Regression-based Business Decision Support: Application in Online Retail

B. Khatri, M. Rungi
Estonian Entrepreneurship University of Applied Sciences, Tallinn, Estonia
(bhavesh.khatri1982@gmail.com, mait.rungi@eek.ee)

2022 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM) | 978-1-6654-8687-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/IEEM55944.2022.9989568

Abstract - **Electronic commerce (e-commerce) opens up various growth possibilities, with product returns being one of the major challenges. Product returns prediction can be beneficial in taking business decisions in e-commerce. This research is based on an e-commerce company in India with nearly 25% product returns as a case study. Logistic-regression-based statistical analysis was used to predict the dichotomous dependent variable (product returns). Product size, payment mode, price, quantity, ZIP codes, and states were considered as independent variables. A decision support system was proposed based on the variables and a machine learning approach. A total of 3,187 past orders were used to train the model, which was applied in a one-month pilot study. The model could predict 60% of correct product returns with an overall efficiency of 87%. The case company used these predictions to take preliminary decisions on predicted return orders, saving 17% of the costs associated with product returns.**

Keywords - **Decision support systems, logistic regression, machine learning, online retail**

I. INTRODUCTION

The growth in the importance of e-commerce has been nothing short of phenomenal [2]. E-commerce involves the operation of a business with the help of the Internet and making use of information technology [3]. There has been a huge increase in worldwide e-commerce, which was valued at USD 150 billion in 1999 [4]. In 2018, over half the world's population (4.5 billion people) were e-commerce customers, with a total sales value of USD 3.45 trillion in 2019 [5].

In e-commerce, there has been a rise in the number of business-to-customer companies over the past few years. Lenient product return policies, cash on delivery (COD), and free shipping have become so-called customer rights. Although these factors seem good for e-business, they also work in another way to generate high product return rates. Even if product quality is good and it is delivered on time, the product can still be rejected.

For sustainable business and progress, it is essential to minimize product returns. This study aims to assess whether a machine learning algorithm that uses logistic regression can predict returned orders, based on historical data. This study also aims to find out how well a logistic-regression-based decision support system (DSS) can predict the returned orders and what attributes of the DSS predict the return rate.

The research was performed using one Indian online retail company.

Paper based on ref. [1].

II. LITERATURE REVIEW

A. E-commerce and its Challenges in India

Low-cost computers, growing Internet use, and an increase in the Internet Service Provider market have helped the growth of e-commerce in India [6]. Due to increasing Internet penetration, 259.14 million of world's 3 billion Internet users are in India [7]. Online customers in India are anticipated to reach 220 million by 2025. In 2019, it was evaluated that one in every three Indians had bought an item by means of a smart-phone [8].

India's e-commerce industry was forecast to grow from USD 600 million in 2011–2012 to almost USD 70 billion by 2020 [7], and this anticipated to reach USD 200 billion by 2027. India's e-commerce business has the potential to develop more than fourfold to USD 150 billion by 2022, backed by rising salaries and a surge in web clients. The Indian e-commerce industry is anticipated to outperform the US to become the second-biggest e-commerce market in the world by 2034 [8].

B. Customer Behavior

Although there is a bright future for e-commerce business, there are also many challenges. There is high rate of product returns regarding online purchases by Indian customers. According to [9], most buyers from an e-business in India are first-time buyers which are not sure about what to expect from e-business websites. When the product is delivered, they may feel remorse and return the product. For first-time buyers, a lenient product returns policy is a major factor in gaining customers' trust and attracting more customers [9]. Product returns are expensive in e-commerce and reverse logistics have challenges [9]. To overcome this challenge, most e-commerce companies in India provide a COD payment option.

The high preference for COD, however, is riskier for online retailers. Manual cash collection is expensive, risky, and painstaking compared to electronic payments [7, 9, 10]. In India, free shipping is also another tool that e-commerce companies use to attract customers to buy online [11].

Payment gateways in India have an unusually high failure rate and many customers do not bother to retry payment after failed transaction [9]. Another challenge is that the quality of connectivity is poor in several regions [9].

Any payment failure or fraud is a loss for the business when they are paid via Internet banking; further, it must pay a significant share of revenue in fees (4% or more).

Fraudulent charges, chargebacks, etc. all become the merchant's responsibility and hence need to be accounted for in the business model [12]. Often, businesses need to take out delivery insurance for high-value products, which increases the delivery cost [12, 13].

Reference [14] has concluded that the attributes displayed for apparel products play an important role in encouraging impulse buying behavior. Some customers also abuse COD and free shipping [15]. E-commerce companies face comparatively high costs when the product is returned if it was purchased via COD. The combination of so-called customer rights with impulsive buying behavior [16] and abusive behavior customer [15] leads to many product returns.

C. Customer Behavior for Product Returns

When it comes to e-commerce, web browsing is a vital part of the online shopping experience for consumers [17]. In fact, the greater the product variety offered, the longer consumers want to browse [18].

Reference [14] stated that price is the deciding factor for online shoppers, followed closely by shipping costs. The fact that around 40% of online shoppers blame their abandoned carts on shipping-and-handling charges illustrates the importance of shipping costs [14]. In contrast, encouragement to purchase is driven by conduct that tends to be impulsive. The definition of impulse buying encompasses sudden buying with no pre-shopping aim either to purchase the item category or to satisfy a particular purchasing need [19]. Erratic behaviors in the proximity of an appealing object activate impulse buying, and such purchases may occur largely without regard to financial or other consequences [14].

Reference [20] stated that customers making online purchases are devoid of the satisfaction obtained from a physical inspection of any product. Additionally, it increases the possibility of customers exercising their company-offered privileges regarding the product returns policy [20]. It has been shown that a good product returns policy has a positive impact on winning customers [21]. An easier product returns policy can also have a positive impact on customer satisfaction [22]. An easy returns policy reduces the risk customers take when purchasing products online [23]. Allowing customers to return products purchased online can be interpreted as a signal of higher product quality [24]. Another problem is the issue of e-commerce fraud; consumers sometimes abuse consumer protection rights or violate social norms [15]. A typical example would be to purchase an item of clothing, wear it once or twice, and then return it to the retailer. In some scenarios, the returned item is not the item that the customer initially purchased [25].

D. Decision Support Systems (DSSs)

Reference [26] stated, a DSS "*is a computer-based system that represents and processes knowledge in ways that allow decision making to be more productive, agile, innovative, and/or reputable*." Knowledge-driven DSSs show the best decision or action to the managers [27]. They have special problem-solving capabilities and have complete knowledge of basic business rules and the domains they work within. According to Reference [28], the following are the major benefits of a good DSS [29]: 1) higher individual productivity; 2) a faster and more efficient system of data analysis; 3) higher accuracy or reliability than human processes or outcomes; 4) faster problem-solving process.

Despite of benefits, a DSS also suffers from the following limitations [28]: 1) it cannot be used as general-purpose software; 2) it is limited to the domain of use; 3) it needs to be integrated separately into decision-making processes; 4) it has technological limitations.

III. METHODOLOGY

A. Description of the Case Company

The authors worked with an e-commerce company from India, named Uniplanet Store. The company is a small-scale Indian online fashion retail company, established in October 2015. It sells printed t-shirts only via e-commerce websites. They are a group of youthful enthusiasts who are purvey unique items at sensible prices. Their preparing unit is in Gujarat, India, from where all the operations, such as designing, printing, packing, and dispatching, are carried out. Uniplanet Store has set up virtual stores on Amazon, Flipkart, Paytm, and its own e-commerce website. The total annual turnover of the company was recorded as approximately €80,000 in 2018.

To gain more orders and garner better customer trust and satisfaction, these marketplaces have liberal rules regarding product returns, such as free shipping, COD payment mode, etc. Uniplanet Store must follow these rules. Uniplanet Store receives a larger number of orders from such big platforms. Hence, although Uniplanet Store gets a larger number of orders from these marketplaces, as many as 25% of orders are returned. This is a challenge for the company because such orders do not provide revenue and represent a loss of time, efforts, and costs. The company does not have control over such orders and must define a strategy to deal with this challenge. If the company can predict returned orders in advance, then it can take further actions on such returns. This can involve contacting a customer to ensure the delivery address, correct size, and his/her willingness to receive an order.

B. Data Collection

Uniplanet Store shared order data from September 2018 to September 2019 (12 months). For this research to include payment type and to analyze the effect of COD on product returns prediction, the authors only considered orders from Amazon. These order data include various fields, including order ID, customer name, customer contact number, stock keeping unit number, price, quantity, payment mode, ZIP code, city, regional state, delivery status, and other fields. For reasons of data privacy, the authors did not analyze personal data and relied on other data. The authors used other data points given in Table 1 as independent variables. The dependent variable is whether the product was successfully delivered or returned to the retail company. Table I shows the details of the independent variables and the number of features each categorical variable holds.

TABLE I
VARIABLES

Independent variable	Category	No. of features
Size	Categorical	5
Price	Numerical	-
Quantity	Numerical	-
ZIP code	Categorical	1,721
Regional state	Categorical	38
Payment mode	Categorical	2

C. *Training and Applying the Model on Live Data*

During the 12 months, the company received orders worth €20,809 (3,187 orders) from Amazon. Of these, 2,276 orders (worth €14,882) were successfully delivered to customers, while 911 orders (worth €5,927) were returned. This means that orders worth 28% of total purchases were returned for one or more reasons. This 28% contributes towards products manufactured but rejected even though the product quality was good, and the orders were dispatched on time. Of these returned orders, €4,725 (80%) were COD, while €1,201 (20%) were prepaid.

To develop a model using Python programming language and the scikit library, the data were loaded into memory in the form of a Python data-frame. This represents the $N \times N$ matrix. Initially, all 3,187 orders were used after encoding to train the machine learning algorithm. Using delivery and returned data gave less accuracy for returned orders. The primary reason for this was the fact that there were more delivered orders than returned orders. Therefore, feeding more delivered orders to the algorithm resulted in improved learning for delivered orders but not returned orders. In this way, using more delivered orders to train the model resulted in higher false-positive errors for returned orders. After using different ratios of delivered orders and returned orders, it was concluded that there should be a good balance between delivered orders and returned orders. Finally, after many attempts, the research ended up using 911 returned orders and only 900 of the most recent, successfully delivered orders to train the model. Before this model was applied to real-life data, the model's accuracy had to be evaluated.

The model was not trained using all data available. The research used 70% of the collected data to train the model, while 30% of the data was used to test the accuracy of the model. Another important fact is randomization of data. That means that 30% of orders were extracted at random from the total order data.

The logistic regression model, trained using the training data set, could predict 95% of returned orders from the test set. This model can be applied to live orders to predict returns. All the 1,811 orders were used to train the model. They were then imported into the Python data-frame. Following this, new orders were imported into the same data-frame. Our new orders were not labeled and thus did not have a delivery status. A dummy delivery status was inserted into new orders. New orders were fed into the model to obtain predictions.

IV. RESULTS AND DISCUSSION

The model was used in October 2019. The company periodically shared details of new orders, the authors responded with predictions from the model, and the company took actions based on these predicted product returns. In the pilot study, a total of 283 orders were considered from Amazon to predict product returns.

Data for the prior 12 months showed that 2,223 (70%) orders out of a total of 3,187 orders were registered as COD. In the pilot study, during October 2019, more than 56% of orders were COD. In the prior 12 months, almost 80% of returned orders were COD. In the pilot study in October 2019, 64% were COD.

In the formulated logistic regression, the COD feature had a coefficient value of 0.17. Even though balanced data were used to train the model, i.e., an almost equal number of COD and prepaid orders, the coefficient for COD was higher than the coefficient for prepaid (0.22). This indicates that COD orders have a higher chance of product returns compared to prepaid orders. Further, the orders coming from certain ZIP codes have a higher probability of returning orders.

In the pilot study, out of total 283 orders, the trained logistic regression model predicted 231 product deliveries and 52 product returns. Their response was to call customers and ensure that they were interested in receiving their orders. They called most of the customers of predicted return products. Out of the predicted 52 products, 19 customers responded to the call, 28 customers did not and 4 remained without call. Out of the 19 customers who responded, three customers agreed that they had changed their mind and wanted to cancel their order. A total of 12 customers agreed to receive the products. Out of these 12 orders, 10 orders were returned even though the customer had verbally agreed to receive the product. Only two customers who agreed accepted the orders. Four customers stated that they had changed their minds and would cancel the order but did not do this. No action was taken on the four predicted returned orders. There were 28 orders with no response from the customer. Out of these 28 orders, 27 orders were returned and only one was successfully delivered. The model had also predicted that a total of 236 orders would be delivered, but only 200 orders were successfully delivered to customers.

To obtain more insight, a confusion matrix (Table II) and classification reports (Table III) were derived at the end of the one-month pilot study.

TABLE II
CONFUSION MATRIX ON PREDICTION FOR OCTOBER 2019

	Predicted	
Orders (*n*=283)	Delivered	Returned
Actual delivered	199	4
Actual returned	32	48

It is clear from the confusion matrix that out of the 80 actual product returns, the model was able to predict 48 product returns. The model had also accurately predicted the successful delivery of 199 orders, out of the 199 delivered orders. However, there were four orders that were incorrectly predicted as product returns. This is a type-I error. Similarly, the model incorrectly predicted 32 orders as delivered that ended up being returned. This is a type-II error. This error is slightly significant as a higher value for this error indicates an increase in the number of product returns that the model did not predict.

TABLE III
CLASSIFICATION REPORT FOR PREDICTIONS IN OCTOBER 2019

	Precision	Recall	F1-score	Support
Delivered	0.86	0.98	0.92	203
Returned	0.92	0.60	0.73	80
Accuracy			0.87	283
Average	0.88	0.87	0.86	283

Table III shows a classification report for the predictions. The recall for the delivered product is 0.98, which shows how significantly accurate the model is. Similarly, the recall of returned products is 0.60. This indicates that the model could successfully predict 60% of the actual product returns; if there are 80 product returns, the model would accurately predict 48 product returns. Assessing the overall efficiency of the model revealed it to be 87%.

In the one-month pilot study the company's store sold approx. €1,815. DSSs' ability to self-learn, identify associations between data points, and perform heuristic operations proved useful. This makes them a potential candidate to provide increased capacity for problem-solving and intelligent suggestions [30]. The machine learning model could correctly predict 60% of product returns. The distribution of the amount of correctly predicted returns was 76% for COD and 24% for prepaid. The value of the actual returned products for the month was €486. The machine learning model could predict correct product returns worth €302. The overall efficiency of the model is 87%. This means that by using this model, the company could predict low-risk delivery orders and concentrate more instead on predicted returned orders through the application of a relevant strategy to save money. During this pilot study, Uniplanet Store was able to save €50 using this model (out of the total of €486). It can be concluded that the company could save 16.5% of the money associated with predicted product returns, which is 60% of actual returns. If a proper strategy is implemented, then the company could save up to 60% of the money associated with product returns.

Summing up, the e-commerce market in India is booming [6], but to remain competitive in the market and attract more and more customers, e-commerce companies need to take many measures. This includes easy product return policies [21], a COD option [11], free shipping [11], etc. Research has confirmed that the COD payment mode has a higher preference than the online prepaid option [9, 10]. Orders from the prior 12 months and the one-month pilot study also showed that there are more COD orders compared to prepaid orders in the company under study. The implementation of all these factors leads to impulse buying behavior and ultimately to a higher product returns rate. In line with [9], the authors found a high rate of returned orders; during the prior 12 months, as many as 28% of total products were returned. Research has confirmed that there is also abusive behavior [15] related to COD [11], free shipping [11], and lenient return product policies [21]. The authors found that, as many of the independent variables are categorical, the logistic regression model is suitable in predicting discrete output values [31], i.e., product returns.

V. CONCLUSION

E-commerce companies should provide a free shipping option to encourage more customers to buy online, as well as a liberal product returns policy. Due to the high failure rate of online payments, customers are inclined to choose the COD option. Although the COD option is costly for any e-commerce business, unfortunately, it cannot be avoided. The collective effect of first-time users, impulsive and compulsive buying, the high chance of fraud or misuse, etc. results in a higher rate of product returns even though the product quality is good, and the product is dispatched on time. As e-commerce companies do not have much control over these factors, predicting product returns can help improve this situation. This research aimed to predict product returns, thus enabling e-commerce companies to take further action regarding predicted product returns. By training a machine learning model based on logistic regression using past order data, the company used for the study could predict 86% of the correct products delivered out of the overall delivered predictions. The overall efficiency of the model proved to be 87%. Company contacted customers to ensure the genuineness of the purchase. In this way, the company effectively reduced the cost of such product returns. Using this application, the company reduced the number of product returns and saved on delivery charges for returned products; the company was able to save 16.5% of the money associated with predicted returns. This research and its application were thus able to help the case company Uniplanet Store in minimizing product returns.

A. Theoretical Implications

The logistic regression model [32] is suitable in predicting discrete output values [31] and reducing the rate of product returns. Some attributes can be irrelevant [33]; the authors discovered few attributes to be less relevant.

B. Practical Implications

For online t-shirt retail, a machine learning DSS based on logistic regression [32] is a beneficial approach to reducing the rate of product returns. The necessary attributes are presented in Table I. Feeding more delivered orders into the algorithm will increase the accuracy.

Working out the machine learning algorithm took approximately 2.5 man-months, and the algorithm helped to reduce the product returns rate by 16.5%.

C. Limitations

The developed machine learning DSS is limited to a specific field (online t-shirt retail), as suggested by [28].

REFERENCES

[1] B. Khatri, “Business decision support using machine learning: Application to online retail in India,” master’s thesis, EUAS, supervisor M. Rungi, Tallinn, Estonia, 2020.

[2] M. Warkentin, *Business to Business Electronic Commerce: Challenges and Solutions*. Hershey, PA: IGI Global, 2001.

[3] D. Chaffey, *E-Business & E-Commerce Management*. Hoboken, NJ: Prentice Hall, 2012.

[4] N. Terzi, “The impact of e-commerce on international trade and employment,” *Proc. Soc. Behav. Sci.*, vol. 24, pp. 745–753, 2011.

[5] L. Fabri and I. Márquez, “Will e-commerce dominate physical store?” *Int. J. Tech. Bus.*, vol. 2, no. 1, pp. 23–29, 2020.

[6] P. Kalia, K. Navdeep, and T. Singh, “E-commerce in India: Evolution and revolution of online retail,” in *Mobile Commerce*, Information Resources Management Association, Ed. Hershey, PA: IGI Global, 2018, pp. 736–758.

[7] B. Avudaiammal, “E-commerce in India: The developments and real challenges,” *International J. of Multidisciplinary Educational Res.*, vol. 9, no. 1, pp. 209-2013, 2020.

[8] India Brand Equity Foundation, “Indian ecommerce industry analysis,” 2020. https://www.ibef.org/industry/ecommerce-presentation

[9] B. Malhotra, “E-business: Issues & challenges in Indian perspective,” *Glob. J. Bus. Manag. Inf. Technol.*, vol 4, no. 1, pp. 11–16, 2014.

[10] H. Kaur and D. Kaur, “E-commerce in India: Challenges and prospects,” *Int. J. Eng. Techniques*, vol. 1, no. 2, pp. 2395–1303, 2015.

[11] N. A. Reddy and R. Divekar, “A study of challenges faced by e-commerce companies in India and methods employed to overcome them,” *Proc. Econ. Financ.*, vol. 11, pp. 553–560, 2014.

[12] S. Ray, “Emerging trend of e-commerce in India: Some crucial issues, prospects and challenges,” *Comput. Eng. Intell. Syst.*, vol. 2, no. 5, pp. 2222–2863, 2011.

[13] N. Chanana and S. Goele, “Future of e-commerce in India,” *Int. J. Comput. Bus. Res.*, pp. 2229–6166, 2012.

[14] E. J. Park, E. Y. Kim, V. M. Funches, and W. Foxx, “Apparel product attributes, web browsing, and e-impulse buying on shopping websites,” *J. Bus. Res.*, vol. 65, pp. 1583–1589, 2012.

[15] A. Heiman, B. McWilliams and D. Zilberman, “Demonstrations and money-back guarantees: market mechanisms to reduce uncertainty,” *J. Bus. Res.*, vol. 54, no. 1, pp. 71–84, 2001.

[16] R. Larose, “On the negative effects of e-commerce: A sociocognitive exploration of unregulated on-line buying,” *J. Comp.-Mediated Commun.*, vol. 6, no. 3, pp. 115–128, 2001.

[17] D. N. Smith and K. Sivakumar K., “Flow and Internet shopping behavior: A conceptual model and research propositions,” *J. Bus. Res.*, vol. 57, pp. 1199–1208, 2004.

[18] E. Y. Kim, “Online purchase intentions for product categories—The function of Internet motivations and online buying tendencies,” *J. Korean Soc. Cloth. Text.*, vol. 32, no. 6, pp. 890–901, 2008.

[19] S. R. Madhavaram and D. A. Laverie, “Exploring impulse purchasing on the Internet,” in *NA - Advances in Consumer Research* (vol. 31), B. E. Kahn and M. F. Luce, Ed. Valdosta, GA: Association for Cons. Res., 2004, pp. 59–66.

[20] R. Yan, “Product categories, returns policy and pricing strategy for e-marketers,” *J. Prod. Brand Manag.*, vol. 18, no. 6, pp. 452–460, 2009.

[21] A. B. Bower J. G. and Maxham III, “Return shipping policies of online retailers: Normative assumptions and the long-term consequences of fee and free returns,” *J. Mark.*, vol. 76, pp. 110–124, 2012.

[22] D. Dissanayake and M. Singh, “Managing returns in e-business,” *J. Internet Commer.*, vol. 6, no. 2, pp. 35–49, 2008.

[23] S. L. Wood, “Remote purchase environments: The influence of return policy leniency on two-stage decision processes,” *J. Mark. Res.*, vol. 38, no. 2, pp. 157–169, 2001.

[24] C. Bonifield, C. Cole and R. L. Schultz, “Product returns on the Internet: A case of mixed signals?,” *J. Bus. Res.*, vol. 63, pp. 1058–1065, 2010. doi: 10.1016/j.jbusres-2008-12-009

[25] K. Wachter, S. J. Vitell, R. K. Shelton, and K. Park, “Exploring consumer orientation toward returns,” *Bus. Ethics: Eur. Rev.*, vol. 21, no. 1, pp. 115–128, 2012.

[26] C. W. Holsapple, “Decisions and knowledge,” in *Handbook on Decision Support Systems 1: Basic Themes*, F. Burstein and C. W. Holsapple, Ed. Berlin, Heidelberg: Springer.

[27] C. Zhengmeng and J. Haoxiang, “A brief review on decision support systems and it’s applications,” in *IEEE International Symposium on IT in Medic. and Education*, 2011.

[28] D. Power, *Decision Support Systems: Concepts and Resources for Managers*. Westport, CN: Quorum Books, 2002.

[29] C. W. Holsapple and M. P. Sena, “ERP plans and decision-support benefits,” *Dec. Supp. Sys.*, vol. 38, pp. 575–590, 2003.

[30] M. M. Hamad and B. A. Qader, “Knowledge-driven decision support system based on knowledge warehouse and data mining for market management,” *Global Journal of Management and Business*, vol. 13, no. 10, pp. 25–41, 2013.

[31] G. James, D. Witten, T. Hastie and R. Tibshirani, *An Introduction to Statistical Learning with Applications in R*. New York, NY, Heidelberg, Dordrecht, London: Springer, 2017.

[32] V. Miskovic, “Machine learning of hybrid classification models for decision support,” in *Sinteza 2014 – Impact of the Internet on Business Activities in Serbia and Worldwide*, 2014, pp. 318-323. doi: 10.15308/sinteza-2014-318-323

[33] S. Shai and B. Shai, *Understanding Machine Learning: From Theory to Algorithms*. Cambridge: Cambridge University Press, 2014.