NEW BEHAVIORAL SEGMENTATION METHODS TO UNDERSTAND CONSUMERS IN RETAIL INDUSTRY

Fahed Yoseph[1], Nurul Hashimah Ahamed Hassain Malim[2] and Mohammad AlMalaily[3]

[1,2]School of Computer Science, Universiti Sains Malaysia
[3]Faculty of Engineering, Mansoura University

ABSTRACT

Behavioral segmentation is considered as one of the most important concepts of modern marketing. Traditional customer segmentation models require months of analytical work, resulting in discrete consumers' insights that are outdated to match the dynamic body of the consumers they are meant to represent. Personalization and consumer experience are make or break factors for the retail industry. This study looks towards maximizing Consumer Lifetime Value (LTV) to accommodates the dynamics in consumer shopping behavior for a medium size retailer. using (LTV) matricto investigate behavioral changes in the consumer shopping history gaining knowledge from behavioral and demographic variables stored in POS database converted into RFM dataset format. In addition, this study applies soft clustering Fuzzy C-Means (FCM) and hard clustering Expectation Maximization (EM) algorithms to classify individual consumers exhibit similar purchase history into specific groups. For measuring the algorithms accuracy, weuse cluster quality assessment (CQA). The CQA shows EM algorithm scales much better than Fuzzyy C-Means algorithm with its ability to assign good initial points in the smaller dataset.

KEYWORDS

Customer Segmentation, Clustering, LTV Matric, Retailing

1. INTRODUCTION

Traditional marketing strategies which are mainly based on marketing experts and sales manager's opinions of the market [17]. For example, Gholamian [22] states that the retail industry is highly competitive, with the number of products often overwhelming. Consumers are faced with a variety of products, causing the demand to be higher and more complex. Also, Traditional market segmentation often fails in granularity and lack of consumers shopping behavior precision. The dynamic transformation of the consumer's shopping behavior pushes the traditional marketing (Mass) to their limits. Traditional marketing methods bases it's reasoning on isolated consumers, the method sees any economic phenomenon only in terms of the shopping behavior of the individuals (consumers) who form the target, without taking in consideration the group effect or the interactions between the (consumers) individuals. This leads the retailer failing to design tailor messaging that is compelling enough and relevant to specific groups of consumers. In light of this trend, modern marketing moves from mass-marketing (products-focus) to target-marketing (consumer-focus). In response, retailers segment their markets to be more strategic in their planning and design and implement successful marketing strategies and retention policies. In this fast-changing and dynamic industry, there is a clear need for advanced methods

DOI: 10.5121/ijcsit.2019.11104

to discover new cosumers segments using essential data trapped in the Point-of-Sales dataset.With consumer segmentation empowering retailers to precisely reach consumers with specific needs and wants, by dividing the market into similar and identifiable segments, to focus on individuals with similar preferences, choices, needs and interests on a common platform [24], [27]. Segmentation evaluates consumers as a segment indirectly, rather than individually or directly. It enables retailers to make full use of their limited resources to serve consumers effectively as consumer sub-groups [10]. Proper mechanisms for treating point-of-sales (POS) events to convert ever-increasing transaction data into knowledge [34]. To take on this challenge one step further, this study (LTV) matric will be applied on POS database converted into RFM (Recency, Frequency, Monetory) dataset.(LTV) and (RFM) are two of the most POS database converted into RFM (Recency, Frequency, Monetory) dataset.(LTV) and (RFM) are two of the most popular techniques of customer segmentation: Consumer Lifetime Value (LTV).

Table 1. Market Segmentation Bases

Base: Description	Objectives Benefits	Ref.
Behavioral; Product attitudes, customer relationship-related features	The goal is to identify behavioral variables such as occasions, benefits, user status, usage rate, buyer-readiness stage, loyalty status and attitude.	9
Demographic; Identifiable population features	The goal is to have a precise customer purchase profile and focuses on measurable criteria of consumers and their households.	20
Geographic; Location-related features	The goal is to have a precise customer purchase profile and focuses on measurable criteria of consumers and their households.	9, 5

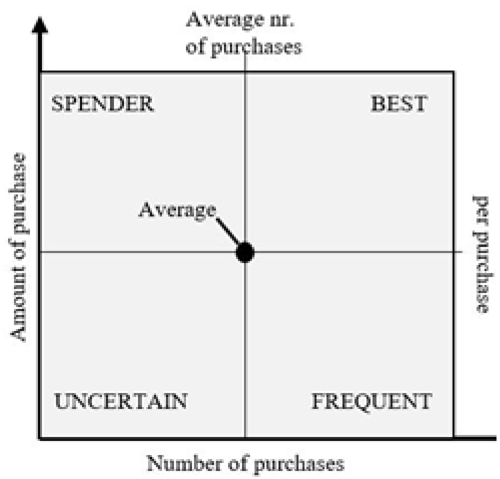

Fig. 1The customer value matrix

LTV is a vital component in Customer Relationship Management (CRM). LTV is a quantitative measurement on the number of sales the consumersare expected to commit to the retailer in their total time [14]. (LTV)is paired to POS Dataset to predict the future cash flows attributed to the consumers during his or her entire lifetime with the retailer before he or she churn [35].

2. IMPERIAL CASE STUDY

A medium size retailer with 65 domestic and international branches. The retailer consumer-base is now diverse, and consumer's demographic and behavioral changes are no longer possible to define consumer shopping profiles based on the traditional known shopping. In the retail industry, having access to customer data is vital to develop a successful marketing strategy, but few retailers are leveraging the information that is available from an integrated POS database. Point of Sales dataset contains a wealth of multitudes of consumer's data points about and key deliverables of the data. This study investigates the evolution of consumer's expectation, the technologies driving him or her, and the balance of trust between consumer and retailer. With more access to information, more choices, and less incentive to be loyal, the consumer nowadays is firmly in control of his or her relationships with the retailers. This study aims to design and develop advanced behavioral segmentation methods. The idea is to show how the customer's behavioral segmentation process can be significantly improved with a hybrid approach that utilizes statistical and machine learning methods. In this study (LTV) Customer Lifetime Value, best fit regression and clustering are used as steps in the analysis of a Point of Sales (POS) dataset, to answer the following study questions:

. 1. Can the Hyped method of statistical (LTV), and machine learning best fit regression and clustering algorithmshelpto revealhidden and vital patterns about consumer shopping behavior?

2. Can the proposed method help to map the entire consumer's journey? In the next sections we will elaborate more on the appropriate literature, in section 2, developing the proposed model in section 3, and comparative analysis of the results in section 4. Conclusions follow in section 5 with answers to the aforementioned study questions.

2.1 SMALL AND MEDIUM SIZE BUSINESSES

The formal definition of a small to medium businesses vary widely, The U.S. Small Business Administration defines as an independently owned business with less or fewer than 1500 employees, with less or more than USD 50 million in annual revenue [41]. Small and medium-size business has long been regarded as the engine of economic development and growth [41]. The small and medium enterprises play a vital role in the global economy with its ability to fuel economic development and growth. Predominantly, the small and medium size business are credited for most of the jobs that are available in developing markets, around 45% of the total employment and 33% of the national income in developing countries [19], [41]. Historically, consumers have had basic expectation like fair pricing and quality service, however in this era of exponentially technological rich in consumer environments, modern consumers have become much more aware, their expectations are higher, and demand more personalized services from the retailer, such as personalized interactions, proactive service, and connected experience across channels [42]. Services and products that are cutting edge today, they will be outdated the next. In this ever-changing context, to win the consumer's heart and wallet, retailers must not only rely on exceptional marketing, sales, and service interactions but also most importantly to prove that the customer is the center of interactions. This can be only achieved by gaining vital Return On Investment (ROI) insights and delivering real-time focused communication with consumers [20].

2.1.1 CUSTOMER SEGMENTATION

Dipanjan, Satish, and Goutam [12] defined market segmentation as the process to divide consumers into similar or homogeneous groups sharing one or more characteristics such as shopping habits, lifestyle, taste, and food preferences. According to McCarty and Hastak [30], these characteristics, such as demographics, age, nationality, gender, geographic, such country or city, behavioral such as interests and spending habits, and firmographic such as organization size and industry are relevant to marketing. Traditional approaches to customer segmentation mainly focused on who consumers are, and segments are based on one characteristic such as demographic, geographic, behavioral or firmographic attributes [44]. However, just understanding who are your consumers is not enough. Behavioral segmentation focuses on understanding consumers by what they do, not only by who they are, using insights derived from consumers' actions. Behavioral segmentation which can be derived from combining the attributes mentioned above looks in studying the behavioral consumer's traits: like occasion, their likes and dislikes, spending patterns, attitude, culture, choice, shopping trend, etc. [43]. Thus, in the long run, the technique helps the retailer develop an intelligent, focused marketing strategy that will help retailers improve and expand their consumers base [43], [44], [45]. Of course, consumers shopping behavior vary in some many ways as per different variables. Some consumers may seek pocket-friendly products or services, while other consumers will look for brand names. Some consumers are always loyal to their retailer's brands; some are loyal to old products or brands, whereas most consumers wish to innovate. Therefore, behavioral segmentation can be defined as sorting and dividing consumers as per their traits or characters, and accordingly using focused marketing techniques that will improve consumer's services and optimize sales [33]. Consumer Behavioral Segmentation is a type of consumer segmentation that is based on patterns of displayed consumer behavior as they interact with an organization, brand, services or make a shopping decision. The method empowers businesses to divide consumers into groups and subgroups according to consumer's knowledge of, use of, attitude towards, or response to a brand, product, service [44]. Behavioral Segmentation main objective is to identify consumers' segments that enable a business to understand how to address the group of consumers particular needs or desires, discover unknown opportunities to optimize consumers' journeys and quantify consumers potential value to the organization's business.

As consumers research a service or product, their shopping behavior can reveal valuable shopping insights into which benefits, features, or problems are most applicable to consumers, and also when a consumer places a higher value on certain benefits over others. these main benefits are the consumers shopping driving. In this study we discuss four main benefits of grouping consumers into different segments using Behavioral and Demographic Segmentation methods[45].

1. Personalization from a retailer perspective, understanding how different groups of consumers should be targeted with different offers, precisely at the most appropriate times and date through their preferred shopping channels, and effectively help consumers advance towards successful shopping outcomes and experience in their journeys, this will trigger the retailer to develop Personalizationservices for each group of consumers [44], [45].

2. Behavioral Segmentation offers retailers the Predictive capability, by analyzing historical consumer's behavioral patterns to predict and influence consumer's future shopping behaviors and outcomes [44], [45].

3. Prioritization, Behavioral Segmentation Methods help retailers making smarter decisions on how to best allocate budget and resources, time by identifying profitable and high-value consumers segments and initiatives with the utmost potential business impact [44], [45].

4. Performance, the segmentation helps retailer Monitor sales growth patterns and changes in key consumers segments over specific time to track performance against the retailer's goals. This leads to quantifying the value and size of consumers segments, and tracking how positive and negative, profitable and not profitable consumers segments are growing or shrinking over a specific period of time [33], [44], [45].

2.1.2 RFM WITH CUSTOMER LIFETIME VALUE (LTV)

According to the literature RFM is defined as the standard analysis approach to assess and understand consumer lifetime value, and it is quite popular, especially in the retail industry. Tsiptsis and Chorianopoulos 33 define RFM as the process involving the calculation and the examination of three variables Recency, Frequency, and Monetary (RFM). RFM segmentation allows retailers and marketers to target specific groups (clusters) of consumers with communication that is much more relevant for their particular shopping behavior [24]. Therefore, RFM generates much higher rates of consumer's response, increase in consumer's lifetime value and loyalty. RFM segmentation is a powerful method to identify groups of consumers for special service [8]. In the retail industry, marketing analysis typically has extensive consumers' data, such as consumer shopping history, prior marketing campaign response patterns. The insights that could be generated using RFM techniques can be used by marketers to identify specific groups of consumers (clusters) that can be addressed with offers very relevant to each group (cluster). While there are many ways to perform consumers' segmentation, RFM analysis popularity credited for three reasons, 1. numerical scales that yield to an informative and concise high-level depiction of consumers. 2. Simplicity, marketing analysis can use the RFM model effectively with the need for sophisticated analytical software or data scientists. Intuitively, the output generated by the RFM segmentation is easy to understand [4], [24].

Recency refers to the inverse of the most recent interval from the time when the latest consuming behavior happens to the present moment. Frequency is the number of events the consumer purchases in a period. Monetary is merely the amount of money consumed during the period. As the weighted average of its individual components, the RFM score is calculated as

$$\text{RFM score}_{\text{scaled}} = \frac{\text{RFM score- min(RFM score)}}{\max(\text{RFM score}) - \min(\text{RFM score})} \tag{1}$$

where *rs* = recency score and *rw* = recency weight, *fs* = frequency score and *fw* = frequency weight, *ms* = monetary score and *mw* = monetary weight.

Dwyer 14 defines consumer lifetime value (LTV) as a quantitative measurement of the amount of sales the consumer is expected to spend with a retailer over their lifetime. Furthermore, Safari 31 considers LTV as the present value of all future profits obtained from a consumer over his or her lifetime relationship with the retailer. Loyal consumers cost retailers less to serve, their shopping consumption is more than other consumers and attract more consumers through word of mouth. To better utilize LTV in every-day decision making, Marcus 29 introduced LTV matric as a variant of the RFM analysis for small-business. In LTV matric, F, the frequency of purchase and M, the average purchase amount are used for the segmenting consumers. The easiness to understand quadrant identifiers is considered as its main advantage. In Marcus' approach, the average values for the number of purchases and the average amount spent per consumer are calculated. After identifying these, each consumer is segmented to one of the four resulting categories (quadrants) based on whether consumers are above or below the axis averages (Fig. 1).

Table 2 Encoding of the Age

Age	1-15	15-30	30-45	45-60	65 +
Category	1	2	3	4	5

2.1.3 DATA MINING

Modern information and communication technology generates massive amounts of data to databases, data warehouses, and other repositories. Transforming the insights about (big) data into knowledge can help retailers to make better business decisions 9. Tufféry 34 sees data mining as a powerful analytical tool for gaining insight into the retail industry. According to Azevedo 2, data mining provides the analyses on product sales, consumer buying habits, data and identify naturally occurring clusters of behavior, which then form the basis of segments. Ramageri and Desai 31 say that in the retail sector, data mining offers insightful measures, taking into account all the factors that affect the value of the consumer to the retailer over the entire course of consumer relationship [34], [16].

Gunaseelan and Uma 21 stated that the main aim of data mining is to discover valuable patterns from a large collection of data for users. It can identify patterns, and apply data analysis and discovery algorithms to produce a data mining model. Models help in generating a model, a hypothesis about the data, that key executives can use to make better-informed decisions [2]Error! Reference source not found. There are two primary data mining process goals, which are verification and discovery. Verification is verifying the user's hypothesis about the data while discovery is automation of finding unknown patterns [28], [6].

2.1.4 CLUSTERING ALGORITHMS

Instead of analyzing the entire consumers base as a whole, it's far more efficient to segment consumers base into homogeneous groups (clusters), understand the consumers' traits of each group, and engage each group with relevant more focused marketing campaigns rather than segmentation on just consumers demographic or geography variables. This study uses the power of machine learning algorithms, using two different clustering algorithms, soft clustering algorithm Fuzzy C-means (FCM), Hard clustering algorithm Expectation-Maximization (EM) to devolve consumers' behavior segmentation. According to Lefait and Kechadi 28, clustering consists of "creating groups of objects based on their features in such a way that the objects belonging to the same groups are similar, and those belonging to different groups are dissimilar. Clustering analysis is one of the most important and prominent customer segmentation techniques, and it has long been the dominant and preferred method for customer segmentation [23]. For example, D'Urso [37] used FCM method to cluster potential Chinese travelers. The FCM method combines partitioning and hierarchical clustering procedures. Khavand and Tarokh 24 proposed a data mining tool to prepare a framework for segmenting consumers based on their estimated future LTV value in an Iranian private bank, and the method was implemented in a health and beauty company, as well 25. In retail sales clustering methods have been applied at least in groceries Error! Reference source not found., online retail 8 and for identifying strategies for new ventures [7].

Mirkin [36] proposed a framework for partitional fuzzy clustering which suggests a model of how the data are generated from a cluster structure to be identified. R.Suganya [41] extend the FCMP framework to a number of clustering criteria and study the FCMP properties on fitting the underlying proposed model from which data is generated. D'Urso, [37] used FCM method to cluster potential Chinese travelers. The FCM method combines partitioning and hierarchical clustering procedures. Consumer segmentation to help to analyze transaction data with Fuzzy C-Means for clustering and Fuzzy RFM toidentifying the consumers who have high and low loyalty.

Fuzzy C-Means algorithm (FCM), is one of the best known and the most widely used fuzzy clustering algorithms [41]. FCM allow each point to have a degree with every cluster center. Each data points are given with a value between 0 and 1 memberships to determine the degree of belonging to each group. The performance of FCM clustering depends on the selection of the initial cluster center and the initial membership value [89]. FCM has a wide domain of applications such as agricultural engineering, astronomy, chemistry, geology, image analysis, medical diagnosis, and target recognition [44].

Expectation-maximization (EM) clustering algorithm 11 is closely related to the K-Means algorithm.

In this algorithm, two subsequent steps are iterated until there are no more changes in the current hypothesis [37]. In the Expectation-step (E-step) the probability that each observation is a member of each of the chosen class is calculated. Maximization-step (M-step) alters the parameters of each class with the objective to maximize those probabilities. The iteration is then repeated until converging to a (local) optimum.

3. RESEACH METHODOLIGIES

This section presents those data mining techniques used in this study. The analysis process takes four phases. The first phase focuses on data preprocessing, data cleansing, and data transformation. Customer lifetime value (LTV) is applied against the (RFM) dataset in the second phase. The output is then fed into the clustering algorithms with RFM data generated by the (LTV) matric. Clustering algorithms are applied in the third phase, respectively. Fuzzy c-means (FCM) and EM clustering algorithms are used for the customers' segmentation. At the final phase, the accuracy of these partitions is measured by the cluster quality assessment introduced byDrăghici 13.

The point-of-sales database consists of all product sales and shows that the client sells diverse products like clothing, shoes, perfumes, and accessories, and Each transaction represents a consumer shopping experience. Transaction data from two years are retrieved and stored into an Oracle Data warehouse.

3.1 PHASE 1: DATA PRE-PROCESSING

To prepare for this-this stage, several interviews are conducted with marketing experts, sales directors, the IT manager, in-store employees, and POS engineers. Interviews intend to maximize variation in responses to gain a deep understanding on the challenges experienced at the client company in general and, more specifically, to find information and company insights about the retail industry, the market, and the consumer base.

3.1.1 DATA TRANSFORMATION AND CONVERSION USING ENHANCED RFM MODEL

Based on the client information, RFM values are assigned. The latest purchase date of the consumer R is found from a set of 730 days (records from 2016 to 2018). After considering some rudimentary questions, as records containing NULL values and records with no customer's identification perceived as a viable consumer's data. we have decided to eliminate null values, missing values and records with no customer's identification. The number of transactions during this period F comes from totally 7307 transactions, and total amount purchased M comes from the total sales of 1,300,000 KD. The RFM attributes are weighted with category 10. In this phase, the string variables are convertedto numeric variables, and the data are transformed into a unified format to make the discovery of patterns easier. Continuous consumer-related attributes are encoded by decreasing the original values into a small number of value ranges. The age of the consumer was encoded into 5 categories shown in table 1.

Occasions attributes are encoded into 5 categories as 1 for Ramadan, 2 for EID holidays, 3. For Christmas, 4. For New Year and 5 for Back to School. The Gender attribute is encoded as 1 for Male, 2 for Female and 3 for Organization, Furthermore, demographic variablesare replaced by higher level concept nationality.

For normalizing the RFMscores,we will the following rescaling formula (2)

$$\text{RFM score}_{\text{scaled}} = \frac{\text{RFM score- min(RFM score)}}{\max(\text{RFM score})-\min(\text{RFM score})}. \tag{2}$$

3.1.2 PHASE 2: GENERATING CUSTOMER LIFETIME VALUEUSING (LTV) MATRIC ON RFM DATASET

In the retail industry consumers acquisition costs maybe equal, less or more than you make from the consumer first purchase, but the important question is, is the retailer still making a profile from that consumer in the long run? Figuring out the lifetime value of a consumer will give the retailer the answer. Performing LTV analysis against RFM dataset can effectively improve consumer acquisition and consumer retention, prevent churn, measure the performance of marketing in more detail, help the retailer plan much more efficient marketing budget, and ultimately identify which acquisition channel produces the highest value consumers. The LTV matric divides the RFM dataset into four segments (Best, Spender, Frequent, Uncertain) in terms of the period since the consumer made his or her last transaction (recency), total purchase frequency and total purchase consumer expenditure (monetary). The LTV matric performs the calculation on the number of purchases by dividing the total number of purchases for the consumer with the total number of consumers in the consumer database. The average purchase amount is derived by dividing the total revenue with the total number of purchases. Comparing the average number of purchases, F between consumers and the average purchase amount, M with total average values is the next step. M and F are used to classify each consumer into one of the four categories (Best, Spender, Frequent, Uncertain). Four potential RFM combinations of inputs can be obtained by assigning upward arrow ↑or downward ↓, according to the (RFM) average values of each segment being greater than or less than the overall average. If the average of (R) value of a segment exceeded the overall average (R), then an upward arrow ↑ is added, otherwise and downward arrow ↓ is added. For example, R↑F↑M↓ resemble that the average recency and frequency values of a consumer segment is greater than the overall average, while the monetary average value is smaller than overall averages. These four customer segments include best consumers (Most valuable), spenders (Valuable consumers, frequent consumers (Less valuable), spenders, and the uncertain consumers are (The least valuable). Formula 1 illustrate the LTV formula used by the LTV matric and Table 3 illustrates the result, listing four segments, each with their actual average of R, F and M values.

Formula 1: LTV Formula

$$LTV = \sum_{t=0}^{t} \frac{(p_t - C_t)r_t}{(1+i)^t} - AC$$

Table 3 The Four RFM Segments Generated by LTV Matric

Avg_Recency	Avg_Frequency	Avg_Monetary	Occasion	AGE_ID	GENDER_ID	No_of Cutomer	%_of Cutomer	Category
↓ 342	↓ 19	↑ 5,465	2 | 4	1	2	1200	8%	Best
↓ 129	↓ 56	↑ 4,214	2	6	1	2673	18%	Spender
↑ 34	↑ 113	↓ 300	5 | 2	3	2	4107	27%	Frequent
↑ 16	↑ 107	↓ 115	5 | 1	5	2	7020	47%	Uncertain
					Total	**15000**	**100%**	

Table 3 illustrate the average Recency, average Frequency, and the average Monterey for (Best, Spender, Frequent, Uncertain). The result reveals important insights as the two most profitable segment (Best, Spender) trailing badly in recency and frequency.

3.2 PHASE 4: CUSTOMER SEGMENTATION USING CLUSTERING ALGORITHMS

Next we use data mining analytical capabilities to group consumers in each segment into different groups (Cluster) based on similarity in consumers shopping behavior and patterns similarities. The RFM dataset generated by the LTV matrix is fed into two unsupervised clustering algorithms. The analysis result will trigger the retailer to harness each cluster with appropriate marketing actions.

3.2.1 FUZZY C-MEANS SOFT CLUSTERING

Fuzzy c-means (FCM) is unsupervised Clustering of numerical data method developed by Dunn in 1973 [41]. The algorithm allows one piece of data corresponding to two or more clusters with the same weight, and it uses the concepts from the field of fuzzy set theory and fuzzy logic [41]. The FCM is frequently used in pattern recognition and employs fuzzy partitioning such that a data point can belong to all groups with different membership grades between 0 and 1 [42]. The Fuzzy c-mean algorithm is composed of the following steps

1. Step 1. Initialize U =[x_{ij}] matric, $U^{(0)}$ (7)
2. Step 2. At k-step: calculate the centers vectors (8)
3. $C^{(k)}$=[c_j] with $U^{(k)}$

4. $9C_i = \dfrac{\sum_{j=1}^{n} u_{ij}^{m} x_j}{\sum_{j=1}^{n} u_{ij}^{m}}$

Step 3. Update $U^{(k)}$, $U^{(k+1)}$ (9)

5. $d_{ij} = \sqrt{\sum_{i=1}^{n} (x_{i} - c_{j})}$ (10)

$$\text{Step 4. } u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m_{-1})}}$$

6. Step 5. if || $U(k+1) - u(k)$ || $< \epsilon$ *then STOP; otherwise return to step 2* (11)

the m is any real number greater than 1, u_{ij} is the degree of memberships of x_i in the cluster j, x_i is the ith of d-dimensional measured data, c_i is the d-dimension center of the cluster. This algorithm works by assigning membership to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point. More the data is near to the cluster center more is its membership towards the particular cluster center. Clearly, summation of membership of each data point should be equal to one. After each iteration membership and cluster centers are updated according to the formula.

3.2.2 EXPECTATION- MAXIMIZATION (EM) HARD CLUSTERING

The Expectation-Maximization (EM) is an iterative estimation algorithm developed by Dempster, Laird, and Rubin 11. EM is historically very important algorithm in market segmentation and data mining. EM algorithm has also proven its efficiency in a good performance, decreasing sensitivity to noise and estimation problem involving unlabeled data.

Step 1. Expectation- Maximization (EM) Clustering Initialization, the E-step
Every class j, of M classes (or clusters), is formed by a vector parameter (θ), composed by the mean (μ_j) and by the covariance matric (P_j), which defines the Gaussian probability distribution (Normal) features used to characterize the observed and unobserved entities of the data set as shown in Equation (11)

$$\theta(t)=\left(\mu_j(t),P_j(t)\right),j=1,\ldots\ldots,M. \quad (12)$$

On the initial instant ($t=0$) the implementation can generate randomly the initial values of mean (μ_j) and of the covariance matric (P_j). The EM algorithm aims to approximate the parameter vector (θ) of the real distribution.

Fraley and Raftery 18 suggest another alternative to initialize (EM) with the clusters obtained by a hierarchical clustering technique. The relevance degree of the points of each cluster is given by the likelihood of each element attribute in comparison with the attributes of the other elements of cluster C_j as shown in Equation (12) The E-step

$$\left(C_j \middle| x\right)=\frac{|\sum_j(t)^{-\frac{1}{2}}|e^{n_j}P_j(t)}{\sum_{k=1}^{M}|\sum_j(t)|^{-\frac{1}{2}}e^{n_j}P_k(t)}, \quad (13)$$

Step 2. M-Step

First is computed the mean (μ) of class j obtained through the mean of all points in function of the relevance degree of each point, as shown in Equation (13)

$$\mu_j(t+1)=\frac{\sum_{k=1}^{N} p(C_j|x_k)x_k}{\sum_{k=1}^{N} P(C_j|x_k)}, \quad (14)$$

To compute the covariance matric for the next iteration the with the *Bayes Theorem,P(A|B) = P(B|A) * P(A)P (B)* conditional probabilities of the class occurrence are calculated, as shown in Equation (14)

$$\sum_j(t+1)=\frac{\sum_{k=1}^{N} P(C_j|x_k)(x_k-\mu_j(t))(x_k-\mu_j(t))}{\sum_{k=1}^{N} P(C_j|x_k)}. \quad (15)$$

The probability of occurrence of each class is computed through the mean of probabilities (C_j)

$$P_j(t+1)=\frac{1}{N}\sum_{k=1}^{N} P(C_j|x_k). \quad (16)$$

Step 3. Cluster Convergence

After performing each iteration, a convergence inspection which verifies if the difference of the attributes vector of an iteration to the previous iteration is smaller than an acceptable error tolerance, given by parameter.

4. ANALYSIS RESULTS AND DISCUSSION

Data analysis, though often rushed, is the most important stage in the customer segmentation solution. The analysis is here conducted under the supervision of internal marketing and sales manager to identify which variable and segment make the most sense to focus all efforts on. Such a framework identifies the model strengths and weaknesses, with special attention paid to all implications stemming from each. The internal marketing and sales managers possess an understanding of client capabilities and resources. It helps the client to focus on the consumer and develop marketing mixes for a very specific customer segment. In this section, we will discuss the analysis result from the Spender segment. This segment is the second most profitable segment for the retailer, and it has the potential in moving to the best segment if the retailer truly understands the consumer's behavioral changes in this segment. Important to note, this study the RFM scores calculation method for each behavioral and demographic variable is based on Weighted RFM model (1 – 5), where 1 is worst average score and 5 is the best average score.

Table 6 shows the four clusters generated by EM algorithm, and includes information such as the number of consumers in each cluster, the average of recency, frequency, monetary, best occasion, age, gender and nationality of each cluster. The EM algorithm analysis shows cluster 1 is the best cluster and it is the most beneficial segment because it is superior to the other in terms of best recency (3), frequency (4), and monetary (5).

The analysis also reveals that Female consumers from the GCC region between the age of 30-45 are the best consumers in the Spender category. Their favorite shopping ocean is the EID holiday, follows by the new year holiday, and lastly to some small percentage is the month of Ramadan. Consumers between the age of 30-45 are the best consumers. Figure 2 show the layout of the RFM dataset before and after applying the EM clustering algorithm.

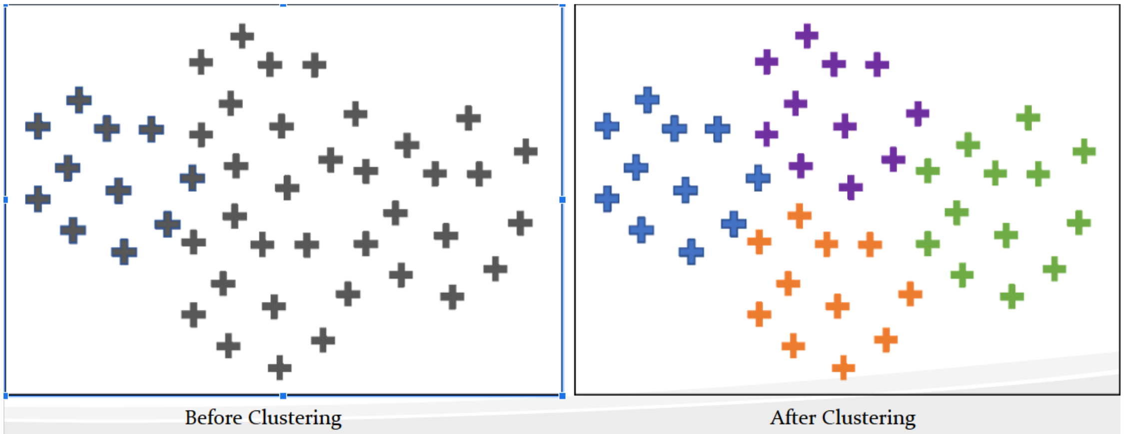

Fig. 2 RFM Dataset before and after applying EM algorithm.

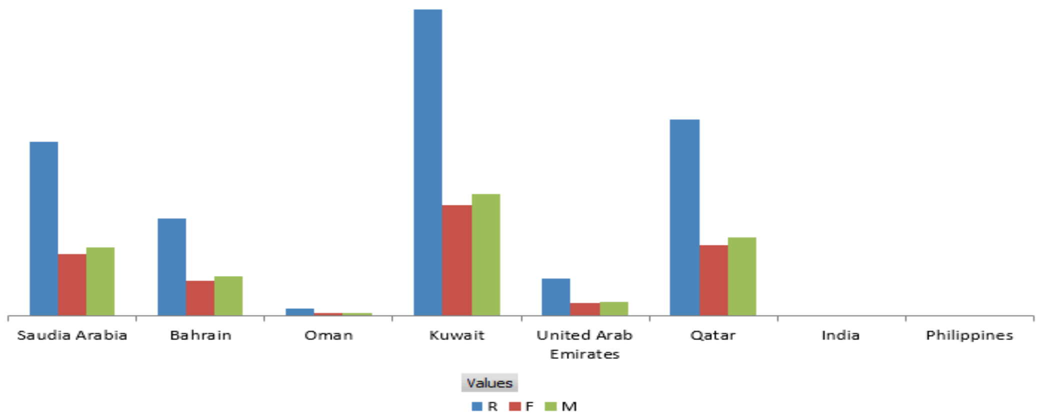

Fig. 3 Top Nationalities with the Best Shopping Expenditure

Table 4. Summary Analysis Of Spender Segment generated By The Em Algorithm Based On The Rfm Dataset

Best Cluster	No. of consumers	Avg. Recency	Avg. Frequency	Avg. Monetary	Avg. Occasion	Avg. Gender	Avg. Age	Avg. Nationality
1	1035	3	3	4	2	Female	30-45	1
2	962	2	3	3	2	Female	30-45	2 | 3
3	378	2	2	3	5	Female	45-60	1 | 3
4	298	2	1	3	5 | 2	Female	1 - 30	11

Table 5 Summary Analysis of Spender Segment Generated by the Fuzzy C-Means (Fcm) Algorithm Based on the Rfm Dataset

Best Cluster	No. of consumers	Avg. Recency	Avg. Frequency	Avg. Monetary	Avg. Occasion	Avg. Gender	Avg. Age	Avg. Nationality
1	835	3	3	4	2	Female	30-45	1
2	797	3	3	4	5	Female	30-45	3 | 5
3	551	2	2	3	5	Female	45-60	2 | 3
4	490	1	3	3	4 | 5	Male	1 - 30	9

4.1 ACCURACY AND EFFECTIVENESS DETERMINATION (INTER-CLUSTER DISTANCE)

RFM consumers' dataset can be clustered in different ways, but how can one know the generated clusters are accurate and meaningful? One of the best techniques to answer the question is using Draghici approach. Clustering has become a key technique in analyzing quality assessment in a variety of recent studies. There are several studies supporting suggestions for measuring the similarity between clustering algorithms. Those measures are used to compare how accurate different clustering algorithms are on a dataset. Accuracy is usually tied to the type of benchmark being considered. The Draghici approach [13] is to compare the size of the clusters vs. the distance to the nearest cluster (the inter-cluster distance vs. the size (diameter) of the cluster). That is the distance between the members of a cluster and the cluster's center, and the diameter of the smallest sphere containing the cluster. If the inter-cluster distance is much larger than the size of the clusters, then the clustering method is trustworthy. Table 4 shows that the quality can be assessed simply by looking at the cluster diameter. The cluster is created by means of a heuristic even when there is no similarity between clustered patterns. This is occurring because the algorithm forces K clusters to be created. Comparing both algorithms using cluster quality assessment on the Spender segment shows EM clustering with the size of the cluster (333.7669039) is more accurate than the Fuzzy c-means (FCM) algorithm with the size of the cluster (276.9641345).

Table 5. Cluster Quality Assessement for the Spender Segment using the EM Algorithm

Spender SEGMENT	EM
Inter-Cluster Distance	
D12	1313039
D13	2366321
D14	2467183
D23	8771418
D24	9410945
D34	10513277
	Size of Cluster (Diameter)
Cluster	Diameter
D1	1129
D2	2328
D3	3569
D4	3934
Cluster Quality	333.7669039

Table 6. Cluster Quality Assessement for the Spender Segment using the (FCM) Algorithm

Spender SEGMENT	Fuzzy C-means (FCM)
Inter-Cluster Distance	
D12	1112011
D13	2154223
D14	2256289
D23	5986325
D24	6852652
D34	6975265
	Size of Cluster (Diameter)
Cluster	Diameter
D1	1524
D2	2659
D3	3624
D4	4015
Cluster Quality	276.9641345

4.2 SUMMARY ANALYSIS

Clustering algorithms successfully divided diverse consumers-base in each segment generated by the LTV into more homogeneous and smaller groups (segments) based on consumer's characteristics such as occasion, age, personality traits, and gender. This can help the retailer build blocks of marketing and initiate a more focused marketing campaign to meet consumer's expectations and identify the likelihood of developing better relationships with potential consumers according to the EM algorithm analysis results of the Best, Spender, Frequent, Uncertain segments. First, we will focus more on the Spender segment, because according to the client it has the best potential and it closely resembles their most loyal consumers.

In the Spender segment consumers in all four clusters in the Spender segment shown different shopping behavior. Cluster 1 ranked first and best segment of consumers. Consumers in this cluster are high in Recency (4), Frequency (4), and strong purchase power with a monetary score (4). Consumers in this cluster have an increasing average purchase power during the period of analysis. Such cluster is very profitable and should be treated accordingly and the potential in moving to the Best segment. Cluster 2 ranked second and second best segment of consumers. Consumers in this cluster are low in Recency (4), acceptable Frequency score (4), and acceptable purchase power with a monetary score (4). Consumers in this cluster have an increasing average purchase power during the period of analysis. The retailer should focus on why is the drop of recency and frequency. This segment has the potential in moving to higher cluster, such cluster is still very profitable and should be approached with a special marketing message. Cluster 3 ranked third with low Recency (2), Low Frequency (2) of visits, however, shown reasonably acceptable score in Monetary (M). The consumers belong in this cluster are expected to be falling from the beneficial segment for a non-beneficial segment. Cluster 4 ranked at the end with low Recency (2), very low Frequency (1) of visits, however reasonably acceptable score in Monetary (3). It can be concluded that the consumers belong in this cluster are at high risk of shifting their business to another retailer. One of the most common indicators of high-risk consumers is a drop off in purchases and decrease of shopping visits. Female consumers are the dominant consumers in this segment, and the EID holidays are their preferred shopping times.

The results generated from EM analysis of the Best segment show consumers in cluster 2 has the best purchase power with high Recency (4), High Frequency (4), and low with a monetary score (2). This ranks the Frequent Segment below the Spender Segment with respect to the consumers total monetary (M). A mix of Male and Female consumers are the best consumers in this segment, and the EID holidays and the month of Ramadan are their preferred shopping times. The results generated from EM analysis of the Frequent segment results show consumers in cluster 1 has the best purchase power (5) with very low Recency (4), Low Frequency (2) scores. This ranks the Best Segment the worst in terms of the consumers' recency and frequency. Female consumers (1) are the best consumers with some major purchases form companies (3). In this segment, and the EID holidays and the month of Ramadan are their preferred shopping times.In Fuzzy c-means (FCM) analysis, the Uncertain segment ranked the lowest average purchase power (M), with high Rececny (4), and Frequency (4) of visits. Male consumers are the dominate consumers in this segment, and the month of Ramadan is their preferred shopping time.

In Fuzzy c-means (FCM) analysis, the Uncertain segment ranked the lowest average purchase power (M), with high Rececny (4), and Frequency (4) of visits, this concluded that consumers in this consumers visit often visit and shop very less. More recommended strategies for the four segments in table 7.

Table 7: Recommended Strategies for The Best, Spender, Frequent, Uncertain Segments

Segment	R	F	M	Recommended Strategy
BEST	V Low	Low	High	▪ Recognizing the importance of these customers ▪ VIP Communication ▪ VIP and special Services ▪ Providing online shopping POS
SPENDER	High	High	High	▪ Communication. ▪ VIP Communication ▪ Informing them about new products and services in a timely fashion.
FREQUENT	Low	High	Low	▪ Frequently, promotion plans ▪ Cross-selling and upselling. ▪ Special discount
UNCERTAIN	High	High	High	▪ Get rid of them

5. SUMMARY

In this study, we examined different methodologies to capture behavioral changes in consumer's shopping behavior.

Firstly, Draghici approach is used to compare the size of the clusters vs. the distance to the nearest cluster.

Secondly, Industry marketing experts as human judgment to validating the accuracy and intelligence of the results.

Analysis and experts agree that classifying consumer purchase behavior using an LTV matric against normally distributed standard RFM Dataset gives some percentage of data accuracy about the entire consumer's journey shopping information. Classification of consumer's shopping power use LTV matric against RFM dataset is to classify the changes in the consumer purchase power over two years' time period. A behavioral variable like Occasion showed a high degree of accuracy, with an estimated accuracy percentage of 78%. Age and gender variables have shown a lower percentage of accuracy related to the consumer purchase expenditure (M) with an estimated accuracy of 53%. However, the nationality variable showed the lowest percentage of accuracy, possibly because many non-citizens are reluctant in revealing their nationality or ethnic background.

5.1 CONCLUSION, CONTRIBUTION, AND ANSWERS TO THE STUDY QUESTIONS

In this study, we proposed the use of Customer Lifetime Value (LTV) matric with standard RFM scores (R, F, M) model to segment a POS converted into RFM dataset using a combination of behavioral and demographic variables. In this study, we proposed consumer behavioral analysis methods integrating LTV matric with standard RFM in two data mining clustering methods. Soft clustering Fuzzy C-Means (FCM) and hard clustering Expectation Maximization (EM) algorithms methods apply separately for every RFM variation. Classifying consumers with different reflected shopping behavior power (Best, Spender, Frequent, Uncertain), and then group each segment based shopping similarity is the objective of this study.

To help mapping the entire consumer's journey, in this study existing consumer purchase profile of behavioral and demographic variables like occasion, age, gender and nationality are segmented and presented accordingly. The study has shown that combining LTV and RFM methods provide an acceptable base for classifying consumer's purchase power and understanding customer segmentation with different values of Recency, Frequency, Monetary. The model using the proposed methodologies can classify profitable consumers from non-profitable consumers, and some to some degree identify those consumers who might be shifted or at the high risk of shifting their business to another competitor. Two clustering algorithms were tested in this study: Hard Clustering EM, and soft clustering Fuzzy c-means. The experiment showed Fuzzy c-means difficulty in selecting the initial cluster centers and required more computation time as it requires more iterations than the EM algorithm. However, it shows lower clustering errors. It is more efficient for large scale clustering, while the EM is more suitable for small scale clustering. The model provided simplicity in classifying consumers based on purchase power, the frequency of making a purchase. However, the proposed model failed to measure consumer shopping trends over the selected periods accurately. The main contribution is a novel and simple customer segmentation method using customer lifetime value matric (LTV) to classify consumers into four quadrants (Best, Spender, Frequent, Uncertain) based on the enhanced RFM variables followed by clustering on behavioral and demographic data. The study provides vital attributes describing consumer shopping behavior in different behavioral and demographic characteristics such as occasion, gender, age, and nationality. Traditional RFM model is essential analysis tool for organizations to design more customized and focused marketing strategies, in this study standard RFM revealed vital information about consumers with significant purchase power, but showed far less precise insights about behavioral changes of consumers during the period of analysis concerning consumers recency and frequency. Second, as a well-known limitation of standard RFM analysis the inability to prospect for new customers, in this study standard, RFM revealed information on the retailer's current consumers. For future research, is to develop comprehensive methodologies by extending the RFM model and developing new RFM variants and more fitting algorithm to capture consumer's shopping trend and consumer's churn probability. Finally, the analytical information gained from extracted data is useful to adopt more targeted marketing strategies and for decision making. We can confirm that proposed customer segmentation using behavioral and demographic variables can contribute to the body of literature in consumer purchase behavior and assist retailers in meeting the needs and preference of consumers.

REFERENCES

[1] APS Meeting Abstracts (Vol. 1, p. 11002).

[2] Azevedo, A. (Ed.). (2014). Integration of Data Mining in Business Intelligence Systems. IGI Global.

[3] Bernard, H. R. (2011). Study methods in anthropology: Qualitative and quantitative approaches. Rowman Altamira.

[4] Birant, D. (2011). Data Mining Using RFM Analysis, INTECH Open Access Publisher.

[5] Broderick, A., & Pickton, D. (2005). Integrated marketing communications. Pearson Education UK.

[6] concept to implementation. Prentice-Hall, Inc.

[7] Carter, N. M., Stearns, T. M., Reynolds, P. D., & Miller, B. A. (1994). New venture strategies: Theory development with an empirical base. Strategic Management Journal, 15(1), 21-41.

[8] Chen, D., Sain, S. L., & Guo, K. (2012). Data mining for the online retail industry: A case study of RFM model-based consumer segmentation using data mining. Journal of Database Marketing & Consumer Strategy Management, 19(3), 197-208.

[9] Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: from big data to a big impact. MIS Quarterly, 1165-1188.

[10] Cleveland, M., Papadopoulos, N., & Laroche, M. (2011). Identity, demographics, and consumer behaviors: International market segmentation across product categories. International Marketing Review, 28(3), 244-266.

[11] Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the royal statistical society. Series B (methodological), 1-38.

[12] Dipanjan, D., Satish, G., & Goutam, C. (2011). Comparison of Probabilistic-D and k-Means Clustering in Segment Profiles for B2B Markets. SAS Global Forum.

[13] Drăghici, S. (2003). Data analysis tools for DNA microarrays. CRC Press.

[14] Dwyer, F. R. (1997). Consumer lifetime valuation to support marketing decision making. Journal of interactive marketing, 11(4), 6-13.

[15] Elby, A. (2015, April). The new AP Physics exams: Integrating qualitative and quantitative reasoning. In APS Meeting Abstracts (Vol. 1, p. 11002).

[16] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. AI magazine, 17(3), 37.

[17] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. Communications of the ACM, 39(11), 27-34.

[18] Fraley, C., & Raftery, A. E. (2006). MCLUST version 3: an R package for normal mixture modeling and model-based clustering. WASHINGTON UNIV SEATTLE DEPT OF STATISTICS.

[19] Gal, A. (2010). Competitiveness of small and medium sized enterprises-a possible analytical framework. HEJ: ECO-100115-A.

[20] Goyat, S. (2011). The basis of market segmentation: a critical review of literature. European Journal of Business and Management, 3(9), 45-54.

[21] Gunaseelan, D., & Uma, P. (2012). An improved frequent pattern algorithm for mining association rules. International Journal of Information and Communication Technology Study, 2(5).

[22] Hossein Javaheri, S., (2008), Response Modeling in Direct Marketing: a data mining based approach for target selection, Masters thesis, epubl.luth.se/1653-0187/2008/014/LTU-PB-EX-08014-SE.pdf.

[23] Kashwan, K. R. & C. Velu (2013). Consumer Segmentation Using Clustering and Data Mining Techniques. International Journal of Computer Theory & Engineering 5(6): 856-861.

[24] Khajvand, M., & Tarokh, M. J. (2011). Estimating consumer future value of different consumer segments based on adapted RFM model in retail banking context. Procedia Computer Science, 3, 1327-1332.

[25] Khajvand, M., Zolfaghar, K., Ashoori, S., & Alizadeh, S. (2011). Estimating consumer lifetime value based on RFM analysis of consumer purchase behavior: Case study. Procedia Computer Science, 3, 57-63.

[26] Kolyshkina, I., Nankani, E., Simoff, S., & Denize, S. (2010). Retail analytics in the context of “Segmentation, Targeting, Optimization” of the operations of convenience store franchises. Anzmac.

[27] Lefait, G., & Kechadi, T. (2010, February). Consumer segmentation architecture based on clustering techniques. In Digital Society, 2010. ICDS'10. Fourth International Conference on (pp. 243-248). IEEE.

[29] Marcus, C. (1998). A practical yet meaningful approach to consumer segmentation. Journal of consumer marketing, 15(5), 494-504.

[30] McCarty JA, Hastak M (2007). Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression. J. Bus. Res., 60: 656-662.

[31] Ramageri, B. M., & Desai, B. L. (2013). Role of data mining in retail sector. International Journal on Computer Science and Engineering, 5(1), 47.

[32] Safari, M. (2015). Consumer Lifetime Value to managing marketing strategies in the financial services. International Letters of Social and Humanistic Sciences, 1(2), 164-173.

[33] Tsiptsis, K. K., & Chorianopoulos, A. (2011). Data mining techniques in CRM: inside consumer segmentation. John Wiley & Sons.

[34] Tufféry, S. (2011). Data mining and statistics for decision making, John Wiley & Sons.

[35] Ziafat, H., & Shakeri, M. (2014). Using Data Mining Techniques in Consumer Segmentation. International Journal of engineering Study and Applications, 1(4), 70-79.

[36] Nascimento, S., Mirkin, B., & Moura-Pires, F. (2000). A fuzzy clustering model of data and fuzzy c-means. In Fuzzy Systems, 2000. FUZZ IEEE 2000. The Ninth IEEE International Conference on (Vol. 1, pp. 302-307). IEEE.

[37] D'Urso, P., Prayag, G., Disegna, M., & Massari, R. (2013). Market segmentation using bagged fuzzy c–means (BFCM): Destination image of Western Europe among Chinese travelers.

[38] K.M. Bataineh, M.Naji, M.Saqer "A Comparison Study between Various Fuzzy Clustering Algorithms" vol.5 no.4 Aug. 2011

[39] Schafer, J. L. (1997). Analysis of incomplete multivariate data. Chapman and Hall/CRC.

[40] Altman, D. G., & Bland, J. M. (1995). Statistics notes: the normal distribution. Bmj, 310(6975), 298.

[41] Hillary, R. (Ed.). (2017). Small and medium-sized enterprises and the environment: business imperatives. Routledge.

[42] Storey, D. J. (2016). Understanding the small business sector. Routledge.

[43] An, J., Kwak, H., Jung, S. G., Salminen, J., & Jansen, B. J. (2018). Customer segmentation using online platforms: isolating behavioral and demographic segments for persona creation via aggregated user data. Social Network Analysis and Mining, 8(1), 54.

[44] Weinstein, A. T. (1994). Market Segmentation: Using Demographics, Psychographics and Other Niche Marketing Techniques to Predict Customer Behavior. Probus Publishing Co.

[45] Brito, P. Q., Soares, C., Almeida, S., Monte, A., & Byvoet, M. (2015). Customer segmentation in a large database of an online customized fashion business. Robotics and Computer-Integrated Manufacturing, 36, 93-100.

AUTHORS

Fahed Yoseph Business Intelligence, Data Mining and Big Data Consultant, Bachelor in Management Information System. Master in computer science in the area of artificial intelligence. Current doctoral student at Åbo Akademi University in Finland. Research interest in Computer Science, database and artificial intelligence.

Dr.Nurul Hashimah Ahamed Hassain Malim (a.k.a Nurul Malim) received her PhD from The University of Sheffield, United Kingdom. She has been serving USM since 2011 where she teaches programming courses on the specialization of Parallel and Distributed Computing and Bioinformatics. Her research focus encompasses algorithms, technologies and infrastructures that support different scales of data as well as parallelization of algorithms on GPU. Her interests include data mining / analytics and sentiment analysis in different datasets such as chemical databases, online social network, Point-of-Sales, Stocks Exchange, News and Online Review databases.

Mohammad AlMalaily Financial, Marketing and BI Consultant, Bachelor in Electronic Engineering. University in Finland. Research interest in Computer Science, Marketing and Finance.