9-2017

Identifying Customer Satisfaction Patterns Via Data Mining: The Case Of Greek E-Shops

Katerina Kalaidopoulou
Athens University of Economics and Business, E-Business Research Center (ELTRUN), akalaidopoulou@aueb.gr

Stratos Triantafyllou
Athens University of Economics and Business, E-Business Research Center (ELTRUN), striantafyllou@aueb.gr

Anastasia Griva
Athens University of Economics and Business, E-Business Research Center (ELTRUN), an.griva@aueb.gr

Katerina Pramatari
Athens University of Economics and Business, E-Business Research Center (ELTRUN), k.pramatari@aueb.gr

IDENTIFYING CUSTOMER SATISFACTION PATTERNS VIA DATA MINING: THE CASE OF GREEK E-SHOPS

Research full-length paper

Track N° 7

Kalaidopoulou, Katerina, Athens University of Economics and Business, E-Business Research Center (ELTRUN), akalaidopoulou@aueb.gr

Triantafyllou, Stratos, Athens University of Economics and Business, E-Business Research Center (ELTRUN), striantafyllou@aueb.gr

Griva, Anastasia, Athens University of Economics and Business, E-Business Research Center (ELTRUN), an.griva@aueb.gr

Pramatari, Katerina, Athens University of Economics and Business, E-Business Research Center (ELTRUN), k.pramatari@aueb.gr

Abstract

In an online marketplace reality in which customer satisfaction emerges as a key success factor for e-retailers, it becomes crucial to better understand whether the shoppers are satisfied and what factors affect their satisfaction experience. As we are in the Big Data era, Business Analytic techniques could assist us to better understand our customers and their respective satisfaction. To this end, this paper presents a data mining based approach to identify different satisfaction patterns/profiles from satisfaction survey responses. This approach was applied on data from over 120 Greek e-shops across 18 industries. Apart from its theoretical contribution, the proposed approach extracts hidden satisfaction patterns with a view to better understand the specific needs and preferences of customers. These insights may be used to support several decisions, ranging from marketing actions per customer satisfaction profile, to actionable decision making and customer-oriented strategies.

Keywords: Customer Satisfaction, Data Mining, Business Analytics, e-business

1 Introduction

In the era of e-business and digital transformation, it becomes crucial for retailers to understand what creates a satisfied customer to fulfill the promises of satisfying online shoppers. Therefore, many companies have identified the need to not only understand customer purchase behavior, but also the customer satisfaction through their online purchase journey. Additionally, technological advances enable direct communication with the customer and mass data collection about their behavior and/or their responses on surveys. To this end, many researches in both academia and industry focusing on analyzing these datasets, to calculate the satisfaction levels and/or identify the factors affecting their satisfaction.

More precisely, many researches aimed to calculate satisfaction in different contexts, such patient (Ortiz and Schacht, 2012) or student (Moro-Edigo and Panades, 2009) satisfaction. The traditional methods of satisfaction measurements include statistics that describe the satisfaction level, but do not identify what affects satisfaction. Therefore, other research used advanced statistical models and data mining techniques, such as regression analysis (Yap et al, 2012, Papaioannou and Martinez, 2016) or classification (Dejaeger et al, 2012) in order to extract the factors that affect satisfaction. However, to the best of our knowledge, not a lot of research has been conducted to extract hidden satisfaction patterns from survey data, with a view to better understand the specific needs and preferences of customers.

To address this gap, this paper presents a Data Mining based approach, which uses cluster analysis on online satisfaction survey data in order to identify customer satisfaction segments and highlight customers' behavior and preferences within each segment. In more detail, the proposed approach is applied to real satisfaction survey data from 120 Greek e-shops across 18 industries supplied by an online survey company. Based on the results of the analysis, it is concluded that customer experiences vary across different industries and there are various levels of satisfaction, with various attributes triggering each satisfaction segment. Moreover, this analysis uses data from many e-shops of the Greek market, hence allowing the generalization of results between different industries. Considering the findings, it provides useful insights about customer satisfaction patterns with a view to understand customers behavior and preferences, and sets forth significant implications on customer loyalty, actionable decision making and customer-oriented strategies.

The remainder of the paper is organized as follows. The 'Background' section summarizes the relevant literature and pinpoints how this study differs from the extant ones. The proposed approach its evaluation using online satisfaction data provided by a major Greek survey company, are described in the next section. Section 4, presents the identified Customer Satisfaction Patterns. Finally, we conclude with the main outcomes of the paper, the theoretical contribution and the practical implications of our approach; and some highlights of further research.

2 Background

Following industrial revolution, more standardized products of varied assortment at affordable prices came to the market satisfying even the most demanding customer needs. However, as far as competition and product variety intensified, the shift to market and customer orientation was inevitable. Keeping customers happy and satisfied is a goal that was included in Customer Relationship Management (CRM) strategies ensuring organization's long-term success. This goal is called customer satisfaction, i.e. "the consumer's fulfillment response, the degree to which the level of fulfillment is pleasant or unpleasant" (Oliver, 1997).

Customer satisfaction has been proven important for companies as it can be served as a competitive advantage (Craig, 1989). If a customer is satisfied then he/she is more likely to buy again and the company holds a competitive advantage against the rivals (Woodruff, 1997). To measure customer loyalty towards the firm, the Net Promoter Score (NPS) Key Performance Indicator (KPI), has been created. NPS is a KPI that represents with a number the company's customer loyalty (Reichheld, 2003). As stated before, customer satisfaction increases customer retention, but it also reduces customer churn, because customers are less likely to abandon the company and turn to a competitor. To this end, a happy customer is more likely to visit the store and make more purchases, thus its lifetime value increases. A study conducted by InfoQuest (2006) concluded that a 'totally satisfied customer' contributes 2.6 times more revenue than a 'somewhat satisfied customer' and 14 times more revenue than a 'somewhat dissatisfied customer'. Also, customer satisfaction reduces the negative word of mouth (WOM), which impacts customer acquisition (Von Wangenheim and Bayon, 2007). According to McKinsey, an unhappy customer tells 9-15 about their people, which indicates that it is important to keep the customer happy.

The above factors show the importance of calculating customer (student, patient, user etc.) satisfaction in both academia and business. To measure and evaluate this metric, most of the times, customers' input is asked regarding their satisfaction during their shopping experience. More precisely, consumers are asked to rate the product or service and a set of factors related to them; these ratings are used to indicate satisfaction. After collecting these data, most researches proceed with calculating descriptive statistics (e.g. average, standard deviation per question) and extract trends and insights. Other researches proceed with analyzing further the satisfaction data, using multivariate linear regression model (Moro-Edigo and Panades, 2009), partial least squares regression (Yap et al, 2012) and logistic regression model (Papaioannou and Martinez, 2016). The aforementioned approaches aim to model customer satisfaction as the dependent variable Y with on one or more explanatory (or independent) variables X. Alternatively, other researches, such as Ortiz and Schlacht (2012) and Subramanian et al. (2014) used exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) which describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called factors.

Apart from the traditional statistical analysis, other researches built models using different data mining techniques to understand and explain customer satisfaction. To this end, the work of McAuley and Leskovec (RecSys 2013) is utilizing latent models in the context of recommender systems in order to predict customer satisfaction/ratings over new products. Barnes et al. (2017) utilize latent dirichlet allocation which allows sets of observations to be explained by unobserved, but controllable groups. Other researches, exploit sentimental analysis or opinion mining (Kang and Yongtae, 2014) in order to study customer satisfaction. Opinion mining refers to the use of natural language processing, text analysis and computational linguistics, to systematically extract, quantify, and study affective states and subjective information. Additionally, classification has been used to identify to which satisfaction category a new observation belongs, based on a training set (Dejaeger et al, 2012). Lastly, clustering or cluster analysis, have been used to segment people based on their demographic data and associate them with their satisfaction (Beynon et al, 2012), but not on the satisfaction ratings themselves. Clustering is a task of grouping a set of items in such a way that items in the same group/cluster are more similar to one another than with the items of a different group. Beyond the customer satisfaction scope, Adamopoulos (2013) combines econometric, text mining, opinion mining, and predictive modeling techniques towards a more complete analysis of the information captured by user-generated content taking the advantage of unstructured data.

Overviewing the existing literature in the e-commerce retail industry, it has been observed that different data mining techniques have been applied to analyze customer and sales data, improve the processes of the industry and support more customer-centric approach. These analyses are useful to gain customer and company insights and support CRM (Anderson et al, 2007). However, they do not take into account customer satisfaction. Some researches attempted to evaluate customer satisfaction on a national basis, and create a customer satisfaction index (Fornell et al, 1996, Fornell, 1992). However, an alternative approach that has been proved successful on identifying patterns in other contexts (e.g. sales data) is to utilize cluster analysis on satisfaction survey responses in order to discover satisfaction patterns.

Overall, it can be observed that there are different data mining techniques that can analyze satisfaction data extracting different knowledge. However, to the best of our knowledge there is not an approach in the literature that uses data mining and focuses on identifying patterns in the satisfaction survey data, with the view to identify different customer behaviors. Focusing on this literature gap, this research aims to identify customer satisfaction types across different industries of the Greek e-commerce market using data from 120 Greek e-shops across 18 industries. In addition, a comparison between the identified customer satisfaction patterns across these industries is also provided.

3 Data Mining Approach to Identify Customer Satisfaction Patterns

In this paper, we propose an approach that employs data mining techniques (clustering) to identify latent customer satisfaction segments from structured survey data and examine the behavior and discriminating characteristics of each segment, as reflected in their responses to a customer satisfaction survey. We have followed the "Design Science" approach (Hevner et al. 2004) that focuses on the development and assessment of artifacts. We evaluate our proposed artifact/approach by applying it in practice in order to prove its sufficiency to manage the original problem i.e. to discover satisfaction patterns. Specifically, we examine patterns and identify segments on actual satisfaction survey responses from over 120 Greek e-shops. Our research contribution is the clustering-based artifact described in this section, but we also provide insights on the identified satisfaction patterns and emergent response segments.

The data for the design and evaluation of our proposed approach were supplied by an online survey company that provides customer satisfaction insights for several e-shops across different industries in Greece. The following subsections summarize our analytical approach for discovering satisfaction patterns from such data, while in Section 4 we discuss the resulting satisfaction segments.

3.1 Dataset Description

The provided datasets consisted of customer responses from a two-part online satisfaction survey. The first part was issued when the customer placed an order (checkout part), and the second one after the order was fulfilled (aftersales part). More precisely, the collaborating company provided data for about 1 million orders, placed in over 120 Greek e-shops across 18 industries, between April 2016 and February 2017. Around 20 percent of these orders were associated with valid recorded responses to either of the survey parts. That happens due to the fact that customers could choose not to complete the survey after their purchase. Hence, the dataset included 202 thousand responses for the checkout part, and 62 thousand responses for the aftersales part (a response rate of 6 percent). For about 26 thousand orders, responses were recorded for both the checkout and the aftersales part. An approximate 80 percent of all responses came from e-businesses operating in the Sports, Shoes, Apparel, Online Pharmacy and Grocery Retail industries.

The data provided by the collaborating survey company included three different datasets:

- A set of any orders placed in the tracked e-shops during the examined period. For each order, provided information included unique identifiers for the order, the e-shop and the customer who placed the order, as well as the date the order was placed and the category (industry) of the e-shop according to the survey company's taxonomy. Further information about the customer was not provided in order to comply with company's privacy regulations.
- A set of responses to the checkout part of the online survey. For each response, provided information included the unique order identifier and the customer's answers to each of the survey items of the checkout part as described in *Table 1*.

Survey Item	Rating Range
Overall Satisfaction	1-10
Recommendation Likelihood	0-10
Product Variety	1-10
Product Availability	1-10
Prices	1-10
Security	1-10
Usability	1-10
Product Presentation	1-10

Table 1. Items of the checkout part of the survey.

- A set of responses to the aftersales part of the online survey. For each response, provided information included the unique order identifier, the date the response was recorded, and the customer's answers to each of the survey items of the aftersales part, as described in *Table 2*.

Survey Item	Rating Range
Overall Satisfaction	1-10
Recommendation Likelihood	0-10
Delivery Time	1-10
Shipping Cost	1-10
Product Quality	1-10
Delivery Options	1-10
Packaging	1-10
Customer Service	1-10

Table 2. Items of the aftersales part of the survey.

The basis for integrating the aforementioned data sources is to be able to create a set of unique identifiers which could indicate that a specific response has been submitted by Customer X after placing an online Order Y. Additionally, if Customer X has responded to both the checkout and the aftersales part of the online survey, then both responses should be associated with Customer X and Order Y. Furthermore, if Customer X has placed additional orders and has responded to any of the parts in each one of them, then all of those responses should be linked to Customer X.

Cleansing tasks applied on the datasets included eliminating records for corrupt, empty or duplicate responses, and handling unique identifier inconsistencies. Thus, we continued the analysis with the 94% of the raw dataset.

3.2 Data Mining Model

In order to be able to discover previously unidentified satisfaction patterns, unsupervised machine learning was employed in the form of cluster analysis. Several pilot data mining models, based on implementations of different clustering algorithms, were developed and tested for interpretability and compatibility of results with prior knowledge of the field, provided by the collaborating survey company. These preliminary results revealed an implementation of the Expectation-Maximization (EM) algorithm to perform better in both business and technical terms, while also providing suggestions concerning the structure of the training data.

In the original dataset, each response included a set of numerical values that represented a customer's answers to each of the survey items. While numerical attributes are widely used with data mining methods, an approach with discrete attributes was found to be more effective during the preliminary tests. Using clustering-based attribute discretization, numerical values were grouped in four satisfaction levels, resembling a four-point Likert scale. The process was performed iteratively for each survey item, so as to control numerical rating level inconsistencies and bring different variables to the same conceptual level. The eventual satisfaction levels for the checkout part are presented in *Table 3*.

Survey Item	"Excellent" Range	"Good" Range	"Average" Range	"Poor" Range
Overall Satisfaction	10	8-9	5-7	1-4
Recommendation Likelihood	10	8-9	4-7	0-3
Product Variety	10	8-9	3-7	1-2
Product Availability	10	8-9	3-7	1-2
Prices	10	7-9	3-6	1-2
Security	10	8-9	5-7	1-4
Usability	10	8-9	4-7	1-3
Product Presentation	10	8-9	4-7	1-3

Table 3. Eventual satisfaction levels for the checkout part.

The data eventually used to train the mining model were structured as a table, with each record representing a customer response. Each column stored a binary value with 1 indicating that the customer responded with that rating, and 0 that not. The columns represented the four different answers on the questions, e.g. "Question 1 – Excellent", "Question 1 – Good", "Question 1 – Average", and "Question 1 – Poor". If the customer has answered this question, then one of the columns has the value 1, else has the value 0. If the customer has skipped this question, then all columns have the value 0. The structure of the training dataset is presented in *Figure 1*. Two mining models were trained, one for the checkout part and one for the aftersales part.

responseID	Overall Satisfaction Excellent	Overall Satisfaction Good	Overall Satisfaction Average	Overall Satisfaction Poor	…
0000001	0	1	0	0	…
0000002	1	0	0	0	…
0000003	0	0	0	0	…
0000004	1	0	0	0	…
…	…	…	…	…	…

Figure 1. Structure of the training dataset for the checkout mining model.

The mining model attempts to group similar survey responses (clusters), e.g. responses in which customers assigned a similar rating to each aspect of the e-shop. Each one of these response groups (clusters) reveals a disparate satisfaction segment, which indicates a distinct customer satisfaction pattern based on similar ratings for each aspect of the e-shop.

4 Identified Customer Satisfaction Patterns

The mining model on checkout data revealed nine response segments (clusters), each survey response belongs uniquely into one segment. Via examining the responses in each segment, we identify and characterize the different customer satisfaction patterns. An example of the detailed clustering results for one satisfaction segment (checkout part) is presented in detail in *Table 4*. The percentages in each cell represents the probability a response that belongs in this segment to be classified as "Excellent", "Good", "Average" or "Poor". This segment accounts for 8.5 percent of all checkout data and its responses tend to include "Good" ratings for most of the items. However, Product Availability is rated

"Average" in nearly all responses, while Product Variety ratings are more mixed, with one half of the responses assigning a "Good" rating and the other half an "Average" rating. Aside from identifying the rating patterns represented in each segment, the calculation of the following descriptive statistics for each segment can help better defining its behavior: the percentage of total responses associated with this segment in each industry, the percentage of responses recorded on mobile devices - smartphones or tablets - as opposed to personal computers, as well as the percentage of total responses associated with impulse purchases as opposed to planned purchases. This segment was specifically associated with the Fashion and Shoes industries, while being largely absent from Books and Electrical Goods e-shops. It included significantly more responses from computers (only 24 percent from mobile devices). A typical 47 percent of responses of this segment were associated with impulse purchases, as shoppers stated.

Survey Item	Excellent	Good	Average	Poor
Overall Satisfaction	18%	67%	14%	1%
Recommendation Likelihood	14%	68%	17%	1%
Product Variety	3%	48%	49%	0%
Product Availability	5%	0%	94%	1%
Prices	0%	100%	0%	0%
Security	9%	76%	15%	0%
Usability	12%	72%	16%	0%
Product Presentation	7%	72%	21%	0%

Table 4. Detailed results of Segment 8.

A challenging task when examining the characteristics of each segment is to actually identify the satisfaction pattern beyond the quantitative results. For this specific segment, lower scores in Product Availability and Variety suggested that customers either were unable to purchase an item of an existing product which was unavailable or out of stock (eventually assigning a lower rating to Product Availability), or were looking for a product that was not available for purchase in the particular e-shop (assigning a lower rating to Product Variety). An evaluation of this segment by the domain experts of the collaborating survey company revealed that this was actually a pattern already reflected in their qualitative measurements, but one they had been unable to associate with specific responses before. This was further reinforced by the fact that many of these responses came from Apparel and Shoes e-shops, where there is a higher level of product customization and shoppers are still likely to substitute one product for another if their preferred color or size is not available. It is also suitable that this segment was less frequent in online bookstores or e-shops with electrical goods, not due to an absence of availability issues in these industries, but rather because shoppers are more likely to switch to another e-shop or not place the order at all if their book or electrical product of choice is not available at the moment of intended purchase.

The nine response segments identified and their corresponding share of all responses are summarized in *Table 5*. The segment described in detail above is Segment 8 ("Substitution-related Issues").

Segment	Segment Name	Percent (%) of all responses
Segment 1	Overall Excited	18.7%
Segment 2	Overall Satisfied	11.7%
Segment 3	Overall Disappointed	7.4%

Segment 4	Pricing Concerns	11.8%
Segment 5	Product Offering Concerns	17.8%
Segment 6	User Experience Concerns	9.2%
Segment 7	Security-driven Satisfaction	7.0%
Segment 8	Substitution-related Issues	8.5%
Segment 9	Selective Answering	7.9%

Table 5. Customer Satisfaction Patterns for the checkout part.

Three of the other, more well-defined segments (Segments 1, 2 and 3) contained responses assigning a similar rating to every survey item and were associated with "Excellent", "Good" and "Average" evaluations of an e-shop respectively. Segment 1 ("Overall Excited") was highly associated with Books and Electrical Goods e-shops, as 24 percent of responses from both industries were included in this segment. Segment 1 responses were less frequent in online supermarkets, accounting for only 13 percent of that industry's responses. Segment 1 included more mobile responses from most of the other segments with 31 percent of its responses recorded from smartphones or tablets. Impulse purchases accounted for 52 percent of all Segment 1 responses. Segment 2 ("Overall Satisfied") was also present in online bookstores, as well as online pharmacies, while being largely absent from the Fashion and Shoes industries. Mobile responses accounted for 25 percent of Segment 2 responses, while 47 percent of Segment 2 responses were associated with impulse purchases. Segment 3 ("Overall Disappointed") was strongly present in online supermarkets and pet stores, while being less frequent in e-shops with office supplies. Responses from mobile devices accounted for 30 percent and impulse purchases for 46 percent of Segment 3 responses.

It is worth mentioning that Segments 1, 2 and 3, accounting for a cumulative 38 percent of all responses, might indicate that it is possible that many customers could be equally satisfied with every aspect of an e-shop and there are no differences between different aspects. While that might be true for some customers, an evaluation of these segments on business terms should raise questions relating to the validity of these responses and the possibility of uniform satisfaction across all aspects. Even more important should be the fact that the remaining 62 percent of all responses, similar to Segment 8 responses as previously described, reflected more complex satisfaction patterns.

Segment 4 ("Pricing Concerns") accounted for 12 percent of all responses and included responses scoring an "Excellent" rating in every survey item except for Prices, revealing pricing concerns. Such responses often came from beauty e-shops, while the segment was expectedly less frequent among e-shops with deals and offers. Mobile responses accounted for 28 percent and impulse orders for 47 percent of segment responses.

A significant 18 percent of all responses were included in Segment 5 ("Product Offering Concerns"), and were associated with "Excellent" or "Good" ratings for items relating to the user interface and content management of the e-shop, but revealed problems with Prices, Product Variety and Availability. This segment had a stronger presence in e-shops with apparel or baby products, while it was mostly absent from the Office Supplies and Grocery industries. Responses of this segment were more frequently submitted from personal computers, with only 23 percent coming from mobile devices.

The opposite situation was observed in Segment 6 ("User Experience Concerns"), where "Good" scores for product-related items and overall were weighed down by "Average" ratings for Usability and Product Presentation. A significant 37 percent of segment responses also rated Product Variety as "Average", which might indicate that problems with product assortment and content organization eventually prevented customers from locating preferred products. Segment 6 was particularly strong in online grocery stores. One in four responses came from mobile devices, while impulse purchases accounted for a comparatively low 44 percent.

Segment 7 ("Security-driven Satisfaction") was associated with very strong scores for Security, Overall Satisfaction and Recommendation Likelihood, indicating customers who were fully satisfied with their online experience mostly because they are used to purchase from or felt safe with a specific e-shop. These responses came from online pharmacies, where purchasing a product might indicate a certain amount of trust, as well as from e-shops with office supplies, which point to loyal customers or habitual buying behavior. In contrast, Segment 7 accounted for only 5 percent of responses in e-shops with deals and offers. About 30 percent of responses came from mobile devices, while impulse purchases accounted for 46 percent of segment responses.

Segment 9 ("Selective Answering") contained responses which included answers to one or two items, and were associated with both satisfied and unhappy customers. This segment was highly associated with impulse purchases, accounting for 64 percent of segment responses, while being distinctly present in accessories e-shops and deals websites. A comparatively stronger 47 percent of segment responses came from mobile devices.

A similar approach was applied to the aftersales dataset, resulting in the ten satisfaction segments presented in *Table 6*.

Segment	**Segment Name**	**Percent (%) of all responses**
Segment 1	Overall Excited	24.1%
Segment 2	Overall Satisfied	6.4%
Segment 3	Overall Disappointed	7.4%
Segment 4	Overall Unhappy	4.0%
Segment 5	Shipping Cost Concerns	12.0%
Segment 6	Packaging & Delivery Concerns	11.7%
Segment 7	Excited Passives	10.8%
Segment 8	Express Delivery	9.3%
Segment 9	Late Delivery	7.6%
Segment 10	Selective Answering	6.7%

Table 6. The ten response segments for the aftersales part.

It is worth noting that these segments also reveal satisfaction patterns across industries. For example, responses from online bookstores tend to be clustered with Segments 1 and 2, indicating both a generally high level of satisfaction in the industry and, more importantly, consistent satisfaction levels across all survey items, revealing book online shoppers to be less demanding or associated with a kind of halo effect. The opposite is observed in the Beauty, Apparel and Shoes industries, where responses were more diverse, with customers satisfied with some aspects of an e-shop and dissatisfied with other aspects at the same time. Online grocery stores were associated with middling customer experiences, while deals responses revealed greater satisfaction with product prices, but also a limited feeling of security.

5 Conclusion

While customers' behavior and expectations are changing over the years, companies have identified the importance of understanding their customer satisfaction in order to better meet their needs. As a result, customer satisfaction surveys are recognized as insightful information sources for monitoring and enhancing satisfaction levels. Motivated by the above-described business need, this research paper presents an approach to identify customer satisfaction patterns by applying cluster analysis on structured satisfaction survey data.

To the best of our knowledge, there is no other approach that utilizes clustering techniques on survey data and describes how to extract hidden satisfaction patterns with a view to better understand the specific needs and preferences of customers. While other researchers aim to find the factors that affect satisfaction, this research evaluates all satisfaction attributes equally and tries to identify if common ratings are given to any of them, hence indicating common beliefs across different customers. In more detail, this research provides a different perspective in satisfaction tracking for e-shops. While other analyses conclude to complex statistics, i.e. numeric means and standard deviation that might be difficult to comprehend by the end business user, this research suggests nine different satisfaction personas/patterns, each one of which representing a satisfaction profile. The goal of this research is to provide a simple-to-comprehend segmentation of the mass survey data available in the business context that can provide actionable insights. Combining them with customer profiles extracted from their demographical and behavioral data, the analysis would lead to a complete customer profiling.

In order to evaluate our proposed approach, we identified customer satisfaction patterns across different industries of the Greek e-commerce market using data from 120 Greek e-shops across 18 industries. By applying our proposed approach to real data, we extracted useful insight about shoppers' behavior, and we also identified that the satisfaction patterns differ across different industries. Considering the findings, we provide useful insights about customer satisfaction patterns and set significant practical implications on customer loyalty, actionable decision making and customer-oriented strategies.

Via comparing the resulting customer satisfaction segments derived from different e-shops of the same market, enables benchmarking across different e-shops between the same industry, as well as benchmarking within different industries. Comparing with the results of the best performing companies, it is possible to set goals and targets that will drive future strategic decision making. Moreover, understanding the satisfaction levels of the customer, can support personalized marketing actions and direct communication, e.g. targeted email campaigns based of the customer satisfaction profile. In more detail, if an e-shop could track directly in which satisfaction patterns/profile a shopper belongs to immediately after he/she completes a survey, then automated marketing messages and content could be triggered. For instance, imagine that we have detected an "Overall Disappointed" shopper, or a shopper with "Pricing Concerns", after his/hers purchase. We could show directly a message to this shopper, e.g. about a price discount for the next purchase, or we even call him/her back. In the same spirit, we could trigger actions when we detect an "Overall Excited" shopper. It could be the right moment to ask them share something about our e-shop at social media, or write a review.

Future research could focus on comparing the customer satisfaction profiles with their purchase profiles. This way we can enable understanding if different product categories drive more satisfaction and how differences occur. Moreover, further research could combine satisfaction analysis with the customer segmentation analysis in order to enrich the insights and the companies obtain a full understanding of their customers' behavior. With such results, an even more targeted communication with the customer and thus better business results can be achieved.

From a technical perspective, it could be interesting to compare different segmentations algorithms, e.g. k-means versus the expectation-maximization algorithm, or different data mining algorithms, such as factor analysis and latent class models versus cluster analysis. Also, studying whether there are alternative ways to identify customer satisfaction patterns it could be of great interest. For instance, we can exploit customer reviews via applying text mining, or we can utilize browsing data and analyze whether different navigation patterns affect customer satisfaction.

REFERENCES

Adamopoulos, P. (2013). What makes a great MOOC? An interdisciplinary analysis of student retention in online courses.

Ahn, K.-I. (2012). Effective product assignment based on association rule mining in retail. Expert Systems with Applications, 39(16): 12551-12556.

Anderson, J. L., Jolly, L. D., & Fairhurst, A. E. (2007). Customer relationship management in retailing: A content analysis of retail trade journals. Journal of Retailing and Consumer Services, 14(6): 394-399.

Armstrong, G. (2014). "Principles of marketing*". Pearson Australia*.

Beynon, M. J., M. Heffernan, and A. M. McDermott (2012). "Psychological Contracts And Job Satisfaction: Clustering Analysis Using Evidential C-Means And Comparison With Other Techniques." *Intelligent Systems in Accounting, Finance and Management* 19(4): 247-273.

Craig C. (1989). "Creating an effective customer satisfaction program*", Journal of Business &* Industrial *Marketing*, 4 (2), 33–42.

Dejaeger, K. (2012). "Gaining insight into student satisfaction using comprehensible data mining techniques." *European Journal of Operational Research* 218(2), 548-562.

Fornell, Claes, et al. (1996). "The American customer satisfaction index: nature, purpose, and findings." *Journal of Marketing*, 60(4), 7-18.

Fornell, Claes. (1992). "A national customer satisfaction barometer: The Swedish experience" the Journal *of Marketing*, 56(1), 6-21.

Guo, Yue, Stuart J. Barnes, and Qiong Jia. (2017). "Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation." *Tourism Management* 59, 467-483.

Hevner, A. R., March, S. T., Park, J., and Ram, S.s. (2004). "Design science in information systems research." *MIS Quarterly* 28(1), 75-105.

Kang, D., and Yongtae P. (2014). "Review-based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach." *Expert Systems with Applications* 41(4), 1041-1050.

McAuley, J., & Leskovec, J. (2013). Hidden factors and hidden topics: understanding rating dimensions with review text. In *Proceedings of the 7th ACM conference on Recommender systems* ACM, 165-172.

Moro-Egido, Ana I., and J. Panades. (2010). "An analysis of student satisfaction: Full-time vs. part-time students." *Social Indicators Research* 96(2), 363-378.

Oliver, Richard L. (1997), "Satisfaction: A Behavioural Perspective on the Consumer", New York: McGraw Hill.

Ortiz, Glorimar, and Lucille Schacht (2012). "Psychometric evaluation of an inpatient consumer survey measuring satisfaction with psychiatric care." *The Patient-Centered Outcomes Research* 5(3), 163-173.

Papaioannou, D., and L. Miguel Martínez (2016). "Measuring Satisfaction with Transit and Car Trips with Use of One Logistic Regression Model" Transportation Research Record: Journal of the Transportation Research Board 2543, 101-107.

Reichheld, Frederick F. (2003). "The One Number You Need to Grow", *Harvard Business Review,* 12, 46-55.

Subramanian, Nachiappan, et al. (2014). "Customer satisfaction and competitiveness in the Chinese E-retailing: Structural equation modeling (SEM) approach to identify the role of quality factors." *Expert Systems with Applications* 41(1), 69-80.

Von Wangenheim, F. and T. Bay (2007). "The Chain from Customer Satisfaction via Word-of-Mouth Referrals to New Customer Acquisition," *Journal of the Academy of Marketing Science*, 35(2), 233–49.

Wah Yap, Bee, T. Ramayah, and Wan Nushazelin Wan Shahidan. (2012). “Satisfaction and trust on customer loyalty: a PLS approach.” *Business Strategy Series* 13(4), 154-167.

Woodruff, Robert B. (1997). “Customer value: the n*ext source for competitive advantage.” Journal of* the *academy of marketing science* 25(2), 139-153.