Science Arena Publications
Specialty Journal of Knowledge Management
Available online at www.sciarena.com
2018, Vol, 3 (1): 1-6

# Online Sales Forecasting of E-Products of Online Store Using the Artificial Neural Network Approach and Customer Sentiment Analysis

**Golaleh Najar vazifehdan[1], Mahmood Alborzi[2*]**

[1] MSc, Department of Information Technology Management, Science and Research Branch, Islamic Azad university, Tehran, Iran
[2*]Associate professor,Department of Information Technology Management, Science and Research Branch, Islamic Azad university, Tehran, Iran.
[3]Assistant professor of computer engineering department,islamic Azad university central tehran branch
***Corresponding Author**

***Abstract:*** *Forecasting is one of the oldest activities and tasks of management and considered as the critical part of every business. Using the experiences and opinions of those who have already purchased a product can be useful in making the right choice. Today, customers use the Internet to buy a product they want. They can find the comments on features, strengths and weaknesses of a product in the Internet. On the other hand, the company and service centers collect comments to forecast the sale of their products. In this regard, the neural networks and data mining models are of useful tools used in developing a high-accuracy model.*
*In present research, a model is presented for forecasting whether the product is purchased by the customer. For this purpose, in the data pre-processing step, the data required are extracted in order to forecast the sales by separating the words and sentences, labeling the components of speech and rooting the words. In the next step, product sales forecasting is performed by using the neural networks and adding 4 effective indicators of advertisements, prices, discounts, free shipping. And finally, the accuracy of 83.65 is reached by boosting the neural networks using the AdaBoost algorithm.*

***Keywords:*** *Sentiment Analysis, Text Mining, Neural Networks, Forecasting*

## INTRODUCTION

Today, the sale is the most important part of a business. Sales forecasting is considered as a key part of financial planning of a business and is a self-assessment tool that uses past and present sales statistics to forecast future sales in a smart way. With an accurate sales forecasting, one can plan for the future accurately. If sales forecasting shows that 30% annual sales are associated to March, production must be raised in this month in order to meet the high demand. A simple sales forecasting can inform the organization about other aspects of its business. On the other hand, this forecasting ca be an important part of a new business. Almost all new businesses require loans or initial capital to provide workplace, inventory, equipment, employee salaries, and marketing. Provision of this capital requires the presentation of figures reflecting durability of the business. In order to obtain competitive advantage in a constantly changing competitive market, the management should make the right decision at the moment regarding the available information. Therefore, making an accurate decision at the best time plays a critical role in gaining

competitive advantage and past data can directly provide the possibility of fair estimate for predictive models. If the trading unit can accurately estimate the amount of sales for the next period, the inventory control unit can effectively control the inventory and achieve the JIT (Time in Just) approach. Additionally, the production unit can correctly plan the activities of the plant. Using such operations, the cost of production can be reduced (XUO & KUO, 2008). Therefore, achieving an accurate forecast seems to be very sensitive and imperative. For today's manufacturers, the key strategy is to manage supply chain efficiently and understand customer demand better (Chong & Zho, 2014). Information technology significantly helps manufacturers improve supply chain management during the year. Using some applications such as organization resource planning, B2B (Business-to-Business) websites, Radio Frequency Identification (RFI) systems for organizing and enhancing supply chain management, and bullwhip effect, can be effective in reducing supply chain management challenges.

One of the important aspects of effective supply chain management is the better sales forecasting so that manufacturers do not want to purchase more or lower than the production line (Huang and Handfield, 2015). Big data and user-generated content are emerging phenomena in the sales forecasting. Recent marketing reports have indicated the impact of the user-generated content. Nielsen (2012), in his report, showed that online consumer surveys are the second most trusted form of surveys. This claim has been investigated by 70% of customers who have trusted this platform. Cao and Colleagues (2009), in their research, have indicated that neural networks have the ability to eliminate the non-linearity of financial data and can provide good predictions about nonlinear complex models. Due to the current development of competitive space, marketing, as a dynamic science fits the world of communications, including the Internet, has changed. Nowadays, researchers find that the unconscious mind of human has a greater impact on his life and choices than his conscious mind. Smart marketers use neuroscience as a new tool in market research and concentrate on the stimulation of the old brain in order to attract more people to their websites and lead more of them into purchasing channels (Alin, 2011). Shojaeiyan et al. (2012,) in their research, showed that the neural network has better predictive power than the ARIMA model. Also, Khan and colleagues (2008) compared the recurrent neural network based on the genetic algorithm and the recurrent neural network on the comparison of stock price changes, and concluded that the recurrent neural network based on the genetic algorithm significantly performs better than the rival algorithm.

Given the user-generated content plays an important role in influencing purchasing decisions and this role becomes more important in helping the organization to understand and forecast product demand. In an e-commerce environment, with an easy access to information, consumers can be easily influenced by it in making their decisions. Markets and e-commerce stores have now become of the main channels of consumer purchasing. Forecasting a potential customer purchasing decision and his/her purchasing performance are currently critical to a company's supply chain management. Now, it is possible to forecast the sales of products and customer demands based on online data on customers' decisions in real time and using various online information in an e-commerce environment (Duan et al., 2008). Therefore, the present study aimed to design a model based on customer sentiment analysis and using artificial neural networks to forecast the sales of a number of e-products on the e-commerce site.

## Method

The present research is a descriptive study in terms of the method and it is an applied study in terms of objective. In present study, library research method was used to collect the data and various sources including specialized articles, Persian and English books, existing related research and theses, searching related topics on websites and historical data in the database of the e-commerce website, were used. In present study, Cross-industry standard process (CRISP) for data mining and programming language of Python were used.

## Research findings

Since it is necessary to obtain the proper knowledge of research data and how to prepare them before applying different models and algorithms, the properties of the data and their preparation steps are discussed in this part. The raw data used in present study were collected from the database of Timche website including 1149 variables related to 726 customers. The number of variables has been reduced with respect to the problem. After data cleaning, the number of customers was 1000. Due to the large amount of data and resources and time limitations, the data has been examined using the various methods and 726 products have been used for modeling and processing during the period studied. The tables, information and comments were stored and collected separately in Excel Software. These tables were merged in Access Software and finally, entered RapidMiner Software for analysis.

### General Data Status

The statistical tests used to analyze the data obtained from a small group (sample) and its generalization to the target population are divided into parametric and nonparametric categories according to the scale of the variables measured. Parametric tests are used to analyze the date on the ratio and interval scales that their least statistical indexes are mean and variance, while non-parametric tests are used to analyze the date on the nominal and ordinal scales and their statistical index are median and mode.

| Variable | Description | Unit | Domain | Min | Max | Mean | Standard deviation | Skewness | Uniqueness | Validity |
|---|---|---|---|---|---|---|---|---|---|---|
| **class** | Classification field which is used for forecasting. | Continuous | (-1,1) | -1 | 1 | 0.238 | 0.972 | -0.490 | -- | 724 |
| **brand** | Brand name | Categorization | - | -- | -- | -- | -- | -- | 14 | 725 |
| **model** | Device model | Categorization | - | -- | -- | -- | -- | -- | 125 | 725 |
| **com** | Customer's comment | Text | - | -- | -- | -- | -- | -- | -- | 724 |
| **price** | Price | Continuous | - | -- | -- | -- | -- | -- | 95 | 700 |
| **number** | Number of comment | Continuous | ( -5,2646 ) | 5 | 2646 | 318.992 | 540.015 | 2.522 | -- | 725 |
| | It has discount | Continuous | (0,1) | 0 | 1 | 0.087 | 0.282 | 2.939 | -- | 725 |
| **freeship** | Shipping is free. | Continuous | (0,1) | 0 | 1 | 0.212 | 0.409 | 1.409 | -- | 725 |
| **advertisements** | Advertisement of the product | Continuous | (-1,1) | 1- | 1 | 0.801 | 0.599 | -2.685 | -- | 725 |

In this part, data mining process is carried out. In this part, we examine the users' comments presented on Timche website.

| |
|---|
| Considering its price, its facilities are well, I did not think it would be so good. |
| I bought it for my father, It's very good. One can easily do his doings using it. It supports 4G which is one of its advantages. |
| Its body is made by poor plastic. It is thin and its body has a good design but low resistance. |
| It has very low internal memory. Its camera has poor performance in the places where there is no light. |
| I bought two of this model with discount. This model is beautiful and has very good facilities and its price is appropriate. |
| Its camera is poor. The photos taken by it has low quality and they are vivid and the videos has low quality as much as vga. |
| I bought it yesterday, this cellphone is great, it's affordable. The LG is a trusted brand. |
| Nowadays, supporting just one SIM card is a big disadvantage. |

**Customers' comments**

AdaBoost algorithm was used to boost neural networks.

**Decision tree accuracy**

After running, it is 76.45%.

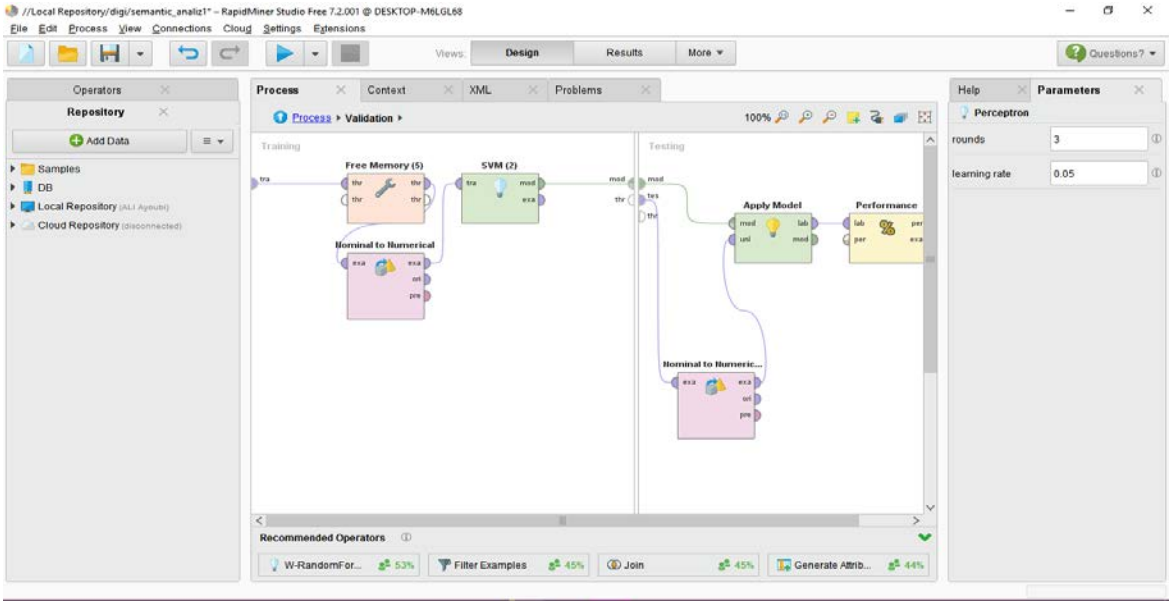
**Perceptron Neural Network**

The prediction is performed by the perceptron neural network.


AUC: 0.766 +/- 0.012 (micro: 0.766) (positive class: 1)

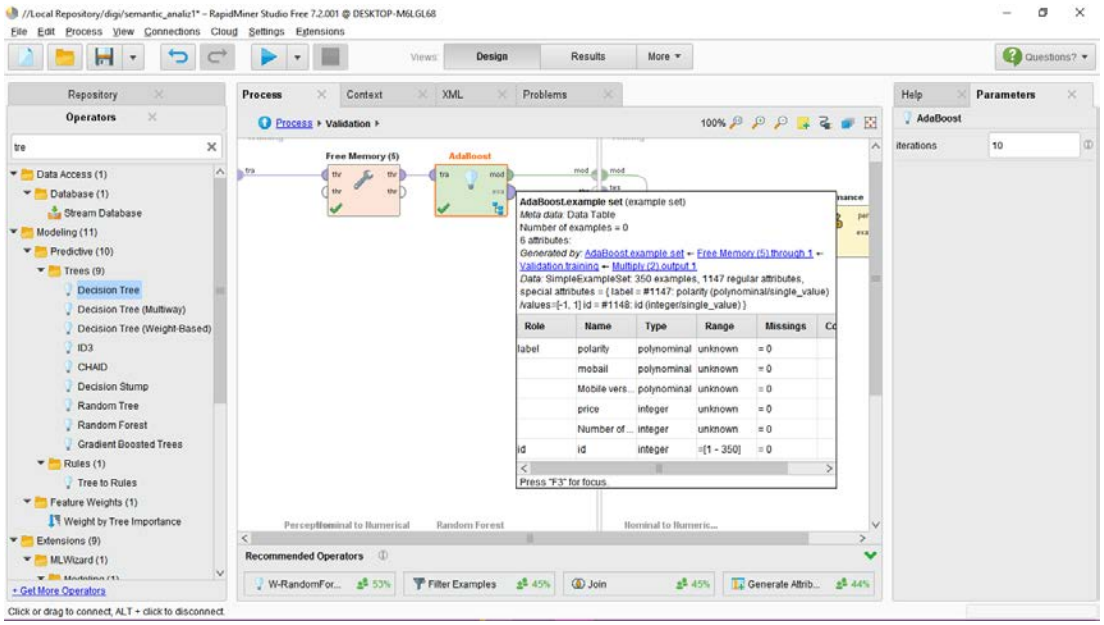
**Decision Support Vector**

The support vector operator was put in the following model and following accuracy was obtained.

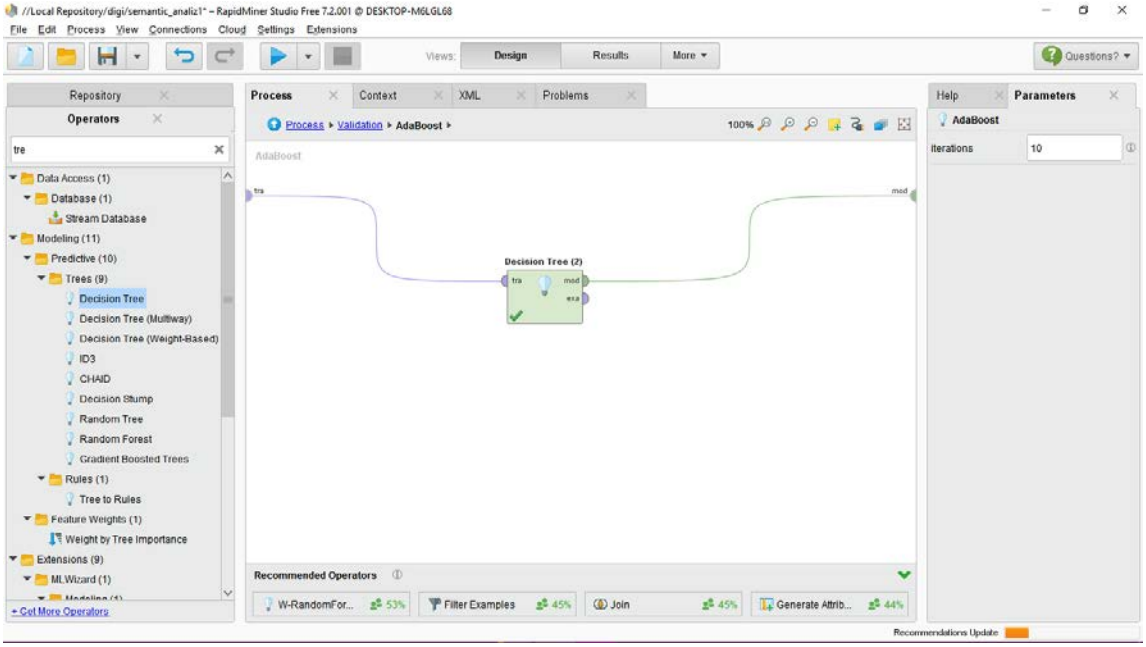


**Boosting Neural Networks using the AdaBoost algorithm**

This method is run in combination with the decision tree and the data are fully examined.




The decision tree is put within this method.




The AdaBoost method is a hybrid method. using this method, several models are produced and the answer to the sample is obtained by averaging the answers of the models produced. This method is more accurate than other methods, followed by the decision.

**Discussion and conclusion**

Based on the results of present study and using above-mentioned comments, one can forecast the sales of e-products using the AdaBoost algorithm and the neural network with an accuracy of about 83.65%. Also, using

the information obtained from the customer, the customer purchasing probability can be estimated. And if this probability is acceptable, some recommendations can be offered to customers to increase their purchasing likelihood and a discount coupon can be made and sent to the customer to attract him/her. In present study, data mining methods were used to increase the accuracy of the model. In the final part, using AdaBoost algorithm along with a neural network and decision tree provided results. Since the neural network is a black box, it cannot be examined to obtain such results, so the results were evaluated to compare the algorithms and it was not feasible to test the model. Also, in this matrix, the confusion of the results of the AdaBoost algorithm shows that this algorithm has some errors in both types of class in a balanced way. That is, both classes have errors, and all errors are not related to just one class. This problem can be seen on the Bayesian network and support vector machine algorithm, that is, all errors are related to one class. Such result shows that the algorithm has not been able to develop an appropriate model for the data.

Direct results of the present research

Overall, the results of the present study can be summarized as follows:

- Predicting customer purchasing using the user's opinion and product information
- Increasing modeling accuracy using Persian texts preprocessing
- Modeling a neural network in order to evaluate the accuracy
- Using decision tree and comparing the methods with each other
- Using the meta-algorithm approach to increase the modeling accuracy
- Using this model, some recommendations can be offered to potential users to buy products.

## References


1. Shojaeiyan, S. F., Abunoori, A.A., and Farrokhi, F. (2012). A Comparative study of Artificial Neural Network Performance and Autoregressive Stacked Moving Average Model in Modeling and Short-Term Forecasting of Exchange Rate in Iran. Master's thesis, Economic Sciences, Tehran: Islamic Azad University, Central Branch.
2. Nadjarzadeh Sharifabad, A. (2014). Sales Forecasting using Neural Network, First Conference on Economics and Applied Management with a National Approach, Babolsar, Shomal Taroud Research Company.
3. Najib Moghadam, S., Asghari Aq Qala, E. (2015). A Conceptual Model of the Relationship between Neural marketing, Branding and Profit: A Study of the Relationship between Neural Marketing, Brand Passion and Profit by Designing the Customers' Mental Map with Structural Equation Modeling Approach, National Conference on Future Studies, Humanities and Development.
4. Chong, A.Y.-L. and Zhou, L. (2014), "Demand chain management: relationships between external antecedents, web-based integration and service innovation performance", International Journal of Production Economics, Vol. science/journal/09255273/154
5. Chong, A.Y.-L. and Zhou, L. (2014), "Demand chain management: relationships between external antecedents, web-based integration and service innovation performance", International Journal of Production Economics, Vol. science/journal/09255273/154
6. Duan, W., GU, B. and Whinston, A. (2008), "The dynamics of online word-of-mouth and product sales – an empirical investigation of the movie industry", Journal of Retailing, Vol. 84 No. 2, pp. 233-242
7. Luca, M. (2011), "Reviews, reputation, and revenue: the case of Yelp.com", working paper,Harvard Business School, Harvard University, Boston, MA, September 16
8. Connor, N. O. M. G., (2006). A neural network approach to predicting stock exchange movements using external factors. *Journal of Knowledge-Based Systems Elsevier*, 3(8), pp. 87-94