

# Comparative Analysis of Machine Learning Techniques in Sale Forecasting

Suresh Kumar Sharma  
Department of Computer Science and IT  
University of Jammu

Vinod Sharma  
Department of Computer Science and IT  
University of Jammu

## ABSTRACT

Forecasting is a systematic attempt to examine the future by inference from known facts. Sales forecasting is an ballpark figure of sales during a specified future period. Formerly, it was a manual process using the mathematical formulas. Due to the advent of computer the process of sale forecasting is fast and accurate. Machine learning, a subfield of Artificial Intelligence, has many algorithms that are used for forecasting. The aim of this research paper is to present a comparative analysis between the traditional methods of forecasting and machine learning techniques. A new technique known as combine approach which constructs from both moving average and ANN and interesting results so obtained are presented here. Experimental setup uses MATLAB.

## General Terms

Machine Learning Techniques, Sale Forecasting. MAE(Mean Absolute Error), MAPE(Mean Absolute Percentage Error), (MSE) Mean Square Error, RMSE(Root Mean Square Error).

## Keywords

Moving Average, EMA (Exponential Moving Average), ANN(Artificial Neural Network), KNN( K-Nearest Neighbor),

## INTRODUCTION

Sale Forecasting is a difficult area of management. Sale forecasting involves predicting the amount people will purchase, given the product features and the conditions of the sale. It is a day-by-day prediction of the level of sales you expect to achieve. In this research paper a step is taken to forecast sale with the help of computer based techniques [1].

### 1.1 Importance of sales Forecasting

Armed with the information about sale in advance can rapidly identify problems and opportunities in business and do something about them. For instances, accurately forecasting sales and building a sales plan can help us manage our production, staff and financing needs more effectively and possibly avoid unforeseen cash flow problems. The importance of sales forecasting can never be overstated. it helps the management to gear up production, ensure capital availability and calibrate marketing strategy, so that the company does not fall short of a sales target, which has been determined through sales forecasting. Sales forecasting helps the company in gauging the market demand for any of the new products that they may be planning to launch in the future [2][3].

### 1.2 Objectives of Sale Forecasting

Some important objectives of sale forecasting include (1) To formulating suitable production policy (2) Planning for Raw Material (3) Pricing Policy (4) Setting of Quotas (5)

Estimating Cash Inflows (6) Manpower Planning (7) Budgetary control over expenses [4].

## 2. DATA SET / HISTORIC DATA

The Historic data used to train and forecast has following important fields as shown in **Table 1**. This research work is general in nature i.e. we can collect the data about any type of item the sale of which is forecast [5][6][7].

Field/Factor	Possible Value
Temperature (a) Minimum Temperature (b) Maximum Temperature	The possible value of this field is the minimum and maximum value of the temperature in degree Celsius.
Day	Monday to Saturday 1,2,3,4,5,6 1 corresponds to Monday and 6 to Saturday
Holiday	0: Holiday 1: No Holiday
Sale of alternate product	Numeric value associated with the sale of the alternate product per day.
Sale of product.	Numeric value associated with the sale of the actual product

Table 1: Fields those are included in Data Set

### 2.1 Reasons for choosing these factors

(1) **Temperature:-** Temperature affects the sales as customers prefer to buy products in moderate temperature. Temperature varies day to day like sunny, rainy, humid, cloudy, hot, cold, hence affecting the sales. Moreover some products are seasonal also.

e.g. Cold drinks, AC's are sold more in summers. Geysers are sold in winters.

(2) **Day:-** Days also affect sales. As Monday comes after one holiday, sales are more on Monday.

(3) **Holiday:-** There is no sale on holiday.

(4) **Sale of alternate product:-** The alternate product also affects sales, as if alternative product has more sales; it means original product has less sale. Usually alternate product is of low cost but also of low quality and it attracts customers.

(5) **Sale of product:-** The sale of the product on current day also helps in forecasting the next day sale.

Figure 1 Shows the data set used in the experimental setup for training purpose.

LOAD DATASET						
VIEW DATA						
LOAD DATA SET						
FILE NAME:		C:\DATA.txt		BROWSE	VIEW DATA	
16.1	27.6	2	1	308	735	
18.4	28.5	3	1	318	832	
18.2	28.2	4	1	328	758	
15.5	28.1	5	1	298	798	
15.3	27.2	6	1	100	228	
11.9	28.3	1	1	288	928	
13.5	28.3	2	1	312	750	
13.1	27.3	3	1	350	800	
14.9	28.1	4	0	0	0 Training Data	
14.8	22.9	5	1	420	768	
18	30	6	1	88	125	
14.4	28.2	1	1	360	850	
14.1	27.4	2	1	320	794	

Figure 1. Data Set/Historic Data

### 3. MODELS USED FOR FORECASTING

In this research paper, A Comparative analysis takes place among Moving Average, Exponential Moving Average, Artificial Neural Network, KNN and Combined Approach. A brief introduction about the models used is as follow:

#### 3.1 Forecasting through Moving Average:

Moving average is also called rolling average, rolling mean or running average. A moving average is commonly used with time series data to smooth out short term fluctuations and highlight longer term trends or cycles. Time series is a collection of data recorded over a period of time i.e. weekly, monthly, and quarterly. In time series data, an analysis of its history can be used by management to make current decisions and plans based on long term forecasting. Time series data usually assumes past pattern to continue into the future. The two most popular types of moving averages are

##### (i) Simple Moving Average

A simple moving average is formed by computing the average price of a security over a specific number of periods. A simple moving average of order k, MA(k) is the value of k consecutive observations. Mathematically, it is

$$F_{t+1} = \frac{(y_t + y_{t-1} + y_{t-2} + \dots + y_{t-k+1})}{k}$$

Here k is the number of terms in the moving average. In simple moving average the value of  $F_{t+1}$  depends upon k. Simple moving average (SMA) is a special case of weighted moving average in which the same weight is assigned to all the data in average [8].

##### (ii) Exponential Moving Average

A type of moving average that is similar to a simple moving average, except that more weight is given to the latest data. This type of moving average reacts faster to recent price changes than a simple moving average. The 12 days and 26 days Exponential moving averages (EMA) are the most popular short term averages. In general, The 50 and 200 days EMA are used as signals of long term trends. This method provides an exponentially weighted moving average of all previously observed values. EMA is appropriate for data with no predictable upward or downward trend. The aim is to estimate the current level and use it as a forecast of future value. Formally, the exponential smoothing equation is

$$F_{t+1} = \alpha y_t + (1 - \alpha)F_t$$

$F_{t+1}$  = forecast for the next period.

$\alpha$  = smoothing constant.

$y_t$  = observed value of series in period t.

$F_t$  = old forecast for period t.

The implication of exponential smoothing can be better seen if the previous equation is expanded by replacing  $F_t$  with its components as follows:

$$F_{t+1} = \alpha y_t + (1 - \alpha)F_t$$

$$F_{t+1} = \alpha y_t + (1 - \alpha)[\alpha y_{t-1} + (1 - \alpha)F_{t-1}]$$

$$F_{t+1} = \alpha y_t + \alpha(1 - \alpha)y_{t-1} + (1 - \alpha)^2 F_{t-1}$$

If this substitution process is repeated by replacing  $F_{t-1}$  by its components,  $F_{t-2}$  by its components, and so on the result is:

$$F_{t+1} = \alpha y_t + \alpha(1 - \alpha)y_{t-1} + \alpha(1 - \alpha)^2 y_{t-2} + \alpha(1 - \alpha)^3 y_{t-3} + \dots + \alpha(1 - \alpha)^{t-1} y_1$$

Therefore,  $F_{t+1}$  is the weighted moving average of all past observations.

The exponential smoothing equation rewritten in the following form elucidate the role of weighting factor  $\alpha$ .

$$F_{t+1} = F_t + \alpha(y_t - F_t)$$

Exponential smoothing forecast is the old forecast plus an adjustment for the error that occurred in the last forecast. The value of smoothing constant  $\alpha$  must be between 0 and 1.  $\alpha$  cannot be equal to 0 or 1. To start the algorithm, we need  $F_1$  because Since  $F_1$  is not known; we can set the first estimate equal to the first observation. Use the average of the first five or six observations for the initial smoothed value [9].

#### 3.2 K-Nearest Neighbor

K-Nearest neighbor technique is one of the prediction methods used in machine learning. It is based on the idea that new object is classified based on attributes and training sample using a majority of k-nearest neighbor category.

In order to apply this technique, it is necessary to have a training set and a test sample, to know the value of k (how many neighbors are used in prediction) and the mathematical formula of the distance calculated between the instances. In general, this distance is determined with Minkowski distance given by the formula

$$\left( \sum_{i=1}^n |x_i - y_i|^p \right)^{1/p}$$

where  $x_i$  represents the test sample,  $y_i$  is the training data, n is the number of features.

The k nearest neighbor classifier is commonly based on the Euclidean distance between a test sample and the specified training samples.

$$\sqrt{\left( \sum_{i=1}^n |x_i - y_i|^2 \right)}$$

It is based on minimum distance from the test instance to the training samples to determine the k nearest neighbors. After k-nearest neighbors are selected, the majority of these k nearest neighbors decides the prediction of the new instance

The general algorithm of computing the k-nearest neighbors is as follows:

**Step1 :** Establish the parameter k that is the number of nearest neighbors;

**Step2:** Calculate the Euclidian distance between the query-instance and all the training samples;

**Step3:** Sort the distances for all the training samples and determine the nearest neighbor based on the k-th minimum distance;

**Step4:** Use the majority of nearest neighbors as the prediction value [10][14].

### 3.3 Artificial Neural Network

The various steps in Artificial Neural Network with backpropagation algorithm are:[5][8][11].

**Step1:** Initialize the weights and biases. The weights in the network are initialized to small random numbers ranging for example from -1.0 to 1.0, or -0.5 to 0.5. The biases are similarly initialized to small random numbers.

**Step 2:** Feed the training sample (record)

**Step 3:** Propagate the inputs forward; we compute the net input and output of each unit in the hidden and output layers.

**Step 4:** Backpropagate the error.

**Step 5:** Update weights and biases to reflect the propagated errors.

**Step 6:** Repeat and apply terminating conditions.

### 3.4 Combined Approach

In the combined approach we first find the moving average of one week, two week, one month, one quarter, half yearly and last week Sale. These six values are act as input to Artificial Neural Network with backpropagation. Thus instead of using all the fields of data set we use only sale of product field in this approach. In this technique we use simple moving average instead of exponential moving average.

## 4. CRITERIA FOR COMPARATIVE ANALYSIS

Following Errors are used to measure the efficiency of the forecasting Model and in turn plays an important role in comparative analysis [12][13][15]

- (i) Mean Absolute Error (MAE)
- (ii) Mean Absolute Percentage Error ( MAPE)
- (iii) Mean Squared Error (MSE)
- (iv) Root Mean Square Error ( RMSE)

### 4.1 Mean Absolute Error

Mean Absolute Error is a common measure of forecast error in time series analysis. The mean absolute is a quantity used to measure how close forecasts or prediction are to the eventual outcomes. As the name suggests, the mean absolute error is an average of the absolute errors.

$$e_i = |(f_i - y_i)|$$

$f_i$  : Prediction

$y_i$  : True Value

$$MAE = \frac{1}{n} \sum_{i=1}^n e_i$$

Lower Values of MAE are better.

### 4.2 Mean Absolute Percentage Error

The mean absolute percentage error is a very popular measure that corrects the “ canceling out” effect and also keeps into accounts the different scales at which this measure can be computed and thus can be used to compare different predictions.

$$|PE| = \left| \left( \frac{Error}{original} \right) \right| * 100$$

$$MAPE = \sum \frac{|PE|}{n}$$

In general MAPE of 10% is considered very good. A MAPE in the range of 20% to 30 % or even higher is quite common. Lower the MAPE, the more accurate the forecast

### 4.3 Mean Square Error (MSE)

It is the arithmetic mean of the sum of the square of prediction error. This error measure is popular and corrects the cancelling out effects.

$$e_i = |(f_i - y_i)|$$

$f_i$  : Prediction

$y_i$  : True Value

$$MSE = \frac{1}{n} \sum_{i=1}^n (e_i)^2$$

The MSE index ranges from 0 to  $\infty$ , with 0 corresponding to the ideal. Lower MSE is better.

### 4.4 Root Mean Square Error (RMSE)

RMSE is a good measure of accuracy. RMSE is the square root of the MSE

$$RMSE = \sqrt{MSE}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (e_i)^2}$$

The RMSE will always be larger or equal to the MAE. It is ideal if it is small i.e. ranges from 5 to 20%.

## 5. EXPERIMENTAL RESULTS

The Experimental Results those include Comparative analysis of various type of errors for all the forecasting models is shown in **figure 2**. The line graph among the actual and forecasting value by using all the models is shown in **figure 3**. In figure 2 , researchers load the data file which contains actual forecasting and forecasting through respective method. For instances the data in the file *annresults.txt* has two columns first column is the forecasted Value for five days and second column contains the actual values. Based on these values MAE, MAPE, MSE and RMSE are computed as shown in the figure. Similar process is used for other techniques. In figure 2, x-axis represents number of days and y axis represents the sale at that day. After loading the data files associated with each method line graph among forecasted values by using these methods is drawn by clicking line graph button. Line graph associated with each method is shown in this figure. When we critically analyze these two figures we concluded that the combined approach is better for sale forecasting among all other machine learning techniques that are considered in this research work.

## 6. CONCLUSION AND FUTURE SCOPE

In this research paper we apply various models of forecasting to the data related to sale. After analyzing experimental results researchers concluded that the combined approach i.e. where we use both moving average and Artificial Neural Network with backpropagation gives accurate results than all other techniques. One more question that fits as a future scope of this research paper is what happens when we use Exponential moving average instead of Simple moving average in the combined approach.

## 7. REFERENCES

- [1] J. Holton Wilson, Barry Keating, 1998, Business forecasting
- [2] TR Jain, SC Aggarwal, 2009, Business Mathematics and Statistics.
- [3] A.K. Sharma, 2005, Text Book Of Business Statistics.

- [4] C.L. Tyagi, Arun Kumar, 2004, Sales Management
- [5] Suresh Kumar Sharma, Vinod Sharma, “ANN Approach for Daily Prediction of Gas Load Forecasting”, 5th National Conf. Computing For Nation Development; INDIACOM.2011
- [6] Chen-Yuan Chena, Wan-I Leeb, Hui-Ming Kuoc, Cheng-Wu Chen, Kung-Hsing Chenf, “The study of a forecasting sales model for fresh food”, Expert Systems with Applications, 2010.
- [7] Yihang Liu, “Sales Forecasting through Fuzzy Neural Networks”, International Conference on Electronic Computer Technology, 2009 .
- [8] Suresh Kumar Sharma, Vinod Sharma, Proficient Prophecy of Foreign Exchange Rate using Artificial Neural Network: A Case of USD to INR. International Journal of Computer Applications, 2012
- [9] Ken Black, 2010, Business Statistics: For Contemporary Decision Making
- [10] Suresh Kumar Sharma, Vinod Sharma Time Series Prediction Using Knn Algorithms Via Euclidian Distance Function: A Case Of Foreign Exchange Rate Prediction, 2012.
- [11] S.Rajasekaran, G.A.Vijayalakshmi Pai, 2004, Neural Networks, Fuzzy Logic and Genetic Algorithms: Synthesis and Applications.
- [12] Thiesing, F.M., “Sales forecasting using neural networks”, International Conference on Neural Networks, 1997.
- [13] Yuquan Qin; Haimin Li, “Sales forecast based on BP neural network”, 3rd International Conference on Communication Software and Networks (ICCSN), 2011.
- [14] Elia Georgiana Dragomir, “Air Quality Index Prediction using K-Nearest Neighbor Technique”, Petroleum - Gas University of Ploiesti Bulletin, 2010

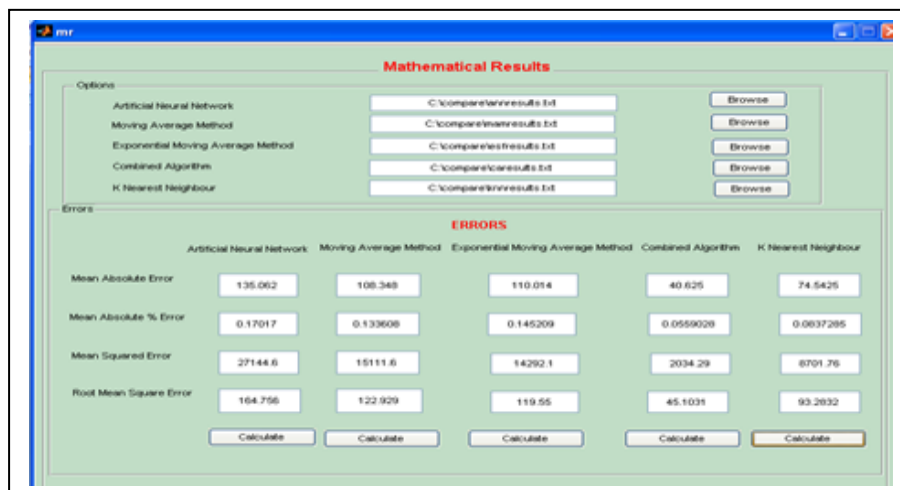


Figure 2: Various type of errors for all the forecasting models

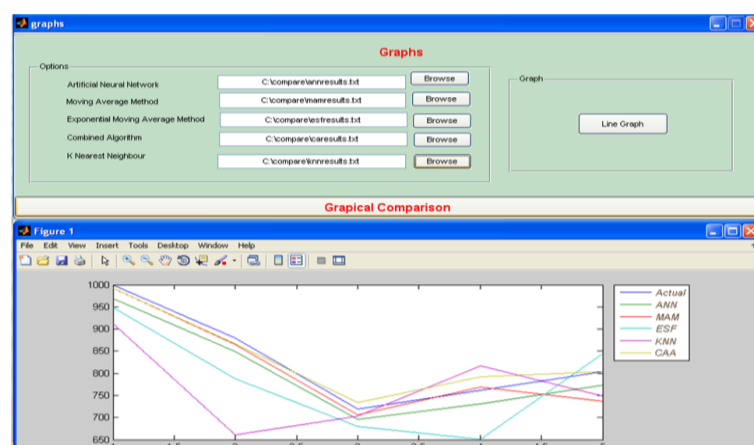


Figure 3: Line graph among the actual and forecasting value by using all the models