TOWARDS SUSTAINABLE AND INNOVATIVE BUSINESS MODELS

EDITOR
Mustafa Latif EMEK

AUTHORS
Eyup ATİOĞLU
Guven PEKDEMIR
K.R.PADMA
K.R.DON
Ozlem ALPU
Tien Dung KHONG

İKSAD
Publishing House

TOWARDS SUSTAINABLE AND INNOVATIVE BUSINESS MODELS

EDITOR

Mustafa Latif EMEK

AUTHORS

Eyup ATİOĞLU

Guven PEKDEMIR

K.R.PADMA

K.R.DON

Ozlem ALPU

Tien Dung KHONG

İKSAD Publishing House

(The Licence Number of Publicator: 2014/31220)
TURKEY TR: +90 342 606 06 75
USA: +1 631 685 0 853
E mail: iksadyayinevi@gmail.com
www.iksadyayinevi.com

ISBN: 978-625-8007-11-4
Cover Design: İbrahim KAYA
September / 2021
Ankara / Turkey
Size = 16x24 c

CONTENTS

PREFACE

Mustafa Latif EMEK …….……..……1

CHAPTER 1

THE EFFECT OF THE HOT AIR BALLOON INDUSTRY ON TOURIST EXPENDITURES

Eyup ATİOĞLU *…….....……..…….……3*

CHAPTER 2

ECONOMIC APPROACHES FOR E-COMMERCE DEVELOPMENT FOR AGRICULTURAL SECTOR IN DEVELOPING COUNTRIES

Tien Dung KHONG …….……27

CHAPTER 3

APPRAISAL OF DATA MINING METHODS ON SALES DATA IN THE E-COMMERCE INDUSTRY

Guven PEKDEMIR & Ozlem ALPU…….…….....……61

CHAPTER 4

THE EXPANSION OF WOMEN ENTREPRENEURSHIP GLOBALLY: A MINI REVIEW

K. R. PADMA & K.R.DON …….……85

PREFACE

Modern problems of a globalized world, such as growing inequality and environmental degradation, as well as the dynamics of the structure of production and consumption create new needs for the development of sustainable business models of organizations, for significant investments in process and product innovations. In the current socio-economic conditions, business organizations are faced with serious challenges that require finding new approaches to stay competitive and increase productivity. Continuous innovation has always been a critical factor in achieving competitive advantage. Today, innovation is recognized as a key mechanism for solving the problems of sustainable development.

It should be admitted that, although research in this direction is being actively carried out by researchers, the available empirical substantiation of the economic efficiency of sustainable innovations is clearly insufficient at the moment. Thus, in this book the authors carried out a systematic analysis of the different aspects of sectoral, marketing and industrial problems of researching sustainable innovation.

I hope you enjoy reading and learning from the studies of best practices of business approaches.

Dr. Mustafa Latif EMEK

CHAPTER 1

THE EFFECT OF THE HOT AIR BALLOON INDUSTRY ON TOURIST EXPENDITURES

Asst. Prof. Eyup ATİOĞLU[1]

[1] Kapadokya University, School of Applied Sciences, Aviation Management Nevsehir/Turkey eyup.atioglu@kapadokya.edu.tr ORCID ID : 0000-0001-9460-4822

INTRODUCTION

In the tourism sector, which is one of the fastest growing sectors in the world, average tourist expenditures are increasing. In competitive global conditions, the growth trend in tourist expenditures is a source of motivation for both policy makers and businesses in the service market. In this context, analyzing the distribution of per capita tourist expenditures such as transportation cost, accommodation, tourist activity, food and beverage is a key element in the formulation and positioning of marketing strategies for a destination. In other words, a destination may design a strategy to attract tourists who will spend more on various activities, goods and services during their stay, rather than attracting many tourists who will spend less (Cai, Hong and Morrison, 1995).

Tourism is a gigantic sector with a predominant socio-cultural aspect, and today it has significant economic effects. For many countries, employment, income and foreign exchange resources are obtained as the return of tourist expenditures. Governments and private sector representatives, who are aware of the economic importance of tourism, transfer their resources to the development of tourism and make significant investments (Choong-Ki, Var and Blaine, 1996). In 2019, 1.5 billion people worldwide became tourism participants (UNWTO, 2020). With its high employment creation and foreign exchange earning effect, the tourism sector has a significant share in most countries' economies (Williams and Shaw, 1992).

Globally, there are many new options for today's tourists in the tourism sector, and there is constant growth, especially in terms of spending on recreation and recreational tourism activities (Mok and Iverson, 2000). Due to the size of tourist expenditures and the importance of its economic effects, the tourism sector is monitored and evaluated by governments, policy makers, destination marketers and researchers. However, from the perspective of tourists, each type of expenditure can become the determinant of the decision to participate in tourism. (Wang, Rompf, Sever, and Peerapatdit, 2006).

When tourists travel to a destination, they spend by demanding a variety of goods and services. The expenditures are a source of income for service providers and stimulate the local economy. In the long run, tourist expenditures have economic benefits at the national level, such as job creation and taxation on income generating businesses. In this context, it would not be wrong to evaluate that tourism can save economies with foreign trade deficit up to a certain level (Engström and Kipperberg, 2015; Marrocu, Paci and Zara, 2015; Paci and Marrocu, 2014). Due to its impact on economic growth, interest in tourism has increased significantly in recent years. The characteristics of the travel destination can directly affect the tourist expenditures. In this context, the identification and analysis of the factors affecting expenditures per tourist has become more important. Indeed, it should be well analyzed how much certain tourism activities can increase the number of tourists or the amount of spending in a destination (Brida, Lanzillotta, Lionetti, and Risso, 2010).

The financial and industrial-based economic crises that emerged in the world economies and the inadequacy of efforts and performance to overcome these crises lead countries to seek new ways to prevent economic recession, ensure economic growth and close the increasing foreign trade deficits. In this context, tourism, which is relatively stable, has become a more popular sector in recent years; The importance attached to tourism investments by the countries of the world has shown itself in the form of an increase in the number of tourists and tourism revenues (Erkan, Kara, and Harbalıoğlu, 2013). Tourism is recognized as one of the catalyst sectors that enables a country to achieve economic and non-economic gains. For this reason, most countries with tourism potential focus on international tourism investments and aim to both increase economic development and spread income to the base. (Çetinbaş and Bektaş, 2008: 37).

In some cases, although the number of tourists increases, the average per capita tourist expenditure may decrease. Neither policy makers nor private sector representatives desire such a situation, they try to expand their market shares by creating new tourism products or improving existing ones, and focus on increasing tourist expenditures. However, tourism businesses are faced with the dilemma of increasing capacity or increasing quality with price in order to increase tourist expenditures. On the other hand, considering the negative effects of tourism activities on natural resources, sustainability becomes important. Therefore, tourists who spend more per capita are targeted rather than more tourists (Aguilo, Rossello and Vila, 2017). Tourist

expenditures, one of the important factors of economic growth, contribute to economic diversity in most developing countries. Foreign exchange revenues from tourism enable countries to finance their budget deficits and contribute to the solution of the unemployment problem. In addition, transportation, trade, construction, accommodation, food and beverage sectors and other service sectors have a strong connection and complementarity with tourism (Aydın, Darıcı and Taşçı, 2015). It is very difficult to measure the expenditures and income in the tourism sector statistically and show them in national accounts. Likewise, the added value that emerges in the tourism sector appears as the income and employment increase created in other sectors (accommodation, transportation, telecommunication, etc.) apart from itself (Dwyer and Spurr, 2011: 2). In general, developments in the tourism sector and increases in tourism revenues have important effects on the economy and development. Increases in tourism revenues positively affect economic growth in a sustainable way. The tourism sector leads to poverty reduction, ensures the creation of rural and local economic infrastructure, and contributes to the reduction of unemployment by creating new job opportunities. It ensures the development and diversification of the socio-cultural level in destinations. It also contributes to the increase in the importance given to environmental protection and ecological balance (Surugiu, Frent and Surugiu, 2009).

Tourist expenditure is defined as the amount paid before or during a tourist visit for goods or services purchased for own use or as a gift.

Tourist expenditures per capita is one of the variables that researchers have been working on in recent years. Tourist expenditures have the ability to create a multiplier effect (Mudarra-Fernandez, Carrillo-Hidalgo and Pulido-Fernandez, 2019). There are many factors that affect the expenditures of tourists, and touristic activities have an important place among these factors. Touristic activities include participation in sports activities, nature-based activities, adventure activities, wildlife observations, etc. (Brida and Scuderi, 2013). The motivation to conduct research on per capita tourist expenditures is that the tourism sector is one of the important sectors of the economy for most countries and provides the largest foreign exchange source for the relevant countries. Therefore, the analysis of per capita tourist expenditures, as well as the total income from the tourism sector, is of great importance (Castro, Molina and Pablo, 2013; Massidda and Mattana, 2013).

Each tourism destination has its own characteristics, and these features affect the decisions of tourists when choosing a destination. In this context, it is unrealistic to think that tourists consider all destinations homogeneously in their destination selection decisions (Phakdisoth and Kim, 2007). In order to increase per capita tourist expenditures, destinations aiming to increase tourist expenditures should develop new tourism products or touristic activities (Kozak, Gokovalı and Bahar, 2008). Cappadocia, one of the most popular destinations in the world with its nature and cultural tourism as well as hot air balloon rides, has almost all the extraordinary features that today's post-

modern tourists are looking for. (Özen and Özdemir, 2019). Cappadocia, which provides a hot air balloon ride experience to approximately 500.000 tourists a year, is the largest hot air balloon industry in the world (SHGM, 2018) (Kapadokya University, 2021).

Studies on tourist expenditures in the literature have generally focused on certain variables that affect expenditures. In some studies, analyzes have been made about certain tourism destinations and it has been seen that there is no empirical study on tourist expenditures related to the Cappadocia region in Turkey. In this study, it is aimed to analyze the effect of the Cappadocia hot air balloon industry on the average tourist expenditure per capita.

1. LITERATURE REVIEW

There are many studies in the literature on tourist expenditures and the factor affecting expenditures. However, it has been observed that the studies on the basis of destination or touristic activity are limited, and this study contributes to the literature in terms of examining the economic effects of the Cappadocia hot air balloon industry.

Abbruzzo, Brida and Scuderi (2014) examined the factors affecting international tourist expenditures in Uruguay through graphical models in their study. In the study, it has been revealed that socio-demographic variables, type of accommodation, destination and duration of stay have a significant effect on tourist expenditures. Aguilo, Rossello, and Vila (2017) investigated the relationship between tourists' personal daily expenditures and their length of stay

with decomposition analysis. In the decomposition study, different types of expenditures such as the place where the expenditure is made, restaurant, transportation and travel are taken into account by distinguishing the expenditures for accommodation and non-accommodation. In the study, a positive relationship was found between the length of stay and tourist expenditures. Nicolau and Mas (2005) examined the factors affecting holiday decision making and the amount of spending with the Heckit model. In the study, it has been determined that factors such as income and the size of the destination to be visited have a positive effect on the decision to go on vacation. It was determined that the amount of expenditure was positively affected by the length of stay, accommodation, the distance between the country of origin and the destination and income.

Cheung and Law (2001) investigated the determinants of tourist expenditures in Hong Kong between 1983-1997 using time series analysis. In the research, a positive relationship was found between average hotel room prices and length of stay and tourist expenditures. Choong-Ki, Var, and Blaine (1996) examined the relationship between international tourist expenditures and the demand function in South Korea using least squares analysis. According to the findings of the study, it was determined that the income factor most affected the spending amounts of tourists coming from 8 countries included in the data set of the study. In addition, it has been revealed that prices and real exchange rate are meaningful and flexible in tourist expenditures. Erol and Hassan (2013) examined the number of tourists coming to

Turkey and tourism revenues between 2001 and 2012. In the study, in which the numerical data compiled from tourism statistics are interpreted, it has been determined that the differentiation in the nationalities of the tourists coming to Turkey affects the tourism incomes.

In their study, Saayman and Saayman (2015) analyzed the relationship between the daily expenditures of tourists coming to South Africa between 2003 and 2010 and the income variable using ARDL bounds test. In the study, it was determined that income is a determining factor in the daily expenditure amounts of tourists. Erkan, Kara and Harbalıoğlu (2013) examined the determinants of tourism revenues in Turkey between 2005 and 2012 with the VAR model and Granger causality analysis. In the study, a bidirectional causality relationship was determined between tourism revenues and the number of tourists in Turkey. In addition, no relationship was found between the real exchange rate and tourism revenues. Phakdisoth and Kim (2007) analyzed tourism demand using the least squares method in Laos for the period 1995-2004, using the number of tourists. In the study, it was determined that the tourism revenues of Laos are not affected by the income level of the tourists and the relative prices.

Kozak (2001) analyzed the individual expenditures of tourists during their holidays in Turkey with a questionnaire including open-ended questions. According to the findings of the study, it was determined that the factors affecting the amount of individual expenditure the most were the length of stay and the type of accommodation. Uysal

and Crompton (1984) analyzed the relationship between the number of tourists, international tourism revenues, income per capita, relative prices, relative exchange rate and promotional expenditures using the least squares method in their study examining the factors affecting the international tourist flow to Turkey. According to the findings of the study, it has been determined that income, prices and exchange rates have an effect on international tourist flows to Turkey. In their study, Saayman and Saayman (2008) examined the factors affecting the number of tourists coming to South Africa from different source markets classified according to continents between 1993 and 2004 with Johansen cointegration analysis. According to the findings of the study, it was revealed that income and relative prices affect the number of tourists. The study also evaluated that climate and capacity play an important role in South African tourism.

Wang, Rompf, Severt, and Peerapatdit (2006) analyzed the factors affecting the amount of spending by international tourists using a survey method. According to the findings of the study, it was determined that income and travel-related characteristics were the most effective factors affecting tourist expenditures. In addition, it has been revealed that the length of stay has a positive effect on accommodation, transportation and touristic activities as well as total expenditures. Kozak, Gokovalı, and Bahar (2008) examined the factors affecting tourist expenditures using the least squares method through the data obtained from the survey method. In the study, the factors affecting tourist expenditures were analyzed. According to the

findings of the study, the income of tourists, the type of vacation they demand and whether they have visited the same location in the past positively affect their average spending. Marcussen (2011) analyzed the factors affecting tourist expenditures in Denmark using the least squares method. While the dependent variable of the study was tourist expenditures, factors such as income, travel distance, length of stay and type of accommodation were determined as independent variables. In the study, it was determined that all independent variables in the model affected the per capita spending per night.

Naude and Saayman (2005) examined the main determinants of travel to Africa for the period 1996-2000 with panel data analysis. In the study, it has been determined that the income level of tourists, relative prices and travel costs are not determinant in the demand for African tourism. Chulaphan and Barahona (2021) examined the factors affecting tourist expenditures per capita in Thailand for the period 2010-2017 with the ARDL bounds test. According to the findings of the study, per capita tourist expenditures are positively affected by income and tourist activity prices. In addition, it has been evaluated that sustainable tourism projects can increase tourist expenditures.

2. DATA AND VARIABLES

The dataset of this study includes 60 observations on a monthly basis between 2014-01 and 2018-12. In the study, Turkey's average tourist expenditure per capita was determined as the dependent variable, while the number of tourists participating in the hot air balloon flight

in Cappadocia and the number of days suitable for flight according to meteorological conditions were determined as independent variables. Information on the variables whose natural logarithms are used is given in Table 1.

Table 1. Variable Information Used in the Study

Variables	Description of the Variable	Analysis Period	Source of Data
lnTUR	Turkey's average international tourist expenditure per capita	Between 2014-01 and 2018-12	TÜİK Database
lnUCUS	Number of tourists participating in the hot air balloon flight in Cappadocia	Between 2014-01 and 2018-12	Kapadokya University Hot Air Balloon and Airship App. and Res. Central
lnGUN	Number of days suitable for flight according to meteorological conditions	Between 2014-01 and 2018-12	Kapadokya University Hot Air Balloon and Airship App. and Res. Central

3. METHODOLOGY AND EMPIRICAL RESULTS

The fact that there is a high level of correlation between time series variables while establishing econometric models is not evidence of a causal relationship between the variables (Holden and Thomson, 1992). This relationship may be due to spurious regression. Cointegration tests such as Engle & Granger (1987) and Johansen &

Juselius (1990) are based on the stationarity of the series at the first difference, which can be an important constraint during the analyses. However, the ARDL method gives valid results regardless of whether there is a level, the first difference, or a combination of both, excluding series that are stationary in the second difference. The ARDL model is defined as below to analyze the relationship between tourist expenditures per capita (lnTUR), the number of tourists participating in the hot air balloon flight (lnUCUS) and the number of flight days (lnGUN).

$$y_t = \alpha + \sum_{i=1}^{p} \gamma_i y_{t-i} + \sum_{j=1}^{k} \sum_{i=0}^{q_j} X_{j,t-i}{}' \beta_{j,i} + \epsilon_t \tag{1}$$

In the model, *p* is the lag number of the dependent variable, q_1 is the lag number of the first explanatory variable, and q_k is the lag number of the k_{th} explanatory variable. In order to accept that there is a cointegration relationship between the variables, the calculated F statistic must be greater than the upper bound critical value (Pesaran, Shin, & Smith, 2001).

There is no need to test the variables for unit roots before ARDL estimation, the model can contain I(1), I(0) or any mutually cointegrated variable. However, a limitation of the ARDL method is that it cannot be predicted with the I(2) series. It should be shown that the lnTUR, lnUCUS and lnGUN series are not I(2), as shown in Table 2.

Table 2. Augmented Dickey–Fuller Unit Root Test

Variables	Status	Intercept	Intercept&Trend	Without intercept&trend
lnTUR	I(1)	-6.6246***	-6.5476***	-7.7101***
lnUCUS	I(0)	-3.7494***	-3.7170***	-0.3188***
lnGUN	I(0)	-3.8943***	-4.3195***	-0.7644***

a: (*)Significant at the 10%; (**)Significant at the 5%; (***) Significant at the 1%

According to the unit root test results given in Table 2, it has been determined that the series can be used in the ARDL limit test method. As seen in Figure 1, the appropriate model was determined as ARDL (1,3,0) using Akaike Information Criteria.

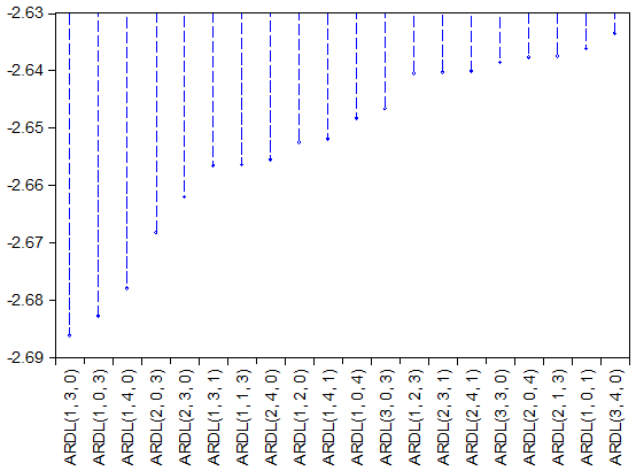

Figure 1. Akaike Information Criteria

The ARDL boundary test results, which were conducted to investigate the cointegration relationship between Turkey's average tourist expenditure per capita, the number of tourists participating in the hot air balloon flight and the number of days available for flight, are given

in Table 3. Considering the upper critical value, the F statistical value (5.0901) shows that there is a cointegration relationship between the variables at the 5% significance level.

Table 3. ARDL Bounds Test

Test Statistic	Value	k
F- statistic	5.0901	2
Critical Value Bounds		
Significance	**I0 Bound**	**I1 Bound**
10%	3.17	4.14
5%	3.79	4.85
2.5%	4.41	5.52
1%	5.15	6.36

The coefficient explaining the long-term relationship variables and the direction of the coefficient are given in Table 4. According to the findings, a 1% increase in the number of tourists participating in a hot air balloon flight increases the average tourist expenditure per capita by 0.07% in the long run. In addition, a 1% increase in the number of days suitable for flight increases the average tourist expenditure per capita by 0.08% in the long run.

Table 4. Long Run Coefficients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
lnUCUS	0.0782	0.0936	2.9327	0.0051
lnGUN	0.0838	0.0403	2.0796	0.0427
C	4.8841	1.1110	4.3960	0.0001

The ECM, which is established after the estimation of the long-term coefficients related to the short-term equilibrium between the variables, expresses the situation in which the short-term imbalances are eliminated. The error correction term coefficient in the model was determined as -0.21. The negative sign before the error correction term indicates that the long-run equilibrium relationship returns to the steady state if the system is shocked.

The results of the tests performed to investigate the normal distribution of the series, the existence of heteroskedasticity in the model and autocorrelation problems in the model results are given in Table 5.

Table 5. Diagnostic Tests

Breusch-Godfrey Serial Correlation LM Test			
F-istatistiği	0.7355	Prob. F(2,34)	0.4846
Obs*R-squared	1.6949	Prob. Chi-Square(1)	0.4285
Heteroskedasticity Test: Breusch-Pagan-Godfrey			
F- istatistiği	1.7142	Prob. F(2,36)	0.1371
Obs*R-squared	9.7248	Prob. Chi-Square(8)	0.1367
Jarque-Bera	2,1218	Prob.	0,3461

According to the findings obtained as a result of the diagnostic tests, no heteroskedasticity and autocorrelation problems were detected in the model, and it was determined that the series had a normal distribution. The ARDL model was examined in terms of structural break with the CUSUM and CUSUMQ structural break tests developed by Brown, Durbin and Evans (1975). If the curve

determined by the CUSUM and CUSUMQ test statistics is between the critical limit of 5% significance, it is accepted that the estimated coefficients do not contain structural breaks in the long run.

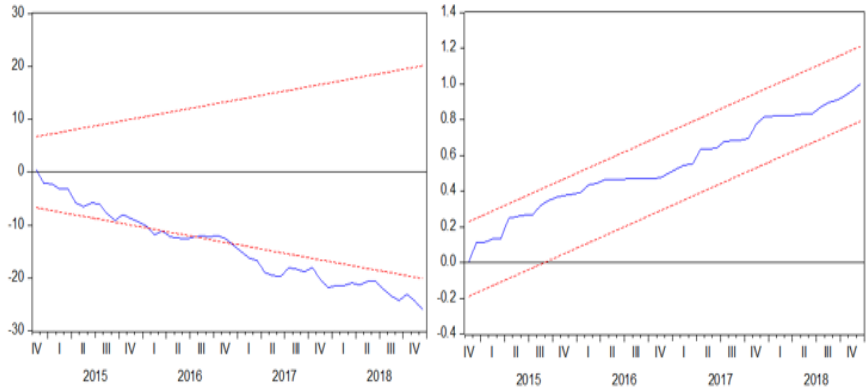

Figure 2. CUSUM (left panel) and CUSUMQ (right panel) Tests

Since the curve determined by the CUSUM test statistics is outside the 5% significance critical limit, it was determined that the estimated coefficients contain structural breaks in the long run.

Table 6. Granger Causality Test

H_0	F-statistic	Prob.
lnGUN → lnTUR	2.84159	0.0673
lnTUR → lnGUN	0.45027	0.6399
lnUCUS → lnTUR	3.19107	0.0491
lnTUR → lnUCUS	0.80841	0.4510
lnUCUS → lnGUN	3.87897	0.0268
lnGUN → lnUCUS	3.51269	0.0369

Finally, as seen in the Table 6, bidirectional causality was found between lnUCUS and lnGUN. In addition, Granger causality was determined from lnUCUS to lnTUR. Finally, a 10% significance level of causality was found from lnGUN to lnTUR.

CONCLUSIONS

The main income in tourism is tourist expenditures per capita, so the first thing to consider in tourism development movements is to seek and find ways to increase the number of tourists. Generally, the purposes and expectations of tourists are known. In order to increase the average per capita tourist expenditure, better and more attractive goods and services should be offered to tourists and detailed statistical analysis should be made for this. Cappadocia plays an important role in increasing tourist expenditures with its hot air balloon rides and its historical and cultural structure.

In this study, it is aimed to analyze the effect of the Cappadocia hot air balloon industry on the average tourist expenditure per capita. As a result of the ARDL bounds test, a cointegration relationship was found between Turkey's average tourist expenditure per capita, the number of tourists participating in the balloon flight and the number of days available for flight. According to the findings, a 1% increase in the number of tourists participating in a hot air balloon flight increases the average tourist expenditure per capita by 0.07% in the long run. In addition, a 1% increase in the number of days suitable for flight increases the average tourist expenditure per capita by 0.08% in

the long run. Although there is a structural break in the estimated long-term coefficients according to the CUSUM test, no structural break was detected according to the CUSUMQ test result.

Granger causality was determined from the number of tourists participating in the hot air balloon flight to the average tourist expenditure per capita. In addition, Granger causality was determined at 10% significance level from the number of days suitable for hot air balloon flight to the average tourist expenditures per capita.

REFERENCES

Abbruzzo, A., Brida, J. G., & Scuderi, R. (2014). Determinants of individual tourist expenditure as a network: Empirical findings from Uruguay. Tourism Management, 43, 36-45.

Aguilo, E., Rossello, J., & Vila, M. (2017). Length of stay and daily tourist expenditure: A joint analysis. Tourism Management Perspectives, 21, 10-17.

Aydın, A., Darıcı, B., & Taşçı, H. (2015). Uluslararası turizm talebini etkileyen ekonomik faktörler: Türkiye üzerine bir uygulama. Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, (45), 143-177.

Brida, J. G., & Scuderi, R. (2013). Determinants of tourist expenditure: A review of microeconometric models. Tourism Management Perspectives, 6, 28-40.

Brida, J. G., Lanzilotta, B., Lionetti, S., & Risso, W. (2010). The tourism-led growth hypothesis for Uruguay. Tourism Economics, 16(3), 765-771.

Brown, R.L., Durbin, J. and Evans, J.M. (1975). Techniques for testing the constancy of regression relationships over time. Journal of the Royal Statistical Society: Series B (Methodological), 37(2), 149-163.

Cai, L.A., Hong, G. and Morrison, A.M. (1995). Household expenditure patterns for tourism products and services, Journal of Travel & Tourism Marketing, Vol. 4 No. 4, 15-40.

Castro, M., Molina, J. A., & Pablo, M. P. (2013). Tourism and GDP. A meta-analysis of panel data studies. Journal of Travel Research, 52(6), 745–758.

Cheung, C., & Law, R. (2001). Determinants of tourism hotel expenditure in Hong Kong. International Journal of Contemporary Hospitality Management, 151-158.

Choong-Ki, L., Var, T., & Blaine, T. W. (1996). Determinants of inbound tourist expenditures. Annals of Tourism Research, 23(3), 527-542.

Chulaphan, W., & Barahona, J. F. (2021). The Determinants of Tourist Expenditure Per Capita in Thailand: Potential Implications for Sustainable Tourism. Sustainability, 13(12), 6550, 1-14

Çetintaş, H., & Bektaş, Ç. (2008). Türkiye'de turizm ve ekonomik büyüme arasındaki kısa ve uzun dönemli ilişkiler. Anatolia: Turizm Araştırmaları Dergisi, 19(1), 37-44.

Dardis R, Soberon-Ferrer H, Patro D. (1994). Analysis of leisure expenditures in the US. Journal of Leisure Research 26(4): 309–321.

Dwyer, L., & Spurr, R. (2011). Tourism Economics Summary. STCRC Centre for Economics and Policy.

Engle R.F. and Granger C.W. J. (1987). Co-integration and error correction: representation, estimation and testing, Econometrica, 55, 251–276.

Engström, T., & Kipperberg, G. (2015). Decomposing the heterogeneous discretionary spending of international visitors to Fjord Norway. Tourism Management, 51, 131–141.

Erkan, B., Kara, O., & Harbalıoğlu, M. (2013). Türkiye De Turizm Gelirlerinin Belirleyicileri. Akademik Bakış Uluslararası Hakemli Sosyal Bilimler Dergisi, (39), 1-20.

Erol, N., & Hassan, A. (2013). Türkiye'ye Gelen Turist Sayısı ile Elde Edilen Turizm Gelirlerinin Türkiye İstatistik Kurumu (TÜİK) Verilerine Göre Karşılaştırmalı Analizi. Journal of Tourism and Gastronomy Studies, 1(2), 3-14.

Holden K. & Thompson J. (1992). Co-Integration: An introductory survey, Brit Rev. Econ. Issues, 14: 1-55.

Jang S, Bai B, Hong GS, O'Leary J. (2003). Understanding travel expenditure patterns: a study of Japanese travellers to the US by income level. Tourism Management 25(3): 331–341.

Johansen, S. & Juselius K. (1990). Maximum likelihood estimation and inference on cointegration with applications to the demand for money, Oxford Bulletin of Economics and Statistics, 52 (2), 169- 210.

Kapadokya University, (2021). Balon uçuş sahasında zipline projesine ilişkin emniyet risklerinin değerlendirilmesi, https://hotairballoon.kapadokya.edu.tr/raporlar-ve-analizler/kapadokya-balon-ucus-sahasinda-zipline-projesine-iliskin-emniyet-risklerinin-degerlendirilmesi (Accessed: 03.07.2021)

Kozak, M. (2001). An analysis of tourist spending and its determinants. Anatolia, 12(2), 196-202.

Kozak, M., Gokovalı, U., & Bahar, O. (2008). Estimating the determinants of tourist spending: A comparison of four models. Tourism Analysis, 13(2), 143-155.

Marcussen, C. H. (2011). Determinants of tourist spending in cross-sectional studies and at Danish destinations. Tourism Economics, 17(4), 833-855.

Marrocu, E., Paci, R., & Zara, A. (2015). Micro-economic determinants of tourist expenditure: A quantile regression approach. Tourism Management, 50, 13–30.

Massidda, C., & Mattana, P. (2013). A SVECM analysis of the relationship between international tourism arrivals, GDP and trade in Italy. Journal of Travel Research, 52, 93–105.

Mudarra-Fernandez, A. B., Carrillo-Hidalgo, I., & Pulido-Fernandez, J. I. (2019). Factors influencing tourist expenditure by tourism typologies: A systematic review. Anatolia, 30(1), 18-34.

Naude, W. A., & Saayman, A. (2005). Determinants of tourist arrivals in Africa: a panel data regression analysis. Tourism economics, 11(3), 365-391.

Nicolau, J. L., & Mas, F. J. (2005). Heckit modelling of tourist expenditure: Evidence from Spain. International Journal of Service Industry Management, 271-293.

Özen, İ. A., & Özdemir, E. G. (2019). Kapadokya sıcak hava balonculuğu üzerine nitel bir araştırma, Nevşehir Hacı Bektaş Veli Üniversitesi SBE Dergisi, 9(2), 579-595.

Paci, R., & Marrocu, E. (2014). Tourism and regional growth in Europe. Papers in Regional Science, 93(Suppl. 1), 25-50.

Pesaran H., Shin Y. and Smith R. J. (2001). Bound testing approaches to the analysis of long run relationship, Journal of Applied Econometrics, 16(3), 289-326.

Phakdisoth, L., & Kim, D. (2007). The determinants of inbound tourism in Laos. ASEAN Economic Bulletin, 225-237.

Saayman, A., & Saayman, M. (2008). Determinants of inbound tourism to South Africa. Tourism economics, 14(1), 81-96.

Saayman, A., & Saayman, M. (2015). An ARDL bounds test approach to modelling tourist expenditure in South Africa. Tourism Economics, 21(1), 49-66.

SHGM, (2018). SHGM'den turizme önemli katkı: Balon faaliyetlerinde yeni noktalar, http://web.shgm.gov.tr/tr/haberler/5874-shgm (Accessed: 03.07.2021)

Surugiu, C., Frent, C., & Surugiu, M. (2009). Tourism and its impact upon the Romanian economy: an input-output approach. Analele Stiintifice ale Universitatii" Alexandru Ioan Cuza" din Iasi-Stiinte Economice, 56, 355-376.

UNWTO, (2020). International tourism growth continues to outpace the global economy, https://www.unwto.org/international-tourism-growth-continues-to-outpace-the-economy (Accessed: 23.06.2021)

Uysal, M., & Crompton, J. L. (1984). Determinants of demand for international tourist flows to Turkey. Tourism management, 5(4), 288-297.

Wang, Y., Rompf, P., Severt, D., & Peerapatdit, N. (2006). Examining and identifying the determinants of travel expenditure patterns. International Journal of Tourism Research, 8(5), 333-346.

Williams, A.M., & Shaw, G. (1992). Tourism and economic development: Western European experiences, 2nd edition, Belhaven, London.

CHAPTER 2

ECONOMIC APPROACHES FOR E-COMMERCE DEVELOPMENT FOR AGRICULTURAL SECTOR IN DEVELOPING COUNTRIES

Tien Dung KHONG[1], Ph.D

[1] School of Economics, Can Tho University, Vietnam
ktdung@ctu.edu.vn.ORCID ID : 0000-0002-2274-2123

While Starting

This research is aimed at analyzing relevant economic theories and approaches on e-commerce to analyze the current situation to improve the effectiveness of e-commerce applications in developing countries. This research employed direct survey data from 164 households in the Mekong River Delta of Vietnam as a case study. The methodologies included descriptive statistics, 5 - point Likert scale, and a Logit model for analyzing households' perception and determinants of e-commerce adaptation. Research results reveal that households are willing to apply e-commerce, especially through several sustainable forms such as linkages with enterprises and apply advanced production techniques. In the case of developing countries where the government is facing financial and resource constraints, in order to solve this issue, the involvement of the community and private organizations including enterprises and cooperatives in the value chain of agricultural products to guide and support people in agriculture digital transformation is essential. The local government can then perform well as an observer or controller to ensure the quality, rights, and obligations of the stakeholders involved.

1. INTRODUCTION

E-commerce is a new field and is increasingly showing a great role in trade promotion, product exchange, and sale through the global internet. In Vietnam, according to a report of the Ministry of Planning and Investment (2020), the application of e-commerce has brought high efficiency to businesses, especially in accessing customers and management and sales activities. The application of e-commerce is an inevitable trend on a global scale. As a country with a proportion of export goods reaching over US$200 billion a year, and a market of over 90 million people (Vietnam General Statistics Office, 2021), Vietnamese businesses cannot ignore this tendency. Customers in major import markets from Vietnam such as the US, EU, Japan, and Korea all have very high rates of using e-commerce. In Vietnam, according to data from the Ministry of Planning and Investment (2020), the application of e-commerce has brought high efficiency to businesses, especially in approaching customers as well as in management and sales activities.

However, one of the causes of the current problems in the agricultural sector is the limited capacity of enterprises in the production, processing, and export of agricultural products, especially in management, research, and forecasting capacity market; Trade promotion has not been paid due attention and investment. The production relations in agriculture were slowly improved. Most especially, there is no connection between sellers and buyers and relevant actors in the value chain of agricultural products. In the

agricultural sector in Vietnam today, communication channels between producers and consumers are still lacking and weak, especially for household or cooperative production units.

In addition, buying and selling through e-commerce websites is only popular for certain types of goods and services such as airline tickets, electronics, tours, fresh flowers, but not clearly shown in the agricultural sector, especially in developing countries. In developed countries, however, through the application of information technology (IT), farm households can access important information related to markets, products, legal and financial regulations, establish and maintain relationships, contact with customers and partners. As a result, realizing the importance and benefits of e-commerce, the application of electronic websites, and the implementation of trade promotion activities for agricultural products associated with farm households is an urgent problem to solve these issues and also improve the efficiency of market access of farm households in the Mekong River Delta (this area is one of the most important areas in Vietnam since it is regarded as the *rice-bowl* of the country, providing more than 45% of the country's agricultural products and 90% for export) (Vietnam General Statistics Office, 2020). Hence, it is necessary to analyse research theories and approaches to expand e-commerce for agricultural products, in order to contribute to bringing e-commerce closer to the households, improving and enhancing profits for stakeholders involved, creating brands for Vietnamese agricultural products to compete in the international markets. This research, therefore, aimed at identifying the relevant economic

theories related to e-commerce in agricultural sectors in Vietnam and the determinants of e-commerce adaptation from households. The findings then are used as a scientific basis to propose policy implications for policymakers in developing countries.

2. GENERAL INFORMATION ABOUT E-COMMERCE THEORIES AND MODELS

In e-commerce, there are three participants including Businesses (coded as B) play the role of driving force for e-commerce development, consumers (coded as C) play a decisive role in determining the success of e-commerce, and the government (coded as G) plays a guiding role for regulations and management. From the relationships between the above subjects, the following types of e-commerce transactions suggested B2B, B2C, B2G, C2G, C2C. In which B2B and B2C are the two most important types of e-commerce transactions (Nguyen Dinh Luan, 2015).

Business-to-Business (B2B): is the implementation of transactions between businesses on the Internet. Parties involved in B2B transactions include online intermediaries, buyers, and sellers. Types of B2B transactions including buying on demand when the price is right and buying on a long-term contract, based on personal negotiations between buyer and seller.

Business-to-Customer (B2C): This is an e-commerce activity that focuses on consumers rather than businesses.

Customer-to-Business (C2B): refers to e-commerce activities, in which using a reverse pricing model where customers will determine the price of products or services. This model of customer rights is more advanced.

Customer-to-Customer (C2C): These are e-commerce activities that use a customer-to-customer auction model.

Among the above-mentioned types of transactions, B2B commercial transactions and B2C commercial transactions are the main forms of e-commerce (Nguyen Dinh Luan, 2015). In this study, a combination of both B2C and B2B models is applied. In which, one side is households as producers who need to sell the products, the other side are enterprises that purchases agricultural products for processing and export.

Table 1. Comparison of traditional and modern e-commerce

Indicators	**Traditional commerce**	**E-commerce**
Meaning	Direct exchange between seller and buyer	Commercial transactions or information exchange electronically via the Internet
For farm households		
Selling products	Selling products directly	Selling many products at the same time online
Operation time	Limited by the store's operating time	24 hours operation automatically by the software
Transaction Time	Slow	Fast

Quality Check	Products are tested by authorities	Products can be checked at the purchasing time
Interact with customers	Directly	Indirect via phone or online
Provide information	Direct exchange	Product information updates continuously
Business scope	Limited to a specific region	Worldwide coverage for ease of access
Marketing	Live product display	Update product images continuously
Payment	Cash	Credit card, money transfer
Delivery	Direct delivery	All purchases and sales take place online regardless of time and place, making it easier to buy and sell, and better meet customer requirements
Cost of selling	Store renting	Significantly reduce production costs
Products	One main product	Diversify product
Business expansion	Require more time	Can increase ad budget when it's doing well without having to worry too much about meeting increased demand
Competitive market	Relatively low competition, restrict in one kind of product	Relatively high, wide range of competition
For Customers		
Accessibility	Slow (because it takes time to get to the store)	Fast
Product selection	Difficult (because you have to move among several areas)	Easy to search through keywords or information portals
Prices	Stable according to the market	Cheaper based on some promotions

Time	Need more time	Saves a lot of time shopping
Strengthen relationships with customers	Without tools to strictly manage customer information, leading to omissions or inaccurate service provision	Customer information is stored in a computer system, from there, information processing, information exchange, and service provision for customers is quick and accurate

Source: Compiled by author, 2021

Moreover, an e-commerce website associated with farm households will be a bridge for domestic and foreign businesses to find information about businesses, exchange, link, and trade products, technologies, and services in the agricultural sector, conducting direct transactions with each other at all times and places. This is also a channel for farm households and businesses to introduce and promote their images, products, and potential for cooperation to a large number of customers, thereby improving competitive efficiency.

3. METHODOLOGY APPROACHES AND RESEARCH FRAMEWORK FOR E-COMMERCE ANALYSIS

The common approach when studying the perception and behavior of the actors directly in the research object is to determine the necessary factors and the current state of the application of e-commerce, thereby as a basis for proposing appropriate recommendations. Therefore, this research employed 03 main methods to analyze the data.

Descriptive statistics: to describe farm households characteristics, providing summaries of earnings data across different modalities. This method is related to collecting data, summarizing, presenting,

calculating, and charting different characteristics to reflect in general the research object in order to analyze information about the respondents, calculate mean, maximum value, minimum value, and standard deviation.

Likert scale: is a unidirectional scale and was invented by an American psychologist namely Likert. Currently, there are several types of scales with Likert that are proposed, however, 5-point Likert scales have been most widely used (Khong et al., 2021). For variables built from a 5-point Likert scale (1 - completely disagree; 2 - disagree; 3 - normal; 4 - agree; 5 - strongly agree), the interval value is determined according to the following formula:

Distance value = (Maximum - Minimum)/n = (5 - 1)/5 = 0.8

Accordingly, the meaning of each interval value is understood as follows:

1,00 - 1,80	Completely disagree
1,81 - 2,60	Disagree
2,61 - 3,40	Normal
3,41 - 4,20	Agree
4,21 - 5,00	Completely agree

Theoretical basis to form a scale to assess people's awareness about the application of e-commerce and the need for e-commerce development based on a pilot survey in the study area, consultation with local agricultural experts based on the conditions in survey areas.

In addition, the specific criteria and characteristics of e-commerce are combined with the content in the previous research. Accordingly, these contents include assessing people's awareness about the application of e-commerce and the need to develop e-commerce for agricultural products in the future. Activities to assess people's awareness about the application of e-commerce and the need to develop e-commerce for agricultural products are summarized in Table 2.

Table 2. Description of the variables in the model to assess households' awareness of the application of e-commerce and the need for e-commerce development

Indicators	Code	Activity descriptions
E-commerce recognition	NB1	Online sales (Livestream sales, Facebook)
	NB2	Connect with businesses to consume products
	NB3	Links with local stores (Green Store, CO.OP FOOD)
	NB4	Towards the implementation of traceability for agricultural products
	NB5	Open an online sales account on the apps (Tiki, Shopee)
Apply e-commerce	AD1	Online sales (Livestream sales, Facebook)
	AD2	Connect with businesses to consume products
	AD3	Links with local stores (Green Store, CO.OP FOOD)
	AD4	Towards the implementation of traceability for agricultural products
	AD5	Open an online sales account on the apps (Tiki, Shopee)
Desired level of e-commerce application	MM1	Online sales (Livestream sales, Facebook)
	MM2	Connect with businesses to consume products
	MM3	Links with local stores (Green Store, CO.OP FOOD)
	MM4	Towards the implementation of traceability for agricultural products
	MM5	Open an online sales account on the apps (Tiki, Shopee)
Evaluate the effectiveness	DG1	Online sales (Livestream sales, Facebook)
	DG2	Connect with businesses to consume products
	DG3	Links with local stores (Green Store, CO.OP FOOD)

of e-commerce	DG4	Towards the implementation of traceability for agricultural products
	DG5	Open an online sales account on the apps (Tiki, Shopee)
Support activities from local government	HT1	Sign up for an advertisement in a local TV, newspapers about household's agricultural products
	HT2	Create favorable conditions to open trade fairs to introduce products and sell products
	HT3	Open an online business store for farm households to participate in trading
	HT4	Open a skill training course on e-commerce (Using the internet, website, paying online, updating information, promoting, customers services)
	HT5	Trade promotion solutions facilitate the expansion of trading markets (bringing products to other places to promote local agricultural products)
	HT6	Directly and widely connecting farm households with consumers and businesses (SAIGON CO.OP retail supermarket model to consume agricultural products)
Desire support from local government	DP1	Sign up for an advertisement in a local TV, newspapers about household's agricultural products
	DP2	Create favorable conditions to open trade fairs to introduce products and sell products
	DP3	Open an online business store for farm households to participate in trading
	DP4	Open a skill training course on e-commerce (Using the internet, website, paying online, updating information, promoting, customers services)
	DP 5	Trade promotion solutions facilitate the expansion of trading markets (bringing products to other places to promote local agricultural products)
	DP 6	Directly and widely connecting farm households with consumers and businesses (SAIGON CO.OP retail supermarket model to consume agricultural products)
	DP7	Support to complete the contract between farm households and enterprises

Evaluating the effectiveness of support activities	HQ1	Sign up for an advertisement in a local TV, newspapers about household's agricultural products
	HQ 2	Create favorable conditions to open trade fairs to introduce products and sell products
	HQ 3	Open an online business store for farm households to participate in trading
	HQ 4	Open a skill training course on e-commerce (Using the internet, website, paying online, updating information, promoting, customers services)
	HQ 5	Trade promotion solutions facilitate the expansion of trading markets (bringing products to other places to promote local agricultural products)
	HQ 6	Directly and widely connecting farm households with consumers and businesses (SAIGON CO.OP retail supermarket model to consume agricultural products)
	HQ 7	Support to complete the contract between farm households and enterprises

Source: Designed and compiled by the author, 2021

The theoretical basis for forming a scale to assess people's perception of the situation of applying e-commerce and the need for e-commerce development mentioned are explained as below:

NB1: This variable represents the awareness of farm households through online sales (Livestream sales, Facebook) using mobile phones to upload images of products for sale (KiotViet, 2020).

NB2: This variable represents the awareness of farm households through connecting with businesses to consume products, which is the form of a One-Commune-One-Product (OCOP) program, producers and traders to find partners to supply raw materials in these models, and the production has to follow the direction of Vietnamese

Good Agricultural Practices - VietGAP signing contracts and purchasing agricultural products from farm households (Thanh Liem, 2020).

NB3: This variable represents the awareness of farm households knowing about stores such as Green Store, CO.OP FOOD. Associate with these stores to sell agricultural products. Green Store will assign commodity standards, price, purchasing - payment methods when cooperating to supply products and goods to supermarkets (Thanh Liem, 2020).

NB4: This variable represents the awareness of farm households that the traceability activity is the activity of quickly retrieving the product origin via QR code (QR code: 2-D barcode or quick response code) printed on product packaging by using a code-scanning application on a smartphone to retrieve the date, month of manufacture, expiry date, and origin of the items, so you can be assured of the quality (Nam Dinh province, 2020).

NB5: This variable transforms farm households' awareness through opening an online sales account on the app or website (Tiki, Shoppe, Food map) is activity towards developing into a farm e-commerce platform products, helping to connect producers and farm households with restaurants, convenience store chains and consumers (KiotViet, 2020).

Moreover, variables AD1, AD2, AD3, AD4, AD5 are the variables showing whether the above activities have pressure or not, respectively.

Variables MM1, MM2, MM3, MM4, MM5 are the variables expressing the desire to apply the above activities to agricultural products of the household, respectively.

Variables DG1, DG2, DG3, DG4, DG5 are the variables to evaluate the effectiveness when applying e-commerce, respectively.

HT1: Signing up for advertisements on local television and newspapers about household agricultural products shows strong cooperation with reputable media units such as Vietnam Television, Voice of Vietnam. Local radio and television stations, the National Office of Intellectual Property of Vietnam to bring information about agricultural products in regions to a large number of consumers. This is a very effective way to contribute to building the brand image of Vietnamese agricultural products and bring that image closer to public perception (Nguyen Quoc Thinh, 2019). It is noted that the definition of support activities from local authorities are these activities include research, consult local agricultural experts, conditions to evaluate the effectiveness of agricultural cooperatives.

HT2: Facilitating the opening of trade shows and exhibitions to introduce Vietnamese products and sell products is an activity that participates in the fair to have the opportunity to exchange, exchange, and learn more about effective business methods such as advertising,

promotion, trade in goods, cooperation, businesses have had a lot of contact with partners who come to work directly for business cooperation (Thanh Liem, 2020).

HT3: Opening an online business store for farm households to participate in trading is a potential effective online sales channel activity that is the design of a sales website. Setting up a sales website offer the best business efficiency from features such as daily product information updates, utilities for easy customer care, and information exchange, link with websites with the same content to develop together to find a large number of potential customers (KiotViet, 2020).

HT4: Open a training course on e-commerce skills (Using the internet, website, paying online, updating information, promoting, promoting, customers services) (Son La Department of Industry and Trade, 2019).

HT5: Trade promotion solutions that facilitate the expansion of trading markets (bringing products to other places to promote local agricultural products) are through trade promotion activities, regional linkages, and connections supply and demand between the locality and the provinces and cities in the region as well as across the country has contributed to promoting the trade of products and goods, participating in fairs and exhibitions to promote, find partners, expanding the market, many products with strengths and specialties

of the province are known to consumers, present in large markets and supermarkets inside and outside the province (Thanh Liem, 2020).

HT6: A direct and widespread bridge between farm households and consumers, businesses (the model of retail supermarket SAIGON CO.OP consuming agricultural products) is the activity of signing commercial contracts to supply agricultural products with cooperatives and businesses in the city to create a stable supply for supermarket systems. The procurement process is organized by SAIGON CO.OP in the raw material areas of the localities, gathered at the central warehouse to control the quality of input goods and ensure distribution to all sales points in the country under the SAIGON CO.OP system (Ngoc Thuy, 2020).

HT7: Supporting the completion of the connection contract between farm households and businesses is the activity of connecting production households and enterprises to purchasing, signing cooperation contracts for production, supply, and consumption of safe agricultural products which are supported by local authorities (Thanh Liem, 2020).

Logit model identifying the determinants of e-commerce adaptation desire

The Logit regression model is a quantitative model in which the dependent variable is a dummy variable, receiving only 2 values of 1 or 0. According to Stock & Watson (2015), the Logit model has the form:

$$L_i = \ln(P_i / 1- P_i) = Z_i = \beta_1 + \beta_i X_i + u_i$$

The model for testing the statistical significance of the coefficients using z-statistics and testing the general significance of the whole model using chi-square statistics. However, the dependent variable in this research takes the value from 1 to 5, representing the level of respondents' desire to participate in e-commerce, so a special form of the Logit model is the ordered Logit model with dependent variable Y is 5 - level Likert scale, X_i is the independent variables, including the age of household head (X_1 – coded as HH-AGE). Younger household heads are more likely to apply e-commerce for agricultural products than older ones due to their awareness, health, and eagerness to learn. The expectation in this variable is negative, which means that the age affects the employee's decision to apply in the opposite direction, if the age increases by one unit (year), the ability for the households to apply e-commerce is expected to be negative, and vice versa. Research by Xiang & Sumelius (2010) showed that age influences the participation of farm households in the management of cooperatives where they can share information.

Experience of the household head (X_2 – coded as HH-EXP): this independent variable represents the number of years of experience in the household heads. It is expected that the number of households with more experience will learn and learn to apply e-commerce to develop agricultural products. The previous study on cooperative participation by Awotide (2012) showed that the more production

experience of the household head, the higher the probability of joining the cooperative.

Using a phone with an internet connection (X_3 – coded as HH-Phone with Internet): This independent variable shows that households with the phone connected to the internet receive the value 1 and vice versa, the value is 0. It is expected that households that own phones connected to the Internet will have access to technology and e-commerce applications to develop agricultural products more efficiently.

Identify activities associated with stores (Green Store, COOP FOOD) to sell local specialties (X_4 – coded as HH-PER3) takes the value 1, and vice versa. It is expected that this variable to be positive and farm households know about local agricultural products stores and connects to consume their farm products.

Identify the activity towards traceability for agricultural products (X_5 – coded as HH-PER 4) receiving the value of 1 and vice versa. Currently, it is observed that consumers focus on the safe origin of products. This variable is expected to be positive, which means that producing households are aware of this activity in order to produce safer products for consumers more efficiently.

Evaluating the effectiveness of online sales activities (Livestream sales, Facebook) (X_6 – coded as HH-EVA 1) is assessed on a 5-point Likert scale. This variable is included in the model to assess whether this activity is effective or not for both households with and without

awareness of this activity. The expectation of the variable is positive, which means that this activity is highly effective on the e-commerce adaptation desire.

Evaluating the effectiveness of connecting with businesses to consume products (X_7 – coded as HH-EVA 2) is assessed on a 5-point Likert scale. The expectation of the variable is also positive. Finally, evaluating the effectiveness of traceability activities for agricultural products (X_8 – coded as HH-EVA 4) is an essential activity for agricultural products to reach consumers to ensure safety in both origin and quality. The expected sign of the variable is positive, which means that when the farmer assesses the high efficiency of traceability for agricultural products, the likelihood of wanting to adapt to e-commerce is higher.

4. CASE STUDY OF E-COMMERCE ANALYSIS IN MEKONG RIVER DELTA VIETNAM

4.1. Data collection method

The primary data used in this research was collected by stratified random sampling method through survey questionnaires and directly interviewed with 164 farm households in the study area in the Mekong Delta of Vietnam. A stratified random sampling method is used in which households will be selected following each specific subdivision. With this sampling method, data analysis is expected to represent the whole area in the Mekong River Delta.

4.2. Results and Discussions

Farm households perceptions and awareness about e-commerce

In order to be able to apply e-commerce in digital transformation, the most necessary condition is access to the Internet. However, through the actual survey results of 164 farm households, the number of households using a phone with an internet connection accounted for just over 50%. Although this result shows that the ability of farm households to access e-commerce is quite high, the remaining 50% still do not have an Internet connection. This is also a big challenge when aiming to increase the level of Internet access to 100% if the local government wants to promote and encourage farm households to apply e-commerce into the agricultural products.

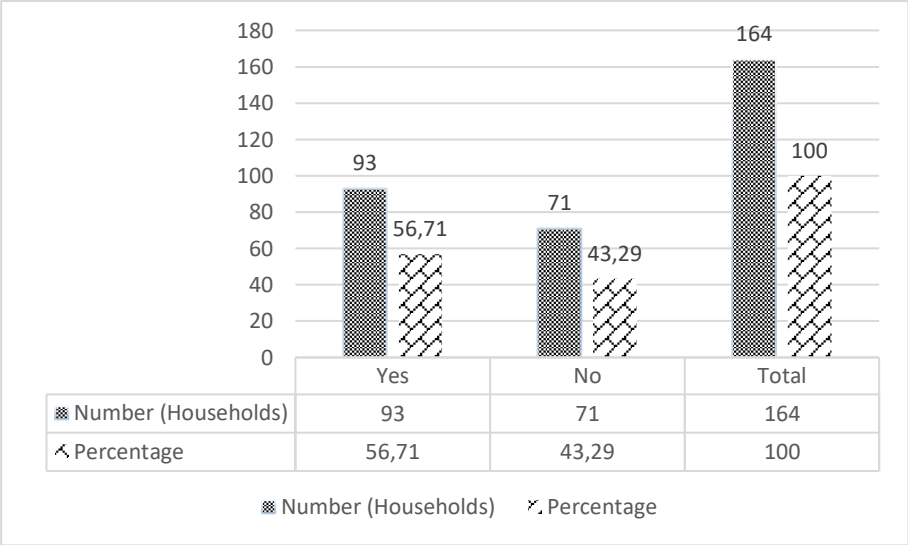

Figure 1. Number of households using mobile phones with an Internet connection

Source: Survey data, 2021

When analyzing the needs of people to apply e-commerce, most of the needs from MM1 to MM5 were agreed by the people, in which, paying special attention to MM2 and MM4 is to connect with consumer enterprises and aim to access the Internet.

Table 3. Farm households demand for applying e-commerce

Variable code	N	Mean	Std. Dev.	Min	Max	Perceive level
MM1	164	2.84	0.961	1	5	Normal
MM2	164	3.68	0.958	1	5	Agree
MM3	164	3.32	1.020	1	5	Normal
MM4	164	3.79	1.072	1	5	Agree
MM5	164	2.85	1.062	1	5	Normal

Source: Survey data, 2021.Note: Refer to Table 2 for variable names

The next question was refer to the evaluation of effectiveness when applying, the research results show that people rate it from normal (neutral) to agree (over 3). This level of efficiency is similar to their desire when the two criteria of association with businesses and traceability are also assessed to bring high efficiency.

Table 4. Households' perception of the significance of efficiency evaluation when applying e-commerce

Variable code	N	Mean	Std. Dev.	Min	Max	Perceive level
ĐG1	164	2.86	0.919	1	5	Normal
ĐG2	164	3.63	0.941	2	5	Agree
ĐG3	164	3.21	0.994	1	5	Normal
ĐG4	164	3.79	1.000	1	5	Agree
ĐG5	164	2.88	1.088	1	5	Normal

Source: Survey data, 2021.Note: Refer to Table 2 for variable names

The results of the evaluation of the desired value to support the application of e-commerce by the households are shown in Table 5. Accordingly, the majority of households were aware that this application does not have much to do with technology because they can be self-taught (DP1 and DP3, DP4). However, what they want to focus on was DP2 - Facilitating the opening of trade fairs and exhibitions to introduce Vietnamese goods and sell products, and DP5, DP6, and DP7, which were related to trade promotion and building new markets in order to achieve more complete linkage model.

Table 5. Households' perception of the significance of the desired level of support for the application of e-commerce

Variable code	N	Mean	Std. Dev.	Min	Max	Perceive level
DP1	164	3.28	0.890	1	5	Normal
DP2	164	3.43	0.873	1	5	Agree
DP3	164	3.38	0.935	1	5	Normal
DP4	164	3.23	1.206	1	5	Normal
DP5	164	3.41	0.952	1	5	Agree
DP6	164	3.69	0.848	1	5	Agree
DP7	164	3.73	0.902	1	5	Agree

Source: Survey data, 2021. Note: Refer to Table 2 for variable names

According to the survey results, the perception level if receiving support from the government on e-commerce development for agricultural products is as follows: Registering for advertising on local newspapers about agricultural products (DP1) with an average value is 3.28, open a training course on e-commerce skills (using the internet,

website, online payment, updating information at reasonable prices, promoting, promoting, customer services (DP4) with an average value of 3.23, opening an online business store for participating farm households (DP3) with an average value of 3.38, all of which are considered normal by farm households, which shows that farm households are still hesitant in wanting to support the application of e-commerce because the above activities are not popular and new to farm households. In addition, the support activities that farm households consider to be most desirable and necessary include facilitating opening trade fairs and exhibitions to introduce Vietnamese products and selling products (DP2) with an average value is 3.43. Trade promotion solutions (delivering products to other places to promote local agricultural products) (DP5) with an average value of 3.41. The connection between farm households and consumers and businesses (DP6) with an average value of 3.69. Finally, support to complete the contract to connect production and consumption between farm households and enterprises (DP7) with an average value of 3.73. These results are in agreement with the desired performance and effectiveness in the previous section.

Table 6. Households perception of effectiveness when local authorities support the application of e-commerce

Variable code	N	Mean	Std. Dev.	Min	Max	Perceive level
HQ1	164	3.23	0.888	1	5	Normal
HQ2	164	3.31	0.762	1	5	Normal
HQ3	164	3.21	0.861	1	5	Normal
HQ4	164	3.10	1.058	1	5	Normal
HQ5	164	3.47	0.962	1	5	Agree
HQ6	164	3.77	0.831	1	5	Agree
HQ7	164	3.77	0.862	1	5	Agree

Source: Survey data, 2021. Note: Refer to Table 2 for variable names

Challenges of farm households when applying e-commerce

Currently, most of the households are observed of lacking the technology adaptation ability and the ability to preliminarily process agricultural products after harvest, lack of market information, and then, decrease market power. Therefore, most agricultural products have not yet added significant economic value in the next stages of the production process to consumers, including design, marketing and consumption, and export. On the other hand, through the actual survey, the households lack concentration, mainly selling agricultural products to traders, farm households are forced to price and the output is not stable. Because of that, it is necessary to apply e-commerce to agricultural products, but in some provinces, e-commerce activities have not been developed yet. Transactions on the Internet are still unfamiliar to people, mainly because of households' awareness of e-commerce, the older households lead to adaptation ability decrease. Moreover, the habits of households in survey areas are not used to

shopping or using online services, so it is difficult for them to adapt to e-commerce too. In terms of infrastructure, services in payment and transportation are not convenient, consumers are confused when buying online, especially agricultural products. Finally, e-commerce activities have not focused on developing in rural areas where the signal is weak. So, although some households are willing to adapt, they do not have access to e-commerce activities in many measures.

Factors affecting the demand for e-commerce development of farm households

Table 7. Estimation results according to the Ologit model

Variable names	Coef. (Std. Err.)	P-value	VIF
HH-AGE	0.020^{ns} (0.013)	0.875	1.08
HH-EXP	- 0.071^{ns}(0.045)	0.109	1.09
HH-PER 3	- 0.852^{**} (0.397)	0.032	1.64
HH-PER 4	- 0.442^{ns} (0.431)	0.305	1.48
HH-Phone with Internet	1.157*** (0.376)	0.002	1.50
HH-EVA 1	0.797*** (0.188)	0.000	1.27
HH-EVA 2	0.404** (0.172)	0.019	1.13
HH-EVA 4	0.253ns (0.161)	0.117	1.15
Observations			164
Chi-square			0.0000
Log-likelihood			- 193.58785

Source: Survey data, 2021

Table 7 reveals that the regression model does not have a multicollinearity issue, the model is statistically significant with the Chi-square test value = 0.0000. Variable 3 (Associating with stores to sell local specialties) is statistically significant at 5%, with a negative

coefficient, so the probability of agreeing to adopt e-commerce decreases. This means that farm households who are interested in linking with stores to sell local specialties will be more effective than other measures. The variable of using a phone to connect to the Internet is statistically significant at a 1% level, with a positive coefficient indicates that online sales are effectively assessed by farm households, so the probability of agreeing to apply e-commerce increases with other conditions constant. Evaluation variable 1 (Online sales) is statistically significant at the 1% level with a positive regression coefficient, online sales are effectively assessed by farm households, so the probability of agreeing to apply e-commerce increases with other conditions unchanged. Evaluative variable 2 (Evaluating the effectiveness of connecting with businesses to consume products) is statistically significant at the 5% level, with a positive regression coefficient connecting with businesses to consume products is evaluated effectively by farm households, so the probability of agreeing to apply e-commerce increases with other conditions unchanged. The remaining variables of age, experience, awareness 4 (Knowledge of traceability activities), rating of 4 (Assessment of efficiency of traceability activities) are not statistically significant.

Recommendation for e-commerce development in the agricultural sector in developing countries

This finding found that there are more than 50% of households use phones with Internet connections. Taking advantage of that,

households can promote and sell their products via mobile phones. In addition to favorable e-commerce development, it is necessary to support from the government and local authorities to support households to consume products as well as develop Vietnamese agricultural products for domestic and foreign markets, specifically as follows:

(i) Support the promotion of agricultural products on television and newspapers in the province: introduce agricultural products on television stations to promote agricultural products to domestic and international markets, creating favorable conditions for market expansion.

(ii) Perform traceability for agricultural products more widely to ensure products meet current consumer requirements.

(iii) Enhancing the benefits of households when applying e-commerce: local governments need to invest in connecting with businesses to consume products, supply inputs, and ensure output, negotiate prices, and payment methods beneficial to farm households.

(iv) Support households to learn and apply e-commerce: offer training courses on e-commerce, introduce online business websites for agricultural products, open online business stores for households to participate, increase the capacity of agricultural product processing, preservation, packaging, and labeling to enhance product added value.

(v) Besides, due to the high demand for fresh food during the period of social distancing because of Covid-19, major e-commerce platforms such as Lazada and Tiki have collaborated with the partners to launch the agricultural product segment, fresh food. Recently, the agricultural product floor FoodMap Asia (foodmap.asia) officially announced the opening of the first experience store in Ho Chi Minh city. This is a step to test the O2O2O model (*online to offline to online*) to increase convenience for customers' shopping experience and increase awareness of FoodMap in the market. The O2O model is a useful solution to help small businesses improve sales efficiency through both online and direct channels. This is how sales attract online channels to their physical stores. Commerce under the O2O model will help customers collect information about products through online spaces such as email, internet advertising, website, sales channels on social networks (Facebook, Zalo), and other channels. From there, it is possible to direct customers to the company's physical stores or make decisions to buy products of the enterprise online.

With the above information, it can be seen that the O2O model is a very suitable business form for Vietnamese people when shopping is done by searching for information on the internet, can order, pay in advance, After that, customers go to physical stores to experience products and buy goods, so customers will no longer worry about the quality of goods. From the strengths of the model, it will help households to have easier access to customers, so local authorities need to give priority to supporting people to learn and apply e-

commerce through training courses on e-commerce, introduced online business websites for agricultural products, introduced the O2O model for people to better understand the model's features, combined with O2O model to improve product consumption.

CONCLUSION AND POLICY IMPLICATIONS

The analysis results reveal that the demand for e-commerce development for agricultural products is quite high, although support from local authorities has been still limited. However, in order to carry out e-commerce development activities for agricultural products, there must be consistent and close coordination between households and local authorities, the government, and enterprise with the following tasks:

For households: It is necessary to change habits in traditional production and receive e-commerce with agricultural products, actively participate in training sessions, trade fairs, and exhibitions in the locality to understand and learn about the production and consumption process.

For local authorities: Establish cooperatives to help households carry out the following activities: (i) Open training courses to introduce agricultural products on e-commerce platforms such as FoodMap; (ii) Support the implementation of activities towards traceability for key agricultural products more widely in the province to ensure quality when bringing agricultural products to the market; (iii) Support to link with agricultural products stores (Green Store, CO.OP FOOD) so that farm households have the opportunity to consume agricultural

products; (iv), support farm households, with typical products of the province to participate in online exhibitions on the e-commerce floor; participate in linking, connecting supply and demand of goods and services; (v) Support to connect with enterprises to sell agricultural products to farm households.

For the government: (i) Deploy human resources to train farm households on e-commerce, introduce online business stores, for people to grasp the current situation of online production and consumption of products. ; (ii) Support to formulate an online agricultural business store for households to participate; (iii) Building standard and featured booths to ensure market access for some key products of the province to evaluate and replicate the O2O model for agricultural products.

REFERENCES

Awotide, D. O. (2012). Assessment of women's participation in cooperative societies and its determinants in Yewa North local government area of Ogun State, Nigeria. Asian Journal of Agriculture and Rural Development, 2(393-2016-23836), 344-350.

Khong, T. D., Loch, A., & Young, M. D. (2020). Perceptions and responses to rising salinity intrusion in the Mekong River Delta: What drives a long-term community-based strategy?. Science of The Total Environment, 711, 134759.

KiotViet (2020), Three channels for selling agricultural products online, 2020. Retrieved at https://kiotviet.vn/3kenh-tinh-doanh-nong-san-online-hang-trieu-khach-dung-dai-bo-qua-neumuon-doanh -thu-dot-pha/ on September 10, 2020.

Ministry of Planning and Investment (2020). Socio-economic overview in the first 9 months of 2020 Hau Giang province. Retrieved at http://www.mpi.gov.vn /Pages/tinbai.aspx?idTin=48081&idcm=503# on 01 October 2020

Nguyen Dinh Luan (2015). E-commerce overview in Vietnam. Financial magazine. Retrived at <http://tapchitaichinh.vn/nghien-cuu--trao-doi/trao-doi-binhluan/tong-quan-ve-thuong-mai-dien-tu-o-viet-nam-100291 .html > on September 8, 2020.

Nam Dinh Province (2020). Traceability of agricultural products Enterprises and consumers mutually benefit, 2020). Retrieved at https://namdinh.gov.vn/ portal/Pages/2020-11-26/Truy-xuat-nguon-gocnong-san-Doanh-nghiep-nguoi-tiiiguwu.aspx on September 11, 2020].

Nguyen Quoc Thinh (2019). Developing the brand of Vietnamese agricultural products based on exploiting regional and regional factors. Financial magazine. Retrieved at <http://tapchitaichinh.vn/nghien-cuu-trao-doi/phat-trien-thuong-hieu-nong-sanviet-nam-dua-tren-khai-thac-yeu-to-vung-mien-306014.html> on September 11, 2020.

Ngoc Thuy (2020). SAIGON CO.OP - "midwife" of agricultural products of Vietnamese cooperatives. Saigon newspaper. Retrieved at https://www.sggp.org.vn/saigon-coop-ba-do-nong-san-cua-hoptac-xa-viet-691476.html on September 10, 2020.

Stock, J. H., & Watson, M. W. (2015). Introduction to Econometrics 3rd ed .

Son La Department of Industry and Trade (2019). Promoting e-commerce application in enterprises. Retrieved at https://socongthuong.sonla.gov.vn /1289/31002/59116/543565/cai-cach-hanhchinh/day-manh-ung-dung-thuong-mai-dien-tu-trong-doanh-nghiep-htx on September 10, 2020.

Thanh Liem (2020). Vinh Long Newspaper, Connecting trade, consumption, and processing of agricultural products). Retrieved at <http://tintucmientay.com.vn/vinh-long-ket-noi-Giao-thuong-tieuthu-che-bien- nong-san-a281362.html> on September 11, 2020.

Vietnam General Statistics Office (2021). Report on social economic situation of Vietnam. Retrieved at https://www.gso.gov.vn/du-lieu-dac-ta/2021/01/noi-dung-muc-dich-tong-dieu-tra-kinh-te-2021/ on September 01, 2021.

Xiang, L. Y., & Sumelius, J. (2010). Analysis of the Factors of Farmers' Participation in the Management of Cooperatives in Finland. Journal of Rural Cooperation, 38(886-2016-64639), 134-155

CHAPTER 3

APPRAISAL OF DATA MINING METHODS ON SALES DATA IN THE E-COMMERCE INDUSTRY

Guven PEKDEMIR[1] & Prof.Dr. Ozlem ALPU[2]

[1] Eskisehir Osmangazi University, Faculty of Science and Letters, Department of Statistics, Eskisehir, Turkey. gpekdemir8@gmail.com, ORCID: 0000-0002-4652-4091

[2] Eskisehir Osmangazi University, Faculty of Science and Letters, Department of Statistics, Eskisehir, Turkey. oalpu@ogu.edu.tr, ORCID: 0000-0002-2302-2953

INTRODUCTION

Nowadays, a great deal of information is collected concerning individuals in different environments and for different purposes. With the help of this information, based on the different characteristics of individuals, predictions are made about their purchasing preferences, needs, and behavioral characteristics (Patel et al., 2014; Bhate & Pasha, 2014; Ismail et al., 2015; Wei et al. 2015; Cirqueira et al., 2020).

It is of great importance to determine which items of the data obtained at this point is meaningful in the decision-making process and which items are worthless. It is also necessary to determine which variables are important for us since the variables related to the characteristics of individuals have a great effect on the accuracy of the results obtained in the estimation and decision-making processes (Wang, Zhang, & Fu, 2010; Brown, 2014; Larose & Larose, 2015).

With the development of computer software, uncovering hidden patterns in a huge number of data, and predicting the future, are concerns researchers and decision-makers are working on. To draw meaningful conclusions from big data, data mining methods analyze the behavioral patterns of individuals and estimate their future behavior (Niu et al. 2002; Vallamkondu & Gruenwald, 2003; Shen et al., 2004; Chen, 2017).

Data mining methods used for classification and prediction are becoming widespread in many disciplines and are being used in the field of e-commerce (Wang et al., 2002; Shen et al., 2004; Raghavan,

2005; Jiang & Yu, 2008; Cheng & Xiong, 2009; Feng et al., 2012, Chajri & Fakir, 2014; Tope-oke et al., 2019).

Electronic commerce, or e-commerce, has become increasingly common in recent years, reaching a high level with the COVID-19 pandemic. Different products and services are offered to people electronically. Companies that conduct the sales of their products and services in the Internet environment store their sales data in their databases due to the nature of the work. They are able to direct their sales through the data in their databases. However, over the years, the recorded data has increased and become uncontrollable. The search for solutions for this and similar problems was started and the idea of data mining was proposed as a result (Banks & Said, 2006).

Although e-commerce and data mining are different areas from each other, it can be said that they also have common points. Thanks to e-commerce, much data emerges, which means input for data mining (Ting, 2014).

With the development of e-commerce, it is necessary to transform the big data formed in the enterprises into useful information to increase the profit of the enterprise. For this purpose, data mining applications on e-commerce data are required. Data mining is the search for relations and rules that will enable us to predict the future from large amounts of data using computer programs (Kantardzic, 2020).

Through the application of data mining on e-commerce data, various studies can be carried out to recognize the customer of the enterprise and to increase its profits, such as estimating potential customers,

determining customer habits, and organizing customer-oriented campaigns. With the infrastructure of electronic commerce, consumers' preferences, habits, and demographics can be monitored and a personalized trade relationship can be established between the purchaser and seller of products or services using this information (Ledolter, 2013).

In short, data mining provides unexpected relationships from big data and summarizes data with new useful information to the data user (Larose, 2005).

In this research, using supervised learning techniques in data mining and the IBM Marketing Customer Value Data set, significant variables are determined to classify the profitable customers of companies. The different supervised learning techniques discussed are compared in terms of accuracy, reliability, and validity. Therefore, in the customer value dataset, the value of data mining classification algorithms in terms of accuracy, validity, and reliability are attempted to be determined.

1. DATA MINING ON E-COMMERCE

The research applies data mining classification techniques on e-commerce data, to examine how the accuracy, reliability, and validity levels of the techniques vary under different conditions, and to reveal significant classification information for the company to provide more profit from its customers.

Within this context, it would be appropriate to examine the data mining techniques used to predict profitable customer response in three phases. The phases of the research were determined as follows:

What are the reliability and validity values obtained using the 10-fold cross-validation method of the results obtained by Naive Bayes, Artificial Neural Networks, J.48, the Logistic Model, Support Vector Machines, Random Forest, and the K Nearest Neighbor methods used in the profitable customer response prediction in the IBM Marketing Customer Value data set?

What are the reliability and validity values of the results obtained by Naive Bayes, Artificial Neural Networks, J.48, the Logistic Model, Support Vector Machines, Random Forest, and the K Nearest Neighbor methods used in predicting profitable customer response in the IBM Marketing Customer Value dataset in case of 5%, 10%, 15%, 20%, 25%, 30% and 35% of the tested data?

What is the level of accuracy of the results obtained by the Naive Bayes, Artificial Neural Networks, J.48, the Logistic Model, Support Vector Machines, Random Forest, and the K Nearest Neighbor methods, which are used in predicting profitable customer response in the IBM Marketing Customer Value dataset, with the help of the confusion matrix?

1.1. Data and Data Preprocessing

In the study, the IBM Watson Marketing Customer Value dataset from the Kaggle database (https://www.kaggle.com/pankajjsh06/ibm-watson-marketing-customer-value-data) was used. It is a dataset publicly available by IBM Watson Analytics to understand customer demographics and purchasing behavior, and to identify the most profitable customers. The dataset consists of 9,134 observations and twenty-four variables that provide information regarding customers' purchasing behavior, customers' sales channels, customer lifetime value, insurance types and coverage, policy offer renewal types, state of residence, size and class of customer vehicles, location code, and customer demographics such as gender, employment status, marital status, education level, and income. Although there is no missing value in the study, profitable customer response was determined as the dependent variable.

During preprocessing performed in the study, ‘customer ID’, ‘effective to date’ and ‘number of open complaints’ variables that were not suitable for the research were ignored from the dataset. After this, outliers were removed from the dataset after applying a logarithmic conversion to ‘customer lifetime value’, ‘income’ and ‘total claim amount’ variables. As a result of the preprocessing procedures, there are 8,976 observations and twenty-one variables in the dataset. The names of the variables used in the study are shown in Table 1.

Table 1: List of Variables Names

Variable Name	Variable Name
1.State	12.Months Since Last Claim
2.Customer Lifetime Value	13.Months Since Policy Inception
3.Response	14.Number of Policies
4.Coverage	15.Policy Type
5.Education	16.Policy
6.Employment Status	17.Renew Offer Type
7.Gender	18.Sales Channel
8.Income	19.Total Claim Amount
9.Location Code	20.Vehicle Class
10.Marital Status	21.Vehicle Size
11.Monthly Premium Auto	

1.2. Data Analysis

After the dataset was prepared with preprocessing, the customer responses were classified into 'yes' and 'no', with 1,284 and 7,692 observations, respectively, according to the results of the two categories of profitable customer responses obtained in the first phase of the analysis. In the second phase of the analysis, a model was created to estimate profitable customer responses by using independent variables. Within this context, Naive Bayes, Artificial Neural Networks, J.48, the Logistic Model, Support Vector Machines, Random Forest, and the K Nearest Neighbor methods, which are popular estimation techniques of data mining, were used. In the third phase of the analysis, the accuracy, reliability, and validity values were examined for profitable customer responses, without separating the dataset as testing and training.

In an experimental study by Kohavi (1995) with data mining techniques, it was stated that cross-validation 10-fold and 20-fold

scores would be a more appropriate option than separating one third of the dataset as training and testing. Additionally, to determine statistically significant variables to predict profitable customer response, the Best First Forward, Best First Backward, and Greedy Stepwise methods, which are variable selection methods, are used in the study.

While the correct classification rate, correct classification number, Cohen's Kappa statistic, mean absolute error, square root error, relative absolute error, and relative square root error values are all used as reliability criteria in the study, the true-positive (TP) rate, false-positive (FP) rate, sensitivity, recall, F- measure, Matthew correlation coefficient (MCC), area under the receiver operating characteristic (ROC) curve, and area under the precision-recall curve (PRC) were used as validation criteria. Detailed formulations of the measures used are given in Witten et al. (2017), Bramer (2020).

2. FINDINGS

The descriptive statistics, missing values, and variable names determined to be statistically significant on the profitable customer response are shown in Table 2 and Table 3.

Table 2: Descriptive Statistics of the Numerical Variables

Variable Name	Min.	Max.	Average	Missing Values (%)
Customer Lifetime Value	3.278	4.865	3.795	0
Income	0	5	3.462	0
Monthly Premium Auto	61	245	91.742	0
Months Since Last Claim	0	35	15.067	0
Months Since Policy Inception	0	99	48.046	0

Number of Policies	1	9	2.967	0
Total Claim Amount	0.732	3.367	2.526	0

Table 3: Descriptive Statistics of the Categorical Variables

Variable Name	Levels of Variable	Count	Missing Values (%)
State	Washington	780	0
	Arizona	1,676	
	Nevada	864	
	California	3,108	
	Oregon	2,547	
Response	No	7,691	0
	Yes	1,284	
Coverage	Basic	5,522	0
	Extended	2,684	
	Premium	769	
Education	Bachelor	2,697	0
	College	2,644	
	Master	727	
	High School	2,576	
	Doctoral	331	

Table 3: Descriptive Statistics of the Categorical Variables (**to be cont.)**

Variable Name	Levels of Variable	Count	Missing Values (%)
Gender	Female	4,563	0
	Male	4,412	
Location Code	Suburban	5,714	0
	Rural	1,682	
	Urban	1,579	
Marital Status	Married	5,196	0
	Single	2,434	
	Divorced	1,345	
Policy Type	Corporate Auto	1,930	0
	Personal Auto	6,671	
	Special Auto	374	
Policy	Corporate L3	994	0
	Personal L3	3,361	
	Corporate L2	582	
	Personal L1	1,218	
	Special L2	162	
	Corporate L1	354	
	Personal L2	2,092	
	Special L1	66	
	Special L3	146	

Renew Offer Type	Offer1	3,686	0
	Offer2	1,412	
	Offer3	2,870	
	Offer4	1,007	
Sales Channel	Agent	5,584	0
	Call Center	2,283	
	Web	430	
	Branch	403	
Vehicle Class	Two-Door Car	1,864	0
	Four-Door Car	4,587	
	SUV	1,778	
	Sports Car	480	
	Luxury Car	123	
	Luxury SUV	143	
Vehicle Size	Medium	6,307	0
	Small	1,735	
	Large	933	

When Table 2 and Table 3 are examined, the number of variables that can be used to predict profitable customer response is twenty-one. Different variable selection algorithms are used to reduce the number of variables and to identify the variables that best predict the profitable customer response. The results obtained using the Best First-Forward, Best First-Backward, and Greedy Stepwise algorithms used for this purpose are shown in Table 4.

Table 4: Results of the Variable Selection Algorithms

Number	Best First-forward	Best First-backward	Greedy Stepwise
1	Employment Status	Employment Status	Employment Status
2	Months Since Policy Inception	Months Since Policy Inception	Months Since Policy Inception
3	Renew Offer Type	Renew Offer Type	Renew Offer Type
4	Total Claim Amount	Total Claim Amount	Total Claim Amount
5	Monthly Premium Auto	Monthly Premium Auto	Monthly Premium Auto
6	Months Since Last Claim	Months Since Last Claim	Months Since Last Claim

As shown in Table 4, it was determined that the six variables predicting profitable customer response did not change even when different algorithms were used. The validation of the result was also tested using the ten-fold cross-validation method.

2.1. Findings of the First Phase

To compare the reliability values of the different data mining methods with 10-fold cross-validation, correct classification numbers, correct classification rates, Kappa statistics, mean absolute error, square root error, relative absolute error, and relative square root error measures were estimated and the results are presented in Table 5.

Table 5: Results of the Reliability Measures from Data Mining Methods with 10-fold cross-validation

Algorithm	Correct Classification Numbers	Correct Classification Rates	Kappa Statistics	Mean Absolute Error	Square Root Error	Relative Absolute Root Eror	Relative Square Root Error
1.Naive Bayes	7835	87.30	0.217	0.210	0.323	85.61	92.23
2. Artificial Neural Networks	7845	87.40	0.259	0.186	0.315	75.95	89.88
3. J.48	8581	95.61	0.838	0.050	0.197	20.35	56.22
4.Logistic Model	7835	87.30	0.217	0.207	0.321	84.27	91.84
5. Support Vector Machines	7824	87.18	0.218	0.128	0.358	52.29	102.28
6. Random Forest	**8914**	**99.32**	**0.976**	**0.042**	**0.097**	**16.93**	**27.74**
7. K Nearest Neighbor	8359	93.13	0.766	0.060	0.228	24.31	65.10

According to Table 5, it can be seen that the best results in terms of the number of correct classification, correct classification rate, Kappa statistics, mean absolute error, square root error, relative absolute

error, and relative root square error values are obtained from Random Forest method. Additionally, it can be seen that the worst results in terms of the number and rate of correctly classified observations, square root error, and relative square root error are obtained using the Support Vector Machines method, while the worst results in terms of average absolute error and relative absolute error are obtained using the Naive Bayes method. Furthermore, it was determined that the correct classification number, correct classification rate, and Kappa statistics results of the Naive Bayes and Logistic regression model were the same.

Validation criteria based on weighted averages obtained from the results of classification as 'yes' and 'no', which is profitable customer response, using the Naive Bayes, Artificial Neural Networks, J.48, Logistic Model, Support Vector Machines, Random Forest and K Nearest Neighbor methods, are shown in Table 6.

Table 6: Results of the Validation criteria from Data Mining Methods

Algorithm	TP Rate	FP Rate	Precision	Recall	F-Score	MCC	ROC Area	PRC Area
1. Naive Bayes	0.873	0.730	0.864	0.873	0.834	0.311	0.759	0.869
2. Artificial Neural Networks	0.874	0.695	0.858	0.874	0.841	0.330	0.798	0.884
3. J.48	0.956	0.033	0.964	0.956	0.958	0.845	0.978	0.967
4. Logistics Model	0.873	0.730	0.864	0.873	0.834	0.311	0.764	0.871
5. Support Vector Machine	0.872	0.726	0.858	0.872	0.833	0.303	0.573	0.784
6. Random Forest	**0.993**	**0.001**	**0.994**	**0.993**	**0.993**	**0.973**	**1.000**	**1.000**
7. K Nearest Neighbor	0.931	0.015	0.953	0.931	0.937	0.787	0.977	0.966

According to Table 6, all of the validation criteria results obtained from the random forest method are found to be higher than the other methods. Compared to the areas under the ROC curve, the results obtained using Support Vector Machines have poor classification in explaining the model, while the results of the Random Forest, J.48, and K Nearest Neighbor methods have perfect classification, and the results obtained using other methods have moderate classification.

When the results obtained regarding the reliability and validation criteria are evaluated as a whole, if it is desired to classify and estimate the profitable customer response as a dependent variable, it can be seen that the best results are the Random Forest, J.48, K Nearest Neighbor, Artificial Neural Networks, Logistics Regression, Naive Bayes, and Support Vector Machines methods, respectively. When sorting, it may be enough to evaluate the correct classification rate and Kappa statistics, but as can be seen in this study, the correct classification number, the correct classification rate, and Kappa statistics of Logistics Regression and Naive Bayes methods are equal. Here, by examining the error statistics, it is preferable to have a low error rate. Compared to the average absolute error, square root error, relative absolute error, relative square root error values of the Logistics Regression and Naive Bayes methods, it can be seen that the Logistic Regression method has fewer errors for each criterion.

2.2. Findings of the Second Phase

Table 7 shows the true positive rates obtained from different data mining methods and the testing ratios to be 5%, 10%, 15%, 20%, 25%, 30%, and 35% of the dataset.

Table 7: Results of the True Positive Rates from Different Ratios of the Testing Data

TP Rate	5%	10%	15%	20%	25%	30%	35%
1. Naive Bayes	88.42	87.29	86.85	87.19	86.85	86.81	87.07
2. Artificial Neural Networks	88.64	87.40	86.40	87.52	86.99	86.96	87.49
3. J.48	94.21	95.76	96.43	95.43	94.03	92.83	92.49
4. Logistics Model	88.42	87.29	86.85	87.19	86.85	86.81	87.07
5. Support Vector Machine	88.42	86.96	86.63	87.02	86.77	86.78	87.07
6. Random Forest	**98.66**	**99.33**	**99.41**	**99.39**	**99.06**	**98.92**	**99.04**
7. K Nearest Neighbor	91.09	92.53	92.94	92.98	91.27	91.23	90.48

When Table 7 is examined, it can be seen that true positive ratios of the Naive Bayes, Artificial Neural Networks, Logistic Regression Model, and Support Vector Machines methods reach the highest value when the testing data ratio is 5%, while the values of the J.48 and Random Forest methods have the highest value when the testing data ratio is 15%. In the K Nearest Neighbor method, on the other hand, when the testing data set ratio is 20%, it can be seen that the true positive ratio reaches its highest value, and then the values decrease.

Table 8 shows how the mean square root of errors varies with different methods and the ratios of testing datasets.

Table 8: Results of Mean Square Root of Errors from Different Ratios of the Testing Data

TP Rate	5%	10%	15%	20%	25%	30%	35%
1. Naive Bayes	0.307	0.321	0.325	0.322	0.326	0.327	0.325
2. Artificial Neural	0.302	0.318	0.320	0.320	0.321	0.322	0.320

Networks							
3. J.48	0.229	0.184	0.178	0.207	0.233	0.250	0.259
4. Logistics Model	0.308	0.321	0.325	0.321	0.326	0.326	0.324
5. Support Vector Machine	0.340	0.361	0.366	0.360	0.364	0.364	0.360
6. Random Forest	**0.105**	**0.093**	**0.094**	**0.106**	**0.114**	**0.114**	**0.121**
7. K Nearest Neighbor	0.259	0.232	0.223	0.231	0.251	0.250	0.260

According to Table 8, it can be determined that the Random Forest method gives the lowest error value at each value of the testing ratio compared to other methods.

When the mean square root of the errors is evaluated, it can be seen that the Naive Bayes, Artificial Neural Networks, Logistic Model, and Support Vector Machines methods give the lowest error value when the ratio of the testing set is 5%, while errors of the J.48 and K Nearest Neighbor methods are lowest in 15% testing data. Furthermore, the Support Vector Machines method gives the highest error values.

2.3. Findings of the Third Phase

The correct classification ratios obtained from the confusion matrix results of the entire dataset, without separating the data as testing and training set, are shown in Table 9.

Table 9: Confusion Matrix and Accuracy Rates for Different Data Mining Methods

Method	**Response Level**	**Confusion Matrix**		**Correct Classification Rate %**
		Yes	No	
1. Naive Bayes	Yes	7639	52	87.92
	No	1032	252	
2. Artificial Neural Networks	Yes	7543	148	87.25
	No	996	288	
3. J48	Yes	7619	72	99.20
	No	0	1284	
4. Logistic Model	Yes	7643	48	87.30

	No	1092	192	
5. Support Vector Machine	Yes	7614	77	87.04
	No	1086	198	
6. Random Forest	Yes	7691	0	100
	No	0	1284	
7. K Nearest Neighbor	Yes	7312	379	95.78
	No	0	1284	

According to Table 9, it has been determined that the Random Forest method has a 100% correct classification rate by correctly classifying the values of the profitable customer response variable on the entire dataset. The Support Vector Machines method has the lowest correct classification rate at 87.04%. The highest correct classification ratio after the Random Forest method is the J.48 method, with its rate being 99.20%. It can be seen that the J.48 method, which has a correct classification rate of 95.61% with a 10-fold cross-validation, has a higher classification ratio of 99.20% when the entire dataset is processed.

CONCLUSION

In this study, comparative information is provided for the determination of profitable customer response using data mining algorithms in the field of e-commerce.

For this purpose, seven classification methods, which are frequently used in the literature, are used and, initially, the most important features in profitable customer response are determined using a 10-fold cross-validation and variable selection algorithms. Variable (Feature) selection is a problem for any organization that profits from

customer segmentation. The benefit of addressing this problem with cross-validation and variable selection methods is obvious.

The study consists of three phases.

In the first phase of the research, data mining reliability and validity values are used with the 10-fold cross-validation method without separating the data as a training and testing set. The Random Forest method was found to be the best classification method in all of the compared reliability measures. Furthermore, the Random Forest method gave the best results for all of the compared validation measures. When the data mining methods are examined together, it is determined that there is no difference in the sorting of the methods in terms of validity and reliability criteria. In addition, the Random Forest, J.48, and K Nearest Neighbor methods are found to be more successful in terms of reliability and validity criteria.

In the second phase of the study, the reliability and validity values of the results obtained in the case of 5%, 10%, 15%, 20%, 25%, 30%, and 35% of the testing set for different data mining methods are compared. It can be seen that the Random Forest method has the highest performance in terms of true positive rates and square roots of mean errors at each method. When the methods are compared on the basis of the square roots of the mean errors, the lowest error values in all sample groups are listed as the Random Forest, J.48, K Nearest Neighbor, Artificial Neural Networks, Logistic Model, Naive Bayes, and Support Vector Machines, respectively. However, when the

methods are evaluated, it can be seen that the mean square root of the errors of the J.48 and K Nearest Neighbor methods fluctuate at different testing rates, so the results are unstable.

When the correct classification rates obtained from the confusion matrix calculated for the entire data set are compared, the Random Forest method is determined to have the highest classification rate by correctly classifying all observations.

Considering the all of the results, more stable and reliable estimations can be obtained by examining the rate of correct classification with various fold cross-validation values of many variables instead of taking a limited number of independent variables to predict profitable customer response.

Based on the findings of the study, the following recommendations are made.

In addition to the measures related to the reliability of the findings obtained in the study, it is suggested that error-based measures, which have an important place in statistics and data mining, should be given.

To cope with the overfitting problem in data mining, it is recommended to use methods such as cross-validation instead of the training and testing set.

In addition to interpreting the results according to the accuracy measures obtained from the confusion matrix, it is important for the validation of the results that criteria such as sensitivity, specificity, the

area under the ROC curve, the Matthew correlation coefficient, F-criterion, and Recall must be reported.

To find the performance of data mining methods and the optimal result of other measures, it is recommended trying k-fold cross-validation values of between 5% and 35% for more stable findings.

REFERENCES

Banks, D. L., & Said, Y.H. (2006). Data Mining in Electronic Commerce, Statistical Science, Vol. 21, No. 2, A Special Issue on Statistical Challenges & Opportunities in Electronic Commerce Research (May 2006), pp 234-246.

Bhate, D. V. & Pasha, M. Y. (2014). Analyzing Target Customer Behavior Using Data Mining Techniques for E- Commerce Data, International Journal of Innovative Research in Computer Science & Technology (IJIRCST), Vol.2, No.1, pp 16-19.

Bramer, M. (2020). Principles of Data Mining. Springer-Verlag London Ltd.

Brown, M. S. (2014). Data Mining for Dummies. Hoboken, New Jersey: John Wiley and Sons, Inc.

Chajri, M., & Fakir, M. (2014). Application of Data Mining in E-Commerce, Journal of Information Technology Research, Vol.7, No.4, pp 79-91, October-December 2014

Cheng, Y., & Xiong, Y. (2009). Application of Data Mining Technology in E-Commerce, 2009 International Forum on Computer Science-Technology & Applications, pp 291-293, Washington, USA.

Chen, Y. (2017). The Application of Data Mining in Electronic Commerce, Advances in Computer Science Research (ACSR), vol.76. 7th International Conference on Education, Management, Information & Mechanical Engineering (EMIM 2017), pp 1017-1020.

Cirqueira D., Hofer M., Nedbal D., Helfert M., & Bezbradica M. (2020). Customer Purchase Behavior Prediction in E-commerce: A Conceptual Framework and Research Agenda. In: Ceci M., Loglisci C., Manco G., Masciari E., Ras Z. (eds) New Frontiers in Mining Complex Patterns, NFMCP 2019. Lecture Notes in Computer Science, vol. 11948. Springer, Cham. doi:10.1007/978-3-030-48861-1_8

Feng Y., Hemin J. & Huirnin Q. (2012). Study on The Application of Data Mining for Customer Groups Based on The Modified ID3 Algorithm in The E-

Commerce, 2012 International Conference on Computer Science & Information Processing(CSIP), pp 615-619, doi:10.1109/CSIP.2012. 6308929.

IBM Watson, Marketing Customer Value Data. (2019). Retrieved from https://www.kaggle.com/pankajjsh06/ibm-watson-marketing-customer-value-data

Ismail, M., Ibrahim, M., Sanusi, Z. & Nat, M. (2015). Data Mining in Electronic Commerce: Benefits and Challenges. International Journal of Communications, Network & System Sciences, 8, pp 501-509. doi: 10.4236/ijcns.2015.812045.

Jiang, Y., & Yu, S. (2008). Mining E-Commerce Data to Analyze the Target Customer Behavior, First International Workshop on Knowledge Discovery & Data Mining (WKDD 2008), pp 406-409.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection, The International Joint Conference on Articial Intelligence, 2, pp 1137-1143.

Kantardzic, M. (2020). Data Mining: Concepts, Models, Methods, and Algorithms, IEEE Press, John Wiley & Sons, Inc.

Larose, D. T. (2005). Discovering Knowledge in Data: An Introduction to Data Mining, John Wiley & Sons, Inc.

Larose, D. T., & Larose, C. D. (2015). Data Mining and Predictive Analytics, John Wiley & Sons, Inc.

Ledolter, J. (2013). Data Mining and Business Analytics with R, New Jersey: John Wiley & Sons, Inc.

Niu, L., Van, X. W., Zhang, C. Q., & Zhang, S. C. (2002). Product Hierarchy-Based Customer Profiles for Electronic Commerce Recommendation, Proceedings of 2002 International Conference on Machine Learning & Cybernetics, vol. 2, pp 1075- 1080.

Patel, M., Karvekar, S., & Mehta, Z. (2014). Customer Behavior Model Using Data Mining. International Journal of Advanced Technology in Engineering & Science, Vol.02, Special Issue No. 01, pp 709-716.

Shen, L., Hawley, J., & Dickerson, K. (2004). E-Commerce Adoption for Supply Chain Management in US Apparel Manufacturers. Journal of Textile & Apparel, Technology & Management, Vol.4, No.1, pp 1-11.

Srinivasa Raghavan, N.R. (2005). Data Mining in E-commerce: A Survey, Sadhana, vol.30, no.2, pp 275-289.

Ting, D. (2014). International Trade E-Commerce Based on Data Mining, 2014 IEEE Workshop on Advanced Research & Technology in Industry Applications (WARTIA), pp 703-705.

Tope-Oke, A. , Afolalu, C. & Omofade, O. (2019). A Data Mining Based Approach to Customer Behaviour in an Electronic Settings. Journal of Computer & Communications, 7, pp 42-53. doi: 10.4236/jcc.2019.75004.

Vallamkondu, S. & Gruenwald, L. (2003). Integrating Purchase Patterns and Traversal Patterns to Predict HTTP Requests in E-Commerce Sites, IEEE International Conference on E-Commerce (CEC 2003), pp 256-263.

Wang, H., Zhang, L., & Fu, H. (2010). The Applications of Data Mining in Electronic Commerce, 3rd International Conference on Advanced Computer Theory & Engineering(ICACTE), pp 604-610.

Wang, J.C., David, C.Y., & Chris, R. (2002) Data Mining Techniques for Customer Relationship Management. Technology in Society, 24, 483-502. doi:10.1016/S0160-791X(02)00038-6

Wei, G. T., Kho, S., Husain, W., & Zainol, Z. (2015). A Study of Customer Behaviour Through Web Mining. Journal of Information Sciences & Computing Technologies, 2(1), pp 103-107.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C.J.(2017). Data Mining: Practical Machine Learning Tools and Techniques (4th ed.). Burlington, MA: Morgan Kaufmann.

CHAPTER 4

THE EXPANSION OF WOMEN ENTREPRENEURSHIP GLOBALLY: A MINI REVIEW

K. R. Padma[1] & K. R. Don[2]

[1]Assistant Professor, Department of Biotechnology, SriPadmavatiMahila VisvaVidyalayam (Women's) University, Tirupati,AP. Corresponding author: Email id: thulasipadi@gmail.com
(Mobile: 8897146167)

[2] Reader, Department of Oral Pathology, Saveetha Dental College, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Velappanchavadi, Chennai, Tamil Nadu, India. Email Id: drkrdon@gmail.com (Moblie :9443215893)

While Starting

The women entrepreneurship rates are expanding each and every year. Today, the success of women in all disciplines is in fact greater in comparison with males. More business is run by women in our Indian society. The women's are more likely to dominate globally in all organizations and revealed to tackle problems related to cultural, financial as well as political. Although, still they are facing gender glitches in several areas of jobs. The development of women tycoons has led to a provoked perception in dominion of entrepreneurship. On the other hand, the progression of women magnates in business and other operational fields can significantly expand our Indian economical growth and further can mitigate poverty, enhance social enclosure. Nevertheless, it is a known reality that in India the population rate of women in more dominant than men yet limitations in entrepreneurial activities. Therefore, our study focussed on establishing gender neutral task in operating business or operating other kinds of ventures. Thus there is vital need to incorporate policy for reservation for women in all cooperation's to encourage women participation to become best entrepreneurs. Hence, our articles throws lights on advantages of women as business persons and challenges they faced to reach greater heights and further future plan directions.

INTRODUCTION

Women Entrepreneurship is an upsurging field of research which has provoked greater interest among several researchers, academicians, institutes, universities plus policy makers. The women entrance in business led to economic extension and prosperity Schramm (2006), and Baumol et al (2007). The accelerating growth of women tycoons contributed to improve poverty levels along with rise of per capita income and employment creation Kumar, S. M. et al 2013, Ogidi, A. E. (2014). Women moguls represent the robust growing assemblage worldwide and have received notice in all fields. In accordance to the earlier literature reports which emphasized the imperative role of women as entrepreneurs (Noguera et al., 2013) who significantly attempted to increase the economy (Kelley et al., 2017; Hechevarría et al., 2019) by providing support to especially women employees, thereby augmenting the gross domestic product (GDP) (Bahmani-Oskooee et al., 2013; Ayogu and Agu, 2015).

Earlier in the 1990s, several academic journals plus eminent newspapers in the US noticed female possessed firms as proprietorship in business (Baker et al. 1997). Although, mostly men focussed firms were most common in the field of business. Various reports have also revealed the gender bias in ventures (Jennings and Brush 2013). Further, in 20th century many articles have highlighted the women involvement at a greater extent to expand the economy as well as provide jobs to unemployed women to mitigate poverty along with supporting women to build their confidence (Sarfaraz et al.,

2014). Although, various prevailing theoretical concepts in literature about women entrepreneurship but still some obstacles were there in journey (Allen and Truman, 1993; SmithHunter, 2006). Therefore, strive for existence should be kept as a key concept to achieve success in life. Though, it is very tough to manage business as well as family life for many women. The main objective in this paper is to highlight the point that strong motivation is the major element to achieve greater heights along with self-confidence to tackle all problems. Hence our article highlights the success achieved as well as challenges faced to become a successful tycoon.

2.Reports on Female Entrepreneurship

The beginning of female entrepreneurs was from 19th century. Schwartz in 1976, contributed his first theoretical paper on women entrepreneurship in the Journal of Contemporary Business. Although, talk on women as entrepreneurs was conducted at the Babson college in 1981. In that conference, Hisrich and O'Brien (1981) made his first presentation about women tycoons. Goffee and Scase, 1985, were initial authors to contribute book to readers on female entrepreneurs. From the 19th century the establishment of women in all disciplines slowly started to increase. Further, in 2000, the second grand academic Diana International conference was held to acknowledge women entrepreneurs. Subsequently, several reports are published in leading journals as well as special issues on women entrepreneurship in 2006 -2012 (de Bruin et al. 2006 and Hughes et al., 2012).

Even, Global Entrepreneurship Monitor (GEM, http://www.gemconsortium.org/) also published a special report on women and entrepreneurship from 2006-2015. As per the literature reports analysis the female entrepreneur's index was enhanced from 2015 in approximately 77 countries. Several studies related to females beginning as entrepreneurs along with reputed journal portraying their career aspects (Shown in table-1 and table-2).

Table-1: First reports on Females as Entrepreneurs

YEAR	REPORTS	AUTHOR REFERENCES
1976	The first report on females as entrepreneurs	Schwartz, E. (1976)
1983	The first conference presentation about women entrepreneurship	Hisrich, R.D., & Brush, C.G. (1983)
1985	The first book publication about female tycoons	Goffee, R., & Scase, R. (1985)
2003	The second international conference about research on women entrepreneurship	Diana International Conference
2006	GEM Report on Women and Entrepreneurship	Global Entrepreneurship Monitor's (GEM)

Table-2: Several journals published and portrayed about females as entrepreneurs

Author and year of Publication	In the Journal Published	Primary factors
Bowen and Hisrich (1986)	Academy of Management Review	Portrayed career perspectives in women
Birley (1989)	Journal of Small Business Management	The diversification between male and females as entrepreneurs and depicted the conditions as well as experiences in between them
Fischer et al. (1993)	Journal of Business Venturing	How effective men and women run business in their own strategies
Mirchandani (1999)	Gender, Work and Organization	Enhancement of female entrepreneurs and investigate the relationship between gender, occupation and structural organization.
Ahl (2006)	Entrepreneurship Theory and Practice	The factors to be considered for necessity to expand the research object primarily focussing on society and gender.
Ahl and Nelson (2010)	International Journal of Gender and Entrepreneurship	The major focus is on the understanding what can be done in understanding the process of "carrying entrepreneurship" in terms of "what women do" and "what men do".

Goyal and Yadav (2014)	Psychosociological Issues in Human Resource Management	The journal report focussed on the issues faced by female tycoons in developing countries
Henry et al. (2016)	International Small Business Journal	Major focus in this journal is related to post-structural feminist approaches
Meyer 2018	Polish Journal Of Management Studies	Main focus on barriers and challenges faced by female entrepreneurs
Irene Kamberidou 2020	Journal of Innovation and Entrepreneurship	The modification of women entrepreneurs in the 21st century and challenging traditional understandings for their professional success.

2.Women's Identity as Entrepreneurship and Challenges faced

Women is regarded as All in All as they have to balance between work-life and family expectations. The social environment interaction primarily provides identity to women. For women's identity among the society, expectations should match such as the household duties balance, if larger family greater obstacles and these are the major challenges related to family life but on the other hand gender based bias is another major problem (Chengadu & Scheepers 2017, Rajani & Panicker 2017 and Anwar & Rashid 2012). Nsengimana et al. (2018) focussed on comprehension of females in patriarchal society and emphasized about the religion, culture plus customs that reinforce

gender gaps. Women's face major challenges and the chief question that arise among the population is "Do you think woman are effective in the running of your business?" Can the balance the work-life with family expectations, The last question that arises is "Are they capable than men?"

Although, the Women's entrepreneurial potential remained unrecognized in India. Nevertheless, the society in India is acting as a chief barrier and restricting them in a sphere and avoiding them to enter as an extrovert person. In India the society frames certain rules among women's as homemaker, child-bearer and as well as to take care of elderly people. The other obstacles in women's barrier is lack of proper education along with basic knowledge in computers plus English knowledge. Thus, these obstacles can be overcome to create their own identity in society to become a renowned entrepreneur.

3.Future directions

The future direction for women to become as a successful entrepreneual is to focus on expansion on economic growth of India. Further, promoting gender integration-inclusion, females must concentrate on upcoming sectors like Blue Economy plus Silver Economy. Nevertheless, the blue growth emphasize the actions related to seas, oceans and coasts which are the major areas to expand economy. The other sectors for expansion of economy of India is to focus on tourism, poultry, fisheries, maritime transport which provides employment to many and further increases the GDP.

CONCLUSION

Our article focussed majorly on the upsurge of women entrepreneurship based on literature reports from the late 1990s to latest 2020s. There is a major drastic development noticed in the published articles, with several exclusive issues to advance the body of knowledge on women entrepreneurship. The success for females as entrepreneurs is based on self-satisfaction which is solely recognized as key element for women entrepreneurs to enhance their own business. Therefore, our current article provide as a broad insight to reader to start-up their own business as well as helps them to gain knowledge about the challenges faced by women plus also have comprehension about dealing the problem with much confidence. Thus, our article provides encouragement to females to become a successful entrepreunal and start their own business without gender bias and support other women by providing employment along with expansion of our economy.

REFERENCES

Ahl, H., & Nelson, T. (2010). Moving forward: Institutional perspectives on gender and entrepreneurship. International Journal of Gender and Entrepreneurship, 2, 5–9.

Allen, S. and Truman, C. (1993). Women in Business: Perspectives on Women Entrepreneurs. Routledge: London. annual global entrepreneurship symposium.

Anwar, M.U & Rashid, A.G (2012). Female entrepreneurs – a review of the literature and proposed conceptual framework. Institute of Business Administration (IBA), Karachi. Female Entrepreneurs. Proceedings of 2nd International Conference on Business Management (ISBN: 978-969-9368-06-6).

Ayogu, D. U., and Agu, E. O. (2015). Assessment of the contribution of women entrepreneur towards entrepreneurship development in Nigeria. Intern. J. Curr. Res. Acad. Rev.3, 190–207.Availableonlineat: http://www.ijcrar. com./vol-3-10/Ayogu,%20Deborah%20U.%20and%20Agu,%20Everistus %20Ogadimma2.pdf.

Bahmani-Oskooee, M., Kutan, M. A., and Xi, D. (2013). The impact of economic and monetary uncertainty on the demand for money in emerging economies. Appl. Econ. 45, 3278–3287. doi: 10.1080/00036846.2012. 705430 .

Baker, T., Aldrich, H. E., & Liou, N. (1997). Invisible entrepreneurs: The neglect of women business owners by mass media and scholarly journals in the USA. Entrepreneurship and Regional Development, 9, 221–238.

Baumol, W. J., Litan, R. E., & Schramm, C. J. (2007). Good Capitalism, Bad Capitalism, and The Economics of Growth and Prosperity. New Haven & London: Yale University Press.

Birley, S. (1989). Female entrepreneurs: Are they really different? Journal of Small Business Management, 27, 32–37.

Bowen, D.D., & Hisrich, R.D. (1986). The female entrepreneur: A career development perspective. Academy of Management Review, 11, 393–407.

Chengadu, S. & Scheepers, C. (2017). Women leadership in emerging markets: featuring 50 women leaders. Routledge, (ISBN 1138188964, 9781138188969) https://books.google.gr/books?id=AplNvgAACAAJ&dq=Women+Leadership+in+Emerging+Markets:+Featuring+46+Women+Leaders&hl=el&sa=X&ved=0ahUKEwi7OTencrVAhWnDsAKHQBvADgQ6AEIJTAA.

de Bruin, A., Brush, C. G., & Welter, F. (2006). Introduction to the special issue: towards building cumulative knowledge on women's entrepreneurship. Entrepreneurship Theory and Practice, 30, 585–593.

Diana International Conference on Women's Entrepreneurship Research (http://www.babson.edu/Academics/centers/blankcenter/globalresearch/diana/Pages/home.aspx).

Fischer, E. M., Reuber, A. R., & Dyke, L. S. (1993). A theoretical overview and extension of research on sex, gender, and entrepreneurship. Journal of Business Venturing, 8, 151–168.

Goffee, R., & Scase, R. (1985). Women in charge: The experiences of female entrepreneurs. London: George Allen and Unwin.

Goyal, P., & Yadav, V. (2014). To be or not to be a woman entrepreneur in a developing country? Psychosociological Issues in Human Resource Management, 2(2), 68–78.

Hechevarría, D., Bullough, A., Brush, C., and Edelman, L. (2019). High-growth women's entrepreneurship: fueling social and economic development. J. Small Business Managem. 57, 5–13. doi: 10.1111/jsbm.12503.

Henry, C., Foss, L., & Ahl, H. (2016). Gender and entrepreneurship research: A review of methodological approaches. International Small Business Journal, 34(3), 217–241.

Hisrich, R.D., & O'Brien, M. (1981). The woman entrepreneur as a reflection of the type of business. In K.H. Vesper (Ed.), Frontiers of entrepreneurial research (pp. 54–67). Boston, MA: Babson College.

Hughes, K. D., Jennings, J. E., Brush, C. G., Carter, S., & Welter, F. (2012). Extending women's entrepreneurship research in new directions. Entrepreneurship Theory and Practice, 36, 429–442.

Irene Kamberidou. (2020). "Distinguished" women entrepreneurs in the digital economy and the multitasking whirlpool. Journal of Innovation and Entrepreneurship. Vol-9:3,2-26.

Jennings, J. E., & Brush, C. G. (2013). Research on women entrepreneurs: challenges to (and from) the broader entrepreneurship literature? The Academy of Management Annals, 7(1), 663–715.

Kelley, D. J., Baumer, B. S., Brush, C., Green, P. G., Mahdavi, M., Majbouri, M., et al. (2017). Global Entrepreneurship Monitor 2018/2017 Report on Women's Entrepreneurship. Babson College: Smith College and the Global Entrepreneurship Research Association.

Kumar, S. M., Mohan, H. C., Vijaya, C., & Lokeshwari, M. (2013). The role of women entrepreneurship in modern world. International Journal of Current Engineering and Technology, 100-104.

Meyer N (2018). Research on female entrepreneurship: are we doing enough? Polish Journal of Management Studies.Vol-17:2, 158-169.

Mirchandani, K. (1999). Feminist insight on gendered work: New directions in research on women and entrepreneurship. Gender, Work and Organization, 6, 224–235.

Noguera, M., Álvarez, C., and Urbano, D. (2013). Socio-cultural factors and female entrepreneurship. Intern. Entrepreneurship Managem. J. 9, 183–198. doi: 10.1007/s11365-013-0251-x.

Nsengimana, S. Chux, G. I., & Robertson, K. T. (2018). The downside of being a female entrepreneur in Kigali, Rwanda. SOCIOECONOMICA - Scientific Journal for Theory and Practice of Socio-economic Development 6 (12), 151-164. http://www.socioeconomica.info/jspui/handle/11171/267, http://hdl.handle.net/11171/267.

Ogidi, A. E. (2014). Women Entrepreneurship and Poverty Reduction. Journal of Business and Entrepreneurship, 1(1), 1-8.

Rajani, S., & Panicker, S. (2017). Encouraging women into entrepreneurship- a case study of Manik Ajay Patwardhan. Zenith International Journal of Multidisciplinary Research, 7(1), 75–97http://www.zenithresearch.org.in/images/stories/pdf/2017/JAN/ZIJMR/8_ZIJMR_VOL7_ISSUE1_JAN_2017.pdf.

Sarfaraz, L., Faghih, N., & Majd, A. A. (2014). The relationship between women entrepreneurship and gender equality. Journal of Global Entrepreneurship Research, 2(1), 1–11.

Schramm, C. J. (2006). The Entrepreneurial Imperative. New York: Collins.

Schwartz, E. (1976). Entrepreneurship: A new female frontier. Journal of Contemporary Business, 5, 47–76.

Smith-Hunter, A. (2006). Women Entrepreneurs Across Racial Lines: Issues of Human Capital, FinancialCapital and Network Structures, Edward Elgar Publishing.

ISBN: 978-625-8007-11-4