Ensemble Approach for Time Series Analysis in Demand Forecasting

Ensemble Learning

A. Okay Akyuz
OBASE, Research & Development Center
OBASE
Istanbul, Turkey
okay.akyuz@obase.com

Berna Atak Bulbul
OBASE, Research & Development Center
OBASE
Istanbul, Turkey
berna.bulbul@obase.com

Mitat Uysal
Dogus University, Department of Computer Science
DOU
Istanbul, Turkey
muysal@dogus.edu.tr

M. Ozan Uysal
Dogus University, Department of Computer Science
DOU
Istanbul, Turkey
ouysal@dogus.edu.tr

***Abstract*—Demand forecasting for replenishment is one of the main issue for retail industry in terms of optimizing stocks, minimizing costs and also for reducing stock out problem. Better forecasting for demands, means maximizing sales and result with more revenue and profit for retailers. An other critical result of the stock out problem is of course dissatisfied customers and customer churn effect to retailers as well. Customers, in general do not wish to buy an equivalent product from different brands instead of their routine selections. There are of course many parameters which affect very seriously forecasting accuracy of consumer demands. For instance; seasonality, promotional effects, social events, new trends, unexpected crisis, terrorism, changes on weather conditions, commercial behavior of competitors at the market etc. In this study, new heuristic approach for ensemble methodology has been proved. It has been implemented in SOK Market. It is one of Turkey's hard discount retail chain with 4000 stores and replenishes 1500 SKUs to stores via 22 regional distribution centers. The results of this approach and how to take benefits of the powerful common minded demand forecasting in time series forecasting analysis have been showed.**

Keywords—ensemble learning; ensemble for time series; demand forecasting; replenishment

I. INTRODUCTION

A supply chain involves the constant flow of information, product, and funds between different stages. Each stage of the supply chain interacts with each other [1]. Meanwhile, the flow of material is an operational level. The flow of material and information is necessary because of the managing the supply chain. Supply chain managers must make sure all the business functions to work harmoniously together [2]. For these reasons, demand-driven inventory management plays a critical role in the retail. Demand forecasting is the period of predicting future customer demands with a certain sensitivity using statistical methods, considering data from past demands and other variable factors. Demand forecasting for many business processes is crucial to organize the resources of companies correctly and to meet customer demands in a cost-effective manner [3]. In the case of food or fashion retailing, it is important to keep demanded products in sufficient amounts in order to avoid excess inventory levels, stock-outs and transportation costs, while meeting the demands of customers at high service levels. Also, minimizing in waste for short shelf-life products is important in food retailing because of operational effectiveness. Customer requests in retail effected by many uncertain factors; promotional effects, seasonality, special days, crisis, terrorism, unexpected weather changes, social events, commercial behavior of competitors at the market and holidays etc. Since demand fluctuations caused by these factors, this situation directly affects the stock performances of the stores and therefore retailers profitability and productivity [4]. It is very important to adjust stock levels of product as necessary sizes by including them in demand forecasting models of the relevant betting factors. In addition to proper management of stocks within retailers, these policies affects their financial performance considerably. Since there is a huge competition at retail industry, most retailers are looking forward to increase their revenue and profit and also they want to decrease their costs with more effective operations [5]. Meanwhile, customers prefer retailers those offer the best price and they can easily find the product they are looking for at an acceptable distance. Due to the high competition in retailing, customers are able to meet their needs by going to another probably a competitor retail store when they can't find the product they are looking for. These dynamics make accurate demand forecasting and inventory control systems more critical for retailers [6]. Finally, our aim is to keep the shelves as full as the highest level, by minimizing disruptions, increasing operational efficiency and managing stocks at the optimum level. Our current system is planned to produce order recommendations periodically and to turn these proposals into warehouse orders after the store managers approvement. In this study, a new ensemble algorithm for retail specific demand forecasting to increase prediction accuracy has been developed. Our model works on top of our time series

978-1-5090-5795-5/17/$31.00 ©2017 IEEE

based forecasting system as an accelerator for increasing accuracy of predictions. We observed our new approach is increasing the prediction accuracy more than %15 to %20 in many product groups at production environment. In Section 2, we are shortly introducing statistical forecasting context with Time Series algorithms, then in Section 3 we are talking about ensemble approach for prediction algorithms and in Section 4 we are giving our heuristic approach for ensemble together with Time Series algorithms in retail specific conditions. At Section 5, you can find our experimental results for real life retail data at production environment.

II. Statistical Forecasting with Time Series Algorithms

In our current solution, we are already using some Time Series algorithms for making forecasting of demand for the next week in each store, product (SKU) couple. Then we are using K-Fold cross validation methodology by taking K=3 for testing data and after that, we infer best accurate one of them for the next weeks forecast. For example, setting $k = 3$ results in 3-fold cross-validation. In 2-fold cross-validation, we orderly select the dataset into two sets d_0 and d_1. We then train on d_0 and test on d_1, followed by training on d_1 and testing on d_0. When $k = n$ (the number of observations), the *k*-fold cross-validation is exactly the leave-one-out cross-validation. This was a general way for making demand forecasting in retail [7]. In time series analysis, the Box–Jenkins method, named after the statisticians George Box and Gwilym Jenkins, applies autoregressive moving average ARMA or ARIMA (ARIMA-Autoregressive Integrated Moving Average) models to find the best fit of a time-series model to past values of a time series [8]. Our current system includes about 10 different Time-Series based algorithms [7]. Regression Models, Exponential Smoothing, Holt-Winters, ARIMA are some of them. Also, products's means of the last 8 weeks are used in this study. In order to evaluate the accuracy and performance of different forecasting models, we are using general indexes for evaluation. These are Mean Absolute Percentage Error and Mean Absolute Deviation [1]. The equations for calculation of indexes are as follows:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \frac{|F_t - A_t|}{A_t} \quad (1)$$

$$MAD = \frac{1}{n}\sum_{t=1}^{n} |F_t - A_t| \quad (2)$$

where F_t is the expected or estimated value for period t and A_t is the actual value for period t, n also takes the value of the number of periods. Accuracy of our model can be calculated as $1 - MAPE$. If the values of the above two indexes are small, that means it is better for the forecasting models accuracy [9].

A. Regression Models

In this project, 3 different regression models are used. R1 shows the variance in weekly customer demand (3).

$$y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \beta_3 X_{3,t} + \beta_4 X_{4,t} + \epsilon_t \quad (3)$$

X1: Special Days (Christmas, mothers day, start of ramadan and other religious days etc.)

X2: The days of promotions

X3: The days when the product is closed for sale

X4: Outlier sales

The R2 model was created by adding weekly partial-regressive terms ($W_{j,t}; j = 1, \ldots, 52$) to the model (4).

$$y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \beta_3 X_{3,t} + \sum_{j=1}^{52} \gamma_j W_{j,t} + \epsilon_t \quad (4)$$

The R3 model is the result of adding the dummy regression terms for the R1 model for each month of the year ($M_{j,t}, j = 1, \ldots, 12$) and for the weeks of the months ($\overline{W}_{j,t}, j = 1, \ldots, 4$) (5). This model has been specially designed to be effective in products (drinks, ice cream, etc.) which are realized according to year-month of seasonality and is included in the system.

$$y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \beta_3 X_{3,t} + \sum_{j=1}^{12} \alpha_j M_{j,t} + \sum_{j=1}^{4} \gamma_j \overline{W}_{j,t} + \epsilon_t \quad (5)$$

B. Exponential Smoothing Models

The simplest of the exponentially smoothing methods is naturally called "simple exponential smoothing" (SES). This method is suitable for forecasting data with no trend or seasonal pattern [1].

$$y_t = \alpha \hat{y}_{t-1} + (1 - \alpha)\, y_{t-1} + \epsilon_t \quad (6)$$

C. Holt-Winters Models

Holt (1957) and Winters (1960) extended Holt's method to capture seasonality. The Holt-Winters seasonal method comprises the forecast equation and three smoothing equations — one for the level L_t, one for trend T_t, and one for the seasonal component denoted by S_t, with smoothing parameters α, β and γ. There are two variations to this method that differ in the nature of the seasonal component. The additive method is preferred when the seasonal variations are roughly constant through the series, while the multiplicative method is preferred when the seasonal variations are changing proportional to the level of the series [1].

$$\begin{gathered} \hat{y}_t = L_t + T_t, \\ L_t = \alpha y_t + (1 - \alpha)(L_t + T_t), \\ T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \\ S_t = \gamma(y_T - L_T) + (1 - \gamma)S_{t-s} \end{gathered} \quad (7)$$

D. ARIMA Models

The Box-Jenkins methods is the most comprehensive of the statistical methods commonly used in forecasting time series. The future value of a variable series of ARIMA, determined history values with a linear function of the random error assumption [8]. ARIMA model shown in (8).

$$\hat{Y}_t = \delta + \phi_1 Y_{t-1} + \psi_1 e_{t-1} + \cdots + \phi_p Y_{t-p} + \psi_q e_{t-q} \quad (8)$$

The variables and parameters of ARIMA:

Y_t : Dependent variable

ϕ_p : Autoregressive process-lag parameter

δ : Constant variable

Ψ_q : Moving average process parameter

e_{t-q} : Error

III. The Usage of Ensemble Learning

Ensemble learning is a learning paradigm that can significantly enhance the generalization ability of the base classifier, where a collection of a particular classifier is trained for the same task. The essential and sufficient condition for a community to perform better than its individual members is the necessity for the correctness and diversity of the basic classifiers [10]. The correct classifier is a classifier that has a better error rate than randomly estimating classes for new samples, and there are two classifiers in case they make different or irrelevant mistakes on common data samples [11]. For this reason, learning a community machine is much less costly than creating a learner with the same sensitivity. The community tendency can greatly increase the generalization ability of the machine learner, but choosing some of them for the community may show even better performance [12]. The ensemble technique, which combines the outputs of several basic classification models to create an integrated output, has become an effective classification method for many areas [13,14].

Ensemble Learning Technics are being used as combining different algorithms power into a single collaborated method philosophy. As a result, we can increase accuracy of our forecasting by means of ensemble our algorithms. Each algorithm can be sensitive under different circumstances. Ensemble them makes more powerful results for decision making process. Ensemble has two main methods at literature, these are called bagging and boosting [15]. To understand bagging, suppose that you are a patient and would like to have a diagnosis made according to your symptoms. At this situation, you may choose to ask a couple of doctors opinion instead of taking only one doctor's opinion. If most of these doctors give a certain diagnosis, you can give a decision according to the final idea or best diagnosis since its most common one among them. You can choose, the final diagnosis according to majority vote, if each doctor gets an equal vote. Intuitively, a majority vote made by a large group of doctors may be more reliable than a single doctor [15]. In boosting methodology, suppose that as a patient, you have certain symptoms again. Instead of consulting one doctor, you again talk with several doctors. Suppose you assign weights to the value or worth of each doctor's diagnosis, according to their accuracies of previous diagnoses they have made. It can easily be said that, this system learns by itself according to previous experiences. At classification and prediction problems final diagnosis then can be calculated as a combination of the weighted diagnoses. This is the main idea behind boosting methodology. Adaboost is one of the most popular boosting algorithm [15].

A. *Ensemble for Replenishment Forecasting*

Ensemble for replenishment demand forecasting needs some modifications in general ensemble strategies, according to retail specific dynamic practise. Ensemble algorithms causes more powerful results on a combined algorithm. We are proposing a new synthesized boosting approach, which takes mostly weighted average of MAPE for each week of previous 4 weeks additional to the contribution of the previous years same week and previous year following weeks trends also. This makes our forecasts more reliable against to changing trends and seasonality behaviors. We are calculating an average MAPE as defined in (1) below and we are changing weights of each algorithm according to their average MAPE for every week in each store and product (SKU) couple. According to MAPE weights our algorithms gains to vote right week by week.

$$M_{avg} = \sum_{k=1}^{n} C_k \, M_k \tag{9}$$

In (9), our constraint is $\sum_{k=1}^{n} C_k = 1$ and M_k means MAPE of the related weeks, for ex. M_1 can be MAPE of previous week and M_2 can be the MAPE of two weeks before related week. Our system can decide automatically the effect of which weeks MAPE to our average MAPE calculation according to some special weeks, which can be previously defined to our calendar by end users. For instance Christmas, mothers day effect, start of ramadan and other religious days etc. You can define special weeks such as a parameter for the system, so for those weeks system can remember the trends of previous years behaviours compared to the ordinal weeks. Eventually, our approach becomes much more sensitive for seasonality and special events. At special weeks C_k constants takes different values for making previous years trends more effective at our forecasts. You can also choose different values for C_k for different special weeks. By means of algorithms, we are not taking into all of the algorithms contribution for final decision as fully democratically. Because some algorithms may gives unreasonable or senseless results for some store, product level. It is not necessary to take their contribution to final decision, in contrast they might give negative contribution to final decision. According to our observations with different sets, we concluded to choose top first %20-%30 of the algorithms as the champions of the week for each store, product level by considering their weigths for their previous experiences. Our aim is to take the opinions of most top level ones in terms of their weights coming from their experiences. While some algorithms stays at the top level for some store, product level for some period, the others may not. Our observations showed us these kind of behaivours can change according to seasonality as well. Eventually, the system becomes an expert system learning from historical data. We are taking final decision according to their weighted contribution by taking into consideration of the participants. For instance; lets say A_1, A_2, A_3 are the algorithms which are champions of the week and their weights are w_1, w_2, w_3 respectively. Each algorithms forecasts multiplied by $(100/(w_1 + w_2 + w_3))$ and finally we take the sum of the results of each champions for the new ensembled forecast. This principle depends the rule $\sum_{k=1}^{n} w_k = 1$ constraint for each store, product couples in each week.

IV. Experimental Results

We made our experiments with weekly period and 18 months of 10 different characteristics stores real life datasets. MAPE distribution is one of the main measure for us to have an opinion for calculating accuracy of our forecasting methodology. We can observe the differences of the mape distributions of our methodology for different product groups at the figures below. Fig. 1 shows us MAPE distribution of our algorithms before implementing our ensemble approach, Fig. 2

shows after ensemble of the algorithms. Replenishing of tobacco products is very hard because of their high price. Incorrect forecasts, either too high or too low, are both economically inefficient and unprofitable. For these products, before implementing of our ensemble algorithm first quartile (first %25) of our estimates MAPE distribution is between %0-%18 and second quartile is distributed in the interval about %18-%38 and the rest %50 is between %38-%80. After implementing our ensemble approach, it becomes %0-%5 for the first quartile, %5-%20 for the second quartile and %20-%50 for the rest of %50 of the MAPE distribution. The results shows us an amazing accuracy increase for our forecasting in each product group below. On the other hand, red meat has a short life. So, it is a perishable commodity and spoils very easily. After applying this approach, forecast accuracy has been increased by %20 in this group. You can easily notice that our prediction accuracy is increasing between %15-%20 in most product groups. On the other hand, according to Fig. 3 below, solid is actual sales data, dotted line is before ensemble our methodology performance and twodash line is ensembled forecast performance. In Fig. 4, the actual sales are shown with the two winning estimating methods frequently. As you can see in this chart, "ensemble" is closer to actual sales than other methods. Given the variety of stores (some stores are located on university campuses, airports, bus terminals etc.) and products, the assumption that a single estimation method can be the best result for all series is far from reality. Instead, it is a more accurate method to assign the most appropriate estimation method to each series. You can easily see that ensemble makes our estimates more powerful, especially for unexpected trends and behaviors of the market.

A. The Study of Implementation in SOK Markets

Some of products are sold continuously by SOK Markets, so they have been sold since the beginning of the time series. Other products are sometimes removed from the portfolio due to different reasons. Different scientific demand forecasting methods are used for each SKU at each store and each product's 3-year historical data is used. Automatically calculate the coefficients of different statistical demand prediction models and perform cross validation tests. Ensemble algorithms are applied on 6 million data set (for 4000 stores and 1500 SKU) to give final decisions for next day forecasts at each night. At next stage, optimization algorithm decides the inventory levels based on inventory holding cost and shortage cost regarding customer service level which is given by supply chain managers. Producing demand forecasts periodically and turning to order is essential for the companies. In this respect, demand forecasts are confirmed by the store manager at the order day. Decisions made replenishment system are integrated with SOK Market’s ERP system. Scenerio based interactive dashboards help supply chain managers to decide the system parameters correctly. Consequently, a data-driven decision support system which is compatible with companies business strategy has been implemented.

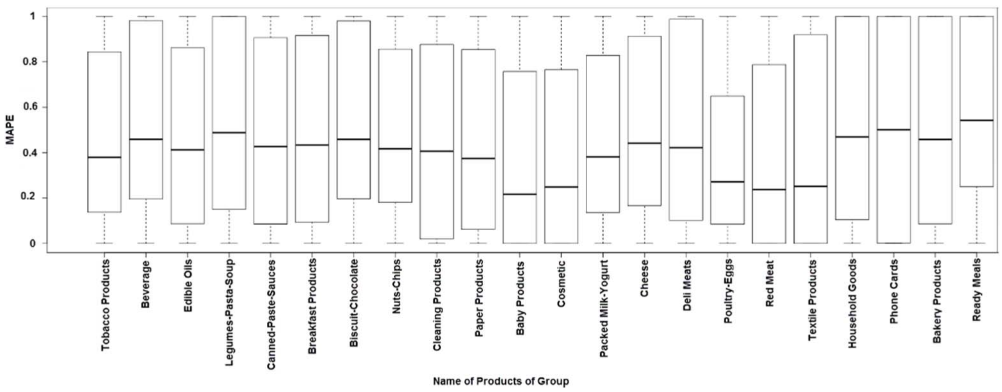

Fig. 1. MAPE distribution according to product groups before ensemble

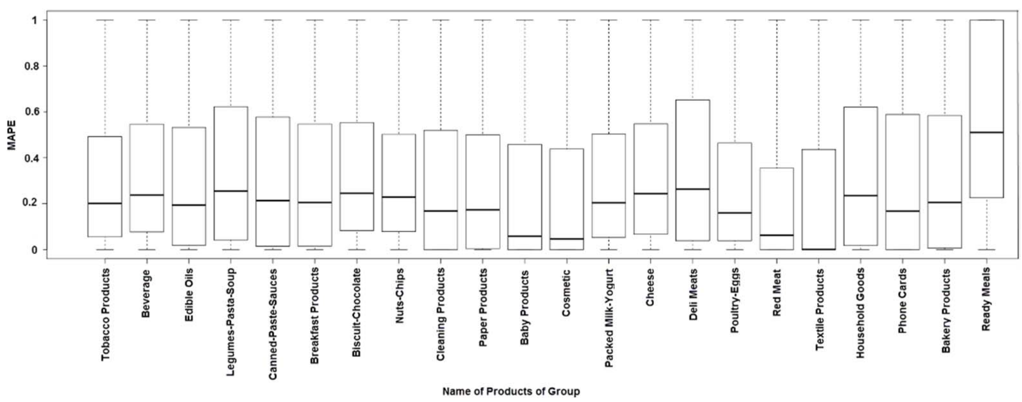

Fig. 2. MAPE distribution according to product groups after ensemble

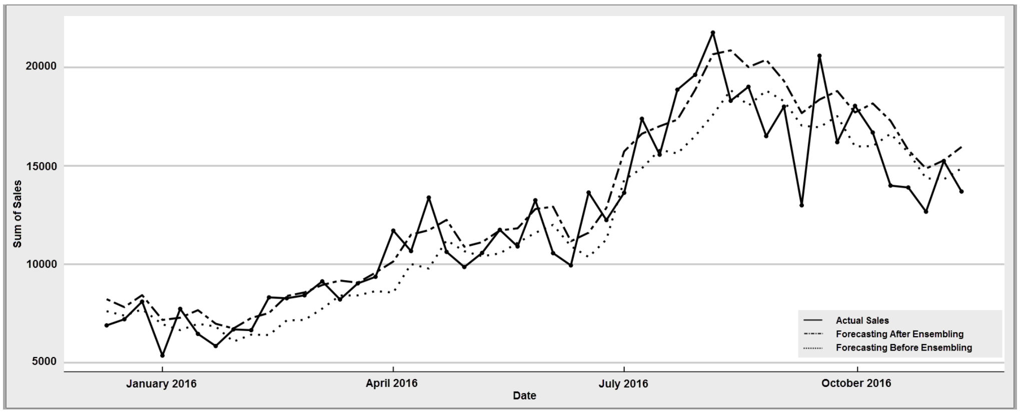

Fig. 3. Forecasting analysis example for one product group

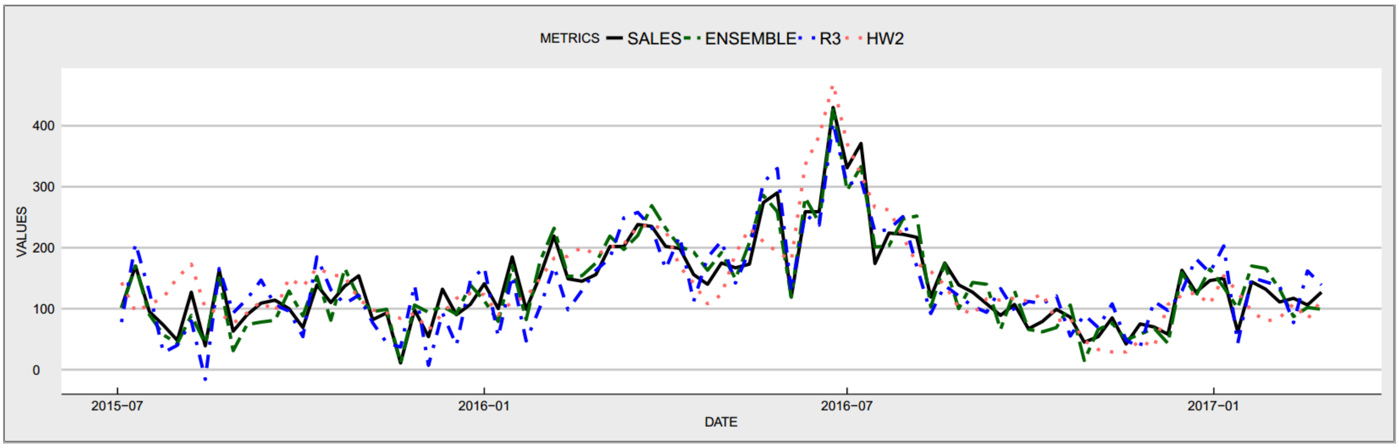

Fig. 4. Consistency Distributions of Demand Forecasting Models

V. CONCLUSION

As a result of our studies, we showed the power of ensemble time series forecasting algorithms using our new heuristic approach. Our approach makes much more reliable and more accurate forecasts according to classical ways of demand forecasting.

Also, at the end of this study, SOK Market has been benefits including: decreasing of stock-outs and stock-days by reducing the percentage of lost sales and increasing overall revenue 30%, a reduction in stock-days of 10% and a reduction in waste for perishable products of 34%.

VI. FUTURE STUDIES

In our study, we observed that if some algorithms has low weights on some store, product level so we can omit them without running them during the process. This behavior will decrease the forecasting time and computing costs. We are also planning to add some more machine learning algorithms into game to gain also their prediction power to our ensemble solution.

REFERENCES

[1] S. Chopra and P. Meindl, "Supply Chain Management: Strategy, Planning, and Operation,". Prentice Hall, 2001, pp.3-63.

[2] G. S. Kushwaha, "Operational Performance through Supply Chain Management Practices," International Journal of Business and Social Science, vol. 3, pp. 222-232, January 2012.

[3] A. A. Syntetos, Z. Babai, J. E. Boylan, S. Kolassa and K. Nikoopoulos, "Supply chain forecasting: Theory, practice, their gap and the future," European Journal of Operational Research, vol. 252, pp. 1-262, November 2015.

[4] J. C. F. Ehrenthal, D. Honhon and T. V. Woensel, "Demand Seasonality in Retail Inventory Management," European Journal of Operational Research, vol. 238, pp. 527–539, October 2014.

[5] M. J. Villena and A. A. Araneda, "Dynamics and stability in retail competition," Mathematics and Computers in Simulation, vol. 134, pp. 37-53, April 2017.

[6] P. J. McGoldrick, "Consumer misbehaviour-Promiscuity or loyalty in grocery shopping?," Journal of Retailing and Consumer Services, vol. 4, pp. 73-81, April 1997.

[7] G. Box, G. M. Jenkins and G. C. Reinsel, "Time Series Analysis: Forecasting and Control,". Wiley, 2008, pp.93-194.

[8] D. Gujarati, "Basics Econometrics,". McGraw Hill, 2003, pp.835-865.

[9] L. Yue, Y. Yafeng, G. Junjun and T. Chongli "Demand Forecasting by Using Support Vector Machine," IEEE, Third International Conference on Natural Computation (ICNC 2007), November 2007.

[10] L.K. Hansen, P. Salamon, "Neural network ensembles," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, pp.993–1001, October 1990.

[11] T.G. Dietterich, "Ensemble methods in machine learning," Proceedings of the First International Workshop on Multiple Classifier Systems, pp. 1–15, December 2000.

[12] Z. H. Zhou, J. X. Wu, W. Tang. "Ensembling Neural Networks: Many Could Be Better Than All," Artificial Intelligence, vol. 137, pp.239-263, 2002.

[13] T. Ho, J. Hull, S. Srihari, "Decision combination in multiple classifier systems", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 16, pp.66–75, January 1994.

[14] J. Kittler, "Combining classifiers: a theoretical framework", Pattern Analysis and Applications, vol. 1, pp.18-27, January 1998.

[15] J. Han, M. Kamber and J. Pei, "Data Mining : Concepts and Techniques,". USA: Morgan Kaufmann Publishers, 2012, pp.327-393.