A Hybrid Statistical Data Preprocessing and Data Forecasting Model on ERP based Supply Chain Management (SCM) Databases

[1] Sam Sekhar M, [2] P.Venkata Chalapathi

[1] *Department of Mechanical Engineering*, Koneru Lakshmaiah Education Foundation, Guntur, A.P, India.
[2] *Departement of Mechanical Engineering*, MVR College of Engineering & Technology, Vijayawada, A.P, India.

***Abstract* - As size of data increases, most of Enterprise Resource Planning (ERP) systems become automated either in standalone environment or cloud environment. These ERP systems have become more complicated and complex when the number of feature space of the Supply Chain Management (SCM) database increases. Most of the traditional ERP tools analyze the SCM data by using the standard data pre-processing and forecasting models. Also, the size of the feature space in the ERP tools is fixed and noisy. In order to overcome these issues in the traditional ERP tools, a novel filtered based forecasting model is required to improve the prediction accuracy on the large feature space and data size. In this paper, a novel statistical kernel data pre-processing based forecasting model is designed and implemented on SCM dataset. Experimental results proved that the present model has high computational forecasting accuracy compared to the traditional forecasting models.**

Keywords* - *Supply chain management, data forecasting, prediction, oracle ERP, machine learning.

I. INTRODUCTION

Machine learning (ML) is a complex technology and when implemented into business processes, it may imply exposing the business to new risks. A human can easily explain and trace her thought of reasoning that derives a decision. This implies an important transparency in the human decision-making process, often taken for granted (Davenport & Ronanki 2018). This is however a large risk with ML. A model's output is often developed through a vast amount of connections in data which makes the reasoning behind the decisions hard to trace and interpret (Brynjolfsson). This is particularly a problem with the extensive networks of deep learning where it is almost impossible to trace the decision-making elements of the algorithms. This phenomenon is often referred to as the black-box dilemma (Davenport & Ronanki 2018) and raises the question of whether the results of a ML algorithm should be trusted or not. This is especially an important question to evaluate in highly regulated industries where it is crucial to maintain a transparency of the decisions made (Davenport & Ronanki 2018). Therefore, this black-box dilemma poses a risk that should be evaluated depending on the intended environment. There is also a risk associated with hidden biases. Even if the algorithms per se do not involve personal bias, there is a risk of bias that stems from the datasets used to train the algorithm. An example of this could be a ML algorithm that is trained by data from manual processes.

Presently, machine learning approach is identified as an efficient and advanced data mining technique which has wide range of applications. The complete working process of data mining is categorized into three sub-stages.

1. Data pre-processing: During the pre-processing stage, different data filtering operations such as filling missing values, data transformation, etc. are carried out.
2. Data Extraction: In this step, patterns are extracted with the help of different efficient extraction algorithms. As we mention here, the initial step enhances the data quality and prepare these data as input to the following step. In the second step, the filtered data are used. Here, various useful patterns are extracted by implementing various extraction algorithms. Thus, these basic concepts of pattern extraction can be easily represented with the help of a model.
3. All required patterns which are successfully extracted in the previous step are used as input to the next step. In the next step, these extracted patterns are evaluated, validated and consolidated. The above patterns must be useful as well as reliable. We must check whether this extracted knowledge is validated with respect to the prior knowledge of the domain. Therefore, most of the conflicts are discarded during this stage. In the final step, all the useful knowledge are consolidated and provided to the end users. Different kinds of machine learning approaches are presented in order to build the above models. These approaches emphasizes on different measures of performance just like accuracy and understandability. The complexity increases with increase of dimensionality.

Hence, the pre-processing step is very important because it decreases dimensionality of SCM data sets. For example, feature selection technique plays significant role throughout the process of data mining. The pre-processing step can be called as the most costly step of mining, as it uses 80% of the total process. Quality of data completely depends upon accurate planning as well as execution the above mention steps. Feature selection technique can be implemented here. It discards duplicate and non-important features. In other

DOI 10.5013/IJSSST.a.19.06.25

ISSN: 1473-804x online, 1473-8031 print

words, it reduces the complexities and makes the resulted features more understandable. Since years extensive amount of researches are carried out in order to develop numbers of different feature selection algorithms.

There are numbers of different feature selection methods. Hence, it is very complicated and hard task to choose the most appropriate and best feature selection technique. As every individual method have its own advantages and disadvantages, we must be very careful while choosing appropriate method in order to result optimized performance. Most of the approaches are inefficient to handle huge quantities of features.

Feature selection process has the prime objective to decrease original set of features. FS approach can be implemented to produce numbers of different subsets having equal or enhanced representation potential. It is also capable to reduce the influences of the curse of dimensionality. There are huge numbers of feature selection algorithm developed by many researchers since last three decades. Among these large numbers of algorithms, appropriate technique must be chosen very carefully.

The process of prediction is very much important during the whole process of decision making. This can be successfully achieved with the help of some predictive technique known as classifier or supervised learner. The SCM predictive model has the responsibility to predict essential and unknown class level in different applications. It is very difficult and challenging task to enhance the predictive accuracy. The classification accuracy completely depends upon the training data sets in the predictive model. The irrelevant and duplicate features of training data set decrease the overall predictive accuracy. In order to enhance the performance of the predictive model, at first the prediction accuracy must be enhanced significantly. After that, decrease the time required for construction predictive model can also improve performance.

The feature selection scheme can be categorized into four sub-categories, those are:-

1. Wrapper
2. filter,
3. embedded and
4. hybrid approaches.

The feature selection technique has the responsibility to choose required featured and discards the duplicate and irrelevant features out of a particular data set with the help of statistical measure. Filter approach does not depend upon the type of classifier, hence it can be implemented in any classification technique. The embedded technique includes training phase of supervised learning scheme. It helps to choose useful features out of training data set. Thus, the performance of the above technique completely depends upon the base classifiers. With the help of classifier, wrapper technique is implemented to identify required features in high dimensional dataset. Filter and wrapper techniques can be merged together in order to implement a hybrid approach.

Predicting data is considered as an efficient measure in order to evaluate the efficiency and effectiveness of a supply chain management system. A large numbers of data predicting models have been proposed in the literature on production, supply chain of manufacturing, engineering, finance, healthcare domain, etc to provide the better efficiency on SCM data. A build-to-order supply chain is a production system which is capable of predicting goods and services according to the customer's requirements. The time constraint and overall costs are considered as vital factors during the process of supply chain management. Build-to-order and configure-to-order strategies are usually carried out through the process of bulk customization and ecommerce. There have been extensive amount of research works carried out in the domain of ERP systems since last decade.

Most of the supply chain management systems are designed and implemented in cloud computing environment in which the resources are not physically exist for SCM models. In fact, the resources are virtualized and distributed. These resources are geographically dispersed which can be accessed with the help of a non-demand, web-based approaches. Most of the classical business applications such as enterprise resources planning, computer aided design, computer aided manufacturing, product data management, depends on a centralized server. All these systems are not autonomous and never support dynamic business environment.

Most of the traditional supply chain management models are failed to analyze the decision patterns on the available data due to noise and uncertainty. In build-to-order supply chain, the complete supply chain is formed from upstream suppliers through downstream order and delivery options. The complete process of build-to-order strategy can be divided into two sub-, those are:- upstream oriented just-in-time method and downstream oriented build-to-order method. A large number of prediction problems present in the current SCM systems which may lead to negative implication on the performance. All of the organizations must have an efficient and effective method in order to manage the risk factors and data prediction models.

Presently, almost all of the companies have shifted from the traditional centralized operations to decentralized operations. Companies have modified most of their strategies, methods and operations in order to deal with the modifying requirements. Also, almost all of the companies are required to compete with each other according to multiple competitive performance objectives such as quality, price, responsiveness, flexibility and dependability. The current market environment mostly changes according to the customers' requirements and preferences. Apart from this, exponential growth of technology can also influence the market environment. Sometimes, these factors may create severe challenges for the production managers. Hence, it is very much necessary to develop different models in order to determine optimal solution. Most of the software development companies presently uses in- memory database that usually supports the concepts of real-time business intelligence. Additionally, it also supports the big data analytics for large complicated data. The in- memory database systems provide fast processing and analysis of vast amount of data. Different data management strategies and multicore hardware architecture are developed in order to manage the in- memory database systems. Enterprise system usually involves different applications such as supply chain management, enterprise resource planning and customer relationship management. ERP system can be defined as a specific kind of information system that is required during the planning and integration of enterprise's subsystems. The prime objective of this ERP system is to merge the interdepartmental operation procedures and management information systems together. Hence, it will decrease the supply chain expenses, decrease the production time, enhances the product quality and provide quality services for its all customers. The implementation procedure of ERP system from buyer to supplier can decrease the overall expenses and the total amount of time. The process of supplier selection and order allocation are considered as the two vital problems that usually occur in the purchasing department of most of the enterprises. These problems can be resolved through the implementation of ERP system. The selection procedure of appropriate supplier brings advantages for the company. Apart from this, it can also increase the customer satisfaction level. Cost reduction is considered as the most vital factor during the process of decision making. However suppliers' price, quality and service are considered as the vital factors during the selection procedure of suppliers. The process of supplier selection is considered as a multiple criteria decision making issue in SCM data prediction.

II. LITERATURE REVIEW OF RELATED WORKS

N. K. Dev, et proposed a model using multi-criteria indicators on SCM data [1]. The major challenge for a big data architect is to incorporate real time predictive analytics ability with the help of offline techniques such as simulation, fuzzy analysis process and order preferences. This proposed approach is an extension of operational units to manage the unstructured relational key performance indicators. This approach is implemented in a big data framework. In other words, presented approach can be defined as a decision support tool which has the responsibility to compute key performance indicators in case of a real-time dynamic system.

S. A. Mansouri, et.al proposed a decision support method for build-to-order supply chain management using a multi-objective optimization [2]. This main objective of this paper is to detect the limitations of decision making support based on multi objective optimisation process. They studied and analysed all the research ideas related to build-to-order supply chain based on multi objective decision making process.

R. Addo-Tenkorang et.al, proposed a data pre-processing and prediction approach on big data applications in supply chain management [3]. Now-a-days, the uses and popularity of big data is increasing rapidly. It is considered as a completely new enterprise system or environment that can offer large numbers of features for acquiring, storing and analysing vast amount of data.

V. Botta-Genoulaz et.al, performed a thorough survey on different ERP systems [4]. An extensive research works have been carried out since last two decades in the field of ERP systems.The complete process SCM is categorized into six types, those are:- ERP implementation, ERP optimisation, ERP management, ERP software and ERP for supply chain management.

S. Bruque-Cámara, et.al, implemented a novel prediction model on SCM data to verify the influences on operational performance [5]. In this work, they analysed and predicted the profit of the SCM data in cloud environment. They implemented a factorial analysis method and data modelling to evaluate the statistical hypothesis.

C. Chen, et.al, implemented a new SCM based cloud computing platform for data prediction [6]. Cloud platform helps for providing ERP services in a distributed manner. According to a recent survey, ERP based on SaaS provides better performance as compared to the IT services. In this research paper, they proposed a novel prediction modelling on cloud-based ERP platform.

W. J. Christensen, et.al, focused on build-to-order and just-in-time predictors in case of applied supply chain knowledge and market performance [7]. The just-in-time method never influences top stream application and it is inefficient to predict the market performance.

M. Giannakis et.al, developed a new multi-agent framework for supply chain risk management process [8]. The high level of complexities in case of supply chains is considered as the major limitation in order to achieve higher supply chain performance. In this work, they implemented a new framework in order to design a multi-agent based decision support system.

A.Gunasekaran et.al, proposed a thorough literature review on different build-to-order supply chain management systems [9].

The build-to-order supply chain or make-to-order system has become more popular since last decade due to large numbers of new data predicting models. Hence, it is necessary to improve the process of traditional modelling and analysis. In most of the scenarios, the conventional operations research approaches are implemented in order to resolve the issues of supply chain management process.

G. J. Hahn and J. Packowski emphasized on different applications of in-memory analytics and supply chain management models [11]. In this work, they have presented a comprehensive view on different applications of in-memory analytics of SCM. The complete working procedure of the above proposed approach can be divided into three phases. In the first phase, they introduced a top-down framework in order to place in-memory analytics applications in SCM. In the next phase, they performed a bottom-up categorisation process on 41 in-memory analytics applications in order to obtain supporting empirical evidence about efficiency of the framework. In the last phase, they obtained different implications for the research and industrial domain.

K. B. Hendricks, et.al, analysed the influence of enterprise systems on corporate performance [12]. They have considered three important systems in their study, those are: ERP, SCM and CRM. In this paper, they have studied and analysed the influences of investments in ERP, customer relationship management systems and supply chain management. They have studied the long term stock price performance and profitability for different companies. In case of traditional ERP systems, better profitability can be achieved as compared to the modified and extended ERP systems. On the contrary, by implementing supply chain management systems, better positive stock returns can be obtained along with improved profitability. On the other hand, in case of customer relationship management systems, neither better profitability nor positive stock returns can be achieved.

S. Kusi-Sarpong, et.al, developed a new strategy for green supply chain evaluation in the domain of mining [13]. They used the basic concepts of rough set and fuzzy TOPSIS approach. This proposed framework is very much beneficial in case of practical managerial decision making processes. They have also presented multiple criteria evaluation in order to test the efficiency of the framework.

C. Lin, et.al, introduced a new ERP model in order to carry out the process of supplier selection [14]. Therefore, the selection process of supplier is very vital for the purchasing department. An enterprise resource planning system can be defined as a specific process selection which will reduce the overall expenses and workload of the company. In other words, an ERP system is defined as an effective tool that is helpful during the process of resource integration and profit creation. By implementing an efficient ERP system, a decision manager will be able to identify the pros and cons of purchasing operation. In order to develop a real time processing environment, a new approach analytic network process is considered.

S. K. Mukhopadhyay et.al, focused on optimal return policy and modular design for build-to-order products [15]. A build-to-order product provides chances to customise the product according to the need of customers. For example we can consider the return policy as an effective tool by which more customer orders can be placed. Apart from this, we can say that, the popularity of internet sale is also growing day by day. But the major disadvantage of internet sale is, the customer cannot verify the product by himself. Hence, most of the companies are offering return policy in order to gain the trust of customers. In this paper, they have considered the theory of modularity as an effective solution for this issue. From the seller's perspective, a proper return policy will no doubt enhance the revenue. They implemented a profit maximization scheme in order to provide optimal policies for both return and modularization.

S. Naciri, et.al, developed an ERP data sharing framework with the help of Generic Product Model [16]. Traditional SCM models are studied and their limitations are summarized on large databases. Due to the large numbers of heterogeneous applications, various hardware and software systems, data management software, data models, schemas, etc are required to analyse the data forecasting models. In order to resolve this issue, Hitachi company proposed a specific modelling language which is termed as generic product model. This models is used to store, share and visualise product data within a specific data warehouse. To extend the range of data those are required to be stored and shared with the help of SCM data warehouse, this work proposed a specific scheme along with a new translator.

J. P. Sepúlveda-Rojas, et.al, proposed a new forecasting scheme for efficient supply chain demand estimation [17]. The main objective of this work is to provide an efficient and effective selection method that will be helpful to predict model during the demand estimation of supply chains. Now-a-days, in order to predict the future demand of a product, different forecasting schemes are developed. Most of the forecasting schemes completely depends upon the historical information (qualitative and quantitative).

Most of these forecast schemes depend upon time series. After that the process of estimation is carried out with the help of forecast error measurement criteria. The forecast error measurement criterion is able to identify the best approach among several approaches. A. Syntetos , et.al, proposed a new supply chain forecasting approach on small size SCM databases [18]. The process of supply chain forecasting is not a simple process. This process includes different complicated issues such as supply chain coordination and sharing of information among different stakeholders.

C. D. Tarantilis, et.al, proposed a web-based ERP system in order to manage different business services and supply chain management [19]. They analysed various applications of real-world process scheduling. An advanced web based ERP system is proposed in order to resolve different business issues. Apart from this, this system is responsible to handle all types of real world business processes. They have introduced a strong workflow engine which is responsible to handle the complete process event flow inside a particular enterprise. The main objective of this process is to enhance the overall efficiency and control. The main problem is to allocate the enterprise resources for data forcasting. The resources are usually located in such a way that the idle time and delays will be reduced.

S. Wadhwa proposed a new technique for supply chain modelling [20]. This proposed approach is an agent based

DOI 10.5013/IJSSST.a.19.06.25 ISSN: 1473-804x online, 1473-8031 print

technique. The most important factor that influences the performance of a supply chain is the modelling of supply chain. A new agent based architecture for modelling of supply chains is implemented in this paper. Agents are considered as the basic and autonomous units those are capable enough to make decisions based on their dynamic business environment. They analysed different feature selection-based ensemble technique in order to resolve data mining issues [8]. All traditional ensemble approaches integrate different machines in order to enhance efficiency in a particular learning task. Some of these approaches are implemented in evolutionary machine learning schemes with an inclusion of learning classifier systems. The working procedure of the above framework can be subdivided into three phases, those are:- pre-gate phase, member phase and post-gate phase. In case of pre-gate phase, data preparation is done. But in case of member phase, various kinds of learning machines are analysed. In the post-gate phase, all the outcomes are merged together in order to form a single ensemble output. Performance of two different kinds of ensembles (random sampling and RSFS feature selection) is studied here. RSFS can be used as the best choice for building ensemble models having irrelevant features. The kind of feature and size of subspace can greatly affect the overall performance of random sampling. The ratio of irrelevant features to relevant features can also influence the performance significantly.

III. PROPOSED MODEL

In the proposed model, a novel filtered based data forecasting approach is designed and implemented on SCM datasets. In this model, initially SCM data is collected from the JD Edwards ERP tool for data pre-processing and analysis. This collected data are pre-processed using the improved normalization method. Improved normalization method is used to filter the noise present in each continuous attribute in order to improve the forecasting accuracy. In the second phase, pre-processed data is given to the principal components analysis to find the weights of each attribute for data forecasting. In the third phase, a multi-layered feed forward neural network is used to improve the prediction of each test instance in the SCM database as shown in figure1.

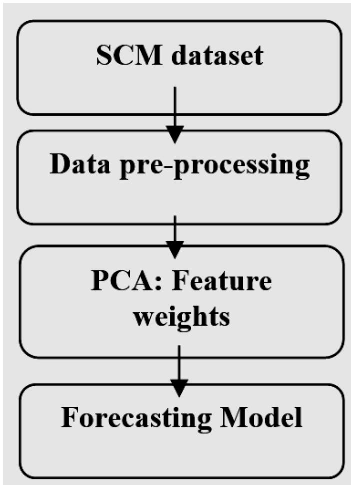

Figure 1-a: Proposed Model

Algorithm 1: Kernel based Data Pre-processing

Input : Training microarray dataset D, F(D): Feature space of D, Max similarity MaxSim[], Threshold T.

Output: Kernel Filtering or Transformed data KD.

Procedure:

Read input data D.

For each pair of feature F[i],F[j] in feature space F(D)

Do

Apply Kernel transformation on I as

$$\text{GeneKernelTransform}(F[i],F[j]) = \frac{1}{1+\eta^2/\cos(\max(\sigma^2_{F[i]},\sigma^2_{F[j]})}$$

Where $\eta = 2*\sum F[i].F[j] - \sum (F[i]+F[j])^2$

If($\text{GeneKernelTransform}(F[i],F[j]) > T$)

Then

Normalize F[i] and F[j] within [0, $\text{GeneKernelTransform}(F[i],F[j])$] using Min-max normalization as KD

Else

Normalize F[i] and F[j] within [0,1] using Min-max normalization as KD.

End if

Done

DOI 10.5013/IJSSST.a.19.06.25 ISSN: 1473-804x online, 1473-8031 print

Algorithm 2: A Novel Weighted Method for FFNN Forecasting

Traditional feature selection measures such as t-statistic, Significance Analysis of Microarray (SAM) and signal to noise ratio (SNR) are used to rank the k-means clustered features on SCM datasets. The main problem in these feature ranking measures is the selection of appropriate features from the high dimensional feature space using wrapper method. These ranking measures incorporate the forecasting accuracy and true positive rate on the selected features (>50) using the t-test, SAM and SNR measures. The modified version of SAM, SNR and t-test measures are summarized below.

Step 1: Compute the attribute weights using the eigen values of PCA method

Step 2: Assign feature weighted using the maximized weights using the (1), (2) and (3) to each feature in feed forward neural network approach.

T-statistic weighting measure is used to find the variation in the features using the standard deviation of the class labels.

W1 is the ratio of difference of the means of the class labels to the maximized standard deviation.

$$W1 = \frac{\mu_P - \mu_N}{\sqrt{\max\{\sigma_P^2 / |P|, \sigma_N^2 / |N|\}}} \qquad (1)$$

Where μ_P is the mean of the positive cluster class samples
μ_N is the mean of the negative cluster class samples.

W2 is the ratio of difference of the means of the class labels to the sum of the standard deviation of the positive and negative classes. Here, the features with highest signal to noise ratio measure is selected as highest weighting measure for data classification.

$$W2 = HSNR = \frac{|\mu_i - \mu_j|}{2(\sigma_P + \sigma_N)} \qquad (2)$$

Where μ_P and σ_P are the mean and standard deviation of the cluster positive class samples
μ_N and σ_N are the mean and standard deviation of the cluster negative class samples.

W3 is the maximization of the correlation between the features, hybrid t-test and hybrid SNR ratio. This ranking measure is used to select the optimal binary class features in each cluster.

$$W3 = MCTSNR = \mathrm{Max}\{\mathrm{Correlation(ClusterFeatures : CF)}, \frac{\mu_P - \mu_N}{\sqrt{\max\{\sigma_P^2 / |P|, \sigma_N^2 / |N|\}}}, \frac{|\mu_i - \mu_j|}{2(\sigma_P + \sigma_N)}\} \qquad (3)$$

Weights W[] = Max{W1,W2,W3}

Step 3: Defining the input, hidden and output layers to each mapper for parallel processing.

Step 4: To each hidden layer apply the logistic activation function for weights and error rate optimization.

Step 5: Classify data using the deep neural network framework until error rate and weights are converged.

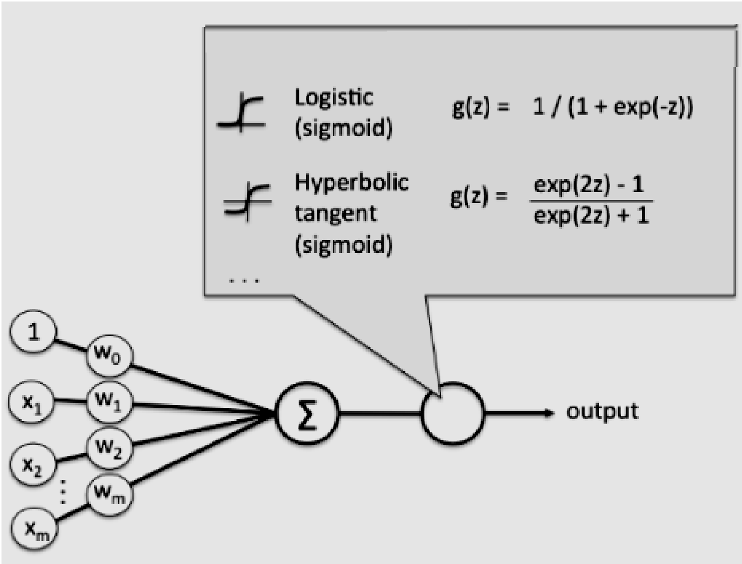

Figure 1-b: Logistic and Hyperbolic output functions

IV. EXPERIMENTAL RESULTS

Experimental results are simulated on real-time oracle ERP supply management dataset with different feature sets. Results are developed in Java environment for data pre-processing, feature extraction and enhanced forecasting model on SCM data.

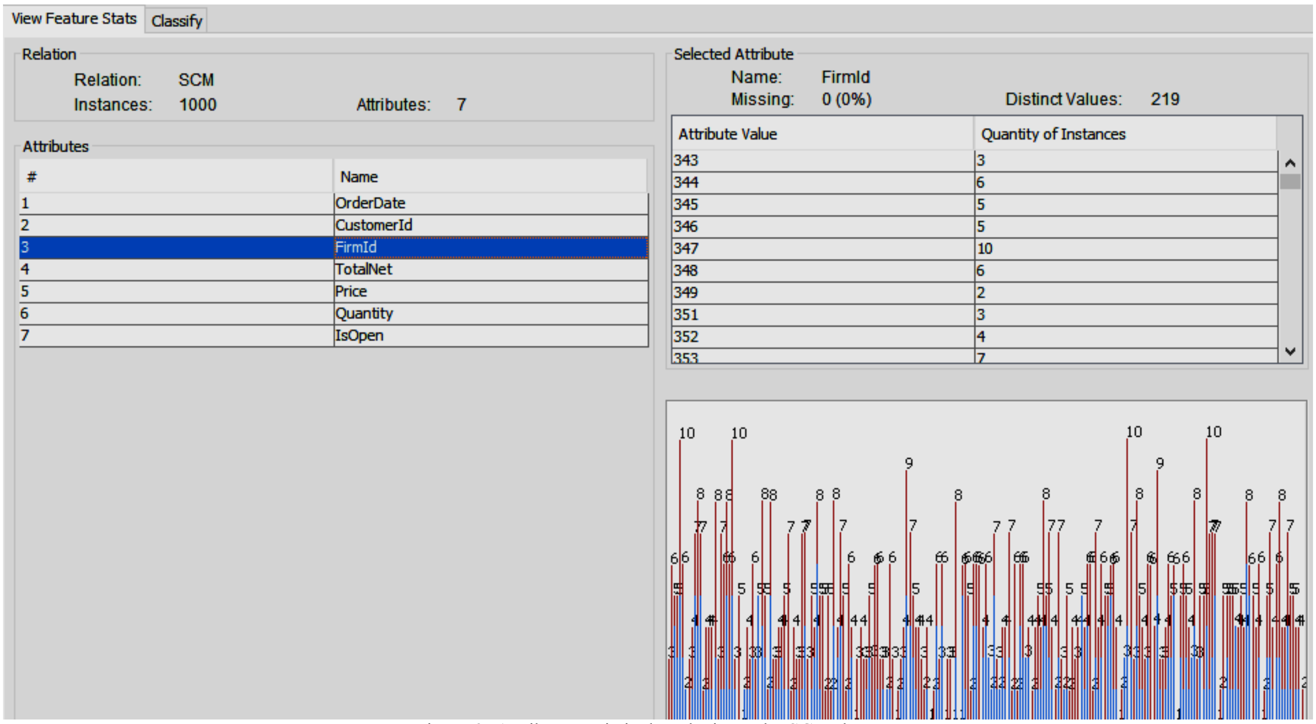

Figure 2: Attribute statistical analysis on the SCM dataset.

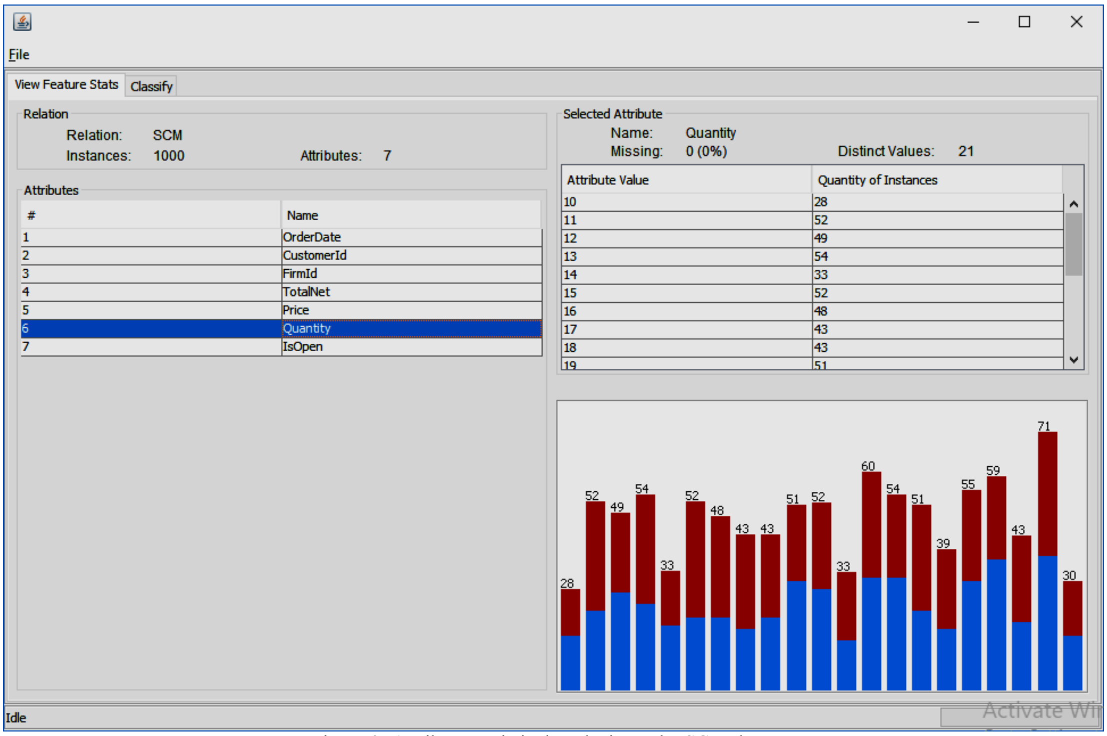

Figure 3: Attribute statistical analysis on the SCM dataset.

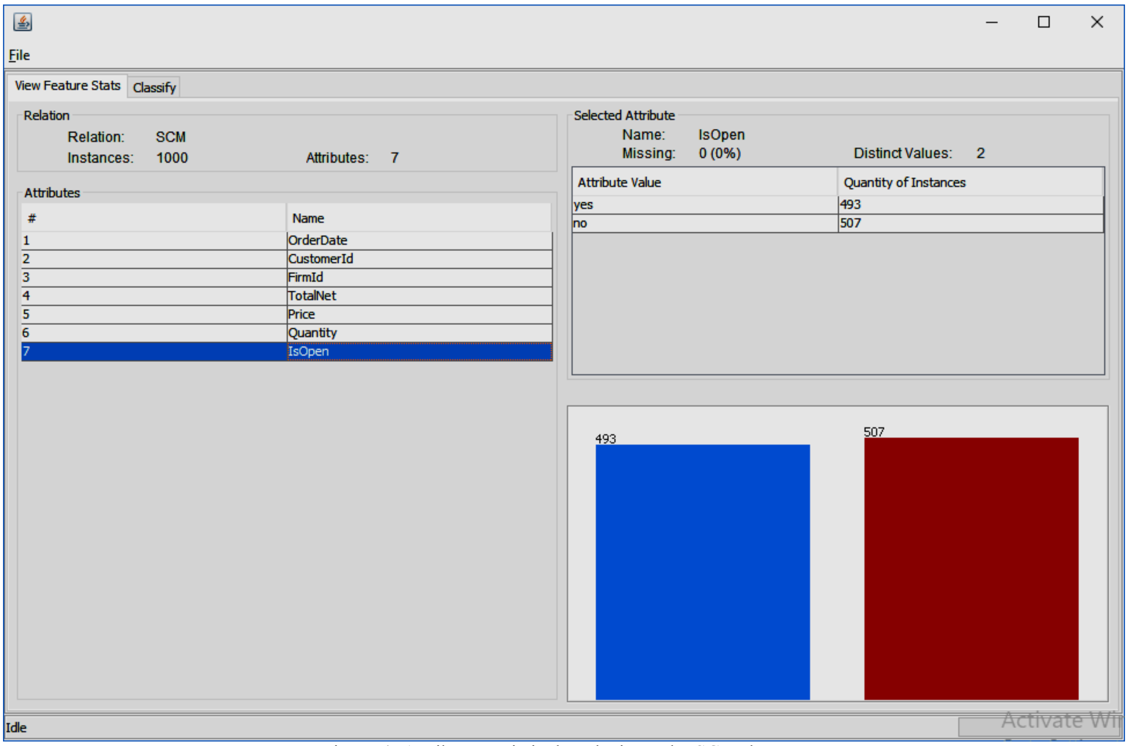

Figure 4: Attribute statistical analysis on the SCM dataset.

Figure 2,3,4 illustrate the statistical analysis of each numerical attirube and its occurences in SCM dataset. From these figures, we can analyze the data distribution of each attribute in the kernel normalization process.

Feature Selection IPCA:

```
eigenvalue      proportion      cumulative
  0.67006         0.03467         0.03467       -0.523a6-0.505a7-0.455a4-0.449a5...
  0.07784         0.00403         0.0387        -0.666a3-0.641a2-0.214a4-0.207a7...
  0.07176         0.00371         0.04242       0.847a1+0.422a2-0.194a4-0.159a6...
  0.06988         0.00362         0.04603       0.731a3-0.53a2+0.325a1-0.201a4...
  0.05175         0.00268         0.04871       0.657a5-0.535a4-0.342a1+0.299a7...
  0.02968         0.00154         0.05024       0.654a7-0.615a6-0.303a5-0.25a2...
  0.02903         0.0015          0.05175       -0.589a4+0.512a6-0.475a5+0.395a7...
```

```
Eigenvectors
 V1      V2      V3      V4      V5      V6      V7
-0.1706  0.1703  0.8471  0.3253 -0.3418 -0.0349 -0.0217 a1
 0.1785 -0.6407  0.4221 -0.5299  0.1648 -0.2503 -0.0945 a2
 0.0431 -0.6661 -0.1374  0.7309  0.006  -0.035  -0.0122 a3
-0.4546 -0.2139 -0.1944 -0.201  -0.5349  0.1912 -0.5888 a4
-0.449   0.1512  0.0628  0.1482  0.6572 -0.3032 -0.4752 a5
-0.5232 -0.0727 -0.1593 -0.0845 -0.2199 -0.6149  0.5119 a6
-0.5052 -0.2073  0.1354 -0.0981  0.2992  0.6545  0.3949 a7
```

```
Ranked attributes
 0.965   -0.523a6-0.505a7-0.455a4-0.449a5...
 0.961   -0.666a3-0.641a2-0.214a4-0.207a7...
 0.958   0.847a1+0.422a2-0.194a4-0.159a6...
 0.954   0.731a3-0.53a2+0.325a1-0.201a4...
 0.951   0.657a5-0.535a4-0.342a1+0.299a7...
 0.95    0.654a7-0.615a6-0.303a5-0.25a2...
 0.948   -0.589a4+0.512a6-0.475a5+0.395a7...

Selected attributes: 1,2,3,4,5,6,7 : 7
```

The above results illustrates the feature weighting process of each attriubute in the PCA method. Here, the eigen value of each attribute is used as weight in the feed forward neural network forecasting model.

TABLE II. PERFORMANCE ANALYSIS OF PROPOSED MODEL TO THE EXISTING MODELS IN TERMS OF TRUE POSITIVE RATE AND RUNTIME(MS)

Model	True Positive	Error Rate	Runtime (ms)
Regression	0.897	0.2492	4526
SVM	0.922	0.1893	4197
Neural network	0.965	0.387	3675
Proposed Model	0.976	0.354	2475

Predicting Results of Forecasting Model on Different Feature Values: Set A

Feature Value set A
Quantity != 30 -> IsOpen != no OrderDate != 09-04-2017 -> IsOpen != no Quantity != 30 AND Price != 34950 -> IsOpen != no Price != 35846 -> IsOpen != no Quantity != 30 AND CustomerId != 12709 -> IsOpen != no Quantity != 30 AND Price != 34950 -> IsOpen != no Quantity != 30 AND OrderDate != 16-11-2015 -> IsOpen != no TotalNet != 11447 -> IsOpen != no OrderDate != 03-11-2015 -> IsOpen != no

Predicting Results of Forecasting Model on Different Feature Values: Sets B, C, D, E

Feature Value Set B
Quantity != 30 AND TotalNet != 12677 -> IsOpen != no
OrderDate != 10-05-2016 -> Quantity != 30
CustomerId != 12323 AND OrderDate != 16-11-2015 -> IsOpen != no
IsOpen != no AND Quantity != 30 -> TotalNet != 11744
IsOpen != no -> Price != 27061
CustomerId != 12238 AND CustomerId != 12709 -> IsOpen != no
FirmId != 434 AND TotalNet != 12677 -> IsOpen != no
Price != 31496 AND CustomerId != 12709 -> IsOpen != no
TotalNet != 12676 AND OrderDate != 16-11-2015 -> IsOpen != no
IsOpen != no -> FirmId != 434
Feature Value Set C
Quantity != 30 AND IsOpen != no -> TotalNet != 12115
IsOpen != no -> OrderDate != 10-05-2016
IsOpen != no -> OrderDate != 26-05-2016
Price != 26043 AND Price != 34950 -> IsOpen != no
IsOpen != no -> CustomerId != 12709
IsOpen != no -> TotalNet != 12454
Feature Value Set D
Quantity != 30 AND IsOpen != no -> OrderDate != 08-07-2016
CustomerId != 12709 AND Price != 34950 -> IsOpen != no
TotalNet != 12676 AND TotalNet != 12677 -> IsOpen != no
IsOpen != no AND Quantity != 30 -> CustomerId != 11127
IsOpen != no AND Quantity != 30 -> Price != 31019
Feature Value Set E
Quantity != 30 -> OrderDate != 20-10-2017
Price != 27189 AND OrderDate != 16-11-2015 -> IsOpen != no
IsOpen != no -> TotalNet != 11998
FirmId != 434 AND OrderDate != 16-11-2015 -> IsOpen != no
IsOpen != no AND Quantity != 30 -> OrderDate != 16-11-2015
TotalNet != 10386 AND CustomerId != 12709 -> IsOpen != no
IsOpen != no AND Quantity != 30 -> FirmId != 434
CustomerId != 11362 AND CustomerId != 12709 -> OrderDate != 08-09-2016
CustomerId != 12607 AND TotalNet != 12677 -> IsOpen != no
Price != 34950 AND TotalNet != 12677 -> IsOpen != no
OrderDate != 08-09-2016 AND CustomerId != 12709 -> IsOpen != no
OrderDate != 25-09-2017 AND CustomerId != 12709 -> IsOpen != no
OrderDate != 26-05-2016 AND Price != 34950 -> IsOpen != no
CustomerId != 10197 AND OrderDate != 16-11-2015 -> Quantity != 30
TotalNet != 10080 AND OrderDate != 16-11-2015 -> IsOpen != no
FirmId != 549 AND TotalNet != 11357 AND CustomerId != 11228 -> IsOpen != no
CustomerId != 12238 AND TotalNet != 11357 AND CustomerId != 11228 -> IsOpen != no

Predicting Results of Forecasting Model on Different Feature Values: Set F

Feature Value Set F
Quantity != 30 AND TotalNet != 10386 AND TotalNet != 10267 AND OrderDate != 20-10-2017 -> IsOpen != no
TotalNet != 12677 AND TotalNet != 11357 AND CustomerId != 11228 -> IsOpen != no
OrderDate != 16-11-2015 AND CustomerId != 12238 AND Quantity != 30 AND OrderDate != 20-10-2017 -> IsOpen != no
Price != 26450 AND TotalNet != 11357 AND CustomerId != 11228 -> IsOpen != no
OrderDate != 26-05-2016 AND TotalNet != 11357 AND CustomerId != 11228 -> IsOpen != no
CustomerId != 12753 AND TotalNet != 11357 AND TotalNet != 10386 -> IsOpen != no
OrderDate != 16-11-2015 AND CustomerId != 12238 AND OrderDate != 15-04-2015 AND OrderDate != 20-10-2017 -> IsOpen != no
FirmId != 434 AND TotalNet != 10386 AND TotalNet != 10267 AND OrderDate != 20-10-2017 -> IsOpen != no
CustomerId != 12709 AND TotalNet != 12067 AND TotalNet != 10267 AND OrderDate != 20-10-2017 -> IsOpen != no
CustomerId != 12709 AND TotalNet != 10386 AND TotalNet != 10267 AND OrderDate != 20-10-2017 -> IsOpen != no
CustomerId != 10433 -> OrderDate != 09-07-2016
Price != 33966 -> CustomerId != 11228
IsOpen != no AND OrderDate != 16-11-2015 -> TotalNet != 12132
IsOpen != no AND Quantity != 30 AND TotalNet != 10080 -> Price != 36188
TotalNet != 11319 -> OrderDate != 20-10-2017
IsOpen != no AND CustomerId != 12709 -> FirmId != 434
IsOpen != no AND CustomerId != 11700 -> TotalNet != 10164
CustomerId != 10950 AND IsOpen != no -> Price != 31019
OrderDate != 16-11-2015 -> Price != 27061
TotalNet != 10386 -> CustomerId != 12709
TotalNet != 10386 -> CustomerId != 12709
IsOpen != no AND OrderDate != 16-11-2015 -> FirmId != 534
IsOpen != no AND OrderDate != 16-11-2015 -> FirmId != 534
Price != 29432 -> TotalNet != 10386
CustomerId != 11038 -> Price != 27061
IsOpen != no AND Price != 31496 -> OrderDate != 08-07-2016
TotalNet != 10386 -> CustomerId != 12709
TotalNet != 12454 AND IsOpen != no AND Quantity != 30 -> OrderDate != 20-10-2017
TotalNet != 12454 -> OrderDate != 26-05-2016
Price != 30492 -> FirmId != 434
TotalNet != 12132 AND IsOpen != no -> OrderDate != 08-07-2016
CustomerId != 12335 -> Price != 27061
IsOpen != no AND Quantity != 30 AND CustomerId != 12238 -> Price != 36188
Price != 29555 -> OrderDate != 26-05-2016
CustomerId != 11038 -> Price != 27061
IsOpen != no AND OrderDate != 16-11-2015 -> CustomerId != 12238
TotalNet != 12856 -> Price != 27061
IsOpen != no AND OrderDate != 16-11-2015 -> CustomerId != 12238
CustomerId != 12440 -> OrderDate != 26-05-2016
TotalNet != 11998 -> OrderDate != 09-07-2016
CustomerId != 12621 -> FirmId != 434

Predicting Results of Forecasting Model on Different Feature Values: Set G and H

Feature Value Set G
Quantity != 30 AND Price != 24994 AND IsOpen != no -> TotalNet != 12067
TotalNet != 12271 AND IsOpen != no -> CustomerId != 12550
IsOpen != no AND Price != 34950 -> CustomerId != 12709
IsOpen != no AND CustomerId != 12753 -> Price != 25593
CustomerId != 12110 -> TotalNet != 12788
FirmId != 434 -> TotalNet != 10386
OrderDate != 10-05-2016 -> FirmId != 434
TotalNet != 12677 AND IsOpen != no -> Price != 31019
OrderDate != 15-10-2017 -> TotalNet != 12454
FirmId != 434 AND IsOpen != no -> TotalNet != 12677
CustomerId != 10950 AND IsOpen != no -> OrderDate != 08-07-2016
IsOpen != no AND FirmId != 434 -> Price != 25593
TotalNet != 12676 -> OrderDate != 26-05-2016
IsOpen != no AND TotalNet != 11744 -> OrderDate != 08-07-2016
IsOpen != no AND OrderDate != 21-07-2016 -> CustomerId != 11228
OrderDate != 16-11-2015 -> TotalNet != 12454
IsOpen != no AND TotalNet != 12271 -> CustomerId != 12238
IsOpen != no AND OrderDate != 20-10-2017 -> Price != 31019
IsOpen != no AND Quantity != 30 AND OrderDate != 06-02-2017 -> Price != 36188
IsOpen != no AND TotalNet != 12454 AND Quantity != 30 -> OrderDate != 20-10-2017
Price != 30507 AND IsOpen != no -> OrderDate != 08-07-2016
TotalNet != 10182 -> FirmId != 434
OrderDate != 20-10-2017 -> CustomerId != 11228
Price != 31019 -> TotalNet != 12454
FirmId != 434 AND IsOpen != no -> Price != 31019
IsOpen != no AND OrderDate != 03-11-2015 -> Price != 31019
OrderDate != 10-05-2016 AND IsOpen != no -> TotalNet != 12677
IsOpen != no AND CustomerId != 12709 -> OrderDate != 16-11-2015
IsOpen != no AND Quantity != 30 AND FirmId != 434 -> Price != 36188
FirmId != 434 -> TotalNet != 12454
OrderDate != 29-03-2017 -> TotalNet != 10386
IsOpen != no AND Quantity != 30 AND TotalNet != 12676 -> Price != 36188
TotalNet != 10080 -> CustomerId != 12709
TotalNet != 11319 AND IsOpen != no -> OrderDate != 26-05-2016
Feature Value Set H
Quantity != 30 AND CustomerId != 12709 -> FirmId != 434
CustomerId != 11356 AND OrderDate != 30-11-2017 AND CustomerId != 12753 AND Price != 28688 -> IsOpen != no
OrderDate != 25-09-2017 AND OrderDate != 30-11-2017 AND CustomerId != 12753 AND Price != 28688 -> IsOpen != no
OrderDate != 22-11-2015 AND OrderDate != 30-11-2017 AND CustomerId != 12753 AND Price != 28688 -> IsOpen != no
CustomerId != 12709 AND TotalNet != 11139 AND TotalNet != 11431 AND TotalNet != 12173 -> IsOpen != no
CustomerId != 11409 AND FirmId != 445 -> IsOpen != no
CustomerId != 12238 AND TotalNet != 12012 AND TotalNet != 11431 AND CustomerId != 12607 -> IsOpen != no
CustomerId != 12709 AND Price != 28688 AND TotalNet != 11431 AND TotalNet != 12783 -> IsOpen != no
OrderDate != 16-11-2015 AND CustomerId != 12238 AND TotalNet != 11431 AND OrderDate != 20-10-2017 -> IsOpen != no
Quantity != 30 AND OrderDate != 20-10-2017 AND TotalNet != 11431 AND FirmId != 528 -> IsOpen != no

V. CONCLUSION

Traditional data forecasting techniques such as Naive Bayes, regression models, neural network etc are used to predict the SCM data class label using the limited data size. These models are highly depend on numerical features and its class distribution. These models are not efficient to predict the outliers using the numerical and nominal

DOI 10.5013/IJSSST.a.19.06.25 ISSN: 1473-804x online, 1473-8031 print

features. Also, as the size of the instances is increasing, these models take high computational time and memory for data prediction. In order to overcome these issues in the traditional ERP tools, a novel filtered based forecasting model is required to improve the prediction accuracy on the large feature space and data size. In this paper, a novel statistical kernel data pre-processing based forecasting model is designed and implemented on SCM dataset. Experimental results proved that the present model has high computational forecasting accuracy compared to the traditional forecasting models.

REFERENCES

[1] N. K. Dev, R. Shankar, R. Gupta and J. Dong , Multi-criteria evaluation of real-time key performance indicators of supply chain with consideration of big data architecture, Computers and Industrial Engineering

[2] S. A. Mansouri, D. Gallear and Md. H. Askariazad , Decision supportforbuild-to-ordersupplychainmanagement through multiobjectiveoptimization, , Int. J.ProductionEconomics135(2012)24–36.

[3] Hardik, S., Tarachand, K., Kailash, Y., Chintan, B., Chainesh, S. and Vinod, R., 2012. Formulation, Development and Evaluation of Floating Matrix Tablets of Atorvastatin Calcium. International Journal, 2(4), pp.35-40.

[4] V. Botta-Genoulaz and P.A. Millet , Supply chain integration through community cloud:Effects on Operational performance, Journal ofPurchasing&SupplyManagement.

[5] S. Bruque-Cámara, J. Moyano-Fuentes and J.M. Maqueira-Marín , A cloud computing platform for ERP applicationsChin, Applied soft computing

[6] Sinha, P., 2013. Particle swarm optimization for solving economic load dispatch problem. Int J Commun Comput Technol, 1, p.63.

[7] W. J. Christensen, R. Germain and L. Birou, A multi-agent based framework for supply chain risk management, Journal ofPurchasing&SupplyManagement17(2011)23–31.

[8] M. Giannakis and M. Louis , Build-to-order supply chain management: a literature review and framework for development, Journal of Operations Management 23 (2005) 423–451

[9] A.Gunasekaran and E. W. T. Ngai, Modeling and analysis of build-to-order supply chains, European Journal of Operational Research 195 (2009) 319–334

[10] A. Gunasekaran and E. W.T. Ngai , A perspective on applications of in-memory analytics in supply chain management, Decisio Support Systems

[11] G. J. Hahn and J. Packowski , The impact of enterprise systems on corporate performance: A study of ERP, SCM, and CRM system implementations, Journal of Operations Management 25 (2007) 65–82.

[12] Anuj, S.A., 2013. Green Synthesis of Silver Nanoparticles by Using Tinospora Cordifolia Stem Powder, Characterization and Its Antibacterial Activity Against Antibiotics Resistant Bacteria. International Journal, 3(4), pp.11-16.

[13] S. Kusi-Sarpong, C. Bai, J. Sarkis, X. Wang, An ERP model for supplier selection in electronics industry, Expert Systems with Applications 38 (2011) 1760–1765

[14] C. Lin, C. Chen and Y. Ting , Optimal return policy and modular design for build-to-order products, Journal of Operations Management 23 (2005) 496–506.

[15] S. K. Mukhopadhyay and R. Setoputro, Optimal return policy and modular design for build-to-order products, Journal of Operations Management 23 (2005) 496–506.

[16] S. Naciri, N. Cheikhrouhou, M. Pouly, J. Binggeli and R. Glardon ERP data sharing framework using the Generic Product Model (GPM), Expert Systems with Applications 38 (2011) 1203–1212.

[17] J. P. Sepúlveda-Rojas, F Rojas, H. Valdés-González and M. S. Martín, Forecasting Models Selection Mechanism for Supply Chain Demand Estimation, Procedia Computer Science 55 (2015) 1060 – 1068

[18] A. Syntetos , Z. Babai , J. Boylan , S. Kolassa and K. Nikolopoulos, Supply Chain Forecasting: Theory, Practice, their Gap and the Future, European Journal of Operational Research.

[19] C. D. Tarantilis, C. T. Kiranoudis and N. D. Theodorakopoulos, A Web-based ERP system for business services and supply

[20] chain management: Application to real-world process scheduling, European Journal of Operational Research 187 (2008) 1310–1326

[21] S. Wadhwa, Supply Chain Modeling: The Agent Based Approach, Information Control Problems in Manufacturing, May 17-19 2006, Saint-Etienne, France

.