S. GÜNEY

A COMBINED APPROACH OF CLUSTERING AND ASSOCİATİON RULE MİNİNG FOR CUSTOMER PROFILING IN VIDEO ON DEMAND SERVICES

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

OF

ATILIM UNIVERSITY

SİNEM GÜNEY

DOCTOR OF PHILOSOPHY THESIS

IN

THE DEPARTMENT OF SOFTWARE ENGINEERING

JUNE 2020

ATILIM UNIVERSITY 2020

A COMBINED APPROACH OF CLUSTERING AND ASSOCIATION RULE MINING FOR CUSTOMER PROFILING IN VIDEO ON DEMAND SERVICES

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

OF

ATILIM UNIVERSITY

BY

SİNEM GÜNEY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR

THE DEGREE OF DOCTOR OF PHILOSOPHY

IN

THE DEPARTMENT OF SOFTWARE ENGINEERING

JUNE 2020

Approval of the Graduate School of Natural and Applied Sciences, Atilim University.

____

Prof. Dr. Ali Kara
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of **Doctor of Philosophy in Software Engineering, Atilim University**.

____

Prof. Dr. Ali Yazıcı
Head of Department

This is to certify that we have read the thesis "A Combined Approach of Clustering and Association Rule Mining for Customer Profiling in Video on Demand Services" submitted by "Sinem GÜNEY" and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Doctor of Philosophy.

__________________________	__________________________
Asst. Prof. Dr. Serhat PEKER Co-Supervisor	Asst. Prof. Dr. Çiğdem TURHAN Supervisor

Examining Committee Members:

Prof. Dr. Ali Yazıcı Software Eng. Department, Atilim University	__________________________
Asst. Prof. Dr. Çiğdem Turhan Software Eng. Department, Atilim University	__________________________
Prof. Dr. Murat Koyuncu Information Sys. Eng. Department, Atilim University	__________________________
Assoc. Prof. Dr. Erol Özçelik Psychology Department, Çankaya University	__________________________
Asst. Prof. Dr. Atila Bostan Software Eng. Department, Ankara Bilim University	__________________________

Date: June 11, 2020

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: Sinem GÜNEY

Signature:

ABSTRACT

A COMBINED APPROACH OF CLUSTERING AND ASSOCIATION RULE MINING FOR CUSTOMER PROFILING IN VIDEO ON DEMAND SERVICES

Güney, Sinem

Ph.D., Department of Software Engineering

Supervisor: Asst. Prof. Dr. Çiğdem TURHAN

Co-Supervisor: Asst. Prof. Dr. Serhat PEKER

June 2020, 67 pages

Today, IPTV (Internet Protocol Television) service providers offer VoD (Video on Demand) services as part of their business initiative toward generating more revenue. To do this, they need to know about customer behaviors and expectations. Such information related to users is stored in CRM (Customer Relationship Management) systems. Against this backdrop, the present work aims to analyze customers in VoD services with applying clustering and Association Rule Mining techniques. The LRFMP (Length, Recency, Frequency, Monetary, and Periodicity) model is applied to find out the customer behaviors, whereas the k-means clustering algorithms allow for determining the number of clusters and customer profiles. As a result, four different customer groups are identified, namely as “consuming and most valuable”, “less consuming and less valuable”, “less consuming but loyal”, and “neither loyal nor valuable”. A major source of information for this study is the content type or genre as regards the content category and rental preferences of subscribers. To this end, the association rule algorithm (Apriori) is employed to predict the customers’ potential rentals. A combined approach as such would be useful for IPTV service providers to further shed light on precise customer behaviors and preferences, thus allowing to create more targeted marketing strategies for each category of subscribers in order to improve customer satisfaction and increase revenues in the long run.

Keywords: Data Mining, CRM, Customer Behavior, Customer Segmentation, Customer Profiling, Association Rule Mining, Marketing Strategies, IPTV Service Provider, Video on Demand Services, RFM Model

ÖZ

A COMBINED APPROACH OF CLUSTERING AND ASSOCIATION RULE MINING FOR CUSTOMER PROFILING IN VIDEO ON DEMAND SERVICES

Güney, Sinem

Doktora, Yazılım Mühendisliği Bölümü

Tez Danışmanı: Dr. Öğr. Üyesi Çiğdem TURHAN

Ortak Tez Danışmanı: Dr. Öğr. Üyesi Serhat PEKER

Haziran 2020, 67 sayfa

Günümüzde IPTV (İnternet Protokol Televizyonu) hizmet sağlayıcıları, daha fazla gelir elde etmek için yaptıkları iş girişimlerinin bir parçası olarak VoD (İsteğe Bağlı Video) hizmetleri sunmaktadır. Bunu yapmak için müşteri davranışları ve beklentileri hakkında bilgi sahibi olmaları gerekir. Kullanıcılarla ilgili bu tür bilgiler CRM (Müşteri İlişkileri Yönetimi) sistemlerinde saklanır. Bu çerçevede, bu çalışma VoD hizmetlerindeki müşterileri kümeleme ve dernek kuralı madenciliği teknikleri uygulayarak analiz etmeyi amaçlamaktadır. LRFMP (Uzunluk, Yenilik, Frekans, Parasal ve Periyodiklik) modeli, müşteri davranışlarını bulmak için uygulanır ve k-küme algoritmaları kümelerin ve müşteri profillerinin belirlenmesini sağlar. Sonuç olarak, dört farklı müşteri grubu, "tüketen ve en değerli", "daha az tüketen ve daha az değerli", "daha az tüketen ama sadık" ve "ne sadık ne de değerli" olarak tanımlanır. Bu çalışma için önemli bir bilgi kaynağı, abonelerin içerik kategorisi ve kiralama tercihleriyle ilgili olarak içerik türü veya türüdür. Bu amaçla, müşterilerin potansiyel kiralarını tahmin etmek için ilişkilendirme kuralı algoritması (Apriori) kullanılır. Bu çalışmada, IPTV hizmet sağlayıcılarının hassas müşteri davranışlarına ve tercihlerine daha fazla ışık tutması için birleşik bir yaklaşım yararlı olacaktır, bu da müşteri memnuniyetini artırmak ve uzun vadede gelirleri artırmak için her bir abone kategorisi için daha hedefli pazarlama stratejileri oluşturulmasına izin verir.

Anahtar Kelime: Veri Madenciliği, CRM, Müşteri Davranışı, Müşteri Segmentasyonu, Müşteri Profilleme, Veri Madenciliği Alanında Birliktelik Kuralları Yönetimi, Pazarlama Stratejileri, IPTV Servis Sağlayıcısı, İsteğe Bağlı İçerik Servisleri, RFM Modeli

I dedicate this thesis to my mother Ayşe Güney and my father Murat Güney for their unconditional love and constant support.

ACKNOWLEDGMENTS

First and foremost, I shall express my special appreciations to Prof. Dr. Ali Yazıcı for his wisdom and inspiration throughout the course of my doctorate program. I aspire to, one day, becoming of help and motivation to others the way he has.

Many thanks I owe to my supervisor, Asst. Prof. Dr. Çiğdem Turhan, for her continuous support and encouragement during my journey. She was a constant force behind widening my research in numerous ways, and I always felt her faith in me.

My further appreciations go to my co-advisor Asst. Prof. Dr. Serhat Peker for encouraging me to do more research, and for all the new things I experienced throughout this period. Without his valuable guidance and insightful advice, it would not have been possible to materialize the present research.

I thank each member of the esteemed jury: Prof. Dr. Murat Koyuncu, Assoc. Prof. Dr. Erol Özçelik and Asst. Prof. Dr. Atila Bostan for their invaluable suggestions and intuitive comments along with compelling questions that further shed light on the aspects of my studies unknown to me.

Many thanks to Mr. Payam Danesh for reviewing the thesis and offering useful improvements. Also, I thank the experienced members of the faculty staff Yasemin Kurt, Ayşegül Özeke, Özgür Soğukpınar Sidar and Çağatay Öztürk for their support during my Ph.D. journey.

I truly believe that this long journey would have been much more challenging had it not been for my devoted parents, Ayşe and Murat Güney, who nourished me with their unconditional love and limitless patience. Thanks also to my brother and his adorable wife and children. I love you so much for you always stand by me, matter not what decisions I make.

I would like to thank my relatives who remained by my side in this journey, my friends who provided sincere support, Mine Genç Tunç and her precious family. I will always regard you as family and my guiding light in life. I would like to express my special thanks to my colleagues for their precious time to help me in gathering data, and for their guidance. I cannot find the right words for your loving hearts, your friendships, and the way you make me feel.

Lastly – and, most certainly, not least - thanks to all others not mentioned here by name, who helped me enrich this study. Beyond a symbol of success, the present initiative became yet more meaningful with all of you who believed in me and shared my happiness. I am glad you are in my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CDN	Content Delivery Network
CH	Calinski-Harabasz
CLV	Customer Lifetime Value
CRM	Customer Relationship Management
DBI	Davies-Bouldin Index
DBSCAN	Density-based Spatial Clustering of Applications with Noise
FCM	Fuzzy c-means
IMDB	Internet Movie Database
IPTV	Internet Protocol Television
LHS	Left-hand-side
LRFMPG	Length, Recency, Frequency, Monetary, Periodicity, Genre
MPLS	Multiprotocol Label Switching
PAM	Partitioning Around Medoids
QoE	Quality of Experience
QoS	Quality of Service
RDBC	Recursive Density Based Clustering
RFM	Recency, Frequency and Monetary
RHS	Right-hand-side
RTI	Rental-Time Interval
Sci-Fi	Science fiction
SCM	Supply Chain Management
SD	Standard Deviation
SIL	Silhouette
SMS	Short Message Service
SOM	Self organizing maps
SSWC	Sum of Squares within cluster
STB	Set-Top Box
WSS	Within-cluster sum of squares

CHAPTER 1

INTRODUCTION

Today, the quality of network has become a main precedence for the broadcasting sector. These firms obtain most of their revenue from the connectivity services, whose provision requires large-scale infrastructural investments [1]. Nowadays, they offer numerous paying services such as Pay-Tv and VoD - video on demand services to subscribers. The IPTV (Internet Protocol Television) includes different service options to subscribers such as Linear TV, which means live TV (television) programming; Time-shifted TV, which provides content being shared for a limited period; and VoD services, which concentrate more on content for subscribers' preferences [2]. With the contribution of rapid changes in both technological developments and customer behaviors, service providers have no option but to evolve accordingly and offer new and more engaging services. In a sense, traditional service providers are increasingly becoming content providers and, henceforth, consider this merchandise as a distinguishing factor for commercial success over rivals. In this way, they can prevent mere revenue generation from connectivity and expand services beyond. Content offerings are now more critical for creating long-term customer relationships and maintaining profitability.

IPTV service providers have even evolved further to acquire – and, perhaps, merge with - content providing firms. With the new technological developments – both in terms of service infrastructure and consumers' end devices such as STB (Set-top Box), mobile phones, tablets, and personal computers, subscribers can receive more volume and faster access to content over different platforms via numerous applications (or apps). Beside wide-spreading IPTV services, the content itself is being used in other ways [3], forcing subscribers to move from the conventional landline and cable broadcasters to IPTV services [4]. Increasing numbers of subscribers are now demanding more and faster services from IPTV service providers, adding to expectations and impatience regarding more relevant contents to suit their choices. This, as a whole, can be both an opportunity and a threat to IPTV service providers depending on how well they know their customers' behaviors and how they offer their services to them.

Another point worth mentioning is the key element that IPTV service providers enhance the usage of smart devices for improvement of VoD service. According to statistics, the allocation of VoD market will be increased from USD 38.9 billion in 2019 to USD 87.1 billion by 2024 [5]. Also, reports state that the number of VoD service subscribers exceeds one billion, and the figures are expected to climb to 1.1 billion subscribers by 2024 [6].

Concerning CRM (Customer Relationship Management), it includes methodology and practices for usage of customer data in the business area, where decisions are made based on such data to gain competitive advantage in the marketing sector [7]. This sector asserts that, to create and implement targeted marketing strategies and promotional, one has to understand precise customer profiles and their rental and viewing preferences. The usage of CRM processes and applying data mining techniques together, then, can help to identify customer value and customer segmentation [8]. Thus, service providers seek to find out about precise customer profiles and their possible relationship with consumption preferences so as to be able to offer more relevant material with the ultimate goal to improve service quality, customer satisfaction and customer loyalty, add to the total revenue and the degree of risk mitigation due to client loss to other rivals.

During the literature review, it is discovered that customer classification and profiling studies in the related sector mostly utilize data mining and clustering techniques in successful CRM applications. Additionally, the RFM (Recency, Frequency, Monetary) model is widely used to identify different customer characteristics and preferences [9]. For this study, the LRFMP (Length, Recency, Frequency, Monetary, and Periodicity) model which is developed by Serhat Peker and his friends, is found to perfectly achieve such a goal as it is beneficial for improved notions concerning the client base, according to which targeted marketing strategies will be then created for each defined group [10]. Furthermore; by understanding the importance of content and the associated impact on IPTV service providers, the study strides to examine customer profiles and content rental transactions to find out about the likely links between these profiles and content category (genre) – a new and unique perspective for both the sector and the field scholar. To materialize the aim of this study, Association Rule Mining is employed to find out subscribers' content preferences in the metadata within the system. In this discipline, the Apriori algorithm, a division of Association Rule Mining algorithms, is used to analyze this categorical data for IPTV customers to uncover the relationship of rented content genres with these profiles.

Based on the facts mentioned above; the objectives of this study are as follows:

- Determining customer behavior by the real-life VoD service subscribers' data with LRFMP model;
- Analyzing customer data to form classifications and identify such profiles for the intended IPTV service provider with the clustering technique; and
- Understanding the relationship between these defined customer groups and the content preferences using Association Rule Mining technique;
- Suggesting targeted marketing strategies and campaigns to IPTV subscribers for increasing their satisfaction;
- Providing appropriate and better services to subscribers whose behaviors are known; and
- Maintaining subscriptions with IPTV for a longer period of time by means of understanding the subscribers' needs and expectations.

1.1 Research Questions

To achieve the objectives herein, the key research questions are:

RQ1: How can the VoD subscribers' behaviors and preferences be best identified?

RQ2: In what way can the customer profiles be identified to better understand their behavior?

RQ3: In what way, if any, do the content types and their interrelations affect the customers' choice?

1.2 Contributions of the Thesis

Some of the contributions of this study are summarized as the followings:

- Helping to uncover some hidden customer data which may help to improve our understanding of customer behaviors;
- Enabling service providers to identify areas needing improvement, such as service quality, closer relationship with customers, etc.;
- Discovering subscribers' content preferences to provide services according to their demands;
- Contributing to investment plans by providing valuable information as to which areas within the platform to invest in and which type of contents and campaigns to be deployed;
- Initiating more relevant campaigns, products and services to increase customer satisfaction and reduce cost and churn;
- Paving the way to creating more targeted marketing strategies for different IPTV customer groups; and
- Leading upcoming studies on IPTV service providers to understand subscribers and carry out more in-depth analyses.

1.3 Structure of the Thesis

The thesis is organized in 5 chapters as follows:

Chapter 2 introduces a comprehensive literature review including CRM (Customer Relationship Management), customer segmentation, the RFM Model, and clustering techniques to be applied to customer behavior.

Chapter 3 presents a combined approach for analyzing and profiling customers in video on demand (VoD) services. The customer is better characterized by clustering in different groups and identifying the relationship between Association Rule Mining and leased content types.

Chapter 4 introduces the application of the combined approach, where major customer profiles are identified and related management results are addressed for IPTV service providers to offer better service to their customers and to propose appropriate marketing strategies accordingly.

Lastly, Chapter 5 summarizes the thesis with limitations and future works.

CHAPTER 2

LITERATURE REVIEW

2.1 Customer Relationship Management

Customer Relationship Management (CRM) systems are major data marketing instruments used by firms often carrying out customer movement analyses, and they contain a mass of related information. Different enterprises create suitable marketing strategies to achieve their objectives and to add to the loyal and lasting client base [11]. In this respect, data mining techniques are employed to determine such factors as customer values to provide quick response to their needs. In general, the history of these movements are assessed for customer profiling and improving the loyalty feature [12]. CRM allows for a better grasp on clients, accessing other potential ones, and adding new loyal customers to the available base. For this reason, data mining allows for useful models and customer classification techniques to unfold any hidden data in this respect. To deal with CR issues, customer value is a number-one priority [13]. Today, CRM helps client-centered enterprises attempting to understand their customers in better detail.

A major factor for effective CRM is corporate consistency [14] as venture profits and cost savings are related with customer value. As a key business practice, CRM aims to add to the profit generated from not only existing clients, but also prospective ones by means of effective communication. Considering the clients' movements and choices, effective CRM establishes not just a client-centered system, but an analytic one to propel the enterprise forward and attract new customers [15]. Ventures with proper customer assessment can offer better services to all groups and improve CRM in the same way [16]. Updated marketing techniques are shaped for similar groups and customer values to offer well-tailored service based on the classification [17]. The literature bears some studies on combining CRM and multi-channel environments, value-creation systems and interactions related to customers' values [18].

Apart from these, one has to identify various CRM attributes to create long-term relationships with customer groups and add to profitability in the stated period [19].

Combining data mining and clustering for client classification, in this sense, sheds light on customer behavior in CRM [20] - which is the practice of determining valuable customers and offering improved service to them against the backdrop of fierce competition among mobile operators. The set of factors employed in clustering and classification optimizes CRM as well as customer analysis [21]. Later - and upon the integration of CRM with the LRFMP model - new recommendations can be made for market strategies to attract potential customers, make already–existing ones feel special, and present content according to every one's taste. In short, applying the LRFMP model to classify users within the IPTV sector can improve the assessment of customers' needs and meeting them in the long term. Upon the deployment of combined clustering techniques for data analysis, scholars have managed to further improve corporate operations based on proposed decision rules [22].

2.2 Customer Segmentation

The first step in CRM, customer segmentation and profiling, is increasingly on the forefront as firms compete for domination. A body of research is available on many data mining technologies applied to these classifications, and then assessed based on client attributes [23]. Customer profiling – otherwise, and with relevantly equal meaning, referred to herein as segmentation, classification, categorization, or grouping – entails dividing the mass of customers sharing identical features into categories to offer material and services fit for their needs [24]. Given various requirements, choices, and attitudes, no similar service can be allotted to different groups. This process constitutes the heart of any marketing initiative as it makes room for learning about individuals and their core values and choices. Profiling in this way is a measurable way of related data analyses and determining specific buying patterns [25].

Marketing with this classification strives to add to financial gains and determining those valuable consumers [26]. Here, demographics and behavioral factors decide how grouping is accomplished [27]. Others point to dividing general and product-specific factors in such classifications, with the general variables being age, gender, nationality, income, educational, etc., and the product-specific ones being the consumers' choices and decisions once purchasing – that is, frequency of purchase, expenditure, etc. [28]. In this way, one can determine more easily the users' contributions based on product-specific factors [29]. As a matter of fact, customer profiling is far more useful for firms that wish to improve product design, add to clients' satisfaction and offer more targeted campaigns. The result is added convenience, reduced time, more relevant sales-based clients' choices, and reduced unfulfilled expectations [30]. Firms analyzing clients based on such intricate profiling tend to empower their ties with them and devise effective marketing strategies accordingly [31]. In this respect, experts assess customer values based on classifying purchase and return patterns, suggesting a separate set of services for each scheme to add to profitability in SCM or supply chain management [32].

With respect to clustering, it is a machine learning method to find the links among various data and to form additional groups accordingly. Segmentation or division is another approach to divide clusters based on similarity within the dataset. Despite clustering and segmenting being identical as the purpose implies, the difference is that clustering detects new customer groups and purchase patterns [33]. In order to finalize the best number of clusters in datasets, one has to determine the count in advance – which is a subjective process and calls for numerous rounds and repetitions. There is also dependency on measuring and separating the sets in the cluster. Reference is made by experts to the techniques for identifying the ideal cluster sales in segmented and hierarchical clustering algorithms [34]. According to Table 2.1, this ideal figure is found upon visual examination and direct and statistical testing. With the direct techniques, measure optimization like the sum of squares or average silhouette in the cluster can be accomplished. In this regard, elbow and silhouette are mostly applied, whereas statistical testing helps to compare evidence and gap statistics. The literature – apart from elbow, silhouette and gap statistics - points to certain index methods chosen to identify the optimum cluster counts.

Table 2.1 Determination of the optimal number of cluster methods

Determining Number of Clusters	**Methods**
Direct Methods	Elbow Methods
	Silhouette Methods
Statistical Testing Methods	Gap Statistic

One can identify the existing customer value, potential customer value and customer loyalty in these classification efforts [35]. Popular among these variables for modelling is the RFM approach [36]. It is paramount to assess the RFM behavior using CLV (Customer Lifetime Value) for classification purposes and determining high-value clientele [37]. What is more, the analysis based on active customer behaviors among enterprises is another initiative that helps to form optimum customer groups. Scholars have also merged time-series clustering with the RFM model to further bring into the spotlight various behavioral patterns within the sequential data [38]. Still, new parameters have been added by researchers to the RFM model to make it more applicable in different fields and understand customers' preferences in a more effective manner.

2.3 RFM Models

Grouping the clients separates them optimally and in accordance to expectations – a task necessary to find out about personal attributes, previous purchases, and possible willingness in future. This helps to analyze behaviors in the RFM model and to offer the right services [39]. RFM, an approach to classify clients over extended periods, is applied to identify specific buying trends and preferences.

The literature contains certain models in accordance to data mining and customer value related to product and shaped for RFM and clustering models. The ultimate goal is to obtain ideal CRM for firms and identify loyalty to products and services [40]. RFM (Recency, Frequency, Monetary) in historical terms goes back to sales practices adopted by firms for direct marketing [41]. This approach - a useful and applicable tool to achieve better marketing via segmentation - is also employed in combination with data mining tools. In 1994, the model suggested by Hughes became the ideal customer value analysis tool, with each factor being equally important and the weights of parameters regarded in the same way [42]. On the contrary, Stone in 1995 put forth the notion that every parameter should be given a different weight with varying significance based on client attributes [43]. Another work asserts that the smallest weight value relies on the *monetary* factor [44].

RFM, as a popular tool in database marketing based on former buying trends among users, contains three dimensions and the categorized customers are rated, thus facilitating the classification of existing and potential clients [45]. This scoring tool represents behavior as figures and identifies the future trends accordingly [46]. Stone suggest that giving separate parameters to RFM is the right practice to apply various elements within various sectors and to add to the precision of segmentation process [47]. The model has been developed over time in terms of simplicity and easy applicability, turning it into a frequently preferred approach in different fields. Additionally, RFM has applications in finance, telecom, marketing and tourism for behavior-based models that forecast certain features based on certain choices and decisions [48]. Apart from these, other areas include promotions related to famous concepts appearing on social media and forecasts related to users' choices and areas of interest [49]. As a result, users' degree of vitality and interactions are examined and introduced to promotional merchandise and other ads that appeal to them.

The RFM model is popular in data mining to determine the retention of existing customers, their loyalty and needs – hence, preferred just as well for the classification of mobile and internet service users in the telecom sector [50]. Consequently, one can simply assign the clients to the right group or segment and carry out a behavior analysis, based upon which specifically tailored products and services are rendered later. The related parameters are key to separating major users by analyzing their habits [51]. *Recency* as a major parameter defines an individual's purchase habit [52]. As for *Frequency*, it establishes all purchases and visits within a given timeline necessary to assess loyalty. Lastly, *Monetary* accounts for all money spent and the average amount of shopping carried out with in a given timeline to assess the client's contribution to the establishment. RFM is beneficial as it can be merged with conventional data mining tools to measure clients based on CLV and the main factor. This is done in an effort to constantly develop better classifications that can lead to added profitability [53]. For precise correlations to be established between the most-popular items in the dataset and users' buying frequencies, the RFM approach is applied as an inventory management system on top of the "*group*" parameter [54]. In other cases, the same model was enriched with *length* and *periodicity* for data concerning customers' grocery shopping habits [10], in this way offering insight into consumption types and devising personalized services based on respective marketing strategies. Plus, certain studies have employed RFM within the financial sector to propose structured and practical ways for conducting trade based on operative marketing for grouping, defining, and understanding users' goals [55].

The addition of new parameters in the RFM model allows for customer segmentation within a variety of sectors - banking, e-commerce, retail, and others – and forming more related policies [56]. These are major research initiatives in the telecom sector, and those offering IPTV and content services can, in the same way, guarantee users' making the most of such content through secure Internet facilities [57]. Accordingly, there are other approaches formed with "*weighted*", "timely type", "amount" and "type of goods" parameters as well, with experts opting for the weighted version in their Internet analyses [58]. The recency, frequency, monetary, time, first lease and churn probability model emerged based on separating the users into various groups [59]. Other models include recency, frequency, and length in measuring IPTV user satisfaction by concentrating on users and detecting personal viewing choices. In this way, different marketing strategies are devised for those opting for certain channels, leading to added satisfaction through the right and targeted promotional initiatives [60]. Yet again, a separate factor, *potential*, targets the users' buying habits within a given span of time applying cross-selling measurement [61]. Henceforth, customer loyalty and related investment levels can be determined. Additionally, the RFM model has been utilized to assess customer value models in the field of logistics, with more focus on *profit* and *frequency* as values that can optimally forecast operations should the RFM parameters be restricted [62]. In detail, the observations in that study were that length, recency and monetary values allow for determining who the potential clients are, whereas frequency is not a major indicator and can demonstrate conflicting results [63]. With *lifetime* and *credit-scoring* parameters supplied to the same model, user classification moved into a new area to determine the optimum cluster account and provide perspective as to the specific customer values in each cluster [64]. So far, though, the most effective approach to identify specific client groups has been known to be the RFM model as it measures behavior and satisfaction with the least number of criteria. According to other studies, social media users sharing the same choices and tastes were defined, on the basis of which services were provided to them along with their attributes and special needs [22].

2.4 Clustering Techniques

Clustering entails the classification of certain items in a way that each category contains similar items with differing properties from those in other groups [65]. According to Aldenderfer & Blashfield, when characterizing all clusters, certain steps are required to choose the instance for clustering, identifying the variables of the entities, separating the cluster by creating classes based on identical entities, and finally validating the final clusters [66].

Partitional clustering algorithms call for defining the number of clusters before running the algorithm. The distance of clusters shifts recursively with every iteration until no more cluster criterion shifts [67]. Depending on the purpose, these algorithms separate the data points into k sections, each representing a cluster. The k-medoids clustering methods are the most common PAM algorithms, which ensure if representative objects are chosen arbitrarily and whether they can be substituted with non-representative objects. The optimum k-medoids in the data are chosen once the items assigned by the algorithm are replaced, up to the point of optimizing the cluster quality. This implies, obviously, that the process is more expensive compared to the k-means algorithm when it comes to large datasets [68].

Density-based clustering algorithms point to areas containing the most density of data objects - known as clusters - and are intended to provide the ideally formed of such clusters. Other well-known algorithms include DBSCAN (Density-based Spatial Clustering of Applications with Noise) and RDBC (Recursive Density Based Clustering) [69].

Hierarchical Clustering offers such levels shaped in the most appropriate way and form - rather distinctive from other algorithms. As to the agglomerative clustering algorithms, we can refer to single linkage, complete linkage, group average linkage, centroid linkage, ward's criterion [70].

As opposed to hierarchical clustering algorithms, partitional algorithms fit large datasets better, the reason being that an added number of data points reduces complexity [71]. They designate items to clusters with iterative sections and devoid of a hierarchical format, with the ultimate purpose of decreasing the square error function and removing the obstacles related to optimization. They are, more precisely, grouped as supervised and unsupervised [72]; in the former case, one has to specify the cluster count, whereas the latter requires no such specification.

Often times, clustering requires two steps, and experts assess the cluster count based on the Ward’s method applied to small datasets, followed by the k-means algorithm [73]. Owing to this method, they can create a dendrogram with the least variance method based on Ward's approach and suitable for data structures. Comparing the clustering approaches, one can see that the k-means algorithm is enough to help identify the precise cluster count for large datasets [74]. This algorithm is widely used for classifying users sharing same lifetime values, and for identifying loyalty based on RFM-assisted performance measurements [75]. As stated before, RFM values are mostly applied for grouping based on the k-means algorithm, whose application leads us to the upcoming batch output and quality depending on choosing the right number of clusters. Optimum Silhouette Index, David-Bouldin and Dunn Indexes are also employed to identify the cluster count ideal to yield meaningful sets of data objects [76].

Using the instruments as stated above, clients are categorized in a way to adjust the marketing policies accordingly and set up rules to determine the purchase tendencies in each group. These rules govern the approach to be adopted towards new customers as well sharing likewise attributes with others [77]. For this ultimate goal, different techniques include clustering, classification, SOM (self-organizing maps) and numerous other evolutionary algorithms. Concerning SOM, it is widely applied and has provided promising outcomes alongside data mining and behavioral scoring models [78]. In most cases, cluster validity measures are deployed in line with SOM), the elbow method and dendrogram techniques. Apart from this, many other efforts have been made to deal with customer profiling, CLV (Customer Lifetime Value) analysis and locating high-value users’ intentions. Using certain methods, one can identify the present client value, potential client value, and overall loyalty in customer segmentation research targeted toward the wireless telecom sector as well [35].

Certain studies also incorporated SOM with LRFM (length, recency, frequency, monetary) for dental services marketing to classify pediatric patients [79], and the conclusions point toward certain suggestions and reduced prices through advertising to add to the frequency of visits paid by clients and precisely determine their expectations. In another work, the RFM technique is assessed in terms of average transaction cancellation and average number of family members traveling with an airline over a time period [80]. Accordingly, marketing schemes were recommended to fit the customers' values at best and based on their socioeconomic standing and travelling tendencies. In recent years, data mining techniques – as referred to previously herein and in relation to RFM - have been specifically applied within the automobile, computer, security and electronics sectors. There, already-existing and future client predictions have been materialized thanks to RFM integration with other data mining techniques as decision tree, rough set theory, neural network, SOM and sequential pattern mining [80]. In a study, a new model was put forth for CLV measurements using fuzzy AHP and fuzzy clustering together. Based on the reports, these measurements, when applied to each customer category and regardless of the entire subject group, point to the inference that it is better to focus on loyal customers who have already and for long been in the system [81]. FCM (Fuzzy clustering method) is a field concerned with justifications related to 0 and 1 in a given operation. The objective function for FCM based on the K-means algorithm has been fine-tuned and optimized based on certain cluster validity indices – with Xie-Beni known to be the ideal index among the rest [82]. Conventional clustering approaches tend to make use of fuzzy clustering to categorize combined sets of client profiles that are hard to separate clearly due to overlapping effect [83]. The fuzzy c-means algorithm reduces the distance-based objective function, being a soft version of k-means itself [84], whereas fuzzy clustering defines the degree of uncertainty in data structures, thus providing firms with precise data pertaining to potential clients. In line with these developments, the online music sector has managed to group its users based on the fuzzy approach, thereby providing services based on taste, approach, and attitude of individuals. The quality of service, when analyzed in this way, assists developers in making sound decisions related to products, promotions, timing and costs, and – by extension – help online services available and marketed to a far larger potential user population [85].

The SOM model is among the unsupervised ones enabling numerous basic operations conducted simultaneously [86] by automatically identifying the dominating attributes in datasets to carry out transfers from a multi-dimensional space to a two-dimensional one, consequently making more clear and comprehensible the extensive and complicated data in the form of graphics generated upon clustering [87]. SOM properties have been used to lead the choice of appropriate marketing goals alongside methods applying the k-means [88]. In the latter case, more precisely, SOM identifies the cluster count in a more rapid and basic way instead of continuous updating of observations and re-delegating to the closest cluster [89]. Additionally, the hierarchical SOM implemented to datasets as regards multimedia market segmentation can justify such classification, given that the related outcomes are easier to comprehend as they are visualized with hierarchical SOM, as opposed to mere segmentation using ordinary SOM [90]. At this point, k-means analyses are formed based on re-iterating trial-and-error tests. The literature also asserts that two-stage clustering can be done in accordance to subscribers' content type viewing habits [91].

Attempt has been made by some to analyze arbitrarily chosen mobile phone users to examine their levels of satisfaction and loyalty, with the results pointing toward the impact of service quality and image on loyalty [92]. Others have clustered users with k-means based on data mining to prove the practicality of the method when forming suggestions for mobile phone product design and packaging to offer variable services based on such classification [93]. Other researches in the sector have addressed the notion of how to keep close ties with present clients, detect new ones and add to revenue by employing mining techniques to reach these goals. The outcomes have been promising concerning customer value with LRFM model and classification with k-means algorithm [94]. As for the present thesis, the number of IPTV subscribers on different services helps to apply these approaches more thoroughly and study is supported more comprehensively with a large amount of subscribers' data.

2.5 Association Rule Mining

Association Rule Mining is used to produce recommendations in data mining techniques with the aim to explore the relationship between two sets of products in the transaction records in the database. The approach also appears in other research efforts that identify client values based on RFM [95] as it is possible to shed light on previously-unknown variables in the dataset and forecast accordingly. In addition to the clustering methods commonly employed to analyze shopping habits, customer value rules are also determined by applying Association Rule Mining algorithms. In this way, marketing strategies can be strengthened accordingly in order to increase customer loyalty based on the observed values. To this end, the required association rules for our purpose can be applied to define other potential customer markets and, hence, offer specifically designed services integrated into the CRM system [96].

Agrawal et al. introduced association rules for discovering the relationship between the data and datasets [97]. Association Rule Mining and clustering define patterns in decision-making based on data mining techniques in different application areas. The method is often used in shopping carts, catalog designs, store shelf arrangements, customer profiling, and market strategies [98]. Apart from being commonly used in market basket analysis, the technique is also used in analyzing various subscribers' preferences such as in the music sector where there is available content with categorical data [99]. In particular – and, to provide a few examples - Amazon, Netflix and Spotify have access to customers' information and other product list reviews. Market basket analysis was proposed to determine the services used by the telecom sector due to concern for customer loss and the packages preferred by them [100].

Apriori algorithm is the most widely used association rules algorithms preferred for various management aspects. It is mostly used in customer shopping behavior research to determine the rules that will help sales firms [101]. It also comprises support, confidence and lift measures. The first two determine the quality of an association rule. In order for the support value to yield appropriate results, the data set should not contain data of different qualities. It is known as the threshold of the support data set, and those below the specified frequency value do not need to be processed [102]. The higher the support value, the stronger the correlation between the product items. Confidence, on the other hand, is a measure of reliability, which indicates the likelihood of a product being purchased in a set of objects, as well as a product purchased [102]. The greater confidence value proves that there is a more pronounced relationship between product items. The lift value controls how popular one product is and indicates the possible purchase of the product. For each dataset, while the support value is calculated for each set, the ratio of total content types to the number of sets is considered as the support value. Confidence, on the other hand, is assigned a value to create the best rules, rather than the most. By the rules determined, the occurrence of unknown relations can be determined. Apriori helps us to find all frequent items using minimum support and minimum confidence value for subsequent decision making [103]. To this end, speed is very important and, as such, the Apriori algorithm stands out as the preferred method in such association analyses.

CHAPTER 3

PROPOSED COMBINED APPROACH

In this chapter, a combined approach is proposed for the analysis of IPTV customer behaviors to identify different customer segments and gain valuable information about different customer groups. The successful implementation of the LRFMP (Length, Recency, Frequency, Monetary, and Periodicity) model, previously applied in the field of retail industry, has guided IPTV service providers to develop and enhance their content offerings to their customers [10]. The model applied to IPTV customer data suggests the proposed model is expected to lead the way for customer segmentation for marketing strategies. One of the most important things for marketing is the data accuracy, because accurate data analysis is key to understanding customer behaviors in order to apply the right marketing strategies to the right customer groups for profitable business strategies. To achieve this, a research procedure is proposed in the study to analyze IPTV customer behaviors by the following: applying the LRFMP model, clustering, creation and classification of different customer groups, proposing to include the genre parameter which determines customer preferences for each content type in offered IPTV services, and using association rules as the basis for content recommendation. The steps followed for the proposed procedure are shown in Figure 3.1.

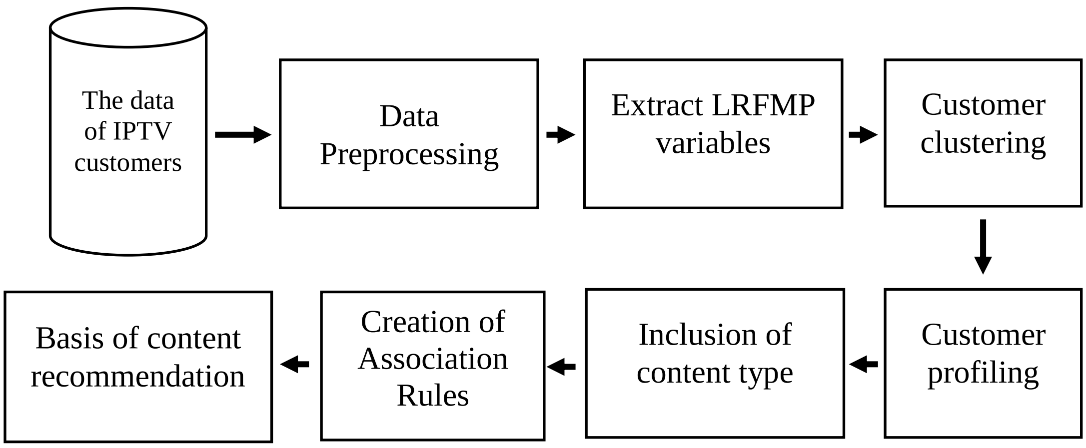

Figure 3.1 Procedure of research

3.1 Background of IPTV and VoD Service

IPTV is the transmission of television contents to customers via the Internet protocol [104]. The emergence of IPTV technology has paved the way for technological development in the telecom sector. Competing IPTV service providers need to pay attention to providing engaging content, transmission quality, and satisfied users. By evaluating user experiences with objective algorithms, researchers can provide recommendations for content offerings to customers according to their subjective characteristics [105].

The advantage of this service over traditional TV broadcasting is the more interactive and personalized customer experience. Increased service quality in IPTV service QoS (Quality of Service) is essential in determining QoE (Quality of Experience) of customers. Understanding customer preferences is the most important criteria to increase customer loyalty through personalized content service delivery. Since IPTV customers' accounts can be integrated with mobile, web, and smart platforms, it becomes more difficult to identify the user [106]. Customer loyalty and expectations have a high impact on the success of IPTV services, whose providers put lots of effort into offering valuable customer-oriented services; for example, continuously improving user interfaces and experiences, providing system reliability and content enrichment. These services help to improve end-to-end service performance, increase market shares and give opportunities to offer new value-added services to have a profitable and sustainable business. IPTV is a relatively new and still rapidly developing technology that offers many live streaming, audio video contents and interactive application services through the internet protocol. Providing all these services over the internet protocol (IP) offers customers a more reliable, secure interactive and personalized user experience compared to traditional TV broadcasting services. For instance, customers can record the current and future programs to watch later, or rent video content at any time [107]. The IPTV system has customer devices such as STB (Set-Top Box) and software applications for mobile phones, tablets, computers, Smart TVs which store a large amount of customer data representing user movements and the quality of service offered by the system.

Figure 3.2 provides a simple description of the IPTV service. IPTV live broadcast channels are created at the IPTV Headend and, then, transmitted to IP - MPLS

(Internet Protocol - Multiprotocol Label Switching) network as multicast streams and the customers get access to the channels with their STB devices. The number of such users has no burden on the headend system. The VoD service is the mostly preferable content part of IPTV; live broadcasts of the web, smart, and mobile platforms are transmitted as unicast streams. With this protocol, VoD contents are streamed directly to each customer so that the customer sends a request to the content to be rented or watched as live streaming. In this vein, CDN (Content Delivery Network) systems are groups of servers that provide these unicast streams to customers.

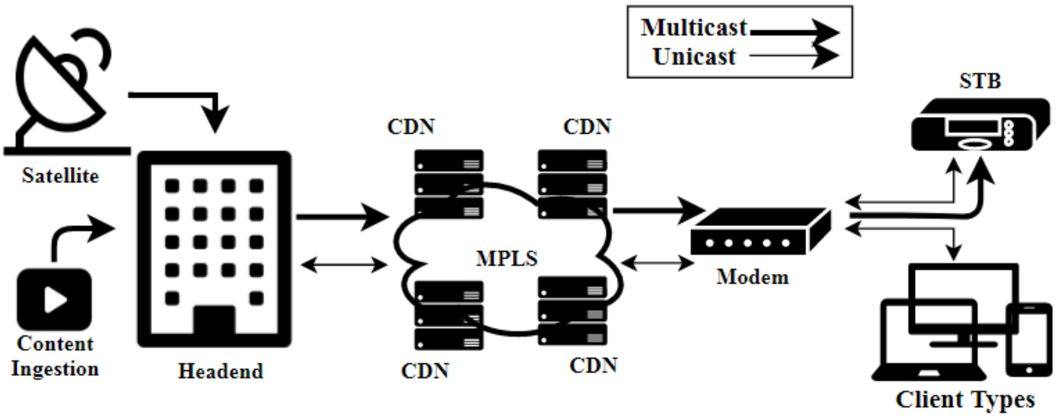

Figure 3.2 IPTV flow diagram

Researchers emphasize that in order to better classify users, an analysis of user comments and user behavior should be performed in social networks [108]. The high demand for optional content when providing IPTV service to customers can cause an unexpected burden. The bandwidth, access, and usage dynamics of the content significantly affects the performance of the system. For these reasons, it improves the quality of service as well as providing personal services by modelling and analyzing the customer's usage habits [109]. Studies also mention a real-time suggestion system to understand when, why and how they change the channel in customer behavior especially through channel transition analysis [110]. Thus, it is possible to present a new channel suggestion list to the user with different techniques as developed based on personal characteristics.

In line with this objective, there are customer behavior analysis studies in the literature, especially pertaining to mobile and internet usage habits for customer segmentation in the telecommunication industry. Customer values and behaviors are the basis of marketing strategies in this sector, which benefits from data mining to improve and increase customer services. It is very important to create the right customer group in determining the appropriate marketing strategies [111]. Some researchers mention that it is necessary to develop appropriate marketing strategies for segmented customers rather than standard market strategies in telecommunication field. The telecom companies aim to increase growth through segmentation by conducting behavior analyses and life time values of the customers to keep them in the system longer [112]. Internet-based applications such as mail, social networking, geo-location and streaming video have increased competitiveness in the mobile market. Researchers have created a user-based service procedure for the evaluation and mapping of such services [113]. In this way, visualized strategies have been developed to meet a wide range of needs and to provide flexibility over time.

CRM and RFM change trends in telecommunications services where user value depends on long-term performances with the aim to develop flexible market strategies. Researchers have analyzed their intensive usage time, content upload processes, mobile phone model information and application data used in customer data analysis on mobile phones [114]. Mobile, fixed line, cellular services and IPTV service customers' data from global telecommunications companies have been tested for this purpose. In a study, service quality concepts were modelled and the indirect effect of corporate brand image was determined in terms of customer satisfaction and repurchase intention [115].

The proliferation of IPTV service resources and the large amount of customer data has led to the need for user behavior analysis. Audience, content analysis, motivation analysis, location and time analysis are now possible. The number of the audience and their demographic characteristics constitute the dimensions of the research. In the literature, contents are categorized and classified into subcategories [116]. Such analyses are applied by researchers to TV programs to devise a user preference model and determine the user habits. The RFM model, as a result, was used to classify the data before customers were clustered. Despite these efforts, their application areas are still in the discovery stage and are open to further developments.

There is also research into the impact of the IPTV service adoption and its integration with communication and media technologies for customers. In the surveys, customers' habits of using IPTV were determined. The IPTV qualifications included customers' compliance, content diversity, monetary values, innovative approach expectations and social impact [117]. There is also another study attempting to reduce the load on IPTV content servers during peak hours for concurrent users in a broadband environment limited by simulation method [118]. The result is that high volume content demand is possible and it performs better with the model that recommends capacity maximization.

IPTV has been subject to flexibility, quality content delivery, reliable and rich personalized services. It encodes the signal it receives and only transmits it to the owner of the carrier over the network, embedded in multimedia products that include learning and interactive videos [119]. In addition to the RFM model, customer margin and marketing touch methods have been added to the customer's past transaction data. More customer-level data needs to be identified in the multichannel services offered. In multimedia applications, it is also stated that the customer data is needed for at least 2-3 years in order to obtain satisfactory results to recognize the customer [120]. In the telecom industry, grouping customers by categorizing different products and services with all variables is applied to later decide on the marketing strategies accordingly. To this end, innovative solutions in the mobile sector have increased; but there are few studies on multichannel behavior data in the literature. For this reason, a web-based multichannel customer segmentation was carried out to determine the relationship between customer revenue and customer loyalty [121]. Because customers often prefer after-sales services, call center customers data also needs to be included in the measurements.

Some content service providers have offered the opportunity to benefit from all content at no extra cost to customers. These providers focus on customer information because identifying what kind of services customers use may be a good marketing strategy by advertising additional services [122].

In an analysis conducted in some telecom companies, it has been proved that treating all customers equally is not a viable strategy and that it is necessary to provide additional services to some in order to maintain customer relationships. Campaigns offered to customers help to increase their emotional commitment and loyalty to the service provider [123].

In one of the researches in which technological growth was examined, the evolution of the telephone was discussed and product comparisons of domestic companies were made according to different demands [124]. Identifying new and important factors that help identify customers is among the strategies of some firms to provide a framework against increased competition. For this, the system automatically learns how data mining analyses a large amount of data set with new possible patterns. In the telecommunications market, cluster analysis is applied to solve customer segmentation problems, and the k-means clustering algorithm is the most convenient method in this regard. Marketing strategies are, then, recommended to identify customer characteristics and make sense of user groups [125].

Some suggested marketing strategies based on customer segmentation include: making customer-oriented changes with accurate market segmentation, positioning and sustainable development as well as data mining methods [126].

3.2 LRFMP Model

The RFM model is developed with customer data specific to the area of application. In the literature, there are studies in the areas of health, technology, finance, retailing and travel in order to identify and provide services to meet customer needs and expectations [16];[64];[53],[127],[10]. Companies need to get to know their customers better in order to keep them in the system for more time and access new users. In the telecommunication industry, together with the dramatic increase in IPTV service providers, customers who are not satisfied with the services offered can easily look for another service provider. In order to keep the existing customers in the system and predict potential ones, the IPTV service providers need to analyze customer behavior. Each parameter in the LRFMP model is a guide to getting to know the customers of IPTV service providers and to provide them customized service offers. The feature descriptions of the LRFMP model recommended for retailing sector adopted to IPTV customers are:

Length: The number of days between the first and last content rental by the customers. A customer with a high length value represents high customer loyalty.

Recency: Refers to the time in days between the date of observation and the latest rental date in the customer analysis.

Frequency: Refers to the total rentals made by the customer within the specified observation period, with each content rental calculated separately in the event of numerous content rentals in a day;

Monetary: it is the average amount of money (currency in TL) spent per content rental, which is important to measure each subscriber’s contribution to company offering IPTV services.

Periodicity: This is the standard deviation of the RTI (Rental Time Internal) value between two consecutive rental times of customers. It provides information on whether customers make regular rentals or not. If this value is low, it means that the customer rents content regularly and at certain intervals.

To apply the LRFMP model used in retail sector to IPTV subscribers, all the required calculations were performed. Table 3.1 below shows the RTI results of the 4 sample customers according to the last observation date 31.08.2018.

Table 3.1 RTI calculation for 4 sample customers based on rental date

Customer ID	Rental Date	Calculation of RTI	Customer ID	Rental Date	Calculation of RTI
Customer1	17.05.2018	0	Customer4	04.07.2018	13
Customer2	20.05.2018	0	Customer2	06.07.2018	32
Customer2	21.05.2018	1	Customer3	08.07.2018	6
Customer2	22.05.2018	1	Customer3	12.07.2018	4
Customer2	25.05.2018	3	Customer2	21.07.2018	15
Customer2	04.06.2018	10	Customer4	24.07.2018	20
Customer3	11.06.2018	0	Customer3	26.07.2018	14
Customer3	18.06.2018	7	Customer4	05.08.2018	12
Customer4	24.06.2018	0	Customer4	05.08.2018	0
Customer1	27.06.2018	41	Customer3	08.08.2018	13
Customer1	01.07.2018	4	Customer2	14.08.2018	24
Customer3	02.07.2018	14	Customer1	21.08.2018	49
Customer1	03.07.2018	2	Customer4	28.08.2018	23

In the light of these data, the total time interval between the customer's first and last rental dates, the number of days between the rental dates, rental frequencies and monetary of the rental content variables are determined using the LRFMP model. The calculations of the variables of these customers as length, recency, frequency, monetary, periodicity of each customer are shown in Table 3.2.

Table 3.2 LRFMP results for 4 customers

Customer	Length	Frequency	Recency	Monetary	Periodicity
Customer1	96	5	10	1.93	24.48
Customer2	86	8	17	2.09	12.11
Customer3	58	7	23	1.75	4.50
Customer4	68	6	3	3.96	8.91

The data related to VoD services subscribers helps to determine customer behavior by interpreting the results of those involved in the service near the last observation date determined in the analysis of the customers. According to Table 3.1, having the maximum length value of Customer1, it represents customer loyalty to service. This is the frequency value of the total visits during the observation period, and the loyalty increases as such frequency increases. The recency value for Customer1 shows the current interaction of the customer in the VoD service and helps to identify the tendency to rent content again. Due to the cheapness of the content prices, it will not be right to interpret the sample customers as contributing to the company or not. When the periodicity value of Customer1 customer is examined, it can be said that the service does not benefit from the service at regular intervals.

The customers' rental of content on different dates allows us to calculate the possibility of rental again. In addition, the last rental dates at different times give information about the active customer. Therefore, it is useful to get such information about the behavior of IPTV customers. This model can lead to the development of another approach to analyze customers on different platforms purchasing different types of content.

3.3 Customer Segmentation based on LRFMP Model Using Cluster Analysis

Combining the LRFMP model applied to IPTV customer data with different clustering algorithms allows for defining the customer segments. Clustering is the grouping of objects with similar characteristics by dividing them into pieces [128]. There are different clustering techniques depending on the data set and requirements available. The most known techniques are partitioning, hierarchical, fuzzy, density-based and grid-based clustering methods [129].

Partitioning Method: It is given to n parts of a given object partitions called clusters, and it offers the ability to organize the cluster by grouping the data into groups. The most common partitioning methods are k-means and k-medoids [130].

Hierarchical Method: It plays a role in separating data into groups at different levels hierarchically. This method is useful for grouping data objects in tree and hierarchical forms [130].

In this study, customer segmentation is performed in the data set consisting of TV audiences. The features in the model selected for each customer are calculated and the customers are segmented according to the calculated characteristics. In this way, an effective environment has been developed to create more effective marketing strategies. Moreover, k-means clustering algorithm, which is a partitioning model, was applied and the customers were parsed according to the appropriate number of clusters [131]. According to the customer requirements, this study is expected to guide the service providers to present their products more beneficially.

When the RFM model is applied in the data set, new features are needed to measure the customers' behavior and loyalty. Thus, length and periodicity values in the LRFMP model are also included in the calculation.

3.4 Integration of Genre Feature to Customer Segments through Association Rule Mining

Such Association Rule Mining-based recommendation schemes appear quite often in the literature, which points to forecasting based on experimental surveys of service providers [132]. While the number of IPTV customers is increasing rapidly, data mining methods are used to identify the types of content they rent. The association rule method can be used to obtain information that is valuable from the transaction data of the customers, to reveal the relationships between the data and to make predictions when necessary.

As stated before, the aim of this study is to find out the relationship between the types of content that customers rent. With this relationship, it proposes content suitable for the customer's taste and increases the satisfaction of the IPTV service, ensuring that it customers stay in the system for a long time. In IPTV, the information related to products rented by customers and the rented product types are determined by means of purchasing habits analysis. For this dataset, the Apriori algorithm is applied as an agreeable model in IPTV customer analysis. In the study, the subscribers' rental content types are stated and defined in the genre (G) list as:

G = {action, adventure, animation, biography, comedy, crime, documentary, drama, family, fantasy, history, horror, music, musical, mystery, romance, sci-fi, sport, thriller, TV-shows, war, western}

First of all, the most rented five content types are analyzed according to their size because it is impossible to analyze all content types. Then, the Apriori algorithm is applied to discover the relation between four popular genres. In this technique, each association rule is created of two different item sets indicated as X and Y [133]. The X character shows the left-hand side (LHS) as the antecedent item with the most rented content type, while the Y character indicates the right-hand side (RHS) as the consequent item set relevant to X [99]. Thus, the antecedent and the consequent items express the degree of uncertainty about the association rules. In the study, Association Rule Mining allows VoD subscribers to determine their possible rental content types. Support, which is an indication of how frequently the itemset appears in the dataset, simplifies the number of transactions and contains all items in the antecedent and consequent parts of the rule. In the consequent items and the support value to the number of subscribers' transactions, all items in the antecedent are included in the confidence value. Confidence, which represent the frequency in which the rule can be true, includes all items in the consequent and the support to the number of transactions, including all items in the antecedent. Finally, the relation between the antecedent and consequent items is denoted by the lift value. Lift takes into account support as well as the dataset. All the measures and formulas are given below:

$$Support\ (X \rightarrow Y) = \frac{X\ U\ Y}{G} \qquad (1)$$

$$Confidence\ (X \rightarrow Y) = \frac{Support(\ X \rightarrow Y)}{Support(X)} \qquad (2)$$

$$Lift\ (X \rightarrow Y) = \frac{Confidence\ (X \rightarrow Y)}{Support\ (X \rightarrow Y)} \qquad (3)$$

The relationship between the association rule and customer records is examined to determine the events that may occur simultaneously. However, the fact remains that the association rule method based on intensive object set calculation is expensive in large volume data sets. In this study, the analysis was conducted for the clusters in which the customer groups are mentioned and for the types of content that the customers rented, as well as the types of content that could be suggested to them.

CHAPTER 4

APPLICATION OF THE COMBINED APPROACH

This section includes the combined approach results of customer behavior analysis using VoD service as IPTV subscribers, determining customer groups based on clustering and discovering the relation of rental content types with Association Rule Mining. For customer behavior analysis, LRFMP model is applied to identify important customer segments and get to know the customers better. Hence, customer behaviors are profiled to provide them personalized service offers and better service quality.

4.1. The Case Company and Data Set Description

This study covers real customer data from customers benefiting IPTV platform offered by one of Turkey's largest telecom service providers. The study includes only VoD service subscribers' data on the STB devices, and the data set is compiled from the database tables where customer transactions are held, mainly 277.808 individual customer data between 1 September 2016 and 31 August 2018. The VoD subscribers' data is transferred in the form of a spreadsheet to Microsoft SQL Server database. Due to the importance given to data privacy, ordinary numbers were given instead of customer IDs.

4.2. Data Preparation

707 test user data were not included in the study because they do not reflect real customer movements. In addition; in order to avoid incorrect analysis while performing the "length" calculation as one of the parameters used for customer loyalty measurements, 81.517 customer data renting on the same day were excluded from the study -the reason being that, otherwise, the periodicity value would not be computable due to lack of the RTI measure. The 540 customers' transaction data with "monetary" value 0 was not included in the study either.

Moreover, the data obtained by completing the missing subscribers' information and incorrect transaction records were removed in the pre-processing stage of the analysis. Thus, the data set was cleaned up and, then, the analysis was performed with 195,493 customer data in total.

In this data set, each transaction record contains the customer ID of each user, the platform information that it belongs to, rental date, content ID, price, name, and the genre information of the rented content. Furthermore, for each content type, the total numbers and IMDB (Internet Movie Database) sub-genre information of each content as universal content information are included in the data set. Finally, although the platform information that customer belongs to is included in the data set, only the customer transaction records on the IPTV platform are considered for the study due to the fact that the customer information on other platforms is not stored properly. The LRFMP parameters were determined for each customer and the min, median, mean, standard deviation (SD) and max values of these parameters are shown in Table 4.1.

Table 4. 1 The results of LRFMP variables descriptive statistics

	Min	**Med**	**Mean**	**SD**	**Max**
Length	1.0	145.0	220.03	200.2	728.0
Recency	1.0	38.0	105.43	159.9	727.0
Frequency	2.0	7.0	14.61	21.8	733.0
Monetary	0.2	2.4	2.51	1.5	39.0
Periodicity	0.0	15.6	33.6	50.3	502.8

4.3. Grouping Customers

In this study, one of the most popular clustering techniques, the partitioning clustering method was used to divide the data into different groups. Thus, clusters were arranged recursively until the optimum division between clusters and data points was achieved.

In this section, customer segmentation is applied to the data set of IPTV customers. The LRFMP calculations were made for each customer in the model and the customers were clustered according to their characteristics.

Marcus proposes the frequency and monetary values for the customer value matrix, whereas the customer loyal matrix includes length and recency values [134]. He said that the increase in customer loyalty depends on the long relationship of the customer. This statement, which is accepted in the literature, has helped many researchers in grouping customers. In addition, researchers using different symbols in the customer segment definition interpreted them according to the mean value [135]. According to the customer's length, recency, frequency and monetary values, those that contributed to the company were interpreted as potential, lost customers and new customer groups. In practice, IPTV service providers have handled their customers in 4 different groups according to their characteristics: (1) consuming and most valuable subscribers, (2) less consuming and less valuable subscribers, (3) less consuming but loyal subscribers, (4) neither loyal nor valuable subscribers. The results of this individual analysis are detailed in Section 4.4.

4.3.1. Determination of Number of Clusters

In the analysis, the subscribers' transactional records in the data set were standardized for the LRFMP variables using min-max normalization between 0 and 1 prior to clustering. For an ideal number of clusters, the k-means algorithm is applied to the LRFMP values. In the k-means algorithm, which is one of the partitioning clustering methods, the clusters define the total change within the cluster or the minimization of the total intra-cluster validation value called WSS (within-cluster sum of square) [136].

The smallest possible value indicates the compactness of the cluster and is important in determining the optimum number of clusters. The k-means algorithm applied in the R programming language was tested for different k values. Each WSS calculation of the k value between 1 and 10 clusters was made and the WSS curve was plotted according to the appropriate number of clusters. Figure 4.1 shows the WSS value decline to k=4, indicating that there is no significant reduction in the number of clusters, and the optimal number of clusters as 4 according to this curve.

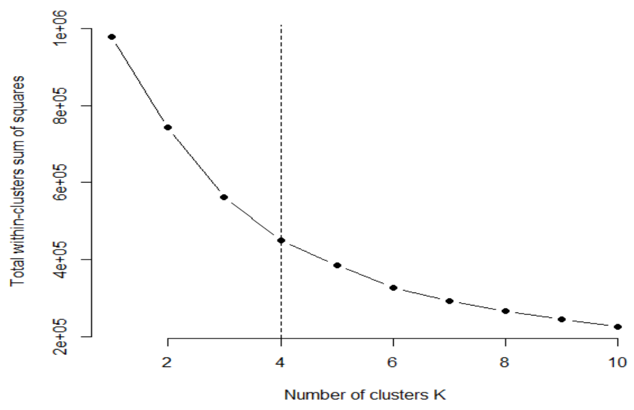

Figure 4.1 WSS results with k-means algorithm

In order to find the ideal set number, iterative k-means algorithm was used to determine the optimal number of sets between 2 and 10 k = 4. In Figure 4.2, each color shows different sets. Furthermore, the most extreme points for these 4 different clusters refer to the most loyal customers of that cluster. Customers at the intersection points of the clusters imply similar behavior to those in other clusters. The mapping of the k-means algorithm to the data set is shown in Figure 4.2.

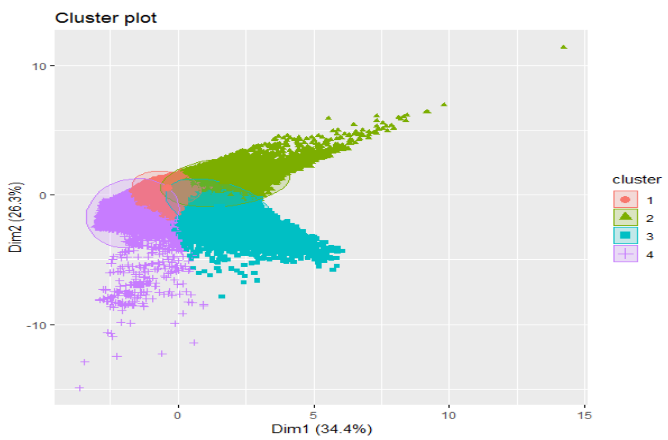

Figure 4.2 Optimal number of clusters with k-means

4.3.2. Performance Comparison of Clustering Techniques

In this study, k-medoids algorithm is used to prove the number of clusters detected in the k-means algorithm. What is important in both algorithms is to perform operations according to the specified number of clusters. The Partitioning Around Medoids algorithm, which means k-medoids, is the drives version of the k-means algorithm. The maximum number of clusters as many as the attributes can be determined via k-medoids. The goal is to check whether each collet creates a distinct set or not. Dissimilarity and the two structures are determined to be separated from each other, which helps to determine the center of the mass. Thus, the medoid vectors appear and the metric account can be used to segment the data more sharply by selecting the metric account [137].

Since it would be more difficult to create a submatrix in the original data set, for k-medoids algorithm used to prove the number of current cluster, the calculation was made in accordance with parameters having higher values than the values determined in the dataset. In the data belonging to 195.493 unique user data, the optimal values were determined based on minimum and maximum values of each parameter. For example, assuming that renting of the customers' content in a month, the size of the data is shortened in accordance with circumstances in which length is greater than 0 and frequency value is greater than 12, periodicity value is over 0, monetary value is over 0 TL or more, and the recency value as min 6. The applicability of the PAM algorithm is shown in Figure 4.3 depicting 43.448 unique users.

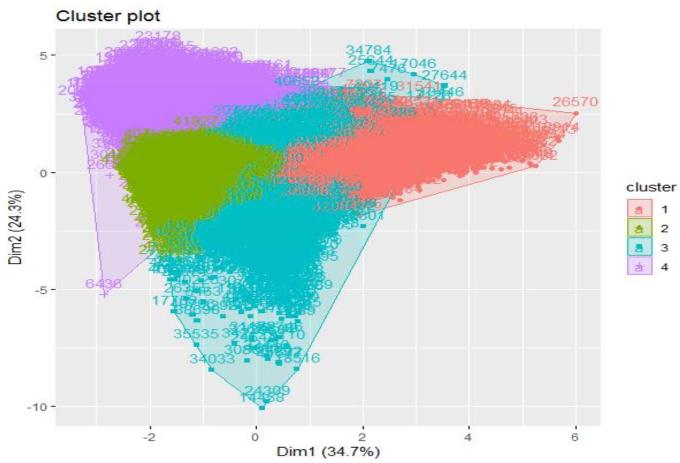

Figure 4.3 Applying k-medoids algorithm to sample data

According to the k-medoids shown in Figure 4.2, the algorithm proves to have low efficiency due to the fact that the customer behaviors close to the intersection points cannot be clearly identified and the data size is less. This algorithm is not preferred because it is slow and does not express customer behavior clearly. In addition, k-means algorithm was re-applied to the same sample data set. As shown in Figure 4.4, the cluster intersections are clearly defined and customer behavior is more easily interpreted.

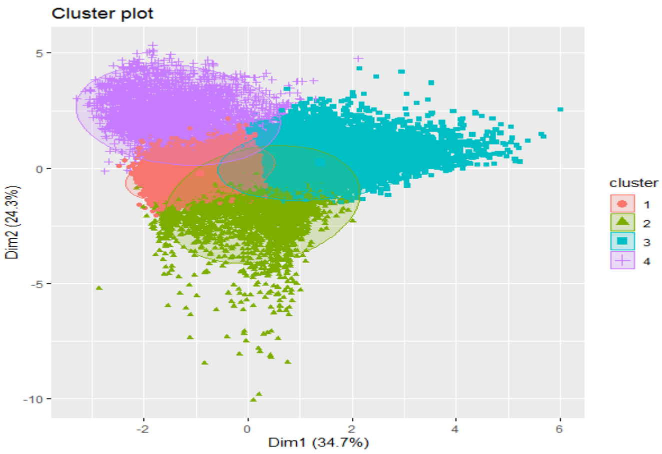

Figure 4.4 Applying k-means algorithm to sample data

Table 4.2 shows the cluster measurements of the k-medoids and k-means algorithms after applying the sample data to the set, revealing the former to have no consistency and that applying the k-medoids algorithm to the original data set is more effective.

Table 4.2 The comparison of the application of the clustering algorithms to data sets

Clusters	**Size of k-medoids clustering**	**Size of k-means results**
Cluster1	10.063	19.561
Cluster2	16.793	4.905
Cluster3	13.099	15.474
Cluster4	3.493	3.508
Total	**43.448**	

Since the k-medoids algorithm could not be adapted to the original data set, the customers were introduced to the profiling stage using the results of the k-means algorithm to determine the optimum number of clusters. Thus, the average values and details of customers' LRFMP values for each cluster are given in Table 4.3. In the literature, the up symbol (↑) is commonly used for above-average values and the down symbol (↓) is used for values below average [135].

Table 4.3 The results of applying k-means clustering algorithm

Cluster	Sample Size	Avg L	Avg R	Avg F	Avg M	Avg P	LRFMP Scores
Cluster1	18.194	440.46	23.71	66.84	2.51	15.54	L↑R↓F↑M↓P↓
Cluster2	28.761	104.64	445.59	5.96	4.03	17.06	L↓R↑F↓M↑P↓
Cluster3	39.925	482.55	63.14	11.19	2.80	102.41	L↑R↓F↓M↑P↑
Cluster4	108.613	117.17	44.59	9.40	2.17	15.71	L↓R↓F↓M↓P↓
Avg		**220**	**105**	**14**	**2.609**	**33.604**	

4.4. Customer Profiling

The evaluation of customers according to the length of time they are in the system is known as creating a customer profile [138]. Companies that pay attention to changing customer needs specifically demonstrate this change in product design.

For the first cluster, the fact that the length value expressing the behavior of the customers in the first cluster is higher than the average indicates that the customers are heavily active between the analyzed dates. Additionally, the fact that recency is less than average shows that these customers rent content at short and similar intervals. Frequency shows that the same customers frequently make leases, but the period between leasing times is not periodic. Since the monetary value was below the average amount, these customers were considered to tend to choose more affordable content. Periodicity, below average, explains that customers have long been content leasing and are, as such, valuable to the system. For the customers in this cluster; it may be advisable to offer a large number of contents at an affordable price, and to apply additional campaigns and discounts for existing and new content. Furthermore, special categories or special packages can be defined for them on special days (such as birthdays and new year. For this purpose, different notification techniques (mail, SMS) can be pre-informed about campaigns.

For cluster 2, the length value is lower than the average, indicating that customers are not heavily active between the analyzed dates. The value for recency is higher than the average, meaning that these customers are renting content at long and different intervals. Frequency indicates that customers rarely rent and that the time between leasing times is not periodic. This information means that such customers are less consuming and less valuable. The monetary value is above the average amount, which means that customers are renting regardless of the content price. Periodicity, which is below average, allows for such customers to be regarded as “valuable” because they have been regularly renting content for a long time. Since the content price is not important for the customers in this cluster, the metadata information of the content plays a critical role for their rentals. In order to increase their frequency and reduce recency of these customers, an analysis of the metadata of rented contents would be conducted. This will help to recommend more relevant contents align with their content preferences by investigating the tendency of these customers for renting their content. In addition, special categories and packages can be created to suit their tastes for certain periods to encourage them toward renting more frequently.

For cluster 3, the length value is higher than the average and the recency value is lower than the average, while the frequency value is lower than the average, which indicates that these customers are less valuable consumers. However, the monetary value is above average, which makes it possible to profit from such customers. A periodicity higher than the average indicates that customers are not very valuable. Therefore, their preferences can be identified by analyzing content metadata and increasing the number of contents that may interest them. In addition, new content availability can be notified to them in advance along with promotional options.

For cluster 4, since each parameter is below the average value, some companies disregard such customer groups and propose marketing strategies aimed at keeping the customers in the other cluster in the long term. However, as shown in Table 4.4, the cluster can be prepared for customers with the largest size of individuals in this group, and surveys can be prepared for such customers to determine the general trends as well as various campaigns to be offered. For example, if it is found that the customer cannot allocate time to the IPTV platform due to other pre-occupations or lack of time, short film proposals can be made for such customer groups. Furthermore, since they prefer cheap content, discount coupons and campaigns can be prepared for the content to change customer habits and tendencies.

Table 4.4 The results of customer groups

Customer Groups	Name of group	LRFMP Scores	Size of group (%)
1	Consuming and most valuable subscribers	L↑R↓F↑M↓P↓	9.30
2	Less consuming and less valuable subscribers	L↓R↑F↓M↑P↓	14.71
3	Less consuming but loyal subscribers	L↑R↓F↓M↑P↑	20.42
4	Neither loyal nor valuable subscribers	L↓R↓F↓M↓P↓	55.55

Such marketing strategies increase the delivery of personalized products to customers and make them feel special. By introducing new marketing strategies, customer satisfaction is increased upon providing content appropriate to the taste.

4.5. Applying Association Rule Mining based on Customer Profiles

In the data set of VoD subscribers, the genre information and subgenres of the rented contents were extracted. The leased contents of each customer's transaction record are transferred to the database with values of 0 and 1. The results of the Association Rule Mining are detailed in this section to reveal the relationship between the types of rented content. The types of content in the customer data are: action, adventure, animation, biography, comedy, crime, documentary, drama, family, fantasy, history, horror, music, musical, mystery, romance, science fiction (sci-fi), sport, thriller, TV shows, war and western. In the study, the Apriori algorithm was applied separately to the customer profiles previously determined, with the purpose to identify the tastes and preferences of the customers in different clusters and to increase the satisfaction by offering them more appropriate content. Apriori algorithm support, confidence and lift values; in determining the types of content that support value in all rentals, it is shown that the customers who rent a specific content type with confidence can be determined to increase such content sale. Moreover, the minimum support value is set as 0.001 as the ratio of the number of subscribers' rental to the total number of rentals; whereas the minimum confidence value is set as 0.9 to discover the best association rules in the large data sets, with min lift value set as 1. The prepared customer number and content type table applied for 777 different contents in the database were included in the analysis. The analysis results for the first 5 most popular content types are included from amongst 22 different genre types. In Table 4.5, the values shown as “1” for consuming and most valuable subscribers profile indicate that the values are the preferences of the customers and “0” is not preferred.

Table 4.5 Summary of consuming and most valuable subscribers’ transactions for each genre types

Genre Types	**0**	**1**	**Genre Types**	**0**	**1**
Comedy	5	18.189	Mystery	4.644	13.550
Adventure	6	18.188	Horror	4.810	13.384
Action	19	18.175	History	7.223	10.971
Drama	397	17.797	Biography	8.814	9.380
Sci-Fi	620	17.574	War	8.865	9.329
Animation	982	17.212	Musical	15.406	2.788
Fantasy	1.181	17.013	Western	15.431	2.763
Thriller	1.358	16.836	Sport	16.575	1.619
Family	1.454	16.740	Music	16.645	1.549
Crime	2510	15.684	Documentary	17.019	1.175
Romance	3.708	14.486	TV Shows	17.605	589

The association rules are limited according to the pre-determined min support and min confidence values, and association rules with values less than these parameter values are not included in the study. Thus, better association rules are created and redundant rules are identified and discarded. The remaining rules are analyzed for samples with a high number of customers and Table 4.6 shows that the most valuable subscribers prefer comedy content, while more than 96% of all rentals prefer adventure, action, drama and sci-fi. Additionally, 99% of the customers renting the mentioned genres show the confidence value that they rented the comedy type. In addition, the lift value shows that the effect of comedy type with popular content genres preferences.

Table 4.6 The relation between comedy and other genres for consuming and most valuable subscribers

No of rules	LHS	RHS	Support	Confidence	Lift	Count
[1]	{Adventure=1}	{Comedy=1}	0.9993954	0.9997251	0.9999999	18183
[2]	{Action=1}	{Comedy=1}	0.9986809	0.9997249	0.9999997	18170
[3]	{Action=1, Adventure=1}	{Comedy=1}	0.9984061	0.9997248	0.9999996	18165
[4]	{Drama=1}	{Comedy=1}	0.9779048	0.9997191	0.9999939	17792
[5]	{Adventure=1, Drama=1}	{Comedy=1}	0.9775750	0.9997190	0.9999938	17786
[6]	{Action=1, Drama=1}	{Comedy=1}	0.9768605	0.9997188	0.9999936	17773
[7]	{Sci.Fi=1}	{Comedy=1}	0.9656480	0.9997155	0.9999903	17569
[8]	{Adventure=1, Sci.Fi=1}	{Comedy=1}	0.9653182	0.9997154	0.9999902	17563
[9]	{Action=1, Sci.Fi=1}	{Comedy=1}	0.9649335	0.9997153	0.9999901	17556
[10]	{Drama=1, Sci.Fi=1}	{Comedy=1}	0.9482797	0.9997103	0.9999851	17253

Each row of the Figures 4.5-4.8 represents the relationship between the common left-hand-side and right-hand-side of the dataset. Color and size represent interest. As the number of rules increases, there is a tendency toward congestion as the visuals are combined. Therefore, applying algorithm figures represents the most preferred types of content in visualizing fewer rules. The green circles show the popular rental content types by the stated subscribers' group and the size of these circles. Also, the size of pink/red circles represents lift, with larger circles implying stronger lift.

As a result of the Association Rule Mining applied to the customer data profiled as consuming and most valuable subscribers, the most rented content types are comedy, adventure, action, drama and sci-fi. Service providers are, then, advised to offer comedy to those selecting these genres.

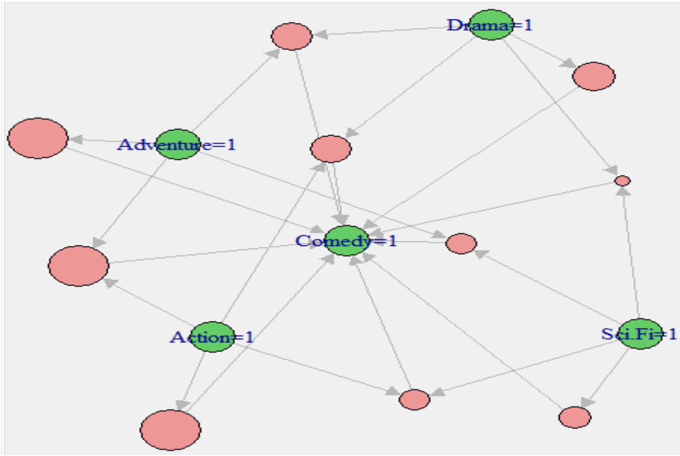

Figure 4.5 Applying Apriori algorithm for comedy consuming and most valuable subscribers

As the rules for all content types are mixed for consuming and most valuable subscribers, the relationship of those expressing only similar content types is shown in Figure 4.5.

In Table 4.7, the values shown as 1 for *less consuming and less valuable subscribers* profile indicate that the values are the preferences of the customers, and 0 means not preferred.

Table 4.7 The summary of less consuming and less valuable subscribers' transactions for each genre types

Genre Types	0	1	Genre Types	0	1
Adventure	4.776	23.985	Horror	22.654	6.107
Action	5.900	22.861	Mystery	24.424	4.337
Comedy	8.140	20.621	Biography	24.488	4.273
Drama	12.791	15.970	History	26.657	2.104
Animation	13.166	15.595	TV Shows	27.468	1.293
Sci-Fi	16.551	12.210	Sport	27.993	768
Thriller	17.138	11.623	War	28.085	676
Fantasy	17.924	10.837	Musical	28.121	640
Crime	20.211	8.550	Documentary	28.313	448
Family	20.653	8.108	Western	28.625	136
Romance	22.525	6.236	Music	28.675	86

The results for the group of *less consuming and less valuable subscribers* shown in Table 4.8 are different. Animation, action, comedy and drama are seen together in more than 26% of those who prefer adventure content. More than 95% of those who prefer animation, comedy, action, drama are excluded from the confidence choices they prefer in adventure. The increase in the rental rate of such contents also explains the lift value, which is greater than 1.

Table 4.8 The relation adventure with other genres for less consuming and less valuable subscribers

No of rules	LHS	RHS	Support	Confidence	Lift	Count
[1]	{Animation=1}	{Adventure=1}	0.5201836	0.9593459	1.150.375	14961
[2]	{Action=1, Comedy=1}	{Adventure=1}	0.5137513	0.9144696	1.096.563	14776
[3]	{Animation=1, Comedy=1}	{Adventure=1}	0.4996697	0.9599225	1.151.067	14371
[4]	{Action=1, Animation=1}	{Adventure=1}	0.4080873	0.9687990	1.161.711	11737
[5]	{Animation=1, Drama=1}	{Adventure=1}	0.2656027	0.9560701	1.146.447	7639

While interpreting the association rules for less consuming and less valuable subscribers, the customers' preferred species are explained. It is expected that the number of leasing results will help deliver the appropriate content.

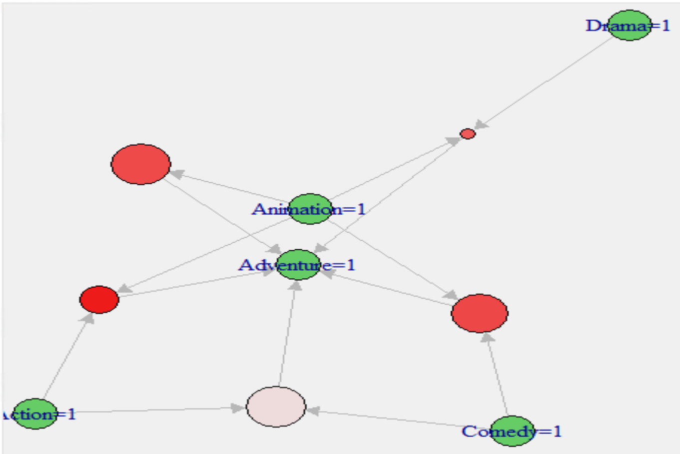

Figure 4.6 Applying Apriori algorithm for adventure in less consuming and less valuable subscribers

Since the rules for all content types are mixed for *less consuming* and *less valuable subscribers*, the relationship of those expressing only similar content types is shown in Figure 4.6. Accordingly, the size of the circles and the intense red emphasizes that the rules should be together.

In Table 4.9, the values shown as 1 for the *less consuming but loyal subscribers* profile indicate that the values are the preferences of the customers, and 0 means not preferred.

Table 4.9 The summary of less consuming but loyal subscribers' transactions for each genre types

Genre Types	**0**	**1**	**Genre Types**	**0**	**1**
Comedy	2.831	37.094	Horror	27.170	12.755
Adventure	4.245	35.680	Mystery	28.691	11.234
Action	5.181	34.744	History	29.424	10.501
Drama	10.400	29.525	War	31.267	8.658
Sci-Fi	12.867	27.058	Biography	32.031	7.894
Animation	15.454	24.471	Musical	38.462	1.463
Thriller	16.017	23.908	Sport	38.959	966
Fantasy	19.529	20.396	TV Shows	39.007	918
Crime	21.904	18.021	Western	39.170	755
Family	23.301	16.624	Documentary	39.418	507
Romance	23.862	16.063	Music	39.632	293

Table 4.10 shows that those who prefer adventure, action, drama and sci-fi have comedy rental above 63%. It is expressed with confidence that at least 92% of these content types prefer comedy. The high value of the lift shows the interestingness of the rule.

Table 4.10 The relation comedy with other genres for less consuming but loyal subscribers

No of rules	LHS	RHS	Support	Confidence	Lift	Count
[1]	{Adventure=1}	{Comedy=1}	0.8335629	0.9327354	10.039.214	33280
[2]	{Action=1}	{Comedy=1}	0.8047339	0.9247352	0.9953106	32129
[3]	{Action=1, Adventure=1}	{Comedy=1}	0.7501816	0.9277351	0.9985395	29951
[4]	{Drama=1}	{Comedy=1}	0.6852599	0.9266384	0.9973591	27359
[5]	{Sci.Fi=1}	{Comedy=1}	0.6326362	0.9334762	10.047.188	25258
[6]	{Adventure=1, Drama=1}	{Comedy=1}	0.6193613	0.9319364	10.030.614	24728
[7]	{Action=1, Drama=1}	{Comedy=1}	0.6082154	0.9253134	0.9959330	24283
[8]	{Action=1, Sci.Fi=1}	{Comedy=1}	0.5912837	0.9304351	10.014.456	23607
[9]	{Adventure=1, Sci.Fi=1}	{Comedy=1}	0.5873513	0.9348961	10.062.471	23450
[10]	{Drama=1, Sci.Fi=1}	{Comedy=1}	0.4945773	0.9382305	10.098.359	19746

The association rules that reveal such content to customers in the group explain the possibility of an increase in content leasing. Figure 4.7 illustrates the relationship between the most popular preferred content and the comedy genre.

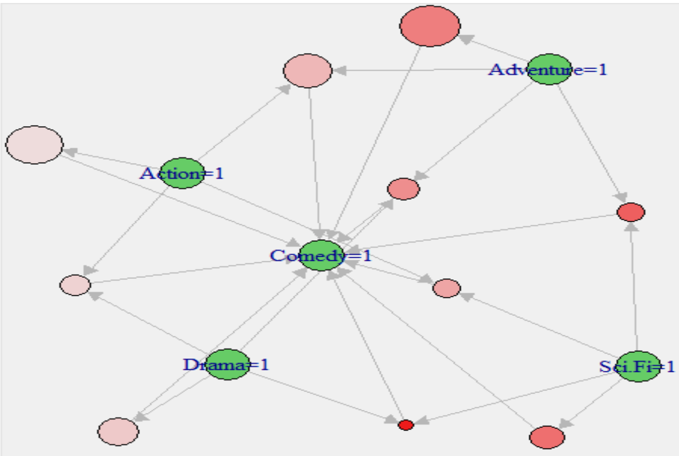

Figure 4.7 Applying Apriori algorithm for comedy in consuming and most valuable subscribers

In Table 4.11, the values shown as 1 for *neither loyal nor valuable subscribers* profile indicate that the values are the preferences of the customers, with 0 meaning not preferred.

Table 4.11 The summary of neither loyal nor valuable subscribers' transactions for each genre types

Genre Types	**0**	**1**	**Genre Types**	**0**	**1**
Comedy	9.221	99.392	History	79.343	29.270
Action	36.127	72.486	Mystery	80.369	28.244
Adventure	36.287	72.326	Horror	81.103	27.510
Drama	40.620	67.993	War	81.418	27.195
Sci-Fi	40.788	67.825	Biography	94.741	13.872
Thriller	58.464	50.149	Western	105.593	3.020
Animation	64.926	43.687	Musical	106.927	1.686
Romance	66.903	41.710	TV Shows	107.072	1.541
Fantasy	71.944	36.669	Music	107.243	1.370
Crime	73.723	34.890	Sport	107.738	875
Family	75.797	32.816	Documentary	107.939	674

When the results in Table 4.12 are considered, *neither loyal nor valuable subscribers* are shown to prefer comedy, while more than half of the ratio of all rentals prefer adventure, sci-fi, drama and action. The majority of those who prefer these genres also rent comedy species. Therefore, increasing the content types offered will likely increase the number of leases.

Table 4.12 The relation comedy with other genres for neither loyal nor valuable subscribers

No of rules	LHS	RHS	Support	Confidence	Lift	Count
[1]	{Adventure=1}	{Comedy=1}	0.6087577	0.9141802	0.9989925	66119
[2]	{Sci.Fi=1}	{Comedy=1}	0.5809710	0.9303502	10.166.625	63101
[3]	{Drama=1}	{Comedy=1}	0.5727859	0.9149765	0.9998626	62212
[4]	{Action=1, Adventure=1}	{Comedy=1}	0.5290987	0.9074788	0.9916693	57467
[5]	{Action=1, Sci.Fi=1}	{Comedy=1}	0.4566765	0.9146919	0.9995516	49601
[6]	{Adventure=1, Sci.Fi=1}	{Comedy=1}	0.4411719	0.9267382	10.127.155	47917
[7]	{Adventure=1, Drama=1}	{Comedy=1}	0.4227947	0.9275096	10.135.584	45921
[8]	{Action=1, Drama=1}	{Comedy=1}	0.4182464	0.9170317	10.021.085	45427
[9]	{Drama=1, Sci.Fi=1}	{Comedy=1}	0.3956248	0.9452474	10.329.418	42970

In Figure 4.8, it is possible that even these types of relationships may not be very strong, but that customer satisfaction may still increase if these types of content are offered more as per the results of the association rules.

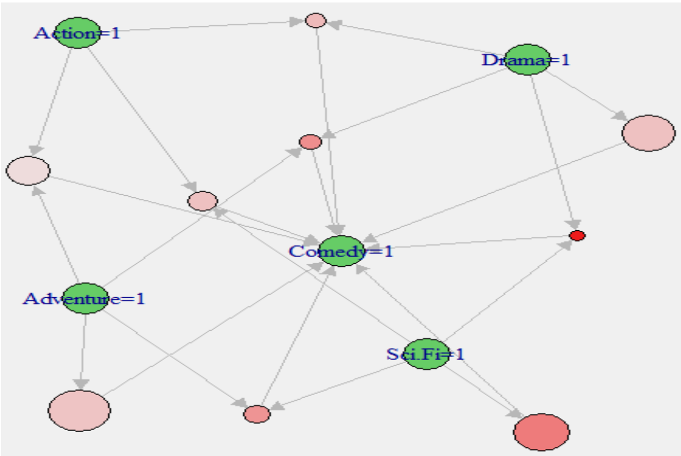

Figure 4.8 Applying Apriori algorithm for comedy in consuming neither loyal nor valuable subscribers

4.6. Managerial Implications

The proposed combined approach helps IPTV service providers to determine customer profiles and develop proper marketing strategies for their subscribers. Since proposing a profile-based marketing strategy to IPTV customers will appeal to the right customer base, individual marketing strategies are recommended for each group.

Accordingly, *consumers and most valuable subscribers* are more likely to prefer comedy, adventure, action, drama and sci-fi. With the help of the values previously used in the grouping, it was determined that the customers preferred the reasonably priced content. For this reason, offering a large number of contents that is reasonable and close to their preferences for those in this group, as well as offering them campaigns about the type of content they like, will increase their commitment to and satisfaction with the IPTV service. In addition, if the content that is offered on special occasions is discounted and media tools are used, further profit can be obtained in this way.

It was determined that *less consuming and less valuable subscribers* do not pay much attention to the content price and, as such, it is important to provide content that is suitable for their taste in order to keep them satisfied with the service. For such customers, categories should be created according to their previous preferences, and the number of contents in their favorite adventure, action, comedy, drama and animation genres should be increased. Furthermore, even providing detailed suggestions to the customers even from the detailed information in the content metadata will make them more satisfied with the service.

Less consuming but loyal subscribers are profitable profiles, and their favorite content types are comedy, adventure, action, drama and sci-fi. This customer group represented 55% of the total customers and a dedicated marketing strategy needs to be followed as they will have a big impact on company’s growth and revenue. If a successful marketing campaign is conducted for this group, it is expected to immediately increase the revenue of the service provider. To do this, massive advertisement campaigns that promote offered contents and services can be done, some surveys via mails, call center representatives, 3rd party companies can be performed to understand reasons behind less consuming behavior. Then, depending on the given customer insights and feedbacks, more attractive content offers, pricing models and promotions may be provided. For such customers, content metadata can be also enriched and new content information added to encourage further rentals.

Neither loyal nor valuable subscribers, comprising almost half of the data set, do not necessarily mean definite gains. This customer group would be the most challenging group to create more revenue from. They may have some technical or quality issues to access content services, or they may not be aware of the offered contents. Perhaps, they have some other concerns such as prices, or not enough attractive content, or simply unwillingness to spend much money or time to watch the offered content. A detailed analysis should be done in this respect; listening to and understanding these customers concerns and issues will give very valuable insights to improve service quality and help to create more awareness of offered services. Based on such insights, closer relationships and interactions should be established, relevant and cheap content packages can be promoted as ads, mail promotions etc. This is important because it will help to reduce churn, which is easier and less costly compared to attracting customers from other competitors.

As a result of this analysis, it has to be stated that the combined approach model proposed here will obviously have to lead to further studies. In the present work, customer profiling and content categorization are combined to better understand the relationship between customers' behaviors, tendency and consumed content categories, allowing more appropriate content recommendations to increase revenue and reduce churn of IPTV service providers. The study can be considered as basis for more advanced recommendation engines through further studies. For the purposes stated in the study, the research questions are laid out and answered below:

RQ1: How can the VoD subscribers' behaviors and preferences be best identified?

Customer behavior analysis was performed with the LRFMP model previously applied in the retail sector in the literature to determine the use of VOD services by the customers.

RQ2: In what way can the customer profiles be identified to better understand their behavior?

After completing the customer behavior analysis, frequently preferred methods in the customer grouping were investigated in the literature, and the final result was obtained by the appropriate k-means algorithm.

RQ3: In what way, if any, do the content types and their interrelations affect the customers' choice?

The size of each of the 22 different content types in the dataset was determined according to the customer groups. In this way, Association Rule Mining was applied to the 5 most popular types of content to reveal the relationship among them and to assert the customers’ choice of the most popular content.

CHAPTER 5

CONCLUSION

With the rapid changes and development of technology in the IPTV, customer diversity increases as well, turning into a challenge of how to satisfy customers with wide-ranging preferences and expectations. In this sector, many firms have attempted to provide suitable services to their customers based on insights into subscribers' behaviors. In the literature, it has been established that customer segmentation and categorization in this sector is primarily conducted for mobile and Internet usage behaviors and habits. On the other hand, due to the enormous diversity of content service providers, such studies for IPTV service agencies are, as such, found to be limited and requiring additional research. Implementing customer segmentation to these service providers allows them to gain a better grasp of their customers' behaviors and expectations.

With this perspective in mind, an analysis was conducted in the present work to identify the customers with different characteristics and to create more clear-cut and distinguished groups. The ultimate goal is for IPTV services to keep their customers subscribed and within the system for longer periods by measuring customer values and identifying valuable customers. Nowadays, IPTV service providers that offer vast amounts of content to their customers require additional surveys to familiarize themselves further with their clientele in a tough and competitive environment.

In the literature, the RFM model is singled out as the most preferred method offering effective advice in customer segmentation and, hence, it is preferred due to its applicability and simplicity. Parallel to this, clustering analyses performed by researchers by adding appropriate parameters to different application areas based on RFM model are on the rise as well. Although the RFM model provides satisfactory results concerning customer categorization, to gain more meaningful insights about the customers, improvements are needed depending on the area in which the model is applied. For IPTV service providers, the body of research is meagre and limited to grouping based on transaction records. Therefore, in the light of the LRFMP model developed within the present study for the retail sector, we attempted to examine the feasibility of segmenting customers for VoD services. In detail, this study was conducted to determine customer behaviors and values according to LRFMP results by using VoD customer data related to a major IPTV service provider in Turkey.

To attain the goals, clustering algorithms, frequent in use throughout the literature, were searched and customer segmentation performed based on partitioning clustering algorithms as the most suitable tool for the available dataset. To prove the effectiveness of the combined approach, the k-means algorithm was applied, and the results were found to be satisfactory for customer profiling. Thus, according to the commonly used WSS result to determine the optimal number of clusters, 4 was obtained as the optimal number of clusters.

The customer value matrix and customer loyal matrix models suggested in this study are shown to be effective and functional in grouping the customers, who are profiled as subscribers that are: *consuming and most valuable; less consuming and less valuable; less consuming but loyal; and finally, neither loyal nor valuable*. In addition, the content metadata included in the data was used to determine customers' tastes and preferences. Apriori, as an Association Rule Mining algorithm with successful results in the analysis of categorical data, was then applied to the VoD customer data to reveal any possible relationships and to analyze the coexistence of the genres in case. The rules generated by Association Rule Mining are applied to analyze customers' rental habits and preferences. In this way, IPTV service providers are expected to provide customized content to their subscribers and to increase their satisfaction, boosting profitability in this course.

With the help of this information, certain practices are encouraged, mainly: providing a large number of alternatives to customers, offering more appropriate content services and developing effective marketing strategies. In this study, individual marketing strategies have been found suitable for each customer group, since it would be impossible to provide special content to everyone within the user base. For the 22 different content types in the dataset, the results of the Apriori algorithm were obtained by taking into account the most popular of the 5 most preferred types in each group. This study is believed to contribute to the literature proposing a combined clustering to the VoD service to determine different customer groups and Association Rule Mining to the subscribers' content preferences. In this sense, the model can be considered as a guide to IPTV service providers for content recommendation on a broad scale and encompassing large number of subscribers. Moreover, equipping service providers with CRM and business intelligent systems can improve their insights about customers’ behaviors, ultimately to devise more efficient and targeted marketing strategies. This will improve service quality and customer loyalty.

Obviously, the findings and contributions in this study are not without certain limitations, among them: the dataset containing only VoD service subscribers' data on IPTV for a definite period; subscribers being STB users, thus narrowing the data rather further; and finally, large dataset rendering impossible the application of hierarchical clustering algorithms. In the latter case, and to define the number of clusters, k-medoids was performed by applying some predefined thresholds and filters to process smaller samples. In this manner, one direction in the future studies is to analyze VoD subscribers' data from different devices like mobile phones, tablets, computers, and smart TVs. It is obvious that there is more data stored in IPTV databases for each customer receiving content services through different applications and devices other than STBs. Such information related to more detailed viewing attributes and properties would lead to deeper analyses of customers’ behaviors. What is more, the outcomes of studies similar to the present thesis work in future would offer more visibility and better accuracy within the dramatically changing markets of today, and the strategies and services they are expected of as providers.

At the sample level, future studies can be enriched by other demographic features such as age, gender, profession, income level and region. Another important point is that the subscribers’ data were provided by only one of the IPTV service provider company in Turkey. In this vein, the study can further enhanced by more than just one VoD or other IPTV service providers in different geographical areas so as to ensure a more comprehensive subscriber behavior analysis. At a more technical level, the present work can be a guide for applying alternative clustering and Association Rule Mining algorithms to compare performance.

At the practical application level, future endeavors may be focused on the notion presented here to advance our knowledge of the impact and relationship of genres and other related topics, with the final aim to make more relevant content investments and predict likely preferences to be made by users in the long run. Lastly, in this respect – and to add to possible future directions - efforts can be made to better understand other competitors’ marketing strategies and how to create more attractive, targeted and interactive services – gamification, to provide an example - that promote VoD services and motivate customers to use IPTV, all leading to improved loyalty and boosted revenues.

REFERENCES

[1] T. Arul, and A. Shoufan, “Subscription-free Pay-TV over IPTV”. *Journal of Systems Architecture*, Vol. 64, pp. 37-49, 2016.

[2] N. M. Dawi, A. Jusoh, K. M. Nor, and M. I. Qureshi, “Service quality dimensions in pay TV industry: a preliminary study”. *International Review of Management and Marketing*, vol. *6* no. 4S, pp. 239-249, 2016.

[3] H. Jeon, W. Seo, E.L. Park and S. Choi, “Hybrid Machine Learning Approach to Popularity Prediction of Newly Released Contents for Online Video Streaming Service,” https://arxiv.org/pdf/1901.09613.pdf, 28 Jan 2019 [17 May 2020]

[4] C. Brantner, “CUT THE CORD: Here's a guide to every live streaming service on the market or coming soon,” https://www.businessinsider.com/live-streaming-services-to-replace-cable-on-the-market-or-coming-soon-2017-1, 21 Jan 2017 [17 May 2020]

[5] Research and Markets, “Video on Demand Market by Solution, Monetization Model, Industry Vertical, and Region - Global Forecast to 2024,” 2020, Available: https://www.researchandmarkets.com/reports/4912113/video-on-demand-VoD-market-by-solution-pay-tv, January 2020 [15 February 2020].

[6] A. Watson, “Number of pay TV subscribers worldwide from 2010 to 2024,” 2019. Available: https://www.statista.com/statistics/825329/pay-tv-subscribers-worldwide/, 4 December 2019 [20 February 2020].

[7] İ. Kabasakal, “Customer Segmentation Based on Recency Frequency Monetary Model: A Case Study in E-Retailing”. *Bilişim Teknolojileri Dergisi*, vol. 13, no. 1, pp. 47-56, Jan. 2020.

[8] W. Y. Chiang, “Establishing high value markets for data-driven customer relationship management systems”. *Kybernetes*, vol. 48, no. 3, pp. 650–662, Mar. 2019.

[9] P.S. Fader, B.G.S. Hardie, and K.L. Lee, "RFM and CLV: Using Iso-Value Curves for Customer Base Analysis". *Journal of Marketing Research*, vol. 42, no. 4, pp. 415–430, Nov. 2005

[10] S. Peker, A. Kocyigit, and P.E. Eren, "LRFMP model for customer segmentation in the grocery retail industry: a case study". *Marketing Intelligence and Planning*, vol. *35, no.* 4, pp. 544–559, 2017.

[11] M.J.A. Berry and G.S. Linoff, Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, 2nd ed., New York, NY: John Wiley and Sons, 2004.

[12] W.-Y. Chiang, "Identifying high-value airlines customers for strategies of online marketing systems," *Kybernetes*, vol. 47, no. 3, pp. 525–538, May 2018.

[13] A. Payne and P. Frow, "Relationship marketing: looking backwards towards the future," *Journal of Services Marketing*, vol. 31, no. 1, pp. 11–15, 2017.

[14] W.-Y. Chiang, "Establishing high value markets for data-driven customer relationship management systems," *Kybernetes*, vol. 48, no. 3, pp. 650–662, Apr. 2019.

[15] N. Azad and F. Ahmadi, "The customer relationship management process: its measurement and impact on performance," *Uncertain Supply Chain Management*, vol. 3, no. 1, pp. 43–50, 2015.

[16] Y.-H. Liang, "Integration of data mining technologies to analyze customer value for the automotive maintenance industry," *Expert Systems with Applications*, vol. 37, no. 12, pp. 7489–7496, 2010.

[17] S. C. Jang, A. M. Morrison, and J. T. O'Leary, "Benefit segmentation of Japanese pleasure travelers to the USA and Canada: selecting target markets based on the profitability and risk of individual market segments," *Tourism Management*, vol. 23, no. 4, pp. 367–378, 2002.

[18] A. Payne and P. Frow, "The role of multichannel integration in customer relationship management," *Industrial Marketing Management*, vol. 33, no. 6, pp. 527–538, 2004.

[19] C.-Y. Chiu, Y.-F. Chen, I.-T. Kuo, and H. C. Ku, "An intelligent market segmentation system using k-means and particle swarm optimization," *Expert Systems with Applications*, vol. 36, no. 3, pp. 4558–4565, 2009.

[20] P. Dhandayudam and I. Krishnamurthi, "Customer Behavior Analysis Using Rough Set Approach," *Journal of theoretical and applied electronic commerce research*, vol. 8, no. 2, pp. 5–6, 2013.

[21] S. M. S. Hosseini, A. Maleki, and M. R. Gholamian, "Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty," *Expert Systems with Applications*, vol. 37, no. 7, pp. 5259–5264, 2010.

[22] Y.-S. Chen, C.-H. Cheng, C.-J. Lai, C.-Y. Hsu, and H.-J. Syu, "Identifying patients in target customer segments using a two-stage clustering-classification approach: A hospital-based assessment," *Computers in Biology and Medicine*, vol. 42, no. 2, pp. 213–221, 2012.

[23] M. Namvar, S. Khakabimamaghani, and M. R. Gholamian, "An approach to optimised customer segmentation and profiling using RFM, LTV, and demographic features," *International Journal of Electronic Customer Relationship Management*, vol. 5, no. 3/4, p. 220, 2011.

[24] M. McDonald, and I. Dunbar, *Market segmentation: How to do it, how to profit from it*, Oxford: Butterworth-Heinemann, 2004.

[25] J. Wu and Z. Lin, "Research on customer segmentation model by clustering," *Proceedings of the 7th international conference on Electronic commerce - ICEC 05*, 2005.

[26] C.-F. Tsai, Y.-H. Hu, and Y.-H. Lu, "Customer segmentation issues and strategies for an automobile dealership with two clustering techniques," *Expert Systems*, vol. 32, no. 1, pp. 65–76, 2013.

[27] Y. Zhang, J. (R. Jiao, and Y. Ma, "Market segmentation for product family positioning based on fuzzy clustering," *Journal of Engineering Design*, vol. 18, no. 3, pp. 227–241, 2007.

[28] M. Wedel, and W.A. Kamakura, *Market segmentation: Conceptual and methodological foundations*. Springer Science & Business Media. 2012

[29] S. H. Ha, "Applying knowledge engineering techniques to customer analysis in the service industry," *Advanced Engineering Informatics,* vol. 21, no. 3, pp. 293–301, 2007.

[30] J. Steenkamp and F. Ter Hofstede, "International market segmentation: issues and perspectives," *International Journal of Research in Marketing,* vol. 19, no. 3, pp. 185-213, 2002.

[31] R.-S. Wu and P.-H. Chou, "Customer segmentation of multiple category data in e-commerce using a soft-clustering approach," *Electronic Commerce Research and Applications,* vol. 10, no. 3, pp. 331–341, 2011.

[32] K. Hjort, B. Lantz, D. Ericsson, and J. Gattorna, "Customer segmentation based on buying and returning behaviour," *International Journal of Physical Distribution & Logistics Management*, vol. 43, no. 10, pp. 852–865, Apr. 2013.

[33] D. Friedman, "Segmentation vs. Clustering," https://dfrieds.com/machine-learning/segmentation-vs-clustering, 9 Jun 2018 [17 May 2020]

[34] A. Kassambara, "Determining the Optimal Number of Clusters: 3 Must Know Methods," https://www.datanovia.com/en/lessons/determining-the-optimal-number-of-clusters-3-must-know-methods/, 2018 [17 May 2018]

[35] H. Hwang, T. Jung, and E. Suh, "An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry," *Expert Systems with Applications*, vol. 26, no. 2, pp. 181–188, 2004.

[36] H. Güçdemir and H. Selim, "Integrating multi-criteria decision making and clustering for business customer segmentation," *Industrial Management & Data Systems,* vol. 115, no. 6, pp. 1022–1040, 2015.

[37] C. Chan, "Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer," *Expert Systems with Applications*, vol. 34, no. 4, pp. 2754–2762, 2008.

[38] H. Abbasimehr and M. Shabani, “A new methodology for customer behavior analysis using time series clustering,” *Kybernetes*, vol. ahead-of-print, no. ahead-of-print, 2019.

[39] L.-C. Cheng and L.-M. Sun, “Exploring consumer adoption of new services by analyzing the behavior of 3G subscribers: An empirical case study,” *Electronic Commerce Research and Applications*, vol. 11, no. 2, pp. 89–100, 2012.

[40] S. M. S. Hosseini, A. Maleki, and M. R. Gholamian, “Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty,” *Expert Systems with Applications*, vol. 37, no. 7, pp. 5259–5264, 2010.

[41] R. C. Blattberg, B.-D. Kim, and S. A. Neslin, *Database Marketing: Analyzing and Managing Customers*. New York, NY: Springer, 2009.

[42] HugheS, A. M. *Strategic database marketing,* Chicago. IL: Probus Publishing Company, 1994.

[43] B. Stone, *Successful direct marketing methods*. Lincolnwood, IL, USA: NTC Business Books, 1994.

[44] A. Z. Ravasan and T. Mansouri, “A Fuzzy ANP Based Weighted RFM Model for Customer Segmentation in Auto Insurance Sector,” *International Journal of Information Systems in the Service Sector*, vol. 7, no. 2, pp. 71–86, 2015.

[45] A. J. Christy, A. Umamakeswari, L. Priyatharsini, and A. Neyaa, “RFM ranking – An effective approach to customer segmentation,” *Journal of King Saud University - Computer and Information Sciences*, 2018.

[46] J. R. Miglautsch, “Thoughts on RFM scoring,” *Journal of Database Marketing & Customer Strategy Management*, vol. 8, no. 1, pp. 67–72, 2000.

[47] B. Stone and R. Jacobs, *Successful direct marketing methods*. New York: McGraw-Hill, 2008.

[48] J. T. Wei, S.-Y. Lin, Y.-Z. Yang, and H.-H. Wu, “The application of data mining and RFM model in market segmentation of a veterinary

hospital," *Journal of Statistics and Management Systems*, vol. 22, no. 6, pp. 1049–1065, 2019.

[49] X. Chen, R. Geng, and S. Cai, "Predicting microblog users' lifetime activities – A user-based analysis," *Electronic Commerce Research and Applications,* vol. 14, no. 3, pp. 150–168, 2015.

[50] G.M. Weiss, "Data Mining in Telecommunications," in *Data Mining and Knowledge Discovery Handbook*, O. Maimon, L. Rokach, Ed. Boston: MA, Springer, 2005, pp. 1189-1201.

[51] R. Carrasco, M. Blasco, J. García-Madariaga and E. Herrera-Viedma, "A Fuzzy Linguistic RFM Model Applied to Campaign Management," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 4, p. 21, 2019.

[52] J.T. Wei, S.Y. Lin and H.H. Wu, "A review of the application of RFM model," *African Journal of Business Management*, vol. 4, no. 19, pp. 4199-4206, 2010.

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

[54] H.-C. Chang and H.-P. Tsai, "Group RFM analysis as a novel framework to discover better customer consumption behavior," *Expert Systems with Applications*, vol. 38, no. 12, pp. 14499–14513, 2011.

[55] A. Sheikh, T. Ghanbarpour, and D. Gholamiangonabadi, "A Preliminary Study of Fintech Industry: A Two-Stage Clustering Analysis for Customer Segmentation in the B2B Setting," *Journal of Business-to-Business Marketing*, vol. 26, no. 2, pp. 197–207, Mar. 2019.

[56] Y.-H. Hu and T.-W. Yeh, "Discovering valuable frequent patterns based on RFM analysis without customer identification information," *Knowledge-Based Systems*, vol. 61, pp. 76–88, 2014.

[57] D.-R. Liu and C.-H. Liou, “Mobile commerce product recommendations based on hybrid multiple channels,” *Electronic Commerce Research and Applications*, vol. 10, no. 1, pp. 94–104, 2011.

[58] E. Gharavi and M. J. Tarokh, “Predicting customers future demand using data mining analysis: A case study of wireless communication customer,” *The 5th Conference on Information and Knowledge Technology*, 2013.

[59] I.-C. Yeh, K.-J. Yang, and T.-M. Ting, “Knowledge discovery on RFM model using Bernoulli sequence,” *Expert Systems with Applications*, vol. 36, no. 3, pp. 5866–5871, 2009.

[60] Y. Chen, Z. Liu, Y. Wang, and Y. Qin, “Research on User Segmentation based on RFL Model and K-means Clustering Algorithm,” *Proceedings of the International Conference on Logistics, Engineering, Management and Computer Science*, 2015.

[61] A. Parvaneh, M. Tarokh, H. Abbasimehr, “Combining data mining and group decision making in retailer segmentation based on LRFMP variables,” *IJIEPR. 2014*, vol. 25, no. 3, pp. 197-206, 2014.

[62] C.-C. Chern, A. J. T. Lee, and C.-P. Wei, “Introduction to the special issue on ‘Data analytics for marketing intelligence,’” *Information Systems and e-Business Management*, vol. 13, no. 3, pp. 399–402, 2014.

[63] K. Chen, Y.-H. Hu, and Y.-C. Hsieh, “Predicting customer churn from valuable B2B customers in the logistics industry: a case study,” *Information Systems and e-Business Management*, vol. 13, no. 3, pp. 475–494, Feb. 2014.

[64] M. Alborzi and M. Khanbabaei, “Using data mining and neural networks techniques to propose a new hybrid customer behaviour analysis and credit scoring model in banking services based on a developed RFM analysis method,” *International Journal of Business Information Systems*, vol. 23, no. 1, p. 1, 2016.

[65] G. Lefait and T. Kechadi, “Customer Segmentation Architecture Based on Clustering Techniques,” in *2010 Fourth International Conference on Digital Society*, 2010 pp. 243-248

[66] C.-Y. Chiu and I.-T. Kuo, “Applying particle swarm optimization and honey bee mating optimization in developing an intelligent market segmentation system,” *Journal of Systems Science and Systems Engineering*, vol. 19, no. 2, pp. 182–191, 2010.

[67] S. Elavarasi, J. Akilandeswari, and B. Sathiyabhama, “A survey on partition clustering algorithms,” *International Journal of Enterprise Computing and Business Systems*, vol. 1, issue. 1, pp. 1–14, 2011.

[68] K. Chitra, and D. Maheswari, “A Comparative Study of Various Clustering Algorithms in Data Mining,” *International Journal of Computer Science and Mobile Computing*, vol. 6, issue. 8, pp. 109–115, 2017.

[69] S.K. Popat, E. Mark, “Review and Comparative Study of Clustering Techniques,” *International Journal of Computer Science and Information Technologies*, vol. 5, issue. 1, pp. 805-812, 2014

[70] S. Patel, S. Sihmar and A. Jatain, “A study of hierarchical clustering algorithms,” in *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, 2015, pp. 537-541.

[71] H. Khobzi, E. Akhondzadeh-Noughabi, and B. Minaei-Bidgoli, “A New Application of RFM Clustering for Guild Segmentation to Mine the Pattern of Using Banks’ e-Payment Services,” *Journal of Global Marketing*, vol. 27, no. 3, pp. 178–190, 2014.

[72] R. J. Kuo, N. J. Chiang and Z.-Y. Chen, “Integration of Artificial Immune System and K-Means Algorithm for Customer Clustering,” *Applied Artificial Intelligence*, vol. 28, no. 6, pp. 577-596, 2014.

[73] D.-C. Li, W.-L. Dai, and W.-T. Tseng, “A two-stage clustering method to analyze customer characteristics to build discriminative customer management: A case of textile manufacturing business,” *Expert Systems with Applications*, vol. 38, no. 6, pp. 7186–7191, 2011.

[74] D. Zakrzewska and J. Murlewski, “Clustering algorithms for bank customer segmentation,” *5th International Conference on Intelligent Systems Design and Applications (ISDA05)*, 2005.

[75] Y.-Y. Shih and C.-Y. Liu, “A method for customer lifetime value ranking — Combining the analytic hierarchy process and clustering analysis,” *Journal of Database Marketing & Customer Strategy Management*, vol. 11, no. 2, pp. 159–172, 2003.

[76] N. J. Momtaz, S. Alizadeh, and M. S. Vaghefi, “A new model for assessment fast food customer behavior case study,” *British Food Journal*, vol. 115, no. 4, pp. 601–613, Dec. 2013.

[77] P. Dhandayudam, “An Improved Clustering Algorithm for Customer Segmentation,” *International Journal of Engineering Science and Technology*, vol. 4, no. 2, pp. 695-702, 2012

[78] N.-C. Hsieh, “An integrated data mining and behavioral scoring model for analyzing bank customers,” *Expert Systems with Applications*, vol. 27, no. 4, pp. 623–633, 2004.

[79] J.-T. Wei, S.-Y. Lin, C.-C. Weng, and H.-H. Wu, “A case study of applying LRFM model in market segmentation of a children’s dental clinic,” *Expert Systems with Applications*, vol. 39, no. 5, pp. 5529–5533, 2012.

[80] W.-Y. Chiang, “Applying data mining with a new model on customer relationship management systems: a case of airline industry in Taiwan,” *Transportation Letters*, vol. 6, no. 2, pp. 89–97, 2014.

[81] F. Safari, N. Safari, and G. A. Montazer, “Customer lifetime value determination based on RFM model,” *Marketing Intelligence & Planning*, vol. 34, no. 4, pp. 446–461, Jun. 2016.

[82] S. Peker, A. Kocyigit, and P. E. Eren, “A hybrid approach for predicting customers’ individual purchase behavior,” *Kybernetes*, vol. 46, no. 10, pp. 1614–1631, Jun. 2017.

[83] C.-H. Wang, "Apply robust segmentation to the service industry using kernel induced fuzzy clustering techniques," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8395–8400, 2010.

[84] C.-H. Wang, "Outlier identification and market segmentation using kernel-based clustering techniques," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3744–3750, 2009.

[85] M. Ozer, "User segmentation of online music services using fuzzy clustering," *Omega*, vol. 29, no. 2, pp. 193–206, 2001.

[86] H. Shin and S. Sohn, "Segmentation of stock trading customers according to potential value," *Expert Systems with Applications*, vol. 27, no. 1, pp. 27–33, 2004.

[87] J.-T. Wei, M.-C. Lee, H.-K. Chen, and H.-H. Wu, "Customer relationship management in the hairdressing industry: An application of data mining techniques," *Expert Systems with Applications*, vol. 40, no. 18, pp. 7513–7518, 2013.

[88] C.-Y. Tsai and C.-C. Chiu, "A purchase-based market segmentation methodology," *Expert Systems with Applications*, vol. 27, no. 2, pp. 265–276, 2004.

[89] S.-C. Huang, E.-C. Chang, and H.-H. Wu, "A case study of applying data mining techniques in an outfitter's customer value analysis," *Expert Systems with Applications*, vol. 36, no. 3, pp. 5909–5915, 2009.

[90] C. Hung and C.-F. Tsai, "Market segmentation based on hierarchical self-organizing map for markets of multimedia on demand," *Expert Systems with Applications*, vol. 34, no. 1, pp. 780–787, 2008.

[91] S.-S. Weng and M.-J. Liu, "Feature-based recommendations for one-to-one marketing," *Expert Systems with Applications*, vol. 26, no. 4, pp. 493–508, 2004.

[92] F. Lai, M. Griffin, and B. J. Babin, "How quality, value, image, and satisfaction create loyalty at a Chinese telecom," *Journal of Business Research*, vol. 62, no. 10, pp. 980–986, 2009.

[93] J. Zhao, W. Zhang, and Y. Liu, “Improved K-Means cluster algorithm in telecommunications enterprises customer segmentation,” *2010 IEEE International Conference on Information Theory and Information Security*, 2010.

[94] V. Babaiyan and S.A. Sarfarazi. “Analyzing Customers of South Khorasan Telecommunication Company with Expansion of RFM to LRFM Model,” *Journal of AI and Data Mining*, vol. 7, no. 2, pp. 331-340, 2019.

[95] M. Chen, A. Chiu, and H. Chang, “Mining changes in customer behavior in retail marketing,” *Expert Systems with Applications*, vol. 28, no. 4, pp. 773–781, 2005.

[96] W.-Y. Chiang, “To mine association rules of customer values via a data mining procedure with improved model: An empirical case study,” *Expert Systems with Applications*, vol. 38, no. 3, pp. 1716–1722, 2011.

[97] Z. Kou, “Association Rule Mining Using Chaotic Gravitational Search Algorithm for Discovering Relations between Manufacturing System Capabilities and Product Features.” *Concurrent Engineering*, vol. 27, no. 3, pp. 213–232, Sept. 2019.

[98] H. Yun, D. Ha, B. Hwang, and K. H. Ryu, “Mining association rules on significant rare data using relative support,” *Journal of Systems and Software*, vol. 67, no. 3, pp. 181–191, 2003.

[99] T. Arjannikov, and J. Z. Zhang, “An Association-based Approach to Genre Classification in Music,” *15th International Society for Music Information Retrieval Conference ISMIR 2014*, pp. 95-100, 2014.

[100] WebFOCUS Release, “Explanation of the Market Basket Model,” https://infocenter.informationbuilders.com/wf80/index.jsp?topic=%2Fpubdocs%2FRStat16%2Fsource%2Ftopic49.htm, [17 May 2020]

[101] W.-Y. Chiang, “Applying data mining for online CRM marketing strategy,” *British Food Journal*, vol. 120, no. 3, pp. 665–675, May 2018.

[102] D.-R. Liu and Y.-Y. Shih, “Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences,” *Journal of Systems and Software*, vol. 77, no. 2, pp. 181–191, 2005.

[103] R. Srikant and R. Agrawal, “Mining generalized association rules,” *Future Generation Computer Systems*, vol. 13, no. 2-3, pp. 161–180, 1997.

[104] S. Park, S.-U. Kang, and H. Zo, “Analysis of influencing factors on the IPTV subscription,” *Information Technology & People*, vol. 29, no. 2, pp. 419–443, Jun. 2016.

[105] S. Zhang, X. Wei, R. Huang and Z. Li, “Data-driven objective evaluation on IPTV user experience,” *2016 8th International Conference on Wireless Communications & Signal Processing (WCSP)*, Yangzhou, 2016, pp. 1-5

[106] Y. Zhang, W. Chen, H. Zha and X. Gu, “A Time-Topic Coupled LDA Model for IPTV User Behaviors,” in *IEEE Transactions on Broadcasting*, vol. 61, no. 1, pp. 56-65, March 2015.

[107] I. Kalbandi, M. V. Pawar, B. Nikhilkumar, and R. Bachate, “IPTV Software Process and Workflow,” *Procedia Computer Science*, vol. 50, pp. 128–134, 2015.

[108] M. Kren, U. Sedlar, J. Bešter and A. Kos, “Determination of user opinion based on IPTV data,” in *2016 18th International Conference on Transparent Optical Networks (ICTON)*, Trento, 2016, pp. 1-5.

[109] Murthy, P. L. Srinivasa and T. V. Gopal. “Modelling and Analysis of IPTV Usage Patterns for Improving Quality of Service,” *International Journal of Computer Applications*, vol. 144, pp. 22-31, 2016

[110] C. Yu, H. Ding, H. Cao, Y. Liu, and C. Yang, “Follow Me: Personalized IPTV Channel Switching Guide,” in *Proceedings of the 8th ACM on Multimedia Systems Conference (MMSys’17). Association for Computing Machinery*, New York, NY, USA, 2017, pp. 147–157.

[111] C. Qiuru, L. Ye, X. Haixu, L. Yijun and Z. Guangping, “Telecom customer segmentation based on cluster analysis,” in *2012 International Conference on Computer Science and Information Processing (CSIP)*, Xi'an, Shaanxi, 2012, pp. 1179-1182.

[112] J. Bayer, “Customer segmentation in the telecommunications industry,” *Journal of Database Marketing & Customer Strategy Management*, vol. 17, no. 3-4, pp. 247–256, 2010.

[113] J. Kim, S. Lee, and Y. Park, “A visual context-based market analysis of mobile application services,” *Management Decision*, vol. 54, no. 9, pp. 2106–2132, 2016.

[114] M. Song, X. Zhao, H. E, and Z. Ou, “Statistics-based CRM approach via time series segmenting RFM on large scale data,” *Knowledge-Based Systems*, vol. 132, pp. 21–29, 2017.

[115] K. Srivastava and N. K. Sharma, “Service Quality, Corporate Brand Image, and Switching Behavior: The Mediating Role of Customer Satisfaction and Repurchase Intention,” *Services Marketing Quarterly*, vol. 34, no. 4, pp. 274–291, 2013.

[116] M. Zhang, Y. Wang, and J. Chai, “Review of User Behavior Analysis Based on Big Data: Method and Application,” in *2015 International Conference on Advances in Mechanical Engineering and Industrial Informatics*, Zhengzhou, Henan 2015, pp. 99-103

[117] D. Lee, I. Son, M. Yoo and J. Lee, "Understanding the Adoption of Convergent Services: The Case of IPTV," *2011 44th Hawaii International Conference on System Sciences*, Kauai, HI, 2011, pp. 1-10.

[118] Y.-F. Chen, Y. Huang, R. Jana, H. Jiang, M. Rabinovich, J. Rahe, B. Wei, and Z. Xiao, “Towards capacity and profit optimization of video-on-demand services in a peer-assisted IPTV platform,” *Multimedia Systems*, vol. 15, no. 1, pp. 19–32, 2008.

[119] T.-C. Lin, S. Wu, J. S.-C. Hsu, and Y.-C. Chou, “The integration of value-based adoption and expectation–confirmation models: An example of IPTV continuance intention,” *Decision Support Systems*, vol. 54, no. 1, pp. 63–75, 2012.

[120] V. Kumar, “A Customer Lifetime Value-Based Approach to Marketing in the Multichannel, Multimedia Retailing Environment,” *Journal of Interactive Marketing*, vol. 24, no. 2, pp. 71–85, 2010.

[121] A. D. Keyser, J. Schepers, and U. Konuş, “Multichannel customer segmentation: Does the after-sales channel matter? A replication and extension,” *International Journal of Research in Marketing*, vol. 32, no. 4, pp. 453–456, 2015.

[122] J. Song, T. Jang, and S. Y. Sohn, “Conjoint analysis for IPTV service,” *Expert Systems with Applications*, vol. 36, no. 4, pp. 7860–7864, 2009.

[123] A. Kuusik and U. Varblane, “How to avoid customers leaving: the case of the Estonian telecommunication industry,” *Baltic Journal of Management*, vol. 4, no. 1, pp. 66–79, Sep. 2009.

[124] Q. Mu and K. Lee, “Knowledge diffusion, market segmentation and technological catch-up: The case of the telecommunication industry in China,” *Research Policy*, vol. 34, no. 6, pp. 759–783, 2005.

[125] L. Ye, C. Qiu-ru, X. Hai-xu, L. Yi-jun and Y. Zhi-min, “Telecom customer segmentation with K-means clustering,” *2012 7th International Conference on Computer Science & Education (ICCSE)*, Melbourne, VIC, 2012, pp. 648-651.

[126] C. Zhao, Y. Wu and H. Gao, “Study on Knowledge Acquisition of the Telecom Customers' Consuming Behaviour Based on Data Mining,” *2008 4th International Conference on Wireless Communications, Networking and Mobile Computing*, Dalian, 2008, pp. 1-5.

[127] M. J. Tarokh and M. Esmaeiligookeh, “Modeling patients value using a stochastic approach: An empirical study in the medical industry,” *Computer Methods and Programs in Biomedicine*, vol. 176, pp. 51–59, 2019.

[128] D. Ren, D. Zheng, G. Huang, S. Zhang and Z. Wei, “Parallel Set Determination and K-Means Clustering for Data Mining on Telecommunication Networks,” *2013 IEEE 10th International Conference on High Performance Computing and Communications & 2013 IEEE International Conference on Embedded and Ubiquitous Computing*, Zhangjiajie, 2013, pp. 1553-1557.

[129] A. K. Jain, “Data clustering: 50 years beyond K-means,” *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.

[130] H. I. Arumawadu, R. M. K. T. Rathnayaka, and S. K. Illangarathne, “Mining Profitability of Telecommunication Customers Using K-Means Clustering,” *Journal of Data Analysis and Information Processing*, vol. 03, no. 03, pp. 63–71, 2015.

[131] M. Khajvand, K. Zolfaghar, S. Ashoori, and S. Alizadeh, “Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study,” *Procedia Computer Science*, vol. 3, pp. 57–63, 2011.

[132] G. Jayalakshmi and K.Nageswara Rao, “Mining Association Rules for Large Transactions Using New Support and Confidence Measures,” *Journal of Theoretical and Applied Information Technology*, pp. 94-100, 2005.

[133] G. W. Milligan and M. C. Cooper, “A study of standardization of variables in cluster analysis,” *Journal of Classification*, vol. 5, no. 2, pp. 181–204, 1988.

[134] D.-C. Li, W.-L. Dai, and W.-T. Tseng, “A two-stage clustering method to analyze customer characteristics to build discriminative customer management: A case of textile manufacturing business,” *Expert Systems with Applications*, vol. 38, no. 6, pp. 7186–7191, 2011.

[135] S. H. Ha and S. C. Park, “Application of data mining tools to hotel data mart on the Intranet for database marketing,” *Expert Systems with Applications*, vol. 15, no. 1, pp. 1–31, 1998.

[136] S. Peker, A. Kocyigit and P. E. Eren, “A methodology for product segmentation using sale transactions,” *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Opatija, 2018, pp. 1249-1253.

[137] A.P. Reynolds, G. Richards, V.J. Rayward-Smith, “The Application of K-Medoids and PAM to the Clustering of Rules,” in *Yang Z.R., Yin H., Everson R.M. (eds) Intelligent Data Engineering and Automated Learning – IDEAL 2004. Lecture Notes in Computer Science*, vol 3177, Berlin, Heidelberg: Springer, 2004.

[138] A. S. Dunk, “Product life cycle cost analysis: the impact of customer profiling, competitive advantage, and quality of IS information,” *Management Accounting Research*, vol. 15, no. 4, pp. 401–414, 2004.