International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 03 | March - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

A Machine Learning Model for Analyzing E-Commerce Supply Chain for Sales Volume Forecast

Esha[1], Prof. Manisha Kadam[2]
Department of CSE, SDBCT, Indore[1,2]

ABSTRACT
Supply chain forecasting has seen a major inclination toward machine learning and evolutionary algorithms due to their capability to analyze large data sets. Sales forecasting is typically complex due to the volatility of the associated variables. This paper presents a machine learning based approach based on the principal component analysis (PCA) and the Scaled Conjugate Gradient (SCG) algorithm to for sales forecasting. The performance of the proposed system is evaluated in terms of the mean absolute percentage error (MAPE) and the regression. It is shown that the proposed system outperforms the previously existing system working on the benchmark datasets [1].

KEYWORDS
Supply chain forecasting, Regression Learning, Principal Component Analysis (PCA), Steepest Descent, Scaled Conjugate Gradient (SCG), Mean Absolute Percentage Error, Regression.

I. Introduction

Global Markets have encountered a lot of volatility in the last decade due to the following reasons:

1) Trade Wars
2) Formation of cartels
3) Outbreaks of new diseases such as Ebola and Covid-19.
4) Global economic slowdown etc.

Some specific products and services account for a large portion for the Gross Domestic Product (GDP) of countries. There is a large diversity in such products and services. For example, Switzerland and Denmark rely heavily in dairy based products, while China relies heavily on Electronic and industrial manufacturing [1]-[2]. Generally, the major export depends on the geographic conditions, manpower and natural resources of the country. Supply chain management plays a pivotal role for such industries in streamlining the processes and deciding the profits. Supply chain management can be defined as the management of the flow of goods and services including all processes which are intertwined with the transformation of raw materials into final products. The different domains affected by supply chain management are [10]:

1) Inventory Management
2) Warehousing and distribution
3) Logistics
4) Procurement
5) Revenue Management
6) Order Management
7) Revenue Generation and Management
8) Information Technology (IT) related to businesses.

II. Supply Chain Forecasting:

Supply chain forecasting can be defined as the prediction of demand metrics based on the previous demands and associated variables. It's a critical domain of supply chain management. Thus apart from increasing profits and streamlining businesses, supply chain management and supply chain forecasting can also have deep impacts on employment, food security, peace and political conditions in a country [5]-[6].

Supply chain forecasting tries to find patterns in previously existing data and forecast future demands [7]. The sales or demands are generally modelled as a function of time and associated variables given by:

$$\boldsymbol{Demand} = \boldsymbol{function}\ (\boldsymbol{time}, \boldsymbol{associated\ variables})$$

The associated variables can be:

1) Current Economic Situation of the Country of export
2) Global Economic condition
3) Political Relations among countries
4) Supply from other countries etc.

Accurate predictions or estimates need to be made considering a pervasive set of associated parameters.

III. Machine Learning for Supply Chain Forecasting

Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. Other advantages include:

1. **Adaptive learning:** An ability to learn how to do tasks based on the data given for training or initial experience.
2. **Self-Organization:** An ANN can create its own organization or representation of the information it receives during learning time.
3. **Real Time Operation:** ANN computations may be carried out in parallel, and special hardware devices are being designed and manufactured which take advantage of this capability.

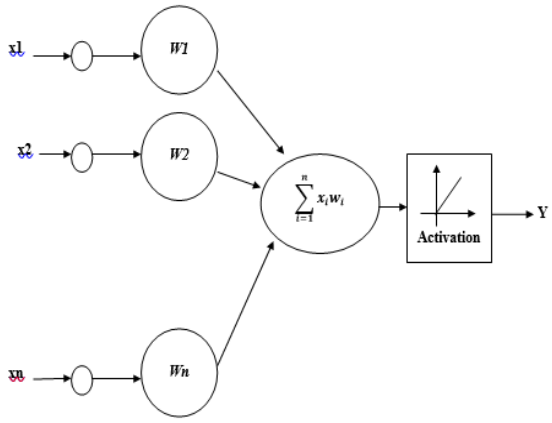

Fig.1 Mathematical Model of Neural Network

The output of the neural network is given by:

$$\sum_{i=1}^{n} XiWi + \Theta \quad (1)$$

Where,
Xi represents the signals arriving through various paths,
Wi represents the weight corresponding to the various paths and
Θ is the bias. It can be seen that various signals traverssing different paths have been assigned names X and each path has been assigned a weight W. The signal traverssing a particular path gets multiplied by a corresponding weight W and finally the overall summation of the signals multiplied by the corresponding path weights reaches the neuron which reacts to it according to the bias Θ. Finally its the bias that decides the activation function that is responsiblefor the decision taken upon by the neuralnetwork. The activation function φ is used to decide upona the final output. The learning capability of the ANN structure is based on the temporal learning capability governed by the relation:

$$w(i) = f(i, e) \quad (2)$$

Here,
w (i) represents the instantaneous weights
i is the iteration
e is the prediction error

The weight changes dynamically and is given by:

$$W_k \overset{e,i}{\rightarrow} W_{k+1} \quad (3)$$

Here,
W_k is the weight of the current iteration.
W_{k+1} is the weight of the subsequent iteration.

(i) Regression Learning Model

Regression learning has found several applications in supervized learning algorithms where the regression analysis among dependednt and independent variables is eeded [31]. Different regression models differ based on the the kind of relationship between dependent and independent variables, they are considering and the number of independent variables being used. Regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a relationship between x (input) and y(output). Mathematically,

$$y = \theta_1 + \theta_2 x \quad (4)$$

Here,
x represenst the state vector of inut variables
y rperesenst the state vector of output variable or variables.
Θ1 and Θ2 are the co-efficents which try to fit the regression learning models output vector to the input vector.

By achieving the best-fit regression line, the model aims to predict y value such that the error difference between predicted value and true value is minimum. So, it is very important to update the θ1 and θ2 values, to reach the best value that minimize the error between predicted y value (pred) and true y value (y). The cost function J is mathematically defined as:

$$J = \frac{1}{n}\sum_{i=1}^{n}(pred_1 - y_i)^2 \quad (5)$$

Here,
n is the number of samples
y is the target

pred is the actual output.

(ii) Gradient Descent in Regression Learning

To update θ1 and θ2 values in order to reduce Cost function (minimizing MSE value) and achieving the best fit line the model uses Gradient Descent. The idea is to start with random θ1 and θ2 values and then iteratively updating the values, reaching minimum cost. The main aim is to minimize the cost function J [31].

IV. The Steepest Descent Algorithm and Principal Component Analysis

The critical aspect about steepest descent is the fact that it repeatedly feeds the errors in every iteration to the network till the errors become constant or the maximum number of allowable iterations are over. This can be mathematically given by:

$if\ PF \neq constant$
$for\ (k = 1, k \leq k_{max} = constant, k = k + 1)$
{
$W_{k+1} = f(X_k, W_k, e_k)$
}
else
{
$W_{k+1} = W_k$ && $training\ stops$
}

Here,
X_k is the input to the kth iteration
W_k is the weight to the kth iteration
W_{k+1} is the weight to the (k+1)st iteration
e_k is the error to the kth iteration
k is the iteration number
PF is the performance function deciding the end of training
k_{max} is the maximum number of iterations

Thus if the error is within tolerance, which is generally not feasible to find beforehand in time series data, the training is stopped if the performance function (which can be the training error) becomes constant for multiple iterations or the maximum number of iterations are over. Now there are various ways in which the error can be minimized. However, the steepest fall of the error with respect to weights is envisaged. It is depicted in the figure below: It can be seen from figure 1 that although the error in training keeps plummeting in all the three cases of gradient descent, the gradient 3 or g3 attains the maximum negative descent resulting in the quickest training among all the approaches and hence the least time complexity. This would be inferred from the number of iterations which are required to stop training. Thus the number of iterations would be a function of the gradient with which the error falls.

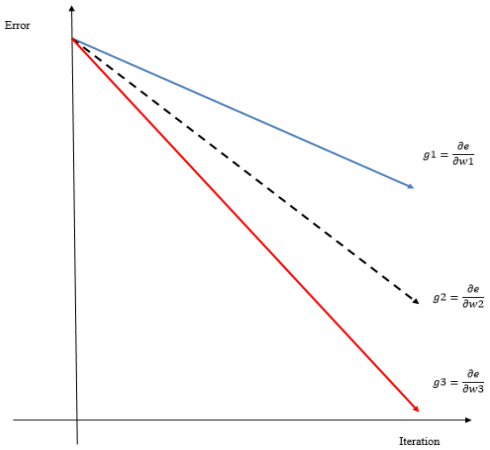

Figure 1. The concept of Steepest Descent

This is mathematically given by:

$$\boldsymbol{k_n = f(g = \frac{\partial e}{\partial w})} \quad \textbf{(6)}$$

Here,
k_n is the number of iterations to stop training.
g is the gradient
w is the weight
e is the error
f stands for a function of

The proposed methodology uses two key components one of which is the training algorithm and the other is the training optimization algorithm. Both are explained in this section

(a) The Scaled Conjugate Gradient (SCG) Algorithm

There are several ways to implement the back propagation technique in the neural networks. One consideration however always remains that of the least time and space complexity so as to reduce the amount of computational cost that is associated with the training algorithm. The essence of the scaled conjugate gradient algorithm is the fact that it has very low space and time complexity making it ideally suited to large data sets to be analyzed in real time applications where the time is a constraint. The training rule for the algorithm is given by:

$$A_0 = -g_0 \quad (7)$$

A is the initial search vector for steepest gradient search
g is the actual gradient

$$w_{k+1} = w_k + \mu_k g_k \quad (8)$$

Here,
w_{k+1} is the weight of the next iteration
w_k is the weight of the present iteration
μ_k is the combination co-efficient

(b) The Principal Component Analysis (PCA)

The principal component analysis (PCA) is basically a dimensional reduction tool which helps to clear out the redundancies in the training data vector in such a way that the training is optimized for lesser number of variables and mean absolute percentage error hits the least values in the lest number of iterations possible. The parameters in the estimation are the various temporal parameters affecting demand are:
Unit Cost, Time, and previous Unit Sales.

The training is stopped based on the mean square error or mse given by:

$$mse = \frac{\sum_{i=1}^{n} e_i^2}{n} \quad (9)$$

The final computation of the performance metric is the mean absolute percentage error given by:

$$MAPE = \frac{100}{M} \sum_{i=1}^{N} \frac{E - E_i|}{i} \quad (10)$$

Here,
n is the number of errors
i is the iteration number
E is the actual value
E_i is the predicted value

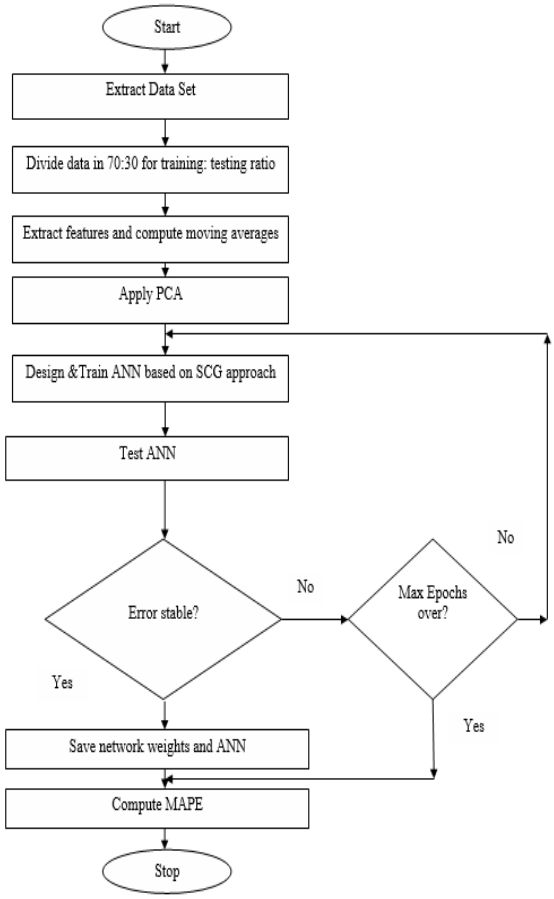

Figure 2. Flowchart of Proposed System

V. RESULTS AND DISCUSSIONS

The data set used in the proposed work is the global sales, as used in [1] available at:
https://data.world/datasets/sales
The training algorithm used is the steepest descent based scaled conjugate gradient algorithm. The accuracy of temporal prediction is computed as:

$$Ac = 100 - \frac{100}{M} \sum_{i=1}^{N} \frac{E - E_i|}{i} \% \quad (11)$$

Here,
Ac is the accuracy computed in %

Basically a 10 hidden layer deep neural network is design based on the scaled conjugate gradient approach. The need for a deep neural network is seen as a necessity since the cloud data in terms of workload is substantially complex and exhaustive.

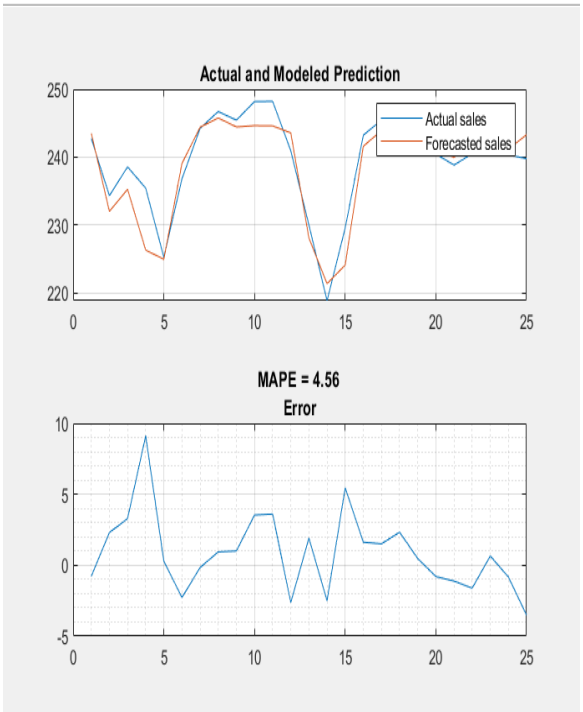

Figure 3. Actual and Modelled values

The figure above depicts the actual and the modelled prediction values in which the red curve is the values corresponding to forecasted values and the blue curve corresponds to the actual values. A difference between them yields the error.

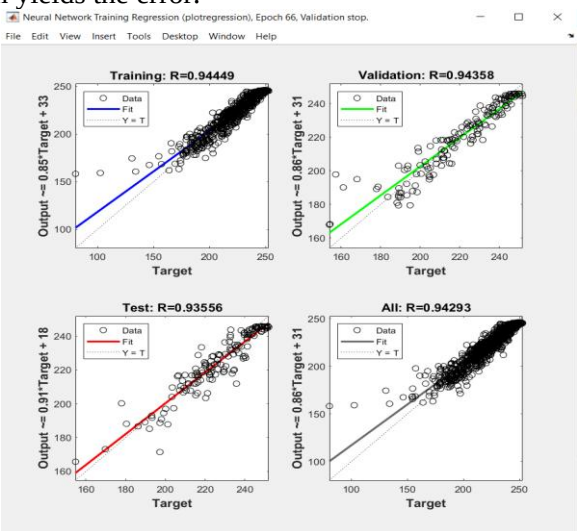

Figure 4. Regression

The figure above depicts the regression obtained in the proposed approach which is a sort of similarity among two random variables. The maximum allowable regression is unity depicting complete similarity.

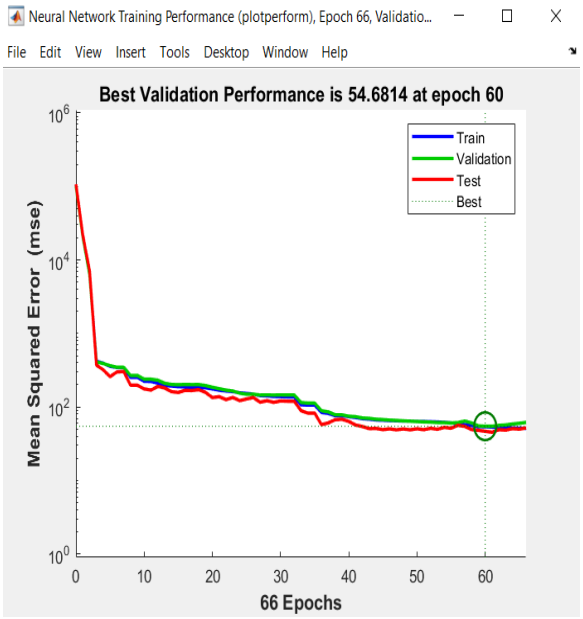

Figure 5. Performance Function

The performance function that decides the culmination of training is the mean squared error in this case given mathematically by equation 5.

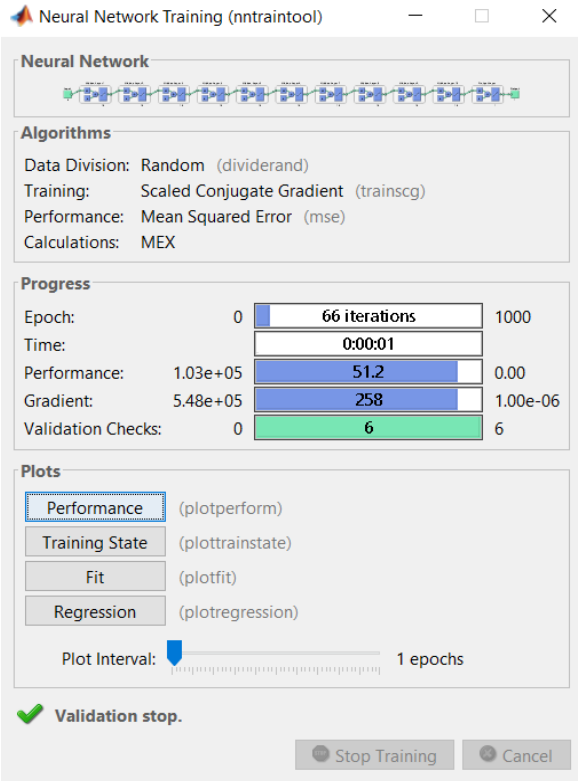

Figure 6. Details of Training

The details of the training are depicted in the figure above, which clearly shows the designed neural network, the training function, the data division and the iterations.

Table.1 Obtained Results

S.No	PARAMETER	VALUE
1.	Algorithm	SCG
2.	Architecture	Back Propagation
3.	Iterations	32
4.	MAPE (proposed work)	4.56%
5.	MAPE (previous work)	10.9%
6.	Accuracy (Proposed Work)	95.44%
7.	Accuracy (Previous Work)	89.1%
8.	Regression	0.9418

The performance of the proposed approach is found better compared to previously existing technique [1] which attains a MAPE of 10.9%.

VI. CONCLUSION

This paper presents a mechanism for supply chain forecasting. The neural network architecture is used to implement machine learning. The architecture of the approach is the use of the back propagation based approaches to utilize the knowledge about the errors in each iteration to affect the weights of each subsequent iteration. The scaled conjugate gradient (SCG) based steepest descent training rule is utilized in this approach to reduce the number of iterations and also the mean absolute percentage error. The principal component analysis (PCA) is used as a data optimization tool. It is shown that the proposed work performs better in terms of mean absolute percentage error (MAPE) compared to the previously existing technique for the same standard dataset. It can be attributed to the back propagation approach as well as the use of the principal component analysis to fine tune and optimize the training process. The proposed approach attains an accuracy of 95.22% in terms of the accuracy and a MAPE of 4.78%. It is substantially improved compared to previous work [1].

References:

[1] Prompatanapak A, Lopetcharat K, (2022) "Managing changes and risk in seafood supply chain: A case study from Thailand", Journal of Aquaculture, Elsevier, Volume-252, pp: 2-7.
[2] Zhang Y., (2021), "Approach Application of improved BP neural network based on e-commerce supply chain network data in the forecast of aquatic product export volume", Journal of Cognitive Systems Research, Elsevier, Volume 57, pp: 228-235.
[3] CH Wang, JY Chen (2020) "Demand forecasting and financial estimation considering the interactive dynamics of semiconductor supply-chain companies", Computers & Industrial Engineering, Elsevier, Volume-138, pp.106104.
[4] R Sharma, P Singhal (2019), "Demand forecasting of engine oil for automotive and industrial lubricant manufacturing company using neural network", Materials Today, Elsevier, Volume 18, Issue 7, pp. 2308-2314.
[5] Rivera-Castro R., Nazarov I., Xiang Y., Pletneev A., Maksimov I., Burnaev E. (2019) Demand Forecasting Techniques for Build-to-Order Lean Manufacturing

Supply Chains. In: Lu H., Tang H., Wang Z. (eds) Advances in Neural Networks – ISNN 2019. ISNN 2019. Lecture Notes in Computer Science, Volume- 11554, pp 213-222.

[6] Singhry H.B., Rahmen A. (2019) "Enhancing supply chain performance through collaborative planning, forecasting, and replenishment", Business Process Management Journal, Emerald Insight, Volume-25, Isse-4, pp:1-23

[7] Bowersox DJ, Closs D, Cooper MB, Bowersox JC, (2018), "Supply Chain Logistics Management", 4th Edition, McGraw Hill Publications.

[8] Mishra, D., Gunasekaran, A., Papadopoulos, T. (2018) Big Data and supply chain management: a review and bibliometric analysis. Ann Oper Res Volume-270, pp.313–336.

[9] Hazen, B.T., Skipper, J.B., Boone, C.A. (2018). Back in business: operations research in support of big data analytics for operations and supply chain management. Ann Oper Res Volume-270, pp.201–211.

[10] Rajesh R. (2016) "Forecasting supply chain resilience performance using grey prediction", Journal of Electronic Commerce Research and Applications, Elsevier, Volume-20, pp: 42-58

[11] Chopra S, Meindl P, Kalra DV, 2016 "Supply Chain Management: Strategy, Planning and Operation", 6th Edition, Pearson Publications.

[12] Jaipuriya S, Mahapatra SS (2015), "Forecasting supply chain demand by clustering customers", Elsevier Volume-48, Issue-3, 2015, pp 1834-1839

[13] Schuh, G., Stich, V., Brosze, T. (2011). High resolution supply chain management: optimized processes based on self-optimizing control loops and real time data. Prod. Eng. Res. Devel, Volume- 5, pp. 433–442.

[14] France RA, Jones EC, Richards CN (2010), "Multi-objective stochastic supply chain modeling to evaluate tradeoffs between profit and quality", International Journal of Production Economics, Elsevier, Volume -127, Issue-2, pp: 292-999

[15] S Thomassey (2010) "Sales forecasts in clothing industry: The key success factor of the supply chain management" International Journal of Production Economics, Elsevier, Volume-128, Issue 2, pp: 470-483

[16] Bojarski A. D., Laínez J. M., Espuña A., Puigjaner L., (2009), "Incorporating environmental impacts and regulations in a holistic supply chains modeling: An LCA approach" Journal of Computers and Chemical Engineering, Elsevier, Volume -33, Issue-10, pp: 1747-1759.

[17] Barbaro, A.F., Bagajewics, M.J., 2003. Financial Risk Management in planning under
uncertainty. In: Grossmann, I.E., McDonald, *C.M. (Eds.), Proceedings of Fourth International Conference on Foundations of Computer Aided Process Operations (FOCAPO).* Coral Springs, FL, ISBN-0965589110

[18] Szidarovszky, F., Gershon, E.M., Lucien, D.,1986. Techniques for Multi-objective Decision Making in Systems Management. *1st ed. vol. 2. Elsevier, West Lafayette, UN*

[19] Hillermeier, C., 2001. Generalized homotopy approach to multi-objective optimization. *Journal of Optimization Theory and Applications* 110, 557–583

[20] Beale, E.M.L., 1955. On minimizing a convex function subject to linear inequalities.
Journal of Royal Statistical Society, Series B 17, 173–184.