Data Mining using Statistical Analysis and Machine Learning Algorithms for Supply Chain Management

Aaditya Desai[1], Shridhar Kamble[2]
[1,2]Department of Information Technology, Thakur College of Engineering, Kandivali (E), Mumbai-400 101, India
E-mail: desaiaaditya@hotmail.com

Abstract— Supply Chain Management (SCM) plays a very vital role in managing and organizing enterprise processes, increasing operational efficiency of the organization. Factors such as product success, customer satisfaction, organization's growth depends upon successful execution of Supply Chain Management (SCM). Supply chain management is becoming a necessity to improve the foundation and infrastructure within societies which in turn increases the economic growth, living standard of society also. The research findings indicate that although it seems that SCM provides many services, but it has some issues too; which includes poor inventory management, bullwhip effect, High cost of logistics, technology usage and inadequate investments in IT. To overcome issues of SCM there is a need of an improved sales forecasting model which will build the reliable and efficient forecasting results. An improved Sales forecasting model is presented in this paper, which is based on kernel based support vector machine regression.

Keywords- SCM, MAPE, MAD, MSE

I. INTRODUCTION

Supply chain encompasses all of the facilities, functions and activities involved in producing a product or service from suppliers to customers. Supply chain functions include purchasing, inventory, production, scheduling, facility location, transportation and distribution [01]. All these functions are affected in the short run by product demand and in the long run by products and processes and changing markets. Forecast of product demand determine how much product to make and how much material to purchase from suppliers to meet forecasted customers' needs. Due to the ever-increasing global competition, sales forecasting plays a prominent role in supply chain management. The availability of huge data demand tools that can extract information from data. Recent research has shown that advanced forecasting tools enable improvements in supply chain performance [02].

A forecasting model which gives accurate results and upon which one can rely, need to be built for improving financial growth of organizations and better decision making. The main objective of this research work is to identify the issues related to supply chain management and how these problems were overcome by using data mining and data warehousing. The issues were identified through a, questionnaires, interviews with apparel manufacturing company with the added support of literature review. Data related to supply chain process were collected and examined with the help of SPSS Statistics 21.

II. FORECASTING

Supply chain management relies on forecasts of future demand for decision making.

Demand forecasts are used in supply chain design, planning as well as in operations. For example, in the production phase of SCM, it is used for aggregate planning, inventory control and scheduling, in marketing it is used for new product introductions, promotions, and sales-force allocation, in finance it is used for plant and equipment investment decisions, operating budgeting and finally in personnel it is used for workforce planning and resulting hiring and layoff [03].

It might be possible that forecasts may always go wrong; hence the forecasted result should include both the expected value and a measure of forecast error. Short-term and aggregate forecasts are more accurate. Forecasting methods are listed as below;

- **Qualitative:** It is based on judgments, expertise, and opinion.
- **Quantitative methods:** These types of forecasting methods are based on mathematical models and rely heavily on mathematical computations.
- **Time Series**: Historical data is used to do forecasting.
- **Causal:** It is based on the relationship between the variable to be forecasted and an independent variable.

III. PROPOSED FORECASTING METHODOLOGY

A good forecasting system has qualities that distinguish it from other systems. The qualities provide a useful basis for understanding why good forecasting systems outperform others over time. If users understand the rationale of a forecast, they can appraise the uncertainty of the forecast, and they will know when to revise the forecast in light of changing circumstances [6] [7]. The more accurate a forecast is, the better are the decisions that depend upon it. Inaccurate forecasts lead to too much or too little capacity and can be very costly. Forecasts cost money, time and effort. Added expense must purchase added accuracy, flexibility, or insight. A sophisticated forecasting system requires ample staff resources and technical skills for maintenance. The choice of a forecasting system must include a commitment to the resources necessary to maintain it and avoid systems the utility will be unable or unwilling to maintain. Earlier statistical methods were used for forecasting. A time series model was constructed solely from the past values of the variables to be forecast. Figure 1 shows the detailed steps of predictive model used to construct the forecasting model based on data mining.

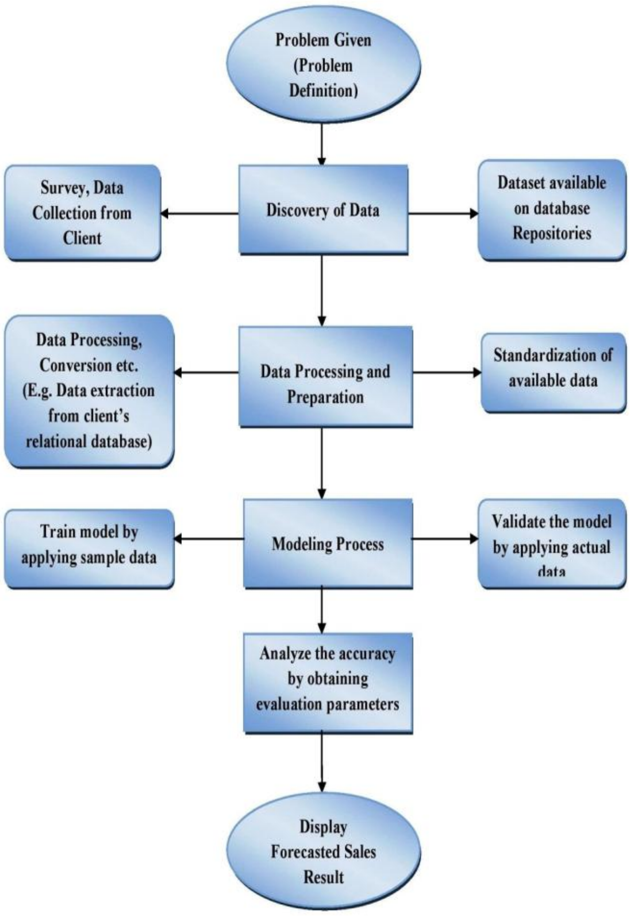

Figure 1: Data Mining based Predictive Modeling

The steps of forecasting methodology are as follows;

- Processing of the problem definition is done and required information related to the problem defined.
- The next step is, Data Discovery. In this step, data are collected from various sources for e.g., Survey data, Data collected from clients and Dataset available in dataset repositories.
- Collected data is pre-processed to remove possible outliers and most relevant features are selected. Data conversion in required format, Data standardization may be a part of this step.
- A Further step is modeling of forecasting system. Training and validation are the sub tasks of this step.
- Forecasting model is build and trained. In this step training of data is done after that, forecasting model is validated.
- Accuracy of forecasting model is analyzed by obtaining evaluation parameters such as MAPE, MAD and MSE.
- Final forecasted results of sales value are displayed.

IV. PROPOSED SYSTEM

Processing of K-SVM Algorithm is depicted in figure 2. It includes many steps such as processing input data, selection of kernel functions, training of forecasting model, validating the forecasting model and finally accuracy calculation.

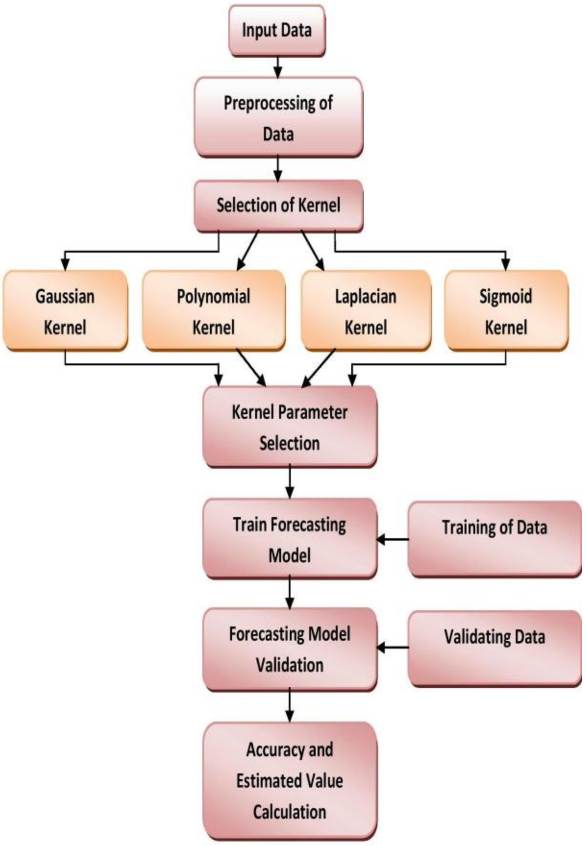

Figure 2: Processing of K-SVM Algorithm

Steps for processing, training and testing of K-SVM Algorithm are as follows:

- Preprocessing of the input data is done and most relevant features are selected, scaling is done in the range [-1, 1] and check for possible outliers.
- The kernel and kernel/regularization parameters can be selected by optimizing a cross-validation based model selection or radius-margin can be used.
- Selection of the appropriate kernel function determines the hypothesis space of the decision and regression function.
- Build and train the forecasting model by selecting the parameters of the kernel function. In this step training of data is done.
- Forecasting model is validated by training unseen test data.
- Desired accuracy is chosen by defining the ‘e-insensitive loss’ and Penalty factor by ‘C’.
- Finally, the accuracy of forecasting model is calculated by accuracy measures such as MAPE, MAD and MSE.

The methodology used for the design, training and testing of Kernel based Support Vector Machine algorithm (K-SVM) algorithm is shown in figure 2.

V. FORECAST EVALUATION

Evaluation of Forecasting techniques can be done on the basis of evaluation parameters such as MAPE, MAD and MSE, etc. A model with minimum MAPE value is considered to be best prediction of performance [07]. Forecast error is the difference between the actual value and the forecast value for the corresponding period [09].

DATA PREPARATION

A detailed survey is carried out in "Creative Handicrafts", Andheri, Mumbai [10] and data collection has been done related to actual sales of manufactured apparels in the company monthly, quarterly and annual wise. There are three categories of apparels are manufactured namely Menswear, Womenware and Kids ware.

The proposed forecasting model has been implemented using C# programming language and .net platform in the visual studio IDE. Machine with windows 7 operating system and 4 GB of RAM is used to conduct the experiments and analyze the results. To ensure that the results are comparable, all experiments were rerun.

Interviews, questionnaires are carried out, in order to find out the need of forecasting in the supply chain system by correlating various questions with each other. It has been observed that efficient forecasting tool is required for better decision making, manage the inventory levels and manage the organization processes.

VI. RESULTS

Actual sales data which is collected from supply chain industry has been tested with proposed methodology. The number of product units that will be sold by monthly, quarterly and yearly basis, depending upon the product and category need to be estimated. Time series data, representing monthly sales data, during the period which runs from June 2009

to June 2014 has been analyzed. Each forecasting model has been

$Et \square Yt \square Ft$

Where, **E** is the forecast error at period **t**, **Y** is the actual value at Period **t** and **F** is the forecast for period **t**.

MAPE stands for 'Mean Absolute Percentage Error' and it is formulated as;

$$\sum_{t=1}^{N} \left| \frac{Et}{Yt} \right|$$

$MAPE \square$

N

(1)

(2) evaluated for identical data and same set of parameters.

Ijfans International Journal of Food and Nutritional Sciences
ISSN PRINT 2319 1775 Online 2320 7876
 Journal UGC CARE Listed (Group-I) Volume 11,Sp Issue 11 A, 2022

The tables I, II and III shows the evaluation result of accuracy error measures such as mean absolute percentage error (MAPE), mean absolute deviation (MAD) and mean squared error (MSE), obtained over the test data and predicting sales of six different products depending upon the category. The best result for each product is highlighted and underlined. For demonstration purpose, values of MAPE, MAD and MSE are calculated for 'menswear' category.

The notations used for table I, II and III are as follows;

MAD stands for 'Mean Absolute Deviation' and it is formulated as;

$$MAD = \frac{\sum_{t=1}^{N}|Et|}{N} \qquad (3)$$

MSE stands for 'Mean Squared Error' and it is formulated as;

P : Product

SA : Simple Moving Average WA : Weighted Moving Average ES : Exponential Smoothing AR : Adaptive Rate Smoothing NB : Naïve Bayes

ANN : Artificial Neural Network SVM : Support Vector Machine K-SVM : Kernel based SVM

$$MSE = \frac{\sum_{t=1}^{N}|Et^2|}{N} \qquad (4)$$

TABLE I. MEAN ABSOLUTE PERCENTAGE ERROR (MENSWEAR)

P	SA	WA	ES	AR	NB	ANN	SVM	K-SVM
P1	0.092	0.084	0.097	0.099	0.079	0.099	**0.074**	0.092
P2	0.086	0.088	0.074	0.096	**0.065**	0.075	0.079	0.069
P3	0.099	0.098	0.096	0.092	0.099	0.098	0.096	**0.088**
P4	0.092	0.095	0.088	0.093	0.091	0.090	0.089	**0.083**
P5	0.091	0.088	0.083	0.097	0.087	**0.072**	0.087	0.074
P6	0.084	0.082	0.089	0.098	0.095	0.099	0.089	**0.082**
Avg.	0.091	0.089	0.088	0.096	0.086	0.089	0.086	**0.081**

IJFANS International Journal of Food And Nutritional Sciences
Official Publication of International Association of Food and Nutrition Scientists

Ijfans International Journal of Food and Nutritional Sciences
ISSN PRINT 2319 1775 Online 2320 7876
, Journal UGC CARE Listed (Group-I) Volume 11,Sp Issue 11 A, 2022

In Table I, the observed value of MAPE for K-SVM is lower than the other techniques except the few cases.

TABLE II. MEAN ABSOLUTE DEVIATION (MENSWEAR)

P	SA	WA	ES	AR	NB	ANN	SVM	K-SVM
P1	111.86	127.37	133.12	130.79	110.67	99.57	79.49	**76.39**
P2	121.1	127.63	127.85	129.41	105.38	88.22	77.1	**71.43**
P3	113.72	126.88	126.57	122.55	108.82	**82.62**	89.83	86.66
P4	97.59	95.72	88.98	109.14	101.78	87.19	82.72	**78.38**
P5	102.17	109.94	109.78	118.37	90.52	85.44	80.96	**78.72**
P6	99.26	105.89	100.47	91.78	88.43	87.39	**83.63**	85.74
Avg.	107.61	115.57	114.46	117.00	100.93	88.40	82.28	**79.55**

Somewhat similar behaviour is observed for MAD value. As we see in the Table II, except the few cases the observed value of MAD for K-SVM is lower than the other techniques.

TABLE III. MEAN SQUARED ERROR (MENSWEAR)

P	SA	WA	ES	AR	NB	ANN	SVM	K-SVM
P1	176.75	293.91	259.65	189.11	187.21	159.56	153.41	**146.72**
P2	194.65	273.45	244.45	227.41	170.21	127.68	118.68	**104.91**
P3	192.98	278.09	248.08	218.41	185.97	151.17	**139.93**	142.36
P4	189.73	212.55	255.76	226.41	**138.21**	149.36	148.58	140.83
P5	190.8	233.9	288.0	199.8	187.2	**162.6**	168.7	164.7

	5	1	9	8	1	**5**	5	2
P6	189.11	285.13	268.64	238.41	189.21	142.91	137.75	**125.48**
Avg.	189.01	262.84	260.78	216.61	176.34	148.89	144.52	**137.50**

In Table III, the observed value of MSE for K-SVM is lower than the other Classical technique and ANN except the few cases. The results obtained are contradictory for some specific products, but the average error levels confirm that the proposed K-SVM methodology consistently outperforms ANN and Classical results, for all the three error measures of MAPE, MAD and MSE. This confirms the ability of the proposed methodology to provide accurate forecasts.

VII. CONCLUSIONS

In the research work, we have observed that K-SVM outperforms the SVM, ANN and Naïve model in various accuracy measures such as MAPE, MAD and MSE. K-SVM require less time to optimally train and gives more consistent and accurate results as compared to ANN and SVM due to optimization of threshold functions. ANN gives improved performance if numbers of iterations are increased with less prediction errors.

In data mining, analysis and preprocessing of input data plays an important role prior to algorithm selection. Proper understanding and interpretation of the problem is required before choosing a technique to predict data, as different forecasting model fits for different conditions. Through data cleansing and preprocessing time series data, pruning decision spaces, improved coding standards, performance of forecasting model can be improved.

It has been observed that every forecasting model has its own significance, one may fit for particular problem but not for all. The proper forecasting model which suits the requirements, needs to be selected and comprehensive understanding of the problem domain is required.

SCOPE FOR THE FUTURE WORK

For future work, the proposed system can be enhanced by optimal design for multiclass SVM classifiers to improve the accuracy and quality of the prediction. It may also possible to enhance the system to keep a track and monitor the current trends and changes in system flow.

REFERENCES

1. Roberta S. Russell, Bernard W. Taylor, "Operations Management", Chapter No. 10, 3rd edition, Pearson publication.
2. Shoumen D. , Clive G., Don G., Nikhil S., Pat D., Reuben S., "Forecasting and Risk Analysis in Supply Chain Management", Managing Supply Chain Risk and Vulnerability, pp 187-203, 2009.
3. Sunil Chopra and Peter Meindl, Supply Chain Management: Strategy, Planning and Operations, Prentice Hall, 2001.

4. Nesreen K. Ahmed, Amir F. Atiya, Neamat El Gayar,and Hisham El-Shishiny, “An Empirical Comparison of Machine Learning Models for Time Series Forecasting”, Journal of Econometric Reviews 29 (5-6), pp.
5. Gianluca Bontempi, Souhaib Ben Taieb, Yann-Aël Le Borgne ,“Machine Learning Strategies for Time Series Forecasting”, Lecture Notes in Business Information Processing Volume 138, 2013, pp. 62-77.
6. Na Liu, Shuyun Ren, Tsan-Ming Choi, Chi-Leung Hui, and Sau-Fun Ng, “Sales Forecasting for Fashion Retailing Service Industry: A Review”, Hindawi Publishing Corporation Mathematical Problems in Engineering, Volume 2013, Article ID 738675.
7. Karin Kandananond, “A Comparison of Various Forecasting Methods for Auto correlated Time Series”, International journal of Engineering and business management, Vol. 4, 2012.
8. Min Li, W.K. Wong, S.Y.S Leung, “Comparison study on univariate forecasting techniques for apparel sales”, Institute of Textiles and Clothing, The Hong Kong Polytechnic University, Hunghom, Kowloon, Hong Kong, China.
9. Jos´e Guajardo, Richard Weber, and Jaime Miranda, “A Forecasting Methodology Using Support Vector Regression and Dynamic Feature Selection”, Journal of Information & Knowledge Management, Volume 05, Issue 04, December 2006.
10. Creative Handicrafts, http://www.creativehandicrafts.org/
11. Tony Van Gestel, Johan A. K. Suykens et al.,“Financial Time Series Prediction Using Least Squares Support Vector Machines Within the Evidence Framework”, IEEE Transactions on Neural Networks, VOL. 12, NO. 4, July 2001.
12. Mouhammd Alkasassbeh,“Predicting of Surface Ozone Using Artificial Neural Networks and Support Vector Machines”, International Journal of Advanced Science and Technology Vol. 55, June, 2013.
13. Debasish Basak, Srimanta Pal, Dipak Chandra Patranabis, “Support Vector Regression”, Neural Information Processing-Letters and Reviews, Vol. 11, No. 10, October 2007.
14. Sansom, D. C., Downs, T., & Saha, T. K., “Evaluation of support vector machine based forecasting tool in electricity price forecasting for Australian national electricity market participants”, Journal of Electrical and Electronics Engineering, Australia, 22(3), pp.227-234, 2003.
15. Neagu C.D., Guo G., Trundle P.R.,Cronin M.T.D, "A comparative study of machine learning algorithms applied to predictive toxicology data mining", Alternatives to laboratory animals : ATLA 35:1, pp. 25-32, 2007.
16. Karpagavalli, Jamuna K and Vijaya MS, "Machine Learning Approach for Preoperative Anaesthetic Risk Prediction", International Journal of Recent Trends in Engineering, Volume No. 1, No. 2, 2009.
17. Sanjeev Kumar Aggarwal, Lalit Mohan Saini, Ashwani Kumar, "Electricity price forecasting in deregulated markets: A review and evaluation, International journal of

production economics, Volume 31, Issue 1, pp. 13-22, Elsevier B.V. 2009.

18. Cem Kadilar, Muammer Simsek and Cagdas Hakan Aladag, "Forecasting the exchange rate series with ann: the case of turkey", Istanbul University Econometrics and Statistics e- Journal, pp. 17-29, 2009.
19. John B. Guerard, Jr., Introduction to Financial Forecasting in Investment Analysis: Springer New York, Online ISBN: 978-1-4614-5239-3, 2013.
20. Hamid R. S. Mojaveri, Seyed S. Mousavi, Mojtaba Heydar, and Ahmad Aminian, "Validation and Selection between Machine Learning Technique and Traditional Methods to Reduce Bullwhip Effects: a Data Mining Approach", World Academy of Science, Engineering and Technology, Volume 25, 2009.
21. Yang LanQin ; Xu Xin, "Research on the Price Prediction in Supply Chain based on Data Mining Technology", Published in International Symposium on Instrumentation & Measurement, Sensor Network and Automation, Vol. 2 , IEEE 2012.
22. John Geweke and Charles Whiteman, Bayesian Forecasting," Bayesian Forecasting", Chapter 1 in Handprojectof Economic Forecasting, vol. 1,
23. pp. 3-80, Elsevier 2006.
24. Réal Carbonneau, Rustam Vahidov, Kevin Laframboise, "Forecasting Supply Chain Demand Using Machine Learning Algorithms", Chapter 6.9.
25. Trigg and Leach, "Exponential Smoothing with an Adaptive Response Rate", Operational Research, Vol. 18, No. 1, pp. 53-59, Mar. 1967.
26. V. Vapnik, S. Golowich, and A. Smola, "Support vector method for function approximation, regression estimation, and signal processing," Neural Information Processing Systems, vol. 9, MIT Press, Cambridge, MA. 1997.
27. Joarder Kamruzzaman and Ruhul A. Sarker, "ANN-Based Forecasting of Foreign Currency Exchange Rates", Neural Information Processing - Letters and Reviews, Vol. 3, No. 2, May 2004.
28. J. Shahrabi, S. S. Mousavi and M. Heydar, "Supply Chain Demand Forecasting- A Comparison of Machine Learning Techniques and Traditional Methods", Journal of Applied Sciences, Volume 9, Issue 3, pp.521-527, 2009.
29. Karin Kandananond, "Consumer Product Demand Forecasting based on Artificial Neural Network and Support Vector Machine", World Academy of Science, Engineering and Technology, 2012.
30. Sapankevych I. and Sankar R., "Time series prediction using support vector machines: A survey," Computational Intelligence Magazine, IEEE, vol. 4, no. 2, pp. 24-38, 2009.
31. Sapankevych I. and Sankar R., "Time series prediction using support vector machines: A survey," Computational Intelligence Magazine, IEEE, vol. 4, no. 2, pp. 24-38, 2009.
32. Learning methods building on kernels, http://www.kernel-machines.org/

33. http://www.saedsayad.com/support_vector_machine_reg.htm.
34. MaRS Library, "Sales forecasting for startups: Cash flow, planning & revenue projection", http://www.marsdd.com/mars-library.
35. K. Au, T. Choi, and Y. Yu, "Fashion retail forecasting by evolutionary neural networks," International Journal of Production Economics, vol. 114,
36. pp. 615- 630, 2008.
37. Z. Sun, T. Choi, K. Au, and Y. Yu, "Sales forecasting using extreme learning machine with applications in fashion retailing," Decision Support Systems, vol. 46, pp. 411-419, 2008.
38. Shridhar Kamble, Aaditya Desai and Priya Vartak, "Evaluation and Performance Analysis of Machine Learning Algorithms" in "International Journal of Engineering Sciences & Research Technology", Vol. 3 Issue 5 - May, 2014, ISSN: 2277-9655.

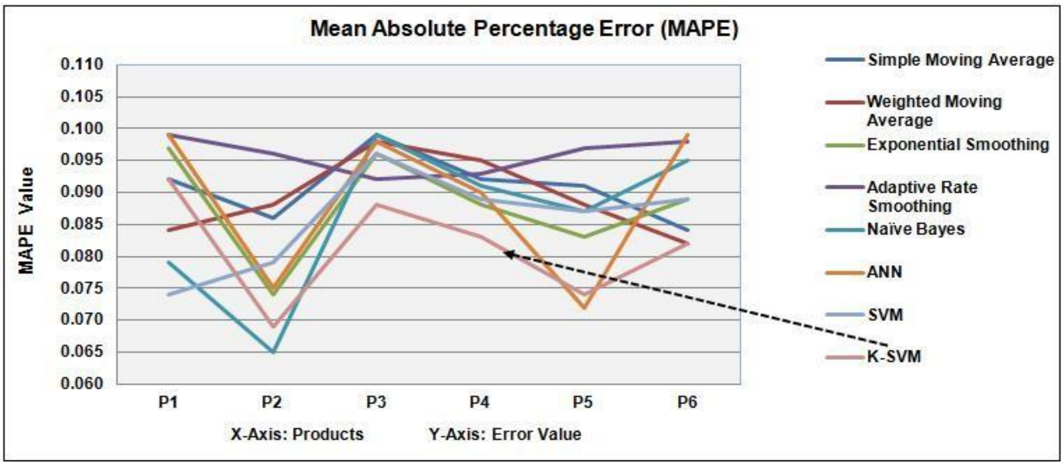

Figure 3: MAPE for All Techniques (Menswear)

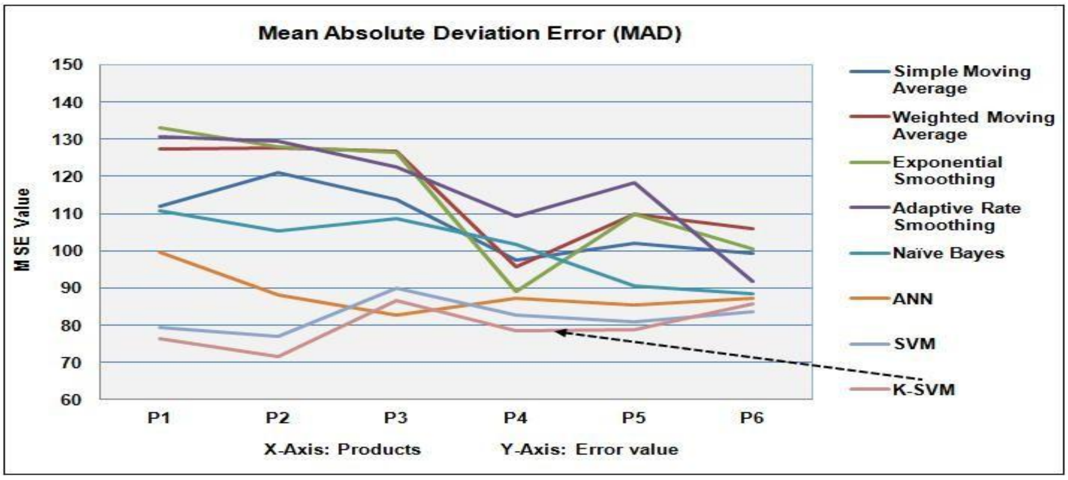

Figure 4: MAD for All Techniques (Menswear)

Figure 5: MSE for All Techniques (Menswear)