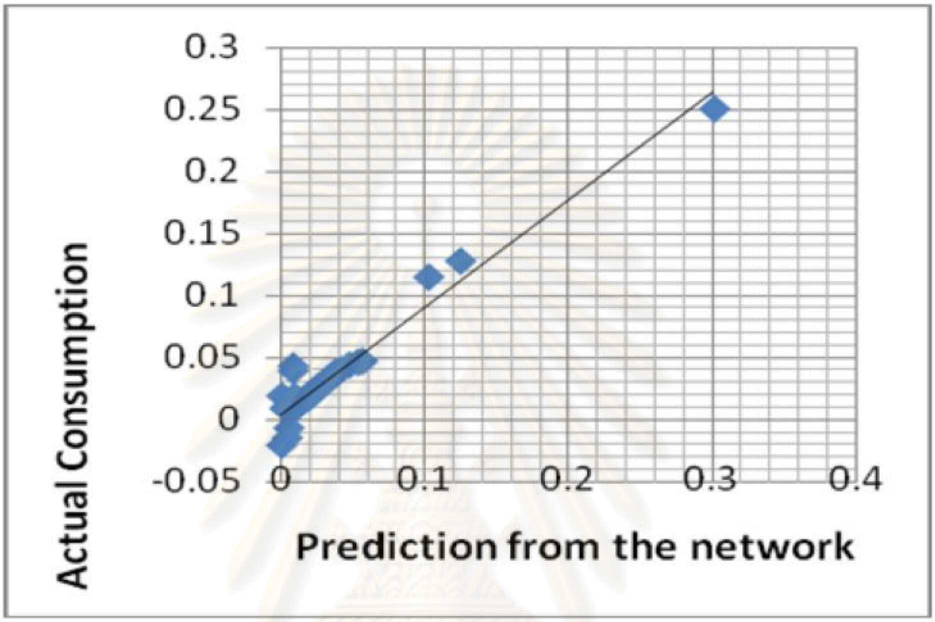


Figure 4.1: Linear relationship plot between actual consumption and predicted consumption from the network

4.2 Calculate Safety stock

To calculate safety stock, the predicted consumption from the network and actual consumption from the company are applied to (1). Two different safety stock values are derived. Here, they are called network safety stock (NW SS) and Ideal SS, respectively. After that, the MAD values are calculated. The first MAD value comes from the difference between NW SS and ideal SS. The second MAD value is the difference between company practice safety stock (CPN SS) and ideal SS. The comparison between the MAD values is illustrated as in Table 4.1. From Table 4.1, it can be seen that the MAD value of NW SS and ideal SS (0.256) is lower than CPN SS and ideal SS (0.298). This implies that the safety stock calculated from the network forecast values is better than the current company practice. This is confirmed by the value of the tracking signal (3). The tracking signal between NW SS an ideal SS is -2.771, which is in the range of 3 MAD. Whereas, the tracking signal between CPN SS and Ideal SS is 64.68. According to [11], the safety stock of the company practice should be reviewed.

Safety Stock Calculation	MAD	Tracking Signal
NW SS VS IDEAL SS	0.256	-2.771
CPN SS VS IDEAL SS	0.298	64.68

Table 4.1: Comparative safety stock calculation.

Safety stock service percentage is used in calculating a statistical safety stock and it is based on allowing an acceptable number of stock-outs within a given number of exposures to stock-out. An exposure to stock-out occurs at the point where stock is the lowest just prior to receiving the next order quantity. The service level is set to 90% in this study to imply that the company will satisfy component stock out 10 times of 100 component shortages. As the company focuses on the high service level of the finished product, they keep the finished product that always available for the end customer requirements. Therefore, 99% of service level for finished product is selected. This means that in term of the component level, the company allows the shortage for 10 out of 100 times but stock-out of finished product can accept the out of stock of finished goods only 1 out of 100 times shortages.

The tracking signal in this research is -2.771, which is under prediction. This can be implied that the quantity of forecast consumption is more than the actual consumption. However, the stock out that falls between -4 and +4 is the normal acceptable limit in which the general business can accept a bias prediction. In other words, the company can accept and allow the under prediction that fall in the acceptable range. A small number under forecasting in one period of time may not cause any problem until an under forecast in the next or later period consumes the safety stock, which will result in a stock out. Otherwise, the company has to pay more cost of keeping the component inventory in the warehouse for over prediction.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

CHAPTER V

CONCLUSION AND FUTURE WORKS

5.1 Conclusions

ANNs were used as a framework for consumption forecast. In this research, the source data from the Jewelry factory database consists of 372 records. The data set is pre-processed using normalization technique called max-min transformation. Then the data is separated in two groups: training data and testing data with the ratio 70:30. Hence, there are 260 and 112 records for the training set and the test set, respectively. The WEKA software was employed as the tool for the network training. Many experiments are run to find the optimal parameters. The ranges of the learning rate is between 0.1 to 0.3. The hidden nodes are between 2 to 10 nodes, and the momentum value is between 0.1 to 0.3. All combinations of these parameters are analyzed. The best architecture of the network is 8-3-1. The optimal parameters from the experiments are 0.1 for the learning rate, and 0.2 of momentum. The optimal number of hidden modes is 3. The number of epochs is set to 50000. The results illustrated that there is a high correlation (0.9737) between the ANN forecast and the actual consumption. This is with the assumption that consumption is a good predictor of the needed safety stock. Therefore, the safety stock can be calculated from the ANN prediction. From the error comparison using MAD and tracking signal, it is easily seen that the safety stock calculated from the forecast values from network is better than the company practice. It is important to note that, the ideal safety stock generally is not known in advance. Therefore, it is crucial for the company to make the precise estimation of the safety stock. Here, the ANNs are employed as a tool to get a better assessment of the safety stock.

Enhancing the forecast performance can help reduce the safety stock, which can benefit the company enormously.

5.2 Discussions

Regarding to the prediction in Jewelry business, the following issues should be highlighted when we apply the methodology in different circumstances:

- 8 input variables are employed in this study. Nevertheless, different choices of variables and numbers of variables utilized may result in the different outcomes.
- Choices of normalization technique could affect the final result.
- Preference in the various parameters such as service level percentage may influence the end result etc.,

Selection hidden unit number is one of the difficult tasks for training the neural network since there is no standard approach for choosing the numbers. In this research, the numbers are experimentally adjusted until the optimal result is discovered.

5.3 Future works

For the future work, the component consumption data are time series. The related data are increasing all the time. How to design a suitable model in time series? And the other factors should take into consideration for even better prediction such as different service level, holding cost etc. In addition, different methodologies should be applied for consumption predictions. The comparisons between techniques are important to find which method performs the best for this particular business.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

REFERENCES

[1] L.Zhange, D. Wang and L.Chang. A Model on Forecasting Safety Stock of ERP based on BP neural Network. Proc.Management of Innovation and Technology (2008): 1418-1422.

[2] Cao Qingkui and Ruan Junhu.Study on the Demand Forecsting of Hosipital Stocks Based on Data Mining and BP Neural Networks. Proc. International conference on electronic commerce and business intelligence. (2009):284-289.

[3] David C. Heath and Peter L. Jackson. Modeling the evolution of demand forecasts with application to safety stock analysis in production/distribution systems. Proc. Of IEEE Transaction. (1994): 17-30.

[4] Jukka Korpela, Antti Lehmusvaaa, Kalevi Kylaheiko and Markku Tuominen. Adjusting Safety Stock Requirements with anAHP-based Risk Analysis. Proc. Of the 36th Hawaii International Conference on System Sciences. (2003): 78a.

[5] Zhang Liangu, Zhao Shuzhi, Wang Min, Zhange Zhihui and Zhu Yonggang. Analyzing on Impact Factors of Safety Stock under Radom Requriement. Proc. International conference on Networks Security. (2009): 744-747.

[6] Qing-Yuan Li and Su-Jian Li. A Dynamic model of the safety stock under VMI. Proc. Of the Eighth International Conference on Machine Learning and Cybernetics. (2009): 1304-1308.

[7] Zhao Ping and Liu Jingjing. The product safety stock prediction method based on Artificial Neural Network. Proc. Of 2010 International Conference of Information Science and Management Engineering. (2010): 299-302.

[8] Houmin Yan, Chelliah Sriskandarajah, Suresh P. Sethi and Xiaohang Yue. Supply-Chain Redesign to reduce safety stock levels: Sequencing and Merging

Operations. Proc. Of IEEE Transactions on Engineering Management. (2002): 243-257.

[9] F. Giannotti and D. Pedreschi. Mobility, Data Mining and Privacy: A Vision of Convergence. Geographic Knowledge Discovery. (2008): 96-106.

[10 Luai Al Shalabi, Zyad Shaaban and Basel Kasasbeh. Data Mining: A Preprocessing Engine. Journal of Computer Science. (2006): 735-739.

[11] Alan G. Konheim. A Geometric Convergence Theorem for The Preceptron. Journal of the Society for Industrial and Applied Mathematics. (1963): 1-14.

[12] Marvin L. Minsky and Seymour A. Papert. Perceptrons: An Introduction to Computational Geometry. Cambridge MIT press. (1988): 292.

[13] S. M. Weiss and C. A. Kulikowski. Computer Systems that Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems. Morgan Kaufman, (1991).

[14] Stuat Russell and Peter Norvig. Artificial Intelligence A Modern Approach. 3rd ED., Prentice Hall. (2009).

[15] Paul Bernard. Integrated Inventory Management. NJ: John Wiley & Sons (1999).

[16] R. Dan Reid and Nada R.Sanders. Operations management An Integrated Approach. 4th Ed., John Wiley & Sons, Inc. (2010).

[17] Open Miner. Practical classification in WEKA [Online]. 2008. Available from: http://www.open-miner.com.

[18] Chun-guang Chang, Ding-wei Wan, Ya-Chen Liu and Bao-ku Qi. Application of Particale Swarm Optimization based BP Neural Network on Engineering Project Risk Evaluating. Proc. International conference on Natural Computation. (2007).

[19] Tadashi Kondo. Feedback GMDH-type neural network using prediction error criterion and its application to 3-dimensional medical image recognition. Proc. Of SICE Annual conference. (2008).

[20] Wilton.W.T Fok, IAENG, Vincent.W.L.Tam and Hon Ng. Computational Neural Network for Global Stock Indexes Prediction. Proc. Of the World Congress on Engineering 2008. (2008).

[21] S.P. Faulhaber. On the compatibility of adaptive controllers. Proc. 4th Annual Allerton Conference Circuits and Systems Theory. (1994): 8-16.

[22] Ian H.Witten and Eibe Frank. Data Mining: Practical machine Learning Tools and Techniques. 2nd Ed. Morgan Kaufmann Publishers. (2005).

[23] Lokad company. Safety Stock Calculation [Online]. Available from: http://www.lokad.com/calculate-safety-stock-with-sales-forecasting-ashx.

[24] Sven f.Crone, Stefan lessmann and Robert Stahlbock. Utility based Data Mining for Time Series Analysis-Cost-sensitive learning for Neural Network Predictions. Proc. Of 1st international workshop on utility-based data mining international conference on knowledge discovery and data mining. (2005): 59-68.

[25] Mark M.Davis and Janelle Heineke. Operations management and integrated Manufacturing and Services. 5th Ed. The McGraw-Hill company. New York. (2005).

[26] Norman Gaither and Greg Frazier. Operations management. 9th Ed. South Western/Thomson Learning. (2002).

Appendices

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Appendix A

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Appendix A

Introduction to WEKA

This chapter describes the WEKA software which is the tool for training neural network.[22] The WEKA was developed by University of Waikato, New Zealand, and WEKA stands for *Waikato Environment for Knowledge Analysis.* Outside the university the WEKA, pronounced to rhyme with Mecca, is a flightless bird with an inquisitive nature found only on the islands of New Zealand. The system distributed under the terms of the GNU General Public License that is written in JAVA. The system can run with any platform such as Linux, Window, and Macintosh operation system and even on a personal digital assistant. It can provide a uniform interface to many different learning algorithms, along with methods for pre and post processing and for evaluating the result of learning schemes on any given dataset.

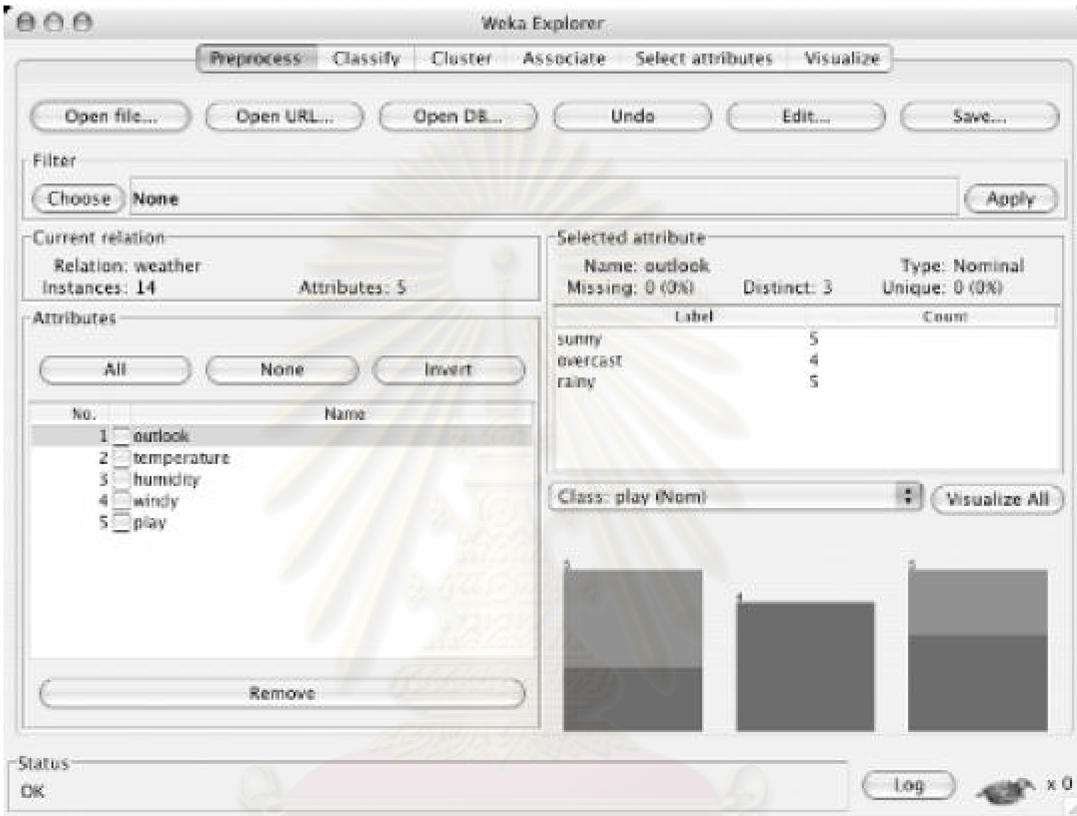
The easiest way to use Weka is through a graphical user interface called the Explorer.

1) Exploring the Explorer

WEKA has six tabs at the top of the explorer window in A.1 and A.2. There are all tabs as following:

1. Preprocess: To choose the dataset and to modify it in various ways.
2. Classify: To train learning schemes that perform classification or regression and to evaluate them.
3. Cluster: To learn clusters for the dataset.
4. Associate: to Learn association rules for the data and to evaluate them.
5. Select attributes: To select the most relevant aspects in the dataset.
6. Visualize: To view different two-dimensional plots of the data and interact with them.

Status box and Log button are at bottom of every panel. The status box displays messages in order to keep informed about what's going on



A.1: The WEKA Explorer.

4.2 Filtering algorithms

Unsupervised and supervised are the filters in WEKA when a filter is selected using choose button. Two filters transform the input dataset in some way. Filters are often applied to a training dataset and then also applied to the test file. When the filters is supervised, it uses values to derive good intervals for discretization, applying it to the test data will bias the results. It is the discretization intervals derived from the training data that must be applied to the test data. It must be careful to ensure that the results are evaluated fairly when supervised filters are used, an issue that does not arise with unsupervised filters.

A.2: Weka's the menu of filters.

Unsupervised and supervised filtering methods were treated separately by WEKA. The attribute filters work on the attributes in the datasets and instance filters work on the instance. There is a further distinction between attribute filters and instance filters.

4.2.1 Normalization

If the neural network were used directly, the effects of some attributes might be completely dwarfed by others that had large scales of measurement. So, it means that different attributes are measured on different scales. Consequently, it is usual to normalize all attribute values to lie between 0 and 1, by calculating

$$a_i = \frac{v_i - \min v_i}{\max v_i - \min v_i} \qquad (4)$$

Where *Vi* is the actual value of attribute *I*, and the maximum and minimum are taken over all instances in the training set.

4.3 Learning algorithms

When select a learning algorithm using the Choose button, Table A.1(a) and (b) lists Wake's classifiers on the Classify panel. There are many categories such as Bayesian classifiers, trees, rules, functions, lazy classifiers, and a final miscellaneous category.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Table 10.5 Classifier algorithms in Weka.

	Name	Function
Bayes	*AODE*	Averaged, one-dependence estimators
	BayesNet	Learn Bayesian nets
	ComplementNaiveBayes	Build a Complement Naïve Bayes classifier
	NaiveBayes	Standard probabilistic Naïve Bayes classifier
	NaiveBayesMultinomial	Multinomial version of Naïve Bayes
	NaiveBayesSimple	Simple implementation of Naïve Bayes
	NaiveBayesUpdateable	Incremental Naïve Bayes classifier that learns one instance at a time
Trees	*ADTree*	Build alternating decision trees
	DecisionStump	Build one-level decision trees
	Id3	Basic divide-and-conquer decision tree algorithm
	J48	C4.5 decision tree learner (implements C4.5 revision 8)
	LMT	Build logistic model trees
	M5P	M5′ model tree learner
	NBTree	Build a decision tree with Naïve Bayes classifiers at the leaves
	RandomForest	Construct random forests
	RandomTree	Construct a tree that considers a given number of random features at each node
	REPTree	Fast tree learner that uses reduced-error pruning
	UserClassifier	Allow users to build their own decision tree
Rules	*ConjunctiveRule*	Simple conjunctive rule learner
	DecisionTable	Build a simple decision table majority classifier
	JRip	RIPPER algorithm for fast, effective rule induction
	M5Rules	Obtain rules from model trees built using M5′
	Nnge	Nearest-neighbor method of generating rules using nonnested generalized exemplars
	OneR	1R classifier
	Part	Obtain rules from partial decision trees built using J4.8
	Prism	Simple covering algorithm for rules
	Ridor	Ripple-down rule learner
	ZeroR	Predict the majority class (if nominal) or the average value (if numeric)

Table A.1(a): Classifier algorithms in Weka.

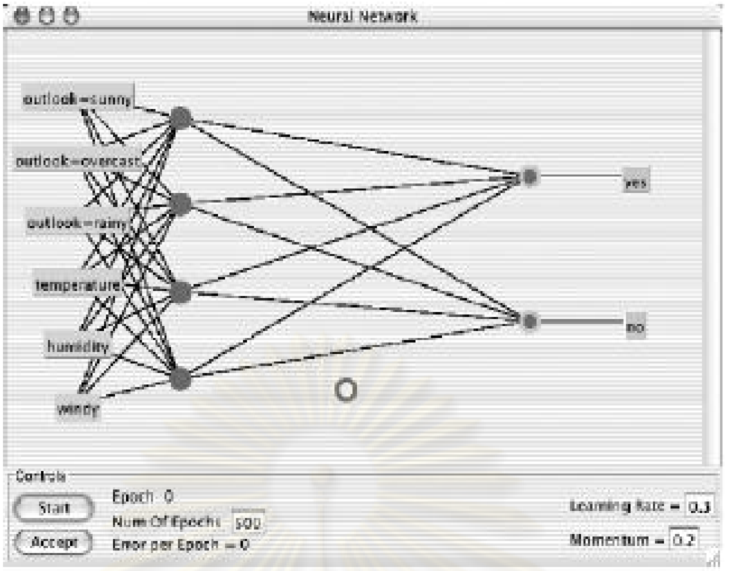
ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Table 10.5	**(continued)**	
	Name	Function
Functions	*LeastMedSq*	Robust regression using the median rather than the mean
	LinearRegression	Standard linear regression
	Logistic	Build linear logistic regression models
	MultilayerPerceptron	Backpropagation neural network
	PaceRegression	Build linear regression models using Pace regression
	RBFNetwork	Implements a radial basis function network
	SimpleLinearRegression	Learn a linear regression model based on a single attribute
	SimpleLogistic	Build linear logistic regression models with built-in attribute selection
	SMO	Sequential minimal optimization algorithm for support vector classification
	SMOreg	Sequential minimal optimization algorithm for support vector regression
	VotedPerceptron	Voted perceptron algorithm
	Winnow	Mistake-driven perceptron with multiplicative updates
Lazy	*IB1*	Basic nearest-neighbor instance-based learner
	IBk	*k*-nearest-neighbor classifier
	KStar	Nearest neighbor with generalized distance function
	LBR	Lazy Bayesian Rules classifier
	LWL	General algorithm for locally weighted learning
Misc.	*Hyperpipes*	Extremely simple, fast learner based on hypervolumes in instance space
	VFI	Voting feature intervals method, simple and fast

Table A.1(b): Classifier algorithms in Weka.

4.3.1 Neural networks

Neural network that trains using backpropagation is *MultilayePerceptron*. It differs from the other schemes because it has it own user interface, although listed under functions in Table A.1(a) and (b). The diagram in A.3 appears in a separate window, when load up the numeric version of the data, invoke MultilayerPerceptron, set GUI to True in it s object editor, and run the network by clicking Start on the Classify panel.



A.3: Weka's neural-network graphical user interface.

There are three layers in this network. Input layer is located on the left with one rectangular box for each attribute. Hidden layer is located next to it which all the input nudes are connected. The last layer is output layer that is located at the right. The output nodes represent the classes that are labels at far right.

Regarding configuring the structure of the network, you can control the learning rate. The momentum and the number of passes will take through the data, called epochs. When you click start button, the network begins to train, a running indication of the epoch, and the error for that epoch is shown at the lower left of the panel in A.3. Note that the error is based on a network that changes as the value is computed. For numeric classes the error value depends on whether the class is normalized. When the specified number of epochs is reached, the network stops at point which you can accept the result or increase the desired number of epochs and press start again to continue training.

These variables set the values by the parameters learning rate and momentum which can be overridden in the graphic interface. The learning rate to decrease with time was caused by a decay parameter that divides the starting value by the epoch number to obtain the current rate. This sometimes improves performance and may stop the network from diverging. When the graphical user interface is not used, the reset parameter automatically resets the network with a lower learning rate and begins training again if it is diverging from the answer.

The number of training epochs was set by the training time parameter. Alternatively, a percentage of the data can be set aside for validation (using validation-SetSize): then training continues until performance on the validation set starts to deteriorate consistently- or until the specified number of epochs is reached. No validation set is used, if the percentage is set to zero. The consecutive times, the validation set error can deteriorate before training is stopped, was determined by the validationThreshold parameter.

Default in the MultilayerPerceptron object editor specify the nominalToBinaryFilter filter; turning it off may improve performance on data in which the nominal attributes are really ordinal. The attributes and numeric class can be normalized. NormalizeAttributes and normalizeNumericClass may improve performance.

4.4 Evaluating numeric prediction

The alternative measures that used to evaluate the success of numeric prediction are as following;

4.4.1 Mean-squared error is the principal and most commonly used measure; sometimes the square root is taken to give it the same dimensions as the predicted value itself. Mean-squared error trends to be the easiest measure to manipulate mathematically when compare with many mathematical techniques. The mathematicians said that it is well behaved. However, mean-squares error has no particular advantage; all the performance measures are easy to calculate.

4.4.2 Mean absolute error is an alternative measures that is an average the magnitude of the individual errors without taking account of their sigh. Mean-squared error tends to exaggerate the effect of outliers-instance, which prediction error is larger than the others. But absolute error does not have this effect that all sizes of error are treated evenly according to their magnitude.

Sometimes mean absolute error is the relative rather than absolute error values that are of importance. For instance, the error of 50 in prediction of 500 or the error of 0.2 in a prediction of 2 that is 10% error is equally important, and then average of absolute error will be meaningless. The relative errors are appropriate. This effect would be taken into account by using the relative errors in the mean-squared error calculation or the mean absolute error calculation.

4.4.3 Relative squared error refers to something quite different that the error is made relative to what it would have been if a simple predictor had been used. The average of the actual values from the training data is the simple predictor in question. So, relative squared error takes the total squared error and normalizes value by divided the total squares error of the default predictor.

4.4.4 Relative absolute error that is the same kind of normalization is the total absolute error. The errors are normalized by the error of the simple predictor that predicts average value in these three relative error measures.

4.4.5 Correlation coefficient is measures the statistical correlation between the a's and the p's. The correlation coefficient value can range from +1 to -1. The meaning of +1 value is the perfectly correlated results, through 0 value is no correlation, then -1 value is the perfectly correlated negatively results. Normally, negative values should not occur for methods of reasonable prediction. Correlation is slightly different from the other measures because of scale independent in when you take a particular set of predictions. The error is unchanged when all the predictions are multiplied by a constant factor and the actual values are left unchanged. This factor appears in every term of SPA in the numerator and in every term of SP in the denominator, thus canceling out. (This is not true for the relative error figures, despite normalization: if you multiply all the predictions by a large constant, then the difference between the predicted and the actual values will change dramatically, as will the percentage errors.) The good performance of correlation coefficient leads to a large value that is different from other methods measure error which the good performance is indicated by small values.

Table 5.8 Performance measures for numeric prediction*.

Performance measure	Formula
mean-squared error	$\frac{(p_1-a_1)^2+\ldots+(p_n-a_n)^2}{n}$
root mean-squared error	$\sqrt{\frac{(p_1-a_1)^2+\ldots+(p_n-a_n)^2}{n}}$
mean absolute error	$\frac{\lvert p_1-a_1\rvert+\ldots+\lvert p_n-a_n\rvert}{n}$
relative squared error	$\frac{(p_1-a_1)^2+\ldots+(p_n-a_n)^2}{(a_1-\bar{a})^2+\ldots+(a_n-\bar{a})^2}$, where $\bar{a}=\frac{1}{n}\sum_i a_i$
root relative squared error	$\sqrt{\frac{(p_1-a_1)^2+\ldots+(p_n-a_n)^2}{(a_1-\bar{a})^2+\ldots+(a_n-\bar{a})^2}}$
relative absolute error	$\frac{\lvert p_1-a_1\rvert+\ldots+\lvert p_n-a_n\rvert}{\lvert a_1-\bar{a}\rvert+\ldots+\lvert a_n-\bar{a}\rvert}$
correlation coefficient	$\frac{S_{PA}}{\sqrt{S_P S_A}}$, where $S_{PA}=\frac{\sum_i (p_i-\bar{p})(a_i-\bar{a})}{n-1}$, $S_p=\frac{\sum_i (p_i-\bar{p})^2}{n-1}$, and $S_A=\frac{\sum_i (a_i-\bar{a})^2}{n-1}$

* p are predicted values and a are actual values.

Table A.2: Performance measures for numeric prediction.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

APPENDIX B

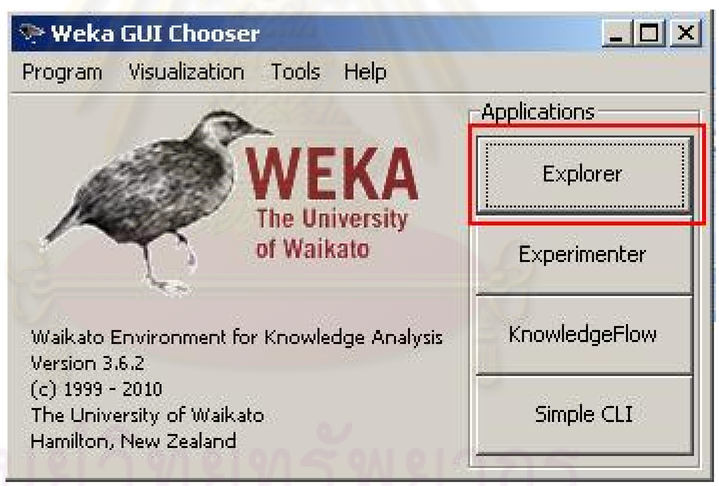
ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Appendix B

WEKA PROCESS

In this Thesis, we used WEKA software to run the data. The steps to use the WEKA software to run the data as the following;

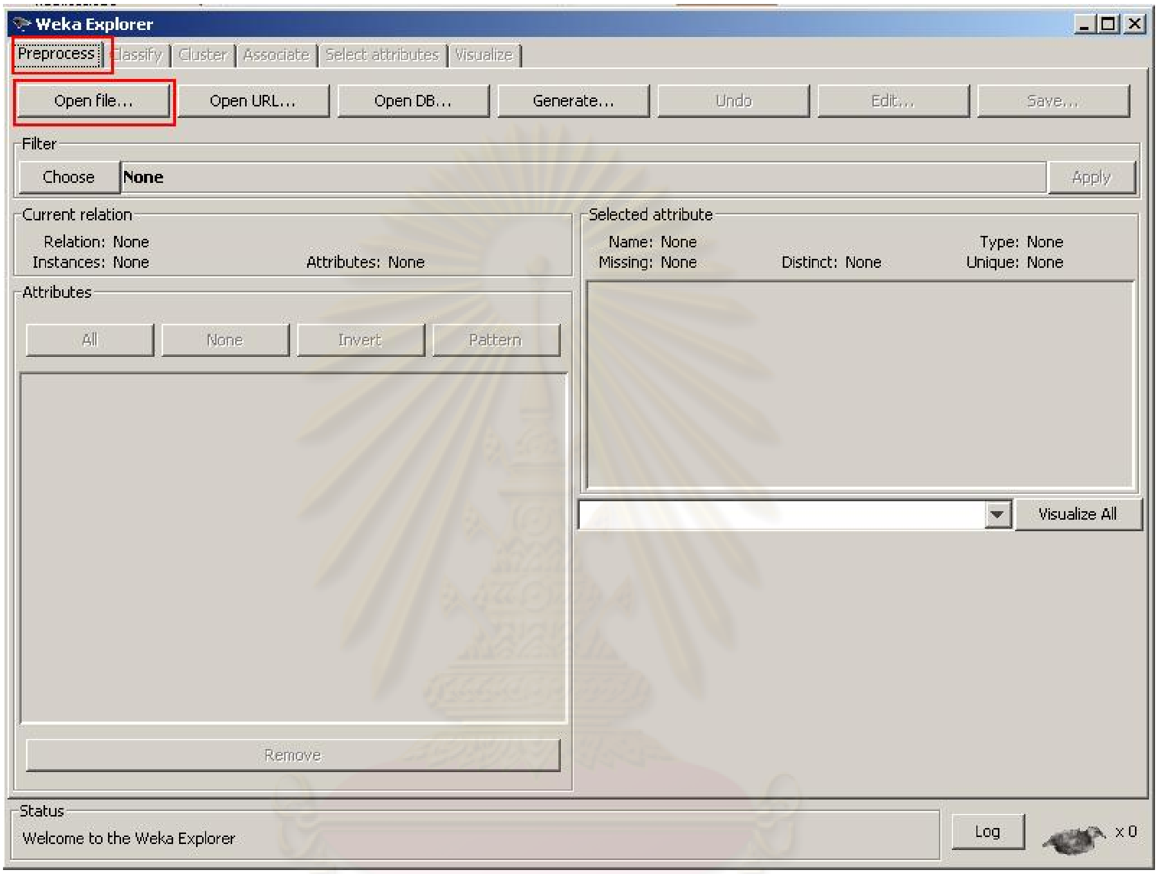
1. Prepare the file: there are 4 files in the various formats which use in the WEKA software such as ARFF, CSV, C4.5 and binary. In this thesis, CSV is selected to be the method to import the file to the software.
2. WEKA interfaces: The explorer is selected;



B.1: WEKA software 3..6.2 interface

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

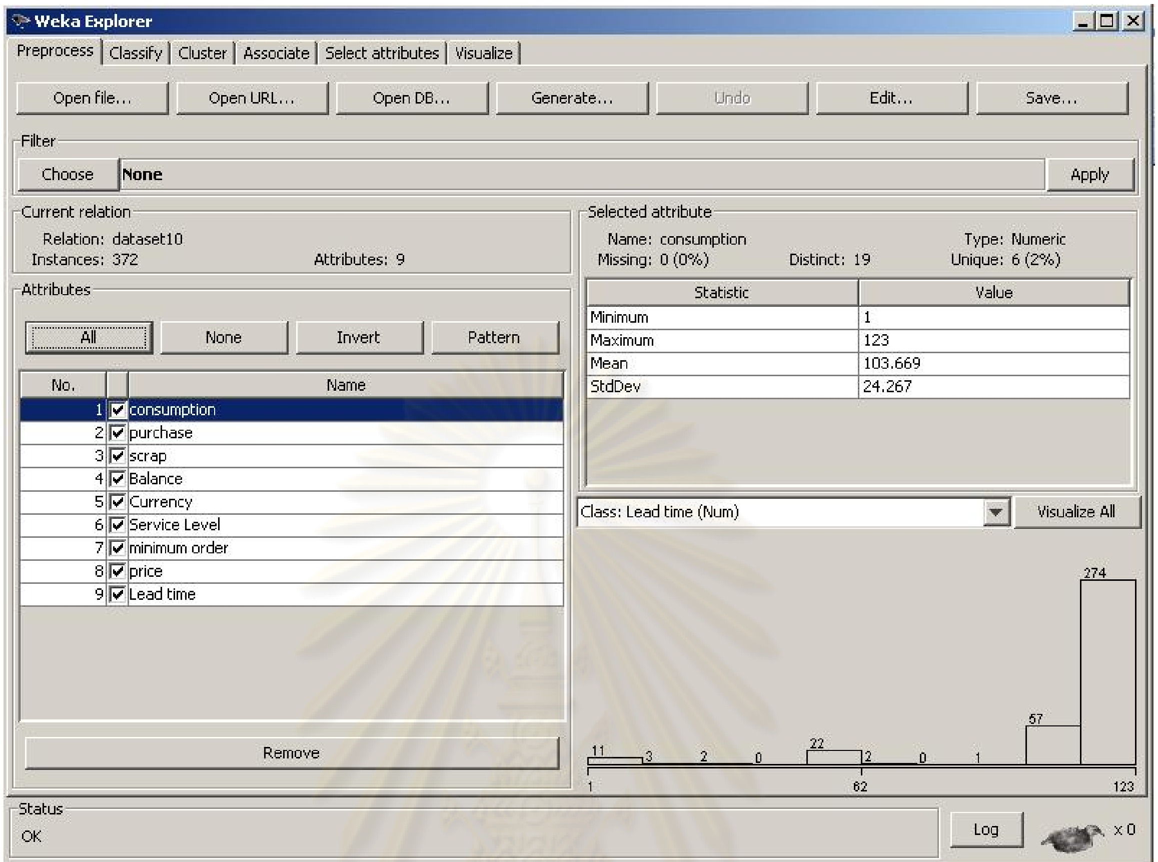
2.1 Preprocessing: Data can be imported from the files in the CSV format by clicking the Open file tab. The preprocessing tools can called “filters”.



B.2 : Preprocess step

Then, the CSV file format was loaded in the software.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



B.3: Data load in the WEKA software in the preprocess

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

2.2 Attribute selection: the artificial Neural Network calculated by the multilayerperceptron method, then set up the value is the essential such as

2.2.1 HiddenLayers: This defines the hidden layers of the neural network. This is a list of positive whole numbers.

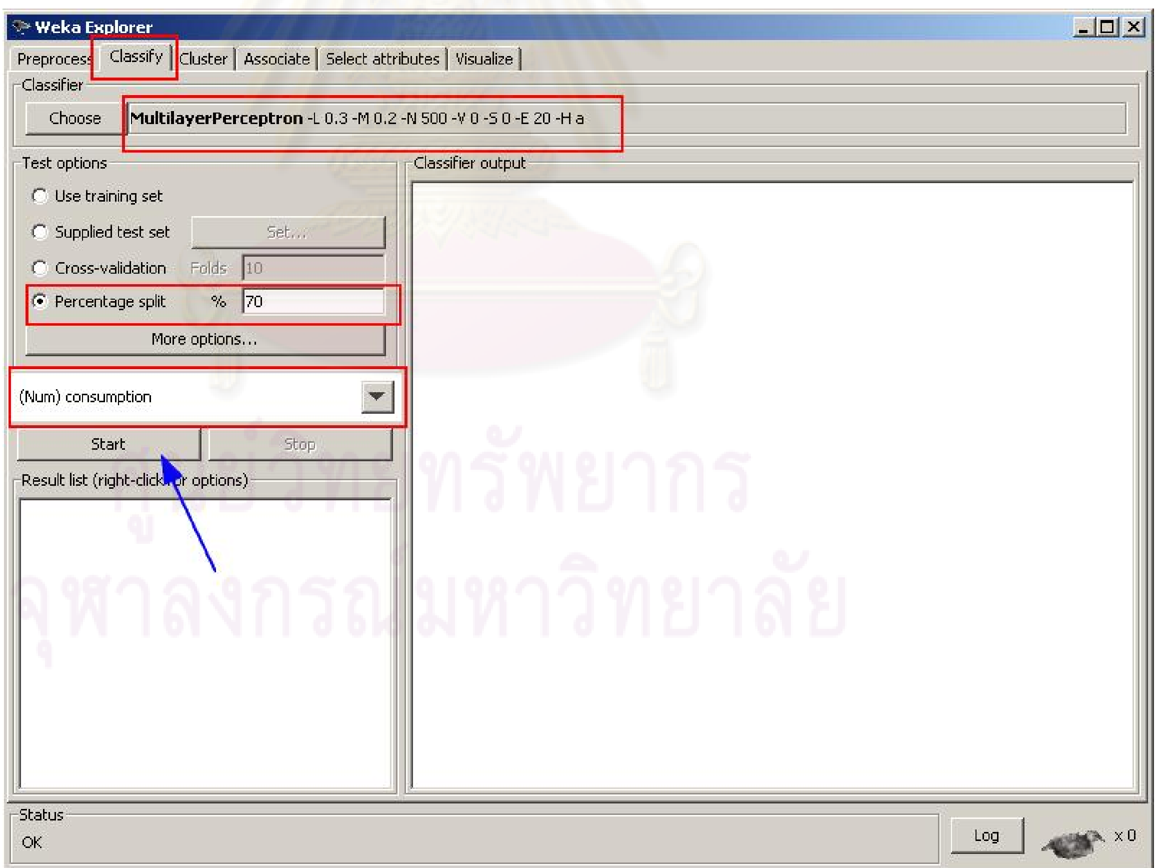
2.2.2 The learning rate: the amount the weights are updated

2.2.3 Momentum applied to the weights during updating.

2.2.4 Training time is the number of epochs to train through. If the validation set is non-zero then it can be terminate the network early.

Then the test options is selected one methods, in thesis, the percentage split 70:30 is selected to experiment.

The consumption is selected to be the output then start to learn the network.



B.4: The Exploer building classifiers by MultilayerPerceptron method

The result can be shown as the following;

=== Run information ===

Scheme: weka.classifiers.functions.MultilayerPerceptron

-L 0.1 -M 0.2 -N 50000 -V 0 -S 0 -E 20 -H 3

Relation: dataset10-weka.filters.unsupervised.instance.Normalize-N1.0-L2.0

Instances: 372

Attributes: 9

consumption

purchase

scrap

Balance

Currency

Service Level

minimum order

price

Lead time

Test mode: split 70.0% train, remainder test

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

Linear Node 0

Inputs Weights

Threshold 1.0057916397340927

Node 1 -0.21321732247259922

Node 2 -0.4280391440591935

Node 3 -1.4065677528593692

Sigmoid Node 1

Inputs Weights

Threshold -3.3748273465773058

Attrib purchase 10.837653005699158

Attrib scrap 3.2027340295930165

Attrib Balance -6.121180967536013

Attrib Currency -7.6738145966091444

Attrib Service Level -5.124757981282398

Attrib minimum order -5.148409662951375

Attrib price -7.719906960003263

Attrib Lead time -0.00913286343619013

Sigmoid Node 2

Inputs Weights

Threshold -2.567480161244726

Attrib purchase 2.7734368893217

Attrib scrap 2.4473539815751506

Attrib Balance 3.2180936995127785

Attrib Currency -8.97746087890873

Attrib Service Level -8.974511187195434

Attrib minimum order -8.975461713630406

Attrib price -8.85489277540162

Attrib Lead time -0.04505786773608095

Sigmoid Node 3

Inputs Weights

Threshold 2.313811041507252

Attrib purchase 11.889458730233207

Attrib scrap 0.38035909768289394

Attrib Balance 11.78136593334059

Attrib Currency -1.2063080622494868

Attrib Service Level 0.43791948155777779

Attrib minimum order 0.43390650297315775

Attrib price -0.9986499061438506

Attrib Lead time 0.02454507301736117

Class

Input

Node 0

Time taken to build model: 162.9 seconds

=== Predictions on test split ===

inst#,	actual,	predicted,	error
1	0.048	0.044	-0.004
2	0.009	0.009	0
3	0.007	0.008	0.001
4	0.019	0.018	-0.001
5	0.016	0.015	-0.001
6	0.008	0.009	0.001
7	0.027	0.026	-0.001
8	0.014	0.014	0
9	0.009	0.039	0.03
10	0.025	0.024	-0.001
11	0.009	0.009	0.001
12	0.042	0.04	-0.002
13	0.012	0.012	0
14	0.019	0.018	-0.001
15	0.009	0.01	0
16	0.01	0.01	0.001

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

17	0.013	0.013	0
18	0.021	0.019	-0.001
19	0.018	0.017	-0.001
20	0.01	0.01	0
21	0.015	0.015	-0.001
22	0.013	0.012	0
23	0.007	0.008	0.001
24	0.004	0.009	0.005
25	0.011	0.011	0
26	0.008	0.009	0.001
27	0.009	0.009	0
28	0.009	0.01	0.001
29	0.009	0.01	0.001
30	0.012	0.012	0
31	0.012	0.012	0
32	0.011	0.011	0
33	0.125	0.128	0.003
34	0.01	0.01	0
35	0.011	0.011	0
36	0.301	0.251	-0.051
37	0.015	0.014	0
38	0.02	0.018	-0.001
39	0.004	0.009	0.005
40	0.039	0.038	-0.001
41	0.008	0.009	0.001
42	0.008	0.009	0.001
43	0.016	0.016	-0.001
44	0.011	0.011	0
45	0.018	0.017	-0.001

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

46	0.015	0.014	-0.001
47	0.004	0.008	0.005
48	0.004	0.009	0.005
49	0.008	0.008	0.001
50	0.009	0.043	0.034
51	0.057	0.047	-0.01
52	0.008	0.009	0.001
53	0.054	0.046	-0.008
54	0.015	0.014	-0.001
55	0.009	0.01	0.001
56	0.017	0.016	-0.001
57	0.009	0.01	0.001
58	0.037	0.037	-0.001
59	0.013	0.013	0
60	0.011	0.011	0
61	0.01	0.011	0
62	0.007	0.009	0.001
63	0.026	0.026	-0.001
64	0.008	0.009	0.001
65	0.01	0.01	0.001
66	0.011	0.011	0
67	0.023	0.022	-0.001
68	0.013	0.012	0
69	0.011	0.011	0
70	0.008	0.009	0.001
71	0.02	0.019	-0.001
72	0.008	0.009	0.001
73	0.005	-0.015	-0.02
74	0.007	0.013	0.006

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

75	0.009	0.01	0.001
76	0.007	0.008	0.001
77	0.013	0.013	0
78	0.015	0.014	-0.001
79	0.01	0.01	0
80	0.008	0.009	0.001
81	0.031	0.03	-0.001
82	0.031	0.029	-0.001
83	0.005	0.01	0.005
84	0.008	0.009	0.001
85	0.014	0.014	-0.001
86	0.035	0.034	-0.001
87	0.013	0.013	0
88	0.007	0.009	0.001
89	0.01	0.01	0.001
90	0.012	0.012	0
91	0.009	0.009	0
92	0.009	0.01	0.001
93	0.001	-0.021	-0.022
94	0.008	0.009	0.001
95	0.02	0.019	-0.001
96	0.014	0.014	-0.001
97	0.021	0.02	-0.001
98	0.018	0.016	-0.001
99	0.008	0.009	0.001
100	0.01	0.019	0.009
101	0.006	-0.007	-0.013
102	0.011	0.011	0
103	0.103	0.115	0.011

104	0.011	0.011	0
105	0.009	0.01	0.001
106	0.013	0.013	0
107	0.001	0.019	0.018
108	0.011	0.011	0
109	0.022	0.021	-0.001
110	0.01	0.01	0.001
111	0.007	0.008	0.001
112	0.007	0.008	0.001

=== Evaluation on test split ===

=== Summary ===

Correlation coefficient	0.9737
Mean absolute error	0.003
Root mean squared error	0.0077
Relative absolute error	24.2183 %
Root relative squared error	24.227 %
Total Number of Instances	112

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

VITAE

Naroumon Yordphet was born on February 21st, 1972, in Phitsanuloke Province. She obtained her bachelor's degree in Public Administration from the Faculty of Art, Chaing Mai University in 1992.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย