

# Predictive and Sentiment Analysis on Factors Influencing E-Commerce Growth in Transition Economies

Joan Leonard Mushi, Haslina Hashim

Mudiarasan Kuppusamy

University of Cyberjaya

drarasan@cybermed.edu.my

## Article Info

Volume 82

Page Number: 772 - 778

Publication Issue:

January-February 2020

## Abstract

With emerging technology in Data Science, every industry is seen to adapt technological trends to benefit their businesses. It's a competition on delivering better quality and high value goods and services to the market and this is only possible through deep analysis in understanding consumer demands and market trends. This research aimed at analyzing the e-commerce industry on how to increase growth through identifying factors that influence online customers purchase intentions and used machine-learning algorithm in predicting online sales. The variable "number of orders" has been used as the dependent variable, which is used in determining the growth of e-commerce from the increase in online sales. The research identified factors such as online reviews, price and product quality (ratings) as main factors pushing online purchases in exploring their impact towards the dependent variable. This research involves three main types of analysis; diagnostic analysis, sentiment analysis and predictive analysis. Findings revealed that online consumers are more drawn to making online purchases if the product description, price, reviews and delivery of products and services match with their needs. The sentiment analysis was able to connect the findings from descriptive analysis by understanding consumers thoughts and discussions on the products through the online reviews. Additionally ANN model was able to accurately predict how number of orders is influenced by factors such as price, online reviews, wishlist, product description, rating, number of feedback, delivery time, delivery price, payment methods, stock available and discount price.

## Article History

Article Received: 14 March 2019

Revised: 27 May 2019

Accepted: 16 October 2019

Publication: 04 January 2020

**Keywords:** Artificial Neural Network, Multiple Linear Regression, and Number of Orders

## I. INTRODUCTION

In the recent decade e-commerce has taken a positive shift and is the fastest growing segment in the retail industry. The introduction of World Wide Web, electronic commerce has reformed traditional commerce and boosted sales as well as exchange of merchandise and information. Particularly, the Internet through its power to enable e-commerce has redefined rules of the entire business industry. Firms that have embraced e-commerce stumbled on new opportunities while firms lagging are becoming irrelevant with the increasing development in the retail industry.

In the research done by [4],[5] and [6] IT has been considered as one amongst the greatest innovations which has further given rise to digital technologies that has been a major backbone for e-commerce growth, it has also enabled

business to be more innovative making businesses more competitive as never before. The use of IT in operations related to economic and commercial processes has given rise to other opportunities under the e-commerce umbrella such as s-commerce and m-commerce, which play an important role in global economic affairs.

Notably growth of e-commerce has led to intense evolutions and changes in commercial affairs. It has transformed the business focus from merely physical goods to delivering information, service and intelligence focus. Moreover [8] states that e-commerce has changed business perspectives by ensuring customer intimacy as opposed to production excellence. Furthermore, e-commerce has reformed the way businesses compete by creating essential operational changes in business-level strategy and business in general. Therefore, reorientation in strategic thinking in business is

therefore necessary to view e-commerce as an engine of profit and growth rather than cost. Gradual growth of e-commerce is an example of how the use of ICT is reshaping trade and production, with major implications for transition economies and developing countries.

## II. MATERIALS

The expansion of Internet adoption and usage over the last two decades has given researchers and academicians new avenues to explore the art of prediction. Tracking diffusion and growth of e-commerce has been difficult since most firms do not precisely track their online sales and in particular firms in the transition and developing countries.

Reference [3] outlined numerous factors for growth of e-commerce such as online security and privacy, Internet penetration, mobile money and social economical aspects such as infrastructure and language. However given we live in a world of data and decisions are based on already existing information, there is lack of information on factors such as pricing effect, online reviews and product quality which are the main purchase drivers for online consumers because they inform online customers information about products prior purchase.

## III. METHOD

This research included both quantitative and qualitative analysis. Quantitative methodology has been used to test hypothesis generated and determine the relationship between predictor variables and response variable where as qualitative methodology has been used to find out if there exists any relationship between the quantitative findings and qualitative findings. This research made use of data from Aliexpress website and relied on the variable “number of orders” as an indicator of e-commerce growth. Specifically the predictor variables that were used were product description, ratings, number of feedback, price, discount price, delivery time, delivery price, wishlist, stock available, return policy, payment methods and online reviews (first 10 reviews of each product).

Throughout the process different analysis was done to mine insights from the data. Multiple linear regression was used to understand the relationship that exists within the variables and how the predictor variables influence the response variable. Sentiment analysis has been performed on the qualitative data to link the qualitative findings together with the quantitative findings. Predictive analysis has also been done to gauge how accurate Artificial Neural Network (ANN) is at predicting the number of orders in an e-commerce platform.

## IV. FINDINGS & DISCUSSIONS

### Multiple linear regression:

Despite having a variety of factors in consideration multiple linear regression has been used to outline the variables that highly affect the number of orders. Moreover multiple linear regression pinpoints factors that affect the dependent variable positively and negatively and through which one is able to understand the crucial variables for predicting the number of orders.

The below figure shows the result obtained from running the multiple linear regression function. Interpretations of the findings have mainly focused on p-value of the variables, parameter estimate, r-square and standard error of each variables.

```
Call:
lm(formula = Number.of.Orders ~ Product.Description + DiscountPrice +
    Price + Rating + Votes + Wishlist + DeliveryTime + Shipping.Cost +
    Stock.Available, data = CapstoneData)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-19920.2  -102.7   -37.5    33.9  30637.9
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  125.756293  150.596409   0.835  0.40373
Product.Description  72.024476  31.696927   2.272  0.02312 *
DiscountPrice  -1.027533   0.487450  -2.108  0.03509 *
Price         -0.084775   0.029932  -2.832  0.00464 **
Rating        -7.935742  12.181283  -0.651  0.51478
Votes         2.345373   0.017417 134.660 < 2e-16 ***
Wishlist      -0.023837   0.003231  -7.377 1.93e-13 ***
DeliveryTime  -2.981747   2.292914  -1.300  0.19353
Shipping.Cost  1.477213  11.808980   0.125  0.90046
Stock.Available -0.029552   0.028515  -1.036  0.30008
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 933.5 on 4284 degrees of freedom
(493 observations deleted due to missingness)
Multiple R-squared:  0.8781, Adjusted R-squared:  0.8778
F-statistic: 3428 on 9 and 4284 DF, p-value: < 2.2e-16
```

### Figure 1: Findings of Multiple Linear Regression

Results obtained from fig.1 above shows the coefficient of determination is 0.8781 which indicates the variability of the response variable is well explained around its mean whereas the adjusted r-squared is 0.8778 indicates 87.78% of the variance in the response variable is explained by the predictor variables. This shows the model fits well with the data. Moreover the difference between r-squared and the adjusted r-squared is minimal (0.003) indicating that there are not many predictor variables used in modeling hence the model is less complex.

The standard error of estimate signifies the average distance the observed values are from the regression line. Correctly it tells on average how wrong the regression model is by using the units of the response variable. Product description, rating

and shipping cost have a slightly large standard error hence far from the regression line compared to other variables. Relying on  $\alpha = 0.05$ , the model shows a general p-value of  $<2.2e-16$  which means the p-value is extremely small, hence making it so significant signifying the predictor variables are statistically significant in predicting the response variable.

Based on the results above the variables that highly affect the response variable is product description i.e. for every unit change in product description, there is going to be an increase of 72.02 in number of orders. Interestingly the variable product description is the only variable that has a huge impact to online purchase i.e. well described products increase online customers purchase intent by 72%. This shows that most online customers rely on product information provided to determine their purchase since the products are not tangible prior purchase. This shows that online customers expect online vendors to provide relevant and accurate information on the product. Additionally this means that online consumers are more satisfied with information that meet their demands through the quantity and credibility of the information provided where quantity refers to the amount of information provided and credibility refers to degree of consumers confidence in the information provided by online vendors. In alignment of the results obtained from the multiple linear regression providing appropriate and adequate information assist in dispelling concerns and fears towards online shopping.

Looking at the key focus variables i.e. price and discount; they have a negative impact to the number of orders. i.e. for every unit decrease of price and discount there is going to be an increase in the number of orders by 0.08 and 1.02 respectively. This implies the lower the price and discount the higher the customers purchase intention and higher the number of orders. This shows that customers are driven by friendlier prices and discounts during online shopping. Results from [7] points out that most online customers expect online products and services to be offered at a lower cost as compared to traditional stores since during online shopping customers are not exposed to the product hence they are not certain on the quality of the product. This rather makes online customers hesitant in spending a lot of money during online shopping therefore seeking for online platforms with friendlier prices for the same product.

For the case of other variables such as votes and shipping costs; for every unit increase in the number of votes and shipping cost there is going to be an increase in the number of orders by 2.34 and 1.47 respectively. This shows that customers are looking into products that are dominant in the market through the number of votes but the shipping expenses are not of high focus however inversely, for every unit decrease in delivery time there is going to be an

increase in the number of orders. This shows that customers are willing to pay shipping expenses but are not ready for longer durations of delivery from the seller. Findings from [2] and [9] align that delivery time is amongst the key factors influencing online shopping. Reference [7] reports that timely delivery of online goods that is reliable and safe is the essential goal for online customers, as online customers believe they are entitled to receive their particular product in the specified time. Delayed deliveries raise customers' dissatisfaction whereas timely and reliable deliveries incite new online sales. Thereafter it can be concluded that quality of delivery service, which includes delivery time, and the state of the product when delivered influences confidence in online shopping, increases e-satisfaction and encourages repurchases.

From the covariance results shown in the figure below, a positive value between two variable show that the two variables tend to increase or decrease together whereas a negative value show that when one variable increases the other decreases. One of the interesting findings from the covariance is the negative relationship between price and wishlist i.e. if the price of a product increases the number of wishlist decreases and the opposite is true. Product price and product description have a negative relationship where if the description of a product decreases the price decreases, this means that products that are well described are sold at a higher price and products with less description are sold at a cheaper price.

Looking at variables that positively impact each other shipping cost and delivery time are among the variables with positive covariance where if the shipping cost increases the time taken to deliver products increases which in reality could mean that higher shipping costs reflect international deliveries which takes longer time to deliver. Ratings and price also have a positive relationship whereby if the rating of a product increases the price increases, this could mean that retailers have the confidence and assurance from customers on the quality of the product hence tend to tweak the prices of the product.

From the covariance results shown in the fig.2, a positive value between two variable show that the two variables tend to increase or decrease together whereas a negative value show that when one variable increases the other decreases i.e. they move in opposite directions. Looking at the fig. 2 below price and discount price have a negative relationship i.e. if the discount of a product increases the price of a product decreases and vice versa is true. Other interesting negative relationship found is between price and wishlist i.e. if the price of a product increases the number of wishlist decreases and the opposite is true. Product price and product description have a negative relationship where if the description of a product decreases the price decreases, this



Fig. 5 below investigates the existing trend between price and the number of orders on the basis of how well the product is described. Despite of products that are “not well described” being sold at a lower price, the number of orders on the products is also very low this explain the fact that consumers are not able to make a purchase decision given that their not able to align their product requirement from the product description.

However, consumers are also price sensitive, in spite of products being well described their purchases are slightly lower due to high prices on such products. However, moderately described products have higher purchases due to the lower price posed to them as opposed to well-described products.

This explains that online consumers are not only looking for goods with sufficient information but also eyeing for products sold at a friendlier price. Fig. 5 below illustrates the existing pattern between the three variables.



**Figure 5: Effect of price on number of orders**

**Predictive Analysis on e-commerce growth:**

In predicting the accuracy of the e-commerce growth through the response variable a series of experiments under the ANN algorithm have been performed. This was done in order to identify the parameters or experiment that will yield the best output hence reckoned reliable. It is important to note that that ANN has been used with variables that were significant in explaining the response variable and this was through the output from multiple linear regressions.

**Table 1: Results of ANN experiments from the data**

ANN	Feature Selection	Hidden layer = 2 i.e. (6,4)	95.6	75.2
	Removing insignificant features from MLR output			
	Hyper parameter tuning	Hidden layer = 1 i.e. 3	94.4%	88.8%
	Hyper parameter tuning	Hidden layer = 2 i.e. (6,2)	98.3%	89.6%
	Cross Validation	K = 10 (i.e. 10 folds cross validation)  7 <sup>th</sup> fold has the lowest MSE		99%

The output obtained in the predictive analysis shows that ANN yields good accuracy in predictions hence is reliable in predicting the number of orders in e-commerce platforms. The models created on neural network throughout all experiments have shown no signs of over fitting or under fitting. The models have also been able to converge within minimum errors. Both the test and train dataset have been able to accurately predict the number of orders with all datasets having an accuracy of more than 70%.

Additionally parameter tuning and cross validation helps in increasing the models accuracy. As there is no single rule in tuning the parameters, the more you tune the model the more likely the model predicts more accurately depending on the different tunings.

**V. CONCLUSION**

This research employed exploratory data analysis, sentiment analysis and predictive analysis through using multi-layered neural network modeling in examining predictors of number of orders on consumer electronic products. The research intended to reveal the scale in which the number of orders is influenced based on a number of predictor variables. Despite that this research focused on three key factors i.e. product price, product quality and online reviews, other factors were also used during analysis and intriguingly came up with new findings. Generally the findings show that the important predictors of number of orders are product description, price, online reviews, delivery time and discount price. Whereas variables such as shipping cost, payment method, rating, stock available and

return policy had no influence to customers purchase decisions in an online shopping context.

One of the interesting findings that came out during analysis is the influence of product description in the number of orders. Having the strongest influence on the number of orders, this explains that customers are influenced in purchasing a product online if the attached description meets their demand and this is determined by the valence and volume (i.e. quality of information and depth of information) the product has. Given that this research had not focused on this variable, there is more room to uncover insights related to product description and online sales in the future.

Amongst other factors influencing consumers' online purchase, product pricing was found to be equally as important. Given that the data was obtained from an e-commerce platform that targets mainly mid to lower income earners, most of the customers in this income band are price conscious looking for products that meet their demands at affordable prices through comparing products from different e-commerce platforms. Additionally pricing has a significant impact on consumer's online purchase, especially in initial phases of consumers' online purchase.

On the contrary customers are not only looking for affordable prices but are also keen on the product description. It was found that customers are likely to buy products which are well described but at an affordable price. However from the analysis, online platform products that are well described are sold at slightly higher prices hence customers opt for moderately described products whose prices are slightly lower. This means that increase in sales orders on consumer electronic goods could be better achieved if prices of well described products could be slightly adjusted or introducing offers such as promotions and discounts on the products could increase the sales orders.

Given that price is associated with the products value and quality, it is important to note excessively low prices on products can generate mistrust from customers especially for first time online customers. Additionally online platforms allow customers to compare prices from different e-commerce platforms, which then raises doubts for excessively low prices. Online customers not only seek to maximize savings but also rely on the confidence built from the projected product online. Henceforth it is important for e-commerce platforms to have a pricing strategy that serves for the organization as well as the end customer.

This research used the variable rating to determine the product quality. Unlike findings from (Chong et al., 2016) which shows product ratings influence the sales by 5-9%

findings in this research show that customers are not keen on the product ratings as opposed to reviews which connect with customers whereby customers are able to have a feel of the product as well as what they should expect from the product. This could mean sentiment analysis is another area of focus for the e-commerce platforms where they could meet customer's demands through accomplishing their shortfalls expressed in the negative comments from customers.

In predicting the number of orders, ANN model has been used and was able to accurately predict the number of orders however; it is safe to note that ANN model was unable to work with all variables in the data. With feature selection where variables, which are not significant and have an undefined relationship with the response variable were removed and ANN model was able to converge. This shows that ANN model is only able to predict the number of variables accurately when significant variables have been used given the data used in analysis.

Lastly, e-commerce is one of the sectors driving big data generation. A lot of user-generated information found online is used as a guide by consumers in decision-making prior purchase. For this case retailers, producers and business owners have to keep in mind on the information uploaded (both by the retailer and consumer) online for their products and businesses in order to increase conversion rates and ultimately growth the e-commerce field.

### **Limitations and future directions**

Despite the conclusions, this research has several limitations worth noting. Primarily this research has focused on consumer electronic goods such as headphones, chargers etc. which are believed to be one of the goods category with a shorter shelf life i.e. goods whose demand is high. The output from the analysis has shown different predictor variables vary in their effect towards the number of orders, which ultimately is used to measure e-commerce growth in e-commerce platforms. Therefore analysis, output and insight generated are only aligned to electronic goods in the e-commerce platform. Future research could look into other products with or without a shorter shelf life and comparisons can be made, additionally incorporating different goods regards of their category is another direction which could yield more actionable insights and bring more value to the e-commerce platform and field as a whole.

Another possible limitation of this research is on the sample size. Given that data is growing at a very fast pace future research could make use of big data in analyzing and evaluating if the model used is applicable in predicting the number of orders.

Lastly this research has focused on ANN algorithm in predicting the number of orders with the predictor variables, given there are quite a number of algorithms in the field of data science there is more room in exploring how accurate and reliable other algorithms such as XG Boost, random forest etc. are in predicting the number of orders and how different variables work in predicting the number of orders.

## REFERENCES

1. A. Y. L. Chong, B. Li, E. W. T. Ngai and F. Lee, "Predicting online sales via online reviews, sentiments and promotion strategies: A big data architect and neural network approach (International Journal of Operations and Production Management)", vol. 36. 4, 2016, pp. 358-383.
2. C. C. Lin, H. Y. Wu, Y. F. Chang, "The critical factors impact on online customer satisfaction (Procedia Computer Science)" vol. 3,2011, pp. 276-281
3. E. O. Ibam, O. K. Boyinbode, M. O Afolabi, "E-commerce in Africa: The case of Nigeria (Research Article EAI.EU)", vol. 4. 15,2018, pp. 1-6
4. J. Kwak, Y. Zhang and J. Yu. "Legitimacy building and e-commerce platform development in China: The experience of Alibaba. (Technological Forecasting and Social change)." Doi.org/10.1016/j.techfore.2018.06.038. 2018. pp. 1-10
5. M. Choshin, A. Ghaffari, "An investigation of the impact of effective factors on the success of e-commerce in small and medium- sized companies (Computers in Human Behavior)". 2017, pp. 67-74.
6. N. Duch-Brown, L.Grzybowski, A. Romahn and F. Verboven. "The impact of online sales on consumers and firms. Evidence from consumer electronics. (International journal of Industrial Organization)." Vol. 52. 2017. pp.30-62
7. N. Vasic, M. Kilibarda, T. Kaurin, "The influence of online shopping determinants on customer satisfaction in the Serbian market (Journal of Theoretical and Applied Electronic Commerce Research)" vol. 14.2,2018, pp. 70-89
8. R. Rahayu and J. Day. Determinant factors of e-commerce adoption by SMEs in developing countries: Evidence from Indonesia. (Procedia- Social and Behavioral Sciences)." vol. 195. 2015. pp. 142- 150
9. R. Shanthi, D. Kannaiah, " Consumers Perception on Online shopping (Journal of Marketing and Consumer Research)" vol. 13.2,2015, pp. 14-21
10. T. Nisar, G. Prabhakar, "What factors determine e-satisfaction and consumer spending in e-commerce retailing (Journal of Retailing and Consumer Services) vol. 39,2017, pp. 135-144

## AUTHORS PROFILE

### Joan Leonard Mushi

Education Background:

Asia Pacific University of Innovation and Technology (APU) - 2019

- Masters in Data Science and Business Analytics

Catholic University of Eastern Africa (CUEA) - 2013

- Bachelor of Arts Degree in Social Sciences. Majors in Economics.

### Dr. MudiarasanKuppusamy

Dr. MudiarasanKuppusamy is the Dean of Faculty of Business & Technology and a Professor of Innovation at University of Cyberjaya, Malaysia. He specializes in the area of digital business and transformation. His email is [drarasan@cybermed.edu.my](mailto:drarasan@cybermed.edu.my)

### Haslina Hashim

Haslina Hashim is a senior lecturer of the School of Marketing & Media of Asia Pacific University. She specializes in the field of innovation and management. Her email is [haslina.hashim@apu.edu.my](mailto:haslina.hashim@apu.edu.my)