NanoWorld Journal

Research Article   **Open Access**

Predicting Profit Ratio for E-Commerce

Amirta V* and Nishigandha Lad

Symbiosis Statistical Institute, Symbiosis International (Deemed University)

*Correspondence to:
Amirta V
Symbiosis Statistical Institute,
Symbiosis International (Deemed University).
E-mail: amritavenkatraman8@gmail.com

Abstract

Sustaining for a longer amount of time is essential for any business in the fiercely competitive E-commerce market. The period of sustainability of the company is directly related to the profitmade by the company. Predicting the profit ratio helps the company to stay one step ahead inthe competition. In this paper three predictive models Multiple, Ridge and Elastic net regression models are presented and compared to choose the best suitable model for predicting the Profit Ratio. The Ridge regression model proves to be the best with 0.8313 R^2 value among proposed models and yields best results.

Keywords

E-Commerce, Profit ratio, Supply chain, Predictive models

Introduction

In the modern world, the online market is flooded with customers since it provides easy access, lower purchase costs, doorstep and easy delivery and heavy discounts on retail purchases. Businesses use internet markets to reach out to their worldwide consumers. B2B (business-to- business) and B2C (business-to-consumers) markets save their time and costs at both ends by using the medium of online markets. Study by Pidada [1], discusses the effectiveness of online marketing in B2B and B2C market. These markets and services intend to expand their business and surge their profit by exhaustive use of efficient supply chains.

The supply chain is an inevitable aspect for any business to flourish. It helps businesses to achieve several business objectives such as improving customer satisfaction, lower operating and production costs, optimized usage of raw material, etc. Studying profit is one of the aspects to know the profile of the business and gives a glimpse on effectiveness of a B2C supply chain inthe online shopping market. But the ultimate goal of a supply chain is elevated profit. Profit shows a rising graph when supply chain management is efficient and achieves all the goals for the growth of a business.

Predicting profit ratio helps top management of the supply chain companies to take strategic decisions, hence it is taken as the research objective. This paper studies the effectiveness of aB2C supply chain in the online shopping market by predicting the profit ratio of an online shopping portal by building machine learning models. Profitability ratios indicate how efficientlya company generates profit [2].

Literature Review

An integrated review technique is followed in this research work. Aforesaid, the paper discusses performance of profit of a shopping website by predicting

the profit ratio. Study by Ishfaq et al. [3] talks about profit margins, costs of multichannel retailers and the profitability by fulfilling online orders which is affected by several aspects of the supply chain. Minimized operating costs has been the primary objective in most of the supply chain network design models. These models typically desire to satisfy every customer demand. Furthermore, if the company is facing com- petition, sometimes it might be more profitable to lose some potential customers to competitors since the cost of maintaining these customers can be prohibitively high. For example, providing heavy discounts can be harmful for the retailer, free and fast deliveries can prove to be a heavy toll on the business.

Most supply chain related costs, such as location, transportation, and inventory costs of an item depend on total demand volume, and no clear method exists for determining a customer's profitability a priori [4]. Authors of [1] show how multiformity in consumers' purchasing attributes lead to important profit intimation. The study also discusses the varied aspects of online business and use of supply chain in product selection, quantity of orders, shipment, delivery, etc. The gap identified from literature review is the comparison of multiple predictive models for predicting the profits for improvement of the supply chain process. It is necessary to compare multiple models to choose the best one as per the necessities of the stakeholders. To study this, machine learning techniques like Exploratory Data Analysis, regression analysis, cross-validation, and few more predicting techniques were incorporated to find the appropriate model that gives the best fit for the data. Methodologies used in research works in [5] and [6] were considered in the model-building stage.

Materials and Methods

The dataset used contains the information of 180520 customers about their online goods purchases. Information like the volume, cost and the profit ratio obtained by the company on their purchase. The objective was to predict the profit ratio, the models were built based on it.

Initially, plots and summary of the predictive variable were studied to explore the data. Duplicate rows and values were removed. Further, feature scaling was done to remove the variables which are least affecting the output variable. Variables which show multicollinearity with the predicted variable were removed. Three predictive models were built for the data and K fold cross validation was done for each model. The models were compared, and the best model was chosen.

Results and Discussion

Profit ratios are the window to companies' ability to convert the sales into profit. The relevance of studying them when learning about a company cannot be overstated. The study's data set displays sales data for a variety of goods that the same firm shipped to various nations and regions around the world. To achieve the goal, it is necessary to comprehend the facts. The histogram was plotted to understand the distribution of the response variable 'Profit ratio'. The profit ratio is left skewed because of the presence of an extremely negative profit ratio. '0.4800' is the highest frequently obtained profit ratio
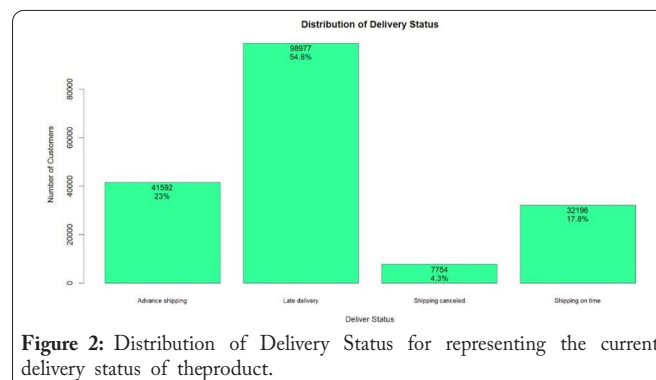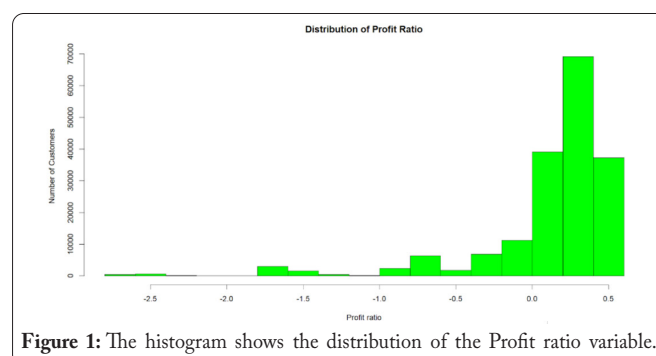
with a frequency of 9197 users. As per the summary of the profit ratio the minimum profit ratio is -2.75, the maximum profit ratio is 0.5 and the average profit ratio is 0.1206.

A bar plot was created to see the nations from which the customers made their purchases in order to further explore the data. Orders came from two locations (Figure 1), with 38.4% of customers coming from Puerto Rico and 61.6% of customers coming from the United States of America (USA). This demonstrates that the Western part of the world has a high level of sales concentration. The items were transported to various locations across the world even though the customers only made purchases from these two nations.

To understand the successful delivery of the product the delivery status variable was visualized (Figure 2). It reveals that 54.8% of the orders were delivered after the scheduled time. Logic dictates that this should have decreased the profit ratio, but the profit ratio distribution refutes this.

To further investigate the relationship between delivery status and profit ratio, box plot was prepared. To have a deeper insight, the boxplot of profit ratio variable was plot- ted after removal of data points below the value of -2. From figure 3 it is concluded that there is no correlation between these two variables. So, Delivery Status cannot be included for predicting Profit Ratio.

The same method was applied to examine the relationships between each categorical variable, including the type of payment used, the nations from which customers made purchases, the markets to which the purchases were delivered, the nations to which the purchases were shipped, whether the purchases had been delivered or not, and the method used to ship the purchases to other nations. The aforementioned variables and the profit ratio did not correlate in any way.

Figure 1: The histogram shows the distribution of the Profit ratio variable.

Figure 2: Distribution of Delivery Status for representing the current delivery status of the product.

Product Name and Profit Ratio were correlated (Figure 4). The box plot is produced for the first 10 products in order to observe the plot precisely. Product name will be taken into account when projecting profit ratio because it has a relationship with profit ratio.

In the predictive models, profit ratio is predicted, and it is called the response variable. The variables which are used as predictors are benefit per order, sales per customer, late delivery risk, order item discount, order item discount rate, order item quantity, sales, product price, difference of days between scheduled and real shipping and product name. The categorical nature of the product name prevents its direct inclusion in the models. As a result, indicator variables are employed to incorporate the information from the variable into the model.

Three predictive models Multiple Linear Regression, Ridge Regression and Elastic Net Regression were used in this paper to predict the Profit Ratio. Uyanık et al. [7] show how Multiple Linear Regression is used in the education sector, following the thread, and applying Multiple Linear Regression for the prediction. In multiple regression when there is a presence of a high number ofpredictors, the chance of occurrence of multicollinearity is high. To overcome this problem as mentioned by McDonald [8], Ridge regression was employed to the data. Although Ridge can improve the prediction in the presence of multicollinearity, it requires all the coefficients roughlyto be in equal size and does not shrink any of the coefficients to zero. To overcome this limitation of Ridge, Elastic Net Regression model which is a hybrid approach that blends the penalization ofboth Ridge and Lasso is used in the data. Han et al. [9] has used Elastic Net Regression to predict changes in civic purpose during emerging adulthood, the same

approach was used in the followingpaper in implementing the model for predicting the response variable.

As authors of [10] suggest, checking the accuracy of models using cross validation techniques is vital. In the paper K-fold cross validation technique is used to validate each model. R², AIC, and BICwere the parameters used to compare the models. The authors of [11] suggest that the model with lower AIC and BIC value among the proposed models are considered to be the best models. table 1 shows the R², AIC and BIC value of all three models.

From table 1 it can be seen that although the AIC and BIC value are lowest for Multiple Linear Regression, Ridge has high R² value and the Elastic Net Regression model with best parameter values applied to the given data converges to Ridge Regression. For predicting the profit ratio of this data Ridge Regression is the most suitable mode.

Table 1: Accuracy of 3 proposed models.

Model Name	Accuracy Measures		
	R² value	AIC	BIC
Multiple Linear Regression	0.6947	–321790.3610	–320535.5434
Ridge Regression	0.8313	-2222.0640	-967.4715
Elastic Net Regression	0.8313	-2222.0640	-967.4715

Conclusion

Predictive models for profit ratio of the website were developed. Using data visualization tools like histogram, bar graphs, box plots, etc. feature scaling was done. Three predictive models viz., Multiple linear regression, ridge regression and elastic net regression were built along with a 10-fold cross-validation technique was carried out to avoid sampling errors. Though the Multiple linear regression model has less AIC and BIC values, R² value for thesame is also lesser than other models. Whereas R² values for ridge regression and elasticnet regression are same along with similar other accuracy measures. After cross-validation, R² values for ridge regression and elastic net regression are 83.13% with -2222.064 and - 967.4715 as AIC and BIC values respectively. It was found that the elastic net regression model converges to ridge regression and then mean squared errors are observed to be low. Thus, results of the analysis showed that ridge regression is the best fitted model for predicting profit ratio of the shopping website. Hence, the novel approach adopted to predict the profitability ratios using machine learning techniques can be deployed so that top management can take strategic decisions to develop the business and improve the supply chain performance further.

Limitation and Further Scope

Online shopping and delivery companies nowadays provide subscriptions and additional benefitson purchase and deliveries are provided to membership holders. The research is based on an assumption that all the deliveries and customers are given equal treatment irrespective of subscriptions. Shipping time and fast deliveries majorly influence the profit
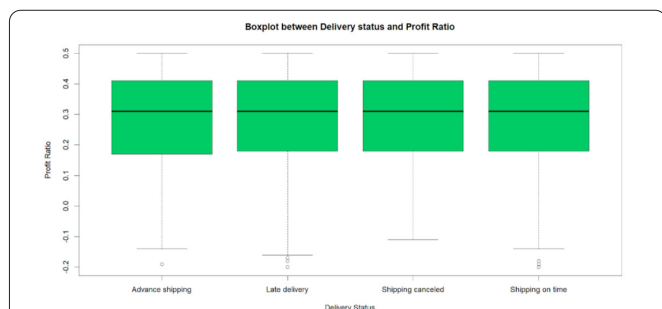

Figure 3: Relationship between Delivery status and Profit ratio. Each box representsthe current delivery status of the product.
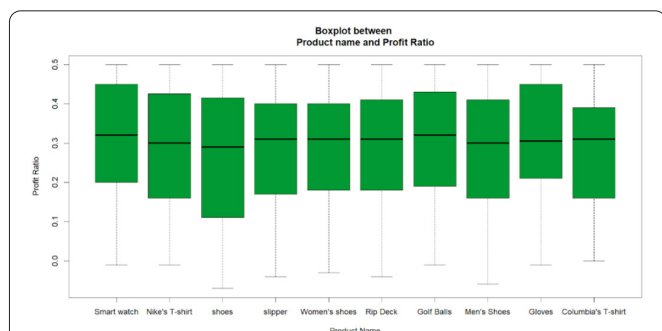

Figure 4: Relationship between Product name and Profit ratio. The first 10 Productsin the respective order is represented in the graph. Each box represents a product.

ratio, which can be considered while model building. This research can be further improved by collecting and analyzing the data considering the influence of the customer-oriented approach of the company on the profit.

Credit Author Statement

Amirta V: Framed the problem statement, Investigation, Analysis; Nishigandha Lad: Analysis, Writing - original draft preparation, Writing - review and editing. All the authors read and approved the manuscript.

References

1. Pidada IAI. 2020. The Effectiveness of Online Marketing Trends: B2B and B2C Application. *Jurnal Ilmiah Manajemen dan Bisnis* 5(1): 90-98.

2. Niraj R, Gupta M, Narasimhan C. 2001. Customer profitability in a supply chain. *J Marketing* 65(3): 1-16. https://doi.org/10.1509/jmkg.65.3.1.18332

3. Ishfaq R, Bajwa N. 2019. Profitability of online order fulfillment in multi-channel retailing. *Eur J Oper Res* 272(3): 1028-1040. https://doi.org/10.1016/j.ejor.2018.07.047

4. Shen ZJM. 2006. A profit-maximizing supply chain network design model with demand choice flexibility. *Oper Res Lett* 34(6): 673-682. https://doi.org/10.1016/j.orl.2005.10.006

5. Doan T, Kalita J. 2015. Selecting machine learning algorithms using regression models. In *IEEE International Conference on Data Mining Workshop (ICDMW),* pp 1498-1505. https://doi.org/10.1109/ICDMW.2015.43

6. Battineni G, Sagaro GG, Nalini C, Amenta F, Tayebati SK. 2019. Comparative machine-learning approach: A follow-up study on type 2 diabetes predictions by cross-validation methods. *Machines* 7(4): 74. https://doi.org/10.3390/machines7040074

7. Uyanık GK, Güler N. 2013. A study on multiple linear regression analysis. *Procedia Soc Behav Sci* 106: 234-240. https://doi.org/10.1016/j.sbspro.2013.12.027

8. McDonald GC. 2009. Ridge regression. *WIREs Comput Stat* 1(1): 93-100. https://doi.org/10.1002/wics.14

9. Han H, Dawson KJ. 2021. Applying elastic-net regression to identify the best models predicting changes in civic purpose during the emerging adulthood. *J Adolesc* 93: 20-27. https://doi.org/10.1016/j.adolescence.2021.09.011

10. Ferdinandy B, Gerencsér L, Corrieri L, Perez P, Újváry D, et al. 2020. Challenges of machine learning model validation using correlated behaviour data: evaluation of cross-validation strategies and accuracy measures. *PLoS One* 15(7): e0236092. https://doi.org/10.1371/journal.pone.0236092

11. Aho K, Derryberry D, Peterson T. 2014. Model selection for ecologists: the worldviews of AIC and BIC. *Ecology* 95(3): 631-636.

12. Wagenmakers EJ, Farrell S. 2004. AIC model selection using Akaike weights. *Psychon Bull Rev* 11: 192-196. https://doi.org/10.3758/BF03206482

13. Koehrsen W. 2018. Overfitting vs. underfitting: a complete example. Towards Data Science. [http://www.pstu.ac.bd/files/materials/1566949131.pdf]