MARKETS AND CONSUMERS

WHAT INFLUENCES ONLINE SHOPPING OF INDIVIDUALS FROM EUROPEAN COUNTRIES?

TINO KUJUNDZIC,
MARIO JADRIC,
MAJA CUKUSIC

Faculty of Economics
University of Split, Croatia

JEL Classifications: O33, L81, L86

Key words: Online shopping, data mining, Orange Canvas, CN2 rules.

Abstract: The trend of broadband Internet expansion in conjunction with the increasing orientation of consumers towards buying via web shops, all combined with increased usage of e-banking services have contributed largely to the growth of online shopping trend. This paper deals with determining the influence of the chosen input variables (reading online magazines and newspapers, searching for product information online, using web TV, radio and e-banking services) on the observed target variable (online shopping, categorized by the level of its development in terms of individuals in European countries). The database was preloaded with data from EUROSTAT consisting of values for the abovementioned variables for 29 European countries in the period from 2007 to 2009. For the data mining process, the open source application Orange Canvas was used.

ISSN: 1804-0527 (online) 1804-0519 (print) Vol.7(1), PP. 16-20

Introduction

The use of information and communication technologies (ICT) is one of the main progress-driving forces within the society and economy for more than a decade. It is important to emphasise that these changes have been taking place almost uniformly in countries going through different phases of development (Carayannis and Sagi, 2002). The official statistics, commonly public and free to use (as are over 90% of business information) (Javorović and Bilandžić, 2007), are crucial for tracking recent changes and conducting further research in this topic.

In 2005, the European Commission set out a new strategic framework for the Information Society titled *i2010 - a European Information Society for growth and employment* (EUROSTAT, 2005). It is a key element of the renewed Lisbon Strategy that offers a comprehensive strategy for the ICT and media sector.

The Eurostat Information Society statistics are the key to monitoring EC's three priorities (ibid.):

a. The completion of a single European information space;

b. Innovation and investment in ICT research;

c. Achieving an inclusive European information society.

The abovementioned *i2010 benchmarking framework* was approved by Member States and the European Commission in 2006, which set out a comprehensive set of benchmarking indicators on Internet and broadband take-up by citizens and businesses, and on the use of computers and online services (European Commission, 2009). With the aid of i2010 benchmarking indicators, this research will strive to determine what influences online shopping of individuals in certain European countries. Although this subject matter is analysed from the aspect of individuals, it does not encompass the usual factors i.e. the causes of online shopping (like in Panian and Klepac, 2003), but it leans solely to correlations of online shopping levels with the observed variables. To reach a conclusion about what influences online shopping several data mining (DM) methods were used. First of all, clustering was done not with the application of K-means or other algorithms (Klepac and Mršić, 2006) but manually. Aside from the usual statistical methods like scatter plotting for shaping our conclusions first parallel coordinates (Huh and Park, 2008) and then CN2 rules (Boswell and Clark, 1991) were considered considerably useful.

Open source, cross-platform data mining application Orange Canvas was used for conducting the research (see Orange website for more details). One of its key advantages is the support of the above mentioned statistical/data mining methods by built-in widgets. Widgets are "interactive graphical user interface components… or graphical user interface wrappers around data analysis algorithms implemented in Orange and Python" (Zupan et al., 2010, p.1).

In a broader sense, the information society is a topic of research in various domains; i.e. it is not fully covered by economic researches exclusively, but also by those of business-sociological (Atak and Erturgut, 2010), or even educational nature (Hosgorur and Bilasa, 2009).

However, in the case of research using DM methods, similar to one presented in this paper, a trend of extensive exploration of consumers' shopping preferences is noticeable. For instance, Rygielski, Wang, and Yen (2002) use DM techniques for Customer Relationship Management (CRM) examination, while Wang and Hong (2006) provide insight to Customer Profitability Management (CPM) systems. There are more studies using DM aiming to analyse process of customer acquisition, retention and growth of its profitability in

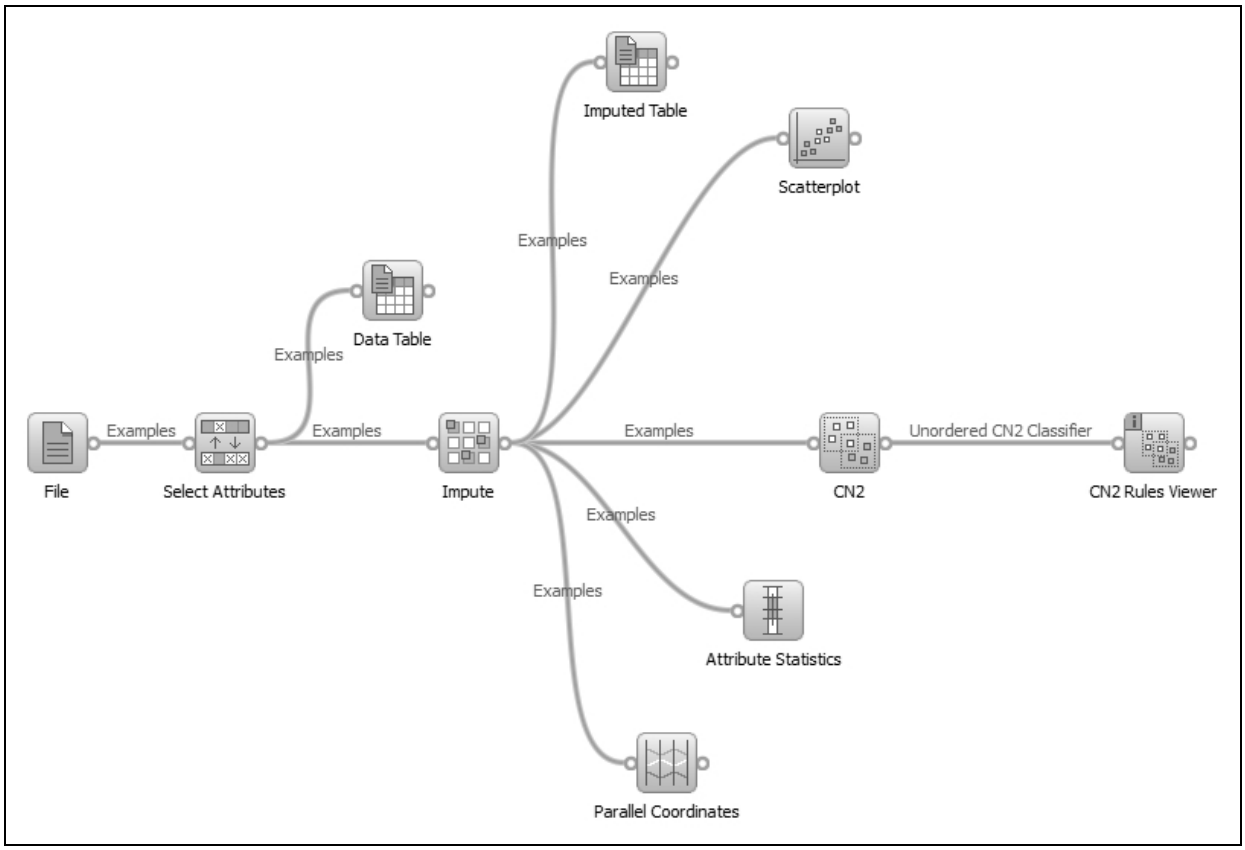
general (see Lessmann and Voss, 2009; Chu, Tsai and Ho, 2007; Chiu and Tavella, 2008).

Methodology and data acquisition

For the purpose of this research the tables containing the intensity of individuals' use of internet services from the mentioned EUROSTAT i2010 Benchmarking indicator category for the years 2007, 2008 and 2009 were acquired. Each of the tables was downloaded into Microsoft Excel (XLS) format along with variable descriptions after which they were all merged into one summary table.

Summary table holds 464 entries with only a small number of missing data (this was resolved later on using Orange Canvas). 29 European countries were listed with data values for total of 16 attributes (5 attributes x 3 years), one of which was the manually calculated class attribute serving as the target variable in the further research. The results of the research are presented through results of the parallel coordinates, CN2 rules and scatter plot widget. This is illustrated in Figure 1 where the DM model prepared in Orange Canvas is shown.

FIGURE 1. THE MODEL CREATED IN ORANGE CANVAS



As the goal of this research is to prove the influence of the observed input variables on the target variable of online shopping, the categorization of the target variable was made first: the average value of online shopping variable for the observed time period was calculated as 28.83%. Then values were used to group observations into three categories: 1 (low): 0-20%, 2 (medium): 20-40%, 3 (high): 40-100%.

The input variables ("influencers") for the three observed years (from 2007 to 2009) are: searching information about products and services online; using E-banking services; using web radio and web TV services; and reading newspapers and magazines online.

Results of the research

Upon loading the input data set, attributes that were to be analysed were selected (via the "Select Attributes" widget). After selecting the class attribute, online shopping, as the target variable and defining the first row from the data table containing the titles of countries as "meta variable" along with the input variables, it is necessary to validate selected data through the widget "Data Table". Issue of few missing entries were resolved by "Impute" widget in the next phase.

Parallel coordinates were used to visualise the influences of input variables to the class target variable of online shopping (for each of the three defined categories). Figure 2 does illustrate that in a rather high extent input variables do influence the target variable. It is important to point out that level of searching information about products and services online, as well as using e-banking services affect the most favourably and in the strongest extent the distribution of target variable values (number of individuals from European countries grouped by the intensity of shopping online).

FIGURE 2. PARALLEL COORDINATES WITH FEATURED MEANS, DEVIATIONS AND CORRELATIONS AMONG THE NEIGHBOURING VARIABLES

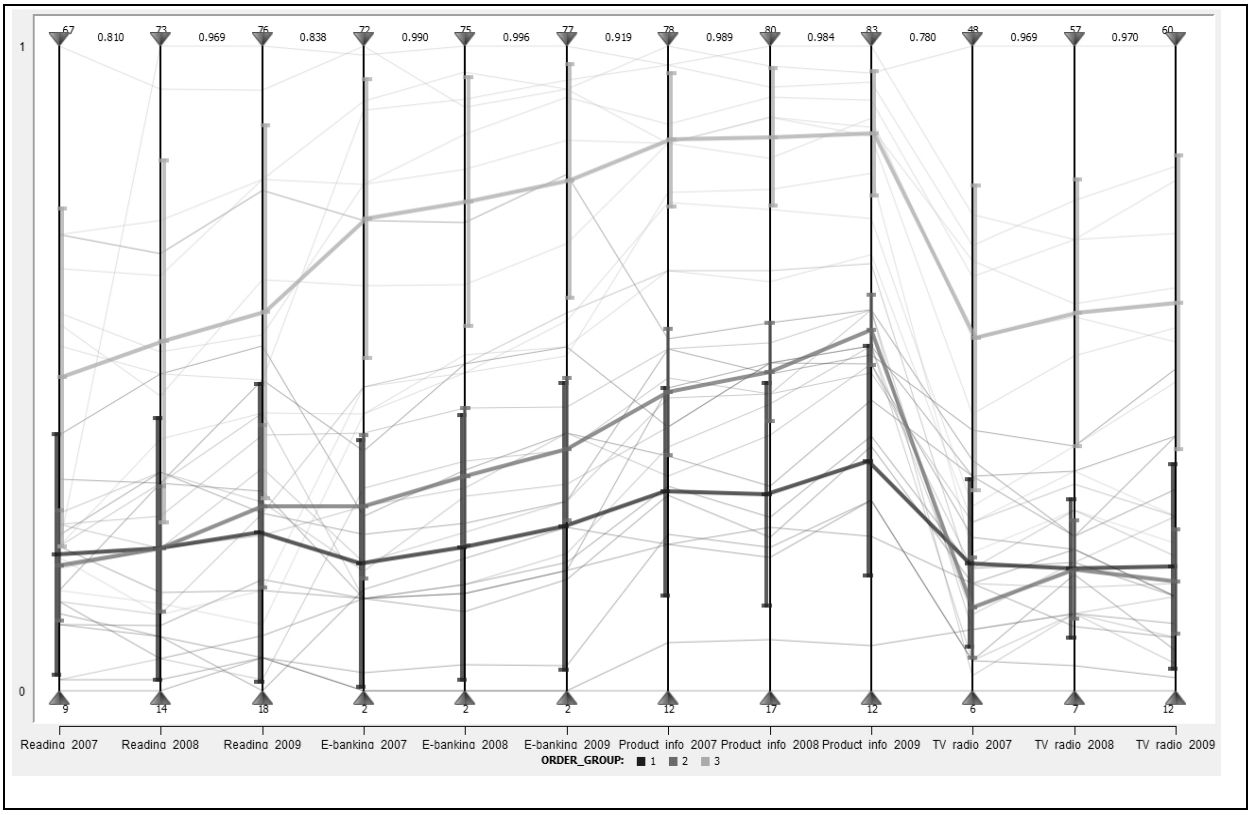


FIGURE 3. THE OVERVIEW OF GENERATED CN2 RULES

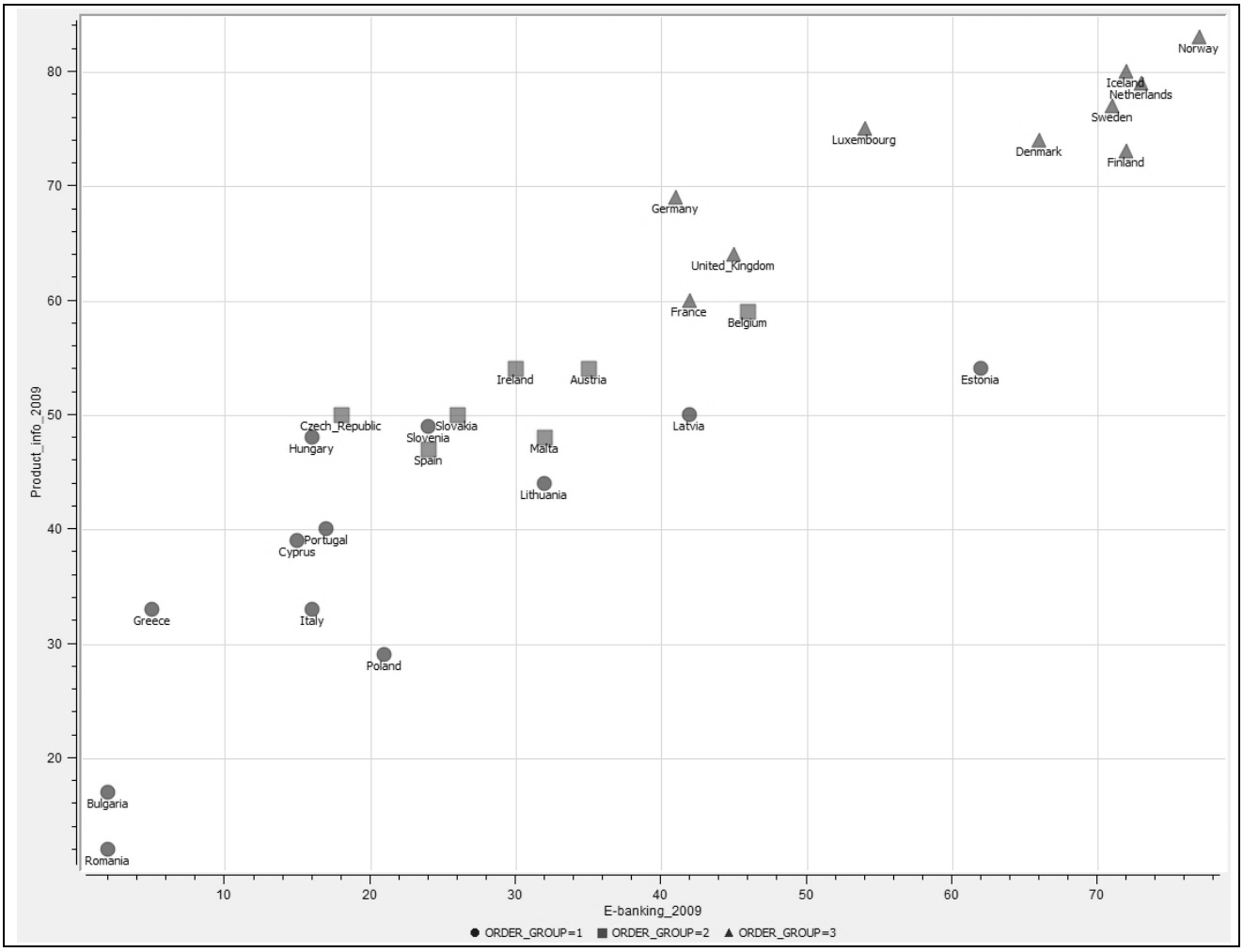
1	0.900	8.0	1	<8.0,0.0,0.0>	IF Product_info_2008.000<=37.000 THEN ORDER_GROUP=1.000
3	0.833	4.0	1	<4.0,0.0,0.0>	IF Product_info_2007.000<=48.000 AND TV_radio_2007.000>14.000 AND Reading_2008.000>27.000 THEN ORDER_GROUP=1.000
2	0.875	6.0	2	<0.0,6.0,0.0>	IF TV_radio_2007.000<=14.000 AND Product_info_2007.000>33.000 THEN ORDER_GROUP=2.000
1	0.917	10.0	3	<0.0,0.0,10.0>	IF Product_info_2009.000>59.000 THEN ORDER_GROUP=3.000

The other two input variables (using web radio and web TV services, and reading newspapers and magazines online) confirm the rule of classifying countries into the third, the most advanced class. However, there is an interweaving between the two less developed classes, especially for the web TV and radio variable, and in a lower extent for reading newspapers and magazines online.

The next method used was CN2 rules. Using CN2 algorithms with Laplace rule quality estimation classification rules can be devised, for this particular case the classification of individuals from European countries into certain classes of the target variables can be done, in accordance to the calculated values of the input variables. Figure 3 clearly shows the overview of generated CN2 rules, from which the most representative ones will be commented.

For example, if the individuals from a certain country were looking for information in 2008 about the products in less than or equal to 37% of the cases, then 8 countries (with the rule quality of 0.9) are classified as group "1", i.e. low online shopping level class.

FIGURE 4. SCATTER PLOT FOR THE VARIABLES USE OF E-BANKING AND SEARCHING FOR PRODUCT INFORMATION IN 2009



On the other hand, if the individuals from European countries were looking for information about products in 2009 in over 59% of the cases, then 10 countries (with the rule quality of 0.917) are classified as group "3", i.e. high online shopping level class.

Furthermore, the rules for the "TV_radio" attribute vary considerably, which indicates once again the weaker correlation between using web TV and radio services and the target class attribute - online shopping (at least in comparison with the dominant attributes "Product_info" and "E-banking").

Finally, the scatter plots clearly show that there is a very strong positive correlation between searching for information about products online and using e-banking services in relation to online shopping.

Also, a relatively high, but not as consistent positive correlation between using web TV and radio services, and online reading (which is the third strongest input variable if we sort by influence) from the aspect of individuals in observed European countries was proven.

Figure 4 illustrates stated very strong positive correlation between input variables of using e-banking and searching for product information online in 2009.

Conclusion

This research confirmed the assumed positive correlation between online shopping and the input variables (using e-banking services, searching for information about products and services online, reading newspapers and magazines online, and using web TV and radio services) signifying certain characteristics of individuals in observed European countries.

It is important to emphasise that the specified input variables vary by intensity of the mentioned correlation: searching for products and services online influences online shopping the most. The use of e-banking services is in a very strong positive correlation with the observed target variable as well. However, reading newspapers and magazines online shows a slightly lower positive correlation, while the use of web TV and radio services can be described only as a moderate positive correlation with regard to online shopping.

The final results are rather logical and understandable because if one wants to buy something online it is necessary (or at least recommendable) to gain some knowledge about the product first - especially via the Internet - and complete the transaction afterwards using an e-banking service.

References

Atak, M., Erturgut, R., 2010. "Importance of educated human resources in the information age and view of information society organizations on human," Procedia - Social and Behavioral Sciences, Vol.2, (2) Innovation and Creativity in Education, pp.1452- 456, DOI: 10.1016/j.sbspro.2010.03.217.

Carayannis, E., Sagi, J., 2002. "Exploiting opportunities of the new economy: developing nations in support of the ICT industry," Technovation, Vol.22 (8), pp.517-24, DOI: 10.1016/S0166-4972(01)00046-3.

Chiu S., Tavella, D., 2008. Data mining and market intelligence for optimal marketing returns, Butterworth-Heinemann.

Chu, B., Tsai, M., Ho, C., 2007. "Toward a hybrid data mining model for customer retention," Knowledge-Based Systems, Vol.20(8), pp.703-18.DOI:http://dx.doi.org/10.1016/j.knosys. 2006.10.003.

Clark, P., Boswell, R., 1991. "Rule induction with CN2: Some recent improvements," Machine Learning - European Working Session on Learning (EWSL-91), Lecture Notes in Computer Science, Vol.482, pp.151-63, DOI: 10.1007/BF b0017011.

European Commission, 2009. i2010 benchmarking framework, available from http://ec.europa.eu/information_society/ eeurope/i2010/benchmarking [accessed 06/03/2010]

EUROSTAT, 2005, Information Society portal, available from http://epp.eurostat.ec.europa.eu/portal/page/portal/informatio n_society/introduction [accessed 06/03/2010]

Hosgorur, V., Bilasa, P., 2009. "The problem of creative education in information society," Procedia - Social and Behavioral Sciences, Vol.1, World Conference on Educational Sciences, Nicosia, North Cyprus, 4-7 February 2009 - New Trends and Issues in Educational Sciences, pp.713-17, DOI: 10.1016/j.sbspro.2009.01.125

Huh, M., Park, D., 2008. "Enhancing parallel coordinate plots," Journal of the Korean Statistical Society, Vol.37 (2), pp.129-33, DOI: 10.1016/j.jkss.2007.10.003

Javorović, B., Bilandžić, M., 2007. Poslovne informacije i business intelligence, Golden Marketing, Zagreb (in Croatian).

Klepac, G., Mršić, L., 2006. Poslovna inteligencija kroz poslovne slučajeve, Lider Press, Zagreb (in Croatian).

Lessmann, S., Voss, S., 2009. "A reference model for customer-centric data mining with support vector machines," European Journal of Operational Research, Vol.199 (2), pp.520-30, DOI: 10.1016/j.ejor.2008.12.017

Orange Canvas website, available from http://www.ailab.si/orange [accessed 06/03/2010]

Panian, Ž., Klepac, G., 2003. Poslovna inteligencija, Masmedia, Zagreb (in Croatian).

Rygielski, C., Wang, J.C., Yen, D.C., 2002. "Data mining techniques for customer relationship management," Technology in Society, Vol.24 (4), pp.483-02, DOI: 10.1016/S0160-791X(02)00038-6

Wang, H., Hong, W., 2006. "Managing customer profitability in a competitive market by continuous data mining," Industrial Marketing Management, Vol.35 (6), pp.715-23, DOI: 10.1016/j.indmarman.2005.06.005

Zupan,B., Leban,G., Demsar,J., Curk,T., 2010. Widgets and Visual Programming, available from http://www.ailab.si/orange/wp/orange%20widgets.pdf [accessed 06/03/2010]