A Comparison of Artificial Neural Network, Double Exponential Smoothing and Expert Judgement for Customer Demand Forecasts In Pharmaceutical Industry : A case study

Muhammad Fansuri Mohd Zahari\textsuperscript{1}, and Zalinda Othman\textsuperscript{2}
\textsuperscript{1}Faculty of Information Science and Technology
Universiti Kebangsaan Malaysia,
43600 Bangi, Selangor, Malaysia.
\textsuperscript{2}muhammadfansuri89@gmailcom, zalinda@edu.ukm.my

ABSTRACT
This paper presents a study of forecasting customer demand in the pharmaceutical industry. A good forecasting could help improve customer service satisfaction and reduce inventory cost. Therefore, three forecasting techniques are chosen which is the Artificial Neural Network, Double Exponential Smoothing and Expert Judgement. At the end of this paper, it shows that Artificial Neural Network prove a better result among the chosen techniques in this demand forecasting research.

Keywords: demand forecast, pharmaceutical, artificial neural network.

I INTRODUCTION
According to Aaron Marquis in his article, forecasting and estimation are interchangeable terms that basically mean predicting what will happen in the future, where the risk is when entering markets that have no need for the business's product.

Throughout our daily live, a famous application such as whether forecasting is something useful and crucial to discover so that we can plan and gain a better decision. Therefore it shows that industries as well should apply in their business process including the focus of this study in pharmaceutical industry which required a forecast involving the demand so that they can plan the inventory for the next production.


In forecasting, there are two groups which based on qualitative and quantitative. A qualitative method are ones that using survey-based or expert judgement such as Delphi method, Market research, Life cycle analogy and Judgment methods. The example of quantitative method is Time Series method such as Simple Moving Averages, Weighted Moving Averages, Simple Exponential Smoothing, Double Exponential Smoothing, while Associative method example such as Simple Linear Regression, Multiple Linear Regression and Nonlinear Regression.

Along this qualitative and quantitative method, Artificial Intelligence (AI) method should not be avoided as an option in forecasting such that we knew that an AI method is where the study involve the resembling of living things throughout their live.

An example of an AI forecasting or prediction method that are widely used are such as Artificial Neural Network (ANN) and Expert System.

As for this study, it will be focusing on applying neural network as the AI forecasting method. Where a comparison between the quantitative statistical method and AI forecasting method will be tested to show which is best predict the actual pattern.

II ARTIFICIAL NEURAL NETWORK
An artificial neural network is an information-processing system that has certain performance characteristics in common with biological neural networks (Fundamentals of Neural Networks). According to Jay Liebowitz, neural networks could loosely fall into the knowledge discovery area to try to infer patterns from data, knowledge, and images.

A biological neural network structure consists of a soma, dendrite, axon and synapse which an artificial neural network resembles as neuron, input, output and weight respectively.
The artificial neuron introduce by McCulloch-Pitts prove to be the earliest and the beginning of the neural net in 1940s. As from the time, the idea have been evolve until they come up with the idea of Hebb learning, perceptrons, Kohonen, Backpropagation and Boltzmann machine.

Neural network since, has been widely accepted and are used in industrial, business, medical, and in the research especially forecasting, pattern recognition, speech recognition, signal processing, control, speech production etc.

Among the few studies that apply ANN in demand forecasting is a study conducted by Wan Azlief Mahrawi (2011) on pharmaceutical, Junhu Qingkui and Ruan (2009) on hospital supplies, Ren Jiafu et al. (2009) on the spare part item, Zhu Ying and Xiao Hanbin (2010) in marketing and CA Mitrea et al. (2009) in inventory management.

According to a study conducted by Imam Santoso et al. (2007), the training shows a hypothesis that the greater the number of neurons in the hidden layer, the lower the error that would result otherwise review of the literature found, Patterson (1996), says that to get the best network, the neurons of the input layer should be greater of the second layer. Zhu Ying and Xiao Hanbin (2010) also suggests roughly where the number of neurons in the hidden layer to double the number of neurons in the input layer. So this study will apply a model with the number of neurons in the hidden layer is two)fold greater than the number of neurons in the input layer.

### III DOUBLE EXPONENTIAL SMOOTHING

One of the very famous statistical technique that have been applied are exponential smoothing (ES) technique. There are three types of ES usually applied which is Single ES, Double or known as Brown ES and finally Holt ES. Single ES involving single constant which is alpha α, Two constant for Brown ES which is alpha α and Beta β, and three constant for Holt with an addition of gamma γ constant.

According to Xiaochen Li (2013), the basic theory of DES or also known as Brown's Exponential Smoothing technique is similar to the average quadratic linear motion, when there is a pattern in which single and double smoothing smoothing, both of which will affect the actual value. Paul A. Jensen (2004) noted this technique estimates both ongoing and expected coefficients for linear prediction equations that model the pattern.


Studies conducted on demand forecasts using a combination of DES techniques is Guo Feng et al. (2012) which focused on spare parts demand forecasting model based on a combination of gray and exponential smoothing and Ayse Karanci (2011) on yatch demand forecasts and spatial statistical approach in the tourism sector.

### IV EXPERT JUDGEMENT

Expert judgment techniques are generally predictive forecasting techniques based on experience practicing experts in a particular field. Nowadays, this technique is seen as outdated and no longer used because most researchers are more concerned with new techniques or technique which is more systematic in addition to the aid of technology.

This view may well be right if judged on the number of studies by researchers who use this technique as a backup of their studies. However, there are still a handful of studies that are still able to prove the success of this technique in certain...
circumstances depending on the number of factors that included the last data, the experience of experts to make judgments, kind of long and short term forecasting, economic conditions and the quality of current health especially the health sector. Previous success using this technique may also be factors in selecting an organization from continuing to use the technique of this expert judgment.

Some related studies are performed by Mounir Ben Ghalia and Paul P. Wang (2000) which include technical considerations support site as intelligent systems to support demand forecasting of hotel rooms while the study that focused on a combination of judgment in their study: Fang-Mei Tseng, Tzung Lin and Hou (2012) research regarding the merger of scenario analysis technique, the Delphi method, and Diffusion of Innovation Model to analyze the development of Light Emitting Diode (Light-Emitting Diode, LED) and a study by Golam Kabir and Razia Sultana Sumi (2012) who conducted a study predictive power engineering companies demand combinations Fuzzy Delphi technique and artificial neural networks. Fuzzy Delphi acts in identifying the criteria that are not needed and thus eliminated from consideration before going through the process.

V METHODOLOGIES

A Data Collection
The data used in this study include data on product demand from customers weekly which is a real pharmaceutical product data. The data have been process manually which the original data are transferred into Spreadsheet (Microsoft Excel) from hardcopies. Data available includes two years data, from 2011 until 2012, then divided by weekly so as to collect a total of 104 weekly data for forecasting purposes. Only one product selected from a number of products produced to represent other products.

B Artificial Neural Network
The architecture of the ANN model used in this study is to have two input neurons which resemble two data which is weekly index and total number of demand, four hidden neurons and one output neuron. Four hidden neurons are choosen based on previous research done by Zhu Ying and Xiao Hanbin (2010) that suggests the number of neurons in the hidden layer to double the number of neurons in the input layer. The activation function used is the log-sigmoid function, learning rate of value 0.01, preliminary weighting of 0.3 and momentum of 0.6.

A learning rate is user-designated in order to determine how much the link weights and node biases can be modified based on the change direction and change rate. The higher the learning rate which is maximum number of 1.0, the faster the network is trained. A momentum rate set at the maximum of 1.0 may result in training which is highly unstable and thus may not achieve even a local minimum. If set at a low of 0.0, momentum is not considered and the network is more likely to settle into a local minimum. (University of Wisconsin System)

Some experiments which involve the repetition of the training is done with the 5000, 500, 300 and 100. Found a value of 100 is sufficient and allows the model is in its ability to reduce error. Figure 2 below shows a model of ANN used in this study.

VI EXPERIMENT RESULT ANALYSIS

Three methods for calculating the error that will be used is the Mean Absolute Percentage Error (Mean Absolute Percentage Error, MAPE), Mean Absolute Deviation (Mean Absolute deviation, MAD) and Mean Square Error (Mean Square Error, MSE). All three of these methods to the principle that the value of the available low is better to compare the forecasting techniques of the same or opposite.

Based on the result, it appears the most lowest average absolute deviation is at beta = 0.9, while the average error squared of the best shows on the beta = 0.8. Therefore, the beta at 0.8 and
0.9 will be tested in this study to get the best result using DES prediction techniques.

The aggregate value of the error calculation using the MAPE, MSE and MAD for each prediction technique is shown in Table 1 where DES1 refers to the DES study using a combination of (alpha, beta) = (0.9, 0.8), DES2 refers to the results of the study using a combination of DES (alpha, beta) = (0.9, 0.9).

<table>
<thead>
<tr>
<th></th>
<th>MAPE</th>
<th>MSE</th>
<th>MAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>12.68%</td>
<td>1671.32</td>
<td>30.95</td>
</tr>
<tr>
<td>DES1</td>
<td>42.24%</td>
<td>5941.80</td>
<td>62.70</td>
</tr>
<tr>
<td>DES2</td>
<td>49.66%</td>
<td>7633.92</td>
<td>71.59</td>
</tr>
<tr>
<td>Expert</td>
<td>53.71%</td>
<td>3575.08</td>
<td>50.03</td>
</tr>
</tbody>
</table>

Based on Table 1, the results obtained with the lowest value for measurement error using the MAPE is in favor of the ANN prediction techniques of 12.68%, followed by DES prediction techniques (alpha = 0.9, beta = 0.8) of 42.24%, DES (alpha = 0.9, beta = 0.9) shed 49.66% and the techniques of expert judgment by 53.71%.

Unlike the results shown using the percentage of MAPE, the results using the MSE calculation shows the arrangement of the other options but still maintained its forecast ANN techniques as a first choice with the smallest error. This arrangement followed the expert judgment forecasting technique, forecasting techniques DES (alpha = 0.9, beta = 0.8) and DES (alpha = 0.9, beta = 0.9).

Value calculation error MAD method also shows a slightly different arrangement. However, ANN prediction technique still shows the smallest error compared to other predictive techniques. Starting with ANN prediction technique that shows the smallest error, the order followed expert judgment forecasting techniques, forecasting techniques DES (alpha = 0.9, beta = 0.8) and DES prediction techniques (alpha = 0.9, beta = 0.9).

Based on figure 3, there exist best forecasts in a sequence showed on forecasting techniques of ANN and expert judgment technique. ANN technique shows the best forecast twice in quick succession in the second week and the third in turn gives the best prediction for three times in a row in week eight to ten weeks and two times in Week 13th and week 14th.

Expert judgement forecasting techniques also tried to show the best prediction of the results sequentially which occur twice in a row in week 23 to week 24, and the sequence of three times in a row in week 30 to week 32.

However, there was a sequence with the best and most improved prediction occurs with the use of ANN forecasting techniques than other predictive techniques as discussed.

VII CONCLUSIONS

In this study, the researchers decided to select forecasting techniques that can show the results of the calculation of the smallest error for two of the three methods used, in between MAPE, MSE and MAD. Once all of the results from each study forecasting techniques are collected, the researchers found that ANN forecasting techniques dominate the three methods of calculating error eligible discussed the dominating two of the three methods of calculation errors. Thus, ANN prediction technique proved to be a suitable technique to forecast the time series data of pharmaceutical products used in this study.

REFERENCES

