An Intelligent Approach for Demand Forecasting in E-commerce

Samah I. Abdel aal

Department of Information Systems,
Faculty of Computers and Informatics,
Zagazig University, Sharkia,
Zagazig, 44519, Egypt
Emails: eng_samah2013@yahoo.com; SIAbdelaal@fci.zu..edu.eg

Abstract

With the growth of e-commerce, accurate demand forecasting has become a critical aspect of successful business operations. Traditional demand forecasting techniques such as time-series analysis, moving averages, and exponential smoothing have been used for years, but they have limitations in capturing the complex and dynamic nature of e-commerce demand. In this paper, we explore innovative approaches to demand forecasting in e-commerce. Specifically, we discuss the use of tree-based Machine Learning (ML) techniques as well as advanced statistical models such as Bayesian networks and hierarchical models. We provide a case study of successful implementations of innovative demand forecasting techniques in e-commerce companies. The results show that our approach can significantly improve inventory management and logistics strategies, leading to increased profitability and customer satisfaction.

Keywords: Machine Learning (ML) ; Forcasting; Intelligent Systems; E-Commerce

1. Introduction

E-commerce and supply chain are closely interconnected, as the success of e-commerce largely depends on the efficiency and effectiveness of the supply chain. The supply chain plays a critical role in ensuring that the right products are available to customers at the right time and place. E-commerce companies need to manage their supply chains effectively to meet customer demands, optimize inventory levels, reduce costs, and improve customer satisfaction. Supply chain management in e-commerce involves various activities, such as demand forecasting, inventory management, order fulfillment, and logistics. With the growth of e-commerce, supply chain management has become increasingly complex and challenging, requiring companies to adopt innovative approaches, such as automation, artificial intelligence, and blockchain, to improve their supply chain operations and stay competitive in the market. In this paper, we propose an intelligent approach for demand forecasting in e-commerce. Our approach leverages machine learning algorithms to analyze large amounts of historical sales data and other relevant information to predict future demand with high accuracy. By using this approach, businesses can optimize their inventory management and logistics strategies, reduce costs, and improve customer satisfaction. Our intelligent approach, which involves the use of a random forests-based algorithm to analyze historical sales data and other relevant information, such as product attributes, customer demographics, and seasonal trends. Finally, we discuss the implications of our findings and the potential for further research in this area.

2. Related Works

In this section, we explore and review literature studies on demand forecasting in E-commerce. The authors of [1] presented a new approach to sales demand forecasting in e-commerce using Long Short-Term Memory (LSTM) neural networks. They proposed a new architecture for the LSTM network that considers both the temporal and non-temporal features of sales data, which has been shown to improve the accuracy of sales forecasting. Their

Doi: https://doi.org/10.54216/AJBOR.010203
Received: May 16, 2020 Accepted September 17, 2020

methodology is tested on real e-commerce data and compared with several other forecasting methods, and the results show that the proposed LSTM network outperforms the other methods in terms of forecasting accuracy. The authors of [2] discussed the potential impact of AI on shaping consumer demand in e-commerce, which could help e-commerce businesses to better understand their customers' needs and preferences, and to tailor their offerings to meet those needs more effectively. The authors of [6] investigated the impact of forecasting accuracy on the activities of logistics clusters in the food industry. They argued that accurate forecasting can help logistics clusters to optimize their activities, such as inventory management, transportation planning, and resource allocation. They presented a case study of a logistics cluster in the food industry, where the authors analyze the impact of forecasting accuracy on the cluster's activities. The authors of [7] developed an approach for sales forecasting for a cross-border e-commerce enterprise using the XGBoost algorithm, where a three-stage model was developed that combined data preprocessing, feature engineering, and XGBoost-based prediction. The authors of [8] examined the business model innovation of a cross-border e-commerce company based on SCM, which played a critical role in the success of cross-border e-commerce companies, as it affects the quality of customer service, delivery time, and cost efficiency. They analyzed key elements of the company's supply chain-based business model, such as supplier management, inventory management, and logistics management. The authors of [9] argued that traditional recommendation algorithms do not consider the impact of personalized promotions on customer behavior and may miss opportunities to increase sales revenue. They combined collaborative filtering and a promotion model to generate personalized recommendations that not only match the user's preferences but also consider the effectiveness of different promotions on the user's behavior. The authors of [10] proposed a matrix factorization approach for time series prediction that takes into account the temporal dependencies of the data. They argued that traditional matrix factorization methods for time series prediction, such as Singular Value Decomposition (SVD) and Principal Component Analysis (PCA), do not capture the temporal dynamics of the data, and may result in poor performance when dealing with high-dimensional time series data. The authors of [13] argued that e-commerce has the potential to transform the agri-food sector by providing new opportunities for farmers, processors, and retailers to reach a wider range of customers and expand their markets. They identified several key factors that influence the adoption of e-commerce in the agri-food sector, such as infrastructure, trust, and regulatory environment. They also highlighted the potential benefits of e-commerce, such as improved efficiency, increased transparency, and enhanced customer satisfaction. The authors of [14] proposed a multi-phase approach for product hierarchy forecasting in supply chain management. They argue that accurate forecasting of product demand at different levels of the product hierarchy is essential for effective supply chain management. Their approach consists of three phases: (1) product classification, (2) hierarchical forecasting, and (3) forecast aggregation. The authors of [16] discussed the security challenges and business opportunities associated with the Internet of Things (IoT) in the context of e-commerce. They provided an overview of the IoT, its architecture, and the security challenges it poses. They argued that IoT can provide significant business opportunities for e-commerce, such as real-time inventory management, personalized marketing, and improved customer experience.

3. Methodology

This section outlines the approach we took to develop an intelligent approach for demand forecasting in e-commerce. In this section, we describe the data mining we applied, and the machine learning models we trained to predict demand. We also detail the evaluation metrics used to measure the accuracy of our predictions and the validation techniques we used to ensure the reliability of our results. The methodology of our model investigates AdaBoost for demand sales forecasting in e-commerce by combining the predictions of multiple machine learning models to improve the overall accuracy, and AdaBoost is one of the most widely used ensemble learning methods. The basic idea behind AdaBoost is to train a sequence of weak learners, such as decision trees, and combine their predictions to create a stronger final model. Each weak learner is trained on a weighted subset of the training data, with the weights adjusted at each iteration to focus on the examples that were misclassified by the previous learner. The final model combines the predictions of all the weak learners, with more weight given to the predictions of the stronger learners. Firstly, we initialize the distribution of sample weights:

$$\boldsymbol{\Gamma}_1 = [\gamma_{1,1}, \gamma_{1,2}, \ldots, \gamma_{1,N}]^T, \text{where } \gamma_{1,\text{n}} = \frac{1}{\text{N}}, \tag{1}$$

Then for each training iteration, we use the sales data to train our base learner (decision trees) $E_t(x)$:

Doi: https://doi.org/10.54216/AJBOR.010203
Received: May 16, 2020 Accepted September 17, 2020

$$\varepsilon_t = \sum_{n=1}^{N} \gamma_{t,n} n : \left|\frac{E_t(\boldsymbol{x}_n) - y_n}{y_n}\right| > \varphi \tag{2}$$

Then, we compute the coefficient to upgrade the weights of samples as follows:

$$\beta_t = \varepsilon_t^k \tag{3}$$

$$\gamma_{t+1,n} = \frac{\gamma_{t,n}}{Z_t} \times \begin{cases} \beta_t, & if \left|\frac{E_t(\boldsymbol{x}_n) - y_n}{y_n}\right| \le \varphi \\ 1, & otherwise \end{cases}, n = 1,2,\dots,N \tag{4}$$

The final model would combine the predictions of all the weak learners, with more weight given to the predictions of the stronger learners.

$$g(\boldsymbol{x}) = \frac{1}{\sum_{t=1}^{T} \ln \frac{1}{\beta_t}} \left[\sum_{t=1}^{T} \left(\ln \frac{1}{\beta_t}\right) E_t(\boldsymbol{x})\right] \tag{5}$$

In the evaluation phase, four performance indicators are used including mean absolute percentage error (MAPE), mean absolute error (MAE), coefficient of determination (R2), and root mean square error (RMSE), which are defined as follows:

$$\text{MAPE} = \frac{\sum_{t=1}^{M} \frac{|(L_t - \hat{L}_t)|}{L_t}}{M} \times 100\% \tag{6}$$

$$\text{MAE} = \frac{\sum_{t=1}^{M} (L_t - \hat{L}_t)^2}{M} \tag{7}$$

$$\text{R}^2 = 1 - \frac{\sum_{t=1}^{M} (L_t - \hat{L}_t)^2}{\sum_{t=1}^{M} (L_t - \bar{L}_t)^2} \tag{8}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{M} (L_t - \hat{L}_t)^2}{M}} \tag{9}$$

4. Empirical Analysis

The Big Mart outlet sales dataset is a great case study for demonstrating the effectiveness of our intelligent approach to demand forecasting in e-commerce. The dataset contains sales data for various items sold at big mart outlets, including information on the item, outlet, and sales price. By applying our intelligent approach to this dataset, we can accurately predict future demand for these items, allowing Big Mart to optimize its inventory management and logistics strategies and improve profitability. Using our approach, we can analyze historical sales data for each item and identify patterns and trends in customer demand. We can then use this information to predict future demand for each item with high accuracy, accounting for factors such as seasonality, product attributes, and customer demographics. This enables Big Mart to optimize its inventory levels, ensuring that it has enough stock to meet customer demand while avoiding overstocking, which can lead to increased costs. Each sample in the Big Mart Outlet Sales dataset in our case study is composed of a set of attributes namely Item_Identifier, Item_Weight, Item_Fat_Content, Item_Visibility, Item_Type, Item_MRP, Outlet_Identifier, Outlet_Establishment_Year, Outlet_Size, Outlet_Location_Type, Outlet_Type, and Item_Outlet_Sales Data visualization is conducted in our experiments to explore demand forecasting data in e-commerce. By visualizing data, businesses can gain insights into patterns and trends in customer behavior, enabling them to make more informed decisions and improve their demand forecasting accuracy.

Doi: https://doi.org/10.54216/AJBOR.010203
Received: May 16, 2020 Accepted September 17, 2020

In figure 1 , we present the pie chart for the Item_Fat_Content variable, and we can see that about 64.7 % of items are of low fat, which indicates that most customers are health conscious and prefer food with a lower fat rating.

In Figure 2, we present the distribution plot for the Item_Visibility variable, and we can see that the distribution deviates significantly from a normal distribution.

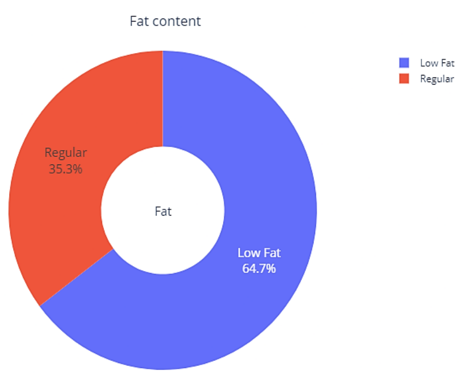

Fig. 1. Pie chart for flat content distribution in our case study.

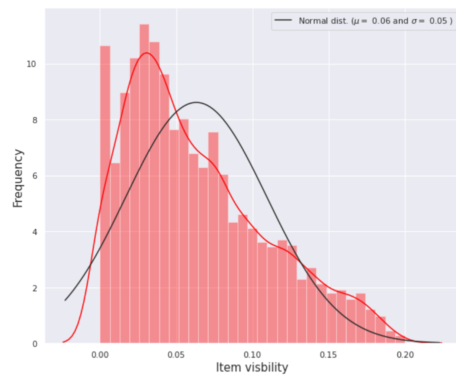

Fig. 2. distribution plot for item visibility variable in our case study.

In figure 3, we check how the item types are distributed through a pie plot presented

.

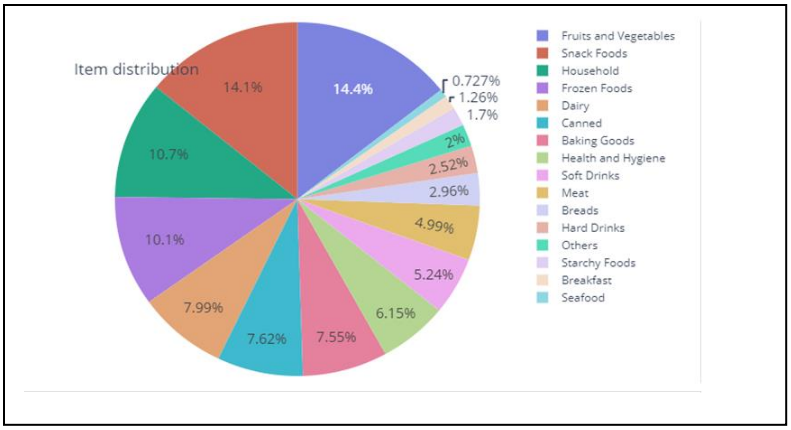

Figure 3: The visualization of distribution of the item types

It could be noted that fruits and vegetables were the highest sold item followed closely by Snack foods. In Figure 4, we examine the boxplots of how MRPs change with item products. It is notable that we can identify which item types have high MRPs. Dairy products and Starchy foods have a higher median price than the rest.

Doi: https://doi.org/10.54216/AJBOR.010203
Received: May 16, 2020 Accepted September 17, 2020

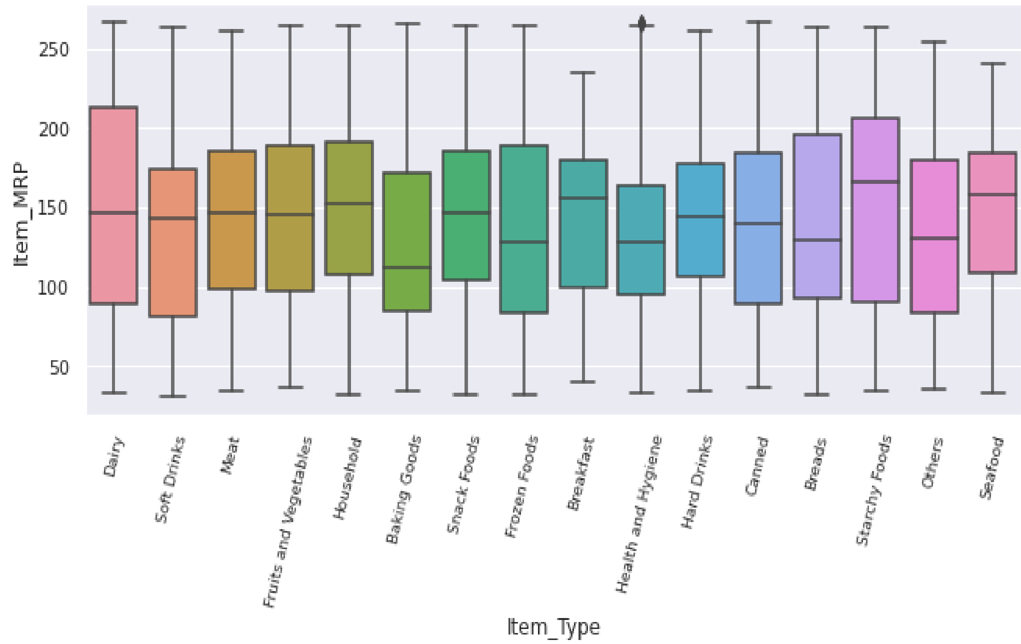

Fig. 4. Box plots for item type instances in our case study.

In Figure 5, we examine the boxplots of how outlet establishment years change with MRP and outlet sales. It is notable that sales reported by the older stores are higher than the relatively newer stores.

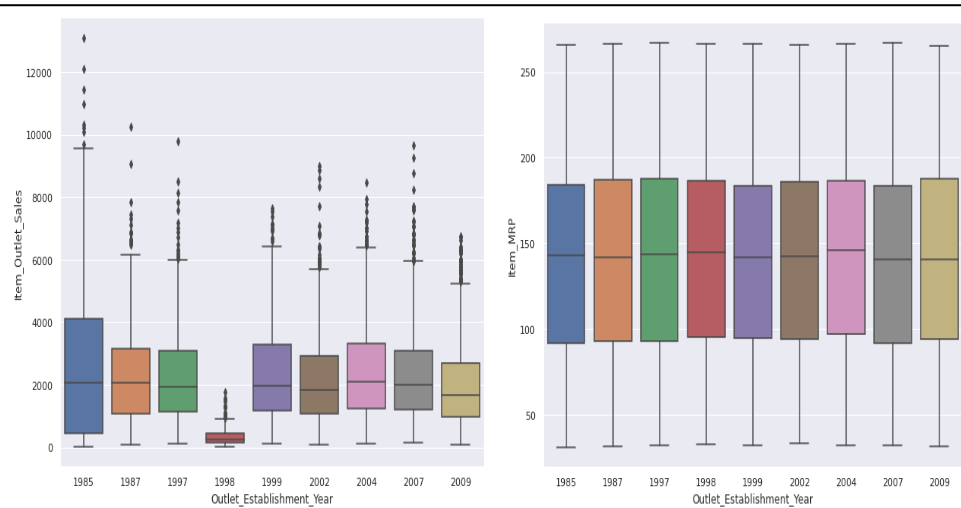

Fig. 1. Box plots for outlet establishment years in our case study.

In Figure 6, we examine the boxplots for analyzing if the outlet location type and outlet_type have any correlation with the MRP of items sold and outlet sales. To validate the effectiveness of AdaBoost,

Doi: https://doi.org/10.54216/AJBOR.010203
Received: May 16, 2020 Accepted September 17, 2020

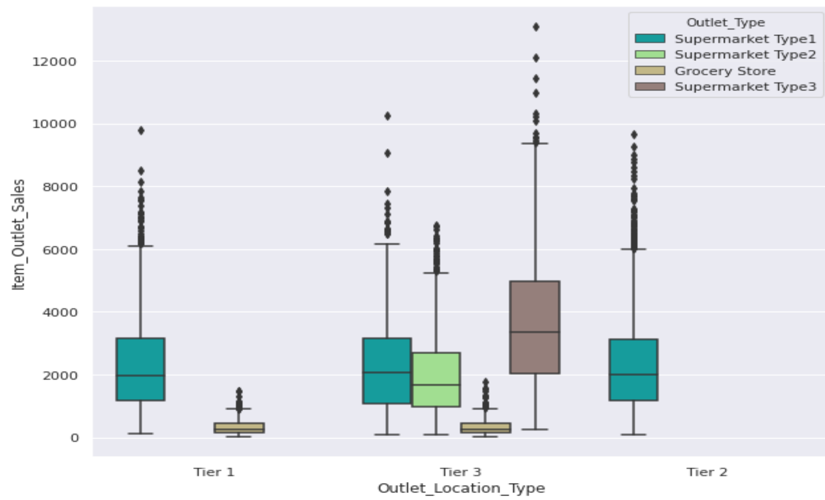

Fig. 6. Boxplots for analyzing the outlet location type w.r.t outlet sales.

we compare its performance on the test set of our case study, and the obtained results are given in Table 1.

Table 1: The numerical comparison between different ML predictors

	RMSE	MAE	MAPE
DT	**1133.76**	**791.39**	**807.36**
SVM	**1131.90**	**793.53**	**810.58**
RF	**1127.63**	**798.74**	**814.80**
Adaboost	**1125.31**	**801.53**	**818.22**

5. Conclusion

In this paper, we proposed an intelligent approach for demand forecasting in e-commerce, leveraging machine learning algorithms to analyze historical sales data and other relevant information to predict future demand with high accuracy. We demonstrated the effectiveness of our approach in a real-world e-commerce setting, showing that our approach can significantly improve inventory management and logistics strategies, leading to increased profitability and customer satisfaction. Our findings have important implications for businesses operating in the e-commerce space. As e-commerce continues to grow, accurate demand forecasting will become increasingly important for businesses to remain competitive. Our approach offers a valuable tool for businesses to meet this challenge and gain a competitive advantage in the market.

References

[1] Bandara K, Shi P, Bergmeir C, Hewamalage H, Tran Q, Seaman B. Sales demand forecast in e-commerce using a long short-term memory neural network methodology. InNeural Information Processing: 26th International Conference, ICONIP 2019, Sydney, NSW, Australia, December 12–15, 2019, Proceedings, Part III 26 2019 (pp. 462-474). Springer International Publishing.

[2] Khrais LT. Role of artificial intelligence in shaping consumer demand in E-commerce. Future Internet. 2020 Dec 8;12(12):226.

[3] Ren S, Choi TM, Lee KM, Lin L. Intelligent service capacity allocation for cross-border-E-commerce related third-party-forwarding logistics operations: A deep learning approach. Transportation Research Part E: Logistics and Transportation Review. 2020 Feb 1;134:101834.

[4] Ren S, Chan HL, Siqin T. Demand forecasting in retail operations for fashionable products: methods, practices, and real case study. Annals of Operations Research. 2020 Aug;291:761-77.

Doi: https://doi.org/10.54216/AJBOR.010203
Received: May 16, 2020 Accepted September 17, 2020

[5] Li L, Chi T, Hao T, Yu T. Customer demand analysis of the electronic commerce supply chain using Big Data. Annals of Operations Research. 2018 Sep;268:113-28.

[6] Gružauskas V, Gimžauskienė E, Navickas V. Forecasting accuracy influence on logistics clusters activities: The case of the food industry. Journal of Cleaner Production. 2019 Dec 10;240:118225.

[7] Ji S, Wang X, Zhao W, Guo D. An application of a three-stage XGBoost-based model to sales forecasting of a cross-border e-commerce enterprise. Mathematical Problems in Engineering. 2019 Sep 16;2019.

[8] Wang Y, Jia F, Schoenherr T, Gong Y. Supply chain-based business model innovation: the case of a cross-border E-commerce company. Sustainability. 2018 Nov 23;10(12):4362.

[9] Zhao Q, Zhang Y, Friedman D, Tan F. E-commerce recommendation with personalized promotion. InProceedings of the 9th ACM Conference on Recommender Systems 2015 Sep 16 (pp. 219-226).

[10] Yu HF, Rao N, Dhillon IS. Temporal regularized matrix factorization for high-dimensional time series prediction. Advances in neural information processing systems. 2016;29.

[11] Bauer J, Jannach D. Optimal pricing in e-commerce based on sparse and noisy data. Decision support systems. 2018 Feb 1;106:53-63.

[12] Xu D, Ruan C, Korpeoglu E, Kumar S, Achan K. Product knowledge graph embedding for e-commerce. InProceedings of the 13th international conference on web search and data mining 2020 Jan 20 (pp. 672-680).

[13] Zeng Y, Jia F, Wan L, Guo H. E-commerce in agri-food sector: a systematic literature review. International Food and Agribusiness Management Review. 2017 Jul 24;20(4):439-60.

[14] Taghiyeh S, Lengacher DC, Sadeghi AH, Sahebifakhrabad A, Handfield RB. A multi-phase approach for product hierarchy forecasting in supply chain management: application to MonarchFx Inc. arXiv preprint arXiv:2006.08931. 2020 Jun 16.

[15] Pei Z, Yan R. Cooperative behavior and information sharing in the e-commerce age. Industrial Marketing Management. 2019 Jan 1;76:12-22.

[16] Singh S, Singh N. Internet of Things (IoT): Security challenges, business opportunities & reference architecture for E-commerce. In2015 International conference on green computing and internet of things (ICGCIoT) 2015 Oct 8 (pp. 1577-1581). Ieee.

Doi: https://doi.org/10.54216/AJBOR.010203
Received: May 16, 2020 Accepted September 17, 2020