MODELING AND PREDICTING CUSTOMER PURCHASE BEHAVIOR IN THE GROCERY RETAIL INDUSTRY

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF INFORMATICS OF

THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

SERHAT PEKER

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

THE DEPARTMENT OF INFORMATION SYSTEMS

JULY 2017

Approval of the thesis:

MODELING AND PREDICTING CUSTOMER PURCHASE BEHAVIOR IN THE GROCERY RETAIL INDUSTRY

Submitted by **SERHAT PEKER** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Information Systems Department, Middle East Technical University** by,

Prof. Dr. Deniz Zeyrek Bozşahin
Director, Graduate School of **Informatics**

Prof. Dr. Yasemin Yardımcı Çetin
Head of Department, **Information Systems**

Assoc. Prof. Dr. Altan Koçyiğit
Supervisor, **Information Systems Dept., METU**

Examining Committee Members:

Assoc. Prof. Dr. Cem İyigün
Industrial Engineering Dept., METU

Assoc. Prof. Dr. Altan Koçyiğit
Information Systems Dept., METU

Assist. Prof. Dr. P. Erhan Eren
Information Systems Dept., METU

Assist. Prof. Dr. Çağdaş Evren Gerede
Computer Engineering Dept., TOBB UET

Assist. Prof. Dr. Rahime Belen Sağlam
Computer Engineering Dept., Yıldırım Beyazıt University

Date: **18.07.2017**

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : **Serhat Peker**

Signature : ______________________

ABSTRACT

MODELING AND PREDICTING CUSTOMER PURCHASE BEHAVIOR IN THE GROCERY RETAIL INDUSTRY

Peker, Serhat

Ph.D., Department of Information Systems

Supervisor: Assoc. Prof. Dr. Altan Koçyiğit

July 2017, 82 pages

In today's business, grocery retail industry companies operate in highly competitive environment. Such an intense competition have compelled companies to develop close and long-term relationships with their customers by implementing more targeted marketing strategies and personalized services. To implement such customized services, modelling and predicting customer purchase behaviors are essential. Accordingly, this thesis mainly aims to model and predict the customers' purchasing behavior in the grocery retail industry using machine learning techniques on past customer purchase logs. To this end, customer segmentation, product segmentation, prediction of customers' individual purchase behaviors, and shopping list prediction are studied and a novel evaluation metric and an approach for determining recommendation list size are proposed. This thesis may serve as a valuable reference for academics and researchers who are willing to investigate customer purchase behavior and identify hidden patterns in their transactional data, and also promises substantial benefits to marketers and decision makers of grocery retailing industry in developing customized services and marketing activities.

Keywords: Knowledge Discovery in Databases, Machine Learning, Data Mining, Customer Modeling, Grocery Retail Industry

ÖZ

PERAKENDE SEKTÖRÜNDE MÜŞTERİ SATIN ALMA DAVRANIŞININ
MODELLENMESİ VE TAHMİN EDİLMESİ

Peker, Serhat
Doktora Bilişim Sistemleri Bölümü
Tez Yöneticisi: Doç. Dr. Altan Koçyiğit

Temmuz 2017, 82 sayfa

Günümüz iş dünyasında, perakende sektöründeki şirketler son derece rekabetçi bir ortamda faaliyet göstermektedir. Böyle yoğun bir rekabet, şirketleri daha hedefli pazarlama stratejileri ve kişiye özgü hizmetler uygulayarak müşterileri ile yakın ve uzun vadeli ilişkiler geliştirmeye zorlamıştır. Bu gibi kişiye özgü hizmetler uygulamak için, müşteri satın alma davranışlarının modellenmesi ve tahmin edilmesi esastır. Buna göre, bu tez, esas olarak, geçmiş müşteri alımları üzerine makine öğrenme teknikleri kullanılarak perakende sektöründe müşterilerin satın alma davranışlarını modellemeyi ve tahmin etmeyi amaçlamaktadır. Bu amaçla, müşteri segmentasyonu, ürün segmentasyonu, müşterilerin bireysel satın alma davranışlarının tahmini, alışveriş listesi tahminini çalışmaları yapılmış, yeni bir değerlendirme metriği ve öneri listesi boyutunu belirleyen bir yaklaşım önerilmiştir. Bu tez, müşteri satın alma davranışını araştırmak ve işlem verilerindeki gizli kalıpları belirlemek isteyen akademisyenler ve araştırmacılar için değerli bir referans olabilir ve kişiye özgü hizmetler ve pazarlama faaliyetleri geliştirilmesinde perakende sektöründeki pazarlamacılar ve karar vericilere önemli faydalar sağlar.

Anahtar Sözcükler: Veri Tabanlarında Bilgi Keşfi, Makine Öğrenimi, Veri Madenciliği, Müşteri Modellemesi, Perakende Sektörü

To my mother, Gülfem Arslanoglu and my wife, Gaye Peker...

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to my supervisor, Assoc. Prof. Dr. Altan Koçyiğit for his continuous support, valuable guidance and motivation. I appreciate all his contributions throughout my PhD thesis study.

I am thankful to the thesis supervising committee members Assoc. Prof. Dr. Cem İyigün and Assist. Prof. Dr. P. Erhan Eren for their invaluable instruction in helping me complete this dissertation. I would also like to thank the examining committee members Assist. Prof. Dr. Rahime Belen Sağlam and Assist. Prof. Dr. Çağdaş Evren Gerede for their valuable comments and help in improving the thesis.

I am grateful to The Scientific and Technological Research Council of Turkey (TUBITAK) for their financial support to this study with 2211-A doctoral research scholarship program.

Special thanks to Ali Mert Ertuğrul for his valuable discussion on various research topics and providing helpful suggestions; Süleyman Özarslan and İdris Songör for providing me a real-life dataset and his support; Mahir Kaya and Emre Sezgin for their continuous supports and motivations; Ayşegül Özkaya Eren, Mehmet Ali Akyol, Özge Gürbüz and Sibel Gülnar for their helps and supports. I would also like to thank all of my collegues in the institute for creating a nice working environment and for their understanding during my research. It is a great pleasure working with them, particularly, Seyma Çavdar, Gökçen Yılmaz, Murat Salmanoğlu, Ali Kaan Sungur, Kerem Kayabay, and Mert Onuralp Gökalp. Throughout my PhD thesis study, I shared so many great moments with all of these people mentioned in this paragraph and they made this long journey more enjoyable.

I am so thankful to my grandmother, Hatice Arslanoğlu for raising me up with love. I know she is smiling up on me from the heaven.

Last but not the least, I would like to express my deepest gratitude to my mother, Gülfem Arslanoğlu and my wife Gaye Peker. They never once stopped believing in me and always be there whenever I need. I thank my lovely wife for her invaluable contributions to my thesis, and her great patience and endless love. I am grateful to my beloved mother whose unconditional love and continuous spiritual support brought me to these days. I am really lucky to have such wonderful women in my life and being aware of this makes me feel perfect. Thus, without their love and encouragement, I would not have finished this dissertation and of course without them all this effort would be worth nothing.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CRM	Customer Relationship Management
RFM	Recency, Frequency, Monetary
KDD	Knowledge Discovery in Databases
LRFMP	Length, Recency, Frequency, Monetary, Periodicity
FMC	Frequency, Monetary, Customer Variety
WSS	Within Sum of Squares
LR	Logistic Regression
DT	Decision Trees
RF	Random Forests
SVM	Support Vector Machines
NN	Neural Networks
MC	Majority Classifier
RBF	Radial Basis Function
MLP	Multi-Layer Perceptron
RNN	Recurrent Neural Networks
MP	Most Popular
ARHR	Average Reciprocal Hit-Rank
IQR	Interquartile Range

CHAPTER 1

INTRODUCTION

In recent years, benefiting from the strengthening global economy, grocery retail industry has grown considerably. According to the United States Department of Agriculture (USDA) (2016), global food retail sales are about $4 trillion annually and Institute of Grocery Distribution (IGD) (2015) expects the value of the world's grocery market to triple within the next five years and reach almost $12 trillion by 2020. With the rapid growth of its economy, it has been one of the most appealing sectors for investors. In this manner, global supermarket chains have expanded their presence in the sector, whereas many new companies have entered into the market in recent years. All of these have led to intense competition among companies in this industry. Therefore, such an increasing market competition in the grocery retail industry puts pressure on firms to effectively manage customer relationships and offer customized services to gain or maintain competitive advantage in the sector.

On the other hand, recent advances in information technology and continuously decreasing costs in data storage and processing have given today's companies the ability to collect and store massive amount of customer transaction data. In grocery retail industry, through loyalty card programs, firms are capturing great volumes of point-of-sale data which hold intrinsically valuable information about shopping trends and patterns of both customers and products. In this context, knowledge discovery in database (KDD) has become highly important for such enterprises to expose the hidden value of transactional data (Fayyad et al. 1996). The core part of KDD process is the application of data mining techniques with which the knowledge is inferred from the data and is transformed it into a human-understandable form for further use (Witten et al. 2016; Hand et al. 2001). Thus, firms can utilize such knowledge in the development of customized services for their customers.

In the literature, studies conducted in the grocery retail industry to extract actionable knowledge from customers' transactional data can be categorized into two major groups. The first group is the studies which use association rule mining (Agrawal et al. 1993) to discover relationships between the products. Several researches (Brijs et al. 2004; Brijs et al. 1999; Brijs et al. 2000; Lawrence et al. 2001; Chen & Lin 2007) have utilized association rule mining for determining which products are purchased together, and turned this knowledge into marketing strategies and product placements. The second group of studies (Wang et al. 2015; Rendle et al. 2010; Wan et al. 2015; Yu et al. 2016) have employed sequence mining based on customers' past purchase sequences for next basket prediction/recommendation.

Despite such a great deal of research activity, only the study (Cumby et al. 2004) has developed individualized prediction models to predict the customers' shopping behavior. Moreover, clustering is one of the most popular techniques in KDD. While many researches have applied customer segmentation in numerous industries including health and beauty (Khajvand et al. 2011), textile (D.-C. Li et al. 2011), healthcare (Wei et al. 2012), hairdressing (Wei et al. 2013), banking (Nikumanesh & Albadvi 2014), and tourism (Dursun & Caber 2016) using different cluster analysis techniques, literature on the use of this technique in the grocery retail industry remains rather limited, and there have been relatively few studies (Lawrence et al. 2001; Chen et al. 2005) utilizing this approach for forming cluster-specific solutions. In addition to these, grocery shopping shows some differences from other fields. For example, customers can buy multiple products in a transaction, and they can also buy similar products in their repeated purchases. Thus, it is essential to consider such issues in designing and evaluation of prediction and recommendation approaches.

Based on the aforementioned research gaps, the objectives of this thesis are:

- To infer actionable knowledge from shopping trends and patterns of both customers and products in grocery retail industry for utilizing it in customized services and marketing strategies.
- To understand and predict customers' individual shopping behaviors and characteristics in grocery retail industry.
- To design and evaluate efficient predictive modeling approaches.

1.1. Research Questions

To achieve above research goals, the following research questions have been pointed out:

- Can we represent customers' behaviors more appropriately and identify customer groups more effectively in grocery retail industry?
- Can we segment products in grocery retail industry using customers' purchase transactions?
- Can combination of individual-level and segment-based predictive modeling approaches improve the performance of prediction of an individual's purchase behavior in terms of prediction accuracy and coverage?
- Which predictive modeling approach is the most effective for shopping list prediction?
- Can we develop a metric for ranking evaluation of multiple preferences in top-n recommendation and prediction lists?
- How can we determine the recommendation list size for customers' multiple item preferences?

In this thesis, six complementary studies are conducted to answer the above research questions. To answer the first question, the existing customer segmentation models are reviewed and the shortcomings of traditional models are identified. Thus, a new customer segmentation model is proposed and its applicability is investigated by conducting a case study on a local grocery chain.

Regarding the second question, a new product segmentation model is developed by getting inspired by the proposed customer segmentation models. Based on this model, a methodology is proposed to segment products by considering customers' purchase transactions. Consequently, the applicability of the proposed methodology is investigated by conducting a case study on a local grocery chain.

In order to answer the third question, the existing predictive modeling approaches of customer behavior are reviewed and their both strengths and weaknesses are identified. To overcome their limitations, a novel hybrid approach, which is combination of these approaches, is proposed to predict individual purchase behaviors of customers in the grocery retailing industry. The proposed approach is evaluated on a real-life dataset by conducting a series of experiments and its performance is also compared to the performances of individual-level and segment-based approaches in terms of prediction accuracy and coverage.

For the fourth question, both conventional predictive modeling approaches, and hybrid approach proposed in this dissertation are applied for the shopping list prediction and the prediction performances are investigated and compared on a real-world dataset obtained from a grocery retailing company.

To answer the fifth question a novel metric is proposed to evaluate the quality of Top-N recommendation and prediction lists for customers with multiple preferences at a time or a specific time interval. The use of the proposed metric is demonstrated in an example and its performance is also investigated by conducting an exhaustive set of experiments on the real-life data from the grocery shopping domain.

For the sixth question, a novel approach to determine the recommendation list sizes for multiple preferences of users is developed. Based on customers' earlier preferences, different machine learning techniques are employed for this purpose. The proposed approach is evaluated on real-life data from grocery shopping domain by conducting a series of experiments.

1.2. Contributions

The main contributions of this thesis are summarized as follows:

- A novel customer segmentation model is proposed and customer types in the grocery retail industry are characterized according to segmentation results.
- A novel product segmentation methodology is proposed based on customers' purchase transactions and different product groups in grocery retail industry are identified according to segmentation results.
- A hybrid approach is proposed for predicting customers' individual purchase behavior.
- A performance comparison is presented for predictive modeling approaches in shopping list prediction.
- A novel metric is proposed for ranking evaluation of multiple preferences in top-n recommendation and prediction lists.
- A novel approach is proposed for adjusting recommendation list size for customers' multiple item preferences.

1.3. Thesis Organization

The thesis is organized in 9 chapters as follows:

Chapter 2 presents a comprehensive literature review including customer segmentation, RFM models, clustering techniques, predictive modelling approaches for customer behavior.

Chapter 3 introduces a novel customer segmentation model and presents a case study in which the customers of a grocery company are clustered into different groups. Each customer group is characterized and corresponding managerial implications are provided.

Chapter 4 presents a product segmentation methodology based on customers' purchase transactions. The proposed approach is applied to real-life data and product groups with different characteristics are identified in the grocery retail industry.

Chapter 5 introduces a hybrid approach for predicting customers' individual purchase behaviors which combines the individual-level and segment-based predictive approaches. The proposed approach is evaluated on a real-life dataset by conducting a series of experiments. The performance of the proposed approach is compared with the individual-level and segment-based approaches and results are discussed.

Chapter 6 presents a comparison of predictive modeling approaches for the shopping list prediction. The comparison is performed on a real-life dataset by conducting a series of experiments, and the results and findings are presented.

Chapter 7 introduces a modified version of average reciprocal hit rank (ARHR) metric for ranking evaluation of multiple preferences in top-n recommendation and prediction lists. The applicability of the proposed metric is demonstrated on the grocery shopping domain by conducting a series of experiments on real-life data, and thus experimental results are reported.

Chapter 8 presents a novel approach to dynamically adjust the recommendation list size for multiple preferences of a customer. The proposed approach is evaluated on real-life data from grocery shopping domain by conducting a series of experiments, and results are reported.

Chapter 9 concludes the thesis with a summary of the main contributions, and a discussion about future research lines.

CHAPTER 2

RELATED WORK

2.1. Customer Segmentation

Customers have varying needs, behaviors and preferences, and it is challenging for companies to serve all customers equally well. Customer segmentation emerged in response to this problem. It was first introduced by Smith (1956), and it has been further supported by many academic studies and has been applied successfully by several companies from various sectors. Customer segmentation is the division of entire customers into distinctive smaller groups, consisting of customers with similar needs and characteristics (McDonald & Dunbar 2004). Today, customer segmentation is a fundamental marketing activity and increasing number of companies are applying it to gain a deeper understanding of their customers' characteristics and needs.

Customer segmentation provides two significant benefits to management and marketing departments. The first one is to allow the effective identification of the key customer groups that include the most profitable and loyal customers (Dibb 1998; Wind 1978; Kotler 1997). Customer segmentation has been employed to determine the profit potential of customers by several research efforts (Wiedmann et al. 2009; Hwang et al. 2004; Chan 2008; Huang et al. 2009; Cheng & Chen 2009; Chuang et al. 2013; Safari et al. 2016; Dursun & Caber 2016; Kao et al. 2011; Chen et al. 2012). With this opportunity, decision makers can target these high-value groups of customers, and thus direct their efforts specifically toward these designated segments. The other important benefit of customer segmentation is enabling management of companies to understand customers' behavior and preferences, and acquire knowledge about different groups of customers. In this way, it is possible to serve each customer segment according to their specific needs and preferences (Dibb 1998). Such knowledge also provides opportunity to more accurately tailor marketing actions and materials to individual customer needs (MacQueen 1967; Dibb & Simkin 1997). In this regard, plenty of studies (Chan 2008; D.-C. Li et al. 2011; Wei et al. 2013; Wei et al. 2012) have applied customer segmentation to produce discriminative customer management and marketing strategies for different types of customer groups.

In customer segmentation, customers are grouped in terms of any kind of variables which can be broadly divided into two groups: general variables and product specific variables (Wedel & Kamakura 2000). The general variables include customer demographics (e.g. sex, age, income, education level, etc.) and lifestyles, whereas the product specific variables include customer purchasing behaviors (e.g. frequency of purchase, consumption, spending, etc.) and intentions. However general variables is easier to apply, the product specific variables are more important for capturing purchase behaviors of customers and more likely to differentiate customer contributions to a business (Tsai & Chiu 2004). In this context, RFM models' attributes are well-known and the most employed characteristics for customer's purchase behavior (Bauer 1988), and Newell (1997) further pointed out that the attributes of these models are very effective for customer behavior analysis and segmentation.

2.2. RFM Models

The RFM model was first proposed by Hughes (1996) to analyze and predict customers' behavior (Hughes 1996). The basic RFM model relies on three attributes which are recency (R), frequency (F), and monetary (M). Recency is the time interval since last purchase (e.g., days or months) and gives information about buying or visiting potential of the customer. If this interval is short, the likelihood of repurchase or revisit is high. Frequency is the number of purchases or visits within a certain period of time, and it is an indication of the customer loyalty. The higher the frequency is, the higher the customer loyalty becomes. Monetary is the total amount spent or the average amount spent per visit during a certain period of time and measures the contribution of the customer to the revenue of a company. The greater the amount spent is, the more the customer contributes to the revenue.

RFM model has evolved during the past two decades. Some previous studies have attempted to develop new RFM models either by considering additional variables or by excluding some of the variables according to the nature of the product or service. Chang and Tsay (2004) established LRFM model by incorporating a new feature, customer relation length (L) into the original RFM model. As mentioned in Reinartz and Kumar (2000), customer relation length plays a vital role in customer loyalty, and thus LRFM model has become one of the most widely used RFM models by receiving considerable attention in recent literature (Hosseini et al. 2010; Wei et al. 2012; D.-C. Li et al. 2011; Kao et al. 2011). Yeh et al. (2009) later proposed RFMTC model by adding two variables: time since first purchase (T) and churn probability (C). In a similar study, Chang and Tsai (2011) has extended RFM model to GRFM model by adding product category group information. A recent study, Chen et al. (2015) considered profit as a new variable besides LRFM variables, while Wei et al. (2012) excluded the monetary variable from the original LRFM model.

RFM models have several advantages. For instance, they are implemented quickly (Kahan 1998) and managers and decision makers can easily interpret the results of these models (Marcus 1998). These models also capture customer characteristics by using only a relatively small number of features (Kaymak 2001). Hence, RFM models are commonly used in customer analysis and customer segmentation. In recent years, it has been shown that different types of RFM models mostly perform well in customer segmentation in a variety of domains, e.g., health and beauty (Khajvand et al. 2011), textile (D.-C. Li et al. 2011), outfitter (Kao et al. 2011), healthcare (Wei et al. 2012), hairdressing (Wei et al. 2013), banking (Nikumanesh & Albadvi 2014) and tourism (Dursun & Caber 2016). These numerous studies on different areas have shown the persistent academic interest in this topic, and the efficiency of RFM models in understanding and segmenting customer behavior.

In RFM models, customers are evaluated according to model features. A common approach to adopt a RFM model is using the actual values of variables for each customer. An alternative way which is called the customer quintile method proposed by Miglautsch (2000) has been preferred in several studies as well (Chen et al. 2015; Cheng & Chen 2009; Hosseini et al. 2010). This approach sorts the customers in descending order according to a specific variable of RFM model and then partitions them into five equal quintiles. The top 20% quintile having highest values is coded as 5. The next 20% quintile coded as 4 and so on. The same are repeated for each feature of the RFM model, and therefore these coded numbers are used as input for the behavior analysis. Another important concept in application of RFM models is the weight of the model variables. There are two broad approaches for this issue in the literature. The first one is that all the variables are equal in the importance, and thus their weights are identical (Hughes, 2005), whereas the other idea is that the importance of the model features varies among industries and they are weighted differently depending on the characteristics of the industry (Stone & Jacobs 2007). Generally, studies that use customer quintile method also prefer to employ weighting strategy for model variables in the implementation of RFM models. On the other hand, studies that use original data instead of coded numbers consider equal weights for the model variables.

2.3. Clustering Techniques

Clustering is an important and widely used tool in customer segmentation (Sarstedt & Mooi 2014; Chiu & Tavella 2008). The objective of clustering is to partition a set of data objects into disjoint groups in which objects share similar characteristics and have greater differences from the objects in other groups (Jiawei & Kamber 2001). There are many clustering techniques proposed in literature and they can be classified into two major groups: hierarchical clustering and partitional clustering (Witten & Frank 2005; Jain & Dubes 1988). Hierarchical clustering algorithms find nested clusters and yields a dendrogram (a tree diagram) that illustrates the arrangement of objects into different clusters, whereas partitional clustering algorithms assign the data objects into non-overlapping clusters such that each data object belongs to only one cluster (Jain et al. 1999; Jain 2010).

Among partitional algorithms, K-means algorithm (MacQueen 1967) is one of the most popular and widely used clustering techniques (Jain 2010). It can easily be implemented and is extremely fast in execution, hence it has been commonly used for clustering in various research fields and application areas including data mining, statistical data analysis, pattern recognition, customer segmentation and other business applications (Cheung 2003; Davidson 2002). This method requires a priori specification of the number of clusters (k). For a given k value, K-means algorithm starts with the random generation of k central points (i.e., centroids). Next, the distance between each instance and each centroid is calculated and then each instance is assigned to the closest centroid. After the clusters are formed, the mean value of each cluster is recalculated based on the current objects in the cluster. The process is iterated until converge occurs, i.e., centroids does not change. One needs to execute the K-means method with different numbers of clusters (k) to determine an optimal k value, because the performance of this algorithm depends on the k value (Michaud 1997).

Ward’s minimum-variance method was first proposed by Ward (1963) and it is one of the most well-known and commonly used hierarchical clustering algorithms. This algorithm is found to be the best among the agglomerative clustering methods by Jain and Dubes (1988). This algorithm basically determines the clusters based on the squared distances between instances different from other linkage methods. Firstly, each instance is considered as a cluster. Then, the sum of squared distances between different clusters are computed and the pair of clusters with the smallest sum of squared distances are merged into one cluster. This process is repeated until all instances are grouped under one cluster.

2.4. Predictive Modeling Approaches for Customer Behavior

Researchers have been trying for decades to predict customer behavior and in the literature, the main predictive modeling approaches for customer behavior fall into three distinct categories: whole customer base, segment-based and individual-level. Whole customer base approach is built based on the data pertaining to all customers (Jiang & Tuzhilin 2006). Segment-based approach is learned from the data pertaining to customers of a particular segment, while individual-level customer approach is learned from the transactional data pertaining to an individual customer (Jiang & Tuzhilin 2006).

Some early studies (Campbell et al. 2001; Bibelnieks & Campbell 2000; Allenby & Rossi 1998; Kamakura & Wedel 1999) have shown that segment-based approach is much more effective than the whole customer base approach. To employ the segment-based approach, customer segmentation is firstly applied. Once customers are partitioned into segments based on their demographic or behavioral characteristics, then the segment-based approach develops separate predictive models for each segment (Apte et al. 2003; Apte et al. 2002). Several research efforts (Apte et al. 2001; Jiang & Tuzhilin 2007; Jiang & Tuzhilin 2006; Jiang & Tuzhilin 2009; Faraone et al. 2012; Palmisano et al. 2008; Faraone et al. 2010; Palmisano et al. 2007; Liu et al. 2013; Lombardi et al. 2013) have employed this approach to provide more targeted and customized services. On the other hand, a separate prediction model is constructed for each individual customer in the individual-level approach. To build more effective predictive models so as to maximize the predictive accuracy, many studies (Jiang et al. 2013; Jiang & Tuzhilin 2009; Jiang & Tuzhilin 2006; Jiang & Tuzhilin 2007; Lombardi et al. 2013; Palmisano et al. 2007; Palmisano et al. 2008) have used this approach.

There are a couple of studies comparing the predictive accuracy performance of segment-based and individual-level approaches in the literature (Palmisano et al. 2008; Palmisano et al. 2007; Lombardi et al. 2013; Jiang & Tuzhilin 2006). These studies have shown that the predictive performance of the individual-level approach usually outperforms that of segment-based approach, but segment-based approaches taken at the optimum granularity levels provide better predictions of customer behavior when there are insufficient data to train the prediction model.

Both segment-based and individual-level approaches have some strengths and weaknesses. Table 1 compares these two approaches by highlighting their main advantages and disadvantages.

Table 1: Segment-based approach vs. individual level approach

	Segment-Based Approach	**Individual-Level Approach**
Strength	No need for a sufficient number of past transactional data for a particular customer-item pair to generate predictions.	Low levels of transaction heterogeneity which leads to a performance improvement in the prediction accuracy.
	Produces predictions for items that are not in the customer's historical data. This leads to a raise in the prediction coverage.	
Weakness	High levels of transaction heterogeneity which causes a performance decline in the prediction accuracy.	The data sparsity problem which occurs when prior transactional data is insufficient for a particular customer-item pair. This makes it difficult to produce predictions and reduces prediction coverage.
		Generates predictions only for items in the past transactional data of the customer. This negatively affects prediction coverage.

Segment-based approaches can generate predictions for a particular customer just after that customer is assigned to a segment and so predictions can be produced for a particular customer-item pair even upon few records of this pair. On the contrary, individual-level approach suffers from data sparsity because it uses a smaller set of data compared to segment-based approach (Hand et al. 2001; Jiang & Tuzhilin 2006). Hence, individual-level approach requires an adequate number of past transactional data for a particular customer-item pair to generate predictions, otherwise it is hard for that approach to build accurate prediction models. Moreover, segment-based approach provides predictions for items that are not in the customer's historical data, whereas individual-level approach is limited to make prediction only for items in the past transactional data of the customer. Therefore, such deficiencies of individual-level approach cause to failure to make predictions and this leads to a considerable decline in the prediction coverage.

Segment-based approach is less exposed to the data sparsity problem, but it has a disadvantage of using more heterogeneous data (Hand et al. 2001). Such heterogeneity formed through combining transactional data of several customers causes low accuracy predictions. On the other hand, data is more homogeneous in the individual-level approach, because only data pertaining to individual customer behavior is considered to build the predictive model. This leads to a reduced bias and thus a better predictive performance in terms of accuracy (Jiang & Tuzhilin 2006).

CHAPTER 3

SEGMENTING CUSTOMERS USING LRFMP MODEL

In this chapter, we propose a new RFM based model called LRFMP (Length, Recency, Frequency, Monetary and Periodicity) for classifying customers in the grocery retail industry; and to identify different customer segments in this industry based on the proposed model. A detailed, in-depth description of this study is in (Peker et al. 2017).

3.1. Introduction

With growing diversity in the preferences and tastes of modern customers, it is impossible for firms to fully satisfy every customer. This challenge is typical in grocery stores where a tremendous diversity of products is available. In this regard, customer segmentation enables companies to divide customers into distinct and internally homogeneous groups and interact with each customer segment separately. Moreover, customer segmentation is a critical success factor for understanding behavior of different groups of customers and evaluating their value (Yao et al. 2014). It also enables companies to identify valuable customers and to retain such customers, which is crucial for the business success in highly competitive industries. (Webster Jr 1992).

In order to perform customer segmentation, there are several techniques proposed in the literature, and among them, clustering is the most commonly used method (Wedel & Kamakura 2000). Moreover, RFM model features are effectively used for understanding and analysing customer behavior characteristics (Newell 1997; Kahan 1998; Fader et al. 2005) and they are also well known and the most widely used methods for customer analysis and segmentation due to its simplicity and applicability (Bult & Wansbeek 1995; Bauer 1988). Different versions of RFM models and cluster analysis have been successfully applied together for the purpose of customer segmentation in different application areas. Although the traditional RFM model performs well in customer segmentation, it ignores periodicity of customer behavior (visit or purchase) which provides significant information about customers. Besides, in classical RFM, the recency variable is defined to consider only the latest transaction of customers, but such a variable may not correctly indicate the customer's repetitive purchasing or visiting tendency in many cases, since customer's behaviors demonstrate temporal variation. In addition to these, despite the existence of several studies in different application domains that focus on how to segment customers for improving customer relationships and marketing strategies, there is little research on customer segmentation in the grocery retail industry.

To make up for above-mentioned shortcomings, this chapter aims to examine the customer segmentation in the increasingly growing Turkish grocery industry. In order to represent customers' behavior more appropriately and to identify customer groups more effectively, this chapter proposes an extended LRFM model, namely LRFMP (Length, Recency, Frequency, Monetary and Periodicity) by adding the periodicity as a new variable and modifying the classic recency variable. Based on LRFMP model features, both hierarchical (i.e., Ward) and partitional (i.e., K-means) clustering techniques are employed individually to cluster customers and intraclass inertia is used to determine the most appropriate one of these clustering algorithms. After the cluster analysis, the customers are segmented into distinct groups and these resulting customer groups are then profiled in accordance with LRFMP characteristics. Finally, unique CRM and marketing strategies for each customer profile are suggested.

The contributions of this chapter are twofold: (1) it proposes a new RFM model, called LRFMP to better represent customers' behavior in the grocery industry; (2) its findings about unique characteristics and behaviors of Turkish customers in the grocery industry and its implications for different customer types provide the managers of grocery companies with directions to enhance customer relationships and develop differentiated services and marketing strategies. The remainder of this chapter is organized as follows: The proposed LRFMP model is presented in Section 3.2. In Section 3.3, we present the methodology of the study. A case study in a grocery retailing company with the results and managerial implications is provided in Section 3.4. Finally, conclusions are in Section 3.5.

3.2. LRFMP Model

Classic LRFM models have mostly performed well in customer segmentation in many different industries (Hosseini et al. 2010; Wei et al. 2012; D.-C. Li et al. 2011; Kao et al. 2011). However, different from the application areas such as hotel, outfitter, dental clinic and hair salon, people tend to visit grocery stores more frequently, and this makes the level of variation in their visiting patterns important in understanding their behavior. According to LRFM model, it is possible that customers with similar profiles may have completely different visiting patterns. Thus, this study introduces periodicity of customers' visits (P) into original LRFM model to characterize customers' behaviors and measure the regularity of them. Furthermore, classical recency feature does not necessarily reflect the customer's actual repurchase or revisit tendency since it is merely based on the last transaction of the customer, thereby causing misleading results in customer segmentation studies. In order to deal with this problem, we modify recency variable in the original model by considering the last N transactions of the customer in its computation, rather than considering solely the last transaction.

The newly proposed feature periodicity (P), the modified feature recency (R), and the other features which are length (L), frequency (F), and monetary (M) are defined as follows:

Length: This feature is the time interval, in days, between the customer's the first and the last visits. It shows the customer loyalty, and the higher the length is, the more loyal a customer is.

Recency: Recency indicates how up-to-date the interaction of a customer with the company is, and gives information about the repeat purchase tendency. In the traditional LRFM model, recency is typically calculated as the time interval (typically in days) between the customer's last visit date and the last date of the observation period. We modify this variable in our model as the average number of days between the dates of the customer's N recent visits and the last date of the observation period. Thus, recency in our model is calculated by the following equation:

$$Recency(n) = \frac{1}{n}\sum_{i=1}^{n} date\_diff(t_{enddate}, t_{m-i+1}) \qquad (1)$$

where $date\_diff(t_{enddate}, t_{m-i+1})$ represents the difference in days between the end date of the observation period, $t_{enddate}$, and the date of (m-i+1)th visit of the customer, t_{m-i+1}; t_m is the last visit of the customer; and *n* is the number of the customer's recent visits considered. Note that for n=1, this newly defined recency variable turns into the traditional recency, and thus our modified recency feature also covers the classical one. In this study, we set *n* to 3, and compute recency for last three visits of customers.

The lower the recency value is, the more recent the customer's visits are and therefore the likelihood of repurchase or revisit is high for that customer.

Frequency: Frequency refers to the customer's total number of visits during the observation period. The higher the frequency is, the higher the customer loyalty becomes.

Monetary: Monetary refers to the average amount of money spent per visit by the customer during the observation period and reflects the contribution of the customer to the revenue of a company. A higher monetary value represents a greater contribution to the company.

Periodicity: This new feature reflects whether customers visit the stores regularly. We define periodicity as the standard deviation of the customer's inter-visit times:

$$Periodicity = stdev(IVT_1, IVT_2, \ldots, IVT_{\mathrm{n}-1}, IVT_{\mathrm{n}}) \qquad (2)$$

where *IVT* denotes the inter-visit time and n represents a customer's number of inter-visit time values. *IVT* is the elapsed time in days between two consecutive visits of the customer, and it is defined as follows:

$$IVT_{\mathrm{i}} = date\_diff(t_{i+1}, t_i) \qquad (3)$$

where $i \geq 1$ and t_i denotes the date corresponding to the i^{th} visit of the customer.

Periodicity indicates the tendency of a customer's visits to occur at regular intervals. If one customer has a low periodicity value, it means that this customer visits or makes purchases at relatively fixed intervals and can be characterized as regular.

In order to clarify and exemplify the shortcomings of the LRFM model and explain the rationale behind the inclusion of periodicity and modification of recency variable, let us consider the grocery store visiting histories of three hypothetical customers in Table 2. Note that each customer in this sample set has visited the store 7 times and their visits are sorted by date. Date columns indicate the visit dates of customers and inter-visit time values (in days) are calculated for each visit of each customer after their first visits. For instance, the inter-visit times for the second visits of the customers are measured in days by getting the difference between their second and first visit dates. Note that to compute length and recency variables, we assumed that the last date of the time period in the example is 30.04.2016. Time (in days) to the last date of the observation period is also given for last three visits of each customer to compute recency variables.

Table 2: Grocery store visiting histories of three hypothetical customers

Customer 1			Customer 2			Customer 3		
Date	IVT	Time to 30/04/16 for last N=3 visits	Date	IVT	Time to 30/04/16 for last N=3 visits	Date	IVT	Time to 30/04/16 for last N=3 visits
11.12.2015	NA		11.12.2015	NA		18.12.2015	NA	
30.12.2016	19		09.01.2016	29		07.01.2016	20	
17.01.2016	18		15.02.2016	37		16.01.2016	9	
02.02.2016	16		29.02.2016	14		10.02.2016	25	
27.02.2016	25	63	27.03.2016	27	34	16.02.2016	6	74
21.03.2016	23	40	13.04.2016	17	17	26.03.2016	39	35
16.04.2016	26	14	16.04.2016	3	14	23.04.2016	28	7

Based on the data in Table 2, length, recency, frequency and periodicity values are computed for each customer and provided in Table 3. There are two recency variables in Table 3. Recency(1) is the traditional recency variable that refers to the number of days between the customer's last visit date and the last date of the time period (e.g. 14 for Customer 1), and Recency(3) is our modified recency variable which is the average of number of days between the dates of the customer's three recent visits and the last date of the observation period (e.g. (63+40+14)/3 = 39 for Customer 1).

Table 3: Values of L, R(1), R(3), F, P variables for three hypothetical customers

	Customer 1	Customer 2	Customer 3
Length	127	127	127
Frequency	7	7	7
Recency(1)	14	14	7
Recency(3)	39	21.67	38.67
Periodicity	4.07	12.21	12.32

Assume that the monetary variable is the same for all customers. According to the traditional LRFM model, Customer 1 and Customer 2 have the same L, R(1), F, M values, and so they have identical characteristics. However, when we look at the inter-visit times of both customers in Table 2, two customers seem to have completely different visiting patterns. Customer 1's visits are more regular than Customer 2's visits over this period. Our new variable, periodicity catches such a difference as shown in Table 3 and periodicity values of both customers are quite different from each other. Customer 1's periodicity value is low which shows that he has a regular visiting pattern, whereas Customer 2 has irregular visit times with a higher periodicity value.

According to Table 3, traditional recency, Recency(1) is the same for both Customer 1 and Customer 2 and their Recency(1) values are higher than that of Customer 3. This indicates that Customer 1 and Customer 2 are more likely to revisit the store in the near future. However, if we consider time intervals between customers' last 3 visits to the last date of the observation period, the situation changes. As shown in Table 2, last 3 visits of Customer 1 and Customer 3 have higher time intervals (14, 40, 63 for Customer 1 and 7, 35, 74 for Customer 3) to the last date of the observation period than the ones of Customer 2 (14, 17, 34). Thus, according to values of our newly defined Recency(3) variable in Table 3, Customer 1 and Customer 3 have the same Recency(3) value and it is higher than the one of Customer 2. This means that Customer 2 have visited the store more recently than the other two customers, and the likelihood of a revisit is higher for this customer. This outcome is completely different from the one that is obtained by using the traditional Recency(1) variable. The reason is that the traditional recency variable ignores the influence of recent visits by considering only the last visit of the customer, and so provides poor information. However, our newly defined recency feature copes with this issue by considering the last N transactions of the customer in its computation.

The given example on three hypothetical customers in Table 2 shows that the periodicity of customers' visits represents their behavior characteristics, as well as the traditional L, R, F, M variables and modified recency feature provide more reasonable information, whether the customer is still active. The above example also shows how our proposed LRFMP model differs from the traditional LRFM model and also enables us to better understand the significance of periodicity and modified recency features to gain more useful insights about customer behavior especially in the grocery industry.

3.3. Methodology

This section presents the systematic methodology of this study for customer segmentation. It is based on an efficient combination of the LRFMP model and the clustering algorithms. Two well-known and commonly used clustering techniques namely K-means and Ward's clustering algorithms are tested to segment customers. This study uses actual customer data for LRFMP model features and assumes that all the model variables are of equal weights (i.e., equally important).

Note that LRFMP model features vary in range, and the scale differences among these attributes obviously influence and distort the results of clustering analysis (Mooi & Sarstedt 2010). Standardization solves this problem by reducing the potential effects of variable differences (Ketchen & Shook 1996; Milligan & Cooper 1988). Therefore, prior to clustering, the LRFMP variables are standardized by using the most widely used scaling technique, simple z standardization, which rescales each variable to have a mean of 0 and a standard deviation of 1.

Both K-means and Ward's method are applied and intraclass inertia is used to determine the most appropriate one of these clustering algorithms. This measure is the average squared Euclidean distance between each observation and its cluster mean, and it shows the compactness of clusters (Michaud 1997). Note that smaller value of this measure indicates a better clustering solution. Prior to applying determined clustering algorithm, finding an optimal number of clusters (k) is a critical issue. For this purpose, various indices are proposed in the literature and they differ in the way they quantify and combine compactness and separation concepts. In this study, total WSS (within sum of squares) is used as the cluster validity index and the elbow method (Thorndike 1953) is employed to determine the optimal number of clusters. In this method, clustering algorithm is executed for different number of clusters (k), and for each k, total WSS (within sum of squares) is calculated. Then, a curve between WSS and the number of clusters is drawn, and the appropriate number of clusters is determined using the location of a knee in the plot. After the proper number of clusters is decided, the results of clustering with the determined number of clusters are analyzed. Then, LRFMP scores are computed for each resulting customer group and based on LRFMP characteristics, customer segments are then profiled. Finally, managerial implications for each customer group are discussed for effective management of customer relationships and marketing strategies.

3.4. Case Study

3.4.1. Sample and Data Description

The grocery company investigated in this study is a local grocery chain that operates more than ten stores in Antalya, Turkey. The original dataset was extracted from this company's loyalty card system and it contains almost 2 million purchase transactions of 16024 customers during the period between October 1, 2012 and August 31, 2014. Within that period, several customers have visited the stores only one or two times. Since periodicity and recency for customers who have visited the stores a few times is considered useless, we excluded the customers who have visited the stores less than three times. A series of data pre-processing tasks including deleting transactions with missing values, removing duplicate records, and aggregating transaction records in the same day for each customer were also performed before analysis. Therefore, the final dataset is left with purchase records of 10471 customers.

In the dataset, each transaction record contains customer's membership number, purchase date, purchase item, item quantity purchased, item category, a set of item sub-categories, unit price. The features of LRFMP model were generated for each customer and the descriptive statistics regarding the maximum, minimum and average values of these attributes are presented in Table 4.

Table 4: The descriptive statistics of LRFMP variables

	Maximum	Minimum	Average	Standard Deviation
Length	699	2	419.81	218.89
Recency	696	1	218.59	195.97
Frequency	586	3	28.74	44.55
Monetary	422.45	0.45	36.76	27.35
Periodicity	454.67	0	43.41	50.88

3.4.2. Performance Comparison of Clustering Techniques

The results of intraclass inertia for each algorithm against varied k values are presented in Figure 1.

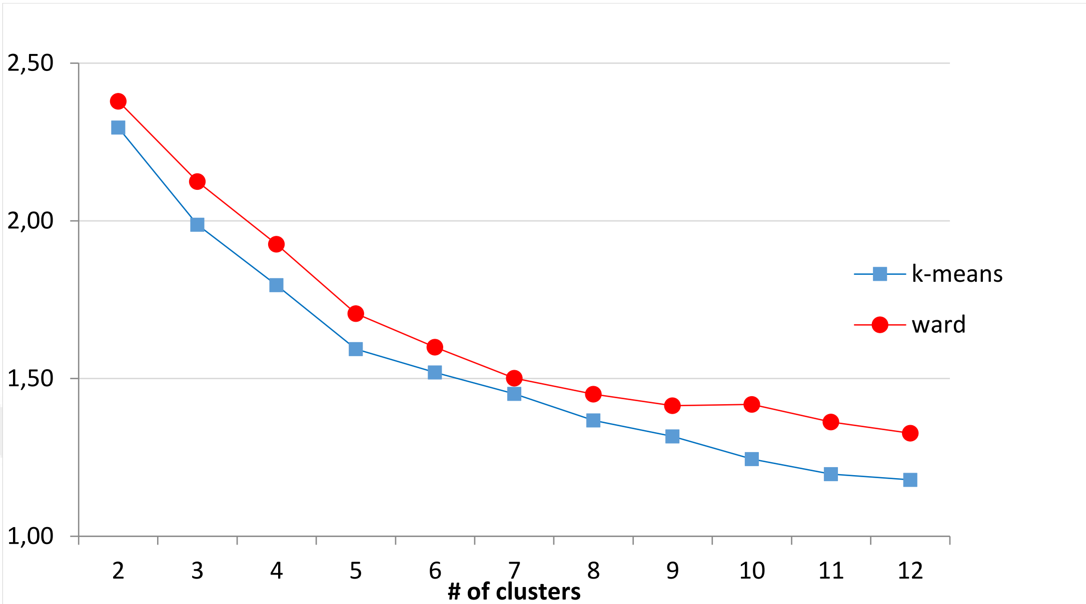

Figure 1: The Intraclass inertia for both clustering techniques

Note that lower values of this measure indicate better clustering performance. As shown in Figure 1, K-means algorithm provides superior results in terms of intraclass inertia than the Ward's method for this dataset. Therefore, we adapt the result of K-means algorithm to characterize customer groups.

3.4.3. Determination of Number of Clusters

Figure 2 shows WSS (within sum of squares) results of K-means algorithm against different number of clusters (k) ranging from 2 to 10.

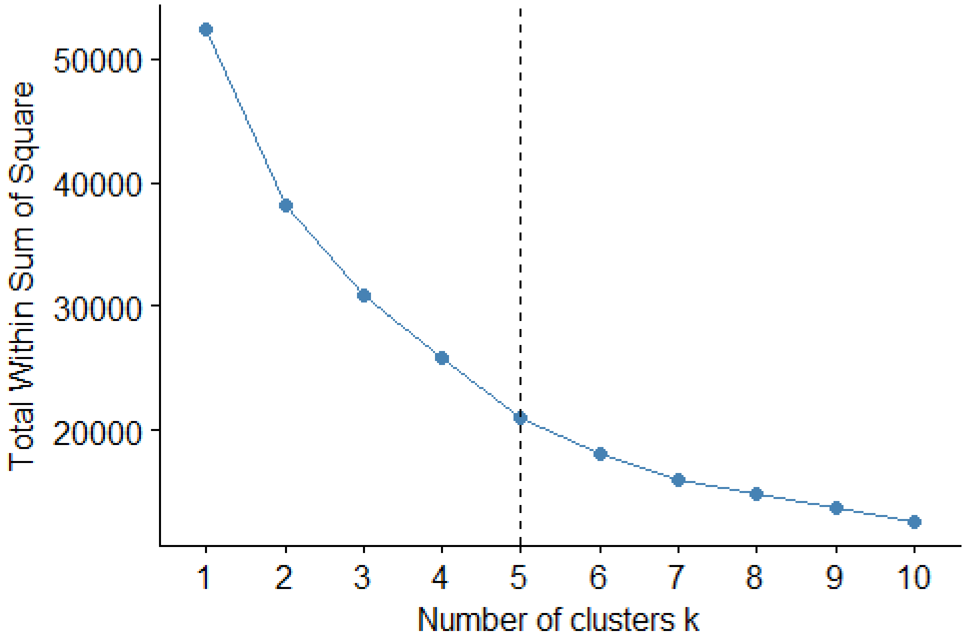

Figure 2: WSS values of the K-means clustering algorithm

The plot shown in Figure 2 indicates that there are bigger decreases in WSS value until k=5, and subsequent clustering with higher number of clusters does not show considerable decreases. Therefore, 5 was chosen as optimum number of clusters and we took the results of clustering with k = 5. Table 5 shows all the details for each cluster such as the sample size, the average values of LRFMP attributes and LRFMP scores. In forming the last column, LRFMP scores, we implemented the technique as suggested by Ha and Park (1998). This technique puts the up symbol (↑), if average L, R, F, M or P value of the cluster is greater than the aggregate average; otherwise uses the down symbol (↓).

Table 5: Clustering results by K-means method

Cluster	Sample Size	Average L	Average R	Average F	Average M	Average P	LRFMP Scores
1	538	633.29	39.67	175.24	24.32	4.99	L↑ R↓ F↑ M↓ P↓
2	4681	564.50	90.19	33.44	31.73	31.49	L↑ R↓ F↑ M↓ P↓
3	1091	482.17	301.18	5.33	34.21	159.32	L↑ R↑ F↓ M↓ P↑
4	818	374.01	220.24	11.85	104.18	45.81	L↓ R↑ F↓ M↑ P↑
5	3343	173.70	399.79	10.34	30.14	27.85	L↓ R↑ F↓ M↓ P↓
Average		419.81	218.59	28.74	36.76	43.41	

3.4.4. Group Profiling and Managerial Implications

In this section, we create descriptive profiles for five groups of customers with different characteristics based on the results of clustering analysis reported in the previous section. Then, various customer relationship management and marketing implications including example industry-specific promotion strategies for each customer segment are discussed.

In order to interpret the results of customer segmentation based on RFM model, Marcus (1998) proposed customer value matrix that is based on frequency (F) and monetary (M) variables and formed four major customer types, including best customers (F↑M↑), spender customers (F↓M↑), uncertain customers (F↓M↓), and frequent customers (F↑M↓). Moreover, Chang and Tsay (2004) created a customer relationship matrix for positioning of customer clusters by using length (L) and the recency (R) variables and proposed four relationship types: close relationship (L↑R↓), potential relationship (L↑R↑), lost relationship (L↓R↑), and new relationship (L↓R↓).

By incorporating periodicity (P) into LRFM model and taking the aforementioned studies as the basis, the clusters determined in Table 5 can be named as shown in Table 6.

Table 6: Group Profiles

Group	Name	LRFMP Scores	Size (%)
1	High-contribution loyal customers	L↑ R↓ F↑ M↓ P↓	5.14
2	Low-contribution loyal customers	L↑ R↓ F↑ M↓ P↓	44.70
3	Uncertain customers	L↑ R↑ F↓ M↓ P↑	10.42
4	High-spending lost customers	L↓ R↑ F↓ M↑ P↑	7.81
5	Low-spending lost customers	L↓ R↑ F↓ M↓ P↓	31.93

Based on the LRFMP scores, Clusters 1 and 2 have L and F values greater than the average, and R and P values lower than the average. These indicate that the customers in both clusters have a close and long-term relationship with the company according to their recent and frequent visits. Thus, the customers in these clusters can be considered as loyal customers. However, the two groups differ in their visiting patterns. P value of the customers in Cluster 1 is much smaller than that of customers in Cluster 2, and the visiting frequency of the customers in Cluster 1 is much greater than that of customers in Cluster 2. Additionally, the customers in Cluster 1 visit the stores more recently, frequently and regularly than the customers in the other clusters. Although Cluster 1 is the smallest segment (5.14%) among all groups, the customers in this group have the highest total expenditure (TRY 4261) during the selected timespan, and their contributions are also higher than the total contributions of other groups. Hence they can be defined as “high-contribution loyal customers”. On the other hand, the customers in Cluster 2 are considered as “low-contribution loyal customers” whose frequency behavior could be further improved.

A grocery retailer should maintain frequent and closer relationships with customers like ones in Cluster 1 who generate the highest benefits for the company to increase its business profit. It is essential to understand shopping habits of these customers and to track purchased items by using point-of-purchase data acquired through store loyalty cards. By using this information, personalized promotions or discounts can be provided for these profitable customers – for example, offering promotions regarding a product of a specific brand only to those customers who purchased this item in one of their earlier visits, or applying a discount on a specific item to those customers who have frequently purchased it previously. Moreover, such regular customers might expect to benefit from special treatment (Wong & Sohal 2002). This can be achieved with offering exclusive services and tangible rewards for them. For instance, providing loyal customers preferred parking spots and giving them priority checkout in large stores can strengthen their emotional bond to the company. Free gifts on special days (e.g. birthdays) might also make these customers feel valued, thereby building long-term customer loyalty.

Customers in Cluster 1 have a low average spending per visit. To increase the total amount spent per visit, conditional promotions can be used. In such promotions, customers have to meet a certain condition to avail them of the discount (Grewal et al. 2011). Two common examples of conditional promotions include multi-purchase discounts or price promotions regarding the shopping amount - for example, buy three items for two or 10% off for spending of TRY 100 or more. Such promotions are effective tools for enticing customers to purchase and spend more in their shopping visits. On the other hand, the customers in Cluster 2 have a higher average amount of spending per visit, although they visit the stores of the company less frequently and regularly. To raise the visit frequency of such customers, a grocery retailer can launch reward programs with a loyalty card system, by which the customer is rewarded with bonus points to redeem for purchases in the retailer's stores (Grewal et al. 2011) - for example, TRY 30 bonus points for 3 purchases; each of which equals TRY 100 or more within a certain period of time, or TRY 50 bonus points for spending TRY 500 or more within a certain period of time. Promotions such as coupons and price-offs, which are offered to be used in the next visits, can also be used for this purpose - for example, 10% off coupon on the next purchase for spending TRY 100. Such rewarding programs and promotions could necessarily encourage customers to visit more frequently and regularly, and to purchase more. Beyond making the purchase more engaging and rewarding, it is also needed to put a strong emphasis on implementing aforementioned customer relationship management and marketing strategies for the customers in Cluster 2 to maintain and increase their loyalty.

For Cluster 3, the average L and R values are greater than the aggregate averages. This indicates that they have been the customers of the company for a long time, but they have not visited the stores recently. Thus, the customers in this cluster can be classified as "uncertain customers". This cluster has the lowest F and highest P values. These attributes indicate that they do not visit the stores frequently and their visits are irregular. Such a low frequency and periodicity can be a warning sign that these customers are likely to leave the company. Despite these, this group has the second largest average spending per visit. Therefore, with their long-term relationships with the company and high spending per visit, they have the potential to be the loyal customers on a long term basis, provided that a company should pay more attention to ways of drawing such customers to its stores. Such uncertain customers could be more focused on prices. Hence, in order to attract this kind of customers to visit more often and more regularly, weekly flyers including recent promotions and store-wide discounts on a wide range of products can be sent to them.

Clusters 4 and 5 have average L and F values lower than the aggregate averages, and average R value greater than the aggregate average. This suggests that the customers in these clusters are lost customers. One major difference between these two clusters is that the customers in Cluster 4 spend extremely higher average amount of money per visit. Moreover, the customers in Cluster 4 spent more money on each visit than the other groups, and this group includes the most profitable customers who make the most contribution after the loyal groups. Thus, winning back such lost customers and converting them into loyal ones are crucial for a company to improve its profitability. To bring back these lost customers, it is important to find out exactly why they left the company. For this purpose, a grocery retailer can benefit from the loyalty card database including customer transactions and dig into customers' purchasing behaviors to infer reasons about why these customers have left. Besides, a company can ask these customers what they would need to buy from the company again. By understanding their shopping patterns and considering their feedback, the corresponding adjustments and strategies can be made for such lost customers - for example, sending recent promotions or discounts to the customers who were not satisfied because of price, or creating new parking lots around a store and informing the customers who were discouraged due to the limited parking area of that store. Therefore, such incentives tend to be more valued and appreciated by such kind of customers, which may in turn boost their willingness to buy from the company again. On the other hand, there is no immediate need to encourage lost customers like ones in Cluster 5 to come back, because they have not bought from the company's stores for a relatively long time and they also have fairly low shopping frequency and low average spending per visit. Because of these, such type of customers has little potential to become loyal and thus a company can exclude such least contributing customers from its promotional campaigns and advertising activities to reduce marketing expenditures.

3.5. Conclusion

It is highly important especially for grocery companies operating in an intensely competitive environment to adjust their marketing strategies and maintain a good relationship with their customers. To deal with these issues, it is essential to effectively classify customers into different groups and understand the characteristics of these distinct groups. In this manner, this chapter proposes an augmented RFM model called LRFMP to gain deeper and reasonable insights about customers and presents an application of this model for the segmentation of customers in the grocery retail industry. LRFMP model is developed by adding the periodicity (P) variable to the traditional LRFM model and redefining the classical recency feature. For our case study, the real-life data from a grocery chain operating in Turkey is used. K-means and Ward's clustering techniques are applied individually to assess their suitability for customer segmentation. The results show that, in this case, K-means clustering technique is more appropriate for segmenting customers. Therefore, customers are grouped into five different segments, and based on the LRFMP model scores, these five groups of customers are then profiled as: "high-contribution loyal customers“, "low-contribution loyal customers“, "uncertain customers“, "high-spending lost customers" and "low-spending lost customers".

CHAPTER 4

A METHODOLOGY FOR PRODUCT SEGMENTATION USING SALE TRANSACTIONS

This chapter presents a novel methodology for product segmentation in the grocery retail industry using customers' purchase transactions.

4.1. Introduction

Clustering which forms meaningful groups by accumulating data objects sharing similar characteristics is one of the most commonly used data mining techniques to extract useful knowledge from large databases. Among the application areas of clustering, customer segmentation is the popular one for discovering useful commercial knowledge in today's competitive environment. Customer segmentation divides entire customers into homogeneous groups based on their behaviors and characteristics utilizing companies' transactional data. Due to its ability to improve decision-making, it has been successfully applied in a variety of industries, e.g., health and beauty (Khajvand et al. 2011), textile (D.-C. Li et al. 2011), healthcare (Wei et al. 2012), hairdressing (Wei et al. 2013), banking (Nikumanesh & Albadvi 2014), and tourism (Dursun & Caber 2016). Customer segmentation enables enterprises to identify different groups of customers, such as loyal, profitable, new, lost, and to understand the needs and characteristics of these customer groups separately. With this knowledge, then they can develop unique customer relationships and marketing strategies for each customer segment (Dibb & Simkin 1997). Hence, companies can improve the quality of the service, retain loyal customers, and increase their profits.

Despite the benefits of customer segmentation and its wide applicability, there has been no study attempting to segment products using transactional data. Nowadays, most retail companies present a tremendous diversity of products to customers, and so product categorization plays a vital role for such enterprises to reduce the complexity in many organizational tasks. Product taxonomy is widely used by grocers, and products are categorized according to their individual characteristics by domain experts. Some category examples from grocery shopping include fruit/vegetable, beverage, breakfast food, and detergent/cleaning. Such a taxonomy can help the grocer in managing products, but the products under the same category may not show a similar pattern in terms of sale transactions. In this manner, product segmentation based on sale transactions can be helpful for managers to discover various product groups with different characteristics. Therefore, such a solution allows supermarkets to obtain more meaningful and homogenous product segments, which can provide more efficient development of marketing and inventory strategies.

Based on the above motivation, in this chapter, we propose a methodology which utilizes sales transactions to cluster products in grocery retailing. For this purpose, this study adapts the two features of well-known RFM model for customer segmentation, and introduces a new feature in order to characterize products more meaningfully and to identify product groups more efficiently. Based on these features, both hierarchical (Ward's method) and partitional (K-means) clustering techniques are tested to divide products into groups. The proposed approach is applied to real-life data from a grocery chain operating in Turkey. The resulting product groups are investigated in terms of determined features to identify their characteristics.

The rest of this chapter is organized as follows. Section 4.2 describes proposed methodology. In Section 4.3, a case study of a grocery retailing company with the results is presented. Finally, Section 4.4 concludes the study.

4.2. Methodology

This study uses sales transaction data and unsupervised learning method to identify different product segments and to obtain valuable insights about these distinct groups. In order to perform product segmentation, the first task is to decide which features are to be used for clustering. The variables of RFM model, which are recency, frequency and monetary, are well known and the most widely used ones for customer segmentation (Bauer 1988; Bult & Wansbeek 1995). Such types of features can easily be extracted from the transactional data of companies. Due to its simplicity, several researchers and practitioners have applied different versions of RFM models for clustering customers, and have achieved useful results.

Based on the aforementioned efficiency of RFM model in customer segmentation, this study incorporates the features of RFM model into product segmentation. In typical RFM model, recency is the time interval since the customer's last purchase (e.g., days or months) and reflects buying or visiting potential of the customer. In customer segmentation, customers' relationship with the company is investigated, and this is one-to-one relationship. However, in product segmentation, the product's characteristics is examined in an aggregated manner by considering whole customers' transactions on that product. Hence, this feature is mostly taking zero value for products. Because of these, recency feature is not proper for product segmentation based on transactional data. Therefore, this research introduces a new model called FMC for product segmentation by incorporating customer variety (C) feature into traditional RFM model instead of recency (R).

In our FMC model, frequency (F) is the total number of times the product has been purchased by the customers in a certain period of time. A product with a high frequency score indicates that it is one of the best-selling products of the company. Monetary (M) is defined as the total amount of money spent for that product by the customers in a certain period of time. The products having high monetary score contribute higher revenue to the business. Finally, customer variety (C) denotes the total number of unique customers who purchased the product in a certain period of time. A product with a high customer variety score means that it is one of the products that is demanded by many customers. In summary, the products having higher scores in terms of these three features are the high value products for enterprises.

In companies' databases, such features like the ones in FMC model are hidden in the transactional data. Based on these transactional data, the FMC scores of each product are computed. After this process, each product has its own frequency, monetary, and customer variety values to identify product's selling characteristic. Since the features vary in range, they are standardized by using min-max normalization (standardization) between 0 and 1 prior to clustering. Two well-known and commonly used clustering techniques namely K-means and Ward's method are applied and intraclass inertia is used to determine the most appropriate one of these clustering algorithms.

After identifying the best clustering algorithm for product segmentation, it is needed to determine the suitable number of cluster (k). For this purpose, we adopted the elbow method. After the proper number of clusters is decided, the results of clustering with the determined number of clusters are analyzed. Finally, FMC scores are computed for each resulting product group and based on FMC characteristics, product segments are then profiled.

4.3. Case Study

4.3.1. Sample and Data Description

The data used in this study was extracted from the loyalty card system of a grocery chain and contains purchase transactions between October 1, 2012 and August 31, 2014. In the dataset, there are many transactions pertaining to the customers who have visited the store irregularly (only a few times during the period). To prevent possible misleading results of such customers, we excluded customers who have visited the store less than 10 times.

In the dataset, product feature includes information about the product brand and amount. For example, the product, "XYZ Ketchup, 600 Grams" in which "XYZ" is the brand of product and "600 Grams" is the amount of it. Product category for this product is "ketchup". Since the product features is highly specific, we considered product categories in this study. Examples of product categories include: milk, apple, sugar, toothpaste, waffle, pasta, etc. There are unique 414 third-level product categories in the dataset.

The features of FMC model were generated for each product category and the descriptive statistics regarding the maximum, minimum and average values of these attributes are presented in Table 7.

Table 7: The descriptive statistics of FMC variables

	Maximum	Minimum	Average	Stand.Dev.
Frequency	38051	2	1166.48	3404.556
Monetary	316114.77	7.29	7419.55	27809.76
Customer Variety	1433	1	128.72	225.93

4.3.2. Performance Comparison of Clustering Techniques

The results of intraclass inertia for each clustering algorithm against varied k values are presented in Figure 3.

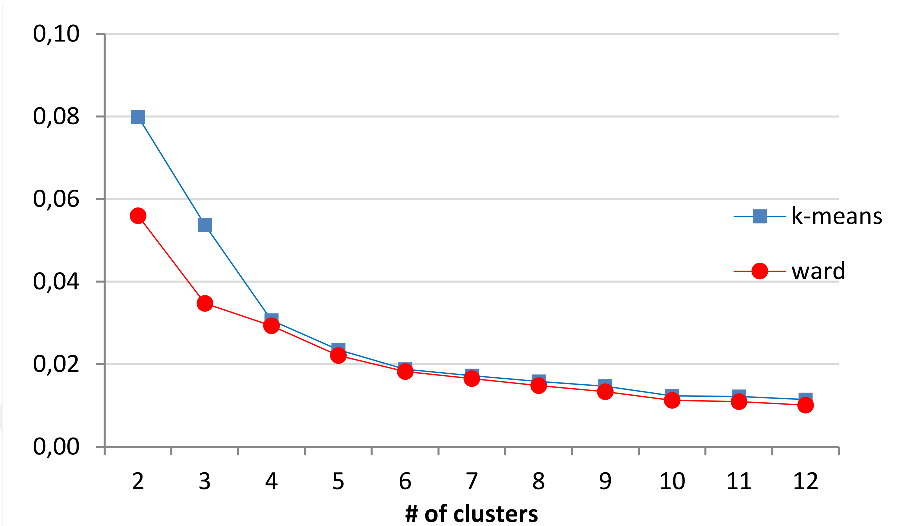

Figure 3: The Intraclass inertia for both clustering techniques

Note that lower values of this measure indicate better clustering performance. As shown in Figure 3, Ward's clustering method provides superior results in terms of intraclass inertia than the K-means algorithm for this dataset. Therefore, we adapt the result of Ward's method to characterize product groups.

4.3.3. Determination of Number of Clusters

Figure 4 shows WSS (within sum of squares) results of Ward's method against different number of clusters (k).

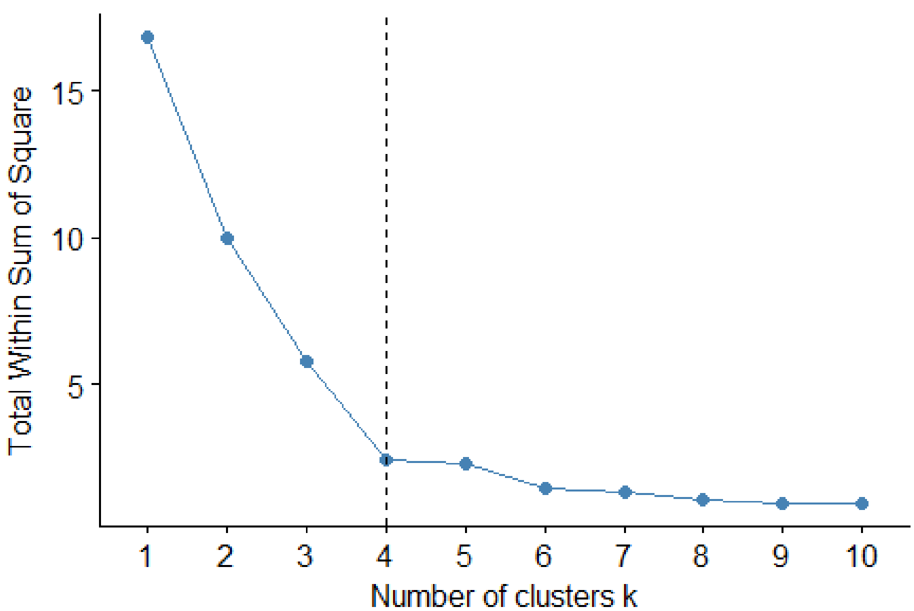

Figure 4: WSS values of the Ward's clustering method

Note that when the number of clusters increases, WSS decreases. As seen in Figure 4, there are significant decreases in WSS value, as increasing the value of k from 2 to 4. For the number of clusters bigger than 4, there is a very slow change in the value of WSS, and its value becomes constant for the higher number of clusters. Therefore, 4 is the elbow point, and we take the results of clustering with *k*=4 to characterize product groups.

4.3.4. Characteristics of Product Segments

We have 4 groups of products with different characteristics and the dendrogram obtained by Ward’s method is shown in Figure 5.

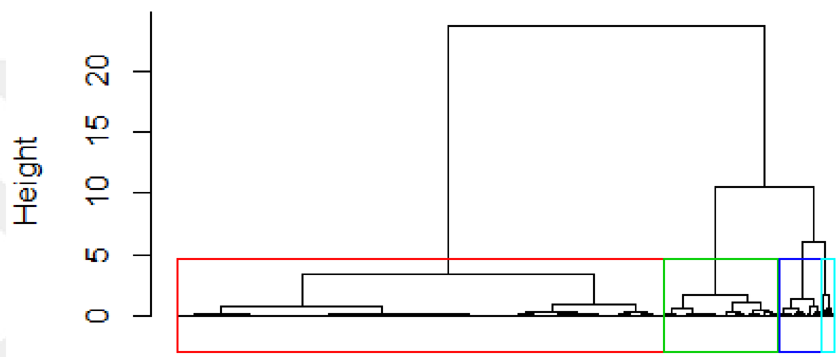

Figure 5: Dendrogram for Ward's method

We can also see from Figure 5 that the best choice for total number of clusters is 4. Table 8 shows all the details for each cluster such as the sample size, the average values of FMC attributes and FMC scores. In forming the last column, FMC scores, we implemented the technique as suggested by Ha and Park (1998). This technique puts the up symbol (↑), if average F, M, or C value of the cluster is greater than the aggregate average; otherwise uses the down symbol (↓).

Table 8: Results of product segmentation based on FMC features

Cluster	Sample Size	Average F	Average M	Average C	FMC Scores
1	8	21753.50	165814.94	1123.87	F↑ M↑ C↑
2	72	1895.87	12945.64	236.73	F↑ M↑ C↑
3	307	184.08	1296.10	30.88	F↓ M↓ C↓
4	27	4291.77	15377.26	658.33	F↑ M↑ C↑
Average		1166.48	7419.55	128.72	

Moreover, Table 9 shows sample products from each segment.

Table 9: Sample products in each segment

Segment	Samples
1	Bread, yogurt, milk, egg, chicken, beef, chocolate, cigarette.
2	Biscuits, liquid oil, butter, meat delicatessen, coffee, tea, sugar, flour, rice, shampoo, paper products (toilet paper, tissue, etc.), etc.
3	Legumes, salt, detergents, cleaning products, personal and baby care products, etc.
4	Non-seasonal vegetables and fruits (e.g. tomato, pepper, potato, onion, apple, banana, etc.), white cheese, pasta, beverages, etc.

As shown in Table 9, cluster 3 has lower F, M and C values compared to average F, M and C values, whereas for clusters 1, 2 and 4, values of F, M and C are above average F, M and C values. However clusters 1, 2 and 4 have identical FMC scores, the magnitude of these features are different for each cluster. In this manner, the values of F, M, C features are the highest for cluster 1 and it is followed by cluster 2 and 4. Products in segment 1 which are bread, yogurt, milk, egg, chicken, beef, chocolate, and cigarette are basic daily necessities of the general population regardless of gender, age, marital status, etc. As a result, they are the best-selling products of the company and make the most contribution to company's revenue at the same time. Thus, for these products, there is no need to provide any promotions or campaigns. However, it is important to effectively manage the inventory for this product group since the consumption rate for products in this segment is considerably high.

On the other hand, products in segment 3 which are nearly 75% of all products are least sold and least contributing products to the company's revenue in terms of FMC scores. Legumes, salt, detergents, cleaning products, personal and baby care products are some of the products placed in this segment. Such type of products has a lower consumption rate, and a long shelf life. Customers usually pursuit promotions or campaigns to buy such products, especially the expensive ones, such as detergents, cleaning products, etc. In that manner, the effect of marketing strategies to these products might become potentially greater than the effect to others.

As seen in Table 9, products may place in different groups, different from their categories formed according to their intrinsic characteristics by domain experts. As an example, "dairy products" is a well-known product category in the supermarket shopping and includes basically the following products: milk, cheese, yogurt, butter, etc. With our proposed methodology for product segmentation, these products are classified into different clusters. According to our clustering results, milk and yogurt are in cluster 1, butter places in cluster 2, and cheese is in cluster 4. Therefore, we can conclude that although products are similar to each other in terms of their intrinsic characteristics, they may have different sale characteristics, so they may be grouped into different segments.

4.4. Conclusion

In this chapter, a novel product segmentation methodology for grocery retail industry is presented. The proposed approach takes customers' transactions on products as input and employs clustering algorithms to segment products. Additionally, this study revises the RFM model used in customer segmentation to be the FMC model for product segmentation. The proposed approach was applied to a supermarket chain. For this case study, the results indicated that Ward's method performs better than K-means algorithm in segmenting products. Further, the obtained product groups showed that products placing in the same category in terms of their intrinsic characteristics may be grouped into different segments according to their sale characteristics.

CHAPTER 5

A HYBRID APPROACH FOR PREDICTING CUSTOMERS' INDIVIDUAL PURCHASE BEHAVIOR

In this chapter, we propose a hybrid approach which predicts customers' individual purchase behaviors and reduces the limitations of the individual-level and the segment-based predictive modeling approaches by combining the advantages of them.

5.1. Introduction

As stated in chapter 1, grocery companies operate in a competitive environment in today's business. Such an environment puts pressure on companies to keep their existing customers and build long-term customer relationship. In this context, customized services and one-to-one marketing actions play an important role for companies to improve the customer satisfaction and loyalty. In order to develop personalized services and marketing strategies effectively, it is required to understand and predict customers' individual buying behaviors and characteristics (Chen et al. 2005; Ha 2007; Kim et al. 2003). Once the customers' behaviors are predicted accurately, firms can integrate this knowledge into their customized services and marketing actions which is designed for satisfying and keeping customers.

In the literature, the individual-level and the segment-based approaches are two major predictive modeling approaches for customer behavior. Individual-level approach constructs a prediction model for each customer by only considering the transactional data of the individual customer. Differing from the former, in the segment-based approach, the predictive models of customer behavior are built for each customer group based on data pertaining to customers of the segment. Although both types of approaches can be used to estimate customers' individual behaviors, they both have some weaknesses. The primary drawback of segment-based approaches is the high levels of transaction heterogeneity which cause the performance of prediction accuracy to drop significantly (Hand et al. 2001). This approach utilizes all the past behavior data of the segment's customers to infer the customer's behavior in an aggregated manner. Thus, due to large variety of transactions of different customers in the segment, it misses unique behavioral patterns of customers, and may be ineffective in predicting the customer's individual behavior accurately. Therefore, the low levels of prediction accuracy cause a serious problem for segment-based approaches by reducing the quality of generated predictions.

Through utilizing a customer's previous transaction records only, individual-level approach usually shows better performance compared to the segment-based one. However, individual-level approach suffers from the data sparsity problem which occurs when prior transactional data is insufficient for generating predictions (Hand et al. 2001). There exist a large number of products and customers in the supermarket shopping domain, but the recorded transactions pertaining to many customer-product pairs are extremely scarce. For such samples with very few records, it is difficult and unreliable to build a prediction model by using individual-level approach. Similarly, the cold-start problem as a special instance of the sparsity problem occurs in individual-level approach and it is the lack of previous transactional records that makes it impossible to build a predictive model. Individual-level approach is not able to generate a prediction for a product which has not been purchased before by the customer. Consequently, such data sparsity problems limit the prediction coverage of individual-level approaches by making many of customer-item pairs hardly predictable, and thus this creates a major negative impact on the effectiveness of individual-level approaches.

As mentioned above, segment-based and individual-level approaches are complementary and can be used together. Although several studies have used these approaches separately to build prediction models for customers' behaviors, as far as we know, there is no study that has combined these two approaches to avoid their aforementioned limitations. Additionally, previous studies have evaluated the performance of these approaches by measuring accuracy-based metrics, but did not consider the prediction coverage metric which is also an important measure for the performance of customer behavior modeling approaches. In this study, we contribute to filling these gaps through proposing a hybrid approach of combining segment-based and individual-level approaches, which provides better performance by overcoming the drawbacks of these approaches, and empirically comparing the proposed approach against segment-based and individual-level approaches in terms of both prediction accuracy and coverage.

To overcome these drawbacks of two existing approaches, this chapter proposes a novel hybrid approach to predict individual purchase behaviors of customers. This approach combines individual-level and segment-based predictive approaches to exploit the advantages of both. It takes customers' historical data as input, and employs features regarding customer purchase behavior and predictive algorithms to make prediction. To use segment-based approach, customer segments are formed based on three novel features regarding customers' purchasing behavior by using two-stage cluster analysis method. The proposed approach is evaluated on a real-life dataset by conducting a series of experiments. Five most widely used and well-known machine learning classifiers – logistic regression, decision trees, support vector machines, neural networks, and random forests – and a baseline classifier are applied to generate predictions.

The performance is compared to the performances of individual-level and segment-based approaches in terms of prediction accuracy and coverage. The performance of machine learning techniques in predicting customers' purchase behavior is also examined. Finally, the results of all experiments are reported and findings are discussed.

The remainder of the chapter is organized as follows: Section 5.2 provides the formulation for the problem we study. In Section 5.3, we present our hybrid approach. The experimental evaluation and results are given in Section 5.4, and finally, the conclusions drawn are stated in Section 5.5.

5.2. Problem Formulation

Let C be the set of customers, P be the set of products in the product catalogue and T be the transactions of these customers for that product set. A customer $c \in C$ can purchase several products and let $P_c \subseteq P$ is the set of products have been already purchased by the customer c. Then, $T_{c,p}$ refers to the transactions of the customer c $\in C$ for the product $p \in P_c$.

Given this notation, our goal is to determine whether a customer c is going to purchase a particular product $p \in P_c$ in a transaction $t \in T_{c,p}$. Then, we can formulate this problem as a binary classification problem where each transaction t is divided into one of two classes: "purchase" or "no purchase".

5.3. Proposed Hybrid Approach

The basic idea of the proposed approach is illustrated in Figure 6. It can be seen that the past purchase transactions and segment information of customers are used as input in the proposed approach. The purchase transactions of customers are stored in the database, but it is needed to cluster customers into groups by using a customer segmentation technique. The customer segmentation approach used in this study is described in detail in the Section 5.3.4.

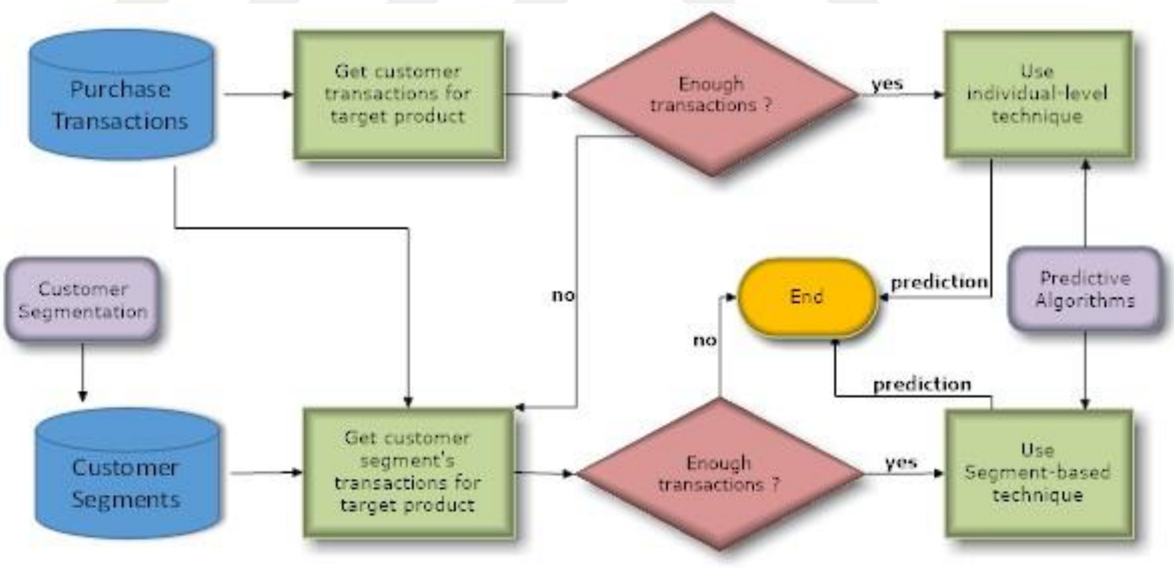

Figure 6: Overview of the proposed hybrid approach

Our hybrid approach combines both individual-level and segment-based techniques by utilizing a switching strategy which is one of the well-known techniques in the context of hybrid recommender systems (Burke 2002). It uses individual-level approach as the primary method, since it is superior to segment-based approach in terms of predicting performance. To build reliable models and improve the prediction accuracy, both techniques are employed only if a particular customer or customer segment has at least five purchase transactions with values of all predictive features for a specific product. In this manner, our hybrid approach first checks whether enough transactional data exist for the target customer-product pair to use individual-level predictive technique. If exists, then our approach triggers individual-level approach to generate prediction. On the other hand, if the target product has not been previously purchased by the customer yet or if previous transactional data is insufficient for that customer-product pair, then the proposed approach switches to segment-based predictive technique and checks for enough data. If the customer's segment has a sufficient number of historical transactional data for the target product, then our hybrid approach utilizes segment-based technique to produce prediction. Note that some products may be rarely purchased. Thus, segment-based technique is not able to generate predictions for such products because of the lack of past data.

The proposed hybrid approach mostly deals with the data sparsity problems of individual-level technique and can produce predictions in cases that individual customer has a few or no historical data for a particular product by putting the segment-based technique into use. Hence, this approach has a potential to increase the overall prediction coverage especially compared to individual-level technique. Additionally, it utilizes the individual-level technique, when there is enough data to build meaningful predictive models. Therefore, it is also expected to maintain a high prediction accuracy like the individual-level technique compared to segment-based technique.

In the subsequent subsections, we will give the details about, customer behavior modeling approaches, predictive features, predictive algorithms, and customer segmentation technique, which are utilized by the proposed approach, respectively.

5.3.1. Predictive modeling approaches for customer behavior

In this study, the predictive models of customer behavior are learnt in the following form:

$$y = f(x_1, x_2, \dots, x_n) \tag{4}$$

where ($x_1,x_2,. . .,x_n$) are all the features which affect and describe a customer's purchase behavior, function f is a predictive model that can be learned via predictive algorithms and $y \in \{0, 1\}$ is the predicted label that indicates whether a particular product p_i is purchased in a particular transaction t.

5.3.1.1. Individual-level Approach

Our individual-level approach builds a predictive model for each customer-product pair using past purchase transactions on a particular product of a particular customer and predictive algorithms. In order to improve the accuracy of the generated predictions, our approach builds a predictive model only if a particular customer has at least five purchase transactions with values of all predictive features for a specific product.

5.3.1.2. Segment-based Aprroach

In this approach, predictive models of customer behavior are built for each <customer segment-product> pair utilizing the entire historical behavior data on the target product belonging to the customers of the target segment in an aggregated manner. Therefore, segment-based approach generates predictions for an individual customer based on the buying behaviors of other customers in the same segment by employing predictive algorithms. Like in our individual-level approach to build reliable models, this approach also builds predictive models for only customer segment-product pairs having at least five past purchase transactions with values of all predictive features.

5.3.2. Predictive Features

Each transaction can be defined by a set of features, which are the predictors for a predictive model. In this study, we identify a set of features capturing purchase behavior of customers. Table 10 shows the features considered with their types and definitions.

Table 10: Predictive Features

Features	**Type**	**Definition**
day	Multi-nominal	Whether the purchase is made on a weekend or a week day (0: weekend; 1: weekday)
time	Multi-nominal	The timespan of the purchase (1: morning, 2: afternoon, 3: evening, 4: night)
purc_interval_time	Numeric	The time elapses in days since the last shopping occasion.
prev_tot_num_products	Numeric	Total number of products purchased in the previous shopping occasion.
last_N_avg_number_products	Numeric	Average number of products purchased in last N shopping occasions.
prev_tot_money_spent	Numeric	Total amount of money spent in the previous shopping occasion.
prod_purc_interval_time	Numeric	The time elapses in days since the last purchase of the product
last_N_num_of_times	Numeric	Number of times the product purchased in last N shopping occasions.
prev_purchase	Binary-nominal	Whether the product was purchased in the previous shopping occasion. (0: not purchased; 1: purchased)

As shown in Table 10 , we represent two temporal features, “day” and “time” as categorical. Accordingly, we categorized “time” feature in this study as follows: values between 08:00 to 11:59 as “morning”, values between “12:00 to 17:59” as afternoon, values between “18:00 to 20:59” as evening, and values between “21:00 to 22:59” as night. Moreover, in this study, we selected N as 5 for the “last_N_avg_number_products” and “last_N_num_of_times” features.

5.3.3. Predictive Algorithms

The proposed hybrid approach incorporates the advantages of the individual-level approach and the segment-based approach and these approaches employ predictive algorithms to build predictive models. A predictive algorithm executes based on the previous transactional data to predict the behavior of the customer and can be any type or ensemble of machine learning classifiers.

5.3.4. Customer Segmentation

As stated before, segment-based approach builds a predictive model for each customer segment. In this manner, meaningful customer groups are needed for this approach, and customer segmentation is performed to create these customer groups. The objective of customer segmentation is to identify groups of customers that have similar purchasing behavior characteristics. In order to perform customer segmentation, we firstly need to decide which features are to be used for clustering. There are several features used in the literature, and among them, RFM model features are well known and the most widely used for understanding and analyzing customer behavior characteristics (Kahan 1998; Newell 1997; Hughes 1996). RFM model relies on three attributes which are recency (R), frequency (F), and monetary (M), and these features measure the recency of customer purchasing behavior, the frequency of purchasing, and the total or average monetary expenditure on purchasing, respectively.

These three indicators relate to customer loyalty and profitability. Hence, in the recent years, different combinations of such features have been successfully applied for the purpose of targeting valuable customers in a variety of industries, e.g., hairdressing (Wei et al. 2013), banking (Nikumanesh & Albadvi 2014), manufacturing (Güçdemir & Selim 2015), telecom (Zabkowski 2016), and grocery retail (Peker et al. 2017). However, in this research, rather than customer value assessment, customer segmentation is applied for identifying customer groups having different purchasing characteristics. Therefore, instead of using RFM model features, we propose three novel segmentation variables related to customers' purchasing behavior namely "average inter-purchase time", "average basket size" and "product variety". Average inter-purchase time refers to the average of customer's inter purchase-times which is the elapsed time in days between two consecutive purchases. Average basket size is the average number of product categories purchased per shopping by the customer. Finally, product variety refers to total number of unique product categories purchased by the customer during his length of relationship with the company.

In this study, we use features describing customers' purchasing behavior and unsupervised learning methods to segment customers. The process is described step by step in Figure 7. It can be seen that the three steps are the following: (1) extracting the identified features for each customer from purchase transactions, (2) deciding on the number of clusters, (3) forming customer groups.

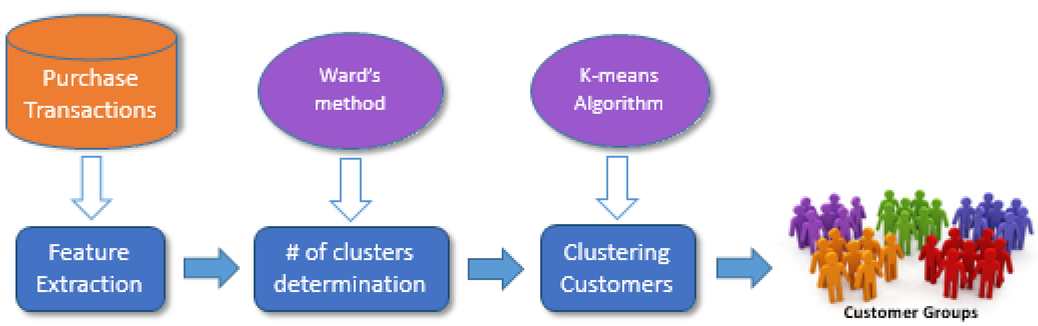

Figure 7: The flowchart of the customer segmentation approach used in this study

To group customers, we employ a two-stage cluster analysis suggested by (Punj & Stewart 1983). In this approach, Ward's hierarchical method is first used to determine the number of clusters (k) and K-means clustering method is then performed to separate the data into k groups. Ward's minimum variance method constructs a dendrogram which provides a better understanding of the data. That way, we decide the optimum number of clusters, and subsequently use K-means clustering algorithm to cluster customers into segments.

Note that customers' buying behaviors and characteristics are subject to change over time, and new customers and products may also be added to the transactional database. Therefore, to deal with these issues, customer segmentation should be repeated periodically.

5.4. Experimental Evaluation

In this section, we compare the performance of our hybrid approach against individual-level and segment-based predictive techniques in terms of accuracy and coverage by conducting a set of experiments on a real-world dataset obtained from a grocery retailing company. We describe each step of this experimental evaluation in detail in the rest of this section.

5.4.1. Data Collection and Pre-processing

The data used in this study was extracted from the loyalty card system of a grocery chain and it contains almost 2 million purchase transactions of 16024 customers between October 1, 2012 and August 31, 2014. Before applying clustering algorithms and building predictive models, some pre-processing procedures were applied. Firstly, missing values and duplicate records were eliminated, and transaction records in the same day were aggregated for each customer. In the dataset, each transaction record contains customer's membership number, purchase date, product purchased, product quantity purchased, product category, a set of product sub-categories and product price. As declared before, product attribute is highly specific, because of that, we considered product category feature in this study.

Since this study focuses on a binary classification problem, a binary target variable called "is_purchased" was created, and the original dataset was transformed into a new form where each data record was labelled with the value of 1 in case of a purchase and 0 in case of no purchase. Moreover, all predictive features and the corresponding features of customer segmentation were hidden in the transactional data. As a result, all predictive features were generated for each record and, similarly, the features of customer segmentation were generated for each customer. However, some transactions and customer may have null values for some features. For example, a customer's first five transactions may not contain a value for the feature "average number of products purchased in last 5 shopping occasions" or customers with only one shopping occasion do not have a value for the segmentation feature "average inter-purchase time". Therefore, we excluded the transactions including predictive features with null values as well as the customers including segmentation features with null values. After performing these pre-processing procedures, the final dataset contained 2705 customers' records on 446 different product categories.

Categorical variables are not supported by some predictive algorithms (e.g. SVM, NN). Thus, multi-nominal predictive features such as the day and time, were transformed through one-hot encoding technique. With this procedure, dummy variables were created through binary values. We generated one dummy variable for day feature and three dummy variables for time feature. Further, it is needed to standardize all predictive features in order to use machine learning algorithms such as SVM and NN. In this context, before training prediction models with such ML algorithms, predictive features were standardized to the interval [0,1] using min-max technique. Likewise, identified features of customer segmentation were also standardized to eliminate scale effects, before feeding data into the clustering algorithms.

5.4.2. Predictive Algorithms

In this study, five different widely used and popular ML classifiers are employed for building predictive models. These are: logistic regression (LR), decision trees (DT), random forests (RF), support vector machines (SVM), and neural networks (NN). Additionally, the majority classifier (MC) is used as a baseline for comparison with above learning algorithms.

The LR is a commonly used method to produce a binary prediction using one or more predictor variables. The DT is a tree-shaped structure that generates a set of rules, distinguishing values in a hierarchical form (Breiman et al. 1984). In this study, we chose the well-known C4.5 algorithm for DT classification. The RF was introduced by Breiman (2001) and is an ensemble method based on creation of many decision trees and combination of their results. Determination of the number of trees is an important step in this method, and thereby it was adjusted in the experiments.

The SVM is an advanced statistical classifier proposed by Cortes and Vapnik (1995) and maps the data to a higher dimensional space. There are several Kernel functions which are being used in SVM. Among them, the radial basis function (RBF) is commonly used and presents less parameters and numerical difficulties than other kernels (Hsu et al. 2003). Thus, this study employs the RBF kernel function for the SVM classifier. This method is affected by two key parameters including the kernel coefficient gamma and regularization term C. Accordingly, they were appropriately selected to obtain the best algorithm performance.

The NN is a popular classifier which consists of a number of simple and highly interconnected neurons to solve problems. In this study, we implemented a multi-layer perceptron (MLP) neural network with the back propagation algorithm which is the most widely used neural network (Rumelhart et al. 1988). Moreover, we chose a single-hidden-layer network topology with a logistic activation function. Input layer of our network consists of nine neurons, one for each of the predictive features while the output layer has a single neuron which represents the target binary variable and uses linear transfer function. The crucial parameters namely number of nodes in the hidden layer and learning rate were varied to determine the optimal values in the experiments. Each network was trained for 100 epochs and the network with a minimum value of SSE was considered.

MC is the simplest classification method which does not perform any learning. It selects the most frequent class as output and then uses that to make all predictions. In our classification problem, for the prediction it uses the label (0 or 1) that occurs more frequently in the data.

5.4.3. Experimental Setup

To conduct experiments, we used time-aware community-centered approach in which the entire dataset is split into test and training sets, as the records in the test set have timestamps more recent than the timestamps of the records in the training set (Campos et al. 2014). In this manner, we sorted all transactions of the dataset according to their timestamps. Then, we picked October 13, 2013 as validation starting date ($t_{validation}$) and March 15, 2014 as test starting date (t_{test}). The period before $t_{validation}$, roughly the first 60% of data was used as the training set, and the period between $t_{validation}$ and t_{test}, roughly the first 20% of data was used as the validation set. The remaining roughly 20% of data after $t_{validation}$ was used as test part.

We built the predictive models using the training set and tuned the predictive algorithms' parameters achieving the best prediction accuracy on the validation set. Predictions were generated for each instance in the test set and the predicted results were compared with the true ones by computing corresponding performance metrics. Both training and validation sets (the first 80% of data) were also used for customer segmentation.

5.4.4. Evaluation Measures

In the experiments, each prediction was compared with the actual occurrence and it was then labeled with one of the entries of confusion matrix as shown in Table 11. A confusion matrix is a tabular representation of prediction model outcomes and "actual" rows represent the true classes for instances, whereas "predicted" columns represent the labels assigned by the classifier.

Table 11: Confusion Matrix

		Predicted	
		Positive	Negative
Actual	Positive	True positive(TP)	False negative(FN)
	Negative	False positive(FP)	True negative(TN)

To exemplify these confusion matrix entries for our study, a predictive model predicts that a customer buys a particular product in a particular transaction and if it is true, then it is counted as true positive (tp), or false positive (fp) otherwise. A predictive model predicts that a customer does not buy a particular product in a particular transaction and if it is true, it is then counted as true negative (tn), or false negative (fn) otherwise.

We evaluated the performance of predictive models by considering F-measure and coverage. F-measure is the harmonic mean of precision and recall. Recall (TPR) is the ratio of the number of correctly classified positive examples to the number of positive examples in the data, while precision measures the proportion of the number of correctly identified positive examples divided to the number of examples classified as positive by the prediction model. These two measures are formulated based on confusion matrix as follows:

$$Recall(TPR) = \frac{tp}{tp+fn} \tag{5}$$

$$Precision = \frac{tp}{tp+fp} \tag{6}$$

Then, F-measure is computed based on the combination of these measures as follows:

$$F - measure = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{7}$$

The overall prediction accuracy was assessed by using micro-averaging method in which tp, fn, tn, fp labels are counted, above performance measures are then calculated by using these aggregated confusion matrix entries (Sokolova & Lapalme 2009).

The prediction coverage is another measure used in study and shows the prediction capacity of the predictive approaches. In this study, we define it as the proportion of products for which a prediction can be formed and it is formulated as follows:

$$Prediction\ coverage = \frac{|P_p|}{|P|} \tag{8}$$

where P_p denotes product set that receives a prediction, and P is the total number of products. Note that this metric is measured over the products in a customer's purchase history rather than all products in the database. Similarly, overall prediction coverage was evaluated via micro-averaging method. In this manner, P_p and P metrics were calculated for each customer separately and these counts were aggregated, and finally prediction coverage was measured.

5.4.5. Parameter Settings

As mentioned before, the predictive algorithms used in this study (except for LR) have several parameters to be tuned before final usage. To determine the best combination of parameter values for each algorithm, we used a trial and error approach in which an exhaustive search is performed with a range of parameter values. Therefore, training-validation procedure was repeated multiple times and best parameter values were picked based on the maximum predictability of the predictive algorithms for the validation set. The results of parameter estimation are discussed in Section 5.4.6.2.

5.4.6. Results

5.4.6.1. Determination of Number of Clusters for Customer Segmentation

In this study, in order to determine the optimal number of clusters for customer segmentation, we used Ward's minimum variance method. It generated a dendrogram as shown in Figure 8.

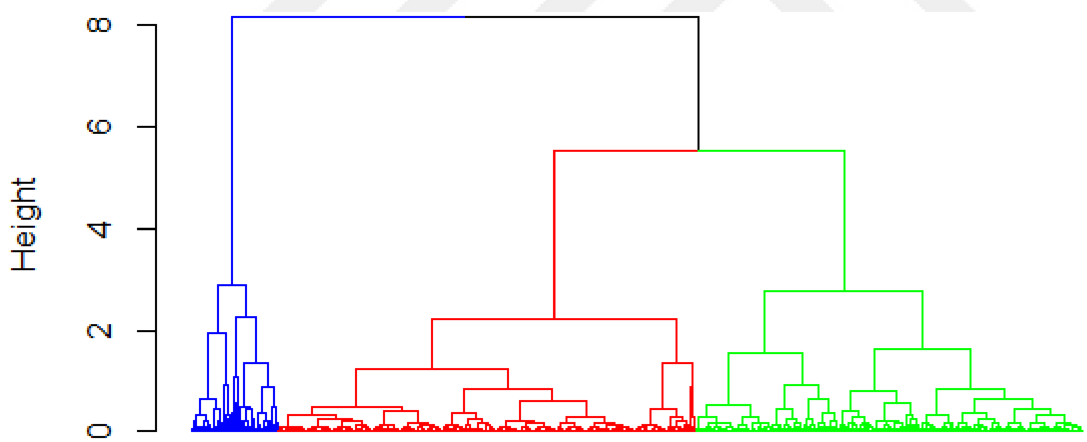

Figure 8: Dendrogram for Ward's method

We can see from the figure that the best choices for total number of clusters is 3 and then customers were segmented into three groups.

5.4.6.2. Parameter Tuning for Predictive Algorithms

The parameter estimation results of four learning algorithms namely, SVM, ANN, DT and RF are presented in Table 12.

Table 12: Parameter Estimation Results

Predictive Algorithms	Parameters	Range of values	Best Value for Individual-level	Best Value for Segment-based
SVM	cost	1, 10-90 (an increment of 10)	70	70
	gamma	0.001, 0.01, 0.1, 0.5, 1	0.001	0.5
NN	# of hidden layer nodes	1-10 (an increment of 1)	3	3
	learning rate	0.1-0.9 (an increment of 0.1)	0.2	0.9
DT	min # of instances per leaf	2-9 (an increment of 1)	2	2
RF	# of trees	50-500 (an increment of 50)	400	400

In Table 12, range of values tried and best values for each technique are given separately. For example, for kernel coefficient gamma of SVM algorithm, we tried 1 and the values varied from 10 to 90 in an increment of 10. Then, according to the results of experiments we determined the optimum value giving best prediction accuracy on validation set and they are 0.001 and 0.5 for the individual-level and the segment-based approaches, respectively. The best values of each parameter given in Table 12 are used for each predictive algorithm in subsequent experiments conducted in testing phase.

5.4.6.3. Comparative Prediction Results

A performance comparison of the individual, segment-based and hybrid predictive with all considered classifiers is presented in Table 13

Table 13: Prediction Results

Classifier	Individual		Segment-based		Hybrid	
	F-measure	Coverage	F-measure	Coverage	F-measure	Coverage
LR	0.308	0.506	0.187	0.871	0.307	0.92
DT	0.295	0.502	0.206	0.866	0.292	0.915
RF	0.289	0.502	0.189	0.866	0.283	0.915
SVM	0.286	0.502	0.222	0.866	0.277	0.915
NN	0.286	0.486	0.210	0.866	0.275	0.913
MC	0.201	0.506	0.006	0.871	0.149	0.920

From the results, we can see that the proposed hybrid approach significantly improves the classifiers' performance, especially the prediction coverage. Hybrid approach is capable of generating predictions for more than 90% of all of all customer-product pairs on average. It achieves substantially higher improvements (more than 80%) in terms of prediction coverage compared to individual-level approach. The proposed hybrid approach also provides 5% increase in the prediction coverage on average, when we compare it with the segment-based approach.

In terms of prediction accuracy, it is obvious that with all classifiers, the predictive performance of individual-level approach is higher than that of the segment-based approach. This is due to the fact that homogeneity of transactions is high in the individual-level approach, and this situation gets distorted in the segment-based approach in which the customers are aggregated into large segments and predictions are generated on those segments. Besides, classifiers with our hybrid approach exhibit an impressive performance versus the ones with segment-based approach, and with machine learning algorithms, our approach provides almost 30% rise in F-measure score on average. When we compare our approach with the individual-level approach, classifiers with our hybrid approach provides a small amount of deterioration in the prediction accuracy. The likely reason behind this result is that the hybrid approach has significantly higher prediction coverage than the individual-level approach.

On the other hand, F-measure scores of LR, DT, RF, SVM, NN, and MC with individual-level approach are .308, .295, .289, .286, .286 and .201, respectively, whereas the same score measures of LR, DT, RF, SVM, NN, and MC with segment-based approach are .187, .206, .189, .222, .210 and .006, respectively. These results indicate that LR and DT performs better with the individual-level approach, while non-linear function methods (SVM and NN) give better performance with segment-based approach. The reason behind this may be that SVM and NN algorithms require massive amounts of training data, and thereby their performance deteriorates with the individual-level approach in which predictive models are constructed based on a smaller amount of data pertaining to a particular customer-product pair. In the same way, these algorithms produced better results with segment based approach since such a model is built based on a large training dataset which is formed by the collection of similar customers' data. Moreover, classifiers with the hybrid approach show similar performance pattern to that with individual-level approach since hybrid approach uses that approach as the primary method. In this context, LR achieves the best F-measure with score of .307, whereas SVM and NN have nearly the same performance and also perform the least favorably with their lowest F-measure scores. In addition to these, as shown in the experimental results, in all predictive modeling approaches, selected machine learning methods significantly outperform the major classifier benchmark in terms of prediction accuracy.

Overall, among the predictive modeling approaches, individual-level approach performs reasonably well in terms of prediction accuracy, but it sacrifices too much coverage. Segment-based approach is highly efficient in terms of prediction coverage by giving prediction for large number of customer-product pairs, but it provides less accurate predictions compared to other approaches. In this manner, hybrid approach improves the performance of the segment-based approach in terms of both the prediction accuracy and coverage. It also provides a significant increase in the coverage with only a small amount of deterioration in the prediction accuracy compared to individual-level approach. Hence, all these results demonstrate that our hybrid approach reduces the shortcomings of the state-of-the-art predictive modeling approaches by utilizing their strengths.

Moreover, regarding the prediction performance of machine learning algorithms, although there is no big difference among them, results suggest that LR outperforms the other four algorithms. Therefore, LR compared to the other classifiers used in this study seems to be the most powerful algorithm for the problem of customers' individual purchase predictions.

5.5. Conclusion

Predicting customers' individual purchase behaviors is crucial in the design of customized services and marketing strategies. Individual-level and segment-based methods are two major predictive modeling approaches of customer behavior that have been successfully employed by several researchers, however both suffer from the certain drawbacks. To eliminate the limitations of the state-of-the-art approaches, in this chapter, we propose a hybrid approach that estimates customers' individual purchase behaviors by exploiting the strengths of these approaches. The proposed approach takes the past transactional data as input and employs predictive algorithms to generate predictions. We evaluated our approach using real-world data and employing five prominent machine learning classifiers and a baseline classifier. Several experiments were carried out to compare the prediction performance of the proposed approach with the other two state-of-the-art approaches in terms of prediction accuracy and coverage, and also to test the performance of different classifiers in the prediction of customer purchase behavior.

The results of this study demonstrate that the proposed approach notably increases the prediction coverage compared to both the individual-level and the segment-based methods, while preserving promising prediction accuracy levels, which are almost the same with the individual-level approach. These findings indicate that the individual-level and the segment-based approaches complement each other really well, and the hybrid approach based on these techniques is able to overcome the shortcomings of the state-of-the-art approaches. Furthermore, the comparative results between the individual-level and the segment-based approaches corroborate the findings of the previous studies (Palmisano et al. 2008; Palmisano et al. 2007; Lombardi et al. 2013; Jiang & Tuzhilin 2006) which have shown that the individual-level technique provides better prediction performance than the segment-based technique. The comparative results also show that among the five algorithms, the logistic regression achieved the highest prediction performance. It seems to be very appropriate to be used for prediction of customer purchase behavior, not only because of its prominent performance, but also due to the ease of implementation and fast computation time.

CHAPTER 6

SHOPPING LIST PREDICTION: AN EMPRICAL COMPARISON OF PREDICTIVE MODELING APPROACHES

This chapter presents an empirical comparison of predictive modelling approaches described in Chapter 5 in the shopping list prediction problem by utilizing different machine learning techniques.

6.1. Introduction

In today's business, customized services and one-to-one marketing actions play important roles for companies to succeed higher customer satisfaction and gain a stronger competitive position. To effectively implement and launch such services, predicting customers' next shopping basket is essential for retail companies. This task is estimating the set of items that a customer could probably buy in the next visit. In recent years, there has been an increasing number of research works done in next basket prediction/recommendation (Wang et al. 2015; Rendle et al. 2010; Wan et al. 2015; Yu et al. 2016).

Most of the existing studies have taken customers' sequential buying behaviors into consideration to predict their next baskets. In this manner, markov chains are popular methods employed for sequential prediction. For example, Rendle et al. (2010) employed both the Markov chain model and the matrix factorization, while (Wang et al. 2015) extended MC based method by utilizing representation learning to predict customers' behaviors in the next transaction. Differently, Wan et al. (2015) and Yu et al. (2016) employed Neural Networks (NN) and Recurrent Neural Networks (RNN) respectively. However such methods have been successfully employed for basket prediction, they are mainly based on transactional data of whole customers.

On the other hand, Cumby et al. (2004) proposed a methodology which predicts customers' shopping lists in a more personalized manner by constructing prediction models relying on their individual past data. This study used individual-level predictive modeling approach mentioned in Section 5.3.1 and built prediction models for reach customer-product pair by utilizing customers' past transactions and employing machine learning techniques. However, this study demonstrated the usage of the individual-level approach with some of primitive classifiers (e.g. Perceptron, Winnow, etc.) in the shopping list prediction, it did not apply any kind of segment based approach or well-known ML algorithms to examine which of these approaches are more effective for shopping list prediction. Therefore, this study aims to empirically compare two conventional predictive modeling approaches with the hybrid approach described in Chapter 5 in the shopping list prediction. For this purpose, machine learning classifiers are used for the prediction generation. The comparison is performed on a real-life dataset by conducting a series of experiments and the results and findings are presented.

6.2. Shopping List Prediction

To create individual shopping lists for customers, this study uses three predictive modeling approaches described in Chapter 5. Both individual-level and segment-based approaches build predictive models by using customers' past purchase transactions and employing predictive algorithms. Similarly, both approaches employ the same feature set and predictive algorithms used in Chapter 5, and segment-based approach also utilizes the same customer segments formed in Chapter 5 in order to generate predictive models.

Different from Chapter 5, the main idea of this chapter is to predict the shopping list of customers. For this purpose, for each transaction, it is needed to rank all the possible products from the most probable one to the least probable one, instead of simply deciding whether a particular product will be purchased in a particular transaction. In this manner, prediction models built by the predictive modeling approaches estimate the probability that a certain product will be purchased in a transaction. The products having a probability value above 0 are sorted according to their probability values in the decreasing order. Thus, shopping list prediction is formed by selecting the top N products from the resulting ranking.

6.3. Comparison of Predictive Modeling Approcahes

After the prediction model is built by any predictive modelling approach, ranked product lists can be generated for new purchase transactions of customers and Top-N ranked products can be formed as a predicted shopping list. However, we compare the predictive modeling approaches in this study, and for a given date, these approaches can generate shopping lists in different sizes. With such situations, it might be impossible to form Top-N ranked products for all approaches, and thus it is unable to make a fair comparison among these approaches. In addition, the individual-level and segment-based approaches may not generate any prediction in some cases, and then again it is impossible to make a comparison among the approaches in such cases. For the above reasons, we proposed a simple baseline method for each predictive modeling approach to add products to the predicted Top-N shopping list, when the number of predicted products is less than N.

For baseline methods, we chose the Most Popular (MP) technique (also known as most-frequent method in the domain of recommender systems (Sarwar et al. 2000)) which sorts the products according to their purchase frequency and simply returns N frequently products from this ranked list. We proposed the MP method in three forms – individual popular (i.e. ranked products of each customer obtained from the customer's past records), segment popular (i.e. ranked products of each customer segment obtained from the customer's past records), and aggregate popular (i.e. ranked products of the whole customer base). All of these methods generate lists containing frequently purchased products that are potentially relevant to customers.

The baseline of each predictive modeling approach utilizes different MP techniques and Table 14 summarizes which predictive approach uses which MP techniques.

Table 14: Baseline methods

MP techniques/Baselines	Individual-level	Segment-based	Hybrid
Individual Popular	1	na	1
Segment Popular	na	1	2
Aggregate Popular	2	2	3

According to Table 14, the baseline method of individual-level approach uses individual popular and segment popular methods respectively to form the shopping list. For the same purpose, the segment-based baseline utilizes segment and aggregate popular techniques respectively, and the hybrid one uses all MP methods which are individual popular, segment popular and aggregate popular ones respectively.

The intuition behind proposing such baselines is using them as the complementary of predictive modeling approaches. The corresponding baseline method is used to reach desired prediction list size (N), and for this purpose, products which are extracted by starting the top of the list by the baseline method are added to the end of the prediction list size of the predictive modeling approach. Thus, all the predictive modeling approaches with their baselines can construct a prediction list which contains a desired number (N) of products. Accordingly, the prediction lists produced by all the approaches are in equal sizes, and it is possible to make a fair comparison among their performances.

6.4. Experimental Evaluation

The performances of all predictive modeling approaches are investigated and compared by conducting a set of experiments on a real-world dataset obtained from a grocery retailing company. All the details related to this evaluation will presented in the following sub-sections.

6.4.1. Data Collection and Pre-processing

This study used the dataset used in Chapter 5 and applied the same data collection and pre-processing processes described in Chapter 5.

6.4.2. Predictive Algorithms

In this study, the same five ML classifiers used in Chapter 5 were employed. Different from Chapter 5, in this study these classifiers generate probability estimates for candidate products. LR and NN techniques are used directly to produce probability estimates. For this purpose, Laplace smoothing was applied for DT and RF methods, since their standard versions do not generate probabilities directly. Likewise, the output of original SVM is not a probability. However, ksvm() implementation in package kernlab (Zeileis et al. 2004) can also compute class-probability output by using an improved implementation (Lin et al. 2007) of Platt's posteriori probabilities (Platt 1999). Thus, ksvm() method was used to obtain class probabilities as output by SVM algorithm. In addition to these, the performance of these machine learning classifiers were benchmarked using baseline methods based on MP techniques introduced in Section 6.3.

6.4.3. *Experimental Setup*

The experimental settings for the evaluation of this study is the same as the one in Chapter 5.

6.4.4. *Evaluation Measures*

For the performance evaluation, we used F-measure and mRHR. F-measure is the harmonic mean of precision and recall. These two measures are defined for our case as follows:

$$Recall@N = \frac{|pred@N \ \cap \ purc|}{|purc|} \tag{9}$$

$$Precision@N = \frac{|pred@N \ \cap \ purc|}{|pred@N|} \tag{10}$$

In these formulas, pred@N donates the Top-N prediction list for a particular transaction, and purc is the actual product set that the customer has purchased in that transaction. Both recall@N and precision@N measures were computed by using micro-averaging method. For his purpose, firstly, pred@N, purc and |pred@N ∩ purc| values were computed for each customer's transaction separately, and then these calculated values of all the transactions were aggregated. Finally, recall@N and precision@N measures were based on these aggregated values, and thus F-measure was calculated using these two measures.

mRHR was another metric used in this study and it is our newly defined measure for the ranking evaluation of in top-N prediction. The detail description of this measure will be given in Chapter 7. Similarly, this measure was calculated via micro-averaging method. The values of number of preferences and (hit_i / $rank_i$) were recorded for each customer's transaction distinctly, and their values pertaining to whole transactions were aggregated. Thus, mRHR metric was computed based on these aggregated values.

6.4.5. *Parameter Settings*

This study used the best values identified in Chapter 5 for the corresponding predictive algorithms in the experiments conducted in testing phase.

6.4.6. *Results*

In this section, we report the experimental results to evaluate the performance of predictive modeling approaches in the shopping list prediction. Figure 9-14 demonstrate the F-measure and mRHR results for each predictive modeling approach with different machine learning techniques across a wide range of prediction list sizes.

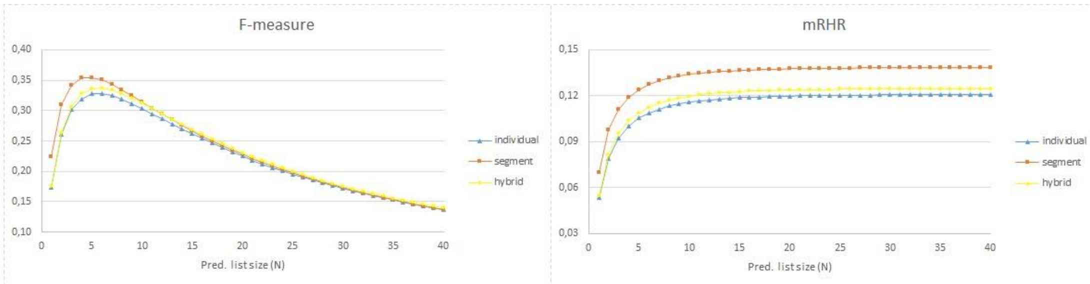

Figure 9: Comparison of predictive modeling approaches with LR

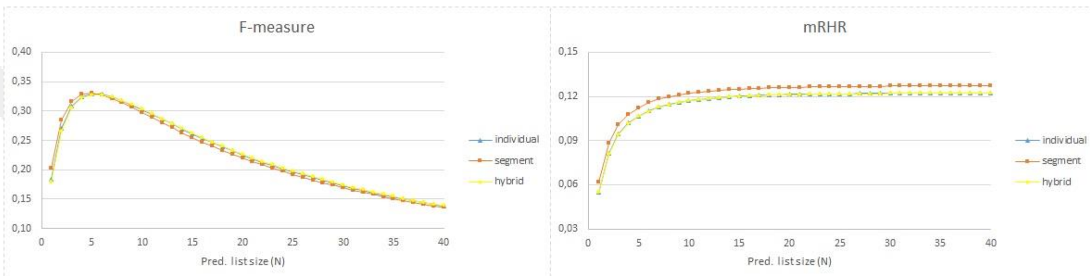

Figure 10: Comparison of predictive modeling approaches with DT

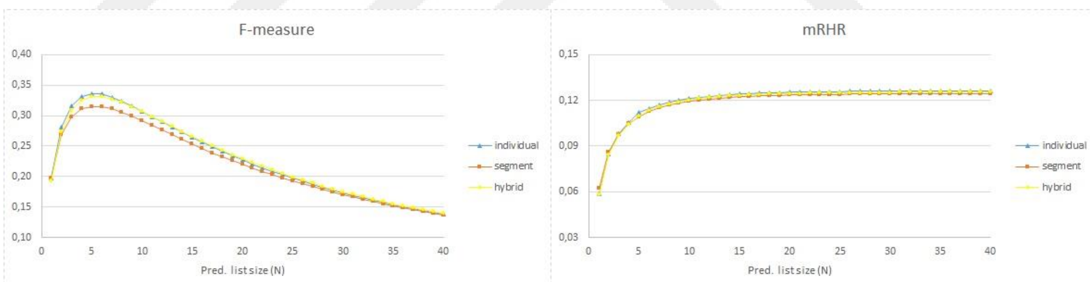

Figure 11: Comparison of predictive modeling approaches with RF

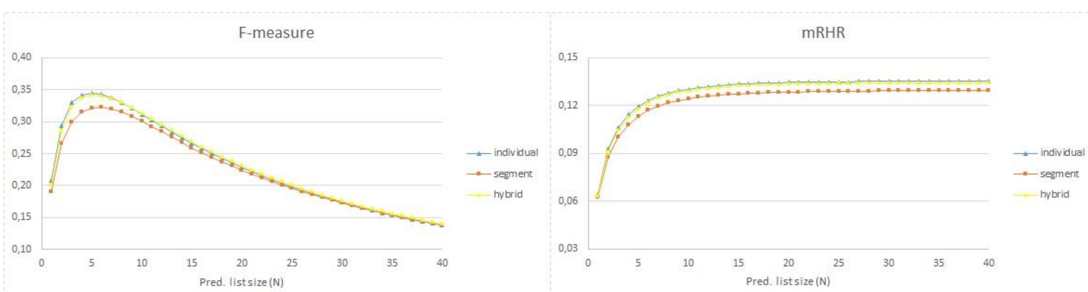

Figure 12: Comparison of predictive modeling approaches with SVM

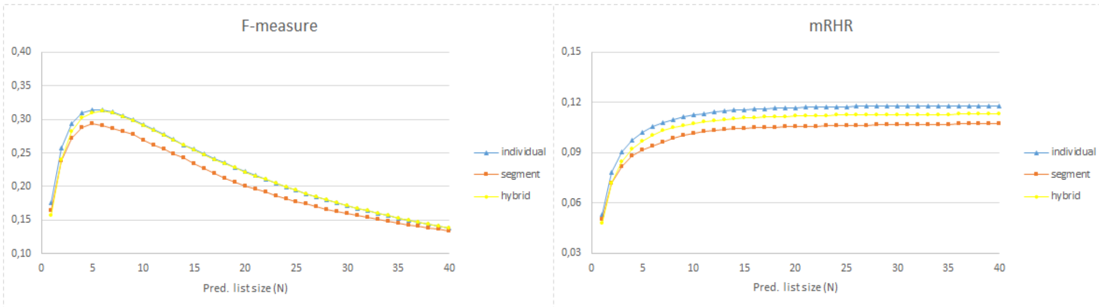

Figure 13: Comparison of predictive modeling approaches with NN

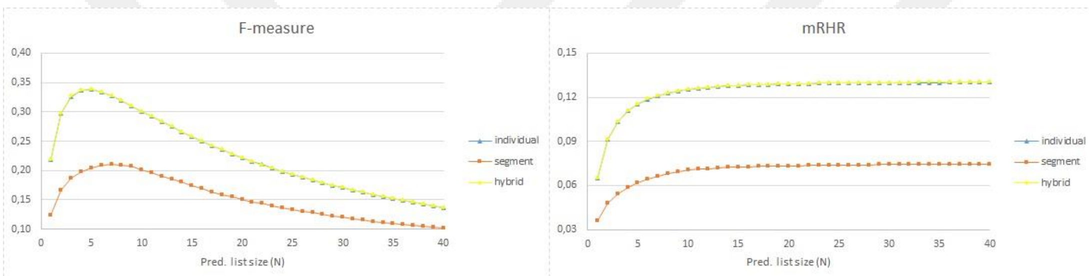

Figure 14: Comparison of predictive modeling approaches with MP

In our case study, the results show that comparison among the predictive modeling approaches varies depending on the machine learning techniques used. It can be observed from the plots that the hybrid and the individual-level approaches dominate segment-based approach when they use RF, SVM, NN and MP algorithms. In contrast, segment-based approach outperforms both the hybrid and the individual-level approaches when LR and DT classifiers are used in these predictive modeling approaches. With RF, SVM and MP algorithms, performances of the hybrid and the individual-level approaches are comparable. The individual-level approach with RF and SVM algorithms perform slightly better than the hybrid approach with the same algorithms, while the hybrid approach with MP method performs slightly better than the individual-level one with that method.

Additionally, it is clear from the curves that prediction list size (N) influences the performance. Performance of both measures exhibits a certain pattern which is consistent for all predictive modeling approaches with all machine learning algorithms. As seen in the plots, when N is small, there is a gradual increase in F-measure. It reaches its maximum value at around N=5 and starts decreasing thereafter. Similarly, for the curves of mRHR, the increases are drastic at the beginning, but gradually slow down, after N exceeds around the values of 5. Further, with higher values of N, the value of mRHR becomes constant. Therefore, results of both F-measure and mRHR imply that 5 would be an optimal value for prediction list size (N) in our case study.

A performance comparison of predictive modeling approaches with each machine learning classifier is presented in Table 15. Note that, for these results N was taken as 5 since this value is the optimum for the prediction list size as stated before.

Table 15: Comparison of predictive modeling approaches with different ML algorithms

Approach	**ML Algorithm**	**F-measure**	**mRHR**
Individual-level	LR	0.328	0.100
	DT	0.329	0.107
	RF	0.337	0.112
	SVM	**0.345**	**0.120**
	NN	0.314	0.102
	MP	0.338	0.116
Segment-based	**LR**	**0.355**	**0.124**
	DT	0.330	0.112
	RF	0.315	0.109
	SVM	0.322	0.113
	NN	0.293	0.092
	MP	0.204	0.062
Hybrid	LR	0.336	0.109
	DT	0.328	0.106
	RF	0.331	0.110
	SVM	**0.341**	**0.118**
	NN	0.310	0.097
	MP	0.338	0.116
(prediction list size (*N*)=5)			

For each predictive modeling approach, ML classifier giving the best performance is bold-faced in the table. As seen in Table 15, segment-based approach exhibits the best performance among the predictive modeling approaches, when it uses LR classifier as the predictive algorithm. Using LR classifier, segment-based approach improves the F-measure from 0.328 to 0.355, and mRHR from 0.1 to 0.124. They are improvements of 8.2% and 24% respectively. The individual-level and hybrid approaches with SVM algorithm follow this best version respectively.

Regarding the performance of machine learning techniques in the shopping list prediction, we can observe from Figure 9-Figure 14 and Table 15 that SVM is the best by providing remarkable performances for all predictive modeling approaches. Moreover, in this study, LR and RF are other algorithms performed well with all predictive modeling approaches.

6.5. Conclusion

Accurate prediction of customers' shopping baskets provides an important contribution to effective customized services and one-to-one marketing actions in retail industry, and thereby can lead to maximize customer satisfaction and company profitability. In this context, this chapter proposes a comparison of conventional predictive modeling approaches and the hybrid approach proposed in Chapter 5 for customer's individual shopping list prediction. The prediction performances of all these approaches were evaluated with different machine learning classifiers.

The results show that the individual-level and the hybrid approaches outperform the segment-based approach when certain machine learning classifiers are used, and are inferior when the other machine learning classifiers are employed. This indicates that the performance of predictive modeling approaches depends on the machine learning algorithm used, and we can conclude that testing different ML algorithms is important, before making a choice for the predictive modeling approach in a certain prediction problem. Moreover, in our case study, the best prediction performance is obtained from the segment-based approach with LR algorithm and SVM compared to the other classifiers mostly perform well for all predictive modeling approaches. Another interesting finding of this study is that the high accuracy can be achieved using a prediction list size (N) covering only a fraction of items (around 5 items).

CHAPTER 7

A MODIFIED RECIPROCAL HIT RANK METRIC FOR RANKING EVALUATION OF MULTIPLE PREFERENCES

In this chapter, we propose a modified version of Average Reciprocal Hit-Rank (ARHR) metric which can be used for the ranking evaluation of in top-N prediction and recommendation lists in the cases where users or customers have multiple preferences at a time or a specific time interval. A detailed, in-depth description of this study is in (Peker & Kocyigit 2016b).

7.1. Introduction

ARHR is a ranking measure that considers the item position in Top-N list. It measures how close the predicted items to the top of the recommendation list. This metric is especially important for the cases in which there are too many alternatives to recommend and the cases in which it is required to present a limited number of items for users. In a grocery shopping, there is a tremendous diversity of products and forming a reasonable recommended item set is extremely difficult. If, for example, the display size of mobile devices is limited, the users may not want to browse through the entire list. It is needed to provide a small set of items that fit in such a small display.

In addition to above cases, users usually investigate items in Top-N list starting at the top and most probably they do not go down to the end of the list, when the list is not short (Bobadilla et al. 2013). Because of that, users do not like long recommendation lists containing too many items. In order to overcome all of the aforementioned problems, the solution is to shorten the recommendation list by identifying the most appealing items for the user. In this manner, it is important to predict highly relevant items in the top positions of recommendation list. Therefore, ARHR measures closeness of correctly predicted item to the top of the list, and thereby perfectly suits for evaluating the quality of Top-N prediction and recommendation lists.

Despite the usefulness of ARHR for the quality assessment of Top-N recommendations and predictions, it cannot be applied for the cases where the user has multiple preferences at a time or a specific time interval. In many of the cases, however, users have multiple preferences at a time or a specific time interval, such as a list of purchased products for a single transaction of grocery shopping, a list of websites being visited throughout the day or a list of TV programs being watched throughout the evening. In such cases, there may be multiple correct matches (hits) and present ARHR metric cannot evaluate these cases because of its shortcomings.

The primary goal of this chapter is to propose a new metric called mRHR to evaluate the quality of Top-N recommendation and prediction lists for users or customers with multiple preferences at a time or a specific time interval. For this purpose, the standard ARHR metric is modified accordingly. The use of the proposed metric is demonstrated in an example and its performance is also investigated by conducting an exhaustive set of experiments on the real-life data from the grocery shopping domain. Experimental and statistical test results are reported to show that our new measure is consistent and appropriate for evaluating quality of Top-N recommendation and prediction lists.

The remainder of this chapter is organized as follows. In Section 7.2, the original ARHR metric is described. Section 7.3 introduces our proposed metric and its usage is illustrated on an example in Section 7.4. Section 7.5 describes comprehensive experimental evaluation. Finally, the chapter is concluded with a summary in Section 7.6.

7.2. ARHR

ARHR is based on reciprocal rank (RR) (Moffat & Zobel 2009) in Information Retrieval (IR) research and it is a popular ranking metric to measure the quality by finding out how far from the top of the list the first relevant document is. ARHR is defined as (Deshpande & Karypis 2004):

$$ARHR = \frac{1}{n}\sum_{i=1}^{h}\left(\frac{1}{p_i}\right) \tag{11}$$

where h is the number of hits which is the number of preferred items that were also present in the top-N recommended items returned for each user, p_i is the position of the item in the ranked list for the i-th hit, and n is the number of users. According to the equation 1, RR, which is equal to $1/p_i$, is calculated for each hit and at the end all computed RR values are averaged.

ARHR rewards each hit based on its position in the recommendation list, and hits that occur earlier in the ranked list are assigned higher weights than that of the ones that occur later in the list. Thus, ARHR measures how close the correctly predicted items to the top of the recommendation list. The value returned by ARHR is between 0 and 1. Higher values of ARHR are more desirable as they indicate that the algorithm is able to predict items in the earlier positions of Top-N lists. Conversely lower values of ARHR indicate that the algorithm predicts items in the later positions of Top-N lists.

Many previous studies (Zheng & Li 2011; Kang & Cheng 2016; Cheng et al. 2014; Ning & Karypis 2011) used ARHR successfully in the evaluation of top-N recommendations. In these studies, ARHR is being used when the user prefers only an item at a time. When the user prefers multiple items at a time, it is not possible to compute and use ARHR metric directly. Therefore, a new metric is required to evaluate the user's multiple preferences at a time and the next section presents mRHR, a modified version of ARHR which can be used for this purpose.

7.3. Proposed Metric

In order to cover multiple preferences at a time or a specific time interval, we proposed a novel metric called *mRHR* which overcomes the shortcomings of the ARHR mentioned in the previous section.

Let the user prefer more than one item in a time interval or at a time, *mRHR* is defined as follows:

$$mRHR = \frac{1}{\#\ of\ preferences} \sum_{i=1}^{N} \left(\frac{hit_i}{rank_i} \right) \tag{12}$$

where hit_i donates if the recommended or predicted item preferred by the user. Like in original ARHR, if the recommended or predicted item is preferred by the user, then it gets true (1), otherwise it gets false (0). *N* is the length of ordered recommendation or prediction list and $rank_i$ is the ranking position of the preferred item in the list. We employ a mapping function for $rank_i$ variable as follows:

$$rank_i = \begin{cases} rank_{i-1}, & hit_{i-1} = 1 \\ rank_{i-1} + 1, & hit_{i-1} = 0 \end{cases} \tag{13}$$

As shown in the function, $rank_i$ value of an item is determined by whether or not the earlier item is preferred by the user. That is, if the hit value pertaining to preceding item in the recommendation or prediction list is true (1), then $rank_i$ is equal to $rank_{i-1}$, otherwise it equals $rank_{i-1}$ plus one. By definition, $rank_i$ for the first recommended item ($rank_1$) is always set to 1 and the above function is computed for the following items.

Our proposed metric, mRHR requires a ranked list of the recommended or predicted items as an input and starts the process from the top of the list. For the top (first) recommended item, *i* is set to 1. Moreover, in the original ARHR, the equation 2 is divided by number of users, since it is only used for users' test sets including single instance and the overall score is computed directly. mRHR can be used when the test sets contain more than one test instance. Thus, it is computed for each test instance of the user. In our proposed metric, the denominator is replaced by the number of preferences of the user at a time, rather than number of users in order to evaluate the multiple items preferred at a time. Then, the overall mRHR of the user is computed by averaging over all computed mRHR values of the user and finally, the overall mRHR for all users is computed by averaging these personal mRHR values of the users.

7.4. Illustrative Example

The following example illustrates the whole computation process of mRHR. Suppose that Alice uses a recommendation agent in her smart phone for the grocery shopping and this application generates a ranked list of predicted products. Suppose further that the application delivers the following recommendation list for the next grocery shopping of Alice: 1.milk, 2.pasta, 3.egg, 4.sausage, 5.ketchup. The numbers indicate the rank of the recommended product in the recommendation list and 1 means the first item in the list. In her next visit, she purchases the following products: egg, cheese, pasta, ketchup. Then, the mRHR metric to evaluate the performance is computed step by step as shown in Table 16.

Table 16. mRHR computation for multiple preferred items by an example

Step	Rec. Rank	Recommended Product	hit_i	$rank_i$	$\frac{hit_i}{rank_i}$
1	1	Milk	0	1	0
2	2	Pasta	1	2	1/2
3	3	Egg	1	2	1/2
4	4	Sausage	0	2	0
5	5	Ketchup	1	3	1/3

Please note that, the computation process starts with the product at the top of the recommended list and goes on until the end of the list. There are three relevant products in the recommendation list and these products are marked as 1 on the hit_i column of Table 16. $rank_i$ is calculated by using Equation 3. For example, $rank_2$ and $rank_3$ values for the second and third recommended products respectively are calculated as follows:

Set $rank_1$=1

Since hit_1=0, then $rank_2$= $rank_1$+1 = 1+1=2

Since hit_2=1, then $rank_3$= $rank_2$ = 2

Finally, by using $hit_i/rank_i$ values, mRHR is computed as:

mRHR = (1/2 + 1/2 + 1/3)/4 = 0.33

The denominator in the above calculation is 4, because the user prefers to purchase four products in her visit. Thus, as shown in the example, our proposed metric, mRHR is capable of evaluating the cases where the user has multiple preferences at a time or a specific time interval.

We used a mapping function for $rank_i$ variable in the computation of mRHR. The following example demonstrates the reasoning behind this strategy. Consider a typical recommender application and its user. Assume that a recommender agent generates a ranked recommendation list containing four items for the next preference of the user. Then, the user prefers four items in that transaction and suppose that Table 17 shows the hits for five different cases. "Rec. Rank" in the table indicates the order of a recommended item. For instance, the second case in Table 17 indicates that the first three items in the recommendation list were preferred by the user, whereas fifth case indicates that the last three items in the recommendation list were preferred by the user.

Table 17. The effect of mapping function for $rank_i$ variable on the mRHR computation

Rec. Rank	***hit_i***				
	(1)	(2)	(3)	(4)	(5)
1	1	1	1	1	0
2	1	1	1	0	1
3	1	1	0	1	1
4	1	0	1	1	1
mRHR without mapping	0.52	0.45	0.43	0.4	0.27
mRHR with mapping	1	0.75	0.63	0.5	0.38

The last two rows in Table 17 show mRHR values calculated by using two different techniques. The first one computes mRHR without using mapping function for $rank_i$ variable in the Equation 3, whereas the second one uses this function to compute mRHR. For example, mRHR values using both options for the first case is computed as follows:

Without using mapping function for $rank_i$: mRHR = (1/1 + 1/2 + 1/3 +1/4)/4 = 0.52

With using mapping function for $rank_i$: mRHR = (1/1 + 1/1 + 1/1 +1/1)/4 = 1

Although all four recommended items are preferred by the user, mRHR value without using mapping function for $rank_i$ is computed as 0.52. Actually, this case should be evaluated as perfect matching. When we look at the mRHR value with using mapping function for $rank_i$, it is 1, which shows the actual evaluation of the case. The similar situation also exists in other cases in Table 17. mRHR values without using proposed technique for cases 2-5 is very low, although three of the four items preferred by the user is also in the recommendation list. On the other hand, mRHR values with using the related function are more consistent and reasonable**.** Therefore, this example demonstrates the effect of using mapping function and clarifies why we employed such a mapping function for $rank_i$ in the computation of our proposed metric, *mRHR*.

7.5. Evaluation

In order to evaluate the performance of newly proposed metrics, most studies explored that whether they correlate well with other standard metrics used for same purpose (Moffat & Zobel 2009; Chapelle et al. 2009; Kekalainen 2005). In this manner, we chose recall which is the most widely used metric for evaluating the quality of recommendations and predictions. We investigated the correlation between recall and our metric in the evaluation of predicting top-N recommendation lists by applying Pearson’s correlation test. Grocery shopping was chosen as application domain and two distinct datasets from two Turkish retail grocery stores were used in the experiments.

In the following experiments, Most Popular (MP) technique was chosen to be the baseline for evaluating and comparing the performance of our proposed metric and recall. The reasons behind this choice were that it is simple and well-known method and it was also employed in the studies of predicting grocery shopping lists (Cumby et al. 2004; Cumby et al. 2005). This approach is not so efficient in the recommendation performance, but it is not our aim in this study. MP method is only considered as a reasonable baseline to predict top-N recommendation lists and to evaluate and compare the performance of corresponding metrics.

In this section, we first describe the datasets used in the experiments. Then, we explain the experimental settings and procedures. Finally, we present the results of the experiments carried out and correlation analysis to explore how well mRHR and recall metrics are correlated.

7.5.1. Datasets

We used two datasets obtained from two different Turkish retail grocery stores. The first company operates in Central Anatolia region, the other one operating in Mediterranean region. Throughout the report, we call the datasets with the region names that they operate in. Mediterranean dataset contains purchase transactions gathered over 100 weeks (January 10, 2012 – August 31, 2014), whereas Anatolia dataset covers the period of 104 weeks (January 1, 2013 – December 31, 2014). In the datasets, many of the customers have visited the stores irregularly (only a few times during the period) and transactions of these customers most probably mislead the results of the study. Thus, we pre-processed the dataset to exclude customers who have visited the store less than 25 times. After this process, the Anatolia and Mediterranean datasets are left with 46 and 2121 customers with 120467 and 534616 purchase records, respectively. Therefore, we have two different legitimate samples to predict top-N recommendation lists. Additionally, as stated in previous chapters, the product feature is overly unique in the datasets. As a result of this, we again only considered product categories, instead of exact products in the experiments.

7.5.2. Experimental Settings and Design

In both datasets, we sorted each customer’s shopping trips according to their timestamps. For each customer’s purchase history, we use the first 80% of the visits as training and the latter 20% as test data. In the testing set, for each visit of the customer a size-N ranked list of predicted products is generated and they are compared with the actual ones purchased by the customer in the test set to compute the corresponding evaluation metrics. We define recall as follows:

$$Recall@N = \frac{|rec@N \ \cap \ purc|}{|purc|} \tag{14}$$

where *rec*@N donates the top-N recommended products for the test instance and *purc* is the actual product set that the customer has purchased in the same test instance.

We first computed results of each metric for each customer separately by averaging all computed values in the customer’s test set, and then average these customers’ personal values to get overall metric values for a given recommendation list size, N for both datasets.

In the experiments, Most Popular (MP) technique method was used to generate top-N recommendation lists for the customers’ visits in the test set. Experiments were repeated 10 times for each dataset by varying recommendation list size, N, from 1 to 10 (in increments of 1) in order to investigate the performance of related metrics for different values of N.

7.5.3. Experimental Results

On the customers’ personal metric values, Pearson’s correlation analysis was applied to examine the relationship between mRHR and recall, and Table 18 shows the correlation coefficient results for different values of N (i.e., 2, 4, 6, 8 and 10) for both datasets.

Table 18. Correlation statistics between mRHR and recall metrics

Dataset	**N**				
	2	4	6	8	10
Anatolia	.967	.893	.867	.826	.776
Mediterranean	.971	.897	.831	.776	.729

According to Table 18, as N increases, the correlation coefficients decrease for both datasets. The main reason behind this is that the difference between the values of two metrics increase, as N increases. Despite this, the output shows that there is a strong, positive correlation between mRHR and recall for both datasets. Note that the significance value is less than .001 for all of the correlation coefficients in Table 18 and this indicates that all of these correlations are statistically significant.

Figure 15 also demonstrates the overall performance results for different values of N for both datasets. It can be observed from the plots that mRHR and recall are positively correlated. This figure also verifies our previous statement, which is that the difference between the values of two metrics raises, as N increases. This is because mRHR is a rank measure, and so it decays faster than recall with the growth of N.

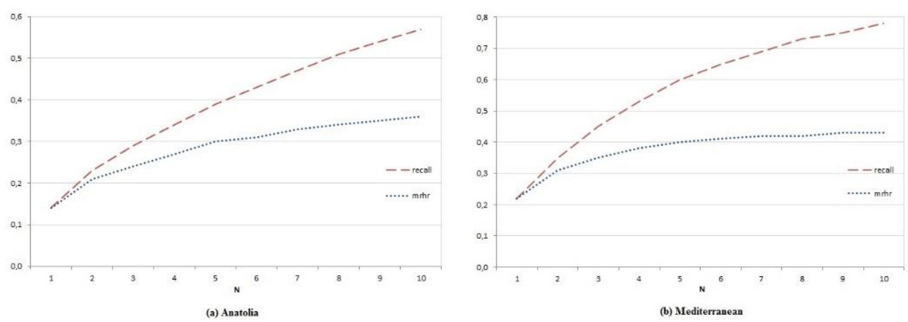

Figure 15: Overall performance results for different values of N

As in many prior studies (Deshpande & Karypis 2004; Ostuni et al. 2013; W. Li et al. 2011; Ning & Karypis 2011; Cheng et al. 2014; Kang & Cheng 2016), with the growth of the recommendation list size N, ARHR and recall or hit rate (HR) generally tend to increase, and many studies (Deshpande & Karypis 2004; Zheng & Li 2011; Ning & Karypis 2011; Cheng et al. 2014; Kang & Cheng 2016) in the literature also show that recall or HR takes mostly higher scores than ARHR for the same N value. In this manner, it is obvious that similar patterns exist in the plots of both datasets in Figure 15, in which our new metric mRHR takes the place of ARHR in the previous studies. Therefore, all of these results indicate that mRHR can take the place of ARHR for evaluating performances of top-N recommendation and prediction lists in the cases where the users have multiple preferences at a time.

7.6. Conclusion

In this chapter, we proposed mRHR, a modified version of ARHR metric that can be used for the purpose of ranking evaluation of multiple preferences in top-n recommendation and prediction lists. The standard ARHR metric is modified in order to make it applicable to multiple preferences of the user in a time interval or at a time. We further demonstrated the applicability of the proposed metric on a grocery shopping domain using real-life data. A series of experiments were conducted to investigate the performance of the proposed metric and the relationship between the performances of the proposed metric and recall was also explored via performing Pearson's correlation analysis.

Experimental results confirmed that our proposed metric, mRHR is significantly correlated with recall and the performance measured by mRHR is consistent with the performance measured by ARHR used in the prior studies in the literature. These findings indicate that our proposed metric overcomes the main shortcoming of standard ARHR metric and supports the ranking evaluation of Top-N recommendations and predictions in the cases where the users have multiple preferences at a time.

CHAPTER 8

AN ADJUSTED RECOMMENDATION LIST SIZE APPROACH FOR CUSTOMERS' MULTIPLE ITEM PREFERENCES

In chapters 5 and 6, different approaches have been introduced to build prediction models for customer purchase behaviors. Using prediction models, it is possible to estimate which products customers might purchase in the future, and based on these predictions, companies can generate personalized promotion and campaign recommendations to customers. In this chapter, a novel approach is presented to dynamically adjust the size of recommendations for multiple preferences of customers. A detailed, in-depth description of this study is in (Peker & Kocyigit 2016a).

8.1. Introduction

By exploiting various prediction models, typical marketing management systems can provide campaign and promotion recommendations to the customers. Prediction models generate either a list of items or a ranked list of items that will be preferred by a particular customer in a particular transaction. In the literature, there are two main recommendation strategies: "find all good items" and "recommend top-N items" (Herlocker et al. 2004). In the "find all good items" approach, all the recommendable items that can suit the customer's tastes are offered, whereas in the "recommend top-N items" approach, only the top ranked N items are recommended to the customer. The latter one is the most common solution for product recommendations and many commercial companies take the take advantage of this technique in their marketing actions.

In "recommend top-N items" strategy, the length of recommendation list usually ranges from 5 to 20 (Pu et al. 2011). It is possible to increase the proportion of items that are correctly identified (recall) by providing more items in the recommendation list. However, users are not likely to be overwhelmed by recommendation lists containing a large number of items, and it is also important to fit recommendation lists in small display devices such as mobile phones (Pu et al. 2011). Moreover, increasing the number of recommended items, N, improves recall, but it is likely to deteriorate precision (Herlocker et al. 2004; Sarwar et al. 2000; Gunawardana & Shani 2009) which is the proportion of recommended items that result in matches. Together with recall, precision shows the quality of the recommendation, and high precision is more preferable in recommender systems using "recommend top-N items" approach (Gunawardana & Shani 2009).

Although, top-N recommendation strategy typically returns a fixed number of items in each recommended list, individuals may have multiple preferences at a time or in a specific time interval, and the number of such preferences may vary depending on cases and context in which the individual is. For example, consider a customer is being time pressed during weekday daytime in the grocery shopping. He prefers to buy a couple of items (e.g., 3 unique items) on weekday evenings for daily needs such as bread, eggs, milk, etc., whereas he has too many items (e.g., 15 unique items) in his shopping basket on a weekend afternoon.

To evaluate the quality of existing recommender systems producing a fixed number of items for users with multiple preferences at a time, let us consider the above example again. Suppose that this customer uses a recommendation agent in his smart phone for the grocery shopping and this application generates a fixed length of recommendation list containing 10 items for his next visit. For the first case, agent may probably perform with high recall and low precision, since it produces a long recommendation list for a number of purchased items, but agent also overwhelms the user with many irrelevant items. For the second case, on the other hand, recall is most probably lower than the one in first case and precision may higher, because the number of recommended items is closer to the number of preferred items. However, recommendation agent also misses some relevant items in this case, since it recommends fewer items than the user prefers.

As explained in the examples, for multiple preferences at a time, recommender systems returning a fixed number of items may cause some issues which are undesirable from the users' point of view, and it is obvious that the length of recommendation list has a significant impact on the recommendation quality. In this respect, this chapter aims to dynamically adjust the recommendation list size of a user with multiple preferences by employing machine learning techniques. The proposed approach dynamically determines the optimal recommendation list size based on the previous preferences of the user. The applicability of the approach is experimentally evaluated by using real-life data obtained from the grocery shopping domain and the results show the effectiveness of our approach.

The remainder of this paper is organized as follows. Section 8.2 describes proposed approach. In Section 8.3, evaluation methodology is presented together with the experimental results. Finally, Section 8.4 concludes the study and points out directions for future work.

8.2. Proposed Approach

The major steps of our approach for adjusted recommendation list size are depicted in Figure 16. First, a predictive model is constructed based on previous preferences of users. Then, identified features' values pertaining to the user and the current recommendation are used as input for the model to estimate an adjusted recommendation list size that will be finally used as a reference in the construction of recommendation list.

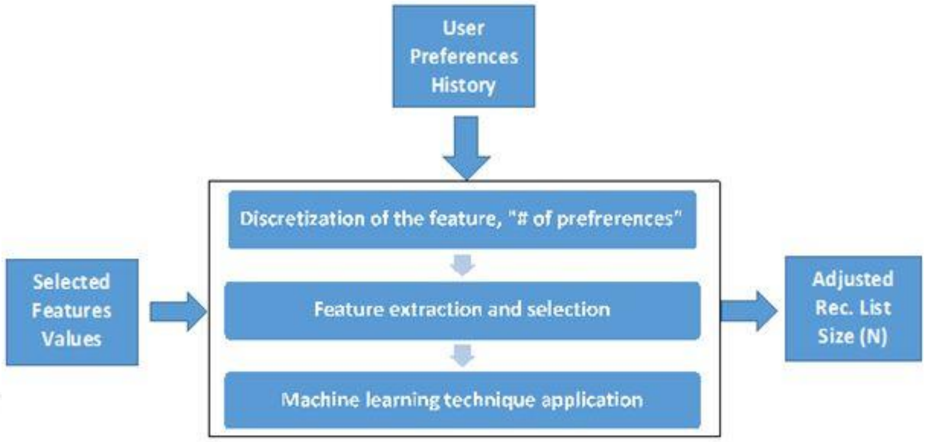

Figure 16: Schematic overview of our approach

In our predictive model, "number of preferences", the numeric target value, is firstly discretized into a set of intervals in order to use machine learning classifiers. The reason behind this step is to construct a categorical variable in order to transform the problem into a classification one, and to employ powerful classification techniques. In case of a continuous target attribute, unsupervised discretization techniques are used to divide the variable into discrete intervals. Famous representatives of unsupervised techniques are equal-width and equal-frequency (Catlett 1991; Kotsiantis & Kanellopoulos 2006). In this study, we use an equal-frequency discretization method, because the equal-width method does not perform well when the variable observations are not distributed evenly, and thereby may cause information loss after the discretization process (Kotsiantis & Kanellopoulos 2006). Equal-frequency as an unsupervised method requires the user to specify the number of discrete intervals and fewer number of intervals is preferable in order to avoid the fragmentation problem occurring in classifiers (Quinlan 2014). Because of that, we set the number of intervals to three (as low, medium and high) in this study.

Not surprisingly, the same number of preferences may mean different things to different people. For instance, in grocery shopping, buying 10 items could be taken as too many to a user with an average number of purchased items of 5, whereas too few to a consumer with an average number of purchased items of 20. Hence, it is important to apply the discretization technique for the number of preferences at the individual level. Therefore, the proposed approach employs equal-frequency technique at the individual level to produce personalized cut points for each user.

After the "number of preferences" is transformed to categorical variable via equal-frequency discretization, features related to users' preferences are identified to build a model that produces better performance. At the final stage of the model construction, a machine learning technique is trained with users' previous preferences with selected features. In this study, we use two classifiers and the linear regression technique in order to compare their performance for recommendation quality. The classifiers used are decision tree (J48) and KNN algorithms. These classifiers are chosen because they are in the top 10 list of classification techniques (Wu et al. 2008). For KNN method, we searched the optimal k value and identified it as 75. We also use multiple regression by coding the values of discretized "number of preferences" variable (low, medium and high) as 1, 2, and 3. In this way, the feature is ordinal and the continuous nature of it is preserved which allow the researchers to employ regression analysis (Rucker et al. 2015).

After training the model, identified features' values of the user and the current recommendation are used as input for the model. Then, it outputs the estimated number of preferences of the user for the recommendation list size. Note that, machine learning classifiers produce the output as a categorical value such as low, medium or high. On the other hand, regression model generates the output as numerical nominal value as 1, 2, or 3. Since, our approach employs the equal-frequency discretization method for "number of preferences", these produced values correspond to bins. Because of that, our approach replaces each bin value by its mean. Note that, if this value is decimal, it is rounded to nearest integer. Therefore, the proposed approach returns this final result as the possible recommendation list size.

8.3. Evaluation

To measure the effectiveness of our approach, grocery shopping was chosen as the application domain and a comprehensive set of experiments were conducted on a real-world dataset obtained from a grocery retailing company. In the following subsections, we first describe the dataset and data pre-processing task. Next, we explain the experimental settings and evaluation measures. Then, the recommendation method that used in the experiments and a benchmark method for the comparison are introduced. Finally, we present the results we obtained.

8.3.1. Dataset

The dataset we used in the experiments contains purchase transactions gathered over 104 weeks (January 1, 2013 – December 31, 2014). In the dataset, many of the customers have visited the store irregularly (only a few times during the period) and transactions of these customers most probably mislead the results of the study. Thus, we pre-processed the dataset by excluding customers who have visited the store less than 25 times. After this elimination, the dataset is left with 46 customers with 534616 purchase records. Moreover, as stated in previous chapters, the product features include both brand and amount information and are highly unique. Therefore, we again only considered product categories, instead of exact products in the experiments. There are unique 335 third-level categories in the target dataset. Customers within our sample bought 47.63 distinct third-level categories on average with the standard deviation of 26.07 and average basket size for our sample is 11.43 with the standard deviation of 6.61.

8.3.2. Data Pre-processing

This process mainly includes outlier elimination and feature extraction steps. One way to define an outlier is using interquartile range (IQR). If a value is more than (Q3 + 3*IQR) or less than (Q1 – 3*IQR), then it is considered as an extreme outlier. Note that, Q1 and Q3 represent the lower and upper quartiles, respectively. Outlier detection is applied to the "number of preferences" (number of unique products purchased). However, this variable may have different distribution for different customers. Because of that, we applied the outlier detection for each customer separately. Therefore, for each customer, we remove the instances having an outlier value for the corresponding variable.

We considered timestamp of transactions as one of the features. Since timestamp is a composite variable in our dataset, we extracted two distinct features from this variable, which are day of the week and time period of the day. We also categorized the values of these variables and Table 19 shows both categorical and actual values of each feature.

Table 19: Time features

Feature	Categorical values	Range of actual values
Day	Weekend	Monday to Friday
	Weekday	Saturday, Sunday
Time	Morning	08:00 to 11:59
	Afternoon	12:00 to 17:59
	Evening	18:00 to 20:59
	Night	21:00 to 22:59

In the experiments, categorical values were used for the features of time and day. Moreover, simple n-visit moving average feature which is the average number of purchased unique products in last n visit is calculated. We selected n as 5. The number of purchased unique products in last visit is also computed and formed as another feature.

8.3.3. Experimental Settings and Design

To generate the training and test sets, we sorted each customer's shopping trips according to their timestamps. For each customer's purchase history, we use the first 80% of the visits as training and the latter 20% as test data for all set of experiments. The training data is utilized to train the predictive model of our approach. In the testing set, for each visit of the customer a ranked list of recommended products is generated and the recommended products are compared with the actual ones purchased by the customer in the test set to compute the corresponding evaluation metric.

8.3.4. Evaluation Metrics

For the performance evaluation, we used F-measure which is the harmonic mean of precision and recall. Recall is computed in formula 14, and precision is defined as follows:

$$Precision@N = \frac{|rec@N \ \cap \ purc|}{|rec@N|} \tag{15}$$

where rec@N donates the top-N recommended products for the test instance and purc is the actual product set that the customer has purchased in the same test instance.

We first computed F-measure for each customer separately by averaging all computed values in the customer's test set, and then average these customers' personal values to get overall F-measure value for a given recommendation list size, N.

8.3.5. Recommendation Method

In the experimental evaluation, Most Popular (MP) technique was used to generate top-N recommendation lists for the customers' visits in the test set. The reasons behind this choice are that it is simple and well-known method and it was also employed in similar studies on predicting grocery shopping lists (Sarwar et al. 2000; Cumby et al. 2004; Cumby et al. 2005). This approach may not be so efficient in the recommendation performance, but it is out of the scope in this study. MP technique is only considered as a reasonable baseline to predict top-N recommendation lists and to evaluate the effectiveness of our proposed approach.

8.3.6. Benchmark Method

To compare the performance of our proposed approach on the recommendation efficiency, we selected last visit's number of products as a benchmark. This method identifies the number of products purchased in the previous visit of the customer as the recommendation list size for the next visit of that customer.

8.3.7. Results

The experimental results of our proposed approach with different ML methods and benchmark methods are listed in Table 20. In the table, there are three different methods that adjust recommendation list size using different machine learning techniques and a benchmark method that adjusts the recommendation list size by using last visit's item count. In addition to these, only two relevant methods using fixed recommendation list size were added to the table for the purpose of comparison.

Table 20: Comparison of methods for recommendation list size

Method	**Avg Rec. List Size**	**F-Measure**
Adjusted with J48	8.96	0.4
Adjusted with KNN	8.94	0.39
Adjusted with Multi. Reg.	11.98	0.44
Last Visit N	11.18	0.4
Fixed N	9	0.37
Fixed N	12	0.38

In Table 20, "Avg Rec. List Size" column indicates the average number of recommended items to per customer. The results listed in Table 20 reveals the following findings:

- Our approach with J48 classifier achieves 8% improvement in terms of F-measure, compared to the standard method using fixed recommendation list size (N=9). Similarly, our approach with KNN classifier also improves F-measure by 5%, compared to related standard method with the same size of recommendation list.
- Our approach with multiple regression achieves 16% improvement in terms of F-measure, compared to the standard method with the same size of recommendation list (N=12).
- Our approach with J48 classifier achieves same F-measure as the method of Last Visit N by shortening the average recommendation list size by nearly 20% (i.e., with average recommendation list size of 8.96, down from 11.18 of Last Visit N approach).
- When we compare our approach based on KNN classifier with the method of Last Visit N, our approach provides 20% reduction in the recommendation list size with only a small amount of accuracy loss (2.5%)
- By increasing recommendation list size by 7% (from 11.18 of Last Visit N approach to 11.98), our approach based on multiple regression achieves 10% improvement in the recommendation performance, compared to the method of Last Visit N.

The above findings indicate that for the same size of recommendation list, our approach provides a better quality of recommendation than the standard one does. It also outperforms the benchmark, Last Visit N method in efficiency by providing reduction in the recommendation list size while preserving the accuracy.

8.4. Conclusion

This chapter proposes an approach to dynamically adjust the recommendation list size of a customer with multiple preferences in order to improve the recommendation quality. By taking advantage of customers' previous preferences and machine learning techniques, the proposed approach adjusts the number of items that will be preferred according to the changing conditions. We evaluated our approach by conducting extensive experiments on a real-life dataset in the grocery shopping domain. According to the experimental results, our approach with all three selected machine learning techniques outperforms the traditional and widely used standard approach in effectiveness and it also provides better performance than the benchmark, Last Visit N method in efficiency by shortening the recommendation list while maintaining the effectiveness.

CHAPTER 9

CONCLUSION

In this thesis, we have investigated how to extract knowledge about customers and products, and predict customers' shopping behaviors and characteristics in the grocery retail industry. For these purposes, we have proposed and evaluated an array of approaches and methods.

In Chapter 3, a new type of RFM model has been proposed for the customer segmentation and it has been applied to grocery retail industry. In this manner, this chapter can contribute to prior literature by providing valuable insights about behaviors of different customer types in the grocery industry. Grocery companies can benefit from the proposed methodology in this study to identify different groups of customers and profile them. This chapter also proposes important practical directions and a wide range of managerial implications including example industry-specific promotion strategies for different customer types.

In Chapter 4, a product segmentation methodology based on customers' purchase transactions has been proposed to group products into meaningful segments in the grocery retail industry. By using this proposed methodology, managers can gain useful insights about products' sale characteristics and identify different product groups. Further, obtained product segments through this methodology can be used by enterprises in streamlining both their marketing and supply chain efforts. In addition to these, the results of this study give a clear view of products' sale characteristics in the grocery retail domain, especially Turkish market.

In Chapter 5, a hybrid approach has been proposed to predict customers' individual purchase behavior. The contributions of this chapter are as follows: (1) as far as we know, the proposed approach is the first hybrid approach combining two state-of-the-art predictive modeling approaches for customer behavior: individual-level and segment based; (2) we introduce three unique features describing customers' purchasing behavior and demonstrate how to utilize them in the customer segmentation; (3) it presents a comparative performance analysis among the proposed hybrid, individual-level and segment-based approaches in terms of both prediction coverage and accuracy; (4) it has shed some light on the performance of popular machine learning techniques for the prediction of customer purchase behavior.

In chapter 6, a comparison of predictive modeling approaches for the shopping list prediction has been proposed. The contribution of this chapter is to demonstrate the performance of predictive modeling approaches with various machine learning classifiers in the shopping list prediction problem. The results and findings can help researchers to understand different aspects of using predictive modeling approaches in the shopping list prediction.

In Chapter 7, a modified version of ARHR metric has been introduced for the purpose of ranking evaluation of multiple preferences in top-n recommendation and prediction lists. By utilizing this proposed metric, it is possible to assess the quality of top-n recommendation and prediction lists.

In Chapter 8, a novel approach has been proposed to dynamically adjust the recommendation list size for customers' multiple preferences. The proposed approach has the advantages of recommending a reasonable number of items without overwhelming the users and increasing recommendation quality. In this manner, recommender systems can utilize this approach to determine the optimal number of items to be recommended for especially the cases where the customer has multiple preferences at a time.

By utilizing the contributions of all these studies, the management of a grocery company can effectively adjust its CRM and marketing actions and efficiently allocate its resources. Thus, such a company can ultimately improve the quality of the service and customer relationships. Such improvements may make the customers highly satisfied, and increase their loyalty, and therefore the company can maximize its profits and maintain its competitive advantage. Despite its notable contributions and benefits, there are some issues and limitations that need to be revisited in future studies.

In Chapter 3, the proposed segmentation model was applied to only the customers of a grocery chain operating in Turkey. The samples from other countries may have different shopping characteristics. In this context, further research might extend this study to different countries to ascertain whether the findings are consistent and what can be done to make them more generalizable. Additionally, for future studies in the grocery retail industry, the proposed LRFMP model may be further enhanced by adding new attributes related to the customer's behavior, such as the number of products purchased, the number of perishable or non-perishable products purchased, to interpret the customer's behavior more soundly. Similarly, the proposed methodology in chapter 4 utilizes the features of FMC model to segment products, and further studies can use different set of features that better represent products' characteristics.

In Chapters 3-5, the performance of K-means and Ward's clustering methods have been tested to segment customers or products into groups. As a future study, it can be helpful to employ different clustering techniques to compare their performances with ones used in this dissertation. Moreover, throughout the thesis, different machine learning techniques have been employed in different studies. A possible direction for future research might be to test the performance of machine learning algorithms that different from the ones used in this dissertation.

Finally, approaches, models and methods proposed in Chapters 3-6 have been applied for the domain of supermarket shopping. However, they can potentially be applied to other domains as well, and the selected feature sets may be different depending on the application domain and the chosen data set. In further studies, it is worthy applying these proposed approaches and methods to other application domains to validate the findings drawn from this dissertation. Likewise, the applicability of the proposed metric in Chapter 7 and the effectiveness of the presented approach in Chapter 8 were demonstrated in the supermarket shopping. Therefore, possible directions in the future work might be to evaluate this proposed metric and approach in other application domains in order to ascertain the results are generalizable.

REFERENCES

Agrawal, R., Imieliński, T. & Swami, A., 1993. Mining association rules between sets of items in large databases. In *Acm sigmod record*. ACM, pp. 207–216.

Allenby, G.M. & Rossi, P.E., 1998. Marketing models of consumer heterogeneity. *Journal of econometrics*, 89(1), pp.57–78.

Apte, C. et al., 2002. Business applications of data mining. *Communications of the ACM*, 45(8), pp.49–53.

Apte, C. et al., 2001. Segmentation-based modeling for advanced targeted marketing. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, pp. 408–413.

Apte, C. V et al., 2003. Data-intensive analytics for predictive modeling. *IBM Journal of Research and Development*, 47(1), pp.17–23.

Bauer, C.L., 1988. A direct mail customer purchase model. *Journal of Direct Marketing*, 2(3), pp.16–24.

Bibelnieks, E. & Campbell, D., 2000. Mail Stream Streamlining. *Catalog Age*, 17(12), pp.118–120.

Bobadilla, J. et al., 2013. Recommender systems survey. *Knowledge-Based Systems*, 46, pp.109–132.

Breiman, L. et al., 1984. *Classification and regression trees*, CRC press.

Breiman, L., 2001. Random forests. *Machine learning*, 45(1), pp.5–32.

Brijs, T. et al., 2000. A data mining framework for optimal product selection in retail supermarket data: the generalized PROFSET model. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, pp. 300–304.

Brijs, T. et al., 2004. Building an association rules framework to improve product assortment decisions. *Data Mining and Knowledge Discovery*, 8(1), pp.7–23.

Brijs, T. et al., 1999. Using association rules for product assortment decisions: A case study. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, pp. 254–260.

Bult, J.R. & Wansbeek, T., 1995. Optimal selection for direct mail. *Marketing Science*, 14(4), pp.378–394.

Burke, R., 2002. Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12(4), pp.331–370.

Campbell, D. et al., 2001. Optimizing customer mail streams at Fingerhut. *Interfaces*, 31(1), pp.77–90.

Campos, P.G., Díez, F. & Cantador, I., 2014. Time-aware recommender systems: a comprehensive survey and analysis of existing evaluation protocols. *User Modeling and User-Adapted Interaction*, 24(1-2), pp.67–119.

Catlett, J., 1991. On changing continuous attributes into ordered discrete attributes. In *Machine learning—EWSL-91*. Springer, pp. 164–178.

Chan, C.C.H., 2008. Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer. *Expert systems with applications*, 34(4), pp.2754–2762.

Chang, H.-C. & Tsai, H.-P., 2011. Group RFM analysis as a novel framework to discover better customer consumption behavior. *Expert Systems with Applications*, 38(12), pp.14499–14513.

Chang, H.H. & Tsay, S.F., 2004. Integrating of SOM and K-mean in data mining clustering: An empirical study of CRM and profitability evaluation.

Chapelle, O. et al., 2009. Expected reciprocal rank for graded relevance. In *Proceedings of the 18th ACM conference on Information and knowledge management*. ACM, pp. 621–630.

Chen, K., Hu, Y.-H. & Hsieh, Y.-C., 2015. Predicting customer churn from valuable B2B customers in the logistics industry: a case study. *Information Systems and e-Business Management*, 13(3), pp.475–494.

Chen, M.-C., Chiu, A.-L. & Chang, H.-H., 2005. Mining changes in customer behavior in retail marketing. *Expert Systems with Applications*, 28(4), pp.773–781.

Chen, M.-C. & Lin, C.-P., 2007. A data mining approach to product assortment and shelf space allocation. *Expert Systems with Applications*, 32(4), pp.976–986.

Chen, Y.-S. et al., 2012. Identifying patients in target customer segments using a two-stage clustering-classification approach: A hospital-based assessment. *Computers in biology and medicine*, 42(2), pp.213–221.

Cheng, C.-H. & Chen, Y.-S., 2009. Classifying the segmentation of customer value via RFM model and RS theory. *Expert systems with applications*, 36(3), pp.4176–4184.

Cheng, Y., Yin, L. & Yu, Y., 2014. LorSLIM: Low Rank Sparse Linear Methods for Top-N Recommendations. In *Data Mining (ICDM), 2014 IEEE International Conference on*. IEEE, pp. 90–99.

Cheung, Y.-M., 2003. k∗ -Means: A new generalized k-means clustering algorithm. *Pattern Recognition Letters*, 24(15), pp.2883–2893.

Chiu, S. & Tavella, D., 2008. *Data mining and market intelligence for optimal marketing returns*, Routledge.

Chuang, Y.-F., Chia, S.-H. & Yih Wong, J., 2013. Customer value assessment of pharmaceutical marketing in Taiwan. *Industrial Management & Data Systems*, 113(9), pp.1315–1333.

Cortes, C. & Vapnik, V., 1995. Support-vector networks. *Machine learning*, 20(3), pp.273–297.

Cumby, C. et al., 2005. Building intelligent shopping assistants using individual consumer models. In *Proceedings of the 10th international conference on Intelligent user interfaces*. ACM, pp. 323–325.

Cumby, C. et al., 2004. Predicting customer shopping lists from point-of-sale purchase data. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, pp. 402–409.

Davidson, I., 2002. Understanding K-means non-hierarchical clustering. *Computer Science Department of State University of New York (SUNY), Albany*.

Deshpande, M. & Karypis, G., 2004. Item-based top-N recommendation algorithms. *Acm Transactions on Information Systems*, 22(1), pp.143–177. Available at: <Go to ISI>://WOS:000188554900006.

Dibb, S., 1998. Market segmentation: strategies for success. *Marketing Intelligence & Planning*, 16(7), pp.394–406.

Dibb, S. & Simkin, L., 1997. A program for implementing market segmentation. *Journal of Business & Industrial Marketing*, 12(1), pp.51–65.

Dursun, A. & Caber, M., 2016. Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis. *Tourism Management Perspectives*, 18, pp.153–160.

Fader, P.S., Hardie, B.G.S. & Lee, K.L., 2005. RFM and CLV: Using iso-value curves for customer base analysis. *Journal of Marketing Research*, 42(4), pp.415–430.

Faraone, M.F. et al., 2012. Using context to improve the effectiveness of segmentation and targeting in e-commerce. *Expert Systems with Applications*, 39(9), pp.8439–8451.

Faraone, M.F., Gorgoglione, M. & Palmisano, C., 2010. Contextual segmentation: using context to improve behavior predictive models in e-commerce. In *Data Mining Workshops (ICDMW), 2010 IEEE International Conference on*. IEEE, pp. 1053–1060.

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

Grewal, D. et al., 2011. Innovations in retail pricing and promotions. *Journal of Retailing*, 87, pp.S43–S52.

Gunawardana, A. & Shani, G., 2009. A Survey of Accuracy Evaluation Metrics of Recommendation Tasks. *Journal of Machine Learning Research*, 10, pp.2935–2962. Available at: <Go.

Güçdemir, H. & Selim, H., 2015. Integrating multi-criteria decision making and clustering for business customer segmentation. *Industrial Management & Data Systems*, 115(6), pp.1022–1040.

Ha, S.H., 2007. Applying knowledge engineering techniques to customer analysis in the service industry. *Advanced Engineering Informatics*, 21(3), pp.293–301.

Ha, S.H. & Park, S.C., 1998. Application of data mining tools to hotel data mart on the Intranet for database marketing. *Expert Systems with Applications*, 15(1), pp.1–31.

Hand, D.J., Mannila, H. & Smyth, P., 2001. *Principles of data mining*, MIT press.

Herlocker, J.L. et al., 2004. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, 22(1), pp.5–53.

Hosseini, S.M.S., Maleki, A. & Gholamian, M.R., 2010. Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. *Expert Systems with Applications*, 37(7), pp.5259–5264.

Hsu, C.-W., Chang, C.-C. & Lin, C.-J., 2003. A practical guide to support vector classification.

Huang, S.-C., Chang, E.-C. & Wu, H.-H., 2009. A case study of applying data mining techniques in an outfitter's customer value analysis. *Expert Systems with Applications*, 36(3), pp.5909–5915.

Hughes, A.M., 1996. Boosting response with RFM. *Marketing Tools*, 5(1), pp.4–10.

Hughes, A.M., 2005. *Strategic database marketing*, McGraw-Hill Pub. Co.

Hwang, H., Jung, T. & Suh, E., 2004. An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry. *Expert systems with applications*, 26(2), pp.181–188.

IGD, 2015. Global grocery markets: our forecasts to 2020. Available at: http://www.igd.com/Research/Retail/Global-grocery-markets-our-forecasts-to-2020/ [Accessed May 30, 2016].

Jain, A.K., 2010. Data clustering: 50 years beyond K-means. *Pattern recognition letters*, 31(8), pp.651–666.

Jain, A.K. & Dubes, R.C., 1988. *Algorithms for clustering data*, Prentice-Hall, Inc.

Jain, A.K., Murty, M.N. & Flynn, P.J., 1999. Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3), pp.264–323.

Jiang, T., Tan, L. & Kim, S., 2013. Personalized defect prediction. In *Automated Software Engineering (ASE), 2013 IEEE/ACM 28th International Conference on*. IEEE, pp. 279–289.

Jiang, T. & Tuzhilin, A., 2007. Dynamic micro targeting: Fitness-based approach to predicting individual preferences. In *Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on*. IEEE, pp. 173–182.

Jiang, T. & Tuzhilin, A., 2009. Improving personalization solutions through optimal segmentation of customer bases. *IEEE transactions on knowledge and data engineering*, 21(3), pp.305–320.

Jiang, T. & Tuzhilin, A., 2006. Segmenting customers from population to individuals: Does 1-to-1 keep your customers forever? *IEEE Transactions on knowledge and data engineering*, 18(10), pp.1297–1311.

Jiawei, H. & Kamber, M., 2001. Data mining: concepts and techniques. *San Francisco, CA, itd: Morgan Kaufmann*, 5.

Kahan, R., 1998. Using database marketing techniques to enhance your one-to-one marketing initiatives. *Journal of Consumer Marketing*, 15(5), pp.491–493.

Kamakura, W.A. & Wedel, M., 1999. Market segmentation: conceptual and methodological foundations. *Kluwer Academic Press, New York*.

Kang, Z. & Cheng, Q., 2016. Top-N Recommendation with Novel Rank Approximation. *arXiv preprint arXiv:1602.07783*.

Kao, Y.-T. et al., 2011. A case study of applying LRFM model and clustering techniques to evaluate customer values. *Journal of Statistics and Management Systems*, 14(2), pp.267–276.

Kaymak, U., 2001. Fuzzy target selection using RFM variables. In *IFSA World Congress and 20th NAFIPS International Conference, 2001. Joint 9th*. IEEE, pp. 1038–1043.

Kekalainen, J., 2005. Binary and graded relevance in IR evaluations - Comparison of the effects on ranking of IR systems. *Information Processing & Management*, 41(5), pp.1019–1033. Available at: <Go.

Ketchen, D.J. & Shook, C.L., 1996. The application of cluster analysis in strategic management research: an analysis and critique. *Strategic management journal*, 17(6), pp.441–458.

Khajvand, M. et al., 2011. Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study. *Procedia Computer Science*, 3, pp.57–63.

Kim, E., Kim, W. & Lee, Y., 2003. Combination of multiple classifiers for the customer's purchase behavior prediction. *Decision Support Systems*, 34(2), pp.167–175.

Kotler, P., 1997. *Marketing management: Analysis, planning, implementation, and control*, NJ: Prentice Hall, Upper Saddle River.

Kotsiantis, S. & Kanellopoulos, D., 2006. Discretization techniques: A recent survey. *GESTS International Transactions on Computer Science and Engineering*, 32(1), pp.47–58.

Lawrence, R.D. et al., 2001. Personalization of supermarket product recommendations. In *Applications of Data Mining to Electronic Commerce*. Springer, pp. 11–32.

Li, D.-C., Dai, W.-L. & Tseng, W.-T., 2011. A two-stage clustering method to analyze customer characteristics to build discriminative customer management: A case of textile manufacturing business. *Expert Systems with Applications*, 38(6), pp.7186–7191.

Li, W. et al., 2011. Design and evaluation of a command recommendation system for software applications. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 18(2), p.6.

Lin, H.-T., Lin, C.-J. & Weng, R.C., 2007. A note on Platt's probabilistic outputs for support vector machines. *Machine learning*, 68(3), pp.267–276.

Liu, X. et al., 2013. Personalized point-of-interest recommendation by mining users' preference transition. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*. ACM, pp. 733–738.

Lombardi, S., Gorgoglione, M. & Panniello, U., 2013. The effect of context on misclassification costs in e-commerce applications. *Expert Systems with Applications*, 40(13), pp.5219–5227.

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Oakland, CA, USA., pp. 281–297.

Marcus, C., 1998. A practical yet meaningful approach to customer segmentation. *Journal of consumer marketing*, 15(5), pp.494–504.

McDonald, M. & Dunbar, I., 2004. *Market segmentation: How to do it, how to profit from it*, Butterworth-Heinemann.

Michaud, P., 1997. Clustering techniques. *Future Generation Computer Systems*, 13(2), pp.135–147.

Miglautsch, J.R., 2000. Thoughts on RFM scoring. *Journal of Database Marketing & Customer Strategy Management*, 8(1), pp.67–72.

Milligan, G.W. & Cooper, M.C., 1988. A study of standardization of variables in cluster analysis. *Journal of classification*, 5(2), pp.181–204.

Moffat, A. & Zobel, J., 2009. Rank-Biased Precision for Measurement of Retrieval Effectiveness. *Acm Transactions on Information Systems*, 27(1). Available at: <Go.

Mooi, E. & Sarstedt, M., 2010. Cluster analysis. In Springer.

Newell, F., 1997. *The New Rules of Marketing: How to Use One-To-One Relationship Marketing to Be the Leader in Your Industry*, New York: McGraw-Hill, Inc.

Nikumanesh, E. & Albadvi, A., 2014. Customer's life–time value using the RFM model in the banking industry: a case study. *International Journal of Electronic Customer Relationship Management*, 8(1-3), pp.15–30.

Ning, X. & Karypis, G., 2011. Slim: Sparse linear methods for top-n recommender systems. In *Data Mining (ICDM), 2011 IEEE 11th International Conference on*. IEEE, pp. 497–506.

Ostuni, V.C. et al., 2013. Top-N recommendations from implicit feedback leveraging linked open data. In *Proceedings of the 7th ACM conference on Recommender systems*. ACM, pp. 85–92.

Palmisano, C., Tuzhilin, A. & Gorgoglione, M., 2007. User profiling with hierarchical context: An e-retailer case study. In *International and Interdisciplinary Conference on Modeling and Using Context*. Springer, pp. 369–383.

Palmisano, C., Tuzhilin, A. & Gorgoglione, M., 2008. Using context to improve predictive modeling of customers in personalization applications. *IEEE transactions on knowledge and data engineering*, 20(11), pp.1535–1549.

Peker, S. & Kocyigit, A., 2016a. An Adjusted Recommendation List Size Approach for Users' Multiple Item Preferences. In *International Conference on Artificial Intelligence: Methodology, Systems, and Applications*. Springer, pp. 310–319.

Peker, S. & Kocyigit, A., 2016b. mRHR: A Modified Reciprocal Hit Rank Metric for Ranking Evaluation of Multiple Preferences in Top-N Recommender Systems. In *International*

Conference on Artificial Intelligence: Methodology, Systems, and Applications. Springer, pp. 320–329.

Peker, S., Kocyigit, A. & Eren, P.E., 2017. LRFMP model for customer segmentation in the grocery retail industry: a case study. *Marketing Intelligence & Planning*, 35(4), pp.1–16.

Platt, J., 1999. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, 10(3), pp.61–74.

Pu, P. et al., 2011. Usability guidelines for product recommenders based on example critiquing research. In *Recommender Systems Handbook*. Springer, pp. 511–545.

Punj, G. & Stewart, D.W., 1983. Cluster analysis in marketing research: Review and suggestions for application. *Journal of marketing research*, pp.134–148.

Quinlan, J.R., 2014. *C4. 5: programs for machine learning*, Elsevier.

Reinartz, W.J. & Kumar, V., 2000. On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing. *Journal of marketing*, 64(4), pp.17–35.

Rendle, S., Freudenthaler, C. & Schmidt-Thieme, L., 2010. Factorizing personalized markov chains for next-basket recommendation. In *Proceedings of the 19th international conference on World wide web*. ACM, pp. 811–820.

Rucker, D.D., McShane, B.B. & Preacher, K.J., 2015. A researcher's guide to regression, discretization, and median splits of continuous variables. *Journal of Consumer Psychology*, 25(4), pp.666–678.

Rumelhart, D.E., Hinton, G.E. & Williams, R.J., 1988. Learning representations by back-propagating errors. *Cognitive modeling*, 5(3), p.1.

Safari, F. et al., 2016. Customer lifetime value determination based on RFM model. *Marketing Intelligence & Planning*, 34(4).

Sarstedt, M. & Mooi, E., 2014. A Concise Guide to Market Research. *The Process, Data, and*.

Sarwar, B. et al., 2000. Analysis of recommendation algorithms for e-commerce. In *Proceedings of the 2nd ACM conference on Electronic commerce*. ACM, pp. 158–167.

Smith, W.R., 1956. Product differentiation and market segmentation as alternative marketing strategies. *The Journal of Marketing*, pp.3–8.

Sokolova, M. & Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), pp.427–437.

Stone, B. & Jacobs, R., 2007. *Successful direct marketing methods*, McGraw Hill Professional.

Thorndike, R.L., 1953. Who belongs in the family? *Psychometrika*, 18(4), pp.267–276.

Tsai, C.-Y. & Chiu, C.-C., 2004. A purchase-based market segmentation methodology. *Expert Systems with Applications*, 27(2), pp.265–276.

USDA, 2016. Global Food Industry. Available at: http://www.ers.usda.gov/topics/international-markets-trade/global-food-markets/global-food-industry.aspx [Accessed May 30, 2016].

Wan, S. et al., 2015. Next Basket Recommendation with Neural Networks. In *RecSys Posters*.

Wang, P. et al., 2015. Learning hierarchical representation model for nextbasket recommendation. In *Proceedings of the 38th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, pp. 403–412.

Ward Jr, J.H., 1963. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301), pp.236–244.

Webster Jr, F.E., 1992. The changing role of marketing in the corporation. *The Journal of Marketing*, pp.1–17.

Wedel, M. & Kamakura, W.A., 2000. *Market segmentation: Conceptual and methodological foundations*, Kluwer.

Wei, J.-T. et al., 2012. A case study of applying LRFM model in market segmentation of a children's dental clinic. *Expert Systems with Applications*, 39(5), pp.5529–5533.

Wei, J.-T. et al., 2013. Customer relationship management in the hairdressing industry: An application of data mining techniques. *Expert Systems with Applications*, 40(18), pp.7513–7518.

Wiedmann, K., Hennigs, N. & Siebels, A., 2009. Value-based segmentation of luxury consumption behavior. *Psychology & Marketing*, 26(7), pp.625–651.

Wind, Y., 1978. Issues and advances in segmentation research. *Journal of marketing research*, pp.317–337.

Witten, I.H. et al., 2016. *Data Mining: Practical machine learning tools and techniques*, Morgan Kaufmann.

Witten, I.H. & Frank, E., 2005. *Data Mining: Practical machine learning tools and techniques*, Morgan Kaufmann.

Wong, A. & Sohal, A., 2002. An examination of the relationship between trust, commitment and relationship quality. *International Journal of Retail & Distribution Management*, 30(1), pp.34–50.

Wu, X. et al., 2008. Top 10 algorithms in data mining. *Knowledge and information systems*, 14(1), pp.1–37.

Yao, Z. et al., 2014. Combining visual customer segmentation and response modeling. *Neural Computing and Applications*, 25(1), pp.123–134.

Yeh, I.-C., Yang, K.-J. & Ting, T.-M., 2009. Knowledge discovery on RFM model using Bernoulli sequence. *Expert Systems with Applications*, 36(3), pp.5866–5871.

Yu, F. et al., 2016. A dynamic recurrent model for next basket recommendation. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, pp. 729–732.

Zabkowski, T.S., 2016. RFM approach for telecom insolvency modeling. *Kybernetes*, 45(5), pp.815–827.

Zeileis, A. et al., 2004. kernlab-an S4 package for kernel methods in R. *Journal of statistical* *software*, 11(9), pp.1–20.

Zheng, N. & Li, Q.D., 2011. A recommender system based on tag and time information for social tagging systems. *Expert Systems with Applications*, 38(4), pp.4575–4587.

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name	: Peker, Serhat
Nationality	: Turkish
Date and place of birth	: 13.01.1986, Edremit
Marital Status	: Married
Phone	: +90 312 210 7869
E-mail	: serhatpeker77@gmail.com

EDUCATION

Degree	Institution	Year
Ph.D	Middle East Technical University, Information Systems	ongoing
MS	Middle East Technical University, Information Systems	2011
BS	Izmir University of Economics, Business Administration	2009
BS	Izmir University of Economics, Computer Engineering	2009

WORK EXPERIENCE

Year	Place	Enrollment
2009-2017	Middle East Technical University	Research and Teaching Assistant
2012-2015	Stratek Strategic Technologies	Researcher and Analyst

PUBLICATIONS

SCI & SCI-E & SSCI & AHCI International Refereed Journals

Peker S., Kocyigit, A., & Eren, P. E. (2017). A hybrid approach for predicting customers' individual purchase behavior. *Kybernetes*, (Accepted and Revisions Underway).

Peker S., Kocyigit, A., & Eren, P. E. (2017). LRFMP model for customer segmentation in the grocery retail industry: a case study. *Marketing Intelligence and Planning*, 35(4), 544-559.

Peker, S., Kucukozer-Cavdar, S., & Cagiltay, K. (2016). Exploring the relationship between web presence and web usability for universities: A case study from Turkey. *Program*, 50(2), 157-174.

International Conference Proceedings

Peker, S., & Kocyigit, A. (2016, September). mRHR: A Modified Reciprocal Hit Rank Metric for Ranking Evaluation of Multiple Preferences in Top-N Recommender Systems. In *International Conference on Artificial Intelligence: Methodology, Systems, and Applications* (pp. 320-329). Springer International Publishing.

Peker, S., & Kocyigit, A. (2016, September). An Adjusted Recommendation List Size Approach for Users' Multiple Item Preferences. In *International Conference on Artificial Intelligence: Methodology, Systems, and Applications* (pp. 310-319). Springer International Publishing.

Peker, S., Gurbuz, O., & Temizel, T. T. (2014, June). A SOAp-based social traffic network system for smart phone users. In *Innovations in Intelligent Systems and Applications (INISTA) Proceedings*, 2014 IEEE International Symposium on (pp. 293-298). IEEE.

Peker, S., & Arkilic, I. G. (2013). Developing A Modular Web based System for Special Education: Advantages and Challenges. *Procedia-Social and Behavioral Sciences*, *83*, 1091-1094.

Arkilic, I. G., **Peker, S.**, & Uyar, M. E. (2013). Students' preferences of communication tools for group projects in a computer-supported collaborative learning environment: A survey. *Procedia-Social and Behavioral Sciences*, *83*, 1121-1125.

Medeni, I. T., **Peker, S.**, & Uyar, M. E. (2012). Assessment of online news agencies by using a knowledge visualization model: expected benefits. *In 2nd World Conference on Information Technology (WCIT) Proceedings*, *1*, 1202-1207.

Medeni, I. T., **Peker, S.**, & Uyar, M. E. (2011, May). A knowledge visualization model for evaluating Internet news agencies on conflicting news. In MIPRO, 2011 Proceedings of the 34th International Convention (pp. 850-853). IEEE.

Peker S., Arkilic I.G., Bekiroglu B., Ergun E. & Sahin Y. G. (2010). MANSE: Managing Special Education and Rehabilitation Centers via Central Database. In *10th International Educational Technology Conference (IETC) Proceedings*, *2*, 920-926.