A product-centric data mining algorithm for targeted promotions

Raymond Moodley[1,3], Francisco Chiclana[2,3], Fabio Caraffini[3] and Jenny Carter[4]

September 25, 2019

Abstract

Targeted promotions in retail are becoming increasingly popular, particularly in UK grocery retail sector where competition is stiff and consumers remain price sensitive. Given this, a targeted promotion algorithm is proposed to enhance the effectiveness of promotions by retailers. The algorithm leverages a mathematical model for optimising items to target items and fuzzy c-means clustering for finding the best customers to target. Tests using simulations with real life consumer scanner panel data from the UK grocery retailer sector show that the algorithm performs well in finding the best items and customers to target whilst eliminating "false positives" (targeting customers who do not buy a product) and reducing "false negatives" (not targeting customers who could buy). The algorithm also shows better performance when compared to a similar published framework, particularly in handling "false positives" and "false negatives". The paper concludes by discussing managerial and research implications, and highlights applications of the model to other fields.

Keywords: Association rule mining, targeted marketing, clustering

1 Introduction

The European grocery retail sector has been undergoing a major shift in recent years with most large, traditional supermarkets (TS) scrambling to retain or grow market share in the face of stiff competition from hard discounters (HD), particularly Aldi and Lidl [21]. In the UK, the "Big Four" TS brands (Tesco, Sainsbury's, ASDA and Morrisons) have all acknowledged pricing pressure from the HDs as a major contributor to softer revenue growth, profitability and market share losses and have consequently stepped-up efforts in both price and

[1]Corresponding author: raymond.moodley@my365.dmu.ac.uk
[2]Dept. of Comp. Science and AI, University of Granada, Spain
[3]Institute of AI, De Montfort University, Gateway House, Leicester, LE1 9BH
[4]Dept. of Comp. Science, University of Huddersfield, Queensgate, Huddersfield, HD1 3DH

non-pricing competitive pursuits [27][26]. The HDs responded by perfecting their pricing strategies and carefully expanding their product range. This has intensified competition even further and as a result, there is now a greater emphasis on customer development and customer data analytics across the sector as grocery retailers jostle for gains [59][19][39]. Indeed data collection by supermarkets has moved beyond mining store-based loyalty cards and now includes mining bank transaction data and social media [31][26].

There have been several recent studies on the retail sector from a marketing and retail research perspective with the aim of helping retailers navigate the changing market environment and remain relevant. These include understanding consumer choice, the impact of promotions and the drivers for store switching by consumers [25][21][15][58]. Concurrently, research on the use of artificial intelligence (AI) in retail remains active, notably on the impact of using recommender systems (RS), predicting consumer behaviour and enhancing the algorithmic efficiency of AI models [55][5][46]. Reutterer *et al.* in [50] proposed a data mining framework for targeted promotions that first clusters customers within a store (based on their purchase history) and then determines the best target items to promote within each cluster. The methods employed in [50] leveraged concepts from several previous studies notably: customer segmentation as detailed in [7], [10] and [37] and association rule mining (for finding frequent itemsets - the same approach used in this study) which was first introduced by [3] and well-documented in [60], [72] and [68]. Whilst there is merit in the approach detailed by Reutterer *et al.* in [50], for example finding lesser-known rules in customers who exhibit similar behaviour, it also has weaknesses, particularly in eliminating / reducing "false positives" (targeting customers who do not buy a product) and "false negatives" (not targeting customers who could buy). "False positives" and "false negatives" can be troublesome for retailers as it reduces their ability to maximise sales whilst causing anger and distrust amongst customers [42][69][53]. By clustering the customers first and applying the same treatment (for example a voucher) for everyone within the cluster, the customer segmentation framework in [50], [7] and [10] targets customers in the cluster who may have not purchased the product (and probably never will) whilst not offering the treatment to customers in other clusters who purchased the product (and probably will in the future either at this or other stores). Further the framework in [50] does not take into account household size, which is especially important as it applies a pruning of 25% of the lowest transactions and 5% of the highest transactions by size. It is likely that there could be several single-household customers in the lowest 25% who are extremely loyal but get eliminated due to low transaction sizes.

This study addresses some of the weakness cited above and contributes to the growing body of research in that it introduces a novel algorithm and mathematical model to help retailers enhance both its itemset and customer targeting for marketing promotions on complementary purchases and uplifts. It proposes a mathematical framework for itemset targeting which is conducted prior to customer targeting. Itemset targeting first leverages the Apriori association rule mining algorithm (introduced in [4]) to find frequent itemsets and then prunes these frequent itemsets, using the mathematical model developed as part of this study, to find the best itemsets to target for marketing promotions. Customer targeting leverages the popular Fuzzy C-Means (FCM) clustering algorithm as initially detailed in [8]. Added practical features of the proposed algorithm include customer pruning to eliminate customers who have not bought the target itemset thereby removing the possibility of "false positives" and factors in customer household sizes to ensure that loyal, small households are not erroneously pruned due to their relatively small transaction sizes thereby reducing "false negatives". FCM remains popular today and has been shown to be more accurate than K-Means clustering with a better replication of real-life scenarios [11][34]. This study also contributes to retail and marketing research in that it uses the UK grocery retail sector as a case study and leverages large cross-store consumer scanner panel data to validate the algorithm and simulate the impacts of varying take-ups in complementary purchases including store switching.

The remainder of the paper is divided as follows: An overview of the relevant literature is provided in Section 2 followed by a detailed description of the problem statement and development of the underlying mathematical model in Section 3. The research methodology is outlined in Section 4 with a presentation of the results and discussion in Section 5. Finally, conclusions and future work are detailed in Section 6.

2 Literature Review

2.1 Market Basket Analysis (MBA) in Customer Relationship Management (CRM)

The definition of customer knowledge in [35] was found to be comprehensive and consisting of three aspects: (1) knowledge for customers (product knowledge, store layout knowledge, etc.); (2) knowledge about customers (who they are, what they buy, when they buy, etc.); and (3) knowledge from customers (what they know about products, what they know about the competition, etc.). The role of MBA (analysing customer transactions) in the retail sector has been well-described in [44] as part of the customer development phase in CRM and whose purpose it is to consistently expand customer transaction intensity, transaction value and individual customer profitability. This is consistent with the definition of MBA provided in [56] where the underlying idea of MBA is that consumers rarely make purchases in isolation and by carefully studying their purchases, managers can develop interventions to influence their purchasing behaviour and enhance sales. Similar conclusions were drawn by [62] noting that the focus of MBA is on purchasing patterns and effective mining of such patterns can lead to predicting future behaviour. Given this, MBA may be considered an effective tool in CRM as it reveals regularities in the purchasing behaviour of customers.

On a more practical level the findings in [23], obtained through surveys with leading retailers showed that MBA is being used by most progressive retailers to understand customer habits and adjust operations to obtain maximum success. This fact was also noted more recently in [12] and [65]. Further, all three studies ([12],[23] and [65]) noted that providing retail operational teams with real-time MBA is allowing for rapid adjustments to be made to the business thereby capitalising on every opportunity and enabling the entire organisation to work smarter. In line with this, software/ business consulting companies and large retailers (e.g.Walmart who owns ASDA) have realised the commercial opportunities in MBA and consequently have developed software and tools to help extract the benefits from MBA [12][65][2].

MBA is synonymous with data analytics concepts of Frequent Itemset Mining (FIM) and Association Rule Mining (ARM) first introduced by [3] and later popularised in several key studies that proposed new data mining algorithms including [4], [9], [28] and [71]. Algorithms based on FIM and ARM analyse the buying habits of customers by finding associations between the items they place in "shopping baskets" and, using this knowledge of associations, retailers can then develop marketing strategies to enhance customer development. These algorithms remain popular today and formed the basis for many recent studies including [50], [57] and [62] and has dedicated chapters in popular data mining books including in [29], [72] and [68].

2.2 Targeted Promotions

According to Reutterer *et al.* [50], there has been an increase in research on the effectiveness of targeted promotions, particularly on its impact in increasing take-up compared to conventional promotions. In this regard the conclusions in [64] are particularly pertinent: "customized coupon campaigns are more effective if they provide more discounts, are unexpected, and are positioned as specially selected for and customized to consumer preferences". A similar conclusion was drawn in [18]: "Targeted promotions based on individual purchase histories are known to increase promotional response...". Whilst targeted promotions are impactful, research on RS found that "false positives" (offering promotions to customers that they would not buy) and "false negatives" (not offering promotions to customers that are likely to buy) should be avoided as much as possible [42][53]. In particular, it was noted that "false positives" can be damaging as it angers the customer and may result in a loss of trust by the customer for the retailer's ability to tailor offerings to the customer's individual preferences [42][53].

2.3 Identifying Target Itemsets

Association Rule Mining (ARM) proposed in [3] and [4], is one of the most popular data mining techniques for MBA today thanks to the substantial growth in research and practical applications that leverage the mathematical framework and various algorithms for finding associations between independent items in a single transaction [68][62]. Whilst considerable research has been conducted on ARM algorithms, the three most popular remain Apriori, ECLAT and FP-Growth, with Apriori still considered the benchmark and widely used [29][62][50]. A review of previous studies pertaining to these algorithms is provided in Section 2.3.1. In contrast with ARM algorithms, the mathematical underpinning of ARM has remained fairly consistent with the notions of support, confidence and the Apriori principle still used across most research and data mining algorithms in MBA [72][68][50].

The task of leveraging ARM for practical retail applications has been widespread - gaining interest from academia and retailers, who have seen the potential for gaining a competitive advantage. Consequently, a good proportion of use cases in ARM remains confidential and inaccessible as retailers conduct proprietary research and engage commercial software companies under non-disclosure agreements [39][12][62]. A key consideration in this regard is identifying the best items to target that fulfils retailers' objectives [12][50][18]. Frequent Itemset Mining (FIM), which consists of finding sets of items that are frequently bought together, is considered to be a subset of ARM and remains a typical starting point for frameworks that identify items to target and consequently up-sell and cross-sell [65][18][50]. The task of finding frequent itemsets is typically achieved by using an ARM algorithm, however, shortlisting the set of all frequent items to find the best itemsets to promote can be tricky and is context-dependent [72][29][9]. This is particularly true in large databases like those typically found in grocery retail [72][29]. The shortlisting of frequent itemsets to find the best itemsets to promote cannot be achieved by FIM/ ARM algorithms alone, and this prompted a need to include further analytical elements to achieve this task [50][41]. In [41] the uninorm was shown to be effective in selecting the best rule for a given antecedent (for example: $A \rightarrow C$ or $A \rightarrow D$) to aid shortlisting of frequent itemsets. However the authors cited the more generic case of $(A \rightarrow C)$ or $(B \rightarrow D)$ as future work. Reutterer *et al.* [50] detailed an approach that achieves shortlisting, however this approach does have drawbacks with regards to the elimination/reduction of "false positives" and "false negatives".

2.3.1 ARM Algorithms

The Apriori algorithm introduced in [4] remains popular in ARM-based data mining as it is efficient and robust [60][50]. However, it has one major drawback in that it is time consuming due to its breadth-first computational approach [22][28][38]. Consequently, several other algorithms have been developed to enhance the efficiency of ARM, most notably the ECLAT and FP-Growth algorithms proposed by [70] and [30] respectively. Unlike the Apriori, which uses a breadth-first approach, the ECLAT algorithm performs a depth-first scan of the database to identify all frequent, 1-item sets and then uses this result and the Apriori principle to generate larger frequent item sets. However, the FP-Growth algorithm (which uses a tree structure) may be seen as a hybrid approach, with a breadth-first scan to establish nodes (e.g. frequent 1-item sets) followed by a depth-first scan to find all subsequent frequent itemsets (branches and leaves). Studies show that both the ECLAT and FP-Growth algorithms have increased computational speed when compared to the Apriori algorithm, but can be more memory intensive, particularly on large databases, as is the case in grocery retail [29][9]. This increased memory requirement is due to the need to scan the entire database into main memory before processing and this is further compounded in the ECLAT algorithm due to the creation of potential larger, frequent itemsets that may not exist in the database [9][29].

Attempts have been made to address the issue of increased memory requirements of both algorithms as well as simplifying the sophisticated data structure of the FP-Growth algorithm, however these were marginally successful for large, sparse datasets as is the case in grocery retail. The "Split and Merge (SaM)" algorithm, detailed in [9], adopted a simplified structure and has lower memory requirements but was not suited to sparse datasets. Similarly, the dECLAT algorithm proposed in [71] attempted to address the memory requirements of ECLAT by considering the absence of an itemset as opposed to its presence, but it too was found to be unsuitable for large, sparse datasets [29]. Given this, it is clear that the ARM algorithms each have their pros and cons and whilst the Apriori is slow compared to the others, it is relatively simple to implement and has lower memory requirements [9]. As a result, the Apriori algorithm was selected for this study, as was the case in [50]. For completeness, it should be noted that irrespective of the algorithm used, the quality of the frequent itemset mining output is equally high [29].

2.4 Fuzzy C-Means Clustering

Clustering is considered to be a fundamental process in data mining with the most popular and widely used clustering methods today being KM and FCM [68][24] [61][43]. FCM, first introduced in [8] and based on the work in [13] was created to overcome some of the problems commonly associated with the crisp clustering approach of KM, including optimal cluster selection and the need for multiple passes to improve accuracy [11][24][34]. Unlike KM, FCM uses a soft clustering approach in which data points on the boundaries are not forced into a single cluster but rather they are allowed to be members of multiple clusters with varying degrees of membership such that the total membership of a data point across all clusters equals one. This approach not only improves clustering accuracy but also closely resembles everyday life [11][43]. There are several other fuzzy clustering algorithms that exist, but FCM remains popular as its relatively stable, reliable and fast [43].

However there are three main, well-known problems with FCM: (1) CPU usage as a result of speed; (2) too many iterations as a result of sub-optimally selecting the fuzzifier "m"; and (3) high dimensionality/ too many initial clusters [61][43][67] . The speed benefits of FCM was noted to be computationally expensive, in particular for large data sets, and there have been several variations of FCM to improve on this over the years [43]. The generally used value of 2 for the fuzzifier "m" is not optimal for all applications and a sub-optimal "m" can be time consuming due to the increased number of iterations required to reach convergence. However a large number of applications use "m" = 2 and this will be used in this study as well [11][61]. Whilst FCM is not without its problems, its accuracy is superior to KM and hence formed the basis for clustering in this study [11].

2.5 Identifying Target Customers

Grouping customers for targeted marketing is very common, done by most major retailers and usually employs some form of clustering [39][6][59]. Clustering using KM was used to create target customer groups in [50]; however the authors did recognise the common challenges with this method including selecting the optimal number of clusters and the need to perform multiple passes for improved accuracy. These problems are not new and were some of the key considerations in the formulation of FCM.

The operational methods for targeting customers is beyond the scope of the present study, however it is worth noting that recent studies have demonstrated that the psychological targeting of customers has proved effective [40]. Whilst the typical method for targeted promotion has been "price" through some form of coupon redemption, it is possible (and may become increasingly used in the future) to target customers with products using their psychological preferences as the key criteria [50][40].

3 Problem Statement and Mathematical Model

We commence by detailing several definitions which will be used throughout this study. We then define the problem and develop the mathematical model that will be used as part of the problem resolution. The mathematical model is based on two parts namely: identifying target itemsets and identifying target customers. These are detailed separately.

3.1 Definitions

The following definitions are used throughout this study:

Items: Leveraging the definition in [3], let $I = \{I_1, I_2, \ldots, I_m\}$, be a set of all items with the assumption that quality and quantity of items $(I = 1, \ldots, m)$ remain constant across all stores and customers do not stockpile.

Customers: Denoted by U and represents a household with size $f; f > 0$ with U purchasing subsets of I, known as baskets.

Transactions: All purchases are in the form of transactions (also called baskets) and contain a subset of I, for example $T_S = I_2,\ I_9,\ I_{11},\ \ldots,\ I_x$ is a single transaction from store S. Itemsets are defined as subsets of items in a transaction. Consumers make one transaction per time period W per store and this study uses $W = 1$ week. This assumption takes into consideration the practical aspects of shopping where the generally accepted length of a shopping period is one week as it is in line with how most people plan their household activity [33][14][51].

3.1.1 Support and Confidence

The standard definitions of support and confidence as outlined in [3] are used in Equations (1) and (2).

$$\text{support of item, } \mathrm{I_i}|_S = \mathrm{supp}(I_i)|_S = \left.\frac{\text{Number of transactions containing } \mathrm{I_i}}{\text{Total number of transactions}}\right|_S \tag{1}$$

$$\begin{aligned}\text{confidence of item } \mathrm{I_i} \text{ leading to items } \mathrm{I_iI_j}|_S &= \mathrm{conf}(I_i \to I_iI_j) \\ &= \left.\frac{\text{Number of transactions containing } \mathrm{I_i} \text{ and } \mathrm{I_j}}{\text{Number of transactions containing } \mathrm{I_i}}\right|_S \end{aligned} \tag{2}$$

Two additional user-defined concepts are defined as follows: Minimum support (minsup) is defined as the minimum support required for an itemset to be frequent whilst minimum confidence (minconf) is defined as the minimum confidence required for two or more items to be associated.

3.1.2 Apriori Principle

The Apriori principle, first detailed in [4], states that for a given set of transactions, $\mathrm{supp}(I_i) \geq \mathrm{supp}(I_i, I_j)$. This is consistent with probability theory where $P(A) \geq P(A \cap B)$ as well as in practical retail terms where the transactions that contain *bread* are always greater than or equal to transactions that contain both *bread* and *eggs*.

3.2 Problem Statement

It is well-documented that targeting the right customers with the right incentives results in increased transaction sizes thereby enhancing the primary objectives of stores, which are to increase revenue and profitability while maximising loyalty from its customer base [63][39][44]. Targeting the right customers with the right incentives is not always easy and has formed the basis of several studies including this study [42][50][1].

Thus the problem being addressed by this study may be stated as follows: For a given store, S, and all its customers U, there may exist several itemsets J in all transactions, T that are not frequent in S but where all subsets of J are frequent in S. Further given that store management are intent on maximising the primary objectives of S, while minimising the associated costs (largely marketing costs), it becomes necessary to optimise the targeting of J and U. Thus the aim of this study is to provide a framework, based on FIM and FCM, for targeting the right customers with the right itemsets thereby providing a useful tool that supports decision making.

3.3 Identifying Target Itemsets

Consider an itemset $J_1 = I_{11}, I_{12}, I_{13}, ..., I_{1x}$ that is not frequent in store S over a time period t but its subset $(J_1 - I_{1x})$ is frequent over the same time period with $\text{conf}((J_1 - I_{1x}) \rightarrow J_1) \geq \text{minconf}$. There may be several such itemsets that fit this criteria and given that most marketing departments have limited budgets the question of which J should be targeted often arises. Intuitively the best target should be that which has the largest $\text{supp}(J)$ as well as the largest $\text{conf}(J - I_x \rightarrow J)$ in S. However, cases do exists where $\text{supp}(J_c) > \text{supp}(J_k)$ but $\text{conf}(J_c - I_{cx} \rightarrow J_c) < \text{conf}(J_k - I_{kx} \rightarrow J_k)$ in which case, the choice between J_k and J_c is not obvious. This problem is not unique to retail and often requires further analysis before a decision can be made. A generalised model is proposed to speed up the decision making process.

3.3.1 Comparing (A $\rightarrow$ C) and (B $\rightarrow$ D) combinations

Let items A, B, C and D be frequent items (itemsets) in store S. Hence $P(A), P(B), P(C), P(D) \geq \text{minsup}$. Further, let $P(A, C), P(B, D) < \text{minsup}$. (A, B) and (C, D) combinations are ignored for now. Four confidence equations for the above frequent items are created:

$$\text{Confidence of (A} \rightarrow \text{C) in S} = \text{conf}(A \rightarrow \text{C}) = \frac{P(A, C)}{P(A)} \tag{3}$$

$$\text{Confidence of (C} \rightarrow \text{A) in S} = \text{conf}(C \rightarrow \text{A}) = \frac{P(A, C)}{P(C)} \tag{4}$$

$$\text{Confidence of (B} \rightarrow \text{D) in S} = \text{conf}(B \rightarrow \text{D}) = \frac{P(B, D)}{P(B)} \tag{5}$$

$$\text{Confidence of (D} \rightarrow \text{B) in S} = \text{conf}(D \rightarrow \text{B}) = \frac{P(B, D)}{P(D)} \tag{6}$$

The following Lemma is detailed which will be leveraged during this study:

Lemma 1: *If an itemset (A, C) exists but is not frequent in store S whilst both A and C are frequent with $P(A) = P(C)$ then both A and C are equally attractive products for marketing to target to make (A, C) frequent.*

Proof:

$$\mathrm{conf}(A \to C) = \frac{P(A, C)}{P(A)} \tag{7}$$

$$\mathrm{conf}(C \to A) = \frac{P(A, C)}{P(C)} \tag{8}$$

Re-arranging the above:

$$\frac{P(A)}{P(C)} = \frac{\mathrm{conf}(C \to A)}{\mathrm{conf}(A \to C)} = 1 \tag{9}$$

Hence if both the support and confidence are equal, targets become equally attractive for marketing. ■

The extension of Lemma 1 may be stated as follows: If $P(A) > P(C)$ then $\mathrm{conf}(C \to A) > \mathrm{conf}(A \to C)$, thus it is more attractive to target customers that buy item C with offers for item A as its is a smaller customer base with a higher probability of take-up (higher confidence).

One objective of marketing is to minimise the amount spent on marketing whilst increasing the sales of a product to make a product combination frequent. This can be achieved by minimising the number of transactions targeted (market target), which is expressed mathematically as follows (see Appendix A for a detailed development of equation (10)):

$$\text{market target} \cdot \mathrm{conf}(A \to C) = (\text{minsup} - P(A, C)) \cdot \text{total transactions} \tag{10}$$

Rearranging equation (10) to form equation (11), it is noted that the marketing target is minimised when $P(A, C)$ and/or $P(A)$ is close to minsup, because we are assuming that$P(A) \geq$ minsup whilst $P(A, C) <$ minsup.

$$\text{market target} = P(A) \cdot \left(\frac{\text{minsup}}{P(A, C)} - 1 \right) \cdot \text{total transactions} \tag{11}$$

Equation (11) may be very useful in deciding on combinations within a store with same minimum support, but could this be extended to compare combinations across stores or across minimum support thresholds?
The market target described in equation (11) is an absolute value, however this could be normalised (where mt is the normalised marketing target) by expressing it as a fraction of the absolute value for minimum support (min support).

$$\frac{\text{market target}}{\text{min support}} = P(A) \cdot \left(\frac{\text{minsup}}{P(A, C)} - 1 \right) \cdot \frac{\text{total transactions}}{\text{min support}} \tag{12}$$

Equation (12) may be rearranged as follows (defining $\text{mt} = \frac{\text{market target}}{\text{min support}}$ and noting that $\text{minsup} = \frac{\text{min support}}{\text{total transactions}}$):

$$\text{mt} = \frac{P(A)}{P(A,C)} - \frac{P(A)}{\text{minsup}} \tag{13}$$

Note that $P(A,C) < \text{minsup}$, hence $\text{mt} > 0$. By normalising the marketing target, it then becomes possible to compare unrelated combinations. For example: is it better to target selling *cheese* to people that buy *butter* where $\text{minsup} = 0.2$ or target selling *nappies* to people that buy *beer* where $\text{minsup} = 0.3$? The answer to this question will be that combination which minimises mt in equation (13). The absolute value of the marketing target however is given by equation (11).

The value of minsup may be adjusted to any value such that $P(A) \geq \text{minsup}$ as minsup sets the threshold for frequency and item A was assumed to be a frequent item as noted earlier. Further, $P(A,C)$ was assumed to be infrequent, hence $P(A,C)$ is initially lower than minsup and $P(A)$. For some applications it becomes necessary to target a market such that $P(A,C)$ must become equal to $P(A)$, hence minsup is set to $P(A)$ while $P(A,C)$ is initially lower than both minsup and $P(A)$. Note that in these applications $P(A,C)$ increases as the treatment of the target market progresses while the market target as described in equation (13) decreases. Eventually $P(A,C) = P(A)$ at which point treatment stops as $\text{mt} = 0$. For example in public health, authorities may want all people with compromised immune systems, A, to be vaccinated with some drug, C, and eventually after treatment, all people, A, would be vaccinated by C, hence $P(A,C) = P(A)$. Similarly in public safety, authorities may want all dangerous criminals, A, to be tracked using a wearable tracking device, C, leading to $P(A,C) = P(A)$.

3.4 Identifying target customers

The well-known RFM (Recency, Frequency and Monetary) framework for customer targeting in a retail setting was used as the starting point [16]. "Recency" is an elimination variable with the assumption that customers must have made at least two purchases during the period with at least one containing the target item else they are eliminated. Indeed, it is difficult to target customers who have not bought the target item as this introduces several variables that are difficult to measure (including taste, allergy/ intolerance and cultural aversion) and may result in increased "false positives" as detailed in [42] and [53]. "Frequency" is item specific and is based on the support, $\text{supp}(C)$, that the user has for the target item C. "Monetary" is defined as the average user transaction size for store S divided by household size, $\overline{|T_{U_S}|}/f_U$, where $\overline{|T_{U_S}|}$ is the average transaction size for user U in store S and f_U is the family size.

The inter-play between the RFM dimensions and how this translates into customer clusters for targeted promotions (as shown in Table 1) is better illustrated through an example. Consider the scenario of a grocery store attempting to make *(jam, peanut butter)* a frequent itemset by promoting *(peanut butter)* to customers that already buy *(jam)*. In this case, it is assumed that both *(peanut butter)* and *(jam)* are frequent but the combination is not frequent. There are several "RFM" combinations for customers in this scenario, including customers who buy *(jam)* but are severely allergic to peanuts and hence will never buy *(peanut butter)*. Such customers will be eliminated early-on in the "recency" stage as they will show zero purchases for *(peanut butter)*, irrespective of their "monetary" spend or loyalty to the store. Any attempt to target these customers will result in a "false positive" and all of its consequences. At the other end of the spectrum, there may be very loyal customers that buy large quantities of *(jam)* and *(peanut butter)*. These customers will be classified as: loyal customers ("high monetary") with a high consumption of *(peanut butter)*, hence "high frequency". Note that there are also several combinations that exist between these two endpoints including "high frequency, low monetary" etcetera.

In line with the above, customers can be clustered into one of nine "buckets" as shown in Table 1 based on "Frequency" and "Monetary" which are considered to be two linguistic variables of granularity three ("Low", "Medium", "High"). "Low","Medium" and "High" are assigned to both "Frequency" and "Monetary" based on the FCM clustering approach which generates three clusters for each linguistic variable. Hence, nine clusters are created in total. These assignments are dynamic and will change based on the itemset targeted as each targeted itemset may have a different support and most likely would attract a different customer base. Customers within each cluster are considered similar to each other for the targeted itemset. Similarly, clusters themselves may have the same trend pattern to other clusters (e.g. Clusters 2 and 6 in Table 1), hence can be treated or approached in a similar manner. We define cluster approaches as follows: (1) "Switchers" if the cluster frequency is less than its monetary spend, except where both frequency and monetary spend are "Low", as is the case in Cluster 1 in Table 1; (2) "Loyal" when frequency and monetary spend are similar and at least "Medium"; and (3) "Drop Out" if clusters have a "High" or "Medium" frequency with a "Low" monetary spend. It should be noted that the use of nine clusters falls within the typical range for market segmentation. Early studies including in [54] used five clusters as the typical value, while more recently nine clusters were used in [52] and eleven in [50].

Note that Cluster 6 is considered to be "Switcher" because its frequency is less than its monetary spend implying that customers who belong to this cluster choose to purchase the target item elsewhere even though they are essentially loyal to the store for other purchases. Consequently the right incentives may enable customers in Cluster 6 to switch purchases of the target item away from other stores to this store.

Cluster	Frequency	Monetary	Approach
1	Low	Low	Switcher
2	Low	Medium	Switcher
3	Low	High	Switcher
4	Medium	Low	Drop Out
5	Medium	Medium	Loyal
6	Medium	High	Switcher
7	High	Low	Drop Out
8	High	Medium	Loyal
9	High	High	Loyal

Table 1: Target Clusters

3.4.1 Treatment Approaches

Uplift theory, first discussed in [49], can be used to identify the "treatment" for each cluster with the treatment varying from "Switcher" to "Avoid Drop-Out" and "Leave Alone / Light Touch" for "Loyal".

"Switcher"
Shoppers in this category either conduct a large proportion of their shopping at other stores and/or have a low take-up of the target item despite having high transaction volumes at the chosen store. An aggressive marketing campaign may be the suggested treatment to "break the habit" and force the shopper to switch stores [45]. From Table 1 clusters that are identified as "Switcher" fall into this category.

"Avoid Drop Out"
Shoppers in this category have a high affinity for the target product from the chosen store but conduct a large proportion of their shopping at others stores. Given this, they can be easily enticed to try an alternative from another store and "drop out". A gentle marketing campaign may be enough to ward off any pricing comparisons. Clusters identified in Table 1 that belong to this treatment approach are in the "Drop Out" clusters.

"Leave Alone" / "Light Touch"
Shoppers that are highly loyal to both the store and target item fall into this category. These shoppers may respond to marketing initiatives to increase their take-up but are usually "set in their ways" and could stockpile to save money which will lead to downstream decline in purchases or ignore the initiatives completely [45]. In any event promotions to this group may drive down prices unnecessarily, consequently a "Light Touch" approach is recommended for clusters 5 and 8 where uplifts can be possible due to the medium valuation in one or both variables "Frequency" and "Monetary". Shoppers in this category are identified in Table 1 as "Loyal".

3.4.2 Creating Clusters for Treatment

Nine clusters were created as shown in Table 1 using a two-step fuzzy clustering process to enable targeted treatment. The fuzzy clustering process used the Fuzzy C-means (FCM) iterative algorithm as initially detailed in [8] and is as follows:

Algorithm 1: Generic FCM Algorithm

---

1 Randomly generate the membership matrix μ using the constraint $\sum_{j=1}^{c} \mu_j(x_i) = 1$ where c is the number of clusters, x_i is the data point and μ_j is the fractional membership of x_i in cluster, j

2 Calculate the cluster centres, $c_j = \frac{\sum_{i=1}^{u} \mu_{ij}^m \cdot x_i}{\sum_{i=1}^{u} \mu_{ij}^m}$ where u is the total number of items and m is the "fuzzification" parameter with a typical value of 2

3 Compute the objective function, $F = \sum_{i=1}^{u} \sum_{j=1}^{3} \mu_{ij}^m \cdot d(x_i, c_j)$ where $d(x_i, c_j)$ is the Euclidean distance between x_i and the centre of the cluster, c_j

4 Recalculate μ using $\mu_j(x_i) = \frac{[\frac{1}{d_{ji}}]^{\frac{1}{(m-1)}}}{\sum_{k=1}^{j} [\frac{1}{d_{ki}}]^{\frac{1}{(m-1)}}}$ where d_{ji} is the Euclidean distance of each data point x_i to each cluster, j

5 Repeat steps 2 to 4 until F is minimised.

The generic algorithm presented above was used to create the two-step FCM clustering process that resulted in the nine clusters.

Step 1: For each target item A in store S

- Cluster $\frac{\overline{|T_U|}}{f_U}$ into 3 clusters by minimising the objective function F_A. Note that $\frac{\overline{|T_U|}}{f_U} \in \mathbb{R}$ which is well-suited to FCM.

- The objective function $F_A = \sum_{i=1}^{u} \sum_{j=1}^{3} \mu_{ij}^m d(\frac{\overline{|T_{U_i}|}}{f_{U_i}}, c_j)$ where μ_{ij} is the degree of membership of $\frac{\overline{|T_{U_i}|}}{f_{U_i}}$ and the cluster, c_j, with $1 < m < \infty$ and $d(\frac{\overline{|T_{U_i}|}}{f_{U_i}}, c_j)$ is the Euclidean distance between the object, $\frac{\overline{|T_{U_i}|}}{f_{U_i}}$ and the centre of the cluster, $c_j = \frac{\sum_{i=1}^{u} \mu_{ij}^m \frac{\overline{|T_{U_i}|}}{f_{U_i}}}{\sum_{i=1}^{u} \mu_{ij}^m}$ and u is the total number of users being clustered

- The typical value for $m = 2$ was used as noted in [8], [11] and [61]

Step 2: For each target item A in store S and Cluster, c_j created in Step 1

- Cluster the elements into 3 further clusters, based on the value of supp(A), by minimising the objective function E_A

- The objective function $E_A = \sum_{i=1}^{n} \sum_{j=1}^{3} \mu_{ij}^{m} d(\text{supp}(A), c_j)$ where μ_{ij} is the degree of membership of supp(A), and the cluster, c_j; n is the cardinality of each $\frac{\overline{|T_{U_i}|}}{f_{u_i}}$ cluster , with $1 < m < \infty$ and $d(\text{supp}(A), c_j)$ is the Euclidean distance between the object, supp(A) and the centre of the cluster, $c_j = \frac{\sum_{i=1}^{n} \mu_{ij}^{m}.\text{supp}(A)}{\sum_{i=1}^{n} \mu_{ij}^{m}}$

- The typical value for $m = 2$ was used as noted in [8], [11] and [61]

3.5 Simulating marketing initiatives

Markov chains, as discussed in [48], were used to simulate the time series impacts of marketing. Given a set of transactions at the start of time period W_t where A and C have been purchased frequently in store S then the four possible "treatment" scenarios that exist for the purchasing of item C by customers that purchase A to make (A, C) frequent (as outlined in Table 1) are: Buying C frequently with large basket sizes ("Leave Alone"), Buying C frequently but basket size is medium ("Light Touch"), Buying C frequently but basket size is small ("Avoid Drop Out") and Buying C at rate lower than the basket ("Switcher").

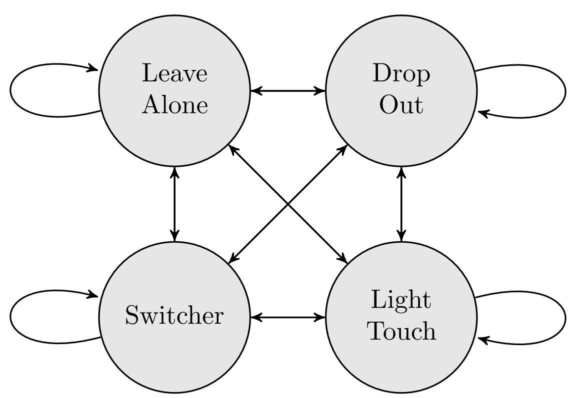

Figure 1: Markov Model for Simulations

This process was modelled using a Markov chain (see Equation (14)) where P_{W_t} is the 1×4 proportion vector of users buying A in one of the four states with regards to C. The process is ergodic (as shown in Figure 1) where users can move from one state to another based on their choice. Hence the proportion vector at time W_{t+1} given by $P_{W_{t+1}}$ could be estimated using the Markov equation described in Equation (14), where M_{ACS} is the 4×4 Markov transition matrix that describes the probability of a user buying A in store S and moving from one state to another with regards to C over a time period. The values of entries in M_{ACS} are "user-defined" and can be adjusted to simulate a variety of scenarios including conservative or aggressive marketing campaigns

$$P_{W_{t+1}} = P_{W_t} \cdot M_{ACS} \tag{14}$$

.

3.6 Bringing it all together - Algorithm for enhancing the purchasing of itemset (A, C) amongst frequent shoppers U in store S

The steps of the proposed algorithm are detailed below:

Algorithm 2: Enhancing the purchasing of itemset (A, C) amongst frequent shoppers U in store S

1. Create a set L, using the Apriori algorithm with very low support, containing all itemsets (A, C) that are not frequent in S but where its subsets A, C are frequent in S
2. Create a shortlist of L using Equation (13) and the extension of Lemma 1, selecting those combinations that minimise the value of the market target
3. For each C in the shortlist, create clusters as outlined in Section 3.4.2
4. Order the clusters based on Table 1
5. Run simulations as outlined in Section 3.5 re-prioritising the list based on the shortest time to frequency

4 Experiments

4.1 Experimental Process

Experiments were conducted based on the well-known Knowledge Discovery in Databases (KDD) process first outlined in [17] and more recently in [60]. Tan (2018) noted that the heart of the KDD process is the data mining phase which leverages models and algorithms to process data into information [60]. In this regard the Apriori and FCM algorithms form part of the data mining phase while the proposed market target model and the simulation forms part of the post processing phase, where patterns are interpreted to select the best information that contributes to overall knowledge [60].

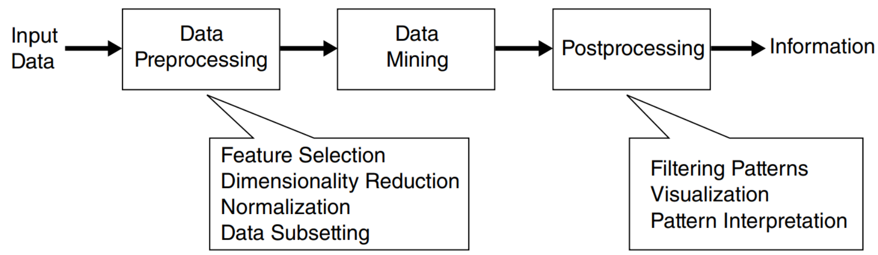

Figure 2: KDD Process as outlined in [60]

From a process implementation perspective, experiments were conducted to validate the proposed algorithm outlined in Section 3. Experiments were divided into four parts: (a) Identifying target itemsets (using Apriori, market target model and FCM); (b) Identifying target customers (using FCM); (c) Simulating the impacts of the proposed algorithm, and (d) Comparing the proposed model against the model detailed in [50] using the same dataset.

4.2 Experimental conditions

The 2012 scanner panel dataset obtained from [33] was used as the basis for the experiments. The data set contained 32,726 unique users and over 51 million individual scanned items across 21 stores in the UK. All store formats of the same store were also combined (e.g. internet, express, garage shop etcetera). Three stores were chosen, a TS store (Store 9), a HD store (Store 13) and a high-end grocery retailer (Store 21). Frequent itemsets were mined using the Apriori algorithm as detailed in [4] for all itemsets with support ≥ 0.02 and confidence ≥ 0.1. Computer programs were written using R software to mine and analyse the data whilst Microsoft Excel was used to compute the mt values and perform simulations. Note that this study does not perform inter-store comparisons, hence the size of each selected stores is not important. Indeed intra-store comparisons are made and whilst it is likely that some goods are purchased more frequently in-store or online than others, it is assumed that the pricing policy is the same across all formats and that customers have full choice in selecting a format that best suits them. These assumptions are considered fair given the prominence of multi-channel shopping, the emphasis that stores place on consistency across all channels, and the growing adoption of the "customer is king" mentality of UK shoppers [14][15][59]. For completeness, the model was also tested on large, dense datasets to evaluate its performance. These datasets were created using a synthetic transaction database generator as detailed in [32] with the largest dataset consisting of 5 million transactions, 100 unique items, 5 frequent itemsets, frequency density of 0.5 and a maximum basket size of 50 items. It should be noted that this dataset is similar in transaction volume (but considerably denser) to that of the UK's largest grocery retailer's daily activity, and represents approximately 28% of the UK's grocery retail market share [47].

4.3 Identifying target itemsets

The data for each store, S, was analysed and target itemsets were identified based on the criteria that there exists two items A and C which are frequent, with minsup = 0.1, but their combination (A, C) is not frequent. Further (A, C) is the optimal combination to target (based on the marketing target obtained using Equation (13)) from all identified targets. As discussed in [41][4][29] models that target itemsets must obey the Apriori and probability principles. These principles are detailed as follows:

P1: $\mathrm{mt}|_{(A,C)} > \mathrm{mt}|_{(B,D)} \quad \forall (A,C),(B,D);\ \mathrm{supp}(A,C) > \mathrm{supp}(B,D)$ and $\mathrm{conf}(A,C) > \mathrm{conf}(B,D)$

P2: $\mathrm{mt}|_{(A,C)} > \mathrm{mt}|_{(A,D)} \quad \forall (A,C),(A,D);\ \mathrm{supp}(A,C) > \mathrm{supp}(A,D)$

P3: $\mathrm{supp}(A) \geq \mathrm{supp}(A,C) \quad \forall A, C$

P4: If $\mathrm{supp}(A,C) > \mathrm{supp}(B,D)$ and $\mathrm{conf}(A,C) < \mathrm{conf}(B,D) \quad \forall (A,C),(B,D)$ then mt (as outlined in Equation (13)) determines priority

4.4 Identifying target customers

Target customers for each targeted (A, C) in each store, S, were identified based on the customer's purchase history and household size using the FCM algorithm detailed in Section 3.4.2. Customer clusters were then classified based on the criteria outlined in Table 1. To eliminate "false positives", customers had to have visited the store at least twice in the year and purchased the antecedent, A, at least once.

4.5 Simulating the impacts of the proposed model

An ergodic Markov model was created using Microsoft Excel to simulate the impact of marketing interventions on the shopping behaviour of the identified target customers. The model was based on the concepts outlined in Section 3.5. Two marketing campaigns ("Conservative" and "Aggressive") were used, with their corresponding proportion vectors given in Tables 2 and 3, respectively. From Table 2, a "Conservative" campaign assumed that all customers that have high monetary and high frequency (e.g. Cluster 9 in Table 1) will remain loyal in the future and hence may be left alone, that is, no treatment is required. On the other hand, treatment of the "Switcher" clusters (e.g. by offering customers money-off vouchers) will result in 90% of the customers who are currently "Switchers" remaining as "Switchers" in the future with the other 10% progressing to a more loyal or frequent state (5% elevated to "Drop Out" and 5% to “Light Touch”). Both the “Drop Out” and “Light Touch” clusters were assumed to show similar trends with 90% of the customer base remaining unchanged, whilst 10% being elevated to more loyal states. The “Aggressive” campaign detailed in Table 3 followed a similar approach to the “Conservative” campaign except that a greater percentage of customers were elevated to more loyal or frequent states as result of larger incentives (e.g. larger money-off vouchers than those offered in the “Conservative” campaign).

In the main, it should be noted that the assumptions of both the “Conservative” and “Aggressive” campaign around increased customer purchases and hence elevations to more loyal or frequent states are broadly in line with current grocery retail practice. In [66] it was found that over 27% of the consumers surveyed would be convinced to buy more or shop in stores that they would not normally use if they were incentivised by a money-off voucher. A more recent study in the U.S. found that over 75% of the consumers surveyed would use paper-based money-off vouchers as a key deciding factor when choosing a store at which to shop [20].

		Future			
		Leave Alone	Light Touch	Drop Out	Switchers
Current	Leave Alone	1	0	0	0
	Light Touch	0.1	0.9	0	0
	Drop Out	0	0.1	0.9	0
	Switcher	0	0.05	0.05	0.9

Table 2: Conservative Marketing Campaign

		Future			
		Leave Alone	Light Touch	Drop Out	Switchers
Current	Leave Alone	1	0	0	0
	Light Touch	0.2	0.8	0	0
	Drop Out	0	0.2	0.8	0
	Switcher	0	0.1	0.1	0.8

Table 3: Aggressive Marketing Campaign

4.6 Model Performance

The performance of the proposed model was compared with the model detailed in [50]. The two models were compared using four tests: (1) Ability to target customers who will buy the product and eliminate “false positives”; (2) Ability to target most or all customers who want to buy the product and reduce “false negatives”; (3) Ability to offer customised treatment and avoid targeting loyal customers (as this drives down price); and (4) Ability to enhance the frequency of the target itemset (increase purchases by the target customers).

5 Results and Discussion

5.1 Identifying Target Items

The support, confidence and marketing target, given by Equation (13), was computed for all associated itemsets for Stores 9, 13 and 21. The total number of infrequent itemsets (where the support of the itemset is less than minsup) exceeded one hundred per store as is often the case in real life. Choosing the best itemset(s) to target for marketing purposes is a well-known managerial challenge, particularly with large datasets, and is well-documented in the literature [68][72][29]. The market target model, given by Equation (13), can be an effective tool in addressing this managerial challenge. Consider the selection of results presented in Table 4 which includes the support, confidence and market target (mt) for several infrequent itemsets (where minsup = 0.1). Clearly the best itemset to target is that itemset which has both the highest support and confidence as it is the most popular and requires the least "effort" (marketing investment: people, money, space etcetera) to increase sales [18][63]. From Table 4 the best itemset for Store 9 is (156,277) and (135,163) for Store 21. It can also be seen that these best itemsets have the lowest market target (mt) for the respective stores, hence they will require the smallest number of customer transactions to be targeted for these itemsets to become frequent. Similarly itemsets (107,270) in Store 21 and (68,88) in Store 9 are considered the worst itemsets for targeting (from the given set in Table 4) as they have the lowest support and confidence. The market target model also confirms this result, as these two itemsets have the highest mt values for their respective store in Table 4. The rows highlighted in yellow are of particular interest as the optimal choice cannot be easily made by merely inspecting both the support and confidence. Hence this choice is best determined by using Equation (13) as detailed in P4 in Section 4.3 - where the best choice has the lowest market target (mt). Consequently, (68,270) and (57 to 274) in Stores 9 and 21 respectively are the better choices from the rows highlighted in yellow as they have smaller market targets. These results and the market target model obey all well-known data mining principles including Apriori and the laws of probability as evidenced by comparing the results presented in Table 4 against P1, P3, and P4 outlined in Section 4.3 which are underpinned by all previous studies on MBA and FIM including [3],[9], [28], and [68].

Itemset Monotonicity

The market target (mt) was tested for monotonicity in line with the practical considerations outlined in [4] and the conclusions in Lemma 1 and Moodley *et al.*,[41]. The monotonicity property of comparing two itemsets with the same

Itemset	Supp(A,C)	Conf(A,C)	mt	Target Priority
156 to 277 in Store 9	0.072	0.696	0.402	Priority 1
68 to 270 in Store 9	0.063	0.593	0.623	Priority 2
156 to 268 in Store 9	0.062	0.600	0.632	Priority 3
68 to 88 in Store 9	0.058	0.541	0.784	Priority 4
135 to 163 in Store 21	0.074	0.677	0.384	Priority 1
57 to 274 in Store 21	0.056	0.593	0.825	Priority 2
107 to 268 in Store 21	0.058	0.508	0.830	Priority 3
107 to 270 in Store 21	0.051	0.444	1.115	Priority 4

Table 4: Target Itemsets

antecedent is detailed in P2 of Section 4.3. Frequent antecedents, A, were selected from the three stores (9, 13, 21) with the market target calculated for a variety of (A, C) combinations and presented in Figure 3. It can be seen that the market target monotonically decreases for increasing supp(A, C) which is consistent with Equation 13, Lemma 1, the practical considerations outlined in [4] and the conclusions in [41]. This implies that the higher the frequency of an infrequent itemset, the fewer the customer transactions required (for targetting) for such itemset to be frequent.

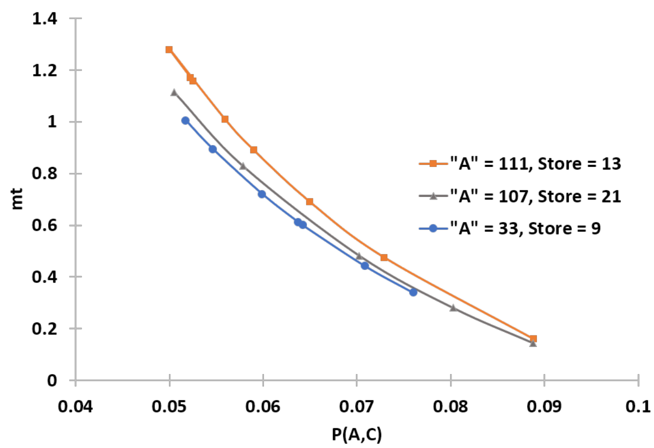

Figure 3: Monotonic Property of Proposed Model

Tests with Other Datasets

Tests were conducted on synthetic transaction data created using the data generator described in [32]. The output file from the frequent itemset mining algorithm (Step 1 in Section 3.6) for both a medium dataset (1 million transactions) and a large dataset (5 million transactions) were similar in format and complexity to the 2012 scanner panel dataset which had 2.6 million transactions. In all cases, the computational approach and the applicability of the mt model was the same, that is, the best itemsets to target where those itemsets that had the smallest mt value.

It should be noted that the large dataset was denser than the 2012 scanner panel dataset with the average number of items per transaction being 29 and 11 respectively. As a result, the large dataset did take longer to process with the runtime for the Apriori algorithm taking 748.8 seconds (approximately 12.5 minutes) to generate the list of frequent itemsets compared with 40.5 seconds for the 2012 scanner panel dataset, using a personal computer that had two 2.66 GHz Intel Xeon 5150 processors with 32GB of RAM. Similarly, the frequent itemset output file for the large dataset was larger than the 2012 scanner panel dataset which increased processing time, but yielded the same results, that is, the best itemsets to target were those that had the smallest mt value.

5.2 Identifying Target Customers

Target customers were identified using the principles outlined in Section 4.4 for six itemsets across three stores. The number of customers together with their "false positive" potential is presented in Table 5. Notably that the "false positive" potential for all customers is zero as only customers that have purchased the antecedent during the year have been selected, hence every target customer has the potential to purchase the itemset (A, C) if incentivized with offers for (C) or both (A, C). Thus the model satisfies the requirements of Test 1 in Section 4.6. "False positive" is a significant issue in retail and was cited as a key challenge in [42] and [53]. This approach ensures that this issue is eliminated altogether. Moreover, this approach reduces both data mining effort and unnecessary marketing spend as it initially prunes all customers with "NULL" values for the target item thereby saving computational effort in clustering and avoiding artificially inflated customer target groups who potentially could receive expensive marketing incentives that are not taken up.

Target Itemset	Target Customers	"False Positive" Potential
68 to 270 in Store 9	8117	0%
156 to 268 in Store 9	7824	0%
284 to 163 in Store 13	4801	0%
153 to 268 in Store 13	5845	0%
57 to 274 in Store 21	1289	0%
107 to 268 in Store 21	1239	0%

Table 5: Target Customers

Clustering Customers

The data in Table 5 were clustered using the FCM principles outlined in Sections 4.4 and 3.4.2 with the results of the clustering presented in Figure 4. Clearly most customers fall into the "Switcher" cluster because it comprised of four of the nine clusters in Table 1 and the selected antecedents are marginally frequent (having frequencies just over the targeted minsup of 0.1) implying that these item are not "top sellers" in these stores. In any event, if the antecedents were "top sellers", for example *bread* or *milk*, then it will be likely that the clustering mix will be swayed towards loyalty. Whilst it is not in the best interest of stores to promote "top sellers" as there is always strong demand and it can cannibalise revenue, it should be noted that this model can effectively separate customers under these conditions as well, thereby allowing for the execution of different treatment measures to different clusters or no treatment at all [21]. This is an important consideration as it meets the requirements of Test 3 in Section 4.6.

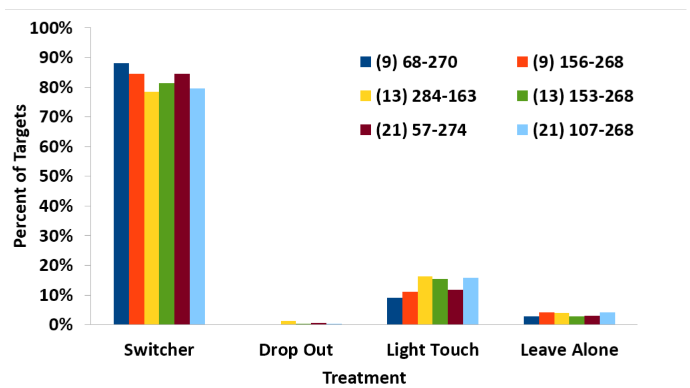

Figure 4: FCM clustering of target customers

In general, having a larger "Switcher" group and smaller loyalty-based groups ("Drop Out", "Light Touch", "Leave Alone") is ideal from a marketing perspective as it allows stores to spend most of its marketing spend on attracting customers who would otherwise spend their money elsewhere, thereby driving up the customer base and increasing revenue. This is consistent with marketing studies that prioritise marketing spend on attracting new customers. It is well-documented that spending marketing money on "Loyal" customers is not ideal as it not only lowers the price expectations with the most frequent / high value shoppers but it also erodes revenue as a large proportion of these customers would not necessarily switch to other stores in the absence of promotions. Given this, it can be concluded that the proposed market target (mt) model and FCM clustering is an effective approach to optimising the targeting of customers for marketing promotions.

5.3 Simulating the impacts of the proposed model

The Excel-based Markov simulator was run using the conditions outlined in Section 4.5 with the various cluster start sizes given in Table 6 and the simulation output shown in Figures 5, 6, 7 and 8, respectively.

Cluster	Frequency	Monetary	Approach	(9) 68-270	(13) 284-163
1	Low	Low	Switcher	20.22%	28.97%
2	Low	Medium	Switcher	46.57%	40.24%
3	Low	High	Switcher	11.67%	5.25%
4	Medium	Low	Drop Out	0.04%	1.21%
5	Medium	Medium	Light Touch	8.91%	14.60%
6	Medium	High	Switcher	9.61%	4.06%
7	High	Low	Drop Out	0.00%	0.00%
8	High	Medium	Light Touch	0.15%	1.60%
9	High	High	Loyal	2.85%	4.06%

Table 6: Target customer cluster sizes at start

The simulations show that it is possible to convert "Switchers" to loyal shoppers for the various itemsets with both conservative and aggressive campaigns using customised treatment for the different clusters. For conservative treatment, the "Switcher" category reduced to approximately 10% after 20 weeks of sustained treatment whilst the "Leave Alone" loyal category increased to over 50% for both stores 9 and 13 during the same period. Aggressive campaigns at both stores reduced the time taken to achieve the 10% "Switcher" and 50% "Leave Alone" levels by 8 weeks.

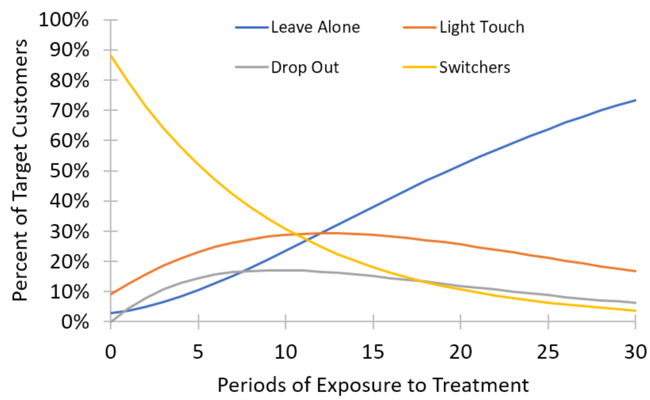

Figure 5: Conservative Campaign in Store 9

Note that there is a gradual shift from "Switchers" to "Leave Alone" loyalty across all marketing campaigns and this will remain the case provided that the

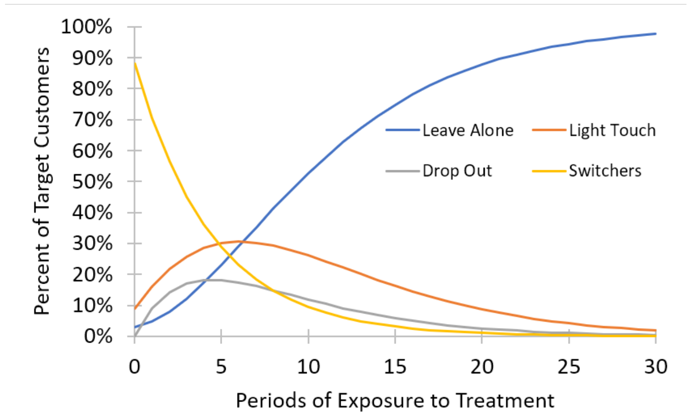

Figure 6: Aggressive Campaign in Store 9

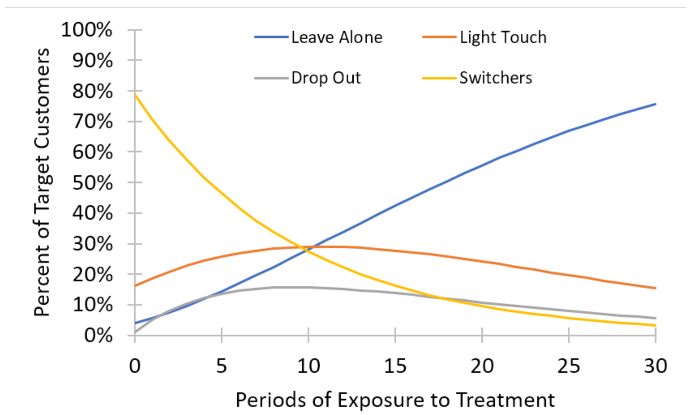

Figure 7: Conservative Campaign in Store 13

promotion impact and/or the utility derived from the switch remains in place. This is consistent with numerous studies on the implications of switching including in [36], [51] and [45]. On a contemporary and practical level, this switching behaviour by consumers is consistent with the rise of the German discounters in the UK and across Europe for that matter, where shoppers have switched to these discounters from traditional stores and remain loyal to these discounters as the price has renamed low while the quality has increased. To further retain customers, the German discounters are retaining customers by rapidly expanding their footprint across Europe and enhancing their stores with value-added features like in-store bakeries, multiple payment methods (thereby increasing localisation and shopper convenience while neutralising any differentiation that the traditional stores may have over them) [21][27].

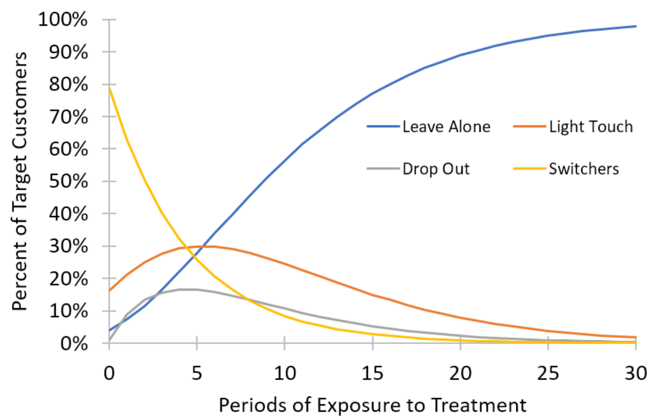

Figure 8: Aggressive Campaign in Store 13

5.4 Model Comparison

The model detailed in [50] was run on the same dataset, itemsets and stores with the analysis of the results given in Table 7. A total of 11 clusters were created and the most popular cluster for each itemset was selected for treatment. From Table 7 it can be seen that following this approach results in a much smaller customer base being selected (compared with cluster sizes in Table 5) which in itself is not a problem, provided that every customer is a highly likely candidate to act upon the treatment. However, only a fraction of the total target customers are known purchasers of the antecedent and in some cases this fraction is as low as 26%, resulting in a very large "false positive" potential which is considered bad for marketing as it angers customers [42][53]. Further, a large proportion of potential, good targets are left untreated as they fall in other clusters (as evidenced by the difference in targets between Table 5 and Table 7), which is a missed opportunity and an increase in "false negatives". Consequently the model proposed in this study performs better than the model outlined in [50] in terms of Test 1, minimising "false positives", and Test 2, minimising "false negatives".

The "Known Purchasers" detailed in Table 7 were classified based on the proposed customer clustering model. Whilst the overall pattern is similar to the results obtained for the proposed model, there is a slight bias towards loyalty as shown in Figure 9. This is expected as the model proposed in [50] prunes the lowest 25% and highest 5% of transactions (based on size), thereby favouring loyal customers who typically have higher transaction sizes. However there is no further segregation of customers and consequent treatment in the model proposed in [50] resulting in all customers (including loyal customers) receiving the same treatment which is likely to drive down prices. Being able to attract new customers, while retaining revenue spend from loyal customers is a chal-

Target Itemset	Target Customers	Known Purchasers	"False Positive" Potential
68 to 270 in Store 9	958	423	56%
156 to 268 in Store 9	200	90	55%
284 to 163 in Store 13	1066	378	65%
153 to 268 in Store 13	1395	551	61%
57 to 274 in Store 21	376	107	72%
107 to 268 in Store 21	411	108	74%

Table 7: Target Clusters

lenge for many retailers as noted in [12], [26], [39] and [21] consequently the model in [50] has limitations in this regard as it could lead to unnecessary price reductions. However the model proposed in this study performs better than the model proposed in [50] in terms of Test 3, ability to offer customised treatment and avoid targetting customers that are loyal.

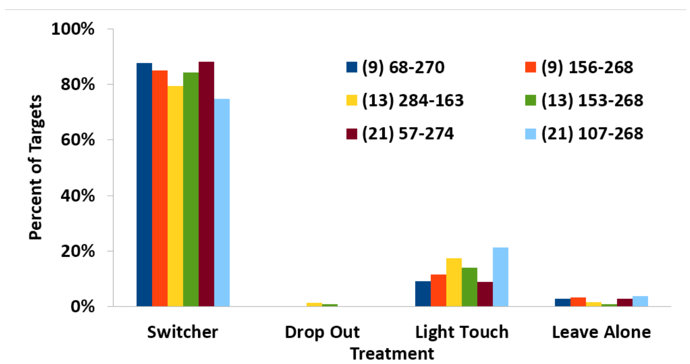

Figure 9: Clustering Known Purchasers

Comparing the models for Test 4, ability to enhance the frequency of the target itemset could not be done given that there is a large proportion of customers that are potential "false positive" in the results obtained from the model detailed in [50]. Given that both models demonstrate benefits in increased frequency of purchasing, it is noted that both models pass Test 4.

6 Conclusions

A mathematical model and algorithm was proposed to target and enhance the frequency of itemsets for a given store. The optimal target itemset minimises the marketing target as given by Equation (13). Experiments conducted using the 2012 dataset obtained from [33] showed that the proposed mt equation satisfied four key principles which were based on widely accepted and used concepts including the Apriori principle.

Enhancing the frequency of the optimal target is possible through marketing treatment on customers who are known purchasers of the antecedent. For enhanced targeted marketing, customers were clustered into four groups with varying loyalty and simulated marketing initiatives were conducted using a Markov-based simulator. Results showed that it was possible to cluster customers based on loyalty and achieve significant enhancements in the frequency of itemsets through both conservative and aggressive, targeted marketing campaigns.

The effectiveness of the proposed model was compared to the model detailed in [50] in the context of previous studies in MBA, FCM and marketing and using four tests. Results showed that the proposed model outperformed the model detailed in [50] in three of the four tests while adhering to MBA, FCM and marketing principles outlined in key studies including [3], [42] and [21]. Comparisons using the fourth test was not possible, however simulation tests conducted in this study and in [50] showed that both models passed test 4 by demonstrating an enhancement in frequency of targeted itemsets.

6.1 Summary of Theoretical and Practical Implications

Theoretical Implications

The proposed algorithm which includes the market target (mt) model and the FCM clustering approach has been developed mathematically in Section 3 using well-grounded theory and concepts [3][8]. The model not only adds to the body of knowledge on MBA but also provides a novel, simple, yet effective way to tackle a well-known problem of finding the best itemset(s) to target in very large datasets [4][29]. The several tests performed on the model against well-known principles, as outlined in Section 4.3, have all yielded good performance in that the model output was consistent and robust.

Practical Implications

The proposed algorithm and model enables management of stores to easily take marketing decisions that not only minimises their marketing spend but also supports customer expansion and prevents unnecessary revenue erosion.

On minimising marketing spend, the market target model can be easily computed for all infrequent itemsets and those that have the lowest market target (from all potential targets) can be selected for promotional activity. This is particularly useful in large datasets and/or where support and confidence values are not both high or both low and/or where management have a desire to target increased sales on a specific consequent to a limited number of customers.

On supporting customer expansion and preventing unnecessary revenue erosion, the FCM clustering approach based on customer loyalty for the store and support for the targeted itemset allows management to isolate all customers who have been known purchasers of items in the itemset into four categories ("Switcher", "Drop Out", "Light Touch", "Leave Alone") where a variety of specific treatments may be applied. The "Switcher" category is likely to be the most valuable as it allows management to target customers who do not frequently purchase all items in the itemset more aggressively than their loyal customers, thereby attracting them to increase purchases within the store and potentially spend less at their rivals. At the same time, by isolating "Leave Alone" and "Light Touch" loyals, management can model the amount of revenue erosion that is likely given that they may incentivise this group with lower prices - a group that already spends large volumes of their shopping budget at the store.

Based on the above, the proposed algorithm and model may have significant benefits at a practical level and may be incorporated into stores' strategic and operational decision support systems.

6.2 Future Work and Other Applications

Experiments on the marketing target model were conducted on a quantity basis in this study. Future work will be largely in finding new applications for the model. Within grocery retail, this could be extended to include other variables such as revenue, profit, environmental friendliness, calories, etcetera. Experiments beyond the grocery retail sector could extend into medicine where patients presenting with symptoms could be clustered based on risk of a consequent disease and offered tailored treatment plans. Further, the prioritisation of public health initiatives could be done using the model, for example choosing how to prioritise between "stop smoking" initiatives or "obesity reduction" or flu vaccine campaigns given limited budgets. In education where the underlying causes for pupil absenteeism which are not obvious could be isolated and whole school initiatives undertaken to address some of these causes, thereby improving pupil attendance and consequently pupil performance. The model could be used to isolate the biggest drivers of overall poor attendance and initiatives could be planned to reduce the impacts of these drivers, for example before school clubs for pupils to discourage late attendance, or free breakfast club on certain mornings (where attendance is generally poor) to encourage pupils to arrive earlier.

Appendix

A Development of Equation (10)

The number of physical transactions required for an itemset to be frequent is given by min support. $P(A, C)$ is always initially less than minsup, hence the number of physical transactions required to make (A, C) frequent is given by:

$$\text{Transactions to make (A,C) frequent} = (\text{minsup} - P(A, C)) \cdot \text{total transactions} \tag{15}$$

The probability of a customer having (A, C) in their basket after initially purchasing (A) is predicated by purchasing (C) and is by given by:

$$P(A, C) = P(C|A) \cdot P(A) \tag{16}$$

Hence the probability of purchasing (C) after purchasing (A) given by:

$$P(C|A) = \frac{P(A, C)}{P(A)} \tag{17}$$

From equation (17), not everyone that purchases (A) will go on to purchase (C). Hence to achieve the required number of transactions to make (A, C) frequent will require a higher number of transactions to be targeted. This number of transactions is called the market target. Hence:

$$\text{Transactions to make (A,C) frequent} = \text{market target} \cdot \frac{P(A, C)}{P(A)} \tag{18}$$

Finally by combining equations (18), (15) and (3), equation (10) results:

$$\text{market target} \cdot \text{conf}(A \rightarrow C) = (\text{minsup} - P(A, C)) \cdot \text{total transactions} \quad \blacksquare$$

References

[1] Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions on*, 17(6):734–749, 2005.

[2] Chain Store Age. Five ways walmart uses big data. https://www.chainstoreage.com/operations/five-ways-walmart-uses-big-data/, 2017. Accessed: 2019-02-28.

[3] Rakesh Agrawal, Tomasz Imieliński, and Arun Swami. Mining association rules between sets of items in large databases. In *ACM SIGMOD Record*, volume 22, pages 207–216. ACM, 1993.

[4] Rakesh Agrawal, Ramakrishnan Srikant, et al. Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB*, volume 1215, pages 487–499, 1994.

[5] Shahriar Akter and Samuel Fosso Wamba. Big data analytics in e-commerce: a systematic review and agenda for future research. *Electronic Markets*, 26(2):173–194, 2016.

[6] Azarnoush Ansari and Arash Riasi. Customer clustering using a combination of fuzzy c-means and genetic algorithms. *International Journal of Business and Management*, 11(7):59, 2016.

[7] Sugato Basu, Ian Davidson, and Kiri Wagstaff. *Constrained clustering: Advances in algorithms, theory, and applications.* CRC Press, 2008.

[8] James C Bezdek, Robert Ehrlich, and William Full. Fcm: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2-3):191–203, 1984.

[9] Christian Borgelt. Frequent item set mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(6):437–456, 2012.

[10] Yasemin Boztuğ and Thomas Reutterer. A combined approach for segment-specific analysis of market basket data, 2006.

[11] Zeynel Cebeci and Figen Yildiz. Comparison of k-means and fuzzy c-means algorithms on different cluster structures. *AGRÁRINFORMATIKA/JOURNAL OF AGRICULTURAL INFORMATICS*, 6(3):13–23, 2015.

[12] Dunhumby. Heartbeat market basket analytics. https://www.dunnhumby.com, 2015. Accessed: 2019-02-28.

[13] Joseph C Dunn. Well-separated clusters and optimal fuzzy partitions. *Journal of cybernetics*, 4(1):95–104, 1974.

[14] Fiona Ellis-Chadwick, Neil F Doherty, and Leonidas Anastasakis. E-strategy in the uk retail grocery sector: a resource-based analysis. *Managing Service Quality: An International Journal*, 17(6):702–727, 2007.

[15] Jonathan Elms, Ronan De Kervenoael, and Alan Hallsworth. Internet or store? an ethnographic study of consumers' internet and store-based grocery shopping practices. *Journal of Retailing and Consumer Services*, 32:234–243, 2016.

[16] Peter S Fader, Bruce GS Hardie, and Ka Lok Lee. Rfm and clv: Using iso-value curves for customer base analysis. *Journal of marketing research*, 42(4):415–430, 2005.

[17] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. From data mining to knowledge discovery in databases. *AI magazine*, 17(3):37, 1996.

[18] Nathan M Fong, Yuchi Zhang, Xueming Luo, and Xiaoyi Wang. Targeted promotions and cross-category spillover effects. 2016.

[19] Forbes. Big data at tesco: Real time analytics at the uk grocery retail giant. http://www.forbes.com/sites/bernardmarr/2016/11/17/big-data-at-tesco-real-time-analytics-at-the-uk-grocery-retail-giant/43bd506a519a, 2016. Accessed: 2017-01-30.

[20] Forbes. How retailers can maximize the power of coupons. https://www.forbes.com/sites/bryanpearson/2017/03/15/research-reveals-how-retailers-can-maximize-the-power-of-coupons/7afaa4e2f016, 2017. Accessed: 2019-08-26.

[21] Els Gijsbrechts, Katia Campo, and Mark Vroegrijk. Save or (over-) spend? the impact of hard-discounter shopping on consumers' grocery outlay. *International Journal of Research in Marketing*, 2018.

[22] Bart Goethals. Frequent set mining. *Data mining and knowledge discovery handbook*, pages 321–338, 2010.

[23] L Gordon. Leading practices in market basket analysis: How top retailers are using market basket analysis to win margin and market share. *Factpoint Group*, 2008.

[24] Anjana Gosain and Sonika Dahiya. Performance analysis of various fuzzy clustering algorithms: a review. *Procedia Computer Science*, 79:100–111, 2016.

[25] Dhruv Grewal, Anne L Roggeveen, and Jens Nordfält. The future of retailing. *Journal of Retailing*, 93(1):1–6, 2017.

[26] Guardian. How supermarkets get your data. https://www.theguardian.com/money/2013/jun/08/supermarkets-get-your-data, 2013. Accessed: 2019-02-28.

[27] Guardian. Uk grocery sales in decline for first time in 20 years. https://www.theguardian.com/business/2014/nov/18/uk-grocery-sales-decline-price-war-asda-sainsburys-morrisons-tesco, 2014. Accessed: 2018-11-30.

[28] Jiawei Han, Hong Cheng, Dong Xin, and Xifeng Yan. Frequent pattern mining: current status and future directions. *Data Mining and Knowledge Discovery*, 15(1):55–86, 2007.

[29] Jiawei Han, Micheline Kamber, and Jian Pei. Mining frequent patterns, associations, and correlations-6: Basic concepts and methods. 2012.

[30] Jiawei Han, Jian Pei, Yiwen Yin, and Runying Mao. Mining frequent patterns without candidate generation: A frequent-pattern tree approach. *Data mining and knowledge discovery*, 8(1):53–87, 2004.

[31] Wu He, Jiancheng Shen, Xin Tian, Yaohang Li, Vasudeva Akula, Gongjun Yan, and Ran Tao. Gaining competitive intelligence from social media data: evidence from two largest retail chains in the world. *Industrial Management & Data Systems*, 115(9):1622–1636, 2015.

[32] Jeff Heaton. Comparing dataset characteristics that favor the apriori, eclat or fp-growth frequent itemset mining algorithms. In *SoutheastCon 2016*, pages 1–7. IEEE, 2016.

[33] Kantar. Kantar's consumer panel shopping data - uk. Private Communication, 2017.

[34] Orhan Kesemen, Özge Tezel, and Eda Özkul. Fuzzy c-means clustering algorithm for directional data (fcm4dd). *Expert Systems with Applications*, 58:76–82, 2016.

[35] Farnoosh Khodakarami and Yolande E Chan. Exploring the role of customer relationship management (crm) systems in customer knowledge creation. *Information & Management*, 51(1):27–42, 2014.

[36] Paul Klemperer. Price wars caused by switching costs. *The Review of Economic Studies*, 56(3):405–420, 1989.

[37] Friedrich Leisch and Bettina Grün. *Extending standard cluster algorithms to allow for group constraints.* na, 2006.

[38] Nizar R Mabroukeh and Christie I Ezeife. A taxonomy of sequential pattern mining algorithms. *ACM Computing Surveys (CSUR)*, 43(1):3, 2010.

[39] Manthan. Food and grocery analytics in 2018 – the new age realities. https://www.manthan.com/blogs/food-and-grocery-analytics-in-2018-the-new-age-realities/, 2018. Accessed: 2019-02-15.

[40] Sandra C Matz, Michal Kosinski, Gideon Nave, and David J Stillwell. Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the national academy of sciences*, 114(48):12714–12719, 2017.

[41] Raymond Moodley, Francisco Chiclana, Fabio Caraffini, and Jenny Carter. Application of uninorms to market basket analysis. *International Journal of Intelligent Systems*, 2018.

[42] María N Moreno, Saddys Segrera, Vivian F López, María Dolores Muñoz, and Ángel Luis Sánchez. Web mining based framework for solving usual problems in recommender systems. a case study for movies recommendation. *Neurocomputing*, 176:72–80, 2016.

[43] Janmenjoy Nayak, Bighnaraj Naik, and HS Behera. Fuzzy c-means (fcm) clustering algorithm: a decade review from 2000 to 2014. In *Computational intelligence in data mining-volume 2*, pages 133–149. Springer, 2015.

[44] Eric WT Ngai, Li Xiu, and Dorothy CK Chau. Application of data mining techniques in customer relationship management: A literature review and classification. *Expert systems with applications*, 36(2):2592–2602, 2009.

[45] Yue Pan, Simon Sheng, and Frank T Xie. Antecedents of customer loyalty: An empirical synthesis and reexamination. *Journal of Retailing and Consumer Services*, 19(1):150–158, 2012.

[46] Marina Paolanti, Daniele Liciotti, Rocco Pietrini, Adriano Mancini, and Emanuele Frontoni. Modelling and forecasting customer navigation in intelligent retail environments. *Journal of Intelligent & Robotic Systems*, 91(2):165–180, 2018.

[47] Tesco PLC. Tesco - open heart surgery in public. https://www.tescoplc.com/media/264601/, 2019. Accessed: 2019-07-03.

[48] Lawrence Rabiner and B Juang. An introduction to hidden markov models. *ieee assp magazine*, 3(1):4–16, 1986.

[49] Nicholas J Radcliffe and Patrick D Surry. Differential response analysis: Modeling true response by isolating the effect of a single action. *Credit Scoring and Credit Control VI. Edinburgh, Scotland*, 1999.

[50] Thomas Reutterer, Kurt Hornik, Nicolas March, and Kathrin Gruber. A data mining framework for targeted category promotions. *Journal of Business Economics*, 87(3):337–358, 2017.

[51] Hongjai Rhee and David R Bell. The inter-store mobility of supermarket shoppers. *Journal of Retailing*, 78(4):225–237, 2002.

[52] Natsuki Sano, Reo Tsutsui, Katsutoshi Yada, and Tomomichi Suzuki. Clustering of customer shopping paths in japanese grocery stores. *Procedia Computer Science*, 96:1314–1322, 2016.

[53] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. Analysis of recommendation algorithms for e-commerce. In *Proceedings of the 2nd ACM conference on Electronic commerce*, pages 158–167. ACM, 2000.

[54] Catherine M Schaffer and Paul E Green. Cluster-based market segmentation: some further comparisons of alternative approaches. *Market Research Society. Journal.*, 40(2):1–10, 1998.

[55] Brent Smith and Greg Linden. Two decades of recommender systems at amazon. com. *Ieee internet computing*, 21(3):12–18, 2017.

[56] David Solnet, Yasemin Boztug, and Sara Dolnicar. An untapped gold mine? exploring the potential of market basket analysis to grow hotel revenue. *International Journal of Hospitality Management*, 56:119–125, 2016.

[57] Zeynab Soltani and Nima Jafari Navimipour. Customer relationship management mechanisms: A systematic review of the state of the art literature and recommendations for future research. *Computers in Human Behavior*, 61:667–688, 2016.

[58] Herb Sorensen, Svetlana Bogomolova, Katherine Anderson, Giang Trinh, Anne Sharp, Rachel Kennedy, Bill Page, and Malcolm Wright. Fundamental patterns of in-store shopper behavior. *Journal of Retailing and Consumer Services*, 37:182–194, 2017.

[59] Charlotte Sturley, Andy Newing, and Alison Heppenstall. Evaluating the potential of agent-based modelling to capture consumer grocery retail store choice behaviours. *The International Review of Retail, Distribution and Consumer Research*, 28(1):27–46, 2018.

[60] Pang-Ning Tan. *Introduction to data mining*. Pearson Education India, 2018.

[61] Pham Huy Thong et al. Picture fuzzy clustering: a new computational intelligence method. *Soft computing*, 20(9):3549–3562, 2016.

[62] Mauricio A Valle, Gonzalo A Ruz, and Rodrigo Morrás. Market basket analysis: Complementing association rules with minimum spanning trees. *Expert Systems with Applications*, 97:146–162, 2018.

[63] Harald J van Heerde and Scott A Neslin. Sales promotion models. In *Handbook of marketing decision models*, pages 13–77. Springer, 2017.

[64] Rajkumar Venkatesan and Paul W Farris. Measuring and managing returns from retailer-customized coupon campaigns. *Journal of marketing*, 76(1):76–94, 2012.

[65] Neha Verma and Jatinder Singh. A comprehensive review from sequential association computing to hadoop-mapreduce parallel computing in a retail scenario. *Journal of Management Analytics*, 4(4):359–392, 2017.

[66] Marketing Week. Using vouchers and incentives to add brand value. https://www.marketingweek.com/using-vouchers-and-incentives-to-add-brand-value/, 2015. Accessed: 2019-08-26.

[67] Roland Winkler, Frank Klawonn, and Rudolf Kruse. Problems of fuzzy c-means clustering and similar algorithms with high dimensional data sets. In *Challenges at the Interface of Data Analysis, Computer Science, and Optimization*, pages 79–87. Springer, 2012.

[68] Ian H Witten, Eibe Frank, Mark A Hall, and Christopher J Pal. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.

[69] Shuo Yang, Mohammed Korayem, Khalifeh AlJadda, Trey Grainger, and Sriraam Natarajan. Combining content-based and collaborative filtering for job recommendation system: A cost-sensitive statistical relational learning approach. *Knowledge-Based Systems*, 136:37–45, 2017.

[70] Mohammed J Zaki. Scalable algorithms for association mining. *Knowledge and Data Engineering, IEEE Transactions on*, 12(3):372–390, 2000.

[71] Mohammed J Zaki and Karam Gouda. Fast vertical mining using diffsets. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 326–335. ACM, 2003.

[72] Mohammed J Zaki, Wagner Meira Jr, and Wagner Meira. *Data mining and analysis: fundamental concepts and algorithms*. Cambridge University Press, 2014.